

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO & ORGANIZAÇÃO DO
CONHECIMENTO

HEBER TORMENTINO DE SOUSA

**ANÁLISE DE ASSUNTOS ASSISTIDA POR COMPUTADOR:
Proposta Metodológica de um Modelo de Inteligência Aumentada**

Belo Horizonte

2022

HEBER TORMENTINO DE SOUSA

**ANÁLISE DE ASSUNTOS ASSISTIDA POR COMPUTADOR:
Proposta Metodológica de um Modelo de Inteligência Aumentada**

Tese apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Doutor, área de concentração Ciência da Informação.

Linha de Pesquisa: Arquitetura e Organização do Conhecimento.

Orientador: Profa. Dra. Gercina Ângela de Lima.

BELO HORIZONTE

2022

S725a Sousa, Heber Tormentino de.

Análise de assuntos assistida por computador [recurso eletrônico] : proposta metodológica de um modelo de inteligência aumentada / Heber Tormentino de Sousa. - 2022.

1 recurso online (243 f. : il., color.) : pdf.

Orientadora: Gercina Ângela de Lima.

Tese (Doutorado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 152-164.

Apêndices: f. 165-243.

Exigência do sistema: Adobe Acrobat Reader.

1. Ciência da informação – Teses. 2. Cabeçalhos de assunto - Teses. 3. Organização da informação - Teses. 4. Inteligência artificial - Processamento de dados – Teses. 5. Indexação - Teses. I. Lima, Gercina Ângela Borém de Oliveira. II. Universidade Federal de Minas Gerais. Escola de Ciência da Informação. III. Título.

CDU: 025.4:004

Ficha catalográfica: Rosimeire Silva Campos de Lima - CRB: 6/3145

Biblioteca Profª Etelvina Lima, Escola de Ciência da Informação da UFMG



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPGOC

FOLHA DE APROVAÇÃO

ANÁLISE DE ASSUNTOS ASSISTIDA POR COMPUTADOR: Proposta Metodológica de um Modelo de Inteligência Aumentada

HEBER TORMENTINO DE SOUSA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Arquitetura e Organização do Conhecimento.

Aprovada em 18 de novembro de 2022, por videoconferência, pela banca constituída pelos membros:

Prof(a). Gercina Ângela de Lima (Orientadora)
ECI/UFMG

Prof(a). Benildes Coura Moreira dos Santos Maculan
ECI/UFMG

Prof(a). Cíntia de Azevedo Lourenço
ECI/UFMG

Prof(a). Mariângela Spotti Lopes Fujita
UNESP

Prof(a). Edson Romano Nucci
UFSJ

Prof(a). Patricia Nascimento Silva
ECI/UFMG

Prof(a). Webert Junio Araujo
CEFET-MG

Belo Horizonte, 18 de novembro de 2022.



Documento assinado eletronicamente por **Gercina Angela de Lima, Professora do Magistério Superior**, em 25/11/2022, às 17:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

Documento assinado eletronicamente por **Cintia de Azevedo Lourenco, Professora do Magistério Superior**, em 29/11/2022, às 17:15, conforme horário oficial de Brasília, com fundamento no art. 5º



do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mariângela Spotti Lopes Fujita, Usuário Externo**, em 29/11/2022, às 18:15, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 29/11/2022, às 19:43, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Edson Romano Nucci, Usuário Externo**, em 29/11/2022, às 21:37, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Patricia Nascimento Silva, Professora do Magistério Superior**, em 30/11/2022, às 07:54, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Webert Júnio Araújo, Usuário Externo**, em 30/11/2022, às 09:14, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1904182** e o código CRC **066DA38B**.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPGOC

ATA DA DEFESA DE TESE DO ALUNO

HEBER TORMENTINO DE SOUSA

Realizou-se, no dia 18 de novembro de 2022, às 14:00 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada *ANÁLISE DE ASSUNTOS ASSISTIDA POR COMPUTADOR: Proposta Metodológica de um Modelo de Inteligência Aumentada*, apresentada por HEBER TORMENTINO DE SOUSA, número de registro 2019664075, graduado no curso de CIÊNCIA DA COMPUTAÇÃO, como requisito parcial para a obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Gercina Ângela de Lima - ECI/UFMG (Orientadora), Prof(a). Benildes Coura Moreira dos Santos Maculan - ECI/UFMG, Prof(a). Cíntia de Azevedo Lourenço - ECI/UFMG, Prof(a). Mariângela Spotti Lopes Fujita - UNESP, Prof(a). Edson Romano Nucci - UFSJ, Prof(a). Patricia Nascimento Silva - ECI/UFMG, Prof(a). Webert Junio Araujo - CEFET-MG.

A Comissão considerou a tese:

Aprovada

Reprovada

APROVADA COM LOUVOR, COM INDICAÇÃO DOS MEMBROS DA BANCA PARA PARTICIPAR DA PREMIAÇÃO DE MELHOR TESE.

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 18 de novembro de 2022.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Gercina Angela de Lima, Professora do Magistério Superior**, em 25/11/2022, às 17:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cíntia de Azevedo Lourenço, Professora do Magistério Superior**, em 29/11/2022, às 17:14, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mariângela Spotti Lopes Fujita, Usuário Externo**, em 29/11/2022, às 18:15, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 29/11/2022, às 19:43, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Edson Romano Nucci, Usuário Externo**, em 29/11/2022, às 21:38, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Patricia Nascimento Silva, Professora do Magistério Superior**, em 30/11/2022, às 07:53, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Webert Júnio Araújo, Usuário Externo**, em 30/11/2022, às 09:13, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1904150** e o código CRC **739B473A**.

Referência: Processo nº 23072.267804/2022-62

SEI nº 1904150

DEDICATÓRIA

Dedico este trabalho à minha tia Almerinda (*in memorian*), porque não há escuridão para aqueles que são luz; e à minha tia Balbina (*in memorian*), que me deu o luxo de ter uma mãe a mais.

AGRADECIMENTOS

Este trabalho foi feito entre 2019 e 2023, o que incluiu os dois anos de isolamento social em razão da pandemia de Covid-19. Foram tempos difíceis, e como não há nada de novo no rugir da tempestade, soubemos atravessá-los, e do lado de cá, onde o sol da esperança continua a brilhar, eu agradeço àqueles que, a seu jeito, participaram desta longa jornada.

Agradeço especialmente:

À minha mãe, Cilda Rosa, e a toda a minha família (pai, irmãos, sobrinhos e sobrinhas), especialmente à minha sobrinha Angélica e às minhas irmãs Cheila, Valdete e Margarete.

Aos meus colegas, Professor Doutor Gustavo Fernandes Rodrigues e Professor Doutor Moacir de Souza Júnior, porque assumiram compromissos além de suas atribuições para que eu pudesse sair para fazer este doutorado. Gratidão!

À minha orientadora, Professora Doutora Gercina Lima, por toda sua dedicação, paciência e inteligência.

À Professora Doutora Mariângela Fujita pela gentileza e pelas contribuições valiosas.

À Professora Doutora Benildes Maculan, à Professora Doutora Cíntia de Azevedo Lourenço, à Professora Doutora Patrícia Nascimento Silva, e à Professora Doutora Célia Dias por suas contribuições valiosas.

Ao Professor Doutor Edson Nucci por sua paciência e estímulo.

Ao Doutor Webert Araújo por sua contribuição.

Ao meu colega Gustavo Portela Lages.

À Doutora Bruna Rafaela Penido Conrado e à Rochelle Martins Alvorcem.

Aos membros do MHTX, especialmente à Celsiane Araújo.

Aos meus amigos Professor Doutor Jorge Bellido, ao Professor Doutor Alexandre Moreira e ao Engenheiro de Telecomunicações Fernando Resende.

Ao Álvaro, à Gildenara da Costa Gomes, ao Vinícius e a todas as pessoas da ECI.

À Renilda Figueiredo, pela revisão gramatical.

Aos meus colegas do DETEM.

Aos meus colegas da UFSJ.

Às minhas alunas e alunos da graduação.

Ao João Dória.

Ao povo brasileiro.

*As ameaças
e as guerras
haveremos de atravessá-las,
rompê-las ao meio,
cortando-as
como uma quilha corta
as ondas.*

Vladimir Maiakóvski

RESUMO

Tema: A inteligência aumentada é a expansão das faculdades cognitivas humana decorrente do uso de ferramentas inteligentes. A origem do conceito é atribuída a Douglas Engelbart que, em 1962, propôs uma estrutura metodológica de como aumentar o intelecto humano para solucionar problemas complexos com a assistência das ferramentas. **Objeto:** O objeto pesquisado foi a análise de assuntos Assistida por Computador. **Escopo:** No contexto da Organização da Informação, a aplicação da inteligência aumentada procura resolver o dilema entre adotar a rapidez da automatização ou manter a qualidade da indexação de assuntos intelectual ao processar o elevado volume disponível de dados textuais. **Objetivo:** Elaborar uma proposta metodológica de um modelo de inteligência aumentada para viabilizar a análise de assuntos assistida por computador. **Ambientação:** O acervo das bases de dados da pesquisa agropecuária (BDPA), tendo como recorte um conjunto de 7377 artigos científicos na área da pesquisa em agropecuária indexados pelo Tesouro Agrícola Nacional (Thesagro). **Tipo de Pesquisa:** Esta pesquisa é de natureza aplicada; quanto aos objetivos, é uma pesquisa exploratória; e quanto aos procedimentos metodológicos é uma pesquisa experimental. **Coleta de Dados:** Os dados coletados foram artigos científicos e os seus descritores atribuídos por pessoas da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e publicados entre 1998 e 2020. Todos os dados estão disponíveis para acesso público e gratuito. **Métodos:** Utilizou-se métodos de inteligência artificial do processamento de linguagem natural para realizar a sumarização automática, a identificação dos assuntos do texto e a modelagem de tópicos para produzir a análise de assuntos assistida por computador. **Fundamento Teórico:** A inteligência artificial pode produzir bons resultados técnicos baseada em dados e a inteligência humana pode utilizar esses resultados como subsídios para a tomada de decisões no processo de indexação de assuntos de artigos científicos. **Fundamento Histórico/Contextual:** A literatura científica sobre inteligência aumentada foi coletada e selecionada das bases científicas de dados *Library and Information Science Abstracts* (LISA); *Library, Information Science and Technology Abstracts* (LISTA); *Institute of Electrical and Electronics Engineers* (IEEE); *Scopus*; *Web of Science*. **Ponto de Vista do Autor:** O ponto de vista sobre a inteligência aumentada para a análise de assuntos assistida por computador é do próprio autor da pesquisa, pois trata-se de uma abordagem inédita, uma inovação na área. No entanto, o conceito de inteligência aumentada é de Bush (1945) e de Engelbart (1962). **Causa e Efeito:** O uso da inteligência artificial para extrair *insight* sobre os dados de indexação de assuntos de uma base de dados formulada por especialista permitiu produzir uma análise de assuntos com resultados satisfatórios de mais 79%, que associados a inteligência aumentada pode chegar a 100%. **Resultados:** Foram elaborados um protótipo tecnológico e uma Base de Dados apresentados sob uma proposta teórico-metodológica unificada com a finalidade de amparar tecnologicamente a tomada de decisões na etapa de tradução do processo de indexação de assuntos. Assim, a inteligência aumentada, como padrão de projeto, forneceu os construtos teóricos para a implementação do protótipo tecnológico denominado de análise de assuntos assistida por computador (3ApC), que foi escrito em código-fonte aberto na linguagem de programação *Python*. O experimento de validação do modelo desenvolvido utilizou 7073 artigos científicos e seus respectivos termos de indexação de assuntos das bases de dados de pesquisa agropecuária (BDPA). Concluiu-se que o uso do 3ApC assistiu adequadamente a análise de assuntos e pôde dar celeridade ao processo de indexação de assuntos com a mesma qualidade do processo intelectual. Além disso, ampliou o horizonte de conhecimento na Ciência da Informação e acrescentou inovação tecnológica ao conjunto de ferramentas disponíveis para a Organização da Informação.

Palavras-chave: Análise de Assuntos. Indexação de Assuntos. Organização do Conhecimento. Inteligência Artificial. Inteligência Aumentada.

ABSTRACT

Theme: *Augmented Intelligence is the expansion of human cognitive faculties resulting from the use of intelligent tools. The origin of the concept is attributed to Douglas Engelbart who, in 1962, proposed a methodological framework on how to increase human intellect to solve complex problems with the assistance of tools.* **Object:** *The researched object was Computer Aided Subject Analysis.* **Scope:** *In the context of Information Organization, the application of Augmented Intelligence seeks to solve the dilemma between adopting the speed of automation or maintaining the quality of intellectual indexing when processing the high volume of textual data available.* **Objective:** *To develop a methodological proposal for an augmented intelligence model to enable computer-assisted subject analysis.* **Setting:** *The collection of the Agricultural Research Database (BDPA), having as a cut a set of 7377 scientific articles in agricultural research indexed by the Thesaurus Agricola Nacional (Thesagro).* **Research Type:** *This research is applied in nature; as for the objectives, it is exploratory research; and as for the methodological procedures, it is experimental research.* **Data Collection:** *The data collected were scientific articles and their descriptors attributed by people from the Brazilian Agricultural Research Corporation (Embrapa) and published between 1998 and 2021. All data are available for public access and free of charge.* **Methods:** *Artificial Intelligence methods of Natural Language Processing were used to perform automatic summarization, identification of text subjects and topic modeling to produce Computer Aided Subject Analysis.* **Theoretical Background:** *Artificial Intelligence can produce good technical results based on data and Human Intelligence can use these results as subsidies for decision making in the process of indexing scientific articles subjects.* **Historical/Contextual Background:** *Scientific literature on Augmented Intelligence was collected and selected from the scientific databases Library and Information Science Abstracts (LISA); Library, Information Science and Technology Abstracts (LISTA); Institute of Electrical and Electronics Engineers (IEEE); Scopus; Web of Science.* **Author's Point of View:** *The point of view on Augmented Intelligence for Computer-Aided Analysis of Subjects belongs to the author of the research, as it is an unprecedented approach, an innovation in the area. However, the concept of Augmented Intelligence is from Bush (1945) and Engelbart (1962).* **Cause and Effect:** *The use of Artificial Intelligence to extract insight into the indexing data from a database formulated by an expert allowed producing a subject analysis with satisfactory results of over 79%, which associated with Augmented Intelligence can reach 100%.* **Results:** *A technological prototype and a Database were developed, presented under a unified theoretical-methodological proposal to technologically support decision-making in the Translation stage of the indexing process. Thus, Augmented Intelligence, as a design pattern, provided the theoretical constructs for the implementation of the technological prototype called Computer Aided Subject Analysis (3ApC), which was written in open-source code in the Python programming language. The validation experiment of the developed model used 7073 scientific articles and their respective indexing terms from the Agricultural Research Database (BDPA). It was concluded that the use of 3ApC adequately assisted the Analysis of Subjects and could speed up the indexing process with the same quality as the intellectual process. In addition, it broadened the horizon of knowledge in Information Science and added technological innovation to the set of tools available for the Organization of Information.*

Keywords: Subject Analysis. Subject Indexing. Knowledge Organization. Artificial Intelligence. Augmented Intelligence.

LISTA DE FIGURAS

Figura 1 – Visão geral da inteligência aumentada na indexação.....	17
Figura 2 – Esquema hierárquico da análise de assuntos.....	35
Figura 3 – A relação da inteligência artificial e o aprendizado profundo.....	46
Figura 4 – Uma <i>pipeline</i> típica do processamento de linguagem natural.....	50
Figura 5 – Visão geral e esquemática de um modelo conceitual.....	60
Figura 6 – Esquema genérico de modelo conceitual.....	61
Figura 7 – As etapas dos procedimentos metodológicos.....	73
Figura 8 – O volume de artigos científicos coletados por ano de publicação.....	77
Figura 9 – As diferenças entre as duas versões online do Thesagro.....	78
Figura 10 – O ciclo completo de desenvolvimento do 3ApC inspirado no RUP.....	81
Figura 11 – A <i>pipeline</i> para a sumarização Automática.....	85
Figura 12 – A <i>pipeline</i> para a formulação das expressões regulares.....	86
Figura 13 – A pipeline para a identificação dos assuntos do texto.....	86
Figura 14 – A pipeline da seleção dos assuntos relevantes do texto.....	87
Figura 15 – Mapa conceitual para distinguir “termo”, “assunto” e “descriptor”.....	92
Figura 16 – A visão geral e conceitual do 3ApC.....	103
Figura 17 – As funcionalidades apresentadas em um diagrama de casos de uso.....	108
Figura 18 – Fragmento de código-fonte contendo explicações computacionais.....	111
Figura 19 – A interface gráfica de usuário da janela principal do 3ApC.....	112
Figura 20 – O início do percurso do usuário para realizar uma análise de assuntos.....	117
Figura 21 – Caixa de diálogo para o usuário inserir ou digitar o título do Artigo.....	118
Figura 22 – A tela contendo o relatório da análise de assuntos.....	119
Figura 23 – Os assuntos relevantes do texto selecionados pelo 3ApC.....	119
Figura 24 – Os assuntos identificados no texto pelo 3ApC.....	121
Figura 25 – A funcionalidade “Texto” permite editar o texto do artigo.....	123
Figura 26 – A funcionalidade de navegação na estrutura do Thesagro.....	124
Figura 27 – As funcionalidades para conhecer a base de conhecimento.....	125
Figura 28 – Uma nuvem de palavras para exibir informações visuais do texto.....	126
Figura 29 – A classificação dos resultados da sumarização automática.....	133
Figura 30 – A intersecção entre dois conjuntos (diagrama de <i>Venn</i>).....	137
Figura 31 – A classificação da Intersecção entre os conjuntos analisados.....	142

LISTA DE TABELAS

Tabela 1 -- Resultados da busca nas bases de dados sobre a inteligência aumentada.....	26
Tabela 2 – Estatísticas relativas à seleção dos dados.....	94
Tabela 3 – A distribuição dos assuntos, termos e descritores.....	96
Tabela 4 – A distribuição porcentual do número de descritores por artigo científico.....	96
Tabela 5 – Amostra de termos coocorrentes nos descritores da BDPA.....	98
Tabela 6 – Subconjuntos de termos do Thesagro.....	100
Tabela 7 – Informações gerais das expressões regulares para o Thesagro.....	135

LISTA DE QUADROS

Quadro 1 - Paradigmas principais do aprendizado de máquinas	46
Quadro 2 – O mapeamento múltiplo como característica de expressões regulares.	56
Quadro 3 – Exemplo de um tópico contendo o termo “germoplasma”.....	58
Quadro 4 – A tomada de decisão na análise de assuntos.	67
Quadro 5 – Comparação entre os modelos utilizados como fonte.	74
Quadro 6 – A relação dos procedimentos metodológicos com os capítulos da tese.	75
Quadro 7 – As URLs de acesso a diferentes versões <i>on-lines</i> do Thesagro.....	78
Quadro 8 – O ciclo completo de desenvolvimento do 3ApC inspirado no RUP.....	80
Quadro 9 – A correlação entre a análise de assuntos e as metodologias tecnológicas.....	81
Quadro 10 – Questões norteadoras para a avaliação manual de resumos.....	90
Quadro 11 – A expressão utilizada no sistema de busca avançado da BDPA.	93
Quadro 12 – Regras de formação do plural necessárias para o Thesagro.....	99
Quadro 13 – A correspondência entre os estágios da análise de assuntos.	104
Quadro 14 – A adequação de Modelos de inteligência artificial.	113
Quadro 15 – O teste da funcionalidade “sumarização automática” do 3ApC.	120
Quadro 16 – Previsão do 3ApC comparada à indexação da BDPA.	122
Quadro 17 – Correspondência de estágios da análise de assuntos e as funcionalidades do 3ApC.	127
Quadro 18 – Um resumo adequado obtido por sumarização automática.....	128
Quadro 19 – Um resumo ilegível obtido por sumarização automática.....	129
Quadro 20 – Exemplo de um resumo com problemas na sumarização automática.	130
Quadro 21 – Análise descritiva dos resumos que precisam de intervenção humana.	131
Quadro 22 – Uma amostra das expressões regulares formuladas para o Thesagro.....	135
Quadro 23 – Classificação da intersecção entre os conjuntos.	137
Quadro 24 – Uma amostra do APÊNDICE E.	138
Quadro 25 – Caso 1: sem intersecção entre os conjuntos de termos indexadores.	138
Quadro 26 – Percurso no Thesagro entre termos do 3ApC e descritores da BDPA.....	139
Quadro 27 – Caso 2: intersecção mediana entre os conjuntos de termos.....	139
Quadro 28 – Percurso semântico entre termos do 3ApC e descritores da BDPA.	140
Quadro 29 – Caso 3: intersecção Excelente entre os conjuntos de termos.....	141
Quadro 30 – Percursos semânticos entre descritores e assuntos relevantes.	141
Quadro 31 – Comparação entre os modelos utilizados como fonte.	165

LISTA DE ABREVIATURAS

3ApC	Análise de Assuntos Assistida por Computador
AAAI	<i>Association for the Advancement of Artificial Intelligence</i>
BCI	Biblioteconomia e Ciência da Informação
BDPA	Bases de Dados da Pesquisa Agropecuária
BINAGRI	Biblioteca Nacional de Agricultura
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CDD	Classificação Decimal de <i>Dewey</i>
IID	Distribuição Independente e Idêntica
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FAIR	<i>Findable, Accessible, Interoperable e Reusable</i>
GUI	Interface Gráfica do Usuário
IA	Inteligência Artificial
IDE	Ambiente de Desenvolvimento Integrado
ITIL	<i>Information Technology Infrastructure Library</i>
LDA	<i>Latent Dirichlet Allocation</i>
LSI	<i>Latent Semantic Indexing</i>
ML	<i>Machine Learning</i>
MHTX	Modelagem Organização e Representação da Informação Hipertextual
OC	Organização do Conhecimento
OCR	Reconhecimento Óptico de Caracteres
OSI	<i>Open Source Initiative</i>
SNIDA	Sistema Nacional de Informação e Documentação Agrícola
SOC	Sistema de Organização do Conhecimento
PLN	Processamento de Linguagem Natural
pLSI	<i>Probabilistic latent semantic analysis</i>
RNN	<i>Recurrent Neural Network</i>
TF	<i>Term Frequency</i>
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
THES	<i>Thesaurus Agrícola Brasileiro (Normalizador de Termos Agrícola)</i>
THESAGRO	<i>Thesaurus Agrícola Nacional</i>
UNICODE	Códigos (de caracteres) Unificados
URL	<i>Uniform Resource Locator</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Problema	15
1.2	Justificativas	18
1.3	Pressupostos.....	19
1.4	Objetivos	22
1.4.1	Objetivo geral	22
1.4.2	Objetivos específicos.....	22
1.5	Estrutura da tese	22
2	REVISÃO DE LITERATURA SOBRE A INTELIGÊNCIA AUMENTADA.....	24
2.1	Questões preliminares.....	24
2.2	Procedimentos adotados na revisão	25
2.3	Descrição dos estudos	28
2.4	Análise qualitativa dos estudos	31
3	FUNDAMENTAÇÃO TEÓRICA E METODOLÓGICA	34
3.1	Análise de assuntos	34
3.1.1	Breve histórico.....	36
3.1.2	Definições.....	37
3.1.3	Características	38
3.1.3.1	Análise de assuntos intelectual.....	39
3.1.3.2	Análise de assuntos assistida por computador.....	40
3.1.4	Aplicabilidade e importância	41
3.2	Inteligência artificial	41
3.2.1	História	42
3.2.2	Definição	42
3.2.3	Características	43
3.2.4	Relação da inteligência artificial com a Ciência da Informação.....	46
3.3	Processamento de linguagem natural	47
3.3.1	História	47
3.3.2	Definições.....	49
3.3.3	Características	49
3.3.4	Aplicações e importância.....	51
3.3.4.1	Sumarização automática	52
3.3.4.2	Expressões regulares.....	55
3.3.4.3	Modelagem de tópicos	57
3.4	Modelo conceitual.....	59
3.4.1	História	60
3.4.2	Definições.....	61
3.4.3	Características	62
3.4.4	Aplicabilidade e importância	63

3.5	Inteligência aumentada	64
3.5.1	Histórico	64
3.5.2	Definições.....	65
3.5.3	Características	66
3.5.4	Aplicabilidade e importância.....	68
3.6	Considerações	69
4	METODOLOGIA	70
4.1	Caracterização da pesquisa	70
4.2	Universo da pesquisa	71
4.3	Procedimentos metodológicos	72
4.3.1	Etapa 1: Planejamento geral	72
4.3.2	Etapa 2: Elaboração da base de conhecimento	75
4.3.2.1	Bases de Dados da Pesquisa Agropecuária	76
4.3.2.2	Thesaurus Agrícola Nacional	77
4.3.3	Etapa 3: Desenvolvimento do software	80
4.3.3.1	Fase de Concepção	81
4.3.3.2	Fase de Elaboração	82
4.3.3.3	Fase de Construção	84
4.3.3.4	Fase de Teste de funcionalidade	88
4.3.4	Etapa 4: Análise dos resultados	89
4.3.4.1	Corpus do experimento de validação	89
4.3.4.2	Experimento de validação	89
5	ELABORAÇÃO DA BASE DE CONHECIMENTOS	92
5.1	Coleta e seleção dos dados	93
5.1.1	Análise, seleção, armazenamento e distribuição dos dados	94
5.2	Análise dos descritores dos artigos científicos selecionados	95
5.2.1	Análise da frequência	95
5.2.2	Análise do número de descritores	96
5.2.3	Os termos coocorrentes nos descritores	97
5.3	Uma análise do <i>Thesaurus</i> Agrícola Nacional.....	98
5.3.1	Problema dos plurais.....	99
5.4	A base de conhecimento elaborada	100
6	DESENVOLVIMENTO DO SOFTWARE.....	102
6.1	Concepção	102
6.1.1	Definição do objetivo e do escopo.....	105
6.1.2	Definição do perfil de usuário	105
6.2	Elaboração	105
6.2.1	Requisitos funcionais.....	105
6.2.2	Diagramas de casos de uso	108
6.2.3	Requisitos de sistema	109

6.3	Elaboração	110
6.3.1	Criação do ambiente	110
6.3.2	Interface Gráfica de Usuário.....	111
6.3.3	Implementação dos métodos de inteligência artificial.....	112
6.3.3.1	Sumarização automática	113
6.3.3.2	Expressões regulares.....	114
6.3.3.3	Modelagem de tópicos.....	115
6.4	Teste de funcionalidade	116
6.4.1	O ambiente.....	116
6.4.2	Interface Gráfica de Usuário.....	117
6.4.3	Métodos de inteligência artificial.....	118
6.4.3.1	Sumarização automática	120
6.4.3.2	Expressões regulares.....	121
6.4.3.3	Modelagem de Tópicos	122
6.4.4	Informações de suporte à tomada de decisões.....	122
6.4.4.1	Edição do texto.....	122
6.4.4.2	Navegação no Thesagro	123
6.4.4.3	Base de conhecimento.....	124
6.4.4.4	Nuvens de palavras.....	125
7	ANÁLISE DOS RESULTADOS	127
7.1	Sumarização automática	127
7.1.1	Visão geral dos resumos	127
7.1.1.1	Resumo adequado	128
7.1.1.2	Resumo ilegível.....	128
7.1.1.3	Resumos com problemas semânticos.....	129
7.1.2	A avaliação geral dos resumos	131
7.1.2.1	Avaliação do conjunto de resumos.....	131
7.1.2.2	Avaliação do conjunto total de resumos.....	132
7.1.3	A inteligência aumentada na sumarização automática.....	133
7.2	Expressões regulares.....	134
7.2.1	Formulação das expressões regulares.....	134
7.2.2	Identificação dos assuntos do texto.....	136
7.3	Modelagem de Tópicos	136
7.3.1	Caso 1: intersecção Ruim	138
7.3.2	Caso 2: intersecção Mediana	139
7.3.3	Caso 3: intersecção Excelente	140
7.3.4	Avaliação geral dos resultados da modelagem de tópicos.....	141
7.4	A proposta metodológica de inteligência aumentada	143
8	CONSIDERAÇÕES FINAIS	145
	REFERÊNCIAS.....	152
	APÊNDICE A – MODELO CONCEITUAL	165

APÊNDICE B – RECURSOS COMPUTACIONAIS	182
APÊNDICE C – RESUMOS POR SUMARIZAÇÃO AUTOMÁTICA.....	187
APÊNDICE D – TAMANHO DOS CLUSTERS.....	222
APÊNDICE E – ANÁLISE DE ASSUNTOS.....	223
APÊNDICE F – TERMOS COCORRENTES NOS ÍNDICES.....	231
APÊNDICE G – MODELO SEQUENCIAL.....	233

1 INTRODUÇÃO

Esta é uma pesquisa em Gestão e Organização do Conhecimento, sendo seu tema central a Organização e a Representação da Informação com a assistência da Tecnologia. Seu interesse empírico está pautado pela indexação de assuntos em artigos científicos no domínio da agropecuária¹ mediada pelo Thesagro (*thesaurus* agrícola nacional), de forma que, por estar no âmbito da Biblioteconomia e da Ciência da Informação (BCI), seus esforços foram no sentido de produzir métodos e tecnologias que facilitem a criação de pontos de acessos aos conhecimentos contidos nos documentos em consonância com as necessidades de informação dos usuários.

A intensa produção científica atual amplia ainda mais o grande acervo acumulado ao longo de décadas nos repositórios virtuais e bibliotecas. Isso traz consigo o problema de “como” dar acesso à informação relevante pontuada pelo usuário no momento da busca, considerando que, como afirma Edgar Morin, a ciência é libertadora, mas traz, ao mesmo tempo, possibilidades de subjugação. Segundo ele, para conceber e compreender esse problema:

há que acabar com a tola alternativa da ciência “boa”, que só traz benefícios, ou da ciência “má”, que só traz prejuízos. Pelo contrário, há que, desde a partida, dispor de pensamento capaz de conceber e de compreender a ambivalência, isto é, a complexidade intrínseca que se encontra no cerne da ciência (MORIN, 2005, p.16).

Nesse cenário de conceber e de compreender a ambivalência intrínseca à ciência, a recuperação da informação se tornou mais complexa, porque, conforme avalia Lima (2004), os métodos intelectuais de análise de assuntos não são mais suficientes para dar conta do volume crescente de informação necessário para atender às demandas dos usuários. Golub (2021, p. 702) assinala que a adoção de meios automatizados como uma solução para o “número cada vez maior de documentos em face de orçamentos institucionais sobrecarregados” ajuda a alcançar o objetivo da BCI de dar acesso ao conhecimento científico relevante, mas, ressalva ele, os algoritmos são incapazes de substituir inteiramente o trabalho intelectual.

Assim, esta pesquisa lidou com o seguinte dilema: adotar a velocidade da automatização da indexação de assuntos por inteligência artificial ou manter a qualidade da indexação de assuntos intelectual. Os polos antagônicos dessa abordagem decorrem do fato de que a análise de assuntos intelectual faz uso do conhecimento pragmático do contexto

¹ Por continuidade a outros estudos do grupo de pesquisa MHTX.

social da tarefa, codificando a cultura local nos termos da indexação de assuntos, enquanto a automatização não inclui a experiência acumulada nas instituições e ameaça os postos de trabalho de profissionais da informação, mesmo sem produzir resultados satisfatórios. No entanto, a automatização não se deteriora com a repetição nem sofre variação de humor e captura o significado do texto a partir de padrões linguísticos com mais consistência e velocidade do que uma pessoa.

A indexação de assuntos cria “uma linguagem entre o documento e o usuário” (VIEIRA, 1988, p. 43) como uma interface de transição entre o dinamismo das necessidades de informação e o significado estático do discurso científico contido no texto. Essas são as duas faces da mesma moeda, em que a máquina tem velocidade em capturar o significado do texto e o ser humano, a habilidade de interpretar as necessidades de informação dos usuários. Logo, o desafio desta pesquisa foi juntar essas duas faces do processo em uma sequência para que colaborem entre si. Essa colaboração entre as ferramentas e a inteligência humana recebe a denominação de inteligência aumentada, descrita pela primeira vez em Engelbart (1962).

Assim, a seguir, neste documento estão descritos o problema, as justificativas, os pressupostos, os objetivos, bem como a estrutura da tese.

1.1 Problema

Os termos “máquina”, “tecnologia”, “algoritmo” e “*software*” foram utilizados como sinônimos ao longo do texto. O termo “análise de assuntos” aponta uma etapa do processo da indexação, enquanto os termos “indexação de assuntos” e “indexação” referem-se ao ato de indexar, o que inclui a análise de assuntos.

Na Biblioteconomia e na Ciência da Informação, a indexação de assuntos é um processo efetuado em duas etapas: (1) análise de assuntos, e (2) tradução. Nesta pesquisa, propõe-se efetuar a primeira etapa com a inteligência artificial (*software*) e deixar que a segunda etapa, a tomada de decisão, seja realizada por pessoas. A essa arquitetura de solução deu-se o nome de inteligência aumentada.

O problema de pesquisa consiste em dar celeridade ao processo de indexação de assuntos sem perder qualidade. Esse problema ocorre em virtude da produção insuficiente de índices de assuntos feita pelo método intelectual (LIMA, 2021b) e da qualidade inadequada da rápida produção de índices feita por método automático (GOLUB, 2021).

O principal desafio da (nossa) Sociedade do Conhecimento é dar acesso adequado à informação, visto que entre as duas pontas do processo (autor e leitor) nem sempre há uma conexão direta (LIMA, 2004). Para fechar essa lacuna, a indexação de

assuntos cria os pontos de acesso à informação. Ou seja, o ato de indexar produz as condições efetivas para pôr em contato usuário e documento; portanto, é esse ato que viabiliza, melhora e aperfeiçoa a recuperação da informação, que é, essencialmente, dar acesso à informação.

A recuperação da informação científica em um mundo totalmente conectado, face ao elevado ritmo de publicação, tornou-se uma tarefa muito difícil, porque demanda mais qualidade e precisão que os sistemas automatizados podem proporcionar. Assim, os artigos científicos produzidos e publicados com o auxílio da máquina precisam de indexadores humanos.

Publicar com auxílio das máquinas e esperar uma indexação sem elas, resulta em um gargalo de acesso à informação; daí a questão de pesquisa: como aumentar a velocidade da indexação sem perder a qualidade? Para responder a essa questão, é necessário desenvolver métodos tecnológicos de assistência² à análise de assuntos para que a indexação continue sendo um ato intelectual humano, conforme requer a ISO 5963, mas que, ao mesmo tempo, não seja necessariamente um processo manual, pois isso seria inadequado para o volume da produção científica disponibilizado atualmente.

Se algoritmos nem sempre distinguem o falso do verdadeiro nem o relevante do meramente repetitivo, e se os processos intelectuais não são mais suficientes para organizar o conhecimento disponível, então a Organização da Informação necessita da inteligência humana para atingir a qualidade necessária e dos algoritmos para atender à demanda (quantidade) crescente por informações científicas relevantes.

Considerando que, na análise de assuntos, os algoritmos podem chegar aos resultados rapidamente e que as pessoas são mais eficazes em selecionar as melhores soluções, é de se supor que são imprescindíveis, nos processos documentários, a inteligência humana e o suporte técnico das máquinas. Assim, o problema, na essência, é aplicar algoritmos no processo de indexação para produzir mais, e pôr o ser humano no centro das decisões para garantir a mesma qualidade dos processos intelectuais.

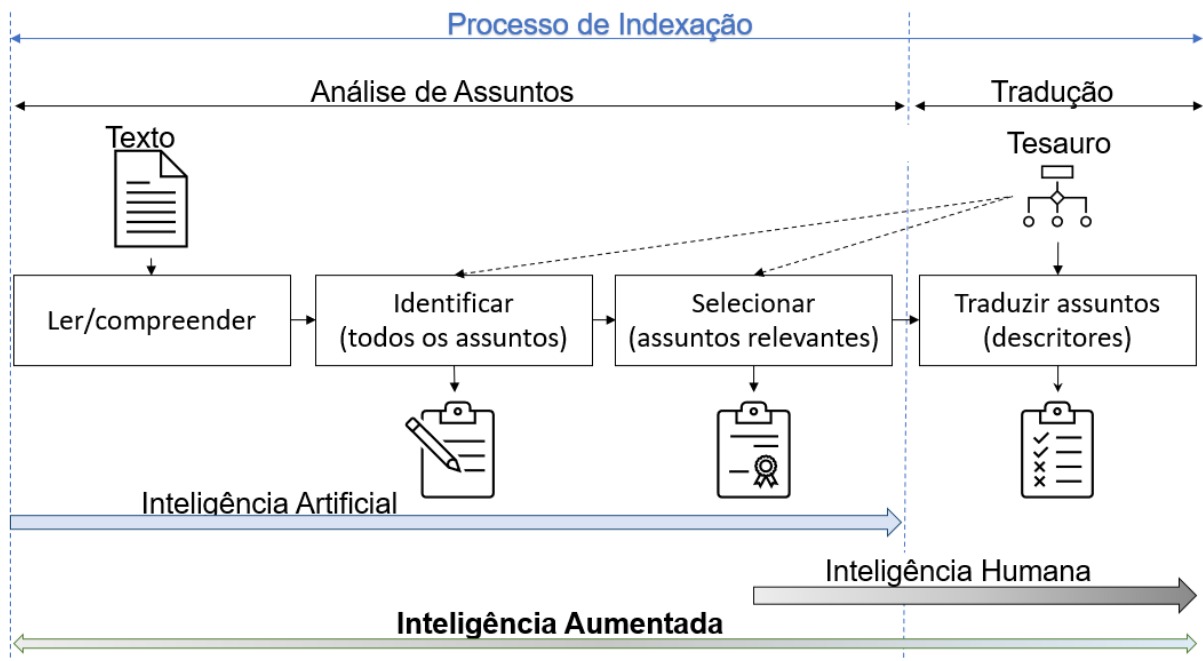
Por hipótese, resolver o problema consiste em programar uma ferramenta tecnológica, amparada em metodologia científica, que faça a análise de assuntos e ofereça um conjunto de respostas elegíveis para o ser humano escolher as melhores. Essa combinação de tecnologia e decisão humana denominou-se de inteligência aumentada. Para viabilizar essa concepção, optou-se por pesquisar a aplicação de métodos de inteligência artificial como técnicas computacionais para indexar os artigos científicos da Base de Dados

² Ver seção 3.1.

de Pesquisa Agropecuária (BDPA), utilizando os artigos e seus respectivos metadados como base de conhecimento e o Thesagro (*Thesaurus* Agrícola Nacional) como o Sistema de Organização do Conhecimento (SOC) mediador.

Nesta pesquisa, o processo de indexação deve combinar o uso simultâneo da tecnologia e da cognição humana, o que corresponde à inteligência aumentada (Figura 1). Justifica-se essa escolha, o fato de que a inteligência aumentada é um padrão de projeto para desenvolver os artefatos de suporte às tarefas, nesse caso específico, documentárias, de forma que a tomada de decisão continua sendo uma incumbência humana.

Figura 1 – Visão geral da inteligência aumentada na indexação.



Fonte: Elaborado pelo autor.

Para chegar à inteligência aumentada na indexação, é preciso responder às seguintes questões:

1. Quais recursos informacionais e métodos computacionais são necessários e suficientes para assistir adequadamente à análise de assuntos?
2. Em que medida a modelagem de tópicos latentes de *Dirichlet* (BLEI *et al.*, 2003) pode contribuir efetivamente para a análise de assuntos?
3. A inteligência artificial aplicada a uma base de conhecimento é suficiente para assistir adequadamente a análise de assuntos?

As justificativas dessas e de outras decisões de pesquisa estão descritas na próxima seção.

1.2 Justificativas

A indexação é o “mecanismo de publicação sistemática de fragmentos de trabalho científico” (SARACEVIC, 1975, p. 323) que pode ter sido o evento-chave do sucesso das ciências modernas (VAKKARI; CHANG; JÄRVELIN, 2022). Nesse sentido, justificam o desenvolvimento desta pesquisa:

1. importância de se estudar a análise de assuntos;
2. necessidade de desenvolver técnicas e tecnologias para auxiliar o trabalho do profissional da informação na análise de assuntos;
3. urgência de criar meios de lidar com os desafios impostos pelo *big data* na indexação da informação científica;
4. ampliação dos horizontes teóricos e tecnológicos da BCI na consecução do seu objetivo de dar acesso ao conhecimento científico relevante.

A análise de assuntos é uma das etapas mais importantes dos processos documentários, porque ela pode determinar a qualidade dos produtos, que, por consequência, vai impactar diretamente a capacidade dos sistemas de recuperação da informação (SRI). Ou seja, quanto melhor a análise de assuntos, mais eficaz e eficiente poderá ser a recuperação da informação. Por isso, para atender à justificativa exposta na Justificativa (1), desta seção, esta pesquisa contribuiu com estudos sobre a inteligência aumentada para a análise de assuntos.

A análise de assuntos implica ler um documento e extrair e traduzir a essência de seu conteúdo em termos de um vocabulário controlado, uma tarefa sujeita à interferência de fatores cognitivos e psicológicos como o conhecimento prévio, a especialização, a formação e o nível de cansaço do profissional da informação (MACULAN; LIMA, 2017). Esses fatores podem tornar o processo documentário lento, o que redundará em aumento dos custos e até ineficiência na recuperação da informação. No entanto, substituir o profissional da informação por ferramentas de automatização não tem resolvido os problemas satisfatoriamente (GOLUB, 2021). Por isso, considerando a Justificativa (2), desta seção, esta pesquisa investigou o uso de algoritmos como uma extensão dos processos cognitivos, uma vez que os algoritmos não estão sujeitos aos fatores citados e que, se usados adequadamente, podem aumentar a velocidade da produção.

O grande volume de dados difíceis de serem analisados, armazenados e validados impõe diversos desafios, e um deles é a capacidade de se indexar a produção científica para viabilizar uma recuperação da informação científica eficaz e eficiente. Assim, a

urgência de criar metodologias de uso e ferramentas tecnológicas (Justificativa 3) que possam lidar satisfatoriamente com as questões decorrentes desse desafio justifica esta pesquisa.

Na Justificativa (4), desta seção, aponta-se a obrigatoriedade de se ampliar os horizontes teóricos e tecnológicos da BCI que leve à consecução do objetivo de dar acesso ao conhecimento relevante, ou seja, favorecer o uso social da informação como preconizam as finalidades declaradas da Organização do Conhecimento (HJØRLAND; ALBRECHTSEN, 1995; ALMEIDA; MARQUES, 2017). Com esse propósito, metodologias e artefatos foram desenvolvidos para extrair percepções dos dados capazes de atender às demandas específicas da comunidade de usuários da informação, sobretudo para acervos escritos em português circunscritos a um domínio do conhecimento.

Enfim, o uso da inteligência aumentada contribui com a extração, por meio das máquinas, de padrões do *big data* que, de outra forma, seriam perdidos. Além disso, reduz o custo, encurta o tempo de processamento, mitiga a subjetividade, a fadiga e a distração inerente aos processos intelectuais e, ao manter a tomada de decisão como um processo cognitivo, garante a qualidade dos produtos documentários.

Além do mais, na Organização do Conhecimento (OC), é preciso considerar a advertência de Birger Hjørland no seu magistral artigo “*Is classification necessary after Google?*”:

Se acreditamos que nós, no campo da Organização do Conhecimento (OC), temos direito a um lugar no mundo acadêmico, bem como na prática do OC, temos que ser capazes de fornecer justificativas tanto acadêmicas quanto práticas para a classificação e para as outras formas de OC (HJØRLAND, 2012, p. 301, tradução nossa).

Essas justificativas levaram em conta os pressupostos, apresentados na próxima seção, que derivam do princípio de aumentar a produtividade sem excluir as pessoas e sem destruir empregos, de promover a inclusão, de incentivar o uso da intuição e da percepção exclusivamente humanas, porque as pessoas são essenciais e insubstituíveis e os *softwares*, ainda que cognitivos, são apenas ferramentas.

1.3 Pressupostos

A análise de assuntos pode produzir respostas diferentes a cada vez que o processo for realizado; mesmo assim, todas essas respostas podem ser consideradas corretas, visto que, na indexação de um texto, não se procura um conjunto de termos que seja a resposta exata; na verdade, nem mesmo existe uma resposta mais correta que outra (LANCASTER, 2004). Isso ocorre em razão da natureza subjetiva dos processos intelectuais humanos e o caráter dinâmico das necessidades de informação das pessoas.

Pressupõe-se que as técnicas computacionais sejam eficientes e capazes de detectar os padrões existentes nos dados e de os utilizarem para classificar os objetos que ainda não estão classificados; e que a inteligência humana pode supervisionar e interferir todo o processo e tomar a decisão final sobre o que deve constar na lista de termos do índice de assunto e, assim, evitar que as respostas de baixa confiabilidade sejam incluídas na base de conhecimento. Assim, para orientar este processo, nesta pesquisa, se admitiu como pressupostos:

1. Uma lista de termos de indexação é uma aproximação e as necessidades de informação dos usuários declaradas nos termos de busca também é uma aproximação (BELKIN, 1980). Nas palavras de Lancaster (2004, p. 285), “aproximações de necessidades de informação” e “aproximações de indexação”.
2. O significado do texto do documento contém toda a informação necessária à análise de assuntos Automatizada. Nesse sentido, Luhn (1957) e Baxendale (1958) afirmam que o significado do texto pode ser deduzido pelo algoritmo computacional do contexto, da frequência e da coocorrência dos termos do texto. Pressupõe-se que os assuntos (tópicos) podem ser representados por termos de um vocabulário controlado (ver o Pressuposto 3) que, no momento da busca por informação, podem ser representados por outros termos pelo usuário, em razão do contexto cultural, mas que, no momento da tradução dos assuntos do texto em termos do vocabulário controlado, o indexador humano seja capaz de antecipar essa diferença (ROSENBERG, 1971).
3. O conteúdo temático do texto pode ser representado por uma lista de termos de um vocabulário controlado que permite o processamento do documento no nível simbólico, porque algoritmos computacionais lidam com representações (símbolos) e não com o significado cultural em si. Isso também explica o porquê de os processos automáticos serem velozes e o porquê de apresentarem inesperadamente respostas de baixa confiabilidade, mesmo em modelos excelentes (HINTON, (2020); DEERWESTER *et al.* (1990); HOFMANN (1999) e BLEI; NG; e JORDAN (2003)).
4. O termo “inteligência aumentada” é polissêmico, pois significa, ao mesmo tempo, padrão de projeto, o processo (uso) e o resultado (consequência). Assim, a inteligência aumentada nomeia o planejamento, o processo e a consequência (expansão cognitiva do usuário) do uso do artefato com inteligência artificial (YAU *et al.*, (2021); PASQUINELLI (2015, p. 203); (KOVARIK *et al.*, 2019); e SADIKU *et al.* (2021)).

A qualidade dos resultados de um sistema computacional é avaliada por meio da comparação de seus resultados com um padrão-ouro, que, no caso das indexações, são os

termos indexadores elaborados manualmente. A importância do Pressuposto (1) na compreensão de que modelos obtidos têm taxas de acerto muito baixas. Segundo Narukawa *et al.* (2009), em uma análise comparativa entre dois sistemas automatizados de indexação, o índice médio de consistência de atribuição de termos na indexação foi em torno de 23%.

O Pressuposto (2) de que o texto contém toda a informação imprescindível, aborda a condição necessária e suficiente para garantir que os algoritmos sejam capazes de elaborar informações documentárias, mesmo sem compreender a informação no nível semântico e sem ultrapassar os limites impostos pelos dados processados. Dessa forma, entende-se que os termos indicadores dos assuntos estarão presentes no texto e, caso contrário, será necessária a intervenção do profissional da informação.

O assunto pode transcender o texto, mas estará contido em um grupo de textos dentro do *corpus*, conforme preconiza o Pressuposto (3). Significa dizer, que o assunto principal, em *corpus* monotemático, ainda que relevante, não acrescenta valor à informação documentária, uma vez que ele recupera todos os documentos. Por exemplo, num *corpus* de Agricultura, o termo “agricultura” é relevante, mas recupera tudo. Um termo compartilhado por vários textos delimita um grupo de documentos (*cluster*), mas precisa estar acompanhado por termos mais específicos para individualizar um documento a ponto de torná-lo recuperável. O Pressuposto (3) implica na necessidade de processar computacionalmente o *corpus* inteiro ao se analisar o assunto de um texto (BENGIO *et al.*, 2003).

Depreende-se do Pressuposto (3), que um termo no vocabulário controlado é um símbolo que representa um assunto e que o índice de assuntos representa o documento, porque para o algoritmo, a informação documentária não tem significado semântico, sendo apenas um dado computável, um símbolo que representa, substitui e identifica o documento na base de dados para permitir a sua recuperação (ALDOUS, 1985).

O Pressuposto (4) assegura o uso adequado do termo “inteligência aumentada” em diferentes situações relatadas nesta tese, tanto no Modelo Conceitual quanto na análise de resultados.

Por fim, esses pressupostos asseguram o êxito da proposta de assistência às atividades documentárias. A aplicação deles viabiliza a utilização da inteligência artificial sob a supervisão humana; ajuda na adoção de uma análise de assuntos menos subjetiva e mais coerente; e corrobora com a ideia de que as teorias, as técnicas e os métodos desenvolvidos no âmbito da BCI são adequados para enfrentar os desafios do *big data*, quando utilizados para alcançar os objetivos descritos na próxima seção.

1.4 Objetivos

Esta pesquisa pretende beneficiar bibliotecários, analistas de assunto, indexadores e pesquisadores que necessitem de processar informações em grandes repositórios textuais (*big data*).

1.4.1 Objetivo geral

Elaborar uma proposta metodológica de um modelo de inteligência aumentada para viabilizar a análise de assuntos assistida por computador.

1.4.2 Objetivos específicos

Os objetivos específicos da pesquisa, como etapas para alcançar o objetivo geral, foram formulados pela combinação dos objetivos conceituais e procedimentais, conforme apresentado a seguir.

1. **Contribuir** com estudos de inteligência aumentada para a Organização da Informação.
2. Listar e articular as melhores práticas de elaboração de Modelos de inteligência artificial para a análise de assuntos: **contribuir** com metodologia eficaz para a análise de assuntos assistida por computador.
3. Implementar um protótipo tecnológico: desenvolver uma ferramenta para **viabilizar** o uso do modelo desenvolvido (Modelo de inteligência aumentada).
4. Avaliar o Modelo de inteligência aumentada desenvolvido por meio de uma simulação computacional (experimento): **validar** a solução desenvolvida.

1.5 Estrutura da tese

Esta tese está estruturada em quatro partes correspondentes às etapas dos procedimentos metodológicos, da seguinte forma:

Na Etapa 1 (planejamento geral), foram abordadas a introdução, a justificativa, os pressupostos e os objetivos concernentes à aplicação do modelo de inteligência aumentada na análise de assuntos (Capítulo 1). Ainda nesta etapa, foram descritas a revisão de literatura (Capítulo 2), a fundamentação teórica e metodológica (Capítulo 3) e a metodologia (Capítulo 4).

Na Etapa 2 (elaboração da base de conhecimento), elaborou-se uma base de conhecimento, cujo relato está no Capítulo 5.

Na Etapa 3 (desenvolvimento do *software*), desenvolveu-se o *software* de **análise de assuntos assistida por computador (3ApC)**, que é um protótipo tecnológico, cujo relato está no Capítulo 6.

Na Etapa 4 (Análise dos resultados), realizou-se um experimento de validação do *software* produzido. Os resultados e as análises estão descritos Capítulo 7.

Apresenta-se ainda as conclusões da pesquisa (Capítulo 8) e as referências bibliográficas.

Um conjunto de sete Apêndices apresenta os dados, resultados e análises que complementam a compreensão da tese.

Considerando que “todo conhecimento comporta o risco do erro e da ilusão” (Morin, 2005), por meio desta breve descrição da estrutura da tese, espera-se que o leitor seja capaz de recompor o percurso da pesquisa.

No próximo capítulo, apresentam-se a Revisão de Literatura sobre a inteligência aumentada e os achados fundamentais na literatura científica sobre o tema.

2 REVISÃO DE LITERATURA SOBRE A INTELIGÊNCIA AUMENTADA

Neste capítulo, apresenta-se a revisão da literatura dos trabalhos recuperados das bases de dados científicas pautados pela aplicação da inteligência aumentada, em que cada documento foi analisado individualmente, porque a pesquisa bibliográfica recuperou poucos itens.

2.1 Questões preliminares

A inteligência aumentada é o moderno processo de sistematizar o uso da inteligência artificial para melhorar a tomada de decisões (SADIKU *et al.*, 2021). Trata-se de um padrão de projeto que combina competências e capacidades complementares de humanos e computadores na resolução de problemas (PAVLOU, 2018).

As percepções (*insights*) extraídas do processamento de dados (*big data*) por meio de técnicas da inteligência artificial ampliam a compreensão humana de um problema por dar uma visão completa do cenário. Mas essa riqueza de informações ocorre pela mediação das ferramentas computacionais que “não tem consciência” (GUIGNARD *et al.*, 2020, p. 29), assim, a função da supervisão humana é evitar que decisões ruins da inteligência artificial sejam executadas, e, assim, impedir que preconceitos e discriminações sejam reforçados e que falhas catastróficas ocorram.

Submeter os achados da inteligência artificial à supervisão humana promoveu uma expansão cognitiva do intelecto (ASHBY, 1956) e produziu uma nova realidade em que a percepção humana da “fronteira entre o ciberespaço e o mundo físico ficou esmaecida” (BOSCHERT, 2019, p. 7). Soma-se a isso o fato de que as tecnologias inteligentes se tornaram muito comuns e as áreas do conhecimento que não as empregarem diligentemente podem se tornar obsoletas. Neste sentido, é importante para a BCI ampliar os estudos sobre a contribuição que essas tecnologias podem dar às atividades profissionais da área. Assim, a função deste estudo é ser informativo sobre o estado atual da inteligência aumentada na BCI, considerando a perspectiva apresentada em Maculan (2015) de que uma revisão da literatura deve preservar a história de um desenvolvimento científico (função histórica) e atualizar a comunidade de pesquisadores (função de atualizar) sobre os avanços científicos sobre a temática revisada.

Na próxima seção, apresentam-se os procedimentos adotados na revisão.

2.2 Procedimentos adotados na revisão

Este trabalho caracteriza-se como uma pesquisa descritiva de abordagem qualitativa, em que a seleção dos dados se realizou por meio de uma pesquisa bibliográfica sobre o uso de da inteligência artificial para assistir à tomada de decisões humana.

A análise dos dados e a síntese dos resultados seguiu o percurso da revisão da literatura narrativa e a metodologia está descrita na seção 4.3.

A coleta e a seleção dos dados têm a seguinte estrutura: 1) estabelecimento dos critérios de seleção dos estudos (seção 2.3.1); 2) busca na literatura (seção 2.3.2); 3) seleção dos dados (seção 2.3.3); 4) avaliação da contribuição de cada estudo selecionado à temática (seção 2.3.4); e 5) análise dos dados selecionados e síntese dos resultados (Seção 2.3.5).

Nesta etapa, definiram-se três questões, a saber: o que é a inteligência aumentada (definição)? Em quais contextos a inteligência aumentada foi utilizada? Quais as consequências do uso da inteligência aumentada?

A partir das questões acima, definiram-se os critérios de inclusão de um estudo no *corpus* da pesquisa: 1) conter a palavra-chave “*augmented intelligence*” ou “inteligência aumentada”; e 2) versar sobre o uso de *softwares* com inteligência artificial supervisionado pela inteligência humana.

São relevantes para esta pesquisa os artigos sobre o uso de *softwares* com inteligência artificial para assistir à tomada de decisão humana, independente da área do conhecimento em que a pesquisa foi realizada. Foram considerados não relevantes para esta revisão da literatura os estudos sobre hiperautomação, teleoperação, transumanismo e realidade aumentada, porque versam sobre o uso de hardware.

Para selecionar os estudos que respondam às questões de pesquisa e atendam aos critérios de elegibilidade, foi realizada uma busca sistemática na literatura, conforme descrito na próxima seção.

Para buscar e coletar os estudos, foram selecionadas um conjunto de bases científicas de dados abrangente o suficiente para cobrir os aspectos transdisciplinares do assunto pesquisado, conforme recomendado pela *International Society For Knowledge Organization* (2007).

A expressão “*augmented intelligence OR inteligência aumentada*” foi utilizada para a busca sistemática nas seguintes bases científicas de dados: *Library and Information Science Abstracts* (LISA); *Library, Information Science and Technology Abstracts* (LISTA); *Institute of Electrical and Electronics Engineers* (IEEE); *Scopus* (base de dados de resumos e citações de literatura revisada por pares); *Web of Science* e, também, no sistema automatizado de recuperação da informação Google Acadêmico.

A coleta de dados resultou na recuperação de 105 documentos, dos quais apenas 18 foram selecionados. Uma parte dos artigos estava inacessível, outra parte apareceu em mais de uma base (duplicados), ou não eram relevantes para a pesquisa.

Tabela 1 -- Resultados da busca nas bases de dados sobre a inteligência aumentada.

Base de dados	Encontrados	Relevantes	Selecionados
LISA	12	0	0
LISTA	8	5	3
<i>Web of Science</i>	12	5	2
<i>Scopus</i>	12	5	0
IEEE	33	12	5
Google Acadêmico	28	15	8
Total	105	42	18

Fonte: Elaborado pelo autor.

Na literatura em português, os trabalhos apenas citavam o termo “inteligência aumentada”, mas não eram estudos sobre a temática, de forma que não foram incluídos nesta revisão de literatura.

O *corpus* obtido não é exaustivo, devido à natureza interdisciplinar, multidisciplinar e multiparadigmática do assunto pesquisado, no entanto, suficiente para se obter uma compreensão do conceito e do contexto implicados na inteligência aumentada.

As publicações coletadas foram submetidas aos critérios de elegibilidades e verificadas quanto à pertinência delas em relação à inteligência aumentada, conforme está relatado na próxima seção.

Foram encontradas pesquisas em múltiplas áreas do conhecimento e em suporte a um grupo bastante diverso de profissionais, que incluiu oncologistas, psiquiatras, operários, educadores e decisores operacionais. No todo, o *corpus* coletado permitiu avaliar as questões epistemológicas, éticas e filosóficas do uso da inteligência aumentada, bem como os efeitos produzidos por esse uso.

Em relação às fontes de informação, o IEEE apresentou o maior conjunto de documentos relacionados à temática e o Google Acadêmico, o maior número de trabalhos relevantes.

Os estudos avaliados como pertinentes à inteligência aumentada foram classificados em dois grupos: (1) estudos práticos; e (2) estudos teóricos.

Os autores Golestan, Mahmoudi-Nejad e Moradi (2019), Ko, Yen e Chen (2020), Frownfelter *et al.* (2020), Xu *et al.* (2020), Thamm *et al.* (2020) e Liao *et al.* (2020)

apresentaram em seus estudos a metodologia de uso, os *softwares* (Quadro 1) e os impactos da aplicação da inteligência aumentada, portanto, pertencem ao Grupo 1.

Quadro 1 – Estudos práticos de inteligência aumentada encontrados na literatura.

Descrição do estudo	Autoria
Um <i>framework</i> para dar autonomia às crianças em jogos eletrônicos.	Golestan, Mahmoudi-Nejad, Moradi (2019)
Um <i>software</i> para identificar automaticamente as varreduras de ventilação pulmonar prejudicadas por aprisionamento de aerossol. A decisão final ficou reservada às pessoas.	Ko, Yen e Chen (2020)
Um <i>software</i> para analisar dados clínicos e antecipar quais pacientes de quimioterapia oncológica eram susceptíveis à depressão. A adoção de terapias preventivas ficou a cargo dos médicos.	Frownfelter <i>et al.</i> (2020)
Um <i>software</i> para tomar decisões operacionais, sob a supervisão dos funcionários, em uma empresa de despacho de pedidos.	Xu <i>et al.</i> (2020)
Um hardware de assistência ao trabalhador na linha de montagem manual, em que o sistema detecta um erro e comunica ao operário, tanto o erro, quanto a solução.	Thamm (2021)
Um <i>software</i> distribuído para viabilizar o trabalho coletivo de 30 <i>designers</i> .	Liao <i>et al.</i> (2020)

Fonte: Elaborado pelo autor.

Os autores Cerf (2013), Pasquinelli (2015), Zheng *et al.* (2017), Rui (2017), Fong *et al.* (2019), Kennedy e White (2020), Tan e Teo (2020), Romero-Brufau *et al.* (2020), Akata *et al.* (2020), Guignard *et al.* (2020), Rivera e Gleeson (2020), Liu (2020) e Wójcik (2020) apresentaram em seus estudos os fundamentos teóricos (Quadro 2) na forma de discussões e análises do uso da inteligência aumentada. Portanto, pertencem ao Grupo 2.

Quadro 2 – Estudos teóricos da inteligência aumentada encontrados na literatura.

Descrição	Autoria
Discute a relação entre o artigo de Vannevar Bush, " <i>As We May Think</i> " e o conceito de inteligência aumentada de Engelbart.	Cerf (2013)
Discute inteligência aumentada desde um ponto de vista teórico-filosófico para superar as visões de racionalidade instrumental.	Pasquinelli (2015)
Apresentam novas perspectivas para a inteligência artificial em que se investigou cientificamente como aumentar o intelecto humano sem substituir o ser humano pela máquina.	Zheng <i>et al.</i> (2017)
Discute o impacto do uso da inteligência artificial e sugere "IA + humanos" no lugar de "IA x humanos".	Rui (2017)

Descrição	Autoria
Apresenta o método de inteligência aumentada, por meio de raciocínio semântico na automação, que modela a vagueza inerente entre os conceitos e os fenômenos no mundo real.	Fong <i>et al.</i> (2019)
Discute as questões relacionadas ao futuro da inteligência aumentada com base no ciclo de representação, predição e comunicação.	Kennedy & White (2020)
Analisa a aprendizagem com dados e inteligência artificial.	Tan & Teo (2020)
Pesquisa como os profissionais da saúde distinguem inteligência aumentada de inteligência artificial na prática médica.	Romero-Brufau <i>et al.</i> (2020)
Discute a ética do uso da inteligência aumentada	Akata <i>et al.</i> (2020)
Analisa eticamente a inteligência aumentada.	Guignard <i>et al.</i> (2020)
Propõe um modelo de inteligência aumentada para democratizar o acesso dos funcionários aos dados da empresa para que se tornem versões melhores de si.	Rivera & Gleeson (2020)
Discute o uso da inteligência aumentada nos cuidados médicos críticos.	Liu (2020)
Apresenta uma revisão de literatura sobre a inteligência aumentada na Biblioteconomia e Ciência da Informação.	Wójcik (2020)

Fonte: Elaborado pelo autor.

Até aqui, descreveram-se os procedimentos sistemáticos para a coleta e a seleção dos dados. Foram encontrados três trabalhos relacionados à BCI, quatro pesquisas na área da medicina, seis artigos específicos para as engenharias e cinco voltados para a ciência da computação.

Na próxima seção, estão descritas a análise dos dados coletados e a síntese dos resultados sob as diretrizes de uma revisão de literatura narrativa.

2.3 Descrição dos estudos

Na descrição dos trabalhos, na medida do possível, apresentaram-se as informações conforme as questões de pesquisas propostas na metodologia que abrangem as definições, os contextos de uso e os benefícios da inteligência aumentada.

O volume de dados selecionado não permitiu uma análise sistemática por meio de métodos estatísticos, por isso, empregou-se uma análise narrativa da literatura, em que cada trabalho foi analisado individualmente.

As análises dos dados selecionados e a síntese dos resultados estão apresentadas na ordem cronológica das datas de publicação. Começando por Cerf (2013)

que traçou uma linha temporal do desenvolvimento do conceito e do termo inteligência artificial, que, segundo ele, começou em 1945 com Vannevar Bush e adquiriu o significado atual no trabalho de Douglas Engelbart, em 1962.

Em Pasquinelli (2015), o autor afirma que a inteligência computacional expande social e politicamente as faculdades cognitivas humanas, mas que o assunto está sendo subestimado na teoria da mídia, nas ciências cognitivas, nas neurociências e nas filosofias da mente e da política e que essa falta de compreensão da complexidade da inteligência aumentada e dos seus impactos pode trazer consequências sérias para o futuro da humanidade.

As três afirmações teóricas, abaixo, encontradas em Zheng *et al.* (2017, p. 1) são importantes para a conceituação e a compreensão da inteligência aumentada: (1) o objetivo de longo prazo da inteligência artificial é fazer a máquina “aprender” e “pensar” como um ser humano; (2) os altos níveis de incerteza e de vulnerabilidade da vida, associados à natureza aberta dos problemas humanos, deixa a máquina incapaz de substituir completamente os seres humanos, independentemente do quão inteligente ela seja; e (3) combinar a capacidade cognitiva humana com a capacidade da máquina nos processos e atividades humanos desenvolve um aumento de inteligência na solução dos problemas.

Corroborando essa perspectiva, Rui (2017, p. 3) analisa o impacto da inteligência artificial na sociedade humana e aponta uma ruptura dos meios tradicionais de trabalho, sugerindo uma colaboração, “IA + humanos”, para evitar o confronto “IA x humanos”.

Para resolver as questões relacionadas à automação, Fong *et al.* (2019) desenvolveram um método de inteligência aumentada, que, por meio de raciocínio semântico, lida com a vagueza inerente dos conceitos para descrever coisas e fenômenos físicos. O método associa aprendizagem ontológica a objetos reais, encapsula referência no mundo e informações contextuais, para permitir o deslocamento seguro de entes robóticos.

As questões relacionadas ao futuro da inteligência aumentada foram discutidas em Kennedy e White (2020) com base na influência mútua de três conceitos fundamentais: representação, previsão e comunicação. A representação do conhecimento na mente inteligente, mediante o uso da inteligência artificial, faz a previsão de um estado futuro do mundo e comunica essa predição a outras mentes inteligentes. A realimentação constante desse ciclo virtuoso leva ao aumento da inteligência, pois impulsiona mudanças disruptivas, reforça a percepção sensorial (representação), expande a cognição (previsão) e melhora a interação humana (comunicação).

O modelo de inteligência aumentada proposto por Golestan, Mahmoudi-Nejad e Moradi (2019) constitui-se na estrutura computacional de desenvolvimento de jogos inteligentes que dá autonomia à criança para configurar os mecanismos complexos dos jogos

inteligentes. O objetivo foi facilitar a vida da criança com deficiência que nem sempre pode contar com a assistência de um adulto.

O modelo de inteligência aumentada analítico multidimensional elaborado por Xu *et al.* (2020) foi desenvolvido para investigar a tomada de decisão operacional. Na empresa analisada, a inteligência artificial desempenhou papel importante nas decisões e ampliou a capacidade cognitiva dos funcionários.

Um sistema de assistência ao trabalhador na linha de montagem é necessário, porque, segundo Thamm *et al.* (2021, p. 2), os “operários em países em desenvolvimento têm baixo nível de formação”. Duas câmeras de vídeo com inteligência artificial monitoram a atividade industrial e, quando identificam um erro, o *software* inteligente do sistema elabora uma solução e a comunica ao operário. Os autores não apresentaram a viabilidade técnica do uso desse sistema em um cenário real.

A argumentação ética de Akata *et al.* (2020, p. 20) sobre o uso da inteligência aumentada sustenta-se em quatro princípios: “colaboração, adaptação, responsabilização e explicação” e recomenda prudência no uso da inteligência artificial, porque ainda não se compreende a consciência, os objetivos e as estratégias desses sistemas.

O maior desafio da atualidade é lidar com o *software* inteligente, que pode ser visto como uma ferramenta a mais, no entanto, uma diferente de todas as outras, pois tem capacidade cognitiva de tomar decisões e não tem responsabilidade pelas consequências. Portanto, cabe à sociedade disciplinar o uso da inteligência artificial, assumir a responsabilidade pelas decisões automáticas e tratar essas ferramentas como o que elas realmente são: próteses “que estendem a cognição da mesma forma que um martelo estende a ação corporal” (GUIGNARD *et al.*, 2020, p. 2).

A inteligência aumentada pode melhorar a produtividade dos trabalhadores ao ampliar o poder de processamento, análise e compartilhamento de dados no local de trabalho. Assim, organizações inovadoras precisam democratizar o acesso aos dados da empresa para que seus funcionários se tornem “versões mais eficazes e capacitadas de si mesmos” (RIVERA; GLEESON, 2020, p. 12).

O *framework* de inteligência aumentada desenvolvido por Liao *et al.*, (2020) para as atividades genéricas de design promoveu uma colaboração efetiva entre 30 *designers* e entre eles e as máquinas, alavancando a criatividade e a produção.

Uma pesquisa (*survey*) realizada por Romero-Brufau *et al.* (2020) investigou como os profissionais da saúde distinguem os sistemas inteligentes, porque não estava claro se a terminologia usada afetava as atitudes e as percepções na prática médica. As conclusões da pesquisa foram de que os termos “inteligência artificial” e “inteligência aumentada” são entendidos como sinônimos para o corpo clínico pesquisado.

O sistema computacional desenvolvido por Frownfelter *et al.* (2020) identificou o risco de depressão em pacientes sob quimioterapia contra o câncer. A solução consistiu em deixar para o médico a decisão final sobre tratamento contra a depressão, mas a partir das recomendações que o *software* com inteligência artificial forneceu após avaliar os dados do paciente. A pesquisa concluiu que esse modelo de inteligência aumentada preveniu a depressão e melhorou a qualidade de vida dos pacientes.

Outro modelo de inteligência aumentada na área médica foi apresentado por Ko, Yen e Chen (2020), que fez previsões, em tempo real, de defeitos em equipamentos hospitalares provocados por aerossóis. Nesse modelo, os dados coletados por sensores foram analisados para emitir alertas antecipados que permitiram a equipe médica intervir preventivamente. Assim, humanos e máquinas colaboraram para salvar vidas.

A inteligência artificial já afeta a vida de pacientes médicos e, no futuro próximo, vai alterar completamente os cuidados com a saúde, inclusive os intensivos. Devido aos perigos potenciais dessa nova realidade, todos os procedimentos devem ser planejados com antecedência para produzir inteligência aumentada como garantia de benefícios para médicos e pacientes:

trabalhar para garantir que os recursos formidáveis da inteligência artificial sejam usados para produzir inteligência “aumentada”, em vez de apenas “artificial”. A inteligência aumentada coloca os médicos e, em última análise, os pacientes, em vez de algoritmos, no centro. Onde conecta-se com sucesso a interface de inteligência clínica e a de máquina, tem um vasto potencial para tornar a assistência médica mais eficaz, eficiente e sustentável. Isso também garantirá que a inteligência artificial em saúde seja segura, confiável e equitativa para todos os pacientes (LIU, 2020, p. 1, tradução nossa).

Duas revisões de literatura foram encontradas sobre a inteligência aumentada. Uma revisão de literatura sistemática feita por Tan e Teo (2020) para analisar a aprendizagem humana mediada por inteligência artificial e a outra feita pela professora polonesa Magdalena (WÓJCIK, 2020) sobre a inteligência aumentada na perspectiva da BCI. Ambas concluem que a temática é pouco pesquisada tanto na área da educação quanto na literatura da BCI.

Na próxima seção, apresenta-se a discussão geral dos dados desta revisão de literatura narrativa.

2.4 Análise qualitativa dos estudos

A inteligência aumentada, por convenção, concretiza-se no uso de inteligência artificial associado à inteligência humana, mas não se limita a isso, porque a expansão cognitiva ocorre sob o uso de ferramentas. Algumas ferramentas atuais possuem capacidades

cognitivas que podem até substituir as pessoas em atividades laborais, mas elas não acrescentam os valores humanos aos resultados, por isso há necessidade de combinar essas duas inteligências, a natural e a artificial.

Colocar a Inteligência Natural para supervisionar e controlar a inteligência artificial produz soluções mais seguras, eficientes e humanizadas. Confirmam essa proposição as áreas do conhecimento beneficiadas pela inteligência aumentada, como o transporte (FONG *et al.*, 2019), a educação (TAN; TEO, 2020), a medicina (FROWNFELTER *et al.*, 2020; KO; YEN; CHEN, 2020; LIAO *et al.*, 2020), o suporte a crianças com necessidades especiais (GOLESTAN; MAHMOUDI-NEJAD; MORADI, 2019), as decisões operacionais (XU *et al.*, 2020) e as atividades de *design* coletivo (LIU, 2020).

A tomada de decisões alicerçadas nos valores humanos universais, como a veracidade da informação, a autodeterminação política, a proteção de dados individuais e a confiabilidade técnica é, provavelmente, mais necessária hoje do que em qualquer outra época (BOSCHERT, 2019). Haja vista, as empresas multinacionais, que, por razões comerciais, usam apenas inteligência artificial na recuperação da informação e o faz com tanto êxito que rivalizam em poder com os estados nacionais. No entanto, por não incluir a supervisão humana no processo, graves consequências podem estar em curso, geradas pela incapacidade das máquinas em excluir as informações falsas. A inclusão da inteligência humana no processo pode ajudar, porque ela é melhor em distinguir as informações verdadeiras (KORTELING *et al.*, 2021).

Conclui-se que os riscos associados ao uso de inteligência artificial não devem ser subestimados, pois a natureza opaca desses sistemas associada aos conjuntos de dados tendenciosos pode agravar problemas sociais. Por isso, submeter as decisões da inteligência artificial à supervisão da inteligência humana produz soluções mais seguras e eficazes para uma quantidade maior de problemas práticos do que usar uma dessas inteligências isoladamente.

A colaboração entre os humanos e os sistemas inteligentes exige novas pesquisas e descobertas. Áreas como a análise de assuntos que demandam senso crítico, criatividade e imaginação não podem ser inteiramente delegadas aos sistemas inteligentes, porque eles são produzidos em contextos culturais humanos e, por isso retêm, replicam e reforçam os vícios culturais (preconceitos, discriminações e crenças exclusivistas). Logo, pelo menos por enquanto, as pessoas são indispensáveis na tomada de decisão nos processos que atendem ou lidam diretamente com as pessoas.

A inteligência aumentada pode ser introduzida no domínio da Organização do Conhecimento como um princípio metodológico do uso da inteligência artificial necessário à elaboração, manutenção e atualização de Sistemas de Organização do Conhecimento (SOC),

principalmente quando essas atividades necessitam de processamento de *corpus*, uma vez que os textos, mesmos os científicos, podem conter viés discriminatório que só podem ser percebidos pela inteligência humana.

A inteligência aumentada adiciona inteligência artificial aos processos conduzidos por pessoas, e essa característica coloca a inteligência aumentada no campo das Ciências Sociais Aplicadas. Essa afinidade epistemológica soma-se a possibilidade de se aumentar a eficiência dos meios de produção e atualização de vocabulários controlados, de tesouros, de classificações e indexação de assuntos, o que, certamente, faz da inteligência aumentada uma ferramenta valiosa para os profissionais da informação.

Este estudo contribuiu com a compreensão do conceito, apresentou a terminologia do campo e mostrou que a inteligência aumentada é uma realidade em muitos campos científicos. Com isso, entende-se que, na BCI, também é necessário incorporar as tecnologias inteligentes, uma vez que o volume de publicações (a matéria-prima da BCI) cresce num ritmo acima do que os métodos manuais de indexação, por exemplo, conseguem processar.

Concluiu-se que a literatura sobre inteligência aumentada ainda é incipiente na área da Biblioteconomia e Ciência da Informação, o que aponta à necessidade de desenvolver novos estudos para incutir os valores da BCI na solução dos problemas da área com a assistência da inteligência artificial, e, com isso, atingir o objetivo de ampliar e favorecer o acesso do usuário às informações relevantes.

3 FUNDAMENTAÇÃO TEÓRICA E METODOLÓGICA

Neste capítulo, apresentam-se as temáticas para a realização e a avaliação da pesquisa, incluindo o desenvolvimento do modelo conceitual, do artefato computacional e do experimento de validação sob os princípios da inteligência aumentada.

Neste sentido, este capítulo foi estruturado da seguinte forma: 3.1 análise de assuntos: o objeto da pesquisa; 3.2 inteligência artificial (ia): área que congrega os princípios teóricos da metodologia utilizada; 3.3 processamento de linguagem natural: campo da inteligência artificial da metodologia desenvolvida; 3.3.4.1 sumarização automática: produz resumos automáticos; 3.3.4.2 expressões regulares: identificam os assuntos do texto; 3.3.4.3 modelagem de tópicos: selecionam os assuntos relevantes; e 3.7 inteligência aumentada: princípios para o planejamento, o desenvolvimento e o uso de artefatos computacionais com inteligência artificial.

3.1 Análise de assuntos

A análise de assuntos é a primeira etapa do processo de indexação. Quanto a terminologia utilizada, admite-se que haja uma certa “confusão conceitual” (DIAS; NAVES, 2004, p. 6), assim a análise de assuntos pode ser denominada na literatura científica da área de “análise temática”, “análise documentária” e “análise de conteúdo”. Nesse sentido, Lima (2021a, p. 8) afirma que:

Existe na literatura da área da Biblioteconomia e Ciência da Informação uma variação terminológica em relação ao termo análise de assuntos, que também pode ser denominado de análise temática ou análise conceitual; entretanto, [...], o termo análise de assuntos é o que melhor representa o processo de representação de conteúdo de um documento.

Assim, nesta pesquisa, optou-se pelo termo “análise de assuntos” com o mesmo sentido de “análise conceitual”, conforme explicado em:

Deve-se esclarecer que mesmo o termo “análise de assuntos” pode ser usado por abordagens diferentes dentro da área de biblioteconomia e ciência da informação. [...], duas diferentes abordagens: uma como uma área do conhecimento que estuda a construção e uso das linguagens de indexação e sistemas de classificação; outra como o processo de análise conceitual, primeira etapa da indexação de um documento (LIMA; MACULAN, 2016, p. 25).

Considerou-se que o processo de indexação envolve as etapas: (1) análise de assuntos; e (2) Tradução (Figura 2). A etapa de Tradução caracteriza-se pela decisão do indexador de quais (e como os) assuntos serão traduzidos para os termos de um vocabulário

controlado e corresponde a etapa de tomada de decisões da inteligência aumentada (ver seção 3.5).

Figura 2 – Esquema hierárquico da análise de assuntos.



Fonte: elaborado pelo autor.

A etapa de análise de assuntos do processo de indexação pode ser dividida em três estágios: (1) ler o texto/compreender o texto; (2) identificação dos assuntos do texto; e (3) seleção dos assuntos relevantes (LIMA; MACULAN, 2014). Depois do estágio 3, aplica-se a etapa de tradução do processo de indexação.

Conforme orientação por escrito da professora Dra. Gercina Ângela de Lima, em 2022, “nem sempre os autores dão ênfase à análise de assuntos, a maioria enfatiza a etapa de tradução, em que ocorre a representação da informação”.

Nesta pesquisa, considera-se que o processo de análise de assuntos pode ser:

1. **intelectual:** todas as etapas são feitas por pessoas;
2. **automatizado:** todas as etapas são realizadas por máquinas. Mas essa abordagem “totalmente automatizada” ocorre apenas sob “uma concepção simplista”, segundo Albrechtsen (1993, p. 223);
3. **semiautomático:** todas as etapas são realizadas pela máquina, mas quando ocorre uma resposta de baixa confiabilidade, o sistema interrompe o processo e aguarda até que uma pessoa decida o que será feito; e
4. **assistido por computador:** todas as etapas são feitas por pessoas, que recebem a assistência da inteligência artificial para realizar as etapas repetitivas. Todas as decisões são tomadas por pessoas.

A análise de assuntos de um documento deve ser orientada para o significado do conteúdo de um documento de modo a suprir as necessidades de informação dos usuários. Para tanto, ela responde duas questões fundamentais: (1) do que trata o documento? esta é a indexação orientada ao documento; e (2) a qual assunto buscado pelo usuário o documento responde? esta é a indexação orientada ao usuário (SOUSA; FUJITA, 2014).

Assim, a análise de assuntos orientada ao documento deriva do significado do texto e pode ser obtida pelo processamento simbólico mais afeito à inteligência artificial, enquanto a análise de assuntos orientada ao usuário depende do contexto cultural (LIMA; FUJITA; REDIGOLO, 2021) e, portanto, tem mais afinidade com o indexador.

Na próxima seção, apresentam-se uma breve história da análise de assuntos.

3.1.1 Breve histórico

Nesta seção, aborda-se a análise de assuntos, dentro da perspectiva histórica, como um processo implícito à indexação, ocorrendo intuitivamente ainda antes de que lhe fosse atribuído a terminologia e as etapas dos dias de hoje.

Dessa perspectiva, a análise de assuntos ocorreu na Biblioteca de Alexandria e nas bibliotecas sumérias dois séculos antes da era cristã. Deduz-se isso do fato de que Calímaco, no século III a.C., em Alexandria, catalogou as obras da época por assuntos (SILVA; FUJITA, 2004); e de que, nas ruínas das bibliotecas sumérias, foram encontrados resumos dos textos das tábuas de argila (KOBASHI, 1996). Em ambos os casos, a análise de assuntos foi utilizada intuitivamente, considerando-se que a etapa de análise de assuntos é necessária para catalogar por assuntos e resumir um documento.

A Documentação, cujos processos incluem (mas não se limitam) a análise de assuntos implicitamente, começou a ser sistematizada para a circulação pública de informações, em 1665, em Paris, com a edição do "*Le Journal des Sçavans*" da Academia Francesa de Ciências. Essa e outras publicações semelhantes "cresceram em número" e aperfeiçoaram seus métodos até alcançar o formato atual (KOBASHI, 1996, p. 7).

Com a revolução tecnológica do século XX, a partir dos trabalhos de Luhn (1957) e Baxendale (1958), a análise de assuntos recebeu aportes dos métodos computacionais para a automatização, chegando inclusive, segundo Albrechtsen (1993, p. 220), a "dominar as pesquisas" de recuperação da informação durante as décadas de 1970 e 1980.

Atualmente, a indexação automatizada não utiliza a análise de assuntos, aplica técnicas de inteligência artificial diretamente nos textos, numa abordagem simplista, que, conforme vaticinado por Albrechtsen (1993, p. 223), além de produzir resultados pobres na pesquisa científica, elimina empregos e não transfere e nem usa otimamente o conhecimento.

Por fim, a análise de assuntos pode ser reconhecida nos primeiros processos de organizar o conhecimento, ainda que praticada intuitivamente; na Documentação, que a partir do século XVII adquiriu uma metodologia própria; e, no formato atual, que, desde o século XX, recebe o aporte tecnológico para a sua automatização.

As definições de análise de assuntos estão descritas na próxima seção.

3.1.2 Definições

Nesta seção, serão abordadas as definições da análise de assuntos como processo crucial da indexação, em que o processo consiste em ler e compreender o texto (ALBRECHTSEN, 1993).

A principal finalidade da indexação é representar a informação significativa de um documento para viabilizar a sua posterior recuperação. Para tanto, ela utiliza a análise de assuntos para reconhecer e selecionar os assuntos mais significativos do texto e traduzi-los para os termos de um vocabulário controlado para criar uma “linguagem intermediária entre o usuário e o documento” (VIEIRA, 1988, p. 43).

A análise de assuntos foi definida em Albrechtsen (1993, p. 220, tradução nossa) como um "processo intelectual ou automatizado pelo qual os assuntos de um documento são analisados para, posteriormente, serem expressos por termos de indexação”.

Em Fujita *et al.* (2009, p. 23), a análise de assuntos foi definida como a “parte mais importante da análise documentária”, sendo inclusive o “condicionador do valor do sistema documentário”.

O problema crucial de identificar todos os assuntos de um documento, segundo Hutchins (1977) é tratado na análise de assuntos. Mas segundo Mai (2001), jamais se descobrirão todos os significados associados a um documento, portanto uma tarefa irrealizável. Nesse sentido, Lima (2021, p. 7) afirma que a análise de assuntos é “uma arte e não uma ciência”, devido à sua complexidade.

Segundo Lancaster (2004, p. 9), “não há um conjunto *correto* de termos de indexação para documento algum” dada a natureza dinâmica da análise de assuntos, que precisa saber: do que se trata no documento; e qual é o “provável interesse” de um determinado grupo de usuários.

O processo da análise de assuntos é complexo devido ao seu caráter subjetivo, às multiplicidades de percursos e à diversidade de interesses dos grupos de usuários atendidos. Nesse sentido, afirma Foskett (1973, p. 40), que “a decisão sobre o que o documento é, além de ser pouco discutida, não pode ser reduzida a regras”, pois as motivações, as atitudes e as decisões do profissional da informação ao escolher os termos representativos dependem do seu estado cognitivo e emocional.

As definições da análise de assuntos evidenciam os aspectos práticos da atividade. Assim, ela pode ser definida tanto como a etapa intelectual, por excelência, do trabalho do profissional da informação; quanto como “o processo de ler um documento para extrair conceitos que traduzam a essência de seu conteúdo” (DIAS; NAVES, 2007, p. 5).

Em síntese, a complexidade do processo de análise de assuntos ocorre em virtude da natureza subjetiva dos processos intelectuais, da diversidade de formas que o processo pode ser efetuado, das prioridades baseadas em grupos de usuários específicos e do tipo de serviço que é oferecido.

Na próxima seção, apresentam-se as características da análise de assuntos.

3.1.3 Características

A análise de assuntos intelectual é um processo intelectual, logo, subjetiva e complexa. Assim, a busca da representação do significado do texto na análise de assuntos está circunscrita à linguagem natural, dependendo do contexto cultural do usuário e dos fatores psicológicos e cognitivos do profissional da informação (FARROW, 1991).

Outra característica da análise de assuntos é que seus resultados ocorrem como processos probabilísticos, portanto, não determinísticos, em que se procura antecipar os termos usados pelos usuários em suas expressões de busca por informações (BATES, 1986). Em outras palavras, o termo que representa o conteúdo do documento foi escolhido por suposição de que existe uma probabilidade de algum usuário utilizá-lo nas expressões de buscas, o que está mais próximo de “adivinhação” do que de previsão por regras formais (FOSKETT, 1973, p. 40). Embora essa visão possa ser contestada, uma vez que os recursos tecnológicos atuais permitem analisar os registros de busca dos usuários e entender qual é o vocabulário que utilizam nas buscas.

Selecionar apenas os assuntos relevantes do texto exige do indexador a capacidade de identificar **todos os assuntos** do documento, pois, só assim, ele poderá escolher quais são significativos e quais não são. Não obstante, um mesmo assunto pode ser significativo em um contexto e irrelevante em outro, a depender do potencial interesse dos usuários. Mas jamais se descobrirão todos os significados associados a um documento e que, portanto, não é possível identificar todos os assuntos do texto para selecionar os relevantes (MAI, 2001).

Lancaster (2004, p. 284) caracteriza a análise de assuntos como um serviço da seguinte forma:

Este (serviço de informação) cria representações dos textos, armazena-os numa base de dados e oferece um dispositivo que possibilita que sejam feitas buscas nessas representações. O serviço de informação também cria representações dos pedidos (enunciados de buscas) e os processa em cotejo (estabelece comparação) com a base de dados, para recuperar as representações de textos que coincidam, ou mais se aproximem das representações dos pedidos.

A expectativa é que o profissional da informação leia o documento inteiro, mas raramente ele tem tempo disponível para isso, o que não chega a ser um problema, pois a leitura mais aprofundada não produz resultados melhores, tanto que a NBR 12676/1992 recomenda, para fins de indexação, que a leitura, apenas:

garanta que nenhuma informação foi negligenciada. Para isso deve-se considerar especialmente: a) título e subtítulo; b) resumo; c) sumário; d) introdução; e) ilustrações, diagramas, tabelas e seus títulos explicativos; f) palavras ou grupos de palavras em destaque; e g) referências bibliográficas (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 1992, p. 2).

Em síntese, a análise de assuntos pode ser caracterizada como um processo, em que o profissional da informação lê um texto e dele extrai os assuntos em linguagem natural que serão traduzidos em termos de um vocabulário controlado.

Na próxima seção, descreve-se a análise de assuntos Intelectual.

3.1.3.1 Análise de assuntos intelectual

O processo de análise de assuntos Intelectual é inteiramente realizado por uma pessoa seguindo o fluxo de atividades linear, que pode ser descrito como:

1. **Compreensão do conteúdo do documento**, através da leitura completa do texto ou do título, do resumo e de outras partes que compõem um documento. [...] 2. **identificação dos conceitos**, estabelecendo o ambiente lógico e as diferenças entre os fenômenos, os processos, as propriedades, as operações, os equipamentos [...]. 3. **seleção dos conceitos**, observando alguns fatores como: exaustividade, especificidade e consistência (VIEIRA, 1988, p. 43-44, grifo nosso).

Assim, todo o processo ocorre na mente do indexador: compreensão, identificação e seleção de conceitos; portanto, processos intelectuais, seguidos pela fase de tradução, que pode ter regras bem definidas na política de indexação, mas ainda assim continua sendo um processo intelectual.

Por depender de processos cognitivos naturais, a consistência dos resultados da análise de assuntos Intelectual fica sujeita aos fatores emocionais do indexador. Nessa abordagem, a subjetividade refletindo-se no desempenho, na regularidade e na parcialidade do ser humano em perceber, identificar e selecionar os conceitos. Assim, a subjetividade pode ser um fator determinante na qualidade da indexação (VIEIRA, 1988).

A análise de assuntos Intelectual vem se revelando inadequada, devido aos custos, ao tempo e à subjetividade intrínseca ao fator humano (MACULAN; LIMA, 2017). Os processos linguísticos, cognitivos e racionais do profissional da informação e a mudança contínua do estado do conhecimento do campo ao qual o documento pertence formam uma

conjunção de subjetividades, que se influenciam mutuamente e exigem muito esforço para serem mitigadas (BORKO, 1977; e FERREIRA; MACULAN, 2019).

Outros fatores capazes de influenciar a qualidade da análise de assuntos são: a qualidade do texto (autor); o conhecimento prévio (profissional da informação); e as estratégias exigidas para o texto (usuário). No entanto, durante a análise de assuntos não é possível modificar nenhum deles, de forma que eles são determinantes e não há nada que possa ser feito para reduzir a sua influência (CINTRA, 1987).

Em síntese, revela-se inadequada a análise de assuntos Intelectual por causa da subjetividade inerente ao processo, do volume elevado de documentos que precisam ser analisados e a alta demanda por informações. Uma alternativa é extrair e selecionar os assuntos do texto por meio da tecnologia. Segundo Lancaster (2004), o computador pode ser programado para extrair o conteúdo informacional relevante do texto da mesma forma (mas não exatamente igual) que um indexador faz.

Na próxima seção, discorre-se sobre o uso da tecnologia para assistir a análise de assuntos por computador.

3.1.3.2 Análise de assuntos assistida por computador

Na análise de assuntos assistida por computador, utilizam-se os computadores para assistir o ser humano no processo de indexação. Nessa abordagem, não se procura substituir o ser humano, como se faz nos processos de análise de assuntos automatizados (seção 3.1), mas busca-se fornecer informações que lhe facilitam e melhoram a tomada de decisões do indexador na representação dos assuntos do documento.

Assistir o processo de tomada de decisões na análise de assuntos consiste em identificar os assuntos do texto, selecionar os relevantes e sugeri-los ao indexador. Assim, as decisões tomadas na etapa de tradução são amparadas pela lista de assuntos identificados no texto, pela lista de assuntos mais relevantes do texto e por informações adicionais sobre os termos, as indexações anteriores e as relações dos termos dentro do tesauro que dão ao indexador uma visão mais ampla do contexto (GRAY; HARLEY, 1971).

Os métodos desenvolvidos no campo da inteligência artificial denominado de processamento de linguagem de natural (ver seção 3.3) podem ser empregados para assistir a análise de assuntos, com isso, essas metodologias ajudam a realizar os processos técnicos atuais da BCI quanto à indexação.

A produtividade e a consistência da análise de assuntos assistida por computador aumentam porque o indexador precisa apenas traduzir os assuntos relevantes do texto, selecionados pelos processos computacionais, em termos do vocabulário controlado que ele

julgue mais compatíveis, considerando o contexto social e cultural das necessidades de informações do usuário. Além disso, são oferecidas ao indexador informações adicionais que tornam a escolha dos descritores mais consistentes e confiáveis.

Na próxima seção, descreve-se a aplicabilidade e a importância da análise de assuntos.

3.1.4 Aplicabilidade e importância

A análise de assuntos produz uma indexação melhor, aumentando a eficiência e a eficácia da recuperação da informação, o que é “uma justificativa econômica” da importância da análise de assuntos (GRAY; HARLEY, 1971, p. 173).

A recuperação da informação pode até parecer um ato mágico para o usuário, mas resulta de esforços técnicos individuais e coletivos de profissionais da informação dedicados a elaborar indexações, resumos, catálogos e classificações, que são processos documentários que incluem a análise de assuntos.

A recuperação eficiente da informação traz benefícios que vão além da satisfação das necessidades de informação individuais, ela alcança e impacta positivamente toda a comunidade servida pela biblioteca.

A representação da informação no contexto da BCI, que é feita por meio da análise de assuntos, cria o acesso à informação relevante, proporcionando encontrabilidade aos artigos científicos e contribuindo, assim, com os avanços científicos e tecnológicos.

Na próxima seção, apresenta-se a inteligência artificial.

3.2 Inteligência artificial

O termo “inteligência artificial” aponta uma área de pesquisa e, num sentido amplo, designa o uso de quaisquer técnicas computacionais inteligentes. Seus métodos aparecem na literatura sob outras designações como aprendizado profundo, aprendizado profundo e processamento de linguagem natural. Ao longo do tempo, o termo perdeu popularidade para os termos “*Machine Learning*”, “Aprendizado de Máquina” e “Aprendizado Automático”.

A inteligência artificial é uma das áreas-chave de interesse da Ciência da Informação, pois compartilha os processos cognitivos como fonte e alvo da informação na mente humana (SARACEVIC, 1996).

Na próxima seção, aborda-se sucintamente a história da inteligência artificial.

3.2.1 Breve histórico

O termo “inteligência artificial” foi criado em 1956 na Conferência de Dartmouth para se referir à “aspiração inebriante de realizar em *software* e hardware uma inteligência em nível humano” (JORDAN, 2018, p. 1) por John McCarthy, que à época era um pesquisador iniciante, procurando distinguir suas pesquisas da “cibernética” do pesquisador sênior Norbert Wiener. No entanto, investigavam o mesmo campo científico: o processamento cognitivo por meio dos algoritmos.

Dois avanços importantes para a inteligência artificial, foram o desenvolvimento das redes neurais artificiais (RNA) e a consolidação do algoritmo de retropropagação³. Não se trata de descobrimentos fortuitos; o arcabouço teórico e metodológico da área recebeu contribuições de várias gerações de cientistas dos mais variados campos do saber humano e nacionalidades.

Os sistemas de inteligência artificial para a recuperação da informação, a classificação de textos, o diálogo artificial (*chatbots*) e a recomendação de conteúdos têm sido a base do sucesso comercial das *Big Techs*⁴ (BIRD; KLEIN; LOPER, 2009; LANE; HOWARD; HAPKE, 2019; SEGARAN, 2007). Um sucesso baseado na disponibilidade e no engajamento de usuários sem exigências quanto à qualidade da informação e impulsionado pela capacidade da inteligência artificial de processar, produzir, empacotar e transmitir um volume elevado de informações em alta velocidade (GOLDBERG, 2017; SEGARAN, 2007).

A perspectiva empreendedora capturou a inteligência artificial, especialmente nos Estados Unidos, no Vale do Silício, em que empreendedores de computador “ganham dinheiro com ideias” e que “é uma cultura em que a quantidade de dinheiro a ganhar e a velocidade em que isso ocorrerá são os valores supremos” (CASTELLS, 2003, p. 61).

Na próxima seção, aborda-se as definições da inteligência artificial.

3.2.2 Definição

A inteligência artificial utiliza um conjunto métodos computacionais para simular a inteligência humana. Dito de outra forma, a inteligência artificial desenvolve *softwares* que: 1) pensam e agem como um ser humano (Teoria da mente); ou 2) pensam e agem racionalmente (Teoria da racionalidade) (RUSSELL; NORVIG; DAVIS, 2010).

As primeiras pesquisas da inteligência artificial tinham por finalidade entender (e imitar) os processos cognitivos humanos por meio das simulações computacionais. No

³ *Backpropagation*.

⁴ Empresas de tecnologia transnacionais.

entanto, não se persegue mais essa meta, pois “para voar, a humanidade abandonou o sonho de fazer uma máquina capaz de se passar por um pombo, assim também, para alcançar a inteligência artificial deve-se abandonar a ideia de produzir um artefato capaz de se passar por uma inteligência humana” (DOMINGOS, 2018, p. 62).

Atualmente, o termo inteligência artificial aponta de modo genérico tanto dispositivos, quanto técnicas. Como uma heurística geral, todo artefato qualificado como *smart*, como em *smartphones*, embute alguma técnica de IA, de forma que, ela pode ser encontrada em quase todas as atividades humanas (EKBIA, 2008, 2010), desde aplicativos em educação, medicina e mercado financeiro até em dispositivos físicos como veículos autônomos (robôs, carros e aviões), videogames e sensores (GOLDBERG, 2017).

Todos os processos que envolvem seleção, classificação, tomada de decisão, cognição, reconhecimento de padrões e automação são elegíveis para o uso da inteligência artificial, sobretudo as atividades que utilizam processamento de linguagem natural (PLN) como os navegadores da Internet, os aplicativos móveis, as redes sociais, os processadores de texto, as campanhas publicitárias e os assistentes pessoais inteligentes (Siri e Alexa, por exemplo).

As técnicas da inteligência artificial foram classificadas pela AAAI⁵ em: 1) Pesquisa; 2) *Machine Learning*; 3) *Data Mining e big data*; 4) Planejamento Automatizado; 5) Representação de Conhecimento; 6) Raciocínio (Probabilístico ou não); 7) processamento de linguagem natural; 8) Robótica; 9) Sistema de Agentes e Multiagentes; e 10) Aplicações (ALVES, 2021, p. 1). Essa classificação não é definitiva, tendo propósito didático, uma vez que um agente inteligente pode agrupar em uma mesma aplicação mais de uma técnica.

3.2.3 Características

A perspectiva científica da inteligência artificial propõe desenvolver os fundamentos teóricos para a perspectiva técnica. Esta última, como uma abordagem de engenharia, procura fornecer apoio e substância para a abordagem científica. Esse aspecto dual resulta na prática discursiva peculiar (EKBIA, 2008) apresentada na seção seguinte.

3.2.3.1 Prática discursiva

A prática discursiva da inteligência artificial utiliza como nomenclatura a linguagem cotidiana para descrever o que os algoritmos fazem. Por exemplo, usa o termo “alimentar”

⁵ Association for the Advancement of Artificial Intelligence

para indicar o fornecimento de dados ao algoritmo e o termo “neurônio” para apontar um tipo específico de algoritmo. Quanto a isso, pode-se afirmar que:

Essas práticas puxam o campo em direções diferentes, gerando um corpo de modelos, metáforas e técnicas que, a julgar pela história da IA, continuam se movendo em ciclos de modas passageiras. Isso acontece porque a prática discursiva simplesmente não pode cumprir o papel esperado: não faz a ponte entre a engenharia e as práticas científicas, ao contrário, amplia a lacuna entre elas, introduzindo, cada vez mais, camadas de linguagem antropomórfica e imagens metafóricas [...] (EKBIA, 2008, p. 5, tradução nossa).

Essa prática discursiva acaba sendo um elemento importante dentro da comunidade de pesquisa da IA, porque ajuda na divulgação e na angariação de fundos. Todavia, resulta no fenômeno *Hype* (abreviação palavra *hyperbole*, em inglês), em que ocorre uma divulgação espontânea e viral por meio de sites, blogs e redes sociais de funcionalidades e potenciais com pouca ou nenhuma base na realidade, funcionando como uma poderosa campanha de *marketing* efetuada por amadores.

3.2.3.2 Rede Neural Artificial

Uma Rede Neural Artificial (RNA) é um modelo computacional genérico, instanciado em um algoritmo. A configuração dos parâmetros da RNA é o que a transforma em um modelo útil de IA. Assim, uma atividade essencial da inteligência artificial é obter os parâmetros adequados para a RNA, atividade denominada de “treinamento”. Esse processo computacional adapta os parâmetros livres da RNA ao ponto de torná-la modelo capaz de classificar corretamente dados sem a interferência humana. O processo começa com valores aleatórios para os parâmetros, que, por meio de heurísticas sofisticadas são modificados, a cada insucesso, até obter a solução aceitável (HAYKIN, 1999).

A inteligência artificial resolve problemas de ordem prática, o que não implica, necessariamente, em autonomia, pois o treinamento da RNA requer dados classificados por humanos como exemplo. Essa dependência de dados rotulados torna a aplicação da inteligência artificial dispendiosa, demorada e, às vezes, impossível (BIRD; KLEIN; LOPER, 2009).

A “Aprendizagem Artificial” consiste em utilizar uma heurística para encontrar um conjunto de parâmetros suficientemente bons para tornar a RNA um modelo útil de classificação, em que se procura a melhor resposta e não a resposta ótima. Como todas as heurísticas, esses algoritmos obtêm, às vezes, respostas não ótimas ou de baixa confiabilidade; ainda assim, são as ferramentas humanas mais lucrativas (SEGARAN, 2007).

A análise de textos computacional foi determinante para o sucesso mercadológico dos motores de busca, das redes sociais contemporâneas, dos agentes de diálogos (*chatbots*) entre humanos e máquinas em linguagem natural, e dos sistemas de recomendações de compras e de conteúdos (LANE; HOWARD; HAPKE, 2019). Para Jordan (2018, p. 4), apesar do sucesso, ainda faltam “significado e raciocínio nos sistemas de processamento de linguagem natural”, pois o pensamento de alto nível computacional ainda permanece ilusório e as soluções, cuja configuração supera a capacidade humana de compreendê-la, ao não apresentarem uma explicação, acabam sendo outro problema em áreas como a Medicina, em que a explicação importa mais do que o resultado (FIORINI *et al.*, 2016).

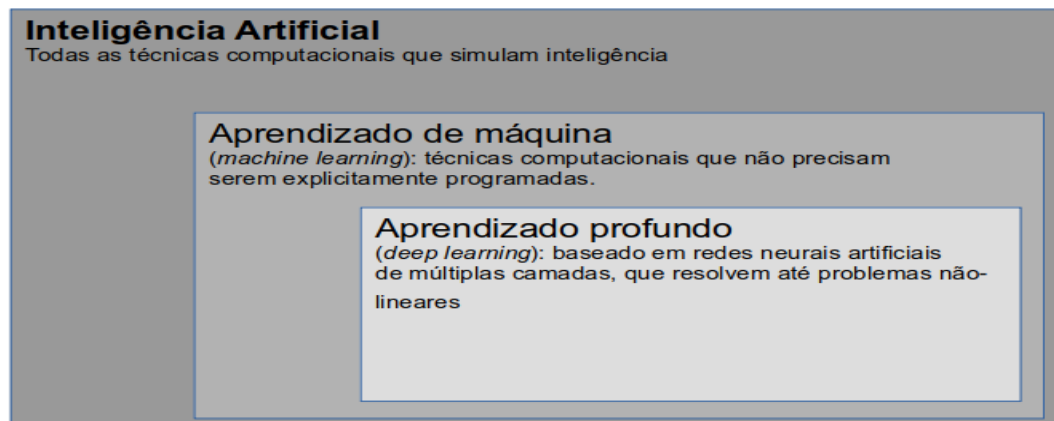
Usam-se dados de entrada rotulados para “ensinar” e “supervisionar” o aprendizado da RNA, técnica denominada Aprendizado Artificial Supervisionado. Por exemplo, no experimento desta pesquisa, será fornecido um conjunto de artigos científicos (dados de entrada) e os respectivos termos de indexação (saída esperada) para servir de “professor” e “supervisor” da Aprendizagem Artificial. Tendo “aprendido”, o modelo de inteligência artificial pode indexar artigos não indexados.

Os dados rotulados submetidos ao Aprendizado Supervisionado são agrupados em três conjuntos: 1) conjunto de treinamento com mais de 75% dos dados para “treinar” o algoritmo; 2) conjunto de teste com 15% dos registros para testar os parâmetros do modelo; e 3) conjunto de validação para avaliar a acurácia do modelo obtido. Essas proporções não são fixas, sendo recomendável reservar mais da metade dos dados para o conjunto de treinamento. O conjunto de testes é utilizado, após o treinamento, para verificar a capacidade de generalização do modelo obtido (SEGARAN, 2007).

Por se tratar de uma heurística, o “treinamento”, geralmente, consiste em pegar um dado, estimar a resposta e compará-la com o valor objetivo (nível de referência). A diferença entre a estimativa e o valor objetivo, se houver, constitui o erro, que, por sua vez, será usado para estabelecer o nível de acurácia. Se o erro for acima do aceitável, o processo é repetido. Esses passos são repetidos até se obter um resultado, cujo erro seja admissível. Se a busca pela resposta certa não convergir dentro de um número de tentativas (épocas), a RNA “desiste” (HAYKIN, 1999).

O Aprendizado de Máquina (ML, do inglês, *Machine Learning*) é um subcampo da inteligência artificial (Figura 3), que, segundo Domingos (2018, p. 27), “foi tão bem-sucedido que ofuscou sua orgulhosa mãe”. A característica distintiva do ML é sua capacidade de si programar. De certa forma, esses algoritmos são capazes de obter dos dados as configurações necessárias para realizar suas próprias tarefas (aprendizagem), sem precisar da intervenção direta de seres humanos.

Figura 3 – A relação da inteligência artificial e o aprendizado profundo.



Fonte: Elaborado pelo autor.

O ML está dividido em cinco paradigmas principais: 1) simbolista, 2) conexionista, 3) evolucionista, 4) analogista e 5) bayesiano. Eles estão elencados no Quadro 5 com os seus respectivos algoritmos, tipo de aprendizado e o campo científico em que se inspiraram.

Quadro 1 - Paradigmas principais do aprendizado de máquinas

Paradigma	Algoritmo	Aprendizado (tipo)	Ciência inspiradora
Simbolista	Dedução inversa	Dedução inversa	Filosofia, Psicologia e Lógica
Conexionista	Retropropagação	Baseado em conexões	Neurociência e Física
Evolucionista	Programação genética	Evolução biológica	Genética e Biologia Evolutiva
Bayesiano	Inferência bayesiana	Teorema de Bayes	Estatística
Analogista	Máquina de vetores de suporte	Extrapolações a partir das semelhanças	Psicologia e Matemática

Fonte: Elaborado pelo autor a partir de (DOMINGOS, 2018, p. 15).

Atualmente, o paradigma conexionista é o mais popular, devido a sua adaptabilidade, flexibilidade e baixa necessidade de fundamentação teórica para explicar os resultados. Além disso, facilita o processamento de amplas bases de dados por sua *práxis* experimentalista baseada em heurísticas. No paradigma conexionista, a propriedade mais importante é a convergência para a solução em um tempo aceitável (RUSSELL; NORVIG; DAVIS, 2010).

3.2.4 Relação da inteligência artificial com a Ciência da Informação

A inteligência artificial interage com pessoas, tecnologias, ideias, ideais e atividades (EKBI, 2008), na medida em que a representação do conhecimento permite aos algoritmos compreender a linguagem natural. Essa compreensão ocorre em sentido duplo entre as pessoas e as tecnologias inteligentes, em que os resultados dessa interação imitam e influenciam os processos intelectuais.

Como ferramenta cognitiva, a máquina memoriza, processa e resolve problemas a uma velocidade acima do que o intelecto humano pode acompanhar (KOBASHI, 1996; SMIT, 1987). No entanto, toda a produção da máquina se destina a expandir e satisfazer as necessidades humanas.

A análise de assuntos pode beneficiar-se da inteligência artificial principalmente por meio dos métodos do processamento de linguagem natural, conforme pode ser visto na próxima seção.

3.3 Processamento de linguagem natural

O processamento de linguagem natural é um campo de pesquisa da inteligência artificial, cuja finalidade é desenvolver técnicas e métodos para extrair informações dos textos escritos em linguagem natural (KHURANA *et al.*, 2022).

O campo de pesquisa do processamento de linguagem natural investiga os algoritmos que permitem a tecnologia processar a linguagem humana com a finalidade de realizar tarefas úteis. Suas abordagens são interdisciplinares, pois utiliza técnicas da linguística, da ciência da computação, da ciência cognitiva e da inteligência artificial.

As duas grandes vertentes do processamento de linguagem natural são: (1) processamento dos textos escritos, e (2) o processamento da fala. Nesta pesquisa, será considerado apenas a primeira vertente, principalmente, no que tange as aplicações computacionais práticas de interações entre a linguagem humana escrita e os computadores.

Na próxima seção, apresenta-se um breve histórico do processamento de linguagem natural como um campo interdisciplinar.

3.3.1 Breve Histórico

Os métodos "modernos" do uso da linguagem natural na recuperação da informação remontam ao sistema descrito por Taube, em 1951, no âmbito do experimento UNITERM, antes do computador pessoal, quando os dados ainda eram escritos à mão ou datilografados em fichas catalográficas (LANCASTER, 2004, p. 250).

A história do processamento de linguagem natural pode ser dividida, segundo Deng e Liu (2018), em três ondas: (1) Racionalista, (2) Empírica, e (3) de aprendizado profundo.

A chamada primeira onda, começou efetivamente em 1954, sob a influência da Guerra Fria, num experimento de tradução automática capaz de traduzir frases do russo para o inglês. Essa abordagem racionalista postulava que os conceitos-chave da linguagem

humana eram programados geneticamente no cérebro das pessoas. As pesquisas dessa onda racionalista produziram algoritmos baseados em regras codificadas explicitamente para incorporar, nos sistemas de inteligência artificial do processamento de linguagem natural, os mecanismos de conhecimento e raciocínio. Esses sistemas são conhecidos Sistemas Especialistas.

O sistema ELIZA representa bem essa onda racionalista, pois estruturava as “informações do mundo real em ontologias conceituais, baseadas em conjuntos complexos de regras manuscritas” (DENG; LIU, 2018, p. 3) para imitar o diálogo entre um ser humano e uma máquina.

A segunda onda do processamento de linguagem natural, a partir dos anos de 1990, baseava-se no chamado aprendizado de máquina, que são técnicas computacionais que aprendem a partir dos padrões latentes nos dados sem propor nenhum modelo que descreva seus achados. Nessa abordagem, utilizam-se os modelos estatísticos e redes neurais artificiais simples para aprender automaticamente os parâmetros que generalizam a resposta, no entanto, esse expediente precisa descartar informações. Caracterizam essa onda os algoritmos de redes Bayesianas, Máquinas de Vetor de Suporte e as Árvores de Decisão. No geral, os sistemas do processamento de linguagem natural da onda empírica têm um desempenho melhor do que os sistemas baseados em conhecimento da onda racionalista anterior.

Na terceira onda, os recursos do processamento de linguagem natural deixam de ser classificados previamente por seres humanos, que era o gargalo das soluções da segunda onda, em que a engenharia do conhecimento criava recursos (dados estruturados e rotulados) manualmente. A partir de 2010, o poder computacional e as grandes bases de dados combinados com aprendizado profundo (ver seção 3.2.3.4) eliminaram as limitações impostas à onda empírica. Os avanços do processamento de linguagem natural atual, de forma mais geral, da inteligência artificial, se devem as “abordagens de aprendizado profundo [que] exploram as redes neurais com várias camadas ocultas para resolver tarefas gerais de aprendizado de máquina dispensando a engenharia de recursos”, a eficácia do aprendizado profundo é menor quando o processamento de linguagem natural se baseia em texto (DENG; LIU, 2018, p. 5).

Em síntese, a abordagem atual do processamento de linguagem natural está fortemente baseada na inteligência artificial do paradigma conexionista do aprendizado profundo. Os métodos orientados aos dados e baseadas no aprendizado de máquina, sem as regras racionalistas da primeira onda, afirmam Deng e Liu (2018, p. 17), fazem o futuro do processamento de linguagem natural “muito mais brilhante do que em qualquer outro momento do passado”.

Na próxima seção, apresentam-se as definições do processamento de linguagem natural.

3.3.2 Definições

O processamento de linguagem natural é um campo de pesquisa científico voltado para o desenvolvimento tecnológico e as suas aplicações estão estreitamente ligadas ao mercado global de informações, sendo, hoje, uma das atividades mais lucrativas quando associada às redes de dados de alta velocidade.

Como campo técnico-científico, o processamento de linguagem natural concentra-se em desenvolver tecnologias para automatizar a análise da estrutura linguística dos textos e pode ser definido como:

um tratado de inteligência artificial e Linguística, dedicado a fazer os computadores entenderem as declarações ou palavras escritas em línguas humanas. Ele surgiu para facilitar o trabalho do usuário e para satisfazer o desejo de se comunicar com o computador em linguagem natural, e pode ser classificado em duas partes, ou seja, Compreensão de Linguagem Natural ou Linguística e Geração de Linguagem Natural que envolve a tarefa de compreender e gerar o texto (BALYAN; MCCARTHY; MCNAMARA, 2020, p. 342, tradução nossa).

Outros autores definem o processamento de linguagem natural como linguística computacional, pois ele usa as técnicas computacionais para aprender, entender e produzir conteúdo de linguagem humana. Com essa definição, o processamento de linguagem natural torna-se um sistema linguístico computacional para auxiliar a comunicação entre humanos e entre a máquina e as pessoas ao analisar e aprender “com a enorme quantidade de conteúdo de linguagem humana que agora está disponível *on-line*” (HIRSCHBERG; MANNING, 2015, p. 261).

Na próxima seção, estão as características do processamento de linguagem natural.

3.3.3 Características

A principal característica do processamento de linguagem natural é processar dados de linguagens naturais, notadamente os textos.

Os textos são formados por uma sequência de palavras em linguagem natural e os algoritmos processam com eficiência os números. Para resolver isso, no processamento de linguagem natural, desenvolveu-se a técnica do “*Word Embeddings*” (BENGIO *et al.*, 2003), que troca os termos da linguagem natural de um texto por números. Assim, os

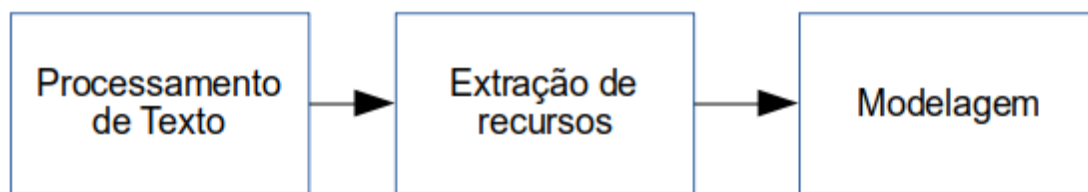
algoritmos processam os números e não as palavras em si, e, com isso, tornam os textos processáveis por computadores.

A menor unidade de informação de uma tarefa de processamento de linguagem natural é chamada de *token* e pode ser um morfema, ou uma palavra inteira, ou sentença, ou até um texto, dependendo da tarefa. Por exemplo, cada descritor de um vocabulário controlado é um *token*, de forma que pode ser formado por um termo único (por exemplo, “arroz”) ou por vários (por exemplo, “Inseto para Controle Biológico”).

O fluxo de atividades de uma tarefa de processamento de linguagem natural é denominado de *pipeline* (Figura 4), que, normalmente, é dividida em três etapas principais:

1. processamento do texto;
2. extração de recursos;
3. modelagem dos dados.

Figura 4 – Uma *pipeline* típica do processamento de linguagem natural.



Fonte: Elaborado pelo autor.

Para um leitor humano, o texto é formado por uma hierarquia de unidades semânticas estruturadas em caracteres, palavras, frases e parágrafos. Para o algoritmo, no entanto, o texto é formado por uma sequência de caracteres. Essa diferença explica, em parte, porque a inteligência humana é capaz de entender o significado do texto, enquanto o processamento de linguagem natural se limita a processar símbolos relacionados entre si por coocorrência, frequência e proximidade.

A primeira etapa da *pipeline* extrai do arquivo eletrônico o fluxo de caracteres sem formatação (*tags*) e sem figuras; por exemplo: de um documento se extrai apenas o texto, que é o dado textual bruto, a matéria-prima das etapas seguintes, e ignora as ilustrações e as formatações (tipo, cor e tamanho da fonte).

Na segunda etapa, as palavras do dado textual bruto, vindas da etapa anterior, são trocadas por representações numéricas adequadas à aplicação do modelo de inteligência artificial, as *Words Embedding*, que se tornam as “variáveis” explicativas (HINTON, 2020). Ou seja, cada termo do texto é armazenado numa estrutura de dados denominada de dicionário; e, nas etapas posteriores, o que se processa são índices numéricos das palavras e não a palavra em si. Os resultados produzidos também serão numéricos, mas eles são

reconvertidos em palavras no processo inverso de consulta ao dicionário, em que entra um número de índice e sai a palavra correspondente.

A estruturação dos recursos computacionais é específica para cada modelo inteligente. Por exemplo, a oração “as pesquisas do grupo de pesquisa” pode ser representada por:

a representação 1:

- [“0, 1, 0”], em um modelo Y, que remova as palavras vazias de sentido e que reduza os termos da frase a sua forma singular, o que tornaria a oração original em “pesquisa grupo pesquisa”; ou

b representação 2:

- “0, 1, 2, 3, 4, 5”, em um modelo X, que não faça pré-processamento do texto.

Essas representações podem ser processadas em dois tipos de estruturas básicas, que permitem prever uma palavra p a partir das palavras circundantes ou prever as palavras circundantes a partir da palavra p . Essas duas estruturas de dados são:

a Saco de palavras (BOW, do inglês, *Bag Of Words*):

- não guarda a ordem dos termos nem preserva a extensão da frase.

b Gramática- n (*n-gram*):

- preserva a ordem das palavras e a extensão da frase.

Nesta etapa, aplicam-se os modelos de inteligência artificial aos recursos. Na próxima seção, apresentam-se as aplicações e a importância do processamento de linguagem natural.

3.3.4 Aplicações e importância

Como uma linguística computacional, o processamento de linguagem natural está incorporado em todas as tecnologias que fazem o uso da linguagem como os assistentes pessoais, os tradutores automáticos e as redes sociais. Toda essa riqueza de recursos tecnológicos faz uso da análise dos assuntos de um texto para dar acesso à informação que satisfaz as necessidades do usuário. Essa descrição demonstra a afinidade do processamento de linguagem natural com a Biblioteconomia e a Ciência da Informação e que esses desenvolvimentos podem e devem ser convertidos em meios tecnológicos para auxiliar a BCI a alcançar seu objetivo primordial de criar acesso à informação relevante.

Nesse sentido, Silva e Correa (2020) afirmam que a BCI pode se beneficiar dos métodos do processamento de linguagem natural em todas as suas atividades, incluindo na análise de assuntos, na extração de informações, na elaboração de resumos, na classificação e no enriquecimento de tesouros.

O processamento de linguagem natural, portanto, é a revolução da inteligência artificial com capacidade de oferecer oportunidades tecnológicas para os que os profissionais da informação tenham ganho de produtividade, de qualidade e de bem-estar. Mas essa é uma questão de interesse multidisciplinar que exige mais pesquisas e mais investimentos.

Na próxima seção, descreve-se a sumarização automática que é um dos métodos do processamento de linguagem natural.

3.3.4.1 Sumarização Automática

Nesta seção, descreve-se a sumarização automática como uma técnica do processamento de linguagem natural, portanto, pertencente à inteligência artificial, que tem por finalidade produzir resumos por meio das máquinas. Lembrando que a sumarização de textos é uma atividade-chave no contexto do acesso à informação (SAGGION, 2008).

Nesta seção, o termo “resumo” se refere aos resumos obtidos por sumarização automática com a finalidade de auxiliar o processo de análise de assuntos, sendo apenas um insumo e não um produto.

Os métodos de inteligência artificial facilitam a geração de resumos de textos, mas a questão crucial, nesse caso, é determinar objetivamente se o resumo produzido tem a qualidade necessária. Quanto a isso, seria necessário responder a seguinte questão: quais critérios objetivos avaliam o conteúdo e a forma do resumo? Porém, não há uma resposta satisfatória para essa questão, porque a sumarização feita por computadores ainda é uma disciplina recente (TORRES-MORENO, 2014).

O primeiro visionário do campo da sumarização automática de texto, segundo Torres-Moreno (2014, p. 3), foi Joseph Joubert, que, em 1780, tinha a “ambição de resumir textos e condensar frases”. Mas foi Peter Luhn quem efetivamente deu início a Sumarização Automática, em Luhn (1957), em que empregou a frequência dos termos para selecionar as frases que deveriam constar no resumo.

Edmundson (1969) fez uma proposta em cima da de Luhn, chamado de seleção automática de sentenças com maior potencial de transmitir a essência do documento.

Na década de 1970, Antonio Zamora e Ricardo Salvador desenvolveram um programa de computador que excluía do texto os clichês, as expressões repetitivas, as tabelas

e as frases triviais, duplicadas e óbvias, obtendo um resumo que continha entre 60 e 70% das mesmas frases escolhidas por resumidores profissionais (TORRES-MORENO, 2014).

Na década de 1970, o processo de pesquisa estagnou, conforme o seguinte relato:

No período entre 1978 e 1993, nenhuma pesquisa importante sobre a sumarização automática foi publicada, porque os periódicos passaram a oferecer resumos em suas próprias publicações e, portanto, não havia mais demandas para a sumarização automática. (MITKOV; HOVY, 2009, p. 554).

A situação mudou, na década de 1990, com a popularização dos computadores. Segundo Mitkov e Hovy (2009, p. 554), o “progresso no processamento de linguagem natural, aliado a grandes aumentos de memória e velocidade do computador”, e a disponibilização dos textos pela Internet renovaram o interesse pela sumarização automática.

Atualmente, a sumarização automática é uma disciplina do processamento de linguagem natural, em que não basta construir um sistema eficiente para produzir resumos, necessita-se também saber avaliar a qualidade dos resumos gerados (GOLUB *et al.*, 2021).

Um “resumo” é uma versão condensada das informações mais relevantes de um documento que ajudam os leitores a avaliar um documento sem lê-lo (SAGGION, 2008). Em Lancaster (2004, p. 100), resumo é definido como uma representação sucinta, “porém exata, do conteúdo de um documento”

A “sumarização” é definida como o processo de elaborar resumos. A sumarização exige esforços para selecionar, reformular e criar um texto coerente contendo os segmentos mais informativos de um documento. Segundo Lancaster (2004, p. 101), o termo “sumarização” designa todos os processos de produção de representações condensadas de textos e, portanto, é aplicável tanto à redação de resumos quanto a de extração.

A “sumarização automática” de texto é definida como “uma tecnologia para produzir resumos”, sendo uma disciplina do processamento de linguagem natural (NLP) para condensar documentos de texto (TORRES-MORENO, 2014, p. 3).

A sumarização automática produz o “resumo de um texto gerado por um *software*, que é coerente e contém uma quantidade significativa de informações relevantes do texto fonte”, cujo tamanho “é menor que um terço do comprimento do documento original” (MITKOV; HOVY, 2009, 554)

Os resumos produzidos por sumarização automática servem como insumos para os processos de análise de assuntos. Essa distinção é importante porque as características como informatividade, relevância e coerência não precisam ser avaliadas com muito rigor em resumos destinados para o uso interno das atividades de indexação.

A sumarização automática produz resumos diferentes dos resumos produzidos por pessoas e não há meios práticos para compará-los. A sumarização automática não leva

em conta o contexto do usuário, por isso seus resumos podem conter incoerências, anáforas sem sentido, lacunas na estrutura retórica e erros ortográficos e gramaticais. Existem métodos para avaliar resumos, mas a utilidade deles é limitada, uma vez que medições de legibilidade com base em comprimento de palavras e verificadores de gramática são avaliações grosseiras.

Avaliar o conteúdo informativo de um resumo implica quantificar o quanto a informação original foi preservada. Logo, é uma tarefa difícil, porque teria que se identificar todas as informações originais, contá-las e depois verificar as informações do texto-fonte remanescentes no resumo. No entanto, ainda não é possível identificar todas as informações presentes em um documento (MAI, 2001).

Existem dois tipos de métodos para avaliar os resumos: 1) intrínseco, que analisa as características internas do resumo, como a coerência gramatical e a legibilidade; e 2) extrínseco, que analisa o resumo quanto ao seu potencial de ajudar o desempenho das tarefas.

Supõe-se que, para um resumo ser útil na indexação, ele deve reduzir a carga de trabalho (o volume e o tempo de leitura) do indexador. Mas é difícil, talvez impossível, medir o tempo que uma pessoa leva para efetuar o processo de análise de assuntos, porque isso depende de fatores ambientais, como a temperatura, os equipamentos de visualização e a qualidade do texto, e de fatores psicológicos e cognitivos, como a capacidade individual de compreender o texto, o nível de estresse e a boa vontade. No entanto a carga de trabalho pode ser estimada transversalmente, medindo-se a taxa de redução do volume de dados a ser analisado. Além disso, pode-se considerar também que:

1. o resumo fornecido pelo autor do artigo científico não facilita a análise de assuntos, porque, na maior parte das vezes, o autor não foi treinado em sumarizar, de forma que seu resumo não é útil para a indexação (BORKO; BERNIER, 1975);
2. não há consenso sobre quais frases são relevantes em um texto, logo não é possível avaliar por comparação resumos de máquinas e de humanos;
3. avaliação da qualidade de resumo é complicada, porque “exatidão, ambiguidade e brevidade são critérios subjetivos” (LANCASTER, 2004, p. 123).

Na inteligência aumentada, que sempre conta com a participação humana, resumos pouco informativos podem ser utilizados, porque a sumarização automática, nesta pesquisa, produz resumos apenas como um insumo do processo de indexação. Além disso, um indexador pode obter informações até mesmo de resumos pouco informativos.

Na próxima seção, apresentam-se as expressões regulares.

3.3.4.2 Expressões regulares

As expressões regulares são métodos indispensáveis no tratamento de texto para a aplicação do processamento de linguagem natural. Trata-se de uma atividade complexa, porque envolve milhares de padrões, o que pode exigir muito tempo se não for feito com eficiência.

A análise de assuntos precisa identificar os assuntos do texto. No texto, diferentemente do vocabulário controlado, um termo pode ser expresso por suas variantes, devido à polissemia e às variações de número (singular e plural) e de gênero (feminino e masculino).

As expressões regulares são métodos genéricos bem estabelecidos que realizam a tarefa de identificar padrões linguísticos de forma otimizada em relação ao esforço computacional e ao tempo necessário para alcançar um resultado.

As expressões regulares surgiram na década de 1950 como uma forma de descrever linguagens regulares e só foram incorporadas pela Informática na década de 1970.

Segundo Kozen (1991), as expressões regulares foram publicadas em 1951 pelo matemático Stephen Cole Kleene e que, a partir de 1968, elas se tornaram muito conhecidas ao serem adicionadas aos analisadores léxicos de compiladores, que foi uma etapa crucial no desenvolvimento da informática.

O essencial da teoria da linguagem formal de Kleene é a demonstração de que linguagens finitas são precisamente o conjunto de linguagens que podem ser descritas por uma expressão regular.

Desde os anos de 1970, os sistemas operacionais e os editores de texto incluem as expressões regulares como ferramentas básicas e nos anos de 1980, os Sistemas de Gerenciamento de Bancos de Dados também adicionaram as expressões regulares às suas bibliotecas (*Library*) padrões, sendo possivelmente um dos métodos mais utilizados na Informática.

Atualmente, além dos sistemas operacionais, bancos de dados e editores de textos, todas as linguagens de programação dispõem de mecanismos para adoção e uso de expressões regulares.

Uma Expressão Regular, segundo Mitkov e Hovy (2009), pode ser definida como uma sequência de caracteres que especifica um padrão de pesquisa no texto.

expressões regulares são “conjuntos de caracteres especialmente codificados para serem usados como padrões para encontrar conjuntos de caracteres em um texto” (FITZGERALD, 2012) e é o recurso computacional essencial dos sistemas de buscas diretas no texto.

As expressões regulares são metodologias inteligentes da Ciência da Computação para encontrar conceitos nos textos expressos por vários termos relacionados por alguma regra latente, especialmente plural, singular e conjugação verbal. Elas são uma resposta técnica a um problema específico, portanto não há variações quanto as suas definições.

As expressões regulares são padrões elaborados especificamente para recuperar, localizar, ou substituir expressões em textos e outros tipos de dados. Uma expressão regular usa uma linguagem formal e não é um código em linguagem de programação.

Por exemplo, a expressão regular “U+0000” é utilizada para encontrar todos o sinal de pontuação “ponto final” em um texto.

Uma característica das expressões regulares é permitir o mapeamento múltiplo de padrões. Observe no Quadro 2, que uma expressão regular composta por letras, parênteses e uma barra vertical, selecionam os caracteres indicados como estão, e usa o par de parênteses para indicar as variações e uma barra vertical para indicar as alternativas.

Quadro 2 – O mapeamento múltiplo como característica de expressões regulares.

Expressão regular	Mapeamentos
Produc(ão ões) de sement(e es)	Produção de semente, Produção de sementes, Produções de semente, Produções de sementes.

Fonte: Elaborado pelo autor.

As expressões regulares são elaboradas a partir de um conjunto de convenções estabelecidas no contexto de uso. Assim, em linguagens de programação diferentes podem apresentar convenções diferentes para um mesmo padrão.

O que melhor caracterizam as expressões regulares é a sua eficiência, sendo insubstituível para localizar e substituir padrões de texto nas tecnologias de informação.

As expressões regulares como recurso das linguagens de programação de computadores permitem encontrar nos artigos científicos os padrões linguísticos (palavras ou expressões) definidos nos vocabulários controlados. O casamento de padrões, nesse caso, realiza operações, cujos resultados equivalem a identificar os assuntos do texto.

As expressões regulares associadas aos tesouros são fundamentais nas aplicações computacionais com o processamento de linguagem de natural, porque essa associação facilita e viabiliza tecnicamente a aquisição automática de termos e de seus relacionamentos. Além disso, as expressões regulares escritas em regras formais permitem o uso de interpretadores que tornam o processo de busca dos termos altamente eficiente.

As expressões regulares, nesta pesquisa, viabilizaram a implantação da etapa de identificação dos assuntos do texto da análise de assuntos.

Na próxima seção, apresenta-se a modelagem de tópicos que é um método do processamento de linguagem natural.

3.3.4.3 Modelagem de tópicos

Um dos métodos do processamento de linguagem de natural que tem afinidade com a análise de assuntos é a modelagem de tópicos. Esse método foi desenvolvido para automatizar a recuperação da informação automatizada, ou seja, é um método que procura recuperar a informação sem a mediação humana.

Nesta pesquisa, adaptou-se a modelagem de tópicos para produzir resultados semelhantes aos da etapa de seleção de assuntos relevantes do texto da análise de assuntos, o que constitui mais inovação nesta pesquisa.

O primeiro método de modelagem de tópicos, proposto em 1990, recebeu o nome de indexação Semântica Latente (LSI). Ele tratava as questões linguísticas dedutíveis do contexto da palavra no texto, como a sinonímia e a polissemia (DEERWESTER *et al.*, 1990). Considerava-se na LSI que a ordem das palavras no documento não era importante para o significado do texto, porque, nesse método, seguia-se a teoria da Permutabilidade dos Termos (ALDOUS, 1985).

Em 1999, foram propostos os algoritmos da família da indexação semântica latente probabilística (pSLI), que aplicavam técnicas da estatística para realizar a modelagem de tópicos. Esse método era de difícil aplicação por ser susceptível a sobreajustes e a sua complexidade aumentava rapidamente com o crescimento do conjunto de dados analisados (HOFMANN, 1999).

Por fim, a técnica alcançou a maturidade tecnológica no modelo de alocação latente de *Dirichlet* (LDA, do inglês, *latent Dirichlet allocation*), com a publicação de Blei, Ng e Jordan (2003). O modelo LDA é atualmente o método padrão de modelagem de tópicos do processamento de linguagem natural.

A distribuição de *Dirichlet* (proposta por Peter Gustav Lejeune Dirichlet) descreve fenômenos estatísticos multivariados, contínuos e parametrizados, sendo a conjugada da distribuição categórica e da distribuição multinomial (KOTZ *et al.*, 2000). Nesse tipo de distribuição de probabilidades, as respostas ficam restritas a um grupo de resultados possíveis (categorias), em que a incorporação de um novo parâmetro e de novos dados não acrescenta mais complexidade ao modelo.

No modelo de alocação latente de *Dirichlet*, considera-se que um tópico seja formado por um ou mais termos associados por relações estatísticas. Assim, o termo “tópico” refere-se a um aglomerado de assuntos relacionados estatisticamente entre si, considerando-se os textos que pertencem a um mesmo *cluster*.

No Quadro 3, pode ser visto o exemplo de um “tópico”, em que o número à esquerda do termo (fator) indica a importância do assunto para o tópico, em que quanto maior o fator, maior a probabilidade do texto que contém o termo pertencer ao *cluster*.

Quadro 3 – Exemplo de um tópico contendo o termo “germoplasma”.

Tópico
0.166 * "germoplasma" + 0.147 * "genética" + 0.052 * "banco de germoplasma" + 0.035 * "983944" + 0.027 * "865355" + 0.027 * "fungo" + 0.027 * "esporo" + 0.027 * "aveia" + 0.027 * "abobora" + 0.027 * " <i>anacardium occidentale</i> "

Fonte: Elaborado pelo autor.

A Alocação Latente de *Dirichlet* (LDA) é um modelo probabilístico generativo de um *corpus*, definido nos seguintes termos:

A Alocação Latente de Dirichlet (LDA) é um modelo probabilístico generativo de um corpus. A ideia básica é que os documentos sejam formados por uma mistura aleatória de tópicos latentes, em que cada tópico pode ser caracterizado por uma distribuição de probabilidades de palavras (BLEI; NG; e JORDAN, 2003, p. 996, tradução nossa).

Supõe-se, no modelo Alocação Latente de *Dirichlet*, que existe uma distribuição de probabilidade que descreve matematicamente as relações dos assuntos dentro dos tópicos, de forma que os documentos parecidos entre si compartilham, em diferentes graus de semelhança, os tópicos. Assim, Blei, Carin e Dunson (2010, p. 55, tradução nossa) definiram o LDA da seguinte forma:

[O LDA é] um modelo probabilístico hierárquico usado para decompor uma coleção de documentos em seus tópicos salientes [os tópicos que se destacam de outros tópicos], em que um “tópico” é uma distribuição de probabilidade sobre um vocabulário.

Em síntese, o LDA é um modelo probabilístico generativo de um *corpus*, em que se usa um modelo hierárquico para decompor os documentos em tópicos. Uma vez que os tópicos de um documento são conhecidos, por hipótese, é possível encontrar e agrupar todos os documentos no *corpus* que compartilhem esses tópicos.

O método padrão de modelagem de tópicos do processamento de linguagem de natural é o modelo de Alocação Latente de *Dirichlet*, que usa a inteligência artificial para descobrir: (1) a probabilidade de um assunto pertencer a um texto; e (2) a probabilidade de um termo pertencer a um tópico.

No nível mais geral, a modelagem de tópicos analisa por métodos da inteligência artificial uma coleção de documentos com a finalidade de encontrar os tópicos que agrupam (*cluster*) os documentos semelhantes.

Os textos são insumos formados por uma sequência de palavras e, o LDA é um algoritmo computacional que lida com números. Para resolver essa questão, desenvolveu-se a técnica do “*Word Embeddings*” (ver seção 3.3.3), em que os termos em linguagem natural do texto são trocados por números, por meio de um processo sistemático chamado de dicionarização. Assim, os algoritmos processam os números que representam as palavras no dicionário (estrutura de dados computacional) e não as palavras em si.

Para um leitor humano, o texto é formado por uma hierarquia de unidades semânticas estruturadas em caracteres, palavras, frases e parágrafos. Para o algoritmo, no entanto, o texto é formado por uma sequência de caracteres. Essa diferença explica, em parte, porque a inteligência humana é capaz de entender o significado do texto, enquanto a inteligência artificial se limita a processar símbolos relacionados entre si por coocorrência, frequência e proximidade.

O modelo de Alocação Latente de *Dirichlet* adequa-se a tarefas como a análise de assuntos, a extração de informações, a indexação, a elaboração de resumos, a classificação, o enriquecimento de tesouros e a modelagem de tópicos (SILVA; CORREA, 2020).

A técnica de modelagem de tópicos pelo modelo LDA utiliza a inteligência artificial, o que significa que não se utiliza modelos dedutivos para calcular seus parâmetros e, assim, não há demonstrações matemáticas das decisões do modelo, o que se emprega nesses casos são métodos de validação por comparação entre o valor estimado e valor correto (padrão-ouro) em experimentos sob controle.

O uso dos métodos de inteligência artificial quando mediado por um modelo conceitual de uso garante que o ser humano esteja no centro da tomada de decisões, que é o que caracteriza a inteligência aumentada.

Na próxima seção, descreve-se o modelo conceitual.

3.4 Modelo conceitual

Nesta pesquisa, optou-se pelo termo “modelo conceitual” para referir-se à metodologia desenvolvida para orientar o desenvolvimento e o uso do protótipo tecnológico. Embora os termos “metodologia de uso”, “modelo de referência” e “modelo de processo” sejam sinônimos de “Modelo Conceitual”, porque apontam atividades de Planejamento (padrão de projeto), Uso (Processo) e Elaboração do *software* (Resultado), não serão utilizados.

Um Modelo Conceitual é uma ferramenta adequada para coletar, aperfeiçoar e listar as melhores práticas no estudo, no desenvolvimento e no uso de artefatos tecnológicos.

Eles são necessários porque artefatos tecnológicos são a essência da sociedade humana, que “é um mundo muito mais artificial do que um mundo natural, que, portanto, é necessário ter conhecimento sobre a ciência do artificial, sobre seus objetos e seus fenômenos” (SIMON, 1970, p. 25).

Na próxima seção, apresenta-se sucintamente a história dos modelos conceituais.

3.4.1 História

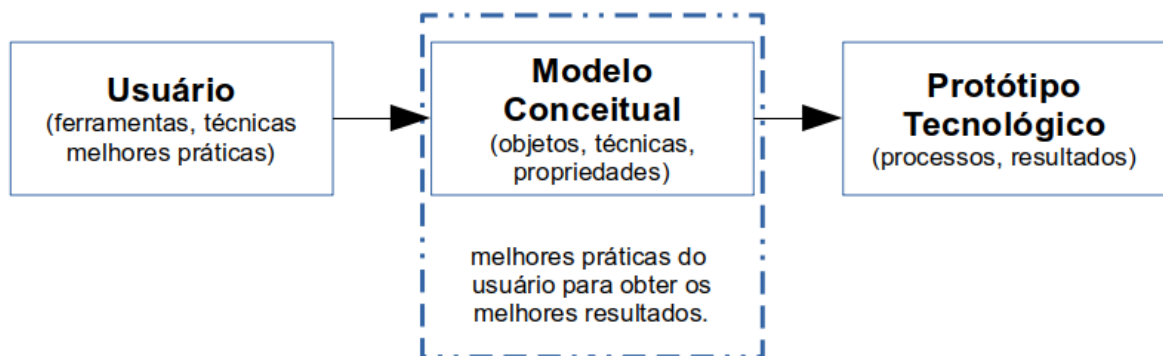
A história dos Modelos conceituais como tal pode ser encontrada a partir da publicação de Norman (1969).

O termo “modelo conceitual” foi adotado por diversas áreas do conhecimento, de forma que se tornou polissêmico, ambíguo e de difícil definição, que pode aparecer nos textos sob a designação simplificada de “modelo”. Entre as décadas de 1970 e 1990, muitas pesquisas foram realizadas sobre a modelagem conceitual, que foram adotadas principalmente pela engenharia de *software* e pelos sistemas de banco de dados.

Ao longo da história, um modelo conceitual já foi: (1) um conjunto de instruções para produzir um objeto real ou intelectual (NORMAN, 1969); (2) um mapeamento dos “conceitos” de um domínio e a relação entre eles (LARMAN, 2007); e (3) uma estrutura de trabalho (*framework*) com as melhores práticas na execução de uma tarefa (PIMENTEL, 2019).

Atualmente, um modelo conceitual pode ser visto como um conjunto de “concepções⁶” de um artefato associado às melhores práticas de seu uso (Figura 5).

Figura 5 – Visão geral e esquemática de um modelo conceitual.



Fonte: Adaptado de Maxey (2002) pelo autor.

⁶ Como se **concebe** um artefato.

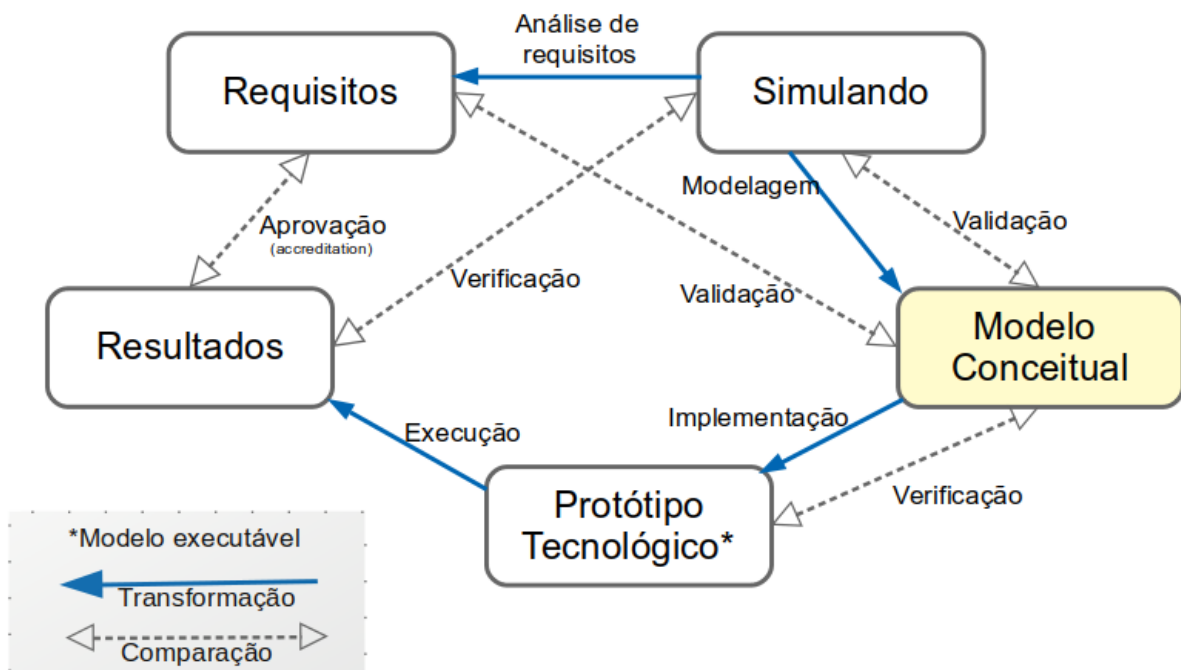
3.4.2 Definições

Segundo Domingos (2018), o modelo conceitual é “o conhecimento preliminar necessário para utilizar uma tecnologia eficientemente” (p. 14). Dessa proposição decorre que existe pelo menos um modelo conceitual para cada tecnologia, ainda que seja só na forma de conhecimento tácito (TAKEUCHI, 2009).

Para a comunidade do *Design* Interativo, os modelos conceituais são representações psicológicas abstratas de **como as tarefas devem ser realizadas** (SANDUSKY; PETTY, 2021). Nesse sentido, se pode citar o “Modelo Conceitual de Referência do Conselho Internacional de Museus” (PADRON; CRUZ; SILVA, 2018, p. 1), que estabelece um “marco de referência para dar entendimento claro, definir com precisão e compartilhar as informações” (RIVA, 2016, p. 3) para atender as necessidades dos usuários dos museus.

Norman (1969) o define como uma explicação útil, não necessariamente precisa, mas simplificada de como um artefato funciona. Esses fatores refletem, em certa medida, o fato de que ao explicar sua necessidade de informação, o usuário tem um conceito na mente diferente do que ele está falando e do que o projetista está entendendo. Uma visão geral da definição do modelo conceitual pode ser vista na Figura 6.

Figura 6 – Esquema genérico de modelo conceitual.



Autor/Copyright holder: Rick Sandusky & Mikel D. Petty. Copyright: Domínio público.

Fonte: Traduzido e adaptado de Sandusky e Petty (2021) pelo autor.

Em Hevner *et al.* (2004), listam-se as funcionalidades que, em tese, deveriam aparecer nos modelos conceituais desde a fase do planejamento, incluindo a implementação e no uso de um sistema artificial: a descrição da interface, o modo de operação, o conjunto de configurações e o padrão de respostas adequados para uma solução ótima. Estas funcionalidades, podem ser novas para o usuário, e, portanto, o uso delas não é intuitivo.

Diante do exposto, compreende-se que as definições variam entre autores, áreas do conhecimento e finalidades, mas em certa medida, define-se modelos conceituais como a ideia descrita em linguagem textual ou gráfica que permite o desenvolvimento, a implementação e o uso das tecnologias. No entanto, não só a terminologia, mas também a forma, as ferramentas e as linguagens artificiais para elaborar um MC modificam-se ao longo do tempo.

Na próxima seção, apresentam-se as características dos modelos conceituais.

3.4.3 Características

Um sistema bem desenvolvido é de uso intuitivo. Esta é uma pré-condição para que uma ferramenta seja capaz de contribuir para o aumento do intelecto humano (ENGELBART, 1962). No entanto, nem sempre é possível desenvolver um sistema inteiramente intuitivo (NORMAN, 1969).

Um Modelo Conceitual é caracterizado como uma solução tecnológica instanciada no artefato para o qual foi desenvolvido. Suas características incluem: 1) a comunicação dos princípios fundamentais e apresentação das funcionalidades do artefato; 2) a compreensão do domínio em que se insere a solução; 2) a transmissão dos detalhes tanto do problema, quanto da solução; 3) a implementação do sistema; 4) a documentação para referências futura; 5) a coordenação do modo de colaboração entre os usuários; e 6) a validação das soluções obtidas por meio do artefato.

A área do *design* de interação (SANDUSKY; PETTY, 2021) elencou outras características dos modelos conceituais, que foram traduzidas e adaptadas ao contexto desta pesquisa e estão listadas a seguir:

1. fornecer compreensão de alto nível do artefato modelado;
2. múltipla modelagem comportamental dos usuários, que possibilita uma solução tecnológica mais utilizável e mais intuitiva;
3. incluir mais de uma trilha de aprendizagem; e
4. alinhamento entre a solução tecnológica e os objetivos estratégicos (refletido na política de indexação) da instituição.

A BCI utiliza os Modelos Conceituais como ferramentas para lidar com os problemas de organização da informação, porque eles permitem conceber duas visões de um objeto informacional, uma física e outra abstrata, facilitando a elicitación dos relacionamentos (PADRON; CRUZ; SILVA, 2018) do todo com suas partes e do objeto com seu ambiente (outros objetos). Isso ocorre porque umas das propriedades intrínseca dos modelos conceituais é a integração por serem flexíveis, adaptáveis, genéricos e abstratos.

3.4.4 Aplicabilidade e importância

Os profissionais da informação propõem, elaboram e utilizam modelos conceituais nas suas práticas profissionais. Os cientistas da computação adotam os modelos conceituais para projetar (*design*) *softwares* e bases de dados da mesma forma que as engenheiras civis fazem com seus desenhos técnicos de plantas baixas e os engenheiros elétricos com seus esquemas visuais de circuitos (PADRON; CRUZ; SILVA, 2018).

O grupo de pesquisa MHTX tem a tradição de elaborar modelos conceituais, haja visto as teses de Lima (2004), Pontes (2013), Silva (2013), Alves (2014), Maculan (2015) e Maia (2018).

Uma outra aplicação dos modelos conceituais é orientar o uso de uma ferramenta (SANDUSKY; PETTY, 2021). Assim, a qualidade da produção por meio da ferramenta melhora porque os modelos conceituais adicionam controle dos padrões de entradas, recomendam e coordenam o uso das melhores práticas, indicam a configuração adequada dos equipamentos e apresentam métodos de avaliação do produto (VAISHNAVI; KUECHLER, 2008).

Os modelos conceituais são desenhados para coordenar, treinar e qualificar pessoas no uso de sistemas artificiais. São eles que fazem a mediação entre a linguagem cotidiana do usuário e a linguagem técnica da nomenclatura. Eles guiam a experiência do usuário, facilitam o uso de interfaces complexas e organizam a sequência de procedimentos que garantem mais eficiência nos processos.

Evitar que os artefatos tecnológicos caiam no paradoxo de ter sido feito para facilitar e acabar dificultando, é uma utilidade dos modelos conceituais, além de conduzir a experiência do usuário e orientar a adoção das melhores práticas no uso de uma ferramenta para produzir respostas otimizadas (NORMAN, 1969).

Em síntese, um modelo conceitual contribui para a compreensão de um problema, que de outra forma poderia ser incompreensível, ao facilitar a visualização dos processos e ajudar na divisão do problema em partes menores e gerenciáveis. De forma que, o uso da

ferramenta acaba por expandir a inteligência do usuário, forjando uma nova experiência e elevando os níveis de compreensão.

Para explicar melhor essa abordagem, na próxima seção, apresenta-se a inteligência aumentada.

3.5 Inteligência aumentada

Uma pesquisa científica contribui para a compreensão de um fenômeno, (KUHN, 2000) e sua abordagem deve considerar os desenvolvimentos anteriores, procurando, em primeiro lugar, conhecer os paradigmas (fundamentos teóricos) e os propósitos práticos utilizados (VAISHNAVI; KUECHLER, 2008), e, com isso, evitar repetições desnecessárias e propor avanços. Por isso, analisou-se tanto o arcabouço teórico quanto os *softwares* utilizados nos estudos que combinaram inteligência artificial e inteligência humana.

Neste contexto, *software* inteligente é um artefato, ou seja, é um objeto humano construído por método científico com uma finalidade bem definida (WIERINGA, 2014). Os termos “inteligência humana”, “pessoa” e “inteligência natural” foram utilizados como sinônimos ao longo deste texto, e, em geral, em contraponto à ideia de inteligência artificial. Enquanto o termo “inteligência aumentada” foi empregado com o sentido de combinar (e não de contrapor) a inteligência humana e a inteligência artificial para alcançar melhores soluções para problemas humanos complexos, conforme os fundamentos teóricos propostos por Engelbart (1962).

Nas próximas seções, apresentam-se o histórico do desenvolvimento do conceito e do termo inteligência aumentada, bem como as suas definições, características e aplicações.

3.5.1 Histórico

O mundo, no pós-guerra, produziu pesquisas científicas num ritmo acima do que se podia processar adequadamente. Com base nesse fato, Vannevar Bush, em 1945, previu corretamente que o volume de dados, em pouco tempo, modificaria a capacidade de se organizar o conhecimento e exigiria o desenvolvimento de novos instrumentos para selecionar as informações relevantes. Então, ele propôs o Memex no seu bem conhecido artigo “*As We May Think*”, no periódico *Atlantic Monthly*, de julho de 1945, da seguinte maneira:

Considere um dispositivo do futuro para uso individual, que é uma espécie de arquivo e biblioteca pessoal mecanizados. Ele precisa de um nome e, para cunhar um ao acaso, “Memex” é o suficiente. Um Memex é um dispositivo no qual um indivíduo armazena todos os seus

livros, registros e comunicações, e que é mecanizado para que possa ser consultado com velocidade e flexibilidade elevadas. É um dispositivo pessoal para aumentar a memória (BUSH, 1945, p. 6, tradução nossa).

Essas ideias visionárias de Vannevar Bush influenciaram Douglas Engelbart na elaboração do modelo conceitual de aumento do intelecto humano proposto no artigo “*Augmenting Human Intellect: a Conceptual Framework*”, que, segundo Cerf (2013), dialoga com o conceito de amplificação da inteligência de William Ross Ashby, no livro “*Introduction to Cybernetics*” (ASHBY, 1956), uma vez que Engelbart (1962, p. 4) também “visa desenvolver meios para aumentar o intelecto humano”.

Essas concepções ajudaram a suavizar a percepção amplamente difundida de que a inteligência artificial é uma substituta da inteligência humana. Além disso, elas garantem a implementação adequada dos sistemas inteligentes ao abordar “diretamente os desafios no planejamento, na avaliação, na implementação e na supervisão” desses sistemas (KOVARIK *et al.*, 2019, p. 1) de forma que não substituam as pessoas.

Na próxima seção, apresentam-se as definições de inteligência aumentada.

3.5.2 Definições

A definição de inteligência aumentada, devido à complexidade do fenômeno e sua novidade na pesquisa científica, ainda não é definitiva (WÓJCIK, 2020), assim ela é apresentada como termo genérico, padrão de projeto, processo e resultado, de forma que é possível afirmar que se trata de um princípio para desenvolver e usar *softwares* com inteligência artificial.

O dicionário *Michaelis* define “princípio” como um conjunto de proposições e de diretivas fundamentais a que se subordinam todas as etapas de desenvolvimento e de uso de um sistema em um campo do conhecimento. Nesse sentido, a inteligência aumentada é, portanto, o princípio para o planejamento, a avaliação, a implementação e a supervisão humana de sistemas com inteligência artificial.

O conceito, não o termo, “inteligência aumentada” foi definido em Bush (1945, p. 6) como “um dispositivo pessoal para aumentar a memória”. Nesse contexto, memória é a habilidade básica do intelecto humano em que “a informação recebida interage com as informações que já residem na mente” (HEYLIGHEN, 2015, p. 14).

Em Pasquinelli (2015, p. 203) define-se a inteligência aumentada como “um termo genérico” utilizado para referir-se à “relação complexa entre a inteligência humana” e “as máquinas computacionais”, e, também, como “uma expansão (também em um grau social e político) das faculdades cognitivas humanas”, de forma que corrobora a seguinte definição:

O termo "aumentar o intelecto humano" significa aumentar a capacidade de uma pessoa de compreender uma situação problemática complexa, [...], e [com isso] derivar soluções para os problemas. O aumento da capacidade, neste caso, significa uma combinação de: compreensão mais rápida e melhor, possibilidade de obter um grau útil de compreensão em uma situação que antes era muito complexa, soluções mais rápidas, melhores soluções e a possibilidade de encontrar soluções para problemas que antes pareciam insolúveis. (ENGELBART, 1962, p. 1, tradução nossa).

Dessa forma, aumentar o intelecto humano significa capacitar uma pessoa a compreender um problema complexo e propor uma solução melhor e em menos tempo do que faria sem o auxílio da ferramenta. Com isso, a inteligência aumentada foi definida como uma consequência virtuosa com impactos positivos sobre a realidade. Yau *et al.*, (2021, p. 1) acrescentam que a inteligência aumentada é o processo em que ocorre uma “parceria entre humanos e máquinas (ou computadores) com a seguinte postulação: 50% inteligência artificial + 50% inteligência humana” (tradução nossa).

Em conjunto, essas definições apresentam a inteligência aumentada como o padrão de projeto centrado no ser humano que permite o planejamento adequado do processo de interação entre as pessoas e a inteligência artificial. Disso decorre uma melhora do desempenho cognitivo humano na tomada de decisões, portanto, torna o trabalho humano mais eficiente e produz soluções mais eficazes para problemas complexos (ZHENG *et al.*, 2017).

Por fim, a inteligência aumentada foi definida como o padrão de projeto, o processo de elaboração e de uso de *software* inteligente e o tipo de resultado esperado. Logo, ela é o princípio para elaborar e usar a inteligência artificial, que acaba por expandir a capacidade cognitiva e intelectual do usuário.

As características da inteligência aumentada estão descritas na próxima seção.

3.5.3 Características

A inteligência aumentada ocorre em duas etapas: (1) utiliza-se a inteligência artificial para descobrir os significados, os padrões e as correlações existentes nos dados em grande escala, e (2), em seguida, encarrega-se a inteligência humana para tomar as decisões, considerando os achados da primeira etapa.

A principal característica da inteligência aumentada é fazer da inteligência artificial uma assistente virtual do ser humano no processamento de *big data* (SADIKU *et al.*, 2021). Assim, o ser humano é quem toma as decisões; e, para esclarecer mais esta questão, na próxima seção, explica-se a relação da inteligência aumentada com a tomada de decisões na análise de assuntos. Quando a inteligência artificial auxilia a inteligência humana

adequadamente, o resultado obtido será maior do que a soma de suas partes, assim as características da inteligência aumentada é a sinergia e a expansão cognitiva.

A inteligência Aumentada proporciona novas possibilidades de aquisição, processamento e uso de informações, aprimorando a experiência humana por meio da interação com a inteligência artificial (WÓJCIK, 2021, p. 1).

As pessoas executam apenas as tarefas intelectuais na inteligência aumentada e os *softwares* realizam as tarefas repetitivas (KOVARIK *et al.*, 2019, p. 3), promovendo a “inteligência artificial colaborando com as pessoas” e não “inteligência artificial substituindo as pessoas”.

Observe no Quadro 4 que a análise de assuntos Intelectual é realizada pela inteligência humana. Ou seja, um profissional da informação faz todo o trabalho e toma todas as decisões, o que torna o processo mais subjetivo, ainda que produza indexações com mais qualidade. Além disso, o processo intelectual exige menos investimento inicial, mas produz menos, faz uso intensivo de mão-de-obra e é uma atribuição do profissional da informação.

Quadro 4 – A tomada de decisão na análise de assuntos.

Característica	Intelectual	Automatizada	Assistida
Inteligência	humana	artificial	Humana + artificial
Subjetividade	Maior	Menor	Menor
Extensão	Um documento	Muitos documentos	Colaborativa
Qualidade	Alta	Desconhecida	Alta
Investimento	Baixo	Alto	Controlável
Velocidade	Baixa	Alta	Intermediária
Formação profissional	Informação	Informática	Informação
Uso de mão de obra	Intensivo	Baixo	Intermediário

Fonte: Elaborado pelo autor a partir de Davenport, Harris e Morison (2010, p. 146).

Segundo Davenport, Harris e Morison (2010), decisões automatizadas se adequam às situações simples com regras bem definidas, e que, por não levantarem exceções, são velozes, mas produzem resultados de baixa qualidade. Do Quadro 4, observa-se que, na análise de assuntos automatizada, a inteligência artificial toma todas as decisões, é mais rápida, processa mais documentos, não usa mão-de-obra humana, exige um investimento mais alto, em que a qualidade das indexações é desconhecida (pode ser baixa) e é uma atribuição laboral da Informática.

Ainda no Quadro 4, observa-se que as decisões assistidas lidam com situações complexas, ocorrências imprevisíveis, beneficiam-se da experiência e da especialização do ser humano (processo intelectual) e da eficiência incansável das máquinas (processo automático). Assim, a análise de assuntos assistida por computador pode ter as melhores características dos dois processos, porque a indexação tem afinidade com as decisões assistidas, considerando que a atribuição dos termos indexadores é uma atividade complexa,

incluindo aspectos além do texto, como preferências culturais, estado do conhecimento do usuário e mudanças linguísticas. Além disso, o processo assistido reduz o trabalho humano, dá emprego para os profissionais da informação, aumenta a qualidade dos resultados e a velocidade em que eles são produzidos.

Na próxima seção, apresenta-se a aplicabilidade da inteligência aumentada na Biblioteconomia e a Ciência da Informação.

3.5.4 Aplicabilidade e importância

A aplicação da inteligência aumentada não se limita ao uso de *softwares*, inclui também o uso de *hardware* para o controle remoto de equipamentos (teleoperação), para a captação de sinais vitais por meio de sensores implantados no corpo humano (transumanismo), para a ampliação da visão virtual para três dimensões (realidade aumentada) e para a automação da linha de produção industrial (hiperautomação). No entanto, o uso de hardware não foi abordado nesta pesquisa, além disso, é o *software* que torna que é inteligente e o *hardware* é apenas a extensão mecânica do artefato.

A inteligência artificial está presente no cotidiano das pessoas. De forma geral, todos os dispositivos denominados de “*smarts*” ou “inteligentes” estão imbuídos de técnicas inteligentes (EKBIA, 2010). Essa presença ubíqua também está disponível para os profissionais da Informação que podem utilizá-la para automatizar, por exemplo, a análise semântica de textos, a produção de resumos, a associação de cabeçalhos de assunto e a catalogação (GRIFFEY, 2019).

Os métodos tradicionais utilizam os modelos matemáticos para explicar os sistemas e fazer previsões, enquanto a inteligência artificial busca a resposta diretamente nos dados, sem hipóteses ou modelos preditivos. Como os dados utilizados são derivados das ações humanas, a inteligência artificial precisa de supervisão humana, porque ela apenas descobre os padrões que já existem nos dados, e, se lhe for dada autonomia, ela pode replicar e reforçar padrões (atitudes e falas) discriminatórios (RUSSELL; NORVIG; DAVIS, 2010). Portanto, a inteligência aumentada é percurso metodológico adequado, pois ela coloca as ferramentas computacionais para realizar as tarefas técnicas sob a tutela da inteligência humana.

Ainda discorrendo sobre o uso da inteligência artificial nas Bibliotecas, Jason Griffey, um bibliotecário, afirma que:

[...] essa metodologia pode interessar às bibliotecas. Mas, à medida que as bibliotecas e seus fornecedores passam a desenvolver sistemas de inteligência artificial e aprendizado de máquina,

umentam-se as implicações quanto à privacidade dos dados coletados e armazenados necessários para treinar e atualizar esses sistemas (GRIFFEY, 2019, p. 8, tradução nossa).

A inteligência artificial é uma boa solução tecnológica para as bibliotecas, mas precisa de supervisão, porque uma das consequências do uso não supervisionado da inteligência artificial é ocasionar problemas como perda de privacidade e a replicação de preconceitos, discriminações e crenças exclusivistas, fatos que se opõem aos princípios fundamentais da BCI.

Assim, o uso da inteligência aumentada pode trazer os benefícios da inteligência artificial para as atividades da BCI e, devido à supervisão humana, o ganho de produtividade não traz consigo os vieses culturais e nem problemas de privacidade comuns aos processos de automatização.

3.6 Considerações

A inteligência artificial “interage com pessoas, tecnologias, ideias, ideais e atividades” (EKBLA, 2010, p. 21), na medida em que a representação do conhecimento permite aos algoritmos “compreender” a linguagem natural. Essa compreensão ocorre em sentido duplo, dos algoritmos para as pessoas e das pessoas para os algoritmos, de tal forma que os resultados artificiais imitam e até influenciam os processos intelectuais. No entanto, toda a produção da máquina se destina a satisfazer as necessidades humanas de informação. Nesse processo, produz o que Engelbart (1962) denominou de aumento do intelecto humano.

Como ferramenta cognitiva, a máquina memoriza, processa e resolve problemas a uma velocidade acima do que o intelecto humano pode acompanhar, assim, a Inteligência aumentada o termo genérico para descrever a relação entre a inteligência humana e os computadores, que pode ser vista como uma expansão informacional, social e política das faculdades cognitivas humanas.

A inteligência artificial é o campo de pesquisas em que se insere a disciplina Processamento de Linguagem natural dos seguintes métodos que foram utilizados nesta pesquisa para assistir a análise de assuntos: (1) Sumarização automática (seção 3.3.4.1); (2) expressões regulares (seção 3.3.4.2); e Modelagem de Tópicos (seção 3.3.4.3).

No próximo capítulo, apresenta-se a metodologia necessária para alcançar a inteligência aumentada como uma metodologia para a análise de assuntos assistida por computador.

4 METODOLOGIA

Apresentam-se, neste capítulo, a caracterização da pesquisa, o universo da pesquisa e os procedimentos metodológicos.

Nesta metodologia, considera-se que a realidade é situada e contextualizada, o que possibilita múltiplas alternativas sociotécnicas, em que o conhecimento científico pode ser obtido pela construção objetiva de um artefato, inscrito em um contexto determinado, em que o conhecimento é obtido pelo fazer iterativo e interativo, e que o desenvolvimento e o aperfeiçoamento se dá pela constante avaliação do impacto do artefato sobre a realidade, e desta sobre o artefato, assim aproximando-se da Pesquisa-ação (VAISHNAVI; KUECHLER, 2004).

Na próxima seção, aborda-se a caracterização da pesquisa.

4.1 Caracterização da pesquisa

Apresenta-se, nesta seção, a caracterização desta pesquisa quanto à sua natureza, aos seus objetivos e aos seus procedimentos. Segundo Gil (2019, p. 43), isso facilita “analisar os fatos do ponto de vista empírico, para confrontar a visão teórica com os dados da realidade”, e, assim “traçar os modelos conceitual e operativo da pesquisa”.

Esta pesquisa é de **natureza aplicada**, pois gerou conhecimento de aplicação prática para solucionar problemas reais da indexação, o que envolveu a análise de artigos científicos e seus metadados (dados de indexação). Com isso, procura-se ampliar o conhecimento sobre a inteligência aumentada e a sua aplicação nos processos técnicos-científicos da Biblioteconomia e da Ciência da Informação.

Classifica-se esta pesquisa como **social aplicada**, pois produziu um artefato tecnológico com o potencial de mudar uma realidade social com impactos positivos para os profissionais da informação, facilitando-lhes a maneira de realizar seus afazeres técnicos; e para usuários das bibliotecas digitais, dando mais qualidade e velocidade nos processos que levam a uma melhor recuperação da informação. Além disso, procurou ativamente métodos e meios tecnológicos que não ameaçam a manutenção de postos de trabalhos.

Quanto aos objetivos, esta é uma pesquisa **exploratória**, pois empregou técnicas e princípios teóricos para serem testados e validados ao longo do desenvolvimento, em que não houve a formulação de hipóteses e, sim, busca por dados, informações e métodos sobre a inteligência aumentada (um assunto determinado) que fossem aplicáveis a análise de assuntos. Essas escolhas foram baseadas nas conclusões da pesquisa bibliográfica (Revisão de Literatura), em que se constatou que inteligência aumentada foi pouco estudada na área da BCI.

Quanto aos procedimentos metodológicos, esta pesquisa é **experimental**, pois utilizou técnicas e fundamentos teóricos em dados reais (artigos científicos produzidos pela Embrapa), cujos resultados (estimados pela inteligência artificial) serão avaliados estatisticamente, quanto à precisão e à acurácia, em relação a dados reais elaborados por seres humanos.

Esta é uma pesquisa Empírica, isto é, baseada em evidências, pois utilizou dados e informações produzidos e acumulados, ao longo de mais de duas décadas (1998 – 2020), em 40 bibliotecas. O que está em conformidade com a afirmação de que a análise preditiva de dados é uma ciência intrinsecamente empírica (TUKEY, 1962), cuja ênfase está no uso de dados sem uma hipótese estatística. Nesse caso, a descoberta de conhecimento está orientada aos dados, em que o poder de processamento disponível procura e testa as respostas por métodos exaustivos. Isso produz boas respostas, cobre mais possibilidades que uma análise humana, mas pode produzir uma resposta de baixa confiabilidade, o que reforça a necessidade da inteligência aumentada, pois o ser humano incluído no processo distingue muito bem se os resultados são bons ou descartáveis.

A interlocução desta pesquisa com a visão do pragmatismo, concepção filosófica que considera o conhecimento como mecanismo de ação, produziu soluções para problemas reais, em que se projetou e construiu artefatos tecnológicos, propôs novos processos e elaborou métodos capazes de produzir resultados com mais velocidade do que os métodos puramente manuais (SIMON, 1970).

Na próxima seção, apresenta-se a delimitação do universo da pesquisa.

4.2 Universo da pesquisa

O objeto desta pesquisa é o processo de análise de assuntos assistida por computador sob os princípios da inteligência aumentada. O universo de aplicação é o acervo da Base de Dados da Pesquisa Agropecuária (BDPA), tendo como recorte um conjunto de 7377 artigos científicos na área da pesquisa científica em agropecuária indexados pelos termos do Thesagro.

O *software* de análise de assuntos assistida por computador (3ApC) desenvolvido nesta pesquisa foi validado com os artigos do *corpus* do experimento selecionado (ver seção 5.4) da BDPA. No experimento de validação desse protótipo tecnológico, avaliou-se qualitativamente o potencial dos assuntos relevantes selecionados pelo *software* serem traduzidos nos respectivos descritores disponibilizados pela BDPA.

As atividades desta pesquisa incluíram o desenvolvimento de um Modelo Conceitual de inteligência aumentada (APÊNDICE A), um protótipo tecnológico (*software*

análise de assuntos assistida por computador (3ApC)) e um experimento de validação do 3ApC.

Na próxima seção, estão descritos os procedimentos metodológicos necessários para alcançar os objetivos desta pesquisa.

4.3 Procedimentos metodológicos

Nesta seção, apresentam-se procedimentos metodológicos necessários ao desenvolvimento da pesquisa e o suporte teórico e prático necessário, que foram obtidos pelo método indutivo de investigação e pela revisão de literatura.

A temática desta pesquisa é por natureza interdisciplinar, de forma que foram consultadas as bases de dados tanto da BCI quanto de outros campos científicos. A busca por literatura científica foi realizada nas seguintes bases de dados: *Library & Information Science Abstracts (LISA)*; *Library and Information Science and Technology Abstracts (LISTA)*; *Association for Computer Machinery (ACM)*; *Institute of Electrical and Electronics Engineers (IEEE)*; e *Web of Science* (disponibilizado pela CAPES) e, também, no sistema automatizado de recuperação da informação *Google Acadêmico*[®].

A seleção dos dados endossou as recomendações metodológicas de Creswell (2018), em conformidade com a redação dada por Sousa, Lima e Araújo (2022, p. 103):

identificar as palavras-chave, realizar pesquisas nos bancos de dados em busca de acervos, coletar a produção científica contendo as palavras-chave e examinar as publicações para entender a contribuição de cada uma para o tema em análise.

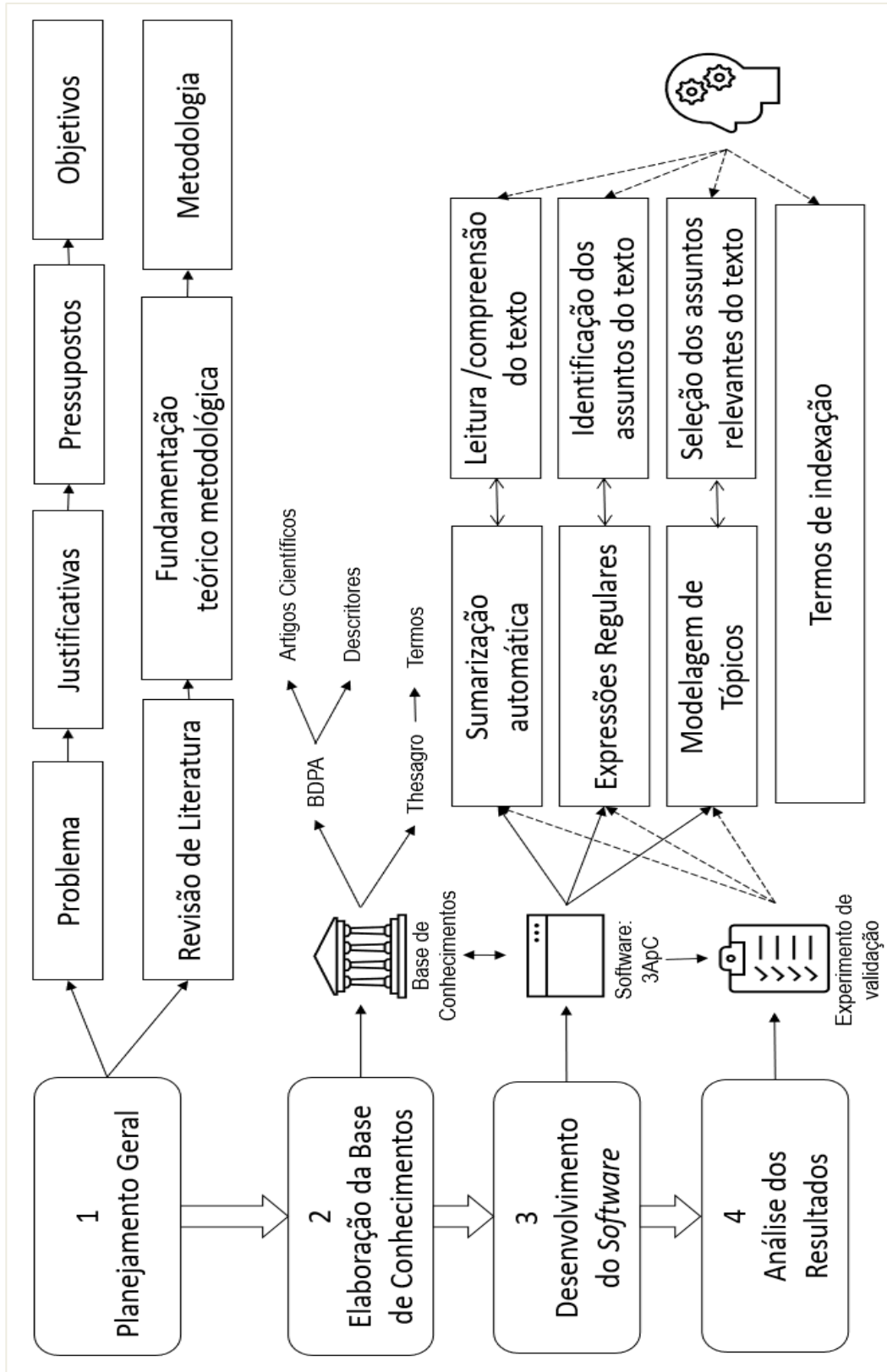
As discussões empreendidas, nesta pesquisa, dão maior ênfase aos aspectos qualitativos, uma vez que se procura a compreensão do fenômeno em uma perspectiva do uso da inteligência artificial supervisionada pela inteligência humana.

Na próxima seção, apresenta-se o planejamento geral dos procedimentos metodológicos.

4.3.1 Etapa 1: Planejamento geral

Os procedimentos metodológicos (Figura 7) visam: (1) elaboração de uma base de conhecimento, (2) desenvolvimento do *software* de análise de assuntos assistida por computador (3ApC), e (3) a realização de um experimento de validação.

Figura 7 – As etapas dos procedimentos metodológicos.



Fonte: elaborada pelo autor.

O levantamento sistemático dos assuntos centrais desta pesquisa incluiu a análise de assuntos, a inteligência artificial, o processamento de linguagem natural e a inteligência aumentada, em que se obteve uma visão abrangente dos campos científicos analisados, o que viabilizou o planejamento da Metodologia. Outras temáticas relacionadas também foram pesquisadas, incluindo a análise de assuntos assistida por computador.

As publicações científicas de Chapman *et al.* (2000), Goodfellow, Bengio e Courville (2016), Zaki e Mccoll-Kennedy (2020) e Studer *et al.* (2021) apresentam procedimentos metodológicos (Modelo Conceitual) consolidados para a mineração de dados que são adaptáveis a esta pesquisa, pelo que foram utilizadas como as referências principais.

De modo geral, os procedimentos metodológicos para a análise de textos por métodos computacionais são divididos em: Compreensão do problema, tratamento dos dados, modelagem e avaliação dos resultados (Quadro 5).

Quadro 5 – Comparação entre os modelos utilizados como fonte.

Modelo Proposto	CHAPMAN (2020)	ZAKI (2020)	STUDER (2021)
Compreensão do problema	Compreensão do negócio	Estudo do conhecimento prévio e compreensão do negócio	Compreensão de negócios e dados
Tratamento dos dados	Compreensão e Preparação dos dados	Compreensão dos dados	Preparação dos dados
Modelagem	Modelagem	Modelagem e Validação dos dados	Modelagem e Avaliação
Avaliação dos resultados	Implantação	<i>Insights</i> obtidos	Implantação, Monitoramento e Manutenção

Fonte: Elaborado pelo autor.

Os procedimentos metodológicos devem contemplar o problema, os dados, o modelo e o uso, pois são eles que delineiam as quatro etapas correspondentes: (1) planejamento do projeto, (2) tratamento dos dados, (3) modelagem da inteligência artificial, e (4) inteligência aumentada.

A relação entre o Modelo, os procedimentos metodológicos e os capítulos da Tese podem ser vistos no Quadro 6 e o diagrama dos procedimentos metodológicos podem ser vistas na Figura 7, e visam alcançar o objetivo geral da pesquisa consubstanciado nos objetivos específicos.

Justifica-se a elaboração da base de conhecimento, a exigência metodológica da inteligência artificial de uma ampla base de dados, informações e conhecimentos para que seus modelos sejam generalizáveis. Além disso, a tomada de decisões (inteligência aumentada) atual beneficia-se dos conhecimentos acumulados nas decisões anteriores.

Quadro 6 – A relação dos procedimentos metodológicos com os capítulos da tese.

Modelo conceitual	Procedimento metodológico	Capítulo da tese
Compreensão do problema	Planejamento geral	1 Introdução (Descrição do Problema; Justificativas; Pressupostos; e Objetivos); 2 Revisão da literatura; 3 Fundamentação teórica e metodológica; 4 Metodologia.
Tratamento dos dados	Elaboração da base de conhecimento	5 Elaboração da base de conhecimento.
Modelagem	Desenvolvimento do <i>software</i>	6 Desenvolvimento do <i>Software</i> .
Avaliação dos resultados	Análise dos resultados	7 Análise dos resultados.

Fonte: Elaborado pelo autor.

A etapa de desenvolvimento do *software* 3ApC justifica-se na necessidade de gerenciar a base de conhecimento, de treinar e de usar os modelos de inteligência artificial, além de ser a Interface entre o indexador e a máquina.

A etapa de análise dos resultados justifica-se na avaliação e validação tanto da base de conhecimento quanto do protótipo tecnológico.

Na próxima seção, descreve-se o planejamento para a elaboração da base de conhecimento.

4.3.2 Etapa 2: Elaboração da base de conhecimento

As Bases de Dados de Pesquisa Agropecuária (BDPA) foram escolhidas por formarem o repositório científico aberto muito importante para as pesquisas agropecuárias no Brasil. Além disso, o grande acervo (mais de um milhão de artefatos) corretamente indexados facilita a realização de experimentos computacionais de indexação.

A base de dados desta pesquisa foi denominada de “base de conhecimento” (em lugar de “banco de dados”), porque, segundo Kobashi (1991, p. 40), os algoritmos de inteligência artificial não lidam apenas com dados, às vezes, eles processam os conhecimentos diretamente.

Nesta seção, o termo “recurso informacional” refere-se ao conjunto de artigos científicos (dados) e seus respectivos descritores (metadados); e aos termos do *Thesaurus* Agrícola Nacional (Thesagro).

Os recursos informacionais foram organizados na base de conhecimento da seguinte forma:

1. *Corpus* da pesquisa:

- a. todos os artigos científicos coletados.
2. Metadados:
 - a. lista de descritores associados corretamente aos seus respectivos artigos;
 - b. quantidade de artigos científicos que cada descritor indexou no *corpus*; e
 - c. os descritores coocorrentes (aparecem juntos nas listas de descritores).
3. Thesagro:
 - a. Termos e suas (meta)informações;
 - b. Expressões regulares;
 - c. Nomes científicos e seus respectivos nomes comuns;
 - d. Termos não preferidos e seus respectivos termos preferidos;
4. *Corpus* do experimento:
 - a. 100 artigos científicos selecionados do *Corpus* da pesquisa.

Na próxima seção, apresenta-se as Bases de Dados da Pesquisa Agropecuária.

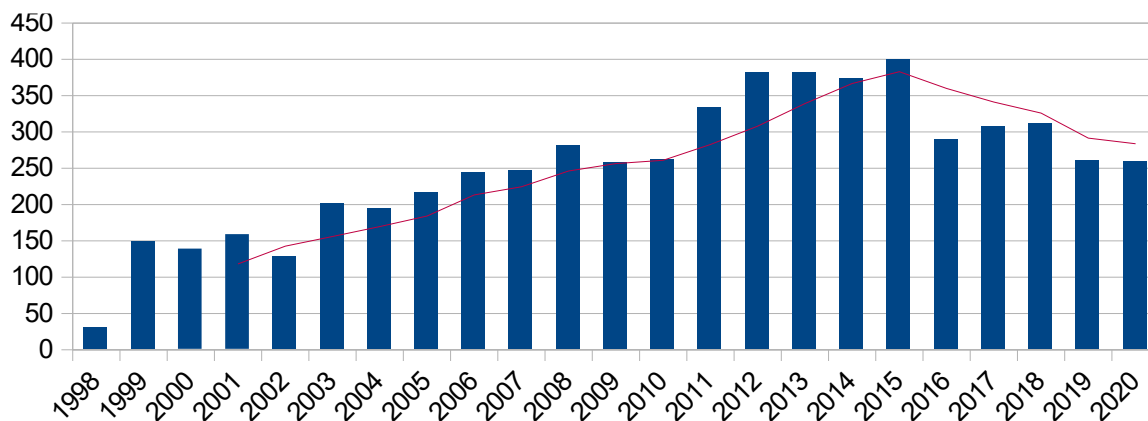
4.3.2.1 Bases de Dados da Pesquisa Agropecuária

Os dados (artigos científicos) e metadados (principalmente os descritores) foram coletados do acervo da biblioteca nacional da agropecuária (BINAGRI) indexados pelas Bases de dados da pesquisa agropecuária (BDPA) gerenciada pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa).

A Embrapa é uma estatal criada em 26 de abril de 1973, vinculada ao ministério da agricultura, pecuária e abastecimento (MAPA), que desenvolve modelos de agricultura e pecuária tropical com características brasileiras considerando as limitações impostas pelo ambiente, pelos efeitos das alterações climáticas e pelas barreiras culturais. A Embrapa contribuiu para transformar a agropecuária nacional em um agronegócio moderno de nível global, sem abandonar a agropecuária familiar, figurando entre os que tiraram o Brasil da condição de importador de alimentos para elevá-lo à condição de produtor e exportador mundial de *commodities*. Este parágrafo é uma adaptação do texto oficial disponível no portal institucional da Embrapa (EMBRAPA, 2022).

Na Figura 8, apresenta-se, no gráfico, o acervo da BDPA em quantidade de artigos científicos coletados nesta pesquisa. A linha de tendência (linha vermelha) mostra a *média móvel de quatro anos* da produção indexada.

Figura 8 – O volume de artigos científicos coletados por ano de publicação.



Fonte: Elaborado pelo autor.

A BDPA é a biblioteca virtual da BINAGRI fundada em 1996, cujo acervo contém 1.015.147 documentos de 40 bibliotecas físicas distribuídas nas unidades da EMBRAPA (conferido em 01/03/2022). A BDPA mantém os documentos (cópias digitais) e seus respectivos metadados acessíveis, públicos e abertos (SILVA, 2022). Este parágrafo é uma adaptação do texto oficial disponível no portal institucional da Embrapa (EMBRAPA, 2022).

Na próxima seção, apresenta-se o *Thesaurus* Agrícola Nacional (Thesagro).

4.3.2.2 *Thesaurus* Agrícola Nacional

Os tesouros são regulamentados e padronizados pela ISO 2788:2020, que os define como “um vocabulário controlado e dinâmico, composto por termos que têm entre si relações semânticas e genéricas, aplicáveis a um domínio particular de conhecimento”. Segundo Hagar Espanha Gomes (1990), o termo “tesouro” tem origem no dicionário analógico de Peter Mark Roget, publicado pela primeira vez em Londres, em 1852.

Um tesouro é uma lista de conceitos de um domínio, estruturada hierarquicamente para representar conteúdos de forma a facilitar a indexação (armazenamento) e a recuperação de documentos. Para Lima e Maculan (2017), são instrumentos de representação do conhecimento com vistas a organizar a informação para recuperá-la. Para Fujita e Leiva (2009, p. 158), os tesouros são “linguagens documentárias”.

O Tesouro Agrícola Nacional (THES) do serviço de Bibliografias Personalizadas em Agricultura (BIP/ACRI) de 1975 foi a fase experimental do Thesagro, restrita a 150 pesquisadores de 39 instituições nacionais.

O THES foi disponibilizado para uso público em 1976 como o Serviço de Disseminação Seletiva da Informação do Sistema Nacional de Informação e Documentação Agrícola (SNIDA), nessa época, os dados gerados mensalmente eram armazenados em arquivos magnéticos (CHASTINET *et al.*, 1978).

A primeira impressão do Thesagro ocorreu em 1979 ainda sob a denominação de “THES – *Thesaurus* Agrícola Brasileiro (Normalizador de Termos Agrícolas)” e a segunda, em 1989. A partir de 2015, o Thesagro passou a ser um serviço *on-line*.

Atualmente, existem quatro versões *on-lines* do Thesagro, que podem ser acessadas pelas URLs listadas no Quadro 7. Não existe nenhum aviso quanto a essa multiplicidade de versões. Por exemplo, a versão oficial tem 12068 relações entre os seus 9857 termos, enquanto a versão 3 tem 12156 relações entre os seus 9573 termos.

Quadro 7 – As URLs de acesso a diferentes versões *on-lines* do Thesagro.

Versão	Descrição	Última Atualização	URL
1	Primeira versão (Desatualizada)	1978	https://snida.agricultura.gov.br/html/Cen_BaseDados1.html
2	Versão com anglicização linguística (Desatualizada)	2015	https://www.vocabularyserver.com/thesagro/index.php
3	Versão em dados abertos para <i>download</i> (XML) (Desatualizada)	2017	https://www.gov.br/agricultura/pt-br/aceso-a-informacao/dadosabertos/thesagro/thesaurusagricolanacional_wxr.xml
4	Versão Oficial do Thesagro	2021	https://sistemas.agricultura.gov.br/tematres/vocab/index.php

Fonte: Elaborado pelo autor.

A primeira versão do Thesagro é de 1976 e, atualmente, não recebe atualizações (Versão 1 do Quadro 13). A Versão 2 (Quadro 13) do Thesagro está disponível *on-line* desde 2015 e apresenta os seus termos em anglicização linguística (Figura 9A) e não recebe atualizações desde 2017.

Figura 9 – As diferenças entre as duas versões online do Thesagro.



(A)



(B)

Fonte: Elaborado pelo autor.

A versão 3 (Quadro 13) do Thesagro está disponível para *download* em um site de dados abertos, desde 2017, em formato XML⁷. A versão oficial do Thesagro (versão 4 do Quadro 13) está atualizada e disponível no servidor da BINAGRI em português padrão (Figura 9B).

Os termos do Thesagro estão distribuídos em 49 níveis hierárquicos, em que 286 termos estão no primeiro nível (raiz) e 6311 (64,61%) termos estão nos três primeiros níveis. São 1165 termos não preferenciais, que fazem com que os termos preferidos correspondam a 88.08% do total de termos. Além disso, são apontadas 111 notas de escopo. Essas informações estavam publicadas na página oficial do Thesagro em 25/09/2022.

O Thesagro ainda não foi ajustado ao Novo Acordo Ortográfico, obrigatório desde janeiro de 2016, de forma que apresenta ditongos acentuados que não se acentuam mais (Exemplo: Nematóide), tremas (exemplo: Frequência Genética) e uso incorreto do hífen (exemplo: Socio-Econômico). Alguns termos estão com acentuação errada, por exemplo, “dicotiledónea”, “Empresa Agrícola” e “Álcalino Terroso”. O termo “hidras te do Canadá” está errado, uma vez que o termo é “Hidraste do Canadá” e o termo “Guacatonga” deveria ter a cedilha e “Motoserra” deveria ser “motosserra” com dois “s”.

O Thesagro, nesta pesquisa, é o mediador especializado entre os artigos científicos da pesquisa agropecuária e o indexador.

Os termos do Thesagro nomeiam os assuntos encontrados no texto e tem uma forma única, enquanto os termos encontrados nos textos podem variar entre masculino, feminino, plural e singular. Isso é um problema no processamento de linguagem natural, porque seus métodos baseiam-se na frequência e na coocorrência dos termos, assim, a contagem das variações terminológicas como um único termo é fundamental.

Para tratar esse problema, nesta pesquisa, optou-se por contar todas as variações de um termo (família semântica) como a frequência de um conceito incluído no Thesagro. Por exemplo: os termos “agricultor”, “agricultora”, “agricultores” e “agricultoras” são contados nos textos como o conceito “agricultor” (termo do Thesagro). Caso contrário, a frequência de ocorrência desse termo não corresponderia à realidade.

Na próxima seção, apresenta-se a etapa 2 do procedimento metodológico: “desenvolvimento do *software*”.

⁷ *Extensible Markup Language*

4.3.3 Etapa 3: Desenvolvimento do software

O planejamento do *software* 3ApC foi inspirado no processo racional unificado de desenvolvimento de *software* (RUP⁸). Inspirou-se no método, mas não o seguiu à risca, porque o RUP foi desenvolvido para grandes projetos com orçamentos elevados que duram anos e são realizados por equipes multidisciplinares (KRUCHTEN, 2004). O ciclo completo de desenvolvimento do 3ApC pode ser visto no Quadro 8.

Quadro 8 – O ciclo completo de desenvolvimento do 3ApC inspirado no RUP.

<p>1 Concepção</p> <p>1.1 Definição do Objetivo e do escopo</p> <p>1.2 Definição dos perfis de usuários</p> <p>2 Elaboração</p> <p>2.1 Requisitos funcionais</p> <p>2.2 Diagramas de casos de usos</p> <p>2.3 Requisitos de sistema</p> <p>3 Construção</p> <p>3.1 Criação do ambiente</p> <p>3.2 Interface Gráfica de Usuário</p> <p>3.3 Implementação dos métodos de inteligência artificial</p> <p>3.4 Informações de suporte à tomada de decisões</p> <p>4 Teste de funcionalidade</p> <p>4.1 Ambiente</p> <p>4.2 Interface Gráfica de Usuário</p> <p>4.3 Métodos de inteligência artificial</p> <p>4.4 Informações de suporte à tomada de decisões</p>

Fonte: Elaborado pelo autor.

As fases do ciclo de desenvolvimento são iterativas (Figura 10), em que o ciclo completo só é encerrado se a fase 4 (teste de funcionalidade) apresentar resultados satisfatórios, caso contrário o ciclo é reiniciado. Com isso, procura-se uma transição suave entre as fases para manter a consistência e a integração da arquitetura planejada para o *software*.

⁸ *Rational Unified Process* patentado pela *Rational Software Corporation*.

Figura 10 – O ciclo completo de desenvolvimento do 3ApC inspirado no RUP.



Fonte: Adaptado pelo autor de Silva (2013).

O desenvolvimento de *software* é feito em ciclos, porque pequenas mudanças em uma das fases podem alterar completamente as outras fases.

Na próxima seção, descreve-se em alto nível a fase de Concepção do ciclo completo de desenvolvimento do *software*.

4.3.3.1 Fase de Concepção

Na fase de **Concepção** do ciclo completo de desenvolvimento, fez-se o alinhamento dos Objetivos do *software* com os objetivos da pesquisa e descreveram-se os perfis de usuários.

Nesta fase, concebeu-se as funcionalidades principais do *software* que correspondem aos estágios do processo de análise de assuntos intelectual (seção 3.1), conforme relacionado no Quadro 9.

Quadro 9 – A correlação entre a análise de assuntos e as metodologias tecnológicas.

Estágio da análise de assuntos intelectual	Metodologia tecnológica (funcionalidade)	Resultados
Ler o texto/compreender o texto	Sumarização automática	Resumo/compreensão do texto
Identificar os assuntos do texto	Expressões regulares	Lista dos assuntos identificados no texto
Selecionar os assuntos relevantes do texto	Modelagem de tópicos	Lista dos assuntos relevantes do texto

Fonte: Elaborado pelo autor.

Assim, os requisitos principais do *software* incluem produzir: (1) um resumo que ajude o indexador compreender do que se trata o texto, (2) a lista dos assuntos identificados no texto, em que os assuntos estão nomeados pelos termos do Thesagro e (3) a lista de assuntos relevantes do texto (antes de ser traduzida na etapa de tradução do processo de indexação), em que os assuntos estão nomeados pelos termos do Thesagro.

Na próxima seção, apresenta-se a fase de elaboração do *software*.

4.3.3.2 Fase de elaboração

Na fase de **elaboração** do ciclo completo de desenvolvimento do *software* foram definidos: (1) os requisitos funcionais, que amparam as funcionalidades do *software*; (2) os casos de usos que demonstram como os requisitos funcionais relacionam entre si e com os usuários; (3) os requisitos de sistema para especificar como o *software* deverá se “comportar” na linha de produção.

Nesta etapa, definiu-se os seguintes recursos computacionais para elaborar o *software*: (1) linguagem de programação: *Python* (seção 4.3.3.2.1); (2) recurso computacional para a sumarização automática: *Library spaCy* (seção 4.3.3.2.2); (3) *corpus* anotado da língua portuguesa para treinar os métodos de inteligência artificial: *pt_core_news_sm* (seção 4.3.3.2.3), e (4) recurso computacional para a modelagem de tópicos: *Library Gensim* (seção 4.3.3.2.4).

Todos esses recursos computacionais são de códigos-fonte abertos, sem restrições de licença quanto ao uso e à distribuição, e estão disponíveis para *download*.

Nos processos de desenvolvimento de *softwares* utilizam-se as *Libraries*, que são módulos computacionais que implementam as funcionalidades genéricas como redes neurais artificiais e modelagem de tópicos. Elas são disponibilizadas para o público em geral.

A vantagem de usar uma *Library* é não ter que desenvolver processos consolidados e obter um recurso computacional de alta qualidade, em que os processos científicos foram implementados, analisados e validados por especialistas.

Na próxima seção, apresenta-se a linguagem de programação *Python*.

4.3.3.2.1 Linguagem de programação *Python*

Os *softwares* são codificados (escritos) em linguagens de programação de alto nível, isto é, usam comandos escritos em inglês, o que facilita escrever, ler e atualizar os códigos dos programas. A escolha da linguagem de programação não afeta a qualidade dos resultados, sendo apenas uma escolha pessoal (e subjetiva) do pesquisador.

A linguagem de programação *Python*, doravante denominada de *Python*, foi desenvolvida sob o conceito de código-fonte aberto, portanto, não se exige qualquer pagamento por sua utilização e não há nenhuma restrição para a distribuição de programas feitos com ela. Assim, os programas desenvolvidos nesta pesquisa podem ser utilizados e distribuídos livremente.

A *Python* se caracteriza por ser de alto nível e de baixo custo de manutenção devido à semântica dinâmica e à orientação a objetos. Além disso, encoraja o uso de pacotes (*package*) e bibliotecas (*library*) prontos, pois isso modulariza as aplicações, facilita a correção e o reuso dos códigos, e reduz os custos de desenvolvimento.

Portanto, a *Python* não interfere na qualidade dos métodos e dos resultados, atuando apenas como um meio de comunicação entre o programador e a máquina.

Na próxima seção, apresenta-se a *Library spaCy*.

4.3.3.2 *Library spaCy*

A *Library spaCy* é um recurso técnico de programação que implementa métodos de processamento de linguagem natural. Trata-se de um projeto de escala mundial, que recebeu inclusive aportes de pesquisas brasileiras na área. Ela está disponível gratuitamente para o processamento de dados textuais em 63 idiomas, incluindo o português.

A *Library spaCy* é definida em sua página oficial como “uma biblioteca de *software* de código aberto para processamento avançado de linguagem natural”, em *Python*.

Na próxima seção, apresenta-se o *corpus* anotado da língua portuguesa para modelos de inteligência artificial.

4.3.3.3 *Corpus anotado da língua portuguesa*

Nesta pesquisa, utilizou-se o modelo de inteligência artificial treinado no *corpus* anotado da língua portuguesa denominado de *pt_core_news_sm*. Trata-se de um recurso computacional *open-source* de elevada qualidade técnica.

O modelo de inteligência artificial treinado no *corpus pt_core_news_sm* está disponível para *download*. Esse tipo de modelo utiliza um grande volume de dados anotados para “aprender” a reconhecer os padrões linguísticos por meio de heurísticas, que são abordagens empíricas, que não carece de explicação matemática para a escolha dos parâmetros (ver seção 3.2.3.4), por isso é classificado como um modelo de aprendizado de máquinas.

A principal utilidade deste modelo, nesta pesquisa, foi viabilizar a sumarização automática. Note que treinar um modelo semelhante ao *pt_core_news_sm* levaria anos, custaria muito e exigiria uma equipe de pessoas para anotar os dados.

Na próxima seção, apresenta-se a *Library Gensim*.

4.3.3.2.4 *Library Gensim*

A *Library Gensim* é um recurso computacional de código-fonte aberto escrito em *Python* que representa documentos como vetores semânticos de forma eficiente (computacional). Ela foi projetada para processar textos digitais brutos (fluxo de caracteres), usando algoritmos de Aprendizado de Máquina Não Supervisionados (seção 3.2.3.3).

Nesta pesquisa, utilizou-se a *Library Gensim* para implementar a modelagem de tópicos, especificamente, a Alocação Dinâmica de *Dirichlet* (LDA).

Na próxima seção, apresenta-se a fase de Construção do modelo de desenvolvimento de *software*.

4.3.3.3 Fase de Construção

Na fase de **Construção** do ciclo completo de desenvolvimento de *software*, foi criado o ambiente (sentido amplo) do 3ApC. Nela foram feitas as escolhas relativas à Interface Gráfica de Usuário (GUI), em que se evitou os excessos de opções para facilitar o acesso aos métodos de inteligência artificial e às Informações de suporte à tomada de decisões.

Na fase de Construção do ciclo completo de desenvolvimento, foram implementados os seguintes métodos de inteligência artificial: (1) sumarização automática (seção 4.3.3.3.1); (2) expressões regulares (seção 4.3.3.3.2); e (3) a modelagem de tópicos (seção 4.3.3.3.3).

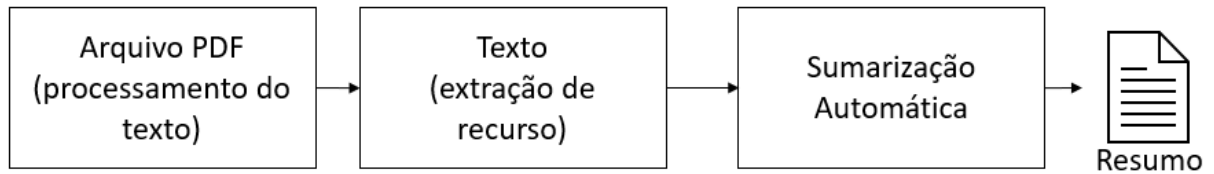
Na próxima etapa, apresenta-se o método de sumarização automática.

4.3.3.3.1 Sumarização Automática

A sumarização automática é a funcionalidade que emula a leitura documentária, porque ela fornece um resumo automático que permite o indexador ter uma noção geral do que trata o documento.

Para a elaboração de um resumo, o arquivo PDF contendo o artigo científico foi submetido no estado bruto à *pipeline* que pode ser visto Figura 11: (1) o Artigo científico (dado bruto) foi submetido ao 3ApC e transformado em um fluxo de caracteres, (2) o fluxo de caracteres foi transformado em um recurso computacional, (3) procedeu-se a sumarização automática sobre o recurso, e (4) produziu-se o resumo.

Figura 11 – A *pipeline* para a sumarização Automática.



Fonte: Elaborado pelo autor.

Essas três etapas são repetidas em todos os processos de processamento de linguagem de natural para a sumarização automática: (1) O arquivo digital contendo o Documento entra como dado bruto e é processado e transformado em um fluxo de caracteres, (2) Do fluxo de caracteres é extraído o recurso informacional por meio de técnicas de limpeza, organização e estruturação dos dados textuais, (3) os recursos informacionais passam pelo processo computacional de sumarização automática que produz o resumo.

O resultado dessa *pipeline* é um resumo. Há muitos métodos de sumarização automática que funcionam bem e não há um que seja melhor do que outro. Além disso, nenhum produz resumos tão bem quanto os seres humanos, mas as máquinas levam muito menos tempo e isso conta muito no ambiente de trabalho.

Na próxima seção, apresentam-se as expressões regulares.

4.3.3.3.2 Expressões regulares

As expressões regulares identificam os assuntos do texto a partir dos termos do vocabulário controlado (Thesagro), para tanto, há dois passos distintos: (1) formulação das expressões regulares (4.3.3.3.2.1), e (2) utilização das expressões regulares para a identificação dos assuntos do texto (seção 4.3.3.3.2.2).

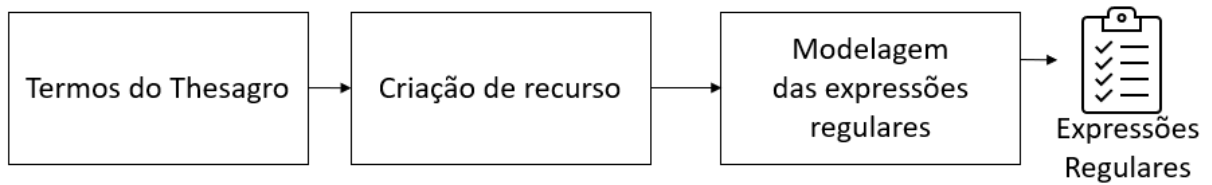
4.3.3.3.2.1 Formulação das expressões regulares

Uma expressão regular é um recurso computacional elaborado a partir dos termos do Thesagro que permite contar todas as variações de um termo no texto analisado.

O primeiro passo para implementar essa funcionalidade é formular as expressões regulares, depois armazená-las na base de conhecimento para serem utilizadas no passo seguinte.

Para elaboração das expressões regulares, os termos do Thesagro são submetidos um-a-um à *pipeline* descrita na Figura 12: (1) os termos passam pela etapa de processamento de texto, (2) criam-se os recursos, e (3) procede-se a modelagem das expressões regulares.

Figura 12 – A pipeline para a formulação das expressões regulares.



Fonte: Elaborado pelo autor.

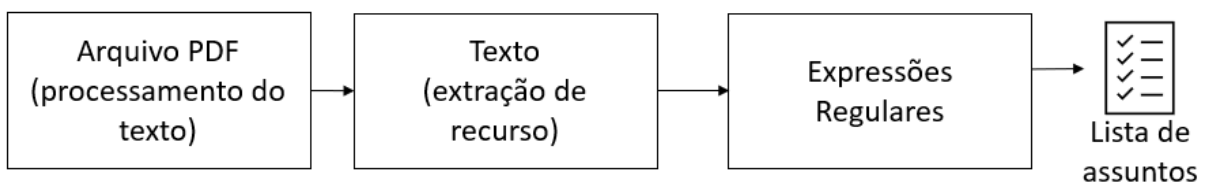
Essas três etapas do processamento de linguagem de natural se repetem para cada termo do Thesagro: (1) um termo entra como dado textual, é processado e transformado em um fluxo de caracteres que é transferido para a “criação de recursos”; (2) a etapa de “criação de recursos” transforma o fluxo de caracteres em um recurso informacional adequado, removendo os sinais diacríticos, trocando as maiúsculas por suas minúsculas correspondentes e delimita-se a terminação do vocábulo, que depende das regras de plural que se aplicam; e (3) na etapa de Modelagem das expressões Regulares, os recursos vindos da “Criação de Recursos” da *pipeline* são transformados em expressões regulares, que são armazenadas na base de conhecimento. O resultado dessa *pipeline* é uma coleção de expressões regulares de todos os termos do Thesagro ordenadas pelo número de palavras do termo (as maiores primeiro).

Na próxima seção, descreve-se a utilização das expressões regulares para a identificação dos assuntos do texto.

4.3.3.3.2 Utilização das expressões regulares

A identificação dos assuntos do texto é feita empregando-se as expressões regulares armazenadas na base de conhecimento conforme a *pipeline* apresentada na Figura 13.

Figura 13 – A pipeline para a identificação dos assuntos do texto.



Fonte: Elaborado pelo autor.

O arquivo PDF contendo o artigo científico é submetido à *pipeline*, em que é processado para a extração do fluxo de caracteres.

O fluxo de caracteres (texto plano) passa pela extração de recursos, em que todas as maiúsculas são trocadas por suas correspondentes letras minúsculas, os sinais diacríticos são removidos e todos os sinais de pontuação são distanciados das palavras por um espaço em branco.

Os recursos criados/extraídos são submetidos às expressões regulares, que identificam e contam todos os termos do Thesagro que estão no texto, incluindo as variações de número e gênero.

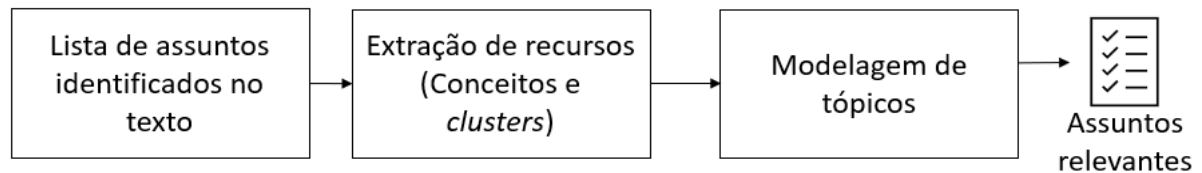
O resultado deste fluxo de processos é a lista de todos os assuntos identificados no texto, ordenada pela frequência do termo e suas variações no texto.

Na próxima seção, apresenta-se a modelagem de tópicos.

4.3.3.3 Modelagem de tópicos

A modelagem de tópicos é utilizada para selecionar os assuntos relevantes do texto, seguindo o fluxo de atividades computacionais descrito no *pipeline* da Figura 14.

Figura 14 – A *pipeline* da seleção dos assuntos relevantes do texto.



Fonte: Elaborado pelo autor.

Na modelagem de tópicos, os assuntos identificados no texto são submetidos ao processo de extração de recursos que consiste em identificar os 20 assuntos (termos do Thesagro) mais proeminentes do texto. Esses assuntos foram denominados de “conceito”. O número 20 é uma escolha de projeto, podendo ser alterado conforme a decisão do usuário.

São considerados assuntos proeminentes do texto aqueles que aparecem no título ou estão tem alta frequência no corpo do texto. Lembrando que o “assunto” é nomeado por um termo no Thesagro.

O termo “conceito” refere-se, portanto, aos 20 assuntos considerados com maior chance de mapear um *cluster* para o artigo na base de conhecimento. Assim, os termos classificados como “conceito” formam uma lista em ordem de maior probabilidade de pertencer ao tópico de um *cluster*, considerando os assuntos identificados no texto, para tanto, um “conceito” é selecionado por uma das seguintes regras:

1. pertence ao título;
2. está entre os 20 assuntos mais frequentes no texto.

Os termos que aparecem no título são incluídos na lista de conceitos, e acrescenta-se à lista os termos mais frequentes até completar 20 conceitos. Por exemplo, se houver apenas um conceito no título, serão acrescentados 19 termos primeiros termos mais frequentes do texto. Note que um “conceito” apenas aponta um caminho com mais chances de ter êxito, estabelecendo uma ordem de busca pelos assuntos relevantes e não afetam a

qualidade dos resultados. Se fossem empregados todos os assuntos do texto a resposta seria a mesma, mas o tempo de processamento seria maior.

A elaboração do tópico é feita pelo algoritmo da inteligência artificial denominado e Alocação Dinâmica de *Dirichlet* ao estabelecer um *cluster*. Nesta pesquisa, considera-se que o termo que pertença ao tópico do *cluster* e ao texto é selecionado como o assunto relevante.

Esse processo é repetido com o próximo conceito, até que se encontre 10 assuntos relevantes no texto ou que toda a lista de “conceitos” seja analisada. Observe que o número 10 foi uma escolha de projeto. O usuário pode escolher outro valor.

Na próxima seção, apresenta-se a fase de teste de funcionalidade do ciclo completo de desenvolvimento de *software*.

4.3.3.4 Fase de Teste de funcionalidade

Na fase de **Teste de funcionalidade** do ciclo completo de desenvolvimento são apresentados os resultados dos testes do Ambiente, do uso da Interface Gráfica de Usuário, do acesso aos métodos de inteligência artificial e às informações de suporte à tomada de decisões.

Os testes aplicados são do tipo *Ad hoc* (LUO, 2001), ou seja, para uma finalidade específica (*Ad hoc*, do Latim, “para esta finalidade”). Esses testes não são estruturados, eles são elaborados para os itens principais da superfície da GUI e testam as funcionalidades quanto a correção, erros de entrada e facilidade de uso.

Os testes foram aplicados como o encerramento de um ciclo completo de desenvolvimento (Figura 10). Foram testados: (1) alinhamento do requisito funcional às exigências dos requisitos de sistema, (2) usabilidade, (3) facilidade de uso (percurso intuitivo), (4) acesso e uso da interface gráfica de USUÁRIO, e (5) coerência do resultado.

Ressalta-se que o planejamento adequado das fases anteriores facilita a aplicação dos testes desta fase.

Em síntese, o *software* foi concebido quanto aos seus objetivos, escopo e necessidades dos usuários, elaborado dentro de um ambiente tecnológico para criar uma GUI adequada ao uso da inteligência artificial e das informações adicionais no processo de tomada de decisão (etapa de tradução).

Na próxima etapa, apresenta-se a etapa de planejamento da análise dos resultados.

4.3.4 Etapa 4: Análise dos resultados

Nesta seção, descreve-se o planejamento de um experimento de validação do 3ApC, utilizando um *corpus* com 100 artigos científicos (amostragem randômica).

No experimento de validação do 3ApC, avaliou-se a sumarização automática, as expressões regulares e a modelagem de tópicos. A principal questão a ser respondida é se os assuntos sugeridos pelo 3ApC são necessários e suficientes para o indexador traduzir os significados do texto em termos do Thesagro.

Na próxima seção, descrevem-se os procedimentos para obter o *corpus* do experimento de validação do protótipo tecnológico.

4.3.4.1 Corpus do experimento de validação

Para avaliar e validar o 3ApC, planejou-se um “ensaio randomizado controlado” (LANCASTER, 2004, p. 126), em que 100 artigos científicos foram selecionados do *corpus* da pesquisa, por um processo aleatório, antes de qualquer procedimento.

O *corpus* do Experimento foi mantido separado de todos os outros processos da pesquisa para preservar as condições reais da linha de produção e, com isso, obter uma amostra com as mesmas características do *corpus* da pesquisa.

Na próxima seção, descreve-se a visão geral do Experimento de validação.

4.3.4.2 Experimento de validação

A análise dos resultados teve o propósito de avaliar e validar o protótipo tecnológico 3ApC, considerando a análise dos dados coletados, a investigação realizada, o desenvolvimento do artefato tecnológico, mediante a aplicação da metodologia. Assim, o planejamento completo do Experimento descrito acima pode ser resumido da seguinte sequência de eventos:

1. *Corpus* do experimento:
 - a. 100 artigos científicos escolhidos ao acaso da base de conhecimento, antes de realizar qualquer procedimento metodológico;
 - b. Análise do *corpus* do experimento para garantir que ele seja uma amostra independente (livre de vieses de escolhas);
2. Sumarização Automática:
 - a. aplicou-se a Sumarização Automática do 3ApC em cada um dos artigos científicos do *corpus* do experimento:

- b. analisou-se a qualidade de todos os resumos obtidos quanto a (1) facilidade de leitura, (2) coerência semântica, e (3) utilidade das informações;
3. Expressões regulares:
- a. Aplicou-se as expressões regulares do 3ApC para identificar todos os assuntos de artigos científicos do *corpus* do Experimento;
 - b. Analisou-se a exaustividade das expressões regulares na identificação dos assuntos dos artigos científicos analisados;
 - c. Comparou-se a exaustividade das expressões regulares com a humana em identificar os assuntos do texto.
9. Modelagem de tópicos:
- a. Aplicou-se a modelagem de tópicos do 3ApC em cada um dos artigos científicos do *corpus* do experimento;
 - b. Comparou-se a lista de descritores do 3ApC com a lista de descritores da BDPA;
 - c. Analisaram-se estatisticamente a porcentagem das listas de descritores consideradas coerentes com a lista de descritores da BDPA.

Foram produzidos resumos pelo método de sumarização automática do 3ApC de artigos científicos do *corpus* do experimento e suas propriedades intrínsecas e extrínsecas foram analisadas.

Os resumos do experimento foram avaliados individualmente quanto às questões listadas no Quadro 10 referentes as seguintes características: (1) facilidade de leitura (inclui estar em português e ter tamanho razoável), (2) coerência semântica e (3) utilidade das informações.

Quadro 10 – Questões norteadoras para a avaliação manual de resumos.

Propriedade	Questões Norteadoras
Facilidade de leitura	As informações não textuais estão em (muito) menor número que as informações textuais? Elementos não textuais superam as informações textuais?
Coerência semântica	As frases são compreensíveis? As frases contêm as relações linguísticas corretas? As informações têm coerência lógica interna?
Utilidade	A informação apresentada é útil para o processo de análise de assuntos?
Idioma	O idioma do resumo está em português?
Tamanho do Resumo	As informações são suficientes para ser útil à análise de assuntos?

Fonte: Elaborado pelo autor.

Os métodos de avaliação seguiram as orientações dos artigos: Gambhir e Gupta (2017), Kiyani e Tas (2017) e Antiqueira *et al.* (2009), conforme descrito nos próximos parágrafos. Segundo Torres-Moreno (2014), os mecanismos da sumarização humana são altamente complexos e ainda não foram compreendidos, de forma que não é possível comparar resumos feitos por pessoas com resumos feitos por máquinas.

A avaliação das características intrínsecas e extrínsecas dos resumos é a forma com que se avalia a qualidade do processo de sumarização de automática.

As expressões regulares são recursos elaborados a partir dos termos dos Thesagro e formam padrões de caracteres que permitem ao algoritmo identificar os termos (assuntos) no texto (ver seção 3.5).

A inspeção uma a uma das expressões regulares geradas a partir dos termos Thesagro não é uma tarefa humanamente realizável. Portanto, foi verificada uma amostragem aleatória de 10 expressões regulares, considerando as seguintes características:

1. termos com menos de 4 letras, porque podem ser confundidos com facilidade com termos vazios de sentido (*stopwords*);
2. termos com terminação em “ão”, devido à multiplicidade de formas plurais;
3. termos compostos; e
4. termos que variam o radical no processo de pluralização.
5. avaliação por amostragem aleatória de artigos científicos do *corpus* do Experimento para verificar manualmente a identificação dos assuntos do texto e comparar com a lista de assuntos obtida por Expressão Regular.

Partindo-se do pressuposto de que não existe uma solução exata para a lista de descritores de um artigo científico, a forma de averiguar a adequação da resposta dos modelos de inteligência artificial (aproximação) é comparar a (aproximação) lista de descritores elaborada pela BDPA e a lista de assuntos relevantes sugerida pela modelagem de tópicos do 3ApC.

Os métodos da inteligência artificial, em geral, incluindo a modelagem de tópicos feita pela Alocação Dinâmica de *Dirichlet*, são métodos empíricos, que fazem inferências diretamente dos dados sem a necessidade de um modelo matemático explicativo (racionalismo). Não há, portanto, testes sistemáticos e precisos para averiguar a correção de modelos inteligentes.

No próximo capítulo, descreve-se a elaboração da base de conhecimento.

5 ELABORAÇÃO DA BASE DE CONHECIMENTOS

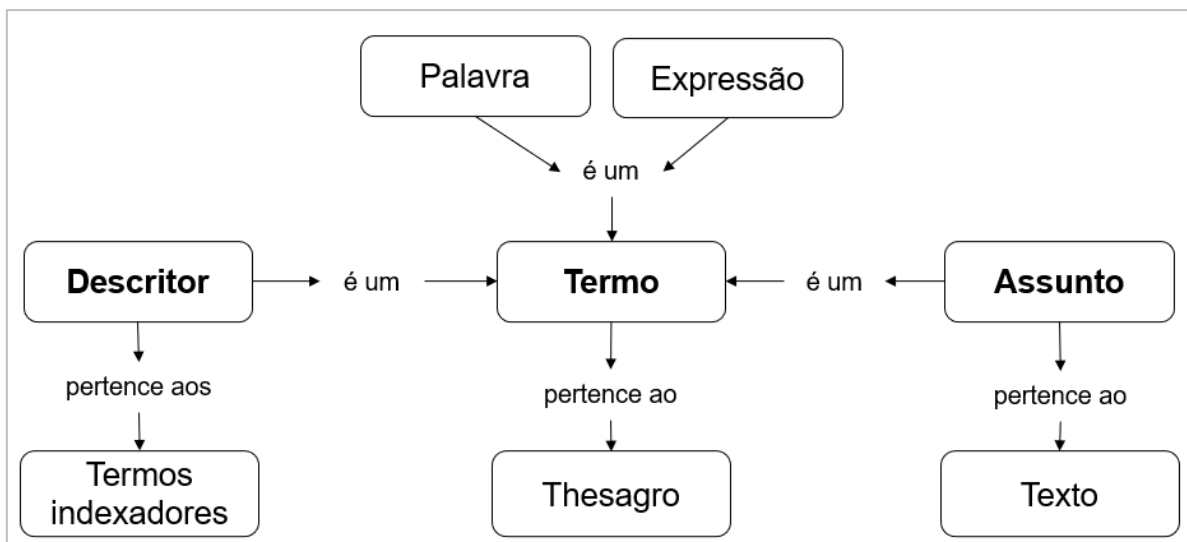
Neste capítulo, descreve-se a etapa 2 dos procedimentos metodológicos (seção 4.3.2). A finalidade da base de conhecimento, nesta pesquisa, é fornecer os insumos necessários para a modelagem da inteligência artificial. Além disso, na inteligência aumentada, os conhecimentos são utilizados para gerar informações adicionais que consubstanciam as decisões tomadas.

A base de conhecimento armazena os “dados brutos” e os “recursos informacionais”, que são os insumos necessários aos algoritmos. O conjunto de dados produzido ao longo de décadas e em diversos locais armazena mais que informações, inclui também o conhecimento. As relações difusas não codificadas explicitamente, mas reconhecidas como padrões apontam o “conhecimento” por trás dessas informações. Por exemplo, a coocorrência frequente dois termos “*Zea Mays*” e “milho” nos descritores de vários artigos, em que um dos termos nem aparece nos textos, indica dois termos fortemente relacionados. A inteligência artificial descobre esse tipo de “conhecimento” e o utiliza para prever “novos conhecimentos”.

Na inteligência aumentada, um ser humano precisa validar um novo conhecimento proposto pela inteligência artificial antes que ele seja armazenado na base de conhecimento.

Neste trabalho⁹, um “assunto” é representado por um termo do Thesagro encontrado no texto e um “descriptor” é representado por um termo do Thesagro utilizado para indexar o artigo científico (Figura 15). Essa terminologia é diferente nos tesauros.

Figura 15 – Mapa conceitual para distinguir “termo”, “assunto” e “descriptor”.



Fonte: Elaborado pelo autor.

⁹ Em um tesouro, um descriptor é um termo preferido.

Como estabelecido nos procedimentos metodológicos (capítulo 4, seção 4.3.2), para a elaboração da base de conhecimento, foram feitos: (1) coleta e seleção dos dados (seção 5.1); (2) análise dos artigos científicos (seção 5.2); (3) análise dos descritores dos artigos científicos (seção 5.3); e (4) análise do *Thesaurus* Agrícola Nacional (seção 5.4).

Espera-se que essas seções sejam suficientes para dar uma visão adequada da elaboração da base de conhecimento, um insumo essencial para elaboração de modelos de inteligência aumentada.

Na próxima seção, apresenta-se a Coleta de dados.

5.1 Coleta e seleção dos dados

Os artigos científicos foram coletados do Sistema de Busca Avançada da Base de Dados da Pesquisa Agropecuária (BDPA), descrita na seção 4.3.2.1, com a expressão de busca¹⁰ do Quadro 11, em 05 de fevereiro de 2021.

Quadro 11 – A expressão utilizada no sistema de busca avançado da BDPA.

Expressão de busca
((embrapa) AND (tipo-material-sigla:"PC" OR tipo-material-sigla:"AP" OR tipo-material-sigla:"DD" OR tipo-material-sigla:"PR" OR tipo-material-sigla:"NT" OR tipo-material-sigla:"RT") AND (idioma:"Português") AND (indicador-meta-id:"1") AND (ano-publicacao:[1850 TO 2022]) AND (ano-meta:[1850 TO 2022]) AND (biblioteca-sigla:"AI-SEDE" OR biblioteca-sigla:"CCPE" OR biblioteca-sigla:"Ebooks" OR biblioteca-sigla:"CPAF-AC" OR biblioteca-sigla:"CNPTIA" OR biblioteca-sigla:"CNPAB" OR biblioteca-sigla:"CNPAE" OR biblioteca-sigla:"CTAA" OR biblioteca-sigla:"CNPAT" OR biblioteca-sigla:"CPAO" OR biblioteca-sigla:"CPAMT" OR biblioteca-sigla:"CNPA" OR biblioteca-sigla:"CPAF-AP" OR biblioteca-sigla:"CPAA" OR biblioteca-sigla:"CPATU" OR biblioteca-sigla:"CNPAF" OR biblioteca-sigla:"CNPc" OR biblioteca-sigla:"CNPc" OR biblioteca-sigla:"CPAC" OR biblioteca-sigla:"CPACT" OR biblioteca-sigla:"CPACP" OR biblioteca-sigla:"CECAT" OR biblioteca-sigla:"CNPf" OR biblioteca-sigla:"CNPgc" OR biblioteca-sigla:"CNPGL" OR biblioteca-sigla:"CNPH" OR biblioteca-sigla:"CNPdia" OR biblioteca-sigla:"CNPmf" OR biblioteca-sigla:"CNPMA" OR biblioteca-sigla:"CPAMN-UEPP" OR biblioteca-sigla:"CPAMN" OR biblioteca-sigla:"CNPMS" OR biblioteca-sigla:"CPAP" OR biblioteca-sigla:"CPPSE" OR biblioteca-sigla:"CPPSUL" OR biblioteca-sigla:"CNPASA" OR biblioteca-sigla:"SIQ" OR biblioteca-sigla:"CENARGEN" OR biblioteca-sigla:"CPAF-RO" OR biblioteca-sigla:"CPAF-RR" OR biblioteca-sigla:"CPATSA" OR biblioteca-sigla:"CNPSO" OR biblioteca-sigla:"CNPS" OR biblioteca-sigla:"CNPS-UEPR" OR biblioteca-sigla:"CNPSA" OR biblioteca-sigla:"CPATC" OR biblioteca-sigla:"CNPm" OR biblioteca-sigla:"CNPT" OR biblioteca-sigla:"CNPuv"))

Fonte: Elaborado pelo autor.

Os documentos coletados cobrem o período de 1998 a 2020 e, como pode ser visto na Tabela 2, somam 7377 arquivos digitais distribuídos da seguinte forma: (1) 1535 arquivos de resumos com menos de 10 kB de dados, portanto sem dados suficientes para uma análise computacional; (2) 804 arquivos com problemas (não podiam ser processados

¹⁰ O sistema da BDPA foi modificado em janeiro de 2022 e, portanto, esta expressão de busca pode produzir resultados diferentes em pesquisas realizadas após esta data.

adequadamente); e (3) 5025 arquivos digitais aptos para a serem utilizados nos modelos de inteligência artificial.

Tabela 2 – Estatísticas relativas à seleção dos dados.

Arquivos	Quantidade
Arquivos coletados	7377
Arquivos com menos de 10 kB de dados	1535
Arquivos com problemas	817
Arquivos adequados para a pesquisa	5025

Fonte: Elaborado pelo autor.

Os problemas encontrados em 817 arquivos incluíam: (1) erros na codificação do arquivo; (2) baixa qualidade da imagem (*scanner*); (3) criptografia; e (4) ausência de descritores.

Dos 7377 artigos científicos coletados, foram selecionados: (1) 4925 artigos científicos e seus respectivos descritores para formar o *corpus* da pesquisa, e (2) 100 artigos científicos e seus respectivos descritores para formar o *corpus* do experimento.

Na próxima seção, descrevem-se os princípios para a análise, seleção, armazenamento e distribuição dos dados.

5.1.1 Os princípios de análise, seleção, armazenamento e distribuição dos dados

Os dados coletados do acervo da Base de Dados da Pesquisa Agropecuária (BDPA) da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) são derivados de pesquisas científicas e estão disponíveis para acesso público, cópia, uso e distribuição, resguardando-se os direitos autorais.

Os dados coletados e os produzidos por esta pesquisa estão disponíveis como dados abertos, em consonância à noção de promover o conhecimento como um bem comum em que todos podem participar, seguindo os padrões FAIR¹¹ de encontrabilidade, acessibilidade, interoperabilidade e reusabilidade (HENNING *et al.*, 2019), descritos a seguir:

1. **Disponibilidade e acesso:** os dados estarão disponíveis integralmente, sem custo e em formato de texto, que os tornam acessíveis e de fácil utilização.
2. **Reutilização e redistribuição:** Os dados desta pesquisa são livres para a reutilização e redistribuição, incluindo a sua combinação com outros conjuntos de dados.

¹¹ Sigla para os termos em inglês: *Findability, Accessibility, Interoperability e Reusability*.

3. **Participação universal:** Todos os dados desta pesquisa podem ser usados, reutilizados e redistribuídos sem restrições para o uso comercial, desde que não haja qualquer tipo de discriminação seja contra áreas de atuação, finalidade, pessoas ou grupos.

Cada arquivo digital selecionado continha um (e somente um) artigo científico, para o qual coletaram-se os seguintes metadados: (1) o identificador do arquivo na BDPA, (2) o título, e (3) os descritores baseados no Thesagro.

Na próxima seção, apresenta-se uma análise dos descritores dos artigos científicos selecionados.

5.2 Análise dos descritores dos artigos científicos selecionados

Os artigos científicos selecionados para a base de conhecimento estavam indexados na BDPA por meio dos termos do Thesagro e, portanto, associados aos seus respectivos descritores.

Os metadados coletados podem ser agrupados em duas categorias: (1) características do artigo (título e identificador), e (2) características do conteúdo do artigo (descritores). Enquanto a primeira caracteriza o artigo do ponto de vista do contexto do autor e é imutável, a segunda expressa a visão do indexador sobre o conteúdo do artigo na perspectiva do usuário e pode mudar a cada vez que o artigo for indexado.

Os descritores foram analisados em conjunto, como um índice, quanto à frequência do descritor no texto (seção 5.2.1), ao número de descritores por artigo científico (5.2.2) e quanto aos termos coocorrentes nos descritores (seção 5.2.3).

Na próxima seção, apresenta-se uma análise das indexações quanto a frequência do termo no texto.

5.2.1 Análise de frequência

A indexação dos artigos científicos selecionados para a base de conhecimento foi realizada ao longo de mais de duas décadas em diversas bibliotecas físicas, por muitas pessoas (logo, culturas diferentes) e sob diversas políticas de indexação. Assim, os descritores refletem essa multiplicidade nas seguintes características apresentadas na Tabela 3.

Tabela 3 – A distribuição dos assuntos, termos e descritores.

Lista de descritores	Número de artigos	Porcentagem
Tem pelo menos um termo do texto	3348	66%
Nenhum termo pertence ao texto	1730	34%
Apenas um descritor aparece no texto	1026	20%
Um descritor aparece uma única vez no texto	598	11%
Um descritor é o termo mais frequente no texto	517	10%

Fonte: Elaborado pelo autor.

Dessa análise, pode-se afirmar que o processo automático que atribuísse o termo mais frequente como descritor, acertaria apenas 10% das vezes.

Essas escolhas dos descritores refletem as visões, as épocas e os locais de indexações das pessoas indexadoras e de como elas perceberam a realidade e o cenário cultural que as envolviam. Não se trata de escolhas determinísticas, nem de respostas exatas, mas são as melhores traduções do significado do texto em necessidades de informações dos usuários.

Na próxima seção, apresenta-se a análise do número de descritores por artigo científico.

5.2.2 Análise do número de descritores

A variação do número de descritores é uma fonte de ruídos nos dados. Nos descritores analisados, o número de descritores por artigos científicos variou entre 1 (um) e 18 termos. A distribuição porcentual dos descritores por artigos científicos pode ser vista na Tabela 4. Note que 20% dos artigos científicos foram indexados com apenas um descritor, que 22% tinham dois descritores, e que 58% dos artigos científicos foram indexados com três ou mais descritores.

Tabela 4 – A distribuição porcentual do número de descritores por artigo científico.

Número de descritores	Número de artigos	Porcentagem (%)
1	1026	20
2	1112	22
3	980	19
4	811	16
5	534	11
6	296	5
7	162	3
8	79	2
> 9	78	2

Fonte: Elaborado pelo autor.

A variação do número termos descritores aponta a multiplicidade de orientações nas políticas de indexação e as modificações inevitáveis que o tempo e a tecnologia exerceram sobre as atividades documentárias das bibliotecas.

Na próxima seção, apresentam-se as informações sobre termos coocorrentes nos descritores.

5.2.3 Os termos coocorrentes nos descritores

Quando se considera o conjunto de descritores de todos os artigos da base de conhecimento, verifica-se a existência de termos coocorrentes, que são os termos que com frequência aparecem juntos nos descritores. Por exemplo, é comum a coocorrência, nos descritores, do nome comum e do seu respectivo nome científico.

O termo comum “milho” aparece em 53% das 205 indexações que contêm o seu nome científico “*Zea Mays*”. Essa coocorrência frequente nos descritores nem sempre ocorre nos textos. Mas considerando que os usuários das bibliotecas da Embrapa sejam pesquisadores, o nome científico nas expressões de busca para encontrar um documento é uma prática provável. No entanto, aqui entra outra questão, que é o controle de vocabulário (equivalências). No caso dos termos “milho” e do “*Zea Mays*”, ambos são termos preferidos¹², ou seja, é permitido que os documentos sejam indexados por eles. Então, toda vez que um artigo for indexado por um termo que tenha um nome científico (planta ou animal), os dois termos deverão ser utilizados. Assim, é possível estimar o nome científico (ou nome comum) a partir da coocorrência nos descritores.

Como o indexador não precisa saber os mais de 2000 nomes científicos do Thesagro, essa pode ser uma informação adicional (suplementar) valiosa para a etapa de tradução.

Coocorrências de termos não são triviais nem para as pessoas, nem para as máquinas, além disso podem aparecer coocorrências não generalizáveis. Por exemplo, “tanque” e “piscicultura” são coocorrentes e “tanque” e “leite” também. Numa proposta metodológica de inteligência aumentada, a inteligência humana inserida no processo distingue bem essas coocorrências.

Na Tabela 5, apresenta-se uma amostra com cinco pares de termos coocorrentes nos descritores da base de conhecimento.

¹² Ambos são descritores em um tesouro.

Tabela 5 – Amostra de termos coocorrentes nos descritores da BDPA.

Termo	Termo coocorrente	Coocorrência
Zea Mays	Milho	53%
Milho	Zea Mays	62%
Alimento	Nutrição Animal	76%
Abelha	Polinização	54%
Anticorpo	Doença Animal	58%

Fonte: Elaborado pelo autor.

Essas coocorrência reforçam os Pressupostos da seção 1.3, de que o assunto de um texto pode ser deduzido do contexto, da frequência e da coocorrência dos termos, porque essas estatísticas apontam relações semânticas. Além disso, elas permitem sugerir termos de indexação que não estejam no texto (transcende o texto), mas que são relacionados ao significado do texto.

Na próxima seção, apresenta-se uma análise do *Thesaurus* Agrícola Nacional.

5.3 Análise do *thesaurus* agrícola nacional

Os termos do Thesagro em conjunto representam os assuntos pesquisados no domínio da agropecuária nacional e foram considerados recursos informacionais valiosos para esta pesquisa por três razões: (1) qualidade da fonte (BDPA), (2) conhecimento acumulado, e (3) especialização no domínio analisado.

A estrutura hierárquica do Thesagro explicita as relações entre termos para as pessoas, mas essa facilidade visual não se aplica aos algoritmos. Por exemplo, a relação entre um nome científico e um nome comum não é simbolicamente assinalada e passa despercebida dos algoritmos. Por isso, foram elaborados algoritmos para identificar as seguintes relações: (1) entre o nome científico e seus respectivos nomes comuns; (2) entre o termo não preferido e seu respectivo termo preferido; e (3) entre os artigos científicos com os mesmos assuntos (que formam um *cluster*).

Os termos do Thesagro mapeados no texto identificam os assuntos do texto e a frequência indica a importância do assunto/termo para o significado do texto. O problema desta abordagem é que, no texto, os termos do Thesagro não possuem uma forma única.

Os textos estão na linguagem natural português, em que pode coexistir até quatro variações de um mesmo conceito: (1) masculina, (2) feminina, (3) singular e (4) plural. Por exemplo: fruta, fruto, frutas e frutos. Portanto, um termo aponta um assunto, mas um Assunto não é necessariamente constituído por apenas um Termo, pode incluir os sinônimos, as quatro variações citadas, o termo preferido, o termo não preferido, o nome científico e o nome comum. Neste sentido, foi necessário criar uma metodologia para tratar essas variações de um mesmo conceito nos textos (seção 5.3.1).

5.3.1 Problema dos plurais

Em português, exceto os nomes próprios e os científicos que não possuem variações, existem várias regras para a formação de plurais, baseadas na terminação da palavra. No Quadro 12 podem ser vistas as regras necessárias aos termos do Thesagro. Essas regras não se aplicam aos verbos.

Quadro 12 – Regras de formação do plural necessárias para o Thesagro.

Terminação da palavra	Regra de formação do plural
Vogal, ditongo oral, N	Acrescentar “S” ao final da palavra
M	Substituir por “NS”.
R, Z	Acrescentar “ES”.
AL, EL, OL, UL	Substituir o “L” por “iS”.
iL	Substituir o “L” por “S”.
S	Substituir por “ES” nas monossílabas e oxítonas.
ÃO	Substituir por “ÕES”, ou “ÃES”, ou “ÃOS”
X	Não varia

Fonte: Elaborado pelo autor.

Por questões de eficiência, as regras de formação do plural foram selecionadas para atender especificamente o Thesagro. Dessas regras, depreende-se que é necessário verificar a última letra de um termo para estabelecer a sua forma plural. Se a palavra terminar em “U”, “S”, “O” ou “L” é necessário verificar o penúltimo carácter também e para as palavras terminadas em “ÃO”, ou ditongo decrescente terminado em “U”, há mais de uma regra, são os “plurais irregulares”.

A característica principal de um plural é a presença da letra “s” acrescida no final da palavra, com raríssimas exceções, por exemplo, a forma plural de “campus” é “campi”. No processamento de linguagem natural, o “problema dos plurais” ocorre durante a contagem dos termos (cálculo da frequência e da coocorrência dos termos), uma vez que o singular e o plural de um termo apontam um só conceito.

Os termos compostos apresentam um desafio adicional, pois não há um padrão geral aplicável a todos eles. Pode ocorrer de alguns termos irem para o plural, a depender do contexto, e outros, não. Por exemplo, o termo “etapa de desenvolvimento da planta” tem, pelo menos, as seguintes formas de plurais:

- a. “etapas de desenvolvimento da planta”;
- b. “etapa de desenvolvimento das plantas”; e
- c. “etapas de desenvolvimento das plantas”.

Para calcular a frequência de um termo no texto, o desafio é encontrar todas as variações do termo (singular, plural, feminino e masculino) e reduzi-las a forma que está no tesouro utilizado, e, então, contar as ocorrências. Para tanto é preciso identificar todos os plurais considerando todos os seguintes subconjuntos de termos do Thesagro (Tabela 6).

Tabela 6 – Subconjuntos de termos do Thesagro.

Termo	Quantidade	Exemplos
Vocábulo	4339	Ampola, da, erva, mate
Simples	3946	Anticorpo, aroeira
Composto	5759	Etapa de desenvolvimento animal
Com menos de 4 letras	50	Ar, lâ, mar
Singular terminado em “s”	172	Arco-íris, antivírus, francês
Nome científico	2186	<i>Apis mellifera</i>
Admite quatro variações	38	Agricultor, agricultores, agricultora, agricultoras

Fonte: Elaborado pelo autor.

As variações de um termo em um artigo científico podem ser reduzidas a sua forma básica por meio da técnica das expressões regulares (ver seção 3.5). Um exemplo: o termo “irrigação por pote de barro” e suas variações podem ser encontrados no texto partir da seguinte expressão regular: “***irrigac(ao|aos|aes|oes) por pot(e|es) de barr(o|os)***”, que testa mais de 16 variações, incluindo formas plurais inexistentes.

As expressões regulares apresentam certas limitações devido a ambiguidades e polissemias da linguagem natural. Por exemplo: (1) termos “lã” e “lá”, devido à normalização que retira os sinais diacríticos, ficam iguais: “la”; (2) O termo “dado” é polissêmico, que pode significar “medição” ou pode ser um pronome indefinido, ou uma forma verbal, ou um adjetivo ou um substantivo comum (um objeto); e (3) O termo “cerca” que significa uma “construção rural”, pode aparecer no texto como parte de uma locução prepositiva.

Na próxima seção, apresenta-se a base de conhecimento elaborada.

5.4 A base de conhecimento elaborada

A base de conhecimento elaborada contém os seguintes dados e informações coletadas da BDPA: (1) 7377 artigos científicos e seus respectivos descritores; (2) 9857 termos do *Thesaurus* Agrícola Nacional (Thesagro); (3) 12068 relações entre termos do Thesagro.

Os artigos científicos estão distribuídos em: (1) *corpus* da pesquisa com 6973 artigos científicos; e (2) *corpus* do experimento com 100 artigos científicos.

A partir dos dados e informações coletados, foram elaboradas as seguintes informações adicionais (conhecimentos): (1) 9857 expressões regulares; (2) 82 termos coocorrentes nos descritores; (3) 2822 nomes científicos e seus respectivos nomes comuns; (4) 1155 termos não preferidos e seus respectivos termos preferidos; e (5) 363 *clusters*.

Os dados, informações e conhecimentos foram armazenados na base de conhecimento como arquivos de dados separados por ponto-e-vírgula (extensão “CSV”), o que inclui: (1) termos do Thesagro, (2) expressões regulares, (3) lista de termos coocorrentes, e (4) lista de nomes científicos. Os resultados produzidos: (1) resumo, (2) lista de assuntos e (3) lista de assuntos relevantes podem armazenadas (salvas) pelo usuário como documentos (*doc*). O *corpus* foi armazenado em três formas: (1) PDF, (2) texto (TXT) e (3) CSV.

Os arquivos em PDF foram armazenados para ter o acervo no estado original. Os arquivos-textos permitem, ao refazer a análise de assuntos, que o processo seja mais rápido porque a etapa de extração do texto já está efetuada. O formato CSV armazena na forma de banco de dados as informações intermediárias, como os assuntos identificados no texto e os descritores, para dar mais celeridade aos modelos de inteligência artificial.

No próximo capítulo, descreve-se a etapa 3 dos procedimentos metodológicos: desenvolvimento do *software*.

6 DESENVOLVIMENTO DO SOFTWARE

Neste capítulo, descreve-se a Etapa 3 dos procedimentos metodológicos (seção 4.3.3), em que os termos “*software*”, “sistema computacional”, “programa”, “artefato tecnológico” e “3ApC” foram usados como sinônimos e de forma intercambiável.

O desenvolvimento do *software* de Análise de assuntos Assistida por Computador (3ApC) inspirou-se no processo de desenvolvimento de *software* RUP¹³, conforme descrito nos procedimentos metodológicos na seção 4.3.3.

O 3ApC foi desenvolvido na linguagem de programação *Python* e todos os seus códigos-fontes e dados estão disponíveis na Internet para *download*¹⁴ sob os princípios FAIR (seção 5.1.1). Assim, o artefato tecnológico desenvolvido é um *software* de código-fonte livre e aberto com permissão para ser copiado, modificado e distribuído livremente sem necessidade de uma licença.

As etapas do ciclo completo de desenvolvimento do 3ApC são padronizadas pelo RUP (seção 4.3.1). O desenvolvimento ocorreu por um processo iterativo e incremental, em que, a cada ciclo de desenvolvimento, uma versão completa do *software* foi produzida e avaliada. Se fossem identificadas deficiências ou necessidades de melhorias em qualquer fase, todo o processo de planejamento era revisado. Quando todos os requisitos foram satisfeitos, o processo foi finalizado. Os detalhes sobre a codificação em linguagem de computador não foram descritos, teve-se as funcionalidades em alto nível.

Na próxima seção, descreve-se as atividades da etapa de concepção.

6.1 Concepção

Nesta seção, apresentam-se a visão geral, o escopo, o objetivo e o perfil do usuário da fase de Concepção do desenvolvimento do *software* 3ApC, conforme foi definida na Metodologia (seção 4.3.3.1).

No diagrama da Figura 16, apresentam-se dois percursos possíveis para indexador durante o uso do 3ApC: (1) análise de assuntos intelectual; e (2) análise de assuntos assistida por computador.

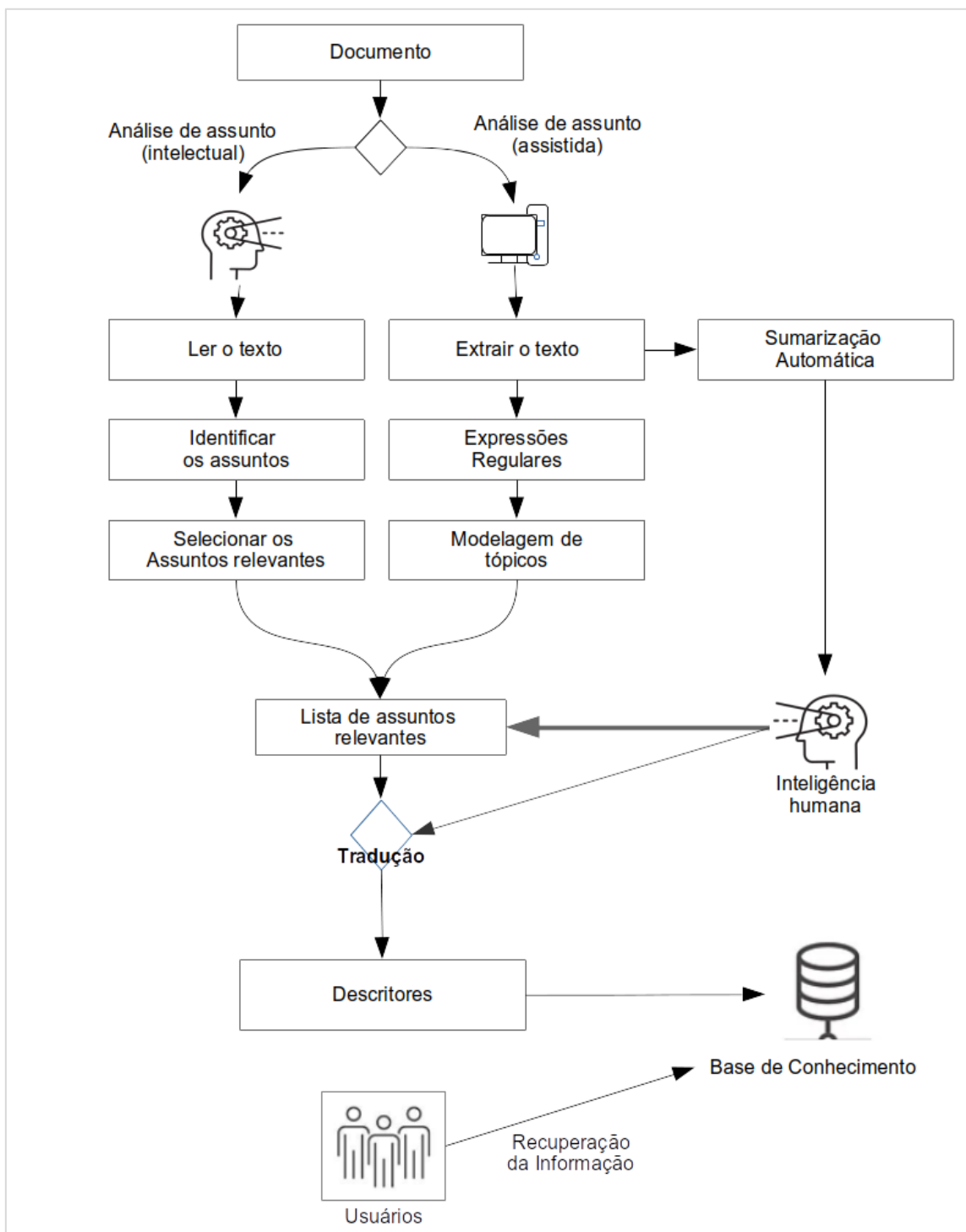
No percurso 1, o usuário pode utilizar o 3ApC para realizar as tarefas da maneira tradicional em que se lê o texto para saber do que se trata o texto e para identificar os assuntos

¹³ *Rational Unified Process* patenteado pela *Rational Software Corporation*.

¹⁴ <https://github.com/3ApC>

do texto. Em seguida, seleciona os assuntos relevantes a partir dos assuntos identificados no texto e passa para etapa de tradução do processo de indexação.

Figura 16 – A visão geral e conceitual do 3ApC¹⁵.



Fonte: Elaborado pelo autor.

¹⁵ Os “assuntos relevantes” e os “descritores” são representados por “termos” do tesauro (ver Figura 15).

No percurso 2, o usuário opta pela assistência do 3ApC, então, obtém o resumo para se inteirar do que se trata o documento, a lista de assuntos identificados no texto e a lista de assuntos relevantes, e acesso a diversas informações suplementares. Com essa assistência, ele pode tomar as decisões (na fase de tradução do processo de indexação) quanto à representação da informação.

Como pode ser visto no Quadro 13, todos estágios da Análise de assuntos Intelectual têm tarefas correspondentes na Análise de assuntos Assistida por Computador, que também produzem os mesmos resultados.

Quadro 13 – A correspondência entre os estágios da análise de assuntos.

Estágio	Análise de assuntos Intelectual	Análise de assuntos Assistida por Computador	Resultado do estágio
0	-	Extrair o texto	Texto plano
1	Ler o texto/compreender o texto	Sumarização automática	Resumo
2	Identificar os assuntos do texto	Expressões regulares	Lista de assuntos
3	Selecionar os assuntos relevantes	Modelagem de tópicos	Lista de assuntos relevantes
4		Funcionalidades complementares	Informações adicionais

Fonte: Elaborado pelo autor.

Note que ao “Ler o texto/compreender o texto”, o usuário tem noção do que trata o documento, da mesma forma, no processo assistido, o indexador tem acesso ao resumo por sumarização automática para saber do que trata o documento, porém com uma quantidade menor de informações a ser analisada. As “expressões regulares” identificam todos os assuntos do texto, e, dessa lista, a “modelagem de tópicos” seleciona os assuntos relevantes. De forma que ao final da etapa análise de assuntos, os processos produzem o mesmo tipo de lista.

Na análise de assuntos assistida por computador do 3ApC, o indexador tem acesso aos resultados da inteligência artificial e às informações adicionais que, combinados, facilitam a tarefa de tradução (a etapa seguinte). Nos dois processos (intelectual e assistido), a etapa de tradução é a mesma, pois é o momento em que o indexador traduz os assuntos relevantes em termos do Thesagro, elaborando, assim, a lista de termos indexadores (descritores). O que diferencia os dois processos (intelectual e assistido) é que, no assistido, o indexador conta com mais informações e resultados com menos esforços intelectuais e tempo.

Na próxima seção, apresenta-se a definição do objetivo e do escopo do 3ApC.

6.1.1 Definição do objetivo e do escopo

O 3ApC tem por objetivo ser uma Interface Humano-Computador (HCI) que utiliza métodos da inteligência artificial para assistir o profissional da informação (usuário) na indexação.

Quanto ao escopo, o 3ApC é uma ferramenta que fornece meios técnicos, sugestões, recomendações e relatórios, mas é o ser humano quem toma as decisões. Uma Descoberta de Conhecimento (KDD) será adicionada a base de conhecimento depois de aprovada pela inteligência humana. Reforça-se que o objetivo do *software* é assistir o processo de tomada de decisões na indexação.

Na próxima seção, descreve-se o perfil esperado do usuário do sistema.

6.1.2 Definição do perfil de usuário

O 3ApC é um protótipo computacional de finalidade bem definida, logo não terá interface complexa, nem será necessário conhecimento avançado no uso *softwares*. No entanto, ele produz resultados que exigem formação acadêmica em Biblioteconomia e Ciência da Informação para interpretá-lo, isto é, espera-se que o sistema seja utilizado por um bibliotecário.

Considera-se que o usuário exerça a função de profissional da informação, tendo treinamento e formação adequados para compreender e realizar uma análise de assuntos, porque o 3ApC faz sugestões técnicas e é o usuário quem vai tomar as decisões.

Na próxima seção, apresenta-se o percurso de elaboração do 3ApC quanto aos requisitos necessários.

6.2 Elaboração

Nesta seção, detalha-se a arquitetura do sistema por meio do levantamento dos requisitos funcionais, dos recursos do sistema e de diagramas.

Na próxima seção, apresentam-se os requisitos funcionais.

6.2.1 Requisitos funcionais

Os requisitos funcionais levam em conta as questões técnicas e não entram em detalhes de atividades básicas como criar, abrir, salvar ou excluir arquivos. Supõe-se que o

usuário do sistema tenha formação acadêmica de nível superior e seja fluente no uso dessas funcionalidades.

O 3ApC foi desenvolvido em um único eixo metodológico que coordena a arquitetura do sistema para atender aos seguintes requisitos funcionais:

1. processar arquivos na extensão PDF:
 - abrir o arquivo e reconhecer a codificação dos caracteres.
2. extrair o texto do arquivo PDF:
 - extrair o fluxo de caracteres do arquivo selecionado.
3. Normalizar e regularizar o texto extraído:
 - excluir caracteres de controle, as marcações de estilo e as informações espúrias.
4. reconhecer e contar os termos do Thesagro presentes no texto:
 - Verificar quais os termos do Thesagro estão presentes nos textos, incluindo as variações de gênero e de número;
 - contar quantas vezes cada termo repete no texto; e
 - elaborar a Lista desses termos com suas respectivas frequências.
5. identificar os nomes científicos:
 - identificar na Lista de termos, quais são nomes científicos; e
 - identificar os respectivos nomes comuns associados ao nome científico.
6. identificar os nomes comuns que tenham nomes científicos:
 - identificar na Lista de termos, os nomes comuns que tenham nomes científicos; e
 - identificar os respectivos nomes científicos associados aos nomes comuns.
7. identificar os termos não preferidos:
 - identificar na Lista de termos, os termos não preferidos e associá-los aos termos preferidos.
8. sumarizar o texto:
 - produzir automaticamente um resumo (por extração) do texto;
 - extrair os termos mais frequentes no texto.
9. selecionar os assuntos representativos do texto:
 - aplicar os métodos de inteligência artificial disponíveis para identificar e selecionar os termos representativos do texto candidatos a descritores.

10. elaborar um relatório com todas essas informações:
 - exibir um relatório contendo o título do artigo, o resumo, a lista de nomes científicos e seus respectivos nomes comuns, a lista de termos não preferidos e seus respectivos nomes preferidos, a lista com 10 termos candidatos a descritores (cada termo deve apresentar, se houver, o respectivo nome científico, nome comum e termo não preferido, além da frequência do termo no texto e quantos textos o termo indexa na base de conhecimento) e a lista de termos presentes no texto ordenados (ordem descendente) pela frequência que aparecem no texto.
11. editar o texto extraído:
 - editor de texto para o usuário editar o texto extraído;
 - salvar o texto editado; e
 - replicar o processo de análise de assuntos no texto editado.
12. navegar na estrutura hierárquica do Thesagro:
 - navegador na estrutura hierárquica do Thesagro para facilitar ao usuário a fase de tradução.
13. gerenciar a base de conhecimento:
 - sistema de gerenciamento na base de conhecimento em que se possa incluir, excluir, editar e buscar conhecimentos na base.
14. fazer análise visual de dados:
 - meio visual para a realização de análise de agrupamento; e
 - nuvem de palavras.

O relatório produzido pelo sistema contém o resumo automático (Requisito funcional 8), os termos mais frequentes no texto, a lista de termos não preferidos e a sugestão dos termos preferidos em relação a eles, uma lista de nomes científicos encontrados no texto e seus respectivos nomes comuns, e a lista de termos sugeridos para a indexação. Apresenta também a lista de conceitos que melhor representam o texto a partir das regras e a lista de descritores presentes no texto e a sua frequência.

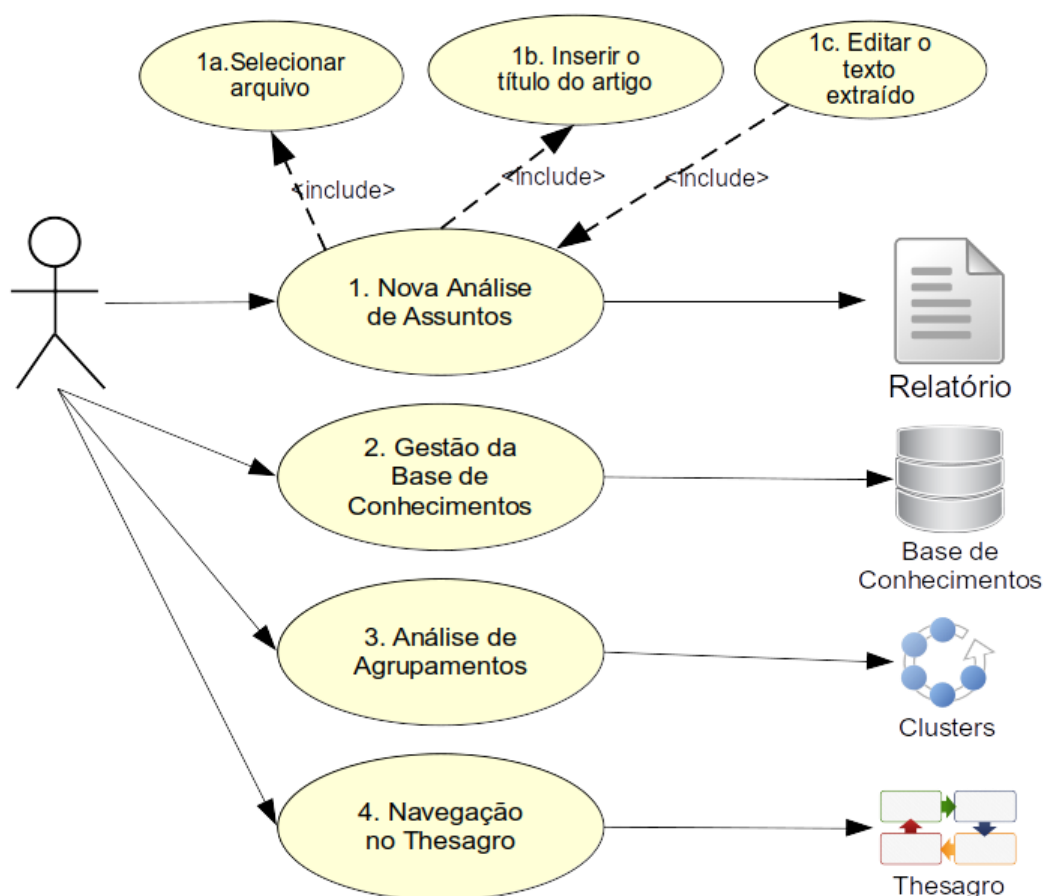
Os mesmos requisitos funcionais do sistema estão apresentados, na próxima seção, por meio de diagrama de casos de uso.

6.2.2 Diagramas de casos de uso

Nesta seção, estão apresentadas as funcionalidades por meio de diagramas de casos de usos (Figura 17), sem descer aos detalhes técnicos, estatísticos ou informáticos. O diagrama facilita a visualização das tarefas a serem executadas pelo usuário sem mostrar a complexidade computacional necessária para produzir o relatório.

Não foram descritas as funcionalidades elementares como abrir, salvar, excluir e imprimir arquivos, por serem fáceis de serem aprendidas/descobertas intuitivamente por um usuário com a formação em Biblioteconomia.

Figura 17 – As funcionalidades apresentadas em um diagrama de casos de uso.



Fonte: Elaborado pelo autor.

O usuário do 3ApC após selecionar a funcionalidade 1 (“nova análise de assuntos”) é solicitado a apontar o arquivo a ser analisado (e clicar em “abrir”), em seguida, é pedido o título do artigo. A partir desse ponto, o sistema extrai os conceitos do texto, seleciona os conceitos e acessa a base de conhecimento para identificar os *clusters* relacionados ao

texto. Assim, o 3ApC faz a modelagem de tópicos, elabora e apresenta o relatório (Requisito funcional 10).

Além de visualizar o relatório e consultar as informações a respeito das sugestões oferecidas pelo *software*, o usuário poderá editar o texto extraído e, a partir do texto editado, solicitar “uma nova análise de assuntos” (Funcionalidade 1).

O usuário pode também gerenciar a base de conhecimento (Funcionalidade 2), ou seja, pode incluir, excluir, editar e pesquisar nela. Além disso, poderá obter análises estatísticas dos termos de indexações existentes, como quantos artigos um termo indexa, ou quais termos aparecem juntos com frequência nos índices.

Poderá também analisar os *clusters* a que o artigo pertence (Funcionalidade 3), visualizar um mapa de tópicos, a nuvem de palavras e, assim, estará bem municiado com informações suplementares que o assista na tomada de decisões.

O usuário poderá navegar pelo Thesagro para localizar um termo, visualizar sua estrutura hierárquica (Funcionalidade 4).

Usando uma analogia, basta “girar a chave” e deixar o resto por conta do “motor sob o capô” para se ter uma análise de assuntos. A simplicidade de uso não deve obscurecer o potencial do *software*. Afinal, o 3ApC oferece processamento computacional cognitivo, projetado organicamente para auxiliar (e não para substituir) o profissional da informação.

Na próxima seção, apresentam-se os requisitos do sistema.

6.2.3 Requisitos de sistema

Para que um *software* se torne útil na linha de produção, ele precisa atender a um conjunto de requisitos denominados na literatura de “requisitos não funcionais” ou “requisitos de sistema”, que é uma lista de ferramentas auxiliares para gerenciar o *software*. No 3ApC, a lista de requisitos de sistema inclui:

1. Pacote (*package*) de instalação:
 - Linux;
 - Unix;
 - Windows.
2. Licença:
 - aberta;
 - gratuita;
 - livre para modificar e redistribuir.

3. Usabilidade:

- simples;
- rápida;
- flexível com uma boa curva de aprendizado.

4. Documentação:

- abrangente.

O pacote de instalação (Requisito de sistema 1) permite instalar o 3ApC como versão *desktop* nos sistemas operacionais Linux, Unix e Windows sem exigir conhecimentos avançados em informática.

A licença de *software* (Requisito de sistema 2) é da modalidade livre, gratuita e aberta, de forma que o 3ApC pode ser usado, modificado e redistribuído livremente.

O 3ApC foi desenvolvido para realizar tarefas complexas, desempenhando funções semelhantes às da cognição humana, o que não significa que a sua usabilidade seja complexa, pelo contrário, é intuitiva.

Seguindo esse raciocínio, planejou-se um 3ApC em que basta o usuário, (se preferir), usar apenas o *mouse*. De forma que, para ter assistência na análise de assuntos, basta selecionar (clique em) “nova análise de assuntos”, apontar o arquivo (clique em “Abrir”) e selecionar (clique em) as linhas do texto que compõem o título do artigo, (clique em “OK”); e todo o processo será efetuado. Espera-se, portanto, que o usuário do sistema tenha apenas discernimento profissional para selecionar a melhor resposta.

Para ampliar a compreensão das funcionalidades do 3ApC, na próxima seção, está delineada e explicada a Interface Gráfica de Usuário (GUI).

6.3 Elaboração

Todos os recursos computacionais utilizados para elaborar o 3ApC estão listados no APÊNDICE B. Os recursos computacionais relacionados a qualidade dos resultados estão descritos e analisados na fundamentação teórica e metodológica (Capítulo 3).

6.3.1 Criação do ambiente

O ambiente do 3ApC foi elaborado em *softwares* (ferramentas de programação) de códigos-fonte livres e abertos. O “ecossistema” de programação foi composto pelo sistema operacional *Linux*[®], a linguagem de programação *Python*[®] e o ambiente de desenvolvimento

(IDE) *PyCharm*® *Community 2022.2*. Observe que essas escolhas não interferem na qualidade do resultado, apenas expressam as preferências pessoais deste autor.

O ambiente foi elaborado visando a portabilidade que é um requisito de sistema (seção 5.2.3). O 3ApC foi desenvolvido (e pode ser utilizado) nos sistemas operacionais derivados do Unix (Linux®, OS® e FreeBSD®) e no Windows® em um computador comum com o microprocessador Intel® *core i3* de 64 bits.

Na próxima seção, apresenta-se a Interface Gráfica de Usuário, que é uma característica indispensável nas aplicações computacionais modernas.

6.3.2 Interface Gráfica de Usuário

A Interface Gráfica de Usuário (GUI) foi elaborada em Python, no *PyCharm*, em que se empregou a *Library PyQt5*® 5.15.6. A escolha desta *library* também é de cunho pessoal e não altera a qualidade dos resultados, sendo um recurso computacional descrito no APÊNDICE B.

A GUI permite acessar as funcionalidades do 3ApC utilizando meios visuais em que se combinam textos e ícones que tornam a utilização do *software* intuitiva, mas que não interferem na qualidade dos resultados, á apenas uma facilitadora de uso.

Os detalhes técnicos de programação estão apresentados nos códigos-fontes. Um exemplo (um fragmento de código) pode ser visto na Figura 18, em que os textos em itálico são comentários explicativos técnicos compreensíveis aos profissionais de programação de computadores.

Figura 18 – Fragmento de código-fonte contendo explicações computacionais.

```

11 from wordcloud import WordCloud
12
13 # Classe processaArquivo:
14 # Inicialização: Abre o arquivo de expressões regulares para extrair Descritores-Thesagro do Texto
15 #               precisa de 'suporte/thesagroComExpressaoRegular.csv', separador de dados = ';'
16 # Extrair_PDF(arquivo): Extrai o texto de arquivos PDF

```

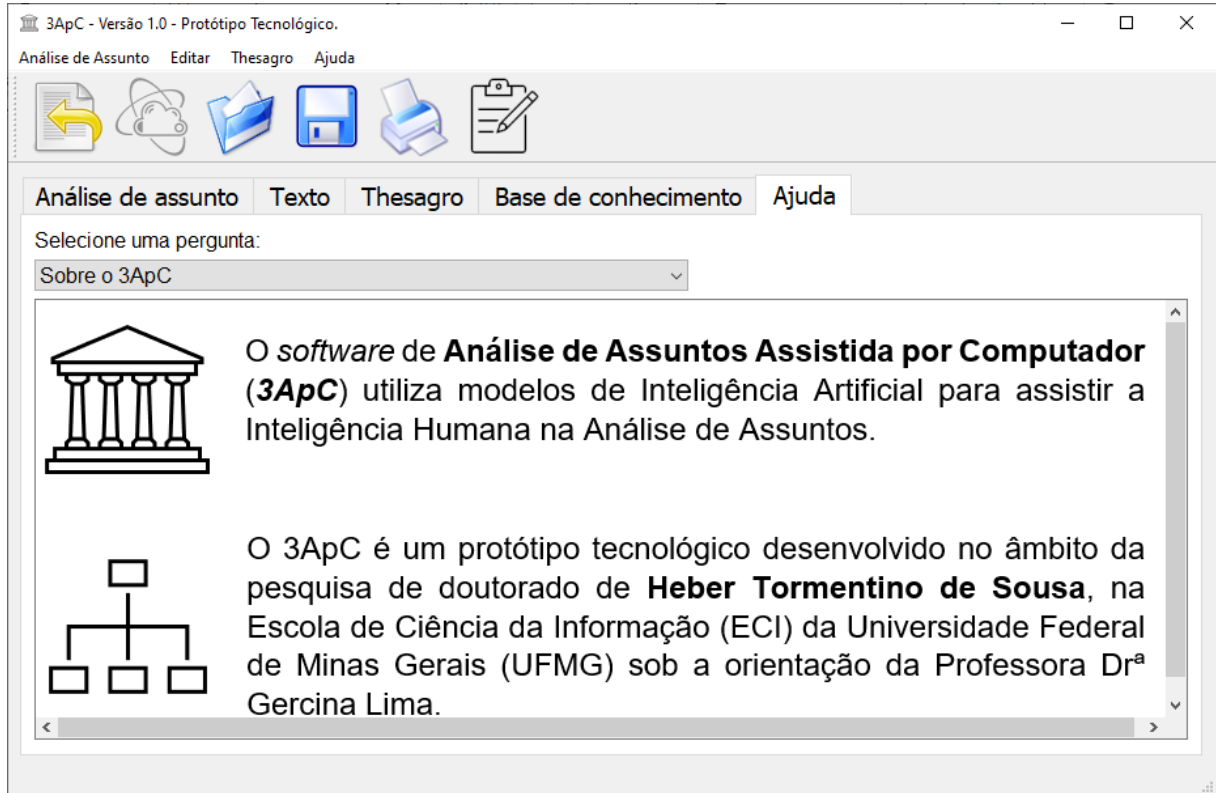
Fonte: Elaborado pelo autor.

A GUI segue o estilo tradicional. O menu principal fica colapsado até que um item seja solicitado, então um submenu é apresentado com as funcionalidades disponíveis. As funcionalidades principais, exatamente igual às disponíveis no Menu, podem ser acessadas por meio de ícones e por combinações de teclas de atalhos. Uma amostra da GUI, A janela principal do 3ApC, pode ser vista na Figura 19.

A GUI do 3ApC obedece às diretrizes propostas nos requisitos de sistema (seção 5.2.3) que inclui interface simples, limpa e eficiente, em que o uso é intuitivo, pois seus elementos são reconhecidos por usuários de outros aplicativos computacionais. Os detalhes

de uso da interface, nessas circunstâncias, são desnecessários, e, portanto, não foram descritos.

Figura 19 – A interface gráfica de usuário da janela principal do 3ApC.



Fonte: Elaborado pelo autor.

Os relatórios produzidos podem ser salvos, exportados e compartilhados. Eles estão discutidos e analisados no Capítulo de Análise dos Resultados.

Na próxima seção, aborda-se a implementação dos modelos de inteligência artificial nos dois métodos principais abordados na fundamentação teórica e metodológica.

6.3.3 Implementação dos métodos de inteligência artificial

Nesta seção, esclarece-se que o termo “rótulo” se refere ao valor que um determinado dado de Entrada deve assumir na saída, ou seja, é a classificação correta do valor de entrada. Por exemplo, para um valor de entrada que seja o texto do artigo científico 311853 da BDPA, o seu rótulo é os seus termos de indexação: “Clima, *Grevillea Robusta*”. O termo “dado rotulado” aponta, ao mesmo tempo, o dado de entrada (artigo) e o seu rótulo (termo de indexação), assim, um artigo e seus termos de indexação formam um “dado rotulado”.

Foram implementados dois métodos de inteligência artificial no 3ApC, que dependem do volume de recursos informacionais disponível. Por isso, a implementação ocorreu por processo ágil¹⁶ de desenvolvimento.

Os dois métodos de inteligência artificial foram selecionados considerando o tipo da “aprendizagem artificial” (Quadro 14), porque essa escolha dependeu da disponibilidade dos dados.

Quadro 14 – A adequação de Modelos de inteligência artificial.

Método	Aprendizagem Artificial	Dados	Método implementado
1	Supervisionada	Textos e Termos de indexação	Modelo Sequencial com mecanismo de atenção
2	Não supervisionada	Textos	Modelagem de Tópicos

Fonte: Elaborado pelo autor.

O Método 1, o modelo sequencial com mecanismo de atenção, é do tipo “aprendizagem artificial supervisionada” de difícil aplicação, porque necessita de dados rotulados em um volume equivalente a mil exemplos por cada padrão reconhecido (GOLUB, 2021). Ele foi removido do 3ApC, mas a sua aplicação e seus resultados podem ser vistos no APÊNDICE G.

O método 2, a modelagem de tópicos, cuja aprendizagem artificial é do tipo “não supervisionada” utiliza a alocação dinâmica de *Dirichlet* (LDA) como algoritmo de inteligência artificial. Ele pode chegar aos resultados com poucos dados de entrada (os textos).

Na próxima seção, apresenta-se a sumarização automática.

6.3.3.1 Sumarização automática

Os resumos produzidos no 3ApC são insumos para assistir à tomada de decisão do profissional da informação (usuário do 3ApC), eles não são destinados para o público (usuário final da informação), devido às limitações da sumarização por inteligência artificial.

Para programar essa funcionalidade foi utilizado a *Libray spaCy* e seu modelo de inteligência artificial, treinado para o português brasileiro (*pt_core_news_sm*).

Um algoritmo, em alto nível, da sumarização automática implementada no 3ApC pode ser descrito, em três etapas, da seguinte forma:

1. Identificar os termos mais importantes para o significado do texto por meio da inteligência artificial.

¹⁶ Metodologia ágil: fluxo de desenvolvimento flexível, interativo e iterativo.

2. Ranquear as frases do texto em ordem ascendente, em que a classificação é feita pelo número de termos importantes que a frase comporta. Quanto mais termos importantes a frase contiver, maior será a classificação dela.
3. Extrair as frases mais bem classificadas que juntas constitui o resumo. Recomenda-se que o número de frases seja, no máximo, cinco, porque acima disso o resumo fica grande.

No 3ApC, a sumarização automática utiliza a inteligência artificial do modelo pré-treinado *pt_core_news_sm da Libray spaCy* para identificar os termos (bigramas, trigramas ou tetragramas) mais importantes no texto, depois avalia e ranqueia todas as frases do texto, em que as frases com mais termos importantes recebem uma pontuação maior. Então, as frases mais bem classificadas são extraídas do texto para formar o resumo (por extração).

As frases são extraídas como estão no texto, assim a qualidade do resumo é a mesma do texto. Com isso, a coerência e a gramaticalidade nas frases do resumo apresentam as mesmas características presentes no texto e não são atributos do *software*.

Ressalta-se que o número de frases que formam o resumo pode ser entre um e o número total de frases do texto. Mas acima de cinco de frases, os resumos ficam extensos e perde-se o sentido de resumir. Por isso, o 3ApC produz resumos com as três frases mais bem classificadas para representar o texto. O número de frases pode ser configurado pelo usuário.

Esta funcionalidade corresponde ao estágio “ler o texto/compreender o texto” da análise de assuntos intelectual.

Na próxima seção, descrevem-se as expressões regulares, que é o método de identificação dos assuntos do texto.

6.3.3.2 Expressões regulares

Para saber quais termos do Thesagro estão presentes em um texto com mil palavras, seriam necessárias 9.706.000 de comparações, além disso, seria necessário considerar que os termos podem variar quanto ao número e quanto ao gênero, por exemplo, “fruta, fruto, frutas e frutos”, o que implicaria no aumento do tempo e do esforço computacional só para identificar os assuntos do texto.

Os termos compostos são avaliados primeiro e, depois de contados, são removidos, para evitar a recontagem dos termos simples presentes no termo composto.

Assim, o algoritmo em alto nível para a identificação dos assuntos do texto pode ser descrito da seguinte forma:

1. **ordenar as expressões regulares** por ordem decrescente do número de termos;

2. **calcular a frequência dos termos no texto**, isto é, contar quantas vezes cada termo do Thesagro aparece no texto analisado. Note que se aplicam sequencialmente, uma por vez, todas as 9706 expressões regulares nessa etapa; e
3. **apresentar a lista de termos**, em ordem decrescente das frequências. Cada termo da lista está na mesma forma que aparece no Thesagro. Por exemplo, “semente” e “sementes” são apresentadas na lista como “semente”.

A aplicação das expressões regulares ocorre como um processo intermediário e sem intervenção do usuário, corresponde à fase de “identificação dos assuntos” da análise de assuntos intelectual. Essa lista de termos identificados no texto, em ordem decrescente de frequência, consta no relatório de resultados.

Na próxima seção, apresenta-se a modelagem de tópicos.

6.3.3.3 Modelagem de tópicos

A implementação da modelagem de tópicos no 3ApC utilizou o modelo Alocação Dinâmica de *Dirichlet* (LDA). Esse modelo não precisa de dados rotulados, porque é um método de inteligência artificial do tipo “aprendizado artificial não supervisionado”. Com isso, contornou-se o problema da ausência de dados previamente classificados por seres humanos em quantidade adequada. O LDA foi implementado a partir da *library Gensim* descrita no Capítulo 3.

Para realizar uma modelagem de tópico, o desafio é encontrar um *cluster* para o artigo na base de conhecimento. Para tanto, seria necessário verificar todos os *clusters*, um-a-um, o que seria contraproducente devido ao tempo e ao esforço computacional necessários. Então, organizou-se a busca por *clusters*, de forma que os termos com maior probabilidade de formar um *cluster* para o artigo são analisados primeiro.

As buscas por *clusters* foram feitas na seguinte ordem de prioridade: (1) os assuntos do título são os primeiros a serem analisados e constituem-se assuntos relevantes; e (2) os assuntos de maior frequência no texto são analisados primeiros. O processo é encerrado quando se seleciona 10 assuntos relevantes, ou quando não há mais assuntos para serem analisados.

Em síntese, o 3ApC utiliza um assunto de cada vez para delimitar um *cluster* na base de conhecimento. O LDA usa esse *cluster* para calcular os tópicos latentes e identificar os assuntos compartilhados pelo tópico e pelo texto. Os assuntos compartilhados são considerados relevantes, e, portanto, são incluídos na lista de assuntos relevantes.

Na próxima seção, apresenta-se o teste de funcionalidade do 3ApC.

6.4 Teste de funcionalidade

Nesta seção, testa-se a Interface Gráfica para verificar a aderência aos requisitos funcionais e intuitividade dos percursos disponíveis para o usuário.

Os testes de funcionalidade são atividades de avaliação, não se trata da análise dos resultados, a qual será apresentada no próximo capítulo. Nesta etapa, os percursos para obter os resultados e coerência na apresentação deles, juntamente com a GUI, a sequência de eventos, e o tratamento de exceções são avaliados.

Na próxima seção, descreve-se o ambiente de utilização do *software* 3ApC.

6.4.1 O ambiente

O ambiente desenvolvido é constituído de uma interface gráfica (GUI), que dá acesso a todas as funcionalidades do 3ApC. Para esclarecer: uma “função” realiza apenas uma tarefa; e uma funcionalidade pode conter várias funções. Por exemplo: a função “salvar” realiza a tarefa de armazenar digitalmente um arquivo no disco rígido.

O ambiente do 3ApC apresenta os seguintes elementos: (1) um menu principal, (2) uma barra de ferramentas, e (3) uma estrutura de abas (recurso visual para separar os recursos e facilitar a análise e a gestão dos resultados).

No menu principal as funções precisam de, pelo menos, dois passos para serem chamadas: (1) primeiro passo: chama-se a funcionalidade (opção visível, por exemplo, “análise de assuntos”) com um clique de *mouse*; e (2) segundo passo: escolhe-se a função (opção colapsada, que só fica disponível depois do primeiro passo) com um segundo clique de *mouse*. A barra de ferramentas apresenta apenas as funções mais utilizadas. Nela, para chamar uma função, basta um único clique de *mouse*.

A estrutura de abas organizam os resultados em: (1) aba de “análise de assuntos”, em que são apresentados todos os resultados referentes a sumarização automática, a identificação de assuntos e a seleção de assuntos relevantes; (2) aba “texto”, em que pode ser visto o texto bruto; (3) aba “Thesagro”, em que se pode ver a estrutura hierárquica dos termos do Thesagro; (4) aba “base de conhecimento”, em que se pode ver o conteúdo da base de conhecimento; (5) aba “tópicos”, em que se apresenta as análises visuais dos *Clusters*; e (6) aba “ajuda”, em que se pode ver as informações sobre o *software* e o tutorial de utilização.

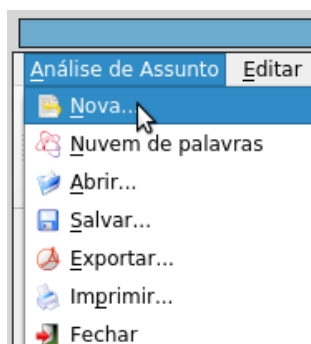
Na próxima seção, descreve-se o teste de funcionalidade da Interface gráfica de usuário.

6.4.2 Interface gráfica de usuário

Nesta seção, para testar a funcionalidade a GUI, utilizou-se de exemplo o artigo científico “Germoplasma de caupi: coleção ativa e de base” (FREIRE *et al.*, 1999).

Considerou-se que o percurso principal do usuário é realizar a análise de assuntos. Para tanto, no 3ApC, o usuário aciona o “Menu Principal → Nova” (Figura 20), e, em seguida, será solicitado o nome do arquivo contendo o artigo científico a ser analisado.

Figura 20 – O início do percurso do usuário para realizar uma análise de assuntos.



Fonte: Elaborado pelo autor.

Após o usuário apontar o diretório e o nome do arquivo, o 3ApC solicitará o título do artigo. Caso se opte por cancelar, será emitido um alerta de que nenhum arquivo foi escolhido e o processo será imediatamente encerrado.

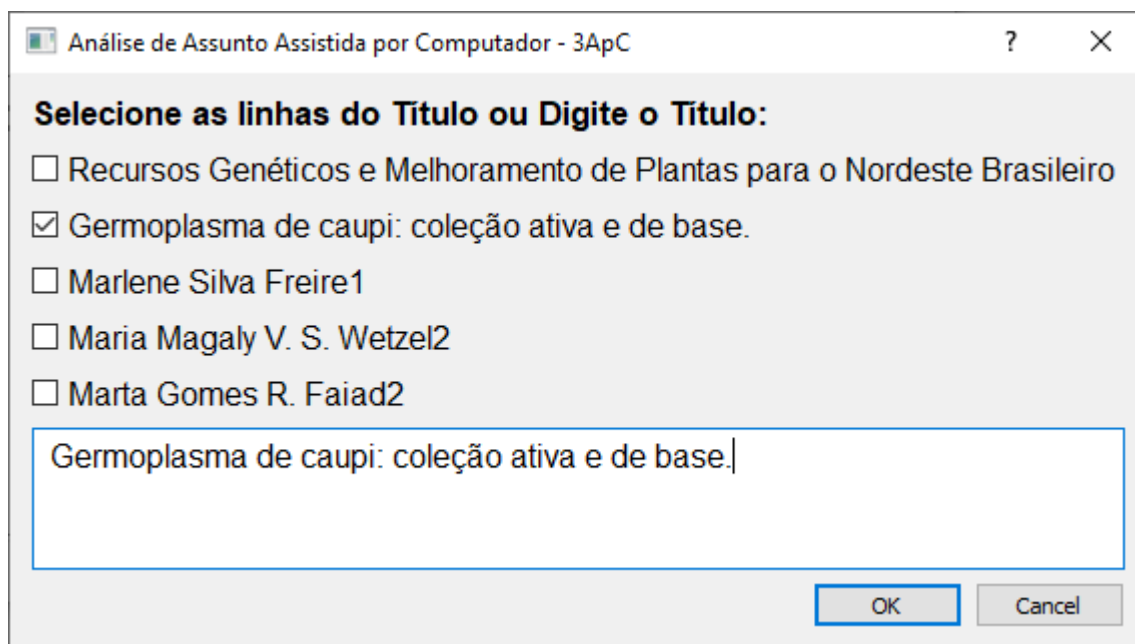
O processo de inserir o título correto do artigo é importante, porque os assuntos relevantes podem estar nomeados no título, logo, a análise de assuntos depende dele.

Os algoritmos não conseguem distinguir os títulos automaticamente no texto extraído, porque os títulos não são identificáveis e nem estão padronizados, de forma que um título pode estar em caixa alta, ou ter apenas a primeira letra em maiúscula. Além disso, o título pode ser a primeira linha do artigo ou pode nem estar entre as primeiras linhas. Assim, o usuário tem as seguintes opções no 3ApC para inserir o título (Figura 21): (1) digitar (ou copiar e colar) o título na caixa de edição, ou (2) marcar as linhas correspondentes ao título (se elas estiverem disponíveis).

As formas de inserir o título não influenciam o resultado, de maneira que o usuário pode escolher o percurso que mais lhe agrada.

As ações do usuário neste percurso, portanto, se resumem a: (1) apontar o arquivo, (2) inserir o título, e (3) aguardar os resultados.

Figura 21 – Caixa de diálogo para o usuário inserir ou digitar o título do Artigo.



Fonte: Elaborado pelo autor.

Na próxima seção, descreve-se o teste de funcionalidade dos métodos de inteligência artificial.

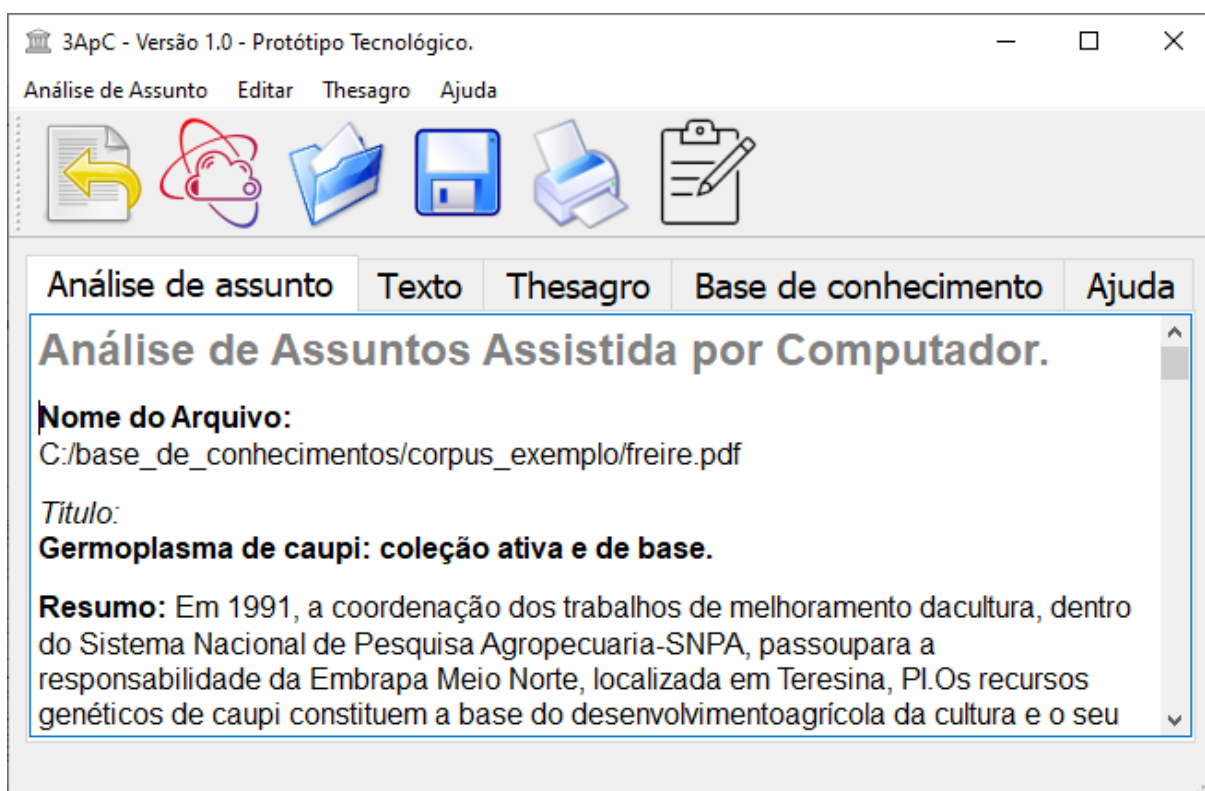
6.4.3 Métodos de inteligência artificial

Os métodos de inteligência artificial são transparentes ao usuário, isto é, não é necessária nenhuma intervenção humana para a sua realização. A duração deles depende dos recursos disponíveis de cada computador, de forma que o tempo médio de execução tende a ser menor em computadores mais potentes.

O percurso de análise de assuntos do 3ApC termina com a apresentação do relatório completo mostrado na Figura 22.

O relatório começa pelas informações contextuais fornecidas pelo usuário, o que inclui o caminho para o arquivo e o nome dele, e o título do artigo científico. Seguidos pelos resultados dos métodos de inteligência artificial: (1) resumo (sumarização automática); (2) os cinco vocábulos mais frequentes no texto; (3) os termos não preferidos e seus respectivos termos preferidos encontrados no texto; (4) a lista de nomes científicos identificados no texto e seus respectivos nomes comuns; (5) a lista de assuntos relevantes (modelagem de tópicos); (6) a lista de assuntos identificados no texto (expressões regulares) com suas respectivas frequências.

Figura 22 – A tela contendo o relatório da análise de assuntos.



Fonte: Elaborado pelo autor.

Na Figura 23, pode ser vista a tabela com os assuntos relevantes selecionados pelo 3ApC de Valladares *et al.* (2003), incluindo as informações adicionais.

Figura 23 – Os assuntos relevantes do texto selecionados pelo 3ApC.

Assunto	Termo Não preferido	Nome Científico	Nome Comum	Termos Coocorrentes	Artigos indexados ⁽¹⁾
Doença					93
Feijão de corda	Caupi	<i>Vigna Unguiculata</i>			9
Genética					16
Genótipo					45
Melhoramento					39
Patógeno					11
Peso					9
População					7
Produção					99
Resistência					55

¹ Artigos indexados: número de artigos indexados com o termo na base de conhecimento.

Fonte: Elaborado pelo autor.

Quando o assunto é simbolizado por um termo não preferido, ele é substituído pelo termo preferido e é movido para a coluna “termo não preferido” (ver Figura 23). O nome comum de um nome científico é apresentado na coluna “nome comum” e o nome científico de um assunto vai para a coluna “nome científico” (ver Figura 23).

A coluna “termo coocorrente” (ver Figura 23) apresenta os termos que coocorrem com uma frequência superior a 90% nas listas de termos indexadores do assunto. A coluna

“artigos indexados” apresenta o número de artigos indexados na base de conhecimento pelo “assunto relevante” (ver Figura 23).

Na próxima seção, apresenta-se o teste da funcionalidade “sumarização automática”.

6.4.3.1 Sumarização automática

Para apresentar o teste de funcionalidade da “sumarização automática” utilizou o artigo científico de Freire *et al.* (1999) e o resumo pode ser visto no Quadro 15.

Quadro 15 – O teste da funcionalidade “sumarização automática” do 3ApC.

Resumo
Os recursos genéticos de caupi constituem a base do desenvolvimento agrícola da cultura e o seu manejo envolve atividades que vão desde o enriquecimento, realizado através da introdução e da coleta de germoplasma, à caracterização e avaliação, e por último a sua conservação a médio prazo, realizada pelo Banco Ativo de Germoplasma (BAG Caupi) na Embrapa Arroz e Feijão e, a longo prazo, na Coleção de Base (Colbase Caupi), realizada na Embrapa Recursos Genéticos e Biotecnologia. Todas as introduções de germoplasma realizadas pela Embrapa Arroz e Feijão, ocorreram através da Embrapa Recursos Genéticos e Biotecnologia, realizada pela Área de Intercâmbio e Quarentena de Germoplasma AIQ, que tem como atribuição, coordenar e processar o intercâmbio de germoplasma para o Sistema Nacional de Pesquisa Agropecuária (SNPA). Outras atividades desenvolvidas no processo de manutenção da coleção ativa se referem à conservação (em câmara controlada para armazenamento a médio prazo), multiplicação (para obtenção de sementes de alta qualidade e em quantidade suficiente para atender à coleção a longo prazo e à demanda dos usuários), regeneração (para manutenção de sua integridade genética na coleção de base), caracterização e avaliação (individualização fenotípica de cada acesso), intercâmbio (distribuição e troca de germoplasma) e banco de dados (disponibilidade informatizada dos dados da coleção) (Goedert, 1988; Breese, 1989; No período 1977/1997, 4.153 acessos foram registrados na coleção ativa de caupi do BAG da Embrapa Arroz e Feijão, sendo, 1.533 amostras oriundas do exterior, perfazendo 36,92% do total do acervo e 2.620 acessos de germoplasma recebidos do Brasil, representando 63,08% da coleção.

Fonte: Elaborado pelo autor.

No corpo do resumo é constituído por três frases extraídas do texto-fonte que são as três mais bem ranqueadas pelo modelo de inteligência artificial. Lembrando que o número de frases (três) foi uma escolha de projeto, porque menos de três frases o resumo ficaria muito curto.

A entrada do algoritmo foi o texto do artigo e a saída foi um resumo coerente do conteúdo do texto, com notável legibilidade e correção gramatical e ortográfica. Por ser um processo de extração, o resumo mantém as referências bibliográficas, as datas, os números e as unidades de medidas, porque as frases foram retiradas do texto no estado em que estavam (isto é, essas frases não foram elaboradas pelo *software*), logo, a qualidade do resumo depende do texto-fonte.

Na próxima seção, apresenta-se o teste da funcionalidade das “expressões regulares”.

6.4.3.2 Expressões regulares

As expressões regulares formuladas fazem parte da base de conhecimento em um arquivo com 9706 registros. No Quadro 4 pode ser vista uma amostra das expressões regulares obtida no teste de funcionalidade “expressões regulares”.

Quadro 4 – Amostra aleatória de expressões regulares.

Termo	Expressão regular
Produção de Sementes	produc(ao aos aes oes) de sement(e es)
Semente Básica	sement(e es) basic(a as)
Tratamento de Semente	tratament(o os) de sement(e es)

Fonte: Elaborado pelo autor.

Os sinais diacríticos e as capitalizações foram retirados do termo em linguagem natural e foram acrescentadas as terminações alternativas para o singular e o plural. Como foram produzidas por um algoritmo, dada a impossibilidade de verificar “manualmente” todas as variações possíveis, algumas alternativas não ocorrem nos textos, mas isso não produz nenhum tipo de erro no processo.

Na Figura 24 pode ser visto uma parte do resultado da funcionalidade expressões regulares, em que apenas os cinco primeiros assuntos estão listados (e suas frequências) dos 155 assuntos identificados no exemplo.

Figura 24 – Exemplo de assuntos identificados no texto pelo 3ApC.

Assunto	Frequência
Semente	59
Caupi	56
Germoplasma	45
Recurso Genético	26
Conservação	19

Fonte: Elaborado pelo autor.

Foram encontradas 65 ocorrências da palavra “semente” no artigo de exemplo (FREIRE *et al.*, 1999), distribuídas da seguinte forma: 51 formas plurais e 14 formas singulares; das quais, três delas faziam parte do termo composto “Produção de Sementes”, outras três pertenciam ao termo composto “Semente Básica” e outra ao termo “Tratamento de Semente”, e as 59 (ver Figura 24) restantes formavam o termo simples “semente” a partir das formas plurais e singulares.

Na próxima seção, apresenta-se o teste da funcionalidade da “modelagem de tópicos”.

6.4.3.3 Modelagem de Tópicos

No Quadro 16, estão confrontados os termos de indexação sugeridos pelo 3ApC e os escolhidos pela BDPA. O conjunto de previsão contém o conjunto de termos da BDPA e apresenta informações adicionais que ajudam na tomada de decisões, incluindo os termos preferidos e os nomes científicos.

Quadro 16 – Previsão do 3ApC comparada à indexação da BDPA.

Artigo	Previsão	BDPA
195124	Feijão de Corda USE Caupi, Semente, <u>Germoplasma</u> , Recurso Genético, <u>Vigna Unguiculata</u> [Feijão de Corda], Dado, Biotecnologia, Conservação	Germoplasma, <i>Vigna Unguiculata</i>

Fonte: Elaborado pelo autor.

Assim, a atividade iniciada em “análise de assuntos → Nova” acionou a modelagem de tópicos e produziu resultados que se adequam aos requisitos funcionais 9 e 10 da seção 5.2.1, apresentando clareza, coerência entre os requisitos funcionais. Esse percurso atende ao Requisito de Sistema 3 (seção 5.2.3) de ter usabilidade simples, rápida e flexível com uma boa curva de aprendizado.

Na próxima seção, apresenta-se a funcionalidade que exibe as informações de suporte à tomada de decisões.

6.4.4 Informações de suporte à tomada de decisões

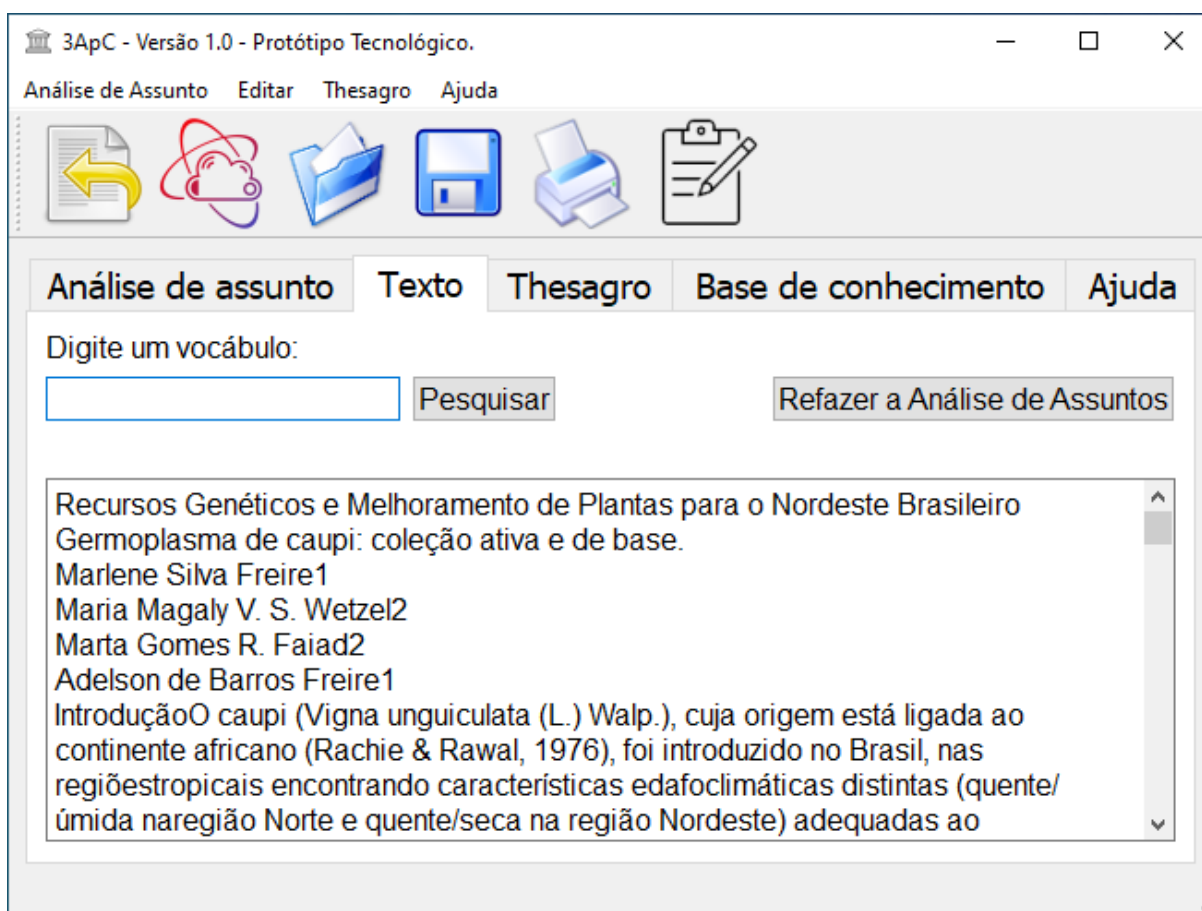
A inteligência aumentada como resultado ocorre na tomada de decisão, porque o usuário tem uma visão maior do cenário. Assim, no 3ApC, foram adicionadas informações de contexto, incluindo o texto bruto, a estrutura do termo no Thesagro, as estatísticas da base de conhecimento e a visualização dos *clusters* e da nuvem de palavras.

Na próxima seção, apresenta-se a funcionalidade que permite a edição do texto extraído do documento.

6.4.4.1 Edição do texto

A funcionalidade apresentada na aba “Texto” (Figura 25) do 3ApC permite a edição do texto extraído. Feita uma inspeção, pode-se optar por remover manualmente as informações irrelevantes. Neste caso, podem ser removidos os caracteres de controle e os conteúdos repetitivos.

Figura 25 – A funcionalidade “Texto” permite editar o texto do artigo.



Fonte: Elaborado pelo autor.

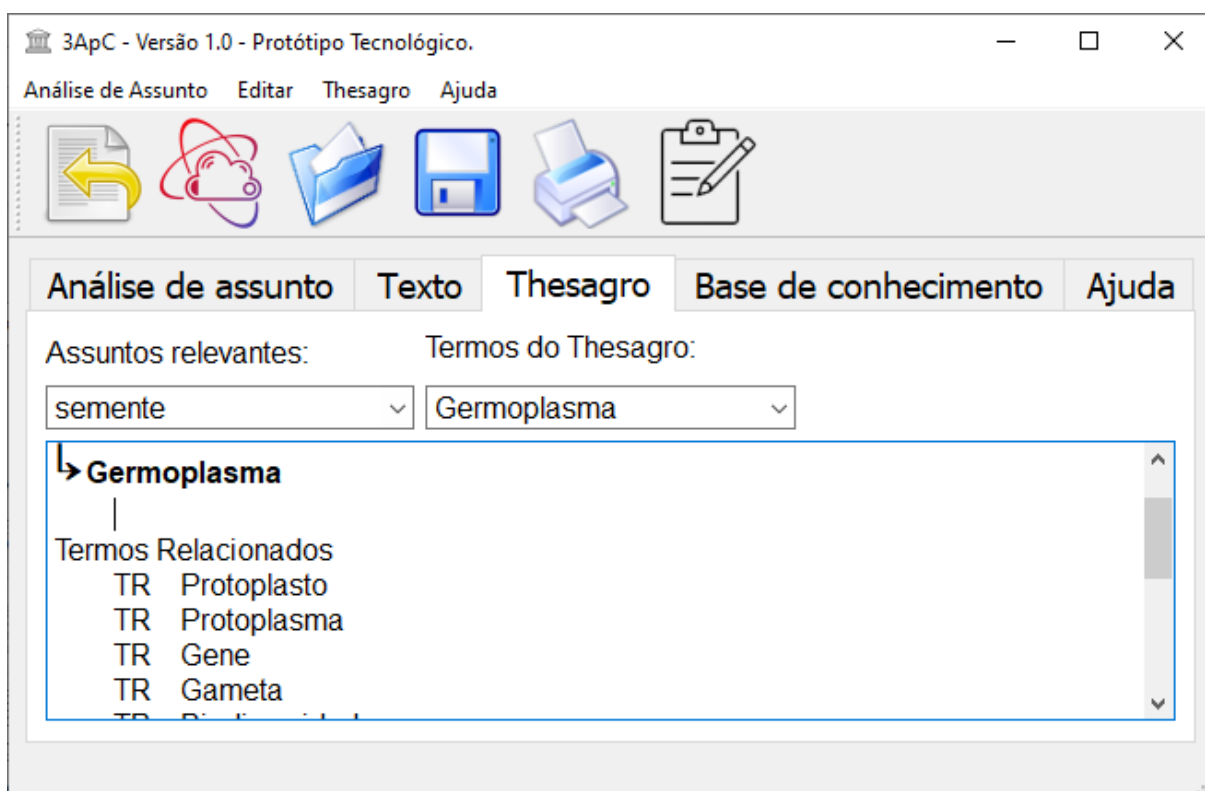
A opção “Pesquisar no texto”, que pode ser vista na mesma Figura 25, permite buscar os termos no texto extraído para facilitar a edição. O texto editado pode ser submetido a um novo processo de análise de assuntos por meio da funcionalidade “Refazer a análise de assuntos”.

Na próxima seção, descreve-se a navegação no Thesagro.

6.4.4.2 Navegação no Thesagro

A aba “tesauro” que pode ser vista na Figura 26 organiza as funcionalidades para a navegação na estrutura hierárquica do Thesagro. Nela, os termos indexadores são sugeridos como opções de navegação, que podem ser selecionadas com um clique de mouse.

Figura 26 – A funcionalidade de navegação na estrutura do Thesagro.



Fonte: Elaborado pelo autor.

Apresentam-se para o usuário a estrutura hierárquica e as listas de termos associados e relacionados do termo escolhido. Na Figura 26, exibe-se a estrutura hierárquica do descritor “semente”, em que se pode verificar que o termo está no terceiro nível hierárquico, sendo “anatomia” seu termo raiz (em negrito) e “anatomia vegetal” seu termo superior imediato; e a sua lista de termos relacionados.

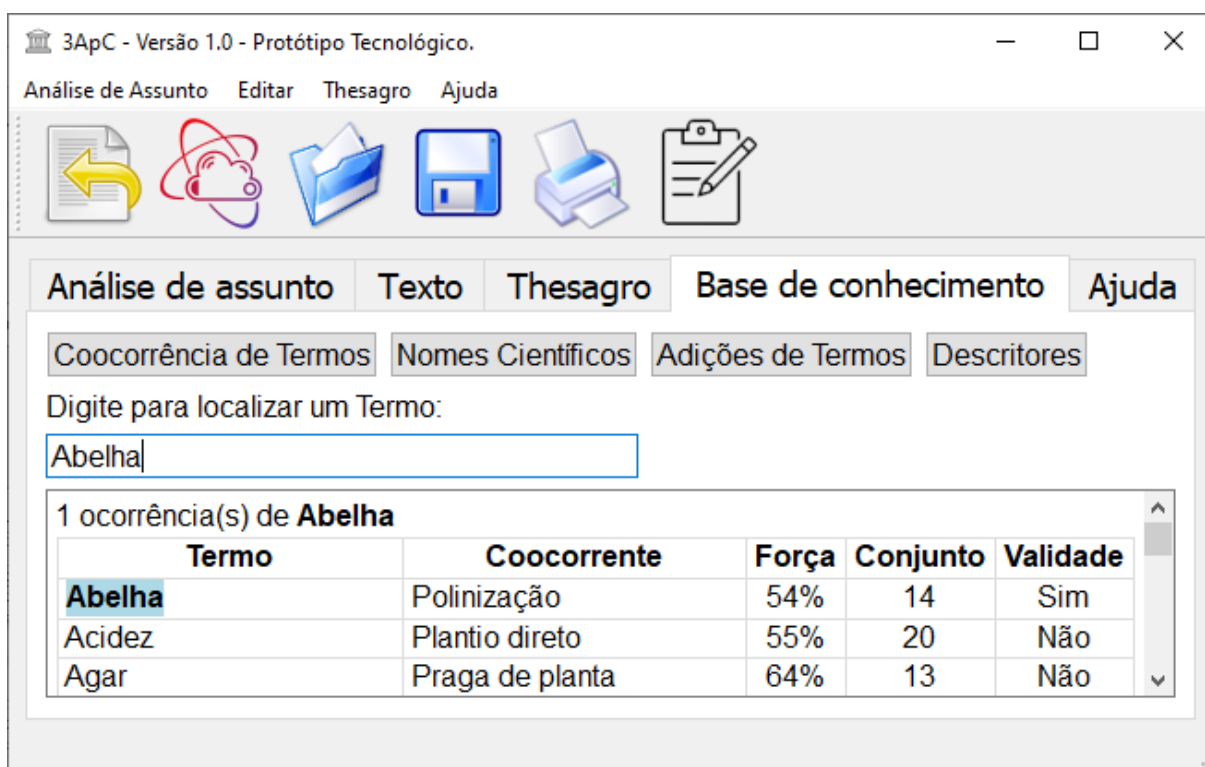
Na próxima seção, descreve-se o teste de funcionalidade da “aba” base de conhecimento.

6.4.4.3 Base de conhecimento

A base de conhecimento foi definida pela *Information Technology Infrastructure Library* (AXELOS, 2019) como centralizadora dos dados e das informações relevantes para facilitar a gestão do “conhecimento”, uma vez que é necessário criar, organizar, compartilhar e utilizar dados e informações para realizar as tarefas.

A aba “base de conhecimento” inclui quatro funcionalidades (Figura 27) que permitem navegar pelas indexações, pelos metadados, pelos nomes científicos e pela lista de termos coocorrentes nos índices.

Figura 27 – As funcionalidades para conhecer a base de conhecimento.



Fonte: Elaborado pelo autor.

Essa funcionalidade permite especificamente gerenciar a base de conhecimento, permitindo que o usuário insira, edite, busque e exclua dados, informações e conhecimentos.

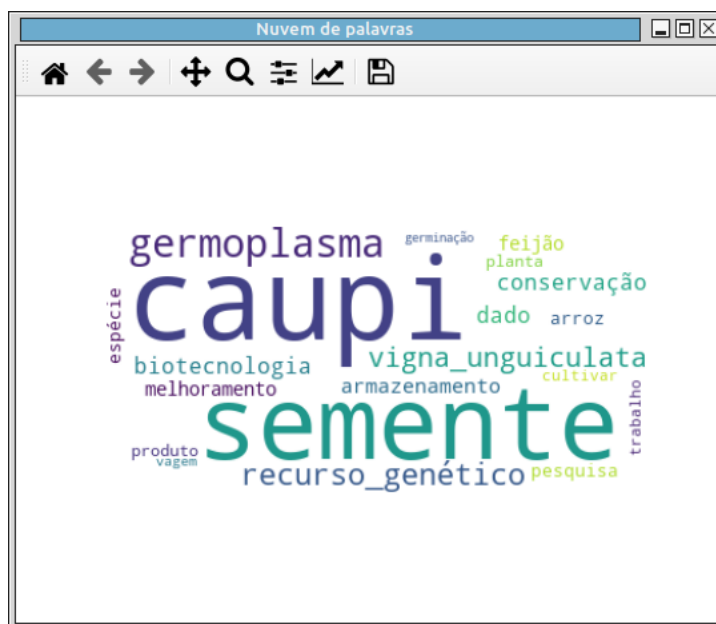
Na próxima seção, descreve-se a funcionalidade de nuvens de palavras.

6.4.4.4 Nuvens de palavras

A visualização baseada em cores, tamanhos e posições facilita a compreensão das informações expostas, por isso foi adicionado o recurso visual denominado de nuvens de palavras, em que as palavras foram desenhadas artisticamente com tamanhos proporcionais às suas frequências.

Na Figura 28, as palavras mais frequentes do arquivo 195124 estão arranjadas artisticamente. Na nuvem de palavras a frequência dos termos do texto é simbolizada pelo tamanho da fonte, quanto maior a fonte usada, maior a frequência do termo no texto. No exemplo, os termos “semente”, “caupi” e “germoplasma” têm as maiores fontes, o que indica que eles são os termos mais frequentes no artigo. Os termos “germoplasma” e “caupi” foram utilizados para indexar o artigo na BDPA, embora “caupi” tenha sido substituído pelo nome científico “*Vigna Unguiculata*”.

Figura 28 – Uma nuvem de palavras para exibir informações visuais do texto.



Fonte: Elaborado pelo autor.

Esse recurso dá uma visão geral do conteúdo do artigo, apresentando as informações de uma forma que exige pouco esforço para ser compreendido, além disso, ele facilita visualizar se houve problemas na extração do texto, ao exibir termos desconhecidos, por exemplo.

Com isso, encerra-se o Teste de funcionalidade do 3ApC. Nesse percurso, foram testadas as funcionalidades para lidar com cada objeto necessário à tomada de decisão em uma tarefa de indexação, que são: (1) artigo (documento), (2) texto (fluxo de caracteres), (3) Thesagro (Tesouro), (4) *corpus*, (5) metadados, (6) termos indexadores (assuntos significativos), (7) informações adicionais (nomes científicos, nomes comuns, termos preferidos e termos não preferidos), (8) agrupamentos, e (9) base de conhecimento. Além disso, também foram apresentadas as funcionalidades de visualização avançada.

A inteligência aumentada ocorre como um princípio no planejamento, na programação e no uso do *software*. Portanto, espera-se que ela ocorra no uso do 3ApC, da mesma forma que se espera que um aplicativo de navegação não dirija o carro, mas que dê as informações necessárias ao motorista para efetuar com êxito a tarefa.

Como pode ser visto ao longo deste capítulo, o 3ApC apresenta qualidades tecnológicas que permitem o uso dos recursos de *hardware* para utilizar com eficiência os métodos de inteligência artificial, de estatística e de edição, além de facilitar o uso de recursos visuais.

No próximo capítulo, apresentam-se as análises dos resultados.

7 ANÁLISE DOS RESULTADOS

Neste capítulo, descreve-se a Etapa 4 dos procedimentos metodológicos (seção 4.3.4). e apresentam-se as análises dos resultados obtidos no Experimento desta pesquisa com a finalidade de demonstrar o conhecimento adquirido, apresentar a solução alcançada, a qualidade dos resultados e a Validação do protótipo tecnológico 3ApC.

No Quadro 17 pode ser vista a relação entre os estágios do processo de análise de assuntos intelectual, as funcionalidades no 3ApC e os tipos de resultados produzidos.

Quadro 17 – Correspondência de estágios da análise de assuntos e as funcionalidades do 3ApC.

Estágio da análise de assuntos	Funcionalidade no 3ApC	Resultados
Ler o texto/compreender o texto	Sumarização automática	Resumo
Identificar os assuntos do texto	Expressões regulares	Assuntos/conceitos
Selecionar os assuntos relevantes	Modelagem de tópicos	Assuntos relevantes

Fonte: Elaborado pelo autor.

Para validar o processo análise de assuntos do 3ApC, foram analisados os métodos de: (1) Sumarização Automática (seção 7.1); (2) expressões regulares (seção 7.2); e (3) modelagem de tópicos (seção 7.3). Na seção 7.4, apresentou-se um panorama da proposta metodológica de inteligência aumentada.

Na próxima seção, analisam-se os resultados obtidos pela sumarização automática.

7.1 Sumarização automática

Nesta seção, analisaram-se os resultados obtidos com a realização de um Experimento com todos os artigos científicos do *corpus* do Experimento com a finalidade de avaliar e validar o método de sumarização automática. Para tanto, analisaram-se os resumos produzidos, conforme as orientações descritas na seção 4.3.4.3. Todos os resumos produzidos no Experimento podem ser vistos no APÊNDICE C.

Na próxima seção, apresenta-se uma visão geral dos resumos obtidos por sumarização automática.

7.1.1 Visão geral dos resumos

Para descrever a análise do método de sumarização automática do 3ApC, foram selecionados três resumos característicos: (1) resumo adequado (seção 7.1.1.1); (2) resumo ilegível (seção 7.1.1.2); e (3) resumo com problemas semânticos (seção 7.1.1.3). Esses três

exemplos foram selecionados por serem ilustrativos, sendo o primeiro tipo o mais frequente (66%), e outros 34% necessitam de intervenção humana.

Na próxima seção, apresenta-se o exemplo de um resumo adequado.

7.1.1.1 *Resumo adequado*

A função dos resumos no 3ApC é permitir que o indexador tenha uma visão geral do texto e compreenda do que se trata o documento sem a necessidade de ler o texto inteiro. O benefício de ler apenas o resumo ao invés do texto-fonte é que se reduz a quantidade de dados e informações que o indexador precisa analisar, assim, ganha-se tempo.

Considerou-se que um resumo adequado apresenta ao indexador facilidade de leitura, fluência semântica e informações úteis.

O resumo do artigo “Adsorção de fósforo em solos de argila de atividade baixa” (VALLADARES; PEREIRA; ANJOS, 2003) produzido pela sumarização automática, que pode ser lido no Quadro 18, é um exemplo de um resumo adequado.

Quadro 18 – Um resumo adequado obtido por sumarização automática.

Resumo
Neste estudo foram avaliadas possíveis correlações entre atributos físicos e químicos do solo e a capacidade máxima de adsorção de fósforo (CMAP) em solos com argila de atividade baixa de diferentes regiões do Brasil, sendo utilizadas amostras de horizontes superficiais (A) e subsuperficiais (B) de 16 perfis de solos. Dentre os horizontes subsuperficiais, os perfis PAd1 e LBd1 foram os que apresentaram maiores valores de CMAP (1.667 mg.kg ⁻¹) quando comparados aos demais, mostrando alta CMAP em solos originários de rochas básicas ou alcalinas; no perfil PAd3, de textura média, verificou-se o menor valor de CMAP (417 mg.kg ⁻¹), mostrando que solos com menores teores de argila apresentam menores CMAP. Os coeficientes de correlação de alguns atributos dos solos com a CMAP (Quadro 3) foram positivos e significativos a 1% ou a 5% de probabilidade no horizonte superficial para o teor de carbono orgânico, superfície específica e teor de argila, sugerindo que para esses horizontes, tais atributos contribuem com elevação da capacidade máxima de adsorção de fósforo, sendo este comportamento também verificado em outros.

Fonte: Elaborado pelo autor.

Esse resumo é formado por três frases extraídas do texto-fonte, que, conforme descrito na seção 4.3.3.1, são as mais bem classificadas pelo modelo de inteligência artificial para representar (resumir) o conteúdo do texto.

O dado de entrada do algoritmo de sumarização automática foi o arquivo digital no formato PDF contendo o artigo, e a saída foi um resumo coerente com notável legibilidade, coerência semântica e correção gramatical e ortográfica.

Na próxima seção, apresenta-se o exemplo de um resumo ilegível.

7.1.1.2 *Resumo ilegível*

O resumo do artigo “Geadas na viticultura e estratégias para prevenção de danos no sul do Brasil” (SANTOS *et al.*, 2020), produzido por sumarização automática, que pode ser

visto no Quadro 19, é um exemplo de um resumo é ilegível. Ressalta-se que os problemas apresentados no resumo não têm relação com a qualidade do texto-fonte.

Quadro 19 – Um resumo ilegível obtido por sumarização automática.

Resumo
UDGLDWLYD HVVDFRPELQDomR GH IDWRUHV DV FRQGLo}HV VmR DV PDLV DGYHUV DV H IDYRUIYHLV D JHDGD QHJUD FRQJHODPHQWR İÜJmRV HFL GRV(VWiGLR GH GHVHQYRO YLPHQWRHPSHUDW XUD & URQFR% UDo RDUPHQWRV 'RUPrQFLD D *HPD GRUPHQWH 'RUPrQFLD D *HPD DOJRGmR DWp SRQWD YHUGH %URWDomR D %URWR FRP IROKD DEHUWD %URWDomR D %URWR FRP IROKDV DEHUWDV &UHVFLPHQWR LQLFLDO D %URWRV DPRV PDLRUHV TXH FP &UHVFLPHQWR LQLFLDO 'DGRV PpGLRV FRPSLODGRV GD OLWHUDWXUD)HQQHOO &HQWLQDUL HW DO D SDUWLU GH UHJLVWURV HP FRQGLo} HV FRQWURODGDV RX D campo, com tempo mínimo de minutos na temperatura especíca. Valores servem apenas como referência, pois as características GR YLQKHGR ORFDO FXOWLYDU HWF SRGHP SURPRYHU YDULDo}HV QHVHV OLPLWHV Os limites de temperatura de cada tecido/estádio podem variar por inÂuência do local de FXOWLYR GR JHQyWLSR H GDV SUIWLFDV GH PDQHMR TXH DQWHFHGHU DP D JHDGD GHQWUH RXWURV IDWRUHV No geral, as partes ligniçadas da videira, como tronco, braços e gemas dormentes, são as que WROHUDP DV WHPSHUDWXUDV FRQJHODQWHV PDLV H[WUHPDV DEHOD DV TXDLV QHP RFRUHP QR VXO GR %UDVLO &RQWXGR DSyV R LQIFLR GH EURW DomR WRGRV RV WHFLGRV HP FUHV FLPHQWR IROKDV UDPRV YHUGHV e inÂorescências) PpWRGRV SDVLYRV GHYHP VHU SULRUL]DGRV SRLV VmR GH PH QRU FXVWR H JDUDQWHP R PHQRU QYHO GH PpWRGRV SDVLYRV 'HQWUH RV PpWRGRV SDVV

Fonte: Elaborado pelo autor.

A sumarização automática falhou devido a erros ocorridos nas etapas anteriores da *pipeline*, (arquivo criptografado), de forma que o fluxo de caracteres extraído não é um texto em português, e, nesse caso, nem mesmo a funcionalidade de edição manual do texto, descrita na seção 6.4.4.1, pode ajudar. Uma alternativa é pedir/solicitar o texto ao autor.

Esse exemplo ilustra bem a importância das metodologias baseadas em inteligência aumentada, uma vez que, se a inteligência artificial falhar, o problema ainda pode ser resolvido pela inteligência humana, mesmo que seja entrando em contato o autor do artigo.

Na próxima seção, apresenta-se um resumo com problemas semânticos.

7.1.1.3 Resumos com problemas semânticos

Os problemas semânticos ocorrem em resumos que não apresentam suas informações em uma ordem lógica, devido a erros ortográficos e ao embaralhamento de fragmentos das frases.

O resumo do artigo “Análise comparativa entre índices de vegetação e sua relação com o balanço hídrico em soja” (SCHAPARINI *et al.*, 2020), pode ser visto no Quadro 20, é um exemplo de um resumo com problemas semânticos, apesar de conter informações úteis e ser compreensível à inteligência humana. Esses problemas estão relacionados à codificação do arquivo em PDF e não à qualidade do texto-fonte.

Quadro 20 – Exemplo de um resumo com problemas na sumarização automática.

Resumo
<p>Sabendo-se da importância econômica da soja para o RS, e do fato de que um dos principais fatores de variabilidade na produtividade é a deficiência hídrica e, ainda, que é possível monitorá-la usando índices de vegetação, o objetivo deste trabalho foi analisar o padrão temporal dos índices NDVI e EVI para três safras soja com condições hídricas distintas durante o ciclo, de forma a possibilitar a avaliação da relação entre IVs e as condições hídricas para a cultura. Para 164plataforma SATveg (Sistema de Análise Temporal da Vegetação) (SATVeg, 2019) que 165 é uma ferramenta destinada Web desenvolvida pela Embrapa Informática Agropecuária, Agropecuária, ao acesso e visualização de per- <u>destinada realizar uma melhor comparação entre os dados mete</u>165é uma ferramenta Web desenvolvida pela Embrapa Informática Agropecuária, destinada fis dos índices vegetativos NDVI e EVI do sensor e IVs, os valores de precipitação pluviual (mm) 166temporais ao acesso e visualização de perfis temporais dos índices vegetativos NDVI e EVI orológicos do 166ao acesso e visualização de perfis temporais dos índices vegetativos NDVI e EVI do MODIS em qualquer local da América do Sul. medidos e os valores de ETc (mm), excesso e deficiência 167sensor MODIS em qualquer local da América do Sul. Perfil temporal dos índices de vegetação Normalized Difference Vegetation Index (NDVI) e Enhanced Vegetation Index (EVI) ao longo do crescimento e desenvolvimento das plantas de soja em três safras no período de 2015 a 2018 [NOV_1 (01/11 a 16/11/2017), NOV_2 (17/11 a 02/12/2017), DEZ_1 (02/12 a 18/12/2017), DEZ_2 (19/12/2017 a 03/01/2018), JAN_1 (04/01 a 19/01/2018), JAN_2 (20/01 a 04/02/2018), FEV_1 (05/02 a 20/02/2018), FEV_2 (21/02 a 08/03/2018), MAR_1 (09/03 a 24/03/2018) e MAR_2 (25/03 a 09/04/2018)].</p>

Fonte: Elaborado pelo autor.

O problema, nesse caso, ocorreu no processo de extração do fluxo de caracteres e pode ser resolvido editando o texto extraído. No entanto, mesmo sem editar o texto é possível que o indexador compreenda do que se trata o documento, pois o resumo tem boa informatividade e os processos cognitivos são bastantes flexíveis (experiência empírica).

Esse resumo apresenta: (1) dificuldade de leitura da última frase pela presença expressiva de informação numérica e (2) problemas semânticos entre seus elementos: (a) erros ortográficos; (b) falta de espaço entre os números e as palavras; e (c) fragmentos de frases fora de ordem.

Os problemas semânticos são difíceis de contornar por meio de algoritmos, mas são facilmente percebidos e resolvidos pela inteligência humana. A funcionalidade descrita na seção 6.4.4.1 foi planejada para resolver dar suporte na resolução desse tipo de falha, porque ela permite editar manualmente o conteúdo extraído e realizar o processo de sumarização automática novamente.

A sumarização automática não identifica nem corrige os problemas semânticos dos resumos automáticos. Na inteligência aumentada, o ser humano incluído no processo pode contornar os problemas semânticos.

A dificuldade de leitura do resumo é a mesma do artigo, pois o autor preferiu apresentar as informações numéricas (datas) diretamente no texto, porque fazia sentido para área científica da publicação, com isto se percebe como o contexto cultural repercute nas atividades automatizadas.

Na próxima seção, apresenta-se a avaliação geral dos resumos.

7.1.2 A avaliação geral dos resumos

Nesta seção, apresentam-se a avaliação do conjunto de resumos que precisariam de intervenção humana (seção 7.1.2.1); e (2) do conjunto total dos resumos (seção 7.1.2.2).

Os resumos foram avaliados manualmente quanto as seguintes características: (1) facilidade de leitura, (2) coerência semântica, e (3) utilidade das informações (ser útil para compreender o documento, estar em português e que o tamanho do resumo seja adequado), em conformidade com a seção 4.3.4.2. Foram consideradas inúteis as informações que não facilitam a compreensão do texto-fonte ou que estão em outro idioma que não seja o português, e os resumo muito curtos.

Os resumos gerados por algoritmos são bem diferentes dos resumos feitos por pessoas e ainda não existem metodologias confiáveis para compará-los. Por isso nenhuma tentativa de cotejamento foi feita entre resumos humanos e resumos artificiais.

Na próxima seção, apresenta-se a avaliação dos resumos que precisam de intervenção humana.

7.1.2.1 Avaliação do conjunto de resumos que precisam de intervenção humana

No Quadro 21, estão listados os 29 resumos (de 100) que precisam de intervenção humana, (ordenados alfabeticamente pela coluna “problema do resumo”).

Quadro 21 – Análise descritiva dos resumos que precisam de intervenção humana.

Ordem	Identificação	Problema do resumo	Causa
1	1122509	Dificuldade de leitura	Erro na extração do texto
2	958618	Dificuldade de leitura	Repetições de itens de tabelas
3	1121184	Dificuldade de Leitura	Repetição de itens
4	1008316	Dificuldade de leitura	Repetição de unidades de medidas
5	133848	Dificuldade de leitura	Muita informação numérica
6	1015548	Dificuldade de leitura	Repetições de itens de tabelas
7	1008415	Dificuldade de leitura	Artigo curto.
8	312244	Dificuldade de leitura	Repetições de itens de tabelas e do rodapé
9	89928	Dificuldade de leitura	Muitas tabelas numéricas
10	157560	Dificuldade de Leitura	Erro na extração do texto
11	105845	Dificuldade de leitura	Repetições de itens de tabelas
12	149762	Dificuldade de leitura	Erro na extração do texto
13	974846	Dificuldade de leitura	Muitas tabelas numéricas
14	1128271	Incoerência semântica	Erro na extração do texto
15	121733	Incoerência semântica	Repetições de itens de tabelas

16	1102669	Incoerência semântica	Repetições de itens de tabelas
17	312244	incoerência semântica	Repetições de itens numéricos de tabelas
18	314117	Incoerência semântica	Só faz sentido para o contexto do artigo
19	1122175	Incoerência semântica	Causa não identificada
20	156709	Incoerência semântica	Muitas tabelas numéricas
21	1126835	Incoerência semântica	Artigo com muitas tabelas numéricas
22	1127048	Incoerência semântica	As referências ocupam 32% do conteúdo
23	1124357	Falta de informatividade	O artigo está em inglês.
24	1126778	Falta de informatividade	O <i>abstract</i> desproporcional ao texto
25	1074242	Falta de informatividade	Artigo curto.
26	162329	Falta de informatividade	Artigo curto.
27	1127555	Falta de informatividade	Artigo curto.
28	1125816	Falta de informatividade	Artigo narrativo
29	315462	Falta de informatividade	Muitas tabelas numéricas

Fonte: Elaborado pelo autor.

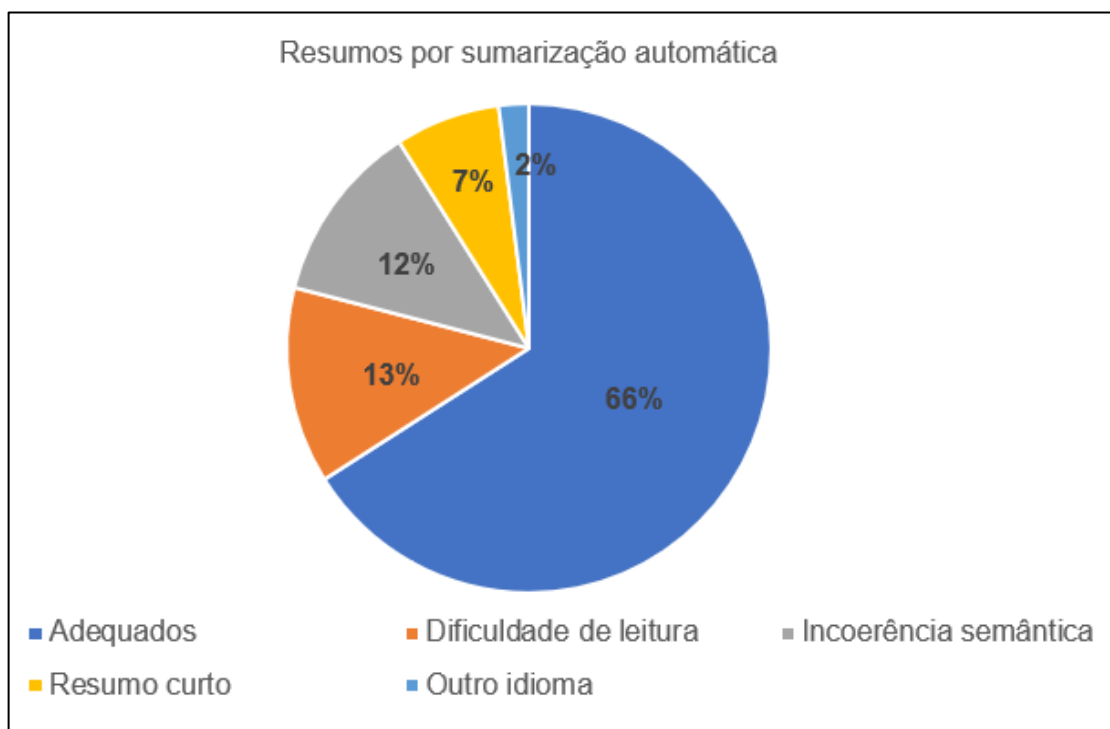
Os 29 resumos (Quadro 21) precisam de intervenção humana para se obter uma versão útil. Desse conjunto: (1) 13 resumos apresentam dificuldades à leitura devido à presença de caracteres espúrios, excessiva informação numérica e repetições descabidas; (2) 12 resumos apresentam incoerência semântica; e (3) 4 resumos têm conteúdo irrelevante.

Na próxima seção, apresenta-se a avaliação do conjunto total de resumos.

7.1.2.2 Avaliação do conjunto total de resumos

Verifica-se, na Figura 29, que a intervenção humana por meio da edição do texto extraído consiste em remover as informações repetidas em 13% dos resumos, e reorganizar os fragmentos de informação em 12% dos resumos. Em 7% dos resumos, os problemas não podem ser corrigidos no momento da indexação, porque decorrem do tamanho reduzido do texto-fonte, e, em 2% dos resumos, o problema decorrem de o *abstract* ser proporcionalmente grande em relação ao texto do artigo.

Figura 29 – A classificação dos resultados da sumarização automática.



Fonte: Elaborado pelo autor.

Assim, os problemas dos 25 primeiros resumos listados no Quadro 21 podem ser resolvidos com o uso da funcionalidade 6.4.4.1 (edição do texto extraído), enquanto os problemas dos 9 resumos que condensam informações de artigos curtos não têm uma solução no local de indexação.

Na próxima seção, avalia-se a importância de uma metodologia de inteligência aumentada na sumarização automática.

7.1.3 A inteligência aumentada na sumarização automática

A colaboração da inteligência artificial com a inteligência humana faz o processo de indexação produzir com mais rapidez as etapas automatizáveis e com mais qualidade as etapas de tomada de decisão, com isso, ela aumenta a produtividade.

Os resumos produzidos por sumarização podem ser classificados em três categorias: (1) puramente automatizados; (2) dependentes de intervenção humana; e (3) puramente intelectuais. Considerando as condições do experimento, a inteligência artificial, resumiria sozinha 66% dos arquivos e dependeria da intervenção humana em outros 25%, o da taxa de sucesso em porcentagem da sumarização automática de 91%. Apenas 9% dos artigos científicos só podem ser resumidos adequadamente exclusivamente pela inteligência humana.

Estas estimativas corroboram a ideia de que o uso combinado da inteligência humana com inteligência artificial viabiliza soluções mais rápidas pela automatização da inteligência artificial e com mais qualidade dada a intervenção humana.

Na próxima seção, apresentam-se as análises dos resultados das expressões regulares.

7.2 Expressões regulares

Nesta seção, apresentam-se os resultados das expressões regulares analisados sob duas perspectivas: (1) **formulação das expressões regulares** a partir dos termos do Thesagro, (seção 7.2.1); e (2) **identificação dos assuntos do texto** por meio de expressões regulares (seção 7.2.2).

Na próxima seção, analisam-se os resultados do processo de formulação das expressões regulares.

7.2.1 Formulação das expressões regulares

Nesta seção, apresenta-se funcionalidade do 3ApC para atualizar expressões regulares da base de conhecimento a partir dos termos do Thesagro, conforme descrito na seção 4.3.3.3.2.1.

Assim, a cada nova versão do tesouro é necessário atualizar as expressões regulares que são parte integrante da base de conhecimento. A atualização das expressões regulares é um processo computacional que só precisa ser realizado quando houver uma atualização do Thesagro.

A base de conhecimento atual do 3ApC contém as expressões regulares correspondentes aos termos da versão atual do Thesagro de 2022.

Não foram formuladas expressões regulares para 151 termos do Thesagro, porque são termos com menos de 4 letras, (exemplos: *lã*, *ar* e *mar*), ou porque são termos raros muito frequentes como *stopwords* e ambíguos (exemplo: “cerca” pode ser uma construção agrícola, ou parte da locução prepositiva “cerca de”). Na

Tabela 7, podem ser vistas informações gerais sobre as Expressões regulares.

Tabela 7 – Informações gerais das expressões regulares para o Thesagro.

Termo	Quantidade	Tipo de expressão regular
Nome Científico	2186	Uma opção
Terminado em “M”	257	Duas alternativas (m ns)
Terminado em “R”	51	Três alternativas (r res ras)
Terminado em “ão”	958	Quatro alternativas (ao aos aes oes)
Compostos	6254	Mais de quatro alternativas

Fonte: Elaborado pelo autor.

O conteúdo entre parênteses foi removido das expressões regulares, porque os qualificadores nunca aparecem nos textos, nos seguintes termos: “Relação de Direito (Relação Jurídica ou Consuetudinária)”, “Imóvel Rural (Conceito Agrário)”, “Parte (em transações que envolvem direito)”, “Proprietário (no âmbito da Governança Fundiária)”, “Família Assentada (em Projetos de Assentamento)”, “Homologação (selecionado do Programa Nacional de Reforma Agrária - PNRA)”, “Beneficiário (do Programa Nacional de Reforma Agrária PNRA)”, “Declarante (no âmbito da Gestão Fundiária)” e “Interessado (no âmbito da Gestão Fundiária)”. Uma amostra das expressões regulares pode ser vista no Quadro 22.

Quadro 22 – Uma amostra das expressões regulares formuladas para o Thesagro.

Termo	Expressão regular	Mapeia
Brassica Campestris	brassica campestris	Uma opção
Traquéia	traquei(a as)	Duas alternativas
Triturador	triturado(r res ras)	Três alternativas
Cipó Limoeiro	cip(o os) limoeir(o os)	Quatro alternativas
Pássaro Predador	passar(o os) predado(r res ras)	Seis alternativas
Cinturão de Defesa	cintur(ao aos aes oes) de defes(a as)	Dez alternativas

Fonte: Elaborado pelo autor.

Considerando todas as variações das palavras que apontam termos do Thesagro, foram formuladas mais de 27770 possíveis alternativas e a exaustividade das expressões regulares alcança 98,5% dos Termos do Thesagro.

Na próxima seção, apresenta-se o processo de identificação dos assuntos do texto por meio das expressões regulares.

7.2.2 Identificação dos assuntos do texto

A avaliação do processo de Identificação dos assuntos do texto se deu por meio de uma amostra aleatória de três artigos científicos do *corpus* do experimento, em que se verificou “manualmente” os assuntos do texto e se comparou com a lista de assuntos proposta pelo módulo.

Os três artigos científicos avaliados foram Freire *et al.* (1999), Martelli, Barros e Assis (2014) e Lemos *et al.* (2020). A avaliação consistiu em aplicar a funcionalidade do 3ApC e contar manualmente todos os assuntos do texto, considerando que os assuntos do texto são nomeados pelos termos do Thesagro.

Nos artigos analisados, a taxa de concordância entre os assuntos identificados manualmente e os identificados por algoritmos é aproximadamente 98,74%. Não chegou a 100% porque o 3ApC não reconheceu os termos que apresentaram erros na extração, exemplo: “sementesob”. Em Freire *et al.* (1999), foram reconhecidos 78 termos de um total de 79 termos, o que equivale a uma taxa de erro de 1,26%. Em síntese, pode-se afirmar que a funcionalidade expressões regulares identifica os assuntos do artigo científico analisado.

Na próxima seção, apresenta-se a análise dos resultados da funcionalidade modelagem de tópicos que corresponde ao estágio “seleção dos assuntos relevantes do texto” da análise de assuntos.

7.3 Modelagem de Tópicos

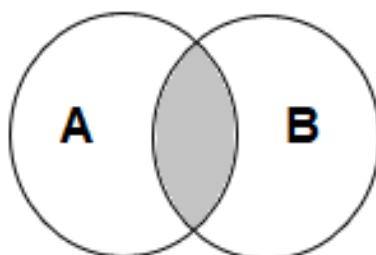
Nesta seção, descreve-se os resultados e a análise sobre a modelagem de tópicos como método para selecionar os assuntos relevantes de um texto. Começa-se pela apresentação de três casos típicos, seguida por uma análise geral dos resultados com a finalidade de validar o 3ApC.

A análise dos resultados da funcionalidade “modelagem de tópicos” foi realizada por meio de um experimento de análise de assuntos assistida por computador nos artigos científicos do *corpus* do experimento.

Nesta seção, utilizou-se o termo “intersecção” (área cinza da

Figura 30) no mesmo sentido matemático dos diagramas de *Venn*. A intersecção de dois conjuntos, representada por $A \cap B$, é um conjunto formado pelos elementos que pertencem ao conjunto A e ao conjunto B simultaneamente.

Figura 30 – A intersecção entre dois conjuntos (diagrama de Venn).



Fonte: Elaborado pelo autor.

O resultado do 3ApC refere-se ao último estágio da análise de assuntos e os descritores da BDPA referem-se à etapa de tradução da indexação. Em outras palavras, o 3ApC produz uma lista de **assuntos relevantes**, enquanto a BDPA apresenta uma lista de **descritores**. Por isso, as duas listas não podem ser comparadas diretamente, pois são conceitualmente distintas. Nesta seção, portanto, avaliou-se a intersecção entre os dois conjuntos, conforme as regras de classificação apresentadas no Quadro 23.

Quadro 23 – Classificação da intersecção entre os conjuntos.

Nº	Categoria	A intersecção entre os assuntos relevantes e os descritores
1	Excelente	Contém 100% dos descritores da BDPA
2	Muito Boa	Contém de 60% a 90% dos descritores da BDPA
3	Mediana	Contém de 10% a 50% dos descritores da BDPA
4	Ruim	Contém 0% de descritores da BDPA
5	Não se aplica	-

Fonte: Elaborado pelo autor.

Nesta seção, “descriptor” refere-se aos termos do Thesagro selecionados na BDPA para indexar o artigo, e “assunto relevante” refere-se aos termos do Thesagro selecionados no 3ApC para indicar os assuntos relevantes do artigo. Cada análise é realizada confrontando-se, para o mesmo artigo, os descritores e os assuntos relevantes. A intersecção é constituída por termos do Thesagro que estão, ao mesmo tempo, nos descritores e nos assuntos relevantes. No Quadro 23, a categoria “não se aplica” é diferente em natureza das as outras categorias, pois ela é usada para indicar que não foi possível “comparar” os dois conjuntos.

O APÊNDICE E (uma amostra pode ser vista no Quadro 24) apresenta o Identificador do artigo na BDPA, os descritores da BDPA, os assuntos relevantes selecionados pelo 3ApC (pela análise de assuntos assistida por computador) e a intersecção entre esses dois conjuntos para todos os artigos do *corpus* do experimento.

Quadro 24 – Uma amostra do APÊNDICE E.

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo 3ApC	Intersecção
207984	Absorção, Absorção de Nutrientes, Adubação, Calagem, Cerrado, Feijão, Grão, Irrigação, Nitrogênio, Nutriente, Nutriente Mineral, <i>Phaseolus Vulgaris</i> , Rendimento, Solo	Adubação, Calagem, Cerrado, Feijoeiro, Solo, Cultivar, Dose, Grão, Nutriente, Produtividade	[Adubação, Calagem, Cerrado, Solo, Grão, Nutriente]
159841	Absorção, Solo	Cádmio, Solo, Planta, Caatinga, Água, Extração, Metal, Pesquisa, Profundidade, Solução	[Solo]

Fonte: Elaborado pelo autor.

Foram analisadas qualitativamente as possibilidades de **existir um percurso** entre assuntos relevantes selecionados pelo 3ApC e os descritores da BDPA para o mesmo artigo científico em três casos emblemáticos: (1) Caso 1: intersecção ruim (seção 7.3.1); (2) Caso 2: intersecção mediana (seção 7.3.2); (3) Caso 3: intersecção excelente (seção 7.3.3); e (4) Avaliação geral dos resultados da modelagem de tópicos (7.4).

Na próxima seção, apresenta-se o Caso 1, em que não há intersecção entre os conjuntos.

7.3.1 Caso 1: intersecção ruim

Nesta seção, analisa-se a modelagem de tópicos do 3ApC, avaliando-se qualitativamente um caso em que não há intersecção entre os descritores e os assuntos relevantes. Portanto, um exemplo em que os conjuntos são disjuntos e a intersecção é ruim.

No Quadro 25, podem ser vistos, para o artigo científico de Alves *et al.* (2019), o conjunto de assuntos relevantes selecionados pelo 3ApC e o conjunto de indexadores da BDPA e a intersecção entre eles.

Quadro 25 – Caso 1: sem intersecção entre os conjuntos de termos indexadores.

Artigo	Termos indexadores da BDPA	Análise de assuntos assistida por computador	Intersecção
(ALVES <i>et al.</i> , 2019)	Agricultura Familiar, Agricultura Sustentável	Agricultor, Bem-Estar, Segurança Alimentar, Assentamento, Solo, Curso, Espécie, Família, Lote, Manejo, Pesquisa	[]

Fonte: Elaborado pelo autor.

Embora não haja uma intersecção entre os dois conjuntos, é possível traçar um percurso entre os termos dos dois conjuntos (Quadro 26). O percurso que o indexador poderia seguir (suposição) é: (1) primeiro termo sugerido pelo 3ApC, “Agricultor” é um termo relacionado (TR) à “Agricultura” que possui 25 termos específicos (TE1), incluindo “Agricultura Familiar” e “Agricultura Sustentável”.

Quadro 26 – Percurso no Thesagro entre termos do 3ApC e descritores da BDPA.

Percurso
Agricultor → (TR) Agricultura → (TE1) Agricultura Familiar
Agricultor → (TR) Agricultura → (TE1) Agricultura Sustentável

Fonte: Elaborado pelo autor.

No artigo, o termo “Agricultor” aparece 28 vezes, enquanto o termo “Agricultura” não aparece. Por isso, faz mais sentido para os processos computacionais (Pressuposto 3, seção 1.3) apontar “agricultor” como um assunto relevante do texto analisado. No entanto, supõe-se que o indexador considerou que “agricultura” descreveria melhor os assuntos do texto e que a representação deveria ser feita pelos termos específicos “Agricultura Familiar” e “Agricultura Sustentável”. Além disso, é possível que a política de indexação da biblioteca possa ter influenciado essas escolhas.

Com isso, se pode afirmar que existem percursos possíveis, entre os termos selecionados pelo 3ApC e os descritores da BDPA para o artigo de Alves *et al.* (2019), mesmo quando não há intersecção.

Na próxima seção, apresenta-se o caso em que ocorre uma intersecção mediana entre os conjuntos.

7.3.2 Caso 2: intersecção mediana

Nesta seção, avaliou-se qualitativamente os percursos possíveis na tomada de decisão para atribuição dos descritores para um caso em que a intersecção entre os assuntos relevantes e os descritores é mediana, ou seja, a intersecção contém entre 60% e 90% dos descritores da BDPA.

No Quadro 27, podem ser vistos para artigo científico de Oliveira *et al.* (2008): (1) conjunto de assuntos relevantes selecionados pela modelagem de tópicos do 3ApC, (2) conjunto de termos indexadores da BDPA e (3) a intersecção entre os dois conjuntos.

Quadro 27 – Caso 2: intersecção mediana entre os conjuntos de termos.

Artigo	Termos indexadores da BDPA	Análise de assuntos assistida por computador	Intersecção
OLIVEIRA <i>et al.</i> (2008).	Absorção, Solo.	Cádmio, Solo, Planta, Caatinga, Água, Extração, Metal, Pesquisa, Profundidade, Solução.	[Solo]

Fonte: Elaborado pelo autor.

Neste Caso 2, há 50% de concordância entre os dois conjuntos analisados, que corresponde ao termo “solo” que aparece no título e repete 55 vezes no texto. Para esse termo, o percurso é direto (Quadro 28).

Quadro 28 – Percurso semântico entre termos do 3ApC e descritores da BDPA.

Percurso
Solo → solo
Extração (por relação semântica) Absorção : a <u>extração</u> foi usada para quantificar a <u>Absorção</u> .

Fonte: Elaborado pelo autor.

O termo “absorção” aparece quatro vezes no texto e foi o termo que a BDPA utilizou como descritor, para o qual não foi encontrado um percurso direto no Thesagro com qualquer um dos termos sugeridos pelo 3ApC. No entanto, é de se observar que o termo “extração”, sugerido pelo 3ApC, aponta o método utilizado para quantificar a “absorção”, tanto que nas palavras-chave atribuídas pelos autores há o termo “extratores”. Assim, por suposição, a mediação humana é que fez a diferença, porque existe uma relação semântica entre “extração” e “absorção” que só pode ser percebida por um especialista.

Os conjuntos de termos analisados apontaram uma relação semântica entre os termos “Extração” e “Absorção”. Explica-se: a extração do elemento químico absorvido pela planta não é uma finalidade em si, o que se quer é quantificar a absorção do elemento químico pela planta, assim a repetição no texto dos termos “extração” e “extrator” apenas indicam que, no artigo, procura-se quantificar a “absorção” do Cádmiu. Essa relação extração/absorção é um percurso entre o termo “extração” proposto pelo 3ApC e o descritor “absorção” da BDPA.

Em síntese, esse Caso 2 demonstra a importância de uma metodologia de inteligência aumentada para os processos de análise de assuntos, pois as sugestões da inteligência artificial capturaram a semântica do texto, mas inteligência humana conseguiu discernir (pragmática) que a “extração” era apenas o meio de quantificar a “absorção”, essa, sim, a área de interesse do artigo científico analisado.

Na próxima seção, apresenta-se o Caso 3: intersecção Excelente entre os conjuntos analisados.

7.3.3 Caso 3: intersecção excelente

Nesta seção, avaliam-se qualitativamente os percursos possíveis entre os assuntos relevantes do texto selecionados pelo 3ApC e os descritores da BDPA para o caso em que a intersecção é classificada como excelente (ver Quadro 23).

No Quadro 29, podem ser vistos para o artigo científico de Pires *et al.* (2004): (1) os assuntos relevantes selecionados pela modelagem de tópicos do 3ApC, (2) os descritores atribuídos pela BDPA, e (3) a intersecção entre os dois conjuntos.

Quadro 29 – Caso 3: intersecção Excelente entre os conjuntos de termos.

Artigo	Termos indexadores da BDPA	Análise de assuntos assistida por computador	Intersecção
(PIRES; BRAGANTINI; COSTA, 2004)	Armazenamento, Feijão, Fungicida, <i>Phaseolus Vulgaris</i> , Revestimento, Semente	Armazenamento, Feijão, Fungicida, Fungo, Germinação, Inoculação, Patógeno, <i>Phaseolus Vulgaris</i> , Revestimento, Semente, Tratamento	[Armazenamento, Feijão, Fungicida, <i>Phaseolus Vulgaris</i> , Revestimento, Semente]

Fonte: Elaborado pelo autor.

Pode ser visto no Quadro 30, que existe um percurso direto entre para cada descritor (BDPA) nos assuntos relevantes (3ApC). Em síntese, a intersecção entre os conjuntos é excelente.

Quadro 30 – Percursos semânticos entre descritores e assuntos relevantes.

Percurso
Armazenamento → Armazenamento
Feijão → Feijão
Fungicida → Fungicida
<i>Phaseolus Vulgaris</i> → <i>Phaseolus Vulgaris</i> ,
Revestimento → Revestimento
Semente → Semente

Fonte: Elaborado pelo autor.

No Caso 3, o conjunto de descritores da BDPA está contido no conjunto de assuntos relevantes selecionados pelo 3ApC. O fato de serem seis termos de intersecção reduz a probabilidade de o resultado ser bom por mero acaso.

Na próxima seção, apresenta-se uma **avaliação geral** dos resultados da modelagem de tópicos do 3ApC.

7.3.4 Avaliação geral dos resultados da modelagem de tópicos

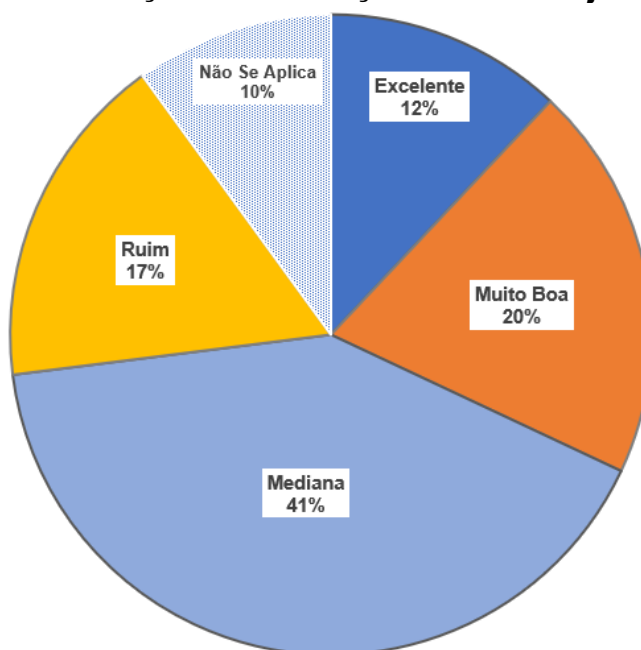
Nesta seção, apresenta-se a avaliação geral dos resultados da modelagem de tópicos do 3ApC para os artigos científicos do *corpus* do experimento.

A intersecção entre o conjunto de assuntos relevantes selecionados pelo 3ApC e o conjunto de descritores da BDPA é: (1) Excelente, em 12%, (2) Muito Boa, em 20%, (3) Mediana, em 41%, (4) Ruim, em 17%, e (5) “Não se aplica” a 10% dos casos. Nessa Análise geral, verificou-se que há **percursos diretos** entre os assuntos relevantes selecionados pela modelagem de tópicos do 3ApC e os descritores da BDPA em 77% dos artigos científicos analisados (Figura 31), que, em 17% dos casos, só há **percursos indiretos** e, nos outros 10% dos casos **não foi possível aplicar o método** devido aos problemas com o arquivo digital.

Essa análise ser numérica deve ser entendida como qualitativa, devido às seguintes razões:

1. A lista de termos da BDPA foi atribuída na **etapa de tradução** da indexação, portanto, são **descritores**. A lista de termos do 3ApC foi seleccionada na **etapa de análise de assuntos** da indexação, que é imediatamente anterior a etapa de tradução da indexação, portanto, são **assuntos relevantes** do texto.

Figura 31 – A classificação da Intersecção entre os conjuntos analisados¹⁷.



Fonte: Elaborado pelo autor.

2. Os assuntos relevantes do 3ApC e os descritores da BDPA são “aproximações” (Pressuposto 1, seção 1.3), porque não é factível antecipar os assuntos dos textos a ponto de adivinhar quais termos o usuário utilizará na busca. Tanto que Lancaster (2004) afirma que não se pode descobrir todas as ideias e significados associados a um documento e Saracevic *et al.* (1988) assevera que a coerência (intersecção), até mesmo entre diferentes especialistas humanos, é reduzida.
3. O número de termos varia nos conjuntos de descritores da BDPA entre 1 e 18 termos, enquanto no 3ApC o conjunto de assuntos tem, em média, 10 termos.

¹⁷ A percentagem de intersecção entre o conjunto de descritores na BDPA e o conjunto de assuntos relevantes sugeridos pelo 3ApC no experimento de validação.

Observação: os 10 arquivos da amostra em que não foi possível aplicar a modelagem de tópicos poderiam ter sido substituídos por arquivos sem problemas. Mas substituí-los ocultaria o fato de que a solução computacional (automatização) apresenta problemas que só o processo Intelectual pode resolver.

Os problemas na extração do texto do arquivo digital podem ser resolvidos pelo ser humano (necessário na inteligência aumentada) com auxílio da funcionalidade “edição do texto extraído” descrita na seção 6.4.4.1. Com isso, afirma-se que uma proposta metodológica de inteligência aumentada pode indexar 100% dos artigos científicos.

Os resultados desta pesquisa **não podem ser comparados** aos de outras pesquisas, porque a atribuição de termos de indexação é uma atividade subjetiva. Mas só para se ter uma ideia, os valores de coerência de pesquisas de métodos de indexação com automatizados são menores que 50%.

Na próxima seção, são feitas algumas considerações sobre esta proposta metodológica de inteligência aumentada.

7.4 A proposta metodológica de inteligência aumentada

Nesta seção, analisa-se as características desta proposta metodológica de inteligência aumentada.

Os algoritmos não percebem os significados, nem mesmo os explícitos, de forma que a inteligência humana é quem deve tomar a decisão final. Com isso em mente, para efetuar uma previsão no 3ApC, aplicam-se as regras elencadas a seguir:

1. Os descritores e os assuntos devem ser termos do tesauro (neste experimento, do Thesagro), obrigatoriamente.
2. Os termos podem ser identificados tanto no título quanto no texto.
3. Os termos nos títulos dos textos podem ser selecionados como assuntos relevantes. Títulos contém assuntos relevantes nas áreas das ciências puras e aplicadas, enquanto nas áreas de ciências humanas, sociais, linguagem e artes esse fato não ocorre sempre, nesse caso, atribuir relevância aos termos dos títulos pode aumentar a revocação. Por isso, no 3ApC, o usuário tem a opção de incluir (ou não) o título como fonte primária de assuntos relevantes.
4. Os assuntos identificados representam (substituem) o texto na modelagem de assuntos.
5. Os descritores embutem as decisões humanas, portanto, são conhecimentos.
6. Os descritores são usados para identificar um *cluster*. Supõe-se que todos os artigos indexados por um mesmo descritor pertencem a um mesmo *cluster*.

7. O indexador é capaz de compreender do que trata o documento a partir do resumo (sumarização automática).
8. O texto passa pela identificação dos termos (expressões regulares).
9. Os assuntos relevantes do texto pertencem ao tópico de um *cluster* (modelagem de tópicos).
10. Um *cluster* do texto pode conter nenhum, um ou mais de um assunto relevante.

Considera-se que os resultados do 3ApC serão sempre analisados por uma pessoa na etapa de tradução da indexação.

A regra 3 é necessária para introduzir entropia nas indexações, isto é, ela funciona como um processo estocástico que sempre adiciona informação nova no processo de indexação. Caso ela não seja adicionada, ao longo do tempo, os novos termos do tesouro e os pouco frequentes nas indexações nunca seriam utilizados. Explicando melhor: os termos mais frequentes são sugeridos mais vezes por causa do reforço estatístico; assim, os termos mais populares ficam ainda mais populares. O processo inverso também ocorre, à medida que um termo não é utilizado, as chances de ele ser indicado diminuem. A regra 3, portanto, mitiga esse viés de popularidade.

Para concluir, na próxima seção, elencam-se as considerações finais deste trabalho.

8 CONSIDERAÇÕES FINAIS

A solução dos problemas práticos muda a realidade e a solução de problemas teóricos modifica o conhecimento. Por isso, na esperança de que esta pesquisa possa alterar beneficentemente o conhecimento e a realidade, apresentam-se, a seguir, as considerações finais.

As redes de alta velocidade associadas aos computadores facilitaram a criação e o acesso ao conhecimento, provocando um crescimento desordenado e muito rápido tanto da demanda quanto da oferta de informações, o que exige da área da Biblioteconomia e da Ciência da Informação o desenvolvimento de meios rápidos de indexação, mas sem perder em qualidade.

Utilizar a automação por inteligência artificial para atender a alta demanda por informação científica nova e relevante, apesar de ser necessária, reduz a eficácia, pois provoca a perda da sensibilidade cultural, dado que a busca por informação contém dois aspectos fundamentais: (1) autor/texto/significado e (2) usuário/necessidade de informação/contexto. A máquina pode descobrir o significado estático do discurso contido no texto (Pressuposto 2 — seção 1.3). Entretanto, o segundo aspecto precisa da contribuição do indexador, porque ele compartilha (portanto, conhece melhor) o contexto social e cultural do usuário.

Considera-se que a análise de assuntos intelectual foi o evento-chave que promoveu o avanço da ciência nos últimos séculos, porque ela organizou as informações relevantes aumentando a qualidade na recuperação da informação. Dessa forma, ela permitiu que as pessoas encontrassem e usassem as melhores publicações para modificar o conhecimento e provocar mudanças no mundo.

Na Revisão de Literatura, percebeu-se que **o dilema** da indexação da Biblioteconomia e da Ciência da Informação entre adotar a velocidade da automatização ou manter a qualidade do processo intelectual pode ser resolvido pela inteligência aumentada, pois ela utiliza os resultados da inteligência artificial (garante mais velocidade) para assistir a Inteligência humana (garante mais qualidade) nas atividades documentárias. Vale ressaltar que não foram encontrados trabalhos publicados sobre a inteligência aumentada aplicada à indexação ou a sua etapa de análise de assuntos.

O problema da indexação Intelectual é que ela demanda mais tempo e maiores custos, além de incluir a subjetividade natural do intelecto humano nos resultados, de forma que ela tem se revelado inadequada para o cenário atual.

A proposta metodológica desta pesquisa baseou-se nos princípios da inteligência aumentada de manter as pessoas no centro das decisões, auxiliadas pelos métodos

automatizados. Essa arquitetura de solução produziu o processo de análise de assuntos Assistida por Computador, o que faz desta uma **pesquisa social aplicada**, em que se planejou e desenvolveu meios para automatizar as tarefas repetitivas das atividades laborais humanas e, com isso, promove uma colaboração entre máquinas e pessoas e afasta a alternativa antagônica de usar as máquinas para substituir as pessoas.

A implementação do protótipo tecnológico denominado de *software* de **análise de assuntos assistida por computador** (3ApC) enquadrou-se na fase de transição da revolução da inteligência artificial, saindo da Era da Especialidade para entrar na Era da Implementação, em que os avanços significativos dependem dos dados (*big data*). Assim, o *software* 3ApC faz parte da era de implementação da inteligência artificial, em que a inovação ocorre no uso que se faz da base de conhecimento e da interação humana.

O 3ApC é uma solução para a área da Organização da Informação em meio a outras soluções, que incluem o MHTX de Lima (2004), a Reengenharia de Tesouro de Maculan (2015), o Semantizar de Maia (2018) e o enriquecimento de ontologias de domínio de Araújo (2021). Essa abundância de soluções ocorre devido à explosão da produção textual e às mudanças linguísticas, sociais e tecnológicas que demandam sempre novos desenvolvimentos.

Esta pesquisa é **empírica**, porque analisou dados reais para extrair *insights* e inovação tecnológica sem uma hipótese estatística. Ou seja, ela é orientada aos dados, às evidências, em que se testou computacionalmente todas as probabilidades em busca das melhores respostas, e, com isso, cobriu mais possibilidades que uma análise puramente intelectual. Nesse sentido, adicionou-se ao desenvolvimento tecnológico a elaboração de uma base de conhecimento, em que se armazenaram os dados e informações elaborados e acumulados na Base de Dados de Pesquisa em Agropecuária (BDPA). Além disso, nesta pesquisa, foram elicitadas informações latentes nas relações entre termos do Thesagro, nos artigos científicos e nos termos indexadores, que adicionadas na base de conhecimento como recursos informacionais adicionais para auxiliar a tomada de decisões na etapa de tradução do processo indexação. Assim, o indexador pode ser municiado com conhecimentos acumulados ao longo de duas décadas em várias bibliotecas da Embrapa. A base de conhecimento elaborada está descrita na seção 5.4.

Dada sua natureza aplicada, a fundamentação teórica e metodológica desta pesquisa desdobrou-se em dois eixos principais: no primeiro, abordou a Análise de assuntos, o objeto da pesquisa; e, no segundo, apresentou as abordagens metodológicas do processamento de linguagem natural da inteligência artificial necessárias ao desenvolvimento da aplicação computacional de assistência ao processo da análise de assuntos.

Esta pesquisa adotou os princípios da **Pesquisa Experimental**, em que os artefatos foram desenvolvidos e testados por métodos experimentais sem a formulação de hipóteses, pois não foram encontrados estudos de inteligência aumentada para a análise de assuntos. Uma das consequências de se ter uma metodologia experimental foi que o modelo sequencial com mecanismo de atenção, desenvolvido e testado nesta pesquisa, foi excluído deste relato, porque não produziu resultados satisfatórios (< 50% de efetividade), apesar de ter consumido muito tempo e muitos recursos computacionais (arquitetura de redes neurais profunda com 64 camadas intermediárias com 1024 neurônios artificiais em cada uma, 65536 neurônios na nuvem computacional do *Google Colab Pro* com *Tensor Processor Unit*). O relato de pesquisa do método de inteligência artificial baseado em modelos sequenciais com mecanismo de atenção pode ser visto no APÊNDICE G.

O planejamento do *software* 3ApC representa uma das facetas da inteligência aumentada: padrão de projeto. A diretriz subjacente de não substituir o trabalho humano, mas usar a máquina para assistir o profissional da informação em suas atividades representa a segunda faceta da inteligência aumentada: processo de uso. O uso combinado do 3ApC e da base de conhecimento dá ao indexador uma visão ampla e abrangente do cenário analisado, facilitando a percepção dos significados do texto associados aos descritores, o que representa a terceira faceta da inteligência aumentada: a expansão cognitiva.

O objetivo geral desta pesquisa foi elaborar uma proposta metodológica de um modelo de inteligência aumentada para viabilizar a análise de assuntos assistida por computador, e, com isso, beneficiar os profissionais da informação (bibliotecários, analistas de assunto e indexadores) e pesquisadores que necessitem de processar informações em grandes repositórios textuais (*big data*). Para alcançar esse objetivo geral, a base de conhecimento (ver capítulo 5) e o protótipo tecnológico (3ApC) (ver capítulo 6) foram avaliados por um experimento de validação, cujos resultados foram analisados no capítulo 7. Assim, as respostas às Questões de Pesquisa podem ser lidas nos próximos parágrafos.

Quais recursos informacionais e métodos computacionais são necessários e suficientes para assistir adequadamente a análise de assuntos? A elaboração da base de conhecimento do 3ApC, que viabilizou o sucesso experimental da pesquisa, permite afirmar que o acervo da BDPA (artigos científicos e seus respectivos descritores, e os termos do Thesagro e suas respectivas relações) são recursos informacionais necessários e suficientes para assistir a análise de assuntos no campo da pesquisa Agropecuária nacional. Os métodos de inteligência artificial que também viabilizaram o êxito experimental da análise de assuntos assistida por computador, nesta pesquisa, foram; (1) a sumarização automática, que permitiu ao indexador apreender o sentido do texto analisado; (2) expressões regulares, que identificaram os assuntos do texto; e (3) a modelagem de tópicos (Alocação Dinâmica de

Dirichlet), que seleciona os assuntos relevantes do texto. Portanto, recursos informacionais e métodos computacionais descritos são necessários e suficientes para assistir adequadamente a análise de assuntos.

Em que medida o método da Alocação Dinâmica de *Dirichlet* (BLEI et al., 2003) pode contribuir efetivamente para a indexação? O Experimento da pesquisa permitiu afirmar que, considerando a comparação entre a indexação da BDPA (feita por pessoas) e os resultados obtidos com o 3ApC, o método da Alocação Dinâmica de *Dirichlet* contribuiu com a indexação em uma medida superior a 79% de efetividade. Importante frisar que a proposta metodológica de inteligência aumentada tem efetividade de 100%, uma vez que nos casos em que a análise de assuntos assistida por computador não for efetiva, a inteligência humana incluída no processo pode tomar todas as providências necessárias para atribuir os descritores adequados ao artigo científico em análise.

A inteligência artificial aplicada a uma base de conhecimento é suficiente para assistir adequadamente o processo da análise de assuntos? O experimento de validação do 3ApC permitiu verificar que os algoritmos processam o número que representam as palavras e não os significados delas, por isso as máquinas não são capazes de substituir inteiramente o trabalho intelectual humano, apesar da *hype* e das promessas em contrário. O que se pode afirmar é que as máquinas são servidoras incansáveis, porém limitadas, e que elas são, essencialmente, boas ferramentas quando submetidas à supervisão humana e, nessa condição, elas podem assistir adequadamente o processo de análise de assuntos. Os resultados analisados no Capítulo 7 demonstraram que o método de modelagem de assuntos do 3ApC apresenta listas de termos descritores surpreendentemente similares às feitas por pessoas e que as informações adicionais disponibilizadas pelo 3ApC dão ao indexador uma visão ampla do contexto, facilitando a próxima etapa da indexação que é a tradução dos assuntos relevantes em termos do Thesagro (atribuição dos descritores).

Não foi possível estabelecer o tempo necessário para realizar uma análise de assuntos intelectual, pois isso depende de fatores extrínsecos ao indexador como qualidade e o tamanho do texto, e de fatores intrínsecos como o grau de especialização, motivação e nível de estresse. Por isso as métricas de ganho temporal com o uso do 3ApC não podem ser aferidas, mas presume-se que utilizá-lo reduz o tempo e os esforços necessários para fazer uma indexação.

As contribuições desta pesquisa incluem (mas não se limitam) a:

1. Compreendeu-se que automatizar resolve o problema da velocidade de produção da indexação, mas que é necessário a presença da inteligência humana no processo para garantir a qualidade da indexação.

2. A análise de assuntos Intelectual resolve as questões relacionadas à qualidade da indexação, mas, dado ao volume de publicações, ela precisa da assistência das máquinas para aumentar a velocidade da produção indexadora.
3. Os métodos de automatização por inteligência artificial incrementam a velocidade da indexação, mas são abordagens simplistas que não favorecem o compartilhamento eficiente das informações, nem a experiência acumulada nas instituições, o que pode acarretar resultados medíocres na pesquisa científica. Por outro lado, esses métodos apresentam características vantajosas como capturar o significado do texto com mais velocidade do que uma pessoa no processo de indexação intelectual.
4. A vantagem do uso dos processos intelectuais é que o indexador, por compartilhar o contexto social e cultural com o usuário, compreende melhor as necessidades de informações, codificando-as com mais precisão do que nos processos puramente automatizados.

Em síntese, a contribuição da proposta metodológica desta pesquisa é propor um processo de interação da inteligência humana com as ferramentas inteligentes, cuja meta principal é pôr o ser humano no centro do processo a fim de que a tomada de decisão seja essencialmente humana. A ideia é que a máquina participa apenas do processo técnico repetitivo, enquanto a inteligência humana participa de todo o processo, da coleta de dados à decisão final. Assim, se obtém a inteligência aumentada que pode promover um aumento da produtividade, da assertividade e do bem-estar humano, enquanto garante robustez à solução técnica.

As limitações desta pesquisa incluem as questões relacionadas à integração do *software* em ambientes reais de trabalho e à natureza específica dos métodos de Inteligência artificial. Quanto à integração do *software*, a limitação reside em desenvolver uma adaptação da base de dados preexistente na instituição que adeque (1) os artigos indexados, (2) as listas de termos indexadores e (3) o tesauro utilizado. Quanto à inteligência artificial, as limitações advêm da natureza empírica dos métodos em que as inferências são feitas diretamente nos dados sem um modelo matemático explicativo de forma que não existem testes sistemáticos e precisos para averiguar a correção dos modelos inteligentes.

As seguintes entregas foram produzidas: um estudo teórico da inteligência aumentada; uma ferramenta tecnológica (3ApC); e um experimento de validação, e com isso, pode-se concluir que:

1. a revisão de literatura contribuiu com um estudo sobre a inteligência aumentada na Organização da Informação, que permitiu compreender como a tecnologia e a inteligência artificial podem auxiliar a indexação, de modo que o Objetivo Específico 1 foi atingido;
2. a elaboração do Modelo Conceitual elencou e articulou as boas práticas disponíveis na literatura científica dos campos científicos correlacionados, ancorado na experiência pessoal dos envolvidos na pesquisa, em que foram necessários conhecimentos avançados em programação de computadores, inteligência artificial, análise de assuntos e indexação. A elaboração dessa metodologia de análise de assuntos assistida por computador significa que Objetivo Específico 2 foi atingido;
3. o desenvolvimento do artefato tecnológico 3ApC viabilizou a inteligência aumentada no processo de indexação temática, conforme demonstrado na análise dos resultados. Assim, atingiu-se o Objetivo Específico 3;
4. o artefato tecnológico 3ApC foi avaliado por um experimento computacional, em que se analisou os assuntos dos artigos científicos do *corpus* do experimento, por meio da inteligência artificial, a partir das informações acumuladas na base de conhecimento. Assim, ao se validar a solução desenvolvida, o objetivo específico 4 foi atingido.

Em relação às pesquisas futuras sugere-se: (1) desenvolver novos modelos de inteligência artificial compatíveis com o 3ApC para outros vocabulários controlados; (2) investigar as repercussões da inteligência aumentada nas tarefas de indexação para descobrir os impactos sociais consequentes do uso de tecnologias inteligentes na recuperação da informação científica tanto no curto quanto no longo prazo; (3) pesquisar a interação dos profissionais da informação com a ferramenta computacional 3ApC para analisar as métricas relativas à precisão e à revocação na recuperação da informação com o intento de propor melhorias à interface gráfica de usuário e aos processos de análise de assuntos; (4) desenvolver técnicas adequadas para integrar múltiplas Bases de Conhecimentos desenvolvidas em locais distintos em que o *software* tenha sido utilizado; e (5) identificar os documentos que o usuário efetivamente recupera com cada termo digitado na expressão de busca para estabelecer/conhecer a relação empírica entre o termo digitado e o documento, na perspectiva do usuário.

Neste trabalho pesquisou-se a análise de assuntos assistida por computador (humano auxiliado pela máquina), em que se evitou as propostas que implicassem em

substituição de profissionais da informação por máquinas, e, ao mesmo tempo, procurou-se o equilíbrio entre manter a qualidade e aumentar a produção de indexações.

Todas essas considerações permitem afirmar que esta pesquisa elaborou uma proposta metodológica de um modelo de inteligência aumentada que viabiliza a análise de assuntos assistida por computador, e a conclusão é que o objetivo geral foi plenamente alcançado.

REFERÊNCIAS

- ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO 12676**: Métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992. 4 p.
- AKATA, Z. *et al.* A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. **Computer**. Long Beach, v. 53, n. 8, p. 18-28. DOI: 10.1109/MC.2020.2996587. Disponível em: <https://dx.doi.org/10.1109/MC.2020.2996587>. Acesso em: 29 nov. 2022
- ALMEIDA, F. R.; MARQUES, M. B. O impacto social da informação: a prestação do serviço de informação à comunidade. **Páginas a&b**: arquivos e bibliotecas. Porto, Portugal, V. 3, n. Especial, p. 3-19, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/69586>. Acesso em: 29 nov. 2022.
- ALBRECHTSEN, H. Subject analysis and indexing: from automated indexing to domain analysis. **The Indexer: The International Journal of Indexing**. Liverpool, UK, v. 18, n. 4, p. 219–224. 10 Jan. 1993. ISSN: 0019-4131. DOI: 10.3828/indexer.1993.18.4.3. Disponível em: <https://dx.doi.org/10.3828/indexer.1993.18.4.3>. Acesso em: 29 nov. 2022.
- ALDOUS, D. J. Exchangeability, and related topics. In: ALDOUS, D. J.; IBRAGIMOV, I. A.; JACOD, J. (Eds.). **École d'Été de Probabilités de Saint-Flour XIII** — 1983. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1985. v. 1117 p. 1–198. DOI: 10.1007/bfb0099421. Disponível em: <http://dx.doi.org/10.1007/bfb0099421>. Acesso em: 29 nov. 2022.
- ALVES, J. C. *et al.* Sistemas agroflorestais biodiversos: segurança alimentar e bem-estar às famílias agricultoras. **Revista GeoPantanal**. Corumbá, v. 26, n. 1, p. 75–94, Jan./jun. 2019. ISSN: 2446-8681. Disponível em: <https://periodicos.ufms.br/index.php/revgeo/article/view/9351/7215#:~:text=Constatou%2Dse%20que%20os%20SAFs,a%20diversifica%20C3%A7%C3%A3o%20de%20atividades%20produtivas%2C>. Acesso em: 29 nov. 2022.
- ALVES, L. L. **Projeto de sistemas de recuperação de informação corporativa: uma abordagem de análise de domínio baseada na análise facetada**. 400 f. Tese de Doutorado. Belo Horizonte: UFMG, 2014.
- ANTIQUUEIRA, L. *et al.* A complex network approach to text summarization. **Information Sciences**. 2009, v. 179, n. 5, p. 584–599. DOI: j.ins.2008.10.032. Disponível em: <https://doi.org/10.1016/j.ins.2008.10.032>. Acesso em: 29 nov. 2022.
- ARAÚJO, W. J. **Proposta metodológica para enriquecimento de ontologias de domínio**. 400 f. Tese de Doutorado — Belo Horizonte: Universidade Federal de Minas Gerais - UFMG, 17 set. 2021.
- ASHBY, W. R. **An introduction to cybernetics**. 2ed. London: Chapman & Hall, 1956. 295 p.
- AXELOS. **ITIL foundation**: ITIL 4 edition. 1 ed. Norwich: TSO (The Stationery Office), 2019. 212 p. ISBN: 978-0-11-331607-6.
- BALYAN, R.; MCCARTHY, K. S.; MCNAMARA, D. S. Applying Natural Language Processing and Hierarchical Machine Learning Approaches to Text Difficulty Classification. **International Journal of Artificial Intelligence in Education**, [s. l.] v. 30, n. 3, p. 337–370, out. 2020.

DOI: 10.1007/s40593-020-00201-7. Disponível em: <https://doi.org/10.1007/s40593-020-00201-7>. Acesso em: 29 nov. 2022.

BATES, M. J. Indexing and access for digital libraries and the internet: Human, database, and domain factors. **Journal of the American Society for Information Science**. Los Angeles, v. 49, n. 13, p. 1185–1205, 1988. DOI: 10.1002/(SICI)1097-4571(1998110)49:13%3C1185::AID-ASI6%3E3.0.CO;2-V. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(1998110\)49:13%3C1185::AID-ASI6%3E3.0.CO;2-V](https://doi.org/10.1002/(SICI)1097-4571(1998110)49:13%3C1185::AID-ASI6%3E3.0.CO;2-V). Acesso em: 29 nov. 2022.

BAXENDALE, P. B. Machine-Made index for technical literature – an experiment. **IBM Journal of Research and Development**. [s. l.], v. 2, n. 4, p. 354-361, out. 1958. DOI: 10.1147/rd.24.0354. Disponível em: <https://doi.org/10.1147/rd.24.0354>. Acesso em: 29 nov. 2022.

BELKIN, N. J. A. Anomalous states of knowledge as a basis for information retrieval. **The Canadian Journal of Information Science**. Toronto, v. 5, n. 1, p. 133-143, 1980. Disponível em: <https://tefkos.comminfo.rutgers.edu/Courses/612/Articles/BelkinAnomalous.pdf>. Acesso em: 29 nov. 2022.

BENGIO, Y. *et al.* A Neural Probabilistic Language Model. **Journal of Machine Learning Research**. Montreal, v. 3, n. 2, p. 1137–1155, 2003. Disponível em: <https://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf>. Acesso em: 22 nov. 2022.

BIRD, S.; KLEIN, E.; LOPER, E. **Natural language processing with Python**. 1 ed. Beijing; Cambridge: O'Reilly, 2009. 479 p. ISBN: 978-0-596-51649-9.

BLEI, D.; CARIN, L.; DUNSON, D. Probabilistic Topic Models. **IEEE Signal Processing Magazine**. Princeton, v. 27, n. 6, p. 55-65, nov. 2010. ISBN: 1053-5888. DOI: 10.1109/MSP.2010.938079. Disponível em: <https://doi.org/10.1109/MSP.2010.938079>. Acesso em: 29 nov. 2022.

BLEI, D. M.; NG, A. Y.; JORDAN, I. M. Latent Dirichlet Allocation. **Journal of Machine Learning Research**. [S.l.], v. 3, p. 993-1022, jan. 2003. Disponível em: <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>. Acesso em: 29 nov. 2022.

BORKO, H. Toward a theory of indexing. **Information Processing & Management**. Grã-Bretanha, v. 13, n. 6, p. 355-365, jul. 1977. DOI: 10.1016/0306-4573(77)90055-3. Disponível em: [https://doi.org/10.1016/0306-4573\(77\)90055-3](https://doi.org/10.1016/0306-4573(77)90055-3). Acesso em: 29 nov. 2022.

BORKO, H.; BERNIER, C. L. **Abstracting concepts and methods**. New York: Academic Press, 1975. ISBN: 978-0-12-118650-0. 250 p.

BOSCHERT, S. *et al.* **Symbiotic Autonomous Systems - White Paper III**. 1. ed. IEEE: Digital Reality, 2019. v. 3. 230 p. Disponível em: https://digitalreality.ieee.org/images/files/pdf/1SAS_WP3_Nov2019.pdf. Acesso em: 29 nov. 2022.

BUSH, V. As we may think. **The Atlantic Monthly**. Atlantic, p. 112 -124, 15 jul. 1945. Disponível em: [http://worrydream.com/refs/Bush%20-%20As%20We%20May%20Think%20\(Life%20Magazine%201945\).pdf](http://worrydream.com/refs/Bush%20-%20As%20We%20May%20Think%20(Life%20Magazine%201945).pdf). Acesso em: 29 nov. 2022.

CERF, V. G. Augmented Intelligence. **IEEE Internet Computing**. [S.l.], v. 17, n. 5, p. 96-96, 2013. ISSN: 1941-0131. DOI: 10.1109/MIC.2013.90. Disponível em: <https://doi.org/10.1109/MIC.2013.90>. Acesso em: 29 nov. 2022.

CHAPMAN, P. *et al.* **CRISP-DM 1.0**: Step-by-step data mining guide. 1 ed. [S.l.]:SPSS, 2000. 76 p. Disponível em: <https://maestria-datamining-2010.googlecode.com/svn-history/r282/trunk/dmct-teorica/tp1/CRISPWP-0800.pdf>. Acesso: 02 fev. 2021.

CASTELLS, M. **The Internet Galaxy**: reflections on the internet, business, and society. 1 ed. New York: Oxford Univ. Press, 2003. 292 p. ISBN:978-0-19-924153-8.

CHASTINET, Y. S. *et al.* Análise da expansão do serviço de bibliografias personalizadas em agricultura (BIP/AGRI) - um serviço brasileiro de disseminação seletiva da informação. **Revista de Biblioteconomia de Brasília**. Brasília, v. 6, n. 2, p. 217–229, 1978. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/72605>. Acesso em: 29 nov. 2022.

CINTRA, A. M. M. Estratégias de Leitura em Documentação. In: SMITH, J. W. **Análise documentária**: a análise da síntese. Brasília – DF: Johanna W. Smit, 1987. 37 p.

CRESWELL, J. W.; CRESWELL, J. D. **Research design: qualitative, quantitative, and mixed methods approach**. 5 ed. Los Angeles: SAGE, 2018. 273 p.

CUNHA, I. M.; KOBASHI, N. Y. Análise Documentária e inteligência artificial. **Revista Brasileira de Biblioteconomia e Documentação**. São Paulo, v. 24, p. 38–52, 1991. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/4469>. Acesso em: 29 nov. 2022.

DAVENPORT, T. H. From analytics to artificial intelligence. **Journal of Business Analytics**. v. 1, n. 2, p. 73–80, 2018. DOI: 10.1080/2573234X.2018.1543535. Disponível em: <https://doi.org/10.1080/2573234X.2018.1543535>. Acesso em: 29 nov. 2022.

DAVENPORT, T. H.; HARRIS, J. G.; MORISON, R. **Inteligência Analítica nos Negócios**. Tradução: Ana Beatriz Rodrigues. 1. ed. Rio de Janeiro: Elsevier, 2010. 237 p.

DEERWESTER, S. *et al.* Indexing by latent semantic analysis. **Journal of the American Society for Information Science**, 1990. v. 41, n. 6, p. 391–407.

DENG, L.; LIU, Y. **Deep Learning in Natural Language Processing**. Singapore: Springer Singapore, 2018. DOI: 10.1007/978-981-10-5209-5. Disponível em: <https://doi.org/10.1007/978-981-10-5209-5>. Acesso em: 29 nov. 2022.

DIAS, E. W.; NAVES, M. M. L. **Análise de assuntos**: teoria e prática. Brasília, DF: Thesaurus, 2007. ISBN: 978-85-7062-620-2. 116 p.

DOMINGOS, P. **O Algoritmo Mestre**: como a busca pelo algoritmo de *machine learning* definitivo recriará nosso mundo. Tradução: Aldir José Coelho Corrêa Silva. 1a ed. São Paulo: Novatec, 2017. 377 p.

EKBIA, H. R. **Artificial Dreams**: The Quest for Non-Biological Intelligence. Cambridge; New York: Cambridge University Press, 2008. 416 p.

EKBIA, H. R. Fifty years of research in artificial intelligence. **Annual Review of Information Science and Technology**. New York, v. 44, n. 1, p. 201–242, 2010. DOI: 10.1002/aris.2010.1440440112. Disponível em: <https://doi.org/10.1002/aris.2010.1440440112>. Acesso em: 29 nov. 2022.

EMBRAPA. **O Sistema - Portal Embrapa**. Institucional. Disponível em: <https://www.embrapa.br/seb/o-sistema>. Acesso em: 1 jun. 2022.

ENGELBART, D. **Augmenting Human Intellect: A Conceptual Framework**. 1. ed. Arlington: [s.n.], 1962. Disponível em: <https://apps.dtic.mil/sti/pdfs/AD0289565.pdf>. Acesso em: 22 nov. 2022.

FARROW, J. F. A Cognitive Process Model of Document Indexing. **Journal of Documentation**. [S.l.], v. 47, n. 2, p. 149–166, 1 fev. 1991. DOI: <https://doi.org/10.1108/eb026875>.

FERREIRA, A. C.; MACULAN, B. C. M. DOS S. Metodologia para a análise de assuntos de acórdãos no contexto do controle externo: proposta de um modelo de leitura técnica. **Em Questão**, 2019. v. 25, n. 3, p. 99–131. DOI: 10.19132/1808-5245253.99-131. Disponível em: <https://doi.org/10.19132/1808-5245253.99-131>. Acesso em: 29 nov. 2022.

FIGUEIREDO, N. Da importância dos artigos de revisão da literatura. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 23, n. 4, p. 131–135, 1990. Disponível em: https://brapci.inf.br/_repositorio/2011/09/pdf_6245ece57c_0018790.pdf. Acesso em: 29 nov. 2022.

FITZGERALD, M. **Introducing regular expressions: unraveling regular expressions, step-by-step**. 1. ed. Beijing Köln: O'Reilly, 2012. 154 p. ISBN: 978-1449392680.

FIORINI, N. et al. Fast and reliable inference of semantic clusters. **Knowledge-Based Systems**, 2016. v. 111, p. 133–143. DOI: 10.1016/j.knosys.2016.08.008. Disponível em: <https://doi.org/10.1016/j.knosys.2016.08.008>. Acesso em: 29 nov. 2022.

FONG, A. C. M.; HONG, G.; FONG, B. Augmented Intelligence with Ontology of Semantic Objects. 2019 International Conference on Contemporary Computing and Informatics (IC3I). **Anais...** In: 2019 INTERNATIONAL CONFERENCE ON CONTEMPORARY COMPUTING AND INFORMATICS (IC3I). Singapore, Singapore: IEEE, dez. 2019. DOI: <https://doi.org/10.1109/IC3I46837.2019.9055577>.

FOSKETT, A. C. **A abordagem temática da informação**. Tradução: A. A. Briquet de Lemos. São Paulo: Polígono, 1973. 437 p.

FREIRE, M. S. et al. Germoplasma de caupi: coleção ativa e de base. In: GONZAGA NETO, L. (Ed.). **Melhoramento genético da aceroleira na Embrapa Semiárido**. In GOEDERT, C.O., RAMOS, S.R.R., ed. Recursos genéticos e melhoramento de plantas para o Nordeste brasileiro. Petrolina, PE: QUEIROZ, M.A. de; Embrapa Semiárido; Embrapa Recursos Genéticos e Biotecnologia, 1999. ISBN: 978-85-7405-001-0.

FROWNELTER, J. et al. Impact of augmented intelligence (AI) on identification and management of depression in oncology. **Journal of Clinical Oncology**, 2020. v. 38, n. 15_suppl, p. e14059–e140590. DOI: 10.1200/JCO.2020.38.15_suppl.e14059. Disponível em: https://doi.org/10.1200/JCO.2020.38.15_suppl.e14059. Acesso em: 29 nov. 2022.

FUJITA, M. S. L.; LEIVA, I. Avaliação da indexação por meio da recuperação da informação. **Ciência da Informação**. Brasília, v. 43, n. 1, p.50-66, 2014. DOI: 10.18225/ci.inf.v43i1.1418. Disponível em: <https://brapci.inf.br/index.php/res/download/55753>. Acesso em: 29 nov. 2022.

FUJITA, M. S. L. *et al.* **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias**. São Paulo – SP: Cultura Acadêmica, 2009. ISBN: 978-85-7983-015-0.

GAMBHIR, M.; GUPTA, V. Recent automatic text summarization techniques: a survey. **Artificial Intelligence Review**. [S.l.], v. 47, n. 1, p. 1–66, 2017. ISSN 1980-6949. DOI:

10.1007/s10462-016-9475-9. Disponível em: <https://doi.org/10.1007/s10462-016-9475-9>. Acesso em: 29 nov. 2022.

GOLDBERG, Y. Neural Network Methods for Natural Language Processing. **Synthesis Lectures on Human Language Technologies**. [S.l.], v. 10, n. 1, p. 1–309, 17 abr. 2017. DOI: 10.2200/S00762ED1V01Y201703HLT037. Disponível em: <http://www.morganclaypool.com/doi/10.2200/S00762ED1V01Y201703HLT037>. Acesso em: 29 nov. 2022.

GOLESTAN, S.; MAHMOUDI-NEJAD, A.; MORADI, H. A Framework for Easier Designs: Augmented Intelligence in Serious Games for Cognitive Development. **IEEE Consumer Electronics Magazine**. [S.l.], v. 8, n. 1, p. 19–24, 2019. DOI: 10.1109/MCE.2018.2867970. Disponível em: <https://doi.org/10.1109/MCE.2018.2867970>. Acesso em: 29 nov. 2022.

GOLUB, K. Automated Subject Indexing: An Overview. **Cataloging & Classification Quarterly**. [S.l.], v. 59, n. 8, p. 702–719, 13 dez 2021. DOI: 10.1080/01639374.2021.2012311. Disponível em: <https://doi.org/10.1080/01639374.2021.2012311>. Acesso em: 29 nov. 2022.

GOLUB, T. P. *et al.* Approaches to Automatic Summarization and Annotation. **Науковий вісник ДДПУ імені Івана Франка. Серія: Філологічні науки (мовознавство)**, 2021. n. 16, p. 12–17. DOI: <https://doi.org/10.24919/2663-6042.16.2021.2>.

GOMES, H. E. **Manual de elaboração de tesouros monolíngües**. 1. ed. Brasília - DF: Imprensa Universitária, 1990. 78 p. Disponível em: <http://www.dominiopublico.gov.br/download/texto/me002423.pdf>. Acesso em: 29 nov. 2022.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. 1 ed. Cambridge, Massachusetts: The MIT Press, 2016. ISBN: 978-0-262-03561-3. 775 p.

GRAY, W. A.; HARLEY, A. J. Computer assisted indexing. **Information Storage and Retrieval**, 1971. v. 7, n. 4, p. 167–174. DOI: 10.1016/0020-0271(71)90002-7. Disponível em: [https://doi.org/10.1016/0020-0271\(71\)90002-7](https://doi.org/10.1016/0020-0271(71)90002-7). Acesso em: 29 nov. 2022.

GRIFFEY, J. **Artificial intelligence, and machine learning in libraries**. Chicago, Illinois: ALA TechSource, 2019. ISBN: 978-0-8389-1814-2. 32 p.

GUIGNARD, J.-B.; PAZ, O.; SAVAROCHE, K. HAI as Human Augmented Intelligence: from Cognitive Biases to the Nature of Cognitive Technology. In: **AI as Augmented Intelligence**. Paris, France: Université Paris sciences et lettres, 2020.

HAYKIN, S. S. **Neural network: a comprehensive foundation**. 2nd ed. Upper Saddle River, N.J: Prentice Hall, 1999. ISBN: 978-0-13-273350-2. 842 p.

HENNING, P. C. *et al.* Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, 2019. v. 14, n. 3, p. 175-192. DOI: 10.22478/ufpb.1981-0695.2019v14n3.46969. Disponível em: <https://doi.org/10.22478/ufpb.1981-0695.2019v14n3.46969>. Acesso em: 29 nov. 2022.

HEYLIGHEN, F. **A cybernetic perspective on the new science of the mind**. 1. ed. Brussel: Vrije Universiteit Brussel, 2015. 171 p.

HEVNER *et al.* Design Science in Information Systems Research. **Management Information Systems Quarterly**. Minnesota, v. 28, n. 1, p. 75-105, 2004. DOI:

10.2307/25148625. Disponível em: <https://doi.org/10.2307/25148625>. Acesso em: 29 nov. 2022.

HIRSCHBERG, J.; MANNING, C. D. Advances in natural language processing. **Artificial Intelligence**. v. 349, n. 6245, p. 261–268. DOI: 10.1126/science.aaa8685. Disponível em: <https://doi.org/10.1126/science.aaa8685>. Acesso em: 29 nov. 2022.

HIRSCHBERG, J.; MANNING, C. D. Advances in natural language processing. **Artificial Intelligence**. [S.l.], v. 349, n. 6245, p. 261–268. Disponível em: <https://nlp.stanford.edu/~manning/xyzzzy/Hirschberg-Manning-Science-2015.pdf>. Acesso em: 29 nov. 2022.

HINTON, G. **The Next Generation of Neural Networks**. In: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 43., Virtual Event China. **Anais [...]**. Virtual Event China: ACM, 2020. DOI: 10.1145/3397271.3402425. Disponível em: <https://dl.acm.org/doi/10.1145/3397271.3402425>. Acesso em: 29 nov. 2022.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: Domain-analysis. **Journal of the American Society for Information Science**. Estados Unidos, v. 46, n. 6, p. 400–425, 1995. DOI: 10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199507\)46:6<400::AID-ASI2>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y). Acesso em: 29 nov. 2022.

HJØRLAND, B. Is classification necessary after Google? **Journal of Documentation**. [S.l.], v. 68, n. 3, p. 299–317, 2012. DOI: 10.1108/00220411211225557. Disponível em: <https://doi.org/10.1108/00220411211225557>. Acesso em: 29 nov. 2022.

HOFMANN, T. **Probabilistic latent semantic indexing**. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99. **Anais...** In: THE 22ND ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE. Berkeley, California, United States: ACM Press, 1999. Disponível em: <http://portal.acm.org/citation.cfm?doid=312624.312649>. Acesso em: 3 jun. 2022.

HUTCHINS, W. J. On the problem of “aboutness” in document analysis. **Journal of Informatics**, East Anglia, v. 1, n. 1, p.17-35, 1977. Disponível em: http://www.scils.rutgers.edu/~muresan/551_IR/Resources/Docs/jinfHutchins1977.pdf. Acesso em: 20 jan. 2022.

INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (EDS.). **La interdisciplinariedad y la transdisciplinariedad en la organización del conocimiento científico**: actas del VIII Congreso ISKO-España, León, 18, 19 y 20 de abr. 2007 = Interdisciplinarity and transdisciplinarity in the organization of scientific knowledge. León, España: Universidad de León.Secretariado Publicaciones, 2007. ISBN: 978-84-9773-333-5.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2788**: Documentation - Guidelines for the development of monolingual thesauri, 1986. 32 p.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 5963**: Documentation – methods for examining documents determining their subjects and selecting indexing terms. Genebra, 1985. 5 p.

JORDAN, M. Artificial Intelligence - The Revolution Hasn't Happened Yet. **Harvard Data Science Review**. [S.l.], v(1), n. 1, jul. 2019. DOI: 10.1162/99608f92.f06c6e61. Disponível em: <https://doi.org/10.1162/99608f92.f06c6e61>. Acesso em: 15 maio. 2021.

KENNEDY, S.; WHITE, C. The future of Augmented Intelligence. **Bell Labs Technical Journal**. [S.l.], v. 25, p. 1–18, 2020. ISSN: 1538-7305. DOI: 10.15325/BLTJ.2020.3015275. Disponível em: <https://doi.org/10.15325/BLTJ.2020.3015275>. Acesso em: 15 maio. 2021.

KHURANA, D. et al. Natural language processing: state of the art, current trends, and challenges. **Multimedia Tools and Applications**. [S.l.], 2022. DOI: 10.1007/s11042-022-13428-4. Disponível em: <https://doi.org/10.1007/s11042-022-13428-4>. Acesso em: 30 nov. 2022.

KIYANI, F.; TAS, O. A survey automatic text summarization. **Pressacademia**, Istambul, v. 5, n. 1, p. 205–213, 2017. DOI: 10.17261/Pressacademia.2017.591. Disponível em: <http://dx.doi.org/10.17261/Pressacademia.2017.591>. Acesso em 30 nov. 2022.

KO, C.-L.; YEN, R.-F.; CHEN, C.-M. Augmented intelligence to predict V/Q-matched defects based on dynamic lung perfusion scan. **Journal of Nuclear Medicine**. [S.l.], v. 61, n. supplement 1, p. 1432-1432, 2020. ISSN 0161-5505, 2159-662X. Disponível em: https://jnm.snmjournals.org/content/61/supplement_1/1432. Acesso em: 29 de nov. 2022.

KOBASHI, N. Y. Análise Documentária e Representação da Informação. **INFORMARE**. Rio de Janeiro, v. 2, n. 2, p. 5-27, 1996. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/40976>. Acesso em: 30 nov. 2022.

KORTELING, J. E. (HANS). et al. Human- versus Artificial Intelligence. **Frontiers in Artificial Intelligence**, 2021. v. 4, p. 622-364. DOI: 10.3389/frai.2021.622364. Disponível em: <https://doi.org/10.3389/frai.2021.622364>. Acesso em: 30 nov. 2022.

KOTZ, S. et al. **Continuous multivariate distributions**. 2 ed. New York: Wiley, 2000. 713 p.

KOVARIK, C. et al. Commentary: Position statement on augmented intelligence (Aul). **Journal of the American Academy of Dermatology**, 2019. v. 81, n. 4, p. 998–1000. DOI: 10.1016/j.jaad.2019.06.032 Disponível em: <https://doi.org/10.1016/j.jaad.2019.06.032>. Acesso em: 30 nov. 2022.

KOZEN, D. **A completeness theorem for Kleene algebras and the algebra of regular events**. [1991] Proceedings Sixth Annual IEEE Symposium on Logic in Computer Science. **Anais...** Em: SIXTH ANNUAL IEEE SYMPOSIUM ON LOGIC IN COMPUTER SCIENCE. Amsterdam, Netherlands: IEEE Comput. Sco. Press, 1991. Disponível em: <http://ieeexplore.ieee.org/document/151646/>. Acesso em: 21 set. 2022.

KRUCHTEN, P. **Introdução ao RUP: rational unified process**. 2 ed. Rio de Janeiro: Ciência Moderna, 2004. 256 p.

KUHN, T. S. **A estrutura das revoluções científicas**. 5 ed. Sao Paulo: Perspectiva, 2000. 324 p. ISBN: 978-85-273-0111-4.

LARMAN, C. **Utilizando UML e padrões: uma introdução à análise e ao projeto orientados a objetos e ao desenvolvimento iterativo**. Tradução: Rosana T. Vaccare Braga; Tradução: João Tortello. 3 ed. Porto Alegre: Bookman, 2008. 696 p. ISBN: 978-85-60031-52-8.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Tradução: Antônio Agenor Briquet de Lemos. Brasília, DF: Briquet de Lemos / Livros, 2004. 237 p. ISBN: 978-85-85637-24-8.

LANE, H.; HOWARD, C.; HAPKE, H. M. **Natural language processing in action: understanding, analyzing, and generating text with Python**. Shelter Island, NY: Manning Publications Co, 2019. 544 p. ISBN: 978-1-61729-463-1.

LEMOS, A. J. et al. Consumo, comportamento ingestivo, desempenho, características de carcaça e rendimento de cortes comerciais de cordeiros em terminação alimentados com feno ou silagem de gliricídia. **Archives of Veterinary Science**. [S.l.], v.25, n.2, p.94-110, 2020. DOI: 10.5380/avs.v25i2.71799. Disponível em: <https://doi.org/10.5380/avs.v25i2.71799>. Acesso em: 21 set. 2022.

LIAO, J.; HANSEN, P.; CHAI, C. A framework of artificial intelligence augmented design support. **Human-Computer Interaction**. Philadelphia. v. 35, n. 5-6, p. 511-544, 2020. DOI: <https://doi.org/10.1080/07370024.2020.1733576>. Disponível em: <https://doi.org/10.1080/07370024.2020.1733576>. Acesso em: 21 set. 2022.

LIMA, G. Â. DE. O ensino da análise de assuntos: em busca de uma metodologia. **Informação & Sociedade: Estudos**. João Pessoa, v. 30, n. 4, p. 1–23, 2021a. DOI: 10.22478/ufpb.1809-4783.2020v30n4.57182. Disponível em: <https://doi.org/10.22478/ufpb.1809-4783.2020v30n4.57182>. Acesso em: 21 set. 2022.

LIMA, G. Â. DE. Gênese da classificação: uma análise de conteúdo a partir da definição. **Perspectivas em Ciência da Informação**. Belo Horizonte, v. 26, n. 1, p. 197–237, 2021b. DOI: 10.1590/1981-5344/32686. Disponível em: <https://doi.org/10.1590/1981-5344/32686>. Acesso em: 30 nov. 2022.

LIMA, G. Â. DE; FUJITA, M. S. L.; REDIGOLO, F. M. Importância do Contexto para a indexação. **PontodeAcesso**. Salvador, v. 15, n. 3, p. 283-302, 2021. ISSN: 1981-6766. DOI: 10.9771/rpa.v15i3.47469. Disponível em: <https://doi.org/10.9771/rpa.v15i3.47469>. Acesso em: 21 set. 2022.

LIMA, G. Â. DE; MACULAN, B. C. M. DOS S. Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. **Ciência da Informação**. Brasília, v. 46, n. 1, p.60-72, jan./abr. 2017. DOI: 10.18225/ci.inf.v46i1.4014. Disponível em: <https://doi.org/10.18225/ci.inf.v46i1.4014>. Acesso em: 30 nov. 2022.

LIMA, G. Â. **Mapa Hipertextual (MHTX): um modelo para organização hipertextual de documentos**. Tese (doutorado em Ciência da Informação). Escola de Ciência da Informação da Universidade Federal de Minas Gerais - Belo Horizonte. p. 207. 2004. Disponível em: <http://hdl.handle.net/1843/LHLS-6BUPG9>. Acesso em: 07 out. 2022.

LIMA, G. Â. B. *Softwares* para automação de bibliotecas e centros de documentação na literatura brasileira até 1998. **Ciência da Informação**, Brasília, v. 28, n. 3, p. 310–321, 1999. DOI: <https://doi.org/10.1590/S0100-19651999000300009>.

LIMA, G. Â.; MACULAN, B. C. M. DOS S. Análise de assuntos a partir de uma perspectiva histórica do ARIST. **Ciência da Informação**. Brasília, v. 41, n. 1, p. 22-35, 2014. Disponível em: <https://brapci.inf.br/index.php/res/download/54450>. Acesso em: 30 nov. 2022.

LIU, V. X. The future of AI in critical care is augmented, not artificial, intelligence. **Critical Care**. [S.l.], nº 673, 2020. DOI: 10.1186/s13054-020-03404-5. Disponível em: <https://doi.org/10.1186/s13054-020-03404-5>. Acesso em: 30 nov. 2022.

LUHN, H. P. A Statistical Approach to Mechanized Encoding and Searching of Literary Information. **IBM Journal of Research and Development**. Miami, 1957. v. 1, n. 4, p. 309–317. DOI: 10.1147/rd.14.0309. Disponível em: <https://doi.org/10.1147/rd.14.0309>. Acesso em: 30 nov. 2022.

MACULAN, B. C. M. DOS S. **Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do THESAGRO**. Tese (doutorado em Ciência da Informação). Escola

de Ciência da Informação da Universidade Federal de Minas Gerais - Belo Horizonte. p. 243. 2015. Disponível em: <http://hdl.handle.net/1843/BUBD-9ZKMUV>. Acesso em: 07 out. 2022.

MACULAN, B. C. M. S. Ambiguidade e o contexto na representação de informações em domínios de especialidade. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 25. p. 98-124, 2020. ISSN 1981-5344. DOI: 10.1590/1981-5344/4300. Disponível em: <http://dx.doi.org/10.1590/1981-5344/4300>. Acesso em: 07 out. 2022.

MACULAN, B. C. M. S.; LIMA, G. A. B. O. Buscando uma definição para o conceito de “conceito”. **Perspectivas em Ciência da Informação**. Belo Horizonte, v. 22, n. 2, p. 54-87, jun. 2017. DOI: 10.1590/1981-5344/2963. Disponível em: <https://doi.org/10.1590/1981-5344/2963>. Acesso em: 30 nov. 2022.

MAI, J. Semiotics, and indexing: an analysis of the subject indexing process. **Journal of Documentation**, 2001. v. 57, n. 5, p. 591-622. ISSN: 0022-0418. DOI: 10.1108/EUM0000000007095. Disponível em: <https://doi.org/10.1108/EUM0000000007095>. Acesso em: 30 nov. 2022.

MAIA, L. S. **Extração e explicitação de relações semânticas para a representação do conhecimento de documentos acadêmicos**: Tese (doutorado em Ciência da Informação). Escola de Ciência da Informação da Universidade Federal de Minas Gerais - Belo Horizonte. p. 263. 2018. Disponível em: <http://hdl.handle.net/1843/VAFA-BB5JHW>. Acesso em: 07 out. 2022.

MARTELLI, M. R.; BARROS, T. T.; ASSIS, O. B. G. Filmes de polpa de banana produzidos por batelada: Propriedades mecânicas e coloração. **Polímeros Ciência e Tecnologia**, 2014. v. 24, n. 1, p. 137–142. DOI: 10.4322/polimeros.2014.062. Disponível em: <https://doi.org/10.4322/polimeros.2014.062>. Acesso em 30 nov. 2022.

MITKOV, R.; HOVY, E. (EDS.). Text summarization. Em: **The Oxford handbook of computational linguistics**. 1 ed. Oxford: Oxford Univ. Press, 2009. p. 553–565. ISBN: 0-19-823882-7.

MORIN, E. **Ciência com consciência**. Tradução: Maria D. Alexandre; Tradução: Maria Alice Sampaio Dória. 82. ed. Rio de Janeiro (RJ): Bertrand Brasil, 2005. 350 p. ISBN: 978-85-286-0579-2.

NARUKAWA, C. M.; FUJITA, M. S. L.; LEIVA, I. G. Indexação automatizada de artigos de periódicos científicos: análise da aplicação do *software* SISA com uso da terminologia DeCS na área de Odontologia. **Informação & Sociedade: Estudos**, João Pessoa, 2009. v.19, n.2, p. 99-118. ISSN 1809-4783.

NORMAN, D. A. **The design of everyday things**. 2 ed. New York, New York: Basic Books, 1969. 347 p.

PADRON, M. F.; CRUZ, F. W.; SILVA, J. R. F. Modelos conceituais na ciência da informação: uma revisão de literatura. **Encontro Nacional de Pesquisa em Ciência da Informação. Anais...** Em: XIX ENANCIB. Londrina - PR: 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/102388>. Acesso em: 30 nov. 2022.

PASQUINELLI, M. (ED.). **Alleys of your mind: augmented intelligence and its traumas**. Lüneburg: Meson Press, Hybrid Publishing Lab, Centre for Digital Cultures, Leuphana University of Lüneburg, 2015. 217 p. ISBN: 978-3-95796-065-8 978-3-95796-066-5.

PASQUINELLI, M. (ED.). Keyword: Augmented Intelligence. Em: **Alleys of Your Mind: Augmented Intelligence and Its Traumas**. DE: meson press, 2015b. p. 203–207.

- PAVLOU, P. A. Internet of Things – Will Humans be Replaced or Augmented? **NIM Marketing Intelligence Review**. 2018. v. 10, n. 2, p. 42–47. DOI: 10.2478/gfkmir-2018-0017. Disponível em: <https://doi.org/10.2478/gfkmir-2018-0017>. Acesso em: 30 nov. 2022.
- PIMENTEL, M. Um Pesquisador em Computação em Busca de um Modo de Fazer Pensar Pesquisas em Informática na Educação. **Revista Brasileira de Informática na Educação**, 2018. v. 26, n. 1, p. 51. DOI: 10.5753/rbie.2018.26.01.51. Disponível em: <https://doi.org/10.5753/rbie.2018.26.01.51>. Acesso em: 30 nov. 2022.
- PIRES, L. L.; BRAGANTINI, C.; COSTA, J. L. DA S. Armazenamento de sementes de feijão revestidas com polímeros e tratadas com fungicidas. **Pesquisa Agropecuária Brasileira**. Brasil, v. 39, n. 7, p. 709–715, 2004. DOI: 10.1590/S0100-204X2004000700013. Disponível em <https://doi.org/10.1590/S0100-204X2004000700013>. Acesso em: 30 nov. 2022.
- PONTES, F. V. **Organização do conhecimento em bibliotecas digitais de teses e dissertações: uma abordagem baseada na classificação facetada e taxonomias dinâmicas**. Tese (doutorado em Ciência da Informação). Escola de Ciência da Informação da Universidade Federal de Minas Gerais - Belo Horizonte. p. 269. 2013. Disponível em: <http://hdl.handle.net/1843/BUOS-97FGQ6>. Acesso em: 07 out. 2022.
- RIVA, P. Il nuovo modello concettuale dell’universo bibliografico: FRBR Library Reference Model. **AIB studi**, 2016. v. 56, n. 2, p. 265-275. DOI: 10.2426/aibstudi-11480. Disponível em: <https://doi.org/10.2426/aibstudi-11480>. Acesso em: 30 nov. 2022.
- RIVERA, M.; GLEESON, P. **Augmented Intelligence Amplifies Performance in the Workplace**. 1. ed. Thinaer, 2020. 14 p. Disponível em: <https://internetofbusiness.com/wp-content/uploads/2020/08/Thinaer-Human-and-Machine-Augmented-Intelligence-Amplifies-Performance.pdf>. Acesso em: 30 nov. 2022.
- ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. **Estudos Avançados em Biblioteconomia e Ciência da Informação**. 1982. v. 1, n. 1, p. 236–274. Disponível em: <https://brapci.inf.br/index.php/res/download/161518>. Acesso: 30 nov. 2022.
- RODRÍGUEZ BRAVO, B.; INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (EDS.). **La interdisciplinarietà y la transdisciplinarietà en la organización del conocimiento científico: actas del VIII Congreso ISKO-España, León, 18, 19 y 20 de abril de 2007 = Interdisciplinarity and transdisciplinarity in the organization of scientific knowledge**. León, España: Universidad de León. Secretariado Publicaciones, 2007.
- ROMERO-BRUFU, S. et al. What ’s in a name? A comparison of attitudes towards artificial intelligence (AI) versus augmented human intelligence (AHI). **BMC Medical Informatics and Decision Making**. [S.l.], v. 20, n. 1, p. 167, 2020. DOI: 10.1186/s12911-020-01158-2. Disponível em: <https://doi.org/10.1186/s12911-020-01158-2>. Acesso: 30 nov. 2022.
- RUI, Y. From Artificial Intelligence to Augmented Intelligence. **IEEE MultiMedia**. [S.l.], v. 24, n. 1, p. 4–5, 2017. DOI: 10.1109/MMUL.2017.8. Disponível em: <https://doi.org/10.1109/MMUL.2017.8>. Acesso em: 30 nov. 2022.
- RUSSELL, S. J.; NORVIG, P.; DAVIS, E. **Inteligência Artificial**. Tradução: Regina Célia Simille. 3 ed. Rio de Janeiro: GEN LTC, 2013. 1115 p. ISBN: 978-0-13-604259-4.
- SADIKU, M. N. O. et al. Augmented Intelligence. **International Journal of Scientific Advances**. [S.l.], v. 2, n. 5, p. 772–776, 2021. ISSN: 2708-7972. DOI: 10.51542/ijscia.v2i5.17. Disponível em: <https://doi.org/10.51542/ijscia.v2i5.17>. Acesso em: 30 nov. 2022.

SAGGION, H. Automatic Summarization: An Overview. **Revue française de linguistique appliquée**, 2008. v. 13, n. 1, p. 63–81. DOI: 10.3917/rfla.131.0063. Disponível em: <https://doi.org/10.3917/rfla.131.0063>. Acesso em: 30 nov. 2022.

SANDUSKY, R.; PETTY, M. D. **What are Conceptual Models?** Disponível em: <https://www.interaction-design.org/literature/topics/conceptual-models>. Acesso em: 16 maio. 2021.

SARACEVIC, T. RELEVANCE: A review of and a framework for the thinking on the notion in information science. **Journal of the American Society for Information Science**, 1975. v. 26, n. 6, p. 321–343. DOI: 10.1002/asi.4630260604. Disponível em: <https://doi.org/10.1002/asi.4630260604>. Acesso em: 30 nov. 2022

SARACEVIC, T.; KANTOR, P. A study of information seeking and retrieving. II. Users, questions, and effectiveness. **Journal of the American Society for Information Science**. Estados Unidos, v. 39, n. 3, p. 177–196, 1988. Disponível em: <https://tefkos.comminfo.rutgers.edu/JASIS1988part2.pdf>. Acesso em: 30 nov. 2022.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. Tradução: Ana Maria P. Cardoso. **Perspectivas em Ciência da Informação**. Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em: https://www.brapci.inf.br/_repositorio/2017/07/pdf_7810a51cca_0000015436.pdf. Acesso em: 31 mar. 2022.

SCHAPARINI, L. P. et al. Análise comparativa entre índices de vegetação e sua relação com o balanço hídrico em soja. **Agrometeoros**. Passo Fundo, v.27, n.1, p.183-190, set 2019. ISSN: 2526-7043. DOI: 10.31062/agrom.v27i1.26574. Disponível em: <https://doi.org/10.31062/agrom.v27i1.26574>. Acesso em: 30 nov. 2022

SEGARAN, T. **Programming collective intelligence: building smart web 2.0 applications**. 1st ed ed. Beijing; Sebastapol [CA]: O'Reilly, 2007. ISBN: 978-0-596-52932-1.

SILVA, A. R. DA. BDPA - **Bases de Dados da Pesquisa Agropecuária**. sistema Ainfo 6. Disponível em: <https://www.bdpa.cnptia.embrapa.br/consulta/creditos>. Acesso em: 31 mar. 2022.

SILVA, M. DOS R. DA; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**. Campinas, v. 16, n. 2, p. 133–161, 2004. Disponível em: <https://www.scielo.br/j/tinf/a/cNngvqQdWfBGrJtLSdLRKnP/?format=pdf&lang=pt>. Acesso em: 30 nov. 2022.

SILVA, M. F. **Proposta de modelo de colaboração para catálogo web facetado**. Tese (doutorado em Ciência da Informação). Escola de Ciência da Informação da Universidade Federal de Minas Gerais - Belo Horizonte. p. 269. 2013.

SILVA, S. R. DE B.; CORREA, R. F. Sistemas de indexação automática por atribuição: uma análise comparativa. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**. Florianópolis, v. 25, p. 01–25, 2020. DOI: 10.5007/1518-2924.2020.e70740. Disponível em: <https://doi.org/10.5007/1518-2924.2020.e70740>. Acesso em: 30 nov. 2022.

SIMON, H. A. **Sciences of the artificial**. 1. ed. Cambridge, Mass.: MIT Press, 1970. 248 p. ISBN: 978-0-262-19051-0.

- SMIT, J. W. Linguagens de Indexação. In: SMIT, J. W. **Análise documentária: a análise da síntese**. Brasília, DF: Johanna W. Smit, 1987. p. 11–26. Disponível em: <https://livroaberto.ibict.br/bitstream/1/1011/1/An%c3%a1lise%20document%c3%a1ria.pdf>. Acesso em: 30 nov. 2022.
- SOUSA, B. P. DE; FUJITA, M. S. L. A análise de assuntos no processo de indexação: um percurso entre teoria e norma. **Informação & Sociedade: Estudos**. João Pessoa, v. 24, n. 1, p. 19–34, 2014. ISSN: 1809-4783. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/92987>. Acesso em: 30 nov. 2022.
- SOUSA, H. T. DE; LIMA, G. Â. DE; ARAÚJO, W. J. Reflexões multidisciplinares sobre o espectro das ontologias. **Investigación Bibliotecológica: archivonomía, bibliotecología e información**. México, v. 36, n. 92, p. 99, 2022. ISSN: 2448-8321, 0187-358X. DOI: Disponível em: <https://doi.org/10.22201/iibi.24488321xe.2022.92.58605>. Acesso em: 30 nov. 2022.
- STUDER, S. et al. Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. **Machine Learning and Knowledge Extraction**. [S.l.], v. 3, n. 2, p. 392–413. DOI: 10.3390/make3020020. Disponível em: <https://doi.org/10.3390/make3020020>. Acesso em: 30 nov. 2022.
- TAKEUCHI, H.; NONAKA, I. **Gestão do conhecimento**. Tradução: Cláudio Reis Gonçalves; Tradução: Ana Thorell. 2 ed. Porto Alegre: Bookman, 2009. ISBN: 978-85-7780-191-6.
- TAN, S. C.; TEO, C. L. Supporting Knowledge Building with Analytics and Augmented Intelligence. **Building on the Emerging Works**. 9 p. 2020. Disponível em: <https://ikit.org/summerinstitute2020/wp-content/uploads/2021/03/137-Tan-Teo-Supporting.pdf>. Acesso em: 30 nov. 2022.
- THAMM, S. et al. Concept for an augmented intelligence-based quality assurance of assembly tasks in global value networks. **Procedia CIRP**, 2021. v. 97, p. 423-428. DOI: 10.1016/j.procir.2020.05.262. Disponível em: <https://doi.org/10.1016/j.procir.2020.05.262>. Acesso em: 30 nov. 2022.
- TORRES-MORENO, J.-M. Why Summarize Texts? Em: TORRES-MORENO, J.-M. (Ed.). **Automatic Text Summarization**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2014. p. 1–21. ISBN: 978-1-119-00475-2 978-1-84821-668-6.
- TUKEY, J. W. The Future of Data Analysis. **The Annals of Mathematical Statistics**. Estados Unidos, v. 33, n. 1, p. 1–67, 1962. ISSN 0003-4851. Disponível em: <http://www.mat.ufrgs.br/~viali/estatistica/mat2274/material/textos/2237638.pdf>. Acesso em: 30 nov. 2022.
- VAISHNAVI, V.; KUECHLER, W. **Design science research methods and patterns: innovating information and communication technology**. Boca Raton: Auerbach Publications, 2008. ISBN: 978-1-4200-5932-8.
- VAKKARI, P.; CHANG, Y.; JÄRVELIN, K. Disciplinary contributions to research topics and methodology in Library and Information Science—Leading to fragmentation? **Journal of the Association for Information Science and Technology**. [S.l.], 2022. v. 73, n. 12, p. 1706–1722. DOI: 10.1002/asi.24690. Disponível em: <https://doi.org/10.1002/asi.24690>. Acesso em: 30 nov. 2022.
- VALLADARES, G. S.; PEREIRA, M. G.; ANJOS, L. H. Adsorção de fósforo em solos de argila de atividade baixa. **Bragantia**. Campinas, v. 62, n. 1, p. 111–118. ISSN:1678-4499, 0006-8705. Disponível em:

<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/105801/1/1223.pdf>. Visto em: 07 out. 2022.

VIEIRA, S. B. Indexação automática e manual: revisão de literatura. **Ciência da Informação**. Brasília, v. 17, n. 1, p. 43–57, 1988. Disponível em: <https://revista.ibict.br/ciinf/article/view/298/298>. Acesso em: 30 nov. 2022.

WIERINGA, R. J. **Design Science Methodology for Information Systems and Software Engineering**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. 332 p. ISBN: 978-3-662-43838-1.

WÓJCIK, M. Potential use of Augmented Reality in LIS education. **Education and Information Technologies**. [S.l.], v. 21, n. 6, p. 1555–1569, 2016. ISSN: 1360-2357, 1573-7608. DOI: 10.1007/s10639-015-9399-z. Disponível em: <https://doi.org/10.1007/s10639-015-9399-z>. Acesso em: 30 nov. 2022.

WÓJCIK, M. Augmented intelligence technology. The ethical and practical problems of its implementation in libraries. **Library Hi Tech**. [S.l.], v. 39, n. 2, p. 435–447, 2021. DOI: 10.1108/LHT-02-2020-0043. Disponível em: <https://doi.org/10.1108/LHT-02-2020-0043>. Acesso em: 30 nov. 2022.

XU, B. *et al.* **Artificial Intelligence or Augmented Intelligence: A Case Study of Human-AI Collaboration in Operational Decision Making**. PACIS 2020 Proceedings. **Anais... In: TWENTY-THIRD PACIFIC ASIA CONFERENCE ON INFORMATION SYSTEMS. DUBAI: 2020**. ISBN: 978-1-73363-253-9. DOI: [et.org/pacis2020/147](https://doi.org/et.org/pacis2020/147). Disponível em: <https://doi.org/et.org/pacis2020/147>. Acesso em: 30 nov. 2022.

ZAKI, M.; MCCOLL-KENNEDY, J. R. Text mining analysis roadmap (TMAR) for service research. **Journal of Services Marketing**. [S.l.], v. 34, n. 1, p. 30–47, 2020. DOI: 10.1108/JSM-02-2019-0074. Disponível em: <https://doi.org/10.1108/JSM-02-2019-0074>. Acesso em: 30 nov. 2022.

ZHANG, A. *et al.* **Dive into Deep Learning**. 1. ed. <https://d2l.ai>: open source, 2021. v. 1. DOI: 10.48550/arXiv.2106.11342. Disponível em: <https://doi.org/10.48550/arXiv.2106.11342>. Acesso em: 30 nov. 2022.

ZHENG, N. *et al.* Hybrid-augmented intelligence: collaboration and cognition. **Frontiers of Information Technology & Electronic Engineering**. Zhejiang, v. 18, n. 2, p. 153–179, 2017. DOI: 10.1631/FITEE.1700053. Disponível em: <https://doi.org/10.1631/FITEE.1700053>. Acesso em: 30 nov. 2022.

APÊNDICE A – MODELO CONCEITUAL

Este modelo conceitual foi elaborado como um guia para o desenvolvimento dos procedimentos metodológicos.

As publicações científicas de Chapman *et al.* (2000), Goodfellow; Bengio; e Courville (2016), Zaki e Mccoll-Kennedy (2020) e Studer *et al.* (2021) apresentam Modelos Conceituais consolidados para a mineração de dados que são adaptáveis a esta pesquisa, pelo que foram utilizadas como as referências principais.

De modo geral, os modelos conceituais de análise de textos (mineração de dados) por métodos computacionais são divididos em: Compreensão do problema, tratamento dos dados, modelagem e avaliação dos resultados (Quadro 5).

Quadro 31 – Comparação entre os modelos utilizados como fonte.

Modelo Proposto	CHAPMAN (2020)	ZAKI (2020)	STUDER (2021)
Compreensão do problema	Compreensão do negócio	Estudo do conhecimento prévio e compreensão do negócio	Compreensão de negócios e dados
Tratamento dos dados	Compreensão e Preparação dos dados	Compreensão dos dados	Preparação dos dados
Modelagem	Modelagem	Modelagem e Validação dos dados	Modelagem e Avaliação
Avaliação dos resultados	Implantação	<i>Insights</i> obtidos	Implantação, Monitoramento e Manutenção

Fonte: Elaborado pelo autor.

Os procedimentos metodológicos devem contemplar quatro dimensões, a saber, Problema, Dados, Modelo e Uso, pois são elas que delineiam as quatro etapas que correspondem ao (1) planejamento do projeto, (2) ao tratamento dos dados, (3) à modelagem da inteligência artificial e (4) à inteligência aumentada.

1 Visão geral

No Quadro 2, está uma visão geral do ciclo de vida de um projeto de inteligência aumentada. Cada projeto deve contemplar as quatro dimensões, a saber, Problema, Dados, Modelo e Uso, pois são elas que delineiam as quatro etapas que correspondem ao planejamento do projeto, ao tratamento dos dados, à modelagem da inteligência artificial e à inteligência aumentada como padrão de projeto, processo de uso e resultado.

Cada etapa tem as atividades gerais que as caracterizam, e estas agregam um conjunto de atividades específicas que produzem um artefato. Assim, descrever o

planejamento produz o projeto escrito; tratar os dados cria a base de conhecimento, treinar algoritmo configura o Modelo de IA.

Os procedimentos metodológicos para elaborar o Modelo Conceitual incluíram as seguintes instruções sugeridas por Zaki e Mccoll-Kennedy (2020):

1. definir o objetivo da etapa;
2. estabelecer a questão norteadora da etapa;
3. arrolar o conjunto de técnicas disponíveis e elegíveis para a etapa;
4. selecionar as técnicas adequadas para a etapa.

Quadro 2 – Visão geral do Modelo Conceitual.

Dimensão	Etapa	Atividade Geral e Tarefas específicas	Artefato
Problema	Planejamento	Descrição: Descrever o problema, estabelecer objetivos, estudar as questões éticas e definir métricas	Projeto escrito
Dados	Tratamento	Estruturação: Definir tipo, estrutura, quantidade e qualidade dos dados; e recursos. Coleta: Encontrar e coletar os dados Criação de recursos: Limpar, organizar e estruturar os dados, incluindo regularizá-los e normalizá-los.	Base de conhecimento
Modelos	Inteligência Artificial	Seleção: Selecionar o modelo de inteligência artificial adequado aos objetivos e aos dados Configuração: Estabelecer as métricas de desempenho e selecionar hiperparâmetros Treinamento: Treinar o modelo	Modelo de inteligência artificial
Usos	Inteligência Aumentada	Avaliação: Avaliar o modelo Implantação: Implantar o modelo e usar o modelo	Produto (índices, resumos, catálogos)

Fonte: Elaborado pelo autor.

O Modelo Conceitual proposto recomenda nove atividades gerais, subdivididas em pelo menos 22 tarefas específicas. Os relacionamentos entre as atividades e as tarefas dependem do problema, dos objetivos e dos dados disponíveis. A sequência sugerida não é rígida, mas é recomendada.

As seções seguintes abordam o Modelo Conceitual quanto a a compreensão do problema, a compreensão dos dados, a compreensão da inteligência artificial, e a compreensão da inteligência aumentada.

Etapa 1: Compreensão do Problema

Nesta seção aborda-se a Compreensão do Problema que envolve compreender toda a conjectura envolvida para a solução, inclusive como serão usados os artefatos prontos, uma vez que a inteligência aumentada, como método de planejamento, enseja ganhos cognitivos na produção. Nesta etapa, procura-se responder à seguinte questão norteadora: como obter o produto informacional (índice, resumo, catálogo) a partir dos dados disponíveis?

O termo “Projeto” (com p maiúsculo) se refere ao planejamento, à execução, aos produtos e ao tempo de vida do projeto, ou seja, refere-se ao ato de projetar, fazer e usar insumos, ferramentas e produtos. Enquanto o termo “projeto” (com P minúsculo) foi usado, nesta seção, para se referir ao documento escrito.

Um projeto deve conter a descrição detalhada de todas as etapas, atividades e artefatos. Ele é a documentação estruturada contendo os objetivos, a descrição do problema, o produto a ser feito, o inventário dos dados, os recursos humanos e financeiros necessários, bem como as datas de início e fim do projeto. Mas escrever um projeto está fora do escopo deste documento, pelo que se recomenda consultar o livro *Project Management Body of Knowledge* (ROSE, 2013) para maiores esclarecimentos.

A contribuição desta etapa para o Projeto é tripla, pois prescreve o roteiro para a compreensão do problema; propõe a estrutura padrão do projeto; e desenvolve a argumentação necessária do planejamento à linha de produção.

A importância de um projeto é maior para Projetos com equipes multidisciplinares, pois ele é o ponto de convergência dos múltiplos saberes envolvidos.

O objetivo da etapa de “Compreensão do Problema” é descrever o problema e planejar a solução. Ambos, problema e solução, dependem do conhecimento prévio da equipe humana envolvida e as descrições são necessárias para transformar esse conhecimento, que está na mente das pessoas, em informação. Como enfatiza a BCI, somente informação pode ser compartilhada e só dessa forma as diferentes visões dos envolvidos podem ser integradas; as dúvidas, dirimidas; e os mal-entendidos, desfeitos.

Como diria Grogan (2001, p. 51): “ninguém é imune a problemas, e, assim, teoricamente, todo ser humano é um iniciador potencial do Modelo Conceitual”. Vale dizer ainda que as equipes multidisciplinares de projetos que envolvam a análise de assuntos assistida por computador devem incluir profissionais da informação com domínio técnico na produção de artefatos informacionais (conteúdo e estrutura) e de profissionais da informática para viabilizar o uso deles. Isso ainda corrobora com a ideia do projeto como o ponto de convergência comunicacional entre as pessoas da equipe e entre elas e as outras pessoas interessadas (*stakeholders*).

Obter o financiamento de um projeto precisa ser justificado e talvez seja esta a motivação principal de escrevê-lo. O projeto escrito, portanto, é o artefato que viabiliza o próprio Projeto, pois facilita a comunicação com os potenciais patrocinadores.

A “Compreensão do Problema” não está restrita ao idealizador e aos planejadores; envolve também a diversidade intelectual e cognitiva da equipe, de modo que tudo deve ser descrito e especificado, incluindo a motivação e o porquê de cada seleção de ferramentas ou técnica. Esse detalhamento depende de muitos fatores, que vão desde a finalidade do produto, passando pela experiência da equipe, a quantidade e a qualidade dos dados e o tempo disponível. Assim, a descrição deve incluir, pelo menos, os objetivos, as restrições, as datas de início e de fim das atividades, a equipe, os recursos financeiros necessários e, se possível, a descrição dos benefícios e dos impactos sociais esperados com o sucesso do projeto.

Escrever um projeto não é uma atividade popular, de modo que esta etapa tende a ser bem negligenciada, mas não se deve perder de vista que o sucesso vai depender em sua maior parte desse planejamento. O gerente de projeto deve fazer um esforço adicional para implementar corretamente esta fase, pois o planejamento bem-feito evita esforços inúteis, otimiza custo, tempo e esforço humano, e produz soluções melhores.

Para esta etapa, Zaki *et al.* (2020) recomenda começar com uma revisão de literatura ou estudo do conhecimento prévio (*background*) disponível. Ng (2015) e Goodfellow; Bengio; Courville (2016) consideram que a descrição do problema orienta a elaboração dos objetivos. Para Chapman *et al.* (2002, p. 13), a “fase inicial concentra-se na compreensão dos objetivos e requisitos do projeto convertendo esse conhecimento em uma definição do problema” que aponta o caminho para atingir os objetivos.

Também devem ser levantadas, nesta etapa, as questões éticas e as possíveis consequências sociais perguntando-se:

1. Em que medida o projeto afeta (piora ou melhora) a qualidade do produto para o usuário final do conhecimento em relação aos produtos existentes?
 - Em que medida os empregos das pessoas serão ameaçados?
 - Quais consequências a curto, médio e longo prazo o emprego da inteligência artificial pode produzir na comunidade?

Estas e outras questões éticas também devem ser bem resolvidas e o documento final deve conter (se houver) as questões éticas envolvidas e as soluções implementadas, uma vez que a inteligência aumentada tem como diretriz principal pôr o ser humano no centro das atividades e dos benefícios.

Identificar e definir os termos por meio da elaboração de um Glossário é fundamental em projetos de todas as áreas do conhecimento, e, na BCI, assume um caráter

impositivo. Assim, recomenda-se organizar e “disponibilizar um glossário relevante dos termos técnicos” (Chapman *et al.* 2020, p. 18) tanto para facilitar a prática discursiva quanto para evitar ambiguidades. Por exemplo, em projetos de inteligência artificial, os termos “treinamento” e “aprendizagem” de algoritmo podem ser confundidos com “treinamento” e “aprendizagem” de pessoas.

Não podem faltar em um projeto nem o cronograma nem a análise de custo-benefício, segundo Chapman *et al.* (2002), pois a explicitação dos benefícios em relação aos custos costuma ser o argumento decisivo na angariação de fundos.

Enfim, a Etapa de Compreensão do Problema implica definir objetivos e metas, impor ou negociar um cronograma, analisar a viabilidade econômica, buscar justiça social e respeitar o meio ambiente, sem negligenciar o como e o quando as etapas devem ser executadas, a duração delas, os recursos necessários, a qualidade dos insumos e do produto, e, sempre que possível, explicitar as iterações em grande escala no projeto. É importante também analisar as dependências entre cronograma e riscos, bem como marcar os resultados dessas análises explicitamente no plano do projeto, com ações e recomendações, caso apareçam os riscos.

A etapa seguinte do projeto é compreender os dados, que é o assunto da próxima seção, em que se discorreu sobre a coleta e o pré-processamento dos dados, considerada a parte mais demorada e difícil nos Projetos que envolvam treinamento de modelos de inteligência artificial.

Etapa 2: Compreensão dos dados

A maior parte do tempo investido em um projeto de inteligência artificial se passa com a coleta, a análise e a preparação dos dados, por isso esta seção está dedicada às inúmeras tarefas de tratamento de dados. Ela começa pelas técnicas de coleta dos arquivos, seguida pela extração do fluxo de caracteres (o texto, propriamente dito) e analisa os erros que ocorrem com frequência nesta fase relacionados à codificação do texto do arquivo e termina na normalização e regularização dos dados. Essas técnicas computacionais transformam os dados brutos multiestruturados em recursos computacionais estruturados especificamente para a entrada de um algoritmo.

Nesta seção, o termo “dados” se refere aos insumos (entrada) necessários aos algoritmos, notadamente os textos, os índices e os metadados; o termo “dado bruto” se refere ao texto do arquivo digital coletado da fonte. O termo “Recurso” se refere aos dados depois de reestruturados para se adequar às estruturas exigidas pelos algoritmos de inteligência artificial. A expressão “criar recurso” se refere ao processo de transformar “dado bruto” em “recurso”. O termo “base de conhecimento” será utilizado no lugar de “banco de dados” ou

“base de dados” por recomendação de Kobashi (1991, p. 40), porque, segundo a autora, os algoritmos de inteligência artificial não lidam apenas com dados, às vezes, eles processam os conhecimentos diretamente.

As questões norteadoras desta etapa incluem:

3. Qual é o volume de dados necessário (Completeness dos dados)?
 - Os dados estão corretos (Correção)?
 - Os erros são aleatórios ou sistemáticos?
 - Existem valores ausentes nos dados?
 - Faltam dados (Exaustividade)?

A coleta de dados produz o *corpus* da pesquisa e todo o esforço posterior será em vão se essa etapa não for executada adequadamente.

Os dados brutos apresentam muitos desafios relativos a: quantidade, qualidade, estrutura, distribuição (várias fontes) e localização (onde estão). Os dados precisam ser armazenados de forma a atender requisitos como aumento de produtividade, facilidade de acesso, segurança, estratégias de busca, consultas, visualizações e análises estatísticas. Nesta atividade, lida-se com as propriedades coletivas dos dados (metadados), incluindo o formato, a quantidade, a descrição dos campos das tabelas (se for o caso) e a relevância.

O primeiro desafio da coleta de dados consiste em identificar as fontes externas e, o segundo, em definir apropriadamente o local de destino dos dados (máquina local ou na nuvem; dispositivo externo ou no disco local), pois essas escolhas definem onde os dados serão armazenados, consultados, processados e utilizados. Essas atividades afetam consideravelmente, a fase de produção, o desempenho dos sistemas e o tempo despendido na execução de tarefas.

Os dados brutos podem estar distribuídos por vários sítios na rede mundial de computadores, desde bibliotecas virtuais e repositórios *on-lines* até acervos *of-flines*, implicando em diferentes estratégias para encontrar, acessar e coletar os dados elegíveis e potencialmente úteis. Por isso, recomenda-se coletar uma amostra dos dados brutos para analisar os principais aspectos do conjunto de dados de uma fonte antes de iniciar a coleta propriamente dita. Trata-se de uma pesquisa exploratória inicial, porque dados brutos não são recursos prontos para o uso, necessitando de pré-processamento antes de se tornarem recursos. Por exemplo, arquivos compactados (ZIP, RAR, GZIP) precisam ser descompactados, dados distribuídos precisam ser integrados e dados incompletos precisam ser mesclados.

O volume necessário de dados costuma ser o maior obstáculo para a elaboração de modelos de inteligência artificial. Quanto maior o volume coletado, melhor; de preferência

na casa dos milhões de arquivos. Segundo Golub (2021), recomendam-se mil artigos para cada padrão de dados, o que implica a necessidade do desenvolvimento de métodos de coleta automática (*scraping* ou raspagem). Além disso, não é possível uma solução técnica definitiva para a coleta de dados, pois cada fonte de dados demanda um *script* específico, sem contar que, ao longo do tempo, as estruturas, a localização e as formas de acesso modificam-se, exigindo sempre novos desenvolvimentos.

Com a coleta dos dados efetuada, o próximo passo é verificar a qualidade dos dados, que é o assunto da próxima seção.

Preparação dos Dados

Um *corpus* precisa ter um volume suficiente para a tarefa a ser executada e ter seus arquivos em uma estrutura conversível para o formato exigido pelo algoritmo. Satisfeitas essas características, procede-se à verificação da confiabilidade da fonte produtora e à pertinência e relevância dos dados para o Projeto.

A qualidade dos dados determina o sucesso ou o fracasso do Projeto. Assim, verificá-la, em relação ao objetivo e ao produto a ser produzido, acaba sendo mais uma necessidade e uma medida de bom senso.

A qualidade dos dados pode ser aferida pelo volume obtido em relação ao necessário, pela completude, pela confiabilidade da fonte e pela integridade. Mas não é o suficiente. Vale lembrar que os dados brutos são oriundos de múltiplas fontes; produzidos em longos períodos; armazenados em tipos e estruturas próprias; e compilados para as mais variadas finalidades. Juntos formam um *corpus* não uniforme, não homogêneo e multiestruturado, de sorte que é bem difícil realizar operações úteis sem um devido pré-processamento. Sempre haverá dados imprestáveis, por essas e outras razões. Por exemplo, modelos de processamento de linguagem de natural não reconhecem textos em formatos de imagens, ou de vídeo ou de áudio, por isso, nessas situações, os arquivos precisam ser descartados ou convertidos em texto escrito. Alguns dados brutos, ao serem convertidos, resultam em textos sem qualquer sentido lógico e, mesmo em textos extraídos corretamente, o sentido lógico pode ser comprometido por marcações de estilos, conteúdos de tabelas, notas de rodapé, cabeçalhos, rodapés, anexos, apêndices e referências. Por isso, os dados coletados devem ser submetidos a processamentos de uniformização da estrutura, conformação do conteúdo e homogeneização.

O próximo passo é preparar os dados, o que exige tempo, esforços e persistência e deve estar direcionado ao propósito do Projeto. Num cenário otimista e desejável, os dados são abundantes, mas, na realidade, isso não é muito factível, principalmente quando se trata

de textos científicos, pois são produzidos no contexto da ciência, que se subdivide em centenas de áreas, campos e disciplinas, com culturas, terminologias e paradigmas restritos a comunidades pequenas de pesquisadores. Além disso, dados textuais podem se tornar obsoletos em pouco tempo, pois a linguagem muda constantemente tanto na grafia, quanto na semântica. Assim, a preparação dos dados envolve selecionar, limpar, integrar e mesclar múltiplos dados e bases de dados.

A natureza da análise, o produto e os objetivos do Projeto juntamente com a metodologia de inteligência artificial delineiam o tipo, a qualidade, a quantidade e a estrutura dos dados; indiretamente determinam as técnicas de preparação. Por exemplo, para a indexação, os dados de entrada, os textos, são formados por um grande conjunto de termos; e a saída, o índice, é formado por um pequeno conjunto de termos. Nessa conjectura, para elaborar um modelo de inteligência artificial com aprendizagem supervisionada, o conjunto de textos e de índices deve ser grande o suficiente para conter todos os padrões relevantes.

Até esse ponto, apresentaram-se os problemas que afetam a qualidade dos dados, que podem ser resumidos em inadequação, obsolescência e fragmentação. Adicionado a eles, está o problema de transformar a estrutura original para a estrutura exigida na entrada do algoritmo. Uma variedade de procedimentos é empregada para tornar os dados brutos recursos computacionais, que é o processo conhecido como “criar recursos” ou “criação de recursos”.

Assim, ao longo desta seção foram discutidas as causas do problema como um todo e do porquê essas questões precisam ser analisadas. Tendo realizado a inspeção qualitativa dos dados, Chapman *et al.* (2002) recomendam elaborar um relatório em que constem os problemas encontrados, as soluções adotadas e o resultado. Logo, esse relatório conterá as preciosas lições aprendidas que podem ser empregadas novamente no futuro, no caso de ser necessário ampliar a coleta dos dados, refazer a pesquisa ou utilizar o *corpus* em outro contexto. Além disso, vai permitir manter um histórico recuperável das decisões tanto para conferir qualidade e segurança quanto para garantir a lisura no processo e a reprodutibilidade dele.

Todas as técnicas e atividades descritas nesta seção implicam no processo de Criação de Recursos, que é a atividade a ser descrita na próxima seção.

A criação de recursos

Os dados coletados, no estado bruto, não se adéquam às estruturas exigidas para as entradas dos modelos inteligentes, até porque existem muitas técnicas disponíveis e os dados podem estar estruturados para qualquer uma delas. Mas transformar o dado bruto

multiestruturado em recurso computacional demanda obviamente reestruturação, mesclagem e integração dos dados de maneira a “produzir novos atributos e novos registros inteiros e a transformar atributos existentes” (CHAPMAN *et al.*, 2002, p. 24). Por exemplo: um dado pode ser um texto em um arquivo e outro pode ser um texto armazenado em um atributo de um registro de um banco de dados; assim, têm-se dois dados em duas estruturas completamente diferentes que precisam de técnicas distintas para serem processados e transformados para uma estrutura única.

A Criação de Recursos é a atividade de transformar dados brutos em registros bem estruturados na base de conhecimento. Ela junta e combina múltiplos fragmentos, formatos, fontes e estruturas de dados por meio de dois processos conhecidos por Integração, que faz a unificação estrutural, e Mesclagem, que dá a coerência estrutural. Tanto a integração quanto a mesclagem aumentam os recursos na base de conhecimento, mas são técnicas genéricas para conformar os dados e deixá-los coerentes, o que depende de cada contexto de pesquisa.

No momento do uso, cada algoritmo exige um tipo de dado em um formato específico, e dar aos recursos a forma estrutural exigida pelo algoritmo é chamado de Formatação, que é o assunto da próxima seção.

Formatação dos dados

Os modelos de inteligência artificial exigem dados sob um determinado formato, por exemplo, um modelo de sequência-a-sequência exige uma sequência na entrada (sequência de caracteres – texto, por exemplo) e uma sequência na saída (sequência de termos de indexação, por exemplo). A formatação, como sugere o nome, transforma a estrutura dos recursos na estrutura específica exigida pela entrada do modelo sem alterar o significado original dos dados.

Normalmente, dados textuais estão na forma de arquivos PDF, embora não precisem estar nesse formato; podem estar em outros formatos, como o HTML e o DOC. O PDF é um formato para impressão, enquanto os algoritmos processam o fluxo de caracteres, isto é, o texto puro ou texto plano (do inglês, *plain text*). Então, transformar o documento em um fluxo de caracteres é apenas o primeiro dos desafios da formatação, porque o texto extraído apresenta maiúsculas, minúsculas, diacrítica, termos técnicos, sinais de pontuação visíveis, sinais de pontuação invisíveis (quebras de linhas), caracteres estranhos e caracteres de controle. Assim, o segundo desafio da formatação é tratar cada um desses problemas individualmente.

Após a extração do texto do documento, a redução dos caracteres a um padrão de capitalização, a remoção dos sinais diacríticos, o acréscimo dos espaços em branco entre os sinais de pontuação e as palavras próximas, a remoção das quebras de linhas e da hifenização, aplica-se o método de formatação apropriado. Com isso, os recursos da base de conhecimento se tornam processáveis pelo algoritmo e o próximo passo é a seleção do modelo de inteligência artificial que é assunto da próxima seção que descreve as atividades específicas para a Etapa Compreensão da inteligência artificial.

Etapa 3: Compreensão da inteligência artificial

Em teoria, as técnicas da inteligência artificial foram escolhidas na Etapa 1. Mas, na prática, não é bem assim, porque as metodologias dependem dos dados obtidos, dos recursos disponíveis, do volume e da formatação, que só serão conhecidos ao final da Etapa 2. Por isso, o Projeto como um todo é um grande processo iterativo em que ajustes precisam ser refinados até se alcançar a melhor solução, considerando todos os fatores. Não é raro uma *pipeline* se mostrar inviável, e, por isso, todo o processo precisa ser reiniciado.

Aprendizagem artificial

Existem apenas dois tipos de técnicas computacionais denominadas de “aprendizagem artificial” para configurar modelos da Inteligência artificial (Quadro 3) e a escolha de uma delas depende do poder computacional disponível, das suposições, restrições e objetivos definidos na Etapa 1 e dos dados disponíveis, coletados, analisados e preparados na Etapa 2.

Quadro 3 – A adequação de Modelos de inteligência artificial.

Aprendizagem	Exigência	Adequação	Exemplo
Supervisionada	Textos rotulados	Difícil de adotar, porque é difícil conseguir textos rotulados em quantidade suficiente.	Modelos sequenciais com mecanismos de Atenção
Não supervisionada	Textos	Fácil de adotar, porque há textos suficientes.	Modelagem de Tópicos: Latent Dirichlet Allocation (LDA)

Fonte: Elaborado pelo autor.

Nesta pesquisa, portanto, neste Modelo Conceitual leva-se em conta a recuperação da informação, em que se fará uso simultâneo de:

1. informação científica;
2. linguagem natural;
3. idioma português;

4. dado textual;
5. vocabulário controlado.

O uso desses elementos afunila as opções disponíveis. As informações produzidas sob o método científico, em linguagem natural, no idioma português formam um conjunto de dados pequeno para o volume esperado para aplicações de inteligência artificial. Em consequência, os métodos elegíveis também diminuem, porque as opções disponíveis são voltadas para a linguagem pragmática, para o mercado e para a interação intensa entre o usuário e o algoritmo, que têm à disposição um volume massivo de dados. Nessas circunstâncias, o popular modelo “intenção + entidade” não é aplicável.

No contexto de dados textuais científicos para a classificação de documentos para busca em bibliotecas, a técnica com mais chances de ser escolhida é a aprendizagem não supervisionada embutida na modelagem de tópicos (DEERWESTER *et al.*, 1990; HOFMANN, 1999; BLEI; NG; JORDAN, 2003). Não há muitas chances de se utilizar a aprendizagem supervisionada, devido à necessidade de um volume de dados rotulados muito grande.

Nesta etapa, enfatizou-se a escolha de modelos de inteligência artificial pelo tipo de aprendizagem artificial, que acaba sendo uma escolha baseada na disponibilidade de dados rotulados. Por exemplo, a aprendizagem supervisionada só é factível se houver um conjunto de dados rotulados grande o suficiente para obter modelos generalizáveis, precisos e estáveis. Geralmente, não há essa quantidade de dados científicos, textuais e rotulados. Só para se ter uma ideia, quando se fala de quantidade razoável está se falando de milhões. Na próxima seção, se discorre quanto aos modelos selecionáveis e as implicações decorrentes de cada escolha.

Seleção da técnica de modelagem

Cada técnica de modelagem precisa de um *pipeline* específico, de forma que o modelo de inteligência artificial escolhido determina o processo subsequente e, dependendo do momento que ocorrer a escolha, a etapa de Compreensão de Dados precisa ser reiniciada. Logo, a seleção da técnica de modelagem deve ser feita o mais cedo possível.

Esta etapa também precisa se adequar à tarefa e ao volume de dados disponíveis. Por exemplo, modelos sequenciais, como todas as técnicas inteligentes de aprendizado supervisionado, precisam de elevado volume de dados rotulados. Se eles não estiverem disponíveis, essas técnicas não são elegíveis. Claro que os profissionais da informação podem se sentirem intimidados com a complexidade e as variedades de modelos de

inteligência artificial disponíveis, mas basta uma análise mais profunda para revelar um campo científico com muita repetição e redundância de técnicas, porque é uma área com muitos grupos concorrentes.

A modelagem inteligente depende da etapa de compreensão dos dados, pois é a qualidade e a quantidade dos dados que vai dizer quais modelos podem ser utilizados.

Por fim, para saber se a técnica selecionada é a correta, avalia-se os modelos na linha de produção por meio de testes bem planejados. Por isso, na próxima seção, discorre-se sobre o planejamento de testes em linha de produção.

Planejamento de testes

Antes de produzir um modelo de inteligência artificial, é preciso planejar os testes de avaliação dos resultados na linha de produção, porque as condições de simulação são sempre diferentes das condições encontradas na linha de produção efetivamente. Não raro, ocorre de bons modelos serem inúteis na execução de tarefas reais, afinal nem sempre os atributos do conjunto de dados estão disponíveis no local de trabalho e nem sempre quem produziu o modelo conhece realmente o que se passa na linha de frente.

Os testes devem ser planejados antes de se ter qualquer resultado para evitar o risco de se elaborar testes para justificar os resultados. E o que se quer é que os testes capturem realmente a qualidade do modelo para o ambiente real. As melhores métricas não garantem melhores resultados, apenas os resultados reais podem dizer se o modelo é válido.

Um modelo de inteligência artificial deve ser avaliado por métodos e técnicas semelhantes às usadas para experimentos reais de outros campos científicos. Os testes devem ser controlados em cenários reais. Uma avaliação não significa uma conclusão, mas uma oportunidade de refinar o artefato por meio de *insights* obtidos com os testes (HEVNER *et al.* 2004). Isso também deve ser levado em conta ao se planejar testes de modelos.

O modelo obtido com o conjunto de treinamento tem sua qualidade e validade avaliadas no conjunto de teste, que deve ser uma amostra aleatória do *corpus* que tenha sido separada antes mesmo de se iniciar o treinamento do modelo de inteligência artificial. Os dados dos conjuntos de treinamento e validação não devem pertencer ao conjunto de testes, pois isso pode afetar todas as métricas de avaliação. Segundo Zhang *et al.* (2021, p. 318), nessa amostragem aleatória:

cada exemplo é uma subsequência capturada arbitrariamente da longa sequência original e que duas amostras aleatórias adjacentes não são necessariamente adjacentes na sequência original. Para modelagem de linguagem, o objetivo é prever o próximo termo com base nos

termos que já estão na base de dados, portanto os rótulos são a sequência original, deslocada por um termo (tradução nossa).

Enfim, antes de ajustar e avaliar um modelo, é preciso planejar a bateria de testes de avaliação e o que se pretende mensurar. Depois de planejar os testes, ainda é preciso ajustar os modelos, que é o assunto da próxima seção.

Ajustar modelos

A tarefa específica da etapa de Compreensão da inteligência artificial consiste em construir um modelo com parâmetros configurados automaticamente por heurísticas aplicadas aos dados de treinamento. Mas o modelo deve executar bem a tarefa que lhe é atribuída, de forma que ele deve ser reajustado tantas vezes quanto necessário até se atingir o nível satisfatório.

Para compreender como isso funciona, é preciso distinguir entre parâmetros e hiperparâmetros. Os dados e as heurísticas alcançam apenas os parâmetros que, na realidade, são coeficientes de equações lineares que podem ter milhares de variáveis, enquanto os hiperparâmetros podem ser definidos pelo pesquisador. São hiperparâmetros: o número de tentativas completas para alcançar uma solução ou desistir (número de épocas), o número de camadas ocultas, o número de neurônios e o tipo da função de ativação. Existem outros hiperparâmetros, mas todos eles são atribuições da informática. Então, ajustar um modelo é basicamente testar centenas de possibilidades de hiperparâmetros, o que exige um amplo conhecimento técnico e muito tempo e recurso computacional. Um modelo ajustado nem sempre é o melhor modelo, pode ser apenas o melhor que se consegue considerando recursos, dados, competência e tempo disponíveis.

Enfim, a tarefa de “modelagem” produz um modelo configurado, logo pronto para uso na perspectiva da Informática. E a próxima tarefa é avaliar esse modelo no ambiente real de trabalho. Assim, na seção a seguir, discorre-se sobre a atividade de avaliar o modelo em uma tarefa real para ser validado na perspectiva da BCI.

Avaliando o modelo

Avalia-se um modelo de inteligência Artificial configurado de acordo com as métricas definidas na Etapa 1 por meio dos testes planejados na atividade específica “planejar testes”, porque um modelo computacional pode ter excelentes métricas computacionais e, mesmo assim, em situações reais, apresentar instabilidade, ou incapacidade de generalização, ou baixa precisão. Desse modo, as questões norteadoras desta atividade são:

1. O modelo está correto?
2. Os atributos usados no treinamento estão disponíveis na tarefa real?
3. As atividades do treinamento podem ser reproduzidas na tarefa real?
4. O Projeto é satisfatório ou deve ser reiniciado?

Modelos de inteligência artificial são métodos empíricos que não explicam a própria configuração. Não existem formas de explicar os parâmetros dos modelos de IA, de forma que é preciso “determinar o quão bem um artefato funciona” (HEVNER *et al.*, 2004). Essa tarefa é importante, uma vez que todo o processo da análise de assuntos futuro depende da correção atual do modelo.

Basicamente, o teste mais simples é escolher uma amostra aleatória de dados já classificados, utilizar o modelo para classificar esses mesmos dados e cotejar as respostas. No entanto, como afirma Lancaster (2004, p. 285), o problema da recuperação da informação é que se “coteja aproximações de necessidades de informação com aproximações de indexações”. Esse grau de incerteza aponta que nem sempre haverá uma resposta exata, mas apenas “aproximações”. Assim, os testes também apresentam resultados aproximados e cabe ao profissional da informação distinguir as melhores “aproximações”. Desse modo, a avaliação deve ser um processo iterativo, paciente e exaustivo, levando em conta a dinamicidade e a multiplicidade de respostas corretas possíveis.

Num cenário ideal, o profissional da informação obtém muitos modelos de inteligência artificial, testa todos e seleciona o melhor. Mas a realidade é que quase sempre existe apenas um modelo disponível, porque a quantidade de dados rotulados é insuficiente. Além disso, há escassez de tempo e limitação de recursos humanos, financeiros e tecnológicos, de forma que haverá poucas alternativas e o profissional da informação contará com sua própria habilidade, experiência e até um pouco de intuição na avaliação dos modelos.

A avaliação dos modelos na perspectiva da informática extrapola o escopo desse trabalho e, por isso, pressupõe-se que os modelos, quando avaliados, estejam computacionalmente corretos. Ainda sobre as avaliações, na perspectiva da BCI, os resultados dependem do contexto e da habilidade técnica do profissional da informação. Em outras palavras, não existe um método universal de testagem dos modelos de inteligência artificial.

Aprovado o modelo de inteligência artificial, a próxima etapa consiste na implantação do modelo, que é a Etapa Compreensão da inteligência aumentada, descrita na próxima seção.

Etapa 4: Compreensão da inteligência aumentada

A inteligência aumentada é um jeito de planejar a execução de tarefas com o auxílio de ferramentas e é também o nome que se dá ao efeito que a ferramenta produz, que é expandir a inteligência do usuário. Por exemplo: a inteligência de um motorista com um

aplicativo de navegação é expandida de tal forma que ele consegue encontrar um endereço em uma cidade totalmente desconhecida (note que o aplicativo dá informações, mas não dirige o carro). Assim, espera-se que a ferramenta de assistência à análise de assuntos forneça informações suficientes para o indexador navegar pelas “vuelas”, “becos” e “grandes avenidas” dos significados do texto, de forma que esse processo resulte em expansão cognitiva. As questões norteadoras desta Etapa são:

1. Como implantar o modelo no ambiente de trabalho?
2. Como treinar o profissional que vai utilizar o modelo?
3. Quais os impactos sociais do modelo?

Uma vez elaborado o Projeto, coletados os dados, configurado um modelo de inteligência Artificial, desenvolvida a aplicação (*software*), resta a implantação do Modelo Conceitual como fluxo normal de trabalho e produção. E esta é exatamente a finalidade desta Etapa, que é discorrer sobre a adoção do sistema e o treinamento do profissional da Informação.

Depois que o sistema estiver pronto, é importante responder às seguintes questões relativas ao Projeto (as respostas devem constar no projeto):

1. O problema foi resolvido?
2. Os objetivos foram alcançados?
3. Todas as lacunas foram tratadas?
4. Houve aumento da produtividade e da produção?
5. As condições de trabalho melhoraram para o profissional da informação?
6. Houve um aumento intelectual da equipe?
7. Quais são os impactos dessa solução na comunidade?

O sucesso em longo prazo da inteligência aumentada depende da estratégia de implementação. Todos os procedimentos devem ser bem documentados e a disciplina (sistematização, capacidade de manter as regras) deve ser mantida. Além disso, é importante atentar para que sistemas de inteligência aumentada, como todos os sistemas tecnológicos, estão em constante evolução, o que implica:

desenvolvimento diário de sistemas de aprendizado de máquina; os profissionais precisam decidir se coletam mais dados, aumentam ou diminuem a capacidade do modelo, adicionam ou removem recursos de regularização, aprimoram a otimização de um modelo, melhoram a inferência aproximada em um modelo ou depuram o *software*. Todas essas operações consomem, no mínimo, tempo para serem experimentadas, por isso é importante ser capaz de determinar o curso de ação correto em vez de [tentar] adivinhar cegamente (GOODFELLOW; BENGIO e COURVILLE, 2016, p. 503).

A estratégia de implantação, “curso de ação correto”, deve incluir como e por quem o *software* será utilizado e como serão treinados os usuários. Deve incluir também o monitoramento e a manutenção do sistema para que os resultados da inteligência aumentada permaneçam coerentes e consistentes.

O estado da arte da sistematização da inteligência aumentada na BCI está em andamento (WÓJCIK, 2021). Assim o enfoque deve abranger as técnicas consolidadas e robustas, incluindo estar preparado para se adaptar ao inesperado. Com isso, pode-se dizer que o Modelo Conceitual está completo.

Considerações gerais sobre o modelo conceitual

Nesta metodologia de uso, existem pelo menos duas contribuições importantes. Primeiro, analisou-se a utilidade de empregar a inteligência aumentada na análise de assuntos e pesquisou-se como utilizar as ferramentas inteligentes para assistir as pessoas (e não para substituí-las), fornecendo meios de usar a tecnologia na linha de produção da informação. Em segundo lugar, produziu-se um Modelo Conceitual para orientar o desenvolvimento de ferramentas inteligentes no contexto da análise de assuntos.

Os Modelo Conceitual não são estranhos à BCI, sendo uma de suas ferramentas metodológicas importantes e populares. Quando bem utilizados na prática, produzem artefatos mais bem elaborados, com maiores garantias de qualidade e agem como o ponto de convergência entre os envolvidos no processo.

A inteligência aumentada como padrão de projeto coloca as necessidades humanas como o centro das atividades ao subsidiar o desenvolvimento e a utilização de ferramentas tecnológicas inteligentes. Ela é sensivelmente humana em contraponto às automatizações insensíveis, pois, ao mesmo tempo, acrescenta os benefícios da automatização sem remover postos de trabalho.

Ao diminuir a carga de trabalho, a inteligência aumentada reduz o estresse e retira da equação a possibilidade de “profissionais da informação substituídos pela inteligência artificial” e coloca nessa mesma equação a possibilidade de “profissionais da informação servidos pela inteligência artificial”. Assim, o profissional da informação só tem a ganhar, pois não se procurou automatizar as respostas, mas desenvolver meios assistir o processo de dar respostas.

Pelo menos em teoria, a inteligência aumentada previne erros, deficiências e ajuda a tomada de decisão, enquanto aprende. Com base nisso, a lógica do Modelo Conceitual, descrito aqui, é colocar humanos e computadores juntos de forma a criar um sistema social e tecnológico colaborativo capaz de superar as limitações atuais das automatizações.

REFERÊNCIAS DO APÊNDICE A

- BLEI, D.; CARIN, L.; DUNSON, D. Probabilistic Topic Models. **IEEE Signal Processing Magazine**, v. 55, n. 4, p. 55–65, nov. 2010.
- CHAPMAN, P. et al. **CRISP-DM 1.0: Step-by-step data mining guide**. [s.l.: s.n.].
- DEERWESTER, S. et al. Indexing by latent semantic analysis. **Journal of the American Society for Information Science**, v. 41, n. 6, p. 391–407, set. 1990.
- GOLUB, K. Automated Subject Indexing: An Overview. **Cataloging & Classification Quarterly**, v. 59, n. 8, p. 702–719, 17 nov. 2021.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge, Massachusetts: The MIT Press, 2016.
- GROGAN, D. **A prática do serviço de referência**. Tradução: Antônio Agenor Briquet de Lemos. 1ª edição ed. Brasília, DF: Briquet De Lemos, 2001.
- HEVNER et al. Design Science in Information Systems Research. **MIS Quarterly**, v. 28, n. 1, p. 75, 2004.
- HOFMANN, T. **Probabilistic latent semantic indexing**. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '99. **Anais...** Em: THE 22ND ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE. Berkeley, California, United States: ACM Press, 1999. Disponível em: <<http://portal.acm.org/citation.cfm?doid=312624.312649>>. Acesso em: 3 jun. 2022
- KOBASHI, N. Y. Análise Documentária e Representação da Informação. **INFORMARE**, v. 2, n. 2, p. 5–27, 1996.
- LANCASTER, F. W. **Indexação e resumos: teoria e prática**. Tradução: Antônio Agenor Briquet de Lemos. Brasília, DF: Briquet de Lemos / Livros, 2004.
- ROSE, K. H. A Guide to the Project Management Body of Knowledge (PMBOK® Guide)-Fifth Edition: Cover to Cover. **Project Management Journal**, v. 44, n. 3, p. e1–e1, jun. 2013.
- STUDER, S. et al. Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. **Machine Learning and Knowledge Extraction**, v. 3, n. 2, p. 392–413, 22 abr. 2021.
- WÓJCIK, M. Augmented intelligence technology. The ethical and practical problems of its implementation in libraries. **Library Hi Tech**, v. 39, n. 2, p. 435–447, 21 jun. 2021.
- ZAKI, M.; MCCOLL-KENNEDY, J. R. Text mining analysis roadmap (TMAR) for service research. **Journal of Services Marketing**, v. 34, n. 1, p. 30–47, 8 jan. 2020.

APÊNDICE B – RECURSOS COMPUTACIONAIS

No desenvolvimento de um *software* utiliza-se uma ampla gama de recursos preexistentes. Os *softwares*, dada a sua natureza virtual, congregam os recursos de programação e os recursos informacionais para processar os dados e produzir a informação. Isso significa que os recursos computacionais utilizam os recursos informacionais para aplicar as regras, os procedimentos e as avaliações sobre os dados (insumos) para produzir as informações (produto).

Neste apêndice, nas seções pertinentes, estão descritos os principais recursos utilizados, em que o relato técnico da relação de cada recurso com esta pesquisa foi ilustrado e reforçado por uma tabela contendo a identidade (nome, marca ou logotipo), a descrição e a URL do recurso. Na próxima seção, estão descritos os recursos computacionais.

1 Recursos computacionais

Os recursos computacionais podem ser classificados em *hardware* (recursos físicos) e *software* (recursos de programação). Os recursos físicos são muito caros e alguns só estão disponíveis na nuvem. Por exemplo, um computador adequado para o treinamento de modelos de inteligência artificial requer uma elevada capacidade de processamento que só pode ser fornecida por processadores com TPU, conforme explicado na próxima seção.

1.1 Processadores

O processamento de grandes volumes de dados textuais exige máquinas muito mais poderosas que os computadores pessoais e que só estão disponíveis na Nuvem. Dito de outra forma, é necessário comprar acesso aos poderosos computadores e *softwares* disponíveis na nuvem para treinar modelos sequenciais, caso contrário, o tempo necessário pode ser proibitivo.

Os computadores comuns utilizam a Unidade de Processamento Central (CPU), enquanto computadores das nuvens usam *Graphics Processing Unit* (GPU) desenvolvidas para a renderização gráfica e o processamento em tempo real de dados. A CPU e a GPU são de propósito geral e se diferenciam apenas pela maior velocidade desta última.

Os centros de processamento corporativos, as nuvens, desenvolveram a *Tensor Processing Unit* (TPU) especificamente para o processamento de dados multidimensionais do aprendizado profundo. Essa especialização em modelos de inteligência artificial baseados em tensores torna esses processadores muito mais velozes, mas os inviabiliza para o uso em

tarefas de propósito geral, por isso eles estão disponíveis apenas nas nuvens. Davenport (2018, p. 78) afirmou que é muito “difícil criar recursos competitivos de inteligência artificial sem a nuvem” e, por isso, na próxima seção, estão descritas as formas de acesso à nuvem computacional.

1.2 Nuvem computacional

Os termos “computação em nuvem” (*cloud computing*) e “nuvem” referem-se aos *data centers* que vendem processamento computacional e espaço de armazenamento virtual como um serviço (*computing as a service*). Eles vendem acesso a supercomputadores, que se dá de qualquer lugar com Internet.

Nesta pesquisa utilizou-se os serviços de nuvens do *Google Cloud (Google Colab)*, cujos servidores (serviço pago) são dotados de processadores com TPU (*tensor processing unit*). Esses serviços não são necessários para utilizar o 3ApC.

Os modelos sequenciais descritos nesta pesquisa precisam ser treinados por aprendizagem computacional, o que demanda o uso de TPU. Assim, foi necessário comprar os serviços de computação de uma Nuvem. Além da capacidade de processamento e o espaço de armazenamento disponibilizados serem adequados ao propósito da pesquisa, as nuvens, ainda possuem a vantagem de custarem bem menos do que a aquisição dos recursos computacionais necessários.

Esta pesquisa, como a maioria das pesquisas que envolvem a elaboração de *softwares*, fez uso de módulos de *software open source* de alta qualidade, as *libraries*, que estão descritas nas próximas seções.


1.3 Iniciativa de códigos-fontes abertos

O termo “*open source*” ou “fonte aberta” refere-se ao recurso virtual que, além de ser de uso livre, também permite a modificação e a redistribuição do conteúdo sob os princípios FAIR.

Softwares livres são muito eficientes e seus códigos-fontes abertos podem ser obtidos, utilizados, modificados e redistribuídos sem a necessidade de pagar por uma licença. As *Big Techs* promovem esse tipo de iniciativa para baratear os *softwares*, porque, pelo menos em teoria, corrigir erros em códigos-fontes de computação publicamente é mais fácil e seguro do que contratar uma equipe privada para essa tarefa.

Os *softwares* livres são amparados pela iniciativa do código-fonte aberto (OSI, do inglês, *Open Source Initiative*), que combate preconceitos, ajuda a reprimir outros crimes cibernéticos e estabelece os padrões de qualidade e de garantias de direitos. No Quadro 1, podem ser vistas as informações relevantes sobre essa iniciativa.

Quadro 1 – Informações da iniciativa de *Software* livre e aberto.

Identificador	Descrição	URL
	A <i>Open Source Initiative (OSI)</i> é uma entidade sem fins lucrativos para defender as boas práticas <i>open sources</i> .	https://opensource.org/

Fonte: Elaborado pelo autor.

Os *softwares* de código aberto são mais transparentes e confiáveis, supondo-se, é claro, que qualquer um pode ler e entender o código do *software*, caso haja alguma dúvida ou armadilha, propor mudanças e correções, e, se necessário, denunciar.


Em computação moderna, na parte de programação, uma *library* é um código-fonte de elevado nível de qualidade e confiabilidade, que contém muitas funcionalidades. Esses valiosos recursos computacionais de programação permitem aproveitar algoritmos especializados elaborados e validados por autoridades científicas da ciência da computação e do campo científico em que a *library* foi desenvolvida.

Nas próximas seções serão apresentados os *softwares open sources* imprescindíveis a esta pesquisa, inclusive a linguagem de programação *Python*, que está descrita próxima seção.

1.4 Linguagem de programação *Python*

Desenvolvida sob uma licença de código-fonte aberto aprovada pela OSI, a linguagem de programação *Python* pode ser utilizada e distribuída livremente mesmo para uso comercial. As informações relevantes sobre esse recurso estão no /quadro 2.

Quadro 2 – Informações relevantes de acesso ao recurso *Python*.

Identificador	Descrição	URL
	<i>Python</i> é uma linguagem de programação de computadores de semântica dinâmica e <i>open source</i>	https://opensource.org/


Fonte: Elaborado pelo autor.

A *Python* se caracteriza por ser de alto nível e de baixo custo de manutenção devido à semântica dinâmica e a orientação a objetos. Além disso, encoraja o uso de pacote (*package*) e biblioteca (*library*) prontos, pois isso modulariza as aplicações, facilitando a correção e o reúso de códigos.

Os *softwares* produzidos em *Python* podem ser distribuídos como *script* no paradigma “interpretado”, em que o código-fonte é interpretado toda vez que o *software* for utilizado. Ou podem ser distribuídos, à maneira tradicional, como um programa executável no paradigma “compilado”. Em ambos os casos, é necessário uso da *Python Virtual Machine* (PVM) para traduzir o código-fonte para a linguagem de máquina e a mesma PVM converte o código de máquina em uma interface humanamente compreensível.

O *software* desta pesquisa foi desenvolvido em código-fonte escrito em *Python* em uma plataforma de programação (Quadro 3) denominada de *Pycharm Community*, versão 2022.2, produzida e disponibilizada pela *Jetbrains*.

Quadro 3 – As informações do recurso computacional *Pycharm Community*.

Identificador	Descrição	URL
	O <i>Pycharm Community</i> é um Ambiente de Desenvolvimento Integrado (IDE) que facilita criar aplicações com Interface Gráfica de Usuário (GUI).	https://www.jetbrains.com/pt-br/pycharm/download


Fonte: Elaborado pelo autor.

O *software* foi escrito nesse ambiente de programação. Por questões técnicas, antes da implementação definitiva, os módulos desenvolvidos foram testados no *Jupyter*, que está descrito na próxima seção.

1.5 A interface de programação experimental

A interface de programação experimental *Jupyter* (Quadro 4) foi utilizada para testar os módulos do *software* desenvolvido nesta pesquisa. A *Jupyter* foi projetada para ser usada na *Web*, diretamente no navegador, em que permite criar códigos, gerar relatórios e visualizar resultados em um único documento, oferecendo uma experiência simplificada para o programador.

Quadro 4 – Os metadados do ambiente de programação experimental.

Identificador	Descrição	URL
	A <i>Jupyter</i> foi projetada para testar e experimentar códigos-fontes diretamente no navegador <i>web</i> .	https://jupyter.org/

Fonte: Elaborado pelo autor.

Uma das vantagens do uso de ambiente experimental é não ter que lidar com a GUI, pois ela pode ser a última a ficar pronta. Além disso, a *Jupyter* permite a execução de fragmentos do código-fonte, o que é muito mais rápido.

O código testado no *Jupyter* pode ser adicionado ao *software* em elaboração, sem alteração, facilitando o processo de integração entre o computador pessoal e a nuvem por


meio do recurso de compartilhamento de arquivos denominado de *Notebook*. Esse recurso aumenta a produtividade, porque favorece o compartilhamento do código-fonte.

Outra vantagem do *Jupyter* é facilitar o reconhecimento e a aprendizagem do uso de *Library* de alta qualidade disponíveis para *download*. Na próxima seção, apresenta-se a *library Tensorflow* especializada no Aprendizado Automático.

1.6 *Library Tensorflow*

A *Library Tensorflow* (Quadro 5) é uma das mais importantes para o aprendizado profundo, cuja ideia central é uma matriz multidimensional denominada “tensor”. A título de esclarecimento, um dado sem dimensão é um escalar; com uma dimensão, é um vetor; com duas dimensões, é uma tabela; e com três ou mais dimensões, é um tensor.

Quadro 5 – Os metadados descritivos da *Library Tensorflow*.

Identificador	Descrição	URL
 TensorFlow	A <i>Library Tensorflow</i> permite criar modelos de aprendizado profundo.	https://www.tensorflow.org/

Fonte: Elaborado pelo autor.

A *Tensorflow* providencia as estruturas e os códigos do arcabouço teórico das redes neurais profundas tradicionais, recorrentes e convolucionais, bem como as funcionalidades mais eficientes para o treinamento artificial dos algoritmos em TPU e em processamento paralelo ou distribuído. Note que a TPU foi desenvolvida especialmente para a *Tensorflow* e é por isso que as duas têm o termo “tensor” no nome.

APÊNDICE C – RESUMOS POR SUMARIZAÇÃO AUTOMÁTICA.

Resumos
<p>1129606 — Óxidos mistos obtidos de compostos tipo hidrotalcita contendo metais de terras raras e seu uso na reforma a vapor de biogás — As conversões de CH₄ e CO₂ na reforma a vapor de biogás simulado obtidas foram em torno de 97% e 58%, respectivamente, mostrando que estes catalisadores são promissores para reforma de biogás. Testes de reforma a vapor de biogás simulado foram realizados em um reator de leito fixo, da marca PID Eng&Tech, na temperatura de 750°C e pressão atmosférica, utilizando 20 mg de OM misturado com 20 mg de carbeto de silício, por um período de 24 h. O biogás foi simulado usando razão molar CH₄:CO₂ de 2,33:1 (70/30%). Conversão de CH₄ e de CO₂, e razão molar Hz/CO, na reforma a vapor de biogás simulado a 750 CC 6 razão S/C de 3,2, usando NiAlPr. Conversão de CH₄ e de CO₂, 6 razão molar Hz/CO, na reforma a vapor de biogás simulado a 750 CC 6 razão S/C de 3,2, usando NiAlGd. A razão molar H₂/CO aumenta com o aumento da razão CH₄/CO₂ na reforma a vapor de biogás, variando entre 2 e 3 para razões CH₄/CO₂ entre 1 e 5 a 700 °C.</p>
<p>161390 — Capacidade predatória de Orius insidiosus predando Aphis gossypii sob o efeito da temperatura e variação da umidade relativa e fotoperíodo — Capacidade predatória de Orius insidiosus predando Aphis gossypii sob o efeito da temperatura e variação da umidade relativa e fotoperíodo J. E. M. OLIVEIRA, S. A. DE BORTOLI, R. F. SANTOS, J. P. BRITO, J. R. MIRANDA Estudou-se a capacidade predatória de Orius insidiosus predando o pulgão do algodoeiro Aphis gossypii sob o efeito da temperatura (25°C) em diferentes umidades relativas (50, 60 e 70%) e fotoperíodos (10: 14; 14: 10 e 12: 12 de luz:escuro). Em levantamentos realizados por DE BORTOLI & OLIVEIRA (2006) e plantios de algodão na região de Jaboticabal-SP, foi constatada a presença da espécie de predador Orius insidiosus (Hemiptera: Anthocoridae) sempre associada à presença de pulgões, presas consideradas adequadas ao desenvolvimento desse predador, devido principalmente ao seu tamanho. A predação é um processo complexo, afetado por fatores básicos, como densidades da presa e do predador, e por fatores secundários, envolvendo as características do ambiente, da presa e do predador (HOLLING, 1961) sendo que variações das condições ambientais como de temperatura, umidade relativa e fotoperíodo, as quais o predador está submetido, por influenciarem na capacidade predatória, sobrevivência e reprodução. Dessa forma, os valores indicam que o predador O. insidiosus em condições de até 70% de umidade relativa com 14: 10h de fotoperíodo possui boa capacidade de busca e consumo de A. gossypii, quando comparados com luminosidade e umidades inferiores. preyed nymphs cotton aphid Aphis gossypii to determine the survival, longevity on temperature of the 25°C under different relative humidity (50, 60 and 70%) and photoperiods (10:14: The effect of selected dietary supplements on survival and reproduction of Orius insidiosus (Say) (Hemiptera: Anthocoridae).</p>
<p>1128271 — Geadas na viticultura e estratégias para prevenção de danos no sul do Brasil — GH WHFLGR (VWIGLR GH UJmRV 7HFLGRV TXH QR SHUtRGR GH GRUPRqFLD DV YLGHlUDV VmR LQtFLR GH EURWDomR DSyV R SHUtRGR E FUTWLFGR GH JHDGDV WDUGLDV GH XPD DPSOLWXGH GH GH medidas de proteção, principalmente preventivas (passivas), correspondem ações e cientes e DMXVWiYHLV D UHDOLGDGH GH</p>
<p>211707 — Comparação da eficiência de métodos de seleção em gerações segregantes de feijoeiro-comum considerando a resistência à antracnose e o rendimento — Considerando conjuntamente a porcentagem de linhagens resistentes à C. lindemuthianum, a produtividade média das linhagens, a variância genética entre as linhagens geradas por cada método, o número e o percentual de linhagens com produtividade superior à testemunha, a eficiência do método variou com o critério utilizado para compará-los. Para a avaliação da eficiência dos métodos de seleção, foi calculada a porcentagem de linhagens resistentes aos quatro patótipos de C. lindemuthianum utilizados, as análises conjuntas da produtividade das linhagens nas três localidades/ano por cruzamento e método de seleção, a porcentagem de linhagens com produtividade superior à testemunha e a estimativa da variância genética entre as linhagens geradas em cada método dentro</p>

de cada cruzamento e por método considerando todos os cruzamentos. As populações segregantes conduzidas pelos métodos DF4 e SMM seguiram a mesma metodologia de seleção até a geração F4, diferindo apenas na geração em que foi realizada a seleção individual das plantas: em F4, no método DF4, e em F5, no método SMM. Ao considerar as variáveis porcentagem de linhagens resistentes aos quatro patótipos de *C. lindemuthianum*, produtividade das linhagens, estimativa da variância genética e porcentagem de linhagens com produtividade superior à testemunha, os resultados foram bastante contrastantes, quando considerados os cruzamentos que deram origem às linhagens, caracterizando a influência dos genitores e do componente interação linhagem x local. Considerando-se conjuntamente a porcentagem de linhagens resistentes à *Colletotrichum lindemuthianum*, a produtividade média das linhagens, a variância genética entre as linhagens geradas por cada método, o número e o percentual de linhagens com produtividade superior à da testemunha, a eficiência do método varia com o critério utilizado.

119110 — Uma nova etapa da Instrução Normativa nº 51: a Região Centro-Sul — Medidas de monitoramento da implementação da IN nº 51/2002 Prevendo-se dificuldades de implementação na Região Centro-Sul, no que se refere aos anexos IV (Regulamento Técnico de Identidade e Qualidade de Leite Cru Refrigerado) e V (Regulamento Técnico de Identidade e Qualidade de Leite Pasteurizado), o Ofício Circular/Dipoa nº 24/2005, de 29/6/2005, estabeleceu que nos próximos seis meses a partir de 1/7/2005 haveria uma fase de transição para a implantação dos novos parâmetros de controle da qualidade do leite produzido e/ou distribuído nas regiões Sul, Sudeste e Centro-Oeste. Regulamento Técnico de Produção, Identidade e Qualidade de Leite Tipo B; III) Regulamento Técnico de Produção, Identidade e Qualidade de Leite Tipo C; IV) Regulamento Técnico de Identidade e Qualidade de Leite Cru Refrigerado; V) Regulamento Técnico de Identidade e Qualidade de Leite Pasteurizado; e VI) Além do tempo necessário à realização do teste redutase de detecção de antibióticos no leite (encontrado principalmente em leite de gado submetido a tratamento contra mastite, que o torna impróprio para o consumo), bem como a realização do teste alizarol 76% no leite tipo C recebido em latões após as 10 horas, o que diferencia os três tipos de leite são os requisitos exigidos para a realização dos dois principais testes (CPP/CBT e CCS), que deverão vir em um máximo de 10 mil UFC/ml e de 600 mil CS/ml para a análise do tipo A; e de 500 mil UFC/ml e de 600 mil CS/ml para a análise do tipo B, respectivamente. Quanto ao teor de gordura, o leite pasteurizado deve ser classificado como: integral (leite pasteurizado integral), padronizado a 3% m/m – três por cento massa/massa (leite pasteurizado padronizado), semidesnatado (leite pasteurizado semidesnatado), ou desnatado (leite pasteurizado desnatado), e rotulado segundo o tipo (A, B ou C). Em laticínios de pequeno porte, poderá ser adotado o processo de pasteurização lenta na produção de leite para abastecimento público, ou mesmo para a produção de derivados lácteos, desde que: 1) o equipamento de pasteurização a ser utilizado atenda aos requisitos determinados no Regulamento de Inspeção Industrial e Sanitária de Produto de Origem Animal (Riispoa, ou especificados em regulamento técnico específico; 2) o envase seja realizado em circuito fechado, no menor tempo possível e sob condições que minimizem contaminações; 3) a matéria-prima satisfaça às especificações de qualidade estabelecidas pela legislação referente à produção de leite pasteurizado, excetuando-se a refrigeração do leite e seu transporte a granel, quando o produto puder ser entregue em latões, ou em tarros, e na temperatura ambiente do estabelecimento processador em, no máximo, até duas horas após o término da ordenha; e 4) o leite a ser submetido à pasteurização lenta não tenha sido previamente envasado em estabelecimento sob inspeção sanitária federal.

158728 — Teor de açúcares solúveis e insolúveis em folhas de videiras, cv. Syrah, em diferentes posições no ramo e épocas do ano — TABELA 1 - Coeficientes de correlação entre variáveis climáticas e do metabolismo de carboidratos em folhas de videiras 'Syrah', durante os ciclos produtivos do primeiro semestre de 2003 (Ciclo 1- acima da linha pontilhada) e do segundo semestre de 2003 (Ciclo 2- abaixo da linha pontilhada). TABELA 2 - Coeficientes de correlação entre variáveis climáticas e do metabolismo de carboidratos em folhas de videiras 'Syrah', durante os ciclos produtivos do primeiro semestre de 2004 (Ciclo 3- acima da linha pontilhada) e do segundo semestre de 2004 (Ciclo 4- abaixo da linha pontilhada). Durante o crescimento dos frutos e durante as fases de crescimento intenso dos ramos, devido a um grande transporte de açúcares das folhas para as bagas ou para os meristemas, o teor de amido nas folhas é baixo. Por outro lado, em fases em que não há grande transporte de açúcares das fontes (folhas fotossintetizantes) para os drenos (ramos e folhas em crescimento e frutos em maturação), como nas fases finais do ciclo produtivo, o amido tende a

acumular nas folhas, regulando a atividade fotossintética (Pimentel, 1998). De acordo com Teixeira (2001), a temperatura média durante esse ciclo foi abaixo da média histórica da região, provocando uma diferença nas respostas metabólicas das videiras, em relação ao ciclo 1, que apresentou temperaturas mínimas, sempre acima de 15oC, e freqüentemente em torno de 20oC e ao ciclo 4, com temperaturas em torno de 15oC e sem temperaturas abaixo de 13oC. Além das temperaturas mais baixas, os valores de radiação, durante o ciclo 2, são inferiores a 400 ly.dia-1, sendo, em média, inferiores aos demais ciclos.

159841 — **Disponibilidade de cádmio em diferentes solos do Rio Grande do Norte** — Com relação ao DTPA-TEA o maior teor do metal foi extraído do Argissolo Vermelho Amarelo (PVA), sendo 222 mg dm-3, na profundidade mais superficial desse solo (0 - 10 cm), enquanto que o menor teor, 81,5 mg dm-3, também no mesmo solo e na mesma profundidade (10 - 40 cm) dos teores retirados pelo Mehlich-1 e Mehlich-3. 0 – 10 10 – 40 0 – 10 10 – 40 0 – 10 10 – 40 0 – 10 10 – 40 0 – 10 10 – 40 O DTPA-TEA extraiu maior teor de cádmio na amostra do Argissolo Vermelho Amarelo (PVA), sendo 498 kg ha-1 na camada subsuperficial desse solo (10 - 40 cm) e o menor teor foi detectado no Latossolo Vermelho Amarelo (LVA), 110 kg ha-1, na profundidade de 0 - 10 cm. Quanto ao CaCl2 pôde-se verificar a menor extração dentre os extratores, detectando apenas 5,7 kg ha-1 na parte mais superficial (0-10 cm) da amostra de Cambissolo Háplico (CX), enquanto que o maior teor que foi de 351 kg ha-1, ocorreu na camada mais profunda (10 - 40 cm) do Argissolo Vermelho Amarelo (PVA), podendo-se constatar que para essa solução, o teor extraído ocorreu no mesmo solo e na mesma profundidade da extrações feitas pelos extratores Mehlich-1, Mehlich-3 e DTPA-TEA. Nas amostras superficiais (0 - 10 cm) do Neossolo Quartzarênico (RQ) e Argissolo Vermelho Amarelo (PVA) encontraram-se maiores teores de cádmio removidos pelos extratores Mehlich-1, Mehlich-3 e DTPA-TEA, podendo essa maior extração ter sido influenciada pelo teor de matéria orgânica, visto que nas características desses solos pôde-se verificar maiores teores (24,1 g kg-1), em relação aos teores das outras profundidades das amostras estudadas.

1110938 — **Ocorrência de insetos nocivos, inimigos naturais e avaliação do nível de doenças em sistema roça sem queimar de produção de cacau** — Manejo de Agroecossistemas e Agricultura Orgânica Ocorrência de insetos nocivos, inimigos naturais e avaliação do nível de doenças em sistema roça sem queimar de produção de cacau Occurrence of harmful insects, natural enemies and evaluation of the level of diseases in the system without burning of production of cocoa ALVES-JÚNIOR, Miguel1; CELESTINO FILHO, Pedro2, AUGUSTO, Sebastião Geraldo3 Professor da Faculdade de Engenharia Agrônômica. Foram considerados dois sistemas de manejo de cultivo, no cacau: a Roça Tradicional (RT) e a Roça Sem Queimar (RSQ). De forma geral, nas Roças Sem Queimar (RSQ) o nível de doença foi menor quando comparado as Roças Tradicionais (RT), principalmente no estrato acima de 6 anos (Tabela 1). As maiores diferenças são encontradas nos extratos acima de 6 anos, no qual o nível de doença sempre foi maior no sistema RT quando comparado Cadernos de Agroecologia – ISSN 2236-7934 – Anais do VI CLAA, X CBA e V SEMDF – Vol. Tema Gerador 9 Manejo de Agroecossistemas e Agricultura Orgânica Referências Bibliográficas AIME, M. C.; PHILLIPS-MOURA, W. The causal agentes of witches broon and frosty pod roto of cacao (Chocolate, Theobroma cacao) form a new lineage of Marasmiaceae.

153839 — **Desempenho de cultivares de alface tipo americana em cultivo de outono no Sul de Minas Gerais** — As cultivares 45-53R7, RPT 2020, Paloma, Raider, RPT 3030, AEL 10 e Rocco, com produtividade de massa fresca comercial oscilando entre 599,1 a 739,1 g planta-1, foram estatisticamente mais produtivas que as demais cultivares. The cultivars 45-53R7, RPT 2020, Paloma, Raider, RPT 3030, AEL 10 and Rocco, with yield of commercial fresh matter ranging from 599.1 to 739.1 g plant-1 were statistically more productive than the other cultivars. Quanto à massa fresca comercial, obtiveram-se dois grupos, em que as cultivares AEL 10, RPT 3030, Sandy, Paloma, Rocco, RPT 2020, 45-53R7 e Raider, com 599,1; 623,3; 663,5; 669,1; 679,7; 696,3; 712,4 e 739,1 g planta-1, respectivamente, sem diferirem estatisticamente entre si, foram as cultivares que apresentaram os melhores resultados. A cultivar Raider, tradicionalmente utilizada pelos produtores da região, neste experimento apresentou circunferência de cabeça de 47,6 cm, ficando no segundo grupo, sendo estatisticamente equivalente às cultivares Sandy, RPT 2020 e RPT 3030, com 46,7; 46,5 e 46,1 cm, respectivamente. CONCLUSÕES Pelos resultados obtidos neste experimento, para as condições e

época em que foi conduzido, indicam-se como sugestão de cultivo para os produtores de Santana da Vargem, MG as cultivares RPT 2020 e RPT 3030, como novas opções de cultivo; também consolidam a cultivar Raider, já tradicionalmente cultivada.

1122509 — **Efeito da escória de siderurgia nos atributos químicos de um Latossolo amarelo distrófico em Terra Alta - PA** — Utilizou-se um delineamento experimental em blocos casualizados, com três repetições, em esquema fatorial 4x3x8+2, sendo quatro granulometrias (material retido entre as peneiras ABNT 20 – 50; 50 – 100; 100 - 200 e < 200), três doses de escória (1,65, 3,81e 5,98 t ha⁻¹) necessárias para elevação das saturações por bases a 40, 70 e 100%, respectivamente, oito períodos de incubação da escória (15, 30 45, 60, 90, 120, 240 e 360 dias) e mais dois tratamentos adicionais, um com calcário e outro sem calcário e escoria (testemunha). Se utiliza un diseño experimental de bloques al azar con tres repeticiones en un esquema factorial + 2, cuatro tamaños de partículas 4x3x8 (material retenido en la ABNT 20-50; 50 a 100; 100 a 200 y <200), tres dosis de escoria (1,65, 5,98 3,81e t ha⁻¹) necesario para la elevación de la saturación de bases de 40, 70 y 100%, respectivamente, ocho períodos de incubación de la escoria (15, 30, 45, 60, 90, 120, 240 y 360 días) y dos tratamientos adicionales, una con y otra sin piedra caliza piedra caliza y escoria (control). Pesquisas realizadas com a escória distâncias entre a fonte produtora e a mostram que sua ação neutralizante na consumidora, acidez o que aumenta do solo consideravelmente o valor final dos assemelha-se à do calcário, entretanto produtos, em decorrência do elevado em alguns experimentos, têm sido constatados que a escória apresenta Agroecossistemas, v. 11, n. 1, p. 97 – 121, 2019, ISSN online 2318-0188 100 reação mais lenta no solo quando produção de culturas de interesse comparada ao calcário (FORTES 1993; com Agroecossistemas, v. 11, n. 1, p. 97 – 121, 2019, ISSN online 2318-0188 112 nas concentrações de Al e de H+Al aumento da dose de escória aplicada, com aplicação de escória em Latossolo, pode ser explicado pela dependência em razão da presença de agente Agroecossistemas, v. 11, n. 1, p. 97 – 121, 2019, ISSN online 2318-0188 120 PRADO, R. M.; FERNANDES, F. M. Escória de siderurgia e calcário na correção da acidez do solo cultivado com cana-de-açúcar em vasos.

1007731 — **Filmes de polpa de banana produzidos por batelada: propriedades mecânicas e coloração** — A melhor combinação entre módulo de elasticidade e máxima elongação foi a composição com 4,5% purê obtido da polpa de banana (g purê seco/100 g de solução filmogênica), 5% glicerol (g glicerol/100 g de purê seco) e 0,5% pectina (g pectina/100 g de solução filmogênica). Outra possibilidade em avaliação recente é a produção de filmes a partir de polpa de frutas (purê), como uma alternativa potencialmente viável para a confecção de filmes com boas propriedades mecânicas e de barreira a gases^{6,7}. Filmes de polpa de banana produzidos por batelada: Propriedades mecânicas e coloração se comparado aos demais filmes inodoros e incolores. No processamento de nossos filmes, vemos que pequenas frações de pectina podem ser empregadas para manter as propriedades mecânicas reduzindo a quantidade de purê no processamento. Comparação das propriedades mecânicas dos filmes de purê de banana com alguns trabalhos encontrados na literatura referentes a filmes processados a partir de purês de outras frutas.

958618 — **Epidemiologia molecular aplicada ao monitoramento de estirpes de Staphylococcus aureus na produção de queijo minas frescal** — Tais amostras foram colhidas da superfície do tanque de recepção, da superfície do tanque de estocagem do leite cru e do tanque de equilíbrio pós-pasteurização, do leite cru, do leite pasteurizado, da água, da porção final das tubulações relacionadas com o processamento, da superfície das formas, tumbler e mesa, das mãos do manipulador responsável pela elaboração do queijo e do queijo após serem embalados para comercialização. Superfície do tanque de recepção Leite cru do tanque de recepção Superfície do tanque de estocagem Leite cru do tanque de estocagem Tubulação de saída do pasteurizador Superfície do tanque de equilíbrio do leite pasteurizado 2013 n (amostra) (tipo de toxina) 2 (51,52) (sea, sea) 1 (22) (sea) 3 (1,14,23) (sec,seb,sea) 3 (3,24,35) (sea/sec,sec,sea) 2 (6,26) (sea,sea) 2 (27,47) (sea,sea) 3 (48,49,53) (seb/sec,sea,sea) 1 (21) (sea) 1 (56) (sec) 4 (7,17,18,74) (sea/seb,seb,sea,sec) 3 (19,20,69) (sea/tsst, sea, sec) 25 Os pontos de colheita de amostras que apresentaram maior frequência de isolamento de estirpes de Staphylococcus aureus foram as mãos do manipulador (19,5%), leite cru do tanque de recepção (14,6%), tanque de expansão de estocagem do leite cru (12,2%) e o leite cru do tanque de estocagem (12,2%). Com relação aos resultados da presença dos

genes das enterotoxinas e toxina TSST-1 em 25 amostras, pode-se observar na Tabela 3 14 amostras com positividade para o gene da enterotoxina SEA (56%), duas amostras com positividade para o gene SEB (8%), cinco amostras com positividade para o gene da enterotoxina SEC (20%), uma amostra apresentando positividade para os genes das enterotoxinas SEA e SEC (4%), uma amostra apresentando positividade para os genes das enterotoxinas SEB e SEC (4%), uma amostra apresentando positividade para os genes das enterotoxinas SEA e SEB (4%) e uma amostra apresentando positividade para os genes da enterotoxina SEA e da toxina TSST-1(4%).

1121184 — Análise comparativa entre índices de vegetação e sua relação com o balanço hídrico em soja — Sabendo-se da importância econômica da soja para o RS, e do fato de que um dos principais fatores de variabilidade na produtividade é a deficiência hídrica e, ainda, que é possível monitorá-la usando índices de vegetação, o objetivo deste trabalho foi analisar o padrão temporal dos índices NDVI e EVI para três safras soja com condições hídricas distintas durante o ciclo, de forma a possibilitar a avaliação da relação entre IVs e as condições hídricas para a cultura. Para 164 plataforma SATveg (Sistema de Análise Temporal da Vegetação) (SATVeg, 2019) que 165 é uma ferramenta destinada Web desenvolvida pela Embrapa Informática Agropecuária, Agropecuária, ao acesso e visualização de per- destinada realizar uma melhor comparação entre os dados 165 é uma ferramenta Web desenvolvida pela Embrapa Informática Agropecuária, destinada fis dos índices vegetativos NDVI e EVI do sensor e IVs, os valores de precipitação pluvial (mm) 166 temporais ao acesso e visualização de perfis temporais dos índices vegetativos NDVI e EVI orológicos do 166 ao acesso e visualização de perfis temporais dos índices vegetativos NDVI e EVI do MODIS em qualquer local da América do Sul. medidos e os valores de ETc (mm), excesso e deficiência 167 sensor MODIS em qualquer local da América do Sul. Índices de vegetação Normalized Difference Vegetation Index (NDVI) e Enhanced Vegetation Index (EVI) ao longo do crescimento e desenvolvimento das plantas de soja em três safras no período de 2015 a 2018 NOV_1 (01/11 a 16/11/2017), NOV_2 (17/11 a 02/12/2017), DEZ_1 (02/12 a 18/12/2017), DEZ_2 (19/12/2017 a 03/01/2018), JAN_1 (04/01 a 19/01/2018), JAN_2 (20/01 a 04/02/2018), FEV_1 (05/02 a 20/02/2018), FEV_2 (21/02 a 08/03/2018), MAR_1 (09/03 a 24/03/2018) e MAR_2 (25/03 a 09/04/2018). Perfil temporal dos índices de vegetação Normalized Difference Vegetation Index (NDVI) e Enhanced Vegetation Index (EVI) ao longo do crescimento e desenvolvimento das plantas de soja em três safras no período de 2015 a 2018 NOV_1 (01/11 a 16/11/2017), NOV_2 (17/11 a 02/12/2017), DEZ_1 (02/12 a 18/12/2017), DEZ_2 (19/12/2017 a 03/01/2018), JAN_1 (04/01 a 19/01/2018), JAN_2 (20/01 a 04/02/2018), FEV_1 (05/02 a 20/02/2018), FEV_2 (21/02 a 08/03/2018), MAR_1 (09/03 a 24/03/2018) e MAR_2 (25/03 a 09/04/2018). Coeficiente de correlação de Pearson entre a Evapotranspiração da cultura (ETc) e os Índices de vegetação (IVs) Normalized Difference Vegetation Index (NDVI) e Enhanced Vegetation Index (EVI) de todas as safras em conjunto (2015-2016, 2016-2017, 2017-2018) para o subperíodo vegetativo, subperíodo reprodutivo e ciclo completo da cultura.

17098 — Adsorção de fósforo em solos de argila de atividade baixa — Neste estudo foram avaliadas possíveis correlações entre atributos físicos e químicos do solo e a capacidade máxima de adsorção de fósforo (CMAP) em solos com argila de atividade baixa de diferentes regiões do Brasil, sendo utilizadas amostras de horizontes superficiais (A) e subsuperficiais (B) de 16 perfis de solos. Estudos têm relacionado a capacidade de adsorção de fósforo dos solos com o teor de argila e superfície específica (L EAL e V ELLOSO, 1973; JUO e F OX, 1977), porém segundo BAHIA FILHO (1982) e KER (1995), pouca ênfase tem sido dada ao conhecimento da natureza da fração argila e da composição mineralógica na capacidade de adsorção de fósforo. Dentre os horizontes subsuperficiais, os perfis PAd1 e LBd1 foram os que apresentaram maiores valores de CMAP (1.667 mg.kg⁻¹) quando comparados aos demais, mostrando alta CMAP em solos originários de rochas básicas ou alcalinas; no perfil PAd3, de textura média, verificou-se o menor valor de CMAP (417 mg.kg⁻¹), mostrando que solos com menores teores de argila apresentam menores CMAP. Os coeficientes de correlação de alguns atributos dos solos com a CMAP (Quadro 3) foram positivos e significativos a 1% ou a 5% de probabilidade no horizonte superficial para o teor de carbono orgânico, superfície específica e teor de argila, sugerindo que para esses horizontes, tais atributos contribuem com elevação da capacidade máxima de adsorção de fósforo, sendo este comportamento também verificado em outros trabalhos (S ANYAL e D E DATTA, 1991; CAMPELLO et al., 1994; Nos horizontes superficiais, a análise de regressão linear múltipla para a CMAP como variável dependente e as variáveis independentes, selecionadas pelo método “backward stepwise”, carbono, argila, Fes, Fed,

Se, k_i e k_r , apresentou ótimo ajuste ao modelo, representado pelo elevado coeficiente de determinação ($r^2=0,9995$) (Figura 1a), mostrando a pequena dispersão dos resultados da CMAP encontrados no laboratório pelo ajuste da isoterma de Langmuir e os resultados estimados pela regressão múltipla da equação 1, apresentada abaixo:

1016654 — **Qualidade do milho doce minimamente processado conservado sob diferentes atmosferas** — Este trabalho teve por objetivo avaliar a influência de três atmosferas diferentes (2% O₂ + 8% CO₂, 4% O₂ + 8% CO₂ e atmosfera ambiente) na qualidade de milho verde do tipo doce minimamente processado, durante sua conservação a 5 °C. Observou-se o aumento da perda de massa do milho doce ao longo da conservação em todas as atmosferas, tendo o controle apresentando valores de perda de massa significativamente maiores em relação às atmosferas controladas 1 e 2 (Figura 1). Para o milho doce minimamente processado Embrapa HT1 doce, a firmeza decresceu, durante o período de armazenamento, com pequena elevação do sexto para o nono dia, independentemente da atmosfera de conservação, conforme pode ser observado na Figura 3. O aumento nos valores da firmeza nos grão de milho doce minimamente processado, Embrapa HT1 doce, a partir do sexto dia de conservação pode ser explicado pela maior resistência ao rompimento das células, que pode ter ocorrido devido a um “emborrachamento” que ocorre nas células do grão de milho em virtude da perda de água, que causa dificuldade no rompimento do tegumento pela sonda do texturômetro. de frutose, glicose e sacarose no milho doce foram maiores no armazenamento com atmosfera controlada (2% O₂ com 0%, 15% e 25% CO₂) quando comparadas às concentrações desses açúcares após 1 semana em atmosfera ambiente.

89545 — **Transgênicos provocam novo quadro regulatório e novas formas de coordenação do sistema agroalimentar** — O alimento GM ou ingrediente alimentar não é substancialmente equivalente, porque diferenças podem ser claramente definidas, ou porque o análogo convencional não existe (Nutti & Watanabe, 2002).^{11 11} É importante ressaltar os potenciais riscos da produção e consumo dos alimentos transgênicos apontados pelos pesquisadores, que vão muito além dos aspectos da segurança nutricional dos alimentos estrito senso: a) a tecnologia terminator que permite a introdução de genes capazes de tornar estéreis uma segunda geração de sementes; b) a tecnologia traidor, que consiste em alterar geneticamente uma planta para que a expressão de determinadas proteínas esteja condicionada à aplicação de uma substância capaz de ativar ou desativar características específicas da planta; c) a eliminação de insetos e microorganismos do ecossistema, em virtude da exposição da planta a substâncias tóxicas; d) a contaminação de culturas convencionais; a transferência horizontal de genes, ou seja, entre espécies que não se relacionam na natureza; Desse modo, a comercialização de alimentos rotulados poderá permitir a escolha entre três categorias de alimentos: novos alimentos geneticamente modificados, alimentos não geneticamente modificados convencionais ou orgânicos e alimentos tradicionais com alteração genética (Comission of the European Communities, 2001). Além disso, reforçar a garantia do direito de informação sobre segurança e qualidade dos alimentos aos consumidores, reduzindo a assimetria de informação pelo fornecimento de informações adicionais sobre qualidade e sanidade dos produtos, de tal modo que o consumidor possa escolher os produtos a Cadernos de Ciência & Tecnologia, Brasília, v. 20, n. 2, p. 263-303, maio/ago. Alimentar aves com rações animais produzidas nas instalações da empresa; garantir o controle dos ingredientes utilizados; utilizar soja importada exclusivamente do Brasil na produção de rações; realizar testes adicionais de detecção de contaminações por transgênicos, assegurando que seus fornecedores de aves utilizem rações não transgênicas; implantar processo de rastreabilidade de soja desde a fábrica de extração de óleo no Brasil Exigir de seus fornecedores o uso exclusivo de ração animal e farelo de soja isentos de transgênicos; adquirir rações animais de zonas livres do cultivo de variedades geneticamente modificadas Constituir programa de eliminação de ingredientes GM dos produtos alimentares brancos; solicitar aos fornecedores de carne fresca que retirem a soja e o milho GM das rações para animais utilizadas na cadeia de fornecimento de carne fresca Produzir toda a sua ração para porcos e aves sem soja geneticamente modificada até junho de 2001; oferecer ração livre de organismos geneticamente modificados para aves e porcos, com a garantia de um rígido esquema de segurança e preservação de identidade; garantir a preservação de identidade da soja não transgênica

280844 — **Seleção de antagonistas para o controle de *Cylindrocladium spathulatum* em erva-mate** — Os isolados de bactérias e de Trichoderma que apresentaram maior percentual de inibição, em teste de produção de antibióticos e no teste de hiperparasitismo, respectivamente, foram selecionados para serem empregados nos testes subseqüentes com papel celofane, placas sobrepostas, inibição de germinação de esporos e, antagonismo em folhas destacadas e em mudas. Vários estudos vêm sendo feitos sobre a seleção e o uso de antagonistas para o controle de doenças em culturas agrícolas (Bettiol & Kimati, 1989), destacandose as espécies de bactérias do gênero Bacillus e de fungos do gênero Trichoderma (Bettiol, 1997; Melo, 1991). O isolamento dos antagonistas foi feito em meio BDA, utilizando-se diferentes procedimentos: (a) peneiramento de solo infestado, proveniente de áreas plantadas com erva-mate e plaqueamento das partículas de matéria orgânica em BDA; (b) lavagem de folhas de erva-mate nativa com água esterilizada e plaqueamento da suspensão em BDA e, (c) purificação de colônias de microrganismos desenvolvidas em isolamento de lesões de pinta-preta, que apresentaram halo de inibição em relação a *C. spathulatum*. (1982), que estabelece o grau de antagonismo, por meio da divisão em cinco classes de notas: nota 1 - antagonista cobrindo a totalidade da superfície da placa, nota 2 antagonista cobrindo ao menos 2/3 da superfície, nota 3 - antagonista cobrindo ao menos 50% da superfície, nota 4 - patógeno cobrindo ao menos 2/3 da superfície e nota 5 - patógeno cobrindo a totalidade da superfície, anulando o antagonista. A inoculação do patógeno foi feita com uma suspensão de conídios, obtida por meio da lavagem com água esterilizada de colônias cultivadas em placas de Petri com BDA, durante trinta dias, em condições de laboratório.

1122642 — **Conservação de água de coco verde por filtração com membrana** — 3 relatam que o tratamento térmico para inativação das enzimas polifenoloxidase e peroxidase, na água de coco verde, só é eficiente acima de 90°C, mas a partir de 90 segundos de exposição do produto a esta temperatura, já ocorrem problemas sensoriais relacionados às mudanças no aroma e sabor da água de coco. A vazão de circulação da alimentação também foi mantida constante e igual a 480L/h. Foram coletadas amostras da água de coco verde que alimentava o sistema de membranas (alimentação) e, ao final do processo, amostras da água permeada. A Figura 3 apresenta os resultados referentes ao comportamento do fluxo permeado de água de coco verde durante os processos de filtração nas membranas com tamanho de poro equivalente a 0,1µm e cut off de 100, 50 e 20kDa. sendo A, o valor do parâmetro (por exemplo pH ou teor de proteína) na água de coco verde que era alimentada ao sistema e P o valor deste mesmo parâmetro na água de coco permeada. Fluxo permeado médio de água de coco verde durante os processos de filtração nas membranas com tamanho de poro equivalente a 0,1µm, e cut off de 100, 50 e 20kDa

204144 — **Resposta diferencial de linhagens de feijoeiro ao nitrogênio** — Visando à identificação de linhagens de feijão que sejam tolerantes ao estresse de nitrogênio (N) e que também apresentem resposta positiva a esse nutriente, foram avaliadas 100 linhagens em presença e ausência de N em cobertura, em quatro ambientes. Desse modo, pode-se inferir que 77% das linhagens não responderam à adubação nitrogenada em cobertura e que 23% tiveram resposta, evidenciando a possibilidade de se identificar linhagens tolerantes à ausência de N em cobertura e também de se identificar linhagens que respondam ao nutriente aplicado. O desempenho médio das 22 linhagens que responderam positivamente ao N em cobertura foi de 29% acima do obtido sem N. Observa-se que entre essas linhagens ocorreu ampla variação na produtividade média (Tabela 3). Entre as linhagens com resposta positiva, o índice de eficiência proposto por THUNG (1990) variou de 11,3 a 18,3, ou seja, o mínimo de resposta foi 11,3kg de grãos por kg de N aplicado em cobertura (CI-257) a 18,3kg de grãos por kg de N (RC-I-3). A produtividade de grãos com N em cobertura foi 12% superior à obtida sem N. Tabela 3 - Produtividade média das linhagens responsivas à aplicação de nitrogênio, na presença e na ausência do nutriente, com seus respectivos índices de eficiência de utilização de nitrogênio (α).

1118211 — **O Brasil rural - do passado agrário ao sistema agroalimentar global (1968-2018)** — Era uma base estrutural que refletia uma história de concentração extrema da propriedade da terra e da riqueza em geral, somada a outros processos históricos que deixaram um grave legado negativo, como a escravidão (quatrocentos anos) que vigorou até o final do século 19.5 Em síntese, um passado

agrário ancorado em dois eixos principais: abundância de terras nas mãos de poucos e uma ampla oferta de trabalho, mas sem direitos em quase nenhuma esfera da vida social. São rápidas ilustrações, econômicas e culturais, do Brasil agrário do passado, as quais poderiam ser repetidas com diversos outros exemplos, inclusive em outros âmbitos sociais, e são citadas aqui apenas para sugerir que esta foi literatura contextualizada em ambientes rurais – a COLÓQUIO – Revista do Desenvolvimento Regional - Faccat - Taquara/RS - v. 17, n. 1, jan./mar. As mais evidentes manifestações desse processo são: um decrescente grau de autossuficiência na produção alimentar e as necessidades dos domicílios rurais, um declínio do trabalho agrícola em relação ao trabalho não agrícola no total de gastos nacionais com trabalho, um decréscimo da produção agrícola per capita na economia nacional em relação à produção não agrícola e também uma diminuição da proporção populacional residente nas áreas rurais” (BRYCESON; JAMAL, 1997, p. 5). É por esta razão que usualmente a expressão derivada, “desenvolvimento agrário”, se refere às interpretações sobre as mudanças na vida rural de algum país ou região na qual um desses dois fatores – trabalho – existiu em magnitudes expressivas, congregando massas rurais pobres, e o outro – terra – usualmente estava nas mãos de uma minoria de proprietários. Nos anos setenta a desigualdade social naquelas regiões já era espantosa, mantendo de um lado uma das maiores concentrações da propriedade da terra do mundo e, do outro lado, uma imensa massa de famílias rurais empobrecidas – pequenos produtores e trabalhadores rurais sem terra. Nesse período à frente sugerido de duas décadas, as regiões rurais brasileiras deverão experimentar um processo similar ao verificado no caso norte-americano, no qual o número de estabelecimentos rurais desabou pela metade, entre 1950 e 1970, e deverá ser observado um dramático processo de redução populacional de diversas regiões rurais.

13919 — Índices indicadores do risco ambiental para compostos orgânicos não-iônicos: o modelo Rachel — Neste modelo são calculados alguns índices clássicos, usados para avaliar o comportamento ambiental de compostos orgânicos no solo e em plantas, os quais indicam o potencial apresentado por estes tipos de compostos no que diz respeito à lixiviação, a persistência no solo, a volatilização a partir da superfície do solo, a bioconcentração na solução do xilema, a bioconcentração na raiz, a volatilização a partir da superfície das folhas, o coeficiente de lavagem foliar, a concentração provável na água subterrânea e o tempo de deslocamento do composto da superfície do solo até a água subterrânea. $30 (Koc)$, o coeficiente de sorção do composto no solo (K_d), o fator de bioconcentração do composto na solução do xilema (TSCF), o fator de bioconcentração do composto na raiz (RCF), o coeficiente de partição do composto entre o ar e a folha (K_{la}), a meia-vida do composto por volatilização da superfície do solo ($t_{sv}/2$), a meia-vida do composto por volatilização da superfície das folhas (t_{svm}), o coeficiente de lavagem foliar e o coeficiente de partição do composto entre o ar e a água (K_m). A equação de Arrhenius (KEEN & SPAIN, 1992), a equação de van't Hoff (PADILLA et al., 1988) e a equação de Clausius-Clapeyron (SAGEBIEL et al., 1992) são usadas, respectivamente, pelo modelo Rachel para estimar o efeito da temperatura diária média do perfil do solo na meia-vida do composto no solo, no coeficiente de partição do composto entre o ar e a água do solo e no coeficiente de sorção do composto no solo. Os dados de entrada, com respeito ao solo, requeridos pelo modelo Rachel para calcular os índices indicadores e simular a lixiviação, a meia-vida do composto no solo dependente da temperatura do solo, o tempo de deslocamento do composto no perfil do solo e o coeficiente de sorção do composto no solo é afetado pela temperatura do perfil do solo são: a área superficial específica média do perfil do solo, a recarga diária líquida, a profundidade da água subterrânea, a temperaturas anual mínima e a máxima da superfície do solo, a densidade total do solo, a porosidade do solo e os conteúdos volumétricos médios de água na capacidade de campo, de argila e de carbono orgânico de cada camada do perfil do solo. Para a superfície do solo e a camada mais profunda de 2 metros introduzida, o usuário pode desejar que o modelo Rachele mostre os gráficos da temperatura do solo, do fator de retardo, do fator de atenuação, da taxa de degradação do composto no solo, do coeficiente de partição do composto entre o ar e a água do coeficiente de sorção do composto no solo, onde todos variam com a temperatura do solo.

961303 — Influência da posição de semeadura na germinação, vigor e crescimento de plântulas de bacabinha (*Oenocarpus mapora* Karsten - *Arecaceae*) — Os resultados indicaram que a germinação de sementes de bacabinha não foi afetada pela posição de semeadura, alcançando valores acima de 92%, porém, quando as sementes foram postas para germinar com o poro germinativo voltado para a superfície e com a rafe na posição horizontal e voltada para cima, as

sementes germinaram em menor tempo. The results showed that the germination of *Oenocarpus mapora*, seeds was not affected by the sowing position, showing by values above to 92%, however, the when seeds were placed to germinate with the germination pore gone back to the surface, they germinated in smaller time. (1999) observaram o efeito da posição da semente no substrato e no crescimento de plântulas de *Euterpe espirotusantensis* Fernandes e verificaram que a germinação é mais rápida quando a semente é feita com o poro de germinação voltado para cima. Foram estabelecidas quatro posições de sementeira do endocarpo em relação ao plano horizontal a superfície do substrato: PS (poro na superfície), PB (poro para baixo), RHC (rafe na horizontal e para cima), RHB (rafe na horizontal e para baixo), esquematizado na Figura 1. As três posições de sementeira testadas (PS, RHC e RHB) não diferiram entre si, podendo-se observar que quando as sementes foram postas para germinar com a rafe na horizontal para cima (RHC), 100% das sementes germinaram.

314935 — **Seleção simultânea para porte reduzido e alta produção de látex em seringueira** — A partir desses caracteres, foram criadas as variáveis relacionais PBA - produção de borracha por unidade de altura, calculada através da razão entre os valores de produção de borracha seca (g/m) por m, proveniente dos valores de altura total das progênes e, PBB produção de borracha por unidade de área basal da planta, obtida por meio da razão entre os valores de produção de borracha seca (g cm⁻²), proveniente dos valores do perímetro do caule das progênes. Os resultados referentes às estimativas dos parâmetros genéticos para os caracteres altura (A), perímetro do caule (C), produção de borracha seca (PB), produção de borracha por unidade de altura (PBA) e produção de borracha por unidade de área basal da planta (PBB) estão relacionados na tabela 1. Estimativas de parâmetros genéticos para os caracteres altura (m), perímetro (cm), produção de borracha (g), produção de borracha por unidade de altura (g/m) e produção de borracha por unidade de área basal da planta (g cm⁻²), média geral e erro-padrão da seringueira, aos três anos de idade, em Votuporanga (SP) Estimativas (1) Valores genéticos (VG) para produção de borracha por unidade de área basal da planta (PBB), bem como os seus valores genéticos para os demais caracteres e, ganho genético direto com a seleção para PBB e os ganhos indiretos nos demais caracteres, com seleção das 5 melhores progênes para PBB, aos três anos de idade, em Votuporanga (SP) A combinação dos caracteres produção de borracha x altura e produção de borracha x perímetro aumentou a variabilidade genética disponível para seleção e pode propiciar um eficiente melhoramento genético simultâneo para o aumento da produção de borracha e redução do porte da planta.

1129147 — **Avaliação do processo de compostagem a partir da mistura de capim-elefante, gliricídia e farelo de trigo** — O composto elaborado com 50% de capim elefante + 10% de palhada de gliricídia + 40% de farelo de trigo apresentou menor perda proporcional de N após 90 dias de incubação, em relação aos compostos formulados com menores proporções de farelo de trigo. T1 – 50% de capim elefante + 50% de palhada de gliricídia, T2 – 50% de capim elefante + 40% de palhada de gliricídia + 10% de farelo de trigo, T3 – 50% de capim elefante + 25% de palhada de gliricídia + Somente o tratamento com maior proporção de farelo de trigo apresentou valores de pH inferiores aos demais tratamentos nos períodos finais da incubação, assim como a condutividade elétrica, cujo tratamento com maior proporção de farelo de trigo apresentou os maiores valores ao final da compostagem. A Figura 2 apresenta os valores de proporção relativa em relação ao conteúdo inicial de N. Observa-se que o tratamento com maior proporção de farelo de trigo apresentou as menores reduções, com perda menor que 40% do N ao longo de 90 dias de compostagem, enquanto que os demais tratamentos apresentaram perdas próximas a 60%. O composto elaborado com 50% de capim elefante + 10% de palhada de gliricídia + 40% de farelo de trigo apresentou menor perda proporcional de N após 90 dias de incubação, em relação aos compostos formulados com menores proporções de farelo de trigo.

125759 — **Silicato de cálcio como amenizante da toxidez de metais pesados em mudas de eucalipto** — solo não contaminado (0% solo contaminado), contaminação moderada (25% solo contaminado), contaminação intermediária (50% solo contaminado) e contaminação severa (100% solo contaminado). Nos solos não contaminados e com contaminação moderada, a adição de silicato não proporcionou efeito significativo na altura de plantas, enquanto que nos solos com contaminação intermediária e severa, houve um incremento na altura das plantas com aumento das doses de silicato

de cálcio aplicadas. Apesar disso, a adição de 6,4 g kg⁻¹ de silicato promoveu aumento de 82% na altura das plantas, sob contaminação intermediária, e de até 720% na produção de matéria seca da parte aérea, em solo com contaminação severa, evidência de que o silicato de cálcio beneficia o crescimento do eucalipto mesmo sob elevada contaminação do solo por metais pesados. Sintomas de toxidez de metais em *Eucalyptus camaldulensis* aos 15, 21 e 50 dias após o transplante, submetido a diferentes graus de contaminação do solo com metais pesados e a diferentes doses de silicato de cálcio. Fator transferência de zinco do solo para a parte aérea de *Eucalyptus camaldulensis* submetido a diferentes graus de contaminação com metais pesados e a diferentes doses de silicato de cálcio no solo.

280598 — Contribuição de fatores climáticos na ocorrência da seca de ponteiros de *Eucalyptus grandis* em Arapoti - PR — Um estudo foi desenvolvido para verificar o efeito de fatores climáticos sobre a ocorrência da seca de ponteiros do eucalipto da região de Arapoti, PR, Brasil (SPEA), em *Eucalyptus grandis*. O eucalipto, sob déficit hídrico, pode apresentar sintomas como: (1) necrose ou coloração marrom arroxeadada ou verde pálida, com formato de V invertido, nas folhas (sintoma marcador); (2) pequenos cancos, geralmente agrupados, que exsudam gomas nos internódios dos ramos e da haste principal; (3) fendas e rugosidades na haste principal e rugosidades nas inserções dos galhos e; (4) necroses irregulares e presença de tripes nas folhas mais velhas (Ferreira, 1989). O eucalipto submetido ao excesso de água, decorrente da alta pluviosidade, pode apresentar anomalias como: (1) necrose foliar em V invertido; (2) seca de ponteiros nos galhos e na haste principal; (3) brotações adventícias ao longo do fuste e dos galhos; (4) lançamento de folhas com limbo afilado, deformado e com sintomas de deficiência nutricional, em árvores adultas; (5) obstrução com tiloses em vasos do xilema da haste, do lenho e das raízes; (6) acúmulo de etileno ou respectivos precursores em raízes; (7) deslocamento de precursores de etileno das raízes para a parte aérea; Foram usados dados de precipitação pluviométrica média anual, número de dias com precipitação, temperaturas absolutas mensais (mínimas e máximas), temperaturas médias anuais e amplitude térmica média mensal, no período de 1967 a 1992, para estabelecer um padrão de normalidade, para estas variáveis climáticas. Com base nos padrões climáticos da região de origem de *E. grandis* (Boland et al., 1984), pode-se considerar as seguintes médias como adequadas ao *E. grandis*: (a) temperatura média anual de 19,9 °C; (b) precipitação média anual de 1.787,79 mm; e (c) temperatura mínima mensal, no inverno de 3 a 8 °C. Entretanto, as ocorrências de geadas severas, as temperaturas médias mensais máximas (24-30°C) e a distribuição irregular das chuvas, na região de Arapoti podem ter ocasionado estresse e predisposição das árvores aos organismos potencialmente agressivos.

214593 — Sementes infectadas: sustentabilidade das lavouras questionada! — Resumo: Assim, caminha-se para uma 9.600 sementes provenientes das parcelas onde foi utilizada semendanda cada vez maior no uso de defensivos agrícolas, não só te contaminada, das quais 2.400 provinham de sementes e plantas para o tratamento das sementes como também para aplicação foliar. Os métodos de irrigação por aspersão são os que apresenNa Embrapa Arroz e Feijão, as duas determinações do númetam maiores desvantagens na produção de sementes, principalro de plantas com sintomas de antracnose foram realizadas aos 30 e mente por proporcionarem condições altamente favoráveis à disse38 DAE (dias após a emergência), após a segunda e terceira pulveminção e ao desenvolvimento de doenças da parte aérea. feijão de alta qualidade sanitária e fisiológica pode ser viabilizada pela produção em várzeas tropicais, com irrigação por subirrigação durante o inverno, juntamente com a utilização das práticas recomendadas para eliminação dos patógenos transmissíveis pela semente, descritas no esquema de produção preconizado pela Embrapa Arroz e Feijão desde 1976. Com o propósito de dirimir de vez as dúvidas relacionadas a essa questão, foi conduzido um outro estudo, no qual foram comparados três lotes de sementes de feijão das cultivares BRS Valente, Jalo Precoce e Carioca, produzidas na várzea da Fazenda Barreira da Cruz, no município de Lagoa da Confusão, TO, com outros três lotes, das mesmas cultivares, produzidas na Embrapa Arroz e Feijão, no município de Santo Antônio de Goiás, GO. • Nas várzeas de Lagoa da Confusão, TO, o uso de sementes de feijão recém-colhidas, que sofreram dessecação em pré-colheita, provocou drástica redução no estande de plantas – das 18 sementes distribuídas por metro resultaram 3-4 plantas emergidas.

1008316 — **Eficiência do fósforo revestido com polímeros na cultura da soja** — O delineamento experimental foi em blocos ao acaso, constituído por duas fontes de fósforo (MAP e MAP revestido com polímeros) em quatro doses (0, 40, 80 e 120 kg ha⁻¹ de P₂O₅) perfazendo um esquema fatorial 2x4, com quatro repetições. The treatments were constituted by: control - without phosphorus, fertilizer Monoammonium Phosphate (MAP) - 40 kg ha⁻¹ P₂O₅; (MAP + Coat) - 40 kg ha⁻¹ P₂O₅; MAP - 80 kg ha⁻¹ P₂O₅; coated MAP - 80 kg ha⁻¹ P₂O₅; MAP - 120 kg ha⁻¹ P₂O₅ and coated MAP - 120 kg ha⁻¹ P₂O₅. The results of the dry mass of roots and shoots of soybean showed response to the use of phosphate fertilizers, however, there is statistical similarity between MAP and Kimcoat P in different levels of phosphorus applied. Do mesmo modo, os resultados de massa de matéria seca da parte aérea foram estatisticamente semelhantes entre MAP e MAP revestido com polímeros nas diferentes doses de fósforo aplicadas, sendo que ambas as fontes apresentaram efeito quadrático em função das doses de fósforo empregadas. Analisando o comportamento médio das fontes de fósforo (Figura 2), observa-se uma superioridade do MAP revestido na produtividade de grãos de soja até a dose tecnicamente recomendada para a cultura (80 kg ha⁻¹ de P₂O₅).

314145 — **Ações de recuperação em área degradada por fogo em Floresta Ombrófila Mista: resultados parciais** — E-mail: augusta@cnpf.embrapa.br, yeda@cnpf.embrapa.br Resumo - O presente trabalho relata as ações iniciais desenvolvidas em um remanescente de Floresta com Araucária, que sofreu perturbação por fogo, no qual se pretende favorecer a recomposição da cobertura florestal por meio do plantio de sementes e mudas de Araucaria angustifolia (pinheiro) e Ocotea porosa (imbuia). Para a araucária, inicialmente, efetuou-se o plantio por sementes, sendo necessário fazer o replantio com mudas devido ao ataque de *Cebus apella nigratus* (macaco-prego) às plântulas recém germinadas. 2007 - Recompôr a cobertura florestal por meio do plantio de sementes e mudas de espécies arbóreas ameaçadas de extinção; - Avaliar a sobrevivência e desenvolvimento de mudas plantadas e os devidos tratamentos culturais; - Acompanhar e analisar o aparecimento de regeneração natural de espécies com importância comercial; - Avaliar a adequabilidade do uso de geotecnologias para o monitoramento das ações. O plantio da imbuia foi efetuado no mês de setembro de 2004, por meio da distribuição de mudas sobre a área (Figura 5), de modo a ocupar as entrelinhas do plantio de araucária e mantendo também, para esta espécie, o espaçamento entre mudas de 5 m x 5 m conforme ilustrado na Figura 6. Foram contabilizados somente 35 indivíduos remanescentes de espécies nativas, sendo, deste total, 45 % de indivíduos de araucária e 55 % de indivíduos de espécies latifoliadas (Figura 7).

312782 — **Propagação de figueira (*Ficus carica* L.) por meio de estacas retiradas durante o período vegetativo** — O uso de estacas semilenhosas de figueira Roxo de Valinhos, sob nebulização intermitente, permitiu a obtenção de percentuais de enraizamento de 100% em estacas com folhas tratadas com ácido indolbutírico a 800 mg L⁻¹ (NUNES, 1981). A constatação de que não há diferença no percentual de enraizamento entre estacas com ou sem folhas, não tratadas com AIB, é importante na medida em que a propagação da figueira por estacas herbáceas sem folhas pode dispensar o uso da nebulização, uma vez que a prevenção do murchamento é especialmente importante nos casos em que são utilizadas estacas com folhas (FACHINELLO et al., 1995). O maior peso da matéria seca das raízes foi obtido com 150 mg L⁻¹ de AIB em estacas provindas de plantas sem frutos, diferindo estatisticamente somente da concentração de 300 mg L⁻¹, enquanto que em estacas de plantas com frutos, o melhor resultado foi obtido na ausência de AIB. Os maiores percentuais de enraizamento, pesos da matéria seca das brotações e das raízes obtidos em estacas de figueira sem folhas, independentemente da concentração de AIB utilizada, podem ser devidos, provavelmente, a um alto conteúdo endógeno de auxina nas estacas, visto que as plantas-matrizes se encontravam em pleno período vegetativo. Os maiores percentuais de enraizamento, pesos da matéria seca das brotações e das raízes obtidos em estacas retiradas de plantas sem frutos apresentam a mesma tendência dos resultados obtidos com estacas sem folhas.

1129435 — **Impactos da correção de viés sobre projeções de mudanças climáticas aplicadas a simulações de rendimento de culturas** — O uso dos dados climáticos dos modelos de clima sem similares aos com os dados originais do RegCM4; correção de viés para aplicação em simulações de produção. Através da correção dos dados do modelo climáticos foi tividade não é indicado, desvios em

relação às simulações possível simular valores médios do rendimento, com o mocom os dados interpolados por Xavier et al. RegCM áticos gerados com os dados: (a) anomalia climática sobreposta dados pelo históricos (“Método 375 aos correção método quantil (REGQ) e (e) RegCM4 com correção pelo método quantil aplicad a”), (b) RegCM4 original (REGO), (c) RegCM4 correção do viés médio (REGB), (d) Percentual de mudanças (desvio da média futura em relação à média do per nte) da precipitação acumulada integrada ao longo do 386 ciclo dapresente) soja simulado (mm)solar em função da radiação global média integrada ao longo do ciclo da soja simulado em função enários climáticos gerados com os dados: (a) anomalia climática sobreposta aos dados históricos 387 cenários climáticos gerados com os dados: (a) anomalia climática sobreposta aos dados histó do Delta”), (b) RegCM4 original (REGO), (c) RegCM4 do viés médio(b) (REGB), (d)original (REGO), (c) RegCM4 correção do viés médio (REGB 388 correção (“Método Delta”), RegCM4 M4 correção pelo método quantil (REGQ) e (e) RegCM4 correção pelo método quantilquantil (REGQ) e (e) RegCM4 com correção pelo método qu 389 com RegCM4 correção pelo método do mensalmente (REGM). Percentual de mudanças (desvio da média futura em relação à média do período 401 presente) do desvio padrão da produtividade atingível (potencial limitada pela deficiência h e) da produtividade média atingível (potencial limitada pela deficiência hídrica) 402 simuladas pelo modelo simuladas CROPGRO-Soybean em função dos cenários climáticos gerados c odelo CROPGRO-Soybean em função dos cenários climáticos gerados com os dados: (a) 403 dados: (a) anomalia climática sobreposta aos dados históricos (“Método Delta”), (b) RegCM4 o ia climática sobreposta aos dados históricos (“Método Delta”), (b) RegCM4 original (REGO), 404 (REGO), (c) RegCM4 correção do viés médio (REGB), (d) RegCM4 correção pelo método gCM4 correção do viés médio (REGB), (d)

119512 — **Agricultura ou pastagem? Papel das coberturas pedológicas na diferenciação e na transformação de sistemas agrários pioneiros no Brasil** — Esse rio drena os solos ferralíticos formados sobre o basalto e sobre o arenito que se sobrepõe ao basalto (Fig. 2b), sendo que o arenito predomina na região norte da zona de estudo, a jusante da bacia, ao passo que o basalto predomina na região sul, a montante. Início do século 20: concessão do Estado e companhias de colonização 1930: Floresta intacta, mas ocupada; numerosos títulos de propriedades - Imigrantes de São Paulo - Madeira - Imigrantes europeus - Agricultura de subsistência, porcos, e implantação de café 1945: Zona inteiramente desmatada; o café entra em produção 80% da população é rural - Café - Agricultura de subsistência - Muita mão-de-obra 1963: - Inicia-se a compactação dos solos sobre o arenito e sobre o basalto Erosão e compactação generalizada dos solos Erosão e compactação dos solos Agricultura ou pastagem? Concluindo, a região da Bacia do Ribeirão Bandeirantes do Norte, próxima a Londrina, PR, sob clima subtropical úmido, é o meio que, na escala regional, orienta os seguintes sistemas de produção: a) a atividade de criação extensiva de animais tem sido reservada aos solos mais frágeis e mais pobres, assim teoricamente protegida pela cobertura vegetal das pastagens; b) a agricultura intensiva tem sido destinada aos solos mais ricos e mais bem estruturados, com o atual desafio de cuidar da gestão durável de sua fertilidade. Estudo na escala dos lotes dos agricultores Sobre uma amostra de cerca de 20 famílias (dez por comunidade), estudou-se, para cada família, o estado de ocupação dos solos de seu lote, a história da exploração, as relações existentes ou não entre a diversidade dos Cadernos de Ciência & Tecnologia, Brasília, v. 22, n. 1, p. 185-206, jan./abr.

133848 — **Geracao de renda no semi-arido do Nordeste brasileiro: o caso do extrativismo vegetal** — No período de dezembro a abril, quando ocorre a colheita do imbuzeiro, as chuvas que caem na região são desfavoráveis para o cultivo das lavouras alimentares, como também para a formação de pastagem para os rebanhos, o que leva a maioria dos pequenos agricultores a dedicar-se ao extrativismo do fruto do imbuzeiro. Este estudo objetivou fazer um acompanhamento do extrativismo do fruto do imbuzeiro em três comunidades de pequenos agricultores, localizadas na região semi-árida do Estado da Bahia, na safra de 1996, procurando identificar a importância que os recursos provenientes desta atividade têm na composição da renda familiar e na ocupação de mão-de-obra dos pequenos agricultores. Na comunidade de Lagoa do Rancho, 28 pessoas participaram do extrativismo do fruto do imbuzeiro, 50% das quais dedicaram, em média, 51 dias à colheita, com uma produção média de 46,90 kg por dia e 2.391,90 kg no período por cada agricultor. Na comunidade de Fazenda Saco, 40 pessoas participaram da colheita do fruto do imbuzeiro, sendo que 27,5% das pessoas dedicaram, em média, 42 dias ao extrativismo, com uma produção média de 48,20 kg por dia e 2.024,40 kg colhidos no período por pessoa, o que proporcionou uma renda média de R\$ 202,44, equivalente a 2,02 salários mínimos ou a US\$ 205,11 dólares, em valores de abril de 1996. Na

comunidade de Lagoa do Meio, 29 pessoas participaram do extrativismo do fruto do imbuzeiro, em 1996, 44,83% das quais dedicaram, em média, 44 dias à colheita, com uma produção média de 41,75 kg por dia e 1.837,00 kg colhidos no período, proporcionando uma renda média de R\$ 183,70, equivalente a 1,84 salários mínimos ou a US\$ 186,12 dólares, em valores de abril de 1996.

121733 — **Análise de risco em sistemas de produção agrícola: uma abordagem heurística** — Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Prob($X < 22,856$) $< 0,006$ Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Preço=R\$ 47,06; Prob($X < 32,052$) = 0,76 Queda de preço de 28,69% relativamente ao observado R\$ 45,58; Produtividade=35,099 sacos/hectare Queda de preço de 34,89% relativamente ao observado Instável. Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) = 0,08 Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Preço=R\$ 24,00; Prob($X < 95,303$) = 0,03 Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) = 0,69 Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(1) Nivelamento τ_1 - Preço observado Nivelamento τ_2 - Preço provável Nivelamento τ_3 - Preço pessimista Menor preço(2)

134513 — **Comportamento inicial de espécies florestais exóticas na região da Mata Atlântica de Sergipe** — O' Ajuda e Umbaúba, em Scrgipc, em abril de 1996, avaliando-se nove espécies, 18 meses após o plantio: mogno-africano (*Khaya ivorensis*), nim (*Azadirachta indica*), contas-de-rosário (*Adenantha pavonina*), *Eucalyptus cloeziana*, *E. citriodora*, *E. tereticornis*, *E. pellita*, *E. camaldulensis* e *E. brassiana*. As condições de tabuleiros costeiros, em solo Podzólico Amarelo Distrófico, foram mais favoráveis ao desenvolvimento das plantas que as de baixada litorânea, com solo de Areia Quartzosa, tendo o *E. camaldulensis*, o *E. citriodora* e o *E. tereticornis* sido as espécies de melhor desempenho e o mogno-africano e as contas-de-rosário, as que apresentaram desenvolvimento mais lento. indicando poucas possibilidades de adaptação na baixada litorânea., 1978) considera 40 espécies com potencialidade para as diversas regiões ecológicas do País e, com base no zoneamento ecológico da Região Nordeste (Golfari & Caser, 1979), para a região bioclimática 2, denominada zona tropical úmida, foram indicadas 15 espécies, sendo as mais promissoras o *Eucalyptus tereticornis*, *E. camaldulensis* e *E. brassiana*. Em tabuleiros costeiros houve comportamento diferenciado, onde os melhores desempenhos foram das espécies *E. tereticornis* e *E. camaldulensis*, embora semelhantes a *E. pellita* e *E. citriodora* e diferindo estatisticamente das espécies *E. brassiana* e *E. cloeziana* além de *Khaya ivorensis* e *Adenantha pavonina*. As espécies *Khaya ivorensis* e *Adenantha pavonina* tiveram desempenhos inferiores, diferindo estatisticamente das demais espécies, tanto em condições de baixada como de tabuleiros costeiros, para altura de plantas e diâmetro do coleto.

1128414 — **Estado nutricional do pessegueiro na principal região produtora de pêssego para indústria no Brasil** — Entretanto, para a maioria das amostras, os teores de potássio encontraram-se na classe “acima do normal”, encontrou-se correlações entre P/Ca, P/Cu, Ca/P, Ca/Mg e Ca/B. É possível constatar a carência do uso de ferramentas para avaliar o estado nutricional dos pomares, principalmente a diagnose foliar, o que levou ao uso inadequado de fertilizantes e conseqüentemente um desbalanço nutricional das plantas. The sample includes 65 peach orchards of the Maciel, Granada, Esmeralda, and Sensação cultivars in the municipalities of Pelotas, Canguçu, and Morro Redondo during the 2018 agricultural cycle. the leaf samples for an analysis of their nitrogen (N), phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), boron (B), copper (Cu), iron (Fe), manganese (Mn), and zinc (Zn) content, the results were presented in frequency distributions divided into classes based on the nutrient requirements indicated for the crop in the region, in addition to the

correlations between the nutrients. O presente trabalho foi realizado no ciclo produtivo de 2018, na região produtora de Pelotas, RS (Pelotas, Canguçu e Morro Redondo) em 65 pomares de pessegueiros das cultivares Maciel, Granada, Sensação e Esmeralda, que segundo MAYER et al. Frequency distribution of macronutrient (A) and micronutrient (B) levels in peach tree leaves in the commercial peach-growing region of Pelotas, RS, divided into classes based on nutrient requirements, adapted from CQFS-RS/SC (2016).

1015548 — **Avaliação fenológica da pereira Triunfo cultivada em clima Semiárido no Nordeste do Brasil na safra de 2012** — Essa baixa produção é atribuída a vários fatores como: indefinição de cultivares e porta-enxertos adaptados às diferentes regiões potencialmente produtoras (MACHADO et al., 2013), resultando em baixa produtividade e qualidade de frutos (FAORO; ORTH, 2010). iii) produção de frutas por planta (kg), medida utilizando-se de balança de precisão (Marca Filizola®, modelo CF15, com 0,5 g de precisão), e iv) produtividade, obtida pela multiplicação da produção de frutos por planta e número total de plantas em um hectare. Triunfo' apresentou ciclo fenológico (Figura 1) de 144 dias (Tabela 1), mostrando-se mais tardia que as cultivares 'Housui' e 'Kousui', que completaram seu ciclo em 128 e 115 dias, respectivamente, quando submetidas às condições de cultivo do semiárido brasileiro (LOPES et al., 2013b). O número de frutos por planta foi de 162, e a produção por planta foi de 26,34 kg, maior que o resultado da pereira 'Carrick' enxertada em 11 marmeleiros e em *Pyrus calleryana* Decaisne, que apresentaram produções que variaram entre 2,58 e 9,05 kg por planta (PASA et al., 2011). Triunfo' em condição do semiárido brasileiro; ii) o ciclo fenológico da pereira 'Triunfo' em condição semiárida tropical é de 144 dias, completando o ciclo no mês de julho, e iii) a pereira '

1123970 — **Estudo químico de chás de folhas de urucuzeiro (*Bixa orellana*L.)** — Destaca-se ainda que os teores de Al não se mostraram ser dependentes da localidade de coleta das amostras, isto é, se eram da Embrapa Amazônia Oriental (ponto de referência), São Miguel do Guamá ou Vigia, conforme teste de ANOVA, seguida de Dunnett, com 95 % de significância, implicando que o fato da planta ter sido cultivada sem qualquer adubação ou tratos especiais não interfere nos teores de Al presentes nos chás de folhas de urucuzeiro (*Bixa orellana* L.), estudando amostras de três plantas medicinais provenientes do mercado do Ver-o-Peso, em Belém do Pará, obtiveram teores médios de Cu em chás de folhas de boldo (*Peumus boldus* Molina) igual a 7,88 mg kg⁻¹, para os chás de folhas de cidreira (*Melissa officinalis* L.) média de 9,57 mg kg⁻¹ e para os chás de folhas de canela (*Cinnamomum Zeylanicum* B.) média de 5,28 mg kg⁻¹. Para o Ca é recomendada uma ingestão diária de 1.000 mg para homens e mulheres (DRI, 1997), todavia nos chás de folhas de urucuzeiro (*Bixa orellana* L.) foram encontrados uma dosagem média numa xícara de chá de apenas 2,47 mg, o que correspondem a 0,25 %, portanto o chá das folhas de urucuzeiro (*Bixa orellana* Os teores de Na, Fe, Cr, Al, Zn e K não se mostraram dependentes da localidade de coleta das amostras, ou seja, se eram originárias da Embrapa Amazônia Oriental (ponto de referência), São Miguel do Guamá ou Vigia, conforme teste de ANOVA, seguida de Dunnett, com 95 % de significância, implicando que o fato da planta ter sido cultivada sem qualquer adubação ou tratos especiais não interfere nos teores de Na, Fe, Cr, Al, Zn e K presentes nos chás de folhas de urucuzeiro (*Bixa orellana* L.). Em termos de Na, Mn, Mg, Cu, Ca, K, Fe e Zn a ingestão de uma xícara de chá das folhas de urucuzeiro (*Bixa orellana* L.) (250 mL) não apresenta nenhum risco a saúde humana, todavia, os teores médio encontrados para Al indicam que uma xícara de chá das folhas de urucuzeiro (*Bixa orellana*

134397 — **Virus no tomateiro** — Estes últimos, particularmente, os geminivírus transmitidos por mosca branca e de ocorrência recente nesta cultura no País, constituem, atualmente, o grupo de doenças de maior importância para o tomateiro, sendo fator limitante à produção. Desde a detecção dos geminivírus nesta região, em 1996, foi observado um rápido aumento na incidência destes vírus em tomateiro, independente da ocorrência de altas ou baixas populações de mosca branca em campo. Os geminivírus constituem o principal problema fitossanitário do tomateiro, considerando a severidade das doenças causadas por estes vírus, a diversidade de espécies destes vírus detectada infectando tomateiro no Brasil, a inexistência de cultivares resistentes aos geminivírus no País, a dificuldade de controle da mosca branca e o grande número de espécies de plantas que esta praga pode infectar. Os danos diretos e/ou indiretos provocados pela mosca branca em tomateiro no Submédio do Vale São Francisco, a redução da área plantada, além da baixa produtividade obtida nos anos seguintes à

detecção destes vírus nesta cultura na região, propiciaram o fechamento de fábricas de processamento. Para o manejo de geminivírus em tomateiro, visando a redução da população de mosca branca e da incidência de plantas infectadas com estes vírus, são recomendadas as seguintes medidas:

1120443 — Sistemas agroflorestais biodiversos: segurança alimentar e bem-estar às famílias agricultoras — Realizou-se um estudo em 2014, envolvendo 25 sistemas agroflorestais (SAFs) biodiversos implantados em Mato Grosso do Sul (55°17' e 55°22' de longitude Oeste e 21°58' e 22°06' de latitude Sul), objetivando caracterizar as espécies vegetais existentes, as práticas de manejo utilizadas nos agroecossistemas e formas de uso das espécies vegetais. Nesse contexto, realizou-se uma pesquisa com o objetivo de identificar iniciativas de SAFs biodiversos implementadas em um assentamento rural situado no município de Dourados, estado de Mato Grosso do Sul, região Centro Oeste do Brasil, visando sistematizar e caracterizar as suas formações, as práticas de manejo adotadas, as espécies vegetais que compõem os sistemas e as principais formas de uso pelos agricultores. De acordo com dados levantados na pesquisa fica evidenciado que as espécies encontradas nos sistemas agroflorestais, 27% apresentam mais de uma forma de uso e 17% possuem três ou mais formas de uso pelos agricultores (Tabela 2), ou seja, quase 50% das espécies que compõem o SAFs tem mais de uma forma de uso. A presença de hortos medicinais nos sistemas agroflorestais permite aos agricultores ter próximo às suas residências espécies que podem ser utilizadas nos cuidados primários à saúde, entendendo e respeitando os conhecimentos tradicionais das famílias construídos ao longo do tempo, favorecendo que ocorra a manutenção dessas espécies nos sistemas. Nossos resultados evidenciam que sistemas agroflorestais biodiversos, do tipo quintais agroflorestais, asseguram a diversificação das atividades produtivas e contribuem estrategicamente à segurança alimentar e nutricional das famílias, reduzem drasticamente a dependência de insumos externos, favorecem o microclima local, proporcionado bem-estar às famílias, além de contribuírem com a restauração e manutenção do equilíbrio biológico no entorno das residências.

151173 — Consumo hídrico da bananeira no Vale do São Francisco estimado pelo método da razão de Bowen — A ETc acumulada entre maio de 1999 (120 dias após o plantio - DAP) até o término da colheita do primeiro ciclo, em abril de 2000 (437 DAP), foi de 1210mm, correspondendo a um valor médio de $3,8 \pm 1,1$ mm/dia; no segundo ciclo, entre abril (438 DAP) e novembro de 2000 (término da colheita aos 658 DAP), o consumo de água foi de 880mm, equivalendo a um valor médio de $4,0 \pm 1,2$ mm/dia. May 1999 (120 days after planting - DAP) and the end of the first harvest in April 2000 (437 DAP) A minimum value of 1.7mm/day occurred in June 1999 (140 DAP) during the vegetative stage of the first growing season, while the maximum (6.3mm/day) occurred in September 2000, during the harvest period of the second growing season (580 DAP). Com a radiação solar global (Rg) obtida na estação agrometeorológica à 500m da área experimental, por registros de actinógrafo bimetalico, e o saldo de radiação (R) e a evapotranspiração da cultura (ET), determinaram-se equações de regressão para estimativa do consumo hídrico da cultura da bananeira em função da radiação solar incidente e de DAP. A evapotranspiração da cultura acumulada entre 120 DAP (maio de 1999) e 437 DAP, no término da colheita do primeiro ciclo (abril de 2000) foi de 1210mm, tendo um valor médio de $3,8 \pm 1,1$ mm/dia.

1129161 — Caracterização da Moringa oleifera Lam e sua utilização na alimentação animal — CARACTERIZAÇÃO DA Moringa oleifera Lam E SUA UTILIZAÇÃO NA ALIMENTAÇÃO ANIMAL (Characterization of Moringa oleifera Lam and its use in animal feed) of Moringa oleifera, and to address the effects of the use of this species as an alternative food in animal nutrition. Os resultados da composição química da gliricídia são semelhantes aos verificados para moringa, as folhas de moringa apresentam cerca de 19,85% de MS e 41,25 de FDN (BAKKE et al. Já a composição do feno de moringa apresenta cerca de 90,12% de MS e 19,47% de FDN (OLIVEIRA et al., 2017), enquanto o feno de gliricídia apresentam um teor de MS de 91,24% e 49,86% de FDN (BAYÃO et al., 2016). Existem diversas pesquisas realizadas em todas as partes do mundo com bovinos, caprinos, suínos e aves, onde os resultados foram positivos na produção após a inclusão da moringa (feno, farinha de folhas, farelo de folhas, silagem de folhas e moringa in natura), considerando a espécie como uma alternativa alimentar para alimentação animal (MENDIETA et al., 2009);

1102669 — **Caracterização fenotípica e diversidade genética de Passiflora spp. baseada em descritores multicategóricos** — MJ-46-02 CPAC MJ-65-01 CPAC MJ-H-44 CPAC MJ-08-05 CPAC MJ-02-16 CPAC MJ-07-04 CPAC MJ-06-05 CPAC MJ-H-44S CPAC MJ-69-01 CPAC MJ-41-01 CPAC MJ-02-23 CPAC MJ-13-05 CPAC CÓDIGO CPAC MJ-02-24 CPAC MJ-06-07 CPAC MJ-50-01 CPAC MJ-02-06 CPAC MJ-H-71 CPAC MJ-53-01 CPAC MJ-50-01 CPAC MJ-10-07 CPAC MJ-81-01 CPAC, 1828 Passiflora edulis Sims, 1818 CÓDIGO CPAC MJ-M-20 CPAC MJ-26-03 CPAC MJ-10-01 CPAC

1017311 — **Efeito da planta matriz, estação do ano e ambiente de cultivo na miniestquia de Pinus radiata** — Três ensaios consecutivos foram feitos: (i) o enraizamento das miniestacas coletadas de três posições nas matrizes (topo, intermediário e basal), (ii) a influência do minijardim (vaso, canaletão e campo) no desenvolvimento de minicepas e (iii) o efeito da estação do ano (inverno, primavera e verão) sobre o desenvolvimento de miniestacas coletadas de minicepas no canaletão, no vaso e em campo. Foram coletadas 324 miniestacas de 10 progênies de Pinus radiata, com 66 FLORESTA, Curitiba, PR, v. 45, n. 1, p. 65 - 74, jan. Para testar essa hipótese, em fevereiro de 2008, as 90 estacas com melhor desenvolvimento no experimento anterior, ou seja, aquelas que se apresentavam saudáveis e cujo comprimento estava situado acima de 10 cm de altura, foram transplantadas para vaso, canaletão e campo em espaçamento de 10 x 20 cm no canaletão, 50 x 50 cm no campo e somente uma estaca por vaso. Nesse estudo, o maior índice de enraizamento das miniestacas retiradas da região mediana da planta matriz deve estar relacionado à maturidade do tecido nessa porção da planta, pois não houve a adição de reguladores de crescimento ao enraizamento das miniestacas. Avaliação da capacidade de produção de brotações em miniestacas de progênies de Pinus radiata desenvolvidas em canaletões, vaso e no campo em diferentes estações do ano Houve diferença significativa entre progênies e interação significativa no desenvolvimento das progênies nos meses de inverno e primavera (Tabela 3).

1123933 — **Destinação diversificada** — produção de trigo com diferentes classes comerciais (melhorador, pão, doméstico, básico e outros usos); com características especiais como trigo branqueador e para produção de biscoitos; uso em modelos destinados à exportação (com prioridade para a Ásia e a África); e emprego na composição de rações para suínos e aves. Por exemplo, BRS Tarumã e BRS Pastoreio, para pastejo e grãos; BRS Reponte, para exportação/ração; BRS Marcante, da Classe Pão; BRS 327, para produção de farinha branqueadora; BRS 374, para produção de biscoitos; e BRS Belajoia, para rentabilidade/menor custo de produção. Alguns países, especialmente da Ásia e da África, demandam trigo com qualidade tecnológica compatível com as possibilidades de produção (com maior segurança) de algumas regiões do Sul do Brasil, em especial do Rio Grande do Sul e parte de Santa Catarina e Paraná. A comparação de sistemas para exportação com o sistema de manejo utilizado pela maior parte dos produtores em cada local proporcionou ganhos de rentabilidade na maioria dos casos (Figura 2) com a produção de trigo compatível com as características desejadas pelos países importadores (mais baseada em porcentagem de proteína do que em força de glúten). É necessário mudar o conceito há muito tempo pensado para o trigo nacional, que tinha foco em garantir a autossuficiência Pires e Caierão demonstram possibilidades de diversificação e agregação de valor do País, para uma produção voltada aos mercados nacional e internacional, que busque contribuir para a viabilidade técnica e econômica das propriedades onde o trigo está inserido, desligando da triticultura brasileira o foco exclusivo do abastecimento interno.

1074242 — **Vulnerabilidade ambiental e conflito no uso da terra no município de Mirassol D'Oeste, Brasil** — A erosão hídrica do solo pode ocasionar diversos problemas de conservação do solo e à água, sendo que os principais efeitos são sentidos nas áreas que possuem usos comerciais como a perda de produtividade agrícola, no tocante as questões ambientais os maiores efeitos são sentidos nos cursos hídricos, onde os processos erosivos irão depositar os sedimentos que foram retirados de outras áreas que possuem relevo movimentado (OLIVEIRA et al., 2011). Revista Brasileira de Cartografia, Rio de Janeiro, No 68/10, p. 1917-1936, Nov/Dez/2016 Vulnerabilidade Ambiental e Conflito no Uso da Terra no Município de Mirassol D'Oeste, Brasil são apresentadas as etapas utilizadas na elaboração dos produtos cartográficos, que possibilitaram a geração de dados e informações que serão apresentados no item resultado e discussão. Dando continuidade à pesquisa foi obtido Revista Brasileira de Cartografia, Rio de Janeiro, No 68/10, p. 1917-1936, Nov/Dez/2016

Vulnerabilidade Ambiental e Conflito no Uso da Terra no Município de Mirassol D'Oeste, Brasil o potencial atual à erosão hídrica, para isto foi feita a compatibilização das classes de susceptibilidade à erosão hídrica com o uso atual. Áreas ocupadas pelas classes de potencial à erosão de Mirassol D'Oeste Classes Descrição do potencial à erosão I Alto potencial: uso atual do solo incompatível com a susceptibilidade à erosão laminar Médio potencial: Revista Brasileira de Cartografia, Rio de Janeiro, No 68/10, p. 1917-1936, Nov/Dez/2016 Vulnerabilidade Ambiental e Conflito no Uso da Terra no Município de Mirassol D'Oeste, Brasil CASTRO, S. S. Análise dos compartimentos morfopedológicos como subsídio ao planejamento do uso do solo em Jataí – GO.

1118707 — **Resistência induzida ao pulgão-do-algodoeiro em cultivares de algodão colorido** — The longest duration of the nymphal period, the lower number of nymphs produced, and the major mortality were observed in the ASM treatment. Dessa forma, hipotetiza-se que a aplicação de silício ou ASM proporciona um aumento nos teores de lenhina e fenóis nas plantas de algodão colorido, induzindo resistência a *A. gossypii*. O delineamento experimental foi fatorial envolvendo três cultivares de algodão (BRS Safira, BRS Verde e BRS Rubi) e dois indutores de resistência (silício e acibenzolar-S-methyl) e a testemunha com 10 repetições por tratamento. Resistência a pulgão em algodoeiro colorido 485 Quadro 1 - Médias (\pm EP1) de períodos ninfal (PN), pré-reprodutivo (PPR), reprodutivo (PR), pós-reprodutivo (PPOR) e longevidade (LG), em dias, número total de ninfas (NTN) e mortalidade ninfal (MN), em porcentagem, de *A. gossypii*, em seções foliares de cultivares de algodão colorido tratadas com silício e ASM ($T=25\pm 2^{\circ}C$, $UR=70\pm 10\%$ e fotofase de 12 horas)² O aumento do teor de lenhina devido à aplicação de silício também foi observado em plantas de batata, visando à indução de resistência ao pulgão *M. persicae* (Gomes et al., 2008).

313801 — **Efeito do sombreamento sobre o teor de clorofila e crescimento inicial do jequitibá-rosa** — Toevaluate the effects of periods of shadeness over the initial growth and concentrations of chlorophyll in *Cariniana legalis*(Mart.) 181 crescimento inicial das mudas de espécies arbóreas têm sido as relações entre a altura da parte aérea, o diâmetro do colo e o peso da matéria seca total (parte aérea e raízes), que servem para predizer o grau de tolerância de diversas espécies ao sombreamento (KOZLOWSKI, 1962; ENGEL, 1989; CHAVES & PAIVA, 2004). As demais avaliações, como matéria seca total (aérea e radicial), área foliar e teores de clorofila a e b, foram realizadas somente aos 180 DAE, nas mesmas 15 plantas utilizadas nas avaliações da altura e diâmetro do colo. Teor de clorofila a, b e total em mudas de Jequitibá-rosa, submetidas a diferentes níveis de luminosidade (RFA), avaliadas aos 180 dias DAE. A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants.

1128737 — **Os "novos" mercados e sua contribuição para a inclusão produtiva dos agricultores familiares no Tocantins** — Los "nuevos" mercados y su contribución a la inclusión productiva de los agricultores familiares en el estado de Tocantins, Brasil Resumen Los "nuevos" mercados son construcciones sociales basadas en especificidades que resultan de los procesos de diferenciación de productos. Se concluye que los mercados institucionales, impulsados por políticas públicas: Programa de Adquisición de Alimentos (PAA) y Programa Nacional de Alimentación Escolar (PNAE), además de las ferias locales, son los principales mercados a los que acceden los agricultores familiares. Em face disso, este artigo analisa a percepção desses agentes sobre a contribuição dos "novos" mercados para a inclusão produtiva dos agricultores familiares, bem como, identifica quais mercados as políticas públicas para a agricultura familiar têm potencializado. E, por último, as conclusões finais apontaram que os mercados institucionais e as feiras locais foram os principais mercados acessados pelos agricultores familiares, além de que os extensionistas rurais são considerados fundamentais para fazer a "ponte" entre esses "novos" mercados e o acesso as políticas públicas. As políticas e os programas que se mostraram mais exitosos na promoção da inclusão produtiva dos agricultores familiares estão mais evidentes na terceira geração de políticas para a agricultura familiar, a qual destaca a construção de mercados para a promoção da segurança alimentar e a sustentabilidade ambiental (GRISA; SCHNEIDER, 2014; SOUSA, 2019b).

134128 — **Aplicacao de CO2 via agua de irrigacao em relacao a produtividade do meloeiro** — A aplicação de CO₂ foi realizada diariamente, três vezes por semana e sem aplicação de CO₂

(testemunha) em três maneiras de condução da cultura: com proteção lateral; solo coberto com plástico e solo sem proteção. A aplicação de CO₂ foi realizada diariamente, três vezes por semana e sem aplicação de CO₂ (testemunha) em três maneiras de condução da cultura (com proteção lateral - quebra vento, solo coberto com plástico - mulch e solo descoberto - solo sem nenhuma proteção), em um delineamento estatístico em blocos ao acaso, com quatro repetições. T1 - proteção lateral com aplicação diária de CO₂; T2 - proteção lateral com aplicação de CO₂ três vezes por semana; T3 - proteção lateral sem aplicação de CO₂; T4 - solo coberto com plástico com aplicação diária de CO₂; T5 - solo coberto com plástico com aplicação de CO₂ três vezes por semana; Comparandose os tratamentos T1 (com aplicação de CO₂ diariamente) e T3 (sem aplicação de CO), verificou-se que a aplicação TABELA 2 - Produtividade total (Pt), produtividade comercial (Pc), produtividade não comercial (Pnc), peso médio de frutos comerciais (Pmf), número total de frutos (Ntf) e número de frutos comerciais (Nfc) em meloeiro cultivado em condições de campo. Em ambiente sem nenhuma proteção e sem aplicação de CO₂ verificou-se menor produtividade, 25,64 t ha⁻¹, sendo o CO₂ mais denso que o ar e a plantas de melão de porte baixo, com folhas que cobrem a superfície do solo evitou o arraste do CO₂ pelo vento, permitindo assim, uma maior interação entre o ambiente e a cultura condicionando maior índice de absorção pelas folhas do meloeiro.

15046 — **Avaliação de bicos de pulverização de agrotóxicos na cultura do algodão** — Como a extração da Rodamina foi realizada com 50 mL de solução extratora, a concentração máxima esperada seria de: 25 µg/50 mL = 0,5 µg/mL ou 500 µg/L. Em função da concentração esperada (500 µg/L) com a deposição total da calda pode ser estabelecida a porcentagem a ser encontrada em cada região da planta. Considerando a amostragem na região apical das plantas, o bico TJ-60 8002 depositou gotas com diâmetro em torno de 452 µm que é aproximadamente 4 vezes maior do que a gota produzida pelo bico TXVK-8 (292 µm) e aproximadamente 8 vezes maior do que a gota produzida pelo bico TXVK-5 (228 µm). 47 pode-se dizer a densidade de deposição (Tabela 4) obtida pelo bico TJ-60 (53 gotas/cm²) foi 2 vezes menor que a do TXVK-8 (122 gotas/cm²) e 4 vezes menor que a do TXVK 5 (195 gotas/cm²). The behavior of the nozzles (TXVK-8, TXVK-4, and TJ-60 8002) were tested in relation to deposition, drop size and density in three parts of the plant (apical, middle, basal). However, the medium deposition of the three nozzles was significantly decreasing from the apical region (45%) to the middle region (18%) and from this to the basal region (7%).

313727 — **Comparação de métodos de estimativa de volume para Pinus oocarpa em diferentes idades e diferente regimes de desbastes** — A análise de variância mostrou pela comparação do valor de F e nível de significância, a não existência de diferenças significativas entre os métodos de estimativa de volume para os 11 regimes de manejo com e sem casca, ao nível de 95% de probabilidade, confirmando a hipótese de que os volumes médios estimados através dos métodos estudados são estatisticamente iguais entre si. A existência de 11 diferentes regimes de manejo, nos quais se variam a idade e o número de desbastes em plantações de Pinus oocarpa da companhia Duraflora, localizada no município de Agudos, em São Paulo, possibilitou a cubagem de 100 árvores por regime de manejo, as quais forneceram conjuntos de dados apropriados para cálculo dos volumes reais, tanto com casca como sem casca para esta espécie. RESULTADOS E DISCUSSÕES Baseado nos resultados desta pesquisa sobre fatores de forma, equações de volume e sobre funções de afilamento, desenvolvidas para a espécie em questão, calculou-se o volume das árvores para cada regime de manejo pelos métodos de avaliação citados. A análise de variância mostrou que não existe diferença significativa entre os métodos de estimativa de volume para os 11 diferentes regimes de manejo, com e sem casca, ao nível de 95% de probabilidade. Com referência ao uso destes métodos, apesar dos fatores de forma serem de mais simples utilização e as equações de volume terem uso consagrado, recomenda-se o uso das funções de afilamento que são flexíveis e permitem a estimativa do diâmetro a uma altura qualquer do fuste, da altura comercial para um dado diâmetro mínimo, e com a integração da função, obter o volume até 96 Bol.

314112 — **Dendrocronologia de espécies da Floresta Ombrófila Mista do Município de Candió, PR** — The trunks of 20 trees were collected in Candió, State of Parana, Brazil, from an area that belongs to ELEJOR, Centrais Elétricas do Rio Jordão, for the following species: Araucaria angustifolia, Clethra scabra, Cedrela lilloi, Ocotea porosa, Podocarpus lambertii and Sebastiania commersoniana (3 to 6, from each species). However, the very extreme growth condition of 1999 and 2000, when

occurred a very low precipitation period followed by a very severe winter, was registered in the growth ring series of most of the disks analyzed. No entanto, em trabalho realizado para caracterizar a regeneração natural, em Floresta Ombrófila Mista na Floresta Nacional de São Francisco de Paula, foi observado que a *Araucaria angustifolia*, apesar de ser a espécie fisionomicamente mais destacada e de maior interesse econômico da FOM, apresentou valores baixos para densidade, frequência e dominância e, conseqüentemente, baixo valor de importância, necessitando a implementação de tratamentos silviculturais e plantios de enriquecimento que beneficiem a sua regeneração (NARVAES et al., 2005). O objetivo desse trabalho é estimar o crescimento passado de espécies de Floresta Ombrófila Mista, nativa do Município de Candói, PR, pela medição dos anéis de crescimento e estabelecer a idade, incremento médio anual e correlacionar o crescimento das espécies *Araucaria angustifolia* se destacou, quando considerado o incremento periódico do DAP dos últimos 10 anos (IPA DAP (10)), apresentando, em média 0,9 cm de incremento ao ano, variando de 0,5 cm a 1,3 cm.

162329 — Resposta da alfaca tipo americana a doses e épocas de aplicação de molibdênio em cultivo de inverno — (2004), verificaram efeitos quadráticos para a matéria fresca comercial, tendo as doses de 236; 212 e 229 g/ha de molibdênio proporcionado os maiores rendimentos comerciais quando o nutriente foi aplicado aos 14, 21 e 28 dias, respectivamente. Os resultados evidenciaram efeitos significativos independentes para as doses de molibdênio e época de Resposta da alfaca tipo americana a doses e épocas de aplicação... aplicação, assim como para a interação entre estes fatores, variando com as características avaliadas. Com o incremento das 145 doses de molibdênio em função das épocas de aplicação (Figura 2), verificaram-se efeitos quadráticos tendo as doses de 77,0; 82,6 e 77,5 g/ha de molibdênio proporcionado as maiores retornos aos 14, 21 e 28 dias após o transplante, respectivamente. Verifica-se pela Figura 2 que as épocas de aplicação apresentaram pontos de máxima produtividade de magnitudes semelhantes com pequenas variações entre doses, onde a aplicação aos 28 dias após o transplante na dose estimada pela equação proporcionou maior matéria fresca com 553,4 g/planta, comparativamente aos 21 dias (516,9 g/planta) e aos 14 dias (540,2 g/planta). Esta característica não apresentou diferenças estatísticas entre as épocas de aplicação que obtiveram 42,7 cm (14 dias), 42,9 cm (21 dias) e 43,8 cm (28 dias) e entre doses de molibdênio que registraram variações de 42,2 a 44,6 cm.

314079 — Estratégia complementar para conservação de espécies florestais nativas: resgate e conservação de ecótipos ameaçados — Do ponto de vista da integridade dos ecossistemas e da viabilidade das populações remanescentes, principalmente das espécies endêmicas, a intervenção Estratégia complementar para conservação de espécies florestais nativas: resgate e conservação de ecótipos ameaçados Na amostragem das matrizes para coleta de sementes, visando à restauração dos ecótipos, como medida preventiva, recomenda-se observar duas distâncias críticas: uma correspondente à distância mínima (a) entre as matrizes para que não haja problema de depressão por endogamia; a outra, correspondente à distância máxima (b) entre as matrizes, para que não haja perda. Estratégia complementar para conservação de espécies florestais nativas: resgate e conservação de ecótipos ameaçados Embora as formações florestais da região possam apresentar uma grande diversidade de espécies, apenas um número limitado delas deverá ser incluído na amostragem e plantio em cada ponto da rede de bancos de conservação pelas seguintes razões: a) o plantio de um grande número de espécies não permite a inclusão do número requerido 23 de indivíduos para formar populações viáveis de cada espécie; b) mesmo com o plantio de somente algumas espécies, o ambiente criado pelo estabelecimento destas proporciona condições favoráveis à nucleação (regeneração natural de diversas outras em seu entorno, ao longo do tempo). Entre os cuidados técnicos a serem observados na instalação e no manejo desses bancos, destacam-se os seguintes aspectos fundamentais: a) produção de mudas, separadamente por matriz, e o seu plantio em um delineamento em blocos ao acaso, contendo parcelas de uma planta, repetidas tantas vezes quantas forem possíveis, limitadas pela disponibilidade de mudas e de área para o plantio; b) plantar as espécies secundárias e as tardias, após as pioneiras, com intervalos variáveis de dois a três anos, ou mais, dependendo dos requisitos das espécies tardias, bem como da velocidade de fechamento do dossel e da longevidade das espécies pioneiras e secundárias iniciais; c) à medida que as plantas forem crescendo, efetuar desbastes periódicos das pioneiras, eliminando-se as plantas de baixo vigor para abrir espaço para as demais, mas, cuidando-se para manter, em alguns blocos, os representantes das progênies eliminadas de outros, de maneira que todas as matrizes amostradas na região continuem representadas nesse povoamento; d) coletar sementes, de preferência, antes

dos desbastes e, posteriormente, sempre que houver possibilidade. É importante seguir os critérios de: 1) distância mínima entre matrizes; 2) distância máxima de 20 km a 30 km do local de plantio e 3) restrição de coleta somente dentro do mesmo estrato de fenofase para que todas as árvores tenham a Estratégia complementar para conservação de espécies florestais nativas: resgate e conservação de ecótipos ameaçados

1123909 — **Precipitação no extremo norte da Amazônia: distribuição espacial no estado de Roraima, Brasil** — Nesses dois últimos casos como ações para prevenir que fogos derivados das queimadas escapem do controle e se transformem em incêndios florestais de grandes proporções em anos de seca severa (BARBOSA; FEARNside, 1999; ALENCAR et al., 2006; ARAGÃO; SHIMABUKURO, 2010; XAUD et al., 2013; BARNI et al., 2015a; A modelagem da distribuição espacial e temporal da precipitação no Estado de Roraima foi realizada a partir da confecção de mapas de precipitação do produto TRMM 3B43 (<http://trmm.gsfc.nasa.gov/>), em Hierarchical Data Format – HDF (formato de arquivo padronizado para armazenamento de dados), com resolução espacial de 0,25 graus (~27,75 km; 1° ~111 km) e de mapas do WORLD CLIM (<http://www.worldclim.org/current>), no formato raster (grade de células), com resolução espacial de 10 minutos ou 0,1666668 graus (~18,5 km). A krigagem ordinária é uma técnica de interpolação geoestatística robusta, fácil de aplicar e facilita a interpretação dos resultados (CAMBARDELLA et al., 1994; BELLO-PINEDA; HERNÁNDEZ-STEFANONI, 2007; BOHLING, 2005; ALVARES et al., 2011; 2014; BARNI et al., 2016). Um semivariograma foi modelado para cada mês em particular, em que os parâmetros observados em cada modelo foram o efeito pepita (erro aleatório, intrínseco à estrutura e distribuição espacial dos dados), alcance (medida que informa até onde o modelo controla as estimativas) e patamar ou sela (limite a partir do qual as estimativas dos valores modelados se tornam aleatórias (e.g. ALVARES et al., 2011; 2014; BARNI et al., 2016). Precipitação no extremo norte da Amazônia Gráfico 3 - Comparação entre os valores anuais observados (Estações) e os valores preditos ou modelados pelo TRMM (TRMM 3B43) e pelo WORLDCLIM (WCLIM), nos 59 postos meteorológicos localizados na área de estudo.

1008415 — **Qualidade de pimentões produzidos em sistema orgânico e armazenados com e sem refrigeração** — choice of cultivar, and this will depend on your performance not only productive, but also qualitative aspects and endurance of the fruit in the postharvest. The aim of this work was to select and assess the quality of 14 peppers cultivars from the collection After the selection of the best cultivars, these were packed in trays, involved with PVC film of 10 microns, being one part stored at room temperature (22 ± 5 o C and relative humidity 55 ± 5) O delineamento experimental foi inteiramente casualizado, em esquema fatorial 14×2 , cujos tratamentos foram as 14 cultivares (I -16, Italiano, Ruby, All big, Fry King, Tico, Casca Dura Ikeda, Margareth, Ambato, Avelar, Magda Super, Lito, Agro Sul Gigante) e dois tipos de armazenamento (ambiente e refrigerado), com três repetições. Fry King Cônico Verde Claro 13,75 a 4,98 d 12 4 Tico Cônico Verde Claro 9,88 c 5,36 d 10 4 Keystone Quadrado Verde 8,65 c 8,17 a 10 4 CD Ikeda Cônico Verde 11,44 b 4,91 d 10

975131 — **Dinâmica do uso e cobertura da terra do município de Floresta, PE** — The objective was to map the dynamic changes of land use and forest cover in the last 21 years (1987 to 2008), the City of Forest-PE. The decrease in forest cover between 1987 and 2008 the city was the main consequence of the interaction of climate and anthropogenic pressure forest resources, need for urban expansion and agricultural Portanto, neste estudo objetivou-se mapear e criar um banco de dados sobre a dinâmica de mudança do uso da terra e da cobertura vegetal nos últimos 21 anos (1987 a 2008), no município de Floresta, sertão de Pernambuco, visando subsidiar a gestão florestal, bem como avaliar suas causas e efeitos. Houve também uma mudança na posição da vegetação densa em relação ao ano de 1987, uma vez que, na obtenção da imagem de 1997, o solo encontrava-se úmido naquela região, pois nesse ano as precipitações ocorreram de uma forma bem mais distribuída ao longo do ano, o que pode ter causado uma migração e conseqüentemente um aumento da vegetação densa em relação ao período anterior (IBGE, 2010). Ainda na tabela 1, para 2008, observa-se que todas as classes de vegetação diminuíram: a vegetação densa perdeu 4.224 ha e a vegetação semidensa foi suprimida em 37.520 ha, o que totalizou uma perda de 11,24% de vegetação em comparação com o período de 1997, passando de 38,66% para 27,38%.

312244 — **Vigor juvenil em progênies de erva-mate (*Ilex paraguariensis* A. St.-Hil.) nativas do Estado de Mato Grosso do Sul** — Valores fenotípicos, genéticos aditivos, ganhos genéticos preditos e nova média da população dos dez melhores indivíduos para o caracter altura (cm) em erva-mate, no município de Ponta Porã, MS. TABLE 2: Phenotypic and additive genetic values, genetic gain and new mean of population of the best ten individuals for the trait height (cm) in erva-mate in Ponta Porã, MS. Valores fenotípicos, genéticos aditivos, ganhos genéticos preditos e nova média da população dos dez melhores indivíduos para o caracter diâmetro (mm) em erva-mate, no município de Ponta Porã, MS. TABLE 3: Phenotypic and additive genetic values, genetic gain and new mean of population of the best ten individuals for the trait diameter (mm) in erva-mate in Ponta Porã, MS. 3 7,82 8,54 27,40 17 3 7,73 8,48 27,34 29 7 7,62 8,43 27,29 22 6 7,58 8,37 27,23 23 0 7,55 8,32 27,18 23 5 7,48 8,28 27,13 24 4 7,47 8,24 27,09 11 5 7,36 8,19 27,05 É importante salientar que há três estratégias que podem ser adotadas para a formação de pomares de sementes, usando-se a metodologia de modelos lineares mistos ou Reml/Blup: a) pomar de sementes testado (1,5 geração), com seleção de genitores; b) pomar de sementes, com seleção de indivíduos no experimento
Ciência Florestal, v. 17, n. 1, jan-mar, 2007 Vigor juvenil em progênies de erva-mate (*Ilex paraguariensis* A. St.-Hil.)

975338 — **Dinâmica do herbicida METRIBUZIN aplicado sobre palha de cana-de-açúcar (*Saccharum officinarum*)** — DO HERBICIDA METRIBUZIN APLICADO SOBRE CANA-DE-AÇÚCAR (*Saccharum officinarum*)1 PALHA DE Performance of Metribuzin Applied on Sugarcane Straw ROSSI, C.V.S.2, VELINI, E.D.3, LUCHINI, L.C.4, NEGRISOLI, E.5, CORREA, M.R.5, PIVETTA, J.P.6, COSTA, A.G.F.7 e SILVA, F.M.L.8 RESUMO - O herbicida metribuzin é recomendado para o controle de plantas daninhas na cultura de cana-de-açúcar. No segundo, avaliou-se a lixiviação do metribuzin em 5, 10, 15 e 20 t de palha por hectare sob simulação de chuva de 2,5, 5, 10, 15, 20, 35, 50 e 100 mm, um dia após a aplicação (DAA). No tocante ao parâmetro "c" do modelo de Mitscherlich, verificou-se que no tratamento de 20 t ha⁻¹ houve aumento na taxa de extração do herbicida da palha de cana-de-açúcar pelo incremento de cada milímetro de chuva simulada, de tal forma que, de acordo com o modelo aplicado, mais de 99% do metribuzin da máxima saída foi extraído quando a lâmina de chuva alcançou 30 mm. (2007), em estudo com metodologia semelhante, avaliando a dinâmica do herbicida amicarbazone em aplicação sobre diferentes quantidades de palha de cana-de-açúcar, demonstraram que 76, 75, 63 e 58% do herbicida aplicado foi lixiviado por uma lâmina de 20 mm de chuva para 5, 10, 15 e Planta Daninha, Viçosa-MG, v. 31, n. 1, p. 223-230, 2013 228 ROSSI, C.V.S. et al. A quantidade de metribuzin extraída com a precipitação de 100 mm aos 0, 1, 7, 14 e 28 DAA e a precipitação de 20 mm aos 7 e 14 dias após esta primeira chuva, nos diferentes tratamentos de períodos de permanência do produto na palha, está demonstrada na Tabela 4.

1074877 — **Efeito de micorrizas e da fertilização no crescimento de mudas de *Campomanesia Xanthocarpa* (mart.) O.berg., produzidas em diferentes substratos** — As mudas adubadas apresentaram maior crescimento nos substratos com 40% e 50% de vermicomposto, enquanto que nas mudas não adubadas, foram os substratos com 30% a 50% de vermicomposto (Figura 3). FIGURA 3: Biomassa seca da parte aérea (g) de mudas de *Campomanesia xanthocarpa* com e sem adubação e biomassa seca radicial (g) de mudas com inoculação micorrízica e com adubação, com inoculação micorrízica e sem adubação, sem inoculação micorrízica e com adubação e sem inoculação micorrízica e sem adubação. Shoot dry biomass (g) of seedlings of *Campomanesia xanthocarpa* with and without fertilization and root dry biomass (g) of seedlings with mycorrhizal inoculation and fertilization, with mycorrhizal inoculation and without fertilization, without mycorrhizal inoculation and with fertilization and without mycorrhizal inoculation and fertilization. Verificou-se que este valor se aplica bem para a espécie em estudo, pois independentemente da inoculação micorrízica, as mudas sem adubação apresentaram valores inferiores aos das mudas adubadas, como também, verificou-se que o substrato comercial e os com 5% e 10% de vermicomposto não se mostraram adequados para produzir mudas de Ci. A inoculação com fungos micorrízicos arbusculares em mudas de *Campomanesia xanthocarpa* não substitui a adubação de cobertura, a qual aumenta o crescimento das mudas, demonstrando ser imprescindível para a obtenção de mudas de *Campomanesia xanthocarpa* com qualidade.

1125800 — **Quilombolas e indígenas: análise dos públicos da agricultura familiar excluídos das políticas públicas em uma região da Amazônia legal** — Os quilombolas e os indígenas foram considerados os públicos da agricultura familiar do Tocantins (e também da região Amazônica) mais preteridos pelas políticas públicas, pois são afetados por diferentes problemas sociais no meio rural, por viverem em territórios com áreas aquém de suas necessidades ou, ainda, não estarem dotados de aparatos jurídico-formais e itinerários técnicos que lhes assegurem a regularização territorial. 409 Revista Humanidades e Inovação v.7, n.16 - 2020 lhes sobre esta situação e os possíveis problemas que podem acarretar aos quilombolas, uma extensionista menciona que... se você fosse fazer esse levantamento e comparar os dados, por exemplo, a Fundação Cultural Palmares, inclusive tem a Coordenação Estadual dos Quilombolas aqui do Tocantins, se você fosse comparar os dados que esta instituição tem da quantidade das comunidades quilombolas, do número de famílias que têm, com o número de DAP, você ia perceber que tem uma diferença muito gritante nesses dois dados por causa dessa identificação que o próprio técnico lá da assistência técnica está fazendo. As experiências de inserção de quilombolas em políticas de Compras Governamentais de alimentos são limitadas, ou seja, o acesso às políticas ocorre mais no campo individual.3 Mesmo não sendo de forma coletiva, esta condição de acesso contribuiu para que houvesse o escoamento da produção, embora seja difícil a comercialização na própria localidade, visto que os quilombolas produzem gêneros alimentícios semelhantes. Esta narrativa trata, provavelmente, de enunciar o quão restrito é o acesso dos quilombolas as políticas públicas para a agricultura familiar, não devido a inexistência destas, já que se constata que as feiras são os mercados mais imbricados com os quilombolas e que não é necessária política pública para participar de tal canal de comercialização. Deste modo, a comercialização dos produtos dos quilombolas torna-se um entrave... por falta de documentação, pois ficam fora das políticas públicas como um todo, da assistência técnica, do crédito, da vigilância sanitária, dos programas de comercialização, como PAA e Pnae, a questão de acesso a editais de infraestrutura para instalar dentro da propriedade, então como um todo, as políticas, elas acabam não sendo acessadas então pelo quilombola (GE03, gestor, 2018).

962648 — **Commodities agrícolas e preço do petróleo** — Variações no preço desta influenciam a produção e o preço das commodities e, também, variações no preço de qualquer uma das commodities poderão influenciar o preço da ureia. Nota-se também que, no mesmo período, apenas 10% de variância do preço do petróleo é explicada pelo preço da ureia, indicando que o preço do petróleo é uma variável exógena. Pode-se observar que o preço da ureia, 24 meses após o choque, explica 24% do erro de previsão do preço do arroz e que, no mesmo período, o arroz explica 20% da variância do preço da ureia. A variância do preço da ureia sofre uma influência maior da variância no preço do milho: ao final de 24 meses, 23% dessa variância é explicada pelo preço do milho. ças do preço do milho sobre o preço da ureia, mas não rejeita a hipótese de ausência de causalidade das primeiras diferenças do preço da ureia sobre o preço do milho.

314117 — **Caracterização de substratos para produção de mudas de espécies florestais elaborados a partir de resíduos orgânicos** — Resultados e Discussão Tendo-se como referência limites estabelecidos para a interpretação de resultados de análise de substratos constituídos de compostos orgânicos e utilizados para produção de mudas em tubetes (VALERI; CORRADINI, 2000), pode-se observar que os substratos formulados com as misturas de bagaço de malte + serragem - S2, lixívia celulósica + serragem (1:4) - S5 e lixívia celulósica + serragem (3:2) misturado com casca de pínus (1:1) - S7 apresentaram baixo pH; Essas situações geralmente ocorrem em condições de elevada acidez do solo, o que não é o caso para o Al nos substratos avaliados, em que o teor de Al trocável foi baixo nos substratos compostos pelas misturas de Plantmax Florestal® + casca de pínus (1:1) - S1, lodo celulósico + serragem (1:1) - S3, lodo celulósico + serragem (1:1) + serragem (1:1) - S4, lixívia celulósica + serragem (1:1) misturado com casca de pínus (1:1) - S7 e esterco de caprino compostado - S8 e médio nos substratos compostos pelas misturas de bagaço de malte + serragem (1:4) - S2, lixívia celulósica + serragem (4:1) - S5 e lixívia celulósica + serragem (3:2) - S6 (TOMÉ JÚNIOR, 1997), e a saturação do Al na CTC efetiva foi baixa em todos os substratos (m = 0 a 15 %), conforme Malavolta (1989), citado por Tomé Júnior (1997, p. 134). Tomando-se o valor 18 para a razão na relação C:N, como referência para a indicação de uma adequada compostagem dos resíduos utilizados como substrato neste estudo, conforme Especificações...(1999), pode se observar que os substratos compostos pelas misturas de Plantmax Florestal® + casca de pínus (1:1) - S1, lodo

celulósico + serragem (1:1) misturado com casca de pínus (1:1) - S4, lixívia celulósica + serragem (4:1) S5 e lixívia celulósica + serragem (3:2) - S6 apresentaram valores da relação C: Embora as considerações do parágrafo anterior sejam válidas, pela dependência de estudos adicionais, os substratos compostos pelas misturas de lixívia celulósica + serragem (4:1) - S5 e lixívia celulósica + serragem (3:2) - S6 não seriam adequados para a produção de mudas de Pinus e Eucalyptus em razão dos baixos teores de P, e o substrato composto pela mistura de lodo celulósico + serragem (1:1) - S3 em função do elevado valor de pH em CaCl₂, pela possibilidade da insolubilização de micronutrientes. Os substratos formulados com a mistura de lixívia celulósica + serragem (4:1) e lixívia celulósica + serragem (3:2) além da restrição na relação C:N, apresentam baixos teores de P; Com relação à quantidade de água disponível a tensões entre -3 kPa e -100 kPa, os melhores substratos são os formulados com as misturas de lixívia celulósica + serragem (4:1); lixívia celulósica + serragem (3:2) e o lixívia celulósica + serragem (3:2) misturado com casca de pínus (1:1) e Os substratos formulados com as misturas de Plantmax Florestal® + casca de pínus (1:1); bagaço de malte + serragem (1:4); lodo celulósico + serragem (1:1) misturado com casca de pínus (1:1) e esterco de caprino compostado podem ser utilizados na produção de mudas de espécies florestais tolerantes à acidez, adotando os protocolos de manejo recomendados, tanto para irrigação quanto para a fertilização.

313739 — Liquidambar para produção de madeira no sul e sudeste do Brasil — O liquidambar é uma espécie de ampla distribuição natural, abrangendo as regiões leste e sudeste dos Estados Unidos, avançando para o sul, pelo México, até a América Central. Sweetgum is a North and Central American tree species well known for the healing gum since the times of the Aztec civilization. Experiments in several sites have demonstrated its potential value for wood production in small wood lots in the South and Southeastern Brazil. Plantios experimentais desta espécie, no Brasil, envolvendo sementes coletadas de diferentes regiões, confirmaram a maior taxa de crescimento e produtividade de madeira nas procedências da América Central, quando comparadas com as do México e dos Estados Unidos. as espécies folhosas do sudeste dos Estados Unidos, o liquidambar é considerado de rápido crescimento e produtora de madeira de alto valor econômico, com múltiplos usos como matéria-prima para celulose, dormentes, madeira serrada para embalagens, estrados, móveis e até como lenha.

1110441 — Solos arenosos para sistemas de integração lavoura-pecuária-floresta em Arez, Rio Grande do Norte — A retenção de água como atributo limitante ao desenvolvimento dos vegetais está sendo seriamente comprometida pelo processo de degradação do solo em terras agrícolas, em função do padrão de manejo do solo, com baixa biodiversidade e alto grau de revolvimento pelas operações constantes de preparo do solo. As amostras de solo com estrutura indeformada foram analisadas no Laboratório de Análises Físicas de Solo do Centro de Ciências Agrárias da Universidade Federal da Paraíba, para a determinação das variáveis: capacidade de aeração do solo, teor de água na capacidade de campo, ponto de murcha permanente, água disponível, capacidade de água disponível, densidade do solo, porosidades (total; macro e microporosidade), condutividade hidráulica saturada, granulometria, argila dispersa em água e grau de floculação. Entre os benefícios que os sistemas de integração promovem ao solo podem-se citar o favorecimento à infiltração e retenção de água, incremento e manutenção da matéria orgânica, desenvolvimento do sistema radicular dos vegetais (PEZARICO et al., 2013), aumento da porosidade total e agregação do solo (LOSS et al. Quanto a granulometria dos perfis avaliados, o perfil PA2-TM apresentou teores consideráveis da fração areia dos 0 aos 50 cm de profundidade, sendo, portanto, classificado como areia; contudo dos 50 aos 180 cm, a classificação textural do perfil mudou de areia para areia franca, em função dos teores de silte e argila encontrados em profundidade, caracterizando um horizonte B textural com caráter fragipânico. O aumento de densidade do solo traz como empecilho o aumento dos níveis de compactação do solo, comprometendo a resistência à penetração de raízes e as taxas de infiltração de água no solo (MELO et al.

1122380 — Características agronômicas de genótipos de sorgo mutantes BMR e normais utilizados para corte e pastejo — RESUMO - O experimento foi conduzido nas dependências da Embrapa Milho e Sorgo, em Sete Lagoas, MG, objetivandose avaliar as características agronômicas de genótipos de sorgo mutantes BMR e normais, sendo 11 mutantes BMR e nove normais. Os sorgos

para corte e pastejo, vêm se destacando como mais um recurso disponível para manter a estabilidade da produção de forragem ao longo do ano, por serem plantas adaptadas à baixa disponibilidade de água, apresentarem elevado rendimento forrageiro, bom valor nutricional, alta capacidade de germinação, grande velocidade de crescimento, boa habilidade para perfilhamento e rebrota; após o corte ou pastejo, capacidade elevada de produção de matéria seca sob condições de estresse ambiental, facilidade e baixo custo de implantação (RODRIGUES, 2000). Foram utilizadas as duas fileiras centrais de cada parcela, referente aos 42 dias de rebrota após o primeiro corte, onde foram avaliadas as seguintes características: altura das plantas: que foi obtida através da medida do nível do solo à extremidade superior da planta, em 20% das plantas de cada parcela; produção de matéria verde: que foi obtida a partir da pesagem de todas as plantas da área útil da parcela, realizada após o corte a 15 cm do solo; produção de matéria seca: que foi obtida a partir da Scientia Agraria Paranaensis – SAP; Valores médios de altura em metros (m), produção de matéria verde em toneladas por hectare (PMV t ha⁻¹) e produção de matéria seca em toneladas por hectare (PMS t ha⁻¹) de vinte genótipos de sorgo para corte e pastejo, mutantes BMR e normais, avaliados no segundo corte em Sete Lagoas, MG, ano agrícola de 2011. Os genótipos normais IS10428xTX2784, IS10252TX2784 e o isogênico normal BR007AxTX2784 são os mais produtivos, pois apresentam maiores valores de altura, produção de matéria verde e produção de matéria seca, em relação aos demais grupos de genótipos avaliados, nas condições de realização deste experimento.

955737 — **Vigor, floração, produção e qualidade de pêssegos 'Chimarrita' e 'Maciel' em função de diferentes porta-enxertos** — 'Maciel' 2008 Aldrighi Capdeboscq Flordaguard Nemaguard Okinawa Tsukuba 1 Umezeiro Seleção Viamão CV (%) 'Chimarrita' 2008 Aldrighi Capdeboscq Flordaguard Nemaguard Okinawa Tsukuba 1 Umezeiro Seleção Viamão CV (%) 3,34 3,65 3,55 3,29 3,05 2,57 3,30 18,12 3,91 a 2,38 ab 2,61 a 2,47 a 2,41 ab 2,32 ab 1,35 b 3,15 a 15,59 5,13 a 4,50 ab 5,61 a 5,24 a 4,68 ab 5,74 a 3,29 b 4,82 a 10,6 4,71 a 4,42 a 4,93 a 4,41 a 3,92 ab 4,13 ab 2,17 b 4,22 a 17,07 6,50 a 6,54 a 6,63 a 6,67 a 6,85 a 6,45 a 4,71 b 6,43 a 7,54 6,04 a 6,30 a 6,04 a 5,64 a 6,06 a 5,73 a 4,50 b 6,07 a 5,80 8,07 a 7,91 a 8,10 a 7,98 a 7,83 a 7,64 a 5,98 b 7,64 a 7,48 7,16 a 7,57 a 7,68 a 6,86 ab 7,33 a 6,97 ab 5,78 b 7,14 a 5,92 5,30 b 8,98 ab 5,24 b 6,29 ab 5,50 b 11,39 a 8,39 ab 11,32 a 25,18 8,89 n.s. Provavelmente, o retardo no IF da cultivar 'Maciel' sobre 'Flordaguard' e 'Okinawa' em 2008, da PF, quando enxertada sobre Okinawa (2008) e Flordaguard (2009) e a antecipação do IF, quando enxertada sobre Nemaguard (2009), Seleção Viamão e Capdeboscq (2008 e 2009) e antecipação da PF quando enxertada sobre Nemaguard (2009), sejam em função do vigor destes porta-enxertos. Para a cultivar 'Chimarrita', a antecipação do IF foi verificada quando enxertada sobre Seleção Viamão (2008), Aldrighi (2009) e Nemaguard (2008 e 2009) e a antecipação da PF foi observada quando enxertada sobre Nemaguard (2008 e 2009) e Seleção Viamão (2008 e 2009). Os frutos da cultivar 'Maciel' enxertada sobre Umezeiro, produzidos no ano de 2008, apresentaram menor acidez titulável (AT), diferindo apenas de Aldrighi e Nemaguard; maior relação SS/AT, não diferindo de Tsukuba 1 e pH mais elevado dentre os porta-enxertos avaliados (Tabela 2).

1122175 — **Efeitos do clima sobre o incremento diamétrico de Paricá (*Schizolobium parahyba* var. *amazonicum* - Fabaceae) em plantios comerciais** — Monocultivo de Paricá (Moncp); (ii) Sistema Agroflorestal Paricá e Curauá (SAFpc); (iii) Consórcio Florestal Paricá e Freijó (CFpf); (iv) Sistema Agroflorestal Paricá, Freijó e Curauá (SAFpfc); (v) Monocultivo de Paricá (Moncp); Sistema Agroflorestal Paricá e Curauá (SAFpc); Consórcio Florestal Paricá e Freijó (CFpf); Sistema Agroflorestal Paricá, Freijó e Curauá (SAFpfc); Consórcio Florestal Paricá, Mogno, Freijó (CFpmf) e Sistema Agroflorestal Paricá, Mogno, Freijó e Curauá (SAFpmfc), Campo Experimental Tramontina Belém, S.A, Aurora do Pará, PA, Brasil. Agroforest System of Paricá and Curauá (SAFpc); Forest Consortium of Paricá and Freijó (CFpf); Agroforest System of Paricá, Freijó, and Curauá (SAFpfc); Forest Consortium of Paricá, Mahogany, and Freijó (CFpmf), and Agroforest System of Paricá, Mahogany, Freijó, and Curauá (SAFpmfc), Campo Experimental Tramontina Belém, S.A, Aurora do Pará, PA, Brazil. matrix among the climate characteristics and the periodic semester increment in diameter (IPSDAP) of Paricá, Campo Experimental Tramontina Belém, S.A, Aurora do Pará, PA, Brazil. height (DAP) and mean annual increment in diameter (IMADAP) of Paricá in consoriated forest plantings, agroforests, and in monoculture, Campo Experimental Tramontina Belém, S.A, Aurora do Pará, PA, Brazil.

1129319 — **Análise exploratória de dados espaciais da produção de café no estado do Paraná, 1980-2018** — Ideias centrais • Inovação tecnológica e produtividade na cultura cafeeira do Paraná • Análise exploratória de dados espaciais da cultura do café no Paraná • Tendências da cultura do café no estado do Paraná 1980-2018 Recebido em 13/08/2020 Considerando-se a importância da cultura do café na economia do estado e o desenvolvimento tecnológico da produção, o objetivo do presente estudo foi realizar a análise exploratória dos dados espaciais, da produção de café dos municípios do estado do Paraná no período 1980-2018. Especificamente pretendeu-se: a) estimar e analisar a evolução da área colhida, produção e produtividade da cultura do café no Paraná, no período 1980-2018; e b) realizar a análise exploratória de dados espaciais da área colhida e a produtividade dos municípios do Paraná, para identificar agrupamentos espaciais (clusters) e analisar sua evolução. É importante destacar a participação do Instituto Agrônomo do Paraná (Iapar), Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e Universidade Estadual de Londrina (UEL) no desenvolvimento de novas técnicas de produção e melhoramento genético para a cultura do café, com o objetivo de aumento da produtividade, resistência a pragas, qualidade da bebida e redução do custo médio de produção (Martin et al., 1995; Demoner et al., 2007). Estas tecnologias trouxeram melhoria para a qualidade do café produzido e mudança no processo de comercialização, pois, na década de 1990, cerca de 55% dos produtores realizavam colheita no pano, 60% vendiam seu produto já beneficiado, e 55% verificavam o tipo de bebida antes da comercialização (Demoner et al., 2007) Para obter os novos derivados de café demandados pelos consumidores, foram necessárias adaptações dos equipamentos e das linhas de produção, em geral para reduzir custos, e torradores automatizados e mais modernos, com maior precisão de controle da torra, modernização do processo de moagem, com a aquisição de moinhos menos agressivos aos grãos, e empacotadoras automatizadas e mais precisas (Santini et al., 2006).

1124357 — **Comparison between barley malt protein quantification methods** — Considering the importance of protein to the brewing process, four protein quantification methods were compared: combustion methods - carbon, hydrogen and nitrogen elemental analysis and Dumas - and digestion ones - Kjeldahl and flux injection analysis (FIA). The Kjeldahl method was created in 1833 and meant for the brewing industry, as it can be used to measure the differences in protein content in grains throughout the malting and fermentation processes (Kjeldahl, 1883; The genotype BRS Brau was cultivated in the Midwestern part of Brazil, both in the Federal District (CFI) and in the State of Goiás (CFN); the other, BRS Cauê, was grown in the Federal District (CFE) and in the Southern part of Brazil (CP) and donated by Fundação Agrária de Pesquisa Agropecuária (FAPA) located in Guarapuava, Paraná, Brazil. Kjeldahl and Carbon, Hydrogen and Nitrogen (CHN) elemental analysis, (b) flux injection analysis (FIA) and CHN, (c) Kjeldahl and FIA, (d) Dumas and CHN, (e) Dumas and Kjeldahl and (f)

57793 — **Adaptabilidade e estabilidade de cultivares de milho no nordeste brasileiro** — As cultivares BR 106, BR 5011, BR 5004 e BR 5033 têm recomendação justificada, especialmente, para pequenos e médios produtores rurais, onde certamente contribuirão para a melhoria da produtividade do milho. BR 106 e BR 5011 (de porte normal e ciclo semitardio), BR 5028 e BR 5033 (de porte baixo e ciclo precoce) e a BR 5037 (de porte baixo e ciclo superprecoce), de boa adaptação e de alta estabilidade de produção na região (1998a) verificaram, em cinco ambientes do Estado de Sergipe, que as cultivares BR 106, BR 5011, BR 5033 e BR 5037 repetiram o bom comportamento apresentado nos trabalhos anteriores, justificando suas recomendações para o estado. Relacionando-se a estabilidade das cultivares avaliadas neste trabalho com suas respectivas bases genéticas (Tabela 5), nota-se que todas as cultivares mostraram a mesma resposta à estabilidade, à exceção das BR 473 e BR 5033, independentemente de suas bases genéticas (híbrido triplo, híbrido duplo e cultivares), apesar de o híbrido triplo BR 3123, dos híbridos duplos As cultivares BR 106, BR 5011, BR 5004 e BR 5033 podem melhorar substancialmente a produtividade média dos sistemas de produção dos pequenos e médios produtores rurais.

1126778 — **Práticas agroflorestais no Rio de Janeiro: interação de saberes para a conservação ambiental e valorização da campesinidade** — Práticas agroflorestais no Rio de Janeiro: interação de saberes para a conservação ambiental e valorização da campesinidade Agroforestry in Rio de

Janeiro state: interaction of knowledges for environmental conservation and valorisation of peasantry
 MATTOS, Claudemar¹; TUBENCHLAK, Fernanda²; MORAES, Luiz Fernando³ 1 PPGCiAC - UFRJ Macaé / Centro Tiê de Agroecologia, claud3mar@gmail.com; 2 Instituto Internacional para Sustentabilidade, f.tubenchlak@iis-rio.org; Palavras-Chave: Agrofloresta; Sistematização; Mapeamento; Restauração de Paisagens Keywords: Agroforestry; Systematization; Mapping; Restoration of Landscapes Abstract: The geographic and ecological characteristics, the dominant agrarian structure, together with the knowledge accumulated by public and civil society institutions, give the state of Rio de Janeiro a natural aptitude to the adoption and development of agroforestry practices. Este processo teve início em 2010 e foi apresentada em junho de 2012, no I Seminário de práticas Agroflorestais e de Pousio no Rio de Janeiro, e resultou na publicação da Resolução Inea n. 86/2014, que dispunha sobre critérios e procedimentos para a implantação, manejo e exploração de sistemas agroflorestais e a prática do pousio no Estado do Rio de Janeiro (Nazário et al., 2016). Neste contexto, o presente relato descreve o resultado do processo de mapeamento de práticas agroflorestais no estado do Rio de Janeiro, buscando contribuir com o registro histórico de parte das práticas agroflorestais fluminense, bem como contribuir com o debate dos avanços e desafios para que as práticas agroflorestais sejam mais utilizadas nos manejos de agroecossistemas agroecológicos, a fim de sistematizar e promover o intercâmbio de conhecimentos sobre o uso atual de Sistemas Agroflorestais (SAF) por agricultores familiares no Rio de Janeiro. O mapeamento de experiências agroflorestais foi realizado no âmbito do Projeto SEISAF (Sistematização participativa de experiências e intercâmbio de conhecimentos em sistemas agroflorestais voltados à agricultura familiar em regiões da Mata Atlântica no sul e sudeste do Brasil) coordenado pela Embrapa Meio Ambiente, sendo dinamizado no Rio de Janeiro pela Embrapa Agrobiologia e ASPTA, entre outras entidades.

89928 — Tomada de decisões gerenciais no âmbito de duas fazendas familiares no sul goiano

— Uma fazenda de pequena escala (61 ha) tem a produção de leite como atividade principal, e uma outra, de tamanho médio (422 ha), que incorpora os cultivos de grãos, cereais e tomate industrial, bem como a atividade produção de leite, são estudadas como sistemas representativos de produção agropecuária. Nestas tabelas estão detalhados os custos de produção (direto, indireto e fixo), os resultados da pecuária leiteira (venda de leite, descarte de animais e ganho de capital) e os resultados de venda (produção média diária (L/dia), preço médio recebido por litro de leite, custo do litro de leite e lucro da venda). (L/dia) da venda do leite Produção média diária 345,0740 Preço médio recebido pelo litro de leite Custo do litro de leite Lucro da venda de leite 338 % do total Lucro da atividade (R\$) Nessa análise técnico-financeira, foram considerados os resultados de 2 anos agrícolas em que o mercado do leite apresentou preços que remuneravam todos os custos de produção das fazendas estudadas, mas, no final de 2001, verificou-se queda de preços do produto sem, na opinião dos produtores e dados preliminares, a correspondente redução nos custos de produção (Fig. 0,5 L/ha 1,39 L/ha 92,47 kg/ha 49,5 kg/ha 39,5 kg/ha 200 kg/ha 82 kg/ha 172 mL/ha 146 mL/ha 86 mL/ha 258 mL

1127555 — Tendência em séries hidrológicas e de mudanças no uso e cobertura da terra da bacia hidrográfica do rio Guanhães, Minas Gerais

— Tendência em séries hidrológicas e de mudanças no uso e cobertura da terra da bacia hidrográfica do rio Guanhães, Minas Gerais Trend in hydrological series and changes in land use and land cover in the Guanhães, Minas Gerais river basin DOI: 10.34188/bjaerv3n2-004 Recebimento dos originais: Fernanda Laurinda Valadares Ferreira Doutoranda em Engenharia Agrícola pela Universidade Federal de Viçosa Instituição: Universidade Federal de Viçosa – Departamento de Engenharia Agrícola Endereço: Av. Peter Henry Rolfs - Campus Universitário, Viçosa - MG, 36570-900 E-mail: fernanda_laurinda@yahoo.com.br Lineu Neiva Rodrigues Doutor em Engenharia Agrícola pela Universidade Federal de Viçosa Instituição: Embrapa Cerrados Endereço: Parque Estação Biológica – PqEB, Brasília, DF, 70770-901 E-mail: lineu.rodrigues@embrapa.br Laura Thebit de Almeida Doutoranda em Meteorologia aplicada pela Universidade Federal de Viçosa Instituição: Universidade Federal de Viçosa – Departamento de Engenharia Agrícola Endereço: Av. Peter Henry Rolfs - Campus Universitário, Viçosa - MG, 36570-900 E-mail: l.thebit@gmail.com David Bruno de Sousa Teixeira Doutorando em Meteorologia aplicada pela Universidade Federal de Viçosa Instituição: Universidade Federal de Viçosa – Departamento de Engenharia Agrícola Endereço: Av. Peter Henry Rolfs - Campus Universitário, Viçosa - MG, 36570-900 E-mail: data of the hydrographic basin of the Guanhães River, Minas Gerais and to correlate with

the use and occupation of the land, between ISSN 2595-573X 452 Brazilian Journal of Animal and Environmental Research Para analisar a tendência foram utilizados dados de vazões e precipitações, sendo aplicados os testes não paramétricos de Mann Kendall e Pettitt, considerando-se um nível de significância de 5%. ISSN 2595-573X 457 Brazilian Journal of Animal and Environmental Research KARMESHU, N. Trend detection in annual temperature & precipitation using the mann kendall test—a case study to assess climate change on select states in the Northeastern United States.

1126373 — **Como manejar doenças foliares em milho** — As principais doenças do milho no Brasil são a mancha branca (*Pantoea ananatis*), a ferrugem polisora (*Puccinia polysora*), a ferrugem tropical (*Puccinia sorghi*), a ferrugem branca (*Physopella zaeae*), a helmintosporiose (*Exserohilum turcicum*), a cercosporiose (*Cercospora* Revista Plantio Direto - Edição Especial Centro-Oeste - Outubro de 2020 *zea-maydis*), as podridões de espigas e grãos (*Fusarium* sp. Para os enfezamentos, haverá um tópico específico por se tratar de doenças que vem ganhando importância em diversos estados brasileiros e que tem seu controle baseado em ações específicas quando comparadas as demais doenças em milho. Em caso de dúvida ou diante de sintomas atípicos, consultar especialistas é importante, pois doenças menos comuns ou novas doenças podem ocorrer, como o caso da estria bacteriana causada pela bactéria *Xanthomonas vasicola* pv *vasculorum*, relatada em 2018 no estado do Paraná (Leite Junior et al. 36 Revista Plantio Direto - Edição Especial Centro-Oeste - Outubro de 2020 Controle químico de doenças foliares Após identificar as doenças que prevalecem na região, avaliar as condições climáticas e as fases do ciclo da cultura e o nível de resistência dos híbridos para o desenvolvimento das doenças, é hora de avaliar a necessidade de realizar a aplicação de fungicidas. Além das práticas já conhecidas para redução de doenças, a eliminação de tigueras passa a ser considerada como ação importante do manejo de doenças, pois o milho que cresce nas áreas aumenta a ponte verde e favorece a manutenção de inóculo de patógenos e insetos entre safras.

1074606 — **Zoneamento edafoclimático da acácia-negra para o extremo sul do Brasil** — A integração dos zoneamentos edáfico e climático, de A interação entre os zoneamentos edáfico e climático, nominada zoneamento edafoclimático, foi elaborada com denominado zoneamento edafoclimático, foi elaborada o uso de ferramentas de sistemas de informações geográficas com o uso de ferramentas de sistemas de informações geoclas, por meio de operações entre imagens geradas em cada gráficas, por meio de operações entre imagens geradas em tipo de zoneamento - climático e edáfico, conforme mencada tipo de zoneamento (climático e edáfico), conforme Agrometeoros, Passo Fundo, v.25, n.1, p.59-69, ago 2017. A interação entre os zoneamentos edáfico e climático, denominado zoneamento edafoclimático, foi elaborada com o uso de ferramentas de sistemas de informações geográficas, por meio de operações entre imagens geradas em cada tipo de zoneamento (climático e edáfico), conforme mencionado anteriormente, obtendo-se as Tabelas 9, 10 e 11, onde são apresentadas as áreas, na região de abrangência do extremo sul do Rio Grande do Sul, RS, relativas às classes de preferência. solos estão contidas em mapas escalas mais 53.006 de solos em 1,4 por risco de geada Desenvolveu-se uma estratégia de interpretação do genéricas (escala > 1: 50.000), onde as unidades 465.021 12,5 de mapea Restrição edáfica e climática por risco de geada mapa Restrição de solos disponível mento são de composição mais heterogênea, 33,1 é necessário o 1.226.736 edáfica para o zoneamento edáfico da acácia-negra para a região de abrangência do Corede Sul, agrupamento das unidades de mapeamento em categorias, RS. O cultivo da acácia-negra no Riodas Grandedulado, do Sul, municípios dopor Corede Sul, Portanto, o unidades de mapeamento a serem avaliadas tendem a forpresente zoneamento poderá apresentar diferenças Serra do Sudeste, é viável, desde que seguidas as orientações deste zoneamento de en66 edafoclimático, realizando os plantios em zonas com riscos menores de geada e com solos profundos, de boa drenagem e menor pedregosidade. On the other hand, the region of the Extreme South of the State of Rio Grande do Sul is lacking in investments, counting on favorable areas for commercial plantations of *Acacia mearnsii*, a species with great potential for development in some regions of the state.

207984 — **Adubação e calagem para o feijoeiro irrigado em solo de cerrado** — no segundo, a resposta de cinco cultivares/linhagens de feijoeiro (Aporé, Carioca, Novo Jalo, MA534657 e MA534666-2) à aplicação no sulco de plantio de 200, 400 e 600 kg ha⁻¹ do fertilizante 4-30-16 mais 30 kg ha⁻¹ de N em cobertura aos 25 dias após germinação, e no terceiro, as doses de N (0, 30, 60, 90 e 120 kg ha⁻¹) aplicadas em cobertura a intervalos de dez dias, a partir dos 20 até aos 70 dias

após a emergência das plântulas. Quanto à calagem e à adubação de plantio, os mesmos autores observaram aumentos de até 54% na produtividade do feijoeiro (cultivar Aporé), decorrentes da calagem e de uma dose econômica de adubação NPK, no plantio, de 360 kg ha⁻¹ do formulado 4-30-16, para uma produtividade de 2.234 kg ha⁻¹. O primeiro experimento recebeu uma adubação de 16-120-64 kg ha⁻¹ de N-P₂O₅-K₂O, respectivamente, no ADUBAÇÃO E CALAGEM PARA O FEIJOEIRO IRRIGADO plantio, e 30 kg ha⁻¹ de N em cobertura aos 25 dias após a emergência das plantas. Com base na quantidade total absorvida (parte aérea + grãos) na média dos tratamentos com calcário (Tabela 2), calcula-se que, para a produção de uma tonelada de grãos, houve necessidade de 47 kg de N, 5,0 kg de P, 37,0 kg de K, 11,9 kg de Ca, 3,9 kg de Mg, 17 g de Cu, 500 g de Fe, 94 g de Mn e 72 g de Zn. Com exceção da cultivar Carioca, a dose de 400 kg ha⁻¹ de fertilizante 4-30-16 mais 30 kg ha⁻¹ de N em cobertura aumentou a produção de grãos das cultivares/linhagens em relação à dose de 200 kg ha⁻¹ mais 30 kg ha⁻¹ em cobertura, embora não houvesse diferença estatisticamente significativa entre as doses.

157560 — Desempenho de ovinos terminados em confinamento com resíduo desidratado de vitivinícolas associado a diferentes fontes energéticas — Na região do Vale do São Francisco, o resíduo de vitivinícolas está sendo em grande parte desperdiçado, podendo ser aproveitado na alimentação animal sob as formas de silagem ou resíduo desidratado, garantindo um bom aporte de nutrientes para os animais, principalmente no período de maior escassez de forragem. A terminação de ovinos exclusivamente a pasto, praticada na maioria das propriedades rurais do semi-árido nordestino, tem-se mostrado ineficaz em grande parte dos aspectos de produção, pois este processo está submetido à irregularidade na disponibilidade de forragem da caatinga, ocasionando longos períodos para os animais alcançar em peso o de abate. **DISCUSSÃO** Observa-se que o tipo de concentrado energético associado ao resíduo de vitivinícolas influenciou ($P < 0,05$) o ganho de peso diário dos animais aos 63 dias, com maiores respostas em ganho de peso para os animais alimentados com as dietas compostas por resíduo e grão de milho moído e resíduo e feno de palma, que não diferiram ($P > 0,05$) entre si (Tabela 3). A igualdade de ganho de peso verificada entre as dietas com feno de palma e grão de milho moído, mesmo com as diferenças em consumo, pode ser explicada pela digestibilidade dos ingredientes, visto que o grão de milho moído apresentou maiores valores de DI VMS (Tabela 1) e maiores valores de NDT quando combinado ao resíduo de vitivinícolas (Tabela 2). Segundo NEUMANN (1977), maior densidade energética resulta em maior ingestão de energia, por conseguinte, menos alimento é requerido para o ganho de peso, o que pode ter ocorrido no presente trabalho, no qual os melhores resultados de conversão foram obtidos para as dietas com grão de milho moído e feno de palma, as quais apresentaram as maiores ingestões de CHOT (Tabela 3).

156709 — Consumo de frutos do imbuzeiro (Spondias tuberosa Arruda) pelos caprinos na caatinga — Número de animais (A), peso vivo dos animais (PV), número de sementes consumidas por dia por animal (SD), número de sementes consumidas por animal na safra (SF), período de consumo de frutos pelos animais (PD), peso médio de sementes consumidas por animal (PMS), peso médio dos frutos consumidos por animal (PMF), peso total estimado dos frutos consumidos pelos animais (PMTF), peso médio dos frutos consumidos pelos animais por dia (PMFCAD), e porcentual de consumo de frutos em relação ao peso vivo dos animais (PCPV) na comunidade de Alto do Angico, em 2003 - Number of animals (A) of food consumption by the animals (PO), mean weight of seeds consumed per animals (PMS), mean weight of seeds consumed per animal (PMF), estimated total weight of the fruits consumed by the animals (PMTF), mean weight of the fruits consumed by the animals per day (PMFCAD) and fruit consumption percentage in relation to live weight of the animals (PCPV) in Alto do Angico, em 2003 Contudo, o peso total estimado de 144 kg de frutos consumidos pelos animais dessa comunidade foi menor que na comunidade do Alto do Angico, onde os animais consumiram, em média, 163 kg de frutos no período de avaliação. - Quadro 2 Número de animais (A), peso vivo dos animais (PV), número de sementes consumidas por dia por animal (SD), número de sementes consumidas por animal na safra (SF), período de consumo de frutos pelos animais (PD), peso médio de sementes consumidas por animal (PMS), peso médio dos frutos consumidos por animal (PMF), peso total estimado dos frutos consumidos pelos animais (PMTF), peso médio dos frutos consumidos pelos animais por dia (PMFCAD), e porcentual de consumo de frutos em relação ao peso vivo dos animais (PCPV) na comunidade de Fazenda Saco, em 2003 - Number of animals (A), animal live weight (PV), number of seeds consumed per animal (SO), number of seeds consumed

per animal in the crop (SF), period of food consumption by the animals (PO), mean weight of seeds consumed per animal (PMS), mean weight of seeds consumed per animal (PMF), estimated total weight of the fruits consumed by the animals (PMTF), mean weight of the fruits consumed by the animals per day (PMFCAO) and fruit consumption percentage in relation to live weight of the animals (PCPV) in Fazenda Saco, 2003 - Quadro 3 Número de animais (A), peso vivo dos animais (PV), número de sementes consumidas por dia por animal (SD), número de sementes consumidas por animal na safra (SF), período de consumo de frutos pelos animais (PD), peso médio de sementes consumidas por animal (PMS), peso médio dos frutos consumidos por animal (PMF), peso total estimado dos frutos consumidos pelos animais (PMTF), peso médio dos frutos consumidos pelos animais por dia (PMFCAD), e porcentual de consumo de frutos em relação ao peso vivo dos animais (PCPV) na comunidade da Lagoa do Meio, em 2003 number of animals (A), animal live weight (PV), number of seeds consumed per animal (SD).

105845 — **A agricultura orgânica e seu potencial para o pequeno agricultor** — Este trabalho faz uma breve abordagem sobre as diferentes correntes da agricultura alternativa e aprofunda o tema nos seguintes pontos: a) situação atual e perspectivas do mercado interno e externo de produtos orgânicos; b) condições de certificação e comercialização de produtos orgânicos no país e legislação que disciplina esses processos; Além da característica técnica, a agricultura biodinâmica fundamenta-se na ciência ritual antroposófica¹. A agricultura ecológica ou agroecologia vai além das outras correntes, considera que as lavouras são ecossistemas nos quais os processos intrínsecos são fundamentais, como a interação da vida e do espírito, assim como a interação natural e o mundo como um todo. A natureza é a natureza e o espírito a direção e sua visão ao mundo a partir de fatos e fenômenos,..." (Klett, 1999, p.29). O diferencial no preço de mercado dos produtos orgânicos em relação aos produtos convencionais tenderá a desaparecer à medida que a quantidade ofertada de produtos orgânicos aumente e atenda a quantidade demandada desses produtos. Uma questão que fica é: se os preços de mercado dos produtos agrícolas orgânicos se igualarem aos dos produtos convencionais, será que os pequenos produtores orgânicos poderão arcar com os custos do processo de certificação orgânica e com todos os outros custos adicionais que ele envolve? O conflito existente entre a agricultura orgânica concebida como um setor ou indústria de produção e o movimento social ambientalista é evidente em toda a história recente de regulação nos EUA, iniciando com a regulação voluntária dos agricultores de padrões orgânicos nos anos 70 e culminando com a aprovação do Decreto Federal de Produção de Alimentos Orgânicos de 1990 (Goodman, 2000).

315462 — **Resposta de Pinus taeda L. à aplicação de cinza de biomassa vegetal em Cambissolo Húmico, em vaso** — Foram analisadas as acículas e o solo (camadas 0 a 10 cm, 10 cm a 20 cm e 20 cm a 30 cm). A aplicação da cinza diminuiu a acidez e o teor de Al, e aumentou os teores de Ca, Mg, K e P, com efeitos mais pronunciados na camada de 0 a 10 cm; não houve efeito dos tratamentos no desenvolvimento das mudas; o teor de Cu nas acículas foi diminuído com o aumento da dose de cinza. Amostras de solo foram coletadas nas camadas de 0 a 10 cm, 10 cm a 20 cm e 20 cm a 30 cm, sendo determinados o pH em CaCl₂, a acidez potencial (H+Al) e os teores de Ca, Mg, K, P, Na e C, conforme metodologia descrita em Silva (1999). Esse valor foi superado com a aplicação de 10 t.ha⁻¹, nas camadas de 0 a 10 cm, enquanto que nas camadas 10 cm a 20 cm e 20 cm a 30 cm, a saturação de K foi maior do que 2,5 % a partir da dose 20 Resultados analíticos de características químicas de amostras de um Cambissolo Húmico (camadas 0 a 10 cm, 10 cm a 20 cm e 20 cm a 30 cm), submetidos a doses de cinza de caldeira.

1125816 — **O almoço do Círio de Nossa Senhora de Nazaré em Belém do Pará: experiência cultural e gastronômica** — De acordo com o Instituto do Patrimônio Histórico Artístico Nacional (IPHAN) (2006:13),... o Círio de Nazaré é um acontecimento que envolve, direta ou indiretamente, toda a população paraense, estendendo sua influência para além dos limites do estado do 4 Research, Society and Development, v. 9, n. 8, e621985669, 2020 (CC BY 4.0) | ISSN 2525-3409 | DOI: <http://dx.doi.org/10.33448/rsd-v9i8.5669> Pará... configurando-o como um dos fenômenos religiosos mais importantes do Brasil. Em 2004, se iniciou o processo no IPHAN que culminou com o reconhecimento do Círio de Nazaré como patrimônio cultural brasileiro, constando no registro os seguintes elementos sobre os quais recaiu o título de patrimônio cultural brasileiro: Procissão

Principal do Círio; as Imagens Original e Peregrina de Nossa Senhora de Nazaré; a Trasladação; Não há alusão as práticas alimentares peculiares nas narrativas que descrevem a origem e os primórdios do Círio de Nazaré (Frugoli; Bueno, 2014), no entanto, mesmo sem poder determinar ao certo sua origem, o Almoço do Círio é um dos elementos reconhecidos como patrimônio cultural, como pertencente à identidade das festividades de Nossa Senhora de Nazaré. Apesar de não apresentar caráter marcadamente ritualístico ou não estar diretamente 9 Research, Society and Development, v. 9, n. 8, e621985669, 2020 (CC BY 4.0) | ISSN 2525-3409 | DOI: <http://dx.doi.org/10.33448/rsd-v9i8.5669> ligado à porção sagrada dos festejos, o Almoço do Círio, flutua no limiar imperceptível entre o sagrado e o profano, cumprindo sua função social como a parte mais íntima da festividade na qual, celebra-se o Círio entre amigos e familiares, apresentando uma “função ritual e simbólica de uma unidade social mais ampla com o ideal de confraternização e convivialidade que se opõe à realidade do individualismo cotidiano” (Frugoli & Bueno, 2014:149). Henrique (2011) um dos relatores do processo de requerimento do status de patrimônio cultural do Círio de Nazaré narra que durante suas pesquisas de campo constatou que o almoço do Círio é um elemento que ultrapassa a questão religiosa, tendo entrevistado, um jovem evangélico que participava da festa em casa de amigos católicos, identificando aquele momento como uma importante confraternização paraense, afirmando que “nesse ponto eu não sou nem evangélico, nem católico, entende?”

215848 — Trinta e dois anos do programa de melhoramento genético do feijoeiro comum em Minas Gerais — Conduziu-se este trabalho, com o objetivo de proceder à análise crítica do Programa de melhoramento genético do feijoeiro comum em Minas Gerais nos últimos 32 anos com ênfase nas implicações da interação linhagens x ambientes, além de avaliar se houve progresso genético. A média da cultivar Carioca, comum em todos os experimentos, foi utilizada como estimador do efeito ambiental e a média das cinco melhores linhagens sob avaliação de cada biênio, como efeito fenotípico. Progresso Genético, Interação Genótipos x Ambientes, Índice de Confiança, *Phaseolus vulgaris* L. ABSTRACT The objective of the present work was to make critical analysis of the dry bean genetic breeding Para isso foi estimado o desvio genético do biênio j (dj) pela diferença entre a produtividade média de grãos das cinco melhores linhagens avaliadas (L_j), no biênio j e a média da Carioca (C_j), no respectivo biênio, ou seja: $dj = \text{Linhagens} \times \text{Locais} - \text{Linhagens} \times \text{Anos} - \text{Linhagens} \times \text{Safras} - \text{Linhagens} \times \text{Anos} \times \text{Safras} \times \text{Locais}$

1118305 — Contribuições do geoprocessamento à compreensão do mundo rural e do desmatamento no bioma Amazônia — Para contribuir para uma melhor compreensão do mundo rural na Amazônia, em bases quantificadas e territoriais, este estudo teve como objetivos identificar, qualificar e quantificar o universo atual dos produtores rurais lato sensu e suas unidades de produção no bioma Amazônia, em bases territoriais e analisar em bases numéricas e cartográficas a participação efetiva do mundo rural nos processos de desmatamento no bioma Amazônia. Os cálculos evidenciaram a existência de três subpopulações de produtores rurais: a dos estabelecimentos agropecuários do IBGE coincidentes territorialmente com os imóveis rurais do CAR e as dos estabelecimentos agropecuários do IBGE não correspondentes aos imóveis rurais cadastrados no CAR e vice-versa. O MUNDO RURAL E O DESMATAMENTO NO BIOMA AMAZÔNIA Qual a participação desse universo de produtores rurais e unidades de produção no processo de desmatamento no bioma Amazônia? Diante do perene desafio de conhecer o universo de produtores, unidades e áreas de produção, a equipe da Embrapa Territorial percebeu, em 2018, a oportunidade criada por duas iniciativas públicas independentes: os dados geocodificados dos imóveis rurais do Cadastro Ambiental Rural (CAR) e as coordenadas geográficas dos estabelecimentos agropecuários recenseados pelo IBGE em 2017. Ele resulta de três subpopulações: a dos estabelecimentos agropecuários do IBGE coincidentes territorialmente com os imóveis rurais do CAR, a dos estabelecimentos agropecuários do IBGE não correspondentes a imóveis rurais cadastrados no CAR e a dos imóveis do CAR aparentemente não visitados por recenseadores do IBGE.

212670 — Armazenamento de sementes de feijão revestidas com polímeros e tratadas com fungicidas — A avaliação do efeito do armazenamento na qualidade de sementes de feijão foi efetuada logo após o revestimento ou tratamento químico e ao longo do armazenamento (um, dois e quatro meses), mediante teste de germinação e teste de sanidade de sementes, conforme as regras para análises de sementes (Brasil, 1992). As sementes tratadas com o fungicida captan (pó seco),

associado ou não aos polímeros, mostraram menor vigor, principalmente após dois meses de armazenamento, em relação às sementes tratadas com os fungicidas benomyl e carbendazin. O vigor, avaliado pela primeira contagem de germinação, das sementes de feijão tratadas com fungicidas e revestidas com polímeros não superou o das sementes tratadas apenas com fungicidas, principalmente até o primeiro mês de armazenamento (Tabela 1). A porcentagem de sementes com bactérias (método do rolo de papel) aumentou ao longo do armazenamento das sementes, o que pode ter mascarado os efeitos dos tratamentos químicos, e pode ter sido uma das causas de redução da germinação das sementes (Tabelas 1 e 4), conforme também verificado por Menezes & Mohan (1982), no controle com inoculação (48,5%), nas sementes revestidas com polímeros (38,0%), e nas sementes tratadas com benomyl (16,5%) e com benomyl misturado aos polímeros (7,0%).

313756 — Desempenho de três mecanismos sulcadores de semeadura na presença de diferentes culturas destinadas à cobertura de inverno — No presente trabalho foram estudados os efeitos de três mecanismos de semeadura (duplo disco defasado, cinzel e duplo disco na germinação e a emergência da soja e de invasoras) na presença de diferentes coberturas de solo (testemunha, 3 e 6 t ha⁻¹ de ervilhaca peluda, e 4 e 8 t ha⁻¹ de aveia preta). Foram avaliadas a cobertura do solo antes e após a escarificação, a largura e a profundidade de trabalho dos mecanismos sulcadores de semeadura e a emergência de plântulas de soja e de invasoras na linha de semeadura. No presente trabalho foram estudados os efeitos de três mecanismos de semeadura (duplo disco defasado, cinzel e duplo disco na germinação e a emergência da soja e de invasoras) na presença de diferentes coberturas de solo (testemunha, 3 e 6 t ha⁻¹ de ervilhaca peluda, e 4 e 8 t ha⁻¹ de aveia preta). Efeito dos sulcadores de semeadura duplo disco defasado (DDD), cinzel (C) e duplo disco (DD) e de diferentes culturas de cobertura vegetal, sobre a emergência de plântulas de invasoras. As alternativas de cobertura do solo e os mecanismos sulcadores estudados apresentaram interação significativa sobre a emergência das invasoras: a maior emergência de invasoras ocorreu na linha de semeadura do mecanismo sulcador tipo cinzel.

955145 — Resultados da pesquisa: como interpretá-los? — Nessa visão, a taxa de crescimento da produção é equivalente à taxa de crescimento dos insumos. As pesquisas empíricas têm mostrado que, taxa (PTF) > 0, ou seja, a produção cresce à taxa mais alta que a taxa de crescimento dos insumos. Sendo taxa (PTF) > 0, a contribuição de conhecimento existe e é dada pela taxa (PTF); quando a taxa (PTF) = 0, significa que inexistente a contribuição de conhecimentos. Na Tabela 1, a PTF está grafada sobre fundo amarelo, para ressaltar que ela é a diferença entre a taxa de crescimento do produto e a taxa de crescimento do insumo. Sabendo-se que taxa(insumos) + taxa (PTF) = taxa (produtos) cabe, assim, perguntar quantos por cento do crescimento do produto, portanto, os conhecimentos, a PTF explica.

957259 — Aplicação de técnicas de sensoriamento remoto e geoprocessamento como apoio a estudos geológicos e geotécnicos para gestão ambiental do município de três cachoeiras, litoral Norte do RS, Brasil — O Município de Três Cachoeiras, situado no Litoral Norte do Rio Grande do Sul faz parte da "Reserva da Biosfera da Mata Atlântica" e tem conflitos ambientais diversos, como áreas para disposição de resíduos sólidos, assentamento de loteamentos residenciais e industriais, jazidas de extração de material para construção, fontes de abastecimento de água e áreas de preservação ambiental. Assim, a partir de levantamentos das características do meio físico, tais como declividade, litologia, formações superficiais, águas subterrâneas e uso do solo, recomendaram-se zonas favoráveis à extração de materiais para construção civil, à implantação de obras de 90 infraestrutura, à disposição de resíduos sólidos, ao assentamento de loteamentos, à agricultura e à implantação de áreas de preservação ambiental. (1988), depósitos de encosta, incluindo tálus e leques aluviais antigos e atuais, depósitos lagunares, incluindo depósitos fluviais e de retrabalhamento fluvial, depósitos eólicos de dunas litorâneas antigas e depósitos marinhos praias antigas são encontrados na região de Três Cachoeiras. Ressalta-se que, devido à cobertura vegetal, os alvos investigados neste tema (solos e rochas), também foram identificados por meio de aspectos de contexto, pois seus atributos espectrais puderam ser observados apenas em locais que continham solos expostos, pedreiras, etc. Considerando-se limitações de resolução espacial e radiométrica da imagem, foram separadas as classes depósitos flúvio-lagunares; depósitos alúvio-deltaicos; depósitos de colúvio; depósitos de elúvio; rochas aflorantes + depósitos de elúvio; água fluvial e água lagunar.

4.6.1.5 – Carta de recomendação à implantação de áreas de preservação ambiental e/ou à recuperação Para este produto ser viabilizado, consideraram-se as informações atualizadas pela imagem de satélite, que indicou as áreas de mata nativa e zonas alagadas (banhados), as altas declividades dadas pela carta específica e outras áreas definidas como reservas ecológicas pela Resolução CONAMA n° 004 de 18.11.85.

1120755 — **Progressos na vitivinicultura brasileira** — Como exemplos, podem ser citadas a seleção de clones e novas cultivares adaptadas às diferentes regiões, a definição de diferentes tecnologias de manejo especialmente para as regiões tropicais e subtropicais e a certificação de produtos vitivinícolas, como produção integrada, indicações geográficas e produção orgânica. Nos anos de 1970, outras uvas americanas, como Seyve Villard 5276 e Couderc 13, deram suporte à expansão do mercado de vinhos brancos, enquanto as cultivares Bordô e Concord tiveram grande aumento de área plantada para a produção de vinhos tintos e suco de uva, respectivamente. O setor vitivinícola brasileiro avançou significativamente nos últimos anos através da produção integrada de uvas finas de mesa, da definição das primeiras indicações geográficas para a produção de vinhos finos e da produção orgânica de uva, vinho e suco de uva. Os anos de 1990 marcam um novo período na produção de vinhos finos com o início do desenvolvimento de indicações geográficas, a partir de um trabalho da Embrapa Uva e Vinho, estimulando a valorização dos vinhos nacionais através deste instrumento de propriedade industrial (TONIETTO, 1993), cujo amparo legal veio com a entrada em vigor da Lei de Propriedade Industrial (LPI) – número 9.279 (BRASIL, 1996), possibilitando o reconhecimento de proteção jurídica de indicações geográficas brasileiras, seja como Indicação de Procedência, seja como Denominação de Origem (INPI, 2000). No Rio Grande do Sul, que é o maior produtor de uvas do Brasil, a Emater-RS tem acompanhado a produção orgânica de uvas e, de acordo com dados extraoficiais levantados, a área de produção de uva orgânica no Estado passou de 90 ha, em 2005, para 517 ha, em 2011.

149762 — **Avaliação de estratégias com agroquímicos no controle de *Bemisia argentifolii* Bellows & Perring (Hemiptera: Aleyrodidae) em tomate** — As estratégias utilizadas para o controle da mosca branca não afetaram a porcentagem de frutos com amadurecimento irregular, nem os resultados para o °Brix e acidez nos híbridos de tomate Heil'iz-2710 e Gem Pride. No híbrido Heinz-2710, a relação custo/benefício variou de R\$ 1,11 (estratégia 2) a R\$ 1,38 (estratégia 4), e com Gem Pride a variação foi de R\$ 1,74 (estratégia 1) a R\$ 3,33 (estratégia 2), indicando que há maior retorno financeiro com o híbrido Gem Pride (estratégia 2), em relação ao Heinz-2710 (estratégia 4). Este trabalho teve como objetivo avaliar estratégias de uso de agroquímicos no controle da mosca branca, *B. argentifolii*, em tomate e seus efeitos sobre o amadurecimento irregular dos frutos, °Brix, acidez, produtividade e relação custo/benefício. Acefato + Buprofezím + espalhante adesivo Deltametrina + Triazofós + espalhante adesivo Lam bdacialotrina + Piriproxifem + espalhante adesivo Deltametrina + Triazofós + espalhante adesivo Fempropatrina + Metamidofós + espalhante adesivo Buprofezím + Metamidofós + espalhante adesivo Acafata + Fempropatrina + espalhante adesivo Lambdacialotrina + Piriproximifem + espalhante adesivo Deltametrina + Triazofós + espalhante adesivo Lambdacialotrina + Piriproximifem + espalhante adesivo Deltametrina + Triazofós + espalhante adesivo Detergente neutro

974846 — **Características micromorfológicas de biótipos de capim-arroz resistente e suscetível ao quinclorac** — Diferenças morfofisiológicas entre biótipos de azevém resistente e suscetível ao herbicida glifosato, por exemplo, resultaram em menor taxa de transporte de água das raízes à parte aérea de plantas do biótipo resistente (Concenço et al., 2007). Na lâmina foliar, foram determinadas as seguintes características: espessura da epiderme adaxial (EED), espessura da epiderme abaxial (EEB), espessura da lâmina foliar (ELF), diâmetro do feixe central (DFC), diâmetro do feixe secundário (DFS), distância entre feixes (DEF), porcentagem de epiderme adaxial (%EPD), porcentagem de epiderme abaxial (%EPB), porcentagem de células buliformes (%CBL), porcentagem de bainha do feixe vascular (%BFV), porcentagem de feixes vasculares (%FXV), porcentagem de esclerênquima (%ESC) e porcentagem de mesofilo (%MES). Seção transversal de raízes de biótipos de capim arroz resistente (A) e suscetível (B) sem herbicida, e resistente (C) e suscetível (D) com herbicida quinclorac. Valores médios de espessura da epiderme adaxial (EED), espessura da epiderme abaxial (EEB), espessura da lâmina foliar (ELF), diâmetro do feixe central (DFC), diâmetro do feixe secundário (DFS), distância entre feixes (DEF) e porcentagem de epiderme adaxial (%EPD), de epiderme abaxial

(%EPB), de células buliformes (%CBL), de bainha do feixe vascular (%BFV), de esclerênquima (%ESC) e de mesofilo (%MES) das folhas de biótipos de capim-arroz resistente (R) e suscetível (S) ao quinclorac, sem (SH) e com (CH) aplicação de herbicida(1). Valores médios de diâmetro da raiz (DAR), espessura do córtex ou aerênquima (ECX), espessura dos feixes vasculares (EAF), diâmetro da medula (DAM) e de percentagem de epiderme (%EPI), de córtex ou aerênquima (%CTA), de feixes vasculares (%FVS) e de medula (%MED) das raízes de biótipos de capim-arroz resistente (R) e suscetível (S) ao quinclorac sem (SH) ou com (CH) aplicação do herbicida(1).

315464 — Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético — Posteriormente, foi proposto um novo método de seleção denominado seleção genômica ampla (genome wide selection – GWS), o qual apresenta alta acurácia seletiva para a seleção, baseada exclusivamente em marcadores, após terem seus efeitos genéticos estimados a partir de dados fenotípicos em uma amostra da população de seleção. (2001) propuseram um novo método de seleção denominado seleção genômica (GS) ou seleção genômica ampla (genome wide selection – GWS), a qual pode ser aplicada em todas as famílias em avaliação nos programas de melhoramento genético de espécies alógamas, apresenta alta acurácia seletiva para a seleção baseada exclusivamente em marcadores (após terem seus efeitos genéticos estimados a partir de dados fenotípicos em uma amostra da população de seleção) A partir dos dados fenotípicos gerados, foram estimados os efeitos de cada um dos 30 locos marcadores por meio do procedimento BLUP/GWS e esses efeitos foram somados para fornecer os valores genéticos genômicos preditos. 2 em que rmq é a proporção da variação genética explicada pelos marcadores, rm \hat{m} é a acurácia da predição dos efeitos dos marcadores que estão em desequilíbrio de ligação com os QTL's e 2 h m^2 \square (\square g^2 rmq / n) / (\square g^2 rmq \square \square e^2). Esse benefício só se concretizará se houver um alto grau de desequilíbrio de ligação envolvendo marcadores SNPs proximamente espaçados e se os efeitos genéticos dos marcadores nas características sob melhoramento forem estimados (a partir dos fenótipos) com alta acurácia e usados na própria população e ambientes em que forem estimados.

154151 — Variabilidade genética de procedências e progênes de umbuzeiro via metodologia de modelos lineares mistos (REML/BLUP) — O presente trabalho teve como objetivos estudar a variabilidade genética, estimar parâmetros genéticos e realizar a predição de valores genéticos dos indivíduos, utilizando a metodologia REML/BLUP a partir da avaliação de procedências e progênes de umbuzeiro. the use of the mixed model methodology trough the REML/BLUP (restricted maximum likelihood/best linear unbiased prediction) procedure is an essential feature for selecting individuals based on their predicted genetic additive and genotypic values, at the intra and inter population levels. This paper aimed to study the genetic variability, estimating genetic parameters and predicting individual breeding values in the umbu tree, through the REML/BLUP procedure applied in the evaluation of populations and progenies. Segundo Resende (2000), no melhoramento de espécies perenes, o uso de técnicas de avaliação genética, com base em modelos mistos do tipo REML/BLUP (máxima verossimilhança restrita/melhor predição linear não viciada), são fundamentais para a predição de valores genéticos aditivos e genotípicos de indivíduos com potencial para 1 (Trabalho 096/2003). O presente trabalho teve como objetivos estudar a variabilidade genética, estimar parâmetros genéticos e realizar a predição de valores genéticos dos indivíduos para fins de seleção, utilizando a metodologia REML/BLUP a partir da avaliação de procedências e progênes de umbuzeiro.

314160 — Recuperação ambiental em áreas de estepe do Primeiro Planalto Paranaense, mediante plantio de espécies arbóreas — Para a efetivação de ações de recuperação ambiental, existem diferentes técnicas e entre essas podem ser citados os plantios de espécies arbóreas, que é a mais difundida (KAGEYAMA et al., 2003; CARPANEZZI, 2005), a semeadura direta (ARAKI, 2005), a instalação de poleiros (ALMEIDA, 2000), a transposição de camada de solo (REIS et al., 2003), a coleta de chuva de sementes (ALMEIDA, 2000) ou mesmo o simples isolamento da área (RODRIGUES; GANDOLFI, 2000). Tais plantios, realizados entre os anos de 2003 e 2006, numa área de 20 ha localizada nas margens da Represa do Iraí (em aproximadamente 25 % da área de entorno), foram feitos por meio da parceria instituída entre a Companhia de Saneamento do Paraná (Sanepar), a Embrapa Florestas e a Universidade Federal do Paraná (UFPR), que apresentou como primeiro objetivo a adequação da área de entorno da Represa do Iraí aos termos da legislação ambiental, que

imagens aéreas e produtos derivados, como modelagem 3D e mapas de hipsometria e exposição/cobertura do solo (D'OLEIRE-OLTMANN et al., 2012; PEREZ e GARCIA, 2017; KRENZ e KUHN, 2018; ERCOLE, 2019; , 2017; KRENZ e KUHN, 2018; ERCOLE, 2019; LOUREIRO, 2019; KRENZ et al., 2019). SPAIRANI-BERRIO et al., 2019; ZEIDAN et al., 2019)

103355 — **A psicometria linear da escalagem ordinal: uma aplicação na caracterização da importância relativa de atividades de produção em ciência e tecnologia** — Neste artigo nossa intenção é mostrar, através de uma aplicação de interesse estratégico para a Embrapa Cerrados, como o método de máxima verossimilhança pode ser utilizado na análise de dados gerados pela versão linear da Lei dos Julgamentos Categóricos e indicar como a mesma pode ser colocada num contexto semelhante ao estudado por Grizzle et al. Estes indicadores são classificados em quatro categorias principais de produção: produção técnico científica (publicação de artigos em periódicos indexados, de artigos em periódicos nãoindexados, de capítulos de livros técnico-científicos, de artigos em anais de congresso e de resumos em anais de congresso e a orientação de teses de pósgraduação), produção de publicações técnicas (publicação de circulares técnicas, de comunicados, instruções e recomendações técnicas, de boletins de pesquisa, de periódicos, organização e editoração de livros e publicação de artigos na mídia), desenvolvimento de tecnologias, produtos e processos (lançamento e geração de cultivares, teste e recomendação de cultivares, evento elite, geração de raças, práticas e processos agropecuários, insumos agropecuários, processos agroindustriais, metodologias científicas, produção de máquinas, equipamentos e instalações, estirpes, monitoramento e zoneamento, produção de *software* e manutenção de bases de dados) e transferência de tecnologia e promoção de imagem (dias de campo, organização de eventos, palestras, participação em exposições e feiras, cursos oferecidos, estágios de nível médio, estágios de graduação, estágios de pós-graduação, produção de folders, produção de vídeos, unidades demonstrativas e de observação e reportagens tecnológicas). A análise da eficiência de produção de cada centro de pesquisa bem como o estabelecimento de suas metas de produção, no contexto do modelo da Embrapa, é determinada por um conjunto de indicadores quantitativos associados as variáveis de produção do sistema. Cerrados com os de P&D da Embrapa, uma vez que é de interesse gerencial da Embrapa valorizar também as outras categorias, especialmente num centro como a Embrapa Cerrados, onde as atividades de desenvolvimento e de transferência de tecnologia são fundamentais. Note que, na parte de entrada de dados, seguindo a instrução "cards", as categorias de produção são representadas por a (produção técnico científica), b (produção de publicações técnicas), c (desenvolvimento de tecnologias produtos e processos) e d (transferência de tecnologia e produção de imagem)

160276 — **Variabilidade genética para o período de vernalização em diferentes populações de cebola no Nordeste brasileiro.** O porcentual de bulbos brotados aos dias foi superior para os produzidos em Mandacaru para as populações cascuda Roxa , Alfa São Francisco, Franciscana IPA , ValeOuro IPA , Brisa IPA , enquanto as populações cascuda Roxa , The percent value of growing bulbs after days was superior to the bulbs produced at Mandacaru to the population purple Valenciana type , Alfa São Francisco, Franciscana IPA , ValeOuro IPA , Brisa IPA , while the population purple Valenciana type , yellow Valenciana type , yellow Valenciana type , Granex and TEG PRR did not present changes in function of the local of production. O porcentual de bulbos brotados foi superior para os bulbos produzidos em Mandacaru para as populações cascuda Roxa , Alfa São Francisco, Franciscana IPA , ValeOuro IPA , Brisa IPA , enquanto para as demais populações não houve grandes diferenças em função do local de produção (Tabela). As populações experimentais cascuda Roxa , Cascuda Roxa , Cascuda amarela e Cascuda Amarela , resultante de seleções dentro de um cruzamento de Valcatorce x Baía Periforme, bem como as populações do grupo Grano, apresentaram menores porcentuais de bulbos brotados quando comparados com as demais populações avaliadas, indicando que as mesmas precisam de maior tempo de armazenamento em câmara fria e/ou de seleções para redução do período de vernalização.

FONTE: ELABORADO PELO AUTOR.

APÊNDICE D – TAMANHO DOS CLUSTERS.

Termo	Cluster
<i>Zea Mays</i>	205
Solo	201
Milho	185
Produtividade	162
Feijão	148
Irrigação	140
Semente	139
Variedade	133
Ovino	130
Caprino	125
Agricultura Familiar	124
<i>Phaseolus Vulgaris</i>	124
Uva	116
Plantio Direto	113
Qualidade	107
Arroz	103
Adubação	101
<i>Oryza Sativa</i>	100
Produção	99
Cerrado	98
Soja	93
Doença	93
Germinação	88
<i>Glycine Max</i>	86
Manejo	85
Doença de Planta	81
Fungo	80
Nitrogênio	79
Caatinga	76
Pastagem	75
<i>Sorghum Bicolor</i>	73
Melhoramento Genético Vegetal	70
Agricultura	67
Pós-Colheita	67
Fruticultura	65
Rendimento	64
Nutrição Animal	62
Controle Biológico	61
Praga de Planta	60
Eucalipto	59
Viticultura	55
<i>Vigna Unquiculata</i>	55
Armazenamento	55

Fonte: Elaborado pelo autor.

APÊNDICE E – ANÁLISE DE ASSUNTOS.

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo (3ApC)	Intersecção
207984	Absorção, Absorção de Nutrientes, Adubação, Calagem, Cerrado, Feijão, Grão, Irrigação, Nitrogênio, Nutriente, Nutriente Mineral, <i>Phaseolus Vulgaris</i> , Rendimento, Solo	Adubação, Calagem, Cerrado, Feijoeiro, Solo, Cultivar, Dose, Grão, Nutriente, Produtividade	[Adubação, Calagem, Cerrado, Solo, Grão, Nutriente]
159841	Absorção, Solo	Cádmio, Solo, Planta, Caatinga, Água, Extração, Metal, Pesquisa, Profundidade, Solução	[Solo]
1074606	<i>Acácia Mearnsii</i> , Crédito agrícola, Risco climático, Seguro rural	Solo, Água, Muda, Temperatura, Clima, <i>Acácia Mearnsii</i> , Agrometeorologia, Dado, Espécie, Geda, Plantio, Profundidade, Risco	[<i>Acácia Mearnsii</i>]
1122509	Acidez do Solo, Latossolo Amarelo, Nutriente, Saturação	Latossolo Amarelo, Químico, Terra, Solo, Acidez do Solo, Agroecossistema, Calcário, Corretivo, Dose, Granulometria, Incubação, Prado, Saturação, Tratamento	[Latossolo amarelo, Acidez do solo, Saturação]
149762	Acidez, Controle Químico, Efeito Residual, Inseto, <i>Lycopersicon Esculentum</i> , Mosca Branca, Praga, Produtividade, Tomate	<i>Lycopersicon Esculentum</i> , Hemiptera, Tomate, Híbrido, Fruto, Meio Ambiente, Custo, Espalhante, Mosca branca, Pesticida, Produtividade, Óleo vegetal	[Tomate, Mosca branca, Produtividade, <i>Lycopersicon Esculentum</i>]
158728	Açúcar, Análise Química, Carboidrato, Fenologia, Folha, Uva, Variedade, Viticultura, <i>Vitis Vinifera</i>	Açúcares, Folha, Ramo, Videira, Temperatura, Amido, Cacho, Carboidrato, Crescimento, Fruto	[Folha, Carboidrato]
134513	<i>Adenantha Pavnina</i> , <i>Azadirachta Indica</i> , Eucalipto, <i>Eucalyptus Cloeziana</i> , Nim	Espécie Florestal, Mata, Planta, Crescimento, Eucalipto, Altura, Desempenho, Diâmetro, Empresa, Espécie, Pesquisa, Umbaúba, Árvore	[Eucalipto]
1128414	Adubação, Análise Foliar, <i>Prunus Persica</i>	Estado Nutricional, Indústria, Pessegueiro, Pêssego, Nutriente, Adubação, Cultura, Deficiência, Folha, Planta, Pomar, Produtor, Solo	[Adubação]
1110441	Aeração, Água, Condutividade Hidráulica, Porosidade, Produção Integrada	Integração, Solo Arenoso, Solo, Água, Porosidade, Areia, Argila, Ciência do solo, Horizonte, Profundidade, Volume	[Água, Porosidade]
1120443	Agricultura Familiar, Agricultura Sustentável	Agricultor, Bem-Estar, Segurança Alimentar, Assentamento, Solo, Curso, Espécie, Família, Lote, Manejo, Pesquisa	[]

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo (3ApC)	Intersecção
1128737	Agricultura Familiar, Desenvolvimento Rural, Distribuição de Renda, Mercado, Políticas Públicas	Agricultor, Mercado, Agricultura Familiar, Desenvolvimento Rural, Produção, Extensionista, Feira, Políticas públicas, Produto, Rede	[Mercado, Agricultura familiar, Desenvolvimento rural, Políticas públicas]
1125800	Agricultura Familiar, Desenvolvimento Rural, Políticas Públicas	Agricultura Familiar, Análise, Políticas Públicas, Inovação, Comercialização, Agricultor, Assistência técnica, Extensionista, Política, Produto, Terra, Território	[Agricultura familiar, Políticas públicas]
133848	Agricultura Familiar, Fruta, Renda, <i>Spondias Tuberosa</i> , Umbu	Imbu, Fruto, Agricultor, Renda, Colheita, Pessoa, Lagoa, produção, Safra, Mão-de-obra, <i>Spondias Tuberosa</i>	[Umbu, Fruta, Renda, <i>Spondias Tuberosa</i>]
1129147	Agricultura Orgânica, Composto Orgânico, Fertilidade, Nitrogênio	Compostagem, Farelo, Gliricidia, Mistura, Trigo, Incubação, Perda, Tratamento, Temperatura, Massa, Agroecologia, Gliricidia	[]
105845	Agricultura Orgânica, Mercado	Agricultor, Agricultura Orgânica, Produção, Agricultura, Alimento, Consumidor, Mercado, Produto, Produtor, Venda	[Agricultura orgânica, Mercado]
119512	Agricultura Sustentável, Fronteira Agrícola, Manejo do Solo	Agricultura, Papel, Pastagem, Solo, Floresta, Agricultura familiar, Café, Lote, Pecuária, Tecnologia	[]
1118211	Agricultura, Desenvolvimento Rural	Produção, Agricultura, Economia, Cadeia Produtiva, Tecnologia, Campo, Capital, História, Terra, Trabalho	[Agricultura]
13919	Agrotóxico, Análise de Risco, Evaporação, Herbicida, Impacto Ambiental, Lixiviação, Modelo Matemático	Composto Orgânico, Risco, Solo, Água, Temperatura, Dado, Folha, Sorção, Velocidade, Volatilização	[Risco]
15046	Agrotóxico, Pulverizador	Agrotóxico, Algodão, Cultura, Pulverização, Planta, Amostragem, Meio ambiente, Perda, Pesticida, Praga, Produto, Tomate	[Agrotóxico]
153839	Alface, <i>Lactuca Sativa</i> , Produção	Alface, Cultivar, Desempenho, Outono, Colheita, Cabeça, Caule, Comprimento, Massa, Planta, Temperatura, <i>Lactuca Sativa</i>	[Alface, <i>Lactuca Sativa</i>]
161390	Algodão, Condição Ambiental, Controle biológico, Percevejo, Predador	<i>Orius Insidiosus</i> , Temperatura, Umidade Relativa, Planta, Algodão, Algodoeiro, Criação, Hemíptera, Ninfas, Oliveira, Predador, Pulgão, Sobrevivência	[Algodão, Predador]
156709	Alimentação, Caatinga, Caprino, Umbu	Animal, Fruto, Peso, Umbu, Semente, Caprino, Consumo, Safra, Folha, Caatinga, <i>Spondias Tuberosa</i>	[Caatinga, Umbu Caprino]

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo (3ApC)	Intersecção
1129161	Alimento Alternativo, Biomassa, Forragem, <i>Moringa Oleifera</i>	Alimentação Animal, <i>Moringa Oleifera</i> , Folha, Produção, Planta, Animal, Dieta, Espécie, Moringa, Semente	[<i>Moringa oleifera</i>]
154151	<i>Anacardiaceae</i> , Melhoramento Genético Vegetal, <i>Spondias Tuberosa</i> , Umbu	Genética, Umbu, Progênie, Melhoramento, Seleção, Fruto, Diâmetro, Espécie, Herdabilidade, Planta, População, <i>Spondias Tuberosa</i>	[<i>Spondias Tuberosa</i> , Umbu]
103355	Análise de Dados, Pesquisa	Produção, Tecnologia, Cerrado, Transferência de Tecnologia, Gestão, Dado, Distribuição, Julgamento, Método, Peso, Pesquisa, Pesquisador, Universidade	[Pesquisa]
211707	Antracnose, <i>Colletotrichum Lindemuthianum</i> , Feijão, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i> , Produtividade	Antracnose, Método, Rendimento, Resistência, Seleção, Cruzamento, Família, Genética, Linhagem, Planta, População, Produtividade, <i>Colletotrichum Lindemuthianum</i> , <i>Phaseolus Vulgaris</i>	[Antracnose, <i>Colletotrichum Lindemuthianum</i> , Produtividade, <i>Phaseolus Vulgaris</i>]
1118707	<i>Aphis Gossypii</i> , Silício	Algodão, Cultivar, Resistência, Planta, Inseto, Algodoeiro, Fenol, Pulgão, Silício	[Silício]
1008415	Armazenamento, <i>Capsicum annum</i> , Pimentão, Preservação de alimento, Produção orgânica	Pimentão, Qualidade, Refrigeração, Fruto, Hortaliça, Agroecologia, Armazenamento, Casca, Coloração, Comprimento, Cultivar, Solido	[Pimentão, Armazenamento]
212670	Armazenamento, Feijão, Fungicida, <i>Phaseolus Vulgaris</i> , Revestimento, Semente	Armazenamento, Feijão, Fungicida, Semente, Patógeno, Fungo, Germinação, Inoculação, Revestimento, <i>Phaseolus Vulgaris</i> , Tratamento	[Armazenamento, Feijão, Fungicida, Semente, <i>Phaseolus Vulgaris</i> , Revestimento]
1127555	Bacia Hidrográfica, Precipitação Pluvial, Vazão	Bacia Hidrográfica, Terra, Água, Floresta, Gestão, Animal, Análise, Dado, Redução, Universidade, Vazão	[Bacia hidrográfica, Vazão]
151173	Balanco de Energia, Banana, Evapotranspiração	Bananeira, Consumo, Método, Vale, Evapotranspiração, Colheita, Cultura, Dado, Água	[Evapotranspiração]
1129606	Biogás, Níquel, Oxido	Biogás, Metal, Oxido, Terra, Carbono, Catalisador	[Biogás, Oxido]
1123970	<i>Bixa Orellana</i> , Metal, Urucum	Folha, Químico, Planta, Planta Medicinal, Adubação, <i>Bixa orellana</i> , Dosagem, Espécie, Ingestão, Metal, Ponto	[<i>Bixa orellana</i> , Metal]
975131	Caatinga, Desertificação, Recurso natural, Sensoriamento Remoto	Dado, Floresta, Terra, Vegetação, Solo, Caatinga, Florestal, Mapa, Mata ciliar, Uso da terra, Água	[Caatinga]

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo (3ApC)	Intersecção
1110938	Cacau, Doença, Manejo, Praga, <i>Theobroma Cacao</i>	Cacau, Doença, Inimigo Natural, Inseto, Produção, Roça, <i>Theobroma Cacao</i>	[Cacau, Doença, <i>Theobroma Cacao</i>]
125759	Cádmio, Eucalyptus Camaldulensis, Poluição do Solo, Solo, Zinco	Cálcio, Eucalipto, Metal Pesado, Muda, Toxidez, Contaminação, Dose, Metal, Parte aérea, Planta, Solo	[Solo]
975338	Cana de Açúcar, Chuva, Herbicida, Lixiviação	Açúcar, Metribuzin, Palha, Chuva, Herbicida, Dado, Daninha, Produto, Pulverização, Simulação, Solo, Tratamento, Vazão	[Chuva, Herbicida]
1102669	Características Agronômicas, Maracujá, Melhoramento Genético Vegetal, Parâmetro Genético	Genética, Folha, Cerrado, Cruzamento, Maracujá, Antocianina, Coloração, Espécie, Flor, Peciolo, Ponto, Trabalho	[Maracujá]
1074877	Casca de arroz, Espécie nativa, Guabiroba, Micorriza vesicular arbuscular, Substrato de cultura	Crescimento, Fertilização, Micorriza, Muda, Adubação, Casca de arroz, Espécie, Fungo, Inoculação, Planta, Produção, Qualidade	[Casca de arroz]
1120755	Certificação de Produto, Suco, Vinho	Vinho, Produção, Viticultura, Manejo, Mercado, Cultivar, Produto, Produtor, Suco, Vale, Videira	[Vinho, Suco]
1124357	Cerveja, Cevada, Malte, Proteína	Agricultura, Cerrado, Forragem, Fungo, Grão, Zootecnia	[]
280598	Chuva, Doença, Temperatura	Seca, Temperatura, Chuva, Inseto, Planta, Dado, Etileno, Florestal, Haste, Sintoma, Árvore	[Temperatura, Chuva]
1122175	Clima, Diâmetro, Paricá, <i>Schizolobium Parahyba</i>	Clima, Paricá, Plantio, Crescimento, Planta, Curaua, Diâmetro, Espécie, Freijó, Sistema de cultivo, Temperatura, Árvore	[Clima, Paricá, Diâmetro]
1128271	Clima, Geada, Mudança Climática, Viticultura	Dado, Geada, Prevenção, Viticultura, Temperatura	[Geada, Viticultura]
1017311	Clonagem, Propagação Vegetativa	Estação do Ano, Matriz, Enraizamento, Planta, Progênie, Brotacao, Campo, Estaca, Floresta, Primavera, Produção	[]
313756	Cobertura do Solo, Solo	Cultura, Desempenho, Inverno, Semeadura, Sulcador, Emergência, Resíduo, Solo	[Solo]
157560	Confinamento, Consumo, Conversão Alimentar, Ganho de Peso, Ovino, Terminação	Confinamento, Desempenho, Ovino, Resíduo, Dieta, Alimentação, Animal, Anta, Idade, Iodo, Mandioca, Milho	[Confinamento, Ovino]
314079	Conservação, Endogamia, Germoplasma	Espécie Florestal, Conservação, Conservação, Semente, Genética, Amostragem, Espécie, Florestal, Matriz, Plantio, População	[Conservação]

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo (3ApC)	Intersecção
1125816	Consumo Alimentar, Culinária	Alimentação, Alimento, Mandioca, Alho, Floresta, Cidade, Cultura, Família, Maniçoba, Pato, Pesquisa, Sabor, Tucupi	[]
280844	Controle Biológico, Doença, <i>Ilex Paraguariensis</i> , <i>Trichoderma</i>	Seleção, Patógeno, Folha, Bactéria, Germinação, Colônia, Conídio, Doença, Fungo, Muda, <i>Trichoderma</i>	[Doença, <i>Trichoderma</i>]
1126373	Controle Químico, Doença de Planta, Enfezamento, Resistência Genética, Zea Mays	Doença, Milho, Sorgo, Híbrido, Manejo, Cota, Enfezamento, Fungicida, Lagoa, Produto	[Enfezamento]
134128	<i>Cucumis Melo</i> , Dióxido de Carbono, Fertirrigação, Fotossíntese, Irrigação, Melão	Produtividade, Água de Irrigação, Fruto, Solo, Planta, Colheita, Cultura, Folha, Melão, Qualidade, Tratamento	[Melão]
1121184	Deficiência Hídrica, Evapotranspiração, Soja	Análise Comparativa, Balanço Hídrico, Soja, Vegetação, Planta, Cultura, Dado, Deficiência hídrica, Safra	[Soja, Deficiência hídrica]
1118305	Desmatamento	Desmatamento, Geoprocessamento, Produção, Agricultura, Assentamento, Análise, Censo agropecuário, Dado, Florestal, Imóvel rural, Pesquisa, Produtor, Terra	[Desmatamento]
134397	Doença, <i>Lycopersicon Esculentum</i> , Tomate	Tomateiro, Vírus, Planta, Tomate, Produção, Cultura, Infecção, Mosca branca, Plantio, Redução, Sintoma	[Tomate]
974846	<i>Echinochloa Frumentacea</i> , <i>Oryza sativa</i>	Biotipo, Herbicida, Raiz, Planta, Folha, Célula, Daninha, Diâmetro, Epiderme, Espessura, Tratamento	[]
312782	Enraizamento, <i>Ficus Carica</i> , Figo, Regulador de Crescimento	Estaca, Ficus Carica, Figueira, Folha, Fruto, Enraizamento, Matéria seca, Peso, Planta, Raiz, Ácido indolbutírico	[Ficus carica, Enraizamento]
962648	Estatística agrícola, Petróleo, Preço internacional, Produto, Ureia	Petróleo, Preço, Milho, Alimento, Arroz, Decomposição, Inovação, Soja, Trigo, Ureia	[Petróleo]
314117	Esterco de Caprino, Produção, Resíduo Industrial, Resíduo Orgânico	Espécie Florestal, Muda, Produção, Resíduo Orgânico, Resíduo, Bagaço, Casca, Esterco de caprino, Florestal, Mistura, Serragem, Água	[Produção, Resíduo orgânico, Esterco de caprino]
958618	Estirpe, Segurança alimentar, <i>Staphylococcus aureus</i>	Epidemiologia, Estirpe, Produção, Queijo, Leite, Colheita, Estocagem, Gene, Produto, Tanque, Toxina	[Estirpe]
204144	Feijão, Melhoramento Genético Vegetal, Nitrogênio, <i>Phaseolus Vulgaris</i>	Feijoeiro, Linhagem, Nitrogênio, Produtividade, Grão, Adubação, Cultura, Nutriente, Safra, Solo	[Nitrogênio]

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo (3ApC)	Intersecção
215848	Feijão, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i>	Feijoeiro, Melhoramento, Feijão, Resistência, Patógeno, Cultivar, Dado, Estimativa, Linhagem, Safra	[Feijão, <i>Phaseolus Vulgaris</i>]
214593	Feijão, <i>Phaseolus Vulgaris</i> , Produção, Qualidade, Semente	Lavoura, Semente, Sustentabilidade, Feijão, Fungicida, Antracnose, Arroz, Cultivar, Patógeno, <i>Phaseolus Vulgaris</i> , Planta, Produção, Solo, Várzea	[Semente, Feijão, <i>Phaseolus Vulgaris</i> , Produção]
1015548	Fenologia, Pêra, <i>Pyrus Communis</i>	Clima, Safra, Produção, Planta, Fruto, Cultivar, Floração, Frutificação, Gema, Oliveira, Produtividade, Temperatura	[]
315462	Fertilidade do Solo	Biomassa, Cambissolo, Cinza, Solo, Muda, Caldeira, Dose, Planta	[]
1008316	Fertilizante fosfatado, Nutrição vegetal, Plantio direto	Cultura, Fósforo, Soja, Solo, Produtividade, Dose, Massa, Matéria seca, Planta, Raiz	[]
955737	Floração, Fruticultura, Pêssego, Produção, Produtividade, Variedade	Floração, Produção, Pêssego, Qualidade, Seleção, Cultivar, Fruto, Pessegueiro, Planta, Produtividade	[Floração, Produção, Pêssego, Produtividade]
314145	Fogo	Floresta, Fogo, Muda, Semente, Crescimento, Araucária, Dano, Espécie, Imbuia, Pinhão, Plantio	[Fogo]
1122380	Fornagem, Ruminante	Características Agronômicas, Corte, Genótipo, Mutante, Pastejo, Gene, Matéria seca, Planta, Produção, Sorgo	[]
17098	Fosfato	Adsorção, Argila, Fósforo, Solo, Carbono, Ferro, Fosfato, Horizonte, Método, Rocha, Universidade	[Fosfato]
961303	Germinação, Palmito, Semente	Crescimento, Germinação, Plântula, Semeadura, Semente, Velocidade	[Germinação, Semente]
314935	<i>Hevea Brasiliensis</i> , Seringueira	Latex, Produção, Seleção, Seringueira, Genética, Altura, Borracha, Hevea, <i>Hevea Brasiliensis</i> , Planta, Progenie	[Seringueira, <i>Hevea Brasiliensis</i>]
162329	<i>Lactuca Sativa</i> , Produtividade, Variedade	Alface, Dose, Inverno, Molibdênio, Época de Aplicação, Caule, Comprimento, Cultura, <i>Lactuca Sativa</i> , Planta, Produtividade, Transplântio	[<i>Lactuca Sativa</i> , Produtividade]
119110	Legislação, Leite Pasteurizado	Leite, Qualidade, Produção, Consumo, Mercado, Leite pasteurizado, Produto, Produtor, Tanque, Transporte	[Leite pasteurizado]

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo (3ApC)	Intersecção
315464	Marcador Molecular	Melhoramento, Seleção, Variação Genética, Genótipo, Genética, Alelos, Dado, Distribuição, Herdabilidade, Método, População, Zootecnia	[]
1016654	Milho Doce, Pós-Colheita, Processamento Mínimo, Qualidade, Zea Mays	Atmosfera, Milho Doce, Qualidade, Armazenamento, Conservação, Carotenóide, Massa, Milho, Perda, Produto	[Milho doce, Qualidade]
57793	Milho, Variedade	Cultivar, Milho, Híbrido, Produtividade, Produção, Análise de variância, Carvalho, Estimativa, Genética, Genótipo, População, Trabalho	[Milho]
313801	Muda, Plântula, Radiação Solar	Clorofila, Crescimento, Sombreamento, Muda, Planta, Diâmetro, Espécie, Folha, Matéria seca, Ponto	[Muda]
1129435	Mudança Climática, Soja	Cultura, Mudança Climática, Rendimento, Simulação, Produtividade, Anomalia, Clima, Dado, Método, Soja, Temperatura	[Mudança climática, Soja]
1122642	Peroxidase, Polifenoxidase	Conservação, Filtração, Membrana, Água de Coco, Processamento, Fruto, Peroxidase, Produto, Proteína, Redução, Tamanho	[Peroxidase]
1126835	Piscicultura	Tambaqui, Tanque, Resistência, Peixe, Bactéria, Animal, Antibiótico, Fazenda, Florfenicol, Lesão, Penicilina, Piscicultura	[Piscicultura]
313739	Procedência	Liquidâmbar, Madeira, Produção, Crescimento, Semente, Espécie, <i>Liquidâmbar styraciflua</i> , Povoamento, Produtividade, Solo	[]
121733	Produtividade, Risco, Sistema de cultivo	Análise de Risco, Sistema de Produção, Produtividade, Custo, Feijão, Distribuição, Preço, Vale, Valor da produção, Venda	[Produtividade]
89545	Segurança Alimentar	Alimento, Soja, Tecnologia, Qualidade, Grão, Consumidor, Empresa, Produto, Rastreabilidade, Rotulagem	[]
957259	Sensoriamento remoto	Geoprocessamento, Gestão, Litoral, Sensoriamento Remoto, Solo, Campo, Dado, Drenagem, Planície, Relevo, Rocha, Terraço, Terreno, Água	[Sensoriamento remoto]
89928	Sistema de Produção	Fazenda, Tomada de Decisão, Custo, Leite, Produção, Cerrado, Madeira, Pecuária, Produtor, Receita, Tecnologia, Venda	[]

Identificador	Descritores da BDPA	Assuntos relevantes selecionados pelo (3ApC)	Intersecção
1129319	Tecnologia Agrícola	Análise, Café, Dado, Produção, Produtividade, Cultura, Produtor, Qualidade, Rendimento, Tecnologia	[]
314112	Temperatura	Espécie, Floresta, Crescimento, Temperatura, Conservação	[Temperatura]
1123933	Trigo	Trigo, Produção, Exportação, Qualidade, Alimento, Cultivar, Demanda, Grão, Perda, Produtor, Tecnologia	[Trigo]
160276	<i>Allium Cepa</i> , Cebola	<i>Allium Cepa</i> , Bulbo, População, Bebedouro, Vernalização, Cebola, Produção, Produção de Sementes, Juá, Umidade Relativa, Variedade, Semente	[<i>Allium Cepa</i> , Cebola]

Fonte: Elaborado pelo autor.

APÊNDICE F – TERMOS COCORRENTES NOS ÍNDICES.

Termo	Termo coocorrente	Coocorrência (%)
Mangifera Indica	Manga	97
PH	Feijão	94
Phaseolus Vulgaris	Feijão	93
Vitis Vinifera	Uva	93
Melancia	Citrullus Lanatus	93
Allium Cepa	Cebola	92
Cenoura	Daucus Carota	92
Batata	Solanum Tuberosum	91
Spondias Tuberosa	Umbu	91
Pupunha	Bactris Gasipaes	89
Cucumis Melo	Melão	88
Alimento	Ovino	88
Arroz	Oryza Sativa	87
Viticultura	Uva	87
Psidium Guajava	Goiaba	86
Allium Sativum	Alho	86
Lactuca Sativa	Alface	85
Competição De Variedade	Produtividade	85
Arroz Irrigado	Oryza Sativa	83
Umbu	Spondias Tuberosa	83
Oryza Sativa	Arroz	80
Alho	Allium Sativum	80
Cebola	Allium Cepa	79
Citrullus Lanatus	Melancia	79
Feijão	Phaseolus Vulgaris	78
Lycopersicon Esculentum	Tomate	78
Solanum Tuberosum	Batata	77
Melão	Cucumis Melo	76
Alface	Lactuca Sativa	76
Bactris Gasipaes	Pupunha	76
Alimento	Nutrição Animal	76
Diagnóstico	Caprino	75
Tambaqui	Colossoma Macropomum	75
Alimento Para Animal	Ovino	75
Doença Animal	Caprino	73
Leite De Cabra	Caprino	73
Cupuaçu	Theobroma Grandiflorum	73
Cordeiro	Ovino	71
Goiaba	Psidium Guajava	69
Palmito	Bactris Gasipaes	69
Diagnóstico	Doença Animal	67
Alimento Para Animal	Nutrição Animal	67
Maçã	Fruticultura	65
Euterpe Oleracea	Açaí	65
Mamona	Ricinus Communis	64

Brusone	Arroz	64
Brusone	Oryza Sativa	64
Lagarta	Praga de Planta	64
Milho	Zea Mays	62
Água	Irrigação	62
Balanço de Energia	Evapotranspiração	62
Theobroma Grandiflorum	Cupuaçu	62
Bulbo	Cebola	62
Palmito	Pupunha	62
Germinação	Semente	61
Colossoma Macropomum	Tambaqui	60
Vinho	Uva	59
Vinho	Enologia	59
Potássio	Nitrogênio	59
Fertilizante	Adubação	58
Anticorpo	Doença Animal	58
Pastejo	Ovino	58
Época de Plantio	Produtividade	58
Rotação de Cultura	Plantio Direto	57
Fertirrigação	Irrigação	56
Confinamento	Ovino	55
Manga	Mangifera Indica	54
Recurso Natural	Caatinga	54
Planta de Cobertura	Plantio Direto	54
Híbrido	Variedade	54
Fertilizante	Nitrogênio	54
Bulbo	Rendimento	54
Abelha	Polinização	54
Zea Mays	Milho	53
Manihot Esculenta	Mandioca	53
Pennisetum Purpureum	Capim Elefante	53
Alho	Produtividade	53
Ricinus Communis	Mamona	52
Açaí	Euterpe Oleracea	52
Enologia	Vinho	52
Brachiaria	Pastagem	52

Fonte: Elaborado pelo autor.

APÊNDICE G – MODELO SEQUENCIAL.

Na inteligência artificial, uma Rede Neural Artificial (RNA) não considera a influência da ordem em que os eventos ocorrem para estimar um fenômeno, de forma que as entradas são tratadas como episódicas. No entanto, séries temporais e outros dados sequenciais descrevem fenômenos em que os eventos anteriores influenciam os próximos eventos e, para capturar e estimar esse comportamento, é que se desenvolveu a Rede Neural Artificial Recorrente (RNN).

Uma RNN pode ser definida como uma classe de Rede Neural Artificial, em que as conexões entre os nós simulam o comportamento dinâmico adequado ao tratamento de dados sequenciais (RUMELHART; HINTON; WILLIAMS, 1986)¹⁸. Ela simula a influência da sequencialidade na produção dos fenômenos. Quanto a isso, pode-se afirmar que a inteligência artificial tradicional assume tacitamente que todos os dados são retirados de alguma distribuição independente e idêntica (IID). Mas isso não é verdade para todos os dados. Por exemplo: em um texto as palavras estão escritas em uma sequência e seria muito difícil decifrar o significado do texto se elas fossem permutadas aleatoriamente. Portanto, é razoável supor que modelos sequenciais terão um desempenho melhor ao descrever dados textuais (ZHANG *et al.*, 2021).

Os modelos sequenciais são do tipo Aprendizado Artificial Supervisionado e, por isso, a sua fragilidade é a necessidade de dados classificados manualmente, pois rotular dados é uma atividade cara e demorada.

A eficiência dos modelos sequenciais no processamento de dados textuais em que o tamanho da sequência de entrada é sempre maior que o tamanho da sequência de saída pode ser melhorada ao se adicionar uma RNA interna, denominada de “mecanismo de atenção”, que “aprende” quais subsequências da entrada têm maior influência na resposta.

Nesta pesquisa, adotou-se o modelo sequencial (ou sequência-a-sequência) com mecanismo de atenção para fazer previsões dos termos do índice de assuntos dos artigos científicos. Ele requer os textos e a sua lista de termos de indexação para aprender, sem nenhum modelo de suposições teóricas para explicar como ele produz a indexação. Ele simplesmente pega o texto na entrada, calcula a saída, depois compara a própria resposta com o índice de assunto pronto. Repete esse processo até sua configuração interna fazer previsões acuradas. Ele tira vantagem do poder computacional ao realizar um número

¹⁸ RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, out. 1986.
DOI: <https://doi.org/10.1038/323533a0>

humanamente incontável de tentativas e correções. Nesse processo, o modelo descobre os valores para milhões de parâmetros capazes de produzir um índice de assunto a partir dos termos do texto.

Em síntese, os modelos sequenciais estão cientificamente consolidados, não precisam de aportes teóricos extras e estão disponíveis como *software* livre. Portanto, os modelos treinados não precisam de explicações metodológicas de sua validade, basicamente, é necessário apenas averiguar o desempenho e a generalização do modelo em um conjunto de dados de teste.

Análise dos resultados

Nesta seção, serão analisados os resultados obtidos com o Modelo Sequencial com Mecanismo de Atenção.

Nesta análise, consideraram-se os princípios da Teoria da Conservação da Informação Computacional (CICT) de Fiorini (2016) sobre o fato de que o conteúdo completo das informações possui acoplados um componente linear (desdobrado) e outro não linear (dobrado), de forma que a representação simbólica (linear) deve conservar a informação original (não linear).

O modelo sequencial com mecanismo de atenção configurado para o 3ApC, denominado de AX-1 (Artificial X – modelo 1), utilizou uma arquitetura de redes neurais profunda com 64 camadas intermediárias com 1024 neurônios artificiais em cada uma, o que deu 65536 neurônios no total. Para treinar essa rede neural, foi necessário usar os recursos da nuvem no serviço denominado de *Google Colab Pro*, que deu acesso à TPU (ver Seção 5.2.1).

Cada registro utilizado tinha dois campos: um para a entrada, constituído pelo conjunto de descritores encontrado no texto sob análise e, outro, para a saída, ou rótulo, formado pelos termos do índice de assuntos feito pela BDPA para o artigo da entrada. No Quadro 24, podem ser vistos a sequência de entrada (na verdade, uma amostra dela formada por 19 termos dos 142 presentes no texto) e o índice de assunto do artigo 311853. Frisa-se que cada registro pode ter até 4000 termos na entrada e de 3 a 6 termos na saída e que o conjunto total de treinamento tinha 4078 registros.

Quadro 24 – Um dado utilizado no treinamento do modelo de IA.

Amostra da entrada	Rótulo
... araucaria areiafina areiagrossa argila arvore cafe campo carbono casuarina caule cerca cienciaflorestal clima climatemperado climatropical corte deficienciahidrica desbaste desenvolvimentoflorestal ...	Clima, GrevilleaRobusta

Fonte: Elaborado pelo autor.

No Aprendizado Supervisionado, a rede neural “calcula” o valor da saída e o compara com o rótulo. Se a estimativa não estiver correta, são feitas novas estimativas e comparações, até que se obtenham estimativas corretas, na verdade, até que se obtenham estimativas aceitáveis, considerando uma margem de erro. O mecanismo de atenção é um algoritmo especial (uma rede neural dentro da rede neural principal) que calcula quais partes da entrada contribuem para gerar a saída. Assim, o modelo AX-1 produziu, na saída, uma pequena sequência a partir de uma grande sequência de entrada, mas, internamente, uma parte dos dados de entrada foi ignorada devido ao mecanismo de atenção que apontou as partes não necessárias à solução.

Todo o processo de treinamento artificial do algoritmo ocorreu por meio da linguagem de programação *Python* no *software Jupyter Notebook*, na nuvem. O código-fonte e o modelo AX-1 treinado estão disponíveis em www.github.com/3ApC e são de acesso, cópia, alteração e redistribuição livres.

O modelo AX-1 configurado, em teoria, pode ser generalizado para indexar os dados que ainda não foram indexados, porque a taxa de erro foi menor que 1% no processo de treinamento e no de validação.

Para analisar a efetividade e a capacidade de generalização do modelo AX-1, simulou-se a linha de produção de índices de assunto, por meio do 3ApC, para os artigos do *corpus* do Experimento. Assim, implementaram-se os testes planejados no Projeto.

A avaliação dos resultados foi feita da seguinte forma: o conjunto de termos previsto pelo 3ApC foi cotejado com o índice de assuntos feito pela BDPA.

Duas previsões de índices de assunto feitos pelo 3ApC pelo AX-1 para cada artigo do *corpus* do Experimento e o respectivo índice de assunto da BDPA, podem ser ao final deste APÊNDICE G.

Nos próximos parágrafos, serão apresentados e analisados três casos ilustrativos da previsão de termos de indexação feitos pelo 3ApC comparados aos índices de assunto da BDPA:

- a. conjunto de previsão sem intersecção (disjunção);
- b. conjunto de previsão com intersecção;
- c. conjunto de previsão exatamente igual.

O primeiro caso ilustrativo de previsão de termos de indexação pelo modelo AX-1 está apresentado no Quadro 25. O artigo “Germoplasma de caupi: coleção ativa e de base”, identificado pelo número 195124 na BDPA, ao ser submetido ao modelo de inteligência artificial AX-1, recebeu a seguinte previsão: “feijão, germinação, melhoramento genético

vegetal, *Phaseolus Vulgaris*”, que é claramente distinto do conjunto de termos de indexação da BDPA, que é “Germoplasma, *Vigna unguiculata*”.

Quadro 25 – Previsão disjunta do 3ApC em relação à indexação da BDPA.

Artigo	Previsão	BDPA
195124	Feijão, Germinação, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i>	Germoplasma, <i>Vigna Unguiculata</i>

Fonte: Elaborado pelo autor.

Ainda que o modelo inteligente tenha predito termos diversos daqueles esperados pelo conjunto de termos de indexação feito para esse artigo, o conjunto previsto encaixa no que poderia ser denominado de “adequado para as necessidades dos usuários”, porque o modelo AX-1, sem qualquer informação adicional além do conjunto de descritores do artigo, deduziu corretamente que o Caupi é um tipo de Feijão, e cabe ao indexador, uma vez que se trata de um modelo de inteligência aumentada, decidir endossar ou não essa previsão.

O segundo caso típico está apresentado no Quadro 26 por meio de duas previsões feitas no 3ApC pelo modelo AX-1. Os conjuntos de previsão são parcialmente iguais aos da indexação da BDPA para os artigos 1128737 e 56709. Note que foram previstos menos termos do que há na indexação-alvo para o primeiro artigo e mais termos para o segundo.

Quadro 26 – Previsões parcialmente semelhantes à indexação da BDPA.

Artigo	Previsão	BDPA
1128737	Agricultura Familiar, Desenvolvimento Rural	Agricultura Familiar, Desenvolvimento Rural, Distribuição de Renda, Mercado, Políticas Públicas
56709	Caatinga, Agricultura Familiar, Caprino, Ovino, Leite de Cabra, Umbu	Alimentação, Caatinga, Caprino, Umbu

Fonte: Elaborado pelo autor.

O terceiro caso, apresentado no Quadro 27, mostra uma previsão feita no 3ApC pelo modelo AX-1 exatamente igual à indexação do artigo 215848.

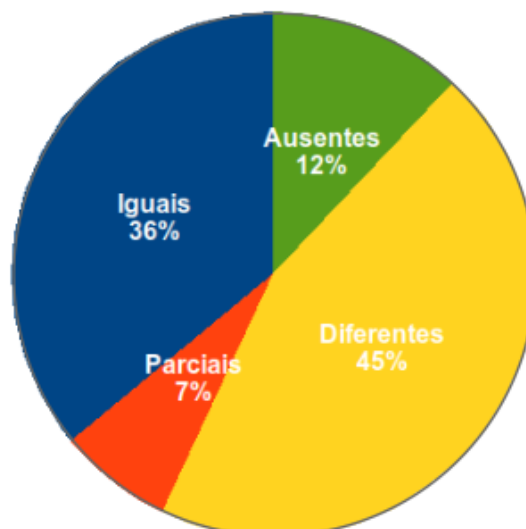
Quadro 27 – Previsão do 3ApC em relação à indexação da BDPA.

Artigo	Previsão	BDPA
215848	Feijão, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i>	Feijão, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i>

Fonte: Elaborado pelo autor.

No geral, o modelo AX-1 não generalizou bem para o *corpus* do experimento, em parte devido a problemas na extração dos textos. As previsões do AX-1 (Figura 19) se mostraram efetivas em apenas 36% das vezes, parcialmente efetivas em 7% das ocasiões, e errou em 45% das previsões. Em 12% das vezes, não foi possível avaliar o método devido à ausência de índices de assuntos para os artigos.

Figura 19 – A proporção das indexações em relação ao padrão-ouro.



Fonte: Elaborado pelo autor.

O modelo não funcionou, apesar de generalizar bem a solução para 36% do conjunto de dados. Os 4078 dados de entrada parecem muitos, humanamente falando, mas representam pouco para o algoritmo que necessita de mais de nove milhões de dados rotulados. Essa é a mesma conclusão de Waltinger *et al.* (2011, p. 38) sobre a classificação hierárquica de Dewey (DDC), de que os resultados mostram uma “distribuição assimétrica” em razão das “questões de escassez de dados”. A escolha de termos para o índice de assunto, na perspectiva algorítmica, apresenta-se como aleatória, devido à quantidade insuficiente de dados em relação ao número de padrões existentes nos dados.

O modelo AX-1 foi eficiente em menos da metade dos casos, o que não o recomenda para a atividade de produção e, portanto, ele não foi incluído no 3ApC.

Assim, pode-se concluir que é possível aplicar esse tipo de modelo à tarefa de indexação de assunto, desde que se consigam milhões de artigos indexados (pelo Thesagro, considerando as condições selecionadas nesta pesquisa).

Quadro dos Resultados obtidos com modelo sequencial.

ID	BDPA	Previsão 1	Previsão 2
207984	Absorção, Absorção de Nutrientes, Adubação, Calagem, Cerrado, Feijão, Grão, Irrigação, Nitrogênio, Nutriente, Nutriente Mineral, <i>Phaseolus Vulgaris</i> , Rendimento, Solo	Feijão, Nitrogênio, <i>Phaseolus Vulgaris</i> , variedade	Feijão, nitrogênio, <i>Phaseolus Vulgaris</i> , plantio direto
159841	Absorção, Solo	Milho, Solo	Milho, Plantio Direto
1074606	<i>Acácia Mearnsii</i> , Crédito agrícola, Risco climático, Seguro rural	<i>Acacia mearnsii</i>	Cartografia, Floresta
1122509	Acidez do Solo, Latossolo Amarelo, Nutriente, Saturação	Adubação, calagem, calcário, cálcio, fertilizante, fósforo, nitrogênio, potássio	Adubação, calagem, fertilizante, fósforo, nitrogênio, nutrição vegetal
149762	Acidez, Controle Químico, Efeito Residual, Inseto, <i>Lycopersicon Esculentum</i> , Mosca Branca, Praga, Produtividade, Tomate	Acidez, fécula, <i>zea Mays</i>	Bebida, nutrição vegetal, silício
158728	Açúcar, Análise Química, Carboidrato, Fenologia, Folha, Uva, Variedade, Viticultura, <i>Vitis Vinifera</i>	Adubação, capim elefante, nutrição vegetal, uva, viticultura	Adubação, fermentação, nitrogênio, uva, viticultura
134513	<i>Adenantha Pavonina</i> , <i>Azadirachta Indica</i> , Eucalipto, <i>Eucalyptus Cloeziana</i> , Nim	<i>Grevillea Robusta</i> , Madeira, Serapilheira, Volume	<i>Grevillea Robusta</i> , Reflorestamento, Semente Florestal, Silvicultura
1128414	Adubação, Análise Foliar, <i>Prunus Persica</i>	Adubação	Adubação
1110441	Aeração, Água, Condutividade Hidráulica, Porosidade, Produção Integrada	Manejo do solo, plantio direto, solo, uso da terra	Manejo do solo, desmatamento, pastagem, matéria orgânica
1120443	Agricultura Familiar, Agricultura Sustentável	Ecossistema, manejo	Ecossistema, modelo de simulação, sistema de cultivo
1128737	Agricultura Familiar, Desenvolvimento Rural, Distribuição de Renda, Mercado, Políticas Públicas	Agricultura Familiar	Agricultura Familiar, Desenvolvimento Rural
1125800	Agricultura Familiar, Desenvolvimento Rural, Políticas Públicas	Agricultura familiar, desenvolvimento rural, integração, políticas públicas, renda, sociologia rural	Agricultura Familiar, Desenvolvimento Rural, Mercado, Política Governamental, Programa de Nutricao
133848	Agricultura Familiar, Fruta, Renda, <i>Spondias Tuberosa</i> , Umbu	Agricultura Familiar, Produção Orgânica, Produtividade, Renda, <i>Spondias Tuberosa</i> , Programa de Nutrição	Agricultura Familiar, Produção Orgânica, Tomada De Decisão

ID	BDPA	Previsão 1	Previsão 2
1129147	Agricultura Orgânica, Composto Orgânico, Fertilidade, Nitrogênio	Adubação, Compostagem, Dióxido de Carbono, Composto Orgânico, Hortaliça, Melão, Nutrição Animal, Ovino	Composto Orgânico, Compostagem, Hortaliça, Planta Medicinal, Zinco
105845	Agricultura Orgânica, Mercado	Comercialização, pós-colheita	Comercialização, fruticultura
119512	Agricultura Sustentável, Fronteira Agrícola, Manejo do Solo	Agricultura Familiar, Hortaliça, Adubação Verde, Cebola, Fósforo, Potássio	Adubação Verde, Cebola, Fósforo, Potássio
1118211	Agricultura, Desenvolvimento Rural	Agricultura, produção	Agricultura, agricultura familiar
13919	Agrotóxico, Análise de Risco, Evaporação, Herbicida, Impacto Ambiental, Lixiviação, Modelo Matemático	Doença, milho	Controle biológico
15046	Agrotóxico, Pulverizador	Microclima, uva, <i>Vitis Vinifera</i>	Microclima, uva, <i>Vitis Vinifera</i>
153839	Alface, <i>Lactuca Sativa</i> , Produção	Alface	Alface
161390	Algodão, Condição Ambiental, Controle biológico, Percevejo, Predador	Agronegócio, Extrusão, <i>Sorgh Umbicolor</i>	Agronegócio, Exportação, Mosca Das Frutas, Irrigação, Tecnologia, Temperatura
156709	Alimentação, Caatinga, Caprino, Umbu	Agricultura Familiar, Caprino, Comercialização, Ovino	Caatinga, Agricultura Familiar, Caprino, Ovino, Leite De Cabra, Umbu
1129161	Alimento Alternativo, Biomassa, Forragem, <i>Moringa Oleifera</i>	Bactéria, Biometria, Forragem, Folha, <i>Rotylenchulus Reniformis</i> , Pastagem, Granulometria, Nutrição Animal, Ovino, <i>Panicum Maximum</i> , Rúmen	Metabolismo, Gramínea Forrageira, Aveia, Solo, Arenoso
154151	Anacardiaceae, Melhoramento Genético Vegetal, <i>Spondias Tuberosa</i> , Umbu	<i>Arachis Hypogaea</i> , Floração, Floração, Fruticultura, Maçã, Polinização	<i>Arachis Hypogaea</i> , Crescimento, Doença, Germoplasma
103355	Análise de Dados, Pesquisa	Trigo	Trigo
211707	Antracnose, <i>Colletotrichum Lindemuthianum</i> , Feijão, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i> , Produtividade	Antracnose, <i>Colletotrichum Lindemuthianum</i> , Feijão, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i>	Feijão
1118707	<i>Aphis Gossypii</i> , Silício	Herbicida, Impacto Ambiental	Herbicida

ID	BDPA	Previsão 1	Previsão 2
1008415	Armazenamento, <i>Capsicum Annuum</i> , Pimentão, Preservação de alimento, Produção orgânica	Controle químico, armazenamento	Armazenamento, <i>Coffea Arabica</i> , desidratação, processamento
212670	Armazenamento, Feijão, Fungicida, <i>Phaseolus Vulgaris</i> , Revestimento, Semente	Armazenamento, Germinação	Armazenamento, herbicida, erva daninha, pastagem natural
1127555	Bacia Hidrográfica, Precipitação Pluvial, Vazão	Biomassa, desertificação	Sensoriamento remoto
151173	Balanço de Energia, Banana, Evapotranspiração	Banana, Evapotranspiração, Irrigação	Bulbo, Evapotranspiração, Irrigação, Solo, Tensiômetro
1129606	Biogás, Níquel, Oxido	Impacto ambiental, mineração, qualidade da água	Impacto ambiental, mineração, qualidade da água
1123970	<i>Bixa Orellana</i> , Metal, Urucum	Marcador genético	Biotecnologia
975131	Caatinga, Desertificação, Recurso natural, Sensoriamento Remoto	Caatinga, Energia, Irrigação, Produção Animal, Programa De Pesquisa	Agricultura Familiar, Caprino, Dieta, Nutrição Animal, Melhoramento Genético Animal, Óleo Vegetal, Produção Animal, <i>Syagrus Coronata</i>
1110938	Cacau, Doença, Manejo, Praga, <i>Theobroma Cacao</i>	Propagação Vegetativa	Cigarrinha, Flutuação Populacional
125759	Cádmio, <i>Eucalyptus Camaldulensis</i> , Poluição do Solo, Solo, Zinco	Análise, Graviola	Resíduo
975338	Cana de Açúcar, Chuva, Herbicida, Lixiviação	Herbicida, soja	Herbicida, soja
1102669	Características Agronômicas, Maracujá, Melhoramento Genético Vegetal, Parâmetro Genético	Banco de Germoplasma, Características Agronômicas, <i>Ipomoea Batatas</i>	Melhoramento, <i>Lirio Myzatriifolii</i>
1074877	Casca de arroz, Espécie nativa, Guabioba, Micorriza vesicular arbuscular, Substrato de cultura	Adubo orgânico, nutrição vegetal, <i>Panicum Maximum</i> , patógeno	Cálcio, Invertase
1120755	Certificação de Produto, Suco, Vinho	Alimento, Fruticultura, Maçã, Produção	Pós-Colheita
1124357	Cerveja, Cevada, Malte, Proteína	<i>Boophilus Microplus</i> , Gramínea Tropical, Bezerro, Digestibilidade, Galinha para Postura, <i>Pennisetum Purpureum</i> , Valor Nutritivo	Bezerro, Confinamento, Feno, Gramínea Tropical, Nutrição Animal, Bezerro, Imunidade, Gramínea Tropical, Nutrição Animal, Valor Nutritivo
280598	Chuva, Doença, Temperatura	<i>Ilex Paraguariensis</i>	Crescimento
1017311	Clonagem, Propagação Vegetativa	Propagação vegetativa	Propagação vegetativa

ID	BDPA	Previsão 1	Previsão 2
313756	Cobertura do Solo, Solo	Arado de aiveca, manejo do solo, plantio direto, solo	Arado de aiveca, manejo do solo, plantio direto, solo
157560	Confinamento, Consumo, Conversão Alimentar, Ganho de Peso, Ovino, Terminação	Parasitismo, Praga de Planta, Uva, Viticultura	Parasitismo, Praga de Planta, Tripes, Uva, <i>Vitis Vinifera</i>
314079	Conservação, Endogamia, Germoplasma	Clonagem, <i>Ilex Paraguariensis</i>	Clonagem, eucalipto, propagação vegetativa
1125816	Consumo Alimentar, Culinária	Consumo alimentar, culinária	Açaí, economia, mercado
280844	Controle Biológico, Doença, <i>Ilex Paraguariensis</i> , Trichoderma	Controle Biológico, Fungo	Controle Biológico
1126373	Controle Químico, Doença de Planta, Enfezamento, Resistência Genética, Zea Mays	doença de planta, híbrido, sorghum bicolor, variação genética	doença de planta, híbrido, sorghum bicolor, zea mays
134128	Cucumis Melo, Dióxido de Carbono, Fertirrigação, Fotossíntese, Irrigação, Melão	Calcário, controle químico	Adubação, <i>Ilex Paraguariensis</i> , soja
1121184	Deficiência Hídrica, Evapotranspiração, Soja	Radiação solar, soja	Fitotoxicidade, planta oleaginosa, produtividade
1118305	Desmatamento	Açaí, espécie nativa	Agricultura familiar, pesquisa agrícola
134397	Doença, <i>Lycopersicon Esculentum</i> , Tomate	Qualidade	Qualidade
974846	<i>Echinochloa Frumentacea</i> , <i>Oryza sativa</i>	Fenologia	Doença de planta
312782	Enraizamento, <i>Ficus Carica</i> , Figo, Regulador de Crescimento	Vitamina C	Amido, Mirtilo
962648	Estatística agrícola, Petróleo, Preço internacional, Produto, Ureia	Geada, Grão, População de Planta, Produtividade, <i>Vigna Unguiculata</i>	Colheita Mecânica, Feijão, Irrigação por Aspersão
314117	Esterco de Caprino, Produção, Resíduo Industrial, Resíduo Orgânico	<i>Cedrela Odorata</i> , Consumo Alimentar, Poluição, Peixe	<i>Cedrela Odorata</i> , Impacto Ambiental, Poluição, Rede, Tanque
958618	Estirpe, Segurança Alimentar, <i>Staphylococcus Aureus</i>	Antibiótico, fermentação, <i>Ralstonia Solanacearum</i> , respiração, tomate, traça, tuta absoluta	Bebida, fermentação, fermentação, processamento mínimo, tecnologia de alimento, <i>Xylella Fastidiosa</i>
204144	Feijão, Melhoramento Genético Vegetal, Nitrogênio, <i>Phaseolus Vulgaris</i>	Aclimatação, <i>Asparagus Officinalis</i> , Aspargo, Híbrido, Irrigação, Produtividade, Variedade	Aclimatação, <i>Asparagus Officinalis</i> , Aspargo, Híbrido, Irrigação, Produtividade, Variedade
215848	Feijão, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i>	Feijão, melhoramento genético vegetal, <i>Phaseolus Vulgaris</i>	Feijão, melhoramento genético vegetal, <i>Phaseolus Vulgaris</i>

ID	BDPA	Previsão 1	Previsão 2
214593	Feijão, <i>Phaseolus Vulgaris</i> , Produção, Qualidade, Semente	Algodão, <i>Anthonomus Grandis</i> , <i>Gossypium Hirsutum</i> , pulverização	<i>Cucumis Melo</i> , Melão
1015548	Fenologia, Pêra, <i>Pyrus Communis</i>	Ameixa, fruticultura, produção, <i>Prunus Salicina</i> , Variedade	Ameixa, banana, fruticultura, radiação, variedade
315462	Fertilidade do Solo	Adubação, produtividade	Adubação, fósforo
1008316	Fertilizante fosfatado, Nutrição vegetal, Plantio direto	Adubação, produtividade, safrinha, safrinha, sistema de cultivo	Adubação, produtividade, safrinha, safrinha
955737	Floração, Fruticultura, Pêssego, Produção, Produtividade, Variedade	Fenologia	Pimenta
314145	Fogo	Manejo	Manejo
1122380	Fornagem, Ruminante	<i>Sorghum Bicolor</i> , sorgo	Adubação, <i>Sorghum Bicolor</i> , sorgo
17098	Fosfato	Solo	Classificação do solo
961303	Germinação, Palmito, Semente	Germinação	Germinação, germoplasma, semente
195124	Germoplasma, <i>Vigna Unguiculata</i>	Feijão, Germinação, Melhoramento Genético Vegetal, <i>Phaseolus Vulgaris</i>	DNA, Genética, Marcador Molecular
314935	<i>Hevea Brasiliensis</i> , Seringueira	Alho, características agrônômicas, clima, produtividade, rendimento	Alho, <i>Allium Sativum</i> , crescimento, cultura in vitro, meristema, produção
162329	<i>Lactuca Sativa</i> , Produtividade, Variedade	Água, Moringa, Planta, Planta, Suplemento Alimentar, Ureia	Cálcio, Clorofila, Deterioração, Fósforo, Semente
119110	Legislação, Leite Pasteurizado	Armazenamento, conservação, pós-colheita, qualidade	Diagnóstico
315464	Marcador Molecular	Marcador molecular	Assentamento
1016654	Milho Doce, Pós-Colheita, Processamento Mínimo, Qualidade, <i>Zea Mays</i>	Qualidade	Uva
57793	Milho, Variedade	Híbrido, Produtividade, Qualidade, Variedade, <i>Zea Mays</i>	Milho Doce, Variedade, <i>Zea Mays</i>
313801	Muda, Plântula, Radiação Solar	Calcário, Crescimento, Espécie Nativa, Herbicida, Planta, Solo	Alface, Crescimento, Caule, Crescimento, <i>Lactuca Sativa</i>
1129435	Mudança Climática, Soja	Soja	Glycinemax, <i>Solanum Tuberosum</i>
1122642	Peroxidase, Polifenoloxidase	Amido, Mandioca, <i>Manihot Esculenta</i> , Patente, Subproduto	Densidade De Plantio, Processamento, Processamento, Silagem
1126835	Piscicultura	Compostagem, resíduo orgânico	Compostagem

ID	BDPA	Previsão 1	Previsão 2
313739	Procedência	<i>Grevillea Robusta</i> , Regeneração	<i>Grevillea Robusta</i>
121733	Produtividade, Risco, Sistema de cultivo	Capim, fertilidade, produção orgânica, proteína bruta, rotação de cultura	Capim, cerrado, políticas públicas, planta de cobertura, plantio direto, rotação de cultura
89545	Segurança Alimentar	Segurança alimentar	Segurança alimentar
957259	Sensoriamento remoto	Sensoriamento remoto	Sensoriamento remoto
89928	Sistema de Produção	Carne	Carne
1129319	Tecnologia Agrícola	Glycine Max	Coffea Arabica
314112	Temperatura	Fenologia	Armazenamento, <i>Coffea Arabica</i> , Fenologia
1123933	Trigo	Trigo	Safrinha, trigo

Fonte: Elaborado pelo autor.