

Ontologias e vocabulários controlados: comparação de metodologias para construção

Daniela Lucas da Silva

Mestre em ciência da informação. Escola de Ciência da Informação (ECI/UFMG).

E-mail: danielalucas@hotmail.com

Renato Rocha Souza

Professor adjunto da Escola de Ciência da Informação (ECI/UFMG).

E-mail: rsouza@eci.ufmg.br

Maurício Barcellos Almeida

Professor adjunto da Escola de Ciência da Informação (ECI/UFMG).

E-mail: mba@eci.ufmg.br

Resumo

O artigo se propõe a apresentar um estudo analítico sobre metodologias e métodos para construção de ontologias e vocabulários controlados mediante análise da literatura sobre metodologias para construção de ontologias e de normas internacionais para construção de *software* e de vocabulários controlados. Por meio de pesquisa teórica e empírica, foi possível construir um panorama comparativo que pode servir de apoio na definição de padrões metodológicos para construção de ontologias através da integração de princípios teóricos e metodológicos da ciência da informação da ciência da computação, bem como de contribuições de metodologias e métodos conhecidos para construção de ontologias.

Palavras-chave

Ontologias. Vocabulários controlados. Construção de ontologias. Construção de vocabulários controlados. Tesouros.

Ontologies and controlled vocabularies: comparison of methodologies for construction

Abstract

This article presents an analytical study about methodology and methods to develop ontologies and controlled vocabularies, built through the analysis of the literature about methodologies for ontology engineering and international standards for the construction of software and controlled vocabularies. Through theoretical and empirical research it was possible to build a comparative overview that helps as a guideline for building ontologies using theories from computer and information science along with empirical research with well known methods and methodologies for building ontologies.

Keywords

Ontologies. Controlled vocabularies. Ontology building. Ontology engineering. Thesauri.

1. INTRODUÇÃO

A organização da informação tornou-se um processo fundamental à medida que vem crescendo exponencialmente o volume de informações disponível, resultando muitas vezes na desorganização de acervos informacionais e conseqüentemente na dificuldade de se encontrar o que se procura em determinado sistema de recuperação de informação. Nesse sentido, pesquisas têm sido desenvolvidas progressivamente visando ao desenvolvimento de mecanismos de indexação, organização e recuperação de informações, com o objetivo único de melhorar a eficácia dos sistemas de recuperação de informação. Podemos citar, dentre outras, algumas pesquisas nessa perspectiva voltadas à exploração semântica da informação: a) a Web Semântica, que pretende criar metodologias, tecnologias e padrões de metadados para aumentar o escopo das atividades desempenhadas automaticamente (BERNERS-LEE; HENDLER; LASSILA, 2001); b) a utilização semântica embutida nos próprios documentos com o uso de estruturas da linguagem natural como os sintagmas nominais e verbais (SOUZA, 2005); c) instrumentos de representação de relacionamentos semânticos e conceituais como as ontologias (GRUBER, 1993) e os tesouros (DAHLBERG, 1978), objetivando evitar problemas relacionados à ambigüidade inerente às palavras da linguagem natural.

Esse fato contribui para a atenção dada às ontologias, cuja origem se dá no campo teórico da filosofia (CORAZZON, 2008), sendo ainda pesquisadas e desenvolvidas como instrumento de representação de conhecimento nos campos das ciências da computação e da informação. Para a ciência da informação, as ontologias são de interesse pela potencialidade que elas têm em organizar e representar informação. Segundo Almeida e Bax (2003), as ontologias podem melhorar os processos de recuperação de informação ao organizar o conteúdo de fontes de dados em um determinado domínio.

Borst (1997, p.12) apresenta uma definição muito aceita pela comunidade de ontologia: “uma especificação formal e explícita de uma conceitualização compartilhada”, em que “formal” significa legível para computadores;

“especificação explícita” estaria relacionada a conceitos, propriedades, axiomas explicitamente definidos; “compartilhado” seria conhecimento consensual; e “conceitualização” diz respeito a um modelo abstrato de algum fenômeno do mundo real. Como componentes da ontologia (GRUBER, 1993a; GOMEZ-PEREZ; FERNANDEZ; VICENTE, 1996), têm-se o seguinte: a) classes conceituais, que organizam os conceitos de um domínio em uma taxonomia; b) atributos de classes, que são propriedades relevantes do conceito; c) instâncias, que são utilizadas para representar objetos específicos de um conceito; d) atributos de instâncias, que são propriedades relevantes que descrevem as instâncias de um conceito; e) relações entre classes, que representam o tipo de interação entre os conceitos de um domínio; f) constantes, que têm sempre o mesmo valor e, geralmente, são usadas em regras ou fórmulas para inferir conhecimento na ontologia; g) termos, que designam os conceitos de um domínio; h) axiomas formais, que restringem a interpretação e o uso dos conceitos envolvidos na ontologia; i) regras, que determinam condições ao domínio, além de inferir valores para atributos.

Entretanto, dentro do domínio de desenvolvimento de ontologias, as abordagens para a sua construção, na maioria das vezes, são específicas e limitadas. Um problema, do ponto de vista metodológico, é que não há um padrão para sua construção (FERNÁNDEZ *et alii.*, 1999; USCHOLD; GRUNINGER, 1996). Apesar de grande quantidade de ontologias já ter sido desenvolvida por diferentes comunidades – na química (GÓMEZ-PEREZ; FERNANDEZ; VICENTE, 1996) e na modelagem de negócio (GRUNINGER e FOX, 1995), só para citar alguns exemplos – sob diferentes abordagens e usando diferentes métodos e técnicas, não há consenso sobre uma metodologia para o processo de construção (FERNANDEZ; GÓMEZ-PEREZ; JURISTO, 1997). A consequência, segundo Fernandez, Gomez-Perez e Juristo (1997), é a ausência de atividades padronizadas, cujas realizações são conduzidas de forma artesanal e não como uma atividade científica. Além disso, verifica-se a falta de explicação sistemática de como e onde serão usadas as abordagens teóricas dentro de seu processo de elaboração.

Pesquisadores da ciência da informação (SOERGEL, 1997; VICKERY, 1997; GILCHRIST, 2003) apresentam similaridades entre vocabulários controlados utilizados na biblioteconomia, como os tesouros e as taxonomias, e instrumentos utilizados na inteligência artificial, como as ontologias. As similaridades estão principalmente na

forma de elaboração da estrutura desses instrumentos, que demanda a organização de conceitos em processos que incluem categorização e classificação de conceitos, definição das relações entre esses conceitos e tratamento da terminologia empregada nos conceitos e relações da estrutura.

Diante do cenário exposto, este artigo se propõe a apresentar um estudo analítico sobre metodologias e métodos para construção de ontologias mais comumente encontrados na literatura e metodologias e normas para construção de vocabulários controlados disponíveis, de modo a delinear um panorama comparativo sobre a construção de tais instrumentos. Tal panorama pode contribuir na definição de padrões metodológicos para construção de ontologias por meio da integração de princípios teóricos e metodológicos da ciência da informação, da ciência da computação, bem como de contribuições de metodologias e métodos conhecidos para construção de ontologias.

2. METODOLOGIA

O objetivo geral deste artigo é apresentar um estudo comparativo entre metodologias para construção de ontologias e vocabulários controlados. Para tal, os passos metodológicos adotados na pesquisa foram os seguintes: i) identificação de documentos que contemplassem o assunto sobre metodologias para construção de ontologias; ii) análise, interpretação e seleção de tais documentos para identificação das metodologias discutidas nos mesmos; iii) seleção de metodologias para construção de vocabulários controlados; iv) definição das categorias de análise das metodologias, métodos e normas, de modo a coletar os dados pertinentes à pesquisa; v) análise comparativa das metodologias, dos métodos e da norma.

Para a identificação de documentos, foram consultadas bases de dados de documentos científicos e empregada a técnica de busca por palavras-chave que refletissem o universo do assunto. O processo de análise e interpretação dos documentos recuperados se deu mediante a utilização de técnicas de análise de conteúdo, as quais permitiram a seleção dos documentos. Após a obtenção da amostragem documental, pôde-se, finalmente, identificar as metodologias discutidas nos documentos e justificar quando da inserção de uma delas no material empírico destinado à análise. Tal identificação também foi realizada com técnicas de análise de conteúdo. A análise de conteúdo contempla um conjunto de técnicas de análise das comunicações

visando a obter indicadores (quantitativos ou não) que permitam a inferência de conhecimentos presentes nas mensagens (BARDIN, 1977). O emprego de tais técnicas é descrito nas seções 2.2 e 2.3.

Essa seção apresenta a descrição dos passos metodológicos que tornaram possível a concretização deste trabalho de pesquisa. No tópico 2.1 são apresentadas as bases de dados, que determinaram a recuperação dos documentos pesquisados. Em 2.2 são descritos os passos e os critérios utilizados na seleção das metodologias para construção de ontologias e das metodologias para construção de vocabulários controlados, ambos compondo o material empírico para análise. No tópico 2.3 são justificadas as categorias de análise do material empírico. E, finalmente, o tópico 2.4 apresenta como foi conduzida a atividade de análise comparativa dos materiais envolvidos na pesquisa.

2.1 Seleção das fontes de pesquisa

As fontes de obtenção dos artigos científicos para o tema *construção de ontologias* foram o portal de periódicos da Capes* e a biblioteca digital Citeseer**, sendo que ambas as fontes disseminam literatura científica nos campos da ciência da informação e da ciência da computação. Além disso, proporcionam análises estatísticas das citações para todos os documentos do repositório, o que possibilita ao pesquisador tomar conhecimento da opinião dos pares em relação ao artigo em questão e, em consequência, de sua aceitação pela comunidade.

Os livros sobre *construção de ontologias* foram pesquisados no catálogo de bibliotecas da Universidade Federal de Minas Gerais (UFMG) pelo sistema Pergamum***, seguindo o critério de maior disponibilidade.

O acesso aos documentos relacionados à *elaboração de vocabulários controlados* se deu a partir das citações de artigos procurados através dos descritores “tesauros” e “vocabulários controlados” na base de dados Citeseer. Pesquisas também foram realizadas nas revistas classificadas na categoria *Qualis* da ciência da informação e na biblioteca Digital Brasileira de Teses e Dissertações (BDTD****).

* <http://www.periodicos.capes.gov.br/portugues/index.jsp>

** <http://citeseer.ist.psu.edu>

*** <http://webpergamum.adm-serv.ufmg.br/biblioteca/php/opcoes.php>

**** <http://bdttd.ibict.br>

A indexação das citações no Citeseer é feita de forma autônoma, ou seja, sem intervenção humana (LAWRENCE, BOLLACKER; LEE GILES, 1999). Quando da execução deste estudo, foi constatada a atualização da base de dados do Citeseer até o período de setembro de 2006[†]. No que diz respeito ao portal de periódicos da Capes, a editora selecionada para consultas foi a Association Computing Machinery, conhecida como Portal ACM, que dissemina literatura de diversas áreas da Ciência da Computação, destacando a inteligência artificial, área na qual foram selecionados artigos científicos sobre *construção de ontologias*.

2.2 Seleção do material empírico

A presente seção apresenta os passos que determinaram a seleção do material empírico nesse estudo e se organiza da seguinte forma: na seção 2.2.1 é descrita a estratégia de pesquisa utilizada na atividade de seleção das metodologias para construção de ontologias; e a seção 2.2.2 apresenta os critérios adotados na seleção das metodologias para construção de vocabulários controlados.

2.2.1 Seleção das metodologias para construção de ontologias

Na atividade de seleção das metodologias para construção de ontologias, foi realizada uma investigação na literatura das áreas da computação, como também da ciência da informação, tendo-se mais sucesso na primeira. Tal fato pode ser explicado em razão de as ontologias de sistemas de informações serem objetos de estudo proeminentes da ciência da computação.

A estratégia de pesquisa para a seleção das metodologias se deu pelos seguintes passos, que serão detalhados nas próximas subseções:

1. seleção de artigos e livros que tratam do assunto “metodologias para construção de ontologias” através da recuperação por palavras-chave determinadas;
2. identificação de metodologias para construção de ontologias nos documentos selecionados;
3. quantificação de ocorrências de citação das metodologias identificadas nos documentos;
4. finalmente, determinação das metodologias que foram analisadas comparativamente na pesquisa.

[†] <http://citeseer.ist.psu.edu/source.html>

É válido destacar neste ponto a diferença entre os termos *metodologia* e *método*, visto que, na investigação realizada na literatura, ambos os termos são usados por vezes de maneira indiscriminada. Contudo, constatou-se que alguns dos objetos a serem identificados e analisados são considerados metodologias, outros são considerados métodos, conforme será visto na seção 3, que apresenta os objetos empíricos.

Segundo definições do IEEE (1990)^{*}, *apud* Fernandez, Gómez-Perez e Corcho (2004), uma *metodologia* seria “uma série integrada de técnicas ou métodos criando uma teoria geral de sistemas de como uma classe de pensamento pode ser executada”^{***}. Um *método* “é um conjunto de processos ou procedimentos ordenados usados na engenharia de um produto ou na realização de um serviço”^{****}, tendo-se em vista que tais processos são compostos de atividades, que, por sua vez, são compostas por tarefas atribuídas a um ou mais membro do projeto. Geralmente, tarefas relacionadas são agrupadas para formar atividades. E uma técnica seria “um procedimento técnico e gerencial usado para alcançar um dado objetivo”^{*****}, que em outras palavras significa dizer um modo pelo qual o método é executado.

Segundo de Hoog (1998)^{*****}, *apud* Fernandez, Gómez-Perez e Corcho (2004), metodologia e método são conceitos distintos, pois uma metodologia refere-se a conhecimento sobre métodos, isto é, determina “como” e “quando” uma dada atividade pode ser realizada. Assim, uma metodologia é composta de métodos que possuem suas próprias técnicas.

2.2.1.1 Passo 1 – Seleção dos documentos

O primeiro passo foi selecionar os artigos científicos que fossem passíveis de análise para levantamento das metodologias discutidas na literatura. O processo de recuperação de artigos na biblioteca Citeseer se dá por palavras-chave ou por *links*^{*****} de citação (LAWRENCE; BOLLACKER; LEE GILES, 1999). A biblioteca oferece

uma opção de busca de seu conteúdo por meio do mecanismo de busca do *Google* (denominada *Google Docs*). Essa estratégia de busca é interessante, pois, para uma maior precisão, o *Google* pesquisa exatamente as palavras ou expressões que forem indicadas no campo de pesquisa por aspas. O *Google* também adiciona automaticamente um “e” entre as palavras inseridas na expressão de pesquisa. O resultado mostra páginas onde ocorram todas as palavras indicadas na busca.

Diante disso, as palavras-chave precisaram ser determinadas de forma a representar o assunto, ou seja, “metodologias para construção de ontologias”. As palavras-chave para a atividade de recuperação foram então escolhidas em função do conhecimento prévio adquirido na leitura de documentos sobre o assunto: “*methodology*”, “*methodologies*”, “*method*”, “*design*”, “*building*”, “*development process*”, “*life cycle*”, em conjunto com “*ontologies*” ou “*ontology*”. É importante salientar que as palavras-chave estão em inglês em virtude de os materiais publicados sobre ontologias se encontrarem, em sua maioria, nesta língua.

A consulta montada com a expressão “*methodology for building ontologies*” resultou na recuperação de 70 páginas no *Google Docs*, já “*methodologies for building ontologies*” resultou na recuperação de 22 páginas. Além da estratégia de busca por expressão, foi pesquisada cada palavra-chave em conjunto com “*ontologies*” ou “*ontology*”. Na análise de busca das palavras-chave, constatou-se que a palavra “*ontology*” em conjunto com as outras palavras-chave, foi a que resultou em maior retorno de páginas. Na análise realizada nos dois contextos, percebeu-se uma coincidência nas páginas recuperadas, isto é, a mesma página era retornada tanto com “*ontologies*” quanto “*ontology*”, constatando uma variação no uso dos termos nos artigos.

Do universo de páginas recuperadas, foram investigadas as 30 primeiras, pelo fato de, na maioria dos casos, as demais páginas retornarem documentos repetidos ou fora do escopo, tendo em vista que os resultados são listados em ordem de importância^{*}. Foram selecionados os documentos que mais se aproximavam do objeto de investigação. Para efetuar a escolha, foi necessário analisar o título mencionado na página, fazer uma leitura do resumo (*abstract*) do documento^{**} e analisar as palavras-

^{*} IEEE Standard Glossary of Software Engineering Terminology. IEEE Computer Society. New York. IEEE Std 610.121990, 1990.

^{**} “[...] integrated series of techniques or methods creating a general systems theory of how a class of thought [...] be performed.”

^{***} “[...] is a set of orderly process or procedure used in the engineering of a product or performing a service.”

^{****} “[...] a technical and managerial procedure used to achieve a given objective”

^{*****} de HOOG R. Methodologies for Building Knowledge Based Systems: Achievements and Prospects. In: Liebowitz J (ed) Handbook of Expert Systems. Chapter 1, Boca Raton, Florida: CRC Press. 1998.
^{*****} “CiteSeer allows the location of papers by keyword search or by citation links”

^{*} Informações sobre o PageRank em <http://www.google.com/technology/>

^{**} A página recuperada no *Google Docs* remete ao documento indexado no Citeseer. A biblioteca disponibiliza um recurso de *download* para visualização do documento em vários tipos de arquivo.

chave do mesmo. Além disso, para uma compreensão macro do objeto de investigação, fez-se breve leitura da seção envolvida com o assunto. Tal leitura foi embasada em uma técnica da análise de conteúdo conhecida como *leitura flutuante*, na qual consiste em estabelecer contato com os documentos a serem analisados, obtendo impressões e orientações acerca do conteúdo (BARDIN, 1977, p. 96-98). Algumas regras advindas da análise de conteúdo também foram consideradas na seleção dos documentos e são descritas a seguir:

- regra da exaustividade: “não se pode deixar de fora qualquer elemento por esta ou por aquela razão que não possa ser justificável no plano do rigor”. Muitos dos documentos filtrados não abordavam diretamente o tema “metodologias para construção de ontologias”, mas outros aspectos relacionados, tais como “ferramentas”, “linguagens”, “aplicações”, “engenharia”, dentre outros, que poderiam servir para a análise;

- regra da representatividade: “a amostragem diz-se rigorosa se a amostra for parte representativa do universo inicial”. Nesse caso, representando os documentos que tratassem diretamente de metodologias para construção de ontologias.

Além das técnicas expostas acima, foram levados em consideração na análise de cada artigo: a) o número de citações recebidas; b) citação de autores mais produtivos na área de ontologias, como Nicolas Guarino, Thomas Gruber, Fernandez Lopes, Asuncion Gómez Perez, Mike Uschold, Michael Gruninger, dentre outros.

É válido ressaltar que, a partir dos documentos selecionados, foi possível identificar novos conceitos e autores para realização de novas buscas no Citeseer, bem como na editora ACM do portal da Capes, que permitiram a identificação de outros artigos relevantes para a pesquisa. Tais artigos também foram considerados na análise de identificação de metodologias para construção de ontologias e, quando considerados relevantes, foram inseridos na amostra de documentos.

Dando continuidade ao processo de seleção de documentos, o próximo passo foi recuperar livros que tratassem de metodologias para construção de ontologias. Para tal, foi acessado o sistema de gestão de acervos das bibliotecas da UFMG (Pergamum). As referências foram coletadas a partir de bases secundárias que refletem a literatura nacional e internacional nos campos de ciência da informação e ciência da computação. Investigou-se o acervo *on-line* das bibliotecas da Escola de Ciência da

Informação (ECI) e do Instituto de Ciências Exatas (ICEX), por serem considerados campos mais pertinentes em relação ao assunto e para diminuir o “ruído” causado pelas acepções de ontologia no campo da filosofia. As palavras-chave determinadas para a busca foram: “ontologia”, “ontologias”, “ontology”, “ontologies”, “engineering” e “web”.

Foram recuperadas obras que tratavam de assuntos relacionados a comércio eletrônico, *web* semântica e recuperação de informação. Foi constatado pouco volume de obras relacionadas à temática “ontologias de sistemas de informações” nos acervos da ECI, bem como nos do ICEX, evidenciando que o assunto de pesquisa ainda é recente nas duas instituições. Foi feita uma análise no sumário de cada obra a fim de verificar a existência do assunto investigado.

Após a realização da atividade de seleção dos documentos descrita anteriormente, foi obtida uma amostra de 25 documentos representativos de análise, os quais podem ser conferidos em Silva (2008, p. 119-120). Foram selecionados 23 artigos científicos e dois capítulos de livros.

2.2.1.2 Passo 2 – Identificação das metodologias para construção de ontologias

O segundo passo foi identificar as metodologias para construção de ontologias abordadas nos 25 documentos. No passo de seleção dos documentos, pôde-se perceber uma carência de metodologias representativas. Segundo a literatura pesquisada, do ponto de vista metodológico, não há um padrão amplamente aceito para construção de ontologias (JONES; BENCH-CAPON; VISSER, 1998; FERNÁNDEZ, GÓMEZ-PÉREZ; JURISTO, 1997; USCHOLD; GRUNINGER, 1996). Este fato demonstra que “*metodologia para construção*” é um assunto ainda em amadurecimento. Uschold e Gruninger (1996) comentam da dificuldade em encontrar, na literatura da área, assuntos sobre “metodologias para construção de ontologias”. Segundo os autores, “[...] não há muitas publicações na área, mesmo na literatura pesquisada”*. Mesmo em função de essas afirmações terem sido feitas a cerca de 10 anos, não foi constatado nessa pesquisa um avanço em relação a propostas de metodologias para construção de ontologias. Isso pode ser conferido no mapeamento temporal das metodologias abordadas nos documentos (1992 até 2007) e em publicações recentes

* “[...] nor is there much published in this área, even in the research literature”.

envolvendo o assunto (FERNÁNDEZ, GÓMEZ-PÉREZ e CORCHO, 2004; BREITMAN, 2005; CARDOSO, 2007).

Todos os fatos expostos anteriormente, além de frisar a importância deste estudo, reforçam o argumento de que a fase de seleção das metodologias para construção de ontologias apresentou dificuldades. Os motivos foram os seguintes: a) as metodologias discutidas na literatura eram quase sempre as mesmas; b) poucos trabalhos publicados sobre o tema. Diante disso, julgou-se pertinente, em termos de obtenção de metodologias representativas de análise, obter a **freqüência de ocorrência** de citação nos documentos selecionados. Através da tabela de freqüência de ocorrência, disponível em Silva (2008, p.123), tornar-se-ia possível identificar as metodologias de construção mais discutidas e dar ênfase na metodologia que mais se destacou na literatura.

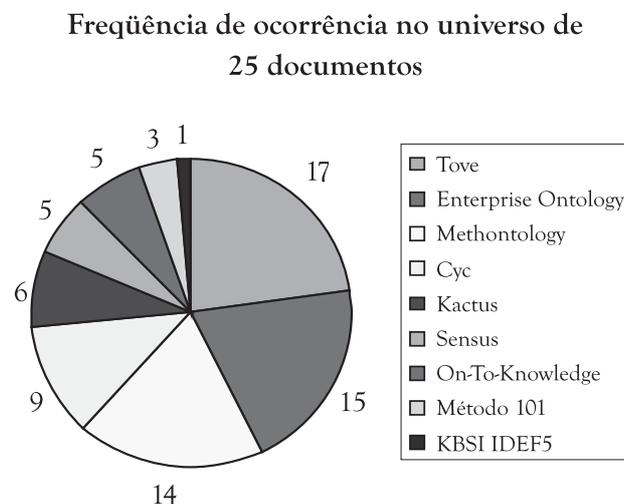
2.2.1.3 Passo 3 – Apresentação da freqüência de ocorrência de citação

As metodologias e os métodos com maior freqüência de ocorrência foram a metodologia usada no projeto **Tove** (68%), o método proposto para o projeto **Enterprise Ontology** (60%) e a metodologia **Methontology** (56%), sendo abordadas em mais de 50% dos trabalhos selecionados. Em seguida, vêm o método **Cyc** com 36%, o método **Kactus** com 24%, o método **Sensus** e a metodologia **On-To-Knowledge** com 20%, o método **101** com 12% e o método **KBSI IDEF5** com 4%.

Para melhor visualização dos resultados, o gráfico da figura 1 mostra as metodologias e os métodos que mais se destacaram nos documentos selecionados.

Na intenção de obter um conjunto razoável de metodologias e métodos para a análise comparativa, decidiu-se como **primeiro** critério de seleção: escolher as metodologias e métodos mais discutidos na literatura. Assim, a tabela de freqüência de ocorrência foi utilizada como referência para a escolha. Dentre as metodologias candidatas à inserção no conjunto, o interesse da pesquisa se concentrava em metodologias que contemplassem a construção de **ontologias de domínio**, isto é, um tipo de ontologia que descreve um vocabulário relacionado a um domínio genérico, tal como medicina (GUARINO, 1998). Tal classificação ou tipo de ontologia foi escolhido pelo fato de a análise comparativa envolver um tipo de instrumento da ciência da informação que objetiva construir linguagens documentárias (DODEBEL, 2002) acerca de um assunto específico ou domínio de

FIGURA 1
Metodologias e métodos por ordem decrescente de freqüência de ocorrência



Fonte: Silva (2008, p.124)

conhecimento, isto é, por meio de vocabulários controlados. Diante disso, o **segundo** critério de seleção de metodologias foi investigar a ontologia resultante da aplicação de cada metodologia. Tal investigação é descrita na subseção a seguir.

2.2.1.4 Passo 4 – Determinação das metodologias e dos métodos

A investigação da ontologia resultante de cada metodologia foi feita mediante a análise de conteúdo nos documentos representativos, bem como em outros identificados em referências bibliográficas. No caso do projeto **Tove**, verificou-se que a metodologia foi derivada da experiência dos autores no desenvolvimento de ontologias para domínios corporativos (GRUNINGER; FOX, 1995). O método proposto por **Uschold e King** foi baseado na prática da construção da ontologia de alto nível Enterprise (USCHOLD; KING, 1995). Ambos são capazes de construir ontologias de domínios sobre negócios. A metodologia **Methontology** já foi aplicada no desenvolvimento de uma ontologia no domínio da química (GÓMEZ-PÉREZ; FERNANDEZ-LOPES; VICENTE, 1996). O método **Cyc** considera o conhecimento consensual sobre o mundo e pode ser utilizado na criação de ontologias para fundamentar diferentes sistemas inteligentes (REED; LENAT, 2002). O método aplicado no projeto **Kactus** resultou numa ontologia no domínio de redes elétricas (BERNARAS; LARESGOITI; CORERA, 1996). O método baseado na

ontologia **Sensus** foi aplicado no desenvolvimento de uma ontologia no domínio de planejamento de operação militar aérea (VALENTE *et alii*, 1999). A metodologia **On-to-Knowledge** visa a construir ontologias para aplicações de gestão do conhecimento e são altamente dependentes da aplicação (FERNANDEZ; GÓMEZ-PEREZ; CORCHO, 2004). O método **101** já foi aplicado na construção de uma ontologia no domínio de vinhos (NOY; GUINNESS, 2001). E, finalmente, o método **KBSI IDEF5**, que auxilia a criação, a modificação e a manutenção de ontologias (JONES; BENCH-CAPON; VISSER, 1998).

Mediante os critérios estabelecidos, apenas dois objetos, dentre os nove apresentados, não foram selecionados para análise comparativa nesta pesquisa. A metodologia **On-to-Knowledge** depende altamente da aplicação na qual a ontologia será modelada, mostrando-se, portanto, inviável na construção de ontologias de domínio. Além disso, não foram identificadas na pesquisa ontologias resultantes da aplicação do método **KBSI IDEF5**. Assim sendo, as metodologias e os métodos escolhidos para a atividade de análise comparativa foram os seguintes: a) metodologia de **Gruninger e Fox**; b) método de **Ushold e King**; c) metodologia **Methontology**; d) método **Cyc**; e) método **Kactus**; f) método **Sensus**; g) método **101**.

Acreditou-se que, analisando as metodologias mais discutidas na literatura, poder-se-ia obter um resultado razoável no que diz respeito a um estudo comparativo de metodologias para construção de ontologias.

2.2.2 Seleção das metodologias para construção de vocabulários controlados

Na atividade de identificação de metodologias para construção de vocabulários controlados, foi realizada uma investigação na literatura do campo da ciência da informação, conforme explanado na seção 2.1. Na investigação, constatou-se a existência de normas para construção de tesouros, taxonomias, enfim, vocabulários controlados, criadas por entidades com aceitação internacional, propiciando, aos desenvolvedores, subsídios metodológicos para melhor orientá-los na confecção desses instrumentos. Segundo definição da Associação Brasileira de Normas Técnicas, uma *norma* é

um documento estabelecido por consenso e aprovado por um organismo reconhecido que fornece, para uso comum e repetitivo, regras, diretrizes ou características para atividades ou para seus resultados, visando à obtenção de um grau ótimo de ordenação em um dado contexto (ABNT ISO/IEC GUIA 2:1998).

Afora as normas, guias nacionais também têm sido criados para ajudar na elaboração de vocabulários controlados. Algumas dessas normas e guias são descritas a seguir:

- Normas e diretrizes da **Unesco** de 1973: *Guidelines for the establishment and development of monolingual thesauri*.
- Manual de Elaboração de Tesouros Monolíngües de 1990: publicado em Brasília pelo Programa Nacional de Bibliotecas das Instituições de Ensino Superior.
- A versão em português da ISO 2788: Diretrizes para o estabelecimento e desenvolvimento de Tesouros Monolíngües, publicada pelo **Senai/IBICT** em 1993.
- Norma elaborada pela American National Standards Institute National Information Standards Organization (U.S.): *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* de 2005 ou ANSI/NISO Z39.19-2005.

Nessa pesquisa, foi utilizada como referência uma norma que se justificou pelos seguintes critérios: a) norma mais atual (de 2005), aceita e consolidada na comunidade de ciência da informação; b) norma que apresenta abordagem interdisciplinar das teorias advindas da ciência da informação e da terminologia - Teoria da Classificação Facetada; Teoria do Conceito; e Teoria da Terminologia - na construção de vocabulários controlados. Essa norma é a **ANSI/NISO Z39.19-2005**, que foi construída tendo como base as diversas normas americanas e internacionais sobre a criação de tesouros, incluindo a ISO 2788. Deu-se maior ênfase à norma ANSI/NISO Z39.19-2005, por considerá-la mais atual em relação à norma ANSI/NISO Z39.19-2003.

Destaca-se que no período de execução da atividade de coleta e análise dos dados, ficou constatado que a norma ANSI/NISO Z39.19-2005 não seria suficiente para classificar seus conteúdos nas categorias de análise fundamentadas na norma IEEE-1074 (1997), conforme poderá ser conferido na seção 3.8. Tal evidência resultou na seleção do manual disponível no sítio da Biblioteconomia, Informação e Tecnologia da Informação - **BITI** (CAMPOS; GOMES; MOTTA, 2004) sobre elaboração de tesouros. É válido ressaltar que a metodologia explicitada no manual da BITI é consequência de trabalhos da professora Hagar Espanha Gomes e equipe (GOMES *et alii.*, 1990), os quais se mostram como referência para cursos de biblioteconomia e ciência da informação, além de bibliotecas preocupadas

com o desenvolvimento de terminologias adequadas para a representação do conteúdo de seus acervos.

A justificativa da escolha de dois objetos empíricos para análise de elaboração de vocabulários controlados (diferentemente da seleção de seis objetos para construção de ontologias), deve-se ao fato de que, como existem normas, os vocabulários controlados não padecem do mesmo problema de falta de consenso de metodologias para construção de ontologias. Desse modo, acredita-se que a investigação literária no contexto dos vocabulários controlados possa ser realizada de uma maneira menos exaustiva em relação às ontologias.

2.3 Coleta e análise dos dados

Esta seção descreve como foi conduzida a atividade de coleta e análise do material empírico. O primeiro passo desta etapa foi definir as categorias de análise que pudessem subsidiar a análise de conteúdo dos materiais envolvendo as metodologias e os métodos para construção de ontologias e a metodologia e a norma para construção de vocabulários controlados.

Na área de engenharia de *software* existem metodologias já consolidadas, difundidas e bem aceitas na indústria de *software* (PRESSMAN, 2002), como é o caso do *Extreme Programming*, conhecido como XP*, e do *Unified Process* (BOOCH; JACOBSON; RUMBAUGH, 2006), ou Processo Unificado, o qual serve de base para a criação de outros processos de *software* devido a sua extensibilidade. Exemplos de processos de *software* baseados no Processo Unificado são o RUP** (*Rational Unified Process*), a sua extensão EUP*** (*Enterprise Unified Process*) e o *Praxis* (PAULA FILHO, 2003), ou PRocesso para Aplicativos eXtesíveis e Interativos. Entretanto, percebe-se um cenário relativamente diferente na Engenharia ontológica, na qual diversas metodologias têm sido apresentadas e discutidas para construção, reutilização e avaliação de ontologias (GUARINO; WELTY, 2000; FERNÁNDEZ *et alii.*, 1999; JONES; BENCH-CAPON; VISSER, 1998; USCHOLD; GRUNINGER, 1996; GÓMEZ-PEREZ; FERNANDEZ; VICENTE, 1996; GRUNINGER; FOX, 1995), mas apresentam abordagens e características diversas, sendo direcionadas a diferentes propósitos e aplicações, ou seja, não possuem propostas unificadas, sendo que grupos diferentes utilizam diferentes abordagens (FERNÁNDEZ *et al.*, 1999).

Fernández (1999) considera que o processo de desenvolvimento de ontologias deveria ser fundamentado no padrão internacional IEEE-1074, norma para desenvolvimento de *software* advinda da área de engenharia de *software*. O autor ressalta que algumas extensões na norma IEEE-1074 em conjunto com algumas adaptações particulares às ontologias deveriam ser levadas em consideração no processo de construção das mesmas. E justifica a aplicação da norma no processo de construção de ontologias por considerá-las como um componente de produtos de *software**.

Na análise de conteúdo feita nos documentos, descrita na seção 2.2, percebeu-se certa semelhança entre algumas fases de desenvolvimento dos instrumentos (ontologias e vocabulários controlados) e outras advindas do processo de desenvolvimento de *software*. Algumas dessas semelhanças foram identificadas principalmente nas atividades de análise de domínio e nas abordagens técnicas para criação de modelos conceituais. O modelo de ciclo de vida de *software*, proposto na norma IEEE-1074, é considerado um padrão aceitável para análise e comparação das fases de desenvolvimento de ontologias e vocabulários controlados. A justificativa para tal consideração é de que o modelo de ciclo de vida descreve um processo estruturado e metódico para desenvolvimento de produtos (PRESSMAN, 2002), e advém da engenharia de *software*, uma disciplina considerada madura no sentido de possuir metodologias amplamente aceitas, conforme já elucidado. Decidiu-se, portanto, na perspectiva deste trabalho, utilizar como subsídio para definição das categorias de análise o padrão aceito internacionalmente para desenvolvimento de *software*, a norma IEEE-1074 (1997).

As categorias de análise de conteúdo do material empírico foram definidas a partir de princípios elucidados por Bardin (1977), que advoga o uso de categorias para procedimentos de análise qualitativa. Segundo o autor, a escolha de categorias pode envolver vários critérios: i) semântico (temas); ii) sintático (verbos, adjetivos, pronomes); iii) léxico (juntar pelo sentido das palavras, agrupar os sinônimos, os antônimos); iv) expressivo (agrupar as perturbações da linguagem, da escrita). Neste trabalho, o critério de escolha das categorias foi o semântico, ou seja, de acordo com a norma IEEE-1074 (1997), com a literatura das áreas de ontologias e

* <http://www.extremeprogramming.org/>

** <http://www.ambysoft.com/unifiedprocess/rupIntroduction.html>

*** <http://www.enterpriseunifiedprocess.com/>

* “[...] ontologies are part (sometimes only potentially) of *software* products. [...] ontologies should be developed according to the standards proposed for *software* generally, which should be adapted to the special characteristics of ontologies.”

vocabulários controlados. Tais categorias foram, então, adaptadas diante dos processos extraídos da norma para construção de *software*, características particulares às ontologias (formalização e integração), e característica peculiar aos vocabulários controlados (formas de apresentação). São elas: i) **gerenciamento do projeto**; ii) **pré-desenvolvimento**; iii) **especificação de requisitos**; iv) **modelagem conceitual**; v) **formalização**; vi) **implementação**; vii) **forma de apresentação**; viii) **manutenção**; ix) **integração**; x) **avaliação**; xi) **documentação**. A seguir, cada categoria é fundamentada: i) de acordo com a norma IEEE-1074 (1997); ii) de acordo com princípios metodológicos para construção de ontologias (FERNANDEZ; GOMEZ-PEREZ; JURISTO, 1997; USCHOLD; GRUNINGER, 1996); iii) de acordo com particularidades na construção de vocabulários controlados (CAMPOS; GOMES; MOTTA, 2004).

- **Gerenciamento do projeto**: ocorrem atividades relacionadas ao início de um projeto, como criação do processo e ciclo de vida, ao planejamento da gestão de um projeto e ao monitoramento e controle do projeto em todo o seu ciclo de vida.

- **Pré-desenvolvimento**: consiste em analisar idéias ou conceitos de um sistema e, em função de problemas observados no ambiente, alocar requisitos para o sistema antes do início de desenvolvimento do produto. A fase inclui atividades de estudo de viabilidade e análise de requisitos do sistema.

- **Especificação de requisitos**: abrange as restrições ou regras que o produto deve cumprir em função da definição das necessidades do requisitante. Os requisitos devem servir como documento inicial para a realização das tarefas de modelagem e prototipação, e tal processo é geralmente iterativo.

- **Modelagem conceitual**: objetiva desenvolver uma representação bem organizada e coerente do sistema que satisfaça os requisitos de produto especificados nas atividades de requisitos.

- **Formalização**: consiste em transformar o modelo conceitual da ontologia (ou conceitualização) em um modelo formal a fim de definir de forma precisa o seu significado. O profissional envolvido na construção da ontologia concentra-se no processo de modelagem computacional do problema, usando, por exemplo, a lógica de primeira ordem e suas extensões (sistemas de representação baseados em frames, redes semânticas,

lógica descritiva etc.). As técnicas empregadas nessa fase são oriundas da área de inteligência artificial.

- **Implementação**: resulta na transformação da representação do projeto da arquitetura do *software* em uma linguagem de programação. No caso específico das ontologias, a implementação consiste em mapear o modelo formal em uma linguagem que se adeque às demandas como Web Ontology Language - OWL (DEAN *et alii.*, 2003), por exemplo.

- **Forma de apresentação**: no contexto dos vocabulários controlados, após o modelo conceitual ter sido concebido, a próxima atividade seria direcionada à apresentação do vocabulário, que implicará diretamente a satisfação dos usuários em relação ao manuseio do instrumento. A forma de apresentação inclui como representar os relacionamentos dos termos, formato de distribuição, tipos de exibição, entre outras.

- **Manutenção**: considerada uma etapa pós-desenvolvimento, que consiste em identificar problemas e melhorias nos produtos, podendo resultar em novas versões.

- **Integração**: esta fase considera a reutilização de conceitos existentes em outras ontologias. Segundo Fernandez, Gomez-Perez e Juristo (1997), uma alternativa para facilitar a construção de ontologias é buscar integrar a ontologia em questão a ontologias existentes. A proposta é examinar a conceitualização de meta-ontologias (chamadas ontologias de alto nível) e selecionar (parcialmente ou por completo) aquelas que melhor se ajustarem ao modelo que está sendo construído. No processo de integração, as atividades podem ser realizadas durante a fase de modelagem conceitual e implementação da ontologia, sendo considerada, portanto, um processo integral. Ressalta-se que o fato de considerar a busca por ontologias existentes não implica, necessariamente, integração. Entretanto, o nome original para o processo (com garantia literária) é integração.

- **Avaliação**: suas atividades são executadas ao mesmo tempo com atividades dos processos orientados ao desenvolvimento do produto, como, por exemplo, condução de revisões e auditorias nos processos, desenvolvimento de procedimento de testes, execução de testes e avaliação de resultados.

- **Documentação**: desenvolvimento e distribuição de documentação em cada fase para desenvolvedores e

usuários envolvidos nos processos, a fim de fornecer, em tempo hábil, informações sobre o produto.

Definidas as categorias de análise, o próximo passo foi a elaboração dos instrumentos para coleta e registro dos dados. Para cada metodologia, método e norma envolvida foi elaborada uma tabela contendo um espaço dedicado a cada categoria de análise. Dessa forma, tornou-se possível colher o conteúdo nos documentos envolvidos a partir das categorias e registrá-lo formalmente em cada tabela. Se uma categoria não fosse pertinente a alguma metodologia ou método, a coluna era preenchida com o valor “Ausente”. A coleta dos dados passou pelas atividades de compreensão, análise e síntese de conteúdo, realizadas a partir do conhecimento adquirido na pesquisa. Tais atividades foram necessárias para a realização da etapa seguinte, que consistia em analisar comparativamente as categorias para cada metodologia, método e norma; sob a orientação dos parâmetros estabelecidos para esta análise. Tal etapa é descrita na seção seguinte.

2.4 Análise comparativa das metodologias, dos métodos e da norma

Feita a coleta e registro dos dados, o próximo passo foi o desenvolvimento de um quadro comparativo contendo as fases de desenvolvimento apresentadas nos objetos empíricos analisados. Tal escopo foi concebido por meio de uma estrutura matricial representando em suas colunas os objetos investigados e em suas linhas cada fase do ciclo de vida. A partir do tratamento e organização do conteúdo em tal estrutura, foi possível analisar comparativamente cada metodologia, método e norma e tirar conclusões teóricas e empíricas acerca do processo de construção de ontologias e vocabulários controlados.

3. APRESENTAÇÃO DAS METODOLOGIAS, DOS MÉTODOS E DA NORMA

Esta seção apresenta a coleta de dados proposta nos passos metodológicos descritos na seção 2. Os dados coletados foram obtidos a partir do material empírico selecionado na pesquisa, que contemplou as metodologias e os métodos para construção de ontologias, bem como a metodologia e a norma para construção de vocabulários controlados.

Ressalta-se que, em virtude do limite de extensão do presente artigo, apresentam-se os objetos investigados de modo simplificado. Em Silva (2008, p.132), encontra-se explicitada, em detalhe, a análise de tais objetos. Desse

modo, a seção 3.1 apresenta a ontologia Cyc e o método utilizado para seu desenvolvimento. A seção 3.2 apresenta considerações sobre o projeto Tove e os procedimentos da metodologia de Gruninger e Fox. A seção 3.3 apresenta o projeto da *Enterprise Ontology* e abrange os procedimentos do método de Uschold e King. A seção 3.4 apresenta o método Kactus. A seção 3.5 apresenta os métodos e as técnicas da Methontology. A seção 3.6 apresenta a ontologia SENSUS e o método baseado em tal ontologia, o método Sensus. A seção 3.7 apresenta o método I01. E, finalmente, a seção 3.8 tece considerações sobre a norma para construção de vocabulários controlados e apresenta os procedimentos metodológicos envolvidos no manual da BITI para construção de tesouros.

3.1 Método Cyc

Nos anos de 1980, a Microelectronics and Computer Technology (MCC) deu início a criação da Cyc, uma ampla base de conhecimento que considera o conhecimento consensual sobre o mundo, incluindo regras e heurísticas para dedução sobre objetos e eventos do cotidiano (REED; LENAT, 2002). A linguagem de representação da Cyc é a CycL, considerada híbrida por combinar *frames* com cálculos de predicado. Tal linguagem possui uma máquina de inferência que permite herança múltipla, classificação automática, manutenção de *links* inversos, verificação de restrições, busca ordenada, detecção de contradição e módulo de resolução.

A base de conhecimento Cyc foi desenvolvida em 1990 por Douglas Lenat e Ramanathan Guha (FERNANDEZ; GOMEZ-PEREZ; CORCHO, 2004), em que três processos foram considerados em tal desenvolvimento, a saber: i) extração do conhecimento de senso comum; iii) extração auxiliada por computador; iii) extração gerenciada por computador. No primeiro processo, o conhecimento requerido para a ontologia foi obtido de forma manual em diferentes fontes como artigos, livros e jornais. O segundo processo foi conduzido de maneira automática, isto é, com uso de ferramentas computacionais de processamento de linguagem natural e aprendizado de máquina capazes de usar conhecimento de senso comum suficiente para investigar e descobrir novos conhecimentos. E, finalmente, o terceiro processo foi conduzido por um número maior de ferramentas no sentido de gerenciar a extração de conhecimento de senso comum (partes consideradas difíceis de serem interpretadas nas fontes de conhecimento envolvidas) na base Cyc.

3.2 Metodologia de Gruninger e Fox

A metodologia foi proposta por Michael Gruninger e Mark Fox em 1995 (GRUNINGER; FOX, 1995), tendo como base para o seu desenvolvimento a experiência obtida no projeto Toronto Virtual Enterprise – conhecido como projeto Tove (FOX, 1992), cujos princípios teóricos e metodológicos encontram-se na inteligência artificial.

O objetivo do projeto Tove é criar um modelo de senso comum sobre empresas, isto é, um conhecimento compartilhado sobre o negócio que conduza a deduções de respostas sobre questões acerca do domínio (FOX, 1992). Para tal, ontologias são criadas no sentido de especificar modelos para organizações públicas e privadas, levando em consideração as seguintes características: a) capacidade de fornecer uma terminologia compartilhada para organizações, que possa ser compreendida e utilizada por cada aplicação, isto é, para cada tipo de negócio; b) definição da semântica de cada termo por meio de uma teoria lógica; c) implementação da semântica em um conjunto de axiomas que permita à ontologia deduzir de forma automática respostas às questões comuns no escopo das organizações; d) definição de uma simbologia para representar graficamente termos ou conceitos (GRUNINGER; FOX, 1996).

A metodologia de Gruninger e Fox foi usada no Enterprise Integration Laboratory (Laboratório de Integração de Empresas) da University of Toronto (Universidade de Toronto) para o projeto e avaliação de ontologias integradas, incluindo propostas de construção de novas ontologias e extensões de ontologias já existentes. Os seguintes procedimentos foram propostos na metodologia: i) elaboração de cenários de motivação, que objetivam identificar problemas no ambiente atual; ii) especificação de questões de competência informal, que objetivam especificar em linguagem natural os requisitos que a ontologia deverá ser capaz de atender; iii) concepção da terminologia formal, em que, mediante declarações em lógica de primeira ordem, os conceitos e suas propriedades são organizados em uma taxonomia; iv) especificação de questões de competência formal, em que problemas são definidos de modo consistente perante os axiomas na ontologia; v) especificação de axiomas formais, que restringem a interpretação dos termos envolvidos nas questões de competência formal; vi) verificação de teoremas completos, que determinam as condições sobre as quais as soluções das questões são completas.

3.3 Método de Uschold e King

O método foi proposto inicialmente por Mike Uschold e Martin King em 1995 (USCHOLD e KING, 1995) e estendido em 1996 por Mike Uschold e Michael Gruninger (USCHOLD e GRUNINGER, 1996) na experiência de desenvolvimento da *Enterprise Ontology*. Tal ontologia foi desenvolvida como parte do projeto *Enterprise* por meio do Instituto de Aplicações em Inteligência Artificial da Universidade de Edinburgh e parceiros como IBM, Unilever e outros.

Uschold e King (1995) consideram os seguintes estágios como sendo necessários a uma metodologia abrangente: i) identificação do propósito da ontologia, que objetiva identificar a necessidade de construção, o grau de formalismo (desde o informal com uso de linguagem natural até o rigorosamente formal com uso de declarações lógicas) e as classes de usuários da ontologia, incluindo desenvolvedores, mantenedores e usuários das aplicações; ii) construção da ontologia, que se divide em: a) captura ou concepção da conceitualização da ontologia; b) codificação ou implementação através de uma linguagem de representação de ontologias, e c) integração com ontologias já existentes; iii) avaliação da ontologia através dos requisitos especificados; iv) documentação acerca das pretensões da ontologia e das primitivas usadas para expressar as definições na ontologia.

3.4 Método Kactus

A ênfase do projeto europeu *Esprit Kactus* está na organização de bases de conhecimento que podem ser compartilhadas e reusadas em diferentes sistemas baseados em conhecimento. Para tal, utiliza ontologias de domínio para organizar o conhecimento independente da aplicação de *software* que será construída.

Baseando-se no projeto Kactus, Amaya Bernaras e colegas (BERNARAS; LARESGOITI; CORERA, 1996) investigaram a viabilidade da reutilização do conhecimento em sistemas de complexidade técnica, como o domínio de redes elétricas, e o papel das ontologias como suporte a tais sistemas. Tal investigação resultou em um método de construção de ontologias, cujos processos envolvidos estariam condicionados ao desenvolvimento da aplicação, ou seja, toda vez que uma aplicação fosse construída, a ontologia, que representa o conhecimento necessário para a aplicação, seria refinada. Tais processos seriam os seguintes: i)

desenvolvimento de uma lista de necessidades ou requisitos que precisam ser atendidos pela aplicação; ii) identificação de termos relevantes para o domínio da aplicação a partir de tais requisitos, construindo, assim, um modelo preliminar; iii) refinar e estruturar a ontologia a fim de obter um modelo definitivo; iv) buscar por ontologias já desenvolvidas por outras aplicações no sentido de sua reutilização. As ontologias reutilizadas demandariam refinamento e extensão para serem usadas na nova aplicação.

3.5 Metodologia Methontology

A metodologia para construção de ontologias Methontology foi desenvolvida no Laboratório de Inteligência Artificial da Universidade Politécnica de Madri entre 1996 e 1997 pelo grupo de pesquisadores Mariano Fernández, Asunción Gómez-Pérez, Antônio J. de Vicente e Natalia Juristo (GOMEZ-PEREZ; FERNANDEZ; VICENTE, 1996; FERNANDEZ; GOMEZ-PEREZ; JURISTO, 1997).

A Methontology contempla um conjunto de estágios de desenvolvimento (especificação, conceitualização, formalização, integração, implementação e manutenção), um ciclo de vida baseado em evolução de protótipos (PRESSMAN, 2002) e técnicas para realizar as atividades de planejamento, desenvolvimento e suporte. A atividade de planejamento inclui um escalonamento das tarefas e controle sobre as mesmas, no sentido de alcançar a qualidade devida. As atividades de suporte contemplam aquisição de conhecimento, documentação e avaliação, e ocorrem durante todo o ciclo de vida da ontologia. Os estágios iniciais de desenvolvimento (especificação e conceitualização) implicam um grande esforço dentro das atividades de suporte, como a aquisição de conhecimento e a avaliação. Várias são as razões: a) a maior parte do conhecimento é adquirida no início do processo de construção da ontologia; b) deve-se avaliar corretamente o modelo conceitual para evitar futuros erros no ciclo de vida da ontologia. Por fim, a documentação detalhada deve ser produzida após cada estágio previsto no ciclo de vida.

3.6 Método Sensus

A ontologia SENSUS foi desenvolvida pelo grupo de linguagem natural *Information Sciences Institute* - ISI com o propósito de ser usada para fins de processamento de linguagem natural. A ontologia SENSUS possui aproximadamente 70 mil conceitos organizados em uma hierarquia, de acordo com seu nível de abstração que

vai de médio a alto. No entanto, sua estrutura não contempla termos específicos de um domínio (SWARTOUT *et alii*, 1996). Para tal, os termos de domínios específicos são ligados à ampla ontologia SENSUS, de forma a construir ontologias para domínios particulares.

O método Sensus, baseado na ontologia SENSUS, propõe alguns processos para estabelecer as ligações entre os termos específicos e os termos da ontologia de alto nível (SWARTOUT *et alii*, 1996). O resultado de tal processo é uma estrutura de uma nova ontologia, que é generalizada automaticamente através de uma ferramenta denominada OntoSaurus (SWARTOUT *et alii*, 1996; FERNANDEZ; GOMEZ-PEREZ; CORCHO, 2004). De acordo com o método, os processos envolvidos na construção da ontologia de um domínio específico seriam: i) identificar termos-chave do domínio; ii) ligar manualmente os termos-chave à ontologia SENSUS; iii) adicionar caminhos até o conceito de hierarquia superior da Sensus; iv) adicionar novos termos para o domínio; v) adicionar subárvores completas.

3.7 Método 101

O método 101 foi concebido por Natalya F. Noy e Deborah L. McGuinness (NOY; MCGUINNESS, 2001) a partir da experiência no desenvolvimento da ontologia de vinhos e alimentos, utilizando o ambiente de edição de ontologias *Protégé-2000* (HORRIDGE *et alii*, 2004).

O método 101 propõe basicamente quatro atividades para o desenvolvimento de uma ontologia: i) definir classes na ontologia; ii) organizar as classes em uma taxonomia; iii) definir *slots* (ou propriedades) para as classes e descrever seus valores permitidos (denominado *facetas*); iv) adicionar valores de *slots* para as instâncias. Tais atividades implicam decisões de modelagem, dentre as quais o método busca enfatizar, além de se encontrarem dentro de um processo iterativo de um ciclo de vida de ontologia.

3.8 Metodologia e Norma para construção de vocabulários controlados

A norma padrão proposta pela organização norte-americana *National Information Standards Organization* (ANSI, 2005) propõe as linhas gerais para construção, formatação e manutenção de vocabulários controlados monolíngües. Em relação às regras de construção, o padrão ANSI/NISO Z39.19-2005 permite construir diversos tipos de vocabulário controlado, incluindo

tesauros, taxonomias, listas e anel de sinônimos, em uma ordem conhecida e estruturada com o intuito de disponibilizar claramente os relacionamentos de equivalência, associativos e hierárquicos, quando aplicáveis a cada tipo (ver apêndice B da norma, página 135). Essa flexibilidade é importante, pois permite melhor adaptação do instrumento às necessidades dos ambientes informacionais, como a Web, por exemplo.

Uma inovação da ANSI/NISO Z39.19-2005 em relação às normas anteriores é a inclusão da análise facetada, cuja base concentra-se no trabalho de Ranganathan (RANGANATHAN, 1967) e nos refinamentos feitos pelo *Classification Research Group* - CRG (CAMPOS; GOMES; MOTTA, 2004), na Inglaterra, para a elaboração de tesauros em áreas específicas do conhecimento. Outra questão importante é a orientação sobre a interoperabilidade entre vocabulários controlados.

E, finalmente, o padrão ANSI/NISO Z39.19-2005 recomenda uma série de tratamentos em relação ao termo frente à citação de atividades envolvidas no processo de construção. Já o manual disponível no sítio da Biblioteconomia, Informação e Tecnologia da Informação - BITI (CAMPOS; GOMES; MOTTA, 2004) sobre elaboração de um tipo de vocabulário controlado, os tesauros, explicita os procedimentos necessários à sua construção. Desse modo, o manual da BITI mostra-se mais eficiente do que a norma ANSI/NISO Z39.19-2005 no que diz respeito a princípios metodológicos definidos para construção de vocabulários controlados.

Os procedimentos propostos no manual resumem-se em: i) planejamento, que consiste em delimitar a área do assunto a ser coberto pelo vocabulário especializado, definir o público-alvo no qual o instrumento se destina e elaborar o planejamento de futuras manutenções que se fizerem necessárias; ii) levantamento do vocabulário, que consiste em selecionar os termos representativos do assunto e defini-los de acordo com a natureza do assunto; iii) organização dos conceitos, que implica agrupar os termos de mesma natureza em categorias ou facetas, a fim de permitir maior compreensão do conceito e da organização das relações entre os conceitos; iv) apresentação final, que inclui os tipos de exibição (listagens alfabéticas simples ou visualizações gráficas) e o formato que pode ser impresso ou eletrônico; e v) critérios para avaliação, que determina a adoção do vocabulário controlado.

Apresentados os objetos investigados, cabe a realização da classificação de conteúdo em cada categoria

denominada, a partir desse ponto, fase do processo de construção. O quadro 1, a seguir, sintetiza a análise das metodologias, dos métodos e da norma investigados.

4. RESULTADOS DA ANÁLISE COMPARATIVA

Para a execução da análise comparativa, foi necessário um entendimento dos procedimentos metodológicos dos objetos investigados, apresentados na seção 3. Tal entendimento tornou-se possível a partir da análise de conteúdo feita nos materiais empíricos por meio das categorias de análise elucidadas na seção 2.

A partir das informações dispostas no quadro 1 da seção 3, chegou-se a algumas considerações sobre as metodologias e os métodos para construção de ontologias analisados e sobre a metodologia e a norma para construção de vocabulários controlados analisadas. Tais considerações são enumeradas a seguir:

- Existe uma variedade de estratégias para desenvolvimento de ontologias, comprovando a hipótese de que grupos diferentes apresentam abordagens e características diversas, sendo direcionadas a diferentes propósitos e aplicações (FERNÁNDEZ *et al.*, 1999).

- No contexto das ontologias, algumas abordagens seguem um modelo de ciclo de vida, outras não. Nesse quesito, a que mais se destaca é a Methontology por ser praticamente completa em relação a um ciclo de desenvolvimento, não propondo apenas a fase de pré-desenvolvimento. Tal constatação pode ser conferida no quadro 1.

- Em relação a detalhes das atividades e dos procedimentos para sua condução, algumas metodologias e métodos mostram-se superficiais na elucidação dos passos para construção de ontologias. É o caso dos métodos Cyc, Kactus e Sensus, os quais parecem considerar que o ontologista já domina o assunto sobre construção de ontologias e não necessita de detalhes acerca de atividades e procedimentos envolvidos. Já a Methontology se destaca por fornecer, na maioria das vezes, detalhes de como proceder na condução de uma dada atividade.

- Algumas abordagens dão mais ênfase em atividades de desenvolvimento, especialmente a implementação da ontologia (método Cyc e método 101), desconsiderando aspectos importantes relacionados a gerenciamento do projeto, a estudo de viabilidade, à manutenção e à avaliação de ontologias.

QUADRO 1
Quadro sinóptico da análise comparativa das metodologias, dos métodos e da norma

Fases do ciclo de vida	Cyc	Gruninger e Fox	Usehold e King	Kactus	Methontology	Sensus	Método 101	Z39.19-2005	Manual da BITI	
Processos orientados	Gerenciamento do Projeto	Ausente	Ausente	Ausente	Escalonamento de atividades	Ausente	Ausente	Recomenda gerenciamento de projeto	Atividades de planejamento do projeto	
	Pré-desenvolvimento	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	Recomenda estudo de viabilidade	Ausente	
		Especificação de requisitos	Extração manual do conhecimento requerido	Determinar o propósito da ontologia	Lista de requisitos	Definição do escopo da ontologia	Ausente	Questões de competência informal	Ausente	Delimitação da área; definição do público-alvo
		Modelagem conceitual	Ausente	Construção do vocabulário consensual	Reuso de conhecimento de alto nível	Métodos para construção da conceitualização da ontologia	Reuso de conhecimento de ontologias de alto nível	Definição de classes, <i>s/ors</i> , facetas e instâncias	Recomendações para o tratamento da terminologia	Análise do domínio segundo as teorias do conceito e da classificação
	Pós-desenvolvimento	Formalização	Ausente	Ausente	Ausente	Através de ferramentas de edição de ontologias	Através de rede semântica	Através de <i>frames</i>	Ausente	Ausente
		Implementação	Codificação e extração do conhecimento	Prolog e Ontologia	Ausente	Crítérios para escolha de ferramentas	Ferramenta OntoSaurus	Ferramenta <i>Protégé</i>	Ausente	Ausente
		Forma de Apresentação	Ausente	Ausente	Ausente	Ausente	Ausente	Ausente	Representação dos relacionamentos dos termos	Listas alfabéticas e visualizações gráficas
	Processos integrais	Integração	Ausente	Ausente	Ausente	Verificações nas representações intermediárias	Ausente	Ausente	Procedimentos relacionados à manipulação de termos	Proposta numa planilha específica
		Avaliação	Integração a microteorias da base Cyc	Integração com ontologias já existentes	Busca por ontologias já desenvolvidas	Documento de integração com meta-ontologias	Integração manual a ontologia Sensus	Considera reutilização de ontologias	Ausente	Ausente
		Documentação	Ausente	Teoremas completos	Questões de competência	Verificação e validação	Ausente	Ausente	Testes de usabilidade	Crítérios de adoção
		Ausente	Primitivas e pretensões	Ausente	Documentação em cada fase	Ausente	Via <i>Protégé</i>	Recomendações sobre o conteúdo	Registro dos dados em planilhas	

Fonte: Adaptado de Silva (2008).

- O manual da BITI apresentou um ciclo de vida praticamente completo segundo a norma IEEE-1074 (1997), conforme pode ser visto no quadro 1. As fases de formalização, implementação e integração não foram consideradas por não pertencerem ao propósito dos vocabulários controlados. Desse modo, pode-se conferir a maturidade da metodologia ante um modelo de ciclo de vida.

- Finalmente, é válido ressaltar que, de acordo com os resultados obtidos em Silva (2008), há indícios de que a metodologia para construção de tesouros apresenta aspectos relevantes que se destacam e podem contribuir com a construção de ontologias. Tais aspectos estariam centrados em princípios teóricos e metodológicos advindos da teoria da classificação (RANGANATHAN, 1967) e da teoria do conceito (DAHLBERG, 1978) nas atividades de identificação, definição e organização de conceitos. Tais princípios seriam pertinentes em metodologias para construção de ontologias, visto que ambos os instrumentos, ontologias e tesouros, representam relacionamentos semânticos e conceituais. Entretanto, a verificação da relevância de tais aspectos diante de outros observados será objeto de trabalho futuro.

5. CONCLUSÕES

Este artigo possibilitou elucidar as metodologias e os métodos para construção de ontologias mais representativos na literatura, bem como apontar similaridades entre padrões de construção de *software* (norma IEEE-1074) e princípios metodológicos empregados na elaboração de ontologias e vocabulários controlados. Tal similaridade ficou evidente na análise das metodologias, métodos e norma investigados, apresentada no quadro sinóptico exibido na seção 3.

A metodologia exposta no manual da BITI para construção de tesouros mostrou-se madura no que diz respeito a um modelo de ciclo de vida, visto que os processos para construção se enquadraram na maioria das categorias de análise advindas da norma IEEE-1074 (1997). Apesar de as recomendações da norma ANSI/NISO Z39.19-2005 terem sido classificadas nas categorias de análise, tal norma não objetiva propor um ciclo de atividades para a construção de vocabulários controlados, e sim recomendar uma série de tratamentos em relação ao termo em atividades que se enquadram no processo de construção. No caso das metodologias para construção de ontologias, a Methontology foi a que mais se destacou no quesito maturidade ante a norma IEEE-1074 (1997),

tendo em vista que apenas a categoria pré-desenvolvimento não foi considerada em seu processo de desenvolvimento.

O artigo também comprovou alguns problemas relacionados à falta de um padrão para construção de ontologias e na falta de explicações sistemáticas de como, onde e sob quais limites podem ser utilizadas as abordagens teóricas dentro do processo de elaboração. Tal fato foi constatado na análise das metodologias e dos métodos para construção de ontologias investigados na pesquisa, que, na maioria dos casos, mostraram-se pouco eficientes na exposição clara dos procedimentos de construção. Desse modo, a solução para tais problemas estaria centrada em uma proposta metodológica fundamentada em princípios teóricos e metodológicos que dessem sustentação científica no processo de construção de ontologias. Finalmente, a apresentação de uma análise comparativa como um passo preliminar pode ser usada para a definição de padrões metodológicos para construção de ontologias.

Artigo submetido em 27/10/2008 e aceito em 26/12/2008.

REFERÊNCIAS

- ALMEIDA, M. B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003.
- AMERICAN NATIONAL STANDARDS ORGANIZATION - ANSI. *ANSI/NISO Z 39.19: 2005: guidelines for the construction, format, and management of monolingual controlled vocabularies*. Bethesda: NISO Press, 2005. 176 p. Disponível em: <<http://www.niso.org/standards/resources/Z39-19-005.pdf?CFID=2436046&CFTOKEN=81298864>>. Acesso em: 27 jun. 2007.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *ABNT ISO/IEC guia 2:1998: guia 2. normalização e atividades relacionadas: vocabulário geral*. Rio de Janeiro, 1998. 21 p.
- BARDIN, Laurence. *Análise de conteúdo*. Lisboa: Edições 70, 1977.
- BERNARAS, A.; LARESGOITI, I.; CORERA, J. Building and reusing ontologies for electrical network applications. In: THE EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, ECAI, 1996. *Proceedings...* 1996. p. 298-302.
- BERNERS-LEE, T; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, v. 284, n. 5, p. 34-43, May 2001.
- BOOCH, G. JACOBSON, I. RUMBAUGH, J. *UML guia do usuário*. Rio de Janeiro: Campus, 2006
- BORST, W.N. *Construction of engineering ontologies for knowledge sharing and reuse*. 1997. Tese (Doutorado). Disponível em: <<http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>>. Acesso em: 03 out. 2005.
- BREITMAN, Karin. *Web semântica: a internet do futuro*. Rio de Janeiro: LTC, 2005.

- CAMPOS, M. L. A.; GOMES, H. E.; MOTTA, D. F. *Tutorial de tesouro*. 2004. Disponível em: <<http://www.conexaorio.com/bitit/tesouro>>. Acesso em: 30 abr. 2008.
- CARDOSO, J. The semantic web vision: where are we?. *IEEE Intelligent Systems*, p. 22-26, Sept./Oct. 2007.
- CORAZZON, R. What is ontology?: definitions by leading philosophers. In: ONTOLOGY: a resource guide for philosophers. 2008. Disponível em <http://www.formalontology.it/section_4.htm>. Acesso em: 02 abr. 2008.
- DAHLBERG, Ingetraut. Teoria do conceito. *Ciência da Informação*, Rio de Janeiro, v. 7, n. 2, p. 101-107, jul./dez. 1978.
- DEAN, M. et al. *OWL web ontology language 1.0 reference*. 2003. W3C Working Draft 21.
- DODEBEI, Vera Lucia Doyle. *Tesouro: linguagem de representação da memória documentária*. Rio de Janeiro: Interciência, 2002.
- FERNÁNDEZ, M. et al. Building a chemical ontology using methontology and the ontology design environment. *Intelligent Systems*, v. 14, n. 1, p. 37-46, Jan./Feb.1999.
- _____; GOMEZ-PEREZ, A.; JURISTO, H. *Methontology: from ontological art towards ontological engineering*. 1997. Disponível em: <<http://citeseer.ist.psu.edu/context/544607/0/>>. Acesso em: 20 ago. 2007.
- _____; CORCHO, O. Methodologies and methods for building ontologies. In: _____. *Ontological engineering*. London: Springer, 2004. p. 107-153.
- FOX, M. S. *The TOVE Project: towards a common-sense model of the enterprise*. 1992. Disponível em: <<http://www.eil.utoronto.ca/enterprise-modelling/papers/index.html>>. Acesso em: 20 nov. 2007.
- GILCHRIST, Alan. Thesauri, taxonomies and ontologies: an etymological note. *Journal of Documentation*, v. 59, n. 1, p. 7-18, 2003.
- GOMES, H. E. et al. *Manual de elaboração de tesouros monolíngües*. Brasília: CNPq/PNBU, 1990. 78 p.
- GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. J. Towards a method to conceptualize domain ontologies. In: ECAI WORKSHOP ON ONTOLOGICAL ENGINEERING, 1996, Budapest. *Proceedings...* Disponível em: <<http://citeseer.ist.psu.edu/483876.html>>. Acesso em: 12 nov. 2007.
- GRUBER, T. *A translation approach to portable ontology specifications*. 1993a. Disponível em: <ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-92-71.ps.gz>. Acesso em: 03 ago. 2006.
- _____. *What is an ontology?*. 1993. Disponível em: <<http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 03 ago. 2006.
- GRUNINGER, M.; FOX, M. S. *The logic of enterprise modelling*. 1996. Disponível em: <<http://citeseer.ist.psu.edu/523974.html>>. Acesso em: 10 nov. 2007.
- _____. *Methodology for the design and evaluation of ontologies*. 1995. Disponível em: <<http://citeseer.ist.psu.edu/grninger95methodology.html>>. Acesso em: 10 nov. 2007.
- GUARINO, N. *Formal ontology in information systems*. 1998. Disponível em: <<http://citeseer.ist.psu.edu/guarino98formal.html>>. Acesso em: 3 set. 2007.
- _____; WELTY, C. *Ontological analysis of taxonomic relationships*. 2000. Disponível em: <<http://citeseer.ist.psu.edu/guarino00ontological.html>>. Acesso em: 12 set. 2007.
- HORRIDGE, M. et al. *A practical guide to building OWL ontologies using Teh Protégé-OWL Plugin and CO-ODE Tools*. 1. ed. [S. l.]: The University of Manchester and Stanford University, 2004. Disponível em: <<http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>>. Acesso em: 7 fev. 2007.
- Institute of Electrical and Electronics Engineers: IEEE standard 1074 standard for developing software life cycle processes*. 1997. Disponível em: <<http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel4/5984/16018/00741936.pdf?temp=x>>. Acesso em: 20 jul. 2007.
- JONES, D.; BENCH-CAPON, T.; VISSER, P. *Methodologies for ontology development*. 1998. Disponível em: <<http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-ontodev.pdf>>. Acesso em: 20 nov. 2007.
- LAWRENCE, Steve S.; BOLLACKER K.; LEE GILES, C. Digital libraries and autonomous citation indexing. *IEEE Computer*, v. 32, n. 6, p. 67-71, 1999.
- NOY, F. N.; GUINNESS, D. L. *Ontology development 101: a guide to create your first ontology*. 2001. Disponível em: <<http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.doc>>. Acesso em: 3 out. 2006.
- PAULA FILHO, Wilson de Pádua. *Engenharia de software: fundamentos, métodos e padrões*. 2. ed. Rio de Janeiro: LTC, 2003. 602 p.
- PRESSMAN, Roger S. *Engenharia de Software*. 5. ed. Rio de Janeiro: McGraw-Hill, 2002. 843 p.
- RANGANATHAN, S. R. *Prolegomena to library classification*. Bombay: Asia Publishing House, 1967. 640 p.
- REED, S. L.; LENAT, D. B. *Mapping ontologies into cyc*. 2002. Disponível em: <http://www.cyc.com/doc/white_papers/mapping-ontologies-into-cyc_v31.pdf>. Acesso em: 20 abr. 2008.
- SILVA, Daniela Lucas da. *Uma proposta metodológica para construção de ontologias: uma perspectiva interdisciplinar entre as ciências da informação e da computação*. 2008. 286 f. Dissertação (Mestrado em Ciência da Informação)- Escola da Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2008.
- SOERGEL, Dagobert. *Functions of a thesaurus: classification: ontological knowledge base*. [S. l.]: College of Library and Information Services, University of Maryland, 1997.
- SOUZA, Renato Rocha. *Uma proposta de metodologia para escolha automática de descritores utilizando sintagmas nominais*. 202 p. 2005. Tese (Doutorado em Ciência da Informação)- Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.
- SWARTOUT, B. et al. *Toward distributed use of large-scale ontologies*. 1996. Disponível em: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html>. Acesso em: 10 abr. 2008.
- USCHOLD, M.; KING, M. *Towards a methodology for building ontologies*. 1995. Disponível em: <<http://citeseer.ist.psu.edu/uschold95toward.html>> Acesso em: 10 nov. 2007.
- _____; GRUNINGER, M. Ontologies: principles, methods and applications. *Knowledge Engineering Review*, v. 11, n. 2, 1996.
- VALENTE, A. et al. Building and (re)using an ontology of air campaign planning. *IEEE Intelligent Systems*, Jan./Feb. 1999. Disponível em: <<http://portal.acm.org/citation.cfm?id=630305.630437&coll=GUIDE&dl=GUIDE>>. Acesso em: 12 abr. 2008.
- VICKERY, B. C. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 277-286, 1997.