

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Escola de Ciência da Informação

Programa de Pós-graduação Gestão e Organização do Conhecimento

Gustavo Portela Lages

**INTELIGÊNCIA DE ESTADO E DOCUMENTOS DESCLASSIFICADOS DA CIA:
uma abordagem de aprendizado de máquina**

Belo Horizonte

2020

Gustavo Portela Lages

**INTELIGÊNCIA DE ESTADO E DOCUMENTOS DESCLASSIFICADOS DA CIA:
uma abordagem de aprendizado de máquina**

Dissertação apresentada ao Programa de Pós-Graduação Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais como instrumento de obtenção do grau de Mestre em Ciência da Informação.

Orientador: Renato Rocha Souza

Belo Horizonte

2020

L174i

Lages, Gustavo Portela.

Inteligência de Estado e documentos desclassificados da CIA [recurso eletrônico] : uma abordagem de aprendizado de máquina / Gustavo Portela Lages. - 2020.

1 recurso online (126 f. : il., tab., color.) : pdf.

Orientador: Renato Rocha Souza
Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.
Referências: f. 116-126.
Exigências do sistema: Adobe Acrobat Reader.

1. Ciência da informação – Teses. 2. Gestão do conhecimento – Teses. 3. Aprendizado do computador – Teses. 4. Recuperação da informação – Teses.
I. Título. II. Souza, Renato Rocha. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 659.2:007.52

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
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO
CONHECIMENTO



ATA DA DEFESA DA DISSERTAÇÃO DO ALUNO **GUSTAVO PORTELA LAGES**

Realizou-se, no dia 28 de agosto de 2020, às 09:00 horas, Videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada Inteligência de Estado e documentos desclassificados da CIA: uma abordagem de Aprendizado de Máquina, apresentada por GUSTAVO PORTELA LAGES, por videoconferência, número de registro 2018666856, graduado no curso de RELACOES ECON.INTERNAZIONALEIS/N, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Renato Rocha Souza - FGV/RJ [por videoconferência] (Orientador), Prof(a). Benildes Coura Moreira dos Santos Maculan - ECI/UFMG [por videoconferência], Prof(a). Cátia Rodrigues Barbosa - Aposentada/UFMG [por videoconferência], Ricardo Rodrigues - Aposentado/UFMG [por videoconferência].

A Comissão considerou a dissertação:

(x) Aprovada

() Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.
Belo Horizonte, 28 de agosto de 2020.

Prof(a). Renato Rocha Souza

Prof(a). Benildes Coura Moreira dos Santos Maculan

Prof(a). Cátia Rodrigues Barbosa

Prof(a). Ricardo Rodrigues Barbosa



UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO
CONHECIMENTO



FOLHA DE APROVAÇÃO

Inteligência de Estado e documentos desclassificados da CIA: uma abordagem de Aprendizado de Máquina

GUSTAVO PORTELA LAGES

Dissertação 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 Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Gestão e Tecnologia.

Aprovada em 28 de agosto de 2020, pela banca constituída pelos membros:

Prof(a). Renato Rocha Souza (Orientador)
FGV/RJ [por videoconferência]

Prof(a). Benildes Coura Moreira dos Santos Maculan
ECI/UFMG [por videoconferência]

Prof(a). Cátia Rodrigues Barbosa
Aposentada/UFMG [por videoconferência]

Prof(a). Ricardo Rodrigues Barbosa
Aposentada/UFMG [por videoconferência]

Belo Horizonte, 28 de agosto de 2020.

Dedico às minhas belas e amadas, Sofia e Mariana.

AGRADECIMENTOS

A Deus por guiar meus caminhos, me amparar e fortalecer nos desafios e me inspirar coragem acreditando na fraternidade e no amor.

A minha família por sempre estar presente e acreditar no meu potencial.

Ao colega Rodrigo pelas ideias e contribuições referentes ao projeto.

Ao colega Heber por suas importantes contribuições e críticas durante a execução.

Ao meu orientador Renato Rocha Souza pelos decisivos apontamentos e direcionamentos na pesquisa.

A ECI/UFMG, desde professores a funcionários, por me acolherem e oportunizarem esse período de aprendizado.

Pedi e vos será dado; buscai e encontrareis; batei e será aberto para vós.
Pois todo aquele que pede recebe, e aquele que busca encontra, e ao que bate será aberto.
(Mateus)

A ignorância é audaz; a sabedoria, reservada.
(Tucídides)

Inteligência de Estado e documentos desclassificados da CIA: uma abordagem de
Aprendizado de Máquina.

Gustavo Portela Lages

State Intelligence and CIA declassified documents: an approach of Machine Learning

Resumo: A Inteligência de Estado é uma atividade de assessoramento do poder decisório central dos Estados Nacionais. A produção de conhecimento da atividade de Inteligência de Estado gera documentos classificados em graus de segredo e temporalidade. Percorridos o tempo definido os documentos são disponibilizados para consulta pública. Assim, a proposta desta pesquisa foi recuperar os principais termos/temas/tópicos acompanhados pela Inteligência de Estado Externa Norte-Americana em um corpus relacionados a assuntos econômicos no âmbito internacional. A premissa é que os resultados encontrados expressam interesses internacionais do Estado norte-americano, uma vez que a Inteligência de Estado existe em função do tomador de decisão em última instância. A etapa analítica aplicou técnicas de processamento de linguagem natural e aprendizado de máquina para executar a análise documental de forma a descobrir e discutir os assuntos, localidades e entidades mais relevantes que foram objetos de observação da Inteligência dos Estados Unidos, se e como se relacionaram ao Brasil. A linguagem *Python* foi o meio utilizado para elaborar algoritmos capazes de executar a recuperação e caracterização do corpora dos documentos, através dela foram aplicados contagem de termos, *word-embeddings*, clusterização, culminando na modelagem de tópicos. Foram definidos seis tópicos finais que abordaram temas ligados a URSS, poder militar, político e econômico, petróleo e derivados, grãos, transporte e serviços, aviação e marinha, China, Brasil, África, Mundo Árabe e Europa.

Palavras-Chave: inteligência de estado; gestão do conhecimento; aprendizado de máquina.

Abstract: State Intelligence is an activity of advising the central decision-making power of National States. The production of knowledge of the State Intelligence activity generates documents classified in degrees of secrecy and temporality. After the defined time, the documents are made available for public consultation. Thus, the purpose of this research was to retrieve the main terms/themes/topics followed by the North American External State Intelligence in a corpus related to economic issues at the international scope. The premise is that the results found express international interests of the North American State, since the State Intelligence exists due to the decision maker in the last resort. The analytical step applied to natural language processing techniques and machine learning to perform documentary analysis in order to discover and discuss the most relevant subjects, locations and entities that were the object of observation by the United States Intelligence, if and how they related to Brazil. The Python language was the means used to develop algorithms capable of performing the recovery and characterization of the document's corpora, through which term counting, word-embeddings and clustering were applied, culminating in the topic modeling. Six final topics were defined that addressed themes related to the USSR, military, political and economic power, oil and its products, grains, transport and services, aviation and navy, China, Brazil, Africa, Arab's world and Europe.

Keywords: state intelligence; knowledge management; machine learning.

LISTA DE FIGURAS

FIGURA 1 - Categorias de Inteligência.....	26
FIGURA 2 - Espiral do Conhecimento.....	34
FIGURA 3 - Conteúdos de conhecimento.....	34
FIGURA 4 - Processos de criação do conhecimento.....	35
FIGURA 5 - Mapa Conceitual da Gestão do Conhecimento.....	37
FIGURA 6 - Pirâmides de Inteligência.....	39
FIGURA 7 - Dados transformam-se em Inteligência.....	40
FIGURA 8 - Elementos de um Sistema de Inteligência.....	42
FIGURA 9 - Vínculos da GIC com outras áreas do conhecimento.....	44
FIGURA 10 - Modelo ICC.....	51
FIGURA 11 - Esquema de metodologia do trabalho.....	64
FIGURA 12 - Data Mining / KDD.....	65
FIGURA 13 - Diferença entre <i>CBOW</i> e <i>Skip-gram (Word2Vec)</i> e <i>Skip-grae</i>	71
FIGURA 14 - Mapa Mental da Metodologia Map.....	76
FIGURA 15 - Termos mais frequentes no corpus.....	80
FIGURA 16 - Gráfico dos termos mais frequentes no corpus.....	81
FIGURA 17 - WordCloud dos termos do corpus.....	81
FIGURA 18 - Perfil de aparição da palavra “soviet(s)”.....	83
FIGURA 19 - Perfil de aparição da palavra “economic”.....	84
FIGURA 20 - Perfil de aparição da palavra “brazil”.....	85
FIGURA 21 - Dispersão Lexical.....	85
FIGURA 22 - Nós perante o <i>token</i> “ussr” em <i>CBOW</i>	88
FIGURA 23 - Nós perante o <i>token</i> “ussr” em <i>Skip-gram</i>	88
FIGURA 24 - Nós perante o <i>token</i> “soviet” em <i>CBOW</i>	90
FIGURA 25 - Nós perante o <i>token</i> “soviet” em <i>Skip-gram</i>	90
FIGURA 26 - Nós perante o <i>token</i> “brazil” em <i>CBOW</i>	92
FIGURA 27 - Nós perante o <i>token</i> “brazil” em <i>Skip-gram</i>	92
FIGURA 28 - Plot dos clusters em <i>K-means</i>	95
FIGURA 29 - Clusters em Ward.....	96
FIGURA 30 - Modelagem de Tópicos LDA (a) / <i>Gensim</i>	97
FIGURA 31 - Modelagem de Tópicos LDA (b) / <i>Gensim</i>	98
FIGURA 32 - Modelagem de Tópicos LDA / <i>Scikit Learn</i>	99
FIGURA 33 - Modelagem de Tópicos NMF / <i>Scikit Learn</i>	100

LISTA DE TABELAS / QUADROS

TABELA 1 - Resultados obtidos com a Revisão Bibliográfica no Portal Capes.....	18
TABELA 2 - Aparição dos termos no corpus.....	82
TABELA 3 - Valores perante o <i>token</i> “ussr”.....	87
TABELA 4 - Valores perante o <i>token</i> “soviet”.....	89
TABELA 5 - Valores perante o <i>token</i> “brazil”.....	91
TABELA 6 - Matriz dos novos Tópicos.....	100
QUADRO 1 - Dado, Informação e Conhecimento.....	41
QUADRO 2 - Interfaces entre Inteligência de Estado e GIC.....	48
QUADRO 3 - Tipologia de Inteligência.....	50
QUADRO 4 - Termos recuperados nos <i>clusters</i> em <i>K-means</i>	94

SUMÁRIO

1 INTRODUÇÃO.....	12
2 REVISÃO DE LITERATURA.....	16
3 REFERENCIAL TEÓRICO.....	21
3.1 Inteligência de Estado.....	22
3.2 Gestão do Conhecimento.....	29
3.3 Inteligência de Estado e Gestão do Conhecimento.....	38
3.4 O papel da Inteligência de Estado nas Relações Internacionais.....	52
3.4.1 O período de Guerra Fria.....	60
4 METODOLOGIA.....	63
4.1 Visão Geral.....	63
4.2 Ferramentas de PLN – Técnicas de Aprendizado de Máquina.....	67
4.2.1 Word-embeddings.....	69
4.2.2 Clusterização.....	71
4.2.3 Modelagem de Tópicos.....	72
4.3 Execução.....	75
5 RESULTADOS E DISCUSSÃO.....	80
6 CONSIDERAÇÕES FINAIS.....	114
REFERÊNCIAS.....	116

1 INTRODUÇÃO

A pesquisa acadêmica sobre Inteligência de Estado (IE) possui suas limitações ligadas às fontes de informação. A academia no Brasil ainda não se dedica muito a esses estudos, embora haja crescimento recente dado a temas como terrorismo, cyber segurança, deep web, big data, entre outros. Os dados ligados à IE disponíveis em veículos de grande circulação, muitas vezes são produções jornalísticas e, em vários casos, dúbias, não se sustentando fidedignamente. As fontes oficiais provêm de documentos desclassificados¹ - aqueles documentos sigilosos tornados acessíveis ao público quando findado o prazo de classificação de cada país; as oficiosas são oriundas de vazamentos como, por exemplo, os proporcionados pela organização *Wikileaks* e o ex-agente da NSA (Edward Snowden).

O ordenamento jurídico brasileiro define, pela Lei 9883, de 07 de dezembro de 1999, em seus artigos 1º, §§3º e 4º:

Inteligência como “a atividade que objetiva a obtenção, análise e disseminação de conhecimentos dentro e fora do território nacional sobre fatos e situações de imediata ou potencial influência sobre o processo decisório e a ação governamental e sobre a salvaguarda e a segurança da sociedade e do Estado” e Contrainteligência como “a atividade que objetiva neutralizar a inteligência adversa”.

Esse é o contexto do trabalho, este tipo de inteligência ligada e executada pelos governos nacionais será o tema base desta pesquisa. Não obstante, há um flanco da e-Science, próprio da Ciência da Informação (CI), conectado a este trabalho. A e-Science pode ser compreendida pelo “uso intensivo de dados na pesquisa científica, o uso de infraestrutura tecnológica para a colaboração entre cientistas e os processos de colaboração científica em si, mediados pelas tecnologias.” (APPEL, 2014, p.8). A técnica de coleta de dados – análise documental – estará presente ao aplicar algoritmos de *Machine Learning* (Aprendizado de Máquina) para delinear os termos/temas/tópicos mais prevalentes no corpus da pesquisa.

A coleta de dados compreende o conjunto de operações por meio das quais o modelo de análise é confrontado aos dados coletados. Ao longo dessa etapa, várias informações são, portanto, coletadas. Elas serão sistematicamente analisadas na etapa posterior. Conceber essa etapa de coleta de dados deve levar em conta três questões a serem respondidas: O que coletar? Com quem coletar? Como coletar? (GERHARDT; SILVEIRA, 2009, p. 56)

Este trabalho, além de analisar documentos produzidos de acordo com a ótica epistemológica da CI e da IE, utiliza artifícios tecnológicos computacionais de programação

¹Art. 2º, inciso V, Decreto 2.134 de 1997 e Art. 24, Lei 12.527 de 2011.

para realizar ou empregar técnicas de organização do conhecimento e representação da informação. Pois faz uma análise documental descobrindo tópicos que se agrupam em temas mais relacionados dentro da corpora do material de análise. O Processamento de Linguagem Natural (PLN) através de aprendizado de máquina permite recuperar e organizar as informações em uma espécie de indexação automática.

Indexação “é uma técnica de análise de conteúdo que condensa a informação significativa de um documento, através da atribuição de termos, criando uma linguagem intermediária entre o usuário e o documento. É um dos processos básicos de recuperação da informação.” (VIEIRA, 1988, p.1). A indexação automática, feita por computador, é mais objetiva e nessa pesquisa se situa, principalmente, no nível de categorização (representa o assunto que predomina) e em certa medida superficial (representa os conceitos principais de forma geral) (ROBREDO, 1982).

A linguagem de programação *Python* será utilizada como um sistema para realizar o processamento da linguagem natural dos documentos do corpus, que no caso está no idioma inglês. Por envolver vários estágios de processamento, com *inputs* (entradas) e *outputs* (saídas), um sistema de processamento de linguagem natural é geralmente chamado de pipeline (LANE; HOWARD; HAPKE, 2019).

“O processamento de linguagem natural é uma área de pesquisa em ciência da computação e inteligência artificial (IA) relacionada ao processamento de idiomas naturais, como inglês ou mandarim. Esse processamento geralmente envolve a tradução de linguagem natural em dados (números) que um computador pode usar para aprender sobre o mundo.”(LANE; HOWARD; HAPKE, 2019, p.4)

Feitas as considerações iniciais, a questão de pesquisa é saber (recuperar) quais foram os principais termos/temas/tópicos (por exemplo: assuntos, localidades e entidades) acompanhados pela Inteligência de Estado Externa Norte-Americana desde o fim da Segunda Guerra Mundial até o início da década de 1990, do ponto de vista econômico, presentes em documentos de inteligência desclassificados.

Como objetivo geral, espera-se caracterizar (interpretar e organizar) o cenário econômico internacional monitorado pela inteligência norte-americana, a partir da análise textual quantitativa e qualitativa dos documentos desclassificados (corpus) coletados no site da *Central Intelligence Agency* (CIA). Os objetivos específicos são: verificar se os temas recuperados têm relação com o contexto histórico vivenciado pelos EUA no período de

produção dos documentos e detectar se e quais temas estavam relacionados ao Brasil. A justificativa deste trabalho está baseada na importância da IE para um país em uma perspectiva realista das relações internacionais e na verificação da aplicabilidade de técnicas de aprendizado de máquina para recuperar e organizar informações em documentos de Inteligência.

Realismo é uma corrente de pensamento de uma das escolas das Relações Internacionais. No realismo, o Estado é o ator central; as decisões são objetivas e pragmáticas para atingir os interesses nacionais; utiliza-se do poder e da força bélica em um sistema internacional considerado anárquico; com ou sem alianças para garantir a sobrevivência do Estado (NOGUEIRA; MESSARI, 2005). O realismo desenha um cenário internacional de anarquia e os países estariam ou em guerra; no dilema de segurança²; ou *détente*, em português, distensão, relaxamento ou diminuição da tensão (SOUSA, 2005). Portanto, necessitam a todo o momento estarem preparados para o conflito (*si vis pacem, para bellum*³). Sun Tzu (2006), referência para o realismo e IE, elenca circunstâncias estratégicas basilares para vencer o inimigo.

Conhece teu inimigo e conhece-te a ti mesmo; se tiveres cem combates a travar, cem vezes serás vitorioso. Se ignoras teu inimigo e conheces a ti mesmo, tuas chances de perder e de ganhar serão idênticas. Se ignoras ao mesmo tempo teu inimigo e a ti mesmo, só contarás teus combates por tuas derrotas. (TZU, 2006, p.23).

Sendo assim, ter claramente definidos os planos e metas de um país é fundamental, mas conhecer os outros é igualmente relevante e necessário. Isso mostra que recuperar e organizar os termos/temas/tópicos dos documentos desclassificados da *Central Intelligence Agency* (CIA) ou Agência Central de Inteligência dos EUA é um bom caminho para traçar um perfil da agenda geopolítica dos EUA. Isso pode favorecer outros serviços de inteligência/países, sobretudo, o próprio Brasil.

A atividade de inteligência (inteligência e contrainteligência) brasileiras terão ganhos ao saber o que monitora e como pensam seus parceiros/concorrentes em âmbito internacional e avaliar o que fez e fará para alcançar seus objetivos e metas internacionais e,

² “Conceito introduzido por John Herz, que traduz os sentimentos de insegurança resultantes de políticas de fortalecimento da capacidade militar de um Estado, com propósitos de aumentar a sua segurança em termos defensivos, interpretadas como ameaçadoras e encorajando o rearmamento dos outros Estados.” (Dicionário de Relações Internacionais, 2005).

³ “Se queres a paz, prepare a guerra”. Disponível em: <https://www.dicionariodelatim.com.br/si-vis-pacem-para-bellum/>. Acesso em 03/12/2018, às 12h09.

sobretudo, se proteger frente a ações externas, ainda que esses documentos se refiram ao passado.

Esta pesquisa é também a continuidade de estudos de Inteligência de Estado iniciados pelo autor durante sua graduação em Relações Econômicas Internacionais – FACE/UFG, que culminou em uma monografia intitulada “A Influência da Inteligência Governamental sobre a Política Externa Brasileira” no ano de 2017. Nela foram discutidos aspectos sobre a importância/necessidade de fortalecer a IE brasileira com vistas a auxiliar no alcance de planos e metas internacionais.

Essa dissertação está dividida em sete capítulos, que contêm: no primeiro, a introdução, com aspectos teóricos gerais, o problema, objetivos geral e específicos da pesquisa; no segundo, a revisão de literatura com a metodologia adotada e seus resultados e discussões; no terceiro, o referencial teórico, subdividido em uma abordagem sobre Inteligência de Estado, outra sobre a conexão epistemológica entre IE e Gestão do Conhecimento (GC), e uma terceira sobre o papel da IE nas Relações Internacionais (RI); no quarto, a Metodologia de Pesquisa, subdividida em três, uma visão geral sobre a condução e aspectos teóricos das técnicas aplicadas na obtenção dos resultados, uma abordagem teórica sucinta sobre modelagem de tópicos e a execução, descrita de forma detalhada e sistemática; no quinto, os resultados e discussões derivados da execução delineada na metodologia, com apresentação de gráficos, tabelas e outros tipos de visualizações que auxiliaram na interpretação dos resultados; no sexto, as considerações finais do trabalho, utilizando os resultados e retomando pontos do problema, objetivos e justificativa da pesquisa, além de críticas e perspectivas futuras para novos trabalhos; por fim, no sétimo, as referências bibliográficas.

A seguir, portanto, está descrito o procedimento e resultados da execução da revisão de literatura com base nos principais temas de fundo teórico da pesquisa através de um procedimento de revisão sistemática narrativa em bases de dados relacionadas à CI, com detalhamento de aspectos de inclusão e exclusão para os resultados recuperados e breve discussão de artigos mais significativos encontrados pela revisão.

2 REVISÃO DE LITERATURA

A revisão sistemática de literatura está dividida didaticamente em categorias ou tipologias. De acordo com a proposta deste trabalho, o enquadramento mais adequado para o que foi conduzido nesse caso seria a revisão narrativa.

A revisão narrativa é empregada para caracterizar o estado da arte de um assunto específico, sob o ponto de vista teórico ou contextual (ROTHER, 2007). Segundo essa autora, tal revisão possibilita a recuperação e atualização de conhecimento sobre um determinado tema em curto período de tempo; entretanto, não possui metodologia que sustente a reprodução dos dados e não traz respostas quantitativas para determinados questionamentos.

Sendo assim, selecionou-se algumas das principais base de dados em ciência da informação e outra mais geral (*Emerald Insight*) e através do portal de periódicos Capes⁴ a pesquisa ocorreu após a definição dos critérios descritos na sequência.

Feita uma análise conceitual dos grandes temas de pesquisa definiu-se três principais: Inteligência de Estado, Mineração de Textos e Modelagem de Tópicos. Em seguida os termos foram traduzidos para a língua inglesa e utilizaram-se os sinônimos mais prevalentes na literatura, quando existentes, para ampliar o alcance das buscas.

Os critérios de inclusão iniciais foram:

- Fonte – periódicos revisados por especialistas;
- Tipo – artigos;
- Período – últimos quinze anos (desde 2004) para limitar o universo e concentrar em publicações mais recentes, considerando a natureza recente das temáticas de *Machine Learning* e possíveis aplicações ligadas a IE.

Foram utilizadas as bases de dados *Library and Information Science Abstracts – LISA (ProQuest)*, *Information Science and Technology Abstracts – ISTA (EBSCO)*, *Emerald Insight*, *Scopus* e *Web of Science*.

Os itens foram pesquisados separadamente desta forma:

⁴ “O Portal de Periódicos foi oficialmente lançado em 11 de novembro de 2000, na mesma época em que começavam a ser criadas as bibliotecas virtuais e quando as editoras iniciavam o processo de digitalização dos seus acervos. Com o Portal, a Capes passou a centralizar e otimizar a aquisição desse tipo de conteúdo, por meio da negociação direta com editores internacionais.”
<https://bit.ly/3difVVy>. Acesso em 28/11/2018, às 11h06.

- a) os termos “Machine Learning”, “Expert Systems” e “Artificial Intelligence” foram combinados com o booleano “OR”;
- b) os termos “State Intelligence” e “Governmental Intelligence” foram combinados com o booleano “OR”, acrescido do booleano NOT e o termo “Competitive Intelligence”;
- c) os termos “Topic Modeling” e “Text Mining” foram combinados com o booleano “OR”.

Os itens a, b e c foram pesquisados em todas as bases de dados citadas, isoladamente e, em seguida, combinados da seguinte forma:

- d) *(Item a) AND (Item b) AND (Item c)*;
- e) *(Item a) AND (Item b)*;
- f) *(Item b) AND (Item c)*;
- g) *(Item a) AND (Item c)*.

Os critérios de exclusão foram:

- resumos de artigos que não apresentassem temática de Inteligência de Estado ou algum tipo de Inteligência Governamental (que envolvessem ações estatais, serviços públicos etc);
- duplicidade de artigos entre as bases de dados.

A revisão sistemática demonstrada acima segue os parâmetros discutidos e exemplificados em dois artigos seminais sobre revisão de literatura, KITCHENHAM et al (2009) e Kitchenham (2010)⁵. Estudos em que se avaliam revisões de literatura manuais e utilizando softwares e as compara.

Os resultados encontrados estão abaixo relacionados para cada critério e respectiva base.

⁵ KITCHENHAM et al. *Systematic literature reviews in software engineering – A systematic literature review*, 2009, p. 7-15.

Id. *Systematic literature reviews in software engineering – A tertiary study*, 2010, p. 792-805.

TABELA 1 - Resultados obtidos com a Revisão Bibliográfica no Portal Capes

Bases de Dados					
Critérios	LISA	ISTA	Emerald Insight	Scopus	Web of Science
Item a	40	15	10.014	40.213	4.997
Item b	0	0	22	21	59
Item c	12	0	972	3.177	11.973
Item d	0	0	6509	0	0
Item e	0	0	6510	0	0
Item f	0	0	151	0	0
Item g	2	0	7012	460	8

FONTE: Elaborado pelo autor.

Os resultados apresentados no *item d* (utilizando a combinação do item “a”, “b” e “c”) representa a combinação ideal, ou seja, aquele em que se objetiva recuperar artigos para subsidiar a literatura de referência para o trabalho. Não obtendo resultados que trouxessem os três principais campos de conceitos da pesquisa combinados, optou-se por executar a pesquisa para combinações dos itens dois a dois, a fim de ampliar as possibilidades para construir um referencial teórico. Assim, utilizando as combinações em blocos de dois campos de conceitos as chances poderiam aumentar e trazer resultados mais próximos do que se desejava.

Na etapa de avaliação dois a dois, as bases de dados pesquisadas que recuperaram artigos passaram por uma avaliação subsequente. Os resumos dos artigos foram lidos a fim de selecionar quais atenderiam aos critérios deste trabalho. Após a varredura na base *LISA* os dois artigos foram selecionados; a base *ISTA* não recuperou artigos de acordo com os critérios definidos; na base *Emerald Insight*, o número excessivo de resultados nos critérios combinados demonstrou uma falta de seletividade e precisão, sendo, portanto, selecionados apenas dois artigos por relevância de citações; a base *Scopus* apresentou dois artigos relevantes em um universo de trinta mais relevantes dos quatrocentos e sessenta para o *item g*, o que remete ao mesmo problema da base anterior; por fim, na base *Web of Science* nenhum dos oito artigos apresentou conteúdo compatível com os critérios estabelecidos.

Os resultados preliminares demonstraram a dificuldade mencionada na introdução de se obter documentos relacionados à Inteligência de Estado (tratar com dados secretos e documentos classificados), sobretudo, quando conjugada aos demais temas principais da pesquisa. Isso levou a uma adaptação da estratégia, diminuindo a combinação dos itens em

dois a dois e, mesmo assim, os resultados não foram tão satisfatórios. Artigos científicos que considerem a análise de documentos desclassificados não trarão em sua maioria o uso de *Machine Learning*, uma vez que as tecnologias envolvidas são mais recentes e a desclassificação cumpre uma temporalidade.

Os artigos selecionados pela base LISA foram baseados em análise de sentimentos – mineração de opiniões para dar suporte a decisões governamentais. O artigo de Xu, Liu e Shang⁶ abordou o desenvolvimento de um método para detectar possíveis eventos emergenciais através do cruzamento de informações de mídias sociais para subsidiar a tomada de decisões (gerenciamento de crises) em tempo real.

O trabalho é relevante quando proporciona capacidade de ação emergencial com fins de evitar ameaças à vida dos cidadãos, prejuízos financeiros para os particulares e para o Estado, convulsões sociais, desastres entre outros problemas públicos. “Caracterizadas por relevância, complexidade, inconstância e imprevisibilidade, as emergências públicas costumam causar uma grave ameaça à estabilidade social e à segurança nacional.” (XU; LIU; SHANG, 2017, p. 488, tradução nossa).

O artigo de Kumar e Sharma⁸ faz uma revisão de literatura para detectar o que tem sido publicado com iniciativas governamentais na área de mineração de opiniões e suas diferentes técnicas para favorecer a inteligência governamental: “a pesquisa proposta é um estudo exaustivo dos artigos publicados de aplicações de mineração de opinião de 2011-2017 com o uso de vários conjuntos de dados (*big data*) no campo da governança digital e suas aplicações.”⁹ (KUMAR; SHARMA, 2017, p. 8, tradução nossa). Da mesma forma que o artigo anterior, este avalia a mineração de opiniões em mídias sociais, o que os autores definem como o novo paradigma tecnológico *SMAC* (*Social media, Mobile, Analytics and Cloud*). Kumar e Sharma reforçam que a mineração de opiniões de *big data* oferecida pelas *SMAC's* pode ser usada para aprimorar a tomada de decisão governamental. Essa governança

⁶ XU; LIU; SHANG. *Leveraging cross-media analytics to detect events and mine opinions for emergency management*, 2017, p. 487-506.

⁷ *Characterized by relevance, complexity, infrequency and unpredictability, public emergencies often cause a serious threat to social stability and national security.*

⁸ KUMAR, AKSHI; SHARMA. *Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence*, 2017, p. 6-34.

⁹ “[...] the proposed survey is an exhaustive study of the published articles of opinion mining applications from 2011-2017 with the usage of various datasets (*big data*) in the field of digital governance and its applications.”

digital tem papel fundamental na elaboração, condução e avaliação dos programas e iniciativas do Estado (KUMAR; SHARMA, 2017).

Os artigos selecionados pela base *Scopus* tratam estritamente de modelagem de tópicos. Uma das publicações de 2014¹⁰ de *Jeffrey M. Binder* e *Collin Jennings*, ambos da Universidade de Nova Iorque, EUA, outra de 2017¹¹ de *Akira Murakami*, *Paul Thompson*, *Susan Hunston* e *Dominik Vajn*, de diferentes universidades do Reino Unido.

O artigo de 2014 aborda a modelagem de tópicos não supervisionada e apresenta um algoritmo, *Networked Corpus*, cuja finalidade é permitir a visualização dos tópicos ao lado dos textos. Essa ferramenta permitirá comparar a modelagem atual com a estrutura de tópicos elaborada no século XVIII. Desta forma, permitirá aos pesquisadores encontrar divergência e congruências com a indexação feita no passado, aprimorar métodos e verificar quais eram as prioridades no contexto histórico naqueles documentos.

O texto de 2017 aplica a modelagem de tópicos em um corpus acadêmico específico, *Global Environmental Change (GEC) journal*, totalizando seiscentos e setenta e cinco artigos analisados. Nesses serão identificados os principais tópicos, as mudanças cronológicas e a variedade de trabalhos presentes no GEC. A modelagem também é comparada com outras formas tradicionais de indexação, anotação semântica e análise de palavras-chave, e aponta os pontos fortes da modelagem frente a esses outros. Para os autores a modelagem é “particularmente útil no início da exploração de um corpus.” (MURAKAMI et al., 2017)

Com esta breve descrição do material recuperado na revisão bibliográfica, o referencial teórico teve de seguir técnicas adicionais/complementares para ser redigido. Uma vez que a complexidade dos temas não permitiu recuperar material multidisciplinar que relacionasse, inicialmente, os três grandes temas do trabalho e, depois, o material pesquisado dois a dois foi insuficiente. Sendo assim, as relações entre eles serão estabelecidas ao longo do referencial basicamente por características e vínculos percebidos pelo autor deste trabalho.

¹⁰ BINDER; JENNINGS. *Visibility and meaning in topic models and 18th-century subject indexes*, 2014, p. 405-411.

¹¹ MURAKAMI et al. *What is this corpus about?: using topic modeling to explore a specialised corpus*, 2017, p. 243-277.

3 REFERENCIAL TEÓRICO

Considerando os resultados apresentados na revisão de literatura quando se coletava referências que conjugavam os três principais temas, a opção para elaboração do referencial teórico foi utilizar acervo de artigos e livros da temática de Inteligência de Estado que o pesquisador já possuía e novos adquiridos por pesquisas pontuais, decretos, portarias e leis que dizem respeito à atividade de inteligência. Quanto aos temas ligados ao Aprendizado de Máquina e Modelagem de Tópicos, artigos seminais como dos autores *Allison J.B. Chaney*, *David M. Blei* e *Chong Wang*, bem como artigos posteriores que citaram esses autores indicaram os rumos das leituras subsequentes. As referências basilares e mais utilizadas nos artigos e livros selecionados serviram como referências complementares para o trabalho. De toda sorte, as pesquisas por temas separados (itens ‘a’, ‘b’ e ‘c’ na revisão de literatura retornaram resultados que auxiliaram pontualmente na elaboração deste referencial. A conexão entre os principais temas foram estabelecidas na discussão dos resultados com base no problema de pesquisa, nos objetivos e na justificativa da pesquisa.

Este capítulo está distribuído em quatro seções: Inteligência de Estado, em que se trata da teoria, conceitos, categorias, divisões, aplicações, fins; Gestão do Conhecimento, seguindo a mesma abordagem da seção anterior; em seguida a Inteligência de Estado e Gestão do Conhecimento, que traz as relações transdisciplinares entre as duas áreas, apresentando, principalmente, como a construção científica entre elas segue processo metodológico e conceitual semelhante; por fim, o papel da Inteligência de Estado nas Relações Internacionais, momento em que se relaciona finalidade e importância da IE para os objetivos de um país, como a produção de conhecimento de Inteligência é fundamental ao processo decisório em nível estratégico internacional para um país, por exemplo, o Brasil.

A parte teórica que envolve a metodologia de execução e análise (ferramentas de PLN e as técnicas de Aprendizado de Máquina) dos dados coletados foi tratada em uma seção específica dentro do capítulo de Metodologia.

Optou-se por tratar separadamente IE e GC para consolidar teoricamente cada campo e, posteriormente, facilitar o entendimento das relações estabelecidas de acordo com semelhanças e diferenças, além de demonstrar como essa pesquisa se insere no contexto do Programa de Pós-Graduação Gestão & Organização do Conhecimento. O último capítulo do

referencial vincula-se às justificativas de pesquisa mencionadas e acompanhado da seção teórica sobre o ferramental metodológico proporcionaram junto da interpretação dos resultados discussões e considerações sobre as contribuições desta pesquisa para IE, GC e RI.

3.1 Inteligência de Estado

A história da Inteligência relaciona-se com a necessidade informacional desde períodos bíblicos e em outros como visto no livro “A Arte da Guerra” de Sun Tzu, general do exército chinês no século 500 antes de Cristo. Moisés teria empregado espiões na busca da “terra prometida” e Tzu instruiu a espionagem em seu livro (BRITO, 2011). Não obstante, o termo *intelligence* ser um eufemismo anglo-saxão para a espionagem, esta é apenas uma parte do processo de inteligência” (ANTUNES, 2001, p. 21). Embora a inteligência não se resuma a espionagem, utiliza-se de seu expediente até a presente data. Fato é que a inteligência institucionalizada em organização, métodos e produto é contemporânea – inicia sua estruturação com formação dos Estados Nacionais Europeus no séc. XIX (BRITO, 2011) e consolida-se no século XX no pós Segunda Guerra Mundial.

A IE desenvolve-se ao mesmo tempo de Organizações Internacionais como a ONU, FMI, Banco Mundial e o fenômeno da Guerra Fria. Enquanto Organização, as agências estatais de inteligência “... tornaram-se os olhos e ouvidos do Estado em questões relativas à segurança, guerra e relações internacionais.” (BRITO, 2011, p. 30). Essas questões estão distribuídas em três esferas, que segundo Brito, sustentaram a institucionalização da atividade de inteligência: policiamento político, a guerra tal citado e a diplomacia.

O marco do policiamento político acontece nos movimentos anarquistas, comunistas e socialistas na Rússia czarista; a guerra no período napoleônico com a modernização das tropas e do combate e a necessidade de informação do adversário, tais como quantitativo de armas, soldados, quem eram os líderes e suas táticas; a diplomacia que inicialmente serviu para coletar informações estratégicas dos países em que estavam localizadas as embaixadas, mas que pelo contexto não puderam empregar métodos dos serviços secretos e serviram como base para o pessoal especializado (BRITO, 2011).

É importante a compreensão sobre o processo histórico de institucionalização da atividade de inteligência, a partir das questões relativas à diplomacia, à polícia política e à guerra, na medida em que nos permite identificar dois componentes fundamentais que a caracterizam: a) a presença

de adversários, dos quais se necessita obter informações com vistas a melhor combatê-los, maximizando as chances de ação do Estado; b) a necessidade do desenvolvimento de técnicas e métodos exclusivos por parte de órgãos especializados (CLARK, 2004, p. 13), uma vez que as informações a serem obtidas são naturalmente protegidas por seus proprietários. (BRITO, 2011, p. 32)

Têm-se como termos sinônimos de IE encontrados na literatura: Inteligência Clássica, Inteligência Estratégica e Inteligência Governamental. Priscila Brandão elabora uma definição de IE que corrobora a abordagem de *Michael Herman* (2003) - autor clássico do novo debate anglo-saxão sobre inteligência. Para Herman, inteligência se diferencia em instituição (quem a executa) e prática (atividades realizadas por essas instituições). *Sherman Kent*¹² (1955) diferencia a inteligência em organização, produto e processo, acrescentando à Herman o produto que seriam as análises e estimativas e até mesmo as coletas, buscas e operações encobertas.

Desta forma, quando falamos de atividade de inteligência, ou simplesmente *intelligence*, referimo-nos a certos tipos de informações relacionadas à segurança institucional e à defesa do Estado, às atividades desempenhadas no sentido de obtê-las ou impedir que outros países a obtenham e às organizações responsáveis pela realização e coordenação da atividade na esfera estatal. O papel essencial da atividade de inteligência é determinar as capacidades e as intenções de nações potencialmente adversárias e antecipar aos *policymakers* informações e análises sobre possíveis ameaças. (BRANDÃO, 2010, p.25).

A lei 9883, de 07 de dezembro de 1999, que instituiu o Sistema Brasileiro de Inteligência (SISBIN) e criou a Agência Brasileira de Inteligência (ABIN), em seus artigos 1º, §§3º e 4º, define como inteligência “a atividade que objetiva a obtenção, análise e disseminação de conhecimentos dentro e fora do território nacional sobre fatos e situações de imediata ou potencial influência sobre o processo decisório e a ação governamental e sobre a salvaguarda e a segurança da sociedade e do Estado” e Contraineligência como “a atividade que objetiva neutralizar a inteligência adversa”.

¹² Sherman Kent é uma das referências mundiais no estudo de Inteligência, senão a maior. Ele é o precursor de uma doutrina de Inteligência e atuou junto da CIA no período da Segunda Guerra Mundial. Professor de História na Universidade de Yale na década de 1930, Kent se viu como outros cidadãos americanos a contribuir com os esforços de Guerra de seu país. Foi, então, que em 1941 se juntou a um braço de Pesquisa e Desenvolvimento do Escritório de Serviço Estratégico (futuramente conhecida como CIA) e demonstrou extrema competência em adaptar métodos acadêmicos para análises de inteligência, em gerenciar equipes de acadêmicos renomados e cumprir prazos e entregar resultados satisfatórios aos tomadores de decisão. Publicou obras como *Inteligência Estratégica para Política Mundial Americana* (1949) e *A Necessidade de uma Literatura de Inteligência* (1955). (<https://www.cia.gov/library/kent-center-occasional-papers/vol1no5.htm>. Acesso em 21/01/2020, às 20:10 – adaptado).

Retomamos a referência já feita na introdução para reforçar o que o ordenamento jurídico principal da área no Brasil entende como Inteligência.

Outros autores, leis e decretos também conceituam a atividade de Inteligência. Além da busca, coleta e análise, a Inteligência é “... uma síntese de conhecimentos que se utiliza, inclusive, do julgamento e da intuição. Visa à antecipação e à previsão.” (NASCIMENTO, 2005, p. 259).

O Decreto nº 4.376 de 2002 repete a Lei 9883 ao definir inteligência e amplia na contrainteligência.

Art. 3º Entende-se como contrainteligência a atividade que objetiva prevenir, detectar, obstruir e neutralizar a inteligência adversa e ações de qualquer natureza que constituam ameaça à salvaguarda de dados, informações e conhecimentos de interesse da segurança da sociedade e do Estado, bem como das áreas e dos meios que os retenham ou em que transitem. (Decreto nº 4.376, 2002, p. 01).

A Política Nacional de Inteligência(PNI)¹³, publicada pelo Decreto Nº 8.793, de 29 junho de 2016, complementa e amplia os conceitos da atividade de Inteligência, essa definição é referenciada na Estratégia Nacional de Inteligência (ENINT), publicado pelo Decreto de 15 de dezembro de 2017:

I – Inteligência: atividade que objetiva produzir e difundir conhecimentos às autoridades competentes, relativos a fatos e situações que ocorram dentro e fora do território nacional, de imediata ou potencial influência sobre o processo decisório, a ação governamental e a salvaguarda da sociedade e do Estado;

II – Contrainteligência: atividade que objetiva prevenir, detectar, obstruir e neutralizar a Inteligência adversa e as ações que constituam ameaça à salvaguarda de dados, conhecimentos, pessoas, áreas e instalações de interesse da sociedade e do Estado.

A Inteligência e Contrainteligência são complementares, adjacentes como yin-yang. Na prática, “do ponto de vista operacional, enquanto a principal missão da área de inteligência é tentar conhecer o outro”, a principal missão da área de infosec¹⁴ é garantir que

¹³ <http://www.abin.gov.br/aceso-a-informacao/legislacao-de-inteligencia/coletanea-de-legislacao/politica-nacional-de-inteligencia/>. Acesso em 20/01/2020, às 21h29.

¹⁴ Infosec, informations security ou segurança de informações é a proteção de informações, que uma vez obtidas por um inimigo tornariam O Estado e seus cidadãos vulneráveis e inseguros (CEPIK, 2003). Para tanto, o autor divide infosec em: a) contramedidas de segurança; b) contrainteligência e c) segurança de operações.

os “outros” só conhecerão o que quisermos que eles conheçam sobre nós mesmos.” (CEPIK, 2003, p. 57).

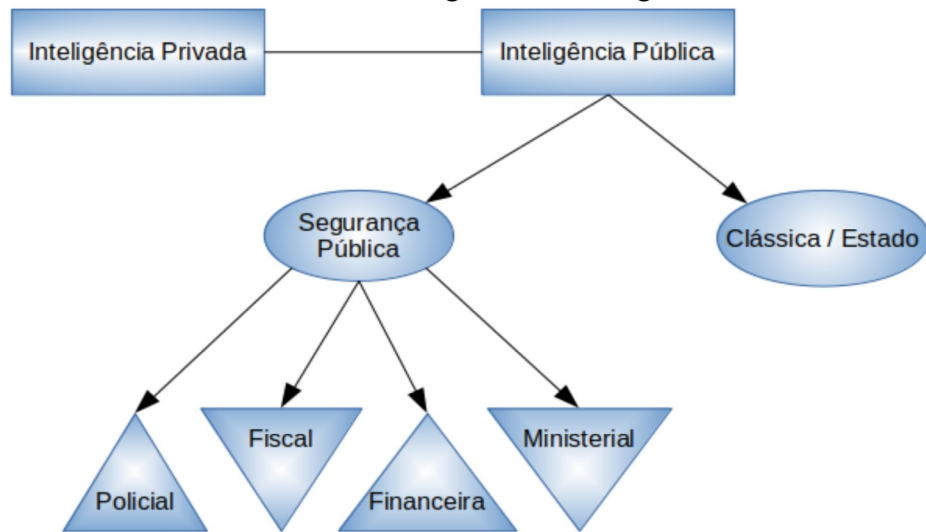
Os elementos essenciais da inteligência dividem-se em “coleta ostensiva e/ou clandestina de informações, análises e estimativas, ação encoberta e contrainteligência, (...) ou simplesmente coleta, análise, contrainteligência e ação encoberta.” (BRANDÃO, 2010, p. 25). Sendo essas as funções da atividade de inteligência, a doutrina brasileira entende o “ciclo de inteligência” ou “processo de inteligência” como: orientação, seguida da produção do conhecimento, depois pela difusão que retorna à orientação.

Dentro da produção do conhecimento há um ciclo próprio: do planejamento, a reunião, processamento (exame, análise, integração, interpretação), difusão, retornando ao planejamento. Cepik (2003), Herman (1996) entre outros pesquisadores propõem mais ou menos etapas nesse ciclo, mas basicamente todos compreendem o que se chama de “ciclo básico.” (GONÇALVES, 2016).

O processo acima exigirá diferentes habilidades técnicas, metodológicas, tecnológicas e de conhecimento: coletores de inteligência são especializados em disciplinas e tipos de fontes e analistas de inteligência são especializados em temas, regiões e tipos de problemas. (ANTUNES; CEPIK, 2003, p. 111)

Almeida Neto faz uma diferenciação da atividade de Inteligência didática e hierarquizada. A Inteligência privada estaria para competitiva/empresarial, enquanto a pública serviria à sociedade através dos servidores. A Inteligência Clássica, subgrupo da pública, assessoraria o chefe de estado nas questões estratégicas (políticas públicas, defesa nacional e política externa). Já a de Segurança Pública assessoraria em nível estratégico e tático as autoridades voltadas à persecução penal (ALMEIDA NETO, 2009).

FIGURA 1 - Categorias de Inteligência



FONTE: Elaborado pelo autor.

Como a Inteligência de Estado é uma atividade de assessoramento ao poder decisório, como visto na definição de Brandão, espera-se, suspeita-se que ela influencie ações dos países em âmbito internacional. Desta forma, *policymakers* (formuladores de políticas ou tomadores de decisão) estariam amparados pela inteligência produzida por órgãos especializados e a levaria em conta, mesmo que inconscientemente, no momento de formulação de políticas, definições e votações domésticas e internacionais, sejam nas relações bi ou multilaterais (ex.: em Organizações Internacionais).

A atividade de IE deve ser encarada como política de Estado, permanente e descolada de programas políticos e/ou políticas públicas de prestação de serviços diretos ao cidadão, sobretudo, desvinculada de interesses particulares de governantes e sem atuar como aparato de policiamento político. É “... reservada para o exterior, das questões mais importantes das políticas exterior, econômica e de defesa; e para o interior, àquelas ameaças susceptíveis de destruir o Estado e o sistema democrático.” (UGARTE, 2003, p. 93).

Brandão organiza oito itens em que a IE deve estar presente e como deve atuar:

[...] espera-se que a inteligência contribua para tornar o processo decisório governamental nas áreas relevantes de envolvimento (política externa, defesa nacional e ordem pública) mais racional e realista, ou seja, menos baseado em intuições e convicções preconcebidas e mais baseado em evidências e reflexão.

Em segundo lugar espera-se que o processo interativo entre policymakers (responsáveis pelas políticas públicas, sejam eles funcionários de carreira, dirigentes nomeados ou políticos eleitos) e oficiais de inteligência produza

efeitos cumulativos de médio prazo, aumentando o nível de especialização dos tomadores de decisões e de suas organizações.

Em terceiro lugar espera-se que a inteligência apoie diretamente o planejamento de capacidades defensivas e o desenvolvimento e/ou aquisição de sistemas de armas, de acordo com o monitoramento das sucessivas inovações e dinâmicas tecnológicas dos demais atores relevantes.

Em quarto lugar, que apoie mais diretamente as negociações diplomáticas em várias áreas, não tanto afetando a definição de política externa, as propiciando ajustes táticos derivados da obtenção de informações relevantes.

Em quinto lugar, que a inteligência seja capaz de subsidiar o planejamento militar e a elaboração de planos de guerra, bem como suportar as operações militares de combate e outras operações cada vez mais relevantes e complexas (operações de paz, assistência, missões técnicas etc.).

Em sexto lugar, a inteligência existe para alertar os responsáveis civis e militares contra ataques surpresa, surpresas diplomáticas e graves crises políticas internas que podem nunca ocorrer, mas para as quais os governantes preferem “assegurar-se” em vez de arriscar.

Em sétimo lugar, sistemas de inteligência monitoram os alvos e ambientes prioritários para reduzir a incerteza e aumentar o conhecimento e a confiança, no caso de implementação e acordos internacionais.

E por fim, sistemas de inteligência servem para preservar o segredo sobre as necessidades informacionais, as fontes, fluxos, métodos e técnicas de inteligência diante da existência de adversários interessados em saber tais coisas. (BRANDÃO, 2010, p. 33-34).

Nos países democráticos que possuem serviços de inteligência, existem controles interno e externos que fiscalizam tais serviços. Ou seja, eles não atuam livremente sem prestação de contas alguma, apesar de terem como mística o segredo. Quanto mais consolidado é o setor de IE em um país, tão mais amadurecido são seus controles. No Brasil o papel de controle e fiscalização é feito sob dois aspectos: ações de inteligência e recursos orçamentários.

No primeiro caso o controle interno da ABIN é feito Câmara de Relações Exteriores (Creden), do Conselho de Governo, também responsável pela supervisão da execução da Política Nacional de Inteligência (PNI), enquanto o externo é feito pela Comissão Mista de Controle das Atividades de Inteligência (CCAI), do Congresso Nacional. No segundo caso, o controle interno é feito pela Secretaria de Controle Interno da Presidência da República (CISSET/PR), que avalia os resultados da gestão financeira e patrimonial nos órgãos integrantes e entidades vinculadas à Presidência da República; enquanto o externo é pelo

Tribunal de Contas da União (TCU), órgão auxiliar do Congresso Nacional, que analisa e julga as contas dos administradores de recursos públicos federais¹⁵.

Os EUA têm um dos serviços de inteligência mais antigos e desenvolvidos do mundo contemporâneo. “Foram décadas empregando os serviços secretos, dentre outras organizações, com vistas a fomentar e fazer prevalecer um ambiente informacional positivo aos interesses estratégicos dessa potência.” (BRITO, 2011, p. 217). A CIA¹⁶ é responsável pela inteligência externa, enquanto a *Federal Bureau Agency*¹⁷ (FBI) é responsável pela inteligência doméstica.

O foco deste trabalho está na CIA e seus documentos desclassificados, pois o intuito é avaliar a visão e atuação internacional do serviço de inteligência norte-americano. A CIA foi criada no pós Segunda Guerra Mundial, em 1947, a partir do *National Security Act*, assinado pelo presidente Truman. O embrião da CIA está em 1941 com o *Office of the Coordinator of Information* (COI), sucedido pelo *Office of Strategic Services* (OSS) em 1942, o qual ofereceu um terço de seus veteranos para CIA. Após o OSS vieram o *Strategic Services Unit* (SSU) em 1945 e o Central Intelligence Group (CIG) em 1946, os quais não existiram por mais de um ano e meio. Por fim, a CIA já completou mais de setenta anos e é uma das agências de inteligência mais consolidadas e atuantes pelo mundo¹⁸.

“A CIA constituiu-se como o principal expoente no combate à União das Repúblicas Socialistas Soviéticas – URSS, em uma disputa de poder e influência mundial” (BRITO, 2009, p. 11). Essa dinâmica marcada pela Guerra Fria impulsionou a atuação internacional da CIA, que através de treinamento e recursos cooperou na criação de serviços de inteligência por diversos países, principalmente na América Latina; de ações encobertas influenciou no surgimento de vários regimes autocráticos pelo mundo. A sua

¹⁵ Ver controle e fiscalização. <http://www.abin.gov.br/aceso-a-informacao/controle-e-fiscalizacao/>. Acesso em 21/01/2020, às 22h05.

¹⁶ Ver visão e missão da agência. <https://www.cia.gov/about-cia/cia-vision-mission-values>. Acesso em 30/05/2019, às 17h20.

¹⁷ Ver prioridades da agência. <https://www.fbi.gov/about/mission>. Acesso em 30/05/2019, às 17h22.

¹⁸ Disponível em: <https://www.cia.gov/about-cia/history-of-the-cia>. Acesso em 30/05/2019, às 17h35.

contrainteligência defende ativamente o país de espionagem estrangeira e consegue alterar a percepção de realidade dos mesmos através de desinformação (BRITO, 2009).

Os mecanismos de controle interno e externo também funcionam com mais desenvoltura, a CIA possui em seu próprio site uma seção, *Freedom of Information Act Electronic Reading Room(FOIA)*, que disponibiliza documentos desclassificados relacionados à IE. Essa é a fonte da qual a pesquisa utilizou os documentos submetidos à análise.

Estudar e entender as práticas de um dos melhores serviços de inteligência do mundo é muito importante como ação de *benchmarking*¹⁹(guardadas a devidas analogias e proporções com a Administração) de inteligência e como projeção de cenários para a contrainteligência.

Portanto, a compreensão e análise dos documentos selecionados ainda que não contemporâneos são uma peça importante e ponto de referência de práticas, intenções e fonte de conhecimento para evolução em eficácia e efetividade. Isso pode proporcionar uma avaliação, adequação e melhoria de serviços de outros países, inclusive do Brasil, e auxiliar nas projeções e ações internacionais dos mesmos, bem como na proteção dos seus interesses intra e extralimites territoriais.

Tão importante é compreender a conexão científica entre IE e CI através da Gestão do Conhecimento, como o ciclo de produção de conhecimento é semelhante, como as pirâmides agregadoras se estratificam se relacionam, entre outros vínculos presentes. Contudo, antes abordar-se-á a Gestão do Conhecimento isoladamente, conceitos básicos e fundamentais, um resgate breve, sucinto, mas com elementos suficientes que permitirão assimilar melhor as conexões propostas com a IE.

3.2 Gestão do Conhecimento

Tal como feito para IE, esta seção pretende orientar em torno de conceitos e teorias fundamentais sobre a Gestão do Conhecimento (GC). Em seguida, serão demonstradas as relações entre IE e GC como metodologias que manipulam dados, informações, produzem

¹⁹ “O **benchmarking** é uma das mais relevantes estratégias para aumentar sua eficiência. Em tradução livre, pode ser traduzido como “ponto de referência”. Trata-se um minucioso processo de pesquisa que permite aos gestores compararem produtos, práticas empresariais, serviços ou metodologias usadas pelos rivais, absorvendo algumas características para alçarem um nível de superioridade gerencial ou operacional.” <https://endeavor.org.br/estrategia-e-gestao/benchmarking/>. Acesso em 30/05/2019, às 18h29.

e gerenciam conhecimentos que tem finalidades de assessoramento aos tomadores de decisões de organizações.

Para Choo (2003), as organizações (do conhecimento) operam em ambiente dinâmico e devem estar preparadas para se sustentar e desenvolver ao risco de perecerem se não tiverem capacidade de adaptação e de anteverem às mudanças com antecedência. “Num nível mais profundo, a organização do conhecimento possui informações e conhecimentos que lhe conferem uma vantagem, permitindo-lhe agir com inteligência, criatividade e, ocasionalmente, com esperteza.” (CHOO, 2003, p. 17).

Para o autor, uma organização do conhecimento é aquela que atualiza constantemente seus conhecimentos (inovando e reaprendendo), imersa em ambiente volúvel (interpretado e analisado para seu contexto de atuação), processando ativamente informações voltadas à tomada de decisões ou melhores e mais apropriados curso de ação. Sinteticamente, o autor adota um modelo tripartite para GC: construção de sentido (*sensemaking*), criação de conhecimento e tomada de decisão.

É neste cenário (organizações do conhecimento) que a GC se insere. Nele são incentivados a criação de novos conhecimentos através da conversão, construção e conexão²⁰, além do gerenciamento a aplicação dos mesmos diante da necessidade de escolha e do formato/modo de tomada de decisões que podem ser: racional-limitado, conduzido por programas de desempenho e procedimentos operacionais padrão; processual, com objetivos estratégicos e claros, porém técnica de ação incerta; político, no qual os objetivos são avaliados de acordo com interesses de diversos grupos de interesse e as decisões e ações alcançadas após barganhas entre eles e, por fim, anárquico, sem uma técnica bem definida, os objetivos são buscados discricionariamente e as decisões e ações são circunstanciais (CHOO, 2003). Nonaka e Takeuchi (1995) também entendem que a base da GC é a criação de conhecimento, tanto na forma de interação quanto no nível.

Alvarenga Neto, Barbosa e Pereira (2007) defendem o pressuposto que a GC transcende a Gestão da Informação (GI) - a GI seria um dos componentes da GC. A amplitude

²⁰ Choo (2003) estratifica os autores que trabalham cada uma das relações: conversão – em *The Knowledge-Creating Company: how Japanese Companies Create the Dynamics Innovation*. Takeuchi e Nonaka (1995); construção – em *Wellsprings of Knowledge: Building and Sustaining the Sources Innovation*. D. Leonard-Barton (1995); conexão – em *The Knowledge Link: how Firms Compete through Strategic Alliances*. J. L. Badaracco (1991) e *Knowledge and Value: a New Perspective on Corporate Transformation*. S. Wikström & R. Normann (1994).

da GC é maior e mais complexa, abarca mais contextos e considerações e tem outros objetivos e metas para uma organização, a exemplo citado anteriormente como criação de conhecimento e seu uso. A GC inclui e incorpora “[...] outros aspectos, temas, abordagens e preocupações, como questões de criação, uso e compartilhamento de informações e conhecimentos, além da criação de contextos organizacionais favoráveis, dentre outros.” (ALVARENGA NETO; BARBOSA; PEREIRA, 2007, p. 8).

Porém antes da junção dos termos “Gestão” e “Conhecimento”, o que eles significam nesse contexto separadamente, do que tratam? Gestão está além de simples controle, é algo mais profundo. Na visão de Alvarenga Neto, Barbosa e Pereira (2007) é relacionado ao conceito denominado por Krogh, Ichijo & Nonaka (2001) de “contexto capacitante” - promoção/geração de atividades criadoras de conhecimento nas organizações. O Conhecimento tratado por Drucker (1993) como único recurso econômico na sociedade pós-capitalista (ou do conhecimento) é para Davenport e Prusak (1998) uma informação mais valiosa, munida de contexto, significado, interpretada – conhecimento ressignificado com aplicação de sabedoria daquele que o analisou. “O conhecimento é uma mistura fluida de experiência estruturada, valores, informações contextuais e conhecimentos especializados que fornecem uma estrutura para avaliar e incorporar novas experiências e informações.” (DAVENPORT; PRUSAK, 1998b, p. 5) E aqui pode ser entendida como a tríade de conversão, construção e conexão.

Nonaka e Takeuchi (1995) discutem a complexidade do conhecimento, os paradoxos ambientais com os quais as organizações têm trabalhado e tentado se adaptar, as aparentes contradições organizacionais que as empresas enfrentam. A sociedade do conhecimento não evita ou elimina os paradoxos, ela os incorpora e estimula. Nela [...] as contradições, as inconsistências, os dilemas, as dualidades, as polaridades, as dicotomias e as oposições não são alheios ao conhecimento.” (NONAKA; TAKEUCHI, 1995). Para os autores o conhecimento é constituído de dois componentes dicotômicos e aparentemente opostos: o conhecimento explícito e o tácito.

O conhecimento explícito pode ser expresso em palavras, números ou sons, e compartilhado na forma de dados, fórmulas científicas, recursos visuais, fitas de áudio, especificações de produtos ou manuais. O conhecimento explícito pode ser rapidamente transmitido aos indivíduos, formal e sistematicamente.

O conhecimento tácito, por outro lado, não é facilmente visível e explicável. Pelo contrário, é altamente pessoal e difícil de formalizar, tornando-se de

comunicação e compartilhamento dificultoso. As intuições e os palpites subjetivos estão sob a rubrica do conhecimento tácito. O conhecimento tácito esta profundamente enraizado nas ações e na experiência corporal do individuo, assim como nos ideais, valores ou emoções que ele incorpora. (NONAKA; TAKEUCHI, 1995, p. 19)

Choo (2003) destaca um terceiro tipo de conhecimento, além do tácito e explícito, com base no trabalho de Boisot (1995)²¹ – o conhecimento cultural. Para o autor, o conhecimento tácito é implícito aos membros da organização, não codificado e de difícil divulgação, aprendido com a experiência de anos de execução de certa atividade, o que amplia a capacidade de análise e discernimento de forma intuitiva; é através do incentivo do desenvolvimento deste tipo de conhecimento que as empresas aprendem e inovam.

O conhecimento explícito é o formalizado em um sistema de símbolos (palavras, números ou fórmulas) ou em objetos físicos (equipamentos, documento e modelos), portanto, mais facilmente disseminado e explicado; quando codificado em procedimentos operacionais padrão, regras ou rotinas, o conhecimento explícito está baseado em regras. “O conhecimento cultural consiste em estruturas cognitivas e emocionais que habitualmente são usadas pelos membros da organização para perceber, explicar, avaliar e construir a realidade.” (CHOO, 2003, p. 190).

Nesse sentido, está para o conhecimento de senso comum de Boisot – as suposições e crenças, convenções e expectativas que descrevem e explicam uma realidade, bem como agregam valor e significado a uma nova informação.

É preciso considerar que, em que pese a sua essência ontológica, que se caracteriza pela intangibilidade e pela produção exclusiva do indivíduo, o conhecimento se manifesta ou se materializa no desenvolvimento do conjunto de práticas e ações organizacionais cotidianas. Pode-se dizer que sua tangibilidade se insere nos diversos processos e práticas organizacionais, e nos produtos e serviços deles decorrentes. (SOUZA; DIAS; NASSIF, 2011, p. 68)

Embora a intenção seja focar na GC, é importante destacar que há certa ambiguidade entre Gestão da Informação e do Conhecimento, tanto conceitualmente como na prática. Há ainda a utilização da expressão composta, Gestão da Informação e do Conhecimento (GIC), que se refere à dinâmica do processo de conhecer, na relação da informação com o conhecimento (SOUZA; DIAS; NASSIF, 2011). A GIC trabalhada por Barbosa (2008) resgata origens e conceitos sobre Gestão da Informação (GI) e Gestão do

²¹ Boisot, M. H. *Information Space: a Framework for Learning in Organizations, Institutions and Culture*. Londres, Routledge, 1995.

Conhecimento (GC), que “embora compartilhando importantes aspectos comuns, possuem suas particularidades” (BARBOSA, 2008, p. 13). A GIC seria um processo complementar e multidisciplinar entre as áreas citadas que possuem componentes indissociáveis. Para o autor a GI tem como fenômeno central a informação ou conhecimento explícito, enquanto a GC teria o conhecimento tácito, as competências pessoais.

A sinergia entre esses aspectos seria a expressão da GIC, que na visão de Barbosa dependeria da competência informacional dos elementos humanos e organizacionais. “Descobrir, organizar, disseminar e utilizar esse conhecimento de maneira efetiva constitui o grande desafio da gestão da informação e do conhecimento.” (BARBOSA, 2008, p. 25).

A correlação entre gestão da informação e gestão do conhecimento, e, por conseguinte, a conjugação de seus estudos, parece ter por fundamento o entendimento de que as estruturas de informação disponíveis nos diversos suportes e recursos informacionais são também estruturas de conhecimento. De outra forma, da mesma maneira que uma estrutura de informação se apresenta como potencialmente informacional, uma estrutura de informação se apresenta também, embora que em outro nível, como estrutura de conhecimento e, por conseguinte, como possibilidade de gerar novo conhecimento. (SOUZA; DIAS; NASSIF, 2011, p. 60)

Souza e Alvarenga Neto (2003), através de revisão de literatura, elaboram um mapa conceitual a partir da perspectiva de elaboração de um modelo geral de GC o pode ser expressa através da convergência de três pilares: o modelo proposto Choo (2003); a ideia do contexto capacitante como sugerida por Krogh, Ichijo & Nonaka (2001); e a metáfora do ‘guarda-chuva’.

A criação de significado é processo social contínuo em que os indivíduos observam fatos passados, recortam pedaços da experiência e selecionam determinados pontos de referência para tecer redes de significado, resulta em um ambiente interpretado ou significativo, que é uma tradução razoável e socialmente crível do que está acontecendo. Tem como principal problema reduzir ou eliminar a ambiguidade e criar significados comuns para que a organização possa agir coletivamente (CHOO, 2003).

Porém, o autor ressalta a necessidade da organização não tentar erradicar a ambiguidade ou estabelecer um consenso. É importante mantê-la em certo nível acomodando as diversas interpretações, proporcionado um espaço cognitivo para que seus membros procurem, experimentem e descubram, fundamental ao processo criativo.

A etapa subsequente é a criação do conhecimento, que envolve, além das dimensões de construção e conexão, a conversão do conhecimento em seus processos de socialização, exteriorização, combinação e internalização, trabalhados por Nonaka e Takeuchi (1995) e explicitadas nas figuras abaixo na espiral do conhecimento e nos conteúdos intrínsecos dos quatro modos de conversão caracterizando os tipos de conhecimento envolvidos em cada um. Percebe-se que o processo de conversão é contínuo e cumulativo, enquanto os tipos de conhecimento são específicos em cada fase.

FIGURA 2 - Espiral do Conhecimento

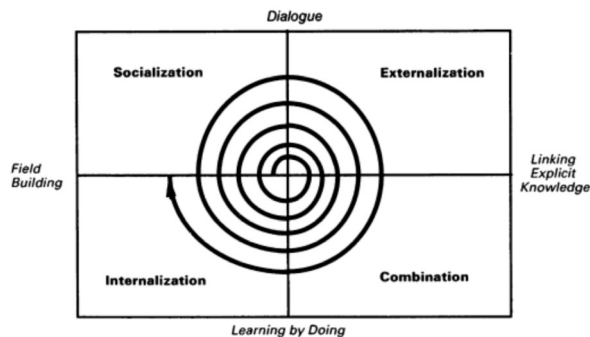


FIGURA 3 - Conteúdos de conhecimento

		Tacit knowledge	To	Explicit knowledge
From	Tacit knowledge	(Socialization) Sympathized Knowledge		(Externalization) Conceptual Knowledge
	Explicit knowledge	(Internalization) Operational Knowledge		(Combination) Systemic Knowledge

FONTE: Nonaka e Takeuchi, 1995, p. 71-72.

Nesta etapa as organizações criam ou adquirem, organizam e processam a informação com o propósito de gerar novo conhecimento através da aprendizagem organizacional. “O novo conhecimento gerado permite que a organização desenvolva novas habilidades e capacidades, crie novos produtos e serviços, aperfeiçoe os antigos e melhore seus processos organizacionais.” (SOUZA, R. R.; ALVARENGA NETO, 2003, p. 6).

Choo (2003) define que as organizações criam e exploram conhecimento para desenvolver novas capacidades e inovações por meio de três atividades concomitantes: gerar e compartilhar conhecimento tácito; testar e criar protótipos de conhecimento explícito; e extrair e aproveitar conhecimento externo. Nesse sentido o autor elabora um quadro sintético em que relaciona os processos de criação de conhecimento na perspectiva de Wikstrom & Normann (1994), Nonaka e Takeuchi (1995) e Leonard-Barton (1998).

FIGURA 4 - Processos de criação do conhecimento

Processos de conhecimento (S. Wikström & R. Normann, <i>Knowledge and Value: a New Perspective on Corporate Transformation</i> , cit.)	Fases de conversão do conhecimento (I. Nonaka & H. Takeuchi, <i>The Knowledge-Creating Company: how Japanese Companies Create the Dynamics of Innovation</i> , cit.)	Atividades de construção do conhecimento (D. Leonard-Barton, <i>Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation</i> , cit.)
Processos gerativos geram novos conhecimentos	Partilhar o conhecimento tácito	Solução compartilhada de problemas
	Criar conceitos	Experimentação e prototipagem
Processos produtivos operacionalizam novos conhecimentos	Justificar conceitos	Implementação e integração de novos processos e ferramentas
	Construir um arquétipo	
Processos representativos difundem e transferem novos conhecimentos	Disseminar o conhecimento	Importação do conhecimento

FONTE: Choo, 2003, p. 219.

A terceira etapa para Choo (2003) é a tomada de decisão, que está alicerçado no pressuposto de racionalidade limitada. Esse conceito trabalha com a incerteza, pois na maioria das vezes, senão em todas, não se conhecerá todas alternativas para uma certa escolha, nem mesmo se saberá a dimensão das possíveis consequências.

O autor reforça que não se tem tempo e recursos para atingir esse conhecimento total, que a alternativa colocada em prática produzirá consequências desejadas e indesejadas, e as não previstas podem se revelar altamente significativas: raramente um conjunto definido ou totalmente coerente de preferências ou critérios pelos quais analisar todas as alternativas e escolher a mais desejável estará presente.

Sendo assim, a tomada de decisão leva a um curso de ação baseado nos propósitos da organização em ambientes complexos, dinâmicos e cheios de incertezas. Por isso, as “[...] organizações elaboram e implementam regras e rotinas para simplificar e orientar o comportamento decisório, de modo que ele seja coerente e coordenado, ao menos num nível mínimo.”(CHOO, 2003, p. 325). É nessa tentativa que se instala o paradoxo explicado pelo autor, na tentativa de alcançar as decisões mais favoráveis.

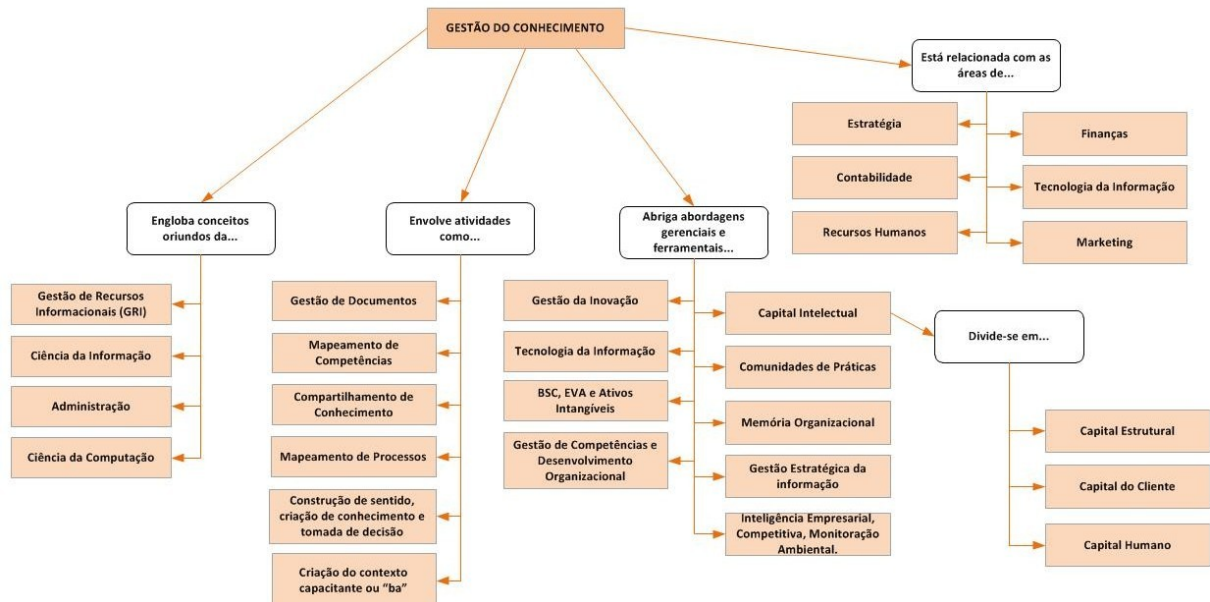
Embora esses valores sejam necessários para manter a legitimidade e credibilidade da organização, em algumas situações de decisão (como, por exemplo, quando os níveis de tensão, incerteza e complexidade são altos) a busca obstinada da conformidade pode restringir o uso da informação e fazer com que os que devem tomar as decisões deixem de registrar informações vitais. Daí podem resultar decisões errôneas [...]. Em segundo lugar, os procedimentos decisórios padronizados armazenam e operacionalizam aquilo que a organização aprendeu com a experiência. Embora os procedimentos padronizados tenham o propósito de permitir que mais sistemas funcionem no nível mais alto de inteligência aprendida, sua promulgação pode ter o efeito contrário quando uma adesão acrítica à tradição desestimula a inovação e a improvisação e impede que a organização se livre de pressupostos superados. Em terceiro lugar, uma organização não é um monólito que defende um conjunto universal de preferências e valores. Objetivos e metas não se alinham ordenadamente em hierarquias, mas estão constantemente mudando em razão de barganhas e da formação de coalizões. As escolhas às vezes são feitas mais com base em concessões ou competição do que numa avaliação racional dos méritos dos resultados. (CHOO, 2003, p. 325).

Seguindo a proposta de Souza e Alvarenga Neto, o próximo item trata-se do “Contexto Capacitante” trabalhado por Krogh, Ichijo & Nonaka (2001), no que tange ao emprego da palavra “Gestão” no conceito de GC. Gestão é a “[...] promoção de atividades criadoras de conhecimento em nível organizacional. A questão colocada é como criar o contexto que capacite as pessoas a compartilhar mais informações e conhecimentos”(SOUZA, R. R.; ALVARENGA NETO, 2003, p. 7), dinâmicas exemplificadas nas **Figuras 2 e 4**.

Finalizando, antes de apresentar o mapa conceitual, os autores discutem a metáfora do “guarda-chuva”, sob o qual a inter-relação e a permeabilidade de temas como gestão estratégica da informação, gestão do capital intelectual, aprendizagem organizacional, inteligência competitiva, comunidades de prática, etc, possibilitam e delimitam a formação de um possível referencial teórico de sustentação que se intitulou Gestão do Conhecimento.

Os elaboradores do mapa conceitual analisam a representação e concluem que quaisquer atividades desenvolvidas em organizações representativas da “era da informação” são relacionadas à GC, que se apresenta como uma filosofia a embasar e motivar as instituições internas, atividades, processos e relacionamentos das modernas organizações. (SOUZA; ALVARENGA NETO, 2003).

FIGURA 5 - Mapa Conceitual da Gestão do Conhecimento



FONTE: Souza e Alvarenga Neto, 2003, adaptado.

A Gestão do Conhecimento deve ter como pressuposto que, pelo menos o produto final seja o conhecimento, “[...] deve ser concebida num espaço epistemológico amplo dedicado à compreensão da relação mente-mundo, onde se dá o processo de conhecer.” (SOUZA; DIAS; NASSIF, 2011, p. 57). Para os autores a GC se encontra em um estágio posterior à da GI, posto que o conhecimento está, qualitativamente, em escala superior à informação; acontece no ambiente organizacional em meio a suas práticas e cultura e depende do planejamento e da construção desse espaço.

A GC é atualmente tida como principal vantagem competitiva de uma organização, tem foco no conhecimento que permeia sua estrutura. “Está enraizado em todas as disciplinas, incluindo negócios, economia, psicologia e gestão da informação. [...] A gestão do conhecimento envolve pessoas, tecnologia e processos em partes sobrepostas²².” (AWAD; GUAZIRI, 2003, p. 26, tradução nossa). Paiva, Aragão e Pereira (2005) analisam diversos conceitos e modelos de GC e detectam dois elementos recorrentes: pessoas e tecnologias. Para os autores, o que se busca em relação as pessoas é mapear os conhecimentos, as experiências e as habilidades, a fim de melhor aproveitá-los, enquanto as tecnologias da informação servem

²² It is rooted in any disciplines, including business, economics, psychology, and information management. [...] Knowledge management involves people, technology, and processes in overlapping parts.

como ferramentas para a captação, o armazenamento e o compartilhamento de informações e de conhecimentos nas organizações. “A gestão do conhecimento corresponde a uma série de ações que buscam coordenar desde os processos de geração até a utilização do conhecimento, passando pelo seu armazenamento e o compartilhamento entre os indivíduos” (PAIVA; ARAGÃO; PEREIRA, 2005, p. 38). Para Souza, Dias e Nassif (2011) a importância da GC se encontra no fato de diversos setores e profissionais da economia compreenderem o conhecimento como principal ativo das organizações, uma vez que este se apresenta como o principal elemento na competitividade sustentável.

Com essa breve abordagem sobre GC é possível avançar para o paralelo com a IE. A apresentação de conceitos, relações e análises comparativas na próxima seção se dará com mais tranquilidade e fluidez, uma vez que já foram apresentados os dois campos em separado. As informações contidas nessas seções proporcionam um conhecimento básico e elementar necessário para compreender a integração disciplinar que se pretende estabelecer entre IE e GC.

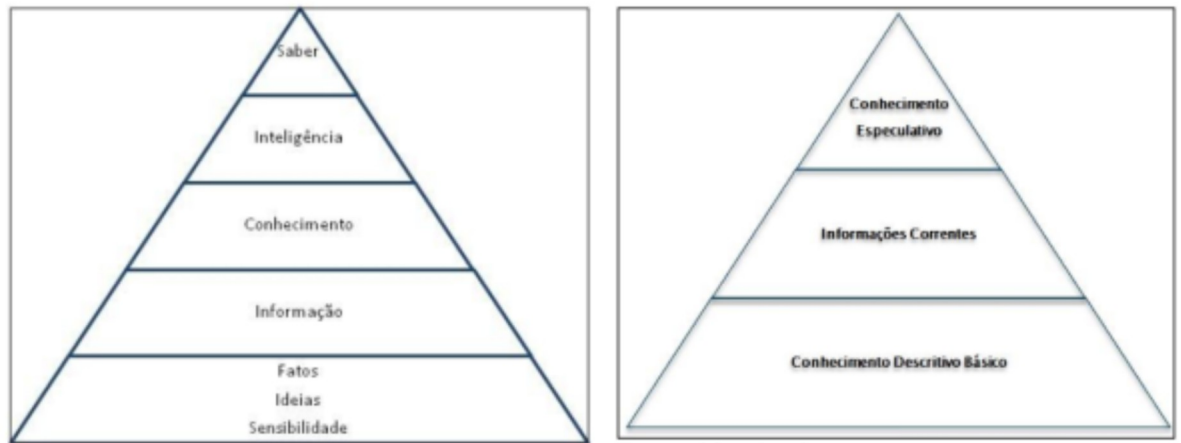
3.3 Inteligência de Estado e Gestão do Conhecimento

Primeiramente, deve-se estabelecer que existe relação transdisciplinar e/ou interdisciplinar entre Ciência da Informação e Inteligência de Estado, em seguida os aspectos da Gestão do Conhecimento (conteúdo da CI) conectados à IE, mostrarão como está estabelecida essa relação. Pombo (1994) trabalha o modelo em tríade (pluri, inter e transdisciplinaridade) constituinte da integração disciplinar, que aplicada à relação que se estabeleceu entre IE e GC, segue basicamente o paralelo da abordagem defendida pelo autor. Para ele a integração disciplinar mais que outras práticas científicas, situa-se, epistemologicamente, na coordenação de conceitos, métodos e teorias, e em parte isso é o que se pretende apresentar em seguida demonstrando a relação entre IE e GC.

A transdisciplinaridade presente entre CI e a IE foi trabalhada por Brito (2009) quando relacionam a Inteligência Governamental e a Inteligência Competitiva (objeto de estudo da CI). Ambas teriam semelhantes processos de produção de conhecimento, além de terem objetivos de subsidiarem o poder decisório em suas instâncias, guardadas devidas especificidades históricas, culturais, metodológicas e de finalidade.

Maciel e Pinheiro (2013) relacionam duas pirâmides de Inteligência, uma elaborada por Barreto (2002), que tem uma abordagem mais genérica (representativa da Ciência da Informação) e a pirâmide elaborada pelo autor com base nos conceitos de Inteligência de Estado de *Sherman Kent*.

FIGURA 6 - Pirâmides de Inteligência



FONTE: Barreto, 2002, p. 5 (esq) / Maciel e Pinheiro, 2013, p. 37 (dir.)

Os autores concluem sobre a diferença nos objetivos da produção de conhecimento delinearem a forma de trabalhar e enxergar a informação, a importância de se entender os diferentes contextos, no caso de Barreto - o empirismo acadêmico – e no de *Kent* – seu trabalho de assessoria política no contexto da Segunda Guerra Mundial e Guerra Fria.

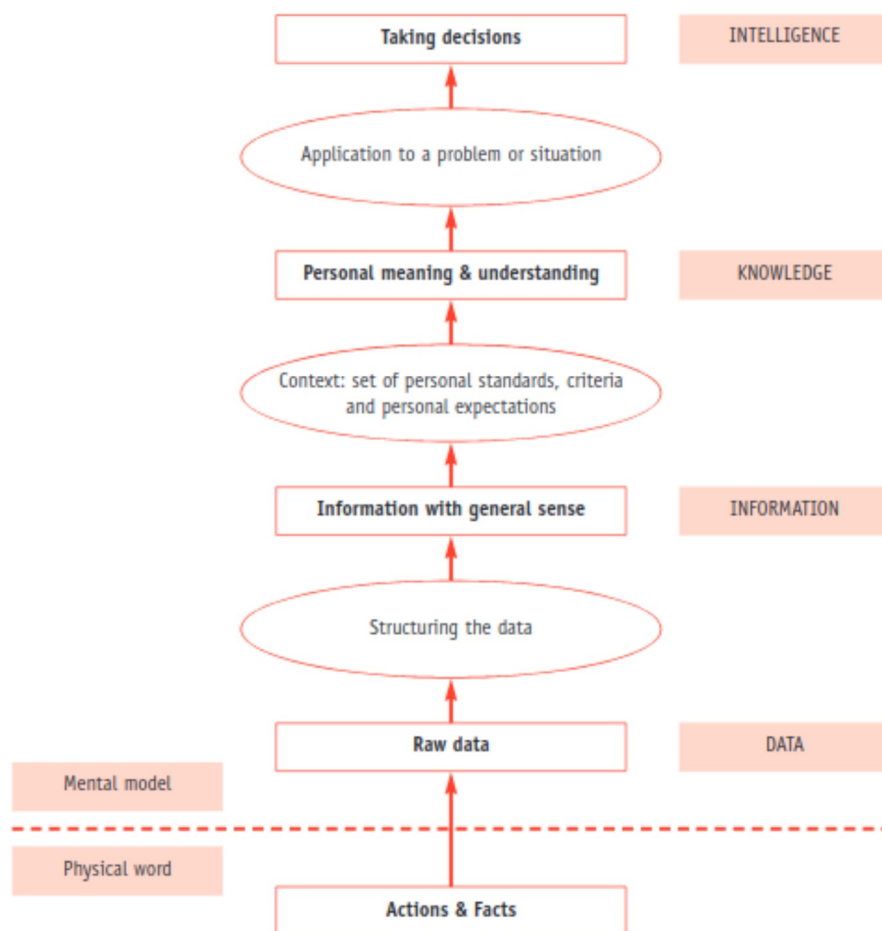
“Desse modo, podemos articular que as hierarquias do conhecimento se apresentaram diferentes devido ao fato de que cada um dos formuladores das pirâmides aqui analisadas tinham finalidades díspares.” (MACIEL; PINHEIRO, 2013, p. 38). Todavia, é possível perceber a construção acumulativa, na transformação de elementos mais básicos até os mais elaborados e analíticos nas duas pirâmides e isso é a congruência metodológica entre a CI e a IE.

Em ambas atividades percebe-se a metáfora da pirâmide representando a forma pela qual os profissionais devem abordar a informação como meio para se alcançar conhecimento preditivo a ser utilizado no processo decisório (MACIEL; PINHEIRO, 2013). Importante congruência detectada pelos autores entre *Sherman Kent*, *Karl Popper* e *Anthony Olcott* é o fazer científico, o caráter objetivo que envolve a coleta de informações, a possibilidade de verificação ou prova dos dados/informações, de mínima subjetividade e

incerteza. Para Kent os “... elementos descritivos básicos e as informações correntes conteriam dados cuja verificação possui constatação e significado que possam ser compreendidos por todos e não haja dúvida a respeito de sua existência, ou sequer outra interpretação a respeito deles.” (MACIEL; PINHEIRO, 2013, p. 36).

A abordagem de Cetisme (2002) é semelhante à de *Kent* e podemos perceber as coincidências metodológicas sobre como fazer “inteligência”. Qual é a sequência a seguir, já que o processo é cumulativo, pragmático e sempre orientado a resultados. Através da figura abaixo percebe-se a sequência demonstrada por Cetisme em relação à pirâmide de *Kent*.

FIGURA 7 - Dados transformam-se em Inteligência



FONTE: Cetisme (2002, p. 80)

Antes de continuar a tratar de Inteligência é necessário realizar um aparte conceitual, pois a Inteligência como dito e visto na figura acima está alicerçada em outros elementos, que precisam ser conhecidos.

A figura de Cetisme (2002) mostra uma evolução agregadora iniciando pelo dado, passando por informação, seguido por conhecimento até o ápice em inteligência. É fundamental caracterizar os elementos agregadores no ciclo da produção de conhecimento em inteligência. Davenport e Prusac (1998) diferenciam em uma tabela cada um dos três. Os autores ressaltam a dificuldade de distingui-las na prática, embora seja útil.

Por exemplo, “informação, além do mais, é um termo que envolve todos os três, além de servir como conexão entre os dados brutos e o conhecimento que se pode eventualmente obter.” (DAVENPORT; PRUSAK, 1998, p. 18). A precisão entre elas é difícil de definir, pois a transição se faz com facilidade desde que obedecido alguns critérios: dado é uma observação crua, uma fotografia; informação, segundo Peter Drucker, é o dado dotado de relevância e propósito; conhecimento é a informação interpretada em um contexto, com significado, é a agregação de valor (DAVENPORT; PRUSAK, 1998).

QUADRO 1 - Dado, Informação e Conhecimento

Dados	Informação	Conhecimento
Simples observações sobre o estado do mundo Facilmente estruturado • Facilmente obtido por máquinas • Frequentemente quantificado • Facilmente transcrível	Dados dotados de relevância e propósito • Requer unidade de análise • Exige consenso em relação ao significado • Exige necessariamente a mediação humana	Informação valiosa da mente humana Inclui reflexão, síntese, contexto • De difícil estruturação • De difícil captura em máquinas • Frequentemente tácito • De difícil transferência

FONTE: Davenport e Prusac (1998, p. 18)

A abordagem sobre o excesso e entropia da informação trabalhada por Maciel e Pinheiro (2013) é relacionada a um fenômeno similar e que precedeu o que se chama hoje de *Big Data*.

Uma quantidade crescente de informação produzida nas décadas de 1940-50 que necessitava ser trabalhada e burilada para gerar conhecimento útil; um processo de racionalização: técnico (SHANNON; WEAVER, 1964) e quantitativo (CORNELIUS, 2002) de tratamento da informação, possível com uma visão materialista, que enxergasse a

informação como objeto e/ou coisa passível de ser manipulado como qualquer outro bem de consumo (BUCKLAND, 1991) são os contextos de surgimento da CI e de uma visão mais tecnológica sobre a gestão da informação (SARACEVIC, 1996). Essa situação era enfrentada por *Sherman Kent* no mesmo período histórico em seu trabalho na CIA (MACIEL; PINHEIRO, 2013).

Embora a lógica de Sherman Kent se baseie numa acumulação frequente de dados, há que se fazer a ressalva de que em determinados casos o excesso de informação e sua massificação não reduzem incertezas (CAPURRO, 2007; CORNELIUS, 2002; WERSIG, 1997), mas por seu caráter entrópico, retiram a sua capacidade de construção e de representação em conhecimentos. (MACIEL; PINHEIRO, 2013, p. 33).

Um sistema de inteligência possui quatro elementos (o usuário da atividade de Inteligência; as políticas que orientam a busca e a coleta de informações; o produtor que elabora o conhecimento e o conhecimento propriamente como produto) e suas relações: formulação e planejamento de políticas pelo usuário; a orientação sobre o que e como reunir na forma de conhecimento para tal usuário; a produção do conhecimento pelo responsável de etapa; por fim, a difusão e utilização deste conhecimento pelo usuário (SANTOS, 2016). A figura abaixo permite a compreensão do processo.

FIGURA 8 - Elementos de um Sistema de Inteligência



FONTE: Santos (2016, p. 28)

Tarapanoff (2006) é uma das defensoras da interdisciplinaridade da CI, haja vista a informação como elemento da cultura. Santos (2016) reitera tal assertiva e sintetiza o contexto informacional estratégico através da interface entre IE e CI visto no processo de produção do conhecimento, no sistema de informação, no produto conhecimento, no processo decisório e no usuário da informação.

Wickramasingue *et al.* citam Davenport e Prusak(1998) e Swan *et al.* (1999) e trazem a seguinte definição de GC, muito próxima de IE:

[...] um conjunto de ferramentas, técnicas, táticas e tecnologias destinadas a maximizar os ativos intangíveis de uma organização por meio da extração de dados relevantes, informações pertinentes e conhecimento pertinente, para facilitar a tomada de decisão superior, para que uma organização atinja e mantenha vantagem competitiva sustentável. (WICKRAMASINGHE *et al.*, 2009, p. 1).

A GIC assim como a Inteligência trabalha com uma abordagem científica: produção, organização e aplicação da informação e do conhecimento. Nesse sentido, a informação e conhecimento servem como acessórios, meio a um fim, qual seja subsidiar estrategicamente algum elemento humano a tomar decisões em seu campo de poder/ação (organizações privadas ou estatais).

A visão estratégica da GIC sintetiza o trinômio – Informação, Conhecimento e Inteligência – como um produto de apoio ao processo de tomada de decisão estratégica, denominado Conhecimento de Inteligência, que se caracteriza por sua natureza proativa, ou seja, antecipa-se aos fatos e busca obter, analisar, interpretar e disseminar conhecimentos aos tomadores de decisão da alta cúpula para a defesa dos interesses da segurança da sociedade e do Estado. (SANTOS, 2016, p.30).

Argumento defendido por Santos (2016) com base no artigo de Barbosa (2008) - do qual concordamos - são os vínculos com as áreas de conhecimento (Recursos Humanos, Finanças, Estratégia, Computação e CI) apresentados na representação multidisciplinar integrativa da GIC com a Inteligência de Estado. Visto que a IE é um processo complexo e cíclico, que bem como a GIC envolve a Informação e o Conhecimento de forma dinâmica.

FIGURA 9 - Vínculos da GIC com outras áreas do conhecimento



FONTE: Santos (2016, p. 32)

A autora trabalhou de forma mais extensa as relações com a IE em cada campo do saber. Aqui, porém citou-se apenas uma em cada área com base na figura acima: no quesito Recursos Humanos a Inteligência utiliza-se das competências; nas Finanças, do capital intelectual; na Ciência da Informação, da organização e do conhecimento; na Computação, dos sistemas de informação e, sem dúvida, na Estratégia, da informação, conhecimento e TI como vantagens competitivas.

O mundo vivencia uma era chamada “sociedade do conhecimento” ou “sociedade pós-industrial”, na qual as transformações são intensas e rápidas, as informações são estratégicas e o capital humano é fundamental para o desenvolvimento econômico (LUCCI, 2009). O capital humano é o responsável por transformar a informação em conhecimento e o “conhecimento passou a representar o elemento central das novas estruturas econômicas e a exercer forte influência sobre o nível de desenvolvimento de cada país.” (BALUÉ; NASCIMENTO, 2007, p. 2).

A competitividade internacional entre países seja através de estatais ou suas multinacionais é tratada em nível estratégico e a proteção do conhecimento “sensível” é imprescindível à garantia da soberania e ao potencial de ganho econômico (BALUÉ;

NASCIMENTO, 2007). Nascimento defende uma cultura de proteção do conhecimento atual com maior rigor em relação às tradicionais práticas de segurança, sobretudo quando a evolução das TIC's revolucionaram as formas de obter, registrar, armazenar e compartilhar informações (NASCIMENTO, 2008).

A pesquisa científica que obtém informações capazes de se tornarem conhecimento e tecnologia vive uma era dúbia em que ao mesmo tempo que há um movimento por dados abertos e compartilhados, há também a proteção visando ganhos para os países/empresas de competitividade e isso se materializa em segredos (fórmulas e/ou procedimentos produtivos não divulgados) e em patentes e direitos de propriedade intelectual (BALUÉ; NASCIMENTO, 2007). Logo, uma cultura de proteção de conhecimento tem sido difundida nos últimos vinte anos no Brasil a fim de resguardar as áreas de maior competitividade e intensivas em tecnologia que o país possui.

No enfoque da Inteligência Econômica, atividades de monitoramento do ambiente externo devem ocorrer paralelamente às atividades de proteção do conhecimento, a fim de manter vantagens competitivas adquiridas e de evitar que concorrentes obtenham informações relevantes. O acesso não-autorizado a know-how, a processos de inovação, pesquisa e desenvolvimento, bem como a planos e estratégias pode comprometer a consecução de objetivos nacionais e resultar em prejuízos expressivos no campo socioeconômico. (BALUÉ; NASCIMENTO, 2007, p. 4)

O Brasil está inserido no contexto como potência regional e global e, embora não seja o líder na produção de conteúdo próprio da sociedade do conhecimento, é partícipe e destaque em alguns setores estratégicos e, por isso, alvo dos demais países que dominam a geopolítica internacional e propagam as TIC's.

Tecnologias competitivas são desenvolvidas no Brasil por empresas genuinamente nacionais, tais como: Petróleo Brasileiro S.A. (Petrobras), a Empresa Brasileira de Aeronáutica (Embraer), Instituto Tecnológico de Aeronáutica (ITA), Instituto Nacional de Pesquisas Espaciais (INPE), Empresa Brasileira de Pesquisa Agropecuária (Embrapa), Indústrias Nucleares do Brasil (INB), entre outras tantas. [...] Esse grupo de empresas, que mantém sob sua guarda conhecimentos considerados sensíveis, integra o rol de setores que necessitam de atenção especial do governo federal. Para obter vantagens estratégicas para o país e garantir sua soberania, o Estado precisa estabelecer uma espécie de barreira de proteção em assuntos relativos à segurança nacional, contra possíveis invasões. (CASTRO, 2009, p. 35-36)

A dissertação de Santos (2016) referencia material distribuído pela SEMG/ABIN no “Encontro com Gestores das Universidades Públicas em Minas Gerais” em 2014 destacando o Brasil como detentor da maior biodiversidade do mundo que, somada aos

conhecimentos tradicionais aliadas ao patrimônio genético, possibilita um potencial elevado de desenvolvimento de produtos comerciais, farmacêuticos, químicos, alimentícios etc. A argumentação reconhece o conhecimento como o ativo mais importante das instituições, e como vantagem competitiva, deve ser protegido.

Neste contexto, a Lei n. 9.883/1999, que instituiu o SISBIN e criou a ABIN, estabeleceu, entre outras atribuições, que compete à Agência: “planejar e executar a proteção de conhecimentos sensíveis, relativos aos interesses e à segurança do Estado e da sociedade”.

Para cumprir essa atribuição a ABIN, por meio do Departamento de Contrainteligência (DCI) e das Superintendências Estaduais, desenvolve e executa ações preventivas e pró-ativas para obstrução de ações adversas, como o Programa Nacional de Proteção do Conhecimento Sensível (PNPC), destacado instrumento da atividade de inteligência do governo federal. Antes de explanar sobre o PNPC propriamente dito, é preciso entender o significado de “conhecimento sensível” e também o de “proteção de conhecimentos sensíveis”. A portaria ministerial nº 42/2009, que instituiu de vez o programa em agosto de 2009, o “conhecimento sensível” foi apresentado, em seu artigo 2º, como sendo “[...] todo conhecimento, sigiloso ou estratégico, cujo acesso não autorizado pode comprometer a consecução dos objetivos nacionais e resultar em prejuízos ao País, necessitando de medidas especiais de proteção [...]” (CASTRO, 2009, p. 36).

Santos faz relação entre os aspectos presentes na Contrainteligência (focada no PNPC) e a CI, “a partir da Gestão das informações e dos conhecimentos científicos e tecnológicos sensíveis que, aliada à Inteligência Econômica, subsidiará o processo decisório de instituições nacionais, públicas ou privadas” (SANTOS, 2016, p. 55) nas principais áreas que o programa está inserido²³:

- Defesa Nacional;
- Pesquisa, desenvolvimento e inovação científica e tecnológica;
- Energia, incluídas as fontes alternativas;
- Minerais e materiais estratégicos;
- Conhecimentos dos povos indígenas e das comunidades tradicionais;
- Agropecuária;
- Desenvolvimento socioeconômico;

²³ Disponível em: <http://www.abin.gov.br/atuacao/programas/pnpc>. Acesso em 05/01/2020, às 20h05.

- Educação e promoção de cultura de proteção do conhecimento sensível.

As dimensões da Gestão do Conhecimento trabalhadas na CI estão intrinsecamente presentes nas características da Inteligência, sobretudo, a de Estado. O trabalho de obtenção de informações seja sob coleta ou busca, o registro, a análise (produção do conhecimento), o armazenamento e a difusão vistos no ciclo de Inteligência são reflexos e ações de Gestão do Conhecimento. A Gestão envolve também avaliação e aprimoramento de sistemas e processos na melhoria contínua em busca de resultados e esses são objetivos de ambos, Gestão e Inteligência.

As políticas de informação e de inteligência são essenciais para a formulação de diretrizes que promovam essa cultura de proteção, mantendo o estímulo à inovação como interface com a Inteligência de Estado, pelas iniciativas relacionadas à melhoria dos processos, ao aprimoramento dos sistemas, ao fortalecimento das culturas e à criação de conhecimentos. (SANTOS, 2016, p. 51)

Contudo, é fundamental ressaltar também aspectos que diferenciam a atividade de Inteligência das demais. Como dito no início desta seção, trata-se de atividade especializada com peculiaridades exigindo técnicas e ações sedimentadas em valores e doutrina compatíveis com atividades de Estado e de interesse da sociedade.

A atividade de Inteligência exige o emprego de meios sigilosos, como forma de preservar sua ação, seus métodos e processos, seus profissionais e suas fontes. Desenvolve ações de caráter sigiloso destinadas à obtenção de dados indispensáveis ao processo decisório, indisponíveis para coleta ordinária em razão do acesso negado por seus detentores. Nesses casos, a atividade de Inteligência executa operações de Inteligência – realizadas sob estrito amparo legal –, que buscam, por meio do emprego de técnicas especializadas, a obtenção do dado negado. (Decreto Nº 8.793, Item 2.4, 2016)

A ABIN, segundo Santos, enquanto organização, “possui papel informacional fundamental para integração estratégica do processo que representa a síntese do trinômio Informação, Conhecimento e Inteligência como um produto de apoio ao processo de tomada de decisão.” (SANTOS, 2016, p. 39). A autora relaciona como componentes integrados à GIC, produto, processo, sistema e organização de Inteligência e, consequentemente, à Ciência da Informação. Nesse sentido, identifica sete abordagens como interfaces entre a IE e CI.

QUADRO 2 - Interfaces entre Inteligência de Estado e GIC

Abordagem	Componentes
Estratégica	informação, conhecimento e TI como vantagem competitiva; inteligência empresarial e/ou organizacional e alinhamento estratégico.
Humana	competências, cultura, compartilhamento e aprendizagem organizacional.
Econômico-financeiro	propriedade intelectual, capital intelectual e ativos intangíveis.
Metodológica	Sistêmico, que utiliza técnicas de diversas áreas do conhecimento para a produção do Conhecimento de Inteligência.
Estrutural	organizacional, compreendendo produto, processo, sistema e organização relativos à Inteligência de Estado como componentes integrados à Gestão da Informação e do Conhecimento.
Mediadora	mediador de produto, processo, sistema e organização, objetivando o compartilhamento do fluxo informacional, centralizado pela Agência Brasileira de Inteligência (ABIN), disseminado através do Sistema Brasileiro de Inteligência (SISBIN) e articulado com os órgãos componentes do sistema, pelo Departamento de Integração do Sistema Brasileiro de Inteligência (DISBIN).
Integrativa	integrador de produto, processo, sistema e organização, objetivando a produção do Conhecimento de Inteligência como suporte ao processo decisório Estratégico, inerente à atividade de Inteligência de Estado.

Fonte: Elaborado pelo autor com base na dissertação de Santos (2016, p. 40).

Angelis (2016) avalia o impacto da Cultura Nacional e da Gestão do Conhecimento sobre a Inteligência Governamental (IG). Ele conclui que o impacto da cultura organizacional/nacional no Brasil é muito maior que na Alemanha, que a IG no Brasil é mais influenciada pela cultura que pelo conhecimento. “Isso está relacionado ao fato de que a cultura alemã, em oposição ao Brasil, é orientada para o futuro e para o desempenho, obtendo informações de fatos, livros e estatísticas, em vez de ser orientada para as pessoas, obtendo informações em primeira mão de fontes humanas/orais.” (ANGELIS, 2016, p. 240, tradução nossa²⁴).

O autor frisa sobre a incapacidade de alguns Estados (Brasil, Turquia e os países que passaram pelo fenômeno da Primavera Árabe) em captar informações da sociedade e transformá-las em Inteligência de forma a antecipar ações coletivas que podem abalar o poder gerencial e estratégico do Estado. O setor público ainda não foi capaz de absorver e

²⁴ *This is related to the fact that German culture, in opposition to Brazil, is future and performance-orientated, getting information from facts, books and statistics, instead of being people-oriented, getting the first-hand (oral) information.*

implementar práticas modernas de gestão, como gestão do conhecimento (GC) e inteligência organizacional (OI) e sua complementaridade com eficácia satisfatória. Isso leva um tempo de amadurecimento e nem todas áreas têm o mesmo interesse de aplicar tais práticas. O setor privado tem trabalhado melhor a GC e IO através de consultorias ligadas a *competitive e business intelligence* (ANGELIS, 2016).

A influência da cultura nacional nos comportamentos é fundamental para entender falhas nas decisões, principalmente devido à dificuldade das pessoas em aprender por comparação e colaboração através de uma visão e uma ação global, participativa e integrativa. (ANGELIS, 2016, p. 245).

De acordo com o autor, os latinos são culturalmente mais direcionados por palavras e emoções, enquanto os germânicos por fatos e dados. Na visão de Lewis (2006), os primeiros teriam uma abordagem múltipla (extrovertidos, impacientes, falantes, gregários etc) e os últimos seriam lineares (introvertidos, pacientes, quietos, solitários etc). A abordagem latina dificulta a valorização e aplicação da Inteligência do ponto de vista estratégico, baseado em análises racionais, apenas em dados e informações. A GC assim como a IE busca dar um enfoque mais científico e analítico ao tratamento das informações para que as decisões sejam mais objetivamente/cientificamente substantiadas.

Porém, como De Angelis ressalta: nenhum extremo é o ideal e mesmo os germânicos carecem de flexibilidade que os proporcionaria criatividade, inovação, certa perspicácia necessária a contextos imprevisíveis como os cenários econômico e geopolítico transnacionalmente interdependente. Um cenário ótimo seria aquele que compusesse as diversas inteligências trazidas pelo autor no **Quadro 3** e tal equilíbrio proporcionaria uma IE bem “afiada” ou ajustada para os objetivos e metas do Estado.

Marco Cepik, uma das referências nacionais em estudos de inteligência governamental, em uma manobra didática restringe o conceito de Inteligência para a aplicação na esfera Estatal daquilo tratado como segredo ou informação secreta - conjuntos mais delimitados de fluxos informacionais. O autor quando aborda inteligência não se limita espionagem e trata duas dimensões da atividade de inteligência para explicá-la: uma que destaca meios especiais utilizados para coletar informações sem a cooperação e/ou conhecimento do adversário; a outra, analítica, destaca a capacidade explicativa e/ou preditiva do conhecimento produzido (CEPIK, 2003). Pela definição desse autor em relação ao quadro

de Angelis (2016), pode-se situá-la em *Governmental Intelligence*, que neste trabalho chamamos de IE.

QUADRO 3 - Tipologia de Inteligência

Spiritual intelligence	Cultural intelligence	Emotional intelligence	Rational intelligence	Governmental intelligence
Connection Inspiration	Adaptation Learning by comparison	Patience Motivation	Quick thinking Connection of ideas	Analysis of knowledge Prediction and prevention
Purpose	Examination of the environment	Empathy	Capacity of synthesis	Combining strategy, foresight and action
Meaning	Communication	Managing emotions	Complex models	
Sensitivity Discernment Great worldview				

FONTE: Angelis, 2016, pg. 255.

Isso reforça a importância da IE como política pública necessária para conferir competitividade ao Estado, atuando com princípios de GC também em suas principais empresas multinacionais, privadas ou públicas. Daí, programas capitaneados pela ABIN como o PNPC, já que “em nível macro (governamental), o estudo de IO é conhecido como IG²⁵” (ANGELIS, 2016, p. 253, tradução nossa).

Esse autor desenvolve um modelo de Inteligência do Conhecimento Cultural (ICC, tradução nossa) e demonstra que a cultura nacional e organizacional (CNO, tradução nossa) impacta a GC e a IG. Abaixo podemos visualizar o modelo na figura criada pelo autor.

²⁵ “In the macro level (government), the study of the OI is known as GI.”

FIGURA 10 - Modelo ICC



FONTE: De Angelis, 2016, p. 258.

Assim, Angelis (2016) defende que se pode abstrair do estudo a alta correlação entre a CON, a GC e a IG, além da importância da integração dos diversos tipos de inteligência para potencializar a IG.

Na concepção do modelo ICC, a GC proporciona metodologias para identificar, armazenar, compartilhar e criar conhecimento, enquanto a IO o integra e interpreta (ERICKSON; ROTHBERG, 2003). No entanto, utilizar o conhecimento produzido com efetividade não é trivial ou automático mesmo quando a GIC ou a IE (no caso do Estado) é bem executada.

Assim, entende-se que o investimento em IE é fundamental para alcançar metas e objetivos do Estado, nacional e, sobretudo, internacionalmente. Traçar planos estratégicos, táticos e operacionais para alcançar vantagens competitivas são fundamentais para o crescimento e desenvolvimento econômico de uma nação. Portanto, portar-se político-economicamente com inteligência nos organismos internacionais, nos blocos econômicos ou nas relações bilaterais devem transcender qualquer ideologia e focar em um programa/plano de desenvolvimento de Estado. Nesse sentido é imprescindível que exista algum mais perene que mandatos presidenciais e é por isso que discutiremos um pouco do papel da IE sobre as relações internacionais.

3.4 O papel da Inteligência de Estado nas Relações Internacionais

O autor desta pesquisa defendeu em sua monografia²⁶ de conclusão de graduação em Relações Econômicas Internacionais que um país (no caso o Brasil) para alcançar seus planos e metas de Política Externa precisa, entre outros fatores, de uma IE eficiente, efetiva e ativa. Isso pode não ser uma condição suficiente, mas é, sem dúvida, necessária. O Brasil sendo um líder regional e se posicionando com “*player*” global precisa ter uma IE forte e atuante, assim como as principais potências internacionais (EUA, Alemanha, Rússia, China etc) possuem.

“As táticas adotadas pelos países estudados por Sandoval (2006) apresentam pontos em comum, tais como: a utilização estratégica da informação; o desenvolvimento de um sentimento coletivo de patriotismo econômico; a criação de estruturas governamentais voltadas para a gestão e preservação de conhecimentos nacionais; e, a sinergia com atores econômicos públicos e privados.” (BALUÉ; NASCIMENTO, 2007, p. 5).

A atuação da atividade de inteligência está sustentada, principalmente, pelo paradigma realista das Relações Internacionais como foi mencionado na introdução deste trabalho. “[...] o Estado é o ator central, as decisões são objetivas, pragmáticas para atingir os interesses nacionais, utilizando do poder e da força bélica no sistema internacional, considerado anárquico.” (LAGES, 2017, p. 35).

Os conceitos que fundamentaram o Realismo/Neorrealismo podem ser encontrados em obras como a de Tucídides²⁷, na qual relata o conflito entre Atenas e Esparta em 431 A.C motivado pelo crescimento e projeção de poder de Atenas; a de Thomas Hobbes²⁸, cuja a descrição do estado de natureza do homem – um ambiente de guerra de todos contra todos, em que o homem é mal por natureza e utiliza todos os meios ao seu alcance para atingir seus objetivos; e a de Clausewitz²⁹ em que a Guerra é uma relação de poder, para o autor:

A guerra é, portanto, um ato de força para obrigar o nosso inimigo a fazer a nossa vontade. [...] O propósito político - a razão inicial para a guerra -

²⁶ A Influência da Inteligência Governamental sobre a Política Externa Brasileira (2017).

²⁷ *The History of the Peloponnesian War*. Atenas, 431 a.C.

²⁸ *Leviatã*, 1651.

²⁹ [Carl Von Clausewitz](#), *Da Guerra*, 1837

determinará assim, tanto o propósito militar a ser atingido como a intensidade do esforço que ele exige. [...] Vemos, portanto, que a guerra não é meramente um ato de política, mas um verdadeiro instrumento político, uma continuação das relações políticas realizada com outros meios. (CLAUSEWITZ, 2005, p. 75, 82, 91).

Outros autores complementam e corroboram essa visão são Hans Morgenthau³⁰(realista) e Kenneth Waltz³¹ (neorrealista). Para Morgenthau “[...] os Estados disputam constantemente o poder enquanto mecanismo garantidor de sua autopreservação. [...] A política internacional é apresentada como a arena inevitável de luta entre os diversos atores estatais com vistas à realização de seus interesses nacionais.” (BRITO, 2011, p. 87). Enquanto para Waltz, a guerra é quase inevitável e acontece por não existir um árbitro internacional, um ente mediador de conflitos com poder coercitivo, entre as nações (Waltz, 1959); o sistema internacional pressiona os Estados constantemente a estarem atentos e preparados para a guerra, mesmo sendo contra seu código de comportamento. “Dessa maneira, um Estado pode ser forçado a violar seus valores, seja fazendo uma aliança com outra nação, indo à guerra “preventivamente” ou assinando tratados, sob a lógica de sua preservação.” (BRITO, 2011, p. 88).

Michael Howard, em sua aula inaugural na “King’s College London” em 5 de maio de 1964, argumentou que a ordem internacional não se baseia no reconhecimento de diferenças e limitação das próprias capacidades pelos atores estatais, mas sim, no entendimento de que a capacidade de impor e estender sua própria ordem é limitada pela vontade, eficácia e capacidade de outros estados de impor a deles. “A condução das relações internacionais deve ser, portanto, um delicado ajuste de poder, uma exploração mútua de intenções e capacidades, a fim de encontrar e preservar uma ordem que, embora individualmente não totalmente satisfatória, é tolerável para todos³².” (HOWARD, 2009, p. 153, tradução nossa). Howard ressalta que o poder que os estados exercem na esfera internacional é composto de muitos atributos: econômicos, diplomáticos, culturais e ideológicos e militares. Porém tal qual a visão realista reforça que o poder militar, a

³⁰ (MORGENTHAU). Política entre as nações – a luta pelo poder e pela paz, 2003, 1093p.

³¹ (WALTZ). *Man, the State and War – a theoretical analysis*, 2001, 263p.

³² *The conduct of international relations must therefore always be a delicate adjustment of power to power, a mutual exploration of intentions and capabilities, so as to find and preserve an order which, though fully satisfying to nobody, is just tolerable to all.*

capacidade de usar a violência para a proteção, execução ou extensão de autoridade, continua a ser um instrumento com o qual nenhum Estado pode dispensar completamente.

Considerando este cenário, é importante salientar que a Inteligência tem como missão assessorar o poder decisório com os conhecimentos necessários e possíveis para proporcionar uma melhor tomada de decisão e quando se trata do nível estatal, a IE que detém essa responsabilidade. Isso quer dizer que em sendo a IE valorizada e respeitada, um Estado sempre atua nas Relações Internacionais, subsidiado por seu serviço de Inteligência, considerando em alguma extensão os conhecimentos que a(s) agência(s) proporcionou(nam) a ele. “Os interesses nacionais estão inseridos nos planos estratégicos de inserção internacional, e para alcançá-los é interessante desenvolver as potencialidades para atuar nas esferas multilaterais e bilaterais.” (LAGES, 2017, p. 31).

Não compreender a relevância do Estado enquanto ator privilegiado nas relações internacionais pode significar a aceitação de um papel secundário nessas relações, em que os projetos nacionais deixariam de existir de maneira autônoma. Sob o viés dos serviços de inteligência estatais, no momento em que a complexidade envolvendo a atuação dos diversos atores globais vai aumentando, a necessidade do Estado em identificar e neutralizar novas e velhas ameaças à sua segurança pode ser primordial. (BRITO, 2011, p. 20)

Brito (2011) utiliza de Norberto Bobbio *et al* para sustentar a necessidade/utilidade da IE nas Relações Internacionais entre Estados Nacionais. Os autores ressaltam o caráter de previsibilidade nas relações jurídicas entre os homens na esfera interna do Estado (eventualmente ignoradas), enquanto nas relações Estado-Estado, ou seja, internacionais, a racionalidade perseguida pela previsibilidade é quase impraticável, considerando normas jurídicas internacionais impositivas (BOBBIO; MATTEUCCI; PASQUINO, 2006).

Nesse sentido, o máximo de informações fidedignas e integras em relação as diversas circunstâncias político-econômicas que se consiga numa relação interestatal é fundamental para tomada de decisões. Embora, essas sejam as informações de mais difícil acesso, mais restritas e sigilosas, o que reitera a pertinência da atividade de inteligência.

Essa “natureza estruturalmente aleatória” faz com que o processo decisório na esfera das relações políticas com outros países seja acentuadamente mais desafiador sobre o prisma informacional, do que na esfera nacional. A partir dessa inerente desordem nas relações internacionais, tem-se um desafio para a escolha das melhores políticas envolvendo a obtenção de informações para suportar o processo decisório, gerando, portanto, um paradoxo. Ao mesmo tempo em que se busca a melhor decisão pautada pela racionalidade, em detrimento do Estado de natureza, não se tem todas as informações necessárias a esse processo de escolha ideal. Então, embora os realistas

apreçoem o processo de escolha das ações estatais a partir de uma citada racionalidade, a mesma só pode ser alcançada com a coleta de informações sobre a realidade internacional circundante, bem como sobre as pretensões dos adversários, o que é uma tarefa de grande envergadura e de difícil execução. (BRITO, 2011, p. 90).

Na esfera internacional os Estados competem por recursos minerais, vegetais, capital humano, mercados consumidores, ciência e tecnologia entre outros ativos. Fato é que em todos esses, a base é a informação agregada inerente a cada um – o que se fará com cada recurso depende do que ele é, do que é capaz, mas, sobretudo, de quem o possui. Neste momento, a IE atua em cima da coleta, busca, análise do dado negado e isso muitas vezes requererá obter informações sem o consentimento de outro Estado. Nesse ambiente, a prática de inteligência é essencial não só à segurança, mas à competitividade das estruturas econômicas e ao desenvolvimento das diferentes nações (PINHEIRO, 2012)

A França, em 2004, estabeleceu formalmente o tema da Inteligência Econômica como prioridade nacional. Alain Juillet (2004), nomeado pelo presidente da república como responsável pelo assunto naquele país, definiu, em entrevista à LExpansion.com, as ações desenvolvidas pelo governo francês para evitar a “pilhagem” de tecnologias francesas, enfatizando a necessidade de recuperar o atraso em que país se encontrava nesse campo.” (Balué e Nascimento, 2007, pg. 5). “A Lei de Espionagem Econômica (ESTADOS UNIDOS, 1996) permitiu aos EUA tratar a obtenção ilegal de segredos comerciais como crime federal, estabelecendo penas que podem alcançar até 15 anos de detenção e multa de 10 milhões de dólares. O Código Penal português (PORTUGAL, 1995) também estabeleceu o delito de espionagem, responsabilizando nacionais que, ao colaborarem com governo, associação, organização ou serviços de informações estrangeiros, tornem acessível fato, documento, plano ou objeto que, em defesa dos interesses do Estado Português, devam ser mantidos em segredo. (BALUÉ; NASCIMENTO, 2007, p. 6)

A espionagem econômica é uma das principais preocupações dos Estados no mundo competitivo. “O Brasil abriga a maior diversidade biológica do planeta que, agregada aos conhecimentos tradicionais associados ao patrimônio genético, resulta em elevado potencial para o desenvolvimento de produtos comerciais” (SANTOS, 2016, p. 52), além de

possuir multinacionais estratégicas nas áreas de alimentação (AMBEV), cosméticos (NATURA) e aviação (EMBRAER). O Brasil foi em 2019 a nona economia mundial, além de ser a primeira da América Latina, isso demonstra o potencial que o país tem de crescimento e, por suposto, o alvo que se torna em um ambiente hostil e competitivo.

Para exemplificar, citamos o ano de 2013, quando houve vazamentos do ex-técnico da CIA, *Edward Snowden*. *Snowden* denunciou a atuação da espionagem norte-americana por meio do vazamento de documentos classificados da *National Security Agency*, pra quem trabalhou. Os documentos davam conta de que muitos países estavam sendo monitorados por escutas telefônicas, inclusive, seus chefes de Estado, por parte do governo americano. A presidente Dilma Rousseff teve seus telefonemas monitorados pelos EUA, evidenciando uma grave falha da contrainteligência brasileira. Se consideramos os dados disponíveis no site www.wikileaks.org/plusd/ - o qual oferece mais de dois milhões de documentos dos EUA sobre geopolítica global - veremos em uma busca simples por tema (*Traffic Analysis by Geography and Subject—TAGS*) 26.273 documentos relacionados ao Brasil, desde janeiro de 1973. (LAGES, 2017, p. 56)

De acordo com o Guia de Inteligência do Consumidor elaborado pela equipe de assuntos públicos da CIA, Inteligência Econômica e Espionagem Econômica envolvem, respectivamente: recursos, atividades e políticas econômicas internacionais, incluindo produção, distribuição e consumo de bens e serviços, mão de obra, finanças, tributação, comércio, negócios e outros aspectos do sistema econômico internacional; e tentativas clandestinas ou ilícitas de adquirir inteligência econômica para ajudar em seus interesses econômicos internacionais.³³ (tradução nossa).

Outra consideração sobre inteligência econômica é trazida em *Economic Intelligence – a Guide for Beginners Pratitioners*.

A Inteligência Econômica (IEc) visa desenvolver melhores métodos para a identificação de fontes relevantes de informação, a análise das informações coletadas e sua manipulação para fornecer o que o usuário precisa para a tomada de decisões. [...] IE refere-se ao conjunto de conceitos, métodos e ferramentas por trás das ações do ciclo de inteligência, apoiando a tomada de decisões dentro da estrutura de uma estratégia organizacional estabelecida. Mercado, tecnologia, questões legais, macroeconômicas e outras que afetem

³³ [...] intelligence regarding foreign economic resources, activities, and policies including the production, distribution, and consumption of goods and services, labour, finance, taxation, commerce, trade, and other aspects of the international economic system and the Economic espionage is clandestine or illicit attempts by foreign interests to assist their economic interests by acquiring economic intelligence. *Consumer's Guide to Intelligence*, Office of Public Affairs, CIA, 1995.

a operação de uma organização são cobertas pela IEc.³⁴ (CETISME, 2002, p, 9, 106, tradução nossa)

Johnson (1997) frisa o envolvimento da CIA e suas diversas agências de IE executando as diversas etapas da atividade de inteligência: coleta, análise, contrainteligência e operações encobertas. O entendimento norte-americano é que a segurança nacional está diretamente ligada à segurança econômica, logo, justifica-se o emprego da IEc. Tanto que em 1993, Bill Clinton funda o Conselho Econômico Nacional como o mesmo status do Conselho de Segurança Nacional.

Diversos interesses/temas foram monitorados no pós-Guerra Fria pela CIA e agências ligadas à IEc, dentre eles: se governos estrangeiros estão negociando honestamente com os EUA; estudos elaborados para o governo norte-americano utilizar nas conferências de comércio internacional indicando como os outros países poderiam agir, com quem se aliar etc; acompanhamento do andamento/resultado de sanções econômicas a outros países; acompanhamento de investimento estrangeiro direto nos EUA; manutenção de um banco de dados de atores econômicos internacionais (governos, instituições ou indivíduos) que violem sanções, lavem dinheiro, vendam armas e, especificamente, ligados ao narcotráfico (JOHNSON, 1997).

Porém, discussões são levantadas sobre o limite de participação das agências de IEc: fariam espionagem/sabotagem econômica e interfeririam no livre comércio/concorrência internacional? A posição institucional é que não realizariam tais ações, por serem antiéticas, contra seus valores e provocarem instabilidade nas relações entre os parceiros dos EUA.

Nossa segurança nacional é inseparável da nossa segurança econômica”, Secretário de Estado Warren Christopher presidente do Comitê de Relações Exteriores do Senado em 4 de novembro de 1993. É por isso que as agências secretas continuarão a buscar inteligência econômica, enquanto ao mesmo tempo – esperamos – rejeitem operações injustificadas, solicitadas por alguns que evidentemente coloquem a margem de lucro de curto prazo acima de todas as outras considerações. (JOHNSON, 1997, p. 512)

A execução de IEc é uma das formas para exercer um poder mais suave sobre outros países. Na visão de Brito (2011), o *hegemon* das Relações Internacionais são os EUA e os serviços de inteligência são um dos principais instrumentos para manutenção dessa

³⁴ “Economic intelligence (EI) aims to develop better methods for the identification of relevant sources of information, the analysis of the collected information and its manipulation to provide what the user needs for decision-making. [...] set of concepts, methods and tools behind the intelligence cycle actions, supporting decision-making within the framework of an established organisational strategy. Market, technology, legal, macroeconomic and other issues affecting the organisation operation are covered by EI.

posição. “Os serviços de inteligência são um dos instrumentos fundamentais a serem empregados por parte do Estado informacional estadunidense. [...] dando suporte e promovendo os interesses do Estado patrocinador, sobretudo a partir da faceta informacional, que permeia todas as esferas.” (BRITO, 2011, p. 214).

Os EUA se inserem num grupo chamado de Estados Centrais, que são aqueles com maior protagonismo internacional e a construção desse diferencial comunga com a evolução e profissionalização da atividade de inteligência no século XX. Além dos altos padrões de desenvolvimento, [...] “a noção de centro, contudo, não se limita ao desenvolvimento, mas também denota vantagem em termos militares, culturais, políticos, ambientais, etc” (RUVALCABA, 2019, p. 139). Para o autor citado, os países considerados centrais são industrializados de forma equilibrada pelos diversos setores, estão continuamente em processo criativo, buscando economia de energia, metodologia e materiais mais produtivos e superando em ciência e tecnologia outras sociedades.

Ruvalcaba (2019) caracteriza os Estados centrais pelas dicotomias econômico-militar, socioinstitucional e comunicativo-cultural. A primeira é composta pelos elementos econômicos de produção, comércio e finanças e militares de aparato militar, território e investimento em ciência e tecnologia. “A história das relações internacionais mostrou que a produção econômica e o poder militar são necessários e se retroalimentam. [...] Embora elas não necessariamente ocorram em paralelo, a centralização da riqueza contribui para o aumento do poder militar.” (RUVALCABA, 2019, p. 145).

O poder econômico dos EUA tem sustentação também na atividade de inteligência e é característica do poder instrumental (BRITO, 2011), cuja a expressão se dá através de pesquisa e tecnologia na academia e no setor privado, empresas multinacionais, acordos interestatais multi e bilaterais e subsídios à Organizações Internacionais.

[...] a informação dos serviços secretos subsidia as duas principais facetas do exercício do poder nas relações internacionais, os poderes instrumental e simbólico. Como foi analisado no decorrer deste trabalho, o poder instrumental ou coercitivo relaciona-se ao emprego (mesmo que potencial) da força, seja esta econômica ou militar. As informações obtidas pelas agências de inteligência subsidiam o processo de tomada de decisão. Em relação à capacidade de influenciar, moldar a percepção do mundo e o horizonte ideológico, que são características do poder simbólico, os serviços secretos também têm intervindo decisivamente para promover os interesses norte-americanos. (BRITO, 2011, p. 215)

A qualificação da população implica uma melhor produtividade per capita e está relacionada a dicotomia socioinstitucional levantada por Ruvalcaba. Uma quantidade expressiva de habitantes, bem nutridos, com boa formação educacional, um alto nível de emprego refletem em ganho de desenvolvimento econômico, militar e social. Isso denota alta produtividade individual e geral. Todavia isso não ocorre sem que haja um planejamento político-institucional, sem que o Estado entenda a importância desse estado de bem-estar social para seu desenvolvimento e que para isso precisa aplicar políticas públicas de serviços (saúde, educação e pesquisa) aos cidadãos e outras que suportem o setor privado a gerar emprego e renda (RUVALCABA, 2019).

O poder comunicativo-cultural de Ruvalcaba está para o poder simbólico de Brito ou o *Soft Power* de Nye (2004). “À medida que os países centrais avançam no desenvolvimento econômico, os valores e estilos de vida em suas sociedades vão sendo modificados, estabelecendo modas e tendências em termos culturais, intelectuais e ideológicos.” (RUVALCABA, 2019, p. 149).

A diferença é que Brito contextualiza dentro da esfera de ação dos serviços de inteligência, enquanto Ruvalcaba não menciona a atividade de inteligência diretamente. Isso quer dizer que esse tipo de poder não é exercido aleatoriamente, mas de forma pensada e planejada.

A questão, então, é como os Estados centrais exercitam seus efeitos de demonstração, aumentam seu soft power ou poder simbólico e se superam geoculturalmente? Existem alguns aspectos que escapam ao aparato estatal, pois estão enraizados na cultura popular ou no desenvolvimento histórico; mas existem outros meios que dependem diretamente dos governos e da gestão dos recursos públicos.[...] algumas áreas estratégicas, nas quais os gastos do governo intervêm, são: defesa, educação, saúde, energia, pesquisa e desenvolvimento. Certamente, o destino final dos gastos públicos é a sociedade nacional, mas sua incidência vai muito além das fronteiras: Estados com altos gastos do governo têm a possibilidade não apenas de influenciar alguns setores-chave para o desenvolvimento, mas também podem intervir em outros – como turismo, diplomacia, comunicação, criatividade e tecnologia – que contribuem diretamente para transmitir no exterior imagens ou projeções específicas da sociedade nacional. Uma característica dos Estados centrais é que, historicamente, eles tiveram gastos do governo que lhes permitiram exercer efeitos demonstrativos, fortalecer seu soft power e influenciar outro geoculturalmente. (RUVALCABA, 2019, p. 150).

A apresentação das dicotomias dos Estados Centrais foi feita para perceber em quais áreas a IE potencialmente atua no sistema internacional, ou seja, a IE tem participação

no exercício de poder pelas três frentes dicotômicas apresentadas, ela promove, auxilia e sustenta a ação do Estado por meio daquelas. Isso justifica o porquê de escolher documentos desclassificados da principal agência de inteligência norte-americana (Estado Central) para pesquisar documentos vinculados a economia.

Os documentos, coincidentes com o período de Guerra Fria, podem nos dar um panorama de quais eram os temas que perpassaram pela demanda informacional da CIA e talvez até da capacidade preditiva com base na perspectiva de Sherman Kent.

Se as informações preditivas não delimitam precisamente quais tecnologias irão vingar, ao menos permitem antever o contexto em que surgirão e de que maneira podem ser mais adequadas à nação que as encomenda. Se a CIA não antecipou a guerra da Coreia, e o NIC não precisou corretamente os mísseis em Cuba ou mesmo a queda do muro de Berlim, a simples tentativa de antever e projetar realidades décadas à frente já se constitui como um grande diferencial. Em um mundo imediatista, o Estado norte-americano tenta antever o panorama em que disputará a hegemonia com meio século de antecedência. Mais do que tentar prever a partir de inferências abstratas, são empregados imensos recursos das agências de inteligência nessa tentativa. Como já dito, toda a estrutura de produção analítica e de coleta informacional estadunidense tem em seu vértice nos relatórios de estimativas. Derivado desta tentativa de detectar previamente as tendências que moldarão o amanhã, têm-se os EUA como um ator relevante no impulsionamento tecnológico e ideológico desse citado futuro. (BRITO, 2011, p. 217)

3.4.1 O período de Guerra Fria

O objeto de estudo desta pesquisa são documentos desclassificados pela CIA que compreendem o período da Guerra Fria e eles são produtos de inteligência gerados por analistas da agência norte-americana. Desta forma, retratam do ponto de vista de IE (com foco econômico) quais temas, localidades e atores os EUA monitoravam, ou seja, quais *players* globais e os aspectos mais relevantes do cenário econômico perpassavam seus interesses.

Historiograficamente, há divergências sobre o que foi a Guerra Fria, período entre o fim da Segunda Guerra Mundial e 1989, caracterizado pela queda do Muro de Berlim, e duas posições antagônicas são postas: “[...] 1ª foi uma construção soviética, que queria expandir o comunismo para o resto do mundo; 2ª foi uma construção norte-americana, para justificar suas ações e consequentes intervenções nas nações que estivessem fora da “esfera” do domínio soviético [...]” (BIAGI, 2007, p. 63).

A CIA atuou de diversas formas por todo o mundo, agentes trabalharam de forma encoberta como jornalistas, missionários, estudantes e empresários pelo mundo³⁵. Em 1974 estima-se que os EUA contavam com duzentos agentes “camuflados” com homens de negócio espalhados por outros países. Operações encobertas ocorrem no período da administração de Nixon no Chile, quando governado pelo socialista Salvador Allende, mais tarde sob Reagan na Nicarágua Sandinista (JOHNSON, 1997). Todas essas ações transcorreram para desestabilizar regimes que potencialmente ou diretamente ofereceram riscos aos interesses estadunidenses e de suas empresas multinacionais.

Nos documentos analisados constatou-se termos bastante frequentes como “USSR”, “soviet” e sua variante plural *soviets*, condizendo com o período de produção dos mesmos. Além disso, outras técnicas executadas corroboraram o interesse norte-americano na União Soviética/Russos, tais itens estavam presentes nas redes de nós, *clusters* e tópicos demonstrando serem o plano de fundo em que os outros temas se relacionavam.

Diversos cenários internacionais tiveram influência da disputa entre as superpotências que capitaneavam a Guerra Fria (EUA e URSS). A disputa por tecnologia esteve entre os principais temas, tanto na corrida espacial como na aviação. Kolnogrov (2005) e Engel (2007) ilustram como a aviação militar e civil, respectivamente tiveram pontos importantes no período que mereciam acompanhamento de IE.

O Oriente Médio e mundo Árabe foram palco também da influência das duas potências (VALBJØRN; BANK, 2012), a despeito de toda um conflito étnico cultural-religioso preexistente (KARSH, 1997). África e Ásia não passaram despercebidas pelos analistas da CIA, desde a África Subsaariana (SCHRAEDER, 2000) passando pelo Chifre da África (SCHWAB, 1978) até o estreitamento de relação entre os EUA e Paquistão, influenciando nas tensões com a Índia à época (MCMAHON, 1988).

A Europa foi uma das regiões com forte atuação no jogo de poder da Guerra Fria, sobretudo com o comércio de *commodities* (petróleo, gás, grãos), tendo os EUA mais poder de ação na região (PAINTER, 2009), pelo menos até os choques do Petróleo na década de 1970 (PAINTER, 2014). Quanto aos grãos, os norte-americanos também foram mais felizes em

³⁵ Eugene Fodor, the president of Fodor's Travel Guides, Inc., once proclaimed that “...I was glad, and proud to be of help” [in providing cover for CIA activities overseas] (Van Gelder 1975, Al). (JOHNSON, 1997, p. 506)

atuar em programas de recuperação da agricultura europeia, principalmente, em um cenário de recuperação econômica dos países europeus (PERKINS, 1997).

O Brasil figurou entre os alvos de busca e coleta da CIA no período da Guerra Fria, temas como finanças, reformas, petróleo e combustíveis estavam presentes nos documentos pesquisados. Períodos de distanciamento (HILTON, 1981) e o chamado ‘pragmatismo responsável’ (FONSECA JR., 2018) marcaram a relação do Brasil com os EUA. O proÁlcool foi um programa de sucesso de alternativa de combustível desenvolvido no Brasil (SOCCOL et al., 2005) durante os períodos de crises do Petróleo internacional. E ao fim do período de ditadura militar no país, crises financeiras na balança de pagamentos e monetária com a inflação descontrolada o atingiram e marcaram sua relação com o FMI e organizações financeiras internacionais (ARIDA; RESENDE, 1985; MARKOFF; BARETTA, 1990).

A China teve um papel a parte durante a Guerra Fria, iniciando o período com um alinhamento com os soviéticos, passando por uma transição com distanciamento e uma aproximação com os norte-americanos. Entre meados da década de 1950 e meados de 1960 houve a separação sino-soviética marcada por divergências político-ideológicas, o que abriu oportunidade para uma crescente relação sino-americana favorecendo a entrada e o comércio chinês na Europa Ocidental e outras regiões do planeta como África e América Latina. (SCHAUFELBUEHL; WYSS; ZANIER, 2019).

Os autores ressaltam que não houve um alinhamento direto na relação com os EUA, mas uma relação independente e equidistante perante as duas potências, seguindo um planejamento estratégico nacional específico modulado pelas mudanças pós-Revolução Cultural (1966-1976).

As referências feitas nessas seções permitem avançar para metodologia compreendendo por que os objetivos da pesquisa foram estabelecidos e como serão perseguidos. Teorizar e conceituar IE e GC individualmente, estabelecer as relações epistemológicas entre as duas áreas e em seguida apresentar o papel da IE nas RI sedimenta o caminho para apresentar como se planeja obter os resultados, um pouco da teoria envolvida nesta metodologia e em qual sequência foi executada. Desta maneira, o próximo capítulo tratará do percurso metodológico dividido em três seções didaticamente posicionadas para o acompanhamento.

4 METODOLOGIA

Este capítulo está dividido três seções: a primeira contém uma breve abordagem da metodologia desta pesquisa, uma prévia do desenvolvimento do trabalho executado; a segunda contém um tratamento teórico sobre as ferramentas de PLN e técnicas de Aprendizado de Máquina utilizadas, o *core* da pesquisa, o conceitual embutido nos algoritmos aplicados sobre os dados (documentos desclassificados) da pesquisa; a terceira e última relata a execução metodológica propriamente dita, a sequência seguida no tratamento dos dados coletados.

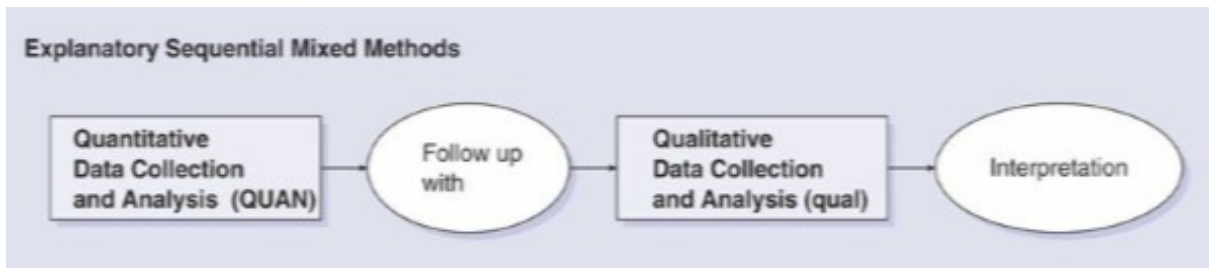
4.1 Visão Geral

Este trabalho tem em sua base, técnicas de coletas de dados, principalmente, a análise documental. É uma técnica que realiza a confrontação dos dados coletados junto ao modelo de análise. Sendo assim, é muito importante saber o que coletar, com quem e como, além de, posteriormente, na posse dessa gama de informações avaliá-las e analisá-las de maneira científica e sistemática (GERHARDT; SILVEIRA, 2009).

A análise documental consiste, basicamente, na coleta de dados a partir da leitura e análise de documentos tais como livros, jornais, registros, fotos, filmes, dentre outros, que são obtidos de maneira indireta pelo pesquisador (BARBOSA, 1998; GIL, 2008). Esse procedimento possibilita ao pesquisador coletar dados em “quantidade e qualidade suficiente para evitar a perda de tempo e o constrangimento que caracterizam muitas das pesquisas em que os dados são obtidos diretamente das pessoas” (GIL, 2008, p. 147). Além disso, a utilização de dados já existentes a partir de documentos auxilia na diminuição de custos e otimização do tempo da pesquisa (BARBOSA, 1998).

Caracteriza-se por uma metodologia mista de pesquisa abordando tanto técnicas quantitativas como qualitativas. Do ponto de vista de Creswell (2014) será uma pesquisa mista exploratória sequencial, já que primeiro se executará uma pesquisa quantitativa e depois os seus resultados serão explicados e debatidos com a etapa qualitativa.

FIGURA 11 - Esquema de metodologia do trabalho



FONTE: Creswell, 2014, p. 270

Foram aplicadas ferramentas de *Data Mining*³⁶ (Mineração de Dados) na execução da pesquisa em documentos que apresentam o teor de IE. Utilizou-se como tema base a Inteligência de Estado aliada à aplicação de técnicas computacionais de análise de documentos, recuperação da informação e indexação automática. A mineração de dados é conceituada em ciência da computação como o processo de descoberta de padrões e relacionamentos interessantes e úteis em grandes volumes de dados. O campo combina ferramentas de estatística e inteligência artificial (como redes neurais e aprendizado de máquina) com gerenciamento de banco de dados para analisar grandes coleções digitais, conhecidas como *data sets*³⁷.

Nesta pesquisa, a mineiração de dados tem uma concepção voltada à GC, que é um processo de descoberta de informações substanciais acerca de um dado domínio (neste caso, IE) em grandes volumes de dados, atribuindo significados aos dados coletados. Ou seja, os documentos desclassificados da CIA sobre assuntos econômicos foram analisados, geraram informação que proporcionou a produção de conhecimento através da interpretação e discussão contextualizada.

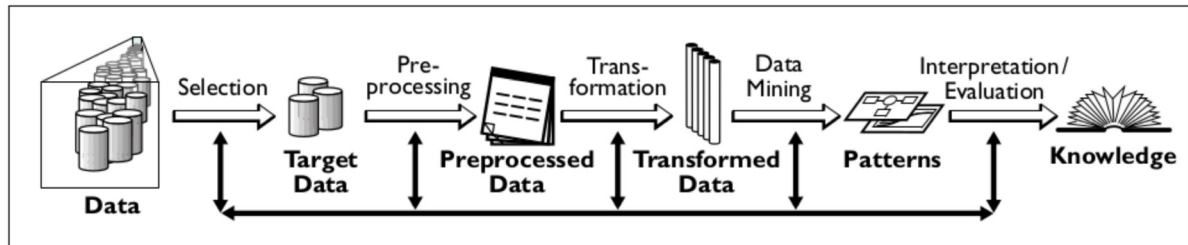
Para Fayyad, a mineração de dados é uma etapa particular de um processo maior - knowledge discovery in databases (KDD): a mineração de dados seria a aplicação de algoritmos específicos para extrair padrões de dados, enquanto o KDD, o processo não trivial de identificar padrões válidos, novos, potencialmente úteis e, finalmente, compreensíveis nos dados (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996). Embora na prática, atualmente

³⁶ *Data mining: includes searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression, clustering, sequence modeling, dependency, and line analysis.* (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996, p. 31)

³⁷ Adaptado de <https://www.britannica.com/technology/data-mining>. Acesso em 25/06/2020 às 04h53.

data mining é o termo mais empregado, e entendido como todo o processo explicitado pelo KDD. A figura abaixo exemplifica a intenção do autor em diferenciar os conceitos.

FIGURA 12 - Data Mining / KDD



FONTE: Fayyad, Piatetsky-shapiro, Smyth, 1996, p. 29.

Serão explorados como fonte de informação documentos desclassificados publicados pela *Central Intelligence Agency* (CIA), a agência central de inteligência norte-americana. Utilizar-se-á como ferramenta de extração de dados destes documentos, algoritmos de programação baseados na linguagem *Python*³⁸ utilizando como ambiente para “escrevê-los” a plataforma *Jupyter Notebook*³⁹, que é um ambiente computacional web, interativo para criação de documentos.

Os documentos criados no ambiente *Jupyter* estão formatados em *JavaScript Object Notation* (JSON⁴⁰), com um esquema e uma lista ordenada de células que podem conter código, texto, fórmulas matemáticas, gráficos, tabelas e imagens. JSON é um formato de texto de troca de dados independente da linguagem de programação, com isso, adapta-se a várias delas (C e familiares, incluindo C++, C#, Java, JavaScript, Perl, Python e muitas outras).

A etapa inicial (pré-processamento) consistiu em realizar a coleta dos documentos nos repositórios presentes no site da CIA, na seção: *Freedom of Information Act Electronic*

³⁸ “*Python* é uma linguagem de programação de alto nível, interpretada, de script, imperativa, orientada a objetos, funcional, de tipagem dinâmica e forte. Foi lançada por Guido van Rossum em 1991.[1] Atualmente possui um modelo de desenvolvimento comunitário, aberto e gerenciado pela organização sem fins lucrativos *Python Software Foundation*. Apesar de várias partes da linguagem possuírem padrões e especificações formais, a linguagem como um todo não é formalmente especificada. O padrão de facto é a implementação Cpython. A linguagem foi projetada com a filosofia de enfatizar a importância do esforço do programador sobre o esforço computacional. Prioriza a legibilidade do código sobre a velocidade ou expressividade. Combina uma sintaxe concisa e clara com os recursos poderosos de sua biblioteca padrão e por módulos e frameworks desenvolvidos por terceiros.” Wikipédia – <https://pt.wikipedia.org/wiki/Python>. Acesso em 08/07/2018, às 16h19.

³⁹ Disponível em: <https://jupyter.org/>. Acesso em 19/05/2020, às 04h19.

⁴⁰ Disponível em: <https://json.org/json-pt.html>. Acesso em 19/05/2020, às 04h23.

Reading Room (FOIA). Uma técnica conhecida como *Crawler*⁴¹ foi utilizada no processo de recuperação do corpus do trabalho. Um programa foi elaborado através da linguagem *Python* que permitiu a captura dos documentos presentes dentro da seção FOIA:

– *Economic Issues*, filtrado pela coleção da *National Intelligence Concil* (NIC)⁴²;

Os documentos recuperados com o *crawler* na categoria *Economic Issues* consistem de assuntos econômicos no período de aproximadamente 45 anos pós Segunda Guerra Mundial. Como tais se estendem até a década de 1990, são contemporâneos do período da chamada Guerra Fria (tensões na disputa pela hegemonia ideológica no cenário internacional pela prevalência de um ou outro modelo de governo/produção político-econômico, consistindo de apoio armamentista e financeiro em diversos conflitos territoriais para angariar apoio/aliados) entre os EUA e URSS.

Em seguida, foi aplicado a tecnologia *Optical Character Recognition* (OCR), pois os documentos coletados na extensão .pdf são imagens escaneadas e precisavam ser convertidos em formato de texto para serem lidos pelo código que executou a modelagem de tópicos.

Esta é a função do OCR, tornar estes documentos passíveis de leitura ao reconhecer opticamente os caracteres presentes. Esta técnica é uma alternativa quando não se têm os documentos no formato apropriado, porém a sua taxa de acerto/recuperação da informação varia conforme as condições visuais dos documentos, sendo dificilmente alcançada a excelência. Em alguns casos os documentos são disponibilizados em formato *machine readable*, passível de leitura por máquinas, como *eXtensible Markup Language* (XML) - uma linguagem de marcação recomendada pela W3C para a criação de documentos com dados organizados hierarquicamente, tais como textos, banco de dados ou desenhos vetoriais.

A linguagem XML é classificada como extensível porque permite definir os elementos de marcação⁴³. Isso simplificaria a execução do trabalho, porém os documentos extraídos da seção FOIA não se apresentaram nesse formato.

⁴¹ “**Web Crawling**: o ato de baixar automaticamente os dados de uma página web, extrair os hiperlinks contidos nela e segui-los. Os dados baixados são geralmente armazenados em um índice ou banco de dados para facilitar sua busca”. <https://bit.ly/33MY3z8>. Acesso em 28/05/2019, às 20:42.

⁴² [https://www.cia.gov/library/readingroom/search/site/?f\[0\]=im_field_taxonomy_nic_function:2](https://www.cia.gov/library/readingroom/search/site/?f[0]=im_field_taxonomy_nic_function:2). Acesso em 08/04/2019, às 13:20.

⁴³ <https://www.tecmundo.com.br/programacao/1762-o-que-e-xml-.htm>. Acesso em 23/08/2018, às 19h55.

A etapa analítica inicial aplicou uma varredura nos documentos através da contagem de termos mais frequentes, *word-embeddings* e clusterização, finalizando com *Topic Modeling* (Modelagem de Tópicos). Tratou-se os documentos através de técnicas diferentes, porém complementares, encontrando termos mais frequentes e seus relacionamentos, além do perfil de agrupamento dos documentos, caracterizando-os gradativamente. Sequencialmente, aplicou-se a modelagem de tópicos para definição/confirmação dos temas mais recorrentes/prevalentes e representativos nos documentos e, desta forma, os resultados estariam validados pelas etapas iniciais. Depreende-se daí que os assuntos, localidades e entidades (tais como setores econômicos, Estados Nacionais, organizações não-governamentais e privadas, dentre outros) objetos de observação da IE dos Estados Unidos no período de cobertura desses documentos, estão de acordo com o conteúdo do corpus representados nesses tópicos. A partir disso contextualizou-se esses termos/temas/tópicos recuperados de acordo com a conjuntura internacional aplicando uma visão de Inteligência de Estado.

Essa foi uma visão metodológica geral sobre a pesquisa e antes que se explore a execução técnica será tratado teoricamente alguns conceitos/aplicações sobre o ferramental de aprendizado de máquina e modelagem de tópicos para fornecer um cabedal conceitual minimamente suficiente para um melhor acompanhamento da etapa analítica.

4.2 Ferramentas de PLN – Técnicas de Aprendizado de Máquina

Na introdução desta pesquisa, assim como na seção anterior, foi mencionado a utilização de PLN com o emprego de técnicas de aprendizado de máquina – o que será detalhado na seção seguinte. Ferramentas de PLN têm sido utilizadas em pesquisas de Inteligência por diversas organizações, incluindo Estados Nacionais. O volume de informações (*big data*) produzidas em um mundo cada vez mais tecnológico enseja um tratamento igualmente avançado para permitir um processamento informacional e uma produção de conhecimentos tempestivos e eficazes.

O termo *Big Data* tem fundamento tanto no volume quanto na distribuição, autores como Borgman destacam que a onipresença dos dados é um fator importante para tal grandeza. A vida moderna é mediada e conduzida com a ajuda de tecnologias da informação, aspectos do comportamento humano são facilmente monitorados e registrados. Hoje, os

indivíduos possuem diversos dispositivos de comunicação, personalizados e identificáveis, passíveis de serem rastreados. São eles, celulares, televisões, ipads, notebooks, computadores de bordo de veículos automotores etc. Esses dispositivos podem sentir, comunicar e computar, capturaram e distribuem textos, imagens, áudio e vídeo. As ações podem ser registradas com coordenadas de tempo e lugar, criando registros contínuos de atividade (BORGMAN, 2015). “Esses rastreamentos podem ser combinados para criar modelos ricos de atividade social. Os dados e os usos para os quais eles podem ser utilizados estão proliferando muito mais rapidamente do que a lei de privacidade ou a política de informações podem alcançar.” (BORGMAN, 2015, p. 22).

A título de exemplo, há um trabalho em que foi feita uma mineração de opiniões ou a chamada análise de sentimentos - técnicas automáticas capazes de extrair informações subjetivas de textos em linguagem natural, como opiniões e sentimentos, a fim de criar conhecimento estruturado que possa ser utilizado por um sistema de apoio ou tomador de decisão. (BENEVENUTO; RIBEIRO; ARAÚJO, 2015, p. 2) - na qual os autores avaliam a emergência do *Big Data*, sua evolução e seu impacto para a construção de um eficiente e efetivo sistema de suporte de decisões e uma poderosa ferramenta estratégica para a inteligência governamental (KUMAR, AKSHI; SHARMA, 2017).

O objetivo específico, neste caso, foi avaliar como as técnicas de *Machine Learning* estão sendo utilizadas para contribuir com Governos/Estados Nacionais, do ponto de vista de avaliação da qualidade e proposição de políticas e serviços governamentais.

Pode-se observar pelo exposto acima que foram aplicadas técnicas de aprendizado de máquina para análise de *big data* com objetivo de produção de conhecimento (Inteligência). Sendo assim, torna-se fundamental saber também o que se entende por “Aprendizado de Máquina”.

Segundo Mitchell (1997), o aprendizado de máquina é o uso de recursos de algoritmos/programas computacionais que aprendam/melhorem com a experiência, ou seja, com a sua execução. Para o autor é fruto de diversos campos do conhecimento: estatística, inteligência artificial, filosofia, teoria da informação, biologia, ciência cognitiva, complexidade computacional e teoria de controle (MITCHELL, 1997). Aprendizado de Máquina refere-se a “... técnicas que empregam um princípio de inferência denominado indução, no qual se obtém conclusões genéricas a partir de um conjunto particular de

exemplos. O aprendizado indutivo pode ser dividido em dois tipos principais: supervisionado e não supervisionado.” (LORENA; CARVALHO, 2007, p. 44).

O supervisionado pressupõe uma definição de partida/referência de análise, exemplos que sirvam de ponto de partida, o modelo é previamente treinado com determinado rótulo. Enquanto o não-supervisionado não possui tal rótulo para o treinamento, o modelo é treinado apenas com os dados brutos, os padrões serão definidos sem prévia referência e quanto maior a quantidade de dados, mais preciso/inteligente fica o modelo (LANE; HOWARD; HAPKE, 2019).

4.2.1 Word-embeddings

Nesta pesquisa trabalhou-se com algumas ferramentas como *word-embeddings* e clusterização antes de aplicar a modelagem de tópicos. Como já mencionado, a intenção foi executar uma varredura nos documentos e conhecê-los melhor para oportunizar uma capacidade melhor de interpretação e análise dos tópicos, ou seja, os resultados dessas etapas proporcionaram interpretações que se acumularam e se somaram ao da modelagem dando mais robustez a análise final. Desta forma, abordou-se cada uma das técnicas na sequência em que foram executadas finalizando com a modelagem de tópicos.

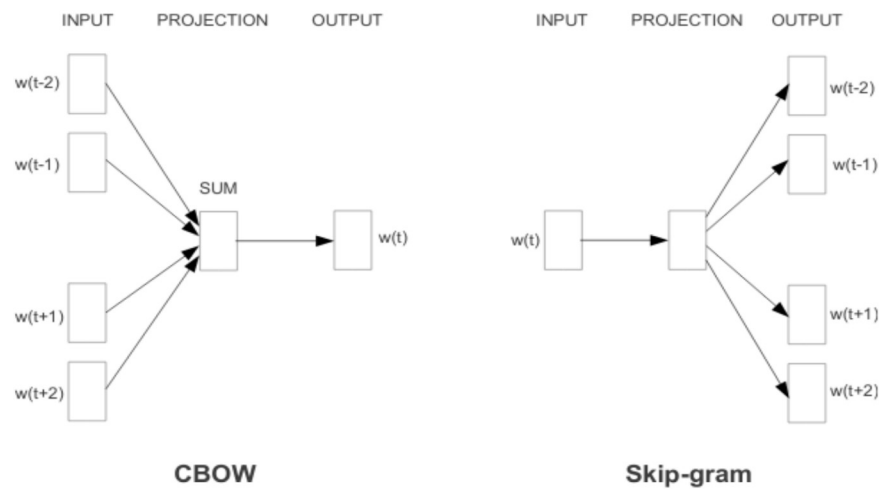
Algoritmos de aprendizado de máquina não-supervisionado realizam *word-embeddings* em modelos de redes neurais. *Word-embedding* é uma representação de palavras de um texto na forma de vetores⁴⁴ matemáticos. Os valores desses vetores acompanhados de vetores de outras palavras no texto são analisados com o objetivo de encontrar sentido e contexto nos textos de uma coleção de documentos. Os algoritmos de *word-embeddings* podem ser treinados pelo próprio corpus da pesquisa, mas também existem conjuntos de vetores pré-treinados a disposição para serem usados em seu *pipeline* (GOLDBERG, 2017). Várias técnicas são empregadas para analisar textos e a representação de suas palavras na forma de vetores como: *n-grams*, *Continuous Bag-of-words (CBOW)* e *Skip-Gram* que fazem

⁴⁴ *Word vectors are numerical vector representations of word semantics, or meaning, including literal and implied meaning. So word vectors can capture the connotation of words, like “peopleness,” “animalness,” “placeness,” “thingness,” and even “conceptness.” And they combine all that into a dense vector (no zeros) of floating point values. This dense vector enables queries and logical reasoning.* (LANE; HOWARD; HAPKE, 2019, p. 182)

parte do pacote *Word2Vec* criado por Mikolov et al. (2013); além da LDA de modelagem de tópicos.

N-grams é uma técnica utilizada para auxiliar na categorização de textos de forma automática em documentos eletrônicos (CAVNAR; TRENKLE, 2001). Os *n-grams* são termos adjacentes presentes no texto, esses termos são as palavras de uma oração e o estudo da ordem e a posição dessas podem revelar informações importantes, indicando que tais podem estar relacionadas diretamente por composicionalidade ou afinidade ou indiretamente por semelhança (SILVA; SOUZA, 2014). Pares de palavras são chamados de bigramas, trios de trigramas e quartetos são os quadrigramas - os *n-grams* também são chamados de expressões multipalavras. O estudo de *n-grams* é importante, pois não apenas uma palavra individualmente apresenta sentido no texto, mas um conjunto de palavras forma um contexto semântico (ZHANG et al., 2009) e isso “as tornam relevantes no tratamento dos recursos lexicais, os quais são importantes insumos informacionais para muitas aplicações relacionadas ao PLN, tais como: Recuperação da Informação (RI), tradução automática, sumarização de texto, etc” (SILVA; SOUZA, 2014, p. 2).

Mikolov et al. (2013) propôs duas arquiteturas de modelos para minimizar a complexidade computacional envolvida na vetorização de palavras tendo em vista a maior complexidade dos modelos baseados em redes neurais com camadas ocultas não lineares. A proposta foi um modelo de linguagem de rede neural que pode ser treinado em duas etapas: primeiro, os vetores de palavras contínuas são aprendidos usando o modelo simples (proposto pelo autor) e, em seguida, o *n-gram* do modelo de rede de linguagem neural é treinado sobre essas representações distribuídas de palavras. Em relação a *Word2Vec*, a principal diferença entre os algoritmos é que o *CBOW* usa o contexto (as palavras em torno da palavra de interesse) para prever a palavra de saída e o *Skip-gram* usa a palavra de entrada (interesse) para prever o contexto como esquematizado na figura abaixo (MIKOLOV et al., 2013).

FIGURA 13 - Diferença entre *CBOW* e *Skip-gram* (*Word2Vec*)

FONTE: Mikolov et al., 2013, p. 5

4.2.2 Clusterização

Clusterização também faz parte de um aprendizado de máquina não-supervisionado, “[...] é uma técnica de organizar dados em grupos cujos membros apresentam alguma similaridade.” (MAGALHÃES; SOUZA, 2019, p. 1), tem sido fundamental em áreas como PLN, Mineração de Dados, Recuperação da Informação (mais rápida) e Reconhecimento de Padrões, Organização da Informação, Extração de Tópicos automática ou mesmo, etapas de pré-processamento para outras técnicas, como Sumarização (ALIGULIYEV, 2009; MAGALHÃES; SOUZA, 2019).

É uma ferramenta eficaz para gerenciar grandes quantidades de informações (*big data*). Ao agrupar documentos (clusterizar) semelhantes, é possível navegar de forma mais dinâmica e celere por coleções de documentos, distinguir mais facilmente seus tópicos e subtópicos e consultá-los com eficiência, entre outras aplicações. O agrupamento pode ser realizado em dois modos diferentes, rígido e flexível: no rígido, são desarticulados e não se sobrepõem por natureza, qualquer padrão pode pertencer a uma e apenas uma classe; no flexível, um padrão pode pertencer a qualquer uma ou a todas as classes com vários graus de associação, pretende-se construir um conjunto de *cluster*, bem como o relacionamento entre eles (ALIGULIYEV, 2009; DE AMORIM, 2015).

A clusterização pode ser categorizada como hierárquica ou particionada, nesta pesquisa utilizamos uma abordagem de cada: o algoritmo *K-means* faz parte da particionada, enquanto o *Ward* da hierárquica.

No *K-means* cada observação é atribuída a um *cluster* (atribuição/particionamento de *cluster*) para minimizar a soma dos quadrados dentro do *cluster*. Em seguida, a média das observações em cada *cluster* é calculada e usada como o novo centroide do *cluster*. Em seguida, as observações são reatribuídas para *clusters* e centroides recalculados em um processo iterativo até que o algoritmo alcance convergência. Além do número de *clusters* de documentos escolhidos, também é possível escolher a quantidade de palavras que o algoritmo retorna para cada *cluster* – o que indiretamente é uma forma de modelagem de tópicos para cada subconjunto (*cluster*) de documentos e de classificação: “[...] outra tarefa comum é agrupar os vetores de documentos de um corpus com algo como *k-means* para criar um classificador de documentos.⁴⁵” (LANE; HOWARD; HAPKE, 2019, p. 217, tradução nossa).

A clusterização hierárquica de Ward é também conhecida como método da mínima variância, pois a medida de distância entre os *clusters* é calculada pelas somas dos erros quadrados de cada *cluster* (DE AMORIM, 2015). Normalmente visualizado em formato de dendrograma, esse algoritmo começa com todos os pontos de dados atribuídos a um *cluster* próprio, sequencialmente, dois *clusters* mais próximos são mesclados no mesmo *cluster* e assim por diante até restar apenas um. Com essa dinâmica, o número ideal de *clusters* é encontrado avaliando a maior distância intersector/longitudinal que atravessa o máximo de linhas representativas de *clusters*⁴⁶.

4.2.3 Modelagem de Tópicos

A Modelagem de Tópicos é uma técnica de aprendizado de máquina que envolve a pesquisa em larga escala de documentos em diversos formatos para descobrir automaticamente seus tópicos (BLEI, 2012), pode ser uma ferramenta útil para a análise estatística de coleções de documentos e outros dados discretos (BLEI; LAFFERTY, 2005).

⁴⁵ Another common task is to cluster the document vectors of a corpus with something like *k-means* to create a document classifier.

⁴⁶ Disponível em: <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>. Acesso em 26/06/2020, às 06h09.

Cientistas desenvolveram algoritmos que são capazes de vasculhar os documentos e encontrar os temas mais presentes, provavelmente, mais relevantes, sejam em coleções de documentos, sites, artigos, blogs etc. Isto posto, tópicos são estabelecidos pela prevalência de palavras, em uma coleção de documentos os temas/tópicos poderão ser os mesmos, porém cada documento exibe determinado tópico, em determinada proporção (BLEI; NG; JORDAN, 2003).

A utilidade da modelagem de tópicos deriva da propriedade que a estrutura oculta inferida assemelha-se à estrutura temática da coleção. Esta estrutura oculta interpretável faz anotações em cada documento da coleção — uma tarefa difícil de executar manualmente — essas anotações podem ser usadas para auxiliar tarefas como recuperação de informações, classificação e exploração de corpus. Desta forma, a modelagem de tópicos fornece uma solução de algoritmo para gerenciar, organizar e registrar anotações em grandes arquivos de textos⁴⁷ (BLEI, 2012, p. 79, tradução nossa).

A modelagem de tópicos está em constante pesquisa e evolução, os modelos se aprimoram e se ajustam a novas necessidades. Os tópicos em um corpus podem sofrer mudanças, evoluir e considerar tais mudanças ao longo do tempo é fundamental para compreender a determinados tipos de corpora. A modelagem de evolução de tópicos é a que considera as mudanças nos tópicos ao longo do tempo, pode-se observar o surgimento de novos tópicos ou a evolução de alguns se o documento é analisado cronologicamente (ALGHAMDI; ALFALQI, 2015).

O *Latent Dirichlet Allocation* (LDA) é a forma mais consagrada da modelagem de tópicos – um modelo probabilístico que acompanha e agrupa palavras por semelhança temática. Inspirado ou derivado de um modelo mais antigo o *Latent Semantic Analysis* (LSA) ou *Latent Semantic Indexing* (LSI), um método que mede o nível de similaridade de palavras e trechos de textos após analisar um corpus, adequado para usos em situações nas quais parte das palavras-chave já são conhecidas antes da extração (CÂNDIDO JUNIOR et al., 2015).

O LDA pertence a uma classe mais ampla de modelos de tópicos probabilísticos que usam a análise bayesiana hierárquica para encontrar a estrutura semântica subjacente dos documentos textuais e duas distribuições estatísticas para descrever o processo de geração de dados latentes em cada parágrafo (JEGADEESH; WU, 2017), esse modelo “[...] permite que

⁴⁷ *The utility of topic models stems from the property that the inferred hidden structure resembles the thematic structure of the collection. This interpretable hidden structure annotates each document in the collection—a task that is painstaking to perform by hand—and these annotations can be used to aid tasks like information retrieval, classification, and corpus exploration. In this way, topic modeling provides analgorithmic solution to managing, organizing, and annotating large archives of texts.*

cada documento exiba vários tópicos com diferentes proporções e, assim, pode capturar a heterogeneidade em dados agrupados que exibem vários padrões latentes” (BLEI; LAFFERTY, 2005, p. 2).

LDA é um modelo simples, porém versátil e potente, uma das suas vantagens é poder ser ajustado, utilizando em modelos mais complexos para objetivos mais complicados, como, por exemplo: a apresentação de correlação entre os tópicos e a inclusão dos metadados na análise (BLEI, 2012). As palavras de cada documento surgem de uma mistura de tópicos, esses são compartilhados por todos os documentos da coleção; as proporções de tópicos são específicas a cada documento e extraídas aleatoriamente de uma distribuição de Dirichlet. “A distribuição Dirichlet é uma distribuição multivariada contínua, sendo uma generalização da distribuição Beta. Em estatística bayesiana, a distribuição Dirichlet é comumente utilizada por ser uma distribuição, a priori, conjugada da distribuição Multinomial.” (RODRIGUES, 2019).

As questões de pesquisa geralmente direcionam a escolha de métodos e dados, mas o inverso também pode ser verdadeiro. A disponibilidade de dados pode direcionar as perguntas de pesquisa e os métodos a serem aplicados. A escolha dos dados também depende de outros recursos à disposição de pesquisadores, incluindo teoria, especialização, laboratórios, equipamentos, redes técnicas e sociais, locais de pesquisa, funcionários e outras formas de investimento de capital (BORGMAN, 2015).

Não menos importante é o direcionamento da pesquisa, as análises e discussões - recolher/coletar os dados é apenas a primeira etapa, o pós-reunião de dados é o que definirá se a escolha foi acertada, a efetividade do tratamento das informações levantadas é o propósito. Frické (2015) aponta que diversas pesquisas em Big Data estão simplesmente coletando mais dados, oferecendo listas, catálogos e classificações e não cunhando teorias, enfrentando problemas ou proporcionando explicações.

O ponto central é que os dados não falam e é preciso uma base de conhecimento e pesquisas prévias para subsidiar uma capacidade de abstração e análise dessas informações. Do contrário, Frické (2015) aponta que muitas das pesquisas com foco em *big data* poderão encontrar relações/conexões espúrias.

A capacidade de coletar grandes quantidades de dados de maneira barata e fácil tem vantagens: os tamanhos das amostras podem ser maiores, o teste de teorias pode ser melhor, pode haver avaliação contínua e assim por diante. Mas a ciência baseada em dados, o “quarto paradigma”, é uma quimera. A ciência precisa de problemas, pensamentos, teorias e experimentos

planejados. Na verdade, a ciência precisa de mais teorias e menos dados.⁴⁸ (FRICKÉ, 2015, p. 660, tradução nossa)

Com essas breves explanações conceituais é possível seguir à próxima seção para acompanhar a sequência de técnicas realizadas na condução dessa pesquisa e partir para a apresentação dos resultados e das discussões de acordo com os objetivos estabelecidos.

4.3 Execução

Como apresentado anteriormente, essa seção detalhará o percurso metodológico realizado desde a escolha dos dados, a coleta até a obtenção dos resultados, ou seja, os passos e etapas analíticas de tratamento dos dados coletados. A execução se deu através da linguagem de programação *Python* em uma plataforma chamada *Jupyter Notebook*. Os notebooks desta pesquisa podem ser visualizados em conta hospedada no repositório do *GitHub*⁴⁹, uma plataforma de hospedagem de código-fonte.

A fim de facilitar o entendimento da sequência metodológica, foi elaborado um mapa mental visto na figura a seguir. Na sequência, o processo foi detalhado por extenso desmembrando as etapas e fases contidas no mapa.

Primeiramente, o corpus da pesquisa foi extraído categoria *Economic Issues* na seção FOIA do site da CIA. Para a extração dessa coleção de documentos foi escrito na linguagem *Python* um código com um mecanismo de *crawler* que recuperou cento e quinze arquivos. Em seguida, verificou-se pelos títulos que haviam quatro arquivos redundantes: dois deles eram cópias com versões mais curtas dos outros dois. Sendo assim, dois foram descartados e restaram cento e treze documentos.

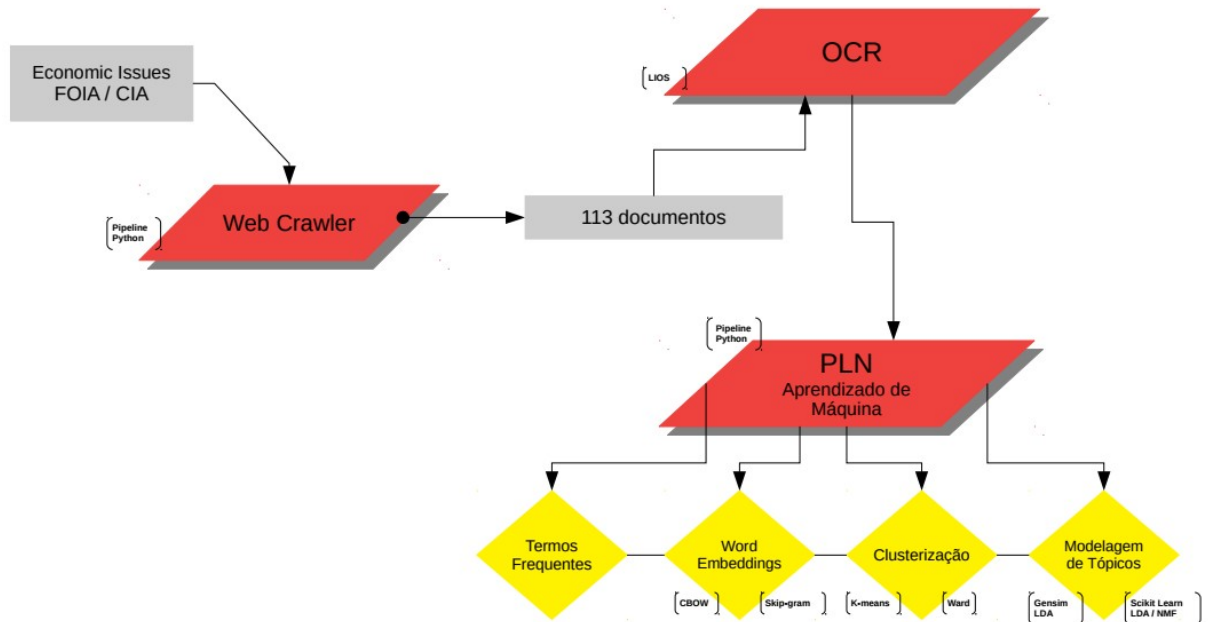
Uma vez feita essa análise para detectar duplicidades e evitar sobrevalorização de resultados, executou-se a conversão dos arquivos para um formato passível de leitura pelo código escrito. Por se tratar de documentos não tão recentes, os arquivos recuperados estavam em formato .pdf escaneados (eram imagens de documentos textuais datilografados à época de suas elaborações) e necessitaram ser convertidos em formato .txt. Para tal conversão foi

⁴⁸ *The ability to cheaply and easily gather large amounts of data does have advantages: Sample sizes can be larger, testing of theories can be better, here can be continuous assessment, and so on. But data-driven science, the “fourth paradigm,” is a chimera. Science needs problems, thoughts, theories, and designed experiments. If anything, science needs more theories and less data.*

⁴⁹ Disponível em: <https://github.com/guplages/Master-Degree>

utilizado um programa de execução OCR: *Linux-Intelligent-Ocr-Solution* (LIOS⁵⁰), um *software* livre e de código aberto.

FIGURA 14 - Mapa Mental da Metodologia



FONTE: Elaborado pelo autor.

Após a conversão realizou-se uma etapa de exploração de conteúdo do corpus, o que se chamou de uma análise textual de varredura. Antes da análise descritiva, porém foi realizada a importação de funções, pacotes e bibliotecas em *Python* que serão utilizadas na análise e visualização dos resultados, tais como *numpy*⁵¹, *pandas*⁵², *matplotlib*⁵³, *nlTK*⁵⁴, *gensim*⁵⁵, *wordcloud*⁵⁶ entre outros que podem ser verificados na pasta do repositório *Github* referenciado em nota de rodapé anteriormente.

Na sequência, foi feito o tratamento dos textos, a limpeza e normalização removendo espaços e linhas desnecessárias, as *stopwords* (artigos, preposições, símbolos etc) previamente inseridas em uma lista padrão e outras acrescentadas e atualizadas conforme

⁵⁰ Disponível em: <https://sourceforge.net/projects/lios/>. Acesso em 19/05/2020, às 05h02.

⁵¹ Disponível em: <https://numpy.org/>. Acesso em 21/07/2020, às 05h23.

⁵² Disponível em: <https://pandas.pydata.org/>. Acesso em 21/07/2020, às 05h25.

⁵³ Disponível em: <https://matplotlib.org/>. Acesso em 21/07/2020, às 05h27.

⁵⁴ Disponível em: <https://www.nltk.org/>. Acesso em 21/07/2020, às 05h29.

⁵⁵ Disponível em: <https://radimrehurek.com/gensim/>. Acesso em 21/07/2020, às 05h30.

⁵⁶ Disponível em: <https://pypi.org/project/wordcloud/>. Acesso em 21/07/2020, às 05h33.

verificação nas etapas subsequentes de análise. Então, obteve-se os termos mais frequentes em cada documento e no corpus inteiro; uma visualização em *Wordcloud*, que explicitou os termos em tamanho diretamente proporcionais a sua frequência; recuperou-se de acordo com a abordagem de *n-grams* os bigramas, trigramas e quadrigramas; Alguns termos foram escolhidos mediante sua relação com a corpora para verificar o contexto apresentando-os centralizados em vinte frases em que ocorriam.

Outra etapa ainda considerada dentro da análise de varredura foi realizada para verificar o comportamento dos termos sob vetorização e transformação em *tokens* pelos modelos *Continuous Bag-of-Words (CBOW)* e *Skip-gram*, utilizados em processamento de linguagem natural e recuperação de informações. Tal como a etapa anterior, antes de proceder à análise, as funções, pacotes e bibliotecas foram importadas, bem como o texto passou pelo mesmo tratamento, limpeza e normalização para remover espaços e linhas dispensáveis, além das *stopwords*.

Então, seguidamente, alguns termos foram escolhidos de acordo com sua frequência de aparição e relevância para a pesquisa e comparados com os *tokens* de maior afinidade/proximidade. Tal análise de semelhança e nós de relação pôde ser observada por resultados matemáticos e visualizações gráficas.

A abordagem de clusterização foi executada com dois algoritmos *K-means clustering* e *Ward*⁵⁷, sendo o último uma clusterização hierárquica. O primeiro é mais apropriado para manipular dados grandes, enquanto o segundo não. Os resultados em clusterização hierárquica são reproduzíveis, já o *K-means* pode variar se o algoritmo for executado várias vezes. O *K-means* o número de *cluster* é definido previamente assim com o número de tópicos no LDA e, conforme a necessidade e resultados observados, ajustado sucessivamente (LANE; HOWARD; HAPKE, 2019). Já em *Ward* o número ideal é avaliado conforme distâncias longitudinais dos *clusters*.

Passadas essas análises prévias, entrou-se propriamente na etapa de modelagem de tópicos. Como de praxe, primeiramente, importou-se as funções, pacotes e bibliotecas e tratou-se o texto com limpeza, normalização e remoção das *stopwords*. Um dos dicionários de *tokens* foi criado, que consiste em identificar cada termo do conjunto de documentos e atribuir

⁵⁷ Disponível em: <https://portaldatascience.com/introducao-a-clusterizacao-e-os-diferentes-metodos/>. Acesso em 20/05/2020, às 13h03.

a eles um número de identificação. Então a função *doc2bow* é aplicada vetorizando cada uma dessas palavras do dicionário atribuindo-lhes valores.

Assim, executou-se a modelagem de tópicos utilizando duas bibliotecas: *Gensim* e *Scikit Learn*⁵⁸. A diferença entre elas é o arcabouço matemático, na segunda há a opção de utilizar a fatoração de matriz não negativa (NMF) para encontrar os tópicos, além da *LDA* - lembrando que a execução em NMF requer os parâmetros de similaridade do modelo *term frequency-inverse document frequency (tf-idf)*. A similaridade diz quanto um documento ou *token* está próximo de outro qualitativamente, enquanto o *tf-idf* é uma função logarítmica do produto do *tf* e *idf*. *Tf* é a frequência de um termo no documento, já o *idf* é o inverso da frequência de aparição daquele termo nos diversos documentos do corpus (KULKARNI; SHIVANANDA, 2019)– o objetivo é equalizar o impacto de *tokens* com alta frequência, pois com essas características tendem a ser menos informativos que os com menor frequência.

Através da *Gensim*, o modelo *LDA* foi executado duas vezes: a - tratamento previamente o texto com lematização e parâmetro de treinamento de cinquenta *passes*; b - em com “estemização” e remoção de nomes próprios no singular e plural⁵⁹ e parâmetro de treinamento de cem *passes*.

Foram pré-determinados cinco tópicos e quinze termos por tópico. Lematização refere-se a tratar os termos/palavras representando-os em sua forma canônica (singular masculino para substantivos e adjetivos e infinitivo para verbos), “consiste em reunir todas as ocorrências da mesma palavra sob uma única forma, o lema, como acontece num dicionário, em vez de as apresentar tal como aparecem nos textos, com variações no gênero, no número ou na grafia” (LUCCA, 2001, p. 45). Enquanto “estemização” o tratamento visa representar os termos/palavras em sua raiz ou radical, em inglês o processo é conhecido como *stemming*, que consiste em encontrar “a raiz ou parte principal de uma palavra à qual são adicionadas inflexões ou elementos formativos⁶⁰” (tradução nossa), para este trabalho se enquadra na “remoção de sufixos das palavras por meios automáticos, o que é uma operação especialmente útil no campo de recuperação de informações” (PORTER, 2006, adaptado).

⁵⁸ Disponível em: <https://scikit-learn.org/stable/index.html>. Acesso em 21/07/2020, às 5h36.

⁵⁹ Disponível em: https://sites.google.com/site/partofspeechhelp/home/nnp_nnps. Acesso em 22/05/2020, às 17h47.

⁶⁰ Disponível em: Acesso em 17/06/2020, às 05:48

A *Scikit Learn*, através do NMF, pode gerar tópicos mais significativos em conjuntos de dados menores, enquanto a *Gensim* é mais indicado para conjuntos maiores⁶¹. É importante ressaltar que os parâmetros como número de tópicos, número de palavras por tópico e tipo de treinamento do modelo são especificados pelo pesquisador de acordo com sua preferência/objetivo e ajustes necessários de acordo com a observação dos resultados em ambas bibliotecas.

Em *Scikit Learn* também foram pré-definidos cinco tópicos com quinze termos em casa e os demais parâmetros foram para NMF: `random_state=1, alpha=.1, l1_ratio=.5, init='nndsvd').fit(tfidf)`; e LDA: `max_iter=5, learning_method='online', learning_offset=50, random_state=0).fit(tf)`.

Após apresentado o percurso metodológico da pesquisa, no próximo capítulo serão expostos os resultados encontrados e, concomitantemente, as discussões acerca dos mesmos. A interpretação e visualização dos resultados é de grande valia para compreensão do problema da pesquisa e da dimensão do alcance de seus objetivos.

⁶¹ Disponível em: <https://medium.com/mlreview/topic-modeling-with-scikit-learn-e80d33668730>. Acesso em 17/06/2020, às 04h36.

5 RESULTADOS E DISCUSSÃO

Antes de iniciar a apresentação dos resultados, optou-se por deixar no repositório do *Github* a lista com a relação dos nomes dos cento e treze documentos do corpus. Caso seja de interesse pode ser consultado no link já disponibilizado e, inclusive, pode-se ter acesso ao conteúdo dos mesmos. No notebook Pré-processamento – *N-grams* é possível ver a lista desses documentos com a respectiva contagem dos dez termos mais frequentes em cada um deles.

A apresentação dos resultados obedecerá à mesma sequência da execução dos notebooks, passando pelos tratamentos iniciais e processo de varredura dos documentos, finalizando na modelagem de tópicos. Sendo assim, a figura abaixo apresenta a primeira abordagem feita no corpus, a contagem dos vinte termos mais frequentes em todo ele.

FIGURA 15 - Termos mais frequentes no corpus

```
[('soviet', 5817), ('us', 3030), ('military', 2484), ('economic', 2468), ('ussr', 2260), ('western', 1900), ('countries', 1740), ('production', 1632), ('air', 1589), ('political', 1580), ('percent', 1569), ('intelligence', 1482), ('bloc', 1457), ('oil', 1438), ('soviets', 1413), ('world', 1373), ('probably', 1326), ('may', 1316), ('west', 1297), ('trade', 1288)]
```

FONTE: dados da pesquisa (2020)

Pode-se observar uma certa relação com a antiga URSS, produção de petróleo, blocos econômicos, comércio, política, países, hemisfério ocidental, ar (provavelmente setor de aviação) e Inteligência. Esta foi a primeira recuperação e feita de maneira geral e irrestrita, portanto, possibilita apenas um panorama, uma primeira pincelada sobre o conteúdo. Porém, é um passo para servir de referência para os demais tratamentos, pois, esperar-se que tais termos ou temas a eles relacionados voltem a aparecer.

Representou-se a lista dos vinte termos mais frequentes em um gráfico de colunas para melhor visualização da proporção entre eles. Através do gráfico é possível ver o destaque do termo “*soviet*” que se somado ao seu plural “*soviets*” realça ainda mais a importância do termo baseado em sua prevalência no corpus.

Outra forma de visualização bem interessante é a *WordCloud*, que permite distribuir em um campo bidimensional os termos de seu conjunto de documentos em análise descartando o tamanho diretamente proporcional à frequência.

Na **Figura 16** é possível verificar o destaque de termos já identificados nas análises anteriores, como dito na metodologia isso é uma forma de consolidar e subsidiar aspectos da análise documental e de recuperação de informações, que além de permitir melhor capacidade de avaliação das clusterizações e modelagem de tópicos ulteriores, atuam de forma sinérgica para um possível trabalho de indexação desse corpus.

Tendo em vista a prevalência de alguns termos, optou-se por utilizar uma técnica que permite selecionar palavras específicas, contabilizar as suas aparições e centralizá-las nas frases em que se encontram nos textos dos documentos para entender melhor o contexto em que estão inseridas.

Atendendo também aos objetivos estabelecidos no trabalho, a palavra “brazil” foi arbitrariamente uma das selecionadas para pesquisar e verificar seu perfil de aparição no corpus, na medida do possível compreender o contexto dos “interesses” da Inteligência externa Norte-americana preferencialmente ligados ao nosso país. As outras palavras escolhidas foram: *soviet*, *communist*, *anti-communist*, *capitalist*, *war*, *economic*, *production*, *ussr*, *china*, essas foram escolhidas pela frequência predominante já apresentada ou pela relação semântica (comunismo, capitalismo).

Para efeitos de discussão estão apresentados em tabela a quantidade de aparições dos termos e em relação às frases, apenas para as palavras, *soviet*, *economic* e *brazil*. A fim de visualizar as demais palavras em suas frases, pode-se acessar o repositório já mencionado no *Github*.

TABELA 2 - Aparição dos termos no corpus

Termos	Soviet	Communist	Anti-communist	Capitalist	War	Economic
Nº de aparições	7636	1514	43	21	2780	2583
Termos	Production	Oil	USSR	China	Brazil	
Nº de aparições	1657	1624	2602	645	426	

FONTE: dados da pesquisa (2020)

Os resultados demonstram através do número de aparições, a importância que os termos tinham para a Inteligência no contexto econômico. Os resultados confirmam, por exemplo, o interesse em Soviéticos/URSS quando comparado a outros países como China e Brasil e que a guerra sempre esteve no radar mesmo em documentos relacionados a economia.

Outra característica é que economia e produção têm importância significativa e indicam estar relacionados ao contexto do petróleo, que no período de Guerra Fria esteve bastante ligado aos países do oriente médio, produtores de petróleo e sujeitos a disputa capitalismo/comunismo.

FIGURA 18 - Perfil de aparição da palavra “soviet(s)”

A expressão "soviet" foi encontrada 7636 times.

aly pricing issues resolved additional deliveries soviet gas likely begin 1985 1987 either spare capacity
sting pipelines siberian pipeline completed given soviets need additional markets europe likely price comp
ould leave western europe dependent algeria libya soviet union almost 40 percent gas needs 1990 three supp
recipients nie 11-8-71 subject cost illustrative soviet intercontinental attack force models attached mem
d costs alternative force projections nie 11-8-71 soviet intercontinental attack forces costs calculated o
recipients nie 11-8-71 subject cost illustrative soviet intercontinental attack force models costing meth
costing methodology memorandum presents estimates soviet expenditures illustrative force projec tions desc
orce projec tions described section x nie 11-8-71 soviet intercontinental attack forces dated 21 october 1
t defense effect costing process amounts building soviet military budget intercontinental attack systems l
ling ond 'ibeatlon -segref -segref ze6 -ee0065 72 soviet deployment practices factors manning levels opera
lars ruble figures represent estimate forces cost soviet union useful analyzing trends comparing costs dif
comparing costs different forces programs appear soviet planners leaders dollar figures represent forces
t procured operated us useful primarily represent soviet level effort compared us determining ruble costs
ood evidence pay scales food clothing allow ances soviet armed forces calculate important area personnel e
flect best information available physical make-up soviet weapon systems soviet procedures operating mainta
available physical make-up soviet weapon systems soviet procedures operating maintaining dollar-to-ruble
n factors derived matching known us dollar prices soviet ruble prices similar items conversion factor used
ruble prices similar items conversion factor used soviet radars example based known soviet us prices non-m
ion factor used soviet radars example based known soviet us prices non-military electronic equipment item
ent cost 100 dollars mix goods services purchased soviets changes ratio overall cost expressed rubles expr

FONTE: dados da pesquisa (2020)

Selecionadas vinte linhas para apresentar a aparição da palavra “soviet(s)”, percebe-se um contexto em torno de conteúdos econômicos e estratégicos de energia, sendo avaliada a URSS como fornecedora de petróleo e derivados ao leste europeu; capacidade militar e de ataque soviético intercontinental em relação a orçamento, equipamentos, logística, custos; valor relativo de moedas (dólar-rublo) em relação a serviços e produtos. Com relação a primeira análise, nesta percebe-se que não houve distinção entre a palavra “soviet” e seu plural “soviets”, permitindo ampliar e avaliar melhor o contexto sem restringi-lo a ocorrências apenas no singular, embora a frequência de aparição seja maior que a no plural.

FIGURA 19 - Perfil de aparição da palavra “economic”

A expressão "economic" foi encontrada 2583 times.

ng possibility market weakness due number factors economic growth continues weak recovery expected latter ha
 ation substitution inventory patterns even modest economic growth percent oecd countries demand opec crude o
 83 beginning rise 1982 levels later year response economic recovery expect non-opec supplies rise 500,000 19
 on share real price oil lower c addition expected economic recovery fails materialize oil consumption contin
 ects could occur including lower inflation higher economic growth higher employment lower oil import costs r
 cret/noworn likely prices break near term greater economic growth time higher oil consumption hasten vulnera
 likewise price oil falls c stable market scenario economic growth assumptions energy price trends eritical f
 supplies could support several years fairly rapid economic expansion without strong upward pressure prices p
 ndonesia mexico egypt cut back imports could face economic austerity severe may generate degree political in
 ity mid-1980s capacity cushion likely shrink oecd economic growth rebounds productive capacity erodes opec c
 us allies way united states could insulate fully economic reverberations supply disruption nf gas markets n
 pliers considerable leverage could used political economic advantage nf e76 central intelligence agency cia
 especially among un agencies well local political economic barriers impede delivery aid people critical need
 equiring food relief shortages resulted political economic turmoil disrupted iniports distri bution within
 ountries morocco hit drought year venezuela whose economic disruption curbed production somalia conflict pre
 cropland east european countries subject drought economic disruption countries 1993 production likely 10 pe
 s present handled routine aid programs des perate economic conditions political repression haiti cuba howeve
 currently necessitating drought kind emergency e economic restructuring® humanitarian relief n natural disa
 erman foreign policy con scious country's growing economic military strength eager attain influential positi
 g since passed phase postwar reconstruction major economic power major partner western alliance econamy ix t

FONTE: dados da pesquisa (2020)

Em relação a palavra “economic”, percebe-se contextos ligados a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), Organização dos Países Exportadores de Petróleo (OPEP) e, também, de conteúdos estratégicos sobre petróleo e derivados, bem como sobre assuntos militares.

Tais análises de contexto vão sustentando interpretações quanto a importância da recuperação da informação com base na frequência dos termos e da relação desses entre si como vetores, confirmando a conjuntura político-econômica da época a que o corpus se refere. Indicam que a recuperação quantitativa de informação se mostra relativamente eficiente e as técnicas apresentam potencial de avaliação qualitativa do corpus.

FIGURA 20 - Perfil de aparição da palavra “brazil”

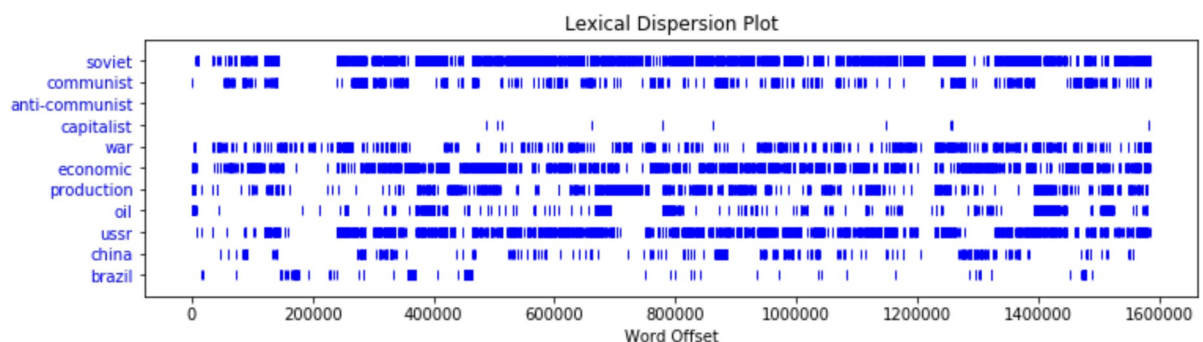
A expressão "brazil" foi encontrada 426 times.

mbourg australia 19 mexico belgium 20 netherlands brazil 21 new zealand bulgaria 22 norway canada 23 pakis
taff national security council acceptable figures brazil's national income later 1944 quasi-military force
ies reporting 100 cases more>-c australia belgium brazil burundi canada central african republic congo fra
xes a- argentina ee 19 b- bolivia ee ee ee 25 c - brazil eee feet neem en d- chile ee 37 e- colombia ee f-
entina fo ee ee ew 88 organization civil aviation brazil «j « 130 airlines brazil se 88 organization civil
anization civil aviation brazil «j « 130 airlines brazil se 88 organization civil aviation chile pte te be
eration bulk latin american fleet owned countries-brazil chile mexico example-tlikely closely associated u
effi ciency hypothetical run us atlantic seaboard brazil example lift capacity 60 c-47's 810 tons addition
tensive well-organized chain airfields facilities brazilian north coast carib bean consideration highest i
apable transporting 2,365 tons per month operated brazil us see p could deliver volume beryl quartz crista
s argentina increase 140,000 miles-1,400 per cent brazil 460,000 miles-over 700 percent venezuela 170,000
s international airline contrast argentina mexico brazil whose principal international carrier well able c
v page-18 country argentina aeroposta bolivia lab brazil aero geral aerovias brasi lap latb british guiana
lia italy iberia spain latin america cruzeiro sul brazil panair brasil causa uruguay lan chile braniff ser
region eastern part network serves lowlands near brazilian border north argentine border south equipment
astern part bolivia two cabotage routes extending brazilian border thence campo grande brazil foreign sche
es extending brazilian border thence campo grande brazil foreign scheduled airlines international routes p
tes panagra con nect bolivia peru chile argentina brazil early 1949 second us airline braniff airways inst
t begun operations route sofas page-29 appendix c brazil government organization civil avia tion order ful
created 20 january 1941 direct aviation activity brazil civil military control civil air affairs vested v

FONTE: dados da pesquisa (2020)

Nessa amostragem sobre o “brazil”, apenas pode-se inferir algum interesse pelo setor de aviação civil(organização, rotas, capacidade etc) e de mineração; sobre limites territoriais e vizinhos continentais, renda nacional e alguma análise comparativa com outros países. Com relação ao Brasil é mais seguro aumentar a quantidade de linhas e a janela de palavras por linha para uma análise mais precisa e profunda.

FIGURA 21 - Dispersão Lexical



FONTE: dados da pesquisa (2020)

Outra consideração em relação a análise de aparição ou frequência dos termos, é a sua importância de acordo com a dispersão em um corpus. A dispersão lexical é uma medida

da homogeneidade de uma palavra entre as partes de um corpus⁶². O gráfico acima mostra a ocorrência das palavras que apresentamos na **Tabela 2**.

A análise do gráfico corrobora os valores relatados na **Tabela 2**. Isto é, palavras cujo número de aparição foram maiores (*soviet, war, economic, production, ussr*) apresentam uma intensidade maior de preenchimento de barras no gráfico e tendem a maior homogeneidade e pontos de coagulação/concentração das barras. Já palavras com menor número de aparições se apresentam com menor intensidade e homogeneidade, isto é, poucas e dispersas “barras”. É o que se esperava e pôde-se observar para *soviet, war e economic* (7636, 2780 e 2583 aparições) - alto preenchimento e alta proximidade; e *china, brazil e capitalist* (645, 426 e 21) - menos barras com maior dispersão das mesmas.

Foi realizada uma análise de semelhança e nós de relação de um determinado *token*/termo perante outros utilizando os modelos *CBOW* e *Skip-gram* presentes na biblioteca Gensim. A título de comparação selecionou-se três *tokens* (*ussr, soviet, brazil*) que podem ser visto nas figuras abaixo, as duas primeiras por serem das mais frequentes e a terceira para explorar mais a visão de Brasil expressada no corpus.

Primeiramente são apresentados os valores e *tokens* em relação aos acima mencionados (quanto mais próximo, mais semelhante ou mais relacionado) em seguida os gráficos de nós⁶³ em que ficou convencionado destacar em primeiro nível os cinco nós principais e deles outros dez *tokens*. Tal qual a análise anterior cujo objetivo foi recuperar o termo/vetor no texto posicionando-o ao centro da linha com diversos outros termos/vetores semelhantes ao seu lado para tentar compreender seu contexto, a demonstração de valores dos *tokens* e nós foi de finalidade semelhante.

⁶² Disponível em: <https://www.scikit-yb.org/en/latest/api/text/dispersion.html>. Acesso em 21/05/2020, às 05h47.

⁶³ Maiores explicações sobre a exibição de nós. Disponível em: <https://networkx.github.io/documentation/stable/tutorial.html>. Acesso em 25/05/2020, às 04h52.

TABELA 3 - Valores perante o *token* “ussr”

<i>CBOW</i>	<i>Skip-gram</i>
ussr	ussr
('satellites', 0.9108450412750244)	('alone', 0.6414995193481445)
('coun', 0.876072108745575)	('zones', 0.629094660282135)
('tries', 0.8752028942108154)	('possibly', 0.6121903657913208)
('orbit', 0.8702796697616577)	('gives', 0.6093751192092896)
('ern', 0.8644266128540039)	('considered', 0.6066901087760925)
('loss', 0.8621011972427368)	('firm', 0.6064708232879639)
('austria', 0.8475918769836426)	('whole', 0.6060540676116943)
('nato', 0.8394609689712524)	('orbit', 0.6059299111366272)
('obtain', 0.8379064798355103)	('relative', 0.6012628674507141)
('japan', 0.8299315571784973)	('constitute', 0.5999215841293335)

FONTE: dados da pesquisa (2020)

Observa-se pela **tabela 3** que os *tokens* em *CBOW* são mais semelhantes à palavra “ussr” que os em *Skip-gram*. Isso pode ser visto considerando que o valor mais baixo no primeiro é maior que o mais alto no segundo, mais ainda, quando comparados os valores do *token* “orbit” em cada um deles. É importante novamente salientar que cada modelo busca relações diferentes (*CBOW*, palavras – *Skip-gram*, contextos), sendo assim, diferem os *tokens* em si, bem como seus valores. Outra observação que merece aprimoramento do código no tratamento das palavras está nos dois *tokens* em *CBOW* – ao que parece é uma apenas um que foi quebrado em dois (*countries* = *coun* + *tries*).

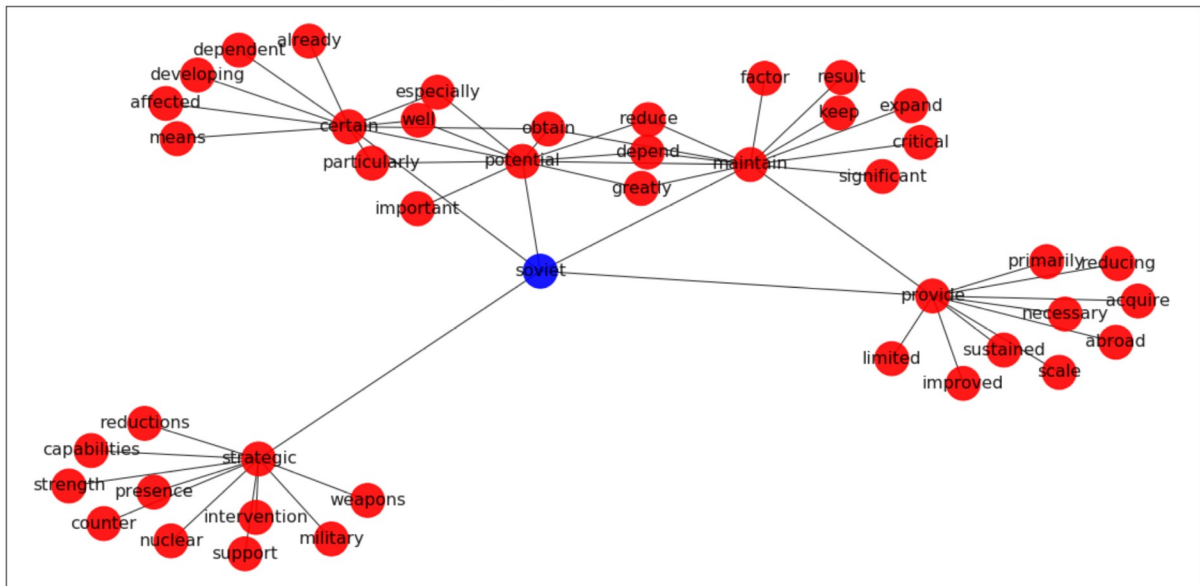
percebe-se maior distanciamento dos nós – como se apresenta na **Figura 22**. Intuitivamente, é esperado que os primeiros nós sejam os *tokens* mais próximos do *token* de referência vistos na **Tabela 3** e em parte são, todavia podem ser encontradas novas relações na rede e outro(s) *token(s)* figurar(em) no primeiro nível dos nós, bem como nos demais níveis.

TABELA 4 - Valores perante o *token* “soviet”

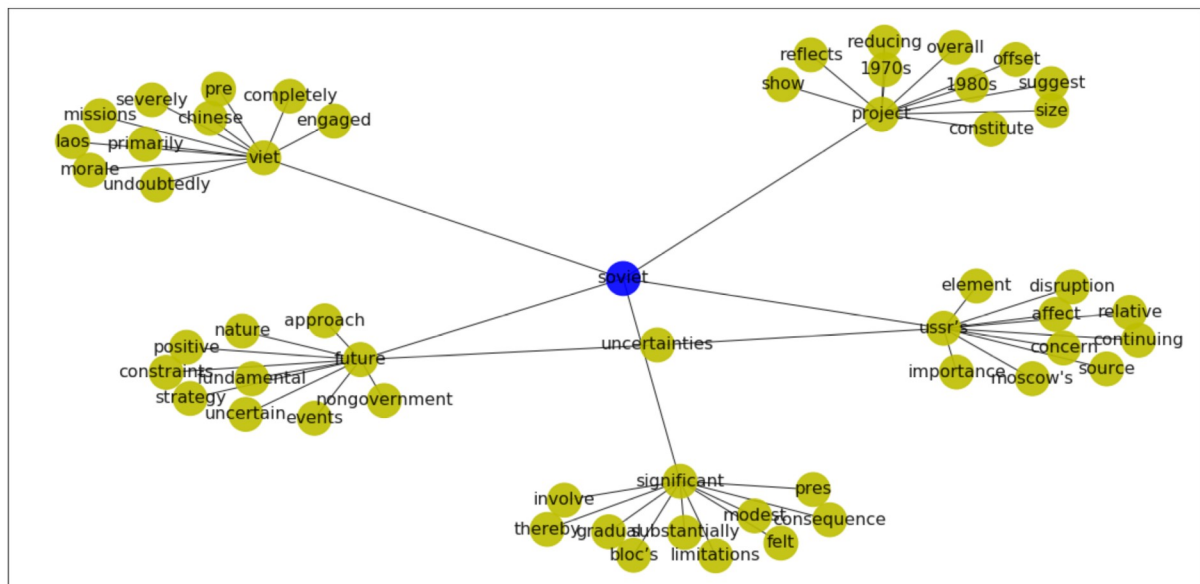
<i>CBOW</i>	<i>Skip-gram</i>
soviet	soviet
('strategic', 0.681423008441925)	('viet', 0.6943706274032593)
('potential', 0.667866587638855)	('ussr's', 0.6749670505523682)
('primarily', 0.6550830602645874)	('moscow's', 0.5667557716369629)
('reduce', 0.6486876010894775)	('bloc's', 0.5646693706512451)
('capabilities', 0.6448086500167847)	('expanding', 0.5635499954223633)
('viet', 0.6415095329284668)	('significant', 0.554236114025116)
('military', 0.6403529644012451)	('modest', 0.5528678894042969)
('strength', 0.6327085494995117)	('related', 0.5494974255561829)
('maintain', 0.6220411658287048)	('pres', 0.5475271940231323)
('expansion', 0.6204479336738586)	('continuing', 0.5470587611198425)

FONTE: dados da pesquisa (2020)

O *token* “soviet” já não apresenta diferença significativa de valores em ambos modelos quando analisada a **Tabela 4**, sendo que os dois primeiros *tokens* são muito próximos (*strategic* e *viet*) e os dois últimos não tão distantes (*expansion* e *continuing*). Dentre os *tokens* apresentados nos dois modelos, há apenas um em comum (*viet*) e, apesar de valores diferentes, eles são relativamente parecidos (*CBOW* – 0.6415, *Skip-gram* – 0.6943). Em valores, a palavra “viet” de *CBOW* estaria em terceiro lugar em *Skip-gram*, muito próximo do primeiro em que se posiciona a mesma palavra recuperada diretamente em *Skip-gram*. As diferentes características de abordagem dos modelos são percebidas quando apenas há um *token* em comum e o *token* que figura em primeiro lugar (*viet*) em *Skip-gram*, só ocupa o sexto em *CBOW*.

FIGURA 24 - Nós perante o *token* “soviet” em CBOW

FONTE: dados da pesquisa (2020)

FIGURA 25 - Nós perante o *token* “soviet” em Skip-gram

FONTE: dados da pesquisa (2020)

Na análise dos nós da palavra “soviet” percebe-se que as distâncias entre os nós de primeiro nível não são visivelmente tão diferentes nos dois modelos como se viu para a palavra “ussr” anteriormente. Há também no primeiro nível de relação da rede de ambos modelos *tokens* que não figuraram na lista de valores da **Tabela 4** – demonstrando que o

algoritmo gerador da rede de palavras não está atrelado àquele gerador da relação de valores dos *tokens* das tabelas, a despeito da rede de nós apresentar mais *tokens* que a tabela.

TABELA 5 - Valores perante o *token* “brazil”

<i>CBOW</i>	<i>Skip-gram</i>
brazil	brazil
('uar', 0.8822234869003296)	('banks', 0.8311260938644409)
('banks', 0.8810585737228394)	('peru', 0.8140648603439331)
('refugees', 0.8803904056549072)	('debt', 0.7888144254684448)
('occurred', 0.8796871304512024)	('creditors', 0.7762261629104614)
('came', 0.8749505281448364)	('imf', 0.7624154090881348)
('followed', 0.8616948127746582)	('brazilian', 0.7611452341079712)
('tween', 0.8554093837738037)	('guatemala', 0.7392417192459106)
('existence', 0.8553272485733032)	('payments', 0.7365383505821228)
('others', 0.8542169332504272)	('bilateral', 0.7011013031005859)
('record', 0.8537944555282593)	('argentina', 0.6869387626647949)

FONTE: dados da pesquisa (2020)

Para a palavra “brazil” observamos comportamentos semelhantes às duas palavras analisadas anteriormente, com características de ambas e em proporções diferentes. Tal como no *token* “ussr”, o “brazil” em *CBOW* apresenta valores mais altos de *tokens* que em *Skip-gram*, sendo o valor mais baixo do primeiro modelo (‘record’, 0.8537) maior que o mais alto do segundo modelo (‘banks’, 0.8311), por outro lado a diferença de valor entre esses *tokens* (*record* e *banks*) é menor que entre os *tokens* (*japan* e *alone*) na análise da palavra “ussr”. Também há uma palavra (*banks*) em comum nos dois modelos: o valor de “banks” em *CBOW* ocuparia o primeiro lugar em *Skip-gram* seguido do mesmo *token* em segundo lugar recuperado diretamente em *Skip-gram* (semelhante ao observado para o *token* “viet” em “soviet”); todavia, considerado em seus respectivos modelos, o *token* *banks* ocupou o segundo lugar em *CBOW* e o primeiro em *Skip-gram*.

Outro aspecto importante é que o contexto prévio em *CBOW* mostra-se mais abstrato se comparado a *Skip-gram*, cujo o contexto financeiro está mais evidente. Há de se registrar pelo *token* em comum e os outros (“debt”, “creditors”, “imf”, “payments”) a relação com temática financeira envolvendo participação de organizações internacionais (*International Monetary Fund* – IMF). Percebe-se a relação com o período histórico, já que o

FONTE: dados da pesquisa (2020)

Os nós relativos ao *token* “brazil” são interessantes quando analisados os primeiros níveis. Além de terem proximidades semelhantes em ambos modelos, é o primeiro caso que apresenta a palavra em comum “banks” nos dois modelos e no primeiro nível dos mesmos. Ademais, a distância do *token banks* em relação ao *brazil* nos dois modelos está bem parecida, sugerindo que a ordem em que os *tokens* aparecem e seus valores (**Tabela 5**) influenciam, respectivamente, no nível em que se apresentam nos nós (primeiro, segundo etc) e na distância em que esses nós estão do nó de referência.

Daí, infere-se que a despeito de não estar diretamente atrelada, a lógica de disposição dos *tokens* nos nós em relação a ordem de valores recuperada guarda semelhança: a ordem dos *tokens* com base nos valores está para o nível de posicionamento dos *tokens* nos nós e os valores em si estão para a distância desses nós. As subsequentes abordagens e técnicas mostram sinergia quanto aos relacionamentos entre os *tokens*, uma vez que se percebe alguns dos *tokens* presentes na rede de nós e tabelas (*CBOW* e *Skip-gram*) também figurarem no perfil de aparição dos *tokens* em linha. Isso demonstra que as diferentes técnicas apresentam complementariedade, corroborando a relação de semelhança vetorial entre os *tokens*.

Até então foram apresentados resultados da etapa que se chamou de varredura dos documentos, uma fase que se estabeleceu antes da modelagem de tópicos para explorar o corpus, conhecê-lo e subsidiar avaliações sobre etapas subsequentes. Foi uma espécie de validação qualitativa, uma referência crítica para os resultados de modelagem de tópicos, uma vez que, durante a varredura também executaram-se técnicas para encontrar relações e semelhanças dentro do corpus. Isso quer dizer que os resultados da modelagem de tópicos não poderiam ser opostos aos da etapa de varredura, em certa proporção eles deveriam se confirmar e ampliar.

Desta forma, para ampliar e consolidar a etapa de “varredura”, serão apresentados resultados de clusterização dos documentos com os algoritmos *K-means clustering* e *Ward*. Esse procedimento como os demais executados no âmbito de PLN faz parte da exploração prévia do corpus, uma maneira de conhecê-lo antes de analisá-lo sob a ótica da modelagem e, assim, ter mais capacidade de inferir sobre o conteúdo dos tópicos recuperados.

Para o *K-means* definiu-se a recuperação de cinco *clusters* com dez termos em cada um e plotou-se os respectivos documentos de cada *cluster* identificados por cinco cores diferentes. Para a visualização sem rótulos (título dos documentos) sobrepostos utilizou-se o D3.js⁶⁵, um navegador baseado em javascript interativo; e o *matplotlib D3* (mpld3⁶⁶), que oferece uma sintaxe *matplotlib* web interativa no qual passar o mouse sobre regiões realça informações sobre as mesmas. Os títulos dos documentos em cada *cluster* podem ser conferidos no repositório do *Github* referenciado na nota de rodapé “42”.

QUADRO 4 - Termos recuperados nos *clusters* em *K-means*

Cluster 0: Palavras: ussr arab moscow europe communist influence chinese leaders armed german		Cluster 1: Palavras: blocs trade soviet credits communist ussr exports china free underdeveloped	
Cluster 2: Palavras: communist ussr italy satellite colonial zone german germany industry transportation	Cluster 3: Palavras: republics brazil aid reformers african africa system south pakistan region	Cluster 4: Palavras: oil production percent expenditures tons million total industry ussr grain	

FONTE: dados da pesquisa (2020)

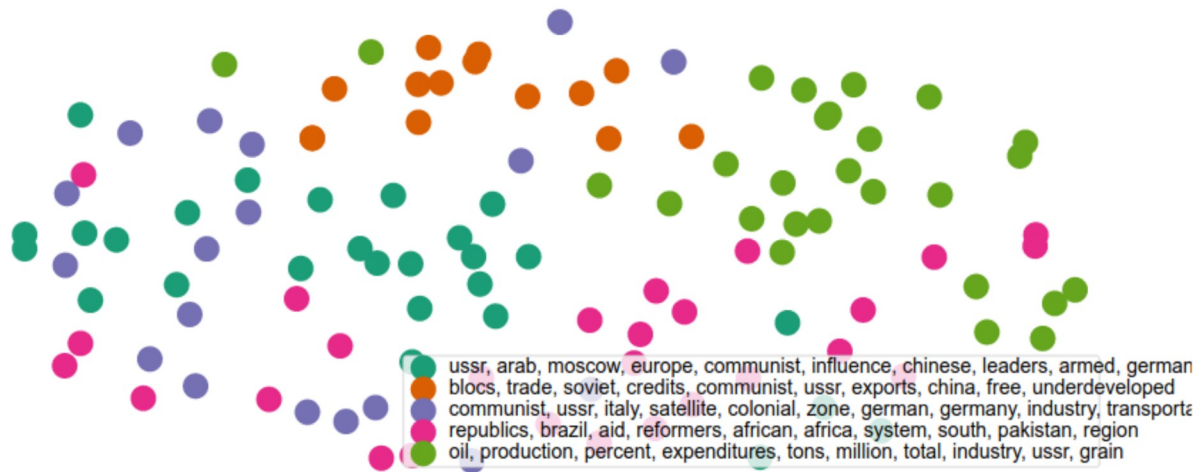
No **Quadro 4** estão relacionados os termos recuperados através do algoritmo do K-means e que foram utilizados como referência na plotagem da **Figura 27**. O *Cluster 0* indica relação com a influência comunista e traz países que adotavam regime de governo

⁶⁵ Disponível em: <http://d3js.org/>. Acesso em 22/05/2020, às 15h43.

⁶⁶ Disponível em: <https://mpld3.github.io/>. Acesso em 22/05/2020, às 16h00.

socialista, como “ussr”, “china” e “german”, provavelmente a oriental. O *Cluster 1* ainda guarda relação com o comunismo e a URSS, porém em contexto mais relacionado ao comércio internacional. O *Cluster 2* também relaciona comunismo e URSS, países da Europa Ocidental (Itália e Alemanha) e indústria, transportes e satélites (um dos grandes temas de disputa na Guerra Fria). O *Cluster 3* parece estar relacionado a países subdesenvolvidos e reformas econômicas a que se submeteram com ajudas financeiras internacionais, com destaque para Brasil e Paquistão e o continente Africano. Por fim, o quinto *cluster* está relacionado a investimentos/gastos com a indústria do petróleo e produção de grãos.

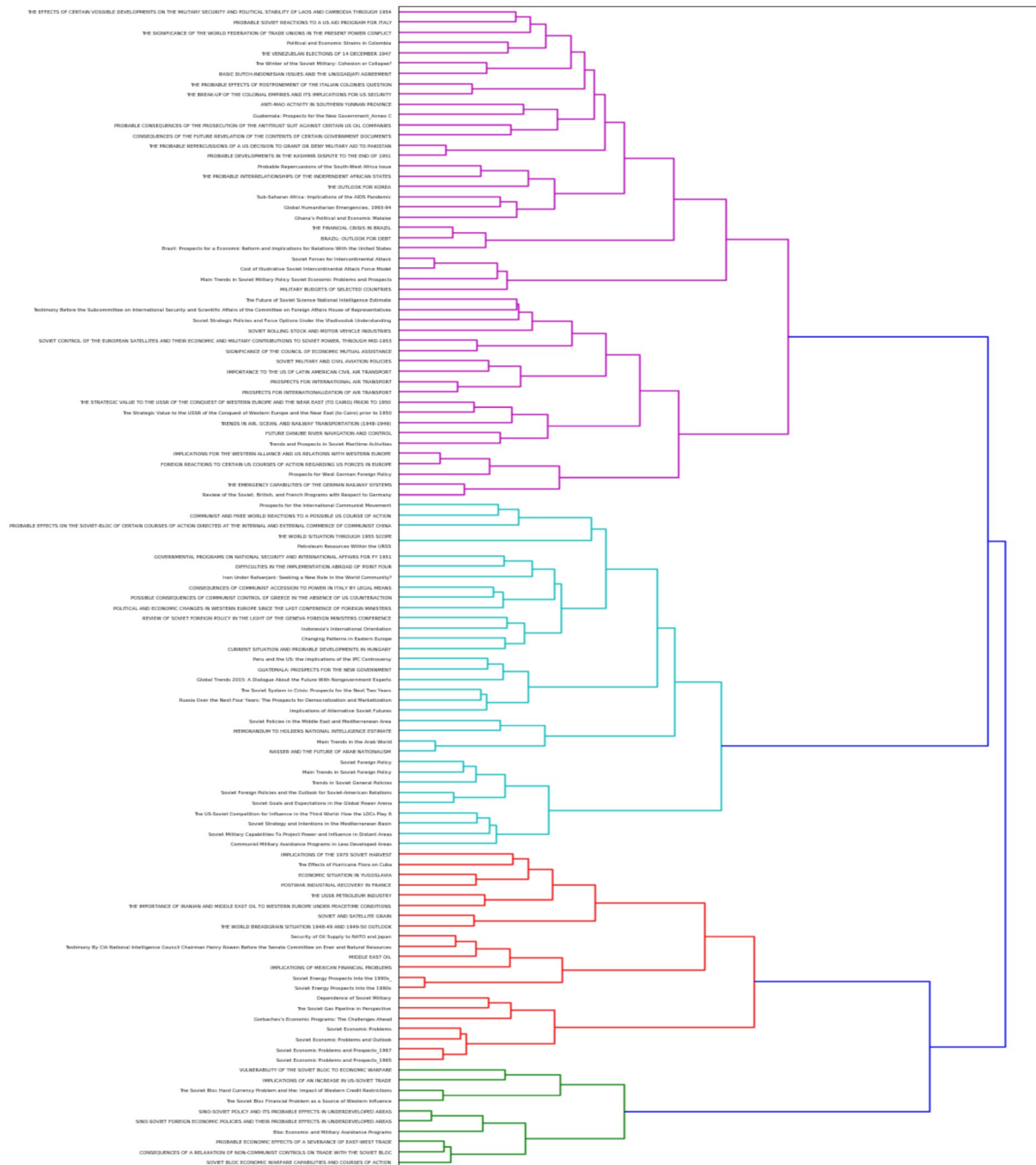
FIGURA 28 - Plot dos clusters em *K-means*



FONTE: dados da pesquisa (2020)

A abordagem da **Figura 28** é uma visualização em dendrograma a partir da execução do algoritmo *Ward*, o qual distribui cada um dos cento e treze documentos em *clusters* vizinhos por afinidade progressivamente até a formação de um só *cluster*. O estudo pode ser feito conforme as cores de distribuição e para a técnica mais comum, a definição da quantidade mais adequada de *clusters* indica um número de quatro referente às linhas azuis na horizontal. Isso porque é onde se encontra a maior distância horizontal que uma linha transversal corta os *clusters*, para os demais as distâncias seccionais são menores. Este número é próximo do número(cinco) pré-definido na abordagem para clusterização em *K-means*. Considerando que na avaliação em *K-means* um dos *clusters* se mostrou disperso, caso ele seja descartado permanecem quatro *clusters* que se compatibilizam com a avaliação do algoritmo em *Ward*.

FIGURA 29 - Clusters em Ward



FONTE: dados da pesquisa (2020)

Depois de executadas as limpezas e tratamentos no texto de todo o corpus e o convertido em uma matriz de termos-documentos, foram aplicadas os modelos (a e b) de LDA presente na biblioteca *Gensim* e os modelos de LDA e NMF da biblioteca *Scikit Learn*. Os resultados são vistos abaixo.

FIGURA 30 - Modelagem de Tópicos LDA (a) / *Gensim*

```
(0, '0.016*air" + 0.010*u" + 0.008*airline" + 0.008*civil" + 0.008*force" + 0.007*aviation" + 0.007*route"
+ 0.006*service" + 0.005*government" + 0.005*aircraft" + 0.004*soviet" + 0.004*international" + 0.004*operat
ion" + 0.004*policy" + 0.004*transport"')

(1, '0.013*soviet" + 0.008*u" + 0.008*economic" + 0.006*state" + 0.006*military" + 0.006*political" + 0.005
*policy" + 0.004*ussr" + 0.004*force" + 0.004*intelligence" + 0.004*power" + 0.004*western" + 0.004*countr
y" + 0.003*government" + 0.003*probably"')

(2, '0.022*soviet" + 0.013*bloc" + 0.007*trade" + 0.006*military" + 0.006*country" + 0.006*force" + 0.006*u
ssr" + 0.005*world" + 0.005*communist" + 0.004*economic" + 0.004*u" + 0.004*naval" + 0.003*control" + 0.003
*ship" + 0.003*china"')

(3, '0.013*soviet" + 0.010*production" + 0.008*oil" + 0.007*ton" + 0.006*year" + 0.006*million" + 0.006*per
cent" + 0.006*western" + 0.006*grain" + 0.005*ussr" + 0.004*total" + 0.004*export" + 0.004*country" + 0.004
*europa" + 0.004*energy"')

(4, '0.006*soviet" + 0.006*oil" + 0.006*percent" + 0.006*total" + 0.005*ton" + 0.004*requirement" + 0.004*m
illion" + 0.004*production" + 0.004*year" + 0.004*gasoline" + 0.004*fuel" + 0.004*car" + 0.004*brazil" + 0.0
04*ussr" + 0.004*1949"')
```

FONTE: dados da pesquisa (2020)

A partir da análise dos cinco tópicos da **Figura 29** percebe-se em todos a presença do termo “soviet” em maior ou menor influência e do termo “ussr” em quatro deles, demonstrando que todos os tópicos estão de certa forma relacionados ao contexto da Guerra Fria. Observando os demais termos dos tópicos é possível estratificar outros contextos embutidos, por exemplo: no tópico ‘0’, aviação civil internacional; no ‘1’, algo em torno de geopolítica de poder político, econômico e militar; o ‘2’, volta-se para economia, comércio internacional, rotas marítimas e china; o ‘3’, aborda exportação, petróleo, grãos e energia; por fim, o ‘4’, recupera petróleo e combustíveis, produção, carros e Brasil.

Infere-se que em certa medida a influência geopolítica soviética em contextos econômicos de comércio internacional, aviação civil, petróleo e derivados, grãos, principalmente, na Europa, China e Brasil era monitorada.

FIGURA 31 - Modelagem de Tópicos LDA (b) / *Gensim*

```
[ (0,
  '0.007*aid" + 0.006*africa" + 0.005*region" + 0.005*third" + 0.005*global" + 0.004*naval" + 0.004*africa
  n" + 0.004*south" + 0.004*india" + 0.004*conflict'),
  (1,
  '0.009*ussr" + 0.006*communist" + 0.004*europ" + 0.004*leader" + 0.004*influenc" + 0.003*parti" + 0.003*i
  nterest" + 0.003*toward" + 0.003*arab" + 0.003*regim'),
  (2,
  '0.014*product" + 0.013*oil" + 0.010*ussr" + 0.009*industri" + 0.009*percent" + 0.007*equip" + 0.006*tota
  l" + 0.005*energi" + 0.005*export" + 0.005*produc'),
  (3,
  '0.023*bloc" + 0.018*ton" + 0.014*trade" + 0.013*million" + 0.013*grain" + 0.011*product" + 0.011*total"
  + 0.009*metric" + 0.006*communist" + 0.006*export'),
  (4,
  '0.013*civil" + 0.013*airlin" + 0.011*rout" + 0.011*aviat" + 0.009*servic" + 0.007*transport" + 0.006*air
  craft" + 0.005*american" + 0.005*schedul" + 0.005*de')] ]
```

FONTE: dados da pesquisa (2020)

A **Figura 30** também recuperou cinco tópicos utilizando o modelo LDA em *Gensim*, porém com outros parâmetros. Ao analisá-los separadamente, observa-se: no tópico ‘0’, algo em torno do hemisfério Sul/“Terceiro mundo”, continente Africano e Índia, ajuda e conflito e naval; o ‘1’, já retornou a URSS e sua influência geopolítica sobre Europa e mundo Árabe; o ‘2’, o contexto energético em torno da indústria e comércio de petróleo ligado a URSS; o ‘3’, a produção e comércio de grãos; e finalmente, o ‘4’, o serviço de aviação civil, aeronave e suas rotas, transporte.

Percebe-se bastante semelhança com os tópicos apresentados na figura anterior, embora termos como “soviet” e “ussr” não fosse recorrente na maioria (talvez pelo tratamento prévio de remoção de nomes próprios). Nesse resultado, percebe-se temas comuns como influência geopolítica da URSS, energia relacionada a petróleo e derivados, produção de grãos, transporte e serviços ligados aviação civil e naval. As variações nos tópicos ficam em torno de regiões que estão presentes em um modelo e não em outro – no modelo “a”, a China e Brasil estão presentes; enquanto no modelo “b”, a África, Índia e países Árabes, embora a região europeia estivesse presente em ambos.

A **Figura 31** traz novamente o modelo LDA, porém em *Scikit Learn*. Nessa recuperação de modelagem de tópicos também ficaram definidos os mesmos parâmetros de tópicos e termos, cinco e quinze, respectivamente. Os resultados aqui acompanham o modelo LDA (a) / *Gensim*, quanto a aparição do termo “soviet” em todos os tópicos e “ussr” em quatro deles, todavia o conteúdo militar está ainda mais presente com o termo “military” em quatro dos tópicos. A análise dos demais termos dos tópicos ainda permite relacionar: no tópico ‘0’, aviação civil e/ou militar; no tópico ‘1’, produção, petróleo e economia; no ‘2’,

europa, oriente médio, árabes e política; no ‘3’, política econômica, bloco comunista, oeste; finalmente no ‘4’, bloco comunista, comércio, produção, grãos e china.

O tópico ‘0’ em *Gensim* que não traz o termo ‘ussr’ está mais ligado à aviação civil e em *Scikit Learn* é o tópico ‘4’, ligado à produção de grãos. Os termos de referências regionais como “brazil”, “africa”, “india” não aparecem em nenhum dos tópicos na **Figura 31**, embora recupere sobre a China, Europa e árabes. No entanto, temas com comércio, petróleo, grãos, disputa geopolítica e aviação são recorrentes.

FIGURA 32 - Modelagem de Tópicos LDA / *Scikit Learn*

```

Tópico 0:
air soviet civil government military aviation ussr international 1948 page airlines policy control foreign 000

Tópico 1:
soviet ussr military western production soviets 000 percent economic oil equipment total west forces east

Tópico 2:
oil soviet arab soviets world ussr western countries military states political europe middle probably economic

Tópico 3:
soviet economic political military ussr west states policy countries probably western bloc communist world for eign

Tópico 4:
bloc tons soviet communist trade million production 000 countries metric grain total percent economic china

```

FONTE: dados da pesquisa (2020)

A **Figura 32**, também com cinco tópicos e quinze termos por tópico refere-se ao modelo NMF em *Scikit Learn*. Os tópicos nesse modelo já não apresentam a mesma presença dos termos “soviet”, “ussr” e “military” vistos no LDA / *Scikit Learn*, embora seja explícito esse conteúdo em dois deles: ‘0’ que trata também de geopolítica e comunismo na europa; e ‘2’, de comércio, bloco comunista, países subdesenvolvidos, importações e china. O tópico ‘1’ refere-se ao contexto energético, produção, preços, exportação, consumo de petróleo e derivados e moedas. O Tópico ‘3’ parece estar bem relacionado ao Brasil em suas dificuldades econômicas, pois traz termos como, deficit, débito, pagamentos, preços, programas, reformas. Por fim, o tópico ‘4’, algo em torno de produção, grãos, ar, quatro diferentes datas (1947-48-49-50), toneladas, pré-guerra, demonstrando ser o de mais difícil caracterização. De forma geral, mantém como os outros modelos temas como energia, produção de petróleo e grãos, aviação e geopolítica da Guerra Fria, retorna com o Brasil e China em diferentes tópicos, porém na análise intratópicos é o modelo com mais dispersão e dificuldade de caracterização.

FIGURA 33 - Modelagem de Tópicos NMF / *Scikit Learn*

Tópico 0:
soviet ussr military economic soviets political western west policy forces communist europe probably page east

Tópico 1:
oil gas energy companies production prices price exports percent consumption 1980s countries western currency producing

Tópico 2:
bloc trade communist countries underdeveloped non economic 1954 credits controls imports west items china west ern

Tópico 3:
brazil brazilian debt economic government payments percent president reform foreign program deficit prices lat in expect

Tópico 4:
tons 000 1949 1948 million production metric 1947 percent total year 1950 prewar grain air

FONTE: dados da pesquisa (2020)

Com os resultados da Modelagem de Tópicos elaborou-se uma matriz para combinar/concentrar os tópicos de todos os modelos para possibilitar uma visualização melhor da quantidade de tópicos representativa e facilitar a análise final realizada com base na relação final de tópicos. Os critérios foram a semelhança de termos/temas dentro dos tópicos; e quando um país figurou sozinho em um dos tópicos lhe foi dado um tópico exclusivo (Brasil e China) – a exceção da URSS, pois entendeu-se como contexto geral de Guerra fria, além de estar presente em vários tópicos. Nas colunas estão intitulados os novos tópicos, enquanto nas linhas os modelos (LDA e NMF) e os seus respectivos tópicos reenquadrados na nova classificação.

TABELA 6 - Matriz dos novos Tópicos

Tópicos	Tópico 1	Tópico 2	Tópico 3	Tópico 4	Tópico 5	Tópico 6
LDA (a)/Gensim	0	1	---	3	4	2
LDA (b)/Gensim	4	1	0	2 e 3	---	---
LDA/Scikit Learn	0	2 e 3	---	1	---	4
NMF/Scikit Learn	---	0	---	1 e 4	3	2

FONTE: Elaborado pelo autor.

Os novos tópicos são descritos abaixo por sua representação nos principais termos:

- Tópico 1: linha aérea; aviação; civil; rota; transporte, serviço; aeronave; soviético; operação; política; militar; governo; URSS; controle;

- Tópico 2: econômico, militar, político (a); Europa, árabe; bloco; inteligência; poder; força; ocidental; oeste; governo; comunismo; soviético; URSS; líder; influência; petróleo;
- Tópico 3: sul; terceiro; região; global; África, Índia; ajuda; conflito; naval;
- Tópico 4: soviético; comunista; URSS; ocidente; Europa; força; militar; econômico; produção; exportação; comércio; consumo; moeda/dinheiro; preços; energia; petróleo; gás; grãos;
- Tópico 5: Brasil; latino; econômico; déficit; débito; pagamentos; preços; programas; reformas; governo; presidente; produção; petróleo; combustível; gasolina; carro; soviético; URSS;
- Tópico 6: econômico; comércio; bloco; comunista; militar; naval; transporte; importação; ocidente; soviético; URSS; China.

Com essa reorganização dos tópicos é possível caracterizá-los melhor, pois concatenou-se aqueles temas que estavam localizados em diferentes tópicos de cada um dos modelos em tópicos únicos finais. Fazendo uma analogia com a pirâmide de produção de conhecimento da IE ou mesmo da GC, cumpriu-se a etapa de transformar os dados em informação e a seguinte seria a produção de conhecimento. Em relação a um processo de IE, foram cobertas as etapas de coleta/busca do dado, a integração/interpretação/organização da informação e a seguir será executada a etapa de produção do conhecimento simulando uma análise de cenário para o tomador de decisão.

Portanto, foram desenvolvidas discussões do ponto de vista de IE para os tópicos elencados acima (resultantes da fusão) relacionados a eventos internacionais no período da Guerra Fria, considerados pelo autor mais estratégicos para os interesses da CIA.

O tópico 1 tem como tema central a “aviação” em meio a conjuntura internacional da Guerra Fria. Existem aspectos militares e civis ligados a aviação nesse contexto, interesses comerciais e políticos balizando decisões dos governantes. Elencamos duas situações envolvendo URSS e EUA para perceber como o tema aviação se envolveu dentro de posicionamentos estatais desses governos.

Há diversos temas controversos dentro da administração do Estado na ex-URSS e um deles é o desenvolvimento da Marinha em relação aos porta-aviões e aviação marítima (KOLNOGOROV, 2005). O autor discute como líderes soviéticos resistiram ao

desenvolvimento de porta-aviões até as últimas décadas da Guerra Fria a despeito da análise de especialistas científicos e autoridades navais sobre a importância do suporte aos combatentes de superfície através da aviação a bordo de aviões. A opinião de Kolnogrov (2005) é que lutas burocráticas e percepções equivocadas levaram a priorizar o desenvolvimento de armas e outras tecnologias e imperativos socioeconômicos ditaram o ritmo lento de desenvolvimento da aviação embarcada em porta-aviões.

A aviação civil estava em crise no pós-Segunda Guerra Mundial, empresas comerciais estavam enfraquecidas e precisavam de manter/recuperar mercado para seus produtos. Uma das gigantes internacionais era a Boeing, empresa norte-americana que passou por riscos financeiros na década de 1960, esteve envolvida em processos de jogos políticos em zonas de influência da URSS, envolvendo China, Paquistão e Índia.

A China demonstrava-se um grande mercado consumidor de aeronaves e ao mesmo tempo, por ser aliada da URSS, interessava aos EUA interferir/enfraquecer essa relação alterando a balança de poder e ainda lucrar com isso (ENGEL, 2007). O autor destaca que as relações sino-americanas foram retomadas no fim da década de 1960, em um momento que a China estava em expansão econômica internacional e as vendas para ela eram vitais ao mercado internacional de aeronaves e essa relação cresceu a ponto de na década de 1990 os EUA autorizar a instalação de montadoras de partes de aeronaves da Boeing em solo chinês. Porém, como ressalta Engel (2007), essa não foi a posição norte-americana de 1946 até o fim da década de 1960. A ameaça comunista e posições ideológicas eram mais fortes a ponto de Washington evitar relações comerciais com a China e outros estados comunistas.

O Tópico 2 refere-se a zonas de influência sobre a Europa e o oriente Médio/Mundo Árabe. Um dos principais itens que dominaram o cenário internacional nessa região foram o conflito Judeus-Palestinos e a produção de petróleo em países como Arábia Saudita, Irã, Kuwait, Iraque e Emirados Árabes Unidos.

Apesar dos conflitos e desavenças regionais percebidas entre os povos judeus, iranianos, curdos e árabes não serem determinados pela Guerra fria em si e mais pela dinâmica local (KARSH, 1997), eles foram marcados pela tentativa de EUA e URSS de alterar a balança de poder na região. As dinâmicas regionais no período da Guerra Fria foram marcadas pelo nacionalismo árabe; golpes de Estado; divergências entre Egito *versus* Arábia Saudita (VALBJØRN; BANK, 2012).

É quase incontestável que as violações (alegadas pelos EUA) à conferência/acordo de Yalta⁶⁷ por parte da URSS deram força ao início do ambiente de Guerra Fria, diversos presidentes e secretários de estado norte-americanos condenavam a dominação soviética sobre o leste Europeu, por outro lado, os EUA atuaram ativamente para evitar a influência da URSS na região e obter mercado consumidor para bens de produção norte-americanos. (DAVIS, 1974).

O autor baliza entre pesquisadores que defendem a responsabilidade da União Soviética em impor regimes comunistas minoritários e aqueles que descrevem como os Estados Unidos buscaram ativamente a diplomacia atômica e a retenção de créditos econômicos para impedir o estabelecimento de uma esfera de influência soviética. Essa situação é agravada quando se coloca no tabuleiro das relações Internacionais a ativação do plano Marshall para a Europa ocidental.

O tópico 3 está relacionado ao hemisfério sul, ao continente Africano, à Índia e a questões navais regionais. A intensificação da Guerra Fria influenciou os movimentos de independência na África e acirrou rivalidades que a divisão territorial (feita desconsiderando questões culturais e étnicas) já havia construído (SCHRAEDER, 2000). O autor destaca a aliança estratégica na África entre EUA e França proporcionou uma zona de influência favorável ao capitalismo sobre a chamada África Francófônica, enquanto os EUA concentravam-se em interesses político-ideológicos, a França detinha-se em questões culturais e econômicas. Schraeder (2000) enfatiza que a relação franco-americana era pragmática e havia tensões, a França por não ter interesses ideológicos tão fortes era mais “indulgente” com nações francôfonas mais alinhadas ao socialismo como Benin, Congo e Madagascar.

Outro aspecto é que doações para África Subsaariana apresentaram condicionantes e obedeceram mais a objetivos geopolíticos estratégicos baseados no contexto da Guerra Fria que o imperativo de implementação de reformas democráticas. “Em outras palavras, o custo geoestratégico da perda de clientes pode substituir qualquer benefício percebido da promoção

⁶⁷ Em **Yalta**, são abordadas questões como o tratamento a ser dispensado à Alemanha após a guerra, incluindo sua divisão em zonas de ocupação e a eliminação de sua indústria bélica, e a perseguição aos criminosos de guerra. Também se decide pela criação do Estado polonês e lançam-se as bases para a criação das Nações Unidas. A URSS concorda em combater o Japão após a derrota final da Alemanha, recebendo em troca áreas de ocupação no Leste Europeu e da Lituânia, Letônia e Estônia. Disponível em <https://www.portalsaofrancisco.com.br/historia-geral/conferencias-de-yalta-e-potsdam>. Acesso em 02/07/2020 às 18h32.

bem-sucedida de reformas democráticas entre os países beneficiários⁶⁸.” (DUNNING, 2004, p. 411, tradução nossa).

Uma das regiões importantes para análise do período é o Chifre da África, que se constitui dos Estados da Somália, Etiópia, Eritreia e Jiboti e por sua localização geográfica estratégica a levou a arena internacional zona de interesse com potencial de crise (SCHWAB, 1978). Para o autor é uma área geopoliticamente fundamental simplesmente por sua localização, sobrepondo-se ao Oriente Médio e ao Oceano Índico, flanqueia Estados ricos em petróleo como Arábia Saudita e Iêmen, controla o Estreito de Bab el Mandeb, passagem para o Mar Vermelho, Canal de Suez e linha de vida de Israel, domina uma parte do Golfo de Áden e do Oceano Índico, rota de petroleiros. A radicalização de regimes políticos na Somália e Etiópia reverberou na região e abriu espaço para influencia, primeiramente, da URSS na Somália e Iêmen do Sul e, conseqüentemente, levaram os EUA a entrarem na balança de poder regional, sobretudo, pela forte presença da marinha⁶⁹ no Oceano Índico (SCHWAB, 1978). A política naval de ambas potências na região era a mesma e baseava-se, segundo Schwab (1978) em quatro princípios ou aspectos: dissuasão estratégica; presença naval; negação ou controle marítimo; e presença de poder sobre o continente/terra.

A influência da Guerra Fria no sul-asiático tem um marco com a assinatura em 2 de maio de 1954 do *Mutual Defense Assistance Agreement* entre EUA e Paquistão e sequencialmente este último foi ingressando em outros tratados e organizações⁷⁰ alinhados aos norte-americanos e se tornou o maior aliado asiático da América (MCMAHON, 1988). Para o autor, o alinhamento com o Paquistão por sua posição estratégica tinha objetivos de fortalecer a força militar para proteção de países do oriente médio através de um pacto com Irã, Turquia e Iraque e ao mesmo tempo conter a influência soviética na região. Contudo, segundo acadêmicos sul-asiáticos, a assistência militar americana não só alienou profundamente a Índia e o Afeganistão e fomentou seus laços com a União Soviética, como também minou fatalmente as perspectivas de estabilidade regional (MCMAHON, 1988). As tensões na região

⁶⁸ *In other words, the geostrategic cost of losing clients may override any perceived benefit from successfully promoting democratic reforms among recipient countries.*

⁶⁹ *In effect a strong American presence on the Indian Ocean was perceived as absolutely necessary if (1) the economic security of the West was to be maintained; if (2) stability and regional security were to be upheld in the Middle East and the Horn of Africa; if (3) a potential blockade of Western oil lanes by the Soviets was to be prevented, and (4) if the Red Sea and Indian Ocean were to be kept open for Israeli and Israeli bound shipping.* (SCHWAB, 1978, p. 7)

⁷⁰ *Southeast Asia Treaty Organization (SEATO); American-sponsored Baghdad Pact*

puderam ser percebidas em episódios como a disputa pelo território de Kashmir, que perduram por mais de 70 anos⁷¹ desde 1947, ano da primeira guerra ente Paquistão e Índia, além das recorrentes tensões entre os dois países.

O tópico 4 está ligado a produção, comércio de commodities (petróleo, gás, grãos) e energia na região da Europa envolvendo o contexto da Guerra Fria. Painter (2009) alega que através do Plano Marshall na Europa os EUA foram mais habilidosos em promover seus interesses energéticos ligados ao petróleo na região que a URSS não obteve o mesmo sucesso nem teve capacidade de utilizar o petróleo como um instrumento de poder e influência na região. O principal aspecto do Plano foi contribuir para a construção de uma nova Europa e seus padrões de uso de energia, bem como uma nova ordem mundial no pós Segunda Guerra, contudo, o papel do petróleo não pode ser negligenciado no desenvolvimento do Plano (PAINTER, 2009).

O autor ressalta que mais de 10% da ajuda total concedida sob o Plano Marshall financiava importações de petróleo em dólares de empresas americanas, uma vez que a Europa e o Japão iniciaram a transição do modal energético do carvão para o petróleo após a Segunda Guerra Mundial. Esse auxílio ajudou a preservar os mercados para as empresas de petróleo dos EUA, especialmente aquelas com concessões no Oriente Médio.

Uma mudança no cenário do petróleo aconteceu com as duas crises na década de setenta (1973 e 1978). As crises abalaram o poder de influência norte-americano, pois o país havia se tornado altamente dependente de importações do óleo e, consequentemente, reduziu sua capacidade de fornecimento para países aliados (PAINTER, 2014). O autor também destaca que em paralelo, o nacionalismo econômico, a guerra e a revolução no Oriente Médio, e a mudança na dinâmica da Guerra Fria levantaram questões sobre a capacidade dos Estados Unidos de terem acesso principal ao petróleo do Oriente Médio, isso minou a confiança no domínio ocidental. Para piorar a situação, a URSS ultrapassou os EUA como o principal produtor de petróleo, e o aumento dos preços permitiu aos socialistas importar grandes quantidades de grãos e máquinas ocidentais, dando a falsa percepção de sucesso do sistema político-econômico soviético.

Os grãos também foram utilizados como instrumento na balança de poder da Guerra Fria. A agricultura é fonte de alimentação e na lógica militar, soldados precisam de

⁷¹ Disponível em: <https://edition.cnn.com/2016/09/30/asia/kashmir-explainer/index.html>. Acesso em 03/07/2020, às 10h59.

alimentação. Assim, “[...] por meio de uma variedade de iniciativas públicas e privadas, o cultivo de trigo e outras linhas da ciência agrícola se tornaram parte integrante do planejamento estratégico americano do pós-guerra.” (PERKINS, 1997). A partir do fim da Segunda Guerra Mundial os EUA iniciaram um procedimento de internacionalização (ao contrário do isolacionismo que predominava) e na adoção desse novo posicionamento estava a assistência internacional na agricultura para países em desenvolvimento e/ou recuperação – isso, certamente, objetivava disseminar os valores americanos presentes no capitalismo (PERKINS, 1997). Tal posição condiz claramente com o plano Marshall e é sabido que a Europa sofria intensamente com desnutrição e má nutrição no pós Segunda Guerra Mundial, as doses diárias de alimentação bem como a variação dos itens era escassa (BADENOCH; FICKERS, 2019), sendo, portanto, um ambiente propício de atuação da internacionalização da agricultura pelos EUA.

Perkins (1997) ressalta o *Point Four Program* lançado em 1949 pelo presidente Harry S. Truman como um marco nas relações internacionais dos EUA e o quarto ponto desse programa durou até 1991 tornando-se a frente tecnocrática da Guerra Fria. Esse ponto representava o programa de ajuda internacional dos EUA ao lado de frente militar representada pela Organização do Tratado do Atlântico Norte (OTAN), representando respectivamente o *soft e hard power* dos EUA na disputa hegemônica com a URSS.

O tópico 5 concentra temas relacionados ao Brasil como finanças, reformas, petróleo e combustíveis. Brasil participa da Segunda Guerra Mundial como coadjuvante, mesmo assim, deixa-a como vencedor ao lado dos Aliados; possuía papel importante no cenário internacional como líder regional e se posicionava como um *player* global, tanto que quase figurou como sexto membro do Conselho de Segurança da ONU por indicação do presidente norte-americano Roosevelt (ERTHAL ABDENUR, 2014). No período da Guerra Fria o Brasil passa por momentos de independência e outros de alinhamento com EUA; sofre um *coup d'etat* e atravessa uma ditadura militar (1964-1985); além de enfrentar momentos de crise e euforia econômicas.

O poder de influência dos EUA sobre a América Latina diminui na primeira década após a Segunda Guerra Mundial, sobretudo pelo distanciamento com o Brasil (HILTON, 1981). Nesse período os Estados Unidos não só negaram ao Brasil manter a assistência militar (solicitação do governo brasileiro para fazer frente a Argentina) nas

mesmas proporções do período da Segunda Guerra, como priorizaram o Plano Marshall (focados na Europa) em detrimento da ajuda econômica ao país.

O autor ressalta que essa postura de distanciamento volta com força em meados da década de 1970, com posicionamentos divergentes nas matérias internacionais, sobretudo o caso do debate em 1975 sobre o acordo nuclear com a Alemanha Ocidental; e na liderança da resistência às políticas norte-americanas na região. Para ele esse segundo momento de “rompimento” ou “não-alinhamento” tem suas raízes no período anterior de clivagem mencionado no início do parágrafo, quando as prioridades nacionais eram divergentes e os conceitos sobre defesa do hemisfério conflitavam-se.

O Brasil adota na gestão de Ernesto Geisel (1974–1979) a política de “pragmatismo responsável” retomando as posições da Política Externa Independente (PEI) e persegue suas próprias prioridades relacionado-se com mais intensidade com Europa e África, enquanto preocupa-se menos com a América Latina sentindo-se seguro quanto à enfraquecida “ameaça comunista” (FONSECA JR., 2018; HARMER, 2012).

Tão importante como destacar a importância do petróleo e seus derivados para o desenvolvimento da indústria e da logística no Brasil é avaliar o contexto energético e a capacidade de alternativas para a cadeia produtiva e o escoamento da produção brasileira seja do setor primário ou secundário. *Diante das crises do petróleo na década de 1970 e a falta de autossuficiência da produção e refino internos, o Brasil desenvolveu um programa baseado em um biocombustível (etanol extraído da cana-de-açúcar) conhecido como “ProÁlcool”* (SOCCOL et al., 2005).

O programa começou em novembro 1975 e como dito foi uma resposta às consequências da primeira crise internacional do petróleo, mas também às flutuações dos preços internacionais do açúcar (HIRA; DE OLIVEIRA, 2009; SOCCOL et al., 2005). O etanol foi utilizado como mistura na gasolina em geral na proporção de 24% e a indústria automobilística brasileira adaptou motores para funcionarem plenamente à álcool impulsionado pela segunda crise do petróleo em 1978 (MATSUOKA; FERRO; ARRUDA, 2009). Através de subsídios governamentais à produção do biocombustível, o país acumulou grandes excedentes de petróleo que tiveram de ser exportados na primeira metade da década de 1980 (SOCCOL et al., 2005), por outro lado na segunda metade da década a indústria

sucroalcooleira sofre um revés com diminuição do preço internacional do petróleo e maior custo de produção do etanol em relação a gasolina (MATSUOKA; FERRO; ARRUDA, 2009).

O Brasil no início da década de 1960, após passar por um bom período de crescimento industrial, encontrava-se em graves dificuldades econômicas como inflação persistente e dívidas internacionais em moeda estrangeira que o levaram a negociações pouco vantajosas com o Fundo Monetário Internacional (MARKOFF; BARETTA, 1990). Os autores ressaltam o ambiente político igualmente conturbado com disputas internas à esquerda e à direita, típicas das percebidas nos países periféricos aos protagonistas da Guerra Fria, atingindo em 1964 o Golpe Militar. As crescentes taxas de inflação, sobretudo na década de 1980, somados aos desequilíbrios na balança de pagamentos levaram a adoção de políticas de austeridade no fim dos anos oitenta do século XX (ARIDA; RESENDE, 1985).

Os autores destacam que essas políticas consistiram de programas que envolveram desde congelamento de preços e salários até reformas monetárias com lançamento de moedas indexadas. A década de 1980 foi o pior momento de crise econômica brasileira durante o período de Guerra Fria, o crescimento anual do Produto Interno Bruto (PIB) ficou em 2,2% ao ano de 1980-1989 (MARQUETTI; MALDONADO FILHO; LAUTERT, 2010).

O último tópico (6) tem como tema central a China e os derivados do contexto de Guerra Fria, como economia, comércio, bloco; comunista; militar; naval; transporte; importação; ocidente; soviético. A relação da China com a URSS foi marcante para a disputa ideológica capitalismo *versus* comunismo, após a vitória comunista em 1949 e o surgimento da RPC. Com Mao Zedong no poder a China recebeu suporte político, econômico, tecnológico e militar da União Soviética por anos (SCHAUFELBUEHL; WYSS; ZANIER, 2019).

Segundo os autores, o auge dessa relação foi na Guerra da Coreia (1950-53), quando os dois países apoiam o lado comunista representado pela Coreia do Norte, porém, poucos anos a frente a relação começa a se deteriorar e de aliados passam a ser adversários.

A chamada *Sino-Soviet Split (1956-1966)* determinou um processo de afastamento entre as duas maiores nações comunistas e vizinhas territoriais (LÜTHI, 2008). Para o autor esse evento foi determinante para segunda metade da Guerra Fria e para Segunda Guerra do Vietnã, teve importância equiparada a outros eventos como construção do muro de Berlim, crise dos Mísseis em Cuba, a própria Segunda Guerra do Vietnã e a reaproximação Sino-

americana. Lüthi (2008) frisa que apesar de soviéticos e chineses serem marxistas-leninistas, discordavam quanto aos métodos de condução da sociedade socialista doméstica, sobretudo na condução econômica; e em relação a política externa frente ao capitalismo/imperialismo.

Mao Zedong, a partir de 1958 forçou o distanciamento com a URSS e adotou uma posição de isolamento da China em relação ao mundo, perseguindo uma pureza ideológica em detrimento de boas relações internacionais e prosperidade econômica (LÜTHI, 2008), pelo menos durante a Great Leap Forward⁷². Contudo, a disputa ideológica somou-se a disputas territoriais (apoio soviético ao *Tibet* em 1959 e à Índia em 1962) e militarização das fronteiras⁷³, quase culminando em guerra em 1969 (confronto em uma ilha no rio Ussuri, chamada Zhenbao) - o que agravou o distanciamento. É importante salientar que o afastamento em relação à União Soviética não significou um alinhamento direto com os EUA, mas sim um posicionamento equidistante dos dois países.

Essa ruptura entre URSS e China ressignificou a relação dos chineses com a Europa oriental e serviu de trampolim para uma relação mais próxima com a parte ocidental europeia (SCHAUFELBUEHL; WYSS; ZANIER, 2019). Os autores pontuam que esse processo se inicia com a importação de grãos e de tecnologia dos países do lado ocidental da Europa, numa clara transposição da Cortina de Ferro na primeira metade da década de 1960, promovido principalmente pelo distanciamento franco-americano seguido de uma aproximação sino-francesa.

Contudo, há outros elos de ligação, por exemplo a Suíça que rapidamente reconheceu a República Popular da China (RPC) e a instalação do consulado chinês em Genebra (onde havia uma sede da ONU), permitiu à RPC estabelecer ligações com a Europa Ocidental, África e América Latina; além disso a China inicia relações diplomáticas com a

⁷² **Great Leap Forward**, in Chinese history, the campaign undertaken by the Chinese communists between 1958 and early 1960 to organize its vast population, especially in large-scale rural communes, to meet China's industrial and agricultural problems. The Chinese hoped to develop labour-intensive methods of industrialization, which would emphasize manpower rather than machines and capital expenditure. Disponível em <https://www.britannica.com/event/Great-Leap-Forward>. Acesso em 05/07/2020, às 21h00.

⁷³ Accidents, such as the U-2 Incident on May 1, 1960, or Soviet Defense Minister Rodion Malinovskii's drunken insult of Mao on November 7, 1964, wreaked havoc on the relationship, but it was the Chinese side that unduly exploited them for its own benefit. (LÜTHI, 2008)
The rivalry between Beijing and Moscow for Hanoi's allegiance during the Second Vietnam War complicated Washington's attempts to contain communist expansion in East Asia, but, in my view, it simultaneously demolished the last vestiges of the alliance by 1966. (LÜTHI, 2008, p. 14)

Comunidade Econômica Europeia em 1975, vindo a assinar um tratado comercial três anos depois (SCHAUFELBUEHL; WYSS; ZANIER, 2019).

As relações entre o leste europeu e Beijing não espelharam igualmente entre todas as repúblicas satélites da URSS as relações Sino-soviéticas pós-ruptura, ou seja, alguns romperam e outros a exemplo de Checoslováquia, Polônia e Alemanha oriental mantiveram relações com RPC (SCHAUFELBUEHL; WYSS; ZANIER, 2019). Os autores destacam que a China enxergava a Europa Oriental como um mercado alternativo para suas commodities e uma boa fonte de produtos industrializados e tecnologia, além de selarem parcerias em logística como a *joint venture* de transporte marítimo (Chipolbrok) com a Polônia.

Com isso, concluímos a reorganização e discussão dos novos tópicos detectando os principais eventos internacionais que pudessem estar relacionados aos termos/temas dos tópicos sem, é claro, ter a pretensão de esgotá-los. Todavia, uma análise de cenário proporcionado pela modelagem de tópicos (assegurada pelas técnicas e resultados anteriores) é útil para a organização da informação e do conhecimento presentes no corpus e pode ainda ser mais explorada para as demais corporas presentes no repositório da CIA.

As análises mostram como a agência norte-americana esteve conectada aos principais acontecimentos internacionais em quatro continentes pelo mundo, demonstrando o alcance geopolítico dos EUA e suas ações na disputa pela hegemonia na medida em que não só os monitorava como participava direta/indiretamente.

Os procedimentos iniciais de varredura dos documentos traçaram e subsidiaram interpretações e deduções sobre os contextos a que os termos/temas/tópicos se referiam acompanhados da percepção de conjuntura histórica. Considerando os propósitos desta dissertação, recuperou-se os principais termos, temas e tópicos monitorado pela inteligência externa norte-americana, presentes nos documentos sobre assuntos econômicos (*Economic issues* / FOIA / CIA), a partir da análise textual quantitativa e qualitativa do corpus desta pesquisa. A alta presença/aparição de termos como “soviet(s)”, “ussr”, “military” e “political” condizem com o período histórico (Guerra Fria) dos documentos. Termos frequentes como “economic”, “countries”, “oil”, “production”, “trade” e “bloc” indicam que principal setor econômico monitorado era ligado à produção de petróleo.

Análises subsequentes mostraram que o termo “soviet” estava ligado ao petróleo e derivados, enquanto o termo “economic” à organizações internacionais e ambos ao contexto

militar. Já quando pesquisou-se o termo “brazil” recuperou-se informações relacionadas à aviação civil, limites/vizinhos territoriais e renda nacional. Ao avançar para análise de rede de palavras (nós em *CBOw* e *Skip-gram*) verificou-se para o termo “ussr” ligações com satélites, órbitas (indicando interesse por temas espaciais); enquanto o termo “soviet” trouxe termos como “strategic”, “viet”, “bloc’s”, “military”, sugerindo possível relação como o conflito no Vietnã (1955-1975). Referente ao termo “brazil”, os nós trouxeram termos como “banks”, “debt”, “creditors”, “imf”, que se relacionam a questões financeiras do país, suas dívidas e credores como o Fundo Monetário Internacional (FMI) vinculando ao tema da renda nacional mencionado anteriormente. Ademais, outros termos como “peru”, “argentina”, “guatemala”, observados nos nós, têm relação com a temática regional, também presente na análise das frases.

Os cinco *clusters* definidos para o algoritmo *k-means* ampliaram e confirmaram alguns dos resultados já discutidos. Por exemplo, a influência soviética do ponto de vista política e econômico em relação a outros países (principalmente, ásia e europa) estava presentes em dois *clusters*; o terceiro estava mais voltado a temas econômicos aliado a países subdesenvolvidos; outro com a produção de petróleo e grãos; e o quinto deles estava muito diversificado, não sendo possível destacar algum aspecto.

A clusterização em *Ward* resultou em um dendrograma que analisado apontou um ajuste ideal de quatro *clusters*. Esse resultado foi compatível com a análise do algoritmo anterior, principalmente, se se descartasse um dos *clusters* (de *K-means*) cuja a definição de conjunto restou prejudicada devido à diversidade de termos *intracluster*. Logo, além da temática semelhante, se teria o mesmo número de *clusters* em cada algoritmo.

Por fim, a modelagem de tópicos, cuja execução representa o fechamento do processo de interpretação e organização da informação nesta pesquisa, a fase final de recuperação dos termos/temas em forma de tópicos encontrados de forma não-supervisionada, ou seja, as relações foram encontradas automaticamente pelo modelos (LDA, NMF). Foi possível nesse formato de PLN reencontrar (como esperado) itens/termos comuns aos vistos nas análises de varredura como: soviéticos e URSS; poder militar; político e econômico; petróleo e derivados; grãos; transporte e serviços; aviação e marinha; China e Ásia; Brasil; África; Mundo Árabe; e Europa. Esses foram termos e temas mais significativos presentes

dentre os quinze recuperados nos cinco tópicos em cada um dos quatro modelos executados (dois LDA em *Gensim*; um LDA e um NMF em *Scikit Learn*).

Em relação ao Brasil, a pesquisa por indução sinalizou e demonstrou o quanto e quais termos/temas estavam relacionados. O termo “brazil” apareceu 426 vezes, aproximadamente seis vezes menos que o termo “ussr” (2602), porém bem próximo do termo “china” (645) e relacionou-se com aviação, renda e território.

Complementarmente, a clusterização recuperou *brazil* no *Cluster 3* acompanhado de outros países subdesenvolvidos. O resultado da modelagem de tópicos também trouxe resultados sobre o Brasil em dois dos quatro modelos: LDA (a)/*Gensim* no tópico 4 e NMF/*Scikit Learn* no tópico 3. Na análise dos tópicos finais, o de número “5” representou o Brasil e temas relacionados como finanças, reformas, petróleo e combustíveis. Isso prova, pelos métodos utilizados que o Brasil foi alvo da CIA no período da Guerra Fria e mostra pelos resultados encontrados e análises desenvolvidas, quais foram os assuntos e entidades provavelmente mais relacionados ao monitoramento formalizado através dos documentos de inteligência que foram desclassificados.

Essas observações são importantes para justificar o investimento em Inteligência de Estado (no Brasil, principalmente) dado que outros países o fazem com seriedade e compromisso tanto para defesa de seus interesses doméstica e internacionalmente. Como defendido por Brito (2011) e Lages (2017), o Estado é um ator privilegiado nas Relações Internacionais e nesse ambiente não possui a previsibilidade administrativa e jurídica como apresenta-se domesticamente. Qualquer ação estratégica racionalmente calculada depende de qualidade informacional tanto para o que busca alcançar como proteger (BRITO, 2011) e a IE é fundamental na persecução desses objetivos (LAGES, 2017).

A IEc, como variante da IE, pode em muitos momentos utilizar artifícios condenáveis de espionagem e/ou sabotagem (JOHNSON, 1997), todavia reais e mais frequentes do que se admite. As dicotomias apresentadas por Ruvalcaba (2019) - econômico-militar, socioinstitucional e comunicativo-cultural – representam meios de ação da IE e apresentam oportunidades de alerta/aprendizado para o Brasil, significando ou sinalizando pela necessidade de um melhor trato de sua inteligência estratégica, menor preocupação com o “inimigo interno” e mais zelo por sua soberania e interesses internacionais.

Logo, conclui-se a etapa analítica desta pesquisa em que se projetou uma análise documental através de algoritmos de PLN. Percorreu-se desde a simples contagem de termos/*tokens*, sua inserção centralizada em linhas nas frases em que ocorriam nos textos do corpus; a rede (nós) de palavras/*tokens* ligados a termos específicos circunstancialmente selecionados; a clusterização de documentos recuperando os principais termos em cada, além de um ajuste automático de “filiação” dos mesmos; por fim, a modelagem de tópicos, em que os termos foram recuperados e alocados conforme sua vinculação no corpus mediante os parâmetros escolhidos, seguido de uma compilação reorganizando os novos tópicos.

Em termos qualitativos foi possível contextualizar historicamente os temas encontrados na fase quantitativa demonstrando que a IE norte-americana atuou ativamente nas RI. A CIA acompanhou estrategicamente os principais países e setores político-econômicos de serviços e comércio mais significantes do ponto de vista do “conflito” internacional da Guerra Fria. Essa observação corroborou a teoria do autor desta pesquisa sobre a necessidade e a importância de uma IE forte e atuante - não por acaso e não só pelas ações de IE, os EUA mantiveram até os dias atuais sua posição de *hegemon* nas RI. A partir da observação, interpretação e discussão desses resultados é possível avançar para as considerações finais avaliando em que medida o problema de pesquisa conseguiu ser respondido, quais e o que dos objetivos foram alcançados e o que se vislumbra de contribuições, falhas, melhorias para esta pesquisa e oportunidades para futuras.

6 CONSIDERAÇÕES FINAIS

Os resultados encontrados - fruto da utilização de ferramentas de PLN e técnicas de aprendizado de máquina e indexação automática para interpretar e organizar informações sobre o corpus - permitiram recuperar e caracterizar elementos essenciais presentes nos documentos. Tais documentos foram obtidos da coleção *Economic Issues*, filtrados pela *National Intelligence Concil* (NIC) nos repositórios presentes no site da CIA, na seção *Freedom of Information Act Electronic Reading Room* (FOIA).

Verifica-se que foi possível executar com o aprendizado de máquina o processo de recuperação dos principais termos/temas/tópicos presentes no corpus. A princípio as relações foram pesquisadas por indução (frequência de aparição do termo, centralizado em frases de ocorrência e nós/rede de palavras), seguido pela clusterização (automática), finalizando na modelagem de tópicos. Infere-se com base nos resultados que o tema central estava em torno da Guerra Fria (relação de tensão geopolítica entre EUA e URSS) abordando as capacidades militares e econômicas, setores produtivos e de serviços (petróleo, grãos, transporte), além de zonas/blocos regionais de influência (Ásia, América do Sul, África, Europa e Oriente Médio).

Quanto ao Brasil, constatou-se que foi objeto de acompanhamento da inteligência externa norte-americana – como já apontava a pesquisa feita no site da organização *Wikileaks* – no âmbito de temas como finanças, reformas, petróleo e combustíveis.

Para uma pesquisa quali-quantitativa em que se utilizou de ferramental matemático-estatístico embutido nos algoritmos de programação para o PLN, logrou-se êxito em recuperar termos/temas/tópicos mais significativos do corpus, o que pode circunstancialmente auxiliar em um processo de indexação.

A parte qualitativa permitiu uma caracterização dos principais temas e em certa medida alguma dedução de contexto considerando a conjuntura histórica e fortalecer a argumentação em torno da importância da IE para planos e metas de política externa de um Estado Nacional. Para maior profundidade é possível explorar mais os termos recuperados isoladamente e executar a leitura completa dos documentos relacionados a cada tópico, porém esse nível de análise não foi objeto deste trabalho.

Portanto, as principais contribuições deste trabalho estão em ampliar os estudos de IE na comunidade acadêmica, sobretudo na ECI/UFMG e as discussões em torno da

importância da IE para os interesses estratégicos de um país ao verificar o teor dos documentos desclassificados pela CIA. E demonstrar as possibilidades das ferramentas de PLN e técnicas de Aprendizado de Máquina para análise documental, proporcionando alternativas tecnológicas de programação para estudos de IE, GC e RI em grandes volumes de documentos (*Big Data*).

Para futuras pesquisas abrem-se oportunidades de melhorar/ajustar e ampliar os algoritmos de PLN com objetivo de executar uma interpretação e organização de informações progressivamente melhor; pesquisar outros documentos desclassificados presentes no mesmo repositório para ampliar o escopo de análise sobre a amplitude de monitoramento da inteligência norte-americana e realizar uma análise comparada em documentos de órgãos decisórios oficiais do governo estadunidense e de organizações internacionais (Organização das Nações Unidas, Organização Mundial do Comércio) para verificar o padrão de termos recuperados e sua relação com os documentos de inteligência.

REFERÊNCIAS

- ALGHAMDI, R.; ALFALQI, K. A Survey of Topic Modeling in Text Mining. **International Journal of Advanced Computer Science and Applications**, v. 6, n. 1, p. 147–153, 2015.
- ALIGULIYEV, R. M. Performance evaluation of density-based clustering methods. **Information Sciences**, v. 179, n. 20, p. 3583–3602, 2009.
- ALMEIDA NETO, W. R. DE. **Inteligência e contra-inteligência no Ministério Público. Belo Horizonte: Dictum, 2009.** 1ª ed. [s.l.] Dictum, 2009.
- ALVARENGA NETO, R. C. D. DE; BARBOSA, R. R.; PEREIRA, H. J. Gestão do Conhecimento ou gestão de organizações da era do conhecimento? um ensaio teórico-prático a partir de intervenções na realidade brasileira. **Perspectiva em Ciência da Informação**, v. 12, n. 1, p. 5–24, 2007.
- ANGELIS, C. T. DE. The impact of national culture and knowledge management on governmental intelligence. **Jornal of Modeling in Management**, v. 11, n. 1, p. 240–268, 2016.
- AN INTRODUCTION TO CLUSTERING AND DIFFERENT METHODS OF CLUSTERING. **Analytics Vidhya**. Disponível em: <<https://bit.ly/3iI4bwP>> . Acesso em: 26 de jun. de 2020.
- ANTUNES, P. C. B. **SNI & ABIN: Entre a Teoria e a Prática**. Rio de Janeiro: [s.n.].
- ANTUNES, P. C. B.; CEPIK, M. Profissionalização da atividade de inteligência no Brasil: Critérios, Evidências e Desafios restantes. In: SWENSON, R. G.; LEMOZY, S. C. (Eds.). . **Intelligence Professionalism in the Americas**. WASHINGTON, DC: Joint Military Intelligence College, 2003. p. 109–154.
- APPEL, A. L. **A e-Science e as atuais práticas de pesquisa científica**. [s.l.] Universidade Federal do Rio de Janeiro, 2014.
- ARIDA, P.; RESENDE, A. L. **Inertial inflation and monetary reform in Brazil**. [s.l.] Pontificia Universidade Católica do Rio de Janeiro, 1985.
- AWAD, E. M.; GUAZIRI, H. M. **Knowledge Managment**. [s.l.] Pearson, 2003.
- BADENOCH, A.; FICKERS, A. **Materializing Europe: transnational infrastructures and the project of Europe**. [s.l.] Palgrave macmillan, 2019.
- BALUÉ, I. G.; NASCIMENTO, M. S. O. DO. **PROTEÇÃO DO CONHECIMENTO – UMA QUESTÃO DE CONTRAINTELIGÊNCIA DE ESTADO**. VII Encontro Nacional de Estudos Estratégicos. **Anais...**2007

BARBOSA, E. F. Instrumentos de Coleta de Dados em Pesquisas Educacionais. **Educativa**, v. out, p. 1–5, 1998.

BARBOSA, R. R. Gestão da informação e do conhecimento: origens, polêmicas e perspectivas. **Informação & Informação**, v. 13, n. 1esp, p. 1–25, 2008.

BARRETO, A. D. A. A condição da informação. **São Paulo em Perspectiva**, v. 16, n. 3, p. 67–74, jul. 2002.

BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para Análise de Sentimentos em mídias sociais. In: **Minicursos**. [s.l.] IEEE, 2015. p. 1–30.

BIAGI, O. L. O Imaginário Da Guerra Fria. **Revista de História Regional**, v. 6, n. 1, 2007.

BINDER, J. M.; JENNINGS, C. Visibility and meaning in topic models and 18th-century subject indexes. **Literary and Linguistic Computing**, v. 29, n. 3, p. 405–411, 2014.

BLEI, D. M. Probabilistic topic models. **Communications of the ACM**, v. 55, n. 4, p. 77–84, abr. 2012.

BLEI, D. M.; LAFFERTY, J. D. **Correlated topic models**. Advances in Neural Information Processing Systems. **Anais...**2005

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent Dirichlet Allocation. **Journal of Machine Learning Research**, v. 3, p. 993–1022, 2003.

BOBBIO, N.; MATTEUCCI, N.; PASQUINO, G. **Dicionário de Política**. 11^a ed. [s.l.] Universidade de Brasília, 2006.

BOISOT, M. H. **Information Space: a Framework for Learning in Organizations, Institutions and Culture**. Londres: Routledge, 1995.

BORGMAN, C. L. **Big Data, Little Data, No Data - Scholarship in the Networld World**. [s.l.] The MIT Press, 2015.

BRANDÃO, P. C. **Serviços secretos e democracia no cone sul: premissas para uma convivência legítima, eficiente e profissional**. Rio de Janeiro: Impetus, 2010.

BRASIL. Decreto de 15 de dezembro de 2017. Aprova a Estratégia Nacional de Inteligência. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 18 dez. 2017.

BRASIL. Decreto 2.134, de 24 de janeiro de 1997. Regulamenta o art. 23 da Lei nº 8.159, de 8 de janeiro de 1991, que dispõe sobre a categoria dos documentos públicos sigilosos e o acesso a eles, e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 27 jan. 1997.

BRASIL. Decreto 4.376, de 13 de setembro de 2002. Dispõe sobre a organização e o funcionamento do Sistema Brasileiro de Inteligência, instituído pela Lei nº 9.883, de 7 de dezembro de 1999, e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 16 set. 2002.

BRASIL. Decreto Nº 8.793, de 29 de junho de 2016. Fixa a Política Nacional de Inteligência. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 30 jun. 2016.

BRASIL. Lei nº 8.159, de 8 de janeiro de 1991. Dispõe sobre a política nacional de arquivos. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 9 jan. 1991.

BRASIL. Lei nº 9.883, de 7 de dezembro de 1999. Institui o Sistema Brasileiro de Inteligência, cria a Agência Brasileira de Inteligência - ABIN, e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 16 set. 1999.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 18 nov. 2011.

BRASIL. Portaria GSIPR nº 42, de 17 de agosto de 2009. Institui, no âmbito da Agência Brasileira de Inteligência - ABIN, o Programa Nacional de Proteção do Conhecimento Sensível - PNPC e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 19 ago. 2009.

BRINGING MATPLOTLIB TO THE BROWSER. **Mpld3**. Disponível em: <<https://mpld3.github.io/>>. Acesso em 22 de maio de 2020.

BRITO, V. D. P. **Sistemas de Inteligência no Brasil e nos Estados Unidos**. [s.l.] Centro Universitário Newton Paiva, 2009.

BRITO, V. D. P. **O Papel Informacional dos Serviços Secretos**. [s.l.] Universidade Federal de Minas Gerais, 2011.

BUCKLAND, M. Information as thing. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 351–360, 1991.

CÂNDIDO JUNIOR, A. et al. Topic Modeling for Keyword Extraction : using Natural Language Processing methods for keyword extraction in Portal Min@s. **REVISTA DE ESTUDOS DA LINGUAGEM**, v. 23, n. 3, p. 695–726, 2015.

CASTRO, A. L. D. DE. **A Proteção de Conhecimentos Sensíveis no Brasil - Uma proposta para ampliar a atuação do Programa Nacional de Proteção do Conhecimento Sensível da Agência Brasileira de Inteligência**. [s.l.] Centro Universitário Newton Paiva, 2009.

CAVNAR, W. B.; TRENKLE, J. M. **N-Gram-Based Text Categorization**. Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval. **Anais...**2001

CEPIK, M. A. C. **Espionagem e Democracia**. [s.l: s.n.].

CETISME. **Economic Intelligence - A Guide for beginners and Practitioners**. [s.l.] Cetisme Partnership, 2002.

CHOO, C. W. **A Organização do Conhecimento - Como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões**. [s.l.] Senac São Paulo, 2003.

CIA VISION, MISSION, ETHOS & CHALLENGES /ABOU CIA. **CIA**. Disponível em: <<https://www.cia.gov/about-cia/cia-vision-mission-values>>. Acesso em: 30 de maio de 2019.
CONTROLE E FISCALIZAÇÃO. **Agência Brasileira de Inteligência**. Disponível em: <<http://www.abin.gov.br/acesso-a-informacao/controle-e-fiscalizacao/>>. Acesso em: 21 de jan. de 2020.

CLAUSEWITZ, C. VON. **Da guerra**. [s.l: s.n.].

CORNELIUS, I. Theorizing information for information science. **Annual Review of Information Science and Technology**, v. 36, n. 1, p. 392–425, 2002.

CRESWELL, J. **Research Design: Qualitative, Quantitative and Mixed Methods Approaches**. [s.l: s.n.].

DATA MINING. **Britannica**. Disponível em: <<https://www.britannica.com/technology/data-mining>>. Acesso em 25 de jun. de 2020.

DATA-DRIVEN DOCUMENTS. **D3js**. Disponível em: <<http://d3js.org/>>. Acesso em 22 de maio de 2020.

DAVENPORT, T. H.; PRUSAK, L. **Working knowledge: how organizations manage what they know**. [s.l.] Harvard Business School, 1998a. v. 35

DAVENPORT, T. H.; PRUSAK, L. **Ecologia da informação - por que só a tecnologia não basta para o sucesso na era da informação**. São Paulo: Futura, 1998b.

DAVIS, L. E. **The cold war begins**. [s.l: s.n.].

DE AMORIM, R. C. Feature Relevance in Ward's Hierarchical Clustering Using the L p Norm. **Journal of Classification**, v. 32, n. 1, p. 46–62, 11 abr. 2015.

DRUCKER, P. F. **Post-Capitalist Society**. [s.l.] Routledge, 1993.

DUNNING, T. Conditioning the effects of aid: Cold war politics, donor credibility, and democracy in Africa. **International Organization**, v. 58, n. 2, p. 409–423, 2004.

ENGEL, J. A. **Cold War at 30,000 Feet: The Anglo-American Fight for Aviation Supremacy**. [s.l.] Harvard University Press, 2007.

ERICKSON, S.; ROTHBERG, H. Big Data and Knowledge Management: Establishing a Conceptual Foundation. **The Electronic Journal of Knowledge Management**, v. 12, n. 2, p. 108–116, 2003.

ERTHAL ABDENUR, A. Emerging powers as normative agents: Brazil and China within the UN development system. **Third World Quarterly**, v. 35, n. 10, p. 1876–1893, 26 nov. 2014.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. **COMMUNICATIONS OF THE ACM**, v. 39, n. 11, p. 27–34, 1996.

FONSECA JR., G. O Pragmatismo Responsável e a Guerra Fria. In: LIMA, S. E. M. (Ed.). . **O pragmatismo responsável na visão da diplomacia e da academia**. Brasília: Fundação Alexandre de Gusmão, 2018. p. 1–476.

FREEDOM OF INFORMATION ACT ELECTRONIC READING ROOM / LIBRARY. CIA. Disponível em: <<https://bit.ly/36YIJCC>>. Acesso em: 08 de abr. de 2019.

FRICKÉ, M. Big data and its epistemology. **Journal of the Association for Information Science and Technology**, v. 66, n. 4, p. 651–661, abr. 2015.

GENSIM: TOPIC MODELLING FOR HUMANS. **Radimrehurek**. Disponível em: <<https://radimrehurek.com/gensim/>>. Acesso em: 21 de jul. de 2020.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. 1ª edição ed. [s.l.] Editora da UFRGS, 2009.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. 6ª ed. São Paulo: Atlas, 2008.

GOLDBERG, Y. **Neural Network Methods for Natural Language Processing**. [s.l.] Morgan & Claypool, 2017.

GONÇALVES, J. B. **Atividade de Inteligência e Legislação Correlata**. 4ª ed. Niterói: Impetus, 2016.

GREAT LEAP FOWARD. **Britannica**. Disponível em: <<https://www.britannica.com/event/Great-Leap-Forward>>. Acesso em 05 de jul. de 2020.

HARMER, T. Brazil's Cold War in the Southern Cone, 1970–1975. **Cold War History**, v. 12, n. 4, p. 659–681, 10 nov. 2012.

HILTON, S. E. The United States, Brazil, and the Cold War, 1945-1960: End of the Special Relationship. **The Journal of American History**, v. 68, n. 3, p. 599, dez. 1981.

HIRA, A.; DE OLIVEIRA, L. G. No substitute for oil? How Brazil developed its ethanol industry. **Energy Policy**, v. 37, n. 6, p. 2450–2456, jun. 2009.

HISTÓRICO. **Portal de Periódicos CAPES/MEC**. Disponível em: <<https://bit.ly/3difVVy>>. Acesso em 28 de nov. de 2018.

HISTORY OF THE CIA / ABOUT CIA. *CIA*. Disponível em: <<https://www.cia.gov/about-cia/history-of-the-cia>>. Acesso em: 30 de maio de 2019.

HOBBS, T. **Leviathan**. New York: Barnes & Noble, Inc, 1873.

HOWARD, M. Military power and international order. **International Affairs**, v. 85, n. I, p. 145–155, 2009.

INTRODUÇÃO AO JSON. **Json**. Disponível em: <<https://json.org/json-pt.html>>. Acesso em: 19 de maio de 2020.

JEGADEESH, N.; WU, D. A. Deciphering Fedpeak: The Information Content of FOMC Meetings. **SSRN Electronic Journal**, p. 734–765, 2017.

JOHNSON, L. K. Economic Intelligence and the CIA. **Southeastern Political Review**, v. 25, n. 3, p. 501–514, 1997.

JUPYTER. **Jupyter**. Página inicial. Disponível em: <<https://jupyter.org/>>. Acesso em: 19 de maio de 2020.

KARSH, E. Cold War, post-Cold War: does it make a difference for the Middle East? **Review of International Studies**, v. 23, n. 3, p. S0260210597002714, jul. 1997.

KITCHENHAM, B. et al. Systematic literature reviews in software engineering – A systematic literature review. **Information and Software Technology**, v. 51, n. 1, p. 7–15, jan. 2009.

KITCHENHAM, B. et al. Systematic literature reviews in software engineering – A tertiary study. **Information and Software Technology**, v. 52, n. 8, p. 792–805, ago. 2010.

KOLNOGOROV, V. To be or not to be: The development of Soviet Deck Aviation. **Journal of Strategic Studies**, v. 28, n. 2, p. 339–359, 2005.

KULKARNI, A.; SHIVANANDA, A. **Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python**. [s.l.] Apress Media, 2019.

KUMAR, A.; SHARMA, A. Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence. **Webology**, v. 14, n. 2, p. 6–34, 2017.

LAGES, G. P. **A Influência da Inteligência Governamental sobre a Política Externa Brasileira**. [s.l.] Universidade Federal de Minas Gerais, 2017.

LANE, H.; HOWARD, C.; HAPKE, H. M. **Natural Language Processing in Action - understanding, analyzing, and generating text with python**. [s.l.] Manning Publications Co, 2019.

LEONARD-BARTON, D. **Nascentes do Saber: criando e sustentando as fontes de inovação**. Rio de Janeiro: Fundação Getúlio Vargas, 1998.

LEWIS, R. D. **When cultures collide**. 3rd. ed. Boston / London: Nicholas Brealey International, 2006. v. 26.

LINUX-INTELLIGENT-OCR-SOLUTION. **Sourceforge**. Página inicial. Disponível em: <<https://sourceforge.net/projects/lios/>>. Acesso em: 19 de maio de 2020.

LORENA, A. C.; CARVALHO, A. C. P. L. F. DE. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, v. XIV, n. 2, p. 25, 2007.

LUCCA, J. L. DE. **Minidicionários da língua portuguesa: análise léxico-estatística contrastiva das macro e microestruturas e sugestão de modelo**. [s.l.] Universidade de São Paulo, 2001.

LUCCI, E. A. A Era Pós-Industrial, a Sociedade do Conhecimento e a Educação para o Pensar. 2009.

LÜTHI, L. M. **The Sino-Soviet Split: Cold War in the Communist World**. [s.l.] Princeton University Press, 2008.

MACIEL, R. F. C.; PINHEIRO, M. M. K. Pirâmides do Conhecimento: abordagens comparativas entre Ciência da Informação e Inteligência de Estado. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 8, n. 1, p. 29–39, 2013.

MAGALHÃES, L. H. DE; SOUZA, R. R. Agrupamento de Notícias de Jornais On-line usando Técnicas de Machine Learning para Clustering de Textos no Idioma Português. **MULTÍPLOS OLHARES EM CIÊNCIA DA INFORMAÇÃO**, v. 9, n. 2, p. 15, 2019.

MARKOFF, J.; BARETTA, S. R. D. Economic Crisis and Regime Change in Brazil: The 1960s and the 1980s. **Comparative Politics**, v. 22, n. 4, p. 421–444, 1990.

MARQUETTI, A.; MALDONADO FILHO, E.; LAUTERT, V. The Profit Rate in Brazil, 1953-2003. **Review of Radical Political Economics**, v. 42, n. 4, p. 485–504, 20 dez. 2010.

MATPLOTLIB: VISUALIZATION WITH PYTHON. **Matplotlib**. Disponível em: <<https://matplotlib.org/>>. Acesso em: 21 de jul. de 2020.

MATSUOKA, S.; FERRO, J.; ARRUDA, P. The Brazilian experience of sugarcane ethanol industry. **In Vitro Cellular & Developmental Biology - Plant**, v. 45, n. 3, p. 372–381, 28 jun. 2009.

MCMAHON, R. J. United States Cold War Strategy in South Asia: Making a Military Commitment to Pakistan, 1947-1954. **The Journal of American History**, v. 75, n. 3, p. 812–840, 1988.

MIKOLOV, T. et al. Efficient estimation of word representations in vector space. **1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings**, p. 1–12, 2013.

MISSION & PRIORITIES / ABOUT. **FBI**. Disponível em: <<https://www.fbi.gov/about/mission>>. Acesso em: 30 de maio de 2019.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997. v. 17.

MORGENTHAU, H. J. **A Política entre as Nações - A luta pelo poder e pela paz**. [s.l.] Universidade de Brasília, 2003.

MURAKAMI, A. et al. ‘What is this corpus about?’: using topic modelling to explore a specialised corpus. **Corpora**, v. 12, n. 2, p. 243–277, ago. 2017.

NASCIMENTO, M. S. O. DO. **Proteção ao Conhecimento: uma proposta de fundamentação teórica**. [s.l.] Universidade de Brasília, 2008.

NASCIMENTO, M. S. O. Compartilhar ou Proteger? Grande desafio no comportamento informacional das organizações. In: STAREC, C.; GOMES, E.; BEZERRA, J. (Eds.). . **Gestão Estratégica da Informação e Inteligência Competitiva**. São Paulo: Saraiva, 2005. p. 255–270.

NATURAL LANGUAGE TOOLKIT. **NLTK 3.5 documentation**. Disponível em: <<https://www.nltk.org/>>. Acesso em: 21 de jul. de 2020.

NOGUEIRA, J. P.; MESSARI, N. **Teoria das Relações Internacionais (capa vermelha)**. [s.l.] Elsevier, 2005.

NONAKA, I.; TAKEUCHI, H. **The knowledge-creating company: how Japanese companies create the dynamics of innovation**. New York: Oxford University Press, 1995.

NUMPY. **Numpy**. Página inicial. Disponível em: <<https://numpy.org/>>. Acesso em: 21 de jul. de 2020.

O QUE É XML?. **Tecmundo**. Disponível em: <<https://bit.ly/3dl1b8t>>. Acesso em: 23 de ago. de 2018.

PAINTER, D. S. The Marshall Plan and oil. **Cold War History**, v. 9, n. 2, p. 159–175, 20 maio 2009.

PAINTER, D. S. Oil and Geopolitics: The Oil Crises of the 1970s and the Cold War. **Historical Social Research**, v. 39, n. 4, p. 186–208, 2014.

PAIVA, S. B.; ARAGÃO, P. O. R. DE; PEREIRA, S. L. Gestão do Conhecimento em uma organização baseada em conhecimento: uma abordagem qualitativa. **Produto & Produção**, v. 8, n. 2, p. 37–56, 2005.

PANDAS. **Pandas**. Página inicial. Disponível em: <<https://pandas.pydata.org/>>. Acesso em: 21 de jul. de 2020.

PERKINS, J. H. **Geopolitics and the Green Revolution: wheat, genes, and the Cold War**. [s.l.] Oxford University Press, 1997.

PINHEIRO, M. M. K. Estado informacional. **Varia Historia**, v. 28, n. 47, p. 61–77, jun. 2012.

PNPC. **Agência Brasileira de Inteligência**. Disponível em: <<http://www.abin.gov.br/atuacao/programas/pnpc>>. Acesso em: 05 de jan. de 2020.

POLÍTICA NACIONAL DE INTELIGÊNCIA. **Agência Brasileira de Inteligência**. Disponível em: <<http://www.abin.gov.br/acesso-a-informacao/legislacao-de-inteligencia/coletanea-de-legislacao/politica-nacional-de-inteligencia/>>. Acesso em: 20 de jan. de 2020.

POMBO, O. Contribuição para um vocabulário sobre interdisciplinaridade. In: POMBO, O.; GUIMARÃES, H. M.; LEVY, T. (Eds.). **Interdisciplinaridade: Reflexão e Experiência**. [s.l.] Lisboa, 1994. p. 92/97.

PORTER, M. F. An algorithm for suffix stripping. **Program**, v. 40, n. 3, p. 211–218, 2006.

PYTHON. **Wikipédia**. Disponível em: <<https://pt.wikipedia.org/wiki/Python>>. Acesso em 08 de jul. de 2018.

ROBREDO, J. Otimização dos processos de indexação dos documentos e de recuperação da informação mediante o uso de instrumentos de controle terminológico. **Ciência da Informação**, v. 11, n. 1, p. 3–18, 1982.

RODRIGUES, A. T. **O Poder da Palavra: aplicação da modelagem de tópicos na construção de indicadores econômicos**. [s.l.] FUNDAÇÃO GETULIO VARGAS, 2019.

ROTHER, E. T. Revisão Sistemática x Revisão Narrativa. **Acta Paulista de Enfermagem**, v. 20, n. 2, p. 5–6, 2007.

RUVALCABA, D. M. Estados Semi-Centrais: uma nova categoria para repensar a estrutura de poder. **Austral: Brazilian Journal of Strategy & International Relations**, v. 8, n. 15, p. 139–167, 2019.

SANTOS, I. M. DOS. **A Ciência da Informação e a Inteligência de Estado: um estudo de Caso do Programa Nacional de Proteção do Conhecimento Sensível da Agência Brasileira de Inteligência**. [s.l.] Universidade Federal de Minas Gerais, 2016.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectiva em Ciência da Informação**, v. 1, n. 1, p. 41–62, 1996.

SCHAUFELBUEHL, J. M.; WYSS, M.; ZANIER, V. Europe and China in the Cold War: Exchanges Beyond the Bloc Logic and the Sino-Soviet Split. In: HANHIMÄKI, J. M.; WYSS, M. (Eds.). **New Perspectives on the End of the Cold War**. [s.l.: s.n.]. p. 242.

SCHRAEDER, P. J. Cold War to Cold Peace: Explaining U.S-French Competition in Francophone Africa. **Political Science Quarterly**, v. 115, n. 3, p. 395–419, 2000.

SCHWAB, P. Cold War on the Horn of Africa. **African Affairs**, v. 77, n. 306, p. 6–20, 1978.

SHANNON, C. E.; WEAVER, W. **The Mathematical Theory of Communication**. [s.l.] THE UNIVERSITY OF ILLINOIS PRESS, 1964.

SI VIS PACEM, PARA BELLUM. **Dicionário de Latim**. Disponível em: <<https://bit.ly/34MNz1T>>. Acesso em: 03 de dez. De 2018.

SILVA, E. M. DA; SOUZA, R. R. Fundamentos em processamento de linguagem natural: uma proposta para extração de bigramas. **Revista eletrônica de biblioteconomia e ciência da informação**, v. 19, n. 40, p. 1–32, 2014.

SOCOL, C. R. et al. Brazilian biofuel program: An overview. **Journal of Scientific and Industrial Research**, v. 64, n. 11, p. 897–904, 2005.

SOUZA, F. DE. **Dicionário de Relações Internacionais**. [s.l.] Centro de estudos da População, Economia e Sociedade - CEPESE, 2005.

SOUZA, E. D. DE; DIAS, E. J. W.; NASSIF, M. E. A Gestão da Informação e do Conhecimento na Ciência da Informação: perspectivas teóricas e práticas organizacionais. **Informacao e Sociedade**, v. 21, n. 1, p. 55–70, 2011.

SOUZA, R. R.; ALVARENGA NETO, R. **A Construção do Conceito de Gestão do Conhecimento: práticas organizacionais, garantias literárias e o fenômeno social**. Encontro da Sociedade Brasileira de Gestão do Conhecimento. **Anais...**2003

TARAPANOFF, K. Informação, conhecimento e inteligência em corporações: relações e complementaridade. In: UNESCO/IBICT (Ed.). **Inteligência, informação e conhecimento**. Brasília: [s.n.]. p. 19–36.

TUCÍDIDES. **História da guerra do peloponeso**. 4ª ed ed. Brasília: Universidade de Brasília, 2001.

TZU, S. **A arte da Guerra**. Porto Alegre: L&PM Pocket, 2006.

UGARTE, J. M. **Controle Público da Atividade de Inteligência: a procura de legitimidade e eficácia**. Anais do Seminário Atividades de Inteligência no Brasil: Contribuições para a Soberania e a Democracia. **Anais...**Brasília: 2003.

UMA ESPIADA NA GRAMA DO VIZINHO, OU: COMO FAZER BENCHMARKING. **Endeavor Brasil**. Disponível em: <<https://endeavor.org.br/estrategia-e-gestao/benchmarking/>>. Acesso em 30 de maio de 2019.

VALBJØRN, M.; BANK, A. The New Arab Cold War: rediscovering the Arab dimension of Middle East regional politics. **Review of International Studies**, v. 38, n. 1, p. 3–24, 1 jan. 2012.

VIEIRA, S. B. Indexação automática e manual: revisão de literatura. **Ciência da Informação**, v. 17, n. 1, p. 43–57, 1988.

VON KROGH, G.; ICHIJO, K.; NONAKA, I. **Facilitando a criação de conhecimento: reinventando a empresa com o poder da inovação contínua**. Rio de Janeiro: Campus, 2001.

WALTZ, K. N. **Man, the State and War - a theoretical analysis**. New York: Columbia University, 2001.

WEB SCRAPING E WEB CRAWLING SÃO LEGAIS OU ILEGAIS?. **Data Science Academy**. Disponível em: <<https://bit.ly/33MY3z8>>. Acesso em 28 de maio de 2019.

WICKRAMASINGHE, N. et al. **Healthcare Knowledge Management Primer**. [s.l.] Routledge, 2009.

WIKSTRÖM, S.; NORMANN, R. **Knowledge and value: a new perspective on corporate transformation**. London: Routledge, 1994.

WORDCLOUD 1.8.0. **Pypi**. Disponível em: <<https://pypi.org/project/wordcloud/>>. Acesso em: 21 de jul. de 2020.

XU, W.; LIU, L.; SHANG, W. Leveraging cross-media analytics to detect events and mine opinions for emergency management. **Online Information Review**, v. 41, n. 4, p. 487–506, 14 ago. 2017.

ZHANG, W. et al. Improving effectiveness of mutual information for substantial multiword expression extraction. **Expert Systems with Applications**, v. 36, n. 8, p. 10919–10930, 2009.