

SEMÂNTICA NA REPRESENTAÇÃO DO CONHECIMENTO: DO VOCABULÁRIO CONTROLADO À ONTOLOGIA

SEMANTICS IN KNOWLEDGE REPRESENTATION: FROM CONTROLLED VOCABULARY TO ONTOLOGY

Jorge Calmon de Almeida Biolchini¹

RESUMO:

A relação dos processos de representação do conhecimento e da informação é analisada, em referência às dimensões da organização do conhecimento no domínio da medicina e à natureza das bases de conhecimento nesta área. O desafio imposto pela explosão informacional crescente frente à necessidade de singularização contextual dos problemas de saúde, na reconstrução do conhecimento especializado pelo profissional, determina requisitos de qualidade e de relevância da informação bem como o desenvolvimento de modelos e sistemas de suporte à inteligência médica. A construção de redes semânticas e ontologias, a partir de bases de conhecimento e vocabulários controlados, é apontada como uma linha de desenvolvimento capaz de fornecer instrumentos para aumentar a recuperação de informação relevante diante da diversidade do uso profissional da informação, permitindo vincular os processos de cognição e de decisão às relações de significação e aos modelos de representação da informação.

¹ Doutor em Ciência da Informação, pelo Programa de Pós-Graduação em Ciência da Informação, do convênio IBICT-UFRJ

jbiol@mtec.com.br

ABSTRACT:

The relation between knowledge and information representation is analyzed, in reference to the knowledge organization dimensions in the domain of medicine and to the nature of the knowledge bases in this area. The challenge of the increasing information explosion to the necessity of health problems contextual singularization, from the part of the professional, when reconstructing specialized knowledge, determines quality and information relevance requirements as well as the development of medical intelligence models and support systems. Constructing semantic networks and ontologies, from knowledge bases and controlled vocabularies, is pointed out as a development line capable to provide the necessary instruments in order to increase relevant information retrieval vis-à-vis the diversity on the professional use of information, allowing to link cognition and decision processes to signification relations and models of information representation.

PALAVRAS-CHAVE:

representação do conhecimento; organização do conhecimento; medicina; base de conhecimento; conceito; ontologia

KEYWORDS:

knowledge representation; knowledge organization; medicine; knowledge base; concept; ontology

Da Organização à Representação do Conhecimento

A representação logicamente formalizada de estruturas de conceitos formadas por conjuntos específicos de elementos e relações conceituais, constitui o campo de investigação conhecido como representação do conhecimento. O corpo do conhecimento em medicina pode ser compreendido, em termos de sua organização, em referência a determinados eixos fundamentais, vinculados à natureza específica dos diferentes conteúdos informacionais deste campo do conhecimento. Esses eixos e seus respectivos sub-eixos de organização do conhecimento em medicina encontram-se relacionados a conceitos oriundos do desenvolvimento histórico e epistemológico dos principais domínios do conhecimento médico. Tais diferentes eixos organizacionais do conhecimento encontram relações de contato e de vínculo entre si, relacionamentos estes que podem representar determinadas unidades de significação que se encontram presentes no conjunto de associações semânticas efetuadas durante o processo de raciocínio de um profissional médico, tal como, por exemplo, durante o desenrolar da leitura analítica de um caso clínico. O conjunto formado por estes eixos organizacionais do conhecimento médico pode, deste modo, exercer a função de um sistema de marcadores lingüísticos de conteúdo semântico de um texto médico, podendo, desta maneira, ser utilizado como um sistema formal voltado para esta finalidade, a qual se encontra compreendida por uma dupla função. A primeira delas, aplicada sobre o sentido vertical de estruturação, lingüisticamente referente ao eixo paradigmático, corresponde a funcionar como um conjunto de identificadores e indexadores de regiões de conceitos e de conteúdos de conhecimento. A segunda, exercida sobre o sentido horizontal, lingüisticamente relativa ao sintagmático, diz respeito à sua função

marcadora da trajetória semântica realizada pelas unidades de conteúdo do texto médico, e referenciada por sua vez aos eixos de organização deste campo de conhecimento especializado. O desenvolvimento da presente pesquisa está voltado no sentido de possibilitar que se avance mais profundamente nesta direção, a partir de um maior detalhamento das unidades conceituais existentes bem como de suas inter-relações em diferentes níveis e direções, juntamente com um maior grau de formalização de seus elementos e características estruturais. Deste modo, possibilita aportar novas funções e possibilidades de representação dos conteúdos informacionais e dos relacionamentos conceituais presentes no conhecimento médico.

Do Léxico ao Conceito

Listas alfabéticas podem permitir a ordenação de léxicos em função da representação gráfica e fonética dos termos empregados na área do conhecimento em questão, sem contudo permitir a recuperação das informações necessárias para o usuário a partir dos conceitos que formam a base do seu raciocínio e de seus processos de tomada de decisão.

Coincidências lexicais, obtidas a partir de procedimentos derivados do estabelecimento de coincidências entre os caracteres pertencentes à cadeia fonética de palavras selecionadas para uma busca, podem permitir a localização de léxicos, e a conseqüente recuperação de dados em função da representação gráfica e fonética dos termos em questão.

A ausência da dimensão conceitual dos termos envolvidos, no entanto, impede o estabelecimento de ordenações de significação entre eles, bem como determina a incidência de falsos positivos, em termos da precisão no processo de recuperação informacional, assim

como a incidência de falsos negativos, no que tange à revocação de elementos a serem recuperados.

Em uma outra esfera, o emprego de procedimentos estatísticos para a recuperação de léxicos contidos em bases textuais permite a recuperação de dados em função do valor ponderal e frequencial dos termos empregados, no nível da representação lexical dos mesmos. Uma maior frequência lexical encontra-se hipoteticamente associada a um maior peso relativo e relevância conceituais dentro de um conjunto de termos.

Ambos os procedimentos acima mencionados trabalham com a representação das informações no nível lexical, diferindo entre si quanto à operação lógica executada sobre este plano de representação gráfico-fonética, bem como quanto à unidade lexical empregada para sua execução. A primeira operação está vinculada à verificação e determinação, seqüencial e linear, da existência de coincidências, as quais incidem de modo sucessivo sobre a estrutura gráfica de cada unidade dos caracteres componentes dos léxicos selecionados. A segunda operação se vincula, por outro lado, ao estabelecimento de cálculos e comparações estatísticas, as quais incidem sobre as unidades compostas, compreendidas por cadeias de caracteres, representadas pelos léxicos e conjuntos de palavras selecionados para a análise.

Estas relações encontram-se esquematizadas na tabela que se segue abaixo:

Métodos de Recuperação da Informação	<u>Lexical:</u>	<u>Estatístico:</u>
Nível de Representação:	<i>Lexical (estrutura gráfico-fonética)</i>	<i>Lexical (estrutura gráfico-fonética)</i>
Operação Lógica:	<i>Determinação Seqüencial de Coincidências</i>	<i>Cálculo Estatístico de Frequências</i>
Unidade de Trabalho:	<i>Caracteres (componentes dos léxicos)</i>	<i>Cadeias de caracteres (de palavras e expressões)</i>

A ausência ou escassez na utilização de relações semânticas nas bases de conhecimento, como instrumentos de tratamento informacional, desenvolvidas no nível dos conceitos associados aos termos contidos nestas bases, assim como no nível de suas relações conceituais, determina uma importante redução no grau de precisão do processo de recuperação informacional através dos instrumentos usualmente utilizados para este fim, de um lado em função da relevância clínica dos termos empregados na busca não ser considerada, bem como, de outro lado, pelo fato das necessidades de busca paciente-específicas por parte dos profissionais não se encontrarem contempladas em sistemas com estas características. (Cooper, 1998)

Em concordância com Cooper, relações de natureza semântica podem vir a ser utilizadas após o emprego de metodologias lexicais e estatísticas, visando aumentar a precisão de informações clinicamente relevantes, como uma espécie de filtro a posteriori, com mínima ou nenhuma influência sobre a revocação.

Além disso, dentro de uma perspectiva ainda mais ampliada, conceitos e relações de significação podem vir a ser utilizados, também, como elementos de partida para o emprego ulterior de metodologias lexicais e estatísticas, determinando-se aprioristicamente um conjunto de possibilidades dos elementos a serem priorizados e uma estruturação de suas relações, coerente com a natureza do conhecimento em questão, e conseqüentemente um direcionamento mais efetivo dos processos de recuperação informacional, consistente com a natureza do raciocínio clínico.

Na medida em que, entre os elementos primordiais que caracterizam o raciocínio humano, encontra-se o campo das relações de significação, capazes de serem estabelecidas entre as unidades de conhecimento e informação, tais relações devem se encontrar no centro das atenções deste campo de desenvolvimento. Uma vez que se destinam a auxiliar os

processos de raciocínio e de tomada de decisão dos profissionais, os sistemas de informação podem ser desenvolvidos visando ampliar as potencialidades, não somente de busca e recuperação de unidades informacionais, como também de organização e apresentação de unidades semanticamente relacionadas, ampliando a recuperação e evocação de unidades de sentido e de conjuntos informacionais relevantes e significativos.

Este desenvolvimento da organização de relações de significação pode ser conduzido através da construção de redes semânticas entre os elementos constitutivos das bases de conhecimento, compreendidos pelos termos que se encontram presentes nas mesmas bem como pelos componentes de suas respectivas definições, de modo a se estabelecer nós e grafos de associação entre tais unidades constituintes. (Sparck Jones & Kay, 1973)

A partir desta linha de desenvolvimento, pode-se proceder à construção de ontologias do conhecimento médico, enquanto especificações explícitas de conceitualizações desenvolvidas neste domínio do conhecimento. Estas conceitualizações são compreendidas como estruturas semânticas formais, definidas por relações intensionais, estruturas estas que codificam as regras implícitas que restringem a estrutura de um elemento da realidade. (Guarino, 1995)

Neste sentido, pode-se proceder ao desenvolvimento de estruturas ontológicas de procedimentos diagnósticos e terapêuticos, organizadas segundo relações interconceituais, definidas a partir das unidades terminológicas propriamente ditas, contidas nas bases de conhecimento médico. Estas unidades, na medida em que apresentam valor de significação referente à terminologia especializada, permitem a construção de novos conjuntos de redes semânticas entre as bases de conhecimento. (Dahlberg, 1978) (Levesque & Lakemeyer, 2000)

O estudo realizado sobre as relações semânticas das bases informacionais que compõem o universo do conhecimento médico pode permitir a redução das diferenças entre as representações e configurações conceituais apresentadas por sistemas de informação e as representações mentais do conhecimento especializado de seus usuários.

A perspectiva de ampliação da dimensão semântica de sistemas artificiais aponta no sentido de contribuir para uma maior aproximação entre os atores humanos e artificiais, otimizando a interação e o diálogo entre ambos, e conseqüentemente influenciando de modo positivo no processo de tomada de decisão.

Ontologia como Campo de Convergência no Tratamento da Informação

A presente abordagem diz respeito, portanto, a uma sistemática de desenvolvimento que implica em uma integração metodológica efetuada entre campos de investigação sobre o conhecimento, os quais se encontram relacionados entre si, ou seja, desde o campo da organização do conhecimento, próprio da Ciência da Informação, até o campo da representação do conhecimento, próprio da Ciência da Computação. Sob esta perspectiva, o presente trabalho se situa na interface entre ambos domínios do conhecimento, pertencentes a campos científicos muito próximos, tanto em termos de um conjunto de interseção composto por objetos de investigação comuns, quanto em termos de suas origens históricas.² Esta pesquisa visa utilizar instrumentos conceituais e metodológicos de ambos

² No nascedouro da Ciência da Informação, o termo “informática” era utilizado por autores russos e franceses, para denominar esta ciência emergente responsável por estudar os fenômenos do universo da informação. Mais tarde, o termo terminou por ser assimilado ao campo da Ciência da Computação, apesar de que posteriormente e mesmo nos dias de hoje possam ainda ser encontradas referências de alguns autores utilizando o termo informática como sinônimo de ciência da informação, em contraposição ao termo computação, como sinônimo de tecnologia da informação. (Gorkova, 1980) (Gross, 2001)

campos de investigação, no sentido de ampliar o ferramental necessário para lidar com a complexidade do universo de estudo em questão, integrando-os no tratamento da informação pertencente ao mesmo.

Sob uma perspectiva cognitivista, a presente linha de investigação diz respeito à relevância central que apresentam as estruturas de representação do conhecimento sobre o desenvolvimento de sistemas de informação inteligentes. Independentemente do tipo de uso e de aplicação projetados para esta espécie de sistema, tipos de estruturas de representação do conhecimento estão sempre e necessariamente presentes, acoplados e subjacentes aos demais módulos funcionais do sistema, visando à realização dos objetivos e funções planejados de modo específico para cada modelo de sistema de informação inteligente. Sempre quando uma interação do gênero humano-máquina é empregada, estruturas semânticas se constituem num dos elementos fundamentais de sua interface. Em domínios especializados de aplicação, tal como o campo médico, a inclusão de estruturas de representação semântica do conhecimento pode ser feita em diferentes níveis de complexidade, de acordo com os usos pretendidos e os usuários potenciais de um dado sistema.

A micro-estrutura semântica das unidades do conhecimento médico, acoplada à multidimensionalidade de relações que dão suporte à complexidade das possíveis redes de significado para diferentes usuários e distintos usos, podem constituir uma base de elementos e relacionamentos semânticos capazes de participar no desenvolvimento de sistemas baseados no conhecimento, os quais se aproximam da estruturação multinível dos processos de tomada de decisão médica. Podem também representar uma sólida base para a integração de diferentes sistemas de informação, em distintos graus de granularidade de dados e de organização da informação, bem como direcionados para o reuso de

estruturas semânticas particulares, desenvolvidas em determinadas aplicações, na concepção de novos tipos de aplicação. (Hjørland & Albrechtsen, 1995)

Na medida em que o conhecimento médico é multi-paradigmático em sua constituição epistemológica e multinível em sua estruturação conceitual, uma perspectiva cognitiva permite contribuir para lidar com os aspectos multidimensionais dos processos de transferência de significado e sentido que estão presentes nas interfaces tanto do desenvolvimento quanto do uso de sistemas de informação. Esta questão assume uma relevância significativa em duas amplas esferas. De um lado, tem uma influência sobre os processos que envolvem as múltiplas possibilidades de interface humano-máquina, num espectro que vai desde o profissional de saúde até o cientista da informação. Por outro lado, exerce uma influência naqueles processos que se encontram relacionados tanto com as interfaces intermodulares dos sistemas de informação, quanto com aquelas pertinentes à intercomunicação entre diferentes bases de dados, assim como entre estas e distintas bases de conhecimento. Portanto, o desenvolvimento de correlações entre processos e estruturas relativas aos campos da cognição, da decisão, e da informação, torna-se necessário para que se possa relacionar semanticamente conteúdo, contexto e propósito.

Visando prover um suporte baseado em conhecimento para a concepção de sistemas e ferramentas direcionados para a tomada de decisão médica, o universo de procedimentos médicos é escolhido pelo autor como um campo central de conceitos a ser trabalhado sob esta perspectiva. Na medida em que os procedimentos constituem as próprias ações que são realizadas pelos profissionais de saúde, tanto no sentido diagnóstico-prognóstico, de se obter conhecimento sobre os estados de saúde dos pacientes, quanto no sentido terapêutico-preventivo, para efetuar mudanças em seus estados e condições de saúde, podem os mesmos ser situados conceitualmente no cerne dos processos das decisões médicas.

O Sistema de Linguagem Médica Unificada (*Unified Medical Language System*), conhecido como UMLS, desenvolvido pela Biblioteca Nacional de Medicina (*National Library of Medicine*), nos EUA, se constitui no conjunto de fontes de conhecimento no seio do qual o presente estudo é conduzido. A escolha deste sistema se justifica face à importância que este vem assumindo ao longo dos últimos quinze anos, enquanto uma base de recursos voltada para o desenvolvimento de distintos sistemas de informação em saúde, assim como para a integração de diferentes tipos de fontes de informação e de bases de dados. Ao relacionar conjuntos terminológicos, a partir de um sistema integrado e composto por mais de trinta vocabulários como fontes de informação, o Metatesouro do UMLS provê um arcabouço para a exploração das possibilidades de recuperação efetiva de elementos informacionais desde múltiplas bases de dados, assim como para a integração de informações baseadas em conhecimento com dados de registros clínicos.³

A multidimensionalidade, a versatilidade e a flexibilidade da representação e da recuperação de informação relevante para a tomada de decisão, as quais são características necessárias para esses dois propósitos, demandam esforços de pesquisa na direção da ampliação do nível de abstração e do grau de complexidade da rede relacional dos conceitos. A provisão de valores agregados de informação, através do desenvolvimento de uma semântica mais profunda das descrições conceituais, pode ser atingida através da evolução efetuada ao se partir da maneira de organização sistemática de um vocabulário controlado até o desenvolvimento de uma ontologia do conhecimento. A conversão do modo alfabético de ordenação dos termos especializados, primariamente concebido como

³ O trabalho de campo da presente pesquisa foi conduzido por ocasião do estágio de doutorado-sandwich, realizado pelo autor no período de janeiro a abril de 2002, no *Lister Hill National Center for Biomedical Communications*, centro de pesquisas vinculado à *National Library of Medicine*, situada no campus dos *National Institutes for Health*, na cidade de Bethesda, EUA.

uma ferramenta de indexação, para uma ordem terminológica estruturada semanticamente, concebida concomitantemente como uma estrutura de conhecimento e como uma interface de significação, multiproposicionais portanto, pode, desta maneira, prover uma base para que sejam desenvolvidos agentes inteligentes de informação apoiados por modelos conceituais específicos relativos ao domínio do conhecimento médico.

O desenvolvimento de uma ontologia, construída sobre o conjunto de conceitos que estão contidos sob os cabeçalhos de assunto do UMLS referentes aos procedimentos médicos, constitui-se, desta forma, no principal objetivo de desenvolvimento da pesquisa.

Princípios da Representação do Conhecimento

A representação do conhecimento apresenta cinco princípios básicos, de acordo com três especialistas nesta área, Randall Davis, Howard Schrobe, e Peter Szolovits, conforme citados por Sowa (2000, p. 134-143). O primeiro princípio refere-se ao fato de que uma representação do conhecimento é um sistema de símbolos e de conexões entre eles, os quais funcionam como substitutos representacionais das coisas do mundo externo, de forma que possam ser armazenados eletronicamente no computador e ser manipulados logicamente, simulando através deste processo o comportamento do sistema externo ao qual representam. O segundo princípio diz respeito à noção de que uma representação do conhecimento é um conjunto de compromissos ontológicos, ou seja, na medida em que a ontologia se constitui em um estudo da existência das coisas no mundo, as categorias de coisas e suas inter-relações conceituais representam os compromissos ontológicos designados pelo modelizador do sistema de representação. O terceiro princípio consiste na

idéia de que uma representação do conhecimento se constitui em uma teoria fragmentária do raciocínio inteligente, ou seja, em uma descrição de comportamentos de coisas no mundo e de suas respectivas interações, sendo esta descrição uma teoria acerca do tipo de coisas em questão, declarada através de axiomas explícitos. O quarto princípio está relacionado ao fato de que uma representação do conhecimento consiste em um meio para se codificar o conhecimento, de maneira a se poder estabelecer um processamento computacional eficiente de seus elementos constituintes. E, finalmente, o quinto princípio refere-se ao fato de que uma representação do conhecimento consiste em um meio para prover a facilitação do processo de comunicação entre os cientistas da informação, os analistas de sistemas e os engenheiros do conhecimento, de um lado, os quais desenvolvem as definições, as representações e as regras, e os especialistas de um determinado domínio do conhecimento, de outro lado, os quais verificam se o sistema de representação constitui uma teoria realista do domínio do conhecimento ao qual pertencem.

Ontologia do conhecimento em medicina

O desenvolvimento de ontologias do conhecimento se apresenta como um campo aberto para a experimentação sobre a conceitualização do mundo, uma espécie de campo-síntese da modelização conceitual do conhecimento que se tem sobre o universo. Além disso, se constitui em um campo experimental de estudo voltado para aplicações práticas em diferentes domínios do conhecimento, servindo como base para uma comunicação mais precisa entre especialistas de um dado domínio, bem como para a construção de sistemas computacionais inteligentes. As questões colocadas pelo campo de investigação da

ontologia dizem respeito à representação de um determinado campo do conhecimento, através da formalização explícita de seus conceitos e relações interconceituais, e da construção de estruturas lógicas que permitam representar as teorias que se possui acerca das entidades e dos processos envolvidos. Uma teoria ontológica contém e inter-relaciona conceitos genéricos que representam características ou componentes de um universo em particular. (Bunge, 1977)

Uma ontologia pode ser desenvolvida em relação a domínios do conhecimento, compreendendo os conceitos e relações conceituais relacionados a conteúdos semânticos, proposições, e teorias acerca do domínio específico a ser representado, assim como pode ser desenvolvida em relação a tarefas executadas por especialistas em determinados domínios, compreendendo a modelagem conceitual dos processos envolvidos, de suas variáveis, e de seus elementos de controle e performance.

Diferentes modelos estruturais de ontologia podem ser construídos, em função do conjunto principal de relações conceituais existentes entre as entidades modeladas. Os principais tipos são constituídos, de um lado, pela taxonomia, a qual é compreendida por relações hierárquicas entre as entidades conceituais de tipo identidade de classes, e de outro lado, pela partonomia, também denominada meronomia, termo derivado do vocábulo grego ‘meros’ que denota ‘parte’. Desta forma, a taxonomia apresenta relações de tipo equivalente à taxonímia (Cruse, 1997), as quais são relacionamentos conceituais da espécie ‘é um(a)’, como, por exemplo, na relação existente entre dois conceitos de uma hierarquia taxonômica constituída por ‘artéria’ e ‘coronária’, onde se pode estabelecer uma relação taxonômica do tipo ‘coronária é um(a) artéria’. No caso da estrutura ontológica de tipo partonômica, ou meronômica, a relação que a fundamenta encontra-se constituída por uma partonímia, a qual é uma relação do tipo ‘parte de’, como, por exemplo, na relação que

pode existir entre dois conceitos de uma hierarquia partonômica constituída por ‘miocárdio’ e ‘coração’, onde se pode estabelecer uma relação partonímica do tipo ‘miocárdio é parte do coração’. Estruturalmente, do ponto de vista da lógica booleana correspondente aos seus respectivos eixos vertical e horizontal, a taxonomia e a partonomia podem ser consideradas como mutuamente inversas, já que o eixo vertical da taxonomia é uma relação “e” entre as características dos conceitos de nível superior e inferior, enquanto que para a partonomia tal relação é do tipo “ou”. Inversamente, a relação correspondente ao eixo horizontal da taxonomia é uma relação do tipo “ou”, enquanto que a relação horizontal da partonomia é do tipo “e”. Assim, por exemplo, em uma estrutura taxonômica, o conceito hierarquicamente superior ‘câmara cardíaca’ se relaciona com os seus conceitos subordinados, ou sendo um ‘átrio’ **ou** sendo um ‘ventrículo’, enquanto que, em uma estrutura partonômica, o conceito hierarquicamente superior ‘coração’ se relaciona com os seus conceitos subordinados, contendo ou sendo composto pelas partes ‘pericárdio’ e ‘miocárdio’ e ‘endocárdio’.

Estruturas ontológicas híbridas podem ser, também, desenvolvidas, contendo regiões ou eixos taxonômicos combinados com regiões ou eixos partonômicos. (Habel, Pribbenow, Simmons, 1995) A escolha da estrutura a ser utilizada para a construção de uma determinada ontologia é determinada pela natureza do universo em estudo, e da modalidade ontológica que seja capaz de melhor servir como modelo representacional daquela realidade em particular. A ontologia de procedimentos médicos em cardiologia, conforme desenvolvida pelo autor com conceitos presentes na fonte de conhecimento médico do UMLS constitui-se em uma ontologia de tipo híbrido, no sentido de melhor representar a caracterização de um procedimento em medicina.

Desenvolvimento da Ontologia de Procedimentos em Medicina

O desenvolvimento de ontologias de domínio é de suma e crescente importância para a construção de muitas aplicações diferentes no campo da informática médica. A relevância de vocabulários controlados e tesouros para o gerenciamento da informação na área médica e para o desenvolvimento da integração de diferentes fontes de dados, de informação e de conhecimento é uma questão central neste campo. A importância crescente das dimensões e dos processos de cognição humana para o entendimento do tratamento informacional por distintos tipos de usuários, motivado por diferentes espécies de necessidades, bem como para a organização e a concepção de sistemas baseados-em-conhecimento, coloca estas questões no coração do desenvolvimento presente e futuro de agentes e sistemas inteligentes.

Posto que o Sistema de Linguagem Médica Unificada (*Unified Medical Language System*) é um conjunto de fontes de conhecimento que apresenta múltiplos propósitos na área médica em termos de informação, foi o mesmo utilizado como base de conhecimento no domínio da medicina no sentido de desenvolver uma ontologia de procedimentos no domínio específico da cardiologia.

O principal propósito de uma ontologia de procedimentos médicos é desenvolver uma conceitualização dos processos que se encontram no cerne das ações médicas, estas concebidas e implementadas no sentido de promover evidência sobre estados de saúde, por um lado, e efetuar mudanças nas condições de saúde, por outro lado. Esses processos constituem uma representação do conhecimento médico, concernente com o campo da intervenção, traduzido em ações, sujeitos, objetos, instrumentos e circunstâncias. Estes

elementos constituintes caracterizam tanto os objetivos como os comportamentos dos atos que são conduzidos no sentido de produzir efeitos no campo da saúde, assim como para revelar o contexto no qual são aplicados.

Os procedimentos são constituídos, em uma base de conhecimento tal como o UMLS, como uma conjunção de um conceito, o qual lhes dá sua designação, e uma definição correspondente, especificamente construída para o conceito. A definição textual específica referente ao termo denota, em uma forma analítica, o conceito a ela associado, o qual é sua contrapartida sintética. As características contidas nas definições textuais dos conceitos, os quais se referem aos procedimentos médicos e estão presentes em uma base de conhecimento médico, representam os elementos relevantes da descrição daquilo que constitui a ação envolvida no procedimento. Estes elementos da definição denotam um conjunto de atributos da descrição do procedimento, e podem então ser definidos como uma estrutura formal de elementos básicos, os quais são os constituintes de todo o processo.

Além da sua função descritiva, o conjunto de elementos contidos na definição dos conceitos é capaz de assumir o papel de ser prescritivo, no sentido de que aos mesmos pode ser atribuída a função de dirigir os profissionais para que estes incluam os elementos informacionais relevantes, o que os constitui como uma peça de representação da base de conhecimento. Para que esta função seja cumprida, o formato da definição necessita preencher os requisitos de ser uma definição canônica do procedimento.

A multidimensionalidade do processo de um procedimento médico pode, então, ser representada como uma estrutura multidimensional dos conteúdos de informação que estão incluídos em sua definição terminológica canônica, permitindo o processamento digital dos elementos envolvidos e das relações entre eles, para diferentes usos e aplicações.

Alguns dos usos que podem ser desenvolvidos ou aprimorados, ao se representar formalmente a complexidade dos procedimentos médicos, são os seguintes: a normalização dos padrões de registro de dados clínicos, a ajuda ao profissional médico iniciante para que considere todo o conjunto de aspectos envolvidos na conduta, a avaliação da performance das ações médicas, a ajuda ao profissional experiente para que explore novas possibilidades de vínculos e relacionamentos entre os conceitos que estão representados, a auditoria da qualidade dos procedimentos realizados, o suporte ao desenvolvimento de sistemas baseados em ontologia para a tomada de decisão médica, a promoção de aperfeiçoamentos evolutivos das bases de conhecimento médico, o suporte à recuperação de informação relevante, a identificação de conceitos e de suas relações inter-conceituais em textos médicos, tais como prontuários clínicos, e laudos de exames complementares, e a integração entre fontes de conhecimento e bases de dados clínicos.

Implicações e Perspectivas

A conclusão do desenvolvimento da ontologia procedimental realizada no domínio do conhecimento médico através da presente pesquisa traz uma série de possíveis desdobramentos relacionados com a organização conceitual no campo da saúde.

A consequência primária é relativa à operacionalização e ao uso dos elementos definidos para o suporte à concepção de sistemas de informação, voltados para diferentes propósitos e utilizáveis em diferentes ambientes de atenção à saúde. Também permite prover elementos de suporte para a modelagem dos processos de raciocínio, conforme executados por

especialistas no campo médico. Estes dois aspectos apontam para sua contribuição no desenvolvimento de sistemas de suporte à tomada de decisão médica.

Por causa de sua derivação das fontes de conhecimento do UMLS, também pode contribuir para uma avaliação da própria organização conceitual deste sistema de linguagem médica.

A natureza dos resultados obtidos, bem como a metodologia desenvolvida para a pesquisa, podem ser estendidas em direção a bases de conhecimento de outras especialidades médicas, permitindo o estabelecimento de vínculos semânticos entre conceitos correlacionados e estruturas e relações significativas, com conseqüências para o manejo de pacientes que apresentam múltiplas doenças e problemas de saúde, assim como para a integração de ambientes clínicos pertinentes a diferentes campos procedimentais e a distintas especialidades médicas.

Como uma conseqüência do refinamento sobre sua especificação da conceitualização do campo, a contribuição para a colocação de padrões pode ser conjeturada.

A implementação da ontologia, compreendida por distintas espécies de representação da informação, tais como tabelas, textos e listas em uma base de dados relacional, com seus diferentes tipos de conteúdo, tais como conceitos, identidades de conceitos, termos, nomes preferenciais, sinônimos, identidades de termos, e relacionamentos, pode produzir elementos de contribuição para preencher as necessidades de informação de diferentes classes de usuários.

Estes produtos também permitem a contribuição tanto para a concepção quanto para a avaliação de diferentes tipos de interface de sistemas, tais como textual, gráfica, linguagem representacional simbólica, de acordo com o uso de informação que é almejado e também com seu potencial usuário. Para o cientista da informação e o engenheiro do conhecimento, as estruturas formais e a axiomatização da ontologia constituem elementos relevantes de

informação. Para o profissional de atenção à saúde, consultas interativas e flexíveis e *displays* individualizáveis de unidades de informação correlacionadas podem representar um incremento diferencial em sua interação com sistemas de informação e de registro de dados clínicos.

Uma das potenciais propriedades emergentes da pesquisa é a possibilidade de sua contribuição para lidar com diferentes níveis de granularidade conceitual relativa a distintos tipos de usuários.

A modularização de seus componentes e a recombinação de seus elementos semânticos específicos permite a construção de padrões de representação do conhecimento diferenciados, com importantes conseqüências sobre os processos cognitivos dos usuários.

Finalmente, a metodologia utilizada e os objetivos delineados podem ser estendidos e empregados na esfera relativa a outros domínios do conhecimento humano especializado.

Tais extensibilidade e generalizabilidade metodológicas dizem respeito tanto aos domínios do conhecimento de caráter propriamente científico, quanto a áreas do saber nas quais a modelização da natureza dos conceitos e das relações entre os mesmos permite uma suficiente formalização lógica e ontológica. Desta maneira, sistemas de informação com propriedades de processamento caracterizando comportamentos de natureza inteligente podem ser desenvolvidos para esses domínios, incorporando o conhecimento especializado no desenvolvimento dos próprios sistemas, e ampliando as capacidades dialógicas e intelectivas dos profissionais e dos usuários que deles se utilizam.

REFERÊNCIAS BIBLIOGRÁFICAS:

BUNGE, M. *Treatise on Basic Philosophy*, v.3, Ontology I: The Furniture of the World. Dordrecht: D. Reidel Publishing Company, 1977

- COOPER, G.F. An Experiment Comparing Lexical and Statistical Methods for Extracting MeSH Terms from Clinical Free Text. *Journal of the American Medical Informatics Association*, v.5, n.1, p.62-75, 1998
- CRUSE, D.A. *Lexical Semantics*. Cambridge : Cambridge University Press, 1997
- DAHLBERG, I. *Ontical Structures and Universal Classification*. Bangalore: Sarada Ranganathan Endowment, 1978
- DAHLBERG, I. Knowledge Organization and Terminology: Philosophical and Linguistic Bases. *International Classification*, v.19, n.2, p.65-71, 1992
- GUARINO, N. Ontologies and Knowledge Bases – Towards a Terminological Clarification. in *Towards Very Large Knowledge Bases, Knowledge Building & Knowledge Sharing*. ed. Mars, N.J.I., p.25-31, Enshede: University of Twante, 1995
- GUARINO, N. *The Ontological Level*. www.ksl.stanford.edu/knowledge-sharing/papers/readme
- GUARINO, N., WELTY, C. A Formal Ontology of Properties. In: *Proceedings of 12th Int. Conf. On Knowledge Engineering and Knowledge Management*. Springer Verlag, 2000
- HABEL, C., PROBBENOW, S. & SIMMONS G. Partonomies and Depictions: A Hybrid Approach. In: *Diagrammatic Reasoning – Cognitive and Computational Perspectives*. Ed. Glasgow, J., Narayanan, N.H. & Chandrasekaran, B. Menlo Park: The AAAI Press, 1995
- HJORLAND, B. & ALBRECHTSEN, H. Toward a New Horizon in Information Science: Domain-Analysis. p.400-422. *Journal of the American Society for Information Science*, 1995
- JOURNAL of the American Medical Informatics Association. *Focus on the UMLS*. v.5 n.1 Philadelphia: Hanley & Belfus, 1998
- LEVESQUE, H.J. & LAKEMEYER, G. *The Logic of Knowledge Bases*. Cambridge: The MIT Press, 2000
- SOWA, J.F. *Knowledge Representation – Logical, Philosophical, and Computational Foundations*. Pacific Grove: Brooks/Cole, 2000
- SPARCK JONES, K. & KAY, M. *Linguistics and Information Science*. ed. Boroko, H. New York: Academic Press, 1973
- UMLS Knowledge Sources. 12th Edition. Bethesda: National Library of Medicine, 2001