

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

JOICE RODRIGUES TEIXEIRA

CONVERSÃO DE TESAUROS EM ONTOLOGIAS:
UM ESTUDO EXPLORATÓRIO

BELO HORIZONTE
2013

JOICE RODRIGUES TEIXEIRA

CONVERSÃO DE TESAUROS EM ONTOLOGIAS:

UM ESTUDO EXPLORATÓRIO

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Mestre em Ciência da Informação.

Linha de Pesquisa: Organização e Uso da Informação

Orientador: Prof. Dr. Renato Rocha Souza

BELO HORIZONTE
2013

T266c

Teixeira, Joice Rodrigues.

Conversão de tesauros em ontologias [manuscrito]: um estudo exploratório / Joice Rodrigues Teixeira. – 2013.
122 f. : il.

Orientador: Renato Rocha Souza.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 116-122

1. Ciência da informação – Teses. 2. Sistemas de Organização do Conhecimento – Teses. 3. Ontologia (Recuperação da informação) - Teses. 4. Tesouro – Teses. I. Título. II. Souza, Renato Rocha. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.4.03



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

FOLHA DE APROVAÇÃO

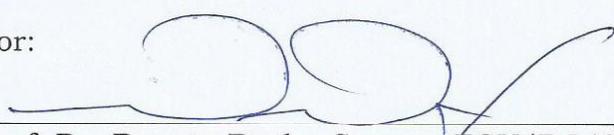
"CONVERSÃO DE TESAuros EM ONTOLOGIAS: UM ESTUDO EXPLORATÓRIO"

Joice Rodrigues Teixeira

Dissertação submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, como parte dos requisitos à obtenção do título de "**Mestre em Ciência da Informação**", Linha de Pesquisa: "**Organização e Uso da Informação - OUI**".

Dissertação aprovada em: 26 de abril de 2013.

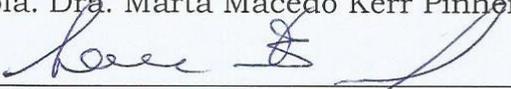
Por:



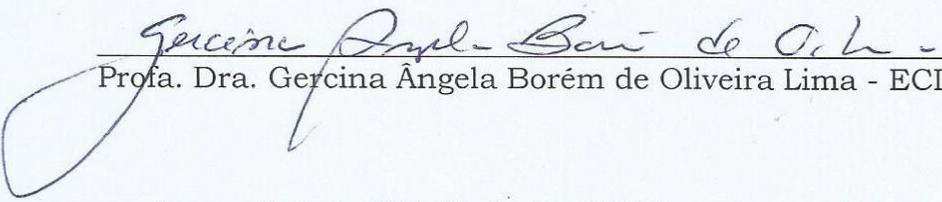
Prof. Dr. Renato Rocha Souza - FGV/RJ (Orientador)



Profa. Dra. Marta Macedo Kerr Pinheiro - Profa. Aposentada - ECI/UFMG

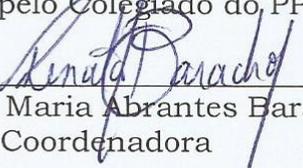


Prof. Dr. Maurício Barcellos Almeida - ECI/UFMG



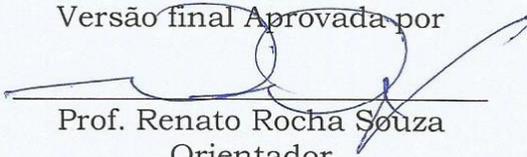
Profa. Dra. Gercina Ângela Borém de Oliveira Lima - ECI/UFMG

Aprovada pelo Colegiado do PPGCI



Profa. Renata Maria Abrantes Baracho Porto
Coordenadora

Versão final aprovada por



Prof. Renato Rocha Souza
Orientador



UFMG

Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Programa de Pós-Graduação em Ciência da Informação

ATA DA DEFESA DE DISSERTAÇÃO DE **JOICE RODRIGUES TEIXEIRA**,
matrícula: 2010654638

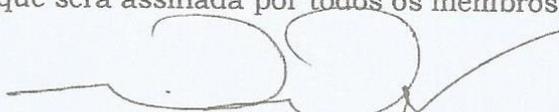
Às 10:00 horas do dia 26 de abril de 2013, reuniu-se na Escola de Ciência da Informação da UFMG a Comissão Examinadora aprovada *ad referendum* pela Coodenadora do Programa de Pós-Graduação em Ciência da Informação em 24/03/2013, para julgar, em exame final, o trabalho intitulado **Conversão de Tesouros em ontologias: um estudo exploratório**, requisito final para obtenção do Grau de MESTRE em CIÊNCIA DA INFORMAÇÃO, área de concentração: Produção, Organização e Utilização da Informação, Linha de Pesquisa: Organização e Uso da Informação. Abrindo a sessão, o Presidente da Comissão, Prof. Dr. Renato Rocha Souza, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Logo após, a Comissão se reuniu sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

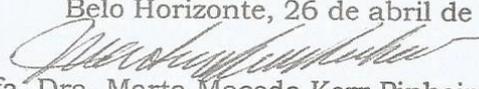
Prof. Dr. Renato Rocha Souza - Orientador	APROVADA
Profa. Dra. Marta Macedo Kerr Pinheiro	APROVADA
Prof. Dr. Maurício Barcellos Almeida	APROVADA
Profa. Dra. Gercina Ângela Borém de Oliveira Lima	APROVADA

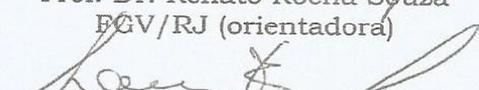
Pelas indicações, a candidata foi considerada APROVADA.

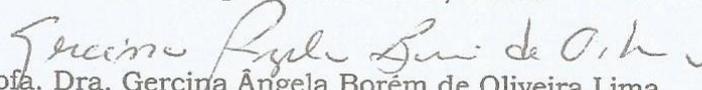
O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 26 de abril de 2013


Prof. Dr. Renato Rocha Souza
ECV/RJ (orientadora)


Profa. Dra. Marta Macedo Kerr Pinheiro
Profa. Aposentada - ECI/UFMG


Prof. Dr. Maurício Barcellos Almeida
ECI/UFMG


Profa. Dra. Gercina Ângela Borém de Oliveira Lima
ECI/UFMG


Prof(a). Renata Maria Abrantes Baracho Porto
Coordenadora do Programa de Pós-Graduação
em Ciência da Informação

Obs: Este documento não terá validade sem a assinatura e carimbo da Coordenadora.

Aos meus pais,

Wilson e Aparecida

“É na educação dos filhos que se revelam as virtudes dos pais”¹.

Por vocês cheguei até aqui, e por vocês quero seguir ainda mais longe.

¹ Henrique Maximiano Coelho Neto (1864 - 1934) escritor, político e professor brasileiro.

AGRADECIMENTOS

Agradeço a Deus, sempre presente. Nele encontrei força, coragem e sabedoria para concluir mais essa etapa da minha vida. Iluminou meu caminho e me abençoou com pessoas especiais, que seguiram comigo ao longo do mestrado.

Diversas pessoas, em momentos distintos, colaboraram de algum modo para que o projeto se concretizasse. Por isso a importância desse espaço para memorar e perpetuar os meus sinceros agradecimentos:

Aos meus pais. Tive a sorte de ter pais sábios, que priorizaram em tudo a minha educação e de minha irmã. Com muito amor, em meio a várias dificuldades, superamos e seguimos juntos. Nos últimos tempos, apoiaram e entenderam enquanto estive ausente.

A minha irmã e amiga Ilseane, por lutar bravamente por dias melhores e por torcer por mim sempre. Agradeço a ela e ao José Miguel pelo presente chamado Douglas, que trouxe muita alegria para nossa família.

Ao Leandro, companheiro fiel nessa fase. Inspira-me a cada dia a viver com objetivo, disciplina e dedicação, em tudo que fizer. Soube entender e apoiar nos momentos mais difíceis. Agradeço ainda a sua família, pelo incentivo e por saberem o valor dessa conquista.

Aos meus avós e familiares, por acompanharem e torcerem, mesmo de longe, a trajetória dos meus estudos. Mesmo que *in memoriam*, agradeço a minha vovó Maria Lopes Teixeira (1898-1996), pela motivação, pois está sempre presente nos meus pensamentos, e a prima Susan Lopes (19xx-2012), pelo apoio incondicional quando me mudei para Belo Horizonte.

Ao meu irmão de coração, Alberth Sant'ana. Exemplo de verdadeiro companheirismo, sou eternamente grata a sua amizade e contribuição na minha trajetória do mestrado.

Ao meu orientador Prof. Renato Rocha Souza, agradeço muito pela oportunidade. Por sua sabedoria e excelência nas orientações, e principalmente pela paciência em meio aos dissabores e obstáculos que enfrentei no período da pesquisa.

Agradeço a minha banca examinadora. À Profa. Gercina Lima pelo imensurável carinho desde a época da graduação, para sempre. Ao Prof. Maurício Almeida por contribuir sabiamente no decorrer de sua disciplina, na qualificação e em minha defesa. E à Profa. Marta Pinheiro, pela disponibilidade e sabedoria admirável. É com muita honra que recebo suas contribuições e participação nesse processo final.

Obrigada a todos os professores do PPGCI, pela troca de experiências e ensinamentos durante as disciplinas do curso.

Aos funcionários do PPGCI pela assistência e orientações, aos profissionais da Biblioteca Etelvina Lima e aos demais colaboradores da Escola de Ciência da Informação.

Aos amigos que conquistei na turma do mestrado, todos foram especiais. E mais que especiais foram Ariane Lemos, Lílian Marques, Luciana Carneiro, Lívia Coutinho, Tatiane Krempser, Raisal Mendes, Soraia de Andrade, Kátia Coelho, Inês Sepúlveda e as companheiras Juliana Moreira, Paula Emanuelle e Fátima Coelho, que nos últimos momentos foram inteiramente solícitas e parceiras. Obrigada pela amizade que estimo manter eternamente.

Aos amigos da turma da graduação em Biblioteconomia (2005), por torcerem e acreditarem na conclusão dessa conquista. Em especial a Renata Almeida, Poliana Moreira e Sandra Vilarino, pela amizade imbatível a qualquer distância.

A todos amigos e vizinhos que me acompanharam... de Guaçuí-ES, de BH-MG, de toda a vida... em especial a Ana Paula Dornelles, grande incentivadora, foi fonte inspiradora ao concluir seu mestrado.

Aos amigos da Kroton Educacional, pela torcida diária. Principalmente ao Luiz Filipe Trivelato, enquanto estive na instituição, meu líder no início das disciplinas, possibilitou que eu conciliasse o trabalho com os estudos. E em especial a Lúcia Zenaro, por também flexibilizar meus horários, parabéns pelo exemplo de pessoa que é. Obrigada pela oportunidade.

Enfim, muito obrigada a todos que torceram positivamente por mim!

[...] Milho de pipoca que não passa pelo fogo continua a ser milho de pipoca, para sempre.

Assim acontece com a gente. As grandes transformações acontecem quando passamos pelo fogo. Quem não passa pelo fogo fica do mesmo jeito, a vida inteira. São pessoas de uma mesmice e dureza assombrosa. Só que elas não percebem. Acham que o seu jeito de ser é o melhor jeito de ser.

Mas, de repente, vem o fogo. O fogo é quando a vida nos lança numa situação que nunca imaginamos. Dor. Pode ser fogo de fora: perder um amor, perder um filho, ficar doente, perder um emprego, ficar pobre. Pode ser fogo de dentro. Pânico, medo, ansiedade, depressão — sofrimentos cujas causas ignoramos. Há sempre o recurso aos remédios. Apagar o fogo. Sem fogo o sofrimento diminui. E com isso a possibilidade da grande transformação.

Imagino que a pobre pipoca, fechada dentro da panela, lá dentro ficando cada vez mais quente, pense que sua hora chegou: vai morrer. De dentro de sua casca dura, fechada em si mesma, ela não pode imaginar destino diferente. Não pode imaginar a transformação que está sendo preparada. A pipoca não imagina aquilo de que ela é capaz. Aí, sem aviso prévio, pelo poder do fogo, a grande transformação acontece: PUF!! — e ela aparece como outra coisa, completamente diferente, que ela mesma nunca havia sonhado. É a lagarta rastejante e feia que surge do casulo como borboleta voante [...].

Rubem Alves.²

² ALVES, Rubem. *O amor que acende a lua*. Campinas: Papyrus, 1999.

RESUMO

O estudo apresenta uma análise da convertibilidade da estrutura e relações conceituais de tesouros no processo de construção de ontologias. Teve-se como interrogativa, a existência de princípios metodológicos que norteiam o processo de conversão, e, como os mesmos podem conduzir e potencializar a transição da estrutura e relações conceituais do tesouro. Sobretudo, o objetivo foi apresentar, com base na literatura científica, as discussões e metodologias relativas ao processo de conversão de tesouros em ontologias. De forma específica buscou-se: identificar na literatura do campo da Ciência da Informação as principais características relativas ao processo de construção das relações conceituais de tesouros e ontologias, a fim de apreender quais relações presentes na estrutura do tesouro atendem a estrutura esperada numa ontologia; identificar na literatura selecionada os métodos e resultados encontrados nas experiências de conversão de tesouros em ontologias, com o intuito de compreender o aproveitamento da terminologia e relações semânticas do tesouro na ontologia; e por fim, apontar a validade e potencial da conversão com base na análise e sistematização dos dados bibliográficos levantados. A pesquisa foi desenvolvida por meio de um processo metodológico de Análise de Conteúdo referenciado por Bardin (2011), dividido em duas etapas: *Etapa 1* - Estudo de Normas de construção de Tesouros e Padrões de construção de Ontologias de Alto Nível, onde se utilizou como referencial para esse percurso metodológico, instrumentos tais como a Norma ANSI/NISO. Z39.19 – 2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, seguida do manual *Manual Ontology for the Twenty First Century: An Introduction with Recommendations* da ontologia de Alto Nível BFO (Basic Formal Ontology); *Etapa 2* - Análise de publicações e experiências relativas à conversão do Tesouro, com a análise empírica da experiência de conversão em quatro publicações na área de CI no intuito de identificar quais procedimentos metodológicos normalmente vem sendo discutidos nos estudos e aplicações práticas. Em resposta às possibilidades e limites na conversão de tesouros em ontologias, os resultados evidenciam que, embora a tipologia e estrutura dos relacionamentos mostrem-se diferentes, a natureza das relações e entidades é fundamentada de forma semelhante, e apontam que a condução do processo pode ter como princípios norteadores as indicações de construção dos relacionamentos, nos manuais e padrões de construção de tesouros e de ontologias. Notou-se ainda, que os métodos propostos estão no nível de experimentação, são pouco aprofundados e carecem de detalhamento sobre as etapas e ações para garantir e certificar a transposição dos conceitos sem se perder o compromisso ontológico esperado na construção de ontologias. Espera-se que tais resultados possam colaborar de forma significativa para a área da Ciência da Informação, por introduzir a temática nas pesquisas de âmbito nacional e cooperar com os estudos para a melhoria das práticas e métodos de construção de ontologia.

Palavras-chave: Sistemas de Organização do Conhecimento. Tesouros. Ontologias. Convertibilidade.

ABSTRACT

The study provides an analysis of the structure and conceptual relations convertibility of thesauri in the construction of ontologies. Had as interrogative, the existence of methodological principles that guide the conversion process and how they can drive and enhance the transition of the structure and conceptual relations of the thesaurus. Above all, the aim was to present, based on the scientific literature, discussions and methodologies relating to the conversion of thesauri in ontologies. Specifically sought to: identify in the literature of the field of information science key features relating to the construction of conceptual relations of thesauri and ontologies, in order to grasp what relationships present in the structure of the thesaurus meet expected structure in an ontology; identify the selected methods in the literature and results found in thesauri conversion experiences in ontologies, in order to understand the use of terminology and semantic relations of the thesaurus in the ontology; and finally, point the validity and potential of conversion on the basis of the analysis and systematization of the bibliographic data raised. The survey was developed through a process of Analysis of methodological Content referenced by Bardin (2011), divided into two stages: Step1-Study of construction standards of Thesauri and Ontologies-building Standards, where if you used as a reference for this methodological course, instruments such as the ANSINISO Standard. Z 39.19-2005, Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies, followed by manual Ontology for the Twenty First Century: An Introduction with Recommendations of high-level BFO ontology (Basic Formal Ontology); Step 2-analysis of publications and experiences concerning the conversion of the thesaurus with the empirical analysis of the conversion experience in four publications in the field of CI in order to identify which methodological procedures normally has been discussed in studies and practical applications. In response to the possibilities and limitations in the conversion of thesauri in ontologies, The results show that, although the typology and structure of relationships are different, the nature of relationships and entities is grounded in a similar way, and pointed out that the process of driving can have as guiding principles the design of the relationships, in the manuals and standards of construction of thesauri and ontologies. It was noted that the proposed methods are at the level of experimentation, are barely fleshed out and lack detail on the steps and actions to ensure and certify the implementation of the concepts without losing the ontological commitment expected in the construction of ontologies. It is hoped that these results can contribute significantly to the area of information science, by introducing the theme for national level and cooperate with studies for the improvement of the practices and methods of construction of ontology.

Keywords: Knowledge Organization Systems. Thesauri. Ontologies. Convertibility

LISTA DE FIGURAS

FIGURA 1	Tipos de Sistemas de Organização do Conhecimento.....	33
FIGURA 2	Taxonomia de entidades SNAP.....	49
FIGURA 3	Taxonomia de entidades SPAN.....	49
FIGURA 4	Modelo de re-engenharia para recursos não-ontológicos.....	53
FIGURA 5	Desenvolvimento de uma análise.....	69

LISTA DE QUADROS

QUADRO 1	Normas e Padrões para construção de Tesouros / Vocabulários Controlados.....	39
QUADRO 2	Componentes de Ontologias.....	43
QUADRO 3	Tipos de Ontologias de Alto Nível.....	46
QUADRO 4	Sinopse dos procedimentos metodológicos: etapas e ações.....	71
QUADRO 5	Sinopse da aplicação do método de Análise de Conteúdo - Etapa 1.....	75
QUADRO 6	Sinopse da aplicação do método de Análise de Conteúdo - Etapa 2.....	81
QUADRO 7	Resultados Etapa 1 – Tipologia de relacionamentos ANSI/NISO Z39:19-2005 x Manual BFO	88
QUADRO 8	Síntese comparativa das iniciativas e metodologias de conversão.....	107

LISTA DE ABREVIATURAS

ABNT	- Associação Brasileira de Normas Técnicas
ANSI	- <i>American National Standards Institute</i>
BFO	- <i>Basic Formal Ontology</i>
CAPES	- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	- Ciência da Computação
CI	- Ciência da Informação
CRG	- <i>Classification Research Group</i>
DOLCE	- <i>Descriptive Ontology for Linguistic and Cognitive Engineering</i>
IBICT	- Instituto Brasileiro de Informação em Ciência e Tecnologia
IEEE	- <i>Institute of Electric and Electronic Engineers</i>
IFOMIS	- <i>Institute for Formal Ontology and. Medical Information Science</i>
ISKO	- <i>International Society for Knowledge Organization</i>
ISO	- <i>International Organization for Standardization</i>
KOS	- <i>Knowledge Organization Systems</i>
LISA	- <i>Library and Information Science Abstracts</i>
MIT	- <i>Massachusetts Institute of Technology</i>
NBR	- Norma Brasileira
NISO	- <i>National Information Standards Organization</i>
OWL	- <i>Web Ontology Language</i>
PERI	- Base Periódicos
PMEST	- Personalidade, Matéria, Energia, Espaço, Tempo
PPGCI	- Programa de Pós-Graduação em Ciência da Informação
RI	- Recuperação da Informação
RDF	- <i>Resource Description Framework</i>
RDFS	- <i>Resource Description Framework Schema</i>
SciELO	- <i>Scientific Electronic Library Online</i>
SKOS	- <i>Simple Knowledge Organization System</i>
SOC	- Sistema de Organização do Conhecimento
SUMO	- <i>Suggested Upper Merged Ontology</i>
TGT	- Teoria Geral da Terminologia

UFMG - Universidade Federal de Minas Gerais
UFO - *Unified Foundational Ontology*
XML - *Extensible Markup Language*
WWW - *World Wide Web*
W3C - *World Wide Web Consortium*

SUMÁRIO

1	INTRODUÇÃO.....	14
1.1	Contextualização.....	15
1.2	Objetivos.....	19
1.3	Justificativa.....	19
2	SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO: abordagem teórica.....	24
2.1	Organização e Representação do Conhecimento: uma introdução..	25
2.2	Sistemas de Organização do Conhecimento: histórico e tipologia....	29
2.2.1	<i>Tesauros.....</i>	34
2.2.2	<i>Ontologias.....</i>	41
2.2.3	<i>SKOS.....</i>	50
2.3	<i>Conversão entre Sistemas de Organização do Conhecimento.....</i>	51
2.3.1	<i>Teorias Subjacentes.....</i>	59
2.4	Relações conceituais em Sistemas de Organização do Conhecimento.....	61
3	PROCEDIMENTOS METODOLÓGICOS.....	65
3.1	Caracterização da pesquisa.....	66
3.2	Condução e recorte da pesquisa.....	69
3.2.1	<i>Etapa 1 – Estudo de Normas de Construção de Tesauros e Padrões de construção de Ontologias de Alto Nível.....</i>	72
3.2.2	<i>Etapa 2 – Análise de Publicações e Experiências relativas a Conversão do Tesauro.....</i>	78

4	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS.....	83
4.1	Resultados: inferências e avaliações.....	84
4.1.1	<i>Resultados – Etapa 1.....</i>	85
4.1.2	<i>Resultados – Etapa 2.....</i>	95
4.1.2.1	<u><i>Experiência Agrovoc.....</i></u>	96
4.1.2.2	<u><i>Método para reengenharia de recursos não-ontológicos.....</i></u>	99
4.1.2.3	<u><i>Método de reengenharia de tesouros em ontologias.....</i></u>	101
4.1.2.4	<u><i>Experiência Tesouro GEM.....</i></u>	104
5	CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS.....	110
	REFERÊNCIAS.....	116-122

1 INTRODUÇÃO

1.1 Contextualização

A tecnologia da informação proporciona a disseminação do conhecimento através da internet, possibilitando uma ampla rede de comunicação eletrônica. Nas últimas décadas, a internet se consolidou como a principal fonte para busca de informações. Nesse ciberespaço, observa-se que para facilitar o processo de busca, diversos Portais e Sites da Web procuram oferecer instrumentos de interação e comunicação com os usuários.

Nesse contexto, novos desafios são apresentados com a crescente complexidade dos documentos armazenados na web. O grande volume de dados e de usuários exigem procedimentos de tratamento e recuperação cada vez mais satisfatórios, onde uma das diretrizes que norteiam o desenvolvimento desses mecanismos de pesquisa refere-se à organização da informação a ser apresentada ao usuário.

Historicamente, pôde-se constatar que, após a Segunda Guerra Mundial, com a identificação da explosão informacional, *Vannevar Bush* produziu o artigo "*As We May Think*"³, publicado em 1945 no *The Atlantic Monthly*, onde descreveu o desafio pós-guerra em tornar acessível o crescente acervo de conhecimento e salientou o papel das tecnologias da informação como meio possível de solução. A partir desse contexto, a Recuperação da Informação (RI) tem como grande desafio atender às necessidades de informação do usuário no meio digital, de forma rápida e precisa.

De acordo com Robredo (2004), no âmbito da RI, os problemas com representação de conteúdo ocorrem à medida que o volume de documentos a processar e a ordenar cresce, de forma que os usuários não mais se contentam com a organização por grandes classes e exigem informações mais detalhadas e concisas.

Nesse aspecto, alguns dos maiores desafios na recuperação da informação são a identificação e localização dos conceitos em um domínio específico de interesse, e, sobretudo a modernização de ambientes para os usuários realizarem consultas.

³ BUSH, V. *As we may thing*. *Atlantic Monthly*, v. 176, n. 1, p. 101-108, 1945.

A partir dessa discussão, observou-se a Web semântica que busca interligar significados de palavras, e tem como finalidade conseguir atribuir um significado (sentido) aos conteúdos publicados na Internet de modo que seja perceptível e processado tanto pelo homem quanto pela máquina.

A Web Semântica surgiu em 2001 - tendo como precursores *Tim Berners-Lee, James Hendler e Ora Lassila* - quando publicaram um artigo na revista *Scientific American*, intitulado: “Web Semântica: um novo formato de conteúdo para a Web que tem significado para computadores vai iniciar uma revolução de novas possibilidades”⁴. De acordo com Ramalho (2008):

O projeto Web Semântica constitui-se como uma tentativa inversa de solução que tem como objetivo desenvolver meios para que as máquinas possam servir aos humanos de maneira mais eficiente, mas para isso torna-se necessário construir instrumentos que forneçam sentido lógico e semântico aos computadores. (RAMALHO, 2008)

Nessa perspectiva as pesquisas em Web Semântica têm como principal enfoque as ontologias, em consequência do desenvolvimento de uma variedade de sistemas e arquiteturas que visam prover a integração semântica através de ontologias. Porém, devemos perceber que o desenvolvimento da Web semântica envolve não somente as novas tecnologias, como os motores de busca e ontologias, mas também as áreas de Ciência da Informação e Biblioteconomia e outras disciplinas que lidam há mais tempo com a organização do conhecimento, especificamente os Sistemas de Organização do Conhecimento (SOC).

Nesse recorte, adotou-se o acrônimo *Knowledge Organizations Systems* (KOS), traduzindo-o para Sistemas de Organização do Conhecimento (SOC), tendo em vista que, de acordo com Brascher (2009) o termo foi proposto em 1998 pelo *Networked Knowledge Organization Systems Working Group*, agrupando sistemas de classificação, tesouros, cabeçalhos de assunto, taxonomias, redes semânticas e ontologias. Esse grupo de pesquisadores em terminologia investiga questões relacionadas ao uso e a interoperabilidade dos SOCs através da Internet.

⁴ Disponível em: <http://www.scian.com/2001/0501issue/0501berners-lee.html>.

Suas tendências de pesquisas estão delineadas pela atualização das tradicionais tabelas de classificação, desenvolvimento de novas tabelas, sua interoperabilidade e, sobretudo o seu emprego em meio eletrônico (catálogos, sites e assemelhados) (GOMES, 2009). A referida autora aponta para as dificuldades desses sistemas face à multidisciplinaridade dos temas de pesquisa e às questões de classificação dos registros, seja em meio convencional, seja em meio digital.

Nota-se ainda que algumas pesquisas da área da Ciência da Computação para aprimorar a recuperação da informação têm levado os pesquisadores a desenvolver ontologias, as quais têm relação com a Biblioteconomia e Ciência da Informação, principalmente no que refere às questões semânticas. Por isso há a necessidade de maior cooperação entre as duas áreas diante da preocupação com as relações semânticas requeridas para a recuperação do conhecimento em meio eletrônico e digital.

Dessa forma, o percurso desta pesquisa perpassa pelas questões da representação e recuperação da informação, e se converge no estudo por melhores técnicas de conversão e compatibilização de vocabulários entre esses Sistemas de Organização do Conhecimento. Para isso teve-se como principal foco a conversão de tesouros em ontologias, fundamentado no estudo comparativo de seus padrões de construção e de pesquisas aplicadas a convertibilidade.

Dentro desse panorama, seguem algumas questões refletidas neste estudo:

As Ontologias são formadas por termos, definições e relações, e possibilitam que a máquina possa processar o raciocínio automatizado, através de regras e inferências (SALES, 2008). O estabelecimento de relações conceituais pode ser útil em circunstâncias como: determinação de sinônimos, atualização do sistema, inclusão de termos em uma só categoria e mapeamento de áreas de assunto. Nota-se que esta afirmação está relacionada e é muito similar ao objetivo da construção de tesouros, onde há a possibilidade de ser estendida também à construção de ontologias.

Partindo do princípio acima relatado, surge então o questionamento da potencialidade e possibilidade de aplicar a conversão de um tesouro para a construção de uma ontologia. Para isso é necessário averiguar as características e os tipos de relacionamentos existentes em cada um desses SOC. Desta forma, a oportunidade de aproveitamento de vocabulários existentes para a criação de ontologias é foco da problematização desse trabalho.

A partir dos enunciados descritos, tendo em vista que os Sistemas de Organização do Conhecimento possuem estruturas distintas no tocante aos seus conceitos e relacionamentos, e considerando a convertibilidade de um tesouro em uma ontologia, teve-se como **problema de pesquisa** investigar as seguintes questões:

- Quais princípios teóricos e metodológicos, discutidos e apresentados na literatura da área da Ciência da Informação, norteiam o processo de conversão de tesouros em ontologias?
- Quais práticas e metodologias conduzem e potencializam a conversão de estruturas e relações conceituais existentes no tesouro para a construção de ontologias?
- Qual o aproveitamento e potencialidade em se converter um tesouro em uma ontologia?

1.2 Objetivos

Assim, fez-se necessário determinar como **objetivo geral** deste estudo, apresentar com base na literatura científica, as discussões e metodologias relativas ao processo de conversão de tesouros em ontologias, visando introduzir a temática nas pesquisas de âmbito nacional e contribuir para a melhoria das práticas e métodos de construção de ontologias.

De forma específica, objetivou-se:

- identificar na literatura do campo da Ciência da Informação as principais características relativas ao processo de construção das relações conceituais de tesouros e ontologias, a fim de apreender quais elementos podem conduzir a identificação de quais relações presentes na estrutura do tesouro atendem a estrutura esperada numa ontologia.
- identificar na literatura selecionada os métodos e resultados encontrados nas experiências de conversão de tesouros e ontologias, com o intuito de compreender o aproveitamento da terminologia e relações semânticas do tesouro na ontologia.
- apontar a validade e potencial da conversão com base na análise e sistematização dos dados bibliográficos levantados, com o propósito de abstrair as possibilidades e os limites para a efetivação da referida conversão.

1.3 Justificativa

Ao longo de sua trajetória a Ciência da Informação vem estudando métodos para o tratamento automático da informação. Geralmente encontramos pesquisas centradas na Recuperação de Informação, área que envolve a aplicação de métodos computacionais no tratamento e recuperação da informação, e que buscam avaliar em que medida a Ciência da Computação contribui para o avanço da Ciência da Informação. Por isso utilizou-se do embasamento teórico-metodológico de

organização do conhecimento já fundamentado pela Ciência da Informação, principalmente pela Biblioteconomia e Documentação, visto que, essas áreas estão sendo redescobertas no ambiente virtual.

Mais especificamente, os Sistemas de Organização do Conhecimento (SOC) são utilizados na representação e configuração de informações para melhorar a recuperação e o gerenciamento de uma coleção na Web. Há um interesse significativo no uso de SOCs para organizar e buscar material na internet. Este interesse resulta em serviços de organização do conhecimento que farão essas fontes mais acessíveis a uma grande variedade de aplicações de softwares e sistemas de recuperação da informação.

O grande volume de material disponível no ambiente Web tem se tornado um problema cada vez mais crítico. Nesse ambiente virtual, como repositório do conhecimento, o conteúdo da internet necessita ser organizado para que haja recuperação da informação com eficácia. Dessa forma, a preocupação nessa área, está voltada para a padronização da terminologia usada pelas pessoas para encontrar e classificar a informação. Nesse sentido, observa-se que são utilizadas técnicas e ferramentas variadas para organização da informação e do conhecimento, inclusive tesouros e ontologias.

Face a esse cenário, a motivação desse trabalho despertou-se na possibilidade do aproveitamento da estrutura terminológica e das relações conceituais de instrumentos de vocabulário controlado para a construção de outros instrumentos. A partir dessa premissa buscou-se fundamentação na literatura científica da área, porém existe uma grande carência da temática nas pesquisas e publicações a respeito da conversão entre SOCs, principalmente na literatura nacional.

Partindo dessa inquietação e da necessidade de aprofundar-se na literatura para buscar as respostas, delimitou-se como escopo do trabalho, analisar as potencialidades da conversão de um tesouro em uma ontologia, tendo em vista os dois pressupostos a seguir:

- A presente pesquisa pode contribuir para o processo de construção de ontologias. Tendo em vista que falta consenso nas publicações e pesquisas sobre ontologias, no que se refere a um padrão metodológico consolidado para sua construção (ALMEIDA, 2005, p. 57).
- A construção de tesouros por ser padronizada e orientada por normas e padrões consolidados e estabelecidos internacionalmente, pode contribuir para as pesquisas de construção de ontologias. Mesmo sendo as ontologias consideradas como um SOC mais complexo em suas relações conceituais.

Tendo como base esses dois pressupostos, delimitou-se como propósito dessa pesquisa a investigação das possibilidades e limites de uma conversão de tesouro em uma ontologia. Dessa forma se fez necessário selecionar instrumentos que dessem condições de conduzir a apresentação do processo de conversão esperada pelo objetivo geral, que serão abordados no Capítulo 3 que relata a metodologia utilizada no estudo.

A justificativa desta pesquisa baseia-se ainda no fato de haver uma carência da temática na literatura nacional da área de Ciência da Informação. Diante do levantamento bibliográfico realizado para a definição do arcabouço teórico, foi possível identificar a ausência da temática em questão nas publicações científicas nacionais da área.

Reforça-se assim a importância de trazer essa discussão para grupos de pesquisa no Brasil, de forma a avançarmos nos estudos dos Sistemas de Organização do Conhecimento face a era tecnológica da informação indicando melhorias e novas perspectivas para o aproveitamento de seus instrumentos de controle de vocabulário já existentes.

A aplicação de métodos oriundos da Ciência da Computação contribui com a Ciência da Informação na medida em que viabiliza a operação de grandes quantidades de dados de uma forma rápida e ágil. No entanto, estas características não necessariamente resultam em processos consistentes ou satisfatórios de recuperação da informação onde a capacidade do computador em operar com

modelos formais poderia ser associada aos procedimentos intelectuais humanos, trabalhando-se com o melhor de cada um para a obtenção de resultados mais satisfatórios e adequados.

Ainda nesse contexto, nota-se que a evolução no processo de desenvolvimento de melhorias para construção de ontologias, contribui para o avanço na RI, haja vista que aposta no desenvolvimento desta nova visão que é a da *Web* semântica. Isto porque as ontologias têm sido adotadas para representar os dados, seus significados e relacionamentos, sendo usada em diversas aplicações.

Em virtude do que foi exposto, o surgimento acelerado de novas tecnologias requer dos profissionais da informação uma pesquisa contínua, lançando sobre tais tecnologias um olhar crítico a fim de avaliar a sua adequação, especificamente no tratamento da informação ou à CI como um todo. Por isso a pesquisa apresentada buscou analisar questões advindas da fusão da Ciência da Informação com a Ciência da Computação no contexto do tema aqui exposto.

Em suma, pretendeu-se aqui, através da literatura consultada na Ciência da Informação e Ciência da Computação, apresentada na próxima seção, estudar e avaliar a contribuição que o tesouro pode dispor na construção de ontologias partindo especificamente das relações conceituais, dos padrões já existentes e de pesquisas realizadas na área. Visou-se assim, contribuir para pesquisas em busca de melhores práticas para a definição da terminologia de um domínio do conhecimento.

O trabalho está organizado tendo como estrutura esta seção introdutória, onde foram apresentados a problemática, os objetivos e justificativa da pesquisa. Em seguida o Capítulo 2, que aborda os princípios teóricos e conceituais através de uma revisão de literatura, dando uma visão geral dos tesouros e ontologias na área de representação e organização da informação, e pontuando a questão das relações conceituais dentro da área macro Ciência da Informação. O capítulo 3 apresenta o percurso metodológico realizado no desenvolvimento dessa pesquisa, que foi baseado na técnica de Análise de Conteúdo e segmentado em duas etapas de acordo com os objetivos propostos:

- **Etapa1** – Estudo de Normas de Construção de Tesouros e Padrões de construção de Ontologias de Alto Nível
- **Etapa 2** - Análise de Publicações e Experiências relativas a Conversão do Tesouro

O Capítulo 4 apresentará os resultados e análises do presente trabalho, e por fim as considerações finais, novos questionamentos, perspectivas de pesquisas futuras, e a contribuição para a área da Ciência da Informação serão indicados no desenvolvimento do Capítulo 5.

2 SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO: abordagem teórica

2.1 Organização e representação do conhecimento: *uma introdução*

Considerando as preocupações relativas a esse trabalho e a proposta de analisar as concepções e características da construção de uma ontologia, a partir de uma conversão dos modelos de relações e categorias de tesauros, o objeto desta pesquisa se encontra articulado nos eixos teóricos apresentados a seguir.

Primeiramente, na articulação entre as Teorias de Representação e Recuperação da Informação, contextualizada no corpo interdisciplinar da Ciência da Informação e da Ciência da Computação, onde se considerou alguns dos elementos básicos da área dos Sistemas de Organização do Conhecimento no processo de recuperação de informação.

Em seguida apresenta-se uma abordagem dos Sistemas de Organização do Conhecimento, sua tipologia, características e tendências em pesquisas na área de Ciência da Informação. Para tanto, fez-se necessário aprofundar somente nos tesauros e ontologias, devido ao enfoque e análise pretendida nesta pesquisa.

O princípio básico que serve de pilar para este trabalho é o da organização do conhecimento ou informação, principalmente no que se refere à sua representação. Pode-se rever a importância dessa abordagem na seguinte colocação de Alvarenga (2003):

A organização da informação ou conhecimento, que como já dito compreenderia um processo de representação, destina-se prioritariamente à recuperação eficaz por parte dos usuários. Para que tal ocorra torna-se necessário que profissionais da informação desenvolvam e implementem sistemas representacionais que estabeleçam a confluência entre a organização cognitiva imposta ao conhecimento pelo seu produtor (representação primária) e a organização conceitual imposta ao documento pelo especialista da informação (representação secundária). (ALVARENGA, 2003).

Assim essa pesquisa está sob a fundamentação teórico-metodológica utilizada para a representação e organização do conhecimento em ambientes convencionais, que também possui embasamento válido para subsidiar a organização do conhecimento em ambientes virtuais. Desde a introdução, essa fundamentação se baseia principalmente nas Linguagens Documentárias, que tem por função tratar o

conhecimento dispondo-o como informação. E diante do grande volume de informações na web, há necessidade de recorrer, para além dos instrumentos clássicos (lista de cabeçalhos de assunto, tesouros e classificações). Nessa abordagem Lara (2004) reforça que:

As metodologias de construção da linguagem documentária são referidas como meio para tratar também universos não exclusivamente bibliográficos, como sites na web, conteúdos de manuais técnicos, de cd-roms de distintos assuntos, etc. Essas metodologias são essenciais à arquitetura da informação. (LARA, 2004, p.232).

Dentro desse espoco, foram consideradas as teorias que fundamentam a área de representação e recuperação da informação, tal como a *Teoria da Classificação Facetada*, a *Teoria do Conceito* e a *Teoria Geral da Terminologia*, devido à importância relatada por Campos:

Uma pesquisa na área de representação e recuperação de informação (CAMPOS, 1994), mostrou a existência de princípios comuns entre aspectos teóricos de instrumentos como esses, que têm por base a Teoria da Classificação Facetada, a Teoria do Conceito e a Teoria Geral da Terminologia, para a formação de estruturas sistemáticas, pois os conceitos de uma área de conhecimento se encontram relacionados entre si. Estes princípios são constituídos por elementos que estão na base da formação de tal estrutura, que são os conceitos, a relações entre os conceitos e a própria apresentação do sistema de conceitos. (CAMPOS, 2001).

No que se refere à *Teoria da Classificação Facetada*, também referenciada como Classificação Facetada, observamos na literatura que esse método é considerado como adequado para a construção de ontologias, porém ainda são bastante superficiais os fundamentos epistemológicos e mais estudos são essenciais para um verdadeiro progresso. Dessa forma nos apoiaremos nos fundamentos da análise em faceta como descreve Tristão (2004):

Desenvolvida por Shiyali Ramamrita Ranganathan na década de 1930, atualmente tem sido largamente discutida na academia como uma solução para a organização do conhecimento, em decorrência de suas potencialidades de acompanhar as mudanças e a evolução do conhecimento. A expressão análise em facetadas foi adotada por Ranganathan para indicar a técnica de fragmentar um assunto complexo* em seus mais diversos aspectos/partes constituintes, que são as facetadas, utilizando, para estabelecer a relação entre eles as “categorias fundamentais”, de noções abstratas, denominadas Personalidade, Matéria, Energia, Espaço, Tempo, conhecidas pela PMEST. Personalidade é a característica distinguindo o assunto; Matéria é o material físico do qual um assunto pode ser composto; Energia é uma ação que ocorre com respeito ao assunto; Espaço é o componente geográfico da localização de um assunto; Tempo é o período associado com um assunto. (TRISTÃO, 2004, p.165)

No trecho exposto por Lima (2004), a autora coloca que “uma das grandes contribuições de Ranganathan foi sua ordem de citação, em que divide todo e qualquer assunto em cinco categorias: Personalidade/Entidade (Personality/Entity), Matéria (Matter), Energia (Energy), Espaço (Space), Tempo (Time), mnemonicamente pela sigla PMEST (LIMA, 2004, p. 78).

A similaridade entre as tabelas que utilizam a classificação facetada e as ontologias está nos princípios de classificação, princípios para definição e princípios para identificação de conceitos, e para a construção dos sistemas de conceitos e termos sua construção é complexa, exigindo a participação de pesquisadores de várias áreas (GOMES, 2009). A produção de instrumentos e serviços é dependente de aporte teórico e metodológico de outras áreas, incorporando a teoria da classificação, que alguns autores consideram como disciplina autônoma, dado o número de pesquisadores, projetos de pesquisa, encontros especializados, organizações científicas e periódicos e diferentes áreas de aplicação.

Para a análise das relações conceituais dentro da metodologia deste trabalho, enfatizou-se a *Teoria do Conceito* (DAHLBERG, 1978) onde a comparação entre as características dos conceitos mostra que dois conceitos diferentes possuem uma ou duas características em comum, então há que falar de relações entre tais conceitos. Tais relações são: relações lógicas, relações hierárquicas, relações partitivas, relações de oposição e relações funcionais.

Ainda de acordo com Dahlberg (1978), notou-se que na categorização formal dos conceitos que tem importância na formação dos sistemas e na combinação dos mesmos conceitos pode ser a seguinte:

- Objetos
- Fenômenos
- Processo
- Propriedades
- Relações

Já na *Teoria Geral da Terminologia*, desenvolvida pelo engenheiro austríaco Eugene Wüster (1981), ponderou-se que o conceito ocupa uma posição central do termo, e busca uma normalização das terminologias existentes nas áreas e o controle da criação de novos termos. No que se refere à organização da informação, a terminologia contribui da seguinte forma, como proposto por Gomes et al (2010):

“Para a criação de instrumentos de controle terminológico voltados para a organização da informação em sistemas de recuperação de informação, a Teoria de Wüster tem aspectos que devem ser considerados. Ao adotar o conceito como unidade de organização da terminologia, Wüster introduz a noção de sistema de conceitos na elaboração de terminologias, devendo-se ter em conta, porém, que esta noção não implica, segundo sua Teoria, numa sistematização global, mas apenas em relação a um conjunto de termos ligados semântica, lógica ou onticamente. Não se encontra em sua Teoria qualquer orientação para uma apresentação sistemática global, ou seja, dedicada a um dado domínio, como um todo.” (GOMES et al, 2010)

O trabalho terminológico na TGT tem por base o conceito (unidade de pensamento), o termo (unidade de comunicação do conceito), a definição (forma de descrição do conceito) e as relações entre os conceitos (relações lógicas ou ontológicas) formando, assim, um sistema de conceitos.

A importância de retratar essas três teorias se dá pelo reconhecimento de seus métodos na área de Ciência da Informação e por serem base no desenvolvimento e construção dos tesouros, taxonomias, ontologias e sistemas de classificação, seja para organização ou representação de documentos em meio convencional e na web.

Apesar dos Sistemas de Organização do Conhecimento apresentarem princípios semelhantes, cada um contém suas particularidades e possuem diferentes funções para determinados contextos. Para tanto, aprofundou-se na tipologia e

características dos Sistemas de Organização do Conhecimento, examinando com mais critério e detalhe os tesouros e as ontologias como segue na próxima seção.

2.2 Sistemas de Organização do Conhecimento: histórico e tipologia

Segundo Hodge (2000), o termo *Knowledge Organizations Systems*, traduzido como Sistemas de Organização do Conhecimento, abrange todos os tipos de esquemas para organizar a informação e promover a gestão do conhecimento, incluindo sistemas de classificação e categorização, cabeçalhos de assunto, catálogos de autoridade, vocabulários controlados, tesouros, e outros instrumentos, tais como redes semânticas e ontologias.

No escopo dos SOC's surgiram algumas iniciativas para reflexões a respeito do desenvolvimento da Organização do Conhecimento. Em 1989 foi fundada a ISKO⁵ - International Society for Knowledge Organization – que é a sociedade internacional líder em discussões sobre organização do conhecimento e reúne profissionais de diversas áreas. Sua missão é promover o trabalho conceitual na organização do conhecimento em todos os tipos de formas e para todos os tipos de fins, tais como bancos de dados, bibliotecas e internet. Conta com mais de 400 membros em todo o mundo, a partir de domínios como a ciência da informação, filosofia, linguística, ciência da computação, etc. De acordo com a ISKO os objetivos dos SOC's são:

- Promover pesquisa, desenvolvimento e aplicações de Sistemas de Organização do Conhecimento que avançam as abordagens filosóficas, psicológicas e semânticas para a ordenação do conhecimento;
- Fornecer os meios de comunicação e de redes de organização do conhecimento para os seus membros;
- Funcionar como um elo entre todas as instituições e as sociedades nacionais, trabalhando com problemas relacionados com a organização conceitual e transformação do conhecimento.

⁵ Disponível em: www.isko.org

Ainda na tentativa de retratar as iniciativas de pesquisa e desenvolvimento dos SOCs, ressaltamos o Manifesto de León⁶, ocorrido em 2007, onde algumas propostas relevantes sobre o futuro da organização do conhecimento surgiram durante a 8ª Conferência na ISKO realizada na cidade de León, na Espanha. Uma das principais temáticas foi a de “gestão do conhecimento multidimensional nos sistemas de organização do conhecimento”, onde as propostas originadas nesse encontro foram tituladas como o "Manifesto de León", e podem ser resumidas nos seguintes pontos:

- Tendência para uma crescente interdisciplinaridade do conhecimento que exige o desenvolvimento de novos Sistemas de Organização do Conhecimento, com base em uma revisão dos princípios subjacentes aos tradicionais SOCs;
- Ampliar as unidades básicas do SOC para além de disciplinas específicas, os novos devem abranger o mundo real como ele é representado no conhecimento humano;
- Os novos SOCs devem permitir aos usuários a passar de uma perspectiva ou ponto de vista para outro, refletindo a natureza multidimensional do pensamento complexo.
- E que as conexões entre os fenômenos e teorias possam ser expressas e gerenciadas por meio de técnicas analítico-sintéticas já desenvolvidas na classificação facetada.

Analisaram ainda o quanto a Biblioteconomia e o campo da Ciência da Informação reagem aos problemas de adaptação ao novo panorama de interdisciplinaridade e dimensão do tradicional ao digital. E o que propõem para tentar resolver as questões de adaptação de SOCs para novas demandas, criação de alternativas de SOCs híbridos e criação de novos SOCs.

Diante dos objetivos desses grupos e manifestos nota-se que há uma grande preocupação em desenvolver a aplicabilidade dos SOCs não só em ambientes convencionais, mas principalmente para o ambiente Web, almejando-se assim melhorias na representação do conhecimento e, por conseguinte na recuperação da informação em seus diversos ambientes.

⁶ Disponível em: <http://www.iskoi.org/ilc/leon.htm>

Porém, diante dessas discussões percebe-se que, ainda não estão concretizadas as iniciativas e aplicações da convertibilidade e compatibilização entre os SOCs já existentes, sendo que essa aplicação encontraria muitas respostas e caminhos a serem seguidos rumo ao desenvolvimento e emprego dos Sistemas no ambiente Web.

Como citado por Robredo (2010), Vickery (2008) agrupa os SOCs em quatro grupos:

- *Era da pré-coordenação*: os SOCs eram estruturas estáticas e atendiam às necessidades dos sistemas manuais de organização e recuperação da informação, como índices e catálogos. Incluem-se aqui as listas de cabeçalhos de assunto e as classificações.
- *Era da pós-coordenação*: os SOCs tornam-se mais dinâmicos e possibilitam que cada um de seus elementos (termos) sejam manipulados de forma independente para representar os assuntos de cada documento. Exemplos de SOCs dessa era são vocabulários controlados (listas de termos autorizados para uso na indexação e recuperação da informação) e tesouros.
- *Era da Internet*: os SOCs que se destacam são as classificações hierárquicas que orientam o usuário na escolha do termo que melhor expressa sua questão de busca; os elos estabelecidos por meio de URL entre itens da Web e os índices das ferramentas de busca, compostos de palavras extraídas dos conteúdos dos objetos informacionais.
- *Era da Web Semântica*: os SOCs dessa era diferenciam-se dos demais por serem projetados para uso por agentes inteligentes. O principal exemplo são as ontologias.

A tipologia dos SOCs é estabelecida por Hodge (2000) da seguinte forma:

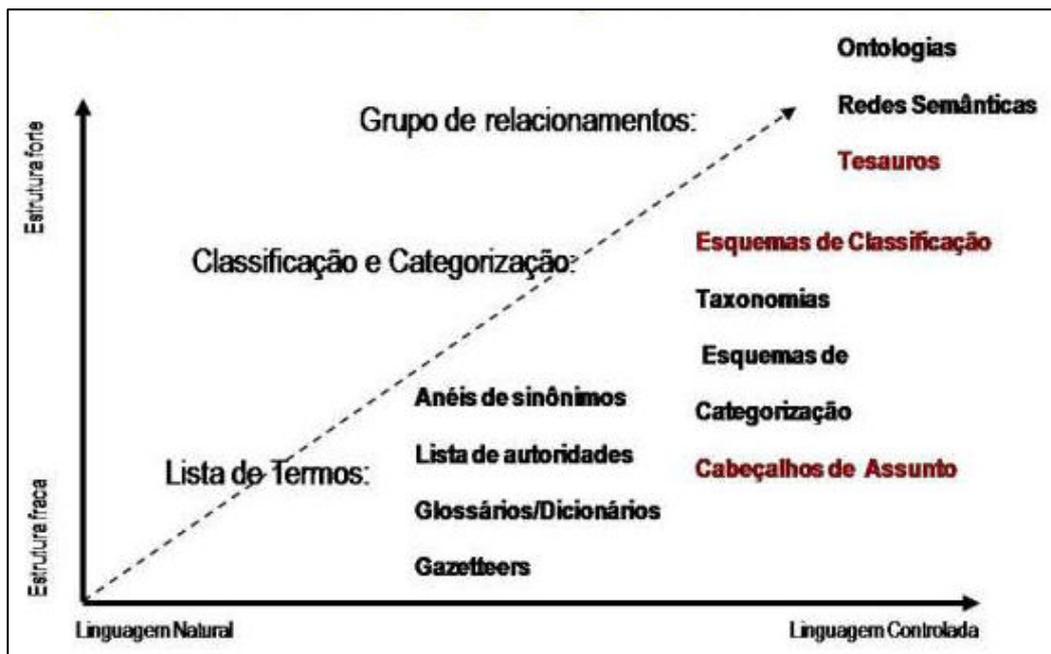
- **Listas de termos**: englobam listas de termos geralmente acompanhados de suas definições. Compreendem:
 - *Lista de autoridades*: listas de termos que controlam as variações de nomes para entidades, como nomes de países, indivíduos ou instituições.
 - *Glossários*: lista de termos com definições, geralmente de um assunto ou domínio específico.
 - *Dicionários*: lista alfabética de palavras e suas definições. Seu escopo é mais geral que o dos glossários.

- *Gazetteers*: dicionário de nomes de lugares e acidentes geográficos, tais como cidades, rios, vulcões. Quando georreferenciados, os *gazetteers* apresentam as coordenadas para a localização de lugares na superfície da Terra.
- **Classificações e categorias**: reúnem SOCs que enfatizam a criação de classes de assuntos. São eles:
 - *Cabeçalhos de assunto*: conjunto de termos controlados que representam os assuntos de uma coleção.
 - *Esquemas de classificação*, taxonomias e esquemas de categorização: esquemas utilizados para agrupar entidades em classes mais gerais. Os esquemas de categorização são geralmente usados para reunir os termos de tesouros em tópicos.
 - As *taxonomias* são usadas em modelos orientados a objeto e em sistemas de gestão do conhecimento, para indicar grupos de objetos baseados em características particulares.
- **Listas de relacionamentos**: agrupam SOCs que privilegiam a conexão entre termos e conceitos. Englobam:
 - *Tesouros*: conjunto de termos que representam conceitos e as relações de equivalência, hierárquicas e associativas que se estabelecem entre eles.
 - *Redes semânticas*: estrutura de conceitos e termos em forma de rede ou teia, os conceitos são nós e os relacionamentos expandem-se a partir dos nós.
 - *Ontologia*: estrutura de conceitos e representação dos relacionamentos complexos entre eles, incluindo regras de inferência e axiomas.

Ressalta-se que os diversos tipos de Sistemas de Organização do Conhecimento, diferenciam-se em alguns aspectos, tal como o grau de controle instaurado (da linguagem natural para linguagem controlada) e o nível de sua estrutura semântica (de pouca a fortemente estruturada), correspondentes às suas principais funções. De acordo com a FIG.1 - Tipos de Sistemas de Organização do Conhecimento,

pode-se perceber a diferenciação e características dos relacionamentos. Quanto maior o controle na linguagem mais forte é a estrutura do Sistema.

FIGURA 1 – Tipos de Sistemas de Organização do Conhecimento



Fonte: ZENG, M. L.; SALABA, A. (2005).

Porém, para Robredo, J.; Brascher, M. (2010) há uma visão que, para a Ciência da Informação, os SOCs são representações de domínios do conhecimento que delimitam o significado de termos no contexto desses domínios, estabelecem relações conceituais que auxiliam a posicionar um conceito no sistema conceitual e são utilizadas como instrumentos de organização e recuperação da informação.

Ainda nesse entendimento (ROBREDO, J.; BRÄSCHER, M., 2010) os dicionários, glossários e gazetteers delimitam conceitos e apresentam alguns tipos de relações conceituais, como as relações de equivalência e associativas. Contudo, não são instrumentos elaborados com objetivo de organização e recuperação de informações.

Sobretudo, na literatura há um consenso de que Tesouros e Ontologias fazem parte da lista de Sistemas de Organização do Conhecimento. Para isso enfatizou-se alguns aspectos das estruturas e relacionamentos desses dois instrumentos que fazem parte do recorte desta pesquisa como seguem nos próximos tópicos.

2.2.1 Tesouros

O termo Tesouro originou-se do latim-*thesauru*, e do grego-*thesaurós*, e tem significado de tesouro, repositório. A partir de 1500 começou a ser empregado fazendo referência a acervos organizados, mas foi em 1852 com a publicação do *Tesaurus of English and Phrases de Roget* que se tornou o marco significativo na concepção desse instrumento, tendo em vista que as palavras e frases, no também chamado Roget's Thesaurus, foram arranjadas não de acordo com sua ortografia, mas sim com o seu significado (DODEBEI, 2002).

Na década de 1940 o termo tesouro passou a ser utilizado na área da Ciência da Informação, principalmente no processo de recuperação da informação, interligando seus conceitos e relações. Com a crescente produção científica, por volta dos anos 50, surgiu a necessidade de um instrumento de representação do conteúdo mais complexo do que as listas de cabeçalho de assunto para proporcionar a recuperação dos documentos.

Nessa perspectiva de evolução dos tesouros, surgem duas vertentes, a norte-americana e a européia. Tesouros construídos na América do Norte foram originados a partir das listas de cabeçalhos de assunto. Durante a década de 1960 vários tesouros foram aperfeiçoados e órgãos oficiais fizeram acordo para reunir esses instrumentos. Para tanto estabeleceram um manual com princípios comuns de construção e uso. Publicado no órgão oficial *Council for Science and Technology* dos Estados Unidos, o manual passou a ser fonte para a construção de tesouros, e serviu de base para a ANSI - *American National Standardization Institute* (CAMPOS, GOMES, 2006).

Na literatura verificou-se que no que se refere à tendência europeia, nota-se o emprego de princípios de classificação, e que construíram vários tesouros e contribuíram para o desenvolvimento teórico da área. Principalmente com a criação do *Classification Research Group* – CRG com o objetivo de estudar e aperfeiçoar as teorias da classificação de Ranganathan. Fizeram parte desse Grupo estudiosos como Vickery, Foskett, Aitchison, Farradane, Austin, dentre outros, que são referências na área de representação e recuperação da informação. A publicação do *Thesaurifacet* foi um grande marco nos estudos dessa vertente, pela integração da tabela de classificação com o tesouro (CAMPOS, GOMES, 2006).

Devido à evolução dos tesouros, surgiram normas e padrões para sua construção e manutenção. Na década de 1970 foi criado o primeiro padrão internacional, a ISO 2788/74, posteriormente revisada em 1986. Na mesma época a NISO - National Information Standards Organization publica a norma ANSI/NISO Z39.19 - *Guidelines for the construction, format, and management of monolingual thesauri*, atualmente está em sua quarta edição.

Especificamente no Brasil, duas publicações se destacaram, a tradução do documento da Unesco, pelo IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia o nomeando de Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngues. E em 1997 a ABNT – Associação Brasileira de Normas Técnicas publicou duas Normas, a NBR 13289/ 97 Terminologia. Princípios. Métodos de elaboração e apresentação de normas de terminologia e NBR 13790/97. Terminologia. Princípios. Métodos de harmonização de conceitos e de termos (BOCCATO, 2011).

Currás (1995) conceitua tesouro como "uma linguagem especializada, normalizada, pós-coordenada, usada com fins documentários, onde os elementos linguísticos que a compõem – termos, simples ou compostos – encontram-se relacionados entre si sintática e semanticamente”.

Considerando que os Tesouros se constituem em uma lista de termos autorizados, pertencentes a um domínio do conhecimento, relacionados semântica e logicamente e seu uso se dá tanto no trabalho de indexação, quanto na elaboração de buscas

pelos usuários, baseou-se na representação clássica dos tesouros, ou seja, naquelas baseadas e referenciadas por normas e padrões consolidados internacionalmente.

Para tal utilizou-se as normas ISO e ANSI / NISO que indicam os tesouros como um conjunto de termos relacionados por diferentes relações hierárquicas, associativas, e de equivalência. Essa abordagem ainda concebe o léxico como um conjunto de conceitos relacionados por relações semânticas, onde os conceitos estão vinculados a um acordo, na forma de etiquetas lexicais, que podem ou não ter relações lexicais estabelecidas entre eles.

De acordo com a norma ANSI-NISO Z39.19-2005 são indicadas 5 funções para os tesouros:

- *Tradução* - Para prover um modo para traduzir a linguagem natural dos autores, indexadores e usuários para um vocabulário controlado usado para indexação e recuperação;
- *Consistência* - Para promover consistência na designação de termos de indexação;
- *Indicação de Relacionamentos* - Para indicar relacionamentos semânticos entre termos;
- *Recuperação* - Para servir como uma ajuda na busca e recuperação de documentos;
- *Nome e navegação* - Provê hierarquias claras e consistentes em um sistema de navegação para ajudar usuários a localizar objetos de conteúdo desejados.

Sua terceira edição, divulgada em 2003, a instituiu como ANSI/NISO Z39.19-2003 – Guidelines for the Construction, Format, and Management of Monolingual Thesauri, e teve o objetivo de apresentar as regras de construção de tesouros monolíngues impressos e eletrônicos, incluindo a relação entre os termos, com exemplos, sugestões de escolha e redação de descritores, entre outros aspectos (BOCCATO, 2011, p. 170).

A norma ANSI/NISO Z39.19 foi elaborada pela National Information Standards Organization (NISO), pertencente ao American National Standards Institute (ANSI), e teve sua primeira edição publicada em 1974, revisada no ano de 1980, sob o título *Thesaurus Structure, Construction and Use*. A segunda edição foi disponibilizada em 1993 - ANSI/NISO Z39.19-1993: *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*, tendo sido baseada nas concepções da norma americana ISO 2788-1986. (BOCCATO, 2011, p. 169).

Boccatto (2011) ainda relata que:

Com a expansão da tecnologia e sua aplicabilidade no acesso e na recuperação da informação viabilizou rapidamente a necessidade da revisão da terceira edição da norma ANSI/NISO Z39.19-2003. A quarta edição, publicada em 2005, nomeada por ANSI/NISO Z39:10-2005 – *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* tem como proposta o estabelecimento de diretrizes e convenções quanto à construção, o formato de apresentação e o gerenciamento de vocabulários controlados, impressos e eletrônicos, focalizando, principalmente, outros tipos de sistemas de organização do conhecimento, além dos tesouros, tais como as listas, os anéis de sinônimos, as taxonomias e as redes semânticas (BOCCATO, 2011).

No que se refere ao seu uso em meio eletrônico, Murakami (2005) enfatiza que a relação entre Tesouros Documentários e a Web é bilateral. No princípio dos anos 90, a Web foi utilizada para a distribuição de Tesouros e posteriormente, os tesouros começaram a ser utilizados para a organização da própria Web, devido à migração de Sistemas de Recuperação da Informação para esse ambiente e a crescente demanda por organização da informação desse ambiente. Para que pudessem ser utilizados na Web, os Tesouros precisaram passar por um processo de adaptação. As atuais pesquisas desse tema estão voltadas para a criação de um padrão estabelecido e uma melhor compreensão das funções dos tesouros na Web, ainda são barreiras para os desenvolvedores de sistemas e softwares.

Buscam-se então, alternativas para a representação de tesouros na Web. O SKOS⁷, que possui como principal elemento o conceito, inclui a proposta principal de organizá-los, onde os descritores e não-descritores são substituídos pela associação com os conceitos preferenciais e alternativos e assim podem ser definidas relações hierárquicas ou associativas entre os conceitos. Esse padrão também prevê

⁷ Disponível em: <http://www.w3.org/TR/skos-reference/skos.html>

relações de correspondência entre os conceitos pertencentes a outros diferentes esquemas conceituais e será abordado em outro tópico.

Ainda nessa vertente, aponta-se o Projeto ISO 25964⁸ que mais especificamente estuda os tesouros e sua interoperabilidade com outros vocabulários. Trata-se de um projeto de desenvolvimento da Norma Internacional ISO Technical Committee 46 (Informação e documentação) Subcomissão 9 (identificação e descrição). O Grupo de Trabalho designado (WG8) procura rever, fundir, e estender duas normas internacionais existentes: ISO 2788 Diretrizes para a criação e desenvolvimento de tesouros monolíngues e ISO 5964 Diretrizes para a criação e desenvolvimento de thesauri multilíngues.

De acordo com os objetivos do Projeto mencionado, as duas normas intimamente vinculadas serão atualizadas para refletir as necessidades do século 21, tendo em conta a demanda de interoperabilidade de uma sociedade em rede. Tudo será mantido e renovado, e os seguintes temas serão adicionados: Orientação sobre as funções eletrônicas; Especificação funcional de software; Interoperabilidade entre os tipos de vocabulários; Modelagem de dados e formatos de intercâmbio de dados.

Cabe ainda citar a BS 8723⁹ - Vocabulários estruturados de recuperação de informação -, uma norma em cinco partes, como seguem:

- Parte 1: Definições, símbolos e abreviaturas;
- Parte 2: Tesouros;
- Parte 3: Outros vocabulários de tesouros;
- Parte 4: Interoperabilidade entre vocabulários;
- Parte 5: Formatos e protocolos para a interoperabilidade

Nota-se que a norma ISO 25964 recorre a BS 8723, visto que abrange todos os aspectos léxicos, monolíngues e multilíngues, incluindo um modelo de dados e formatos e protocolos para troca de dados. E ainda pode possibilitar a

⁸ Disponível em: <http://www.niso.org/workrooms/iso25964>

⁹ Disponível em: <http://schemas.bs8723.org/>

interoperabilidade entre vocabulários, sistemas de classificação, taxonomias e ontologias.

A fim de compilar as normas e padrões e, sobretudo retratar o percurso e evolução das Normas Internacionais de construção de Tesouros e Vocabulários Controlados, apresenta-se o Quadro 1 - Normas e Padrões para construção de Tesouros/Vocabulários Controlados:

Quadro 1 - Normas e Padrões para construção de Tesouros / Vocabulários Controlados

NORMAS E PADRÕES PARA CONSTRUÇÃO DE TESAuros / VOCABULÁRIOS CONTROLADOS		
Mantenedor/ Número	Título	Especificidades
ISO 25964-1:2011	Norma internacional de tesouros e interoperabilidade com outros vocabulários	Estuda os tesouros e sua interoperabilidade com outros vocabulários. Recorre a BS 8723, visto que abrange todos os aspectos léxicos, monolíngues e multilíngues, incluindo um modelo de dados e formatos e protocolos para troca de dados. Substituiu as normas ISO 2788 e ISO 5964.
ISO 2788	Padrão internacional para construção de tesouros monolíngues	Objetiva assegurar uma prática consistente num serviço de indexação ou entre diferentes serviços.
ISO 5964	Padrão internacional para construção de tesouros multilíngues	Objetiva assegurar uma prática consistente num serviço de indexação ou entre diferentes serviços.
BS 8723	Norma Britânica para construção de Vocabulários estruturados de recuperação de informação.	Parte 1: Definições, símbolos e abreviaturas; Parte 2: Tesouros; Parte 3: Outros vocabulários de tesouros; Parte 4: Interoperabilidade entre vocabulários; Parte 5: Formatos e protocolos para a interoperabilidade.
ISO 704	Princípios e métodos para o trabalho de terminologia	Descreve as conexões entre objetos, conceitos, e suas representações terminológicas. Ele também estabelece os princípios gerais que regem a formação de designações e da formulação de definições.
ANSI / NISO Z39.19-2005	Diretrizes para a Construção, Formatação, e Gestão de vocabulários controlados monolíngues	Apresenta diretrizes e convenções para o conteúdo, exibição, construção, testes, manutenção e gestão de vocabulários controlados monolíngues. Concentra-se em vocabulários controlados que são utilizados para a representação de objetos de conteúdo em sistemas de organização do conhecimento, incluindo tesouros.

Fonte: Elaborado pela autora.

É importante mencionar que o foco para este estudo, com relação à estrutura de relações dos tesouros, será a Norma ANSI/NISO Z39:10-2005 – *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. A motivação pela escolha se baseia na visão de que é um padrão aceito e referenciado internacionalmente. Além disso, a ANSI/NISO Z39:10-2005 apresenta abordagem interdisciplinar das teorias originadas da Ciência da Informação e da Teoria da Classificação Facetada, Teoria do Conceito e Teoria da Terminologia para a construção e manutenção de vocabulários controlados.

Entende-se que existe mais de uma norma para construção de tesouros, porém para a metodologia desta pesquisa foi selecionada a Norma ANSI/NISO Z39.19 – 2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* devido a sua alta relevância na literatura da área, pois estabelece diretrizes e critérios quanto a construção, formato, gerenciamento de vocabulários controlados, tanto impressos quanto eletrônicos, focalizando principalmente os tesouros (BOCCATO, 2011).

A ANSI/NISO Z39:19-2005 é composta por onze seções e seis apêndices, assim distribuídas:

1. Introdução
 2. Âmbito de aplicação
 3. Normas de referência
 4. Definições, Abreviaturas e Siglas
 5. Vocabulários Controlados – objetivo, conceitos, princípios e estrutura
 6. Forma, Escopo e Prazo de Escolha
 7. Termos Compostos
 8. Relacionamentos
 9. Apresentação de vocabulários controlados
 10. Interoperabilidade
 11. Construção, Testes, Manutenção e Sistemas de Gestão
- Apêndices A-F

Para tanto, selecionou-se o Capítulo 8 – *Relacionamentos*, tendo em vista que o mesmo trata-se das recomendações para a construção dos relacionamentos em um tesauro, sendo subdividido em três categorias de relações:

- *Equivalência;*
- *Hierarquia;*
- *Associação.*

Tal decisão deu base para a Etapa 1 no desenvolvimento da metodologia dessa dissertação e será apresentada no capítulo de abordagem dos procedimentos metodológicos.

O segundo instrumento de representação do conhecimento a ser enfatizado são as ontologias apresentado na seção a seguir, retratando suas características e conceitos.

2.2.2 Ontologias

No âmbito da elaboração de ontologias, uma problemática que tem se colocado é a ausência de um padrão teórico-metodológico para sua construção. Almeida (2005) salienta que embora haja uma produção significativa de discussões em torno das metodologias para construção de ontologias, essas ainda encontram-se sem consolidação. O referido autor aponta que:

Propostas para avaliação de ontologias são encontradas na literatura, mas parecem existir poucas metodologias formais. A construção de ontologias é ainda mais artesanal do que científica (...) e não existem propostas unificadas, sendo que grupos diferentes utilizam diferentes abordagens (...). Essa diversidade pode ser um fator que dificulta a formulação de metodologias de avaliação formais. (ALMEIDA, 2005, p. 57)

Há uma recente abordagem onde se aplica a utilização de ontologias na organização de conteúdos. Elas são criadas por especialistas e definem as regras que regulam a combinação entre termos e relações em um domínio do conhecimento. Os usuários formulam consultas usando conceitos definidos pela

ontologia, e o que se busca, em última instância, são melhorias nos processos de recuperação da informação.

Na abordagem dessa pesquisa considerou-se ontologia como um tipo de SOC, baseando-se na seguinte concepção de Sales (2008):

As Ontologias são formadas por termos, definições e relações. Devido a estes elementos, a literatura vem definindo ontologia como uma linguagem documentária. Entretanto, apesar de possuir elementos comuns, as ontologias são mais que linguagens documentárias: elas possuem funcionalidades que permitem que a máquina possa processar o raciocínio automatizado, através de regras e inferências. (SALES, 2008)

Na literatura encontram-se várias concepções teóricas a respeito das ontologias. Portanto, optou-se por adotar os conceitos mais referenciados e consolidados, assim como o proposto por Guarino (1998) onde relata que:

[...] uma ontologia se refere a um artefato de engenharia (de software), que é constituído por um vocabulário específico utilizado para descrever certa realidade, mais um conjunto de suposições explícitas a respeito do significado pretendido para as palavras do vocabulário. Esse conjunto de suposições tem em geral a forma da teoria da lógica de primeira ordem, onde palavras do vocabulário aparecem com nomes de predicados unários ou binários, respectivamente chamados conceitos e relações. No caso mais simples, uma ontologia descreve uma hierarquia de conceitos relacionados por relações de classificação; em casos mais sofisticados, axiomas são adicionados à estrutura de forma a expressar outras relações entre conceitos, e para restringir a interpretação pretendida para tais conceitos. (GUARINO, 1998, p.2)

Para SOWA (1999) a ontologia é um “catálogo de tipos de coisas” em que se supõe existir um domínio, na perspectiva de uma pessoa que usa determinada linguagem. O autor enfatiza ainda que é uma teoria que diz respeito a tipos de entidades e, especificamente, a tipos de entidades abstratas que são aceitas em um sistema com uma linguagem.

Na perspectiva de GRUBER (1996) uma ontologia é uma especificação explícita de uma conceitualização. Onde aborda que em uma ontologia, definições associam nomes de entidades no universo do discurso. Exemplificando com *classes-relações-funções*, etc. com textos que descrevem o que os nomes significam e os axiomas formais que restringem a interpretação e o uso desses termos.

Para Noy e McGuinness (2001), a construção de uma ontologia inclui a definição de classes, a organização das classes em uma taxonomia, definição de atributos e a descrição dos valores permitidos para os atributos e o preenchimento dos valores dos atributos para as instâncias. Dessa forma nos basearemos no que as referidas autoras apresentam como os componentes de uma ontologia apresentados no **Quadro 2 – Componentes de Ontologias**, a seguir:

Quadro 2 – Componentes de Ontologias

Componentes de Ontologias	Características
<i>Classes ou conceitos</i>	Descrevem conceitos do domínio, onde as classes são organizadas em uma taxonomia
<i>Relações</i>	Representam as formas de associação entre conceitos de um domínio
<i>Atributos</i>	Descrevem as características que compõem os conceitos e as instâncias
<i>Instâncias</i>	São usadas para representar os elementos de uma ontologia, onde uma instância é um conceito que pertence a uma classe e possui determinadas propriedades
<i>Axiomas</i>	Usam linguagens de lógica para descrever os construtos na ontologia
<i>Regras</i>	Geralmente são usadas para inferir conhecimento na ontologia

Fonte: elaborado pela autora de acordo com Noy e McGuinness (2001).

Com relação a tipologia de ontologias, estas se denominam de acordo com suas características e funções. Guarino (1998) apresenta uma visão ampla de quatro tipos de ontologia, que são:

- **Ontologias de alto nível ou genéricas** – descrevem conceitos de forma bem geral, tais como, espaço, tempo, material, evento, etc., os quais são independentes de um problema ou domínio particular;
- **Ontologias de domínio** – expressam conceituações de domínios particulares, descrevendo o vocabulário relacionado a um domínio genérico, como Medicina;
- **Ontologias de tarefa** – expressam conceituações sobre a solução de problemas, não dependem do domínio, descrevem o vocabulário relacionado a uma atividade ou tarefa genérica, como venda de produtos;
- **Ontologias de aplicação** – descrevem conceitos dependendo de um domínio e de uma tarefa particular. Os conceitos, geralmente, correspondem aos papéis desempenhados pelas entidades do domínio quando realizam alguma atividade.

Para o estudo proposto nesta pesquisa, selecionou-se o tipo *Ontologia de Alto Nível*, também intitulada como *Ontologia de Topo* ou ainda *Ontologia de Fundamentação*. De acordo com Almeida e Bax (2003) não há um consenso quanto aos tipos de ontologias na literatura, porém, as características das tipologias apresentam semelhanças entre as funções e componentes básicos comuns.

Portanto, adotou-se o termo *Ontologia de Alto nível* e utilizou-se essa categoria como foco para a condução dos procedimentos metodológicos descritos no Capítulo seguinte. A decisão é embasada pelo fato de Ontologias de Alto Nível descreverem conceitos genéricos, independentes de um domínio particular, ou seja, por serem passíveis de utilização em domínios diversos.

Utilizou-se como base para os procedimentos metodológicos a BFO – *Basic Formal Ontology*¹⁰, como será relatado no Capítulo 3. Porém cabe citar que ainda existem outras ontologias de alto nível como DOLCE¹¹ - *Descriptive Ontology for Linguistics*

¹⁰ Disponível em: <http://www.ifomis.org/bfo>

¹¹ Disponível em: <http://www.loa.istc.cnr.it/DOLCE.html>

and Cognitive Engineering, UFO¹² - *Unified Foundational Ontology*, , SUMO¹³ - *Suggested Upper Merged Ontology*, GFO¹⁴ - *General Formal Ontology*, dentre outras. Embora não sejam abordadas com profundidade neste estudo, apresenta-se no **Quadro 3 – Tipos de Ontologias de Alto Nível** uma visão geral das principais ontologias de alto nível com suas características e idealizadores.

¹² Disponível em: <http://www.inf.ufes.br/~gguizzardi/TAO-CR.pdf>

¹³ Disponível em: <http://www.ontologyportal.org/>

¹⁴ Disponível em: <http://www.onto-med.de/ontologies/gfo/>

Quadro 3 – Tipos de Ontologias de Alto Nível

ONTOLOGIA DE ALTO NÍVEL	BFO	DOLCE	SUMO	GFO	UFO
	Basic Formal Ontology	Descriptive Ontology for Linguistics and Cognitive Engineering	Suggested Upper Merged Ontology	General Formal Ontology	Unified Foundational Ontology
MANTENEDOR	Barry Smith / Grenon Pierre	Nicola Guarino	IEEE ¹⁵ / Adam Pease	H. Herre / B. Heller	Giancarlo Guizzardi / Gerd Wagner
DESCRIÇÃO	Consiste numa série de sub-ontologias (mais adequadamente concebido como uma série de perspectivas sobre a realidade) SNAP e SPAN, que servem como base para uma série de sub-ontologias em diferentes níveis de granularidade.	Biblioteca de ontologias fundacionais. Trata-se de uma ontologia descritiva de Engenharia linguística e cognitiva.	Lançada em dezembro de 2000, destina-se como uma ontologia fundacional para uma variedade de sistemas de computadores de processamento de informações.	Ontologia para modelagem conceitual. Inclui elaborações de categorias como objetos, processos, tempo e espaço, propriedades, relações, papéis, funções, fatos e situações.	Possui conceitos relativos a eventos, objetos e agentes e, portanto, pode ser utilizada no contexto da modelagem de processos de negócio para prover semântica aos modelos que capturam o conhecimento dinâmico sobre processos.
	O projeto BFO foi iniciado em 2002 e nasce de uma orientação filosófica que coincide com a de DOLCE e SUMO	É primeiro módulo de uma Biblioteca Ontologias Fundacional a ser desenvolvido no âmbito do projeto WonderWeb ¹⁶	O objetivo do IEEE é o desenvolvimento de uma ontologia padrão que irá promover interoperabilidade de dados, pesquisa e recuperação de informações e processamento de linguagem natural.	O projeto GFO começou em 1999 no contexto do projeto GOL (General Ontological Language).	UFO é dividida em três fragmentos denominados UFO-A (Ontology of Endurants), UFO-B (Ontology of Perdurants) e UFO-C (Ontology of Social and Intentional Entities).

Fonte: Elaborado pela autora

¹⁵ Disponível em: Institute of Electric and Electronic Engineers

¹⁶ Disponível em: <http://wonderweb.semanticweb.org/>

Para tanto, faz-se necessário destacar, que a BFO surgiu de uma orientação filosófica concordante com a de DOLCE e SUMO - Suggested Upper Merged Ontology¹⁷. Teve seu projeto iniciado em 2002 e tem como base a teoria desenvolvida por Barry Smith e Grenon Pierre, onde disponibilizaram um manual com recomendações para a estrutura de uma ontologia. Trata-se do *Manual Ontology for the Twenty First Century: An Introduction with Recommendations*, recomendado e publicado pelo IFOMIS - *Institute for Formal Ontology and Medical Information Science*¹⁸.

O Instituto de Ontologia Formal e Ciência da Informação Médica fundado em abril de 2002, compreende um grupo de pesquisa interdisciplinar, com membros da área de Ciência da Computação, Filosofia, Ciência da Informação, Lógica, Medicina e Informática Médica. Possui ênfase em pesquisas teóricas fundamentadas na ontologia formal e aplicada. Tem por objetivo desenvolver uma ontologia formal que será aplicada e testada no domínio da ciência da informação médica e biomédica.

Para Grenon e Smith (2004) a BFO é uma teoria das estruturas básicas da realidade. Os autores enfatizam que a BFO é uma ontologia formal e sua construção é, assim, motivada em por um lado, pelo desejo de ser fiel à realidade, e por outro lado, ter de aceitar uma multiplicidade de perspectivas sobre a realidade. Ainda descrevem que a BFO é projetada para servir como um molde reutilizável, que pode, com algumas modificações, ser usada na construção de ontologias materiais para qualquer domínio de entidades.

De acordo com Grenon e Smith (2004), a BFO adota uma visão da realidade e contém entidades caracterizadas como: *Continuantes* (*dependentes* e *independentes*) e *Ocorrentes*. Eles consideram as *Continuantes* como as que persistem através do tempo (objetos, qualidades e funções), e as *Ocorrentes*, são eventos ou acontecimentos onde acontecem as *Continuantes*.

Para Grenon e Smith (2004) as entidades *Continuantes* podem ser exemplificadas como, o planeta Terra, um pedaço de rocha, etc. No entanto, a BFO subscreve uma

¹⁷ Disponível em: <http://www.ontologyportal.org/>

¹⁸ Disponível em: <http://www.ifomis.org/>

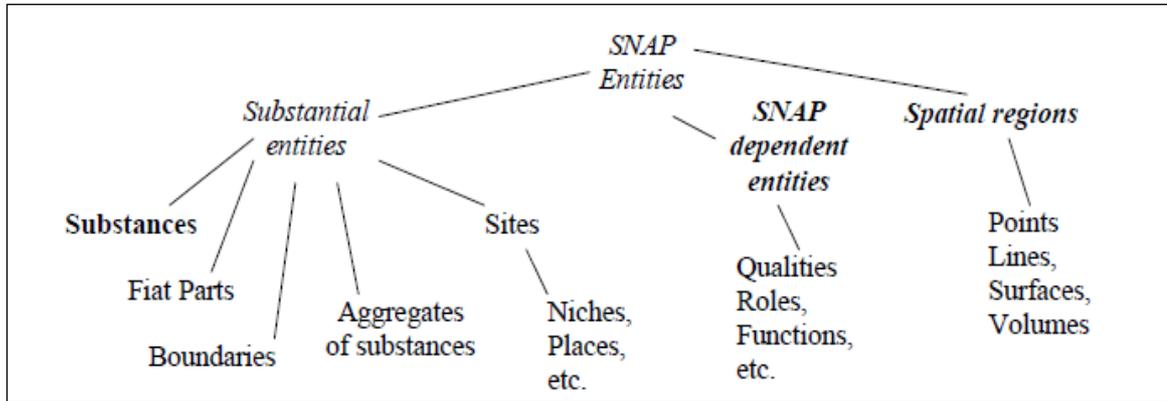
vista segundo o qual o mundo contém Ocorrentes, mais familiarmente conhecidos como processos, eventos, atividades, mudanças. Como Ocorrentes os autores exemplificam como: sorriso, caminhar, o pouso de uma aeronave, a passagem de uma tempestade sobre uma floresta, o apodrecimento das folhas caídas, etc. Mencionam ainda que elas ocorrem no tempo (GRENON, P.; SMITH, B., 2004).

Torna-se um desafio então, construir uma estrutura unificada dentro da qual se pode atender a ambos os modos. Esta estrutura deve manter os dois correspondentes grupos de entidades claramente separados, já que nenhum único inventário pode contemplar ambos. Ao mesmo tempo, no entanto, tem-se de encontrar uma forma de trazê-los juntos, pois Continuantes estão eles próprios sujeitos a constantes mudanças e Ocorrentes dependem de objetos Continuantes como seus portadores (GRENON, P.; SMITH, B., 2004).

Ainda com base nesses autores, o que se refere a diferença no modo Continuantes e Ocorrentes corresponde a uma oposição entre duas diferentes formas de existir no tempo. Assim, podem-se distinguir dois tipos principais de ontologias, SNAP e SPAN, uma para Continuantes, e a outra para Ocorrentes. As relações entre Continuantes e Ocorrentes transcendem a divisão SNAP-SPAN (GRENON, P.; SMITH, B., 2004). Para melhor entendimento os autores relatam que a realidade pode ser dividida ontologicamente de muitas maneiras diferentes e exemplificam como a citação de uma rocha, que pode ser entendida como um objeto, ou como um grupo estruturado de moléculas.

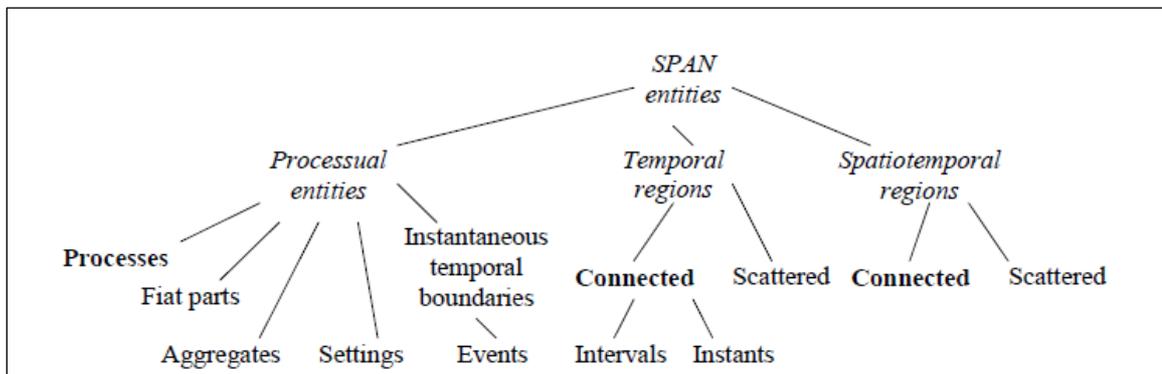
Apresenta-se abaixo as taxonomias de entidades SNAP na FIG. 2 e SPAN na FIG. 3:

FIGURA 2 – Taxonomia de entidades SNAP



Fonte: (GRENON, P.; SMITH, B., 2004)

FIGURA 3 – Taxonomia de entidades SPAN



Fonte: (GRENON, P.; SMITH, B., 2004)

De acordo com Grenon e Smith (2004) nas FIG. 3 e 4, os itens em *itálico* representam categorias que são cumulativas e em **negrito** são as categorias base das entidades.

2.2.3 SKOS

A World Wide Web Consortium (W3C), fundada por Tim Berners-Lee, no MIT¹⁹, é a principal organização internacional de padrões e normas para a World Wide Web (WWW). Dentro dessa organização surgiu o *Simple Knowledge Organization System* – SKOS – traduzido aqui como Sistema de Organização do Conhecimento Simples, sendo um conjunto de linguagens formais criados para a representação de tesouros, sistemas de classificação, taxonomias, dentre outros tipos de vocabulário controlado.

O SKOS – *Simple Knowledge Organization System* consiste em um conjunto de linguagem formais, criado para a representação de tesouros, esquemas de classificação, taxonomias, ou qualquer outro tipo de vocabulário controlado. SKOS são construídas sobre RDF²⁰ e RDFS²¹, e o principal objetivo é permitir a publicação de vocabulários estruturados para a Web Semântica. SKOS atualmente são desenvolvidos dentro do consórcio W3C (W3C, 2012).

Ainda de acordo com a W3C (2012), o modelo de dados SKOS, fornece um padrão, ou seja, um caminho de migração de baixo custo para transportar Sistemas de Organização do Conhecimento existentes para os moldes da Web Semântica. O SKOS também fornece uma linguagem leve e intuitiva para o desenvolvimento e compartilhamento de SOCs e pode ser usado sozinho, ou em combinação com linguagens formais de representação do conhecimento, tais como a linguagem Ontology Web (OWL) (W3C, 2012).

Ressalta-se a importância de mencionar o SKOS nesse estudo, porém o mesmo não será utilizado como base para os procedimentos metodológicos. Isto porque o modelo de dados relaciona-se à temática, contudo sua ênfase se dá na transposição de SOCs para o meio eletrônico, e não propriamente em conversão de um SOC para outro, portanto foge do escopo deste trabalho aprofundar nessa discussão.

¹⁹ Massachusetts Institute of Technology

²⁰ Resource Description Framework

²¹ Resource Description Framework Schema

2.3 Conversão entre Sistemas de Organização do Conhecimento

De acordo com Villazon Terrazas (2011) em 2007 surgiu um novo paradigma de desenvolvimento de ontologias, cuja ênfase é dada sobre a reutilização de recursos de conhecimento, em oposição ao costume de construção de ontologias a partir do zero. No entanto, a fim de apoiar e promover o reuso do vocabulário, novos métodos, técnicas e ferramentas são necessárias para viabilizar o aproveitamento da terminologia e suas relações. Os recursos de conhecimento, como aponta o referido autor, reutilizados nos projetos de construção de ontologias, contém uma riqueza de conceitos e refletem algum grau de consenso da comunidade de determinados domínios.

Em sua tese, refere-se como *recursos não-ontológicos* os recursos de conhecimento cuja semântica ainda não foi formalizada explicitamente por meio de ontologias. Os exemplos apresentados são as tabelas de classificação, tesouros, folksonomias, entre outros. Afirma ainda que esses tipos de recursos codificam diferentes tipos de conhecimento e podem ser implementados em diferentes maneiras.

Diante da análise de literatura apresentada por Villazon Terrazas (2011) os recursos não-ontológicos se categorizam na seguinte proposta: *recursos não-estruturados* (texto livre, por exemplo), *semi-estruturados* (folksonomias, por exemplo) e *estruturados* (bancos de dados, por exemplo). No entanto, ainda não existe um acordo de tipologia para recursos não-ontológicos.

Diante do recorte desta temática, Villazon Terrazas (2011) afirma que um tesouro representa o conhecimento de um domínio com uma coleção de termos e um conjunto limitado de relações entre eles. Relata ainda que, os tesouros são o insumo mais importante para criação, a um custo razoável, de ontologias em muitos domínios. Eles contêm, prontamente disponível, a riqueza de uma categoria de definições mais uma hierarquia, e eles refletem um grau de consenso da comunidade. Desta forma, buscou-se nas definições de tesouros, nas normas existentes para sua construção, e nos modelos de dados para a representação de tesouros, como principal fonte para sistematizar o conjunto de padrões para a conversão de tesouros em ontologias.

Em sua pesquisa sugere padrões para transformar o esquema de recursos não-ontológicos em um esquema ontologia, e o conteúdo desses recursos, em instâncias de ontologias. O método descrito por Villazon Terrazas (2011) em sua tese de doutorado consiste em dois padrões:

- *Padrões para a transformação TBox*

A abordagem Tbox de transformação impõe uma semântica formal para a re-engenharia, mesmo em condições de mudar sua estrutura. Para explicitar a semântica e as relações entre os termos do tesouro, conta-se com um recurso externo, o WordNet²². Para as relações de equivalência sugere a tendência para declarar duas classes equivalentes, quando na verdade seus termos simplesmente expressam sinônimos.

- *Padrões para a transformação Abox*

A abordagem de transformação Abox deixa a semântica informal dos recursos re-engenharia praticamente intocada. A conversão ocorre com o seguintes componentes:

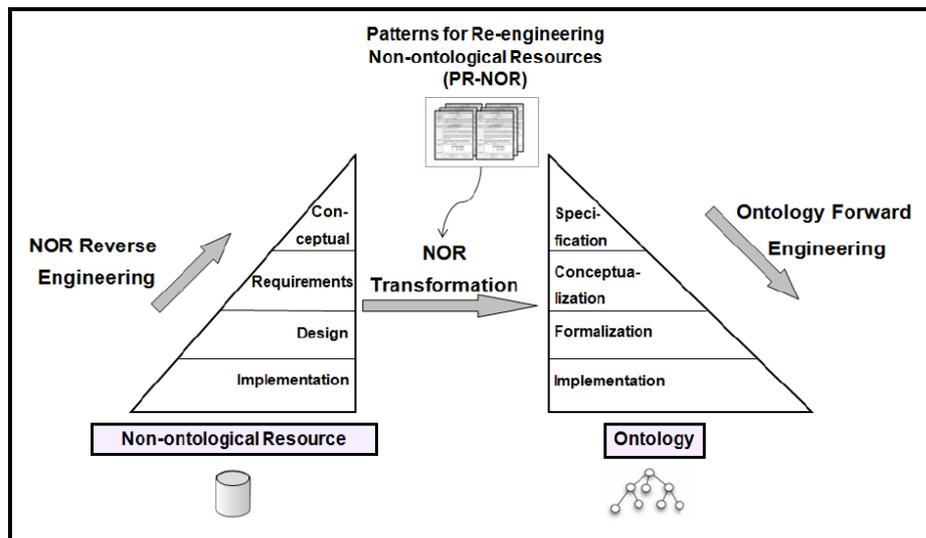
- Termo geral, que será transformado em uma classe;
- Relação hierárquica, será transformada em uma relação de sub classe;
- Relação lógica, que irá ser transformada em uma relação de *ad-hoc*;
- Termos equivalentes, os termos das relações de uso, serão transformados em etiquetas.
- E finalmente, o conteúdo do tesouro será transformado em instâncias de ontologia;

O padrão de re-engenharia do tesouro, fornece um guia para transformá-lo seguindo o modelo de dados de registro baseado em uma ontologia. O padrão transforma o

²² Trata-se de um grande banco de dados léxico para língua inglesa, em que são organizados 70 mil grupos de sinônimos, cada um representando um conceito léxico. Os sinônimos são conectados por relações, e o léxico é dividido em substantivos, verbos, adjetivos e advérbios.
Fonte: MILLER, George A. WordNet: a lexical database for English. *Communications of the ACM*, v. 38, n. 11, p. 39-41, nov. 1995.

esquema de recursos em um esquema de ontologia, e seu conteúdo, em instâncias de ontologias. Uma análise mais profunda do estado da técnica revelou que existem alguns métodos e ferramentas para transformar recursos não-ontológicos em ontologias, mas que eles têm algumas limitações (VILLAZON TERRAZAS, 2011). Para tanto o autor sugere um modelo representado na FIG.4 - Modelo de re-engenharia para recursos não-ontológicos:

Figura 4: Modelo de re-engenharia para recursos não-ontológicos



Fonte: (VILLAZON TERRAZAS, 2011)

Na perspectiva de Soergel (2004) o desenvolvimento de pesquisas para atualização dos SOCs existentes e surgimento de outros atenderia as necessidades do usuário, aprimorando a organização e recuperação da informação em uma maneira que não seria possível somente com os modelos tradicionais. O autor aponta os potenciais benefícios a partir de tais sistemas:

- *Semântica formal*: a definição explícita de conceitos e relações de uma ontologia permite a um identificador único a ser atribuído a cada conceito.
- *Consistência interna*: outro benefício da semântica explícita é a realização de coerência estrutural interna na expressão do conhecimento.
- *Interoperabilidade*: a semântica clara permite a interoperabilidade entre diferentes SOCs, e promove o compartilhamento e reutilização de conhecimento.

- *Maior integração de informações:* a interoperabilidade entre diferentes SOCs torna possível para as máquinas de reconhecer e analisar o significado pretendido dos termos de vocabulários distintos. A capacidade de integrar as terminologias de diferentes fontes maximiza o valor do investimento feito na ontologia.
- *Capacidade de inferência:* o desenvolvimento de novos SOCs tem potencial para expressar o conhecimento além do que está presente na estrutura do sistema.
- *Tratamento automatizado de informação:* descoberta de informações relevantes de fontes diferentes, explorando padrões e filtragem de informação entre os conceitos representados na ontologia. O que permite responder à pergunta, de um ou mais bancos de dados ou, usando processamento de linguagem natural a partir de texto.
- *Processamento de Linguagem Natural:* possibilita dar uma resposta direta a uma pergunta de pesquisa que se expressa em linguagem natural, utilizando-se as relações semânticas em uma ontologia.
- *Compreensão de consulta da pesquisa:* usando o Processamento de Linguagem Natural e processamento semântico, o sistema capaz de compreender uma consulta feita em linguagem natural, determinando os conceitos envolvidos.
- *Pesquisa baseada em Conceito:* uma ontologia pode oferecer capacidades de busca consciente de contexto específico para a área de interesse.
- *Integração de busca de informação e navegação de apoio:* a mineração de textos na Web orientaria o acesso ao significado ocasionando uma organização dinâmica da informação com a possibilidade de ligações entre domínios.
- *Expansão da consulta da pesquisa:* abrangência e clarificação dos termos da consulta do usuário seriam possíveis com a adição de enriquecimento ao contexto específico de informações de domínio.

Soergel (2004) apresenta uma estrutura conceitual e procedimentos de transição para apoiar a conversão de um tradicional SOC para um semanticamente rico, as ontologias. Isso exige relações semânticas que por sua vez são mais expressivas do

que as relações encontradas em tradicionais Sistemas de Organização do Conhecimento. O método apresentado por Soergel indica as seguintes etapas do processo de transformação baseados na experiência da conversão do Tesouro AGROVOC:

- Definir a estrutura da ontologia;
- Estabelecer valores de um ou mais SOC;
- Editar manualmente utilizando um editor de ontologias para tornar as informações mais precisas, adicionando mais informações.

Explorando as relações conceituais no modelo introduzido por Soergel, não existem restrições sobre as potenciais relações a serem aplicadas. O autor afirma que o modelo é extensível e quaisquer possíveis relações específicas podem ser incluídas. Porém a intenção não é obter uma lista completa de tipos de relações, mas apenas dar exemplos para ilustrar as informações adicionais e apontar com clareza a estrutura conceitual que pode ser transmitida através de mais relacionamentos (SOERGEL, 2004).

É quase impossível compilar um inventário completo de tipos de relacionamentos em ontologias. Assim, as relações devem ser acompanhadas de uma nota que especifica o significado ou a intenção do relacionamento. Isto facilitará uma dedução de novos tipos de relacionamento a partir dos relacionamentos originais (SOERGEL, 2004).

Numa terceira perspectiva aponta-se Kless (2012) que também pesquisou e publicou um método para reengenharia²³ de tesouros em ontologias. O autor utilizou o argumento de que a construção de ontologias pode ser facilitada através da adaptação de vocabulários já existentes. Relata ainda que há pouca clareza e consenso nas metodologias de como conduzir a conversão e quais modificações são necessárias nos vocabulários, a fim de transformá-los em ontologias. Sua pesquisa apresenta e fornece passos para seguir na atividade de conversão. O método foi derivado a partir das diferenças estruturais entre tesouros e ontologias,

²³ Entende-se reengenharia como conversão

bem como a partir de melhores práticas de modelagem, alguns dos quais têm sido defendidos na área biomédica (KLESS, 2012).

Cada passo do método foi ilustrado com exemplos do estudo de caso de reengenharia sobre fertilizantes agrícolas AGROVOC do estudioso Dagobert Soergel já mencionado anteriormente. O método citado consiste nos seguintes passos (KLESS, 2012):

- Refinamento de preparação e verificação do tesouro:

As relações hierárquicas podem resumir uma variedade de relações ontologicamente diferentes. As relações Genéricas, Hierárquicas, e Todo Parte devem ser distinguidas. Os relacionamentos inválidos podem ser detectados e não devem ser transferidos para a ontologia. O tesouro também deve ser analisado para conceitos não autorizados, que não fazem parte de sua hierarquia. Enquanto esses conceitos não são aceitos no tesouro, podem ser inseridos na ontologia, desde que eles apareçam como classes de nível superior. Devem ser atribuídos a um lugar apropriado na hierarquia ou tratados como sinônimos de conceitos existentes no tesouro.

- Conversão sintática:

Visa representar o tesouro em uma linguagem formal, de modo que ele pode ser modificado em um editor de ontologias. Para esta etapa, é necessário (preliminarmente) ignorar qualquer semântica das relações em uma linguagem formal e usar essa linguagem para armazenar o tesouro. Todos os conceitos sinônimos e facetas, se houver, serão representados como classes, em OWL por exemplo.

- Definição de prazo / análise da propriedade (em linguagem natural):

Tem a finalidade de preparar a conversão, principalmente para a especificação formal, sugere-se para começar com linguagem natural, especificação das classes, adicionando metadados apropriados. O objetivo desta etapa é identificar características que podem atuar como condições necessárias e suficientes para pertencentes ideais à classe em questão.

- Escolha e alinhamento de ontologias de nível superior e de relações formais:

Etapa que busca alinhar um tesauro para uma ontologia de nível superior. Inclui atribuir conceitos de nível superior ou facetas do tesauro adequados para as classes da ontologia de nível superior como classes equivalentes ou subclasses. Busca o refino das relações hierárquicas e todo-parte do tesauro de acordo com as relações formalmente definidas da ontologia. A etapa pode revelar problemas fundamentais na estrutura semântica que precisam ser resolvidos no decurso do processo de re-engenharia.

- Especificação formal de classes:

Uma das principais características das ontologias é a especificação formal de classes por meio de instâncias. A especificação de uma classe através de condições necessárias é realizada adicionando as restrições como superclasses anônimas usando o axioma subclasse. A especificação de uma classe através de uma propriedade essencial é realizado adicionando as condições necessárias e suficientes como classes equivalentes usando o axioma de classe equivalente.

- Normalização:

O termo foi adotado a partir de modelagem, e tem por objetivo dar apoio à reutilização, modularidade e facilidade de manutenção. O princípio fundamental do método é o de evitar poli-hierarquias na ontologia de modo que elas possam mais tarde ser inferidas automaticamente por algoritmos de raciocínio e tornar-se visíveis. Normalização significa então, mover classes até a hierarquia, sem perda de qualquer informação semântica. Em nenhum momento uma classe é removida ou se torna desnecessária.

- Dissociação de entidades independentes:

Facilita a capacidade de extensão de ontologias a respeito de lidar com as entidades não-essenciais dependentes, muitas vezes entidades como papéis, funções ou disposições. Os portadores de entidades dependentes, devem ser modelados de forma independente de sua respectiva combinação com uma ou mais entidades dependentes. As combinações são modeladas como subclasses definidas das entidades independentes. Para a validação dos resultados do passo de

acoplamento, bem como para validar os resultados a partir dos passos anteriores, a consistência da ontologia deve ser verificada usando um método automatizado.

- *Ajuste de pontuação, ortografia e outros aspectos de rótulos de classe e propriedade:*

Consiste em ajustar a rotulagem de classes da ontologia para melhorar a legibilidade e compreensibilidade por engenheiros e usuários de ontologia. Atualmente, não existem convenções universalmente aceitas sobre como as classes da ontologia devem ser rotuladas, no entanto, práticas comuns podem ser resumidas.

Kless (2012) afirma que a reengenharia de tesouros em ontologias requer uma variedade de habilidades e conhecimentos, assim como aqueles discutidos para a construção “pura” de ontologias. A codificação do conhecimento de um domínio é extensa e complexa para a construção de uma ontologia. Tal conhecimento não está à disposição dos “não-especialistas”, que podem precisar de tempo considerável para adquirir o conhecimento e podem modelar o domínio de forma equivocada. Assim, o autor afirma ser desejável a criação de uma equipe cobrindo a expertise de um domínio. Nota-se que seu experimento não é específico para reaproveitamento de tesouros, e pode ser adaptado a outros tipos de vocabulários controlados, tais como os sistemas de classificação.

Interessante perspectiva encontra-se no caso de conversão do Tesouro GEM (QUIN; PALING, 2001), onde os autores abordam a conversão como meio de aprimorar os instrumentos convencionais de organização do conhecimento. Ou seja, seu foco é aprimorar os instrumentos existentes, e não somente auxiliar na construção de ontologias. Quin e Paling (2001) afirmam que as ontologias tem estrutura que atendem melhor às novas necessidades de informação digital, por isso a justificativa em converter tesouros em ontologias também no sentido de aprimorar os tesouros, e, contudo, permitir que diferentes vocabulários possam ser mapeados em uma ontologia.

Diante das experiências e pesquisas relatadas, observou-se a necessidade de buscar no campo da Biblioteconomia e Ciência da Informação, teorias que pudessem fundamentar e contribuir para a melhoria do processo de conversão de tesouros em ontologias, e serão descritas a seguir.

2.3.1 Teorias Subjacentes

Notou-se que em pesquisas de conversão e reuso de linguagens, é importante que se tenha foco inicial na compatibilidade e convertibilidade entre terminologias, e em instrumentos e métodos que viabilizam a troca e/ou compartilhamento de termos e conceitos. De acordo com essa premissa é de extrema importância mencionar métodos e estudos já consolidados na área da Ciência da Informação, que mesmo tendo sido publicados e desenvolvidos há várias décadas, podem servir de parâmetro e norteador para pesquisas futuras.

No que se refere à compatibilização de linguagens, cita-se dois importantes métodos publicados na área da Biblioteconomia: o *Método de Reconciliação de Tesouros* proposto por Neville (1970) e a *Matriz de Compatibilização Conceitual* elaborada por Dahlberg (1981).

O *Método de Reconciliação de Tesouros* de Neville (1970) consiste no princípio de compatibilizar conceitos e não somente descritores. O que torna o método aplicável é a abordagem do autor em acreditar no fato de que o conceito sobrepõe ao descritor, isto é, tesouros de uma mesma área promovem a indexação de conceitos iguais, ainda que com descritores diferentes. O método propõe níveis de reconciliação, onde faz codificação numérica estabelecendo equivalência entre os conceitos e limita o número de descritores incompatíveis (NEVILLE, 1970).

No que se refere à Matriz de Compatibilização Conceitual de Dahlberg (1983) é compreendida em três etapas para mapear a potencialidade semântica e assim medir a compatibilidade entre linguagens (DAHLBERG, 1983):

- *Coincidência conceitual / grau de equivalência*: quando dois conceitos combinam suas características;
- *Correspondência conceitual / grau de similaridade*: dois conceitos combinam a maior parte de suas características;
- *Correlação conceitual*: dois conceitos são correlacionados através de símbolos matemáticos, estabelecendo uma medida de correlação, quando possuem diferentes níveis de detalhe, ou quando a relação entre eles não é de semelhança.

Em 1986, Frederick W. Lancaster empreendeu em sua obra "*Vocabulary control for information retrieval*"²⁴, que por vezes foi reeditada, um capítulo intitulado "Compatibilidade e convertibilidade" tratando-se do interesse de utilizar conjuntos de termos para integração e cooperação em sistemas de informação e bases de dados distintas. Lancaster (2002) indicou que compatibilidade é o grau de similaridade entre linguagens em seus aspectos estruturais e semânticos. No que se refere a convertibilidade, Lancaster (2002) trata da capacidade de uma terminologia ser convertida em outra, e que a compatibilização de linguagens ocorre através de tabelas de equivalência.

Ainda buscando-se fundamentar e dar embasamento teórico para o trabalho proposto fez-se necessário mencionar as tipologias das relações conceituais nos Sistemas de Organização do Conhecimento, como segue no próximo tópico.

²⁴ LANCASTER, F. *Vocabulary control for information retrieval*. 2. ed. Arlington : IRP, 1986.

2.4 Relações conceituais em Sistemas de Organização do Conhecimento

Os conceitos e relações são a base do conhecimento e do pensamento, e na perspectiva da Ciência da Informação cada vez mais as relações semânticas são utilizadas na recuperação da informação e na construção de ontologias (KHOO e NA, 2006). De acordo com os autores a CI tem centrado nos conceitos e termos, e tem como foco a identificação, o processamento e a gestão das relações para alcançar maior eficácia e aperfeiçoamento de suas técnicas já existentes.

Sales et al (2008) afirma que:

O modelo de relações que aparece em Vocabulários Controlados e em Tesouros é um modelo diádico que revela apenas as categorias às quais os conceitos pertencem, por ex: Coisa-propriedade; Material-produto; Processo-resultado; etc. Nas ontologias, as relações precisam ser explicitadas, pois integram uma proposição. Nas ontologias, o conjunto de relações é mais rico do que nas tabelas de classificação bibliográfica ou nos tesouros, o que permite maior estrutura representativa do conhecimento registrado em um discurso, para que possa ser manipulado pelo computador. É importante ressaltar que tais representações são igualmente relevantes na formulação de buscas nas ontologias. (SALES et al, 2008)

Para Sales et al (2008) as relações entre conceitos apresentam interesse conjunto para a Ciência da Informação, a Terminologia e para a Ciência da Computação. Afirma ainda que seja preciso ter bem definidos os objetivos que se quer atingir através do uso de tal linguagem no momento de elaboração, pois esses objetivos também poderão influenciar no uso de determinadas relações em detrimento de outras. Dessa forma, os objetivos específicos desses instrumentos podem sempre variar de acordo com as expectativas dos usuários e com as metas do sistema que os utilizarão (SALES et al, 2008).

Sales et al (2008) sistematiza as relações conceituais primeiramente como Relações Categorias e Relações Formais:

- As *relações categoriais* são aquelas que revelam duplas de categorias, por exemplo: coisa-processo, material-produto, etc.

- As *relações formais* são aquelas que revelam o tipo de relação existente entre as duplas de categorias, por exemplo: causado_por, ocorre etc.

Ao separar essas duas classes de relações, pode-se observar que ambas poderiam ser subdivididas em relações genéricas, partitivas e funcionais. Sendo assim, foram identificadas seis classes de relações, a saber: Relações Categoriais Genéricas, Relações Formais Genéricas, Relações Categoriais Partitivas, Relações Formais Partitivas, Relações Categoriais Funcionais, Relações Formais Funcionais (SALES et al, 2008).

De acordo com Sales et al (2008) cabe ainda ressaltar a seguinte afirmativa:

Na Ciência da Informação, a ênfase é dada ao desenvolvimento de tesouros, que incluem relações hierárquicas, partitivas, associativas – no plano do conceito – e de equivalência – no plano da língua. Mas ali a natureza das relações não é explicitada, apenas indicada por códigos. Muitos dos autores da área incluem as relações partitivas como hierárquicas, mas é preciso enfatizar o erro desta classificação. As relações hierárquicas são relações lógicas, de abstração. As relações partitivas, ao contrário, são ônticas, pois se dão entre objetos, assim como as relações associativas. Dahlberg (1978) ressalta que o estabelecimento de relações associativas pode variar de acordo com o contexto e, portanto, não se prende a seu detalhamento, mas oferece bases para sua identificação. (SALES et al, 2008)

Considerando as relações conceituais de um tesouro e atendendo a proposta deste projeto abordaremos ainda a estruturação de tesouros no que se referem à compatibilidade, vantagens e limitações em sua utilização para a conceitualização de uma ontologia. Para isso enfatizou-se de acordo com Pizzato (2003) as relações semânticas apresentadas neste instrumento:

- *Homonímia*: ocorre quando dois itens lexicais de significados distintos têm a mesma escrita e pronúncia.
- *Sinonímia*: ocorre quando duas palavras distintas em escrita e pronúncia têm o mesmo significado.
- *Antonímia*: ocorre quando duas palavras têm significados contrários.
- *Hiperonímia*: ocorre quando o significado de um lexema abrange o significado de outro lexema, e o significado de um é mais genérico que o significado de outro.

- *Hiponímia*: é a relação inversa da hiperonímia. Ocorre quando um termo tem significado mais restrito que outro termo.
- *Meronímia*: ocorre quando o significado de um lexema faz parte ou é uma porção do significado de outro lexema.

As relações semânticas oferecem meios para que sejam aplicadas as técnicas de controle de vocabulário. As normas ANSI e ISO afirmam que o emprego de suas relações são úteis para a recuperação da informação. Os três tipos de relações definidas por ambas as normas são: Equivalência, Hierarquia e Associação (PIZZATO, 2003).

- *Relações de Equivalência*: utiliza técnicas de vocabulário controlado.
- *Relações de Hierarquia*: são utilizadas para definir relações de hiperonímia, hiponímia e meronímia.
- *Relações Associativas*: representam uma classe de relações não hierárquicas, para as quais a construção e aplicação na RI é menos clara que para outros tipos de relações.

Assim, a presente proposta teve seu percurso fundamentado com relação aos diversos tipos de relações conceituais, porém deve-se ater que no âmbito das ontologias cada área se manifesta de forma distinta e até inédita, e dessa forma leva-se em conta que as relações serão sempre reflexos do domínio que se está representando.

De acordo com Smith *et al.* (2005) existe uma dificuldade em apontar todos os tipos de relações, acarretando na aceitação de que algumas relações são primitivas e independentes de um domínio. Descrevem três tipos de relações: relação *classe-classe* ("é um"), relação *instância-classe* ("é instância de") e relação *instância-instância* ("é parte de"). Nessa visão a modelagem conceitual de um domínio do conhecimento, preocupa-se em identificar a natureza de seus conceitos, em acordo com a sua materialidade e temporalidade.

Tendo como base toda essa fundamentação teórica, desde a questão das teorias da representação do conhecimento, as caracterizações dos SOCs, em especial tesouros e ontologias, até o relato das experiências de conversão e tipos de relacionamentos conceituais, seguiremos para o próximo capítulo, que descreverá o percurso metodológico desenvolvido na pesquisa.

3 PROCEDIMENTOS METODOLÓGICOS

Uma investigação científica é um processo formal e sistemático para descobrir respostas aos questionamentos por meio de procedimentos metodológicos (GIL, 1991). Complementando esse ponto de vista, “método pode ser compreendido como ‘caminho para’ e metodologia como ‘o conhecimento dos caminhos do processo científico, questionando acerca de seus limites e possibilidades’” (DEMO, 1989).

Face a essa dimensão teórica, esta pesquisa está inserida na concepção da abordagem qualitativa, que ocupa um reconhecido lugar entre as várias possibilidades de se investigar os fenômenos que envolvem as relações sociais. Além disso, trata-se de um estudo na área das Ciências Sociais Aplicadas onde as visões de mundo se baseiam nos interesses, valores e situações do indivíduo, razão pela qual a pesquisa é desenvolvida.

3.1 Caracterização da pesquisa

Tendo como base as formulações de Gil (1991) com relação a classificações de uma pesquisa científica, pode-se afirmar que este estudo apresenta as seguintes caracterizações:

- De **natureza aplicada**, pois almeja conhecimentos para dirigir-se a solução de um problema específico.
- Na perspectiva da abordagem do problema, caracterizada como **pesquisa qualitativa**, pois tem caráter descritivo e foi pautada em análises e interpretações de conteúdos para alcançar os objetivos propostos.
- Do ponto de vista dos objetivos compõe-se como uma **pesquisa de caráter exploratório**, envolvendo pesquisa em literatura científica, em normas e padrões consolidados na área, e análise de exemplos e experiências já obtidos anteriormente com objetivo de proporcionar maior familiaridade com o problema.
- Considerando os procedimentos técnicos, trata-se uma **pesquisa bibliográfica e documental**, tendo em vista que foi elaborada a partir de

publicações já realizadas em livros, artigos de periódicos, normas e padrões, bases de dados científicos e portais disponibilizados na internet.

A pesquisa caracteriza-se por ser um estudo exploratório, definido por Marconi e Lakatos (1990) como sendo o tipo de pesquisa em que, a partir de um problema formulado, familiariza o pesquisador com um ambiente, fato ou fenômeno para realização de futuras pesquisas para modificar ou clarificar conceitos.

O método de pesquisa desenvolvido buscou aliar as concepções teóricas relativas ao tema e as necessidades práticas para a coleta e tratamento dos elementos que subsidiaram o estudo. Foram utilizadas técnicas qualitativas de forma que possibilitasse analisar todas as inter-relações entre os conteúdos pesquisados.

Para fundamentar as ações e o desenvolvimento do percurso metodológico, utilizou-se como norteador as técnicas indicadas no método de Análise de Conteúdo. Na perspectiva dos procedimentos técnicos, por se tratar de uma pesquisa de análise qualitativa, e considerando-se uma pesquisa documental, os principais motivos que justificam a escolha pelo método abordado e referenciado por Bardin (2011) são:

- A possibilidade de efetuar uma análise baseada em interpretações e inferências extraídas de conteúdos de documentos;
- Por se tratar de uma análise documental, possibilita representar o conteúdo de um documento a fim de, num estágio ulterior, facilitar sua consulta e diferenciação.

Portanto, decidiu-se pelo emprego dessas técnicas para seleção, tratamento, análise e inferência dos dados. É válido mencionar que o procedimento de Análise de Conteúdo é segmentado em três fases de acordo com Bardin (2011, p. 125), a saber:

- **Pré-análise:** leitura flutuante²⁵, elaboração do plano de ideias; escolha dos documentos²⁶; formulação de hipóteses e objetivos; elaboração de indicadores;
- **Exploração do material:** sistematização das decisões tomadas na pré-análise e categorização;
- **Tratamento dos resultados, inferências e interpretações:** polo de observação; apresentação das variáveis de inferência.

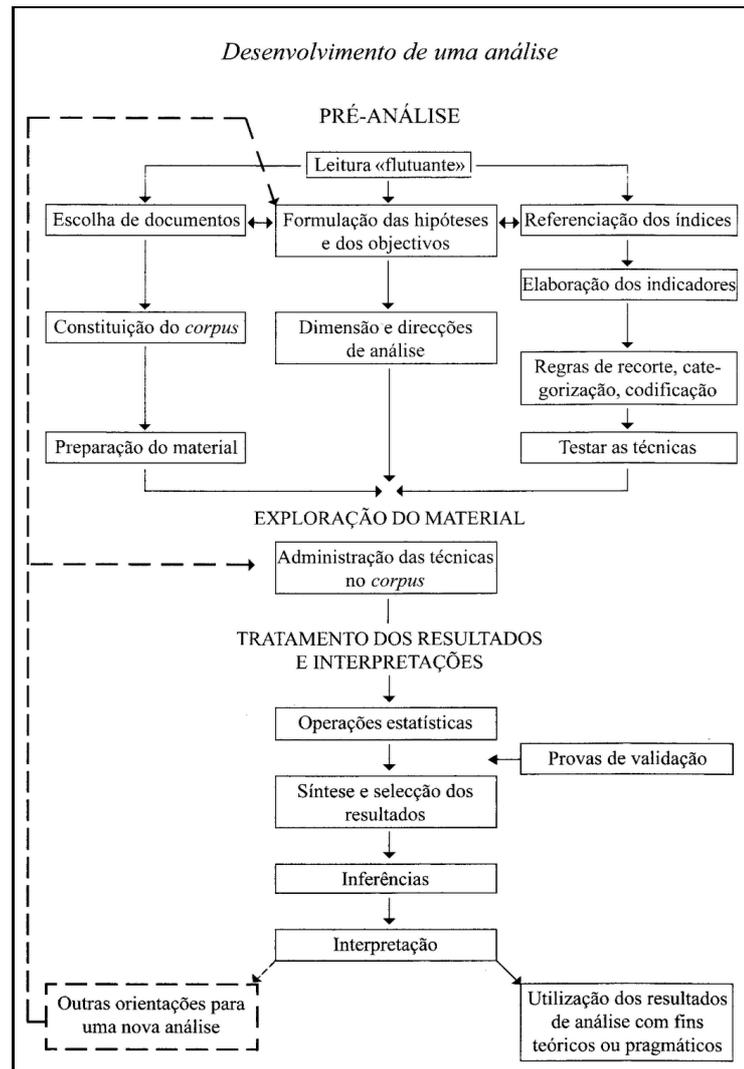
Tendo como base o desenvolvimento de uma análise de conteúdo, como apresentado na FIG. 5, as delimitações, condução da pesquisa, justificativa de seleção dos instrumentos e métodos escolhidos para se chegar aos objetivos propostos neste estudo, serão apresentados nas seções a seguir.

²⁵ Leitura que estabelece contato com os documentos, deixando-se invadir por impressões e orientações. A leitura se torna mais precisa devido ao surgimento de hipóteses ou questões norteadoras, em função de teorias conhecidas (BARDIN, 2011).

²⁶ De acordo com Bardin (2011) a escolha e seleção dos documentos devem seguir as principais regras:

- *Regra da Exaustividade:* é preciso ter-se em conta todos os elementos do *corpus* de documentos, sem deixar de fora qualquer um dos elementos.
- *Regra da Representatividade:* a análise pode efetuar-se numa amostra desde que o material a isso se preste. A amostragem diz-se rigorosa se a amostra faz parte representativa do universo inicial.
- *Regra da Homogeneidade:* os documentos retidos devem obedecer a critérios precisos de escolha e não apresentar demasiada singularidade fora desses critérios.
- *Regra da Pertinência:* os documentos retidos devem ser adequados, enquanto fonte de informação, de modo a corresponderem ao objetivo que suscita a análise (BARDIN, 2011, p.126).

FIGURA 5 - Desenvolvimento de uma análise



Fonte: (BARDIN, 2011, p.132)

3.2 Condução e recorte da pesquisa

A condução do trabalho, seleção dos documentos e direcionamento dos procedimentos metodológicos foram delineados tendo em vista os objetivos, geral e específicos, propostos no desenvolvimento desta pesquisa. Para tanto é válido retomar o objetivo geral de forma a elucidar a meta que se buscou atingir em maior nível de análise:

Objetivo geral - Apresentar com base na literatura científica, as discussões relativas ao processo de conversão de tesouros em ontologias, visando introduzir a temática nas pesquisas de âmbito nacional e contribuir para a melhoria das práticas e métodos de construção de ontologia.

Para alcançar o objetivo geral, e atender ao que foi almejado nessa pesquisa, os procedimentos metodológicos foram divididos de acordo com a proposta de cada objetivo específico. Conduzidos pela descrição desses objetivos fez-se necessário a segmentação em duas etapas que serão apresentadas e detalhadas separadamente nos tópicos seguinte e no **Quadro 4 – Sinopse dos Procedimentos metodológicos: etapas e ações:**

- **Etapa1** – Estudo de uma Norma de construção de tesouros e de um padrão de construção de ontologias de alto nível
- **Etapa 2** - Análise de publicações e experiências relativas a conversão do tesouro

Quadro 4 – Sinopse dos Procedimentos metodológicos: etapas e ações

Descrição da Etapa	ETAPA 1	ETAPA 2	Fechamento
Objetivos específicos	<p>Estudo de uma Norma de construção de tesouros e de um padrão de construção de ontologias de alto nível</p> <p><i>Identificar na literatura do campo da Ciência da Informação as principais características relativas ao processo de construção das relações conceituais de tesouros e ontologias, a fim de apreender quais elementos podem conduzir a identificação de quais relações presentes na estrutura do tesouro atendem a estrutura esperada numa ontologia.</i></p>	<p>Análise de Publicações e Experiências relativas a Conversão do Tesouro</p> <p><i>Identificar na literatura selecionada os métodos e resultados encontrados nas experiências de conversão de tesouros e ontologias, com o intuito de compreender o aproveitamento da terminologia e relações semânticas do tesouro na ontologia.</i></p>	<p>Apontamento da validade da conversão com base na Análise e sistematização dos dados levantados</p> <p><i>Apontar a validade da conversão com base na análise e sistematização dos dados bibliográficos levantados, com o propósito de abstrair as possibilidades e os limites para a efetivação da conversão de tesouros em ontologias.</i></p>
Ações	<p>Análise entre as indicações de relações conceituais na literatura e, sobretudo, em uma das Normas de Construção de tesouro, a ANSI/NISO. Z39.19 – 2005, <i>Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies</i> e no padrão para construção de ontologias de Alto Nível, o <i>Manual Ontology for the Twenty First Century: An Introduction with Recommendations</i> recomendado e publicado pelo IFOMIS - Institute for Formal Ontology and Medical Information Science.</p>	<p>Seleção pelos documentos se deu pela abordagem de relatos de experiência e apontamento de padrões para conversão nas quatro publicações abaixo:</p> <ol style="list-style-type: none"> 1) SOERGEL, Dagobert, ET AL. Reengineering thesauri for new applications: the Agrovoc example. <i>Journal of Digital Information</i>, v.4, n.4, 2004. 2) VILLAZON TERRAZAS, Boris Marcelo. <i>Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies</i>. 2011. 238 f. Tese (Doutorado em Inteligencia Artificial) – Facultad de Informática (UPM), Universidade Politécnica de Madrid, Madrid. 3) KLESS, D., Jansen, L., LINDENTHAL, J., WIEBENSIGN, J.: <i>A Method of Re-Engineering a Thesaurus into an Ontology</i>. In: Donnelly, M., Guizzardi, G. (eds): <i>Formal Ontology in Information Systems - Proceedings of the Seventh International Conference (FOIS 2012)</i>, pp.133–146. IOS Press, Amsterdam (2012). 4) QIN, J.; PALING, S. Converting a controlled vocabulary into an ontology: the case of GEM. 2001. <i>Information Research</i>, v. 6, n. 2. 	<p>Optou-se por analisar e sistematizar a literatura relativa ao objeto, e apresentar um estudo comparativo entre as relações conceituais e as possibilidades e limites de se converter um Tesouro em uma Ontologia.</p>

Fonte: Elaborado pela autora

3.2.1 Etapa 1: Estudo de uma Norma de construção de tesouros e de um padrão de construção de ontologias de alto nível

Objetivo específico 1 - identificar na literatura do campo da Ciência da Informação as principais características relativas ao processo de construção das relações conceituais de tesouros e ontologias, a fim de apreender quais elementos podem conduzir a identificação de quais relações presentes na estrutura do tesouro atendem a estrutura esperada numa ontologia.

Para tanto buscou-se as indicações de relações conceituais em uma das Normas de Construção de tesouro, a ANSI/NISO Z39.19 – 2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, e no padrão para construção de ontologias de Alto Nível, o *Manual Ontology for the Twenty First Century: An Introduction with Recommendations* recomendado e publicado pelo IFOMIS - Institute for Formal Ontology and Medical Information Science.

A escolha pela norma ANSI/NISO. Z39.19 – 2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* é justificada por sua consolidação na literatura da área, por ter abrangência internacional e por ser utilizada mundialmente para orientação na construção de tesouros e outros vocabulários controlados. Especificamente, abordou-se o Capítulo 8 – Relacionamentos - onde se encontram as indicações e recomendações dos tipos de relações a serem construídas.

Com relação ao padrão escolhido para se verificar a construção de uma ontologia, se fez necessário selecionar uma ontologia de alto nível, visto que estas contêm definições abstratas necessárias para a compreensão de aspectos do mundo, como tempo, processos, papéis, espaço, seres, coisas, etc.

No *Manual Ontology for the Twenty First Century: An Introduction with Recommendations*, desenvolvido por Andrew D. Spear, do Institute for Formal

Ontology and Medical Information Science (IFOMIS)²⁷, Saarbrücken, Germany University at Buffalo, Buffalo, New York, U.S.A., selecionou-se o Capítulo 5 como categoria de análise. Intitulado “Introdução às relações básicas de ontologia formal”²⁸, o capítulo trata especificamente da seleção de terminologia, definição e classificação na construção da estrutura hierárquica do domínio pretendido.

A seleção desses dois instrumentos, e a identificação de suas características no que se refere a estrutura e relacionamentos recomendados, conduziu a alcançar o primeiro objetivo específico proposto na pesquisa, onde foram dadas condições de comparar as recomendações da estrutura esperada para um tesouro e a de uma ontologia. Abaixo segue um descritivo do processo de análise de conteúdo, seguido de seu quadro sinóptico. No próximo capítulo serão apresentados os resultados e as inferências advindos dessa técnica.

- Pré-análise: *leitura flutuante; formulação de hipóteses e objetivos; elaboração de indicadores;*

Documentos selecionados e categorias de análise:

- Documento 1: ANSI/NISO Z39:19-2005

- Categoria de análise: Relações conceituais / Capítulo 8 –
Relacionamentos

- Documento 2: *Manual Ontology for the Twenty First Century: An Introduction with Recommendations*

- Categoria de análise: Relações conceituais / Capítulo 5 - *Introdução às relações básicas de ontologia formal*

- Exploração do material: *sistematização das decisões tomadas na pré-análise e categorização:*

²⁷ IFOMIS consiste em um grupo de pesquisa interdisciplinar, com membros de Computação, Filosofia e Ciência da Informação, Lógica, Medicina e Informática Médica, com foco em pesquisa teórica fundamentada na ontologia formal e aplicada. Seu objetivo é desenvolver uma ontologia formal que será aplicado e testado no domínio da ciência da informação médica e biomédica.

²⁸ Tradução de: “*Introduction to Basic Formal Ontology and Relations*”

- Levantamento de todos os tipos de relações indicadas nos capítulos selecionados nos dois instrumentos selecionados na pré-análise. As tipologias e características serão apresentadas no capítulo seguinte no contexto dos resultados e inferências.

- Tratamento dos resultados, inferências e interpretações: *polo de observação; apresentação das variáveis de inferência.*

- Elaborou-se um quadro contendo os tipos de relacionamentos de forma a sintetizar e ilustrar os resultados, as inferências e interpretações, que também serão disponibilizadas no próximo capítulo.

De forma a caracterizar o método aqui adotado, apresenta-se o Quadro 5 - Sinopse da aplicação do método de Análise de Conteúdo – Etapa 1, no intuito de dar um panorama das decisões e do caminho percorrido.

Quadro 5 - Sinopse da aplicação do método de Análise de Conteúdo – Etapa 1

Estudo de Normas de Construção de Tesouros e Padrões de construção de Ontologias de Alto Nível					
FASES	PRÉ-ANÁLISE	KOS	Tesouro	Ontologia de Alto Nível / BFO	
		Objeto selecionado	ANSI/NISO Z39:19-2005 – Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies	Manual Ontology for the Twenty First Century: An Introduction with Recommendations	
	EXPLORAÇÃO DO MATERIAL	DIMENSÕES E VARIÁVEIS DE ANÁLISE	Tipologia	Padrão	Manual
			Abordagem	Apresenta orientações e convenções para o conteúdo, exposição, construção, testes, manutenção e gerenciamento de vocabulários controlados.	Representa uma introdução à ontologia, informações ontologia e ontologias projeto. Baseia-se na abordagem realista à ontologia e no básico associado Basic Formal Ontology (BFO) e princípios de melhores práticas de ontologia
			Objetivos	Alcançar a consistência na descrição de objetos de conteúdo e facilitar recuperação	Estruturar ontologias de domínio, de forma que as torne tratável e interoperáveis por computador
			Mantenedores	ANSI / NISO	IFOMIS
			Unidades de Análise	Capítulo 8 Relacionamentos	Capítulo 5 Introdução às relações básicas de ontologia formal
	INFERÊNCIAS E INTERPRETAÇÕES		Tipos de Relacionamento: Equivalência, Hierarquia, Associativa	Tipos de entidade: Entidades Ocorrentes (SPAN), Entidades Continuantes (SNAP)	

Fonte: Elaborado pela autora.

De forma a viabilizar as interpretações nos documentos selecionados, estabeleceu-se como unidades de contexto e unidades de registro, os títulos, resumos e tipologia dos relacionamentos de cada um dos manuais. A saber:

Norma para Construção de Tesouros	
TÍTULO	ANSI/NISO Z39:19-2005 – Guidelines for the Construction, Format, and Managment of Monolingual Controlled Vocabularies
RESUMO	<p>Esta norma apresenta orientações e convenções para o conteúdo, apresentação, construção, testes, manutenção e gerenciamento de vocabulários controlados monolíngues. A norma se concentra em vocabulários controlados que são usados para a representação de objetos de conteúdo em sistemas de organização do conhecimento incluindo listas, anéis de sinônimo, taxonomias e tesouros. Deve ser considerada como um conjunto de recomendações com base nos procedimentos e técnicas preferenciais. Procedimentos opcionais são, no entanto, às vezes descritos, por exemplo, para a exibição dos termos de um vocabulário controlado. O objetivo principal do vocabulário controlado é alcançar a consistência na descrição de objetos de conteúdo e facilitar recuperação. O controle de vocabulário é realizado por três métodos principais: definindo o escopo, ou significado, de termos; usando a relação de equivalência para vincular sinônimos e quase sinônimos de termos; e fazer a distinção entre homógrafos.</p>
CAPÍTULO 8	<p>Existem três tipos de relações utilizadas nos vocabulários controlados:</p> <ul style="list-style-type: none"> a) Equivalência b) Hierarquia c) Associação <p><i>8.1 Vinculação Semântica</i></p> <p>As relações entre os termos em um vocabulário controlado são indicadas por vinculação semântica. Engloba várias técnicas e convenções para indicar a relação entre termos.</p> <p>Formatos para a exibição de relações semânticas são descritos na secção 9; esta secção está de acordo com a definição e ilustra as relações em si.</p>

Manual de recomendações da BFO	
TÍTULO	Manual Ontology for the Twenty First Century: An Introduction with Recommendations (BFO)
RESUMO	<p>O material apresenta uma introdução à ontologia, informações de ontologia e projetos de ontologias. Esta introdução baseia-se na abordagem realista à ontologia e no associado à Ontologia Formal Básica (BFO), e em princípios de melhores práticas de ontologia que foram desenvolvidos e pesquisados no IFOMIS nos últimos quatro anos. Enquanto se concentra no trabalho feito no IFOMIS, destina-se a maior parte da discussão de uma natureza geral a ser aplicável a um grande número de questões teóricas e aplicadas em ontologia geral. O material disponível é parte de um projeto em andamento para fornecer um recurso <i>one-stop</i> e introdução para a ontologia, com especial ênfase nos princípios de melhores práticas de ontologia. No presente aguardam críticas, sugestões de recursos adicionais para inclusão e <i>feedback</i> quanto a precisão e a utilidade deste material.</p>
CAPÍTULO 5	<p><i>Introdução à ontologia Formal básica e relações</i></p> <p>Estruturação de ontologias de domínio requer o uso de ontologias formais ou de "nível superior", de forma a torná-las tratáveis e interoperáveis por computador, bem como o processamento das informações de forma clara, rigorosa, e inequívocas quanto possível. Além disso, as questões importantes da seleção de terminologia, termo-definição e classificação, no sentido da construção de estruturas de hierarquia, como organizar as informações contidas em um domínio da ontologia, todos beneficiam e podem ser mais bem entendidos no contexto de uma ontologia formal explicitamente definida. Aqui oferecemos uma introdução a alguns dos principais recursos da Ontologia Formal Básica (BFO) que foi desenvolvido no Instituto de Ontologia e Ciência da Informação Médica na Alemanha de Saarbrücken.</p>

3.2.2 Etapa 2: Análise de Publicações e Experiências relativas a Conversão do Tesauro

Objetivo específico 2 - Identificar na literatura selecionada os métodos e resultados encontrados nas experiências de conversão de tesauros e ontologias, com o intuito de compreender o aproveitamento da terminologia e relações semânticas do tesauro na ontologia.

Para atender ao segundo objetivo específico utilizou-se como método a estratégia de busca em bases de dados com o intuito de encontrar relatos de experiências, estudos de casos, padrões e metodologias para a convertibilidade de tesauros em ontologias. O procedimento teve por finalidade a identificação dos pesquisadores da área e a verificação dos motivos e resultados de suas pesquisas. Além disso, buscou-se observar a análise dos conceitos e construtos teóricos adotados nos trabalhos encontrados.

A pesquisa empírica da literatura procura explicar como a temática vem sendo analisada do ponto de vista metodológico, no intuito de identificar quais procedimentos, técnicas e métricas normalmente são empregados nas publicações da área.

A seleção documental da investigação foi constituída por pesquisas da área de Ciência da Informação, encontradas nas bases de dados *Library and Information Science Abstracts* (LISA)²⁹, na Base de Periódicos (PERI)³⁰, no Google Acadêmico³¹ e na Biblioteca Digital de Teses e Dissertações do IBICT³².

Adotou-se como estratégia de busca os termos “Sistemas de Organização do Conhecimento”, em português, “*Knowledge Organization Systems*”, em inglês, assim como “Tesauros”, “Ontologias” e “Conversão” e “convertibilidade” em ambos os idiomas, português e inglês.

²⁹ Library and Information Science Abstracts

³⁰ Disponível em: <http://bases.eci.ufmg.br/peri.htm>

³¹ Disponível em: <http://scholar.google.com.br/schhp?hl=pt-BR>

³² Disponível em: <http://bdt.d.ibict.br/>

Para realizar a seleção dos trabalhos a serem analisados, foram estabelecidos critérios de relevância e categorias de análise para escolha dos documentos a serem estudados, como descritos nos procedimentos a seguir. Embora de acordo com os critérios mencionados o resultado da seleção pelos documentos também se deu pela abordagem de relatos de experiência e apontamento de padrões para conversão.

Caracterizando a Análise de Conteúdo nesta Etapa da pesquisa, o caminho e decisões seguiram o seguinte escopo:

- **Pré-análise:** leitura flutuante, elaboração do plano de ideias; Documento selecionado; formulação de hipóteses e objetivos; elaboração de indicadores;

Categorias de análise:

- Aplicação prática da pesquisa publicada
- Vínculo com instituição de pesquisa em Ciência da Informação
- Reconhecimento da pesquisa diante da área de CI
- Publicação das metodologias e resultados

Documentos selecionados:

- 1) SOERGEL, Dagobert, et al. Reengineering thesauri for new applications: the Agrovoc example. *Journal of Digital Information*, v.4, n.4, 2004.
- 2) VILLAZON TERRAZAS, Borís Marcelo. *Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies*. 2011. 238 f. Tese (Doutorado em Inteligencia Artificial) – Facultad de Informática (UPM), Universidade Politécnica de Madrid, Madrid.

- 3) KLESS, D., Jansen, L., LINDENTHAL, J., WIEBENSIHN, J. *A Method of Re-Engineering a Thesaurus into an Ontology*. In: Donnelly, M., Guizzardi, G. (eds): *Formal Ontology in Information Systems - Proceedings of the Seventh International Conference (FOIS 2012)*, pp.133–146. IOS Press, Amsterdam (2012).
- 4) QIN, J.; PALING, S. Converting a controlled vocabulary into an ontology: the case of GEM. *Information Research*, v. 6, n. 2, 2001.

- Exploração do material: sistematização das decisões tomadas na pré-análise e categorização:

Tendo em vista os diferentes tipos de documentos e abordagens, buscou-se categorizar os padrões e métodos apontados pelos pesquisadores. Para tanto, verificou-se os requisitos básicos de cada padrão e as contribuições para a área, no intuito de medir a complexidade e plenitude das experiências e metodologias propostas nas quatro pesquisas analisadas.

- Tratamento dos resultados, inferências e interpretações: polo de observação; apresentação das variáveis de inferência.

No contexto da análise proposta, serão apresentadas as inferências e resultados no capítulo seguinte.

Previamente apresenta-se a seguir um quadro sinóptico dos procedimentos metodológicos para esta segunda etapa da pesquisa, a saber, **Quadro 6 - Sinopse da aplicação do método de Análise de Conteúdo – Etapa 2:**

Quadro 6 - Sinopse da aplicação do método de Análise de Conteúdo – Etapa 2

Análise de Publicações e Experiências relativas à Conversão do Tesouro						
PRÉ-ANÁLISE	Pesquisa em bases de dados		Portal IBICT / LISA / PERI / Google Acadêmico			
	Objeto selecionado		SOERGEL, Dagobert, et al. Reengineering thesauri for new applications: the Agrovoc example. <i>Journal of Digital Information</i> , v.4, n.4, 2004.	VILLAZON TERRAZAS, Boris Marcelo. Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies. 2011. 238 f. Tese (Doutorado em Inteligencia Artificial) – Facultad de Informática (UPM), Universidade Politécnica de Madrid, Madrid.	KLESS, D., Jansen, L., LINDENTHAL, J., WIEBENSIHN, J..A Method of Re-Engineering a Thesaurus into an Ontology. In: Donelly, M., Guizzardi, G. (eds): <i>Formal Ontology in Information Systems - Proceedings of the Seventh International Conference (FOIS 2012)</i> , pp.133–146. IOS Press, Amsterdam (2012)	QIN, J.; PALING, S. <i>Converting a controlled vocabulary into an ontology: the case of GEM</i> . 2001. <i>Information Research</i> , v. 6, n. 2.
EXPLORAÇÃO DO MATERIAL	DIMENSÕES E VARIÁVEIS DE ANÁLISE	Publicação	Journal of Digital Information / 2004	Universidade Politécnica de Madrid, Madrid / 2011	IOS Press, Amsterdam / 2012	Information Research
		Tipologia do documento	Artigo Científico	Tese de Doutorado	Artigo apresentado em Conferência Internacional sobre ontologia Formal e sistemas de informação	Artigo Científico
		Abordagem	Estudo de caso – Tesouro Agrovoc	Aplicação projetos de Neon (FP6-027595) e SEEMP (FP6-027347)	Definição de etapas para o processo de conversão	Estudo de caso – Tesouro GEM
		Objetivos	Explorar a reengenharia/conversão de KOS tradicionais em Ontologias	Estabelecer padrão de procedimentos	Estabelecer roteiro para conversão	Converter o vocabulário controlado GEM em uma ontologia
		Representantes	Dagobert Soergel / FAO Nações Unidas	VILLAZON TERRAZAS, Boris Marcelo.	Kless, D., Jansen, L., Lindenthal, J., Wiebensohn, J	QIN, J.; PALING, S.
		Unidades de Análise	Caso AGROVOC: princípios que conduziram a experiência	Padrão desenvolvido para conversão	Etapas definidas para conversão de tesouros em ontologias	Caso GEM: princípios que conduziram o experimento

Fonte: Elaborado pela autora.

Em continuidade ao atendimento dos objetivos propostos, é válido retomar a descrição do terceiro objetivo, a saber:

Objetivo específico 3 - Apontar a validade da conversão com base na análise e sistematização dos dados bibliográficos levantados, com o propósito de abstrair as possibilidades e os limites para a efetivação da conversão de tesouros em ontologias.

Coloca-se como resultado do terceiro objetivo uma análise geral de apontamento da validade da conversão, baseando-se nas informações e experiências levantadas. Dessa forma, optou-se por analisar e sintetizar as reflexões sobre a literatura relativa ao objeto de modo mais abrangente, com o intuito de sair do silogismo entre os resultados das etapas 1 e 2 e apresentar uma síntese das inferências e interpretações.

Destina-se assim a apreensão geral dos objetivos, fazendo relato da importância do estudo e apontando suas fragilidades. Seguindo a estrutura textual deste trabalho, as descrições desse fechamento serão apresentadas no final do capítulo 4.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

4.1 RESULTADOS: inferências e avaliações

Dentro do escopo deste estudo exploratório, apresenta-se como resultado da pesquisa, as principais reflexões encontradas nos procedimentos de análise de conteúdo descritos no capítulo anterior.

Conforme os precedentes à estrutura proposta, decorrente dos procedimentos metodológicos estabelecidos, e de acordo com os objetivos específicos já mencionados, a amostra representativa foi analisada em duas etapas, e em sequência fez-se um fechamento, conclusivo, originado da análise das duas primeiras ações. Portanto, este capítulo reúne a análise e interpretação dos dados encontrados nos conteúdos dos documentos examinados, apresentados na sequência das Etapas 1 e 2.

De forma mais específica, pode-se indicar resumidamente os principais resultados para a presente pesquisa, que serão detalhados e apresentados em cada tópico, a saber:

- *Resultados – Etapa 1:* Enumeração das relações conceituais e análise de suas características indicadas na Norma ANSI/NISO Z39:19-2005, no *Manual Ontology for the Twenty First Century: An Introduction with Recommendations*, e na literatura pertinente consultada no percurso da pesquisa.
- *Resultados – Etapa 2:* Sistematização das quatro experiências de conversão de tesouros em ontologias selecionadas neste estudo, sobretudo, o apontamento e avaliação de suas metodologias estabelecidas e principais contribuições para a área de CI.

Por fim, apresenta-se como resultado geral, a indicação das principais contribuições do estudo para o processo de construção de ontologias e da validade e potencialidade do processo de conversão.

4.1.1 Resultados – Etapa 1

Na busca por princípios metodológicos que possam nortear a conversão de tesouros em ontologias, encontrou-se, através dos padrões de construção selecionados e da literatura pertinente ao assunto, a identificação de elementos e orientações que resultaram nas análises e inferências a respeito do processo de transposição dos relacionamentos e conceitos de um instrumento ao outro. Tais elementos evidenciam que, as características e natureza das relações, nos dois instrumentos, podem partir de princípios semelhantes.

A abordagem evidencia que as relações do tesouro podem sim responder à determinação do domínio e escopo da ontologia, entretanto, como dados iniciais, tendo em vista que as classes e relações na ontologia são mais complexas e devem ser revistas por um especialista.

Para fundamentar essa percepção foram destacados os tipos de relações encontrados em ambos os instrumentos selecionados na metodologia, ou seja, em consulta à norma ANSI/NISO. Z39.19 – 2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, e ao *Manual Ontology for the Twenty First Century: An Introduction with Recommendations*.

Em seguida, foi válido retomar algumas das características e, sobretudo, ao entendimento das semelhanças e diferenças entre os dois instrumentos, principalmente no que se refere à estrutura das relações conceituais discutidas na literatura explorada nesta pesquisa.

Através da consulta a ANSI/NISO. Z39.19 – 2005, notou-se que as relações dos tesouros são caracterizadas por três tipos:

- *Equivalência*
 - Sinônimos
 - Variantes Lexicais
 - Sinonímia

- *Hierárquica*
 - Genérica
 - Instância
 - Todo / parte

- *Associativa*
 - Causa / Efeito
 - Agente / Processo
 - Processo / Contra-agente
 - Ação / Produto
 - Ação / Propriedade
 - Ação / Alvo
 - Conceito ou Objeto / Propriedade
 - Conceito ou Objeto / Origem
 - Conceito ou Objeto / Medida ou Mecanismo
 - Matéria-prima / Produto
 - Disciplina ou Campo / Objeto ou Profissional

O padrão ANSI/NISO Z39.19 indica que a definição do escopo dos termos é restrita aos significados que o domínio do tesouro abrange, e, ainda, que os termos devem ser estabelecidos de maneira a evidenciar, a qualquer usuário do instrumento, qual o escopo pretendido.

Nas ontologias as relações são definidas de acordo com o domínio trabalhado, e apresentam estrutura mais abrangente e complexa no relacionamento entre os termos. Contudo, neste estudo em específico, consultou-se o Manual da BFO, que indica os seguintes tipos de entidades:

- ***Continuantes – SNAP:***

- Regiões espaciais
- Entidades Substanciais
- Substâncias
- Partes Fiat
- Limites/Fronteiras
- Agregados
- Lugar
- Entidades dependentes
- Universais

- ***Ocorrentes – SPAN:***

- Entidades Processuais
- Processos
- Partes Fiat
- Eventos
- Agregados de processos
- Configurações – ambiente espaço temporal
- Regiões espaço temporal

Cabe ainda mencionar que as entidades *Continuantes* mantém sua identidade ao longo do tempo e podem ser de três tipos: dependentes, independentes ou região. Já as *Ocorrentes* são processos, eventos, atividades, portanto, ocorrem em um dado momento no tempo. Aparecem em três tipos: entidade processual, região espaço-temporal e região temporal.

Para tanto, sistematicamente apresenta-se no **Quadro 7 - Tipologia de Relacionamentos ANSI/NISO Z39:19-2005 x Manual BFO**, a tipologia dos relacionamentos/entidades recomendados nos dois instrumentos de análise deste estudo.

Quadro 7 – Tipologia de Relacionamentos ANSI/NISO Z39:19-2005 x Manual BFO

Tesauros			Ontologias			
ANSI/NISO Z39:19-2005 – Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies			Manual Ontology for the Twenty First Century: An Introduction with Recommendations (BFO)			
Tipos de Relacionamentos			Tipos de Entidades			
Relações	Equivalência	Sinônimos	Entidades	Continuante	SNAP	Regiões Espaciais
		Variantes Lexicais				Entidades Substanciais
		Sinonímia				Substâncias
	Hierárquica	Genérico				Partes fiat
		Instância				Limites (Fronteiras)
		Todo / parte				Agregados
	Associativa	Causa / Efeito				Lugar (site)
		Agente / Processo				Entidades dependentes
		Processo / Contra-agente				Universais
		Ação / Produto		Entidades Processuais		
		Ação / Propriedade		Processos		
		Ação / Alvo		Partes fiat		
		Conceito ou Objeto / Propriedade		Eventos		
		Conceito ou Objeto / Origem		Agregados de Processos		
		Conceito ou Objeto / Medida ou Mecanismo		Configurações (ambiente espaço-temporal)		
Matéria-prima / Produto	Regiões espaço-temporal					
Disciplina ou Campo / Objeto ou Profissional						
		Ocorrente	SPAN			

Fonte: elaborado pela autora.

Sendo assim, com o propósito de evidenciar a correlação das tipologias e natureza das relações, tornou-se necessário verificar as semelhanças e diferenças entre os dois tipos de SOCs aqui analisados, os tesauros e as ontologias, mais precisamente.

Dessa forma, com base nos documentos selecionados, e, sobretudo na literatura explorada, entendeu-se que as relações são tipicamente distintas, ou seja, apresentam estruturas diferentes, pois cada instrumento possui suas necessidades e objetivos. Mas revela-se que podem originar-se de princípios semelhantes, principalmente no que se refere aos fundamentos de categorização e de classificação.

De acordo com Robredo e Brascher (2010) os Tesauros e Ontologias estruturam, classificam, modelam e representam conceitos e relacionamentos num domínio do conhecimento. Relatam ainda que, são formados, basicamente, por vocabulários controlados acompanhados de relacionamentos semânticos entre os termos. Mantêm pontos em comum, mas diferenciam em alguns aspectos, entretanto, desempenham a função de organizar e recuperar informações (ROBREDO; BRÄSCHER, 2010, p. 161).

Sales e Café (2008) sintetizam as inúmeras definições de tesauros encontradas na literatura, e afirmam que tesauros são vocabulários controlados formados por termos (descritores) semanticamente relacionados, e atuam como instrumentos de controle terminológico. Enfatizam que os tesauros podem estar estruturados hierarquicamente (gênero-espécie e todo-parte) e associativamente (aproximação semântica), e, são utilizados principalmente para indexar e recuperar informações por meio de seu conteúdo (SALES; CAFÉ, 2008).

Tratando-se de ontologias, os autores as definem como o vocabulário usado para compor expressões complexas. Relatando que o objetivo da ontologia é viabilizar um comum acordo no uso do vocabulário compartilhado de uma maneira coerente e consistente (SALES; CAFÉ, 2008).

É oportuno ressaltar que, de acordo com Alvares (2012), recomenda-se que na estrutura do tesauro os termos sejam relacionados semântica e logicamente, formando seus relacionamentos, ligações e vinculações entre os conceitos

representados. A autora indica que basicamente a estrutura dos termos é composta por (ALVARES, 2012):

- *Descritor*: Termo escolhido para representar um conceito no Tesauro. Pode-se utilizar a sigla USE quando houver outros termos que representem o mesmo conceito, antes do termo descritor.
- *Não-descritor* (UP): Embora descreva o mesmo conceito que o descritor, trata-se de um termo não autorizado, para evitar a aparição de sinônimos, e utiliza-se a sigla UP.
- *Nota explicativa* (NE): Indica uma definição do termo ou uma orientação sobre o mesmo.
- *Termo genérico* (TG): Indica que há relação hierárquica entre termos com relação gênero-espécie e que este descritor representa o termo com o conceito mais abrangente.
- *Termo específico* (TE): Refere-se aos termos subordinados ao termo genérico.
- *Termo relacionado* (TR): Aponta relação entre termos que não formam uma hierarquia, mas possuem relação de associação.
- *Categoria* (CAT): Grande grupo ao qual pertence o descritor.

Todavia, segundo Robredo & Brascher (2010) os tipos de relacionamentos e a forma de representá-los variam de acordo com o SOC, e contemplam a noção de hierarquia entre conceitos. Tanto nos tesauros quanto nas ontologias, as hierarquias seguem princípios rigorosos. Apontam que as diretrizes e normas de elaboração de tesauros definem os seguintes relacionamentos hierárquicos (ROBREDO; BRASCHER, 2010):

- Relações gênero/espécie (TG – termo genérico / TE – termo específico). Ex.: instituição de ensino superior / universidade
- Relações todo/parte (TGP - termo genérico partitivo / TEP – termo específico partitivo). Ex.: universidade/faculdade/departamento
- Relação de instanciação (é um, ou IsA). Ex.: universidade /Universidade de Brasília.

Robredo e Brascher (2010) exemplificam em seu trabalho a relação gênero/espécie nas ontologias, utilizando uma hierarquia de classes da ontologia de vinho. Afirmam que nas ontologias, a relação de instanciação ocorre pela ligação de cada indivíduo como membro de determinada classe. Assim, ainda no exemplo do domínio vinho, Casillero del Diablo, Cabernet Sauvignon, é uma instância da classe vinho tinto. Contrapõem afirmando que na estrutura de um tesouro vinho seria um termo específico de bebida, e na ontologia o vinho é uma subclasse de bebida. Dessa forma, a relação todo/parte, assim como outros tipos de relacionamentos entre conceitos, pode ser especificada nas ontologias por meio das propriedades, as quais são definidas segundo os objetivos da ontologia (ROBREDO; BRÄSCHER, 2010, p. 171-172).

Tendo como base as diretrizes de construção e características dos dois SOCs analisados, nota-se que, independente do domínio, a estrutura de relações do tesouro é mais rígida e moderada. Enquanto nas ontologias as relações e classes são mais flexíveis, e, portanto, mais completas, e ainda são variáveis de acordo com o domínio e objetivos da ontologia.

Tal afirmação faz refletir as possibilidades e limites em se converter as relações do tesouro para a construção de uma ontologia. Contudo, há de se considerar que a fundamentação de ambas as estruturas se assemelham no que se refere às bases de categorização em facetas e nos estudos de classificação.

No contexto da fundamentação de faceta e classificação, Lima (2004) evidencia as seguintes questões da análise facetada:

Com relação as categorias fundamentais, o CRG afirma que estas devem ser derivadas da natureza dos assuntos a serem classificados e que nem todos os assuntos possuem categorias fundamentais. O CRG prefere identificar as categorias fundamentais pelo contexto do próprio assunto, concluindo que nenhuma lista de categorias fundamentais deva ser exaustiva ou imposta mecanicamente aos assuntos. Comparadas ao PMEST de Ranganathan, as categorias fundamentais propostas pelo CRG, devido a flexibilidade que permitem, são atraentes aos classificadores, que podem moldá-las à assuntos específicos, permitindo assim, a formação de categorias mais distintas e definidas (LIMA, 2004, p. 78).

Na perspectiva de Ranganathan são propostos princípios classificatórios para a organização de conceitos dentro de classes, que são estruturadas de acordo com categorias do domínio. Como abordado por Campos et al (2010), Ranganathan propõe princípios classificatórios para a organização dos conceitos de um domínio dentro de classe. Tais princípios são (CAMPOS et al, 2010):

“Cânone da Modulação: a ordem dos conceitos em uma classe de conceitos deve ter uma sequência que respeite cada elo da cadeia;

Cânone de Concomitância: duas características distintas (por exemplo, idade e data de nascimento) não devem dar origem a um mesmo renque (série horizontal) de conceitos;

Cânone da Diferenciação: uma característica usada como base para a classificação de um universo deve dar origem a pelo menos duas classes;

Cânone da Exaustividade: o conjunto de elementos de uma classe não deve constituir uma outra classe, ou seja, os elementos de um renque devem ser mutuamente exclusivos;

Cânone para Classes Subordinadas: numa sucessão de classes cada uma deve seguir imediatamente a outra, sem que seja interrompida pela inclusão de uma ou mais classes de natureza distinta.” (CAMPOS et al, 2010).

Os princípios citados fornecem elementos para analisar a formação dos relacionamentos nos tesouros e ontologias, inclusive para validação das relações e classes convertidas.

Analisando a construção de tesouros e ontologias, sabe-se que, para estabelecer os limites do tema do tesouro, faz-se um levantamento prévio de classificação, no intuito de delimitar as grandes categorias e atender a garantia literária do assunto em questão. Já no que se refere ao processo da construção de Ontologias, Campos & Souza (2009) enfatizaram que:

A determinação de conceitos e relações semânticas em ontologias é realizada na etapa de conceitualização, quando há a determinação do domínio e quando se define o escopo da ontologia, de acordo com os comprometimentos ontológicos que a linguagem assume. A delimitação dos conceitos válidos para a representação do domínio deve estar intimamente relacionada com a realidade que esta pretende representar (CAMPOS; SOUZA, 2009, p.116).

No entanto, como destacam Noy e McGuinness (2001), o desenvolvimento de uma ontologia inclui definição de classes, estruturação de classes numa hierarquia taxonômica, definição de atributos e descrição dos valores permitidos para esses atributos, o que difere a cada processo de sua construção.

Sabe-se que nos tesouros há uma preocupação com a *garantia literária* em sua construção, o mesmo ocorre nas ontologias, sendo designado como compromisso ontológico. Garantia literária entende-se como a utilização de literatura para delimitação de um domínio e como fonte para extração e validação da terminologia proposta para a área de abrangência do tesouro. Nas ontologias, o que se assemelha a esse princípio, o compromisso ontológico, consiste em uma base para precisão e conceituação no que se pretende representar na ontologia.

A adoção de ambos os princípios, podem aumentar a precisão da abrangência das classes e entidades e do recorte do domínio nos dois instrumentos. Tanto a garantia literária quanto o compromisso ontológico podem ser viabilizados através da perspectiva classificatória de Ranganathan. Onde a organização dos conceitos pode ser estruturada por categorias advindas da Teoria da Classificação Facetada, ou seja, nos conceitos estabelecidos pela estrutura do PMEST.

Baseando-se nos elementos e princípios relatados, é de extrema importância pontuar e avaliar em que fase do processo de construção de ontologias a conversão dos relacionamentos do tesouro pode ser aplicada.

Considera-se a etapa de conceitualização, onde a estrutura do domínio de conhecimento é traduzida em um modelo conceitual (TEIXEIRA; ALMEIDA, 2009). Pôde-se verificar no trabalho de Teixeira & Almeida (2009) que o referido modelo conceitual descreve soluções relativas ao vocabulário do domínio identificado na atividade de especificação. Relatam ainda que, o processo se dá pela construção de um glossário de termos, em seguida inclui-se os conceitos, verbos, instâncias e propriedades, o que faz identificar e reunir o conhecimento do domínio. Na sequência agrupam-se termos, indicando como exemplo os substantivos, candidatos a conceitos na ontologia, e os verbos, que representam ações no domínio e são candidatos a relações na ontologia (TEIXEIRA; ALMEIDA, 2009).

Por conseguinte, Teixeira & Almeida (2009) colocam a análise do vocabulário de termos, que deve considerar princípios genéricos e casos específicos. E por fim indicam a última tarefa da conceitualização, que é a criação de uma estrutura taxonômica dos termos reunidos, agrupando os que tiverem alguma similaridade. (TEIXEIRA; ALMEIDA, 2009).

Campos & Souza (2009) corroboram a questão de que “em métodos que se utilizam a conversão de tesouros para a construção de ontologias, a escolha das relações não pode ocorrer pela simples transposição de um suporte ao outro, mas prescinde de uma adequação ao domínio da ontologia, já que nos primeiros, as relações se limitam a três tipos” (CAMPOS; SOUZA, 2009).

Tendo em vista que há semelhança na origem das relações, porém há também diferença na tipologia das relações dos dois instrumentos, sugere-se então, uma reclassificação das mesmas nos processos de conversão entre tais sistemas. Ação que pode ajustar as informações que são mais estreitas na estrutura do Tesouro, e reaplicadas no modelo da ontologia, podendo ser ampliadas e readequadas com maior complexidade.

Ainda assim o processo necessita de verificação e validação manual por um ontologista, especialista no domínio em que fizer a aplicação. Isto porque, os tipos de relações nas ontologias são mais complexos e completos, enquanto nos tesouros, se limitam a relações de hierarquia, equivalência, associação e partitivas. E o bom aproveitamento da estrutura e relações ainda não é totalmente garantido no processo de convertibilidade.

Em virtude do que foi mencionado, pode-se afirmar que as relações conceituais presentes na estrutura dos Tesouros podem sim atender aquelas esperadas nas Ontologias, embora necessitem de algumas adaptações. Entretanto cabe mencionar a importância do tesouro estar com a estrutura de relacionamentos bem definidas. Fato que evidencia a importância de atender aos princípios metodológicos na construção do tesouro, justificando assim a análise da Norma ANSI/NISO Z39.19 como base para o presente estudo.

Portanto, acredita-se que os elementos identificados nessa etapa, assim como as características e particularidades de cada instrumentos, servem de subsídio para se avaliar a potencialidade da transposição dos relacionamentos e classes de um tesouro para a estrutura de uma ontologia.

A seguir, relatam-se os resultados obtidos na segunda etapa deste estudo.

4.1.2 Resultados – Etapa 2

Através da literatura selecionada, apresenta-se como resultados, o levantamento das principais experiências de conversão de tesouros em ontologias, a síntese e análise de seus métodos e consequências, tornando-se possível compreender como o aproveitamento da terminologia e relações semânticas do tesouro na construção de ontologias vem sendo praticado e, sobretudo, abordado na literatura.

Os critérios adotados para seleção das publicações já foram indicados no capítulo dos procedimentos metodológicos. Para a análise de conteúdo utilizou-se como definição das unidades de contexto, os títulos, resumos, palavras-chave e métodos desenvolvidos relatados em cada um dos documentos.

Apresentam-se a seguir, de forma sequencial, os quatro documentos selecionados e a interpretação de cada experiência:

4.1.2.1 Experiência Agrovoc

SOERGEL, Dagobert, et al. Reengineering thesauri for new applications: the Agrovoc example. *Journal of Digital Information*, v.4, n.4, 2004.

Título	Reengenharia de tesouros para novas aplicações: o exemplo do AGROVOC
RESUMO	Um problema nos esquemas de classificação e tesouros existentes é a falta de consistência estrutural e semântica bem definida. Capacitar os usuários finais em busca de coleções com desempenho superior a uma pesquisa simples de texto livre (como usado em muitos motores de busca da Web) e desenvolver sistemas que não só encontram, mas também processam a informação para a ação, requer um SOC muito mais poderoso e mais complexo. O presente documento apresenta uma estrutura conceitual e o processo de transição para apoiar a mudança de um tradicional SOC para um SOC semanticamente rico. A estrutura proposta também é compatível com outras abordagens de interoperabilidade como FTR e XML em ambiente Web. O AGROVOC, um tesouro tradicional desenvolvido e mantido pela <i>Food and Agriculture Organization (FAO)</i> das Nações Unidas, serviu como um estudo de caso para explorar a reengenharia de um tesouro tradicional em uma ontologia de pleno direito. Iniciou-se o processo de desenvolvimento de um inventário dos tipos de relações específicas com uma semântica bem definida para o domínio agrícola, para explorar a abordagem de regras que podem agilizar o processo de reengenharia (SOERGEL, 2004).
Palavras-chave	Reengenharia; Tesouro; Agrovoc; Ontologia.
Método apresentado	Inventário completo de tipos de domínio relevantes de entidade e tipos de relações e passos para a reengenharia.

Na literatura da Ciência da Informação, esta experiência é a mais citada e referenciada. Apresenta uma estrutura conceitual e procedimentos de transição para apoiar a mudança de um SOC tradicional para um mais completo nas relações semânticas. A estrutura proposta também se refere às abordagens de interoperabilidade com RDFS e XML no ambiente web (SOERGEL, 2004).

Com base nas ideias apresentadas no artigo selecionado, notou-se que realizar a reengenharia de AGROVOC em um sistema de ontologia, que servirá também as funções de uma enciclopédia tradicional. Isso envolve a criação de um inventário completo de tipos de domínio relevantes de entidade e tipos de relação, que vamos basear em uma análise mais aprofundada de vocabulários AGROVOC e afins e sobre estoques existentes, tais como a Rede UMLS Semantic.

O AGROVOC trata-se de um tesouro tradicional desenvolvido e mantido pela Organização para a Alimentação e Agricultura (FAO) das Nações Unidas. Serviu como um estudo experimental para a reengenharia de um tesouro em uma ontologia. Constatou-se que o processo iniciou-se com um inventário dos tipos de relacionamentos semânticos, específicos para o domínio agrícola e com uma abordagem para explorar as regras *as-you-go* para a racionalização do processo de reengenharia (SOERGEL, 2004).

De acordo com Soergel (2004) as regras de *as-you-go* para reengenharia de um tesouro em uma ontologia implica refino das relações do tesouro, o que consiste em um processo trabalhoso. Em sua pesquisa relata que o processo de reengenharia foi estabelecido de acordo com os seguintes passos:

1. Definir a estrutura da ontologia
2. Preencher os valores (tipos de relações) de um ou mais SOC
3. Editar manualmente utilizando um editor de ontologias:
 - A informação existente mais precisa.
 - Adicionar novas informações.

Pode-se notar, que o método aplicado por Soergel (2004) não é totalmente automatizado, exigindo-se assim a interferência de um especialista para adequar algumas informações, tanto antes da conversão, quanto depois.

O modelo proposto permite que exista qualquer tipo de relacionamento entre os termos na estrutura da ontologia. Enfatiza que é possível introduzir relações entre termos que reflitam com mais precisão a realidade da linguagem do que o mapeamento de termos para os "conceitos". De acordo com o autor, estas duas representações de informações conceituais podem coexistir dentro do mesmo sistema (SOERGEL, 2004).

Com base nos resultados da exploração da aplicabilidade da abordagem do exemplo AGROVOC, pretendeu-se desenvolver um sistema que agilizasse o processo de reengenharia. Nota-se que o experimento almejou a redução no esforço necessário para construção de uma ontologia, e, também, tornar essa reengenharia viável para futuras aplicações.

A pesquisa foi realizada de forma gradual, começando com o teste de conceitos através da aplicação piloto de forma a demonstrar como os benefícios esperados resultarão na prática. O referido teste deu subsídios para o planejamento e implementação eficiente em grande escala para reestruturar o AGROVOC em uma ontologia de pleno direito. Como relata Soergel (2004), esperou-se que o esforço pudesse ser realizado em um ambiente cooperativo, de forma a promover a participação ativa e, finalmente, o uso de um sistema amplamente aceito para o domínio da alimentação e agricultura (SOERGEL, 2004).

Enfim, o experimento de Soergel aponta que a base de conhecimento resultante permitirá (SOERGEL, 2004):

- melhora na interação do usuário com o vocabulário para a formulação de consultas e mais aprendizado sobre o domínio;
- apoio para expansão de consulta;
- suporte inteligente para indexadores humanos e automáticos na indexação e categorização dos sistemas;
- suporte para a inteligência artificial e as aplicações da Web semântica na área de agricultura, processamento de alimentos, e segurança alimentar.

4.1.2.2 Método para reengenharia de recursos não ontológicos

VILLAZON TERRAZAS, Borís Marcelo. *Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies*. 2011. 238 f. Tese (Doutorado em Inteligencia Artificial) – Facultad de Informática (UPM), Universidade Politécnica de Madrid, Madrid.

TÍTULO	Método para reutilização e reengenharia de recursos não-ontológicos para a construção de ontologias
RESUMO	Esta tese é focada sobre a reutilização e a possível reengenharia de recursos de conhecimento, ao contrário do costume de construir novas ontologias do zero. A análise profunda do estado da arte revelou que existem alguns métodos e ferramentas na literatura para transformar os recursos não ontológicos em ontologias, mas com algumas limitações: a maioria dos métodos apresentados é baseada em transformações de <i>ad hoc</i> para o tipo de recurso e a aplicação de recursos. Poucas aproveitam o modelo de dados de recurso, um artefato importante para o processo de reengenharia GGPSFVT08. Não há nenhum quadro integrado, de método ou ferramenta correspondente, que considera os tipos de recursos, modelos de dados e implementações identificadas de forma unificada. No que diz respeito a abordagem de transformação, a maioria utiliza métodos de transformação TBox, outros executam transformação de ABox. Quanto ao grau de automação, quase todos os métodos executam uma transformação semi-automática do recurso. De acordo com o esclarecimento da semântica oculta nas relações dos componentes do recurso, é possível afirmar que os métodos que executam uma transformação TBox tornam explícita a semântica nas relações dos componentes do recurso. A maioria desses métodos identifica relações <i>subClassOf</i> , outros identificam relações <i>ad hoc</i> e alguns <i>id</i> (VILLAZON TERRAZAS, 2011) .
PALAVRAS-CHAVE	Reengenharia; Recursos não ontológicos; Ontologia.
MÉTODO APRESENTADO	Padrão para conversão de Tesouros: <ul style="list-style-type: none"> • <i>Transformação Tbox</i> • <i>Transformação Abox</i>

A tese de Villazon Terrazas (2011) discorre sobre conversão de vários tipos de SOCs para o formato de ontologia. No capítulo específico para tesouros, analisa a definição de um tesouro, incluindo seus componentes. Em seguida, fornece uma definição formal para os tesouros e os modelos de dados identificados, bem como implementações para eles. Por fim, apresentam os padrões para reengenharia de tesouros em ontologias, incluindo os de TBox e transformação ABox. Tem-se como distinção dos dois padrões, a complexidade de tempo do algoritmo de transformação TBox que é polinomial, enquanto que o algoritmo de transformação ABox é linear (VILLAZON TERRAZAS, 2011).

Os dois padrões apresentados para a transformação de tesouros são (VILLAZON TERRAZAS, 2011):

- ***Transformação Tbox***

O padrão para a reengenharia de um tesouro com base um modelo de dados de ontologias. Impõe uma semântica formal para a reengenharia, mesmo na condição de mudar sua estrutura. Para explicitar a semântica de algumas relações associativas entre os termos conta-se com um recurso externo, WordNet.

- ***Transformação Abox***

O padrão para a reengenharia de um tesouro seguindo o modelo de dados e relações baseados em uma ontologia. Neste caso a semântica informal dos recursos praticamente intocada na reengenharia.

Villazon Terrazas (2011) ainda enfatiza que o esquema de um tesouro tem os seguintes componentes principais e indica como os mesmos serão utilizados na estrutura da ontologia:

- Termo Preferencial: será transformado em uma classe;
- Relação hierárquica: será transformada em uma relação *subClassOf*;
- Termos equivalentes: serão transformados em etiquetas;
- Finalmente, os descritores do tesouro são transformados em instâncias na ontologia.

Enfim, a tese propõe um método e seu suporte tecnológico para proporcionar a reengenharia de padrões, a fim de acelerar o processo de desenvolvimento de ontologias. Para atingir este objetivo, necessita-se das seguintes ações são (VILLAZON TERRAZAS, 2011):

- Definição de aspectos metodológicos relacionados com a reutilização de recursos não ontológicos para construção de ontologias;
- Definição da metodologia aspectos relacionados com a reengenharia de recursos não ontológicos para a construção de ontologias;
- Criação de uma biblioteca de padrões de reengenharia de recursos não ontológicos em ontologias;
- Desenvolvimento de uma biblioteca de software que implementam as sugestões fornecidas pelos padrões de reengenharia.

Tendo em mente estes objetivos, Villazon Terrazas (2011) apresentou a definição de recursos não ontológicos e propôs uma categorização deles de acordo com três recursos diferentes: tipo de recurso, modelo de dados e implementação. Ressalta ainda que esta categorização não é exaustiva nem completa. Contudo, argumenta que a solução atende as limitações identificadas no estado da arte da área.

4.1.2.3 Método de Reengenharia

KLESS, D., Jansen, L., LINDENTHAL, J., WIEBENSIHN, J.: *A Method of Re-Engineering a Thesaurus into an Ontology*. In: Donnelly, M., Guizzardi, G. (eds): *Formal Ontology in Information Systems - Proceedings of the Seventh International Conference (FOIS 2012)*, pp.133–146. IOS Press, Amsterdam (2012).

Título	Um método da reengenharia um dicionário de sinônimos em uma ontologia.
RESUMO	A construção de ontologias pode ser facilitada através da adaptação de vocabulários existentes. Há pouca clareza e de fato pouco consenso com relação à quais modificações de vocabulários são necessárias para reengenharia em ontologias. Neste trabalho apresenta-se um método que fornece passos claros a seguir quanto a reengenharia de um tesouro. O método faz uso de ontologias de alto nível e foi derivado das diferenças estruturais entre tesouros e ontologias, bem como das melhores práticas em modelagem, alguns dos quais foram defendidos no domínio biomédico. Ilustramos cada passo do método com exemplos de um estudo de caso de reengenharia sobre fertilizantes agrícolas com base no AGROVOC tesouro. O método deixa claro que a reengenharia de tesouros requer muito mais do que apenas uma conversão sintática em uma linguagem formal ou outras etapas facilmente automatizáveis. O método pode ser usado não só para reengenharia de tesouros, mas também resumir os passos para a construção de ontologias em geral e, portanto, pode ser adaptado para a reengenharia de outros tipos de vocabulários ou terminologias (KLESS, 2012).
Palavras-chave	Reengenharia; Tesouro; Ontologia;
Método apresentado	Etapas para reengenharia

Kless et al (2012) propõe um método que fornece passos claros para conduzir a reengenharia de um tesouro. Segundo o autor, o método foi derivado a partir das diferenças estruturais entre tesouros e ontologias, e também a partir de melhores práticas de modelagem. Ilustraram cada passo do método com exemplos de um estudo de caso de reengenharia AGROVOC (SOERGEL, 2004).

Segundo Kless et al (2012) o método deixa claro que reengenharia de tesouros requer muito mais do que apenas uma conversão sintática em uma linguagem formal ou outros passos automatizados. Propõe ainda que o método pode ser adaptado para a reengenharia de outros tipos de vocabulários ou terminologias. (KLESS et al, 2012).

As etapas do método proposto são (KLESS et al, 2012):

1. Refinamento de preparação e verificação do tesauro
2. Conversão sintática
3. Definição de prazo / análise da propriedade (em linguagem natural)
4. Escolha e alinhamento de ontologias de nível superior e de relações formais
5. Especificação formal de classes
6. Normalização
7. Dissociação de entidades independentes
8. Ajuste de pontuação, ortografia e outros aspectos de rótulos de classe e propriedade.

De acordo com Kless et al (2012) a experiência exige um esforço considerável, e o esforço se justifica quando uma padronização de alta precisão de terminologia é necessária, bem como onde o processamento automatizado de sistemas inteligentes é necessário (KLESS et al, 2012).

Os autores enfatizam que a reengenharia de tesouros em ontologias requer uma variedade de habilidades e conhecimentos relacionados à ontologia, assim como aqueles discutidos em sua construção. Assim, é desejável a criação de uma equipe cobrindo a expertise de domínio (KLESS et al, 2012).

É importante mencionar que, de acordo com Kless et al (2012) o método não foi testado e aperfeiçoado na escala de reengenharia completa, pois trata-se de uma tarefa extremamente demorada, dada a quantidade de conceitos. Dessa forma, relata que o método ainda poderá ser ajustado ou pode ser utilmente alterado com base na experiência de larga escala.

Os autores abordam que em trabalhos futuros, planeja-se uma análise mais detalhada das diferenças práticas entre tesouros e ontologias. E ainda, pretende-se verificar que tipo de artefato é mais apropriado para a aplicação do método (KLESS et al, 2012).

4.1.2.4 Experiência tesauro GEM

QIN, J.; PALING, S. Converting a controlled vocabulary into an ontology: the case of GEM. *Information Research*, v. 6, n. 2, 2001.

Título	Convertendo um vocabulário controlado em uma ontologia: o caso do GEM
RESUMO	O predomínio de informação digital levantou questões sobre a adequação das ferramentas convencionais de biblioteca para organizar informações. A multidimensionalidade de recursos digitais exige uma representação mais versátil e flexível para acomodar a representação e recuperação da informação. Ontologias são usadas como uma solução para esses problemas em muitos domínios de aplicação, principalmente devido à sua capacidade para especificar explicitamente a semântica e as relações e para expressá-los em uma linguagem compreensível por computador. Ferramentas convencionais de organização do conhecimento como classificações e tesouros, se assemelham a ontologias de uma forma que eles definem conceitos e relacionamentos de forma sistemática, mas são menos expressivos do que ontologias quando se trata de linguagem de máquina. Este trabalho utilizou o vocabulário controlado no Gateway de Materiais Educacionais (GEM) como um exemplo para resolver os problemas na representação de recursos digitais. O referencial teórico e metodológico neste trabalho serve como fundamentação e orientação para converter o vocabulário controlado GEM em uma ontologia. Comparado com o modelo semântico original do GEM, a principal diferença entre os dois modelos está nos valores adicionados através da semântica mais profunda em descrever objetos digitais, tanto conceitualmente e relacionalmente (QIN; PALING, 2005).
Palavras-chave	Conversão; Tesauro; Ontologia; Gateway de Materiais Educacionais (GEM).
Método apresentado	Planejamento preliminar e etapas envolvidas na conversão do vocabulário GEM existente para uma ontologia.

A pesquisa utilizou o vocabulário controlado Gateway de Materiais Educacionais (GEM) como um exemplo para resolver os problemas na representação de recursos digitais. Os autores enfatizam que o referencial teórico e metodológico (GRUBER, 1993; SOERGEL, 1999; GUARINO, 1997) utilizado no trabalho serve como fundamentação e orientação para converter o vocabulário controlado GEM em uma ontologia. Comparado com o modelo original semântica do tesouro GEM, a principal diferença entre os dois modelos está nos valores adicionados através da semântica mais profunda em descrever objetos digitais, tanto conceitualmente e relacionalmente (QIN; PALING, 2005).

O estudo discute ainda o planejamento preliminar e etapas envolvidas na conversão do vocabulário GEM existente para uma ontologia. E indica que o objetivo da conversão é não apenas para reduzir a duplicação de esforço envolvido na construção de uma ontologia a partir do zero, usando o vocabulário existente, mas também para estabelecer um mecanismo para permitir que diferentes vocabulários sejam mapeados para a ontologia (QIN; PALING, 2005).

De acordo com Quin e Paling (2005) na transposição dos dados utiliza-se o sistema de *Ontolingua*, que fornece um ambiente de desenvolvimento para construção de ontologias, além de ferramentas de edição, que permite a criação de ontologias modulares e combináveis (QIN; PALING, 2005).

O trabalho aborda que a ontologia deve ter os seguintes valores adicionados em comparação com o vocabulário existente (QIN; PALING, 2005):

1. Níveis mais elevados de concepção do vocabulário descritivo;
2. Semântica mais profunda para a classe/subclasse e entre relacionamentos de classe;
3. Capacidade de expressar os conceitos e relações em uma linguagem de descrição;
4. Reutilização das construções ontológicas em sistemas heterogêneos.

De forma a ilustrar sinteticamente as propostas, objetivos e métodos aqui analisados, fez-se um quadro, Quadro 8 – Síntese comparativa das iniciativas e metodologias de conversão, de forma a sintetizar e facilitar a comparação entre os quatro tipos de experiências e iniciativas.

Quadro 8 – Síntese comparativa das iniciativas e metodologias de conversão

	EXPERIÊNCIA AGROVOC	MÉTODO PARA REENGENHARIA DE RECURSOS NÃO ONTOLÓGICOS	MÉTODO DE REENGENARIA DE TESAuros	EXPERIÊNCIA TESAuro GEM
DOCUMENTO SELECIONADO	SOERGEL, Dagobert, et al. Reengineering thesauri for new applications: the Agrovoc example. <i>Journal of Digital Information</i> , v.4, n.4, 2004.	VILLAZON TERRAZAS, Boris Marcelo. <i>Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies</i> . 2011. 238 f. Tese (Doutorado em Inteligência Artificial) – Facultad de Informática (UPM), Universidade Politécnica de Madrid, Madrid.	KLESS, D., Jansen, L., LINDENTHAL, J., WIEBENSIHN, J.: A Method of Re-Engineering a Thesaurus into an Ontology. In: Donnelly, M., Guizzardi, G. (eds): (FOIS 2012), pp.133–146. IOS Press, Amsterdam (2012).	QIN, J.; PALING, S. Converting a controlled vocabulary into an ontology: the case of GEM. <i>Information Research</i> , v. 6, n. 2, 2001.
INTRODUÇÃO	Apresenta uma estrutura conceitual e procedimentos de transição para apoiar a mudança de um SOC tradicional para um SOC que contenha mais complexidade nas relações semânticas. A proposta está em conformidade com as abordagens de interoperabilidade com RDFS e XML no ambiente web. O AGROVOC é um tesouro desenvolvido e mantido pela Organização para a Alimentação e Agricultura (FAO) das Nações Unidas. Serve como um estudo de caso para explorar a reengenharia de tesouro tradicional em uma ontologia.	No capítulo específico para tesouros, analisa a definição de um tesouro, incluindo seus componentes. Em seguida, fornece uma definição formal para os tesouros e os modelos de dados identificados, bem como implementações para eles. Por fim, apresentam os padrões para reengenharia de tesouros em ontologias, incluindo os de TBox e transformação ABox. Tem-se como distinção dos dois padrões a complexidade de tempo do algoritmo de transformação: TBox que é polinomial, enquanto que o algoritmo de transformação ABox é linear.	Apresentam um método que fornece passos para a reengenharia de um tesouro, derivado a partir das diferenças estruturais entre tesouros e ontologias, bem como a partir de práticas de modelagem. Cada passo do método foi ilustrado com exemplos de um estudo de caso do AGROVOC. O método deixa claro que reengenharia tesouros requer muito mais do que apenas uma conversão sintática em uma linguagem formal. O esforço se justifica quando uma padronização de alta precisão da terminologia é necessária, bem como onde o processamento automatizado de sistemas inteligentes. Tal conhecimento não está à disposição dos não-especialistas. Assim, é desejável a criação de uma equipe cobrindo a expertise de domínio.	O estudo discute o planejamento preliminar e etapas envolvidas na conversão do vocabulário GEM para uma ontologia. O objetivo da conversão não é apenas para reduzir a duplicação de esforço envolvido na construção de uma ontologia a partir do zero, mas também para estabelecer um mecanismo para permitir que diferentes vocabulários possam ser mapeados para a ontologia. Uma questão é apontada para conversão de um vocabulário controlado em uma ontologia: Qual é o valor acrescentado de uma ontologia em comparação com os vocabulários de representação de conhecimento utilizadas em bibliotecas e demais informações?
METODOLOGIA DESENVOLVIDA	O processo desenvolveu-se como um inventário dos tipos de relacionamento específicos para o domínio agrícola. O processo de reengenharia: as regras da reengenharia de um tesouro em uma ontologia implicam refinamento das relações de sinônimos. Os passos do processo são os seguintes: 1. Definir a estrutura da ontologia 2. Preencher os relacionamentos de um ou mais SOC na medida do possível 3. Editar manualmente utilizando um editor de ontologias: <ul style="list-style-type: none"> • Informação existente mais precisa • Adicionar novas informações 	Para Tesouros apresentam dois padrões: <ul style="list-style-type: none"> • Transformação Tbox O padrão para a reengenharia de um tesouro com base em modelo de dados em um esquema de ontologias. Impõe uma semântica formal para a re-engenharia recursos, mesmo à custa de mudar sua estrutura. Para explicitar a semântica da BT, NT e as relações entre os termos RT Thesaurus, conta com um recurso externo, o WordNet. Para as relações UF / USO usa-se o padrão lógico, a SOE. <ul style="list-style-type: none"> • Transformação Abox O padrão para a reengenharia de um tesouro seguindo o modelo de dados e relações baseados em uma ontologia. Deixa a semântica informal dos recursos re-engenharia praticamente intocada.	O método pode ser usado não apenas para reengenharia tesouros, e também resume os passos para a construção de ontologias, em geral, e pode, portanto, ser adaptado para a reengenharia de outros tipos de vocabulários ou terminologias. <ol style="list-style-type: none"> 1. Refinamento de preparação e verificação do tesouro 2. Conversão sintática 3. Definição de prazo / análise da propriedade (em linguagem natural) 4. Escolha e alinhamento de ontologias de nível superior e de relações formais 5. Especificação formal de classes 6. Normalização 7. Dissociação de entidades independentes 8. Ajuste de pontuação, ortografia e outros aspectos de rótulos de classe e propriedade 	Foi utilizado o sistema de Ontolingua para criar a ontologia. Esse sistema fornece um ambiente de desenvolvimento para construção de ontologias. Além de ferramentas de edição, que permite a criação de ontologias modulares, combináveis. Uma vez que as facetas são relativamente independentes, não vão ser ligados por um grande número de axiomas. Tomando os critérios acima em conta, a ontologia deve ter os seguintes valores adicionados em comparação com o vocabulário existente: <ol style="list-style-type: none"> 1. Níveis mais elevados de concepção de vocabulário descritivo; 2. Semântica mais profunda para a classe / subclasse e entre classes relacionamentos; 3. Capacidade de expressar tais conceitos e relações em uma linguagem de descrição e 4. Reutilização e "partes-capacidade" das construções ontológicas em sistemas heterogêneos.

RESULTADOS	<p>Com base nos resultados encorajadores da exploração da aplicabilidade das normas as-you-go para AGROVOC, pretendeu-se desenvolver um sistema que agilizasse o processo de reengenharia. Espera-se que este sistema reduza o esforço necessário e, assim, fazer o esforço de reengenharia viável. A base de conhecimento resultante permitirá:</p> <ul style="list-style-type: none"> • interação melhorada do usuário com o vocabulário para a formulação de consultas e aprendizado sobre o domínio; • apoio para expansão de consulta; • Suporte inteligente para indexadores humanos e automáticos de indexação / categorização dos sistemas; • Suporte para a inteligência artificial e as aplicações da Web semântica para a agricultura de processamento de alimentos, e segurança alimentar. 	<p>Nesta tese fornecem um método e seu suporte tecnológico que contar com reengenharia de padrões, a fim de acelerar o processo de desenvolvimento de ontologias pela reutilização e re-engenharia, tanto quanto possível disponível para recursos não ontológicos. Para atingir este objetivo geral, relatam os seguintes objetivos: (1) a definição de aspectos metodológicos relacionados com a reutilização de recursos para construção de ontologias, (2) a definição dos aspectos metodológicos relacionados com a reengenharia de recursos não-ontológicos para a construção de ontologias; (3) a criação de uma biblioteca de padrões de reengenharia e (4) o desenvolvimento de uma biblioteca de software que implementa as sugestões fornecidas pelos padrões de reengenharia. Introduzem a definição de recursos não-ontológicos e propõem uma categorização deles de acordo com três recursos diferentes: tipo de não-ontológica de recursos, modelo de dados e implementação.</p>	<p>O método proposto é específico para tesouros, mas podem ser adaptados a outros tipos de vocabulários estruturados, tais como os sistemas de classificação. Os passos podem ser considerados passos necessários para a construção de ontologias. Apenas a especificação de classes fornece espaço para as diferenças qualitativas e não podem ser desenvolvidos métodos de avaliação. O método não foi testado e aperfeiçoado na escala de reengenharia completa. Esta é uma tarefa extremamente demorada de vários anos de trabalho, dado o tamanho dos tesouros. O método poderá ser ajustado ou pode ser utilmente alterado com base na experiência de larga escala.</p>	<p>A discussão revela uma intrínseca relação de compromissos ontológicos com tesouros e classificações. Apresenta um conjunto de critérios para orientar o desenvolvimento de ontologias:</p> <ul style="list-style-type: none"> • <i>Clareza</i>: Definição deve ser objetiva, livre de contexto social ou computacional, e documentada com linguagem natural. • <i>Coerência</i>: Os axiomas devem ser logicamente consistentes. • <i>Extensibilidade</i>: A ontologia deve antecipar o uso do vocabulário compartilhado. • <i>Viés de codificação mínima</i>: ocorre quando da escolha da representação é feita simplesmente para a conveniência de notação. A conceituação deve ser especificada em que nível de conhecimento, sem depender de uma codificação particular. • <i>Compromisso ontológico mínimo</i>: quando as ações estão de acordo com as definições da ontologia, que deve fazer um número mínimo de reivindicações sobre o mundo que está sendo modelado (Gruber, 1993).
CONCLUSÕES	<p>A abordagem será gradual, começando com o fornecimento de uma prova de conceito através de uma aplicação piloto que demonstra que os benefícios esperados resultarão de fato. Isto fornece a base para o planejamento e implementação eficiente do esforço em grande escala para reestruturar AGROVOC em uma ontologia de pleno direito. Este esforço pode ser realizado em um ambiente cooperativo, distribuída e coordenada para promover a participação ativa e, finalmente, o uso de um sistema amplamente aceito Conhecimento Organização para o domínio da alimentação e agricultura.</p>	<p>Argumenta que a solução atende as limitações identificadas no estado da arte da área. O esquema de um tesouro tem os seguintes componentes principais:</p> <ol style="list-style-type: none"> (1) Termo Preferencial, que irá ser transformada em uma classe, (2) Relação hierárquica, que irá ser transformada em uma relação <i>subClassOf</i>, (3) Relação lógica, que irá ser transformada em uma relação <i>ad-hoc</i>, (4) Termos equivalentes, os termos das relações de uso, que serão transformados em etiquetas, usando o padrão lógico SOE, proposto por Corcho et al. <p>Finalmente, o conteúdo do tesouro será transformado em instâncias ontologia.</p>	<p>O método revela que há muito mais ações a serem feitas na reengenharia do que uma conversão simples sintática. Cria uma maior sensibilidade para distinguir ontologias de vocabulários, por vezes referido como terminologias. Contribui para uma melhor compreensão do que a construção de ontologias significa e conclui que uma terminologia é uma boa base para a construção de uma ontologia. No processo de reengenharia, é a utilização de lógica de modelação orientada com condições necessárias e suficientes que impõe um nível mais elevado de precisão. Modelagem ontológica desafia e provoca repensar de contradições que permanecem desconhecidos no vocabulário e terminologia. Planeja-se uma análise detalhada e comparação das diferenças práticas entre tesouros e ontologias. Procura-se que tipo de artefato de informação, é apropriado, a fim de saber para qual esforço de reengenharia um tesouro em uma ontologia é justificada.</p>	<p>O referencial teórico e metodológico descrito neste trabalho serviu como fundamentação e orientação para converter tesouro GEM em uma ontologia. Comparado com o modelo original semântica do GEM, a principal diferença entre os dois modelos está nos valores adicionados através da semântica mais profundas em descrever objetos digitais, tanto conceitualmente e relacionalmente. Ao contrário da maioria de esforço ontológico que não fez nenhuma referência a vocabulários controlados criados pelos bibliotecários e cientistas da informação, este estudo reconhece a importância de conectar estruturas de conhecimento existentes, bem como a dificuldade em implementar tais estruturas em ontologias.</p>

Fonte: Elaborado pela autora baseado em (SOERTEL, 2004) (VILLAZON TERRAZAS, 2011) (KLESS; LINDENTHAL; WIEBENSIHN, 2012) (QUIN; PALING, 2001).

Buscando atender ao último objetivo dessa pesquisa, que consiste em apontar a validade e potencial da conversão com base na análise e sistematização dos dados bibliográficos levantados, com o propósito de abstrair as possibilidades e os limites para a efetivação da referida conversão, utilizou-se para esse resultado conclusivo, as inferências obtidas nas etapas 1 e 2 mencionadas anteriormente. Onde, tanto a consulta às normas de construção quanto às experiências relatadas na literatura da Ciência da Informação, dão embasamento para concluir que a transposição dos dados de um tesauro para a construção de uma ontologia é válida sim. Contudo, o que se pode notar, é que a temática ainda é pouco discutida e validada. Os métodos propostos ainda estão no nível de experimentação, e são pouco aprofundados.

Dessa forma, as metodologias apresentadas nas experiências analisadas carecem de detalhamento sobre as etapas e ações para garantir e certificar a transposição dos conceitos sem se perder o compromisso ontológico esperado pela construção de ontologias.

Em virtude da análise encontrada nas duas etapas relatadas, nota-se que tradicionalmente os tesouros são utilizados tanto para controlar e padronizar as linguagens de indexação, quanto para facilitar a realização de consultas pelos usuários. Contudo apesar de permitirem a construção de estruturas lógico-semânticas flexíveis de relacionamentos entre conceitos, os tesouros possuem certas limitações quanto às possibilidades de exploração de determinados conteúdos informacionais, uma vez que não têm suficientemente expressividade que permita representar relações semânticas mais ricas entre os conceitos (BOCCATO; RAMALHO; FUGITA, 2008).

A partir da determinação da amostra selecionada para a análise de conteúdo, pode-se afirmar que, com base no estudo feito sobre a relação dos SOC com os princípios teóricos que orientam a construção de tesouros e ontologias, na área de Ciência da Informação, as ontologias representam o lado objetivo da realidade e requerem um documento que defina formalmente as relações entre os termos (GOMES, 2009). Já os tesouros empregam os termos e apenas indicam a existência de relações entre elas por meio dos conceitos.

5 CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

No intuito de apontar as considerações finais, retoma-se aos pontos principais da pesquisa de forma a avaliar os resultados encontrados, considerando os objetivos propostos e as evidências teóricas que serviram de base para o percurso e condução do estudo. Por fim, serão apresentadas sugestões e expectativas de trabalhos futuros.

Dessa forma, faz-se necessário retornar às questões propostas na introdução deste trabalho, que motivaram o desenvolvimento da pesquisa, de forma a verificar em que medida as mesmas foram atendidas. A saber:

- Quais princípios teóricos e metodológicos, discutidos e apresentados na literatura da área da Ciência da Informação, norteiam o processo de conversão de tesouros em ontologias?
- Quais práticas e metodologias conduzem e potencializam a conversão de estruturas e relações conceituais existentes no tesouro para a construção de ontologias?
- Qual o aproveitamento e potencialidade em se converter um tesouro em uma ontologia?

Para conduzir às respostas das questões apresentadas, o objetivo geral buscou apresentar com base na literatura científica, as discussões e metodologias relativas ao processo de conversão de tesouros em ontologias, visando introduzir a temática nas pesquisas de âmbito nacional e contribuir para a melhoria das práticas e métodos de construção de ontologia.

Portanto, buscou-se aqui, apontar a validade e potencial entre as relações conceituais, e as possibilidades e limites de se converter um Tesouro em uma Ontologia, diante do que se abstraiu na literatura e na análise das experiências, respectivamente observados nas Etapas 1 e 2.

De acordo com a descrição dos objetivos a aplicação dos procedimentos metodológicos as etapas mencionadas consistiram em:

- **Etapa1** – Estudo de uma Norma de construção de tesouros e de um padrão de construção de ontologias de alto nível
- **Etapa 2** - Análise de publicações e experiências relativas a conversão do tesouro

Ainda teve-se como propósito da pesquisa apontar a validade e potencial da conversão com base na análise dos dados bibliográficos levantados, a fim de abstrair as possibilidades e os limites para a efetivação da referida conversão.

Para tanto foi necessário buscar de forma específica a identificação na literatura do campo da CI, características relativas ao processo de construção de tesouros e ontologias, a fim de apreender quais relações presentes na estrutura do tesouro atendem a estrutura esperada numa ontologia. Buscou-se ainda, a identificação de métodos e resultados encontrados em experiências de conversão de tesouros e ontologias, com a finalidade de compreender quais práticas levam ao aproveitamento da terminologia e relações semânticas do tesouro na ontologia.

Os resultados, encontrados em análises e inferências, apontam que a condução do processo de conversão de tesouros em ontologias pode ter como princípios norteadores as indicações de construção dos relacionamentos, tanto nos manuais e padrões de construção de tesouros, quanto nos de ontologias.

Encontrou-se nos padrões de construção analisados e na literatura geral consultada elementos que evidenciam as semelhanças e aproximações na natureza dos relacionamentos, embora a tipologia e estrutura dos relacionamentos mostrem-se diferentes. Questão essa que se fez refletir a potencialidade em se converter a estrutura de um tesouro para a de uma ontologia.

Nessa perspectiva, o que se considera nesse estudo é que pelo fato da natureza das relações e entidades serem fundamentadas de forma semelhante, pode-se sim haver aproveitamento na transposição de um instrumento para o outro. O que se conclui é que, há a necessidade de uma revisão por um especialista para se fazer as adequações dos termos e relacionamentos que não atenderem ao objetivo da ontologia.

Dessa forma, o que se evidencia, é que tal iniciativa, da conversão de tesouros em ontologias, configura-se primariamente em um aproveitamento de terminologia. Tendo em vista que na construção de ontologias há uma fase de levantamento terminológico.

Algumas diretrizes e metodologias foram propostas como relatado na segunda etapa do trabalho, a fim de minimizar os problemas de construção de ontologias, inclusive algumas incluem o estudo e análise do domínio como parte de sua estratégia. Nesse contexto, pode-se aplicar a conversão dos relacionamentos dos tesouros, com o propósito de levantar a terminologia e relações conceituais de um domínio, a fim de facilitar o processo de levantamento terminológico e, portanto, garantir o compromisso ontológico no domínio trabalhado.

No que se refere às experiências consultadas, algumas inconsistências na amostra para a análise de conteúdo foram encontradas. Nesse aspecto notou-se que as metodologias de conversão de tesouro carecem de detalhamento sobre as etapas e ações a serem executadas.

A forma para se conduzir a conversão de estruturas e relações de um instrumento para o outro foi analisada através de quatro experiências como proposto na Etapa 2 dos procedimentos metodológicos.

Considera-se que a conversão de tesouros em ontologias, é um ponto de partida para a construção das relações na ontologia. Dessa forma, evidencia-se que não há aproveitamento pleno das relações, ocasionando na dependência de um processo de revisão e conferência por um especialista do domínio.

Em virtude dos fatos mencionados, e após o percurso teórico e analítico desenvolvido no decorrer da pesquisa, mostra-se que é benéfico sim converter um tesouro em uma ontologia. Contudo, algumas fragilidades nos padrões e metodologias investigados, mostram que o processo de convertibilidade de tesouros requer muito mais do que apenas uma conversão sintática em uma linguagem formal ou outros passos facilmente automatizados.

Diante das inferências advindas desse estudo, considera-se que a construção de ontologias pode ser facilitada através da adaptação de vocabulários existentes. Os vocabulários controlados não padecem do mesmo problema de falta de consenso de metodologias para construção de ontologias.

Acredita-se que, ao identificar nas teorias e normas relacionadas a construção de tesouros e de ontologias, os critérios norteadores desta análise de conversão das relações conceituais, tornaram possível a realização desse estudo. Tomando como base as características que aproximam e distanciam ambos os instrumentos, permitindo alcançar respostas às perguntas por ora indicadas nessa pesquisa.

Nota-se que o estudo proporciona uma contribuição para as pesquisas na área de Ciência da Informação no Brasil, principalmente pelo fato da temática ainda ser pouco discutida por pesquisadores brasileiros.

No intuito de continuar com estudos desta temática, o trabalho sugere para estudos futuros uma investigação mais exaustiva sobre as metodologias de convertibilidade e inclusive uma abordagem a respeito de compatibilização de vocabulários. Pretende-se ainda expandir e dar continuidade a temática com projetos futuros no intuito de:

- Apontar métricas formais para a atividade de conversão de instrumentos de vocabulários controlados em modelos de ontologias.
- Propor um protótipo através da convertibilidade de um tesouro, aplicando ferramentas disponíveis para tal atividade e utilizando as teorias de compatibilização (DAHLBERG, 1983) e reconciliação (NEVILLE, 1970), de forma a atender e aperfeiçoar a fase de levantamento terminológico, avaliação de conteúdo e aquisição de conhecimento na construção de ontologias.
- Validar e estabelecer uma metodologia para conversão através da aplicação em um domínio do conhecimento.

- Verificar se o método não pode ser usado apenas para conversão de tesouros, ou se pode ser adaptado para a convertibilidade de outros tipos de vocabulários ou terminologias.
- Empreender método para reclassificação das relações que se diferem de um instrumento para o outro.

Em suma, almeja-se assim contribuir para o diálogo da Ciência da Informação e a Ciência da Computação, no sentido de automatizar e poupar etapas na construção de ontologias. Por fim, espera-se ter contribuído, através deste estudo exploratório, para se alcançar melhorias no suporte metodológico no processo de construção de ontologias, para sua atualização e, sobretudo, no auxílio da determinação e levantamento terminológico da ontologia.

REFERÊNCIAS

AGÜERA, J.R.P. *Automatización de tesauros y su utilización em la Web Semântica*. 2004. Disponível em: <http://www.w3.org/2001/sw/Europe/events/200406-esp/trabajo_finalextratesauros/trabajo-final-extratesauros.html>. Acesso em: 29 jun 2010.

ALMEIDA, Mauricio B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ci. Inf.*, v. 32, n.3, p. 7-20, 2003.

ALMEIDA, Maurício Barcellos. Uma iniciativa interinstitucional para construção de ontologia sobre Ciência da Informação: visão geral do Projeto P.O.I.S.. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, n. 19, p. 53-72, 2005. Disponível em: <<http://redalyc.uaemex.mx/redalyc/pdf/147/14701905.pdf>>. Acesso em: 10 mar. 2012.

ALMEIDA, Maurício Barcellos; TEIXEIRA, Lívia M.D.; COELHO, Kátia Cardoso; SOUZA, Renato R. Relações semânticas em ontologias: estudo de caso do Blood Project. *Liinc em Revista*, Rio de Janeiro, v. 6, n. 2, p. 384-410, jun. 2010. Disponível em: <<http://revista.ibict.br/liinc/index.php/liinc/article/view/362/250>>. Acesso em: 12 jun. 2011.

ALVARENGA, Lídia. Representação do conhecimento na perspectiva da Ciência da informação em tempo e espaços digitais. *Enc. Bibli: R. Electr. Bibliotecon. Ci. Inf.*, Florianópolis, n. 15, 2003. Disponível em: <http://www.encontros-bibli.ufsc.br/Edição_15/alvarenga_representação.pdf>. Acesso em: 10 set. 2010.

_____. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. *Datagramazero: R. Ci. Inf.*, v. 2, n. 6, dez. 2001. Disponível em: <http://www.dgz.org.br/dez01/Art_05htm>. Acesso em: 10 set. 2010.

ALVARES, Lílian. *Tesauro* (ppt). Brasília: UNB, [s.d.]. Disponível em: <http://www.alvarestech.com/lillian/Analise/Modulo3/Aula31Tesauros.pdf>. Acesso em: 12 de janeiro de 2013.

ANSI/NISO. Z39.19 – 2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda: NISO Press, 2005. 184 p. Disponível em: <<http://www.niso.org>>. Acesso em: 03 fev. 2012.

ASSOCIAÇÃO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO – ANCIB. *GT 2: Organização e Representação do Conhecimento*. Disponível em: <<http://www.ancib.org.br/grupos-de-trabalho/gt-2-organizacao-representacao-do-conhecimento/>>. Acesso em: 22 nov. 2010.

BARBOSA, Alice Príncipe. Classificações facetadas. *Ciência da Informação*, Rio de Janeiro, v. 1, n. 2, p. 73-81, 1972.

BARDIN, Laurence. *Análise de conteúdo*. São Paulo: Edições 70, 2011.

BEMERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. *Scientific American*, 2001, May. Disponível em <<http://www.scian.com/2001/0501issue/0501berners-lee.html>>, Acesso em 20 set. 2009.

BRASCHER, Marisa. *Tesouro, taxonomia e ontologia: uma evolução?* Ciclo de conferências sobre Organização da Informação Legislativa e Jurídica. Brasília, 2009. Disponível em: <http://projeto.lexml.gov.br/args/BRASCHER.pdf>. Acesso em: 29 jun. 2010.

BOCCATO, V. Os sistemas de organização do conhecimento nas perspectivas atuais das normas internacionais de construção. *InCID: Revista de Ciência da Informação e Documentação*, Ribeirão Preto, v. 2, n. 1, p. 165-192, jan./jun. 2011. Disponível em: <http://revistas.ffclrp.usp.br/incid/article/view/44>. Acesso em: 24 Out. 2012.

BOCCATO, V. R. C.; RAMALHO, R. A. S.; FUJITA, M. S. L. *A contribuição dos tesouros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais*. In: García Marco, F. J. (Ed.). *Avances y perspectivas en sistemas de información y documentación - IBERSID*, 2008. Zaragoza: Universidad de Zaragoza, 2008. p. 199-209.

CAFÉ, Ligia ; BRATFISH, Aline . Classificação analítico-sintética: reflexões teóricas e aplicações. *Transinformação*, v. 19, p. 237-250, 2007.

CAFÉ, Ligia ; BRASCHER, M. . Organização do Conhecimento: teorias semânticas como base para estudo e representação de conceitos. *Informação & informação* (UEL. Online), v. 16, p. 25-51, 2011.

CAFÉ, Lígia; MENDES, Fernanda. Uma contribuição para a construção de instrumentos analítico-sintéticos de representação do conhecimento. *Perspect. ciênc. inf.*, Dez 2008, vol.13, no.3, p.40-56. ISSN 1413-9936.

CAMPOS, Julia Aparecida Gonçalves; SOUZA, Renato Rocha UNIVERSIDADE FEDERAL DE MINAS GERAIS. *Análise conceitual sobre as relações semânticas em ciência da informação: contribuições para o desenvolvimento de ontologias*. 2009. 135 f., enc. Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CAMPOS, L.; CAMPOS, M.; CAMPOS, M.. Diretrizes para definição de domínio no reuso de ontologias biomédicas: uma abordagem baseada na análise do compromisso ontológico. *XI ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação*, Brasil, out. 2010. Disponível em: <<http://enancib.ibict.br/index.php/xi/enancibXI/paper/view/163>>. Data de acesso: 17 Jan. 2013.

CAMPOS, M. L. A. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: EdUFF, 2001.

CAMPOS, M. L. A.; GOMES, H.E.; MOTTA, D.F. *Tutorial de Tesauro*. 2004. Disponível em: <http://www.conexaorio.com/bitit/tesauro>. Acesso em: 30 nov 2010.

CAMPOS, M. L. A.; GOMES, H. E. Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n. 3, p.348-359, set./dec. 2006.

CAMPOS, Maria Luiza de Almeida. GOMES, Hagar Espanha; Tesauro e normalização terminológica: o termo como base para intercâmbio de informações. *DataGramaZero*, v.5, n.6, dez. 2004. Disponível em <http://www.dgzero.org/dez04/Art_02.htm>. Acesso em: 01 jul. 2012.

CAMPOS, Maria Luiza de Almeida; MARCONDES, Carlos Henrique. Ontologia e web semântica: o espaço da pesquisa em Ciência da Informação. *Ponto de Acesso*, Salvador, v.2, n.1, p. 107-136, jun./jul. 2008.

CAMPOS, M. L. A. (1994) *Em busca de princípios comuns na área de representação da informação: uma comparação entre o Método de Classificação Facetada, o Método de Tesauro-Baseado-em-Conceito e a Teoria Geral da Terminologia*. Rio de Janeiro: UFRJ/IBICT. 196 p. (Dissertação de Mestrado).

CINTRA, A. M. M. et al. *Para entender as linguagens documentárias*. 2.ed. São Paulo: Polis, 2002. 92p. (Coleção Palavra-chave, 4).

CURRÁS, Emília. *Ontologias, taxonomias e tesauros em teoria de sistemas e sistemática*. Brasília: Thesaurus, 2010. 182 p.

_____, Emília. *Tesauros: linguagens terminológicas*. Brasília: IBICT, 1995. 286 p.

DAHLBERG, Ingetraut. Teoria do conceito. *Ciência da Informação*, Rio de Janeiro, v.7, n. 2, p. 101-107, 1978a.

_____, Ingetraut. Toward establishment of compatibility between indexing languages. *International Classification*, Frankfurt, v.8, n.2, p.86-91, Apr./June. 1981.

DEMO, Pedro. *Metodologia científica em ciências sociais*. São Paulo: Pioneira, 1989.

DODEBEI, V.L.D. *Tesauro: linguagem de representação da memória*. Niterói: Intertexto; Rio de Janeiro: Interciências, 2002.

GARCIA JIMÉNEZ, A. Instrumentos de representación del conocimiento: tesauros versus ontologias. *Anales de Documentación*, [S. l.], n. 2, p. 79-95, 2004. Disponível em <<http://www.um.es/fccd/anales/ad07/ad0706.pdf>>. Acesso em: 10 jan. 2008.

GIL, Antonio Carlos. *Como elaborar projetos de pesquisa*. São Paulo: Atlas, 1991.

GOMES, Hagar Espanha. Tendências da pesquisa em organização do conhecimento. *Tendências da Pesquisa Brasileira em Ciência da Informação*, v. 2, n.1, 2009.

GOMES, Hagar Espanha; CAMPOS, Maria Luiza de Almeida; GUIMARÃES, Ludmila dos Santos. Organização da Informação e Terminologia: a abordagem onomasiológica. *DataGramaZero*, Rio de Janeiro, v. 11, n. 5, out. 2010. Disponível em: <http://dgz.org.br/out10/Art_03.htm>. Acesso em: 21 mai. 2011.

GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human and Computer Studies*, v.43, n. 5-6, p. 625-640, 1995.

GUARINO, N. Formal Ontology in Information Systems. *Proceedings of FOIS'98*, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15, 1998.

_____, N. Understanding, building and using ontologies. *In: PROCEEDINGS OF KNOWLEDGE ACQUISITION FOR KNOWLEDGE-BASED SYSTEMS WORKSHOP*. 10. 1996. Disponível em: <<http://ksi.cpsc.ucalgary.ca/KAW/KAW96/guarino/guarino.html#Heading4>>. Acesso em: 05 set. 2010.

GRENON, P.; SMITH, B. SNAP and SPAN: Towards Dynamic Spatial.Spatial *Cognition & Computation*, v.4, n.1, 2004, p.69-104. Disponível em http://Ontologyontology.buffalo.edu/smith/articles/SNAP_SPAN.pdf. Acesso: Fev. 2012.

GRUBER, T. (1996). *What is an ontology?* [S. l. : s. n.], 1996. Disponível em: <<http://ksl.stanford.edu/kst/what-is-an-ontology.html>> Acessado em: abril 2012.

HODGE, G. *Systems of Knowledge Organization for Digital Libraries*. Washington: Beyond Traditional Authority Files, 2000. The Digital Library Federation Council on Library and Information Resources. Disponível em: <http://www.clir.org/pubs/abstract/pub91abst.html>. Acesso em: 10 set. 2010.

HJØRLAND, B. What is Knowledge Organization (KO)? *Knowledge Organization: international journal devoted to concept theory, classification, indexing, and knowledge representation*, Frankfurt, v. 35, n. 2/3, p. 86-101, 2008.

KHOO, C.S.G.; NA, J. Semantic relations in Information Science. *Annual Review of Information Science and Technology*, v.40, p.157-228, 2006.

KLESS, D., Jansen, L., LINDENTHAL, J., WIEBENSIHN, J.: *A Method of Re-Engineering a Thesaurus into an Ontology*. In: Donnelly, M., Guizzardi, G. (eds): *Formal Ontology in Information Systems - Proceedings of the Seventh International Conference (FOIS 2012)*, pp.133–146. IOS Press, Amsterdam (2012)

LANCASTER, F. W. *El control del vocabulario en la recuperación de información*. 2. ed. Valencia: Universitat de València, 2002.

LARA, Marilda. Linguagem documentária e terminologia. *TRANSINFORMAÇÃO*, Brasília, DF, 16.3, 02 06 2005. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewarticle.php?id=72>>. Acesso em: 30 abr. 2010.

LIMA, G. Â. B. O. *Mapa hipertextual (MHTX) um modelo para a organização hipertextual de documentos*. 2004. 199f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

LUCAS, D.; SOUZA, R. R.; ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. *Ciência da Informação*, v. 37, p. 60-75, 2008.

MARCONI, M.A.; LAKATOS, E.M. *Técnicas de pesquisa: planejamento e execução de pesquisas, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados*. 2. ed. São Paulo: Atlas, 1990.

MOREIRA, Alexandra; ALVARENGA, Lídia; OLIVEIRA, Alcione de Paiva. O nível do conhecimento e os instrumentos de representação: tesouros e ontologias. *Datagramazero: Revista de Ciência da Informação*, v. 5, n. 6, dez. 2004. Disponível em <www.dgzero.org/dez04/Art_01.htm>. Acesso em 10 set. 2010.

MOREIRA, Alexandra. *Tesouros e Ontologias: estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico-sintético*. Belo Horizonte, 2003. 150 f. il. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2003.

MOREIRO GONZÁLEZ, J. A. *La representación de los contenidos digitales: de los tesouros automáticos a las folksonomías*. In: WORKSHOP CALSI, 6., 2007, Valencia, *Actas...Valencia*: [S.n.t.], 2007. Disponível em: <<http://www.calsi.org/2007/wpcontent/uploads/2007/11/jamoreiro.pdf>>. Acesso em: 13 jun. 2010.

NEVILLE, H. H. Feasibility study of a scheme for reconciling thesauri covering a common subject. *Journal of Documentation*, London, v. 26, n. 4, p. 313-336, Dec. 1970.

NOY, N. F.; McGUINNESS, D. L. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University: Stanford, 2001. Disponível em: <<http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noymcguinness.html>> Acesso em: jun 2011.

PICKLER, M. E. V.. Web Semântica: ontologias como ferramentas de representação do conhecimento. *Perspect. ciênc. inf.* [online]. 2007, vol.12, n.1, pp. 65-83. ISSN 1413-9936. doi: 10.1590/S1413-99362007000100006.

PIZZATO, L. A. S. *Estrutura multitesauro para recuperação de informações*. 112 f. Dissertação(Mestrado em Ciência da Computação) – Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2003.

RAMALHO, Rogério Aparecido Sá ; VIDOTTI, Silvana Aparecida Borsetti Gregorio ; FUJITA, Mariângela Spotti Lopes . Web semântica: uma investigação sob o olhar da Ciência da Informação. *Datagramazero* (Rio de Janeiro), v. 8, p. 4, 2007.

ROBREDO, J. Organização dos documentos ou organização da informação: uma questão de escolha. *Datagramazero* (Rio de Janeiro), Internet, v. 5, n. 1, 2004.

ROBREDO, Jaime; BRÄSCHER, Marisa (Orgs.). *Passeios pelo bosque da informação*: estudos sobre a representação e organização da informação e do conhecimento – eróico. Brasília,DF:Ibict, 2010. v + 335 p.ISBN: 978-85-7013-072-3.

SALES, Luana Farias; CAMPOS, Maria Luiza de Almeida and GOMES, Hagar Espanha. Ontologias de domínio: um estudo das relações conceituais. *Perspect. ciênc. inf.* [online]. 2008, vol.13, n.2, pp. 62-76.

SALES, Rodrigo de; CAFÉ, Lúgia. Semelhanças e diferenças entre tesauros e ontologias. *DataGramazero*, Rio de Janeiro, v. 9, n. 4, p. 00, ago. 2008.

SALES, Rodrigo de and CAFE, Lúgia. Diferenças entre tesauros e ontologias. *Perspect. ciênc. inf.* 2009, vol.14, n.1, pp. 99-116.

SARACEVIC, T. Ciência da Informação: origem, evolução, relações. *Perspectivas em Ciência da informação*, Belo Horizonte, v.1, n.1, p.41-62, jan/jun 1996.

SMITH, B. *Ontology and Informations Systems*. 2004. Disponível em: <<http://www.ontology.buffalo.edu/ontology>> Acesso em: 15 nov. 2010.

SMITH, B., *et al.* Relations in biomedical ontologies. *Genome Biology*, 6 (R46). 2005.

SOERGEL, Dagobert, ET AL. Reengineering thesauri for new applications: the Agrovoc example. *Journal of Digital Information*, v.4, n.4, 2004.

SOERGEL, Dagobert. *Organizing information*: principles of data base and retrieval systems. Orlando, Fla.: Academic Press, 1985. 450 p. ISBN 0126542600

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, v. 33, n.1, p.132-141, jan./abr. 2004.

SOUZA JUNIOR, M. B. ; CAFÉ, Ligia . Ontologias: abordagens nas teses e dissertações das universidades públicas brasileiras. *Informação & Sociedade* (UFPB. Online), v. 22, p. 81-98-98, 2012.

SKOS Simple Knowledge Organization System. Reference. W3C Recommendation 18 August 2009. disponível em: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> acesso: 17 nov 2010.

SOWA, J. F. *Knowledge representation*: logical, philosophical and Computational foundations. Pacific Grove: Brooks/Cole, 2000.

_____, J. F. *Ontology*. 2000. Disponível em: <<http://www.jfsowa.com/ontology/>>. Acesso em: 20 nov 2010.

TEIXEIRA, Lívia Marangon Duffles; ALMEIDA, Maurício Barcellos UNIVERSIDADE FEDERAL DE MINAS GERAIS. *Conceitualização na construção de ontologias: relações semânticas no âmbito do Blood Project*. 2009 201 f., enc. Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

TRISTÃO, Ana Maria Delazari; FACHIN, Gleisy Regina Bóries; ALARCON, Orestes Estevam. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. *Ciência da Informação*, Brasília, v.33, n.2, p.161-171, maio/ago. 2004.

TUDHOPE, D.; ALANI, H.; JONES, C. Augmenting Thesaurus Relationships: Possibilities for Retrieval. *International Journal on Computer Science and Information Systems*, 1(8), 2001.

_____, D. & Lykke Nielsen, M. Introduction to special issue: Knowledge Organization Systems and Services. *New Review of Hypermedia and Multimedia*, 12(1), 3-9, 2006.

VICKERY, B. Ontologies. *Journal of Information Science*, v. 23, n. 4, p.272-286, 1997.

_____. On 'knowledge organization'. Última edição 2008. Disponível em: <<http://www.lucis.me.uk/knoworg.htm#start>>. Acesso: 30 mar. 2012.

VILLAZON TERRAZAS, Borís Marcelo. *Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies*. 2011. 238 f. Tese (Doutorado em Inteligência Artificial) – Facultad de Informática (UPM), Universidade Politécnica de Madrid, Madrid.

W3C. Semantic Web Activity. Disponível em <<http://www.w3.org/2001/sw/Activity>>. Acesso em: 20 ago. 2012.

ZAVITOSKI, M. T. *Exploração do uso do tesouro como instrumento de recuperação da informação*. 2001. 144 f. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo.

ZENG, M. L. Knowledge organization systems (KOS). *Knowledge Organization: international journal devoted to concept theory, classification, indexing, and knowledge representation*, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.

ZENG, M. L.; SALABA, A. Toward na International Sharing and Use of Subject Authority Data. IN: FRBR WORKSHOP, OCLC, 2005. Kent State University. Disponível em: <www.oclc.org/research/events/frbr.../zeng/Zeng_Salaba.ppt...> Acesso em: março 2010.