

GT 2 – Organização e Representação do Conhecimento

Modalidade de apresentação: Comunicação Oral

DIRETRIZES PARA DEFINIÇÃO DE DOMÍNIO NO REUSO DE ONTOLOGIAS BIOMÉDICAS: UMA ABORDAGEM BASEADA NA ANÁLISE DO COMPROMISSO ONTOLÓGICO

Linair Maria Campos

Universidade Federal Fluminense/IBICT

Maria Luiza de Almeida Campos

Universidade Federal Fluminense

Maria Luiza Machado Campos

Universidade Federal do Rio de Janeiro

Resumo: Embora tenham evoluído muito nos últimos anos, as metodologias para reúso ainda carecem de maior detalhamento nas ações a serem tomadas para a identificação do compromisso ontológico na etapa de definição do domínio considerado, e a reorganização das ontologias reutilizadas de acordo com uma visão de alto nível, utilizando ontologias de fundamentação como, por exemplo, a BFO. Em nossa proposta, procuramos fornecer uma série de diretrizes voltadas para preencher esta lacuna. Essas diretrizes se apóiam em aportes teóricos da Ciência da Informação (CI), fundamentados em autores como Neville, Dahlberg e Ranganathan, e da Ciência da Computação (CC), em autores como Guarino e Welty e Guizzardi, em uma perspectiva interdisciplinar. Neste artigo, detalhamos uma dessas diretrizes, voltada para compreender a estratégia de organização das classes das ontologias, suas naturezas e seus problemas de organização. Para evidenciar os benefícios dessa proposta, a diretriz foi aplicada por sobre um recorte previamente determinado da Gene Ontology (GO), resgatando-se aspectos de seu compromisso ontológico. Como resultado, obtivemos possíveis classes de conceitos que poderiam ser utilizados para estender a GO para o domínio estudado, definindo-o de forma mais precisa. Os resultados apontam ainda para a possibilidade de padronizar as definições de termos, de modo a favorecer o uso mais preciso de ferramentas de software na manipulação de definições de forma mais estruturada. A partir dos resultados obtidos, concluímos que o conjunto de diretrizes propostas fornece apoio para guiar a definição de domínio de maneira mais precisa, através do resgate do compromisso ontológico e das perspectivas que podem ser adotadas para a complementação do domínio alvo, de acordo com o que é resgatado.

Palavras-chave: Ontologia. Reuso de ontologias. Diretrizes teórico-metodológicas. Compromisso ontológico. Definição de domínio.

1. Introdução

Em domínios de temática complexa, como fruto da sua constante produção científica, os dados e recursos precisam ser articulados e, para isso, descritos de forma padronizada para um melhor proveito de suas comunidades, as quais muitas vezes possuem demandas de informação crescentes e urgentes. Nesse sentido, em algumas áreas, como por exemplo, a Biomedicina, comunidades interessadas têm desenvolvido e divulgado, publicamente, desde a década de 1990 (PEREZ-REY et al., 2004), uma série de *ontologias* para auxiliar na descrição e recuperação do seu conteúdo informacional. De acordo com Guarino: “ (...) no seu uso mais prevalente em inteligência artificial (IA), uma ontologia diz respeito a um *artefato de engenharia*, constituído por um *vocabulário* específico usado para descrever uma certa realidade, mais um conjunto de afirmativas que se referem ao *significado pretendido* das palavras do vocabulário” (GUARINO, 1998, p.2). As afirmativas, referidas por Guarino, são geralmente axiomas utilizados para aumentar a precisão do vocabulário. Entretanto, algumas ontologias (denominadas de ontologias leves, ou *lightweight*) não possuem axiomas, limitando-se, na maioria das vezes a um conjunto de termos com relacionamentos de gênero-espécie (*is-a*) e partitivos. Este é o caso de muitas das ontologias biomédicas, como as encontradas na OBO (Open Biomedical Ontologies) (2009). A OBO é um consórcio que se propõe a coordenar o desenvolvimento ortogonal, não redundante, e padronizado das ontologias da área. As ontologias da OBO, embora estejam sofrendo um processo de reformulação, ainda apresentam uma série de problemas (SMITH, WILLIAMS E SCHULZE-KREMER, 2003), que incluem deficiências na forma como a estrutura hierárquica de seus conceitos é formada, na maneira como são formuladas suas definições, na forma como os conceitos são relacionados, dentre outros.

Estes problemas, que muitas vezes se originam na construção das ontologias, se refletem na articulação de ontologias afins, não só pela constante mudança e pelo grande número de termos, como também pela existência de escopos pouco claros e pela falta de documentação dos pressupostos teóricos que norteiam a organização sistemática das ontologias. A articulação destas ontologias pode ser de grande valia, pois devido à complexidade de suas temáticas, muitas vezes mais de uma ontologia é necessária para um determinado propósito. Por exemplo, pode-se usar a Gene Ontology (GO) (GENE

ONTOLOGY CONSORTIUM, 2000) para anotar aspectos relativos às funções moleculares de um gene, e utilizar a *Pathway* (OBO, 2009) para associar a via biológica onde estas funções estão inseridas. Se essas ontologias estivessem articuladas, o processo de anotação genômica poderia ser mais preciso, ou, pelo menos, mais facilitado. Naturalmente, a interligação de ontologias construídas sob diferentes princípios também apresenta problemas, como, por exemplo, a existência de termos equivalentes, porém considerados segundo diferentes níveis de detalhe (um mais genérico do que o outro, ou vice-versa).

A fim de minimizar os problemas de construção e reúso de ontologias, várias metodologias ou diretrizes (PINTO e MARTINS, 2001; PÂSLARU-BONTAS, 2007; SURE, STAAB E STUDER, 2002; USCHOLD E KING, 1996; FERNÁNDEZ-LÓPEZ, GÓMEZ-PÉREZ, JURISTO, 1997; PLANT E GAMBLE, 2003; GANGEMI, STEVE e GIANCOMELLI, 1996) têm sido propostas. Algumas incluem o estudo ou análise do domínio como parte de sua estratégia, de modo a tomar diretrizes sobre como abordar os documentos e especialistas para extrair o conhecimento ali contido. Este estudo, quando existente, costuma ser adotado após a etapa de planejamento da ontologia. Entretanto, podemos observar que existe uma carência de detalhes entre o estudo do domínio e o produto final esperado desse estudo: um modelo conceitual do domínio, com suas classes de relevância organizadas segundo princípios bem definidos e explicitados. Por exemplo, no contexto da análise de domínio, reconhece-se a necessidade de se *identificar* entidades, relações, funções e axiomas comuns às diversas fontes analisadas (GUIZZARDI, 2000), porém seria útil haver um procedimento sistemático para a *organização* desses elementos, no contexto do reúso. Neste sentido, não encontramos uma proposta metodológica de reúso de ontologias que, a partir das temáticas e conceitos identificados por uma análise prévia do domínio, explique como se deve dar o recorte dessas ontologias para o escopo representativo do domínio desejado.

O objetivo deste trabalho é evidenciar que o reúso de ontologias biomédicas pode ter sua precisão aumentada através da adoção de diretrizes para a definição de recorte de domínio, baseadas na análise do compromisso ontológico destas ontologias. Devido a restrições de espaço, vamos nos limitar a apresentar o conjunto de diretrizes propostas e focar a discussão em uma das diretrizes, voltada para compreender a estratégia de

organização das classes das ontologias, suas naturezas e seus problemas de organização. Nossa proposta é baseada em aportes teóricos das áreas da Ciência da Informação (CI) e da Ciência da Computação (CC), em uma iniciativa de caráter interdisciplinar. O trabalho encontra-se estruturado como se segue: no item 2 apresentamos os fundamentos teóricos utilizados; no item 3 os trabalhos relacionados; no item 4 as diretrizes propostas, detalhando uma delas, a partir de sua aplicação em um experimento realizado; no item 5 um experimento conduzido e, no item 6 as nossas conclusões.

2. Fundamentos teóricos

A diretriz discutida neste trabalho propõe explicitar o *compromisso ontológico* e, neste processo, apontar problemas encontrados na organização das ontologias analisadas.

Guarino e colegas (1994) situam o papel do compromisso ontológico como o de um elemento fomentador da precisão entre a conceituação, que está no plano da semântica do entendimento humano, e a representação de uma visão de mundo. Esta última, um modelo desse mundo, imprecisa em algum grau em relação ao significado pretendido pelo homem. É formada por um conjunto de regras que objetiva minimizar ambigüidades ou entendimentos conflitantes dos conceitos do mundo sendo representado de acordo com a conceituação do homem, e embora seja voltada para o entendimento humano, é também adequada para a conversão formal em uma linguagem que possa ser entendida pela máquina.

Para explicitar o compromisso ontológico, propomos tratar no escopo deste artigo, de dois aspectos: (i) a perspectiva classificatória para analisar o recorte do domínio; (ii) a perspectiva para a organização dos elementos do domínio. Estes aspectos são abordados a seguir. O detalhamento dessa proposta encontra-se no contexto de uma tese de doutorado, cuja qualificação foi aprovada em 2010 (CAMPOS, 2010).

2.2 Perspectiva classificatória do recorte do domínio

A perspectiva classificatória estabelece uma forma sistemática de alto nível para se definir e classificar o conhecimento de um domínio.

Na CI, Dahlberg, com a Teoria do Conceito (1978b), fornece-nos um método de raciocínio analítico-sintético para organizar os conceitos de um domínio em grandes categorias, a partir da análise da sua definição (pensamento analítico), e da descoberta de características comuns que revelam categorias para agrupar esses conceitos (pensamento sintético). Outro teórico importante da CI é Ranganathan (1967), que propõe uma série de princípios classificatórios para a organização dos conceitos de um domínio dentro de classes, as quais são estruturadas de acordo com categorias utilizadas para pensar a sua organização (CAMPOS e GOMES, 2008).

Na CC, especialmente propostas por teóricos que se aproximam da Filosofia e das Ciências Cognitivas, encontramos iniciativas que descrevem as diferentes meta-propriedades dos conceitos, geralmente expressas em ontologias formais de alto-nível ou de fundamentação (GUIZZARDI, 2005; SMITH e GRENON, 2004). Essas ontologias, bem como as noções que as fundamentam, são importantes contribuições para a definição de primitivas que ajudam a identificar diferentes tipos de conceito. Algumas destas noções são explicadas a seguir, observando que possuem muitos aspectos em comum com a CI.

2.2.1 Universais e particulares

A noção de universal diz respeito à de generalização de tipos, categorias, espécies de entidades no mundo. Por sua vez, as entidades individuais no mundo, generalizáveis, são denominadas *particulares*, ou ainda, instâncias das respectivas entidades universais, ou ainda, indivíduos. A noção de universais e particulares é tomada como princípio em ontologias de fundamentação¹ como a Basic Formal Ontology (BFO) (GRENON, SMITH, e GOLDBERG, 2003) e a DOLCE (GANGEMI et al., 2002) e em iniciativas da CC que tratam de aspectos teóricos da construção de ontologias (GUARINO, 1998; GUIZZARDI, 2005; SMITH e GRENON, 2004). Embora as ontologias sejam formadas por termos que representam universais, cada termo deve corresponder à existência de instâncias.

Na Ciência da Informação a noção de universais e particulares também é trabalhada, embora com terminologia por vezes distinta. Dahlberg (1978a) menciona a distinção entre objetos individuais e objetos gerais, com o mesmo sentido:

¹ Ontologias de fundamentação, ou de alto nível descrevem conceitos gerais, tais como: tempo, espaço, matéria, objeto, evento, ação, e são independentes de um domínio particular.

toda vez que o objeto é pensado como único, distinto dos demais, constituindo uma unidade inconfundível (coisas, fenômenos, processos, acontecimentos, atributos, etc) pode-se falar de objetos individuais. Pode-se dizer que o que caracteriza os objetos individuais é a presença de formas do tempo e espaço. Os objetos individuais estão aqui e agora. Ex.: esta casa, esta mesa, este automóvel, esta partida de futebol. (...) Mas, além dos objetos individuais (...) podemos referir-nos a objetos gerais que, de certo modo, prescindem das formas do tempo e do espaço. A estes objetos (...) correspondem os chamados conceitos gerais (...). (DAHLBERG, 1978a).

Dahlberg, porém, traz a preocupação com a definição do conceito, afirmando que com a ajuda da linguagem natural é possível formular enunciados (afirmativas) sobre os objetos (gerais e individuais), sendo cada enunciado um elemento do conceito e sendo que o conjunto de tais enunciados sobre um objeto fornece o conceito deste. Dahlberg afirma ainda que os elementos contidos nos conceitos gerais se encontram também nos conceitos individuais. Através da formulação dos enunciados sobre os atributos dos objetos se obtém as características dos respectivos conceitos (DAHLBERG, 1978a).

2.2.2 Natureza das entidades de um domínio

Para ajudar a organizar as entidades de um domínio por sua natureza, Ranganathan elaborou a Teoria da Classificação Facetada na qual apresenta princípios para a organização de conceitos hierarquicamente estruturados (CAMPOS, 1995). A Classificação Facetada parte da identificação dos domínios de conhecimento que se quer representar. A partir daí, dentro de cada domínio, este é analisado e seus elementos básicos identificados e agrupados em classes de acordo com características comuns. Essas classes são chamadas de facetas. Cada faceta, por sua vez, pertence a uma categoria fundamental. As categorias propostas por Ranganathan são referenciadas pelo acrônimo PMEST (*Personality, Mater, Energy, Space, Time*): Personalidade, Matéria, Energia, Espaço e Tempo.

De acordo com Ranganathan, *Matéria* se refere à manifestação intrínseca de materiais e propriedades do conteúdo sendo descrito (CAMPOS, 2001, p.60). Entretanto, para Ranganathan, as propriedades não fazem parte de uma categoria, sendo consideradas uma manifestação de uma das categorias² (CAMPOS, 2001, p. 131). Assim, por exemplo, a cor de uma cadeira não pertence à categoria de Matéria, como

² Já Dahlberg, no âmbito da Ciência da Informação, atribui às propriedades o caráter de categoria.

uma manifestação isolada, mas ocorre ligada a uma entidade, no caso, a cadeira. *Energia* se refere aos seus processos, *Espaço* a aspectos de sua localização, *Tempo* aos seus atributos temporais, e, finalmente, *Personalidade* às entidades relacionadas, suas partes ou o que não pertence às outras categorias.

De maneira semelhante à proposta de Ranganathan, ontologias de topo como a BFO e a DOLCE distinguem a natureza das entidades do mundo através de categorias que são independentes de domínio. Na BFO esta categorização é dividida em continuantes (continuant) e ocorrentes (occurent), as quais se subdividem em outras. Continuantes e ocorrentes são denominados, respectivamente, endurantes (endurants) e perdurantes (perdurants) na DOLCE. Na proposta de categorias fundamentais de Ranganathan (1967), endurantes corresponderiam à categoria de Personalidade, enquanto que perdurantes corresponderiam à categoria de Energia.

Porém, na BFO e na DOLCE a diferença entre continuantes e ocorrentes está centrada na perspectiva de existência no tempo, enquanto que para Ranganathan esta não é a característica diferencial, haja vista a categoria fundamental de *Espaço*, utilizada para classificar localizações, como, por exemplo, continentes e países (CAMPOS e GOMES, 2003), que seriam considerados continuantes na BFO ou perdurantes na DOLCE.

Continuantes são entidades que preservam sua identidade ao longo do tempo, mesmo que sofram mudanças. Organismos, células e moléculas são exemplos de continuantes. Os continuantes existem por completo durante um determinado período de tempo (o período de sua existência), ou então não existem mais. Ou seja, não é concebível ter um mesmo continuante que exista hoje, não exista amanhã e volte a existir depois de amanhã. E, ainda, em um dado momento, todas as partes do continuante estão presentes, formando um todo.

Continuantes podem ser de três tipos: dependente, independente ou região no espaço. Continuantes dependentes, tais como as qualidades ou papéis, são entidades que para existir necessitam de outras que sejam independentes. Continuantes independentes são portadores dos dependentes. Por exemplo, moléculas (continuante independente) são portadoras de função molecular (continuante dependente).

Ocorrentes, por outro lado, nunca ocorrem por inteiro em um dado momento no tempo. Eles se desdobram em fases ao longo do tempo. Exemplos de ocorrentes são: eventos, processos, atividades. De acordo com Grenon, Smith e Goldberg (2003) “Isto significa que se nós segmentamos o intervalo de tempo durante o qual um ocorrente ocorre então nós segmentamos também o ocorrente.”. Além disso, as partes de um ocorrente são sempre ocorrentes, embora os continuantes possam participar de ocorrentes.

Na BFO, ocorrentes podem ser de três tipos: entidade processual, região espaço-temporal e região temporal. Entidades processuais dizem respeito a processos, suas partes, seu contexto, limites e agregados de processos (conjunto de processos). Região espaço-temporal diz respeito ao tempo-espacó onde um determinado processo se desenrola. Na DOLCE esta noção corresponde às noções de localização espacial (que para Ranganathan corresponde à categoria de Espaço) e localização temporal. Região temporal, na BFO, diz respeito a intervalos de tempo, correspondendo na DOLCE à noção de intervalo de tempo e à categoria fundamental de Tempo, de Ranganathan.

A DOLCE possui outros ramos de primeiro nível, para representar qualidade e entidades abstratas (GANGEMI et al., 2002), entretanto foge do escopo deste trabalho, aprofundar esta discussão.

2.2.3 As noções de identidade, rigidez e unidade

Identidade diz respeito a estabelecer as características únicas que distinguem uma determinada instância de uma classe das outras instâncias (GUARINO e WELTY, 2000; 2000b; 2001). Uma questão ligada ao conceito de identidade diz respeito à noção de rigidez. Uma propriedade é considerada rígida se é aplicável a todas as instâncias de uma entidade ao longo do tempo (GUARINO e WELTY, 2000b), ou seja, se a entidade continua sendo a mesma, durante todo o tempo, mesmo sofrendo mudanças. Por exemplo, costumamos tomar a propriedade PESSOA como rígida, pois podemos entender que uma entidade nunca deixa de possuir a propriedade de ser PESSOA enquanto existir, mesmo que sofra mudanças. Por outro lado, a propriedade ESTUDANTE não é rígida, pois se pode entender uma entidade que possua a propriedade de ser ESTUDANTE apenas em parte de sua existência, embora permaneça o mesmo indivíduo (GUARINO e WELTY, 2000b). Guarino e Welty alertam ainda para a distinção entre

carregar (ou ser portador de) uma condição de identidade e de *fornecer* esta condição. Nesse sentido, podemos dizer que propriedades rígidas *fornecem* uma condição de identidade, enquanto que propriedades que não são rígidas apenas *carregam* essa condição.

Unidade diz respeito a estabelecer quais as partes que caracterizam uma determinada instância de uma entidade, formando um todo (possuindo uma condição de unidade), através de uma relação de unificação que une essas partes, e é válida para todas as instâncias dessa entidade. Dependendo da natureza ontológica de tal relação podem-se ter diferentes tipos de unidade. Por exemplo, é possível distinguir unidade *topológica* (perspectiva da estrutura), unidade *morfológica* (perspectiva da forma) e unidade *funcional* (perspectiva da função). Guarino e Welty propõem que a noção de todo como sendo *essencial*, ou seja, perdurando para uma entidade durante toda a sua existência (GUARINO e WELTY, 2000b).

A partir das noções de identidade, unidade, individualidade e rigidez, Guarino propõe um conjunto de restrições para orientar a formação da *taxonomia central* (*backbone taxonomy*) de ontologias (GUARINO e WELTY, 2000b). Esta taxonomia central consiste apenas de propriedades rígidas, que são de três espécies: *categorias*, as quais não carregam identidade e estão no nível mais alto (por serem muito genéricas); *tipos*, que fornecem identidade; e *quase-tipos*, que carregam, mas não fornecem identidade. As restrições propostas servem como diretrizes para evitar inconsistências. Guizzardi, por sua vez, propõe uma teoria para modelagem conceitual e apresenta um conjunto de quatro postulados para apoiar a construção de modelos conceituais bem fundamentados (GUIZZARDI, 2005) (GUIZZARDI, WAGNER, SINDEREN, 2004). Estes postulados estão em conformidade com as restrições propostas por Guarino e Welty.

O **primeiro postulado** afirma que cada objeto em um modelo conceitual do domínio tem de ser obrigatoriamente uma instância de uma classe representando um *Sortal* (entidade que *carrega* uma condição de identidade, podendo ser ou não rígida). O **segundo postulado** afirma que um Objeto em um modelo conceitual do domínio não pode instanciar mais do que uma classe do modelo conceitual representando um Sortal do tipo *Substância* (*substance sortal*). Um Sortal do tipo *Substância* é o sortal único que *fornecer* o princípio de identidade para suas instâncias (sendo, portanto, rígido). O

terceiro postulado afirma que uma classe de um modelo conceitual representando um classificador rígido não pode ser uma subclasse de uma classe representando um classificador anti-rígido³. O **quarto postulado** afirma que uma classe de um modelo conceitual representando um universal dispersivo (*dispersive universal*) não pode ser uma subclasse de uma classe representando um Sortal. Um universal dispersivo representa um amplo espectro de conceitos com diferentes princípios de identidade. Por exemplo: Coisa, Elemento, Entidade, dentre outros. Desta forma, universais dispersivos não fornecem um princípio de identidade para suas instâncias. Assim como, pelo mesmo motivo, as categorias fundamentais também não fornecem um princípio de identidade.

As abordagens acima explicadas fornecem um princípio ontológico segundo o qual os conceitos devem ser classificados em mais alto nível e ainda aspectos que permitem caracterizar de forma precisa a sua natureza. Entretanto, existem aportes metodológicos da CI que podem apresentar perspectivas equivalentes ou complementares, no escopo de mais de cinqüenta anos dedicados ao estudo de organização de domínios de conhecimento. Estes aportes são discutidos a seguir.

2.2.4 Teoria do Conceito e o Registro do Conceito

A Teoria do Conceito, formulada por Dahlberg (1978), fornece-nos um método de raciocínio analítico-sintético para organizar os conceitos de um domínio em grandes categorias, a partir da análise da sua definição (pensamento analítico) e da descoberta de características comuns que revelam categorias para agrupar esses conceitos (pensamento sintético). De acordo com Dahlberg (1983), de um ponto de vista formal, um conceito consiste de três partes: (A) o referente, (B) as características predicativas de um referente e (C) a expressão verbal ou codificada que denota o referente na forma de um termo, uma descrição ou uma notação. Dahlberg destaca, ainda, a importância das definições dos termos para a construção de qualquer tipo de sistema de classificação ou tesauro, especialmente se estes forem ser comparados: “Os termos são apenas os portadores da informação e só podem exibir a informação sobre o conteúdo do conceito representado se as relações necessárias tiverem sido estabelecidas através de uma definição” (DAHLBERG, 1983, p. 5). No âmbito da compatibilização de linguagens, Dahlberg alerta para a importância de se trabalhar não só no plano do casamento verbal

³ Uma propriedade anti-rígida é *opcional* para *todas* as suas instâncias (GUARINO e WELTY, 2000b).

ou lingüístico dos termos, mas também da análise dos conceitos, para que uma correspondência semântica possa ser estabelecida. Para isso, recomenda que a compatibilização seja apoiada por um *registro do conceito*, o qual contém informações que ajudam a compreender o significado do termo, como, por exemplo: sua definição, o conceito mais alto na hierarquia e a categoria da forma do conceito (form category), a qual indica se este é um objeto, um processo, uma qualidade.

2.2.3 Cânones de Ranganathan aplicados ao reúso de ontologias

Outro teórico importante da CI é Ranganathan (1967), que propõe uma série de princípios classificatórios para a organização dos conceitos de um domínio dentro de classes, as quais são estruturadas de acordo com categorias utilizadas para pensar a sua organização, à semelhança de Dahlberg (CAMPOS e GOMES, 2008). Dentre os princípios classificatórios de Ranganathan (1967), destacamos uma série de cânones, conforme Campos e Gomes (2008), através dos quais propomos analisar a estratégia de formação hierárquica das ontologias que vamos reutilizar: **(i) Cânone da Modulação** - a ordem dos conceitos em uma classe de conceitos deve ter uma seqüência que respeite cada elo da cadeia; **(ii) Cânone de Concomitância** – duas características distintas (por exemplo, idade e data de nascimento) não devem dar origem a um mesmo renque (série horizontal) de conceitos; **(iii) Cânone da Diferenciação** – uma característica usada como base para a classificação de um universo deve dar origem a pelo menos duas classes; **(iv) Cânone da Exclusividade** – o conjunto de elementos de uma classe não deve constituir uma outra classe, ou seja, os elementos de um renque devem ser mutuamente exclusivos; **(v) Cânone para Classes Subordinadas** – numa sucessão de classes cada uma deve seguir imediatamente a outra, sem que seja interrompida pela inclusão de uma ou mais classes de natureza distinta.

A tabela 1 busca ilustrar as abordagens da CI e CC que, embora não sejam exatamente equivalentes, buscam tratar de aspectos semelhantes, e por vezes complementares, da definição de um domínio.

Tabela 1: Abordagens da CI e CC relacionadas à definição de domínio

Perspectiva	CI	CC
Classificatória, para analisar o recorte do domínio	Categorias fundamentais	Ontologias de topo
Organizar elementos do domínio	Cânone da diferenciação	-
	Cânone da concomitância	-
	Cânone da exclusividade	Segundo Postulado de Guizzardi
	Cânone da modulação	-
	Cânone das classes subordinadas	-
	-	Primeiro Postulado de Guizzardi
	-	Terceiro Postulado de Guizzardi
	-	Quarto Postulado de Guizzardi
	Teoria do conceito (foco nas definições, relações e categorização) como base para a formação de estruturas hierárquicas e suas relações	Formalização das noções de Identidade, unidade, rigidez, individualidade, como base para a formação da estrutura taxonômica central de ontologias

De modo geral, podemos observar que a CC pode contribuir para o estabelecimento mais formal da identidade dos conceitos de um domínio, com o foco na sua manipulação por máquinas. Por outro lado, a CI pode ajudar na formação mais sistemática de renques e cadeias, com o foco da sua compreensão e manipulação por pessoas, sem o prejuízo do seu tratamento pela máquina.

3 Trabalhos relacionados a metodologias de reúso de ontologias

Dentre as diversas metodologias de elaboração de ontologias, até onde tivemos notícia, poucas contemplam atividades especificamente voltadas para o reúso de forma detalhada (GANGEMI, PISANELLI e STEVE, 1996, USCHOLD e KING, 1995), embora algumas mencionem a importância do reúso. Por outro lado, existem propostas que possuem um foco bem específico voltado para o reúso, apresentando diretrizes mais detalhadas sobre as etapas e procedimentos a serem adotados para o processo do reúso como um todo (PINTO e MARTINS, 2001; PÂSLARU-BONTAS, 2007). Tanto as metodologias de elaboração quanto as específicas para o reúso nos oferecem uma visão de processo, com etapas e tarefas que devem ser efetuadas para o reúso das ontologias. Dessa maneira, exploramos a literatura relacionada com o objetivo de identificar tais etapas e tarefas de forma geral. Como resultado, obtivemos as seguintes etapas: (i) planejamento; (ii) análise preliminar do domínio; (iii) escolha das ontologias a reutilizar; (iv) definição da estratégia de recorte (o que modularizar); (v) elaboração do modelo

conceptual a partir das ontologias selecionadas (como estruturar e modelar); (vi) definição da estratégia compatibilização; (vii) implementação da ontologia criada; (viii) avaliação da ontologia criada; (ix) documentação da ontologia criada, sendo que esta etapa ocorre ao longo de todo o processo de reúso, sendo que esta última etapa ocorre ao longo de todo o processo de reúso.

Por motivos de espaço, entretanto, vamos detalhar apenas as tarefas da etapa (iv), que vem a ser o foco de nosso trabalho. Essas tarefas encontram-se listadas na Tabela 2 e foram obtidas a partir da revisão de literatura relacionada ao reuso de ontologias.

Tabela 2 - Tarefas do reúso contempladas na etapa de definição da estratégia de recorte

Etapa	Tarefas
(iv) Definição da estratégia de corte e sua obtenção (o que modularizar)	<ul style="list-style-type: none">- Explicitação do compromisso ontológico das ontologias selecionadas;- Adoção de uma ontologia de alto nível para o domínio;- Identificação dos segmentos de interesse (conceitos de interesse, seus relacionamentos, sua natureza) em cada uma das fontes terminológicas a reutilizar, ou seja, delimitação do escopo desejado em cada ontologia;- Identificação dos problemas de compatibilidade a serem tratados;- Captura dos termos que formam o escopo de interesse com as suas definições explicitadas, quando possível.- Estender as ontologias reutilizadas, se necessário.

4. Diretrizes para definição de domínio

O escopo do presente trabalho é voltado para a definição da estratégia de recorte do domínio e sua obtenção. A sua implementação se dá através de um conjunto de diretrizes. De acordo com Nurcan e colegas (1999), diretriz é um conjunto de *indicações* sobre como proceder para atingir um propósito. Desta forma, definimos diretriz como um conjunto de ações *recomendadas* para atingir um objetivo e um resultado esperado, possuindo os seguintes atributos: (i) nome da diretriz; (ii) objetivo esperado; (iii) ações que devem ser realizadas para alcançar o objetivo esperado; (iv) resultado esperado ao aplicar a diretriz.

Dentro da etapa de **definição da estratégia de corte**, identificamos quatro diretrizes, a saber: dir01 - explicitar o compromisso ontológico das ontologias selecionadas; dir02 - adotar uma ontologia de alto nível para o domínio de acordo com o tema alvo; dir03 – identificar os segmentos de interesse nas ontologias selecionadas; dir04 - identificar de forma preliminar os problemas de compatibilidade a serem tratados; dir05 - extrair os termos que formam o escopo de interesse.



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

Para cada uma das diretrizes identificamos um objetivo, uma série de ações para atingir este objetivo e os resultados esperados. Por motivos de espaço, vamos detalhar apenas a primeira diretriz, dir01, correspondente a “explicitar o compromisso ontológico das ontologias selecionadas”. As demais diretrizes estão sendo trabalhadas como parte de uma tese de doutorado, como explicado anteriormente. Essa diretriz, dir01, tem como **objetivo específico** compreender a estratégia de organização das classes das ontologias, suas naturezas e questões associadas. Para atingir este objetivo, um **conjunto de ações e resultados** é proposto, como mostra a Tabela 3.

Tabela 3 – Diretriz para explicitar o compromisso ontológico das ontologias selecionadas.

Ações Sugeridas e Resultados Esperados da Aplicação da Diretriz DIR01
1 - Buscar informações disponíveis sobre as ontologias. Sugerimos: - procurar no site da OBO e na literatura pela documentação das ontologias - de posse da documentação, buscar informações sobre padrões de nomenclatura, padrões de definição, granularidade, estratégia de organização das classes; Resultados: relação de aspectos conhecidos e não conhecidos sobre as ontologias;
2 – Definir critérios para conseguir as informações não obtidas em (1). Sugerimos: - analisar a definição/nomenclatura dos termos das classes de primeiro nível das ontologias e das suas subclasses (para hierarquias muito profundas limitar a análise a apenas um nível de profundidade) para identificar, quando possível, os princípios adotados para a formação das hierarquias (incluindo a granularidade), para a nomenclatura dos termos e para a sua definição. Resultados: relação de critérios para obtenção das informações não disponíveis sobre as ontologias (os critérios sugeridos podem ser adaptados ou substituídos por outros);
3 – Definir critérios para a identificação de problemas na organização das hierarquias. Sugerimos: - analisar as classes de primeiro nível das ontologias, adotando os seguintes cânones de Ranganathan: Diferenciação; Concomitância; Exclusividade; Modulação; para Classes Subordinadas. - analisar as classes de primeiro nível das ontologias, adotando os postulados 1, 3 e 4 de Guizzardi - verificar se existe definição de termo raiz incompatível com definição de termos subordinados. Resultados: relação de critérios para análise de problemas na organização das hierarquias ontologias (os critérios sugeridos podem ser adaptados ou substituídos).
4 – Aplicar os critérios definidos em 2 e 3 para obtenção das informações. Sugerimos: - utilizar apoio de editores de ontologias: Protégé se a ontologia estiver em OWL e OBO-Edit se estiver no formato .obo. Resultados: relação dos princípios de organização das classes de primeiro nível de cada ontologia; relação dos problemas encontrados em cada ontologia, de acordo com a aplicação dos critérios.

5. Aplicação da diretriz para explicitar o compromisso ontológico de ontologias

Tomamos como exemplo o reúso de um subconjunto de 18 termos da ontologia GO (ramo Molecular Function) utilizados para anotação de um experimento genômico (WAGNER, 2006), e da ontologia INOH (Molecule Role). As ontologias estão relacionadas ao tema de função molecular, embora com diferentes objetivos. Enquanto a primeira é voltada para a anotação funcional de genomas, a segunda é voltada para a anotação de vias biológicas (YAMAMOTO et al., 2004). Ao aplicar a **ação (1)** para buscar informações sobre as ontologias, começamos pelo levantamento da documentação disponível na literatura e na web, obtendo alguns aspectos conhecidos sobre as ontologias, dentre eles: (i) padrões de nomenclatura; (ii) nível de granularidade; (iii) padrões de definição.

A documentação encontrada sobre a GO revelou, dentre outras coisas, a existência de alguns padrões de nomenclatura e de definição, mas estes não se aplicam a todos os termos. Por exemplo: os termos do ramo função molecular da GO são, de modo

geral, formados pela adição do sufixo “activity” (com exceção da subhierarquia *binding*) ao produto gênico que é portador da função. Em relação à estruturação, até onde pudemos perceber, não existem princípios detalhados para a formação das hierarquias de função molecular. Existe, no entanto, uma orientação básica para que os termos sejam agrupados de acordo com a sua similaridade funcional, em vez de serem agrupados por pertencerem ao mesmo processo biológico. Dessa forma, as hierarquias devem conter cadeias de funções de mesma natureza e não um grupo de funções de natureza distinta que tenham como traço comum estarem envolvidas em um mesmo processo.

A documentação encontrada para a Molecule Role revelou que esta é uma ontologia de *nomes* de proteínas e de famílias de proteínas. A ontologia contém os nomes dos grupos funcionais de moléculas, os nomes *abstratos* de moléculas e os nomes *concretos* de moléculas, coletados manualmente na literatura. Esta abordagem de categorização, segundo os autores, é baseada em uma classificação de *papéis* (*roles*, em inglês) de moléculas em processos de interações entre proteínas e em processos de transdução de sinal, conforme observado na literatura científica da área.

A documentação geral sobre as ontologias GO e INOH possibilitou-nos um entendimento preliminar sobre as mesmas, porém não encontramos muitos detalhes sobre os critérios adotados para a organização das suas estruturas hierárquicas. Por esta razão, aplicamos a **ação (2)**, para buscar informações complementares. Para isto procedemos à análise da definição dos termos de primeiro nível das ontologias como forma de complementar a compreensão sobre a estruturação destas ontologias. Na INOH, entretanto, esta análise não pode ser aprofundada, uma vez que muito poucos termos possuíam definição. Como exemplo desta ação, tomamos o ramo *binding*, da GO, conforme Tabela 4 abaixo.

Tabela 4 – Resumo da análise do ramo Binding, da GO.

Identificação: GO:0005488 - Binding

Definição: A interação seletiva, não covalente, geralmente estequiométrica, de uma molécula com um ou mais sítios específicos em outra molécula (The selective, non-covalent, often stoichiometric, interaction of a molecule with one or more specific sites on another molecule).

Princípios observados para a organização do ramo binding: Elementos químicos (boron binding), compostos orgânicos (lipid binding) e inorgânicos (nitric oxide binding), papéis de moléculas (antigen binding, neurotransmitter binding), localizações celulares (cell surface binding), organismos (bacterial binding).

Como perspectiva para analisar os problemas encontrados nas hierarquias, aplicamos a **ação (3)** e optamos por adotar os cânones de Ranganathan, explicados no item 2.3.

Após a definição dos critérios (ações 2 e 3) aplicarmos a **ação (4)**. Como resultado, observamos a violação de alguns desses cânones, como, por exemplo, a do cânones para classes subordinadas, uma vez que a definição de *binding* indica que o critério de divisão tem a ver com interação entre moléculas, o que coloca termos como “bacterial binding” e “extracellular matrix binding” fora de contexto, por que uma bactéria (bacterial) e uma matrix extracelular (extracellular matrix) possuem natureza distinta da natureza de uma molécula. A tabulação dos resultados gerais aponta para a pertinência e atualidade do trabalho de Ranganathan aplicado à organização taxonômica das ontologias, conforme podemos perceber pelas violações de seus cânones, observadas ao longo de nosso trabalho. Em especial, destacamos os cânones: exclusividade e classes subordinadas, com maior número de ocorrências (8); modulação (7); diferenciação (4) e, por fim, o cânone da concomitância (1).

Ainda, ao considerarmos os princípios de organização, observamos que estes apontam para possíveis classes de conceitos que poderiam ser utilizados para estender a GO, ou ainda como forma de padronizar as definições de seus termos, de modo a favorecer o uso mais preciso de ferramentas de software, que poderiam manipular as definições de forma mais estruturada. Estes princípios são: (i) Tipo de reação bioquímica; (ii) Elementos químicos; (iii) Compostos orgânicos; (iv) Tipos de papéis de moléculas (antigen, neurotransmitter); (v) Tipos de localizações celulares; (vi) Tipos de enzimas (ex: hydrolase, isomerase); (vii) Tipo de atividade catalisada pela enzima (ex: catalysis of free radical formation, first spliceosomal transesterification activity); (viii) Tipo de canal iônico (ion channel), de acordo com o elemento químico que identifica o ion (ex: cálcio, potássio); (ix) Tipo de regulação (ex: inibição ou ativação) do canal iônico; (x) tipo de transferência de elétrons: “de” um local ou “dentro de” um local (“from” ou “within”); (xi) Tipo de sistema de transporte de elétrons; (xii) Tipo de regulação da enzima; (xiii) Tipo de íon de metal; (xiv) Tipo de entrada (input), que pode ser energia (energy) ou um sinal (signal); (xv) Tipo de regulação de proteasome; (xvi)

Tipo de elemento (de diferentes granularidades; por exemplo, osso e ribossomo) do qual a molécula é componente estrutural; (xvii) O modo pelo qual (como) se dá a regulação da atividade de transcrição; (xviii) O modo pelo qual (como) se dá a regulação da atividade de transcrição e o efeito dessa regulação; (xix) A fase da transcrição (por exemplo, se é de iniciação ou de terminação); (xx) Forma pela qual se dá a atividade de translação; (xxi) Substâncias (podendo ser orgânicas ou inorgânicas).

Ao analisarmos os descritores da INOH (Molecule Role) observamos que o ramo “chemical” pode ser útil para descrever os itens: (ii), (iii), (viii) e (xiii) e (xxi); o ramo “protein” pode ser útil para descrever os itens (iv); o ramo “protein/enzyme” pode ser útil para descrever os itens (vi).

Ao analisarmos a GO (celular component) observamos que o ramo “celular component” pode ser útil para descrever os itens (v) e (xvi); outros ramos da GO poderiam ser úteis, mas devido à complexidade da GO não vamos explorá-los aqui.

De posse das classes de conceitos que podem ser utilizadas para estender a GO, podemos exemplificar uma possível forma de fazer essa extensão, como explicado a seguir.

5.1 Fundamentação da extensão do recorte de origem da GO

A análise das classes de primeiro nível da GO permitiu, dentre outras coisas, obter indícios de possíveis classes de conceitos que poderiam ser utilizados para estender a GO. Em contrapartida, esperamos que o entendimento das classes do recorte de origem de acordo com a noção de função da BFO, permita evidenciar quais destas classes de conceitos vamos adotar para complementar o escopo do recorte inicial. Desta forma, propomos proceder com uma ação que visa explicitar o recorte de origem de acordo com a noção da temática alvo, no caso, a de função molecular. Desta forma, podemos selecionar de forma mais fundamentada, as classes escolhidas para estender a GO, de acordo com nossos propósitos.

Para tal, partimos do descritor e da definição de cada termo e ainda referências, na GO e na INOH, a identificadores de enzimas (EC number) nos metadados do tipo dbxref, buscando identificar: portador da função, função executada, processo onde ela se manifesta. O processo poderia ser entendido como o *processo biológico* (presente no

ramo de processo biológico da GO) ou como a *via* biológica (presente em ontologias com a pathway). O processo biológico pode conter mais de uma via biológica. Para descobrirmos o processo biológico, recorremos à anotação presente no recorte de origem, em conjunto com a de função molecular. Para descobrirmos a via biológica, partimos da busca dos identificadores de enzima (quando presentes) em bancos de enzimas tais como o BRENDA (SCHOMBURG et al., 2004). Estes bancos contêm referências a vias metabólicas associadas às enzimas nele presentes.

Cabe ressaltar que além do portador e do processo, os elementos do contexto biológico da função (moléculas com as quais o portador interage, o local onde a função se reliza, etc) também poderiam ser explicitados. Por exemplo, uma análise mais criteriosa das subclasses do termo *binding* revela uma diversidade de tipos de moléculas, as quais interagem com os portadores das funções moleculares em questão, e que poderiam ser usadas em anotações genômicas com a própria GO ou com outras ontologias, se disponíveis, tais como: elementos químicos (boron binding), compostos orgânicos (lipid binding), compostos inorgânicos (nitric oxide binding), componentes celulares (cell surface binding), dentre outros. Por motivos de tempo, nosso experimento se limita a identificar, na amostra selecionada, quando possível, o **portador da função** e um **processo onde esta pode se manifestar**, sabendo de antemão que é possível que a função se manifeste em mais de um processo.

A seguir ilustramos um exemplo do detalhamento da análise realizada, feito com o termo GO:0008026 - ATP-dependent helicase activity, como mostra a Tabela 5.

Tabela 5 – Explicitação da definição do termo ATP-dependent helicase activity, da GO.

Identificação: GO:0008026 - ATP-dependent helicase activity

Definição: Catalise da reação ATP + H₂O = ADP + phosphate para guiar o desdobramento de uma hélice de DNA ou RNA

Função: Catalisar a reação: ATP + H₂O = ADP + phosphate; **Portador:** helicase (enzima)

Processo (via biológica – obtida a partir de metadados das ontologias): Purine metabolism (só encontrado para termos subordinados) -> PW:0000031 (ontologia pathway) = purine metabolic pathway

A partir do detalhamento pudemos observar que **portadores** de funções moleculares são geralmente enzimas. Este tipo de classe, como vimos, foi constatada como possível candidata para extensão da ontologia GO, estando presente no ramo

protein/enzyme da INOH. Outras classes, tais como as de vias biológicas, que são **processos** onde se manifesta a função molecular, poderiam ser utilizadas para estender a GO, a partir de ontologias como a Pathway (KUSHIDA, TAHAGI, FUKUDA, 2006). Poderíamos ainda desmembrar a **função** propriamente dita, de modo a explicitar a reação química catalisada, mas ontologias de reações químicas adequadas seriam muito complexas e não se encontram disponíveis no momento (SMITH, KUMAR, BITTNER, 2005).

6. Considerações Finais

Até onde pudemos perceber, as metodologias para reúso carecem, de modo geral, de algum detalhamento sobre as etapas e ações a serem tomadas para a identificação do compromisso ontológico do domínio considerado e a reorganização das ontologias reutilizadas de acordo com uma visão de alto nível, como, por exemplo, a BFO. Em nossa proposta, procuramos fornecer uma série de diretrizes voltadas para preencher esta lacuna e detalhamos uma delas.

Como resultado da aplicação das ações da diretriz dir01, observamos princípios organizadores do recorte alvo na GO. Estes apontam para possíveis classes de conceitos que poderiam ser utilizados para estender a GO, ou ainda como forma de padronizar as definições de seus termos, de modo a favorecer o uso mais preciso de ferramentas de software, que poderiam manipular as definições de forma mais estruturada (CAMPOS et al., 2010). Exemplos destes princípios são tipos de: (i) elementos químicos; (ii) compostos orgânicos; (iii) localizações celulares; (iv) enzimas (ex: hydrolase, isomerase), dentre outros. Ao analisarmos os descritores da INOH (Molecule Role), por exemplo, observamos que o ramo “chemical” pode ser útil para descrever os itens: (i) e (ii); o ramo “protein/enzyme” pode ser útil para descrever os itens (iv). Este último, em especial, se aplica a descrever os portadores das funções presentes no nosso recorte de origem, como, por exemplo, a enzima *helicase*. Essas informações foram possíveis a partir da explicitação do compromisso ontológico das ontologias envolvidas, em especial ao recorte de origem da GO, que possui definição para a maioria dos seus termos. Os resultados obtidos apontam para o fato de que o conjunto de diretrizes propostas fornece apoio para guiar a definição de domínio de maneira mais precisa, através do resgate do

compromisso ontológico e das perspectivas que podem ser adotadas para a complementação do domínio alvo.

ABSTRACT

Ontology reuse methodologies still lack improvement in their directions to help identify ontological commitment in domain definition and also in reorganizing ontologies according to high-level perspectives, using, for instance, foundational ontologies like BFO. In our proposal, we aim to provide a series of guidelines to help fulfill this gap. They are based on works from Information Science, in authors like Neville, Dahlberg and Ranganathan; and on works from Computer Science, in authors like Guarino and Welty and Guizzardi, in an interdisciplinary perspective. In this paper we discuss one of such guidelines, aimed at understanding the organizational strategy of ontology classes, their natures and organizational problems. The guideline was applied on a previously chosen set of terms from Gene Ontology (GO), retrieving aspect of its ontological commitment. As result, we get possible classes of concepts that may be used to extend GO, defining the studied domain in a more precise way. Results also points to the possibility of standardize terms definitions, helping the application of software tools in manipulating term definitions in a more structured and precise way. From the results obtained, we conclude that the proposed set of guidelines provide support to guide domain definition in a more precise way, by the retrieval of ontological commitment and the perspectives that can be adopted to extend the target domain.

Keywords: Ontology. Ontology reuse. methodological guidelines. ontological commitment. domain definition.

Referências

ARP R, SMITH B. **Function, role, and disposition in Basic Formal Ontology**. In: PROCEEDINGS OF BIO-ONTOLOGIES WORKSHOP, Toronto, p. 45-48, 2008.

BOECKMANN B, BLATTER MC, FAMIGLIETTI L, HINZ U, LANE L, ROECHERT B, BAIROCH A: **Protein variety and functional diversity**: Swiss-Prot annotation in its biological context. In: C R Biologies, v.328, p. 882-899, 2005.

CAMPOS, L.M. **Metodologia para definição de domínio no reuso de ontologias biomédicas**. Proposta de qualificação de doutorado, UFF, 2010.

CAMPOS, M.L.A. **Perspectivas para o Estudo da Área de Representação da Informação**, Ciência da Informação, v. 25, n. 2, 1995.



CAMPOS, M.L.A. **A Organização de Unidades do Conhecimento em**

Hiperdocumentos: o modelo conceitual como um espaço comunicacional para a realização da autoria. 2001, 198p. Tese de Doutorado – UFRJ, IBICT, Rio de Janeiro, 2001a.

CAMPOS, L.M.; PRAZERES, M.G.C.; CAMPOS, M.L.A.; BRAGANHOLO, V.; CAMPOS, M.L.M. **Exploring GO term definitions to enhance automatic annotation of molecular function.** In: International Workshop on Genomic Databases, Rio de Janeiro, Brazil, 2010.

CAMPOS, M. L. A. ; GOMES, H. E. **Taxonomia e Classificação:** princípios de categorização. Datagramazero (Rio de Janeiro), v. 9, n. 1, 2008.

CAMPOS, M. L. C. e GOMES, H.E. **Organização de domínios de conhecimento e os princípios ranganathianos.** Perspectivas em Ciência da Informação. Belo Horizonte, UFMG, v.8, n.2, jul./dez. 2003.

DAHLBERG, I. **Teoria do conceito.** Ciência da Informação, v. 7, n. 2, p. 101-07, 1978a.

DAHLBERG, I. **A referent-oriented, analytical concept theory of Interconcept.** International Classification, v. 5, n. 3, p. 122-151, 1978b.

DAHLBERG, I. **Conceptual compatibility of ordering systems.** International Classification, v. 10, n. 2, p.5-8, 1983.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. **METHONTOLOGY:** From Ontological Art Towards Ontological Engineering. Spring Symposium on Ontological Engineering of AAAI. Stanford University, California, p 33–40, 1997.

GANGEMI, A., STEVE, G., and GIANCOMELLI, F. , **ONIONS:** An Ontological Methodology for Taxonomic Knowledge Integration, ECAI-96 Workshop on Ontological Engineering, 1996.

GENE ONTOLOGY CONSORTIUM. **Gene Ontology:** tool for the unification of biology. Nature Genetics, n. 25, p. 25-29, 2000.

GRENON, P.; SMITH, B.; GOLDBERG, L. **Biodynamic Ontology:** Applying BFO in the Biomedical Domain, In: PROCEEDINGS OF THE WORKSHOP ON MEDICAL ONTOLOGIES, Rome, Italy, 2003.

GUARINO, N. **Some Ontological Principles for Designing Upper Level Lexical Resources,** In: First International Conference on Language Resources and Evaluation, Granada, Spain, May, 1998.

GUARINO, N.; CARRARA, M.; GIARETTA, P. **Formalizing ontological commitments,** Proceedings of the 12th National Conference On Artificial Intelligence, v.1, p.560-567, 1994.

GUARINO, N. e WELTY, C.. **Identity, Unity, and Individuality:** Towards a Formal Toolkit for Ontological Analysis. In Proceedings of ECAI-2000: The European Conference on Artificial Intelligence. Berlin, Germany, IOS Press, 2000.

GUARINO, N.; WELTY, C. **Towards a methodology for ontology based model engineering.** In Proceedings of the ECOOP-2000 Workshop on Model Engineering, 2000b.



GUARINO N., WELTY, C., **Identity and Subsumption**, LADSEB CNR internal report 01/2001, 2001.

GUIZZARDI, G., **Ontological Foundations for Structural Conceptual Models**, Tese de Doutorado em Ciência da Computação, Universidade de Twente, Enschede, Holanda, ISBN 90-75176-81-3, 416 p., 2005.

GUIZZARDI, G., **Desenvolvimento para e com Reuso**: Um Estudo de Caso com um Sistema de Vídeo sob Demanda. Dissertação de mestrado. PPGI/UFES, 2000.

GUIZZARDI, G.; WAGNER, G; SINDEREN van, M. **A formal theory of conceptual modeling universals**. In: Workshop on Philosophy and Informatics, Cologne, Germany, 2004.

OBO. **Open Biomedical Ontologies**. Disponível em: <http://obo.sourceforge.net>. Acesso em 24 mai. 2009.

PÂSLARU-BONTAS. **A Contextual Approach to Ontology Reuse**: Methodology, Methods, and Tools for the Semantic Web. PhD Dissertation, Freie Universitaet, Berlin, 2007.

PEREZ-REY, D; MAJO, V; GARCIA-REMESAL, M; ALONSO-CALVO, R.

Biomedical Ontologies. In: Post-Genomic Information Systems, Proceedings of the 4th IEEE Symposium on Bioinformatics and Bioengineering, p.207, 2004.

PINTO, S, MARTINS, J. P. , **A Methodology for Ontology Integration**. In: Proceedings of First International Conference on Knowledge Capture, Victoria, B.C., Canada, 2001.

PLANT, R.; GAMBLE, R. Methodologies for the Development of Knowledge-Based Systems: 1982-2002, **Knowledge Engineering Review**, v.18., n. 1, p. 47-81, 2003.

RANGANATHAN, S. R. **Prolegomena to Library Classification**. New York : Asia Publishing House, 1967.

S. NURCAN, J. BARRIOS, G. GROSZ, C. ROLLAND. **Change process modelling using the EKD - Change Management Method**. Proceedings of the 7th European Conference on Information Systems, ECIS'99, Copenhagen, Denmark, June 23-25, p.513-529, 1999.

SCHOMBURG, I; CHANG, A; EBELING, C; GERMSE, M; HELDT, C; HUHN, G; SCHOMBURG, D. **BRENDA, the enzyme database**: updates and major new developments. Nucleic Acids Research, n. 32 (Database issue), 2004..

SMITH, B.; GRENON, P. **The cornucopia of formal-ontological relations**. Dialectica, v.58, n.3, p.279–296, 2004.

SMITH, B.; WILLIAMS, J.; SCHULZE-KREMER, S. **The ontology of gene ontology**. In AMIA Annual Symposium Proceedings, 2003.

SMITH, B; KUMAR, A; BITTNER, T. **Basic Formal Ontology for Bioinformatics**. Journal of Information Systems, 2005.

SURE, Y; STAAB, S; STUDER, R. **Methodology for development and employment of ontology based knowledge management applications**. SIGMOD Record, v. 31, 2002.



XI Encontro Nacional de Pesquisa em Ciência da Informação
Inovação e inclusão social: questões contemporâneas da informação
Rio de Janeiro, 25 a 28 de outubro de 2010

USCHOLD, M.; KING, M. **Towards a methodology for building Ontologies**. In: Proceedings of the IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, 1996.

KUSHIDA, T.; TAHAGI, T.; FUKUDA, K.I. **Event Ontology**: a pathway-centric ontology for biological processes. The Pacific Symposium on Biocomputing (PSB), 2006.

WAGNER, G. **Geração e análise comparativa de seqüências genômicas de Trypanosoma rangeli**. Dissertação de Mestrado, Instituto Oswaldo Cruz, Rio Janeiro, 2006.

YAMAMOTO, S., ASANUMA, T., TAKAGI, T., AND FUKUDA, K., **The molecule role ontology**: An ontology for annotation of signal transduction pathway molecules in the scientific literature, Comparative and Functional