

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS - CAMPUS DE MARÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

ANTONIO MARIN NETO

**ONTOLOGIAS NA REPRESENTAÇÃO DO CONHECIMENTO:
UMA FERRAMENTA SEMÂNTICA PARA A CIÊNCIA DA
INFORMAÇÃO**

MARÍLIA - SP
2018

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS - CAMPUS DE MARÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

ANTONIO MARIN NETO

**ONTOLOGIAS NA REPRESENTAÇÃO DO CONHECIMENTO:
UMA FERRAMENTA SEMÂNTICA PARA A CIÊNCIA DA
INFORMAÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências - Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, campus de Marília, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

ÁREA DE CONCENTRAÇÃO: Informação, Tecnologia e Conhecimento.

LINHA DE PESQUISA: Informação e Tecnologia.

ORIENTADOR: PROF. DR. EDBERTO FERNEDA

MARÍLIA – SP

2018

Marin Neto, Antonio.
M337o Ontologias na representação do conhecimento: uma
ferramenta semântica para a ciência da informação /
Antonio Marin Neto. – Marília, 2018.
82 f. ; 30 cm.

Orientador: Edberto Ferneda.
Dissertação (Mestrado em Ciência da Informação) –
Universidade Estadual Paulista (Unesp), Faculdade de
Filosofia e Ciências, 2018.
Bibliografia: f. 76-82

1. Ontologias (Recuperação da informação). 2.
Representação do conhecimento (Teoria da Informação). 3.
Ciência da informação. I. Título.

CDD 029.7

ANTONIO MARIN NETO

**ONTOLOGIAS NA REPRESENTAÇÃO DO CONHECIMENTO:
UMA FERRAMENTA SEMÂNTICA PARA A CIÊNCIA DA
INFORMAÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências - Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, campus de Marília, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

ÁREA DE CONCENTRAÇÃO: Informação, Tecnologia e Conhecimento.

LINHA DE PESQUISA: Informação e Tecnologia.

ORIENTADOR: PROF. DR. EDBERTO FERNEDA

Dr. Edberto Ferneda (orientador)
Faculdade de Filosofia e Ciências, UNESP-Marília

Dr. Fernando de Assis Rodrigues
Faculdade de Filosofia e Ciências, UNESP-Marília

Dr. Rogério Aparecido Sá Ramalho
Departamento de Ciência da Informação, UFSCar

Marília, agosto de 2018.

À minha esposa e companheira, sem a qual esse trabalho não existiria.

Agradecimentos

Primeiramente gostaria de agradecer a todos os professores, funcionários e colegas da UNESP Marília com os quais tive prazer de conviver nesse tempo de pesquisa, e que de alguma forma me ajudaram em todo esse processo de aprendizado e adaptação ao “mundo acadêmico”.

Em especial alguns amigos da minha primeira tentativa como aluno especial nos idos de 2010, como Fernando Assis e Zaira Zafalon, além dos novos amigos nessa nova jornada, em especial o parceiro de orientador e de vários almoços, Jorge Zanaite.

Gostaria de agradecer ao meu orientador, Prof. Dr. Edberto Ferneda, pela paciência com a minha falta de responsividade em alguns momentos e pelas diversas mudanças e reflexões ao longo do caminho. Muito obrigado por me manter “no foco”, mas ao mesmo tempo me ajudar em algumas “viagens” a respeito da dissertação. As ontologias entraram na minha vida em uma conversa despreziosa durante uma tarde em Marília, e acabaram se tornando uma “paixão” acadêmica na minha vida.

Também não poderia deixar de agradecer a Prof. Dra. Silvana Vidotti, que foi uma das incentivadoras quando tentei o mestrado pela primeira vez, e mais recentemente pelas conversas nas aulas, juntamente com o Prof. Vechiato, e claro, pelo divertido e excelente churrasco de fechamento da disciplina.

A conclusão desse mestrado para mim também é um momento de retribuir aos meus pais, que sempre me deram a melhor educação possível com esforços imensuráveis, mas que infelizmente não puderam ver a minha colação de grau na graduação. Sinto que agora, a possibilidade de os ter presenciando esse momento, é uma forma de pagar essa dívida de gratidão que carregava. Obrigado mãe e pai, amo vocês.

Devo especiais agradecimentos aos meus amigos de Marília, Sandra e Carlos. À Sandra pelas nossas excelentes (e quase intermináveis) conversas, que me ajudavam a relaxar depois de horas de viagem, além da paciência em ouvir sobre a minha pesquisa durante horas, e que me ajudavam a refletir enquanto explicava. E o Carlos, o irmão que não tive, mas que a vida me deu, que além de “pouso” me proporcionava momentos de risada, conversas sérias e claro, cerveja e ótima comida. Vocês são incríveis, obrigado.

Com um misto de agradecimento e um pouco de culpa, agradeço as minhas filhas por mesmo sem saberem, serem grandes responsáveis pela força de vontade em levar esse mestrado até o final. Infelizmente perdemos horas e dias de convivência entre viagens, momentos de dedicação, e etc., mas sempre era recompensado com um abraço carinhoso ou uma gargalhada gostosa quando estávamos aproveitando nossos momentos. Como eu prometi, agora o papai vai ter bastante tempo pra brincar com vocês, obrigado minhas pequenas pela paciência e compreensão, mesmo que sem saber.

E por fim, gostaria de agradecer a minha grande incentivadora e inspiração nessa jornada acadêmica, minha esposa Carolina. Foi através dela que conheci a área da Ciência da Informação, foi com ela que desenvolvi a minha vontade de me tornar um pesquisador e dedicar parte do meu tempo a ciência. Mas acima de tudo, foi ela quem, além de me incentivar, proporcionou a condição para que eu levasse esse mestrado até o fim. Nunca conseguirei expressar o meu amor e agradecimento a você.

“Qualquer novo conhecimento
provoca dissoluções e novas integrações.”

(Hugo von Hofmannsthal)

Resumo

As ontologias surgiram a partir de discussões na Filosofia com o objetivo de discutir a essência das coisas, nos estudos da metafísica por Aristóteles, que dentre os diferentes ramos de investigação filosófica, tratava do conhecimento da essência de toda a realidade. A ontologia trata do estudo do “Ser” enquanto ser. O termo foi posteriormente utilizado em diferentes áreas como a Ciência da Computação (CC), Psicologia, Ciência da Informação (CI), entre outros. Nesse texto é apresentado um levantamento da utilização das ontologias na representação do conhecimento, identificando e destacando a forma de utilização da ontologia como ferramenta na CC e CI. Para tal, o trabalho inicial com uma contextualização do termo ontologia, desde a sua origem na Filosofia até a sua utilização em algumas das áreas do conhecimento. Em seguida é apresentado um levantamento sobre a representação do conhecimento, a destacar a CC e CI, bem como sobre a utilização da ontologia nesse contexto informacional. Afim de evidenciar a importância das ontologias para representação do conhecimento, também são listadas as principais metodologias de criação de ontologia, aprendizado de ontologia, bem como as ferramentas utilizadas para essas tarefas, além do gerenciamento e engenharia de ontologias. E por fim são apresentadas as considerações finais acerca da utilização das ontologias como ferramentas semânticas na CI e como as futuras pesquisas podem se beneficiar desse trabalho como uma base para o entendimento das ontologias como ferramenta semântica na CI.

Palavras-chave: Ontologia, representação do conhecimento, aprendizado de ontologia, criação de ontologia.

Abstract

Ontologies come from discussions in Philosophy to debate the essence of things in Aristotle's studies of metaphysics, which among the different branches of philosophical inquiry dealt with the knowledge of the essence of all reality. Ontology deals with the study of "Being" as being. The term was later used in different areas such as Computer Science (CC), Psychology, Information Science (IS), and others. This thesis presents a survey of the use of ontologies in knowledge representation, identifying and highlighting how ontology is used as a semantic tool in CC and IS. For this, is presented a contextualization of the term ontology, from its origin in Philosophy to its use in some of the areas of knowledge. Next, a survey is presented on the representation of knowledge, highlighting the CC and IS, as well as on the use of the ontology in this informational context. In order to demonstrate the importance of ontologies for knowledge representation, the main methodologies for ontology creation, ontology learning, as well as the tools used for these tasks, as well as the management and engineering of ontologies are also listed. Finally, it's present the final considerations about the use of ontologies as semantic tools in IS and how future researches can benefit from this work as a basis for the understanding of ontologies as a semantic tool in IS.

Keywords: Ontology, knowledge representation, ontology learning, ontology construction.

Lista de Figuras

Figura 1 – Tipos de Ontologias	20
Figura 2 - Exemplo de definição de classe usando OWL.....	24
Figura 3 - Algumas definições de IA organizada em 4 categorias	26
Figura 4 - Cenários para a construção de ontologias e redes de ontologias	51
Figura 5 - Atividades da Engenharia de Ontologias.....	57
Figura 6 - Arquitetura conceitual do aprendizado de ontologia	59

Lista de Tabelas

Tabela 1 - Categorias genéricas de Aristóteles.....	8
Tabela 2 - Tabela de julgamentos e de categorias	9
Tabela 3 - Primitivas da RDFS.....	23
Tabela 4 - Periodos históricos da IA segundo Russel e Norvig (2010).....	27

Sumário

1	INTRODUÇÃO	1
1.1	Objetivos	3
1.2	Metodologia.....	3
1.3	Estrutura da pesquisa.....	5
2	ONTOLOGIAS.....	7
2.1	A Ontologia e os sistemas de categorização na Filosofia	7
2.2	Ontologias para além da Filosofia.....	10
3	ONTOLOGIAS E A REPRESENTAÇÃO DO CONHECIMENTO	17
3.1	Formalização das ontologias	19
3.2	Representação do conhecimento na Inteligência Artificial.....	24
3.3	Representação do conhecimento na Ciência da Informação	34
4	CONSTRUÇÃO DE ONTOLOGIAS.....	40
4.1	Metodologias de construção de ontologias	40
4.2	Ferramentas para construção de ontologias.....	52
4.3	Engenharia Ontológica.....	56
4.4	Aprendizado de Ontologia.....	58
5	APRENDIZADO DE ONTOLOGIA A PARTIR DE TEXTOS.....	62
5.1	Abordagens de aprendizado de ontologia	62
5.1.1	Abordagens de aprendizado baseadas em técnicas linguísticas	63
5.1.2	Abordagens de aprendizado baseadas em técnicas estatísticas	65
5.1.3	Abordagens de aprendizado baseadas em aprendizado de máquina	67
5.2	Ferramentas para aprendizado de ontologias	69

6	CONSIDERAÇÕES FINAIS.....	75
6.1	Trabalhos futuros.....	77
	REFERÊNCIAS	78

1

Introdução

A origem do termo ontologia remonta os estudos da metafísica por Aristóteles, que dentre os diferentes ramos de investigação filosófica, tratava do conhecimento da realidade de todos os seres, ou da essência de toda a realidade. A ontologia trata do estudo do “Ser” enquanto ser (CHAUI, 2013).

Posteriormente ao estabelecimento do estudo das Ontologias pela Filosofia, esse conceito passou a ser utilizado em diferentes áreas como a Ciência da Computação (CC), Engenharia de Software (ES), Sistemas de Informação (SI) e Ciência da Informação (CI).

A CC utiliza então o conhecimento representado pelas ontologias em diferentes aplicações, que de acordo com Gruber (1993), Vickery (1997) e Chandrasekaran, Josephson e Benjamins (1999), se destacam os sistemas de recuperação de informação, compartilhamento e reutilização de informação e processamento de linguagem natural. Russell e Norvig (2010) apontam também a utilização das ontologias nos processos de aprendizagem, tomada de decisão e comunicação dos agentes inteligentes, já que esses dependem de estruturas claras e definidas para poder utilizar e enriquecer o conhecimento a eles oferecido. As ontologias também são utilizadas na ES, que faz uso de abordagens sistemáticas, disciplinadas e quantificadas no ciclo de vida do software: análise e design, implementação, implantação, operação e manutenção.

Na área da CI, o termo ontologia passa a ser utilizado na literatura no final da década de 1990, principalmente por pesquisadores da área de Organização do Conhecimento (FERNEDA, 2013). Também é possível estabelecer uma relação entre ontologias e outros elementos comuns na CI: taxonomias e tesouros. Além disso, segundo Vickery (1997) projetos de ES utilizando

ontologias demonstram paralelos com teorias da CI, como classificação facetada, vocabulários controlados e lexicografia.

A utilização das ontologias na CI não está restrita somente a elementos similares ou paralelo entre teorias de outras áreas, segundo Almeida (2013) os princípios ontológicos são um assunto relevante na CI, pois diversos estudos já vêm sendo conduzidos na área, em diversos campos com diferentes conotações.

Pela própria natureza das ontologias, elas são comumente associadas a representação do conhecimento, algo que é tão antigo quanto a própria civilização humana, conforme demonstram estudos arqueológicos sobre a linguística, o primeiro alfabeto base da linguagem ocidental vem do Egito, datado aproximadamente em 4.000 AC. Para a Filosofia, segundo Mora (1996, apud CAIXETA; SOUZA, 2008), a representação refere-se aos “[...] diversos tipos de apreensão intencional de um objeto”.

Diferentes técnicas de representação do conhecimento foram desenvolvidas em diversas áreas, em geral essas variações são resultantes das necessidades de manipulação de cada área, mas há algumas convergências entre as técnicas, e segundo Vickery (1986) aparentemente cada área pode aprender alguma coisa com as demais.

Na CC, para o campo da Inteligência Artificial (IA) a representação do conhecimento é uma das principais preocupações, pois grande parte dos programas de IA envolvem o processamento de informação simbólica, ou seja, que os programas manipulam símbolos que de alguma forma representam pedaços de informação sobre o mundo, com o propósito que executar uma tarefa que normalmente requer inteligência.

Para a CI, a representação do conhecimento é algo que faz parte do processo de modelagem do conhecimento, que visa a construção de representações que são feitas através de diferentes tipos de Sistemas de Organização do Conhecimento (SOC), sistemas conceituais que representam determinado domínio através da sistematização dos conceitos e das relações semânticas entre si.

Dentre as diversas maneiras de se realizar essas tarefas, a pesquisa trata da utilização das ontologias como ferramentas semânticas na CI, propondo uma investigação através de um levantamento bibliográfico das relações com a CC, principalmente no campo da IA, conforme comentado anteriormente, mas também apresentando as metodologias de criação e aprendizado de ontologia.

Para que então, conhecendo-se a representação do conhecimento através da ontologia, e entendendo o processo de criação das ontologias, permita entender o seu potencial na área da CI, de forma transdisciplinar, propondo uma integração primeiramente dentro da própria CI, onde tecnologia e organização do conhecimento caminham juntas na busca de novas ferramentas e aprimoramento das aplicações, bem como com a CC, com o intuito de desenvolver colaborações no sentido do desenvolvimento de novas técnicas, tecnologias e soluções semânticas para ambas as áreas.

1.1 Objetivos

Baseando-se na utilização das ontologias pelas áreas da Filosofia, CC e CI, apresenta-se como objetivo geral da pesquisa o levantamento da utilização das ontologias na CI, afim da identificação das ontologias como ferramenta semântica na CI, para a utilização em diferentes aplicações da área, no que tange a representação do conhecimento visando identificação de mais aplicações possíveis.

Para o desenvolvimento da pesquisa, buscando alcançar o objetivo geral, foram definidos objetivos específicos, com as ações necessárias com o intuito de direcionar a realização de diferentes fases da pesquisa, obtendo uma maior coesão e coerência entre as etapas de desenvolvimento, conforme listados a seguir:

- Resgatar a origem do termo ontologia na Filosofia e esclarecer como as áreas da CC e CI passaram a usar o termo, citando as principais aplicações em cada área.
- Contextualizar a aplicação das ontologias na representação do conhecimento, estabelecendo aproximações dos conceitos de informação e conhecimento nas áreas da CC e CI, através do detalhamento da representação do conhecimento em cada área.
- Identificar o uso das ontologias como ferramentas semânticas na CC e na CI, fazendo uma relação entre as áreas e listando as potenciais futuras aplicações de estudos na área da CI.

1.2 Metodologia

Para alcançar os objetivos estabelecidos, essa pesquisa foi feita de forma exploratória, visando estabelecer uma maior familiaridade com o tema ontologias e representação do

conhecimento, para um posterior aprofundamento nos assuntos mais relevantes para o alcance dos objetivos.

A abordagem metodológica escolhida foi a de levantamento bibliográfico interdisciplinar, envolvendo as áreas da Filosofia, CC e CI, dividido em 4 etapas de pesquisa (por assunto) afim de possibilitar um maior entendimento da pesquisa e um aprofundamento nos assuntos a cada nova etapa.

No primeiro momento foram feitas buscas simples pelo termo “ontologia” na base *Scielo*, visando publicações da Ciência da Informação e Filosofia e nas bases da ACM (*Association for Computing Machinery*), IEEE (*Institute of Electrical and Electronics Engineers Standards Association*) e o *Google Scholar* para textos e publicações em geral. Essas buscas visavam estabelecer um parâmetro de volume de informação disponível, onde o critério para ordenação dos resultados foi a relevância (oferecida por cada base de dados), assim como o número de citações.

Para compreender melhor o processo de criação das ontologias, foram utilizadas as mesmas bases para buscas pelos termos “*ontology learning*”, “*ontology engineering*”, “*ontology construction*”, “*construção ontologias*”, “*aprendizado ontologia*” e “*engenharia ontologia*”. Esses termos foram definidos com base na pesquisa realizada na primeira etapa, onde esses termos foram citados por autores ao se referirem ao processo de criação de ontologias.

Na etapa seguinte, já com um foco em representação do conhecimento, as mesmas bases foram utilizadas, mas os termos buscados foram:

- *ontology + knowledge representation*
- *ontology + knowledge*
- *ontology + information representation*
- ontologia + representação
- ontologia + representação conhecimento
- ontologia + representação informação

Dos resultados obtidos, novamente usando as ordenações de relevância e número de citações, os que no resumo tratavam de como a ontologia era utilizada na representação de informação ou conhecimento, foram inicialmente selecionados. Posteriormente, a partir dos textos selecionados os autores mais citados ou textos relacionados que não haviam sido ainda

obtidos na pesquisa eram buscados manualmente. E caso fossem relevantes, eram adicionados a pesquisa.

Além dessa busca, para poder contextualizar melhor a pesquisa no campo da IA, foram procurados os principais cursos da área no Brasil (Unicamp e USP) além do *MIT* e *Stanford* nos Estados Unidos. E baseado nas bibliografias desses cursos, foram escolhidos três livros como base para a pesquisa sobre IA, além dos textos encontrados.

E por fim, durante o desenvolvimento do trabalho, novos textos seguiram sendo incorporados ao corpus da pesquisa conforme eram citados pelos artigos que estavam sendo utilizados para elaboração do texto. Isso permitiu um aprofundamento em assuntos mais clássicos, como a documentação, organização do conhecimento e IA, além de levar a pesquisa a chegar em autores renomados da área sem a necessidade de citações indiretas.

1.3 Estrutura da pesquisa

A pesquisa foi estruturada com a intenção de apresentar o assunto de forma sequência e estrutura, para demonstrar a relação entre os tópicos de forma evolutiva. Dito isso, o capítulo 2 visa introduzir o conceito das ontologias, resgatando a origem do termo na Filosofia e sua relação com a categorização introduzida por Aristóteles. O capítulo apresenta também a utilização das ontologias nas áreas da CC (e áreas correlatas, como ES e SI) e na CI.

O capítulo 3 discute a representação do conhecimento através das ontologias, para isso inicia-se com uma introdução geral do que se trata a representação do conhecimento, e assim como no capítulo 2, faz um aprofundamento para as áreas da CC e CI.

Já no capítulo 4, são abordadas as diferentes metodologias de criação de ontologias com o intuito de mostrar o processo de evolução desde as primeiras ontologias documentadas até os dias de hoje, com metodologias mais estruturadas e ferramentas computacionais disponíveis.

No capítulo 5 o assunto da criação de ontologia é aprofundado pela pesquisa sobre o aprendizado de ontologia a partir de texto, que consiste na automação de todo ou parte do processo de criação de ontologias utilizando textos não estruturados para a representação do conhecimento. Buscamos assim mostrar que a representação do conhecimento expresso em documentos pode ser uma maneira de utilização das ontologias como ferramenta semântica para representação do conhecimento humano.

Por fim, no capítulo 6 são apresentadas as considerações finais, incluindo uma reflexão sobre as ontologias como ferramentas semânticas e trabalhos futuros na CI para o empoderamento da área frente a nova era da informação digital.

2

Ontologias

A origem do termo ontologia remonta os estudos da metafísica por Aristóteles, que dentre os diferentes ramos de investigação filosófica, se tratava do conhecimento da realidade de todos os seres, ou da essência de toda a realidade. A ontologia se trata do estudo do “Ser” enquanto ser. Em grego, o termo *onto* significa “Ser”, enquanto *logia*, estudo ou conhecimento (CHAUI, 2013).

Nesse contexto, a Ontologia diz respeito aos tipos de coisas que existem e sua descrição, e foi Aristóteles o primeiro a utilizar a palavra grega *kategoria* como um termo técnico para predicação, sendo que a teoria das categorias elaborada pelo filósofo se tornou um dos mais importantes tópicos do estudo das ontologias (ALMEIDA, 2014). Ademais, a análise de como a filosofia estuda a teoria das categorias é importante para a CI, por ser a base de diversos estudos e ferramentas da área, conforme será visto mais a frente nesse texto.

Além do sistema aristotélico de categorias, segundo Almeida (2014) outros dois também devem ser analisados. Os três então seriam: o aristotélico, o kantiano e o husserliano.

2.1 A Ontologia e os sistemas de categorização na Filosofia

O sistema proposto por Aristóteles, que se tornou o mais conhecido, conta com uma lista de tipos de mais alto nível, que são conhecidos como categorias. Para os que estudam o sistema aristotélico, um sistema de categorias deve ser capaz de fornecer um inventário das coisas que existem, sendo assim, cada entidade possui uma característica fundamental chamada essência real. Assim, a categoria na qual uma entidade vai ser classificada depende da essência que possui, e essa entidade deve compartilhar um conjunto de propriedades suficientes com as

outras entidades da mesma categoria. Na Figura 1 podemos ver as 10 categorias, ou níveis, criados por Aristóteles:

Tabela 1 - Categorias genéricas de Aristóteles

Termo Aristotélico	Significado Moderno	Exemplo
<i>Ti esti, ousia</i>	Substância	Homem
<i>Poson</i>	Quantidade	Cinco metros
<i>Poion</i>	Qualidade	Branco
<i>Pros ti</i>	Relação	Metade
<i>Pou</i>	Local	No mercado
<i>Pote</i>	Data	Ontem
<i>Keisthein</i>	Postura	Sentado
<i>Echein</i>	Estado	Vestido
<i>Poitein</i>	Ação	Queimar
<i>Paschein</i>	Sentimento	Ser queimado

Fonte: (ALMEIDA, 2014)

Kent posteriormente questionou a maneira não sistemática que foi adotada pela teoria aristotélica para a definição das categorias, e propôs então um novo sistema partindo do conceito de julgamento na forma proposicional do sistema lógico aristotélico, *sujeito-cópula-predicado*, utilizando quatro aspectos pra classificar os julgamentos: quantidade, qualidade, relação e modalidade.

Para o aspecto quantidade, o sujeito pode ser universal, particular ou singular. Para o aspecto qualidade o predicado pode ser afirmativo, negativo ou infinito. No aspecto modalidade três tipos de verbos de ligação são propostos para conectar o sujeito e os predicados: problemática, assertória e apodítica. Para o quarto e último aspecto, a relação, julgamentos são combinados em inferências silogísticas, como em lógica, resultando em categórica, hipotética ou disjuntiva. A seguir a tabela de julgamentos e categorias de Kent (ALMEIDA, 2014):

Tabela 2 - Tabela de julgamentos e de categorias

Aspecto considerado		Julgamentos	Categorias
I	Quantidade de julgamentos	Universal	Unidade
		Particular	Pluralidade
		Singular	Totalidade
II	Qualidade	Afirmativo	Realidade
		Negativo	Negação
		Infinito	Limitação
III	Relação	Catagórico	Inferência e Subsistência
		Hipotético	Causalidade e Dependência
		Disjuntivo	Reciprocidade
IV	Modalidade	Problemático	Possibilidade
		Assertório	Existência
		Apodítico	Necessidade

Fonte: (ALMEIDA, 2014)

O terceiro estudo sobre categorias na Filosofia listado por Almeida (2014) é a fenomenologia desenvolvida por Hursserl, que se propõe a estudar os fenômenos, ou quais são as formas nas quais as coisas aparecem para nós em diferentes formas de experiências conscientes. Hursserl utiliza o termo categoria de uma forma particular, como uma entidade do reino das essências formais, dentro da Ontologia proposta por ele, que se organiza em reino dos fatos, reino das essências e o reino dos significados ou sentidos.

No reino dos fatos estão as entidades concretas, que são indivíduos empíricos concretos, estados de coisas empíricos concretos e eventos empíricos concretos. O reino das essências é composto das entidades ideais, ou aquelas não-reais, que não existem no espaço e tempo. E o reino do significado, ou sentido, contém os significados como as pessoas pensam os objetos, e são o conteúdo de experiências intencionais, as formas nas quais objetos são apresentados na consciência humana (ALMEIDA, 2014).

Os três estudos sobre categorias na Filosofia apresentados por Almeida (2014) fortalecem a ideia trazida por Vickery (1997, tradução nossa) de que “filosoficamente falando, Ontologia é o estudo do que existe e do que nós temos que assumir que existe para alcançar uma descrição convincente da realidade”.

Outras áreas acabaram se interessando pelo estudo das ontologias que era desenvolvido pela Filosofia, e partindo da abordagem filosófica transplantaram o termo para diferentes aplicações, como IA e Representação do Conhecimento (RC), tendo como um dos conceitos

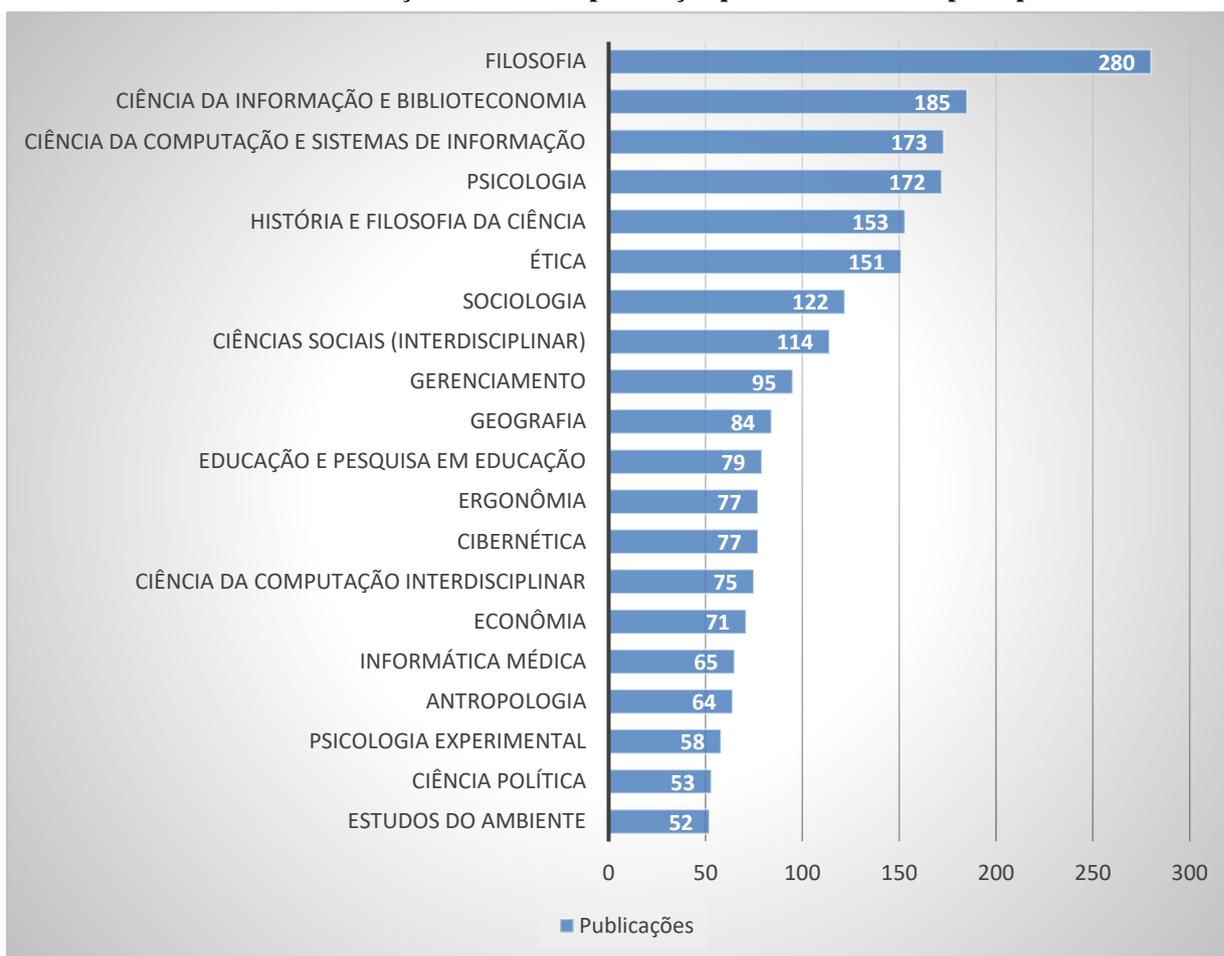
mais amplamente aceitos, o apresentado por Gruber (1993, tradução nossa) de que “uma ontologia é uma especificação explícita de uma conceitualização”, sendo que no mesmo parágrafo, Gruber (1993) deixa claro a origem filosófica do termo e diz que “para sistemas baseados em conhecimento, o que existe é exatamente aquilo que pode ser representado”.

Na seção seguinte iremos apresentar os principais conceitos segundo autores das áreas, quais são as principais aplicações de ontologias no campo da IA e na área da CI, fazendo um paralelo entre as aplicações e definições.

2.2 Ontologias para além da Filosofia

Posteriormente ao estabelecimento do estudo das Ontologias pela Filosofia, esse conceito passou a ser utilizado em diferentes áreas como a CC, ES, SI e CI. Na análise bibliométrica realizada por Yin, Lee e Yang (2009) sobre a utilização do termo “ontologia”, utilizando o *Social Sciences Citation Index* (SSCI) da *Web of Science*, resultou em 1.996 índices na literatura onde os títulos traziam a palavra ontologia de 1956 a 2008, mostrando um aumento significativo e constante desde 1995. Segundo os autores da pesquisa, o estudo das ontologias vem se tornando popular e está em um momento de alta maturidade nos últimos 15 anos, por exemplo, com 1000 citações em 2005 e chegando a 2000 em 2008 em diferentes áreas do conhecimento, como mostrado no Gráfico 1 com a distribuição do índice para os 20 principais assuntos no período entre 1956 e 2008. Apesar da pesquisa ter sido realizada há 8 anos atrás, ela mostra uma progressão constante ao longo de mais de 50 anos, evidenciando assim o crescimento do interesse no assunto.

Gráfico 1 - Distribuição do índice de publicação para os 20 assuntos principais



Fonte: Adaptado de Yin, Lee e Yang (2009)

Nos estudos sobre ontologias na CC, e especificamente no campo da IA, um conceito sobre ontologia constantemente utilizado é o apresentado por Gruber (1993), em que a “ontologia é uma especificação explícita de uma conceitualização”. O autor parte da origem do termo na Filosofia, onde segundo ele “uma Ontologia é a explicação sistemática da existência”, e estabelece uma relação com sistemas baseados em conhecimento, pois para esses sistemas o que existe é exatamente o que se pode ser representado.

Poucos anos após o trabalho de Gruber (1993) ser publicado, Guarino e Giaretta (1995) publicaram um outro estudo sobre ontologias, onde escreveram sobre a terminologia. Segundo Guarino e Giaretta (1995), o termo ontologia é usado de diferentes maneiras na representação e compartilhamento do conhecimento, e se baseiam em tão diferentes definições que o assunto deveria ser melhor esclarecido. Ainda naquela época, com o estudo de ontologias se tornando mais evidente, os autores identificaram 7 diferentes interpretações de ontologia:

1. Ontologia como uma disciplina filosófica.

2. Ontologia como um sistema conceitual informal.
3. Ontologia como uma conta formal semântica.
4. Ontologia como especificação de uma conceitualização.
5. Ontologia como a representação de um sistema conceitual via teoria lógica.
 - a. Caracterizada por propriedades formais específicas.
 - b. Caracterizada apenas pelo seu propósito específico.
6. Ontologia como o vocabulário usado por uma teoria lógica.
7. Ontologia como uma (meta) especificação de uma teoria lógica.

Das 7 interpretações apresentadas acima, a primeira é a que se diferencia mais das demais, por remeter a área original do estudo das ontologias. Guarino e Giaretta (1995) estabelecem que quando utilizam um artigo indefinido antes do termo ontologia (exemplo: *uma* ontologia) com “o” minúsculo, refere-se então a um determinado objeto em particular (que sua natureza pode variar), enquanto quando utilizado com “O” maiúsculo, se trata da Ontologia como a disciplina filosófica, que já foi introduzida anteriormente nessa seção.

As interpretações 2 e 3 trazem as ontologias como uma entidade semântica conceitual, formal ou informal, enquanto as interpretações 5 e 7 trazem as ontologias como um objeto sintático. Já na interpretação 6, ontologia se trata apenas de uma “ferramenta”, o vocabulário que será utilizado pela teoria lógica. Para a interpretação 4, oriunda do trabalho apresentado por Gruber (1993), Guarino e Giaretta (1995), consideram que ela pode ser mal interpretada e acabar sendo utilizada de forma incorreta, por se basear em um espectro extenso de conceitualização (GUARINO; GIARETTA, 1995).

Guarino e Giaretta (1995) restringem então a utilização do termo ontologia em 3 sentidos técnicos. No sentido (a) uma ontologia é um sinônimo para “teoria ontológica”, que difere da teoria lógica pela sua semântica, já que seus axiomas devem ser verdade em todos os cenários possíveis da conceitualização utilizada. No sentido (b) ontologia é sinônimo da especificação de um comprometimento ontológico¹. Por fim, no sentido (c) ontologia é sinônimo de conceitualização, mas diferentemente do apresentado por Gruber (1993) nesse caso se trata de conceitualização mais estrita e com as relações claramente definidas.

Posteriormente a esse esclarecimento de terminologia proposto por Guarino e Giaretta (1995), Chandrasekaran, Josephson e Benjamins (1999) também se preocuparam em falar sobre

¹ Comprometimento ontológico é a descrição semântica da conceitualização pretendida de uma teoria lógica (GUARINO; GIARETTA, 1995).

a utilização das ontologias na IA no texto: “O que são ontologias e porque precisamos delas?”. Os autores começam o trabalho fazendo um posicionamento claro de sua proposta sobre uso de ontologias na IA:

Teorias em IA caem em duas grandes categorias: teorias de mecanismos e teorias de conteúdo. Ontologias são teorias de conteúdo sobre os tipos de objetos e as relações entre objetos que são possíveis em um domínio específico de conhecimento. Elas fornecem potenciais termos para descrever o conhecimento sobre o domínio.

A classificação das ontologias dentro da categoria de teoria de conteúdo é justificada por Chandrasekaran, Josephson e Benjamins (1999), afirmando que uma das principais contribuições das ontologias é identificar classes e objetos específicos, e as relações existentes em um determinado domínio.

De forma mais aprofundada, analogamente ao senso de Guarino e Giaretta (1995), Chandrasekaran, Josephson e Benjamins (1999) apresentam que ontologia é sinônimo da especificação de um comprometimento ontológico, estabelecendo uma relação com a origem do termo na Filosofia, é possível afirmar que ontologia é então uma representação de vocabulário, em geral especializado em um domínio ou assunto específico, ou também a representação de um corpo de conhecimento. A diferença é que o vocabulário fornece um conjunto de termos que descrevem os fatos de um domínio, enquanto um corpo de conhecimento usa um vocabulário para descrever todo o conhecimento do domínio.

Conforme demonstrado pelas definições apresentadas, apesar de não haver uma definição única sobre o que são as ontologias, dentro da IA há um consenso de que uma das principais funções desempenhadas pelas ontologias é a representação do conhecimento. Essa aplicação é embasada por diversos autores como Gruber (1993), Guarino e Giaretta (1995), Vickery (1997), Chandrasekaran, Josephson e Benjamins (1999), Almeida (2014) e Russell e Norvig (2010).

A CC utiliza então o conhecimento representado pelas ontologias em diferentes aplicações, que de acordo com Gruber (1993), Vickery (1997) e Chandrasekaran, Josephson e Benjamins (1999), se destacam os sistemas de recuperação de informação, compartilhamento e reutilização de informação e processamento de linguagem natural. Russell e Norvig (2010) apontam também a utilização das ontologias nos processos de aprendizagem, tomada de decisão e comunicação dos agentes inteligentes, já que esses dependem de estruturas claras e definidas para poderem utilizar e enriquecer o conhecimento a eles oferecido. O detalhamento sobre as

formas de representação do conhecimento e como a CC, e especificamente o campo da IA, se beneficia disso serão discutidos no capítulo a seguir.

As ontologias também são utilizadas na ES, que faz uso de abordagens sistemáticas, disciplinadas e quantificadas no ciclo de vida do software: análise e design, implementação, implantação, operação e manutenção. Happel e Seedorf (2006) apresentaram uma categorização das ontologias no processo de ES que torna mais fácil a compreensão da utilização das ontologias. Essas categorias são:

- **Desenvolvimento orientado a ontologias (ODD):** Considera o uso das ontologias durante o desenvolvimento, e que descrevem o domínio do problema. Exemplo dessa abordagem é a integração com softwares de modelagem de linguagens, uma vez que por permitirem inferências e outros tipos operações lógicas na ontologia do domínio, essas irão auxiliar na criação de modelos de dados, validações lógicas e criação de vocabulários. Para esse cenário, também é considerado a utilização das ontologias no processo de Engenharia de requisitos, na elaboração de documentos de especificação de requisitos, e para a representação formal do conhecimento acerca dos requisitos, possibilitando validação automática de consistência, já que oferecem um formalismo ao modelo.
- **Desenvolvimento habilitado por ontologias (OED):** Também durante a etapa de desenvolvimento do software, mas dessa vez como um acessório na validação e controle para as estruturas de dados, tipos de variáveis utilizadas na programação e busca de componentes que possam ser reutilizados.
- **Arquiteturas baseadas em ontologia (OBA):** Utilização da ontologia como artefato primário durante a operação do sistema, como parte central da lógica da aplicação. Regras de negócio podem ser um exemplo do que é representado na ontologia.
- **Arquiteturas habilitadas por ontologias (OEA):** Utilização das ontologias para fornecer suporte à infraestrutura no tempo de execução de um sistema de software. Um exemplo são os serviços semânticos da Web, onde ontologias adicionam uma camada semântica sobre as descrições de serviço da Web existentes, adicionando funcionalidade para a descoberta, correspondência e composição automáticas de fluxos de trabalho baseados em serviço.

Na CI o termo ontologia passa a ser utilizado na literatura no final da década de 1990, principalmente por pesquisadores da área de Organização do Conhecimento. E foi a partir da fundação da *International Society for Knowledge Organization* (ISKO), em 1989, que a área da Organização do Conhecimento passou a se consolidar (FERNEDA, 2013). Também é possível estabelecer uma relação entre ontologias e outros elementos comuns na CI: taxonomias e tesouros. Nesses casos, a similaridade é pela possibilidade de restringir a linguagem natural, como um vocabulário controlado (ALMEIDA, 2014). Além disso, segundo Vickery (1997) projetos de ES utilizando ontologias demonstram paralelos com teorias da CI, como classificação facetada, vocabulários controlados e lexicografia.

A utilização das ontologias na CI não está restrita somente a elementos similares ou paralelo entre teorias de outras áreas, segundo Almeida (2013) os princípios ontológicos são um assunto relevante na CI, pois diversos estudos já vêm sendo conduzidos na área, em diversos campos com diferentes conotações. Isso pode ser ilustrado por uma reflexão em Almeida (2014):

Ao que parece, os autores de Ciência da Informação nem sempre mencionam explicitamente o termo “ontologia”, mesmo que princípios ontológicos aparecem claramente na literatura daquele campo de pesquisa.

Na tese “Ontoforinfoscience: Metodologia Para Construção De Ontologias Pelos Cientistas Da Informação”, Mendonça (2015) propõe a utilização das ontologias na CI de uma forma mais ampla:

[...] a ontologia como um instrumento para a organização e representação da informação, capaz de capturar, mapear e compartilhar o conhecimento de domínios diversos, de forma a “estabelecer a verdade sobre a realidade, encontrando respostas para a questão ‘o que existe’” (SMITH, 2004, p. 22). Nesse sentido, compreende-se ontologias muito além do que um simples artefato da engenharia de software.

Em seguida, Mendonça (2015) lista algumas das aplicações das ontologias na CI: (i) suporte à extração de informação; (ii) tradução de linguagem natural; (iii) integração automática de um conjunto de vocabulários padronizados ou dicionários de dados relacionados a um domínio específico; (iv) fornecer informações sobre categorias (conceitos) existentes em um domínio; (v) integração de bancos de dados, softwares ou modelos de negócios; (vi) desenvolvimento de recursos para a Web Semântica.

Considerando as aplicações de ontologias na CI levantadas por Mendonça (2015), somadas as relações estabelecidas entre elementos ontológicos da CI com as ontologias feitas

por Vickery (1997) e Almeida (2014), é possível observar o elevado potencial de contribuição das ontologias em diferentes campos da CI.

No capítulo seguinte serão abordadas as questões básicas da representação do conhecimento, passando pelas diferentes formalizações de ontologias, aplicações diretas na CI e as vantagens obtidas pela área no estudo e utilização das ontologias.

3

Ontologias e a Representação do Conhecimento

A representação do conhecimento é um fenômeno tão antigo quanto a própria civilização humana, conforme demonstram estudos arqueológicos sobre a linguística, o primeiro alfabeto base da linguagem ocidental vem do Egito, datado aproximadamente em 4.000 AC. A forma mais próxima do que usamos hoje como estrutura de comunicação escrita vem da Grécia, a partir do uso das primeiras letras alfa e beta, que revolucionaram o modo de informação ocidental, gerando um salto em sua representação (CAIXETA; SOUZA, 2008).

Caixeta e Souza (2008) também apontam o desenvolvimento do conhecimento matemático como um importante impulso na representação dos fenômenos da natureza, pois segundo os autores:

[...] a matemática foi fundamental para estabelecer leis da natureza baseadas em símbolos que representassem a contagem e a maioria das leis de causa e efeito e outras relações que regem elementos da natureza, passíveis de comprovação por uma lógica, científica e racional.

Para a Filosofia, segundo Mora (1996, apud CAIXETA; SOUZA, 2008), a representação refere-se aos “[...] diversos tipos de apreensão intencional de um objeto” e alguns filósofos apresentaram suas visões sobre a representação de formas diferentes. Para Aristóteles, a representação é um fenômeno da fantasia (intelectual ou sensível); os estoicos a viam como impressão, Descartes como imaginação; Spinoza distinguia entre apresentação sensível e conceitual; e por fim, Schopenhauer definia como a forma do mundo dos objetos como manifestações de vontade. Para o autor, o termo é ambíguo especialmente quando se trata de

esclarecer a natureza e as formas do conhecimento do mundo pela epistemologia e a psicologia, e na relação entre as duas.

Na psicologia a representação pode ser vista de 4 formas: (i) objeto efetivamente presente, equiparada a ideia de percepção; (ii) reprodução na consciência de percepções passadas, como memória ou recordações; (iii) como antecipação de acontecimentos futuros baseada em eventos passados, como a imaginação; (iv) união na consciência de várias percepções não atuais e nem passadas ou antecipatórias, que tanto pode levar a ideia de imaginação ou alucinação. Além dessas quatro formas, a psicologia também admite a representação baseada no predomínio de um dos sentidos humanos, como também representações conceituais, afetivas e volitivas (CAIXETA; SOUZA, 2008).

Seguindo o paradigma da representação, Capurro (1991) traz em seu estudo a abordagem de que os seres humanos são conhecedores e observadores da realidade externa, e que o processo do conhecimento se dá através da assimilação das coisas através da representação na mente/cérebro do receptor. Uma vez tendo essas representações processadas e codificadas no cérebro, poderiam ser comunicadas a outras mentes e/ou armazenadas e processadas por computadores. Portanto, para Capurro (1991), as pessoas são processadores biológicos de informação, e essa informação é a codificação dupla da realidade. Os humanos usam as informações para fins racionais específicos, mas não há nada que vá contra a hipótese de que os computadores possam alcançar o mesmo nível de processamento e uso da informação.

Conhecimento, crenças e opiniões são pessoais, subjetivos e conceituais. Mas para que seja possível comunicar esse conhecimento, é preciso que seja expresso, descrito ou representado de alguma maneira física (tangível), como um símbolo ou texto (BUCKLAND, 1991). De certa forma complementar ao conceito de “informação como coisa” apresentado por Buckland, Vickery (1986) afirma que o conhecimento pode ser representado por símbolos combinados de diversas maneiras, tendo como o mais conhecido a linguagem natural.

Tal representação através de símbolos é relevante em diversas situações de diversas áreas, como documentos e índices, estruturas documentárias, estruturas de registros e arquivos em um banco de dados, estruturas de dados em programas de computador, estruturas sintáticas e semânticas em linguagens naturais, representação do conhecimento em IA, redes semânticas, modelos de memória humana, entre outros. O fator comum entre todas essas aplicações ou campos é a necessidade de definir como o conhecimento vai ser representado, para que essa representação possa ser manipulada (VICKERY, 1986).

Diferentes técnicas de representação do conhecimento foram desenvolvidas em diversas áreas, em geral essas variações são resultadas das necessidades de manipulação de cada área, mas há algumas convergências entre as técnicas e segundo Vickery (1986), aparentemente cada área pode aprender alguma coisa com as demais.

Nas seções a seguir serão exploradas a formalização das ontologias como estruturas de representação do conhecimento para diferentes situações, as representações do conhecimento nas áreas da CC (principalmente no campo da IA) e CI, trazendo uma visão geral das aplicações e ferramentas.

3.1 Formalização das ontologias

Guarino (1995) faz uma derivação do conceito introduzido por Gruber (1993), de que uma ontologia é uma “especificação de uma conceitualização”, conforme abordado no capítulo anterior. Nessa derivação, Guarino (1995) chega a hipótese de que ontologias correspondem a teorias lógicas, ou seja, uma caracterização axiomática do significado de um vocabulário lógico.

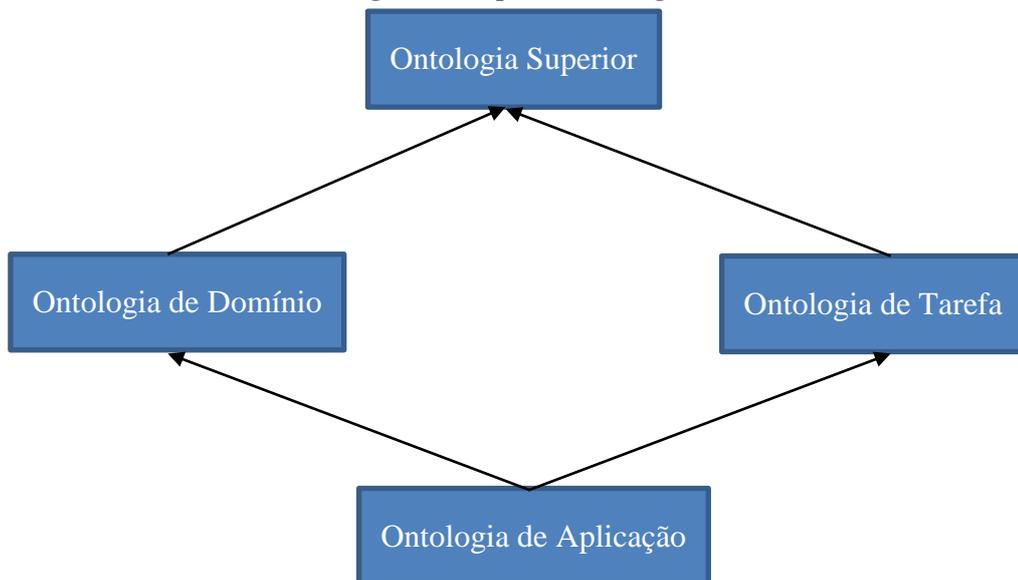
Partindo dessa premissa, o autor explica que a diferença entre uma ontologia e uma base de conhecimento, se caracteriza pelo propósito da ontologia em si, que pode ser dada como uma base de dados particular a uma situação, onde todas as afirmações que descrevem os fatos são assumidas como verdadeiras para um grupo de usuários em virtude de um vocabulário de significado acordado entre todos. Sendo que uma base de conhecimento genérica também pode descrever fatos e afirmações relacionadas a um estado particular.

As ontologias podem então ser classificadas de acordo com duas dimensões, nível de detalhe e nível de dependência. Para a primeira dimensão, uma ontologia se aproxima do significado desejado de um vocabulário (podendo ser utilizada para estabelecer um consenso sobre o vocabulário compartilhado ou uma base de conhecimento que utilize esse vocabulário), mas acabam dependendo de uma linguagem mais rica. Já uma ontologia com baixo nível de detalhamento, pode ser desenvolvida já partindo do pensamento em utilizar serviços de inferência, visando o compartilhamento entre usuários que já possuem um acordo (formalização) do vocabulário a ser empregado (GUARINO, 1995).

Para a segunda dimensão de classificação, nível de dependência, Guarino (1995) enumera 4 tipos de ontologia (Figura 1 –):

- **Ontologias superior (*top-level ontologies*):** Descreve conceitos gerais como espaço, tempo, ações e etc., que são independentes de situação ou um domínio em particular. Mais adequadas para unificação de ontologias para uma grande quantidade de usuários.
- **Ontologias de domínio e de tarefa (*domain ontologies e task ontologies*):** Descrevem, respectivamente, vocabulários a um domínio genérico (medicina ou engenharia) ou uma tarefa genérica (como diagnosticar ou vender), através da especialização de termos da ontologia superior.
- **Ontologias de aplicação (*application ontologies*):** Descrevem conceitos dependentes de tarefa e em um domínio em particular, sendo em geral especializações de ambas ontologias relacionadas. Esses conceitos normalmente correspondem a funções exercidas por entidades do domínio quando executam uma certa atividade.

Figura 1 – Tipos de Ontologias



Fonte: Guarino (1995)

Outra forma de se classificar, ou diferenciar, ontologias é levando em consideração o nível de formalidade. Ontologias podem ser altamente informais se expressadas em linguagem natural; semi-informal se expressada em uma restrita e estruturada forma da linguagem natural; semiformal se expressa em uma linguagem artificial formalmente definida (OntoLingua, OWL, etc.); e fortemente formal se fornecer termos definidos meticulosamente com uma semântica

formal, teoremas e provas de suas propriedades (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004).

Gruber (1993) identificou cinco componentes das ontologias: classes, relações, funções, axiomas formais e instâncias. Esses mesmos componentes são apontados em outros trabalhos, com o manual de Noy e McGuinness (2001) sobre como criar ontologias usando o Protégé, e apesar de completos, esses componentes possuem diversos aspectos técnicos, que estão mais relacionados a IA e representação utilizando lógica. Com o objetivo de facilitar o entendimento, omitindo alguns detalhes técnicos, Ramalho (2010) apresenta uma síntese dos componentes de uma ontologia:

- **Classes e Subclasses:** As classes e subclasses de uma ontologia agrupam um conjunto de elementos, “coisas”, do “mundo real”, que são representadas e categorizadas de acordo com suas similaridades, levando-se em consideração um domínio concreto. Os elementos podem representar coisas físicas ou conceituais, desde objetos inanimados até teorias científicas ou correntes teóricas;
- **Propriedades Descritivas:** Descrevem as características, adjetivos e/ou qualidades das classes;
- **Propriedades Relacionais:** Trata-se dos relacionamentos entre classes pertencentes ou não a uma mesma hierarquia, descrevendo e rotulando os tipos de relações existentes no domínio representado;
- **Regras e Axiomas:** Enunciados lógicos que possibilitam impor condições como tipos de valores aceitos, descrevendo formalmente as regras da ontologia e possibilitando a realização de inferências automáticas a partir de informações que não necessariamente foram explicitadas no domínio, mas que podem estar implícitas na estrutura da ontologia;
- **Instâncias:** Indicam os valores das classes e subclasses, constituindo uma representação de objetos ou indivíduos pertencentes ao domínio modelado, de acordo com as características das classes, relacionamentos e restrições definidas;
- **Valores:** Atribuem valores concretos às propriedades descritivas, indicando os formatos e tipos de valores aceitos em cada classe.

O passo final na formalização da representação do conhecimento utilizando ontologias, se dá pela codificação dos elementos da ontologia em um formato que possa ser processado pelos computadores, sendo o objetivo da ontologia a utilização por alguma tecnologia. Na

literatura das áreas de CC, ES e SI as práticas mais comumente utilizadas são as diretamente relacionadas com a área, como por exemplo *Unified Modeling Language* (UML) e Modelo Entidade Relacionamento (MER).

A utilização do UML pode ser utilizada para a modelagem de ontologias por possuir uma estrutura que prevê os elementos utilizados pelas ontologias, além de ser de fácil entendimento por pessoas externas à área da CC, e por haver diversas ferramentas CASE disponíveis para criação e manipulação dos diagramas UML, que são a base da representação dessa linguagem (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2003).

Outro exemplo citado foi a utilização de técnicas de modelagem de bancos de dados, MER. Nesse caso também há a possibilidade de estabelecer uma relação entre os elementos de uma ontologia e os elementos da MER. O ponto positivo dessa representação é a possibilidade de geração de instâncias através de dados inseridos no banco de dados, e a facilidade de consulta com a linguagem SQL. Assim como é possível a utilização de índices e chaves entre os relacionamentos, permitindo a adequação dos axiomas e regras, mas ainda não é possível fazer a representação de ontologias mais complexas (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2003).

Para a representação mais estruturada e de ontologias mais pesadas, uma alternativa é o formado RDF – *Resource Description Framework* – que foi criado para a descrição de metadados na Web. O modelo de dados RDF consiste em três elementos:

- **Recursos (*resources*):** Qualquer tipo de dado descrito pelo RDF. Recursos são descritos com expressões RDF e remete a uma URI (*Uniform Resource Identifiers*) para mais elementos opcionais de ancoragem.
- **Propriedades ou predicados (*properties ou predicates*):** Definem atributos ou relações que descrevem o recurso.
- **Afirmações (*statements*):** Definem valor a uma propriedade de um determinado recurso.

Os recursos não precisam ser necessariamente objetos, expressões RDF também podem ser um recurso RDF, como uma recursividade. Além disso, a notação RDF não presume nenhuma ordem na estrutura documento que contém as expressões RDF que representam uma ontologia (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2003).

Como extensão ao RDF, a W3C publicou em 2004 a especificação final do RDF *Vocabulary Description Language: RDF Schema*, ou RDFS. Essa extensão adiciona 16 novas

primitivas que trazem uma maior carga semântica a representação utilizando esse esquema. Essas primitivas permitem, por exemplo, que sejam definidos relacionamentos entre propriedades e recursos, e limitar entre quais recursos é uma determinada relação, algo que não é possível com o RDF. Veja na tabela abaixo a estrutura das primitivas disponíveis na RDFS:

Tabela 3 - Primitivas da RDFS

Propriedade	Domínio	Alcance
rdf:type	rdfs:Resource	rdfs:Class
rdf:subject	rdf:Statement	rdfs:Resource
rdf:predicate	rdf:Statement	rdf:Property
rdf:object	rdf:Statement	rdfs:Resource
rdf:value	rdfs:Resource	rdfs:Resource
rdf:first	rdf:List	rdfs:Resource
rdf:rest	rdf:List	rdf:List
rdfs:subClassOf	rdfs:Class	rdfs:Class
rdfs:subPropertyOf	rdf:Property	rdf:Property
rdfs:comment	rdfs:Resource	rdfs:Literal
rdfs:label	rdfs:Resource	rdfs:Literal
rdfs:seeAlso	rdfs:Resource	rdfs:Resource
rdfs:isDefinedBy	rdfs:Resource	rdfs:Resource
rdfs:member	rdfs:Resource	rdfs:Resource
rdfs:domain	rdf:Property	rdfs:Class
rdfs:range	rdf:Property	rdfs:Class

Fonte: W3C

Para melhorar a capacidade de inferência utilizando raciocínio sobre o conhecimento representado pelas ontologias com RDFS, foi criado o OIL (*Ontology Inference Layer*) adicionando mais primitivas ao modelo do RDFS e evitando o mecanismo de reificação (transformar um conceito abstrato em algo concreto, uma coisa). O OIL foi construído na arquitetura de camadas e usa lógica descritiva para criar uma semântica mais clara para os primitivos modelados. Cada camada construída sobre a outra, adiciona uma nova funcionalidade e mais complexidade a camada anterior (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2003).

Assim como o OIL, o DAML+OIL foi criado como uma extensão ao RDFS, porém essa linguagem não é dividida em diferentes camadas, ela provê extensões de lógica descritiva

diretamente ao RDFS, sendo uma linguagem estendida da família da SHIQ que prove tipos de dados e nominais.

Mais recentemente, em 2003, o grupo de trabalho *Web Ontology* (WebOnto) da W3C criou o OWL, uma linguagem que deriva da DAML+OIL e é construída tendo como base a RDFS. Assim como a OIL, a OWL é dividida em camadas: OWL Lite que estende RDFS e reúne as funcionalidades mais comuns da OWL, e é indicada para usuários que desejam criar apenas taxonomias e constantes simples; OWL DL que inclui o vocabulário completo da OWL; e a OWL Full, que oferece mais flexibilidade na representação do que a OWL DL (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2003). A seguir um exemplo de como utilizar OWL para definir a classe *Travel*, que possui um único valor para o atributo *flightNumber*, cujo o tipo é *integer*, e utiliza um outro recurso para definir o atributo *transportMeans*:

Figura 2 - Exemplo de definição de classe usando OWL

```
<owl:Class rdf:ID="Flight">
  <rdfs:comment>A journey by plane</rdfs:comment>
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#Travel"/>
    <owl:Restriction owl:cardinality="1">
      <owl:onProperty rdf:resource="#flightNumber"/>
      <owl:allValuesFrom rdf:resource="&xsd;integer"/>
    </owl:Restriction>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#transportMeans"/>
      <owl:hasValue rdf:datatype="&xsd:string">
        plane
      </owl:hasValue>
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

Fonte: (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2003)

A seguir, será discutido como a representação do conhecimento é tratada em diferentes áreas, dando foco principalmente para IA e CI.

3.2 Representação do conhecimento na Inteligência Artificial

A IA é um dos campos mais novos na ciência e engenharia, tendo início logo após o final da Segunda Guerra Mundial, e teve o seu nome cunhado em 1956. A IA compreende uma grande variedade de subcampos, desde os mais genéricos como aprendizado e percepção, até os mais específicos como jogar xadrez, resolver teoremas matemáticos, carros autônomos ou

diagnóstico de doenças. IA é relevante a qualquer tarefa intelectual, sendo um campo verdadeiramente universal (RUSSELL; NORVIG, 2010).

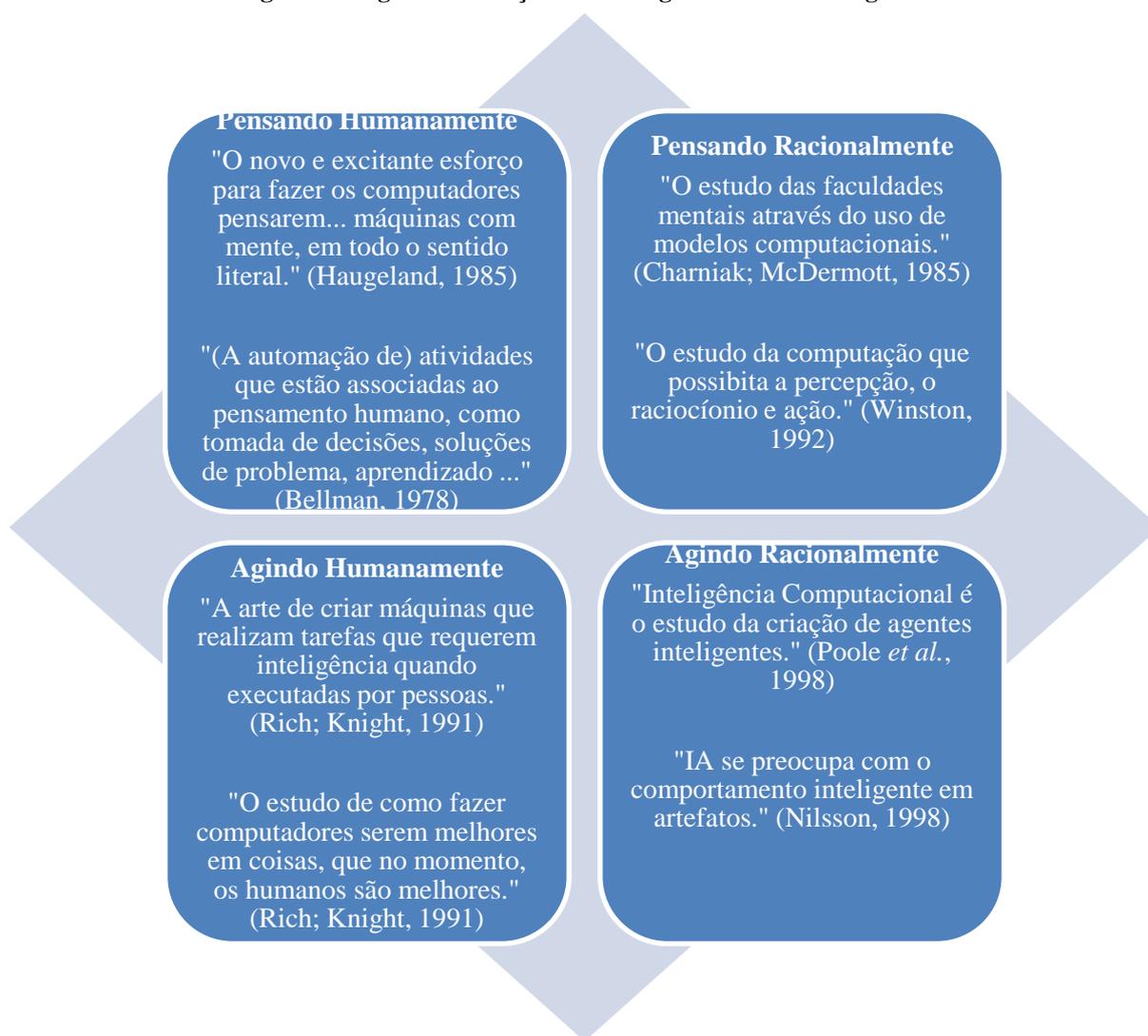
Segundo Russell e Norvig (2010) historicamente existem 4 abordagens que tem sido seguida por diferentes pessoas com diversos métodos, que podem ser brevemente descritas da seguinte forma:

- **Ações Humanizadas (abordagem do Teste de Turing):** Abordagem onde o computador possui capacidades que são relacionadas a ações relacionadas a pessoas, como por exemplo processamento de linguagem natural, representação do conhecimento, raciocínio automatizado, aprendizado de máquina, visão computacional e robótica.
- **Pensamento Humanizado (abordagem de modelo cognitivo):** Para ser possível determinar que um computador pensa como um humano, é preciso primeiro conhecer como o pensamento humano em si funciona. Para isso há 3 maneiras: (i) introspeção no pensamento humano; (ii) experimentos psicológicos; (iii) análise de imagens do cérebro. Uma forma clássica de fazer uma equivalência nessas áreas, seria analisar se as entradas e saídas de um programa são correspondentes a entradas e saídas de pessoas, estabelecendo que há então um bom modelo humano / computador e vice-versa, enquanto autores modernos separam esses dois 2 modelos.
- **Pensamento Racional (abordagem das leis do pensamento):** Abordagem inspirada pelo filósofo grego Aristóteles, que foi um dos primeiros a tentar codificar o “pensamento correto” como um processo irrefutável de raciocínio. Seus silogismos forneceram estruturas de argumentação que sempre levam a resultados corretos quando partem de premissas corretas. Essas “leis dos pensamentos” seriam supostamente a forma de operar do pensamento humano, e seus estudos levaram ao desenvolvimento do campo chamado hoje de lógica. Programas de computador usando essa abordagem seriam, em teoria, capazes de resolver qualquer problema em notação lógica. Porém essa abordagem tem 2 obstáculos: (i) não é fácil transformar conhecimento informal para a forma de notação lógica, principalmente quando o conhecimento não conta com 100% de certeza; (ii) há uma grande diferença entre resolver um problema teórico e um problema na prática.
- **Ações Racionais (abordagem dos agentes racionais):** Agente racional é um agente que atua para obter a melhor saída, ou quando há incerteza, para obter a melhor saída esperada. Dessa maneira, a abordagem de ação racional consiste no comportamento

onde o agente vai realizar ações como operar autonomamente, perceber o seu ambiente, existir durante uma quantidade de tempo maior possível / necessária, se adaptar as mudanças e criar e perseguir seus objetivos. Essa abordagem tem relação com o pensamento racional e as leis do pensamento, mas tem por necessidade um caráter mais generalista.

A Figura 3 a seguir ilustra essas abordagens com algumas definições de IA de alguns autores de cada abordagem.

Figura 3 - Algumas definições de IA organizada em 4 categorias



Fonte: (RUSSELL; NORVIG, 2010)

Além das 4 abordagens da IA, Russel e Norvig (2010) apresentam uma cronologia do desenvolvimento da IA (**Erro! Fonte de referência não encontrada.**) e dentre elas, uma revela como a representação do conhecimento viria se tornar uma disciplina chave no desenvolvimento

do campo da IA, nomeada pelos autores como era dos “Sistemas baseado em conhecimento: A chave para o poder?”, que ocorreu entre os anos de 1969 e 1979.

Essa fase surge em contraposição as tentativas anteriores de solução de problemas, que se baseavam em métodos generalistas de busca da melhor solução através de mecanismos lógicos usando passos elementares para achar a solução final. Essas abordagens foram chamadas de “métodos fracos” pois apesar de serem generalistas, eles não eram capazes de escalar ou resolver problemas mais complexos.

Tabela 4 - Periodos históricos da IA segundo Russel e Norvig (2010)

Nome	Período
Gestação da IA	1943 – 1955
Entusiasmo inicial, grandes expectativas	1952 – 1969
Uma dose de realidade	1966 – 1973
Sistemas baseados em conhecimento: Chave para o poder?	1969 – 1979
IA se torna uma indústria	1980 – Presente
O retorno das redes neurais	1986 – Presente
IA adota a metodologia científica	1987 – Presente
A emersão dos agentes inteligentes	1995 – Presente
Disponibilidade de bases de dados muito grandes	2001– Presente

Fonte: (RUSSEL; NORVIG, 2010)

A alternativa a esses métodos fracos era a utilização de modelos mais poderosos e baseados em conhecimento específico de domínio, que permitiam assim grandes passos de raciocínio que poderiam tratar mais facilmente casos típicos em uma área mais restrita do conhecimento. A importância do conhecimento sob o domínio se mostra importante também na área de processamento de linguagem natural, onde o conhecimento do domínio se torna efetivo para, por exemplo, resolução de ambiguidade e compreensão de referências a pronomes, já que é projetado especificamente para esse fim (RUSSELL; NORVIG, 2010).

Com um crescimento dos programas computacionais para resolver problemas do mundo real com a esse tipo de abordagem, aumentou também a demanda por linguagens para a criação desses sistemas, como o Prolog, PLANNER e outros, além da necessidade de esquemas de representação do conhecimento, como por exemplo a adoção de métodos estruturados que faziam a combinação de fatos sobre um determinado objeto, tipos de eventos relacionados e propriedades em grandes taxonomias hierárquicas, análogas a taxonomias da área de biologia (RUSSELL; NORVIG, 2010).

No campo da IA, a representação do conhecimento é uma das principais preocupações, e que apesar das diferentes definições de IA, algo sobre o que quase todas as abordagens concordam é que programas de IA envolvem o processamento de informação simbólica, ou seja, que os programas manipulam símbolos que de alguma forma representam pedaços de informação sobre o mundo com o propósito que executar uma tarefa que normalmente requer inteligência. E justamente pela importância dessa representação pelos símbolos sobre o mundo, a busca por formalismos que possam ser usados com essa finalidade é essencial para qualquer trabalho na IA (REICHGELT, 1990).

Segundo Sowa (1999) a representação do conhecimento se desenvolveu como um ramo da IA, mas passou a estar presente em diversos sistemas avançados que requerem inteligência humana para executar uma tarefa, tornando-se assim um assunto multidisciplinar, que aplica teorias e técnicas de diferentes campos como lógica, ontologias e computação. A lógica fornece a estrutura formal e regras para inferência; ontologias definem o tipo de coisas que existem no domínio da aplicação; e a computação suporta a aplicação que distingue a representação do conhecimento da filosofia pura. Sowa (1999) explica o relacionamento entre essas disciplinas:

Sem a lógica, a representação do conhecimento é vaga e sem critério para terminar se uma afirmação é redundante ou contraditória. Sem ontologia, os termos e símbolos são mal definidos, confusos e prejudiciais. E sem modelos computacionais, a lógica e a ontologia não podem ser implementadas em programas de computador. **Representação do conhecimento é a aplicação da lógica e ontologia na tarefa de construção de modelos computacionais para algum domínio.**

A visão de Sowa (1999) é complementar a hipótese da representação do conhecimento, apresentada por Brian Smith (1982) citado por Reichgelt (1990), que diz que qualquer processo inteligente possui ingredientes estruturais que observadores externos naturalmente utilizam para representar o conhecimento que aquele processo exhibe. Além disso, esses ingredientes desempenham um papel essencial e causal em criar o comportamento que aquele conhecimento manifesta.

Em outras palavras, se assume que qualquer sistema – seja humano ou um computador – que manifesta um comportamento inteligente, contém uma subestrutura chamada base de conhecimento, sendo essa base de conhecimento diretamente relacionada a codificação do conhecimento que o sistema tem disponível. A base de conhecimento é manipulada então por uma outra subestrutura, em geral chamada de motor de inferência, que processa os símbolos da base de conhecimento para gerar um comportamento inteligente (REICHGELT, 1990).

Segundo Reichgelt (1990), a filosofia utiliza o termo conhecimento no sentido de “crença verdadeira justificada”, mas que isso não se aplica exatamente na IA no contexto da representação do conhecimento, já que as linguagens de representação do conhecimento não se ocupam de afirmar que uma peça de conhecimento representado é verdadeiro ou justificado. Assim como, o fato de algum pedaço de informação estar representado usando linguagens de representação do conhecimento não torna essa informação verdadeira e nem a justifica. O autor inclusive argumenta que possivelmente seria melhor utilizar o termo linguagem de representação da informação ao invés de linguagem de representação do conhecimento.

Quanto ao formalismo na representação do conhecimento na IA, Reichgelt (1990) faz uma distinção em 4 níveis para discussão:

1. **Implementação:** As linguagens de representação do conhecimento são destinadas a utilização na representação da informação em um computador, e portanto, é essencial que seja possível construir um programa de computador subjacente à linguagem de representação do conhecimento. Sob um ponto de vista de notação, esse nível se preocupa com as estruturas de dados utilizadas na representação do conhecimento, e sob o ponto de vista de inferência, a preocupação está na descoberta e implementação de algoritmos que cheguem as inferências desejadas.
2. **Lógico:** Nesse nível, a respeito das notações, a principal preocupação é com as propriedades lógicas da linguagem de representação do conhecimento, onde duas questões devem ser levadas em conta: (i) qual o nível de formalismo permitido pela linguagem em questão; (ii) qual o poder de expressão da linguagem. E em relação à inferência, a preocupação é em relação as propriedades lógicas da linguagem.
3. **Epistemológico:** Para esse nível, as principais preocupações são a respeito dos tipos de primitivas estruturais necessárias para a representação do conhecimento, e quais são as estratégias de inferências que devem estar disponíveis.
4. **Conceitual:** Enquanto o nível epistemológico se preocupa sobre quais os tipos de primitivas estruturais necessárias, no nível conceitual a preocupação é sobre quais são as primitivas em si que devem ser incluídas na representação.

Além dos diferentes níveis de formalismo na representação do conhecimento, Reichgelt (1990) se baseia nas análises feitas por outros dois autores para estabelecer dois diferentes tipos

de conhecimento. De acordo com o autor, essa distinção se dá pelo fato de que um determinado tipo de conhecimento é mais facilmente representado do que o outro, e que isso pode ser levado em conta no momento da escolha da linguagem de representação a ser utilizada.

O primeiro tipo é o conhecimento de domínio, que contém informações sobre um domínio particular e que é usado por programas que precisam manipular informações de um domínio específico. Esse tipo, por sua vez, pode ser dividido em outros dois subtipos: (i) *conhecimento estruturado*, que se preocupa com os tipos das entidades que são relevantes para o domínio, pois em alguns casos um certo domínio possui uma estrutura muito particular a ser representada e a informação deve ser explicitamente descrita. Um exemplo de conhecimento estruturado seria o domínio de medicina interna, com os seus tipos específicos de entidades como pacientes, doenças, órgãos, drogas e etc.; (ii) o segundo tipo é o conhecimento relacional, onde o foco principal é a respeito das relações entre as entidades e fatos do domínio. Um exemplo da aplicação do conhecimento relacional seria a avaliação de que se um paciente está amarelo, ele pode ter doença do fígado, evidenciando assim o relacionamento entre as entidades e fatos acerca do conhecimento do domínio.

À parte do conhecimento de domínio, um sistema especialista pode possuir conhecimento de como utilizar o conhecimento de domínio para resolver um problema particular no domínio. Esse conhecimento de resolução de problemas é chamado por Reichgelt (1990) de conhecimento estratégico. Basicamente, a distinção entre os tipos de conhecimento apresentados está no nível de abstração, onde em geral o conhecimento de domínio está fortemente ligado ao conhecimento estrutural que o sistema ou agente possui.

A representação do conhecimento em bases do conhecimento, anteriormente citada, é o componente central dos agentes baseados em conhecimento. Esses agentes utilizam o conhecimento representado através das linguagens de representação do conhecimento para derivar o conhecimento implícito ali contido, e com isso inferir conclusões que levem o agente a resolver um determinado problema ou chegar a uma determinada conclusão. A base de conhecimento em seu estado inicial pode ser chamada de conhecimento prévio, e a cada passo de raciocínio realizado pelo agente, a base de conhecimento vai sendo atualizada com a nova visão de mundo que o agente possui, melhorando assim a sua capacidade de responder a diferentes questionamentos e situações (RUSSELL; NORVIG, 2010).

De forma geral em IA, uma base de conhecimento é um grupo de sentenças, onde cada sentença está na linguagem de representação do conhecimento escolhida, que reflete uma certa porção do mundo. Essa sentença pode ser chamada de axioma quando ela é formada sem derivar

de nenhuma outra sentença previamente existente. Uma dessas linguagens é a lógica, que se baseia em uma sintaxe que garante a formação correta das sentenças, e definem semântica para as sentenças que são expressadas. A semântica define a verdade a respeito de cada mundo possível representado (RUSSELL; NORVIG, 2010). Essa é uma visão abrangente de lógica, mas existem diversas variações e sintaxes diferentes como a lógica Escolástica, matemática, booleana, de Frege, algébrica, proposicional, etc. (SOWA, 1999).

A utilização de diferentes modelos de lógica é algo já antigo na IA, e há alguns argumentos a favor e contra a sua utilização. Reichgelt (1990) cita a semântica e a expressividade como dois fatores a favor da utilização de lógica. Através da semântica, é possível saber exatamente o que uma sentença representa. Também permite que se saiba se o processo de inferência chegou a uma conclusão livre de ruídos. Já com a expressividade, se apresentam dois fatores relevantes, (i) primeiramente permitir que a lógica expresse conhecimento incompleto; (ii) e a existência de diferentes lógicas que se pode utilizar, conforme citado anteriormente.

Ademais, o autor também cita conhecidos problemas com a lógica e possíveis soluções a serem adotadas. O primeiro problema levantado por Reichgelt (1990) é a ineficiência da lógica para resolução de problemas, pois ao passar do tempo os programas passam a ficar lentos de forma que se torna pouco útil à sua utilização. Isso acontece principalmente por controles mal implementados e bases de dados que não fornecem afirmações suficientes para um raciocínio mais completo.

Outro problema levantado é que alguns problemas em lógica acabam não tendo uma solução definitiva, ou simplesmente não são solucionáveis, o chamado problema de indecidibilidade. Problema esse que pode ser resolvido, por exemplo, tornando a linguagem menos expressiva e a simplificando, tornando assim o raciocínio mais direto. Em seguida, vem o problema do conhecimento procedural, que por se tratar a lógica de um conhecimento declarativo, pode implicar na capacidade de resolver problemas mais complicados (REICHGELT, 1990).

Com certeza há uma outros problemas que podem ser identificados no momento da representação do conhecimento através da lógica, ou mesmo no momento da inferência pela derivação dos axiomas e afirmações contidas na lógica para um determinado problema. Porém, apesar de descrever suficientemente bem fatos e afirmações sobre um domínio, a semântica da lógica não é totalmente completa.

Na lógica, a utilização do símbolo \exists denota a existência de algo no contexto da expressão, mas a lógica em si não possui vocabulário para descrever essa “coisa” que existe, e nesse ponto as ontologias preenchem essa lacuna por se tratarem do estudo da existência, de todos os tipos de coisa que compõe o mundo. Duas fontes para as categorias ontológicas são a observação e o raciocínio. A observação fornece o conhecimento do mundo físico, e o raciocínio faz uso da observação gerando uma estrutura de abstrações chamada metafísica (SOWA, 1999).

Para Sowa (1999), a escolha das categorias ontológicas na representação do conhecimento é o primeiro passo no projeto de um banco de dados ou de um sistema orientado à objetos. A escolha dessas categorias vai determinar tudo que pode ser representado na aplicação em software, ou em toda uma família de aplicações. Qualquer incompletude, distorção ou restrição na estrutura das categorias, inevitavelmente irá limitar a generalidade de qualquer programa ou banco de dados que utilize essas categorias.

A importância das categorias na representação do conhecimento pelas ontologias é reforçada por Russel e Norvig (2010), ao afirmarem que a organização dos objetos em categorias é vital para a representação do conhecimento, e que apesar da interação com o mundo real se dar através de objetos individuais, muito do raciocínio se dá no nível das categorias. As categorias também são importantes para se fazer previsões sobre objetos após eles serem classificados, como por exemplo inferências podem ser realizadas de acordo com a presença de certos objetos com determinadas características em cada categoria, e usando essas inferências, fazer previsões e afirmações sobre os objetos.

Através dos trabalhos de Reichgelt (1990), Gruber (1993), Guarino e Giaretta (1995) e Sowa (1999) – além de diversos trabalhos na CC que usam como base os autores listados - é possível concluir que sistemas mais complexos, ou os que necessitam de uma representação mais semântica do conhecimento sobre o domínio, a ontologia é uma importante ferramenta a ser considerada. Grimm, Hitzler e Abecker (2007) levantam alguns exemplos de como aplicações computacionais podem se beneficiar de modelos conceituais de domínios formalizados em ontologias:

- **Integração da informação:** Integrar fontes heterogêneas de informação através da definição de um esquema, onde a ontologia pode ser o mediador entre os esquemas, permitindo que informações de diferentes bases de dados sejam interpretadas e passadas em ambos os sentidos.

- **Recuperação da informação:** Através do uso de ontologias é possível aumentar a precisão dos resultados levando em consideração informações semânticas contidas em pesquisas e documentos, elevando palavras-chaves a conceitos e relações ontológicas.
- **Gestão de conteúdo melhorada ontologicamente:** Em diversas áreas da computação, os dados de fato processados são marcados com metadados por diversas razões, e ontologias fornecem um vocabulário específico do domínio para a realização dessas marcações. Além disso, a formalidade da representação das ontologias permite a automação desse processo e a sua integração a ontologia do domínio, facilitando a interpretação pela máquina.
- **Gestão do conhecimento e portais comunitários:** Em algumas companhias, associações ou grupos, o conhecimento individual é visto como um recurso estratégico e deve ser compartilhado e sistematicamente mantido, o que é chamado de gestão do conhecimento. As ontologias podem auxiliar essa gestão unificando os esforços de gestão através de um modelo conceitual compartilhado, conectando sistemas técnicos de navegação, armazenagem, recuperação e troca de conhecimento comunitário.
- **Sistemas especialistas:** Para diversos domínios, como medicina ou direito, é desejável simular um especialista no domínio que possa responder a perguntas sofisticadas, e nos sistemas especialistas isso pode ser atingido pela incorporação de uma ontologia de domínio desenvolvida para formalizar conhecimento de um especialista do determinado domínio. Essas perguntas podem então ser respondidas pelo agente inteligente através de raciocínio sobre o conhecimento representado.

Os exemplos apresentados por Grimm, Hitzler e Abecker (2007) vem para complementar a abordagem sobre a importância da representação do conhecimento na CC e como os campos e áreas correlatas, como IA, SI e ES que utilizam as ontologias como uma importante ferramenta de suporte, principalmente quanto a estrutura de dados. A seguir, será explorada a representação do conhecimento com ontologias na CI e como isso é utilizado por outras ferramentas já existentes na área.

3.3 Representação do conhecimento na Ciência da Informação

Vickery (1986), citando Samuel Johnson, traz que o conhecimento é de dois tipos, um assunto que sabemos por conta própria ou que sabemos onde encontrar informação sobre, e Vickery afirma que a documentação é particularmente preocupada com o segundo tipo de conhecimento, e que todo conhecimento é simbólico. E que, portanto, a representação do conhecimento é um assunto que tem ocupado o mundo da documentação desde o princípio.

Ainda sobre a simbologia e representação, ao falar sobre “Informação como Coisa”, Buckland (1991) lança mão de uma abordagem pragmática, onde busca identificar os principais agrupamentos na utilização do termo informação, e mesmo sendo essa uma abordagem completa e que não irá satisfazer a quem deseja obter uma exata definição para o termo, o autor lista a partir disso, os três principais usos do termo informação:

- (1) **Informação como processo:** Quando alguém é informado, o que essa pessoa sabe é modificado. Nesse sentido, “informação” é “o ato de se informar; comunicação do conhecimento ou “notícia” sobre um fato ou acontecimento; ação de contar ou o fato de ser dito sobre algo.” (*Oxford English Dictionary*, 1989, vol. 7, p. 944).
- (2) **Informação como conhecimento:** “Informação” é utilizado para denotar aquilo que é percebido no “processo informacional”: o “conhecimento comunicado se preocupa com um fato em particular, assunto ou acontecimento; daquilo que se é informado ou dito; inteligência, notícias” (*Oxford English Dictionary*, 1989, vol. 7, p. 944). A noção de que informação é aquela que reduz incertezas pode ser visto como um caso especial de “Informação como conhecimento”.
- (3) **Informação como coisa:** O termo “informação” também costuma ser atribuído a objetos, como dados e documentos, que são referenciados como “informação” porque eles são considerados informativos, por “ter a qualidade de transmitir conhecimento ou comunicar informações; instrutivo.” (*Oxford English Dictionary*, 1989, vol. 7, p. 946).

Uma característica chave da informação como conhecimento é ela ser intangível, já que conhecimento, crenças e opiniões são pessoais, subjetivas e conceituais. Então, para que seja possível a comunicação da informação como conhecimento, isso precisa ser expressado, descrito ou representado de alguma maneira em um meio físico como um sinal, texto ou outro

elemento da comunicação. E então, qualquer expressão, descrição ou representação pode ser a “informação como coisa” (BUCKLAND, 1991).

Como é possível notar pelos textos de Vickery e Buckland, há uma clara ligação entre os termos informação e conhecimento, mas que ao mesmo tempo causam discussão acerca de qual seja o mais correto no âmbito da CI, ou em qual situações cada um se aplica. No trabalho “Organização da Informação ou Organização do Conhecimento?”, Brascher e Café (2008) tratam dessa discussão, onde apresentam diversas visões de diferentes autores na busca de delimitar a utilização dos termos organização da informação, organização do conhecimento, representação da informação e representação do conhecimento no domínio da CI.

Para a compreensão dos termos informação e conhecimento, Brascher e Café (2008) estabelecem que é preciso “a) relacionar seus conceitos às funções que damos a eles nos contextos em que se inserem; b) diferenciá-los de conceitos próximos a eles incluídos no sistema referencial”. Citando Fernandez-Molina, Brascher e Café (2008) afirmam ser possível encontrar contribuições que estabelecem distinções mais claras entre os termos:

dados, informação e conhecimento: os dados são informação potencial, que somente são percebidos por um receptor se forem convertidos em informação e esta passa a converter-se em conhecimento no momento em que produz uma modificação na estrutura do conhecimento do receptor (FERNANDEZ-MOLINA, 1994, p.328)

Seguindo a linha de raciocínio que atribui ao conhecimento a condição de informação processada, seja por qualquer meio, as autoras trazem ainda a visão de Burke que compara informação e conhecimento aos ingredientes de uma refeição, sendo a informação algo cru e o conhecimento algo que foi cozido. E complementam com a visão de Fogl, de que a informação compreende três elementos: (i) conhecimento (conteúdo da informação); (ii) linguagem (um instrumento de expressão de itens da informação); e (iii) suporte (objetos materiais ou energia). Fogl também diz não haver uma conexão direta entre o objeto e informação, uma vez que a fonte da informação é o conhecimento humano e a consciência humana, mas não o próprio objeto e que a informação deve ser analisada sobre três aspectos: semântico, pragmático e quanto a fixação do conhecimento e dos juízos de valor e o material de suporte utilizado. Então, Brascher e Café (2008) fazem uma síntese das características que consideram as mais importantes acerca dos conceitos de informação e conhecimento, que são:

- 1) Conhecimento é o resultado da cognição (processo de reflexão das leis e das propriedades de objetos e fenômenos da realidade objetiva na consciência humana);

- 2) Conhecimento é o conteúdo ideal da consciência humana;
- 3) Informação é uma forma material da existência do conhecimento;
- 4) Informação é um item definitivo do conhecimento expresso por meio da linguagem natural ou outros sistemas de signos percebidos pelos órgãos e sentidos;
- 5) Informação existe e exerce sua função social por meio de um suporte físico;
- 6) Informação existe objetivamente fora da consciência individual e independente dela, desde o momento de sua origem.

Tomando como base a diferenciação entre informação e conhecimento exposta por Brascher e Café (2008), é importante também passar pela discussão sobre a representação da informação e do conhecimento, que também podem ser entendidas de formas diferentes. As autoras discordam do pensamento de Hjørland e Anderson de que a organização do conhecimento é a descrição de documentos, seus conteúdos, características e objetivos, e a organização dessas descrições afim de tornar esses documentos e suas partes acessíveis a pessoas que os procuram ou procuram por mensagens neles contidas, e que a organização compreende todo o tipo de método de indexação, resumo, catalogação, classificação, gestão de arquivos, bibliografia e a criação de bases de dados bibliográficas e textuais para recuperação da informação.

Para Brascher e Café (2008) isso se trata de organização da informação, que na visão das autoras são dois processos distintos de organização, sendo o de organização da informação algo que se aplica às ocorrências individuais de objetos informacionais, inclusive a organização desses objetos para arranjá-los sistematicamente em coleções, como em bibliotecas, museus, arquivos, tanto tradicionais quando digitais. Enquanto a organização do conhecimento se aplica a unidades do pensamento (conceitos), visando a construção de modelos de mundo que se constituem em abstrações da realidade.

Sendo assim, os dois processos de representação produzem resultados diferentes, conforme propõem Bascher e Café (2008):

Esses dois processos produzem, conseqüentemente, dois tipos distintos de representação: a representação da informação, compreendida como o conjunto de atributos que representa determinado objeto informacional e que é obtido pelos processos de descrição física e de conteúdo, e a representação do conhecimento, que se constitui numa estrutura conceitual que representa modelos de mundo, os quais, segundo Le Moigne (apud CAMPOS, 2004, p.23), permitem descrever e fornecer explicações sobre os fenômenos que observamos.

A organização do conhecimento é definida pelas autoras então como “o processo de modelagem do conhecimento que visa a construção de representações do conhecimento”. O processo se baseia na análise do conceito e suas características para um posterior estabelecimento da posição que o conceito deve ocupar em um determinado domínio, bem como a relação com os demais conceitos presentes no sistema, e reunir o que se conhece em uma estrutura sistematicamente organizada (BRASCHER; CAFÉ, 2008).

Isso leva então, segundo Brascher e Café (2008), a representação do conhecimento, que é feita através de diferentes tipos de sistemas de organização do conhecimento (SOC), sistemas conceituais que representam determinado domínio através da sistematização dos conceitos e das relações semânticas entre si. Em 1998 o *Networked Knowledge Organization Systems Working Group* propôs a criação do termo *knowledge organization systems* (KOS) para englobar sistemas de classificação, cabeçalhos de assunto, arquivos de autoridade, redes semânticas e ontologias. Hodge (2000), sugere uma taxonomia para os tipos de SOC, ampliando a abrangência inicialmente proposta para contemplar também dicionários, glossários, taxonomias e tesouros.

Brascher e Café (2008) resumem os objetivos dos SOC, propostos por Soergel:

prover uma mapa semântico para domínios individuais e para os relacionamentos entre domínios, fornecendo orientação e servindo como um instrumento de referência; melhorar a comunicação e o ensino; prover uma base conceitual para a boa execução da pesquisa e implementação; prover classificação para a ação, isto é, o uso prático dos SOC em diferentes atividades profissionais, tais como a classificação de doenças para diagnósticos médicos e de mercadorias para o comércio; apoiar a recuperação da informação; prover uma base conceitual para sistemas baseados em conhecimento e para a definição de elementos de dados e hierarquias de objetos na engenharia de software, servir como um dicionário mono, bi ou multilíngue para uso pelo homem ou por sistemas automáticos de processamento da linguagem natural.

No contexto apresentado, considerando a descrição de recursos informacionais, os SOC cumprem o papel de padronizar a representação da informação no que se refere a identificação do assunto. E segundo Vickery (2008), no contexto da recuperação da informação, os SOC são “instrumentos complementares que ajudam o usuário a encontrar seu caminho no texto”.

As conclusões e delimitações apresentadas por Brascher e Café (2008) são de certa forma compatíveis e complementares ao estudo de Alvarenga (2003), que falava sobre a representação do conhecimento acontecer em três momentos distintos: na produção de registros de conhecimento, na organização dos sistemas documentais e no acesso às informações pelos

usuários. Além da separação em representação primária e secundária, onde a primária parte da ideia de que, representar significa o ato de colocar algo no lugar de, e então “classifica-se em nível primário a representação feita pelos autores no momento da expressão dos resultados de seus pensamentos, estes derivados de observações metódicas da natureza e dos fatos sociais, utilizando-se das linguagens disponíveis no contexto da produção e comunicação de conhecimentos”.

Os registros advindos dessa representação passam então integrar acervos de arquivos, bibliotecas, serviços ou centros de documentação/informação, passando por uma nova representação (representação secundária), para que possam então ser incluídos em sistemas documentais referenciais. Ambas as representações são processos cognitivos, já que a percepção, classificação e criação de conhecimento sobre os seres são prerrogativas da racionalidade humana, onde é possível identificar processos cognitivos básicos como levantamento das características percebidas, comparação entre estas e as já identificadas nos seres já conhecidos, caracterizando-se essa ocorrência como processos classificatórios /cognitivos (ALVARENGA, 2003).

Segundo Alvarenga (2003), uma das grandes dificuldades que vem sendo ressaltadas na tentativa de classificação dos seres, coisas e textos se trata do fato de que esse processo se baseia na constatação de que “as coisas e os conhecimentos que lhe são correspondentes não se reduzem ao que deles pode ser visto explicitamente”. Alvarenga (2003) trata então da preocupação que se deve ter ao se trabalhar com a classificação:

não é suficiente que se capte superfícies de emergência, essências, características, ou que se considere nos documentos apenas as unidades lexicais (palavras) deles constantes. Torna-se necessário que se aprofunde no conhecimento das relações entre as similitudes e diferenças entre conhecimentos e objetos integrantes de uma coleção. O processo compreende ainda uma atitude hermenêutica ou interpretativa, igualmente indispensável à análise e à classificação de um universo de coisas. O ato de interpretar é inerente a qualquer abordagem aos conhecimentos, qualquer que seja o grau de suposta clareza com que esses tenham sido produzidos.

Farias e Pinho (2016) destacam então as ontologias como ferramenta na organização e representação do conhecimento, salientando que essas trazem em sua semântica o estudo das coisas, objetos e seres do mundo, abordando suas relações e restrições em um domínio específico. Permitindo então que um determinado contingente informacional venha a ser representado, através de especificação formalmente descrita, viabilizando que essas informações possam ser processadas e decifradas por máquinas, de forma que o domínio que

se deseja abordar possa ser organizado e representado, tendo em vista os diferentes modelos, tipos de relações, sendo um modelo mental de uma dada realidade.

Diante do exposto anteriormente, é possível identificar as ontologias não somente como uma estrutura acessória a representação do conhecimento, mas como uma ferramenta semântica no processo informacional como um todo. Ramalho (2006) identifica o potencial das ontologias ao defini-las como:

um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal, a partir da descrição dos aspectos semânticos de conteúdos informacionais, possibilitando a realização de inferências automáticas [...]

Diferentes instrumentos na área da CI são comparados às ontologias, principalmente tesouros e taxonomias, muito em parte pela possibilidade do controle terminológico e da possibilidade de expressar relação entre termos e/ou conceitos. Ramalho (2010) explicita as diferenças de concepção entre tesouros e ontologias:

Ao longo dos últimos anos inúmeros estudos comparativos entre ontologias e tesouros têm constatado que, apesar de possuírem características comuns, tais instrumentos caracterizam-se como diferentes modelos de representação do conhecimento. Enquanto os tesouros são desenvolvidos como ferramentas de auxílio para os usuários na busca de informações, as ontologias têm como principal objetivo descrever formalmente os recursos informacionais para possibilitar a realização de inferências automáticas.

Em relação às taxonomias, a diferença é percebida na estrutura das ontologias, que são mais completas por apresentarem novos componentes estruturais que vão além de relações hierárquicas. Pode-se dizer que as ontologias permitem uma maior liberdade na representação de tipos de relacionamentos que não seriam possíveis em outros modelos de representação, podendo ser criadas a partir de diferentes técnicas de representação e organização, cabendo a aqueles que as desenvolvem fazer as escolhas de design e projeto no momento da modelagem e implementação.

A seguir, serão listadas as principais metodologias e boas práticas na criação, desenvolvimento e manutenção de ontologias e também serão destacadas quais ferramentas os especialistas tem à disposição para a construção de ontologias.

4

Construção de Ontologias

O campo da engenharia ontológica foi assunto de várias pesquisas nas últimas duas décadas. Esse campo se refere ao conjunto de atividades que se preocupam com o processo de desenvolvimento de ontologias, incluindo o ciclo de vida, princípios, métodos e metodologias para construção de ontologias, assim como as ferramentas e linguagens que suportam essas atividades (CORCHO; FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2006).

Neste capítulo será então abordado o processo de construção, engenharia e aprendizado de ontologia, visando estabelecer o que são exatamente esses termos, quais são as diferentes abordagens, metodologias e ferramentas empregadas em cada fase do processo de criação de novas ontologias.

4.1 Metodologias de construção de ontologias

Desde os anos 1990, várias abordagens foram publicadas sobre o processo de construção de ontologias. No ano de 1990, Lenat e Guha publicaram os passos que seguiram para criar ontologias no projeto Cyc, que se tratava da construção de uma base de conhecimento composta por uma enorme quantidade de conhecimento de senso-comum (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2002).

O método utilizado para construir as bases de conhecimento consistia em 3 fases: A primeira era a codificação manual de artigos e peças que poderiam representar conhecimento, sendo indicado manualmente quais eram considerados senso-comum. As segunda e terceira fases eram para a extração de conhecimento em cima da base já existente, usando linguagem natural e aprendizado de máquina. A diferença entre elas é que na segunda fase, havia uma

maior supervisão humana, enquanto na terceira, era praticamente toda autônoma (CORCHO; FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2003).

Trabalhos na mesma década no campo da IA também vinham explorando o uso formal de ontologias. Em um desses estudos, Gruber (1993) definiu ontologia como "uma especificação explícita de uma conceituação", e usou essa definição para representar conhecimento de um determinado domínio, pois para sistemas de IA algo que exista é necessário que possa ser representado.

A ontologia definia então um vocabulário, o qual era usado para consultas e afirmações que eram trocadas entre agentes. Os relacionamentos ontológicos funcionavam como acordos de compartilhamento de vocabulário coerente e consistente entre esses agentes (GRUBER, 1993).

Como o conhecimento pode ser então representado em uma ontologia, quando se cria uma ontologia está se fazendo escolhas de design. Para guiar essas escolhas de design, Gruber (1993) definiu alguns princípios para a construção de ontologias:

- **Clareza:** Uma ontologia deve comunicar de forma eficaz o significado desejado de cada termo, as definições devem ser objetivas. Enquanto a motivação de definição de um conceito pode vir de uma situação social ou requisito computacional, a definição deve ser independente de contexto social ou computacional. Todas as definições devem ser documentadas em linguagem natural.
- **Coerência:** Deve permitir inferências que sejam consistentes com as definições. No mínimo, os axiomas devem ser logicamente consistentes. E a coerência também deve ser aplicada aos conceitos que são definidos informalmente, como os descritos em linguagem natural. Se uma sentença que pode ser inferida dos axiomas contradiz a definição ou mesmo algum dos exemplos fornecidos informalmente, a ontologia é incoerente.
- **Extensibilidade:** Uma ontologia deve ser desenhada para antecipar o uso de vocabulário compartilhado. Deve oferecer uma base conceitual para antecipar uma série de tarefas, a representação deve ser feita de maneira que qualquer um possa estender e especializar a ontologia de forma monolítica. Como por exemplo, definir novos termos para usos específicos, baseados no vocabulário existente de uma maneira que não necessite uma revisão das definições existentes.

- **Mínimo viés de codificação:** A conceituação deve ser especificada pelo nível do conhecimento, sem depender de um entendimento particular de simbologia do domínio. Assim, a representação da ontologia deve ser feita de forma pura, podendo ser compreendida por diferentes sistemas e estilos de representação.
- **Mínimo comprometimento ontológico:** Uma ontologia deve requerer o mínimo necessário de comprometimento ontológico para suportar o compartilhamento de conhecimento desejado. A ontologia deve depender o mínimo possível da necessidade de o mundo ser modelado, permitindo assim que as partes tenham liberdade de especializar e instanciar a ontologia de acordo com as suas necessidades.

O objetivo da criação desses 5 princípios era servir como uma avaliação das decisões de desenho de uma nova ontologia, ajudando assim a escolher de forma apropriada maneiras de conceituar o conhecimento de um domínio, através de uma modelagem construtiva (GRUBER, 1993).

Em 1995, dois outros projetos que utilizaram ontologias em seus estudos também publicaram seus processos. Esses projetos foram o Enterprise Ontology e TOVE (TOronto Virtual Enterprise) (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2002).

Durante o processo de desenvolvimento do projeto *Enterprise Ontology*, Uschold e Grüninger sentiram a necessidade de usar uma metodologia para o desenvolvimento da ontologia. Metodologia essa, que também é baseada na experiência prática adquirida ao longo do projeto, que consistia na construção de uma ontologia para modelar processos corporativos no *Artificial Intelligence Application Institute*, em Edimburgo (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2002).

A metodologia proposta por Uschold e Grüninger previa 4 atividades:

1. **Identificar o propósito da ontologia:** Tornar claro porque a ontologia está sendo construída e qual o uso pretendido para ela.
2. **Construção da ontologia:**
 - a. **Capturar ontologia:** Identificar os conceitos chave e relacionamentos no domínio de interesse, e assim produzir definições precisas e sem ambiguidades.
 - b. **Codificação:** Representar o conhecimento adquirido na tarefa anterior em uma linguagem formal de representação de ontologia.
 - c. **Integração com ontologias existentes:** Durante a captura e codificação, deve haver o questionamento de como e quando usar ontologias já existentes.

3. **Avaliação da ontologia:** Utilização de um ambiente computacional para validar a ontologia, através de uma janela de referência, comparando com os resultados de acordo com requisitos e questões de competência.
4. **Documentação:** Recomenda-se que o autor defina um conjunto de regras para representação e documentação da ontologia, que pode variar de acordo com o tipo e propósito da ontologia.

Os autores também propuseram três estratégias para identificação dos principais conceitos de uma ontologia:

- **Top-down:** Conceitos mais abstratos são identificados primeiro, e então são especializados em conceitos mais específicos.
- **Bottom-up:** Conceitos mais específicos são identificados primeiro e então generalizados em conceitos mais abstratos.
- **Middle-out:** Os conceitos mais importantes são identificados antes e então, são generalizados e especializados em outros conceitos.

Uschold e Grüninger também mencionaram que somente as fases que haviam descrito não eram suficientes para compor uma metodologia. Pois toda metodologia também deveria incluir um conjunto de técnicas, métodos e princípios para cada uma das quatro fases listadas anteriormente. Além de indicar qual o relacionamento existente entre cada uma das fases propostas (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2002).

Outra metodologia proposta em 1995, por Grüninger e Fox, foi baseada no trabalho realizado no projeto TOVE (*TOronto Virtual Enterprise*), que buscava a criação de uma base de conhecimento usando lógica de primeira ordem (CORCHO; FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2003).

O processo proposto por Grüninger e Fox consistia em identificar o cenário de motivação, que servirá para guiar o conjunto de solução que se deseja alcançar. Depois, baseado nesse cenário de motivação, são estabelecidas perguntas de competência informais (em linguagem natural) para fornecer a justificativa para essa nova (ou extensão da) ontologia.

Depois que as perguntas de competência informal são elaboradas, a terminologia da ontologia deve ser expressa utilizando lógica de primeira ordem, que tem como objetivo determinar os objetos do domínio de discurso. Eles vão ser representados por constantes e variáveis na linguagem.

Após essa representação, serão formuladas as questões de competência formal, usando a mesma linguagem utilizada para a representação da ontologia. Após a criação das questões, são especificados os axiomas e definições, também utilizando a linguagem formal da ontologia. Por fim, estabelecer os teoremas de integralidade, que vão completar a relação entre os axiomas e os elementos da lógica de primeira ordem (GRÜNINGER; FOX, 1995).

Com o crescimento de trabalhos a respeito da construção e utilização de ontologias, Uschold (1996) propôs uma abordagem unificada, pois apesar de já terem sido publicados alguns trabalhos com metodologias, e princípios para criação de ontologias, sendo o autor esses trabalhos se tratavam mais de tentativas de identificar e descrever o processo de desenvolvimento de ontologias, e refletiam estudos de casos de uma única ontologia ou listavam conjuntos de recomendações que eram limitados a um projeto em particular.

Sendo assim, Uschold (1996) definiu como objetivo de longo prazo colocar todo o processo em um *framework* coerente, que poderia fornecer orientação para qualquer um que desejasse criar uma ontologia.

Porém, a maior barreira para a criação de um *framework* coerente e que unificasse todas as técnicas e métodos apresentados até então, dizia respeito a inexistência de uma visão clara de quão genérica cada técnica ou método era. Assim, não havia regras ou suporte na decisão de quais métodos e técnicas deveriam ser aplicadas (USCHOLD, 1996).

A abordagem unificada proposta por Uschold (1996) foi uma combinação das metodologias apresentadas pelos projetos *Enterprise Ontology* e TOVE, com o objetivo de possibilitar aos futuros construtores de ontologias um melhor entendimento de como escolher a técnica mais apropriada para cada circunstância.

Mesmo após a proposta de um *framework* unificado, o processo de criação de ontologias continuava mais artesanal do que uma atividade de engenharia em si. Pois cada equipe de desenvolvimento seguia o seu próprio conjunto de princípios, critérios de *design* e etapas no processo de desenvolvimento de ontologias (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2002).

Inspirados na proposta unificada de Uschold, pesquisadores do Laboratório de Inteligência Artificial da Universidade Politécnica de Madrid (UPM) criaram a metodologia METHONTOLOGY, para a construção de ontologias a partir do zero, pela reutilização de ontologias ou através de processos de reengenharia (CORCHO; FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ, 2003).

De acordo com Fernández, Gómez-pérez e Juristo (1997), o processo de desenvolvimento de uma ontologia pela METHONTOLOGY se refere a quais atividades devem ser cumpridas, entretanto não determinam qual a ordem de execução dessas atividades. Segundo os autores, as atividades do processo de construção de ontologias são:

- **Planejamento:** Antes de construir a ontologia é necessário planejar as principais tarefas a serem feitas, como elas serão organizadas, quanto tempo será gasto em cada uma e quais recursos serão necessários.
- **Especificação:** É necessário conhecer o propósito e o escopo de uma ontologia. Para isso, responda as perguntas: porque está ontologia está sendo construída e quais são os usos desejados e os usuários finais? Depois, especifique ou anote essas respostas em um documento de especificações de requisitos.
- **Aquisição de conhecimento:** Para que a ontologia atenda ao propósito, ela precisa se basear em conhecimento formal. É preciso então adquirir o conhecimento, listando bases de conhecimento e outras fontes que serão utilizadas no processo.
- **Conceitualização:** Uma vez que existe conhecimento suficiente, é necessário construir um modelo conceitual que descreva o problema e a sua solução.
- **Integração:** Ontologias são construídas para serem reutilizadas, sendo assim, tente integrar o máximo possível com ontologias já existentes.
- **Implementação:** Para que a ontologia seja processável por computador, é necessário implementar essa ontologia utilizando uma linguagem formal.
- **Avaliação:** Antes de tornar a ontologia disponível a outros usuários, avalie a ontologia criada. Faça julgamentos técnicos, de acordo com quadros de referência.
- **Documentação:** Documente a ontologia para facilitar o compartilhamento e reutilização da ontologia.
- **Manutenção:** A qualquer momento e em qualquer lugar, alguém pode solicitar para incluir ou alterar termos e definições na ontologia. Então, a manutenção de uma ontologia é algo sério e que deve ser feito com cuidado.

A METHONTOLOGY também se preocupa em falar do ciclo de vida de uma ontologia, mas assim como no processo de desenvolvimento, não é determinado a ordem ou profundidade de cada estado, mas sim a identificação de cada estágio que uma ontologia passa durante sua existência. Além disso, o ciclo de vida de uma ontologia determina quando uma atividade deve

ser executada para que se mova de um determinado estado para outro (FERNÁNDEZ; GÓMEZ-PÉREZ; JURISTO, 1997).

A reutilização de ontologias pré-existentes é uma peça chave na criação de novas ontologias, permitindo que haja uma maior velocidade e refinamento no processo de construção de ontologias. Isso acontece de forma análoga na área de desenvolvimento de software, e o conceito de reengenharia é comumente utilizado na Engenharia de Software e aos poucos vem se movendo para o campo da Engenharia do Conhecimento (GÓMEZ-PÉREZ; ROJAS-AMAYA, 1999).

Quando se tenta definir o conceito de reengenharia, outros conceitos próximos surgem, como engenharia reversa, reestruturação e engenharia direta. O termo engenharia reversa pode ser utilizado para designar o processo de análise de um sistema para identificar seus componentes e relações, e/ou para representar o sistema de uma outra maneira. Já a reestruturação pode ter diversas definições, e a mais representativa foi feita por Chikofsky que definiu a reestruturação como uma transformação no sistema para passar de uma representação para outra em um mesmo nível de abstração, que é independente do projeto de implementação, no sentido físico da implementação de um sistema (GÓMEZ-PÉREZ; ROJAS-AMAYA, 1999).

E assim, Gómez-Pérez e Rojas-Amaya (1999) definem como reengenharia de ontologias o processo de recuperação e mapeamento do modelo conceitual de uma ontologia implementada para outra, mais aplicável a outro modelo conceitual, que é reimplementado. Assim como definido no processo de Engenharia de Software, Gómez-Pérez e Rojas-Amaya (1999) também propõe três passos:

- **Engenharia reversa:** O objetivo é gerar como saída um possível modelo conceitual, baseado no código implementado pela ontologia. Com o propósito de construir um modelo conceitual, é utilizado o conjunto de representações intermediárias proposto pela metodologia METHONTOLOGY.
- **Reestruturação:** O objetivo com a reestruturação é reorganizar o modelo conceitual inicial em um novo modelo que é construído tendo em mente o uso da ontologia reestruturada ou por uma aplicação que a reutilize. Com isso, não há maneira de garantir que a ontologia reestruturada será totalmente válida para todas as ontologias que reutilizem o conhecimento.
- **Engenharia direta** (*Forward Engineering*): O objetivo é fazer uma nova implementação da ontologia, baseada no novo modelo conceitual.

Em 2001 a equipe responsável pelo desenvolvimento e manutenção do software Protégé² publicou um documento com o título “*Ontology Development 101: A Guide to Creating Your First Ontology*”. Segundo os autores, o guia buscou fornecer um ponto de partida para ajudar no processo de criação de ontologias. E apesar de sugerirem uma série de passos e recomendações a serem seguidas, segundo o guia não há uma única e correta metodologia de criação de ontologias definitiva e não é objetivo do guia definir uma. As ideias apresentadas foram as que a equipe entendeu como útil na própria experiência de criação de ontologia (NOY; MCGUINNESS, 2001).

O guia apresenta o que chama de “Uma simples metodologia de engenharia do conhecimento”, onde descrevem uma abordagem iterativa para o desenvolvimento de uma ontologia, iniciando com uma passagem inicial superficial para criar a ontologia, e depois há o refinamento e evolução da ontologia para preenchimento dos detalhes. Para que isso seja possível, Noy e McGuinness (2001) enfatizam o que chamam de “três regras fundamentais no design de ontologias” as quais se referem diversas vezes durante a metodologia. São elas:

- 1) Não existe um único modo correto de se modelar um domínio, há sempre alternativas viáveis. A melhor solução quase sempre depende da aplicação que se tem em mente e as extensões que se consegue antecipar.
- 2) O desenvolvimento de ontologias é necessariamente um processo iterativo.
- 3) Conceitos em ontologias devem se aproximar a objetos (físicos ou lógicos) e a relacionamentos no domínio de interesse. Possivelmente serão substantivos (objetos) ou verbos (relacionamentos) em sentenças que descrevem o domínio.

A metodologia proposta é dividida em 7 passos, conforme detalhamento abaixo:

Passo 1: Determinar o domínio e o escopo da ontologia.

Para iniciar o processo de construção de uma ontologia deve se responder algumas perguntas para definir o domínio e o escopo, como por exemplo: (i) Qual é o domínio que a ontologia irá cobrir; (ii) Para o que será utilizada a ontologia; (iii) Quais tipos de questões a informação contida na ontologia devem ser respondidas; (iv) Quem irá usar e manter a ontologia.

² Protégé é um software *open-source* para edição de ontologia e um *framework* para a criação de sistemas inteligentes e gerenciamento de bases de conhecimento. Mais detalhes serão encontrados na próxima seção deste capítulo.

As respostas a essas perguntas podem mudar durante o processo de design, mas a qualquer momento elas irão ajudar a limitar o escopo do modelo.

Outra maneira de determinar o escopo de uma ontologia é esboçar uma lista de questões que a base de conhecimento da ontologia deve ser capaz de responder, as perguntas de competência como descrito por Gruninger e Fox (1995). Essas questões irão server posteriormente como um teste decisivo, por exemplo: A ontologia possui informação suficiente para responder a esses tipos de questões? As respostas requerem um nível particular de detalhamento ou representação de uma área particular? Essas questões de competência devem ser apenas um esboço e não precisam ser exaustivas.

Passo 2: Considerar a reutilização de ontologias existentes.

Normalmente se deve considerar que alguém já construiu algo similar e que dê para ser aproveitado, pois a reutilização de ontologias pode em certos casos ser um requisito caso o sistema no qual a ontologia vai ser utilizada precise interagir com outras aplicações que já estão comprometidas semanticamente com outras ontologias ou vocabulários controlados.

Passo 3: Enumerar os termos importantes na ontologia.

É útil criar uma lista abrangente com os todos termos os quais se deseja utilizar ao longo do desenvolvimento da ontologia. Inicialmente os termos devem ser selecionados sem a preocupação com sobreposição aos conceitos, relacionamentos entre os termos ou qualquer propriedade que pode haver em algum conceito.

Os passos que serão apresentados a seguir estão muito ligados entre si e normalmente são feitos repetidamente de forma alternada e podem também influenciar a lista de termos, com a inclusão de novos termos conforme as classes vão sendo definidas.

Passo 4: Definir as classes e a hierarquia entre elas.

Para a criação das classes e a hierarquia de classes, Noy e McGuinness (2001) descrevem brevemente as abordagens sugeridas por Uschold (1996) – e já apresentadas anteriormente aqui – como as possíveis escolhas nessa etapa. Os autores também destacam que não há uma melhor abordagem entre as três (*top-down*, *bottom-up* ou *middle-out*), e que a escolha depende fortemente da visão pessoal do especialista sobre o domínio em questão.

Qualquer que seja a abordagem escolhida, deve-se partir dos termos listados no passo anterior e selecionar os que descrevem objetos que existam de forma independente em detrimento dos termos que descrevem esse objeto. Esses termos selecionados vão se tornar classes na ontologia e serão as ancoras na hierarquia. As classes são organizadas

então em uma taxonomia hierárquica de acordo com a resposta para a pergunta “sendo uma instância de uma classe, o objeto vai necessariamente ser uma instância de alguma outra classe”, ou seja, “Se uma classe A é uma superclasse da classe B, então todas instâncias de B são também uma instância de A”.

Passo 5: Definir as propriedades das classes, os *slots*.

Após a definição das classes, ou pelo menos de algumas delas, é preciso descrever também a estrutura interna dos conceitos, pois a classe por si só não vai fornecer informação necessária para responder as perguntas de competência. Após a seleção dos termos listados no passo 3 para a definição das classes, o restante vai ser basicamente propriedades das classes.

Podem haver diversos tipos de propriedades que descrevem uma classe, como por exemplo: (i) intrínsecas, que são propriedades inerentes a classe; (ii) extrínsecas, que são propriedades externas ou que identificam a classe; (iii) partes, caso a classe seja composta de outras classes de forma estruturada; (iv) relacionamentos, onde se deve determinar como uma classe se relaciona com outra. As propriedades serão herdadas em todas as subclasses de uma hierarquia e a propriedade deve estar descrita sempre na classe mais genérica que pode ter essa propriedade.

Passo 6: Definir as facetas dos *slots*.

As facetas de uma propriedade são características que descrevem essa propriedade, como por exemplo tipo do valor, valores permitidos, cardinalidade e outras que podem ser definidos de acordo com a propriedade que está sendo atribuída.

Essas facetas permitirão que as propriedades sejam melhor definidas e possam ser usadas adequadamente por motores de inferência e até mesmo na definição de hierarquia e estruturação da taxonomia.

Passo 7: Criação de instâncias.

O último passo é então criar as instâncias individuais das classes na hierarquia. Para isso deve se escolher a classe, criar uma instância individualizada daquela classe e preencher as propriedades com os valores adequados.

Seguindo esses passos, Noy e McGuinness (2001) afirmam ser possível ter uma ontologia que funcione para o domínio desejado, mas reforça novamente que apesar de resultar em uma ontologia, não há um único modo correto de se criar uma ontologia, e que sendo um processo criativo duas ontologias sobre o mesmo domínio projetadas por pessoas diferentes não devem ser a mesma ontologia. E que além disso, as potenciais aplicações para a ontologia, a

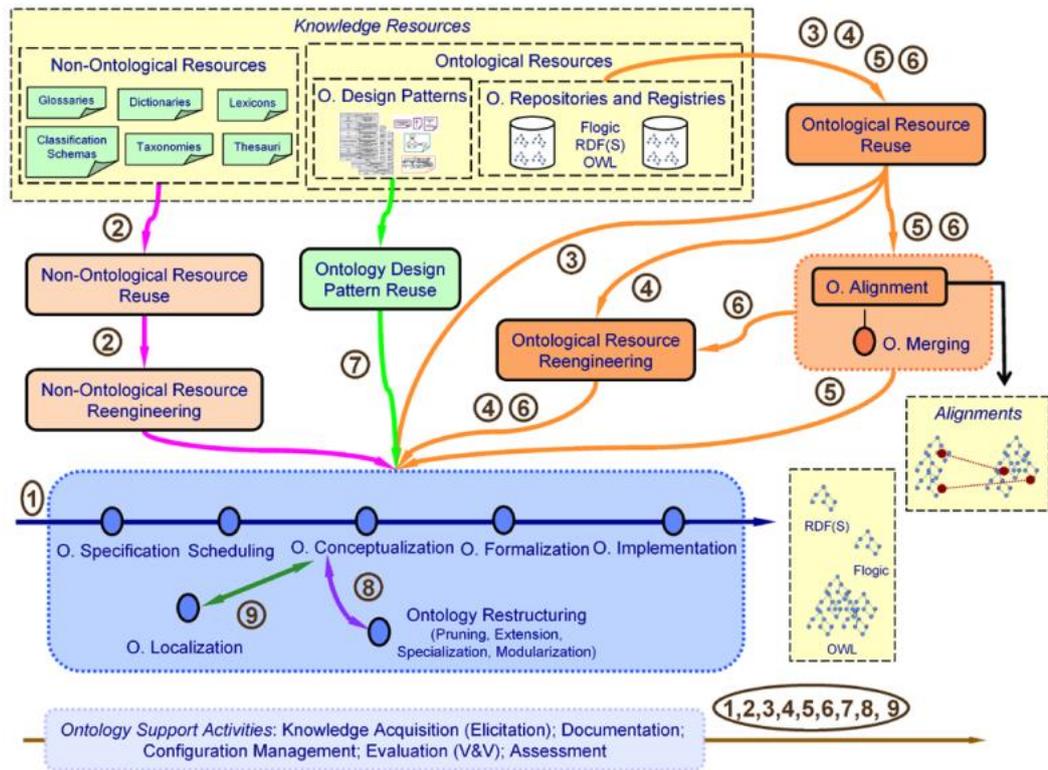
visão e compreensão do domínio do especialista vão seguramente afetar nas escolhas de projeto da ontologia.

Suárez-figueroa (2010) apresenta em 2010 sua tese *NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse* (ou em tradução nossa: Metodologia NeOn para construção de redes de ontologias: especificação, agendamento e reutilização), que apresenta como motivação a falta de abordagens que auxiliem desenvolvedores a construir grandes ontologias que sejam embarcadas em redes de ontologias³, e que possuam configurações complexas onde equipes distribuídas consigam construir ontologias colaborativamente reutilizando, permitindo reengenharia de fontes de conhecimento, mantendo o alinhamento e tendo em mente a evolução contínua dessas ontologias. Além de que as abordagens hoje existentes não são descritas como soluções centradas nos usuários, mas sim muito mais orientadas aos pesquisadores e desenvolvedores das ontologias.

A metodologia NeOn não se concentra em definir fluxos rígidos para construção de ontologia. Sua proposta se baseia em 9 cenários que comumente aparecem no processo de desenvolvimento de uma ontologia, sugerindo então qual caminho seguir em cada situação apresentada nas propostas. Além disso o *framework* da metodologia fornece elementos de apoio para a criação de ontologias: (a) glossário dos processos e atividades envolvidas no desenvolvimento das ontologias; (b) dois modelos de ciclo de vida de ontologias; (c) conjunto de diretrizes metodológicas para diferentes processos e atividades (SUÁREZ-FIGUEROA, 2010).

³ Rede de ontologias é o conjunto de ontologias únicas interconectadas entre si através de diversas meta-relações estabelecidas na definição de cada ontologia (SUÁREZ-FIGUEROA, 2010).

Figura 4 - Cenários para a construção de ontologias e redes de ontologias



Fonte: (SUÁREZ-FIGUEROA, 2010)

Os 9 cenários propostos pela metodologia NeOn são:

- Cenário 1: **Da especificação para implementação.** A rede de ontologia é desenvolvida do princípio, do zero, sem a reutilização de recursos de conhecimento previamente existentes.
- Cenário 2: **Reuso e reengenharia de recursos não-ontológicos.** Casos onde o desenvolvedor da ontologia precisa analisar recursos não-ontológicos e decidir, de acordo com os requisitos da ontologia, se a ontologia pode ser preenchida com os recursos analisados. Esse cenário também cobre a tarefa de reengenharia dos recursos selecionados para dentro da ontologia.
- Cenário 3: **Reuso de recursos ontológicos.** Recursos como ontologias completas, módulos ontológicos e/ou afirmações de ontologias serão utilizados para criar a nova ontologia.
- Cenário 4: **Reuso e reengenharia de recursos ontológicos.** Para esse cenário, além da reutilização, como no cenário 3, há também a reengenharia dos recursos ontológicos utilizados.
- Cenário 5: **Reuso e fusão de recursos ontológicos.** Cenário que ocorre em casos onde diversos recursos ontológicos no mesmo domínio são selecionados para reutilização e

os desenvolvedores da ontologia querem criar um novo recurso ontológico a partir de 2 ou mais recursos.

Cenário 6: **Reuso, fusão e reengenharia de recursos ontológicos**. Cenário similar ao cenário 5, mas ao invés de reutilizar o novo recurso gerado, o desenvolvedor deseja fazer a reengenharia dos recursos resultantes.

Cenário 7: **Reuso de *Ontology Design Patterns*⁴ (ODPs)**. Utilização de padrões de projeto de ontologias para a construção de ontologias.

Cenário 8: **Reestruturação de recursos ontológicos**. Desenvolvedores da ontologia fazem a reestruturação dos recursos ontológicos para serem integrados a rede de ontologias que será construída.

Cenário 9: **Regionalização de recursos ontológicos**. Adaptação dos recursos ontológicos para um certo idioma e cultura, o que produz uma ontologia multilingua.

As questões de aquisição do conhecimento, documentação, gerência de configuração e avaliação devem ser tratados ao longo de todo o processo de criação da ontologia, qualquer que seja o cenário utilizado (SUÁREZ-FIGUEROA, 2010).

Outro fator importante na construção de ontologias é a utilização de ferramentas de suporte no processo, que permitem ao desenvolvedor da ontologia analisar diferentes recursos, visualizar a ontologia e até automatizar partes do processo de construção. Na seção seguinte serão apresentadas as ferramentas mais relevantes encontradas para essa pesquisa.

4.2 Ferramentas para construção de ontologias

No livro *Ontology Engineering* Gómez-pérez, Fernández-lópez e Corcho (2004) falam sobre as diferentes ferramentas utilizadas nas atividades de engenharia de ontologias, e as classificam da seguinte forma:

- **Ferramentas de Desenvolvimento de Ontologias:** Grupo das ferramentas ou suíte de soluções que são utilizadas para construir ontologias desde o princípio ou com reutilizações. Além das funções mais comuns como edição e navegação, essas ferramentas também oferecem suporte a documentação, exportação / importação para diferentes formatos ou linguagens, edição gráfica, gerenciamento de bibliotecas e etc.

⁴ *Ontology Design Pattern* são padrões de projeto adotados para a construção de ontologias e que são maneiras de ajudar o desenvolvedor no momento da modelagem e também uma forma de se apresentar boas práticas para construção de ontologias baseadas em experiências anteriores (SUÁREZ-FIGUEROA, 2010).

- **Ferramentas de Avaliação de Ontologias:** Utilizadas para avaliar o conteúdo das ontologias e suas tecnologias relacionadas.
- **Ferramentas de Fusão e Alinhamento de Ontologias:** Essas ferramentas são utilizadas para fundir ontologias de um mesmo domínio ou alinhar ontologias de domínio diferentes para serem utilizadas em tarefas de construção, manutenção ou avaliação de ontologias.
- **Ferramentas de Anotação em Ontologias:** Permite ao usuário inserir instâncias de conceitos e de relações em ontologias. Muitas dessas ferramentas são recentes e tem relação com a manutenção de dados para a Web Semântica.
- **Ferramentas de Pesquisa em Ontologias e Motores de Inferência:** Facilitam a pesquisa através de *queries* em ontologias, similares as existentes para bancos de dados, e também permite que a utilização dessas pesquisa e suas relações gerem inferências lógicas a partir dos dados da ontologia.
- **Ferramentas de Aprendizado de Ontologia⁵:** São capazes de derivar ontologias de forma semiautomática a partir de textos em linguagem natural, assim como de fontes de dados estruturadas como banco de dado, através de técnicas de processamento de linguagem natural e aprendizado de máquina.

Apesar de ser uma classificação por tipo de tarefa realizada pela ferramenta, existem casos onde uma ferramenta ou suíte pode pertencer a mais de um grupo ao mesmo tempo (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004). Nessa seção iremos abordar o primeiro grupo, as ferramentas de desenvolvimento de ontologias.

Na página *Ontology Editors* do *World Wide Web Consortium* (2010), ou W3C, estão listados atualmente 12 editores de ontologia: Protégé, NeOn Toolkit, SWOOP, Neologism, TopBraid Composer, Vitro, Knoodl, Anzo for Excel, OWLGrEd, Fluent Editor, Semantic Turkey e VocBench.

Na pesquisa realizada por Abburu e Babu (2013) foram apresentadas 10 ferramentas de construção de ontologias: Ontolingua Server, OntoSaurus, OilEd, WebOnto, Protégé, SWOOP, TopBraid Composer, WebODE, OntoEdit e Neon Toolkit. Como conclusão do artigo, Abburu e Babu (2013) apontaram, dentre outras comparações que as ferramentas protégé e NeOn toolkit

⁵ As ferramentas de aprendizado de ontologia serão mais detalhadas no capítulo 6 que trata de aprendizado de ontologias a partir de texto.

oferecem mais funcionalidades para os desenvolvedores de ontologia (tabela 1 apresenta um comparativo das ferramentas pesquisadas).

O Protégé é um software para criação, edição e manutenção de ontologias, e teve sua primeira versão publicada em 11 de novembro de 1999 em uma parceria entre as universidades de Stanford e Manchester. Atualmente é desenvolvido e mantido por uma equipe específica no *Stanford Center for Biomedical Informatics Research*.

Além de ser um editor de ontologias, oferece ao desenvolvedor de ontologias funcionalidades de suporte como motores de busca e inferência, métodos de solução de problemas, visualização e exportação de ontologias. O software é de uso gratuito e está licenciado sob a licença BSD 2-clause⁶, com versões para *desktop* independente de plataforma, além da versão Web que pode ser utilizada diretamente do site do Protégé como pode ser instalado para uso local.

O NeOn Toolkit também é uma ferramenta *open-source* compatível com diversas plataformas e que oferece suporte completo a todo o ciclo de vida para a engenharia de ontologias, incluindo funcionalidades como anotação, documentação, editor gráfico, visualização de ontologias, gerenciamento de conhecimento, modularização de ontologia, avaliação de ontologias e etc. Além de oferecer a possibilidade de outros desenvolvedores estenderem suas funcionalidades através de *plug-ins* (NEON PROJECT, 2011).

⁶ A licença BSD 2-clause (ou também conhecida como Free-BSD) é uma forma de licenciamento de *softwares open-souce*, considerada uma das mais permissivas por permitir que qualquer pessoa utilize, modifique e até distribua o mesmo *software* desde que mantenha informações sobre os direitos autorais do mesmo (OPEN SOURCE INITIATIVE, 1999).

Tabela 1 - Comparação de ferramentas de criação de ontologias

Tool Features	Ontolingua Server	OntoSaurus	OilEd	WebOnto	Protégé	SWOOP	TopBarid Composer	WebODE	OntoEdit	Neon Toolkit
Availability	Free	Free & Open	Free & Open	Free	Free	Free & Open	Commercial	Free	Free	Free & Open
Versioning	No	No	No	Y/N	Y/N	Yes	Y/N	No	Y/N	Yes
Collaborative	Yes	Yes	No	Yes	Yes (Collaborative Protégé)	Yes	Yes	Yes	No	Yes
Graphical Class/Property taxonomy	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Back up manage- ment	No	No	No	Yes	No	Y/N	Y/N	Yes	No	Yes
Support growth of large ontologies	Yes	Y/N	No	Y/N	Yes	Yes	Y/N	Yes	Y/N	Yes
Querying	No	No	No	Y/N	Yes	No	Yes	No	Y/N	Yes
User Interface	No	Y/N	Yes	Y/N	Yes	Yes	Yes	Yes	Y/N	Yes
Consistency check	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
OWL Editor	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Extensibility	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Ontology Libraries	Yes	No	Yes	Yes	Y/N	Y/N	Y/N	No	No	Yes
Architecture	Client/Server	Client/Server	Standalone	Client/Server	Standalone	Standalone	Client/Server	N-Tire	Standalone	Standalone
KR Paradigam of Knowledge model	Frames+, FOL	DL	DL	Frames+, FOL	Frames+, FOL+, Meta classes	DL	DL	Frames+, FOL	Frames+, FOL	DL
Import	Ontolingua, DAML+OIL, CLIPS	LOOM, IDL, KIF, C++	RDF(S), DAML+OIL	OMCL	RDF(S), OWL	RDF(S), OWL	RDBMS, OWL, RDF(S)	RDF(S), DAML+OIL, OWL	RDF(S), DAML+OIL	RDF(S), OWL
Export	Ontolingua, DAML+OIL, CLIPS	LOOM, IDL, KIF, C++	RDF(S), DAML+OIL, OWL	OMCL, Onto- lingua, RDF(s), OIL	RDF(S), OWL, CLIPS	RDF(S), OWL	OWL, RDF(S), XML	RDF(S), DAML+OIL, OWL, CLIPS	RDF(S), DAML+OIL, OWL	RDF(S), OWL
Storage	Files	Files	Files	Files	Files, DBMS(JDBC)	Files	Files	DBMS(JDBC)	Files	Files
Reasoner	JTP, Prolog, CML, Epikit	PowerLoom, Stella	FaCT	-	Pellet	Pellet	Pellet	Prolog	OntoBroker	Pellet2, Hermit, OntoBroker
Merging	Chimaera	None	None	None	Prompt, OWLDiff	Yes	Y/N	ODE Merge	Yes	Yes
Debug/Repair	No	No	Very Little	No	Very Little	Yes	No	No	No	Yes
Built-in Inference	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Y/N	Yes
Implemented in	Lisp	Lisp	Java	Lisp	Java	Java	Java	Java	Java	Java Eclipse

Fonte: (ABBURU; BABU, 2013)

4.3 Engenharia Ontológica

Em domínios complexos, como compras na Internet ou permitir um carro autônomo trafegar em cidades reais, requerem modelos flexíveis e mais genéricos, permitindo que conceitos mais genéricos (eventos, tempo, objetos físicos ou crenças) sejam representados em diferentes contextos e utilizações. Russel e Norvig (2010) afirmam que esse processo pode ser chamado de Engenharia Ontológica (*Ontological Engineering*).

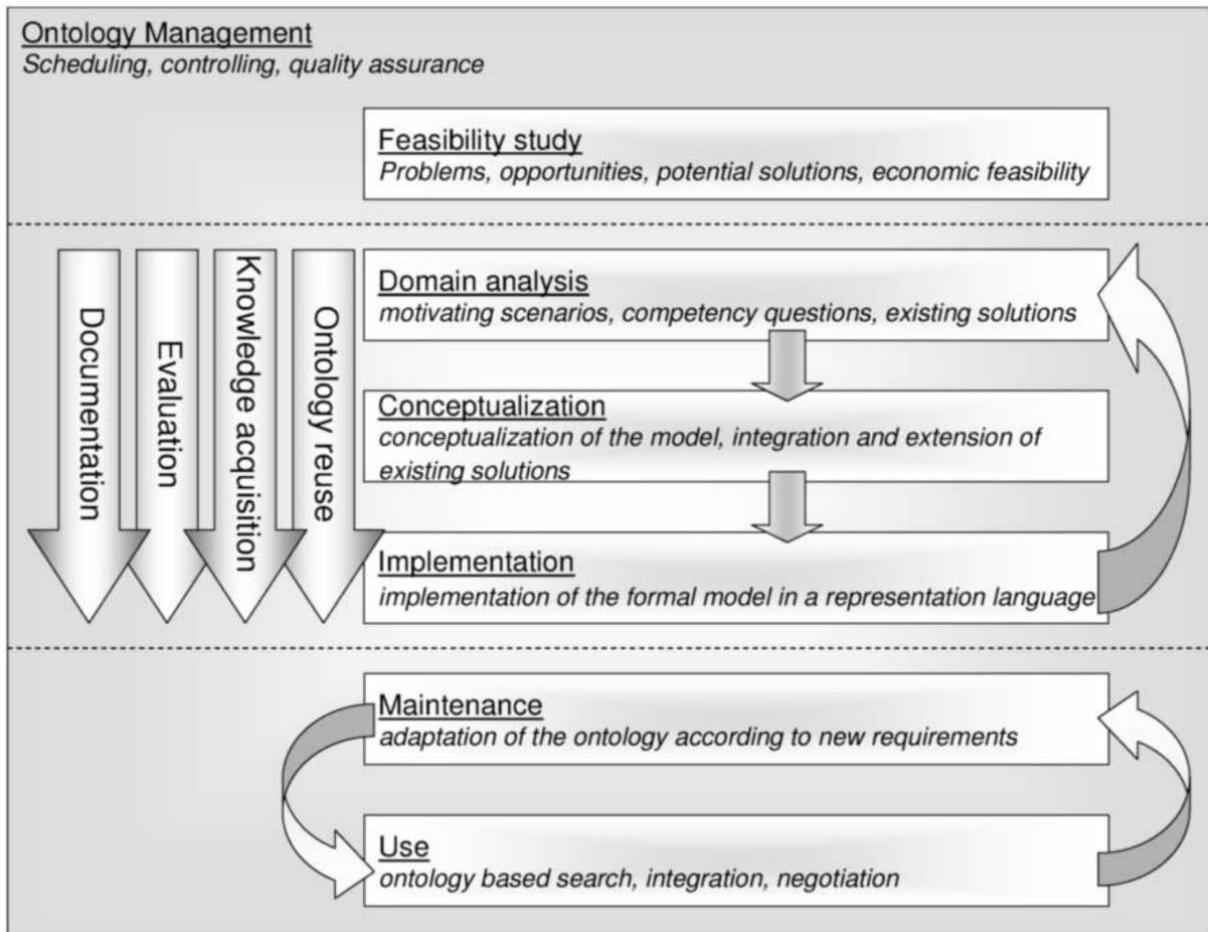
A Engenharia de Ontologias surge como um conjunto de atividades que se preocupam com o processo de desenvolvimento de ontologias como um todo, com o ciclo de vida de uma ontologia e com as metodologias, ferramentas e linguagens usadas na construção de ontologias.

Assim, a engenharia de ontologias se origina em um contexto de uma nova ciência que codifica características de coisas, e por outro lado, busca extrair a essência dessas coisas (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2003).

Gómez-Pérez, Fernández-López e Corcho (2003) dividem o processo de engenharia de ontologias (Figura 1) nas seguintes atividades:

- **Gerenciamento** (*management*): Definição organizacional de todo o processo de engenharia da ontologia.
- **Desenvolvimento** (*development-oriented*): Dividido em duas etapas, sendo a primeira a fase pré-desenvolvimento onde há o estudo de viabilidade se uma aplicação baseada em ontologias, ou o uso da ontologia é a melhor maneira de resolver o problema apresentado. E na fase do desenvolvimento em si, existem as atividades de análise do domínio, conceitualização e implementação. Após a implementação, tarefas de manutenção e utilização são executadas na ontologia.
- **Atividades de suporte** (*support activities*): Aquisição de conhecimento, avaliação, reutilização e documentação são realizadas em paralelo com as atividades de desenvolvimento, por isso são consideradas atividades de suporte

Figura 5 - Atividades da Engenharia de Ontologias



Fonte: (GÓMEZ-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2003)

A engenharia de ontologias vem evoluindo nos últimos anos, mudando do *status* de arte para ciência, principalmente com a evolução dos diversos métodos que foram apresentados anteriormente. Ainda assim a tarefa de construção de ontologias continua sendo algo que requer muitos recursos e possui um custo alto. Dessa forma, são necessárias técnicas que sejam capazes de reduzir os custos relativos às tarefas da engenharia e manutenção das ontologias (CIMIANO et al., 2009).

Devido à alta disponibilidade de dados em diversos formatos, muitos pesquisadores desenvolveram métodos de suporte a engenharia de ontologias baseados nas técnicas de mineração de dados, ou seja, derivando as relações semânticas que possam suportar o engenheiro de ontologia na modelagem do domínio. Essa abordagem orientada a dados como técnica de suporte a engenharia de ontologia ficou conhecida como aprendizado de ontologia (*Ontology Learning*), e tem o potencial de reduzir o custo da criação de uma ontologia, bem como o custo de manutenção (CIMIANO et al., 2009).

4.4 Aprendizado de Ontologia

A construção de ontologias, em geral, esteve baseada na ontologia que se desejava construir, até a apresentação da metodologia METHONTOLOGY. Apesar de ter sido proposta para a criação de ontologias do zero e não estar vinculada a nenhum outro projeto, a METHONTOLOGY (assim como as demais metodologias apresentadas) ainda dependiam da participação de um especialista em todas as suas etapas.

A necessidade de profissionais capacitados, como especialistas de domínio ou engenheiros de conhecimento, torna o processo de construção manual de ontologias caro e demorado. Essa dificuldade na aquisição de conhecimento requerido pelos sistemas baseados em conhecimento, é chamado de “gargalo da aquisição de conhecimento” (TAWFIK; AREF; SALEM, 2011, p. 169, tradução nossa).

Uma abordagem para tornar o processo de criação e manutenção de ontologias mais rápido, é a automação de parte dos métodos ou pelo menos de etapas dessas metodologias. Na literatura, como definido, por exemplo, por Maedche e Staab (2001), Cimiano et al. (2009) e Browarnik e Maimon (2015), o campo conhecido por pesquisar ferramentas e métodos automáticos ou semiautomáticos na criação e manutenção de ontologias é o Aprendizado de Ontologias, tradução do termo em inglês *Ontology Learning*.

O aprendizado de ontologias se concentra na descoberta de conhecimento de diferentes fontes e sua representação através de estruturas ontológicas. Tawfik, Aref e Salem (2011) definem dois aspectos fundamentais no aprendizado de ontologias. O primeiro é a disponibilidade de conhecimento prévio, ou seja, o processo de aprendizado pode ser do princípio (do zero) ou baseado em algum conhecimento disponível, como algo utilizado em uma versão anterior da ontologia (no caso de manutenção ou evolução de uma dada ontologia) ou outra estrutura semântica, como um vocabulário controlado.

Outro aspecto diz respeito ao tipo de entrada no processo de aprendizagem, que define então a abordagem a ser adotada. Cimiano et al. (2009) classificam essas abordagens da seguinte maneira:

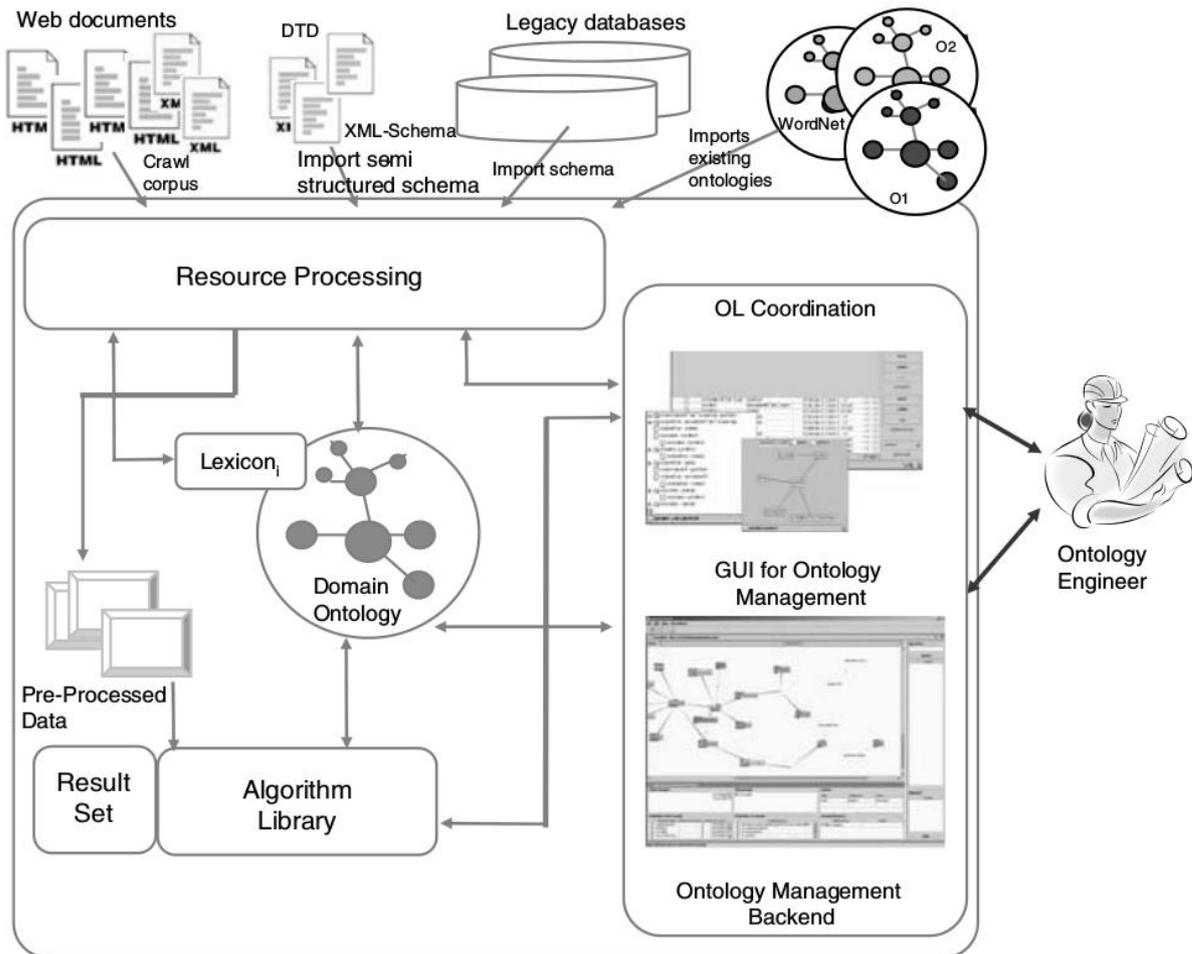
- **Dados estruturados:** bancos de dados, bases de conhecimento e demais estruturas relacionais.
- **Dados semiestruturados:** dicionários (como o *WordNet*) e demais fontes que tenham definições e permitam a obtenção de relações existentes entre os termos (ex. Sinonímia, hponímia, etc.)

- **Dados não estruturados:** textos em linguagem natural, como artigos científicos, revistas, jornais e a maioria das páginas *Web*.

Segundo Cimiano et al. (2009), o aprendizado de ontologia se constrói a partir de técnicas já bem estabelecidas e multidisciplinar, como o processamento de linguagem natural, aprendizado de máquina, aquisição de conhecimento e engenharia de ontologia.

Com o objetivo de melhor representar o aprendizado de ontologia, Cimiano et al. (2009) propõe uma arquitetura genérica e seus principais componentes para o processo de aprendizado de ontologias. Segundo os autores, o processo de aprendizado de ontologia não difere muito do processo clássico de mineração de dados. Os componentes da arquitetura do aprendizado de ontologia são: gerenciamento de ontologia, coordenação, processamento de recursos e a biblioteca de algoritmos.

Figura 6 - Arquitetura conceitual do aprendizado de ontologia



Fonte: Cimiano et al. 2009

O gerenciamento de ontologia se trata do componente que permite a manipulação da ontologia pelo engenheiro de ontologias, tipicamente oferecendo ferramentas de importação, busca, edição, versionamento e avaliação. Além disso, o principal propósito do componente de gerenciamento (no contexto do aprendizado de ontologia) é funcionar como uma interface entre a ontologia e o algoritmo de aprendizado (CIMIANO et al., 2009).

O componente de coordenação é utilizado pelo engenheiro de ontologia para interagir com os componentes de processamento de recurso e com a biblioteca de algoritmos. Para esse componente, é esperado uma interface de fácil compreensão, que dê suporte ao usuário na escolha de dados relevantes para entrada que serão depois utilizados no processo de descoberta. A principal tarefa desse componente é fazer um arranjo sequencial e aplicar os algoritmos escolhidos pelo usuário, passando os resultados uns para os outros (CIMIANO et al., 2009).

O componente de processamento de recursos se trata de um conjunto de técnicas para descoberta, importação, análise e transformação de dados de entrada relevantes para a criação da ontologia. Tem como um de seus principais subcomponentes o sistema de processamento de linguagem natural. A principal tarefa desse componente é a geração de um conjunto pré-processado de dados e que será a entrada para a biblioteca de algoritmos (CIMIANO et al., 2009).

A biblioteca de algoritmos funciona como a espinha dorsal para os algoritmos do *framework*, onde diversos algoritmos de extração e manutenção são fornecidos para todo o processo. Ou seja, a biblioteca de algoritmos contém os algoritmos que de fato serão usados no processo de aprendizado da ontologia e em geral é composta por algoritmos de aprendizado de máquina e outras versões customizadas para atender ao propósito da ontologia (CIMIANO et al., 2009).

Além de algoritmos de aprendizado de máquina, a biblioteca de algoritmos também deve conter uma variedade de implementações para medir a distância e similaridade de termos, para assim suportar o agrupamento semântico e também medições tradicionais de arranjos entre palavras que sejam conhecidos pela linguística computacional. Para que seja possível a combinação das extrações de diferentes algoritmos de aprendizagem, é necessário a padronização do resultado de saída desses algoritmos (CIMIANO et al., 2009).

Cimiano et al. (2009) apresenta uma tabela com os típicos algoritmos de aprendizado de máquina que podem ser utilizados pela biblioteca de algoritmos no aprendizado de ontologias:

Tabela 2 - Algoritmos de aprendizado de máquina usados na biblioteca de algoritmos

Algoritmo	Uso genérico	Uso no aprendizado de ontologia
Descoberta por regra de associação	Descoberta de transações “interessantes” em grupos de itens (ex.: dados de clientes)	Descoberta de associações interessantes entre palavras
Agrupamento (hierárquico)	Descoberta de grupos em dados (não supervisionado)	Agrupamento de palavras
Classificação	Previsão (supervisionada)	Classificação de novos conceitos em uma hierarquia já existente
Lógica indutiva de programação	Indução de regra a partir de dados (supervisionado)	Descoberta de novos conceitos a partir de dados adicionais
Agrupamento conceitual	Descoberta de conceitos	Aprendizado de conceitos e hierarquias de conceitos

Fonte: (CIMIANO et al., 2009)

A Tabela 1 ilustra a variedade de algoritmos que podem ser empregados nos métodos baseados em aprendizado de máquina, onde cada um tem seu objetivo específico e podem ser combinados de várias maneiras em diferentes etapas do aprendizado, como extração conceitos, agrupamento de palavras e aprendizado de hierarquia e conceito.

5

Aprendizado de ontologia a partir de textos

O aprendizado de ontologia a partir de textos, segundo Wong, Liu e Bennamoun (2012), é o processo de identificar termos, conceitos, relações e, opcionalmente, axiomas de informações textuais, e os utilizar para construir e manter uma ontologia. É um processo interdisciplinar que lança mão de técnicas de campos já estabelecidos como a recuperação de informação, a mineração de dados e, principalmente, o processamento de linguagem natural.

Levando em conta a alta disponibilidade de documentos digitais na *Web* e nas organizações, esses podem ser considerados como um recurso útil na aquisição de conhecimento no aprendizado de ontologias (ZAHRA et al., 2014).

Além dos documentos encontrados na *Web*, novos conhecimentos são criados e publicados em meios como livros, artigos, teses, dissertações, *proceedings* e similares. Esses documentos podem ser fortes bases para a construção de conhecimento sem intervenção humana, uma vez que são fontes confiáveis e formalizadas de transmissão do conhecimento (BROWARNIK; MAIMON, 2015).

5.1 Abordagens de aprendizado de ontologia

Por se tratarem de textos escritos em linguagem natural, todas as abordagens de aprendizado a partir de textos têm uma forte relação com o processamento de linguagem natural, uma vez que é necessário o processamento do texto para o aprendizado da ontologia. Os métodos encontrados no levantamento bibliográfico foram divididos em 3 categorias,

segundo classificação feita por Gómez-Pérez e Manzano-Macho (2004): (i) Abordagens de aprendizado baseadas em técnicas linguísticas; (ii) Abordagens de aprendizado baseadas em técnicas estatísticas; (iii) Abordagens de aprendizado baseadas em aprendizado de máquina.

Os métodos listados a seguir foram selecionados de acordo com sua fonte de entrada (textos não estruturados) e a partir dos levantamentos feitos por Gómez-Pérez e Manzano-Macho (2004), Tawfik, Aref e Salem (2011) e Wong, Liu e Bennamoun (2012).

5.1.1 Abordagens de aprendizado baseadas em técnicas linguísticas

Métodos que utilizam como base em sua operação técnicas linguísticas, como por exemplo extração baseada em padrões linguísticos, medições de relatividade semântica entre outras. Abordagens linguísticas são em geral aplicadas juntamente com alguma abordagem estatística para determinar a relevância dos elementos encontrados em recursos textuais do domínio desejado (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

Nos levantamentos realizados por Gómez-Pérez e Manzano-Macho (2004) e Tawfik, Aref e Salem (2011) foram apresentados diversos métodos com abordagens utilizando padrões linguísticos e relações semânticas da língua. Essa característica faz com que muitos desses métodos precisem de uma maior interação humana ou dependem de estruturas pré-existentes como a *WordNet*.

Como **Alfonseca e Manandhar**, que propuseram uma abordagem para a extensão de ontologias existentes, como a *WordNet*. Essa abordagem se baseia em adquirir propriedades contextuais de palavras que ocorrem simultaneamente em um determinado conjunto de conceitos, seguido de um algoritmo de classificação *top-down*, que servirá tanto para o agrupamento quanto para a classificação de conceitos. Essa técnica é baseada na hipótese, dos mesmos autores, de que o significado de uma palavra está altamente relacionado com os contextos em que ela aparece (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

Seguindo a mesma linha descrita acima, o método de **Moldovan e Girju** permite a descoberta de conceitos e relações específicas de um domínio para a expansão de uma ontologia pré-existente, mas com novo conhecimento adquirido de uma outra fonte processada. Essa fonte deve ser um corpus não específico do domínio ampliado por outros recursos léxicos, como dicionários e tesouros, sejam específicos ou genéricos (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

O método de **Roux e colegas** visa enriquecer uma determinada ontologia a partir de conceitos extraídos de um corpus utilizando técnicas de processamento de linguagem natural. O método se baseia em grafos contextuais e utiliza as relações sintáticas entre os conceitos para que, através de padrões de verbos, seja possível construir uma representação semântica (TAWFIK; AREF; SALEM, 2011). Assim, de acordo com esses padrões de verbos e suas correspondências na estrutura semântica, novos conceitos podem ser adicionados à ontologia (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

Para a manutenção e expansão de uma ontologia de domínio, Hahn et al. (2001, apud GÓMEZ-PÉREZ; MANZANO-MACHO, 2004) apresentam um método baseado na compreensão de textos em linguagem natural, onde uma dada taxonomia é incrementada e ampliada de acordo com conceitos extraídos de textos do mundo real (não necessariamente processados). O processo de aquisição dos conceitos é baseado em linguística e na qualidade conceitual de diversas evidências, como a geração e refinamento de hipóteses. Dessa maneira, o aprendizado é atingido através do refinamento de diversas hipóteses sobre o pertencimento de um conceito a uma determinada instância. Novos conceitos são adquiridos tendo como base duas fontes de evidências: conhecimento prévio a partir de textos do domínio e padrões linguísticos em cada item léxico desconhecido que apareça (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

O método de **Hearst** propõe a aquisição automática de relações hiponímias de um corpus para a construção de um tesouro de domínio geral, usando a *WordNet* para validação e melhoria de performance. O objetivo desse método é descobrir instâncias de padrões de relação, podendo assim construir novas relações léxicas. Apesar de focar na construção de tesouros, esses padrões podem ser úteis na expansão léxica e relacionamento semântico da informação (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

Aussenac-Gilles et al. (2000a,b) apresentam um método que permite a criação de modelos de domínio através da análise de corpus com ferramentas de processamento de linguagem natural e técnicas linguísticas, ajudando o especialista na construção de ontologias. O método combina ferramentas de aquisição de conhecimento baseado em linguística com técnicas de modelagem que permitem a criação de relação entre modelos linguísticos e outras fontes de texto. Essas fontes são primariamente textos, mas podem ser também ontologias pré-existentes ou recursos terminológicos. O processo pode ser descrito nas seguintes etapas:

1. **Constituição do corpus:** Os textos devem ser selecionados a partir de documentos específicos do domínio, de acordo com os requisitos da ontologia. Esse corpus deve

cobrir todo o domínio alvo e a seleção dos textos deve ser feita por um especialista do domínio.

2. **Estudo linguístico:** Etapa onde é feita a escolha das ferramentas e técnicas linguísticas adequadas à serem aplicadas nos textos. Como resultado dessa etapa serão extraídos termos do domínio, relações léxicas e grupos de sinônimos. A extração de termos é baseada na análise de frequência de ocorrência dos termos, e a extração de relações é feita através de padrões linguísticos.
3. **Normalização:** Essa etapa inclui uma fase linguística, onde um especialista deve selecionar os termos e relações léxicas que devem ser modeladas. Também nessa fase, o especialista deverá adicionar alguns conceitos em linguagem natural para os termos selecionados, de acordo com o contexto do domínio. Em seguida, na fase conceitual, os conceitos e relações semânticas são definidas e normalizadas de acordo com os rótulos dos conceitos e das relações.
4. **Formalização:** Fase final do método, onde a ontologia deve ser validada pelo especialista e implementada na linguagem escolhida.

5.1.2 Abordagens de aprendizado baseadas em técnicas estatísticas

Grupo de métodos que são baseados principalmente no cálculo de diversas métricas estatísticas, que ajudam o especialista detectar novos conceitos ou relacionamentos entre eles (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

Dentre as métricas mais comuns estão a análise de frequência de repetição de palavras (ou padrão de palavras) e o índice TF-IDF, que normalmente é aplicado no cálculo de relevância para a descoberta de elementos na construção de ontologias de domínio (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

Lopes (2012) apresenta em seu trabalho uma nova proposta de índice que também pode ser utilizado em métodos estatísticos, chamado Frequência de Termo e Disjunção de Corpora (tf-dcf). Esse índice realiza uma indicação primária de relevância do termo de acordo com sua ocorrência no corpus do domínio e posteriormente é adotado um mecanismo de recompensa/penalização conforme a ocorrência do termo em corpora contrastante. Segundo a autora:

A base do índice *tf-dcf* é considerar a frequência absoluta de termo como indicação primária da relevância de um termo. Em seguida, escolhe-se penalizar termos que aparecem nos *corpora* contrastantes dividindo a frequência absoluta do termo no corpus de domínio pela composição geométrica da sua frequência absoluta em cada um dos corpora contrastantes.

O método de Agirre et al. (2000, apud GÓMEZ-PÉREZ; MANZANO-MACHO, 2004) busca enriquecer conceitos em grandes ontologias existentes, controlando a proliferação de significados dentro da ontologia, usando textos da *Web*. O objetivo principal desse método é diminuir duas deficiências de grandes ontologias, como a *WordNet*: a falta de ligação entre os conceitos e a proliferação de diferentes significados para o mesmo conceito. O método se baseia em 4 etapas (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004):

1. **Recuperação de documentos relevantes para cada conceito na ontologia da *Web*:** São construídas consultas para cada significado dos conceitos contidos na ontologia. Documentos que sejam relacionados com mais de um significado são descartados e os documentos relacionados ao mesmo significado do conceito são agrupados para formar coleções para cada significado.
2. **Construção de assinatura de tópicos:** Os documentos agrupados são então processados para a extração de palavras e suas frequências. Utilizando abordagem estatística, as palavras mais próximas ao conceito são selecionadas. Então, os dados de uma coleção são comparados com os de outra coleção e as palavras com frequências distintas em grupos diferentes são agrupadas em uma lista de acordo com a frequência. Essa lista é chamada de assinatura do tópico.
3. **Agrupamento de sentido das palavras:** Para cada palavra, os conceitos relacionados ao significado da palavra são agrupados hierarquicamente através da comparação de diferentes assinaturas dos tópicos, para a descoberta de significados compartilhados. Sendo possível então identificar assinaturas que se sobrepõe.
4. **Avaliação:** Essa tarefa é realizada pelo usuário, comparando os resultados utilizando um algoritmo de desambiguação de significados.

Faatz e Steimmetz (2002, apud GÓMEZ-PÉREZ; MANZANO-MACHO, 2004) apresentam um método que visa melhorar uma ontologia existente extraindo significados da *Web* e comparando estatisticamente o uso da palavra em um *corpus* com a estrutura da ontologia em questão. Para isso, cada conceito na ontologia deve ser associado a uma ou mais sentenças em linguagem natural, para que assim o algoritmo possa calcular a similaridade entre as

palavras e enriquecer a definição do conceito, e criar grupos de palavras relacionadas a um novo conceito. Os novos conceitos devem ser então avaliados por um especialista do domínio que irá decidir se serão adicionados ou não à ontologia.

O método de Xu et al. (2002, apud GÓMEZ-PÉREZ; MANZANO-MACHO, 2004) foi criado para a aquisição de termos relevantes a um domínio e suas relações usando técnicas híbridas não supervisionadas de mineração de texto para o aprendizado de padrões de relações léxicas-sintáticas. Para isso são utilizadas duas técnicas de mineração de texto, onde a primeira utiliza uma ontologia pré-existente como conhecimento inicial para aprender os padrões de relação léxico-sintática, e o segundo é baseado em diferentes métodos de aquisição por co-localização para lidar com palavra que ocorrem em ordens livre.

O processo, de modo geral, inicia com a mineração de termos relevantes no domínio em diversos documentos categorizados para o domínio desejado. Em seguida, é realizado o aprendizado dos padrões léxico-sintáticos entre os termos extraídos do *corpus*, usando as relações contidas na *WordNet*. No final desse processo, dois tipos de padrões podem ser identificados: padrões específicos do domínio e padrões independentes de domínio. Com as relações agrupadas e termos extraídos, grupos de termos são criados e, por último, com uma atividade aprendizado de co-localização de termos, os termos são colocados nos lugares corretos da taxonomia usando os padrões já mencionados e com informações estatísticas calculadas para cada termo.

5.1.3 Abordagens de aprendizado baseadas em aprendizado de máquina

Grupo que inclui os métodos que utilizam diferentes algoritmos de aprendizado de máquina para auxiliar o especialista na tarefa de detectar novos conceitos ou relações entre eles. Em geral esses métodos são utilizados em conjunto com outros métodos, principalmente linguísticos. Segundo Gómez-pérez e Manzano-macho (2004) as pesquisas com métodos baseados em aprendizado de máquina levaram a uma série de técnicas automatizadas para a aquisição e revisão de conhecimento.

Pesquisadores em aquisição de conhecimento vem buscando por abordagens integradas, que se beneficiem da sinergia entre métodos tradicionais de aquisição de conhecimento e as técnicas de aprendizado de máquina, já que essas técnicas oferecem uma série de ferramentas e sistemas que podem ajudar no desenvolvimento de novas técnicas e princípios para a automação da aquisição de conhecimento (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

A abordagem de Hwang (1999, apud GÓMEZ-PÉREZ; MANZANO-MACHO, 2004) visa representar e buscar informação de grandes bases de dados textuais. Esse método se baseia no uso de ontologias dinâmicas que capturam informação semântica de documentos. A ontologia é organizada em uma taxonomia simples no qual os conceitos são identificados nos documentos para facilitar o processo de recuperação da informação. O processo de aprendizagem é baseado em processamento de linguagem natural em conjunto a técnicas de aprendizado de máquina, e pode ser representado pelos seguintes passos:

1. Um especialista fornece ao sistema um conjunto de palavras semente, que representam os conceitos em alto nível. E então, são selecionados automaticamente documentos na *Web* que possuam essas sementes.
2. O sistema processa os textos selecionados e extrai somente as frases que contém as palavras semente informadas e gera uma correspondência entre os conceitos e os coloca no local adequado na ontologia e alerta o especialista sobre as mudanças. As iterações continuam por um número determinado de vezes, e enquanto isso o método indexa os documentos para futura recuperação de acordo com os conceitos identificados em cada texto.
3. Diversos tipos de relação linguística são extraídos, como “é-uma”, “parte-de”, “fabricado-por”, etc.

Em cada iteração, o especialista é consultado para atestar se os conceitos estão corretos e se necessário, o especialista pode fazer os ajustes que desejar e reconstruir a ontologia.

O método Khan e Luo (2002, apud GÓMEZ-PÉREZ; MANZANO-MACHO, 2004) foi construído para a criação de ontologias de domínio a partir de documentos em texto usando técnicas de agrupamento e a *WordNet*. A construção da ontologia é feita através de uma abordagem *bottom-up*, iniciando pela construção de uma hierarquia utilizando técnicas de agrupamento, associando documentos com conteúdos similares aos mesmos conceitos na ontologia. Em seguida, em uma abordagem *bottom-up*, um conceito é associado a cada grupo de documentos relacionados ao mesmo tópico, utilizando o algoritmo de rastreamento de tópico.

Kietz et al. (2000, apud MAEDCHE; STAAB; 2001) apresentam um método genérico para a descoberta de ontologias de domínio a partir de fontes heterogêneas utilizando técnicas de análise de linguagem natural. Em sua abordagem foi adotada um modelo de balanço cooperativo, onde o trabalho de construir a ontologia é distribuído entre diversos algoritmos de aprendizado e o usuário. O método se baseia na ideia de que a maioria dos conceitos e estruturas

conceituais de um domínio a serem incluídos na ontologia e a terminologia de um dado domínio são descritos em documentos. O processo é cíclico no sentido que uma ontologia resultante pode ser refinada aplicando o método repetitivamente. Esse método é parcialmente suportado pela ferramenta Text-To-Onto.

O método de Missikoff et al. (2002, apud. GÓMEZ-PÉREZ; MANZANO-MACHO, 2004), que é suportado parcialmente pela ferramenta OntoLearn, foi desenvolvido para a construção e enriquecimento de ontologias usando técnicas de processamento de linguagem natural e aprendizado de máquina, partindo da *WordNet* como fonte de conhecimento prévio para a construção de um núcleo para uma ontologia de domínio, com a retirada de todos os conceitos não específicos do domínio. Em seguida, são utilizados métodos estatísticos para determinar a relevância de um termo para o domínio e de interpretação semântica baseada em algoritmos de aprendizado de máquina para identificar o sentido correto dos termos e das relações semânticas entre eles. Os autores propõem 3 passos para atingir esses objetivos:

1. **Extração de termos:** Termos e combinações de termos são extraídos de um *corpus* pré-processado utilizando técnicas de processamento de linguagem natural.
2. **Interpretação semântica:** Nesse passo, o sentido correto do conceito é definido para cada componente de um termo complexo, em um processo de desambiguação semântica, então identificar as relações semânticas entre os conceitos para a construção de conceitos mais complexos. Ao final desse processo, uma coleção de conceitos será gerada, demonstrando a taxonomia e as relações entre os conceitos complexos do domínio. Para suportar esse processo, é preciso o uso de recursos semânticos e linguísticos, que foi testado com a *WordNet*, para assistir a interpretação semântica dos termos.
3. **Criação da ontologia de domínio:** Esse passo tem como objetivo integrar a taxonomia obtida no passo anterior com o núcleo da ontologia de domínio. Caso não haja uma ontologia prévia para a criação desse núcleo, os autores sugerem utilizar a *WordNet* como base.

5.2 Ferramentas para aprendizado de ontologias

Ao longo dos anos, houve o aumento da necessidade da construção de novas ontologias para os mais diversos fins, assim uma série de ferramentas foram sendo criadas para fornecer suporte ao processo de aquisição do conhecimento e aprendizado de ontologias.

Por se tratar de uma ferramenta computacional, a apresentação e estudo dessas ferramentas se concentra na área da CC, e principalmente no campo da IA com o uso de técnicas de aprendizado de máquina, como será demonstrado a seguir.

No decorrer do levantamento para essa pesquisa, foram encontradas diversas ferramentas relacionadas ao processo de construção de ontologias. Algumas com seu uso mais específico, como por exemplo a extração de termos e conceitos, e outras com propósitos mais genéricos, tratando de todo o processo de aprendizado.

A ferramenta ASIUM, um acrônimo para “*Acquisition of Semantic knowledge Using Machine learning methods*”, tem como principal objetivo ajudar especialistas na aquisição de conhecimento semântico e relações taxonômicas entre termos extraídos de textos técnicos usando análise sintática (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004). Se trata de um sistema semiautomático de aprendizado de ontologia, que faz parte da infraestrutura do sistema de extração de informação chamado INTEX, do *Laboratoire d'Automatique Documentaire et Linguistique de l'Universite de Paris 7* (WONG; LIU; BENNAMOUN, 2012).

O método de aprendizado do sistema é baseado em agrupamento conceitual e hierárquico, onde grupos básicos são formados por palavras que ocorrem com o mesmo verbo após a mesma preposição. O sistema usa o agrupamento como métrica para calcular a similaridade entre os grupos, e então o especialista decide se um novo conceito deve ser criado. Esses grupos são sucessivamente agregados pelo método de agrupamento conceitual para formar novos conceitos na ontologia, então o especialista define um limite mínimo para converter grupos em conceitos e no final, o aprendizado é validado pelo especialista (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

O CRCTOL (*Concept-Relation-Concept Tuple-based Ontology Learning*) é um sistema construído em 2005 na *National Technological University of Singapore* para a construção de ontologias a partir de documentos específicos de domínio, utilizando técnicas estatísticas para as tarefas de aprendizado de ontologia (WONG; LIU; BENNAMOUN, 2012).

O sistema usa importadores de texto para converter documentos de diferentes formatos (PDF, XML, HTML, etc.), e transformá-los em texto puro. Em seguida, é iniciado o passo de extração de termos e a formação de conceitos, onde os termos são avaliados de acordo com uma métrica chamada DRM (*Domain Relevance Measure*), que determina a relevância do termo pro dado domínio. A partir da relevância do termo e de relações semânticas, é construído a hierarquia dos conceitos e a descoberta das relações não-taxonômicas (WONG; LIU; BENNAMOUN, 2012).

Mo’K Workbench é uma ferramenta configurável que suporta a construção semiautomática de ontologias a partir de um *corpus* usando diferentes técnicas de agrupamento, auxiliando o especialista no processo exploratório de escolha do método de aprendizagem mais adequado. Dessa forma, a ferramenta suporta a elaboração, comparação, caracterização e avaliação de diferentes modelos conceituais de agrupamentos (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

O processo de aprendizado suportado pelo Mo’K tem como entrada um *corpus* e nenhum outro conhecimento adicional é utilizado para rotular a entrada, guiar o aprendizado ou validar o aprendizado. Através de técnicas de processamento de linguagem natural, a ferramenta extrai uma lista de triplas (compostas por um verbo, um termo e a função sintática do termo na sentença) do *corpus* e calcula o número de ocorrência dessas triplas no *corpus*. Triplas com baixa ocorrência ou com muitas ocorrências são removidas da lista e por fim, o Mo’K calcula a distância semântica entre as triplas da lista para formar agrupamentos conceituais (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

O sistema OntoGain, criado pela *Technical University of Crete*, foi projetado para a aquisição não supervisionada de ontologias a partir de textos não estruturados, utilizando técnicas linguísticas e estatísticas nas tarefas de aprendizado (WONG; LIU; BENNAMOUN, 2012).

O processamento dos textos é feito com a ferramenta *OpenNLP*, que cuida das tarefas de tokenização, lematização, rotulagem de discurso e processamento parcial dos textos. Em seguida, termos compostos ou termos com múltiplas palavras aninhadas são extraídos utilizando a métrica C/NC-value, que juntamente com análise de frequência de termos irá finalizar a extração de termos e a formação de conceitos (WONG; LIU; BENNAMOUN, 2012).

O OntoLearn se propõe a extrair termos relevantes em um domínio a partir de um *corpus*, e relacionar esses termos a conceitos em uma ontologia de propósito genérico, e detectar relacionamento entre os conceitos. Para executar essas tarefas, são utilizadas técnicas de análise de linguagem natural e aprendizado de máquina (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

A extração de termos é feita de *corpus* especializados e então, o sistema filtra os termos usando técnicas de processamento de linguagem natural e métodos estatísticos para uma análise comparativa entre diferentes domínios e em *corpora* contrastante. Através dessa análise é possível identificar a terminologia que é utilizada no domínio alvo mas que não é utilizada em outros domínios. Posteriormente são realizadas interpretações semânticas nos termos para que

a ferramenta possa relacionar os termos a conceitos, de acordo com relações linguísticas, gerando assim uma floresta conceitual. E por fim, o OntoLearn faz a integração entre essa floresta conceitual com uma base de conceitos (como a *WordNet*) para a criação de uma ontologia de domínio. A validação é feita por um especialista ao final do processo (TAWFIK; AREF; SALEM, 2011).

O Poronto é uma ferramenta apresentada por Zahra, Carvalho e Malucelli (2014), que consiste na construção semiautomática de ontologias a partir de textos em português na área de saúde e tem seu processo dividido em duas etapas: (i) criação do corpus e (ii) criação da ontologia. A ferramenta foi desenvolvida utilizando a linguagem Java e está disponível na forma de código aberto.

A criação do *corpus* tem o início com o envio de documentos enviados pelo usuário no formato PDF, que são então convertidos em texto puro passando para a fase de processamento, onde o texto é dividido por espaços em branco (tokenização) e em seguida as *stop words* são removidas.

Finalizando a construção do *corpus*, se inicia a construção da ontologia. Primeiramente são extraídos termos simples e em seguida os termos compostos, que são selecionados de acordo com alguns critérios estatísticos. Com os termos selecionados, o sistema inicia uma busca de sinônimos no OpenThesaurusPT, que devem ser validados pelo usuário. Além dos sinônimos do passo anterior, o sistema também busca mais correspondência de termos na lista de Descritores em Ciência da Saúde (DeCS). Após essas seleções, uma lista de termos é apresentada ao usuário que deverá escolher os mais relevantes para que sejam utilizados na montagem da taxonomia, que dará origem a ontologia e que pode ser exportada nos formatos OWL e XLS (Microsoft Excel).

O Prométhée é uma ferramenta baseada em aprendizado de máquina para extração e refinamento de padrões léxico-sintáticos dos relacionamentos conceituais de uma *corpora* técnica. A ferramenta usa uma base de padrões que são extraídos durante o processo de aprendizagem, e para refinar esses padrões os criadores do sistema propuseram o sistema de aprendizagem *Eagle*, que se baseia no paradigma indutivo de aprendizado por exemplos, que consiste na extração de descrição de conceitos alvos de acordo com descrições extensas desses conceitos e também de conhecimento prévio do domínio dado (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

A interface entre os dois sistemas se dá em duas etapas: (i) o Prométhée extrai os padrões sintáticos e léxicos, e então algumas instâncias desses padrões são criadas a partir do *corpus* e

classificadas de acordo com os exemplos e padrões. (ii) A partir desses padrões rotulados o *Eagle* produz as descrições que são interpretadas como restrições para refinamento dos padrões e conceitos (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

A ferramenta SOAT (*Semi-automatic Ontology Acquisition Tool*) possibilita a aquisição semiautomática de uma ontologia de domínio a partir de um *corpus* do domínio desejado. O principal objetivo da ferramenta é extrair relacionamentos de sentenças processadas com a aplicação de regras para a identificação de palavras-chave com alta ligação semântica, como hiperônimos e sinônimos (TAWFIK; AREF; SALEM, 2011).

A intervenção do usuário é necessária para a validação dos resultados da aquisição e para o refinamento e atualização das regras de extração. A ferramenta impõe um alto nível de restrição quanto a qualidade do *corpus* que está sendo utilizado, uma vez que as sentenças devem ser bem exata e precisas para que sejam suficientes para a extração dos principais e mais importantes relacionamentos (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

O SYNDIKATE é um sistema automático *stand-alone* de aprendizagem de ontologia que utiliza apenas técnicas linguísticas para executar as tarefas de aprendizado. A extração de termos é feita através de análises da estrutura sintática e análise de dependência, que consiste em utilizar uma gramática para verificar as valências binárias entre a cabeça sintática, como o pronome, e seus principais modificadores. Para que seja estabelecida uma relação entre a cabeça sintática e seus modificadores, são feitas novas análises de compatibilidade morfológica e alguns critérios semânticos tem que ser respeitados (WONG; LIU; BENNAMOUN, 2012).

A formação de conceitos, hierarquia e a criação das relações não taxonômicas é feita em seguida, utilizando modelos semânticos pré-definidos onde cada termo no grafo sintático de dependência é associado a um conceito na base de conhecimento do domínio. A abordagem então é tentar encontrar todas as ligações de relações entre conceitos correlatos a duas palavras, que ocorram em sub-grafos e que as condições gramaticais e conceituais sejam atingidas (WONG; LIU; BENNAMOUN, 2012).

O TERMINAE integra ferramentas linguísticas e de engenharia do conhecimento. A ferramenta linguística permite a criação de definições terminológicas através da análise de ocorrência de termos em um *corpus*. A ferramenta de engenharia do conhecimento é composta de um editor e navegador para ontologias (TAWFIK; AREF; SALEM, 2011).

Esse sistema ajuda na representação terminológica como conceitos, construindo esses conceitos a partir do estudo da correspondência com termos no *corpus*. A validação é feita com

a apresentação dos termos candidatos a conceito a um especialista, que irá escolher o grupo de termos adequados. Após a escolha, o especialista deve também conceitualizar os termos encontrados e fazer a análise do uso dos termos no *corpus* para definir todos os significados possíveis no domínio (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

O Text-to-Onto cria um ambiente para a construção de ontologias a partir de um núcleo ontológico inicial e também através da descoberta de estruturas conceituais em diferentes fontes em alemão, usando técnicas de aquisição de conhecimento e aprendizado de máquina (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

Para a execução do processo, foram implementadas técnicas de aprendizado de ontologia a partir de textos livres e semiestruturados, e o resultado é uma ontologia de domínio que contém conceitos independentes do domínio. Trata-se de um processo cíclico, uma vez que há a possibilidade do refinamento da ontologia a cada repetição do processo (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

O TextStorm & Clouds é um *framework* para a construção de redes semânticas que contém apenas conceitos e suas relações, usando textos relevantes para o domínio desejado. O *framework* é composto de dois módulos principais: TextStorms e Clouds, que realizam atividades complementares para a construção da rede semântica (TAWFIK; AREF; SALEM, 2011).

O TextStorm é responsável pela tarefa de extração das relações entre os conceitos a partir de textos, usando técnicas de processamento de linguagem natural. Enquanto o Clouds está concentrado em completar essas relações através de regras de extrapolação em conhecimento previamente extraído usando técnicas de processamento de linguagem natural. Após a extração dos conceitos e a criação de uma árvore conceitual, o Clouds também é o responsável pela criação da rede semântica (GÓMEZ-PÉREZ; MANZANO-MACHO, 2004).

6

Considerações finais

A relação entre as áreas da CC e CI é evidenciada por autores de ambas as áreas, como por exemplo quando Vickery (1986) afirma que a representação do conhecimento é relevante em diversas áreas, e que diferentes técnicas de representação do conhecimento foram desenvolvidas em diversas áreas, e em geral essas variações são resultadas das necessidades de manipulação de cada área, mas há algumas convergências entre as técnicas e aparentemente cada área pode aprender alguma coisa com as demais.

Ramalho (2006) também evidência essa relação entre CC e CI quando define a ontologia como “um artefato tecnológico que descreve um modelo conceitual de um determinado domínio em uma linguagem lógica e formal”.

O presente trabalho então se dedicou a mostrar como a representação do conhecimento faz uso intenso das ontologias, tendo o assunto comumente referenciados em trabalhos nos campos da organização do conhecimento e Inteligência Artificial. Inclusive seguimos a linha de argumentação apresentada por Ramalho (2010), que destaca que apesar das semelhanças, as ontologias são mais completas e adequadas ao contexto semântico do que instrumentos como tesouros e taxonomias.

Apesar de comumente citado como mecanismo de controle terminológico, acredito que essa é uma subutilização do potencial semântico das ontologias. Analogamente, seria como utilizar um supercomputador com centenas de processadores e ampla capacidade de memória para “navegar” na Internet.

Assim como compartilho da ideia apresentada por Alvarenga (2003), de que considera que a CI deve avançar na integração com outras áreas:

[...] considera-se que o maior problema da ciência da informação não seria fundamentalmente encontrar as leis da informação, mas avançar na interação entre os conhecimentos teóricos dos campos que a compõem e os conhecimentos de outras áreas diferentes de pesquisa que tenham objetos e processos comuns. Estudos teóricos que tenham por objeto a representação, a ontologia, a epistemologia, o conceito, a cognição e a hermenêutica são apenas alguns dos inúmeros tópicos teóricos que esclarecem facetas fundamentais para se estudar os fenômenos essenciais da área.

Penso que assim se torne possível mais trabalhos onde seja possível ver o arcabouço teórico da CI criar soluções interessantes para diversas áreas. Como por exemplo, quando foram apresentadas as técnicas de criação de ontologias com o objetivo de demonstrar que apesar de inúmeras tentativas e propostas de metodologias unificadas, os aspectos particulares de cada projeto e domínio acabam prevalecendo.

Outros campos na CI que vem utilizando de forma sistemática as ontologias são a Web Semântica e os sistemas de recuperação de informação. Nesses casos a integração com a CC é mais clara e até mais antiga, principalmente porque ambas as áreas (CI e CC) compartilham de muitas das ideias empregadas e implementadas. Inclusive como descrito por Ramalho, Vidotti e Fujita (2007):

Atualmente, verifica-se uma tendência de aproximação entre as áreas de Ciência da Informação e Ciência da Computação, principalmente no que tange ao desenvolvimento de novos instrumentos de representação e recuperação de recursos informacionais. Segundo Saracevic (1996), a recuperação de informação pode ser considerada como a vertente tecnológica da Ciência da Informação.

E por fim, as ontologias podem ser importante aliadas na pesquisa de automação de processos informacionais como classificação, indexação, catalogação e anotação de metadados. Com o poder das redes semânticas, por exemplo, conjugada às tecnologias de aprendizado de máquina, as ontologias se tornam ferramentas importantes para que o conhecimento seja adequadamente representado e expandido, possibilitando que haja a representação e extração do conhecimento sem necessariamente haver a interferência de valores culturais e comportamentais.

Dessa maneira, entendo que apesar das ontologias já serem utilizadas em alguns campos da CI, seja possível através da transdisciplinaridade e aplicação de conceitos computacionais, tornar-se uma ferramenta ainda mais poderosa do ponto de vista semântico para a CI, como foi identificado e descrito pela pesquisa apresentada.

6.1 Trabalhos futuros

A presente pesquisa pode ser tomada como um ponto de partida para a identificação do potencial semântico das ontologias na CI, principalmente quanto a representação do conhecimento. A escolha dessa linha de condução do trabalho foi justamente por considerar que a partir de um melhor entendimento de uma das funções básicas de uma ontologia, seria possível então dar um passo a frente em direção a trabalhos mais profundos e aplicados.

O primeiro campo que fica claro é quanto a representação do conhecimento em si, apesar de já haver vários trabalhos que foram inclusive citados aqui, uma linha pouco explorada ainda é a automação dos processos de representação dos recursos informacionais. Pesquisas já existem, mas acredito que uma maior integração com os últimos tópicos de IA, como *Deep Learning* e avanços do aprendizado de máquina em geral, podem gerar resultados mais sólidos se aplicados aos anos de experiência de especialistas humanos em algoritmos modernos e redes neurais complexas.

Ferneda e Ataíde (2017) propõe o OntoSmart, um modelo de recuperação da informação baseado em ontologias, onde no modelo de recuperação proposto, elementos linguísticos que formam uma ontologia são considerados termos de um vocabulário de domínio, utilizado como ferramenta de padronização terminológica das representações dos documentos e das buscas em um sistema de recuperação de informação. Tais representações utilizam como base formal o Modelo Espaço Vetorial, que fornece uma base matemática consistente e consolidada. E de acordo com os resultados apresentados pela pesquisa, os próprios pesquisadores apontam a necessidade de mais estudos e propostas nessa área.

Referências

ABBURU, Sunitha; BABU, Suresh. Survey on Ontology Construction Tools. **International Journal Of Scientific & Engineering Research**. N.i., p. 1748-1752. jun. 2013. Disponível em: <<https://www.ijser.org/researchpaper/Survey-on-Ontology-Construction-Tools.pdf>>. Acesso em: 28 maio 2018.

ALMEIDA, Mauricio Barcellos. Revisiting ontologies: A necessary clarification. **Journal Of The American Society For Information Science And Technology**, [s.l.], v. 64, n. 8, p.1682-1693, 31 maio 2013. Wiley. <http://dx.doi.org/10.1002/asi.22861>. Disponível em: <<http://mba.eci.ufmg.br/downloads/pos/RevisitingOntologies-Almeida.pdf>>. Acesso em: 22 jun. 2018.

ALMEIDA, Maurício Barcellos. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. **Perspectivas em Ciência da Informação**, Belo Horizonte, Mg, v. 19, n. 3, p.242-258, set. 2014. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1981-5344/1736>. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1736/1448>>. Acesso em: 03 jun. 2018.

BROWARNIK, Abel; MAIMON, Oded. Ontology Learning from Text: Departing the Ontology Layer Cake. In: THE FIRST INTERNATIONAL CONFERENCE ON BIG DATA, SMALL DATA, LINKED DATA AND OPEN DATA, 1., 2015, Barcelona, Espanha. **Proceedings...** . Wilmington, Eua: Iaria Xps Press, 2015. v. 1, p. 62 - 68. Disponível em: <<https://www.iaria.org/conferences2015/ALLDATA15.html>>. Acesso em: 15 set. 2017.

BUITELAAR, Paul; WEBER, Nicolas; CIMIANO, Philipp. Ontology Learning and Population in SmartWeb. In: PHILIPS SYMPOSIUM ON INTELLIGENT ALGORITHMS (SOIA), 3., 2006, Amsterdam, Holanda. **Proceedings...** . [s.i]: [s.i], 2006. p. 1 - 12. Disponível em: <<https://pdfs.semanticscholar.org/9494/f3d4454c317c3d38df794e30ede2243cc9d7.pdf>>. Acesso em: 12 mar. 2017.

CAIXETA, Mario; SOUZA, Renato Rocha. Representação do conhecimento: história, sentimento e percepção.. **Informação & Informação**, Londrina, Pr, v. 13, n. 2, p.34-55, 15 dez. 2008. Universidade Estadual de Londrina. <http://dx.doi.org/10.5433/1981-8920.2008v13n2p34>. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1815>>. Acesso em: 16 jun. 2018.

CAMPOS, Maria Luiza de Almeida. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, Df, v. 33, n. 1, p.22-32, jun. 2004. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1064>>. Acesso em: 29 abr. 2018.

CHANDRASEKARAN, B.; JOSEPHSON, John R.; BENJAMINS, V. Richard. What are ontologies, and why do we need them? **IEEE Intelligent Systems**, [s.l.], v. 14, n. 1, p.20-26, jan. 1999. Bimestral. Institute of Electrical and Electronics Engineers (IEEE).
<http://dx.doi.org/10.1109/5254.747902>.

CHAUI, Marilena. **Convite à Filosofia**. São Paulo, Sp: Ática, 2000. 567 p.

CIMIANO, Philip et al (Org.). Ontology Learning. In: STAAB, Steffen; STUDER, Rudi (Org.). **Handbook on Ontologies**. 2. ed. Berlin: Springer, 2009. Cap. 12. p. 245-267. (International Handbooks on Information Systems).

CIMIANO, Philipp. **Ontology Learning and Population from Text: Algorithms, Evaluation and Applications**. New York, Ny, Eua: Springer Science-i-business Media, 2006. 361 p.

DELLSCHAFT, Klaas; STAAB, Steffen. Strategies for the Evaluation of Ontology Learning. In: BUITELAAR, Paul; CIMIANO, Philipp (Ed.). **Ontology Learning and Population: Bridging the Gap between Text and Knowledge**. Amsterdam, Holanda: Ios Press, 2008. Cap. 5. p. 253-288.

FARIAS, Karla Meneses; PINHO, Fábio Assis. ONTOLOGIAS COMO FERRAMENTA DE ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO: um olhar sobre os laudos médico-legais. **Informação em Pauta**, Fortaleza, Ce, v. 1, n. 2, p.41-65, dez. 2016. Semestral. Disponível em: <<http://periodicos.ufc.br/informacaoempauta/article/view/5431>>. Acesso em: 22 jun. 2018.

FERNEDA, Edberto. Ontologia como recurso de padronização terminológica de um sistema de recuperação de informação. 2013a. 96 f. Relatório de Pesquisa (Pós-Doutorado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2013a.

FERNEDA, Edberto; DIAS, Guilherme Ataíde. OntoSmart: um modelo de recuperação de informação baseado em ontologia. **Perspectivas em Ciência da Informação**, Belo Horizonte, MG, v. 22, n. 2, p.170-187, jun. 2017.

GÓMEZ-PÉREZ, Asunción; FERNÁNDEZ-LÓPEZ, Mariano; CORCHO, Oscar. **Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web..** Londres, Inglaterra: Springer-verlag London, 2003. 404 p. (Advanced Information and Knowledge Processing).

GÓMEZ-PÉREZ, Asunción; MANZANO-MACHO, David. An overview of methods and tools for ontology learning from texts. **The Knowledge Engineering Review**, Londres, Inglaterra, v. 19, n. 3, p.187-212, set. 2004. Disponível em: <<https://doi.org/10.1017/S0269888905000251>>. Acesso em: 26 ago. 2017.

GRIMM, Stephan; HITZLER, Pascal; ABECKER, Andreas. Knowledge Representation and Ontologies: Logic, Ontologies and Semantic Web Languages. In: STUDER, Rudi; GRIMM, Stephan; ABECKER, Andreas (Ed.). **Semantic Web Services: Concepts, Technologies, and Applications**. Berlin, Alemanha: Springer-verlag Berlin Heidelberg, 2007. Cap. 3. p. 51-105.

GRUBER, Thomas R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, Inglaterra, v. 5, n. 2, p.199-220, jun. 1993. Trimestral. Elsevier BV. <http://dx.doi.org/10.1006/knac.1993.1008>. Disponível em: <<http://www.sciencedirect.com/science/journal/10428143>>. Acesso em: 09 jun. 2018.

GUARINO, Nicola. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In: PAZIENZA, Maria T. (Ed.). **Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology**. Berlin: Springer-verlag, 1997. Cap. 8. p. 139-170. (Lecture Notes in Artificial Intelligence).

GUARINO, Nicola; GIARETTA, Pierdaniele. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: MARS, N. J. I.. **Towards Very Large Knowledge Bases, Knowledge Building & Knowledge Sharing 1995**. Amsterdã, Holanda: Ios Press, 1995. Cap. 4. p. 33-45.

GÖCKEL, Rudolph. **Lexicon Philosophicum**. Marchioburgi: Hutwelcker, 1615. 390 p. Disponível em: <<http://www.mdz-nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb10008451-1>>. Acesso em: 10 out. 2017.

HAPPEL, Hans-jörg; SEEDORF, Stefan. Applications of Ontologies in Software Engineering. In: INTERNATIONAL WORKSHOP ON SEMANTIC WEB ENABLED SOFTWARE ENGINEERING, 2., 2006, Athens, Ga, Eua. **Proceedings... . [s.i]**: Springer-verlag Berlin Heidelberg, 2006. p. 1 - 14. Disponível em: <https://km.aifb.kit.edu/ws/swese2006/final/happel_full.pdf>. Acesso em: 21 fev. 2018.

LOPES, Lucelene. **Extração automática de conceitos a partir de textos em língua portuguesa**. 2012. 156 f. Tese (Doutorado) - Curso de Ciência da Computação, Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rs, 2012. Disponível em: <<http://hdl.handle.net/10923/1651>>. Acesso em: 5 mar. 2017.

MAEDCHE, Alexander; STAAB, Steffen. Ontology learning for the Semantic Web. **Ieee Intelligent Systems**, [s.l.], v. 16, n. 2, p.72-79, mar. 2001. Semestral.

MENDONÇA, Fabrício Martins. **ONTOFORINFOSCIENCE: METODOLOGIA PARA CONSTRUÇÃO DE ONTOLOGIAS PELOS CIENTISTAS DA INFORMAÇÃO: Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA)**. 2015. 320 f. Tese (Doutorado) - Curso de Pós-graduação em Ciência da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, Mg, 2015. Disponível em: <<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/BUBD-A35H3K>>. Acesso em: 12 dez. 2017.

NEON PROJECT. **NeOn Toolkit**. 2011. Disponível em: <<http://neon-toolkit.org/index.html>>. Acesso em: 28 maio 2018.

NOY, Natalya F.; MCGUINNESS, Deborah L.. **Ontology Development 101: A Guide to Creating Your First Ontology**. 2001. Coordenada por Stanford Center for Biomedical Informatics Research (BMIR). Disponível em: <<https://protegewiki.stanford.edu/wiki/Ontology101>>. Acesso em: 15 jan. 2018.

Open Source Initiative. **The 2-Clause BSD License**. 1999. Disponível em: <<https://opensource.org/licenses/BSD-2-Clause>>. Acesso em: 28 maio 2018.

RAMALHO, Rogério A. S.. **Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação**. 2006. 120 f. Dissertação (Mestrado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2006.

RAMALHO, R.A.S.. **Desenvolvimento e utilização de ontologias em Bibliotecas Digitais: uma proposta de aplicação**. Tese (Doutorado em Ciências da Informação) – Universidade Estadual Paulista, 2010.

REICHGELT, Han (Ed.). **Knowledge representation: an AI perspective**. New Jersey, Ny, Eua: Ablex Publishing Corporation, 1991. 251 p.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3. ed. Upper Saddle River, Nj, Eua: Prentice Hall, 2010. 1152 p.

SUÁREZ-FIGUEROA, Mari Carmen. **NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse**. 2010. 288 f. Tese (Doutorado) - Curso de Doctorado En Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Madrid, Espanha, 2010. Disponível em: <<http://oa.upm.es/3879/>>. Acesso em: 31 abr. 2018.

TAWFIK, Marco Alfonse; AREF, Mostafa M.; SALEM, Abdel-badeeh M.. An Overview of Ontology Learning From Unstructured Texts. In: **INFORMATICS 2011**, 11., 2011, Rožava, Slovakia. **Proceedings of the Eleventh International Conference on Informatics**. Rožava: Technical University Of Košice, 2011. p. 169 - 174.

VECHIATO, Fernando Luiz; VIDOTTI, Silvana A. B. Gregorio. **Encontrabilidade da informação**. São Paulo, Sp: Cultura Acadêmica, 2014. 198 p. Disponível em: <<https://repositorio.unesp.br/handle/11449/126218>>. Acesso em: 24 out. 2016.

VICKERY, B. C.. Ontologies. **Journal Of Information Science**, Londres, Inglaterra, v. 23, n. 4, p.277-286, ago. 1997. Bimestral. SAGE Publications. <http://dx.doi.org/10.1177/016555159702300402>.

VICKERY, Brian C.. KNOWLEDGE REPRESENTATION: A BRIEF REVIEW. **Journal Of Documentation**, Londres, Inglaterra, v. 42, n. 3, p.145-159, mar. 1986. Emerald. <http://dx.doi.org/10.1108/eb026790>.

WONG, Wilson; LIU, Wei; BENNAMOUN, Mohammed. Ontology learning from text: A Look Back and into the Future. **Acm Computing Surveys**, New York, Ny, Eua, v. 44, n. 4, p.1-36, 1 ago. 2012.

WORLD WIDE WEB CONSORTIUM (W3C). **Ontology editors**. 2010. Disponível em: <https://www.w3.org/wiki/Ontology_editors>. Acesso em: 09 ago. 2010.

YIN, Chi-yen; LEE, Yau-jung; YANG, Jiann-min. Ontology: the historical review and literature productivity analysis using bibliometric methodology from 1956 to 2008. In: INTERNATIONAL CONFERENCE ON INTERACTION SCIENCES: INFORMATION TECHNOLOGY, CULTURE AND HUMAN, 2., 2009, Seoul, Korea. **Proceedings...** . New York, Ny, Usa: Acm, 2009. v. 1, p. 1346 - 1350.

ZAHRA, Faruk Mustafa et al. Ferramentas para aprendizagem de ontologias a partir de textos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 1, p.03-21, mar. 2014. Trimestral. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1413-99362014000100002>.