

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/250991526>

Ontologias de domínio: um estudo das relações conceituais

Article in *Perspectivas em Ciência da Informação* · August 2008

DOI: 10.1590/S1413-99362008000200006

CITATIONS

18

READS

67

3 authors, including:



Luana Sales

IBICT

80 PUBLICATIONS 439 CITATIONS

[SEE PROFILE](#)



Maria Luiza de Almeida Campos

Universidade Federal Fluminense

38 PUBLICATIONS 276 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Gestão de dados de pesquisa FAIR: uma proposta de modelo para aceleração das pesquisas científicas no estado do Rio de Janeiro [View project](#)



Sistema de Informação sobre a pesquisa corrente: um CRIS para a área Nuclear [View project](#)

Ontologias de domínio: um estudo das relações conceituais

Luana Farias Sales

**Mestre em Ciência da Informação;
Analista em C & T da CNEN/IENT**

Maria Luiza de Almeida Campos

**Doutora em Ciência da Informação;
Professora Adjunta do Departamento
de Ciência da Informação da UFF e do
Programa de Pós-Graduação em Ciência
da Informação UFF/IBICT**

Hagar Espanha Gomes

Livre Docente; Consultora

No âmbito da elaboração de Ontologias, uma problemática que tem se colocado é a ausência de um padrão teórico-metodológico. Propõe-se então uma sistematização das relações apresentadas nas literaturas da Ciência da Informação, da Terminologia, da Ciência da Computação. A sistematização realizada busca unir as duplas de categorias (relações categoriais) com as relações potenciais entre elas (relações formais). A partir da seleção de um corpus da Gene Ontology, foi feita uma análise das definições a fim de identificar as relações já mencionadas na literatura ou descobrir a existência de novas relações. Conclui-se que relações diferentes podem surgir em diferentes domínios e que definições sistematizadas são indispensáveis para o estabelecimento seguro de relações formais.

Palavras-chave: *Relações Conceituais; Terminologia, Ontologia; Organização do Conhecimento*

Domain Ontologies: study of conceptual relationships

Lack of theoretical bases and sound methodologies are problems that arise when building Ontologies. The study proposes systematization of relations found in the literature from Information Science, Terminology, Computer Science and Bioinformatics. The systematization proposed aims at linking pairs of categories (categorical relations) with potential relations between them (formal relations). The study shows that different relations may arise in different domains and that systematic definitions are fundamental for the establishment of formal relations.

Keywords: *Conceptual relationship; Terminology, Ontologies, Knowledge Organization.*

Recebido em 03.01.2008 Aceito em 18.06.2008

1 Considerações iniciais

Quando a *Web* foi criada, não se poderia imaginar o tamanho de sua repercussão. O mecanismo que havia sido proposto apenas para troca de informações entre pesquisadores tomou uma proporção tão grande, que atingiu a vida dos mais diversos cidadãos. Com isso, os próprios criadores da *Web* atualmente se questionam e procuram novas ferramentas que possibilitem uma recuperação de informação mais precisa.

Tendo em vista que quanto maior o número de informações publicadas na *Web* maior a dificuldade em acessá-las, e também que quanto maior a utilização desse mecanismo, maior a necessidade de se obter a informação desejada em menor espaço de tempo, Tim Berners-Lee se juntou a outros pesquisadores, em 1994, para a criação de um consórcio Internacional chamado W3C - *World Wide Web Consortium*.

O W3C projeta uma *Web* ideal, em que a recuperação de informação poderá ser feita de forma objetiva e consensual e, mais que isso, poderá oferecer diretamente, através de um processamento inteligente, respostas para muitas questões. Essa *Web* ideal chama-se *Web Semântica*.

Para que esse processamento inteligente possa acontecer, é preciso, entre outras coisas, um instrumento de padronização terminológica, que o W3C vem chamando de Ontologias.

As Ontologias são formadas por termos, definições e relações. Devido a estes elementos, a literatura vem definindo ontologia como uma linguagem documentária. Entretanto, apesar de possuir elementos comuns, as ontologias são mais que linguagens documentárias: elas

possuem funcionalidades que permitem que a máquina possa processar o raciocínio automatizado, através de regras e inferências.

No que diz respeito às ontologias, o uso de relações pode ser justificado através da afirmação de Motta (1987), em seu livro "Método relacional"¹ como nova abordagem para construção de tesauros", de que o estabelecimento de relações conceituais pode ser útil em circunstâncias como: "determinação de sinônimos, atualização do sistema, inclusão de termos em uma só categoria e mapeamento de áreas de assunto" (MOTTA, 1987, p. 61). Apesar de esta afirmação estar relacionada à construção de tesauros, pode ser estendida também à construção de ontologias.

Outros autores como Kietz, Maedache e Volz (2000) afirmam que a determinação de relações conceituais consistentes é útil à elaboração de ontologias, na medida em que estas relações garantem a consistência na adoção de algoritmos baseados em regras de associação. Já Uschold e King (1995), Fernandez, Gomez-Pérez e Juristo (1997) relacionam a importância do uso de relações conceituais bem estabelecidas para a consistência da estrutura terminológica e para a elaboração da taxonomia que deve compor a estrutura da Ontologia (GUARINO, 1995; SURE; STAAB; STUDER, 2002).

O modelo de relações que aparece em Vocabulários Controlados e em Tesauros é um modelo diádico que revela apenas as categorias às quais os conceitos pertencem, por ex: Coisa-propriedade; Material-produto; Processo-resultado; etc. Nas ontologias, as relações precisam ser explicitadas, pois integram uma proposição.

Um importante elemento das ontologias é a representação do conhecimento, que envolve análise semântica, em especial para a organização de sistemas de conhecimento que, nas ontologias, visam otimizar a recuperação, ou seja, prover o computador de mais inteligência. Trata-se de uma área de pesquisa estreitamente ligada à organização do conhecimento, em especial no que tange à rede semântica. Esta representa o conhecimento na forma de nós (conceitos, atributos) ligados por arcos (relacionamentos). No que tange os nós, a terminologia e a teoria do conceito têm sua contribuição a dar, mas disto não nos ocupamos no presente trabalho, privilegiando apenas os relacionamentos semânticos. A literatura tem registrado inúmeros trabalhos sobre esta questão, abordando, no entanto, o processamento automático da linguagem natural. Propostas de estruturas conceituais a partir de análises conceituais têm sido objeto de pesquisas nas áreas de bioinformática, visando a representação dos objetos (em especial suas partes, relação com objetos e seus atributos, relação com outros objetos). Nas ontologias, o conjunto de relações é mais rico do que nas tabelas de classificação bibliográfica ou nos tesauros, o que permite maior estrutura representativa do conhecimento registrado em um discurso, para que

¹ Método analítico que pretende ser instrumento para o estabelecimento de relações em tesauro de forma mais objetiva [...] Consiste na análise das definições dos conceitos que integrarão o sistema, de forma a identificar suas características e na estruturação desses conceitos, tendo em vista as relações existentes entre eles (MOTTA, 1987, p. 39).

possa ser manipulado pelo computador. É importante ressaltar que tais representações são igualmente relevantes na formulação de buscas nas ontologias.

Vickery(1997) ressalta, em seu estudo pioneiro, que o problema da representação é algo de que a Biblioteconomia/Ciência da Informação de há muito se ocupa.

O presente trabalho vem apresentar uma sistematização das relações abordadas na literatura das áreas de Ciência da Informação, Terminologia e Ciência da Computação, e sua aplicação numa amostra da *Gene Ontology*.

A *Gene Ontology* é uma ontologia de domínio, formada por três categorias de conceitos, a saber: Função Molecular, Processo Biológico e Componente Celular. Ela tem por objetivo produzir um vocabulário controlado que possa ser aplicado a todos os organismos, para representação do conhecimento na descrição de genes e papéis de proteínas em células. Faz parte do consórcio internacional OBO - Open Biomedical Ontologies. Esta fonte foi selecionada por sua relevância para as atividades de pesquisa ora em desenvolvimento junto ao Laboratório de Biologia Molecular de Tripanosomatídeo e Flebotomíneos da FIOCRUZ².

2 Teorias para o estabelecimento de relações

O estudo das relações conceituais aqui discutidas é fruto de levantamento realizado na literatura, a partir de revisão de estudos teóricos e de comentadores. Desta forma, a data de publicação passa a ter menor relevância, ou seja, o foco está nos formuladores das teorias, embora se possa adicionar, conforme o caso, comentários a esses autores, estes, sim, mais recentes.

As relações entre conceitos apresentam interesse para a Ciência da Informação, para a Terminologia e para a Ciência da Computação. No entanto, as abordagens para seu estabelecimento diferem, tendo em vista o desenvolvimento de instrumentos distintos. Em todos eles, as relações conceituais fazem parte de sua essência. Sendo assim, é preciso ter bem definidos, no momento da elaboração, os objetivos que se quer atingir através do uso de tal linguagem, pois esses objetivos também poderão influenciar no uso de determinadas relações em detrimento de outras. Os objetivos específicos desses instrumentos podem sempre variar de acordo com as expectativas dos usuários e com as metas do sistema que os utilizarão.

Segundo a ISO 704 (INTERNATIONAL STANDARD ORGANIZATION – ISO, 2000, p. 5):

Na organização do conceito em um sistema conceitual, é necessário trazer à mente o campo do conhecimento que

² As atividades que vêm sendo desenvolvidas pelo Grupo de Pesquisa em Ontologia e Taxonomia, junto ao Laboratório, fazem parte do projeto de pesquisa "Integração de ontologias: o domínio da bioinformática e a problemática da compatibilização terminológica", financiado pelo CNPq.

originou os conceitos e também considerar as expectativas dos usuários e as metas do sistema que utilizarão, pois seus objetivos específicos podem sempre variar de acordo com a área do conhecimento que se visa representar.

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

Esta classificação diferencia as relações que podem ser visualizadas na parte sistemática das linguagens, isto é, as de gênero-espécie e parte-todo, em um Tesouro, das relações que não podem ser visualizadas por falta de uma forma de representação, mas que ainda assim existem, como é o caso das relações associativas (não-hierárquicas).

Na Terminologia, a ênfase é dada ao desenvolvimento de glossários, que são instrumentos de padronização de termos de uma dada especialidade ou domínio. O objetivo destes instrumentos é melhorar a comunicação na comunidade científica. Neste caso, a utilização de relações pode ser importante para apontar outros termos relacionados ao conceito.

A Teoria da Terminologia enfatiza o papel das características para a formação dos conceitos, as quais são fixadas na definição. É no momento da atribuição das características que se observam as relações entre os conceitos, e é neste momento que se aplicam os princípios para estabelecer relacionamentos.

Diferentemente da Ciência da Informação, a Teoria Geral da Terminologia (WUSTER, 1981) classifica as relações em lógicas e ontológicas. Nestas últimas, se incluem as partitivas e as associativas, e Wüster as apresenta detalhadamente.

As Relações de Contato são consideradas como a subcategoria mais importante, e se auto-explicam a partir de suas espécies, que são as Relações de Coordenação e as Relações de Encadeamento. A principal Relação de Coordenação é a relação de parte-todo. Essa relação pode ocorrer entre o todo e suas partes e entre as próprias partes, sendo considerada uma relação espacial e conseqüentemente uma relação de simultaneidade. Outras Relações de Coordenação mencionadas por Wüster são: a Relação de Inclusão e a Relação de Integração. As Relações de Encadeamento são relações temporais e se subdividem em Relação de Antecessão e Relação de Sucessão.

As Relações de Causalidade são relações de parentesco e são subdivididas em: Relações entre Gerações, que devem expressar a relação entre duas gerações diferentes e podem ser ascendentes e descendentes; e Relações entre Estágios, que devem expressar relações entre estágios de evolução de um mesmo e único indivíduo ou de uma substância. Sendo assim, as relações entre estágios podem ser: Relações Filogênicas, Relações Ontogênicas e Relações entre Substâncias, que são apresentadas no quadro de Wüster, porém não são definidas em seu texto.

Sager (1990) propõe uma outra classificação para as relações entre conceitos, a saber: Relacionamentos Genéricos, Relacionamentos Partitivos, Relacionamentos Polivalentes e Relacionamentos Complexos.

Os Relacionamentos Genéricos correspondem às relações lógicas. Os Relacionamentos Partitivos correspondem aos relacionamentos de parte-todo. Já os Relacionamentos Polivalentes correspondem às relações polihierárquicas, mencionadas pela ISO 2788 (ISSO, 1986) e por Aitchison (1987) na categoria relações hierárquicas. Wüster (1981) considera essa relação como lógica ou, mais especificamente, como relação de combinação lógica.

Os Relacionamentos Complexos correspondem a todas as relações associativas na Ciência da Informação e a um tipo de relação Ontológica na Terminologia, representada, por exemplo, pelas relações de causalidade já citadas. Apesar de aparecer em grande parte da literatura da Ciência da Informação e da Terminologia, esse tipo de relação possui uma característica distinta em Sager, que a torna especial para o estudo das relações conceituais em Ontologias. Ela é dividida em Relações Complexas Categoriais e Relações Complexas formais. Esta tipologia de relações ainda não havia aparecido na literatura da área de Terminologia e Ciência da Informação, mas já havia sido apresentada na literatura da Ciência da Computação, como, por exemplo, em Sowa (2000), pode-se acreditar que esta divisão das relações seja de grande serventia para o processamento das relações conceituais pela máquina. Sendo assim, Sager pode ser considerado um autor importante para esses estudos, pois consegue observar que as relações podem ocorrer em duas modalidades.

Como se observa, a literatura da Ciência da Informação apresenta apenas relações entre categorias – Relações Diádicas, mas as ontologias necessitam de relações mais formalizadas – Relações Triádicas.

Na Ciência da Computação, a ênfase é dada ao desenvolvimento de Ontologias. A necessidade de uma linguagem formal para o processamento da máquina faz com que as relações devam ser explicitadas, por exemplo, em forma de verbos (Ex: *affects*), ou verbos preposicionados (Ex: *interacts_with*). Portanto, além de identificar os elementos de relação, é necessário explicitar a própria relação.

As relações complexas mencionadas por Sager (1990) são um reflexo das relações que aparecem na literatura da Ciência da Computação com o nome de função.

No que diz respeito ao estabelecimento de relações no âmbito de Ontologias, considera-se importante a sistematização tanto das relações categoriais como das relações formais.

As relações nas Ontologias se apresentam de duas formas: na terminologia - sistema de termos com definições em linguagem natural; e na parte formal - expressas através de linguagem própria para entendimento da máquina.

Na parte terminológica, o uso das relações categoriais, como apresentadas na literatura da Ciência da Informação e da Terminologia, seria de grande benefício, pois ligaria as classes de conceitos.

Já na parte formal, não basta ligar as classes, mas é preciso também dizer a forma como se dá essa ligação, o que pode ser expresso através das relações sugeridas pela Ciência da Computação, que chamamos aqui de relações formais, já que expressam a forma como as categorias se relacionam. Com as relações estabelecidas nas duas modalidades de forma consistente, podem-se construir definições formais que possibilitem o processamento pela máquina.

Assim, as três áreas podem contribuir de alguma forma.

3 Sistematizando as relações conceituais

Uma revisão da literatura da Ciência da Informação e da Ciência da Computação sobre relações conceituais (SALES, 2006) mostrou como as relações conceituais se apresentavam.

O estudo permitiu classificar as relações para ontologias, primeiramente em: Relações Categoriais e Relações Formais.

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

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

3.1 Relações genéricas

As relações genéricas se dão a partir do princípio lógico de abstração e, desta forma, respondem a pergunta "é um?" Assim, ocorrem sempre no interior de uma mesma categoria. Esse tipo de relação não apresenta grandes problemas, pois não é mutável; ao contrário, se manifestam sempre da mesma forma. Para cada par de relação categorial foi encontrada uma relação formal³.

³ Apesar de, na maioria dos casos, as relações se apresentarem em inglês, elas estão traduzidas da mesma forma como as relações categoriais (Seção 3.3)

QUADRO 1 – Sistematização das relações genéricas

Relações Categoriais Genéricas	Relações Formais Genéricas
Gênero-espécie	Tem-tipo
Espécie-gênero	É_um
Espécie-instância	É_instanciado_por
Instância-espécie	É_instância_de

Fonte: Dados da pesquisa.

3.2 Relações Partitivas

As relações partitivas são aquelas que revelam a relação entre o todo e suas partes. Essas relações se manifestam diferentemente de acordo com a natureza do conceito. Ou seja, quando o conceito é classificado como uma personalidade, que é material, o seu todo e parte são sempre um objeto concreto. Entretanto, quando ele se manifesta como um processo, é um objeto ocorrente ou uma etapa, um estágio, uma fase (SOWA, 2000; SMITH, 2005). Além disso, é importante destacar que elas se dão sempre no interior de uma mesma categoria.

Sobre esse tipo de relação, pode ser observado também que a literatura já evidencia as duplas de ocorrências, que estávamos chamando até agora de relações categoriais, já que esta dupla evidencia categorias de conceitos, porém não apresenta uma sistematização das possibilidades de relações entre essas duplas.

A sistematização das relações partitivas foi realizada da seguinte forma: uma vez identificadas as relações formais partitivas (coluna 2), procedeu-se à sua caracterização (coluna 1). Para cada uma delas foi criada uma relação inversa, mesmo quando não mencionada na literatura, de modo a apresentar uma caracterização mais completa.

Chegamos ao QUADRO 2 abaixo:

QUADRO 2 – Sistematização das relações partitivas

Relações Categroriais Partitivas	Relações Formais Partitivas
• Objeto-constituente • Constituinte-objeto	tem_constituente constituente_tem_todo
• Objeto-unidade • Unidade-objeto	tem_unidade unidade_tem_todo
• Coleção-elemento • Elemento-coleção	tem_elemento elemento_tem_coleção
• Massa-porção • Porção-massa	tem_porção porção_tem_massa
• Objeto-material • Material-objeto	tem-material material_tem_objeto
• Área-lugar • Lugar-área	tem_lugar lugar_está_em
• Zona-região • Região-zona	tem_lugar lugar_está_em
• Caráter-atividade • Atividade-caráter	tem_etapa tem_processo
• Processo contínuo - ato • Ato - processo contínuo	tem_etapa tem_processo
• Processo descontínuo-fase • Fase-processo descontínuo	tem_etapa tem_processo
• Grupo-membro • Membro-grupo	tem_membro membro_tem
• Classe-membro • Membro-classe	tem_membro membro_tem
• Substância-particular • Particular-substância	tem_substância substância_tem
• Conjunto-subconjunto • Subconjunto-conjunto	tem_subconjunto subconjunto_tem
• Conjunto-elemento • Elemento-conjunto	tem_elemento tem_conjunto
• Todo/sub-parte temporal • Sub parte temporal - todo	tem_subparte tem_todo
• Ciência - objeto de estudo • Objeto de estudo - ciência	tem_ramo ramo_tem

Fonte: Dados da pesquisa.

Observa-se que, nas relações em inglês, houve a preocupação em empregar o verbo *ter*, o que foi mantido na tradução, sendo o verbo *ser* empregado na relação de subordinação (*é um*). A padronização destas expressões merece estudo posterior.

3.3 Relações funcionais

As relações funcionais são aquelas que revelam a relação de objeto com o mundo ou a função de um objeto em um determinado contexto. Essas relações podem se manifestar entre categorias e intra-categorias.

O ponto de partida para a sistematização das relações funcionais foram as categorias rangianthianas PMEST, (personalidade, matéria, energia/processo, espaço e tempo). Foram obtidas as seguintes relações:

Personalidade-Personalidade
Personalidade – Matéria
Personalidade-Processo

Personalidade-Espaço
Personalidade-Tempo

Matéria-Personalidade
Matéria-Processo
Matéria-Matéria
Matéria-Espaço
Matéria-Tempo

Processo-Personalidade
Processo-Matéria
Processo-Processo
Processo-Espaço
Processo-Tempo

Espaço-Personalidade
Espaço-Medida
Espaço-Processo
Espaço-Espaço
Espaço-Tempo

Na medida em que a classificação foi sendo realizada, sentiu-se a necessidade de inclusão de outras categorias, que eram previstas por Ranganathan (1967) como subcategorias de alguma dessas cinco. Neste trabalho, ficaria mais clara a separação das mesmas como categorias próprias, como é o caso das categorias medida e propriedade. Desta forma, foram incluídas mais 14 relações:

Personalidade-Medida
Matéria-Medida
Processo-Medida
Espaço-Medida
Tempo-Medida
Medida-Medida
Medida-Propriedade
Propriedade-Personalidade
Propriedade-Medida
Propriedade-Processo
Propriedade-Espaço
Propriedade-Tempo
Propriedade-Propriedade
Propriedade-Medida

Essa distribuição serviu para categorizar as relações apresentadas na literatura (SALES, 2006). Porém, como era esperado, algumas categorias não apresentavam facetas, pois as relações identificadas na literatura não esgotam as possibilidades de manifestação das relações, já que estas podem aparecer de forma diferente dependendo da área.

O passo seguinte foi a identificação da possibilidade de relações formais entre essas categorias. Para a execução desta etapa, utilizou-se a listagem de relações funcionais identificadas na literatura. Para cada dupla de ocorrência foram selecionadas todas as possibilidades de relações formais, chegando assim, à seguinte sistematização:

QUADRO 3 – Sistematização das Relações Funcionais

Relações Categorias Funcionais	Relações Formais Funcionais
Personalidade – Personalidade	<ul style="list-style-type: none"> • é_produto_de • é_instrumento_de • transforma-se_em • deriva_de • derivado_de • interage_com • manifestação_de • analisa • mede • cria • localizado_em • adjacente_a • circunjacente_a • transversal_a
Processo – processo	<ul style="list-style-type: none"> • método_de • interage_com • é_sub_evento_de • é_fase_de • é_etapa_de • causa • impede
Processo – Personalidade	<ul style="list-style-type: none"> • causa • afeta • rompe • destrói • resulta_em • localizado_em
Espaço-Espaço	<ul style="list-style-type: none"> • adjacente_a
Processo – Espaço	<ul style="list-style-type: none"> • ocorre_em
Medida – Personalidade	<ul style="list-style-type: none"> • é_medida_de
Medida – Processo	<ul style="list-style-type: none"> • é_medida_de
Personalidade – Processo	<ul style="list-style-type: none"> • usado_em • afeta • realiza/produz • é_usado_para • analisa • mede • diagnostica
Personalidade – Propriedade	<ul style="list-style-type: none"> • tem_propriedade
Propriedade – Personalidade	<ul style="list-style-type: none"> • é_propriedade_de • é_contraagente_de
Propriedade-processo	<ul style="list-style-type: none"> • é_propriedade_de

Fonte: Dados da pesquisa.

Quem domina a teoria de Ranganathan (1967) sabe que, ao incluir o segundo nível de Personalidade, ele caracterizou Agentes/Instrumentos que afetam Processo. Nesta perspectiva, algumas relações acima ficam mais compreensíveis.

Esta sistematização evidencia as possibilidades de relações formais existentes entre as duplas de ocorrências de relações categoriais funcionais. Enfatizamos que estudos devem ser feitos para padronizar as relações na língua portuguesa.

Conforme pôde ser observado, existem relações que parecem sinônimas, mas que devem ser analisadas em um contexto, para verificar se elas podem ser consideradas equivalentes ou não. Nesse último caso, devem-se criar regras para quando usar cada relação.

4 Aplicação do modelo de relações no âmbito da *Gene Ontology*

Este capítulo trata da análise das definições de uma amostra da *Gene Ontology*, que foi realizada visando a aplicação do modelo de relações sistematizadas no capítulo anterior.

Foram identificados 267 termos. Destes, 20% cobriam aproximadamente 70% das ocorrências e suas definições, na *Gene Ontology*, foram analisadas. Os termos foram reunidos segundo as categorias da *Gene Ontology*: Função molecular, Processo biológico e Componente celular.

A *Gene Ontology* fornece orientação para a definição de cada uma das categorias. Alguns modelos, entretanto, merecem discussão, o que não cabe fazer aqui, mas vale dizer que dificultariam um processo automático de tratamento da informação. Por exemplo, há inúmeras definições circulares. Encontrou-se pelo menos uma definição pela negação.

As relações identificadas na categoria Função Molecular são representadas por verbos ou substantivos (atividade) de catálise ou “*binding*”. Enquadram-se no modelo Processo/atividade – Personalidade. Foram identificadas as seguintes relações: transfere_para, reação_de, interage_com, movimenta-se_ao-longo_de, possibilita_movimento_com/dentro_de/entre/para_fora, contribui_para. Esta categoria requer revisão do modelo de definição, de sorte a prover definições consistentes que facilitarão o tratamento automático. Por exemplo, não se pode iniciar uma definição com um verbo.

As relações identificadas na categoria Processo biológico ligam Processo/Personalidade. Foram identificadas as seguintes relações: progressão_de; aumenta, gera, causa/resulta em.

A Categoria Componente Celular é uma das mais complexas, pois envolve não apenas a relação partitiva mas, ainda, espaço. É evidente que, nesta categoria, a relação é entre Parte X Parte (PersonalidadeXPersonalidade). Foram encontradas as seguintes relações: componente_de, parte_de, parte_externa, parte_interna, parte_em_torno_de, parte_dentro_de.

O estudo mostrou a necessidade de rever as definições, pois elas devem fornecer elementos para se identificar proposições verdadeiras.

A seguir será mostrado um exemplo de conceito de cada categoria, aplicando-se o modelo triádico. Este modelo, diferentemente do modelo diádico dos tesouros, onde se evidenciam pares de categorias (ex: processo-resultado), é composto por três partes, a saber: categoria antecessora, relação formal e categoria sucessora.

Componente celular

QUADRO 4 – Análise de um modelo de relações na categoria Componente Celular da GO

Termo	<i>Membrane</i>		
Definição	<i>"Double layer of lipid molecules that encloses all cells, and, in eukaryotes, many organelles; may be a single or double lipid bilayer; also includes associated proteins."</i>		
Relação Formal	<i>Encloses</i>		
Relação Categorial	Categoria Antecessora: Personalidade	Categoria Sucessora: Personalidade	
Modelo Comprovado	Personalidade (<i>membrane</i>)	<i>encloses</i>	Personalidade [<i>all</i>] <i>cells</i>

Fonte: Dados da pesquisa.

Função molecular

QUADRO 5 – Análise de um modelo de relações na categoria Função Molecular da GO

Termo	<i>Protein kinase activity</i>		
Definição	<i>"Catalysis of the transfer of a phosphate group, usually from ATP, to a protein substrate".</i>		
Relação Formal	<i>"transfer of"</i>		
Relação Categorial	Categoria Antecessora: Energia	Categoria Sucessora: Personalidade	
Modelo Comprovado	Processo/atividade (<i>protein kinase activity</i>)	<i>Transfers to</i>	Personalidade (<i>protein substrate</i>)

Fonte: Dados da pesquisa.

Processo biológico

QUADRO 6 – Análise de um modelo de relações na categoria Processo Biológico da GO

Termo	<i>Positive regulation of growth rate</i>		
Definição	<i>"Any process that increases the rate of growth of all or part of an organism."</i>		
Relação Formal	<i>Increases</i>		
Relação Categorial	Categoria Antecessora: Processo	Categoria Sucessora: Medida	
Modelo Comprovado	Processo/atividade (<i>positive regulation of growth rate</i>)	<i>increases</i>	Processo/atividade (<i>rate of growth</i>)

Fonte: Dados da pesquisa.

5 Considerações finais

Através da análise das definições no *corpus* selecionado, pôde-se encontrar na *Gene Ontology* novos tipos de relações formais, que, apesar de não revelados em sua estrutura, estavam implícitos nessas definições.

Essa análise evidenciou relações que já haviam sido sugeridas pela literatura, mas também encontrou relações que ainda não haviam sido mencionadas. Isso aconteceu porque cada área se manifesta de forma diferente, e as relações serão sempre reflexos do domínio que se está representando. Neste sentido, para cada ontologia a ser construída poderão aparecer relações inéditas, ainda não mencionadas pela literatura.

O estudo permitiu analisar as definições e identificar problemas que devem ser corrigidos. As definições rigorosas são resultadas de análise dos referentes, de modo a estabelecer proposições verdadeiras.

6 Referências

- AITCHISON, J. *Thesaurus construction: a practical manual*. 2. ed. London: ASLIB, 1987.
- DAHLBERG, I. A referent-oriented analytical concept theory of Interconcept. *International Classification*, Frankfurt, v. 5, n. 3, p. 142-150, 1978.
- FERNANDEZ, M.; GOMEZ-PÉREZ, A.; JURISTO, N. From ontological art towards ontological engineering. *AAAI Technical Report*, California, p. 33-40, jun.1997.
- GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human and Computer Studies*, v.43, n. 5-6, p. 625-640, 1995.

INTERNATIONAL STANDARD ORGANIZATION - ISO. ISO 2788: documentation guidelines for the establishment and development of monolingual thesauri. 2.ed. [S.l.], 1986. 32p.

_____. ISO-DIS-704: principles and methods of terminology. Geneva, 2000.

KIETZ, J-U; MAEDCHE, A.; VOLZ, R. A method for semi-automatic ontology acquisition from a corporate intranet. In: EKAW'00: WORKSHOP ON ONTOLOGIES AND TEXT, 2000, Juan-Les-Pins. *Proceedings ...* Berlin: Springer, 2000.

MOTTA, D. F. *Método relacional como nova abordagem para a construção de tesouros*. Rio de Janeiro: SENAI/ DN, 1987

RANGANATHAN, S. R. *Prolegomena to library classification*. Bombay: Asia Publishing House, 1967.

SAGER, J. C. *A practical course in terminology processing*. London: John Benjamins Publishing Company, 1990.

SALES, L. F. *Ontologias de domínio: um estudo das relações conceituais e sua aplicação*. 142 f. Dissertação (Mestrado) – Universidade Federal Fluminense, Niterói, 2006.

SMITH, B. et al. Relations in biomedical ontologies. *Genome Biology*, [S.l.], v.6, n.5, 2005.

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

SURE, Y; STAAB, S; STUDER, R. *Methodology for development and employment of ontology based knowledge management applications*. 2002. Disponível em: <www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2002_sigmod-methodology.pdf>. Acesso em: 15 out. 2005.

USCHOLD, M.; KING, M. Towards a methodology for building ontologies. In: IJCAI-95: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING, 1995, Montreal. *Proceedings...* Mahwah, NJ, USA: CRC, 1998. Disponível em: <<http://citeseer.ist.psu.edu/uschold95toward.html>> . Acesso em: 15 out. 2005.

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

WUSTER, E. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, L'informatique et les sciences des chose. In: RONDEAU, G.; FELBER, E. (Org.). *Textes choisis de terminologie*. Québec: GIRSTERM, 1981. (Fondements théoriques de la terminologie, v. I). p. 57-114.