

Uma comparação de metodologias para construção de ontologias e vocabulários controlados¹

Daniela Lucas da Silva (PPGCI/UFMG)
Renato Rocha Souza (PPGCI/UFMG)
Maurício Barcellos Almeida (PPGCI/UFMG)

Resumo: O artigo se propõe a apresentar um estudo analítico sobre metodologias e métodos para construção de ontologias e vocabulários controlados através da análise da literatura sobre metodologias para construção de ontologias e de normas internacionais para construção de *software* e de vocabulários controlados. Através 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.

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.

¹ Comunicação oral apresentada ao GT-02 - Organização e Representação do Conhecimento.

1. Introdução

A organização da informação tornou-se um processo fundamental na medida em que vêm 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 num 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, tais como: a) a Web Semântica que pretende criar metodologias, tecnologias e padrões de metadados para aumentar o escopo das atividades desempenhadas automaticamente (BERNERS-LEE, HENDLER e 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 tesauros (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 num determinado domínio.

Borst (1997, p.12) apresenta uma definição muito aceita pela comunidade de ontologia: “uma especificação formal e explícita de uma conceitualização compartilhada”, onde “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 e VICENTE, 1996) têm-se: 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; e i) **regras**, que determinam condições ao domínio, além de inferirem 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 al., 1999; USCHOLD e GRUNINGER, 1996). Apesar de uma grande quantidade de ontologias já ter sido desenvolvida por diferentes comunidades - na química (GÓMEZ-PEREZ, FERNANDEZ e 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 e JURISTO, 1997). A conseqüência, 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 disto, 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 ao cenário exposto acima, 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 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.

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 normas 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; e **v)** análise comparativa das metodologias, dos métodos e da norma.

Para a identificação de documentos, foram consultadas bases de dados de documentos científicos (portal de periódicos da Capes¹ e a biblioteca digital Citeseer²) 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 através da utilização de técnicas de análise de conteúdo, as quais permitiram a seleção dos documentos pertinentes à análise. A análise de conteúdo contempla um conjunto de técnicas de análise das comunicações visando obter indicadores (quantitativos ou não) que permitam a inferência de conhecimentos presentes nas mensagens (BARDIN, 1977).

Após a obtenção da amostragem documental, pôde-se, finalmente, identificar as metodologias discutidas nos documentos. As metodologias e os métodos para construção de ontologias escolhidos para a atividade de análise comparativa foram: a) metodologia de **Gruninger e Fox**; b) método de **Uschold e King**; c) metodologia **Methontology**; d) método **Cyc**; e) método **Kactus**; f) método **Sensus**; e 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.

Na atividade de identificação de metodologias para construção de vocabulários controlados constatou-se a existência de normas e manuais criados por entidades com aceitação nacional e internacional. 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; e 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. Esta norma é a **ANSI/NISO Z39.19-2005** (ANSI, 2005), que foi construída tendo

como base as diversas normas americanas e internacionais sobre a criação de tesouros, incluindo a ISO 2788. Além da norma, foi utilizado o manual disponível no sítio da Biblioteconomia, Informação e Tecnologia da Informação – **BITI** (CAMPOS, GOMES e MOTTA, 2004) sobre elaboração de tesouros. Tal manual, mesmo focando em um tipo específico de vocabulário controlado, mostrou-se coerente em relação às recomendações tratadas na norma ANSI/NISO Z39.19-2005.

Na análise preliminar de conteúdo feita nos materiais sobre construção de ontologias e vocabulários controlados, percebeu-se certa semelhança entre algumas fases de desenvolvimento de tais instrumentos 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. 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). A escolha da norma justifica-se pelo fato da mesma descrever um processo estruturado e metódico para desenvolvimento de produtos (PRESSMAN, 2002), e por advir da Engenharia de Software, uma disciplina considerada madura no sentido de possuir metodologias amplamente aceitas (FERNANDEZ, 1999, p.1). E, como as ontologias podem ser consideradas componentes de produtos de *software* (FERNANDEZ, 1999), achou-se pertinente utilizar tal norma como instrumento de análise qualitativa dos dados.

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

- **Gerenciamento do projeto:** ocorrem atividades relacionadas ao início de um projeto como criação do processo e ciclo de vida do *software*; ao planejamento da gestão de um projeto e ao monitoramento e controle do projeto de *software* 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 *software*. A fase inclui atividades de estudo de viabilidade e análise de requisitos do sistema.
- **Especificação de requisitos:** abrange as restrições ou regras que o *software* 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, e que satisfaça os requisitos de *software* 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 pro-

fissional 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 al., 2003), por exemplo.
- **Manutenção:** considerada uma etapa pós-desenvolvimento, em que consiste em identificar problemas e melhorias nos produtos, podendo resultar em novas versões nos mesmos.
- **Integração:** uma alternativa para facilitar a construção de ontologias é buscar integrar a ontologia em questão a ontologias existentes. Esta fase considera a reutilização de conceitos existentes em outras ontologias. A proposta é examinar a conceitualização de meta-ontologias (chamadas ontologias de alto nível) e selecionar (parcialmente ou por completo) àquelas 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.
- **Avaliação:** suas atividades são executadas ao mesmo tempo com atividades dos processos orientados ao desenvolvimento do *software* 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 para desenvolvedores e usuários envolvidos nos processos, a fim de fornecer, em tempo hábil, informações sobre o *software*.

Ressalta-se que a categoria formalização foi direcionada também para o contexto dos vocabulários controlados, quando se define as suas formas de apresentação.

Feita a coleta e registro dos dados, o próximo passo foi o desenvolvimento de um quadro comparativo contendo as fases de desenvolvimento apresentadas nas metodologias e métodos para construção de ontologias e vocabulários controlados. Tal escopo foi concebido através 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, o mesmo se propôs a apresentar 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 101. E, finalmente, a se-

ção 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 e 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 Ramathan Guha (FERNANDEZ, GOMEZ-PEREZ e 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; e 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 gerenciarem 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 e 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 através 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; e d) definição de uma simbologia para representar graficamente termos ou conceitos (GRUNINGER e 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, a saber: 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, através de declarações em lógica de primeira ordem, os conceitos e suas propriedades são organizados em uma taxono-

mia; 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; e 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* através 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; e 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 e 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: 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; e iv) buscar por ontologias já desenvolvidas por outras aplicações no sentido de reutilização das mesmas. 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 e VICENTE, 1996; FERNANDEZ, GOMEZ-PEREZ e 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 alcançarem 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 em 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; e 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 70.000 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 al., 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 al., 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 al., 1996; FERNANDEZ, GOMEZ-PEREZ e 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; e v) adicionar subárvores completas.

3.7 Método 101

O método 101 foi concebido por Natalya F. Noy e Deborah L. McGuinness (NOY e 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 al., 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*); e iv) adicionar valores de *slots* para as instâncias. Tais atividades implicam em 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 tesouros, taxonomias, listas e anel de sinônimos, em uma ordem conhecida e estru-

turada 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 uma 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 e MOTTA, 2004), na Inglaterra, para a elaboração de tesouros 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 e MOTTA, 2004) sobre elaboração de um tipo de vocabulário controlado, os tesouros, explicita em detalhe e em seqüência 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 explicitamente 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 em agrupar os termos de mesma natureza em categorias ou facetas, a fim de permitir uma 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 se o vocabulário controlado está fornecendo resultados de busca satisfatórios, implicando em uma boa relação entre precisão e revocação.

Tendo apresentado os objetos investigados, cabe a realização da classificação de conteúdo em cada categoria denominada, a partir desse ponto, fase do processo de construção. Desse modo, as fases *gerenciamento do projeto*, *pré-desenvolvimento*, *especificação de requisitos*, *modelagem conceitual*, *formalização*, *implementação*, *manutenção*, *integração*, *avaliação* e *documentação* são apresentadas no Quadro 1 a seguir, que sintetiza a análise das metodologias, dos métodos e norma investigados. As fases que não foram propostas nos objetos investigados tiveram suas células preenchidas com “Ausente”; quando contempladas, a célula foi preenchida com princípios metodológicos adotados.

Fases do ciclo de vida		Cyc	Gruninger e Fox	Uschold e King	Kactus	Methontology	Sensus	Método 101	Z39.19-2005	Manual BITI	
Gerenciamento do Projeto		Ausente	Ausente	Ausente	Ausente	Escalonamento das atividades; controle das tarefas e garantia da qualidade.	Ausente	Ausente	Recomenda características necessárias aos sistemas de gerenciamento.	Atividades de planejamento do projeto.	
Processos orientados	Pré-desenvolvimento	Ausente	Cenários de motivação	Ausente	Ausente	Ausente	Ausente	Ausente	Recomenda um estudo de viabilidade para a construção.	Ausente	
	Desenvolvimento	Especificação de requisitos	Extração manual do conhecimento requerido.	Questões de competência informal.	Determinar o propósito da ontologia.	Lista de requisitos.	Definição do escopo da ontologia.	Ausente	Definição de questões de competência	Ausente	i) delimitação da área; e ii) definição do público-alvo.
		Modelagem conceitual	Ausente	Concepção da terminologia formal.	Construção do vocabulário consensual.	Identificação de termos relevantes.	Atividades envolvendo a construção da conceitualização da ontologia.	Identificação de termos-chave do domínio.	Definição de classes, <i>slots</i> , facetas (restrições) e instâncias.	Recomenda a análise facetada na organização do conhecimento; organização das relações em hierárquica, associativa e de equivalência.	Seleção de termos; definição de conceitos e relações (lógica e ontológica); classificação em facetas.
		Formalização	Ausente	Formalização das questões de competência. Especificação dos axiomas.	Ausente	Ausente	Formalizar o modelo conceitual em uma linguagem formal como lógica descritiva.	Através de rede semântica.	Linguagem de representação baseada em <i>frames</i> .	Recomenda tratar a representação dos relacionamentos dos termos.	Listas alfabéticas e visualizações gráficas; formatos impressos ou eletrônicos.

	Implementação	Codificação e extração do conhecimento com apoio de ferramentas computacionais.	As especificações dos axiomas são implementadas na linguagem Prolog.	Requer uma linguagem de representação como Prolog e Ontolingua.	Ausente	Recomenda ambientes de desenvolvimento adequados à meta-ontologias e ontologias selecionadas na fase de integração.	Utiliza a ferramenta Onto-Saurus para desenvolvimento.	Uso interativo da linguagem através da ferramenta <i>Protégé</i>	Ausente	Ausente	
	Pós-desenvolvimento	Manutenção	Ausente	Ausente	Ausente	Ausente	Orientação quanto a ajustes necessários nas atividades de conceitualização.	Ausente	Ausente	Recomenda procedimentos relacionados à adição, à alteração e à exclusão de termos.	A inclusão de novos termos deve ser proposta numa planilha específica.
Processos integrais	Integração	Integração de sistemas computacionais à micro-teorias da base Cyc.	Integração a ontologias de núcleo comum.	Integração com ontologias existentes.	Buscar por ontologias já desenvolvidas.	Documento de integração com meta-ontologias.	Integração manual à ontologia SENSUS.	Considera a reutilização de ontologias.	Ausente	Ausente	
	Avaliação	Ausente	Através de teoremas completos.	Pode ser feita através de questões de competência.	Ausente	Divide-se em verificação e validação da ontologia.	Ausente	Ausente	Recomenda executar testes de usabilidade.	Critérios adotados para a adoção ou não de um tesouro.	
	Documentação	Ausente	Ausente	Pretensões e primitivas.	Ausente	Recomendada em cada fase.	Ausente	Via <i>Protégé</i>	Recomendações quanto ao conteúdo da documentação.	Registro dos dados em planilhas.	

Quadro 1 – Quadro sinóptico das metodologias, métodos e norma frente às categorias de análise predefinidas

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 condução das mesmas, 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, a manutenção e a avaliação de ontologias.
- 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 implementação e integração não foram consideradas por não pertencerem ao propósito dos vocabulários controlados. Desse modo, pode-se conferir a maturidade da metodologia frente a um modelo de ciclo de vida.
- Finalmente, é válido ressaltar que a metodologia para construção de tesouros se destacou perante as metodologias e aos métodos para construção de ontologias no quesito princípios teóricos e metodológicos para identificação, definição e organização de conceitos. Apesar de essas últimas indicarem métodos para identificar conceitos, apresentarem recursos para se definir conceitos e organizá-los em uma taxonomia, não deixam explícitos os princípios teóricos que governam os seus construtos. Já no caso dos vocabulários controlados é explícita a inserção de princípios da teoria da classificação (RANGANATHAN, 1967) e da teoria do conceito (DAHLBERG, 1978) na especificação de seus elementos. Tais princípios seriam pertinentes em metodológicas para construção de ontologias, visto que ambos os instrumentos, ontologias e tesouros, representam relacionamentos semânticos e conceituais.

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 exibido na seção 3. A metodologia exposta no manual da BITI para construção de tesauros 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 frente à 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 usado para a definição de padrões metodológicos para construção de ontologias.

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.

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.

BARDIN, Laurence. *Análise de conteúdo*. Lisboa: Edições 70; 1977.

BERNARAS, A.; LARESGOITI, I.; CORERA, J. *Building and Reusing Ontologies for Electrical Network Applications*. In: PROCEEDINGS OF THE EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, ECAI/96, p. 298-302, 1996.

BERNERS-LEE, T; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American*, vol. 284, n°. 5, maio 2001, p. 34-43.

BORST, W.N. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. 1997. Tese (Phd). Disponível em: <<http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>>. Acesso em: 03 outubro 2005.

CAMPOS, M.L.A.; GOMES, H.E.; MOTTA, D.F. *Tutorial de Tesauro*. 2004. Disponível em: <http://www.conexaorio.com/bit/tesauro> Acesso em: 30 abril 2008.

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 Abril 2008.

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

DEAN M., CONNOLLY D., VAN HARMELEN F., HENDLER J., HORROCKS I., MCGUINNESS D.L., PATEL-SCHNEIDER P.F., and STEIN L.A. *OWL Web Ontology Language 1.0 Reference*. W3C Working Draft 21. 2003.

FERNANDEZ, M.; 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 Agosto 2007.

FERNÁNDEZ, M; GOMEZ-PEREZ, A.; SIERRA, J.P.; SIERRA, A.P . Building a Chemical Ontology Using Methontology and the Ontology Design Environment. *Intelligent Systems*, v. 14, n. 1, p. 37-46, jan./ fev.1999.

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A; CORCHO, O. Methodologies and Methods for Building Ontologies. In: GÓMEZ-PÉREZ, A.; FÉRNANDEZ-LÓPES, M.; CORCHO, O. *Ontological Engineering*. London: Springer, 2004. pp. 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 Novembro 2007.

GILCHRIST, Alan. Thesauri, taxonomies and ontologies - an etymological note. *Journal of Documentation*, v. 59, n. 1, 2003, p. 7-18.

GÓMEZ-PÉREZ, A., FERNÁNDEZ, M., VICENTE, A.J. *Towards a Method to Conceptualize Domain Ontologies*, ECAI'96 - Workshop on Ontological Engineering, Budapest, August, 1996. Disponível em: <<http://citeseer.ist.psu.edu/483876.html>>. Acesso em: 12 Novembro 2007.

GRUBER, T.. *What is an Ontology?* 1993. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 03 Agosto 2006.

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 Agosto 2006.

GRUNINGER, M.; FOX, M. S. *Methodology for the design and evaluation of ontologies*. 1995. Disponível em: <<http://citeseer.ist.psu.edu/grninger95methodology.html>>. Acesso em: 10 Novembro 2007.

GRUNINGER, M.; FOX, M. S. *The Logic of Enterprise Modelling*. 1996. Disponível em: <<http://citeseer.ist.psu.edu/523974.html>>. Acesso em: 10 Novembro 2007.

HORRIDGE, M. et al. *A Practical Guide to Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools*. Ed.1. The University of Manchester and Stanford University, Agosto, 2004. Disponível em: <<http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>>. Acesso em: 07 Fevereiro 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 Julho 2007.

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: 03 Outubro 2006.

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. 640p.

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 Abril 2008.

SILVA, Daniela Lucas da. *Uma Proposta Metodológica para Construção de Ontologias: Uma Perspectiva Interdisciplinar entre as Ciências da Informação e da Computação*. 283 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*. 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*. 202p. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

SWARTOUT, B; PATIL, R.; KNIGHT, K.; RUSS, T. *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 Abril 2008.

USCHOLD, M.; KING, M. *Towards a Methodology for Building Ontologies*. 1995. Disponível em: <<http://citeseer.ist.psu.edu/uschold95toward.html>> Acesso em: 10 Novembro 2007.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods an applications. *Knowledge Engineering Review*, v. 11, n. 2, 1996.

VICKERY, B. C. Ontologies. *Journal of Information Science*, v. 23, n. 4, p. 277-286, 1997.

¹ <http://www.periodicos.capes.gov.br/portugues/index.jsp>

² <http://citeseer.ist.psu.edu>