

## A Teoria do Conceito Revisitada em Conexão com Ontologias e Metadados no Contexto das Bibliotecas Tradicionais e Digitais

*Concept Theory Revisited in Connection with Ontologies and Metadata in Traditional and Digital Libraries Context*

por Lídia Alvarenga

**Resumo:** Ensaio baseado na literatura e na reflexão dedutiva da autora, bibliotecária e atualmente professora da área de tratamento da informação e bibliometria, na Escola de Ciência da Informação da UFMG. O pressuposto básico do trabalho se ancora na afirmação de que o conceito se constitui no componente invariável do processo de organização de bibliotecas tradicionais e digitais. Sua fundamentação conceitual aborda: a) fundamentos da teoria do conceito, sob a ótica da ciência da informação e suas relações com os metadados, ontologia e o processo de classificação; b) peculiaridades e restrições intervenientes no processo de organização e tratamento da informação; c) alguns fatores específicos inerentes à representação e recuperação de informações. O artigo ressalta potencialidades, ainda pouco exploradas, dos recursos tecnológicos na criação de bibliotecas digitais e chama atenção para o fato de que as atuais tentativas de classificação automática de documentos repetem as experiências de muitas décadas, realizadas na área da ciência da informação, alcançando-se resultados ainda bastante limitados. As reflexões empreendidas levam a que se conclua que, para superar o problema do registro e análise conceituais, tornar-se-ia necessária a criação de uma nova linguagem entendida por máquinas e seres humanos e que possibilitasse igualmente a codificação de conceitos discursivos, gráficos e sonoros. O tratamento da web, tal como vem sendo proposto por alguns sistemas de metadados, torna-se factível, caso este se processe em nível da produção do documento primário, na construção de hipertextos, ou seja, aprioristicamente. O tratamento a posteriori ao processo de produção apresenta-se como um recurso pouco exequível, devido a grande quantidade de objetos de diferentes tipos, expostos e dispersos na web, a cada segundo e a impossibilidade de execução pela máquina de processos interpretativos dos conceitos que compõem os objetos digitais.

**Palavras-chave:** Bibliotecas digitais; Metadados; Ontologias; Teoria do conceito; Bibliotecas

**Abstract:** This essay is based on author thoughts originated from scientific literature and her experiences and professional and academic life in library science schools and librarianship context in Brazil. The article basic presupposition posed is that concept, as the principal thought element, represent the unchanging part in the process of migration from traditional to digital libraries. As conceptual foundation the article outstands: concept theory, ontology, metadata and classification. The results point to changes in resources production and technical processing in digital context, concerning with the weakness of actual language, that could have been changed in the future. The present endeavour toward to documents automatic classification repeats old experiences in information science research arena. The information treatment from web documents today only would be possible a priori, i.e., with the author responsibility, in the moment of the genesis of documents; the same task a posterior could get impossible, caused by the large amounts and sorts of digital documents in www.

**Key-words:** Digital libraries; Libraries; Metadata; Ontology; Classification; Concept theory

### Introdução, problema e contexto

Este artigo é constituído de um ensaio em que se propõe delinear uma estrutura teórico-conceitual que vise à fundamentação de pesquisas no campo das bibliotecas digitais. Seu pressuposto básico se ancora na afirmação de que o conceito se constitui no componente invariável do processo de organização de bibliotecas tradicionais e digitais e sua fundamentação conceitual aborda: a) fundamentos da teoria do conceito sob a ótica da ciência da informação e suas relações com os metadados, ontologia e o processo de classificação; b) peculiaridades e restrições intervenientes no processo de organização e tratamento da informação; c) alguns fatores específicos inerentes à representação e recuperação de informações, fatores esses que se localizam no complexo espaço digital do qual participam produtores de documentos, bibliotecários e usuários.

O presente texto emerge de reflexões a partir da literatura publicada e da experiência e vivência da autora no campo da biblioteconomia, como bibliotecária, coordenadora de sistemas de informação nas áreas dos transportes e da educação, pesquisadora e professora universitária, atuando atualmente no ensino e na pesquisa na área do tratamento da informação e bibliometria.

Reportando-se a Goethe, Vickery já afirmava em 1986 que toda a representação do conhecimento é simbólica e constitui-se em matéria que vem preocupando o mundo da documentação desde a sua origem.

Partindo-se da definição de que representar significa o ato de colocar algo no lugar de, pode-se classificar em nível primário a representação simbólica feita pelos autores, no momento da expressão dos resultados de suas observações metodicas sobre a natureza, utilizando-se das linguagens disponíveis no contexto da produção e comunicação de conhecimentos. Nessa representação as línguas dos diversos povos e línguas de especialidades desempenham papel primordial.

Em instância profissional de intermediação entre o conhecimento e seus usuários potenciais, os bibliotecários vêm desempenhando, no exercício de suas funções profissionais, diferentes tipos de representações, que podem ser vistas como de natureza secundária, envolvendo a substituição da informação primária por registros específicos a ela relativos. Nos catálogos dos sistemas e serviços de informação mantidos especialmente no âmbito de bibliotecas, a informação primária textual, sonora ou gráfica encontra-se representada por novos conjuntos de informações, dando origem a produtos que poderiam ser considerados metainformações ou metadocumentos. Na catalogação bibliográfica tradicional, o documento é representado por um conjunto de informações relativas à sua descrição física e pontos de acesso, representação esta preparada e armazenada em um contexto físico independente do documento primário.

Considerando-se que o novo espaço digital tenha se expandido, pode-se dizer ilimitadamente, os acervos de objetos digitais, gerados em âmbitos particulares e institucionais se multiplicaram, tanto no que tange à sua tipologia quanto à sua complexidade. Dentre esses objetos encontram-se dentre outros, textos das mais variadas naturezas, home-pages, listas de discussão, software em geral, padrões, protocolos e outros dispositivos essenciais à implementação das tecnologias digital, objetos esses que passam a requerer diferentes tipos de índices e pontos de acesso, visando-se a uma recuperação bem sucedida.

À medida que as tecnologias da informação foram sendo criadas, disponibilizadas e aperfeiçoadas, os sistemas de representação e recuperação de informações documentais assistiram a uma extração dos limites dos tradicionais catálogos referenciais em fichas, alcançando as bases de dados online. Essa nova situação resultou em uma grande mudança no ambiente dos sistemas de representação, assistindo-se a um gradual envolvimento de bibliotecários com profissionais oriundos de outros campos de conhecimento, destacando-se dentre estes o pessoal da ciência da computação e da lingüística e também os produtores comerciais de bases de dados, nos mais diversos campos de conhecimento.

Estimuladas pelo apelo advindo das novidades tecnológicas, relacionadas com as bibliotecas e arquivos digitais, algumas pessoas costumam ter em nossos dias uma visão equivocada das bibliotecas tradicionais, considerando-as como na situação descrita por Rasmussen (1999, p. 397), em que tais instituições são vistas essencialmente como detentoras de coleções de livros empoeirados, acessíveis através de catálogos em fichas. Segundo o citado autor, essas pessoas demonstram desconhecer o fato de terem sido as bibliotecas as primeiras instituições a fazerem uso dos sistemas automatizados de recuperação de informações.

Embora o reconhecimento formal de bibliotecas digitais seja relativamente novo, é conhecido que algumas bibliotecas tradicionais, especialmente nos países econômica e tecnologicamente mais desenvolvidos, já vinham se aperfeiçoando no que se refere ao uso das tecnologias de processamento eletrônico de dados, desde que começaram a organizar e oferecer acesso a bases de dados referenciais online, e que os OPAC's (on-line public access catalogs) passaram a ter uma maior presença no contexto de bibliotecas, ambos fenômenos que, em escala de maior intensidade, remontam ao final dos anos 60 e início dos anos 70.

O advento do mundo digital ocasionou, entretanto, novas mudanças no trabalho de autores e bibliotecários, fazendo com que estes, de repente, envolvessem-se com novas possibilidades tecnológicas, diretamente incidentes nos processos de produção, armazenagem, tratamento e recuperação de documentos e informações, alterando de forma radical seus processos de trabalho e produtos finais. Em decorrência dessa verdadeira ruptura, compartilham hoje do mesmo espaço de produção os autores de textos, sons e imagens, fixas ou em movimento, geradas em todos os quadrantes do mundo civilizado, juntamente com profissionais dedicados às tarefas de organização de conjuntos específicos de e informações e de criação de meios para recuperá-los. Nesse complexo processo, vem sendo construído um ilimitado e cambiante universo constituído dos denominados objetos digitais.

O meio digital se constituiu, portanto, no espaço sem precedentes para o registro e recuperação de documentos textuais, sonoros e imagéticos e que, ao ensejar uma enorme gama de possibilidades de armazenagem, memória e formatos, passou também a requerer novos elementos facilitadores de sua recuperação.

Pode ser destacada como uma das principais mudanças ocorridas, com a implantação e desenvolvimento da plataforma www e da internet, a contingência denominada de desterritorialização do documento, ou seja, o fato de passar o documento a ter sua materialidade desvinculada da forma física anterior, assumindo a forma digital, que possibilita uma organização espacialmente integrada de textos, imagens e sons.

Com o advento e intensificação do uso da internet, a chamada biblioteca digital tornou-se realidade. Em constante evolução, esse novo e potente meio de comunicação vem possibilitar, finalmente, a concretização do sonho da biblioteca sem paredes, da grande biblioteca universal, preconizado por pensadores, dentre os quais podem ser citados os enciclopedistas franceses e Jorge Luís Borges, que um dia sonharam com a possibilidade de

existência de uma biblioteca que pudesse abarcar a totalidade dos conhecimentos produzidos, em todos os recantos do mundo; a concretização, enfim, de uma situação dantes utópica e recorrente em inúmeros momentos da evolução do pensamento científico e filosófico da humanidade.

Podemos identificar em artigo publicado por Vickery (1986), ainda nos anos de 1980, uma verdadeira antevisão de que, com o advento do computador e a conseqüente mudança de paradigma quanto à própria materialidade dos documentos, o trabalho de tratamento da informação extrapolaria as fronteiras dos campos da biblioteconomia e correlatos, envolvendo novas situações, fatores e elementos diversos inerentes ao trabalho de profissionais vinculados a diferentes campos disciplinares, ampliando o campo de pesquisa relativo à biblioteca. Como exemplos desse tipo de situação o autor destaca:

- a estrutura dos registros e arquivos em bases de dados;
- as estruturas de dados em programas de computador;
- as estruturas semânticas e sintáticas da linguagem natural;
- a representação do conhecimento na inteligência artificial;
- os modelos de memória humana.

Em todas essas instâncias, tornar-se-ia necessário que se decidisse como os vários tipos de conhecimentos, informações e instrumental tecnológico deveriam ser representados [ou identificados] [\[\\*\]](#). Segundo Vickery, diferentes tipos de dados requereriam diferentes técnicas de representação. Daí a diversidade de tipos de "tratamento da informação", desempenhados nos contextos das bibliotecas, arquivos, museus e hoje na www, a grande plataforma de exposição de objetos digitais, com a qual tão rapidamente aprendemos a conviver.

A representação do conhecimento, em nossos dias, não comprehende somente a substituição do documento primário por uma informação catalográfica. Considerando-se que o documento não se acha fisicamente em outro espaço, mas no próprio meio que lhe proporciona materialidade, novas formas de se criar índices de recuperação foram ensejadas.

No novo contexto de produção, organização e recuperação de objetos digitais, as metas de trabalho não se restringem à criação de representações simbólicas dos objetos físicos constantes de um acervo, mas comprehendem estabelecimento dos denominados metadados, muitos dos quais podem ser extraídos diretamente dos próprios objetos, constituindo-se esses em chaves de acesso a serviço dos internautas.

Por biblioteca tradicional, entende-se a instituição social, criada com a finalidade de se adquirir, tratar, armazenar, disseminar e disponibilizar documentos, em sua materialidade convencional, independentemente de sua forma física ou suporte (livro, periódico, mapa, gravura, filme, cd-rom, etc.), armazenados em instituições determinadas. As bibliotecas constituem-se, pois, em sistemas formais de preservação e memória dos saberes e do conhecimento público de natureza científico, tecnológico e cultural, na acepção de John Ziman, caracterizável por possuir as qualificações de conhecimento derivado, acumulado e publicado.

É consenso afirmar que, sem bibliotecas, não haveria ciência, tecnologia ou as demais manifestações, ficando o conjunto do conhecimento restrito aos saberes limitados ao escopo da memória oral coletiva ou dos eventuais registros humanos dispersos.

Por biblioteca digital entende-se, por outro lado, um conjunto de objetos, concebidos em meio digital, desmaterializados de sua condição física tradicional, constituídos de funções inteiramente novas que lhes garantem a hipertextualidade e caráter multimidiático, tornando-os passíveis de acatar novos e peculiares arranjos e tipos de abordagem, no processo de sua recuperação.

De forma simplificada, a biblioteca digital, pode ser definida como um conjunto de objetos digitais construídos a partir do uso de instrumentos eletrônicos, concebidos com o objetivo de registrar e comunicar pensamentos, idéias, imagens e sons, disponíveis a um contingente ilimitado de pessoas, dispersas onde quer que a plataforma www alcance.

A partir do pensamento de autores contemporâneos (Watson-Verran e Turnbull apud Bishop et al, 2000), Dias (2001) destaca do conceito de biblioteca digital a afirmação de que : "[...] a biblioteca digital parece estar se firmando como um conjunto de artefatos, conhecimento, práticas e uma comunidade que engendra compromissos realísticos assumidos por profissionais da informação, analistas de sistemas e usuários". O conceito, segundo a síntese feita por Dias, expressaria um amálgama formado pela mistura de lugares, corpos, vozes, habilidades, práticas, mecanismos técnicos, teorias, estratégias sociais e trabalho coletivo, elementos estes que em conjunto constituiriam a instância dos conhecimentos/práticas tecno-científicas, com as práticas ativas e em evolução

Com inspiração no texto de Jessica Milstead, em seu trabalho no Annual Meeting of The American Society for Information Science, edição 1999, pode-se afirmar que o paradigma da biblioteca digital é diferente da biblioteca tradicional. Se por um lado, as bibliotecas há alguns anos tenham procurado transpor seus próprios muros para atender aos pedidos de seus usuários, já se utilizando de novas tecnologias eletrônicas para a organização de seus catálogos, por outro lado a dependência de acervos físicos dispersos constituía-se no fator primordial de sua existência. Com o advento das redes e sistemas de bibliotecas, ainda na forma tradicional, chegou-se a vislumbrar a existência de bibliotecas sem acervo, sem paredes, sem livros, ou seja, já se podia conviver com a possibilidade de se prestar um serviço de acesso a documentos primários, localizando-se e acessando-se documentos por meio de catálogos online, obtendo-se os textos primários através dos serviços de comutação e empréstimos entre bibliotecas.

Ampliando-se o escopo do conceito, poder-se-ia considerar a biblioteca digital um termo genérico que simboliza todo o conjunto de objetos digitais existentes, aleatoriamente colocados na web. Já num contexto mais limitado, afirma-se que uma biblioteca digital pode estar circunscrita a um segmento definido e identificável de objetos, organicamente articulados e voltados para interesses específicos de um grupo.

Ao invés de se acessar fisicamente os documentos utilizando-se de catálogos, sejam estes processados eletronicamente ou não, os objetos digitais na biblioteca digital se posicionam em sua completude, diretamente no ciberespaço.

As mudanças que ensejaram a existência das bibliotecas digitais constituem-se em verdadeira ruptura com o status quo, considerada por muitos de forma semelhante à invenção da imprensa por Gutenberg, por volta do ano de 1500.

Atualmente, os principais problemas técnicos e de utilização relativos às bibliotecas digitais poderiam ser resumidos: a) nas dificuldades relacionadas ao financiamento de infra-estrutura e de acesso amplo pelas diversas camadas da sociedade, especialmente em países em desenvolvimento; b) na organização do enorme potencial de documentos disponíveis na web; c) na modelagem e arquitetura de dados que viabilizem um melhor acesso dos usuários da rede digital, em meio à massa exponencialmente crescente de recursos disponíveis.

Embora detentora de inesgotável capacidade de memória, socialização e poder de comunicação, antes apenas vislumbrados, a internet vem sendo às vezes desqualificada, por ser composta de fontes transitórias e de proveniência pouco segura. Segundo Woodward, as fontes hoje existentes na internet perfazem numericamente muitos milhões de unidades, distribuídas em milhares de servidores, cada um com seus hardware e software próprios. Constatata-se, entretanto, que esse material é pouco organizado, de qualidade e estabilidade variável, sendo portanto difícil conceituá-lo, acessá-lo, pesquisá-lo, filtra-lo e referenciá-lo utilizando-se dos processos tradicionalmente executados no âmbito da biblioteconomia e da ciência da informação.

Ao prever a evolução do sistema de editoração eletrônica em meio digital, essência das bibliotecas digitais, Lancaster ressalta que, num primeiro estágio, a publicação eletrônica portaria as mesmas características das publicações convencionais, em papel, da mesma forma que os catálogos em linha nada mais são que catálogos em fichas em forma eletrônica. Uma versão eletrônica seria um análogo direto de uma página impressa. As publicações verdadeiramente eletrônicas somente surgiriam, segundo o autor, numa segunda fase,

*"uma fase na qual as novas publicações serão projetadas ad initio para exclusiva distribuição eletrônica. Nessa fase serão exploradas as possibilidades reais da eletrônica e as novas publicações se despojarão por completo das limitações inerentes à impressão e papel, quer dizer um modo de apresentação da informação estático, essencialmente linear".*

Falando de possibilidades das publicações digitais, o autor destaca que:

*"[...] elas serão dinâmicas e multidimensionais, ao invés de estáticas e unidimensionais. Uma descrição narrativa será apresentada como um hipertexto, reorganizável sob o controle do usuário. Muitas ilustrações estáticas serão substituídas por animação ou por modelos analógicos eletrônicos (por exemplo, ilustrações de fenômenos físicos, operações de equipamentos, realização de uma experiência). Também haverá a complementação do texto ou das ilustrações por meio de saída de áudio, caso sejam adotadas certas formas de distribuição eletrônica" (Lancaster, 1991, p.268).*

Cumpre ressaltar que as publicações eletrônicas brasileiras parecem, em sua maioria, ainda se enquadrar na primeira fase prevista em que o uso das novas tecnologias acha-se muito aquém de sua potencialidade.

## Conceito como elemento invariável

O que mudou com o advento das bibliotecas digitais não foram, portanto, os conteúdos nem a essência das mensagens veiculadas, mas a forma e o meio através dos quais os documentos passaram a ser produzidos e registrados: um meio mais leve, ágil e dinâmico em suas possibilidades de processamento e comunicação. O que na realidade mudou com o advento do meio digital foi a possibilidade de disponibilização virtual de documentos multimídia completos, num espaço amplo e interconectado de acesso remoto, assim como a capacidade de memória e a expansão das relações entre documentos e componentes intervenientes nos conceitos neles expressos. Também se constitui em fator de mudança a possibilidade de abordagem intertextual presente nos sistemas de hipertexto.

Os conteúdos dos documentos, sejam eles digitais ou não, não sofreram quaisquer alterações; porquanto suas existências se referem a objetos, pensamentos e idéias presentes no mundo cognoscível e no imaginário da humanidade, em sua necessidade básica de comunicação. Não houve uma alteração no sistema de gênese e registro dos pensamentos, através de signos verbais lingüísticos, sonoros ou gráficos. Os autores continuam produzindo textos, sons e imagens, utilizando-se das linguagens disponíveis e consensualmente aceitas. A parte substancial dos documentos que se refere a seu conteúdo, à sua atinência, ao seu significado, os enunciados que compõem os conceitos neles contidos, tudo isso continua invariável; tudo isso é uma contingência com a qual as máquinas têm que conviver e daí decorre a dificuldade primordial do processo de tratamento da informação, antes em ambientes tradicionais e hoje na web.

De conceitos mais ou menos intensamente construídos são produzidos os documentos que, relativos aos referentes do mundo físico, social e emocional observados, percebidos e sentidos pelo homem, caracterizam-se ao mesmo tempo como pertencentes igualmente às esferas específicas da unidade e da diversidade. Classificando-se ora como entes distintos entre si, ora portadores de características similares que lhes outorgam relações de interdependência e interdisciplinaridade, os conceitos que formam os documentos envolvem fatores que muitas vezes vêm "sacudir" (na acepção foucaultiana do termo) a certeza de especificidade que pode decorrer de objetividades inadvertidas oriunda de nossos sentidos imediatos.

O que na realidade se classifica em uma biblioteca tradicional ou digital não são os documentos, mas os conceitos contidos nesses documentos. Deve-se ressaltar que não somente os textos são formados de conceitos, mas também as imagens fixas e em movimento e os sons, fato que vem dificultar sobremaneira os projetos de classificação automática de documentos na web.

O estudo da gênese e formação dos conceitos vem sendo feito ao longo da história do pensamento filosófico e científico, em diversos campos disciplinares, destacando-se a filosofia, a semântica, a psicologia, a biblioteconomia, a terminologia, sempre envolvendo as bases cognitivas e lingüísticas do conhecimento e saberes humanos.

Por teoria do conceito entende-se, neste trabalho, o conjunto de enunciados oriundos de pesquisas e reflexões pertinentes à complexa região epistemológica interdisciplinar que compreende o ato de representação, comunicação e preservação de objetos e pensamentos e cujo conhecimento integra os campos da linguagem, da psicologia cognitiva, da comunicação e da ciência da informação. Neste estudo, o conceito e a compreensão do que seja conceito é tema crucial, por pertencer à essência do trabalho de tratamento e organização da informação, compreendendo os processos de análise de assunto, classificação e recuperação da informação.

Embora muito se fale sobre o conceito, nem sempre se tem clareza de seu significado. Até mesmo partindo da própria etimologia do termo, existem discussões que merecem ser resgatadas. Dahlberg chama atenção para um fato relacionado aos componentes essenciais de um conceito relativo a um referente qualquer. Considerando-se que, para Aristóteles, o significado do conceito (*horos*) incluía três elementos, *logos*, *pragma* e *noema*, Dahlberg ressalta que, no processo de tradução de *horos*, a partir do pensamento do filósofo grego, Boethius tenha vertido o termo para o latim utilizando-se o vocábulo *terminus*, que privilegia somente o *logos*, o lado lingüístico do conceito; ficou portanto falha a correspondência do termo. Esta confusão ensejou, segundo a renomada teórica do conceito, a que filósofos posteriormente tenham preferido o uso de *terminus*, considerando estar incluído neste termo o elemento conceitual e também o elemento meramente lingüístico. Ainda segundo Dahlberg, é sabido que Christian Von Wolff (1679-1754) tenha traduzido corretamente por conceito o termo *horos*, que quer dizer conjuntamente signo (termo) e conteúdo.

O triângulo de Ogden & Richards, publicado no clássico livro "O significado de significado...", publicado em primeira edição em 1923, edição brasileira de 1972, tem sido muito utilizado para se representar o conceito devido a sua clareza e facilidade de compreensão. O referido diagrama espelha-se em um triângulo em que, no vértice superior encontra-se o objeto da realidade; no vértice da direita o conceito, formado de todos os

enunciados que podem ser proferidos sobre o referente; e no vértice da esquerda encontra-se o símbolo, o signo, termo, número, ícone, designando o conceito sobre o objeto referente.

São recorrentes, desde a antiguidade grega as discussões sobre as conexões entre as palavras e as coisas, entre os referentes (coisas, pensamentos), seus conceitos e respectivos símbolos usados para sua representação. Na Idade Média, as discussões sobre o uso de termos e conceitos, no intento de se definir as coisas, chegaram a dar origem a duas correntes diferentes de pensadores: os nominalistas e os conceitualistas.

No pensamento de Aristóteles, encontram-se contribuições bastante relevantes para a teoria do conceito, envolvendo o ato da cognição humana e a fatoração do conceito em categorias, dentre outras abordagens, ressaltando-se que sua influência nesse campo de conhecimento tenha sido reconhecida e referenciada em quase todas as épocas da evolução do pensamento filosófico e científico. Para Aristóteles, "saber seria ter muitos conceitos e conhecer significava três coisas":

1. formar conceitos, ou seja, constituir em nossa mente um conjunto de notas características para cada uma das essências que se realizam na substância individual.;
2. aplicar esses conceitos que formamos a cada coisa individual, colocar cada coisa individual sob um conceito. Chegar à natureza; contemplar a substância; olha-la e voltar para dentro de nós mesmos para procurar no arsenal de conceitos aquele que se ajustasse a uma singular substância; e formular um juízo;
3. embaralhar entre si esses diversos juízos, em forma de raciocínios que nos permitissem chegar à conclusão acerca de substâncias que não temos presentes.

Esta seria uma forma de se abordar a natureza física e abstrata desvendando-lhe os mistérios e enunciando verbalmente sobre cada faceta ou ângulo relacionado ao objeto observado, reconstruindo-se finalmente o objeto por meio da síntese de todas as facetas a ele relativas.

Um clássico trabalho sobre a gênese dos conceitos na mente humana e sua extração para o campo da organização e recuperação de informações é de autoria de Jesse Shera, um dos clássicos teóricos da biblioteconomia.

Shera publicou importante trabalho nos proceedings da Conferência Internacional de Classificação e Recuperação de Informações, realizada em Londres sob os auspícios da ASLIB, em 1957. Seu trabalho tem por objetivo o aperfeiçoamento da organização e recuperação de registros gráficos e ilustra muito bem como se desenvolve, no cérebro humano, o processo de formação de conceitos. Cada nova sensação quer se origine da experiência ou da leitura de algum texto é classificada no cérebro humano, de acordo com o acervo de experiências passadas, ali registrado, possibilitando a ordem que levaria ao saber.

A ciência seria o conhecimento coordenado, formado pela conjunção de duas ordens de experiências: a observação direta e a classificação dos resultados dessa observação, de acordo com as concepções já adquiridas sobre o tema. A primeira seria interpretada em termos da segunda, ou seja, de acordo com o acervo de conhecimentos (conceitos) já armazenados. Em outras palavras, cada nova experiência, cada nova sensação seria classificada no cérebro, de acordo com um padrão de relação, originário das experiências passadas e dessa forma, haveria incorporação do novo saber, estabelecendo-se a ordem, diminuindo a incerteza, deixando mais satisfeito o homem em seu intento de sobreviver.

A facilidade com a qual as novas percepções pudessem ser organizadas e classificadas, assimilando-se com a experiência passada, determinaria o processo de aprendizagem de uma pessoa. Se ao ler um texto uma pessoa fosse incapaz de classificar novas idéias, pensamentos e sensações, admitir-se-ia a incomprensibilidade do texto para aquela pessoa. Daí afirmação de que os atos de conhecer perceber e ter consciência seriam funções da forma como o sistema nervoso humano se encontra estruturado.

O homem tenderia a encontrar, desse modo, felicidade, conforto, segurança, através do reconhecimento de padrões familiares e, a menos que ele fosse particularmente aventureiro e temerário, ele se retrairia e se afastaria dos padrões que lhe fossem pouco familiares.

Os conceitos ou padrões se constituem na matéria do qual as classificações são feitas. Os conceitos podem se referir às coisas concretas e abstratas. Os conceitos e a formação de conceitos são o material para a construção das classificações bibliográficas.

Segundo Shera,

"Um conceito é uma rede de padrões de inferências, associações e relacionamentos que são predados ou ditos de outra forma trazidos em cena através do ato da categorização" [...] a cristalização ou formalização do pensamento inferencial, nascida da percepção sensorial, condicionada pela operação do cérebro humano e delineada pela experiência humana. Ela repousa na fundamentação de todo pensamento mas ela é pragmática e instrumental. É permanente e efêmera. Permanente porque sem ela, a cognição é impossível; efêmera porque ela pode ser rejeitada quando sua utilidade é esgotada" (Shera, 1957).

Livros e documentos são registros do conhecimento humano que é formado de conceitos. Durante séculos o homem tem categorizado os conhecimentos em uma hierarquia taxonômica que associa semelhanças ou separa diferenças, constantes dos conceitos. Os conceitos e os diversos elementos que contribuem para a formação de conceitos, mais ou menos intensos, constantes dos livros e outros documentos constituem-se no material da classificação. São essas unidades de pensamento e as relações entre elas, como estão presentes nas páginas dos documentos e na mente do usuário, que se constituem na primeira preocupação do bibliotecário e o compele a sucessivas considerações sobre o papel da classificação.

### **Tratamento e organização da informação na internet**

Nas bibliotecas digitais, as fontes não são catalogadas e classificadas como os bibliotecários entendem esses termos. Observa-se que essas técnicas nem sempre são citadas nos contextos de um possível tratamento da internet, por serem ferramentas muito caras, envolvendo um tempo incompatível com o universo de objetos digitais existentes na rede. O volume de informações livremente colocados na web torna impossível um tratamento da informação nos moldes tradicionais.

Em nossos dias, muitos são os profissionais que trabalham no desenvolvimento de ferramentas de organização e busca de informações constantes nos objetos digitais. Como consequência natural advinda da prática em situações similares, os bibliotecários deveriam constituir uma classe profissional com atuação mais marcante, na tarefa de organizar a internet. Na realidade entretanto o pessoal da área da ciência da informação vem se mantendo tímido, apresentando contribuições muito aquém de suas possibilidades.

No contexto da ciência da computação esforços vêm sendo empreendidos no sentido de se classificar automaticamente objetos digitais dispersos, procurando tornar mais eficaz o processo de recuperação de informações. Nesse sentido, muitos pesquisadores vêm desenvolvendo motores de buscas (search engines) e agentes inteligentes para acessar informações, fundamentando-se esse instrumental primordialmente em buscas por palavras, utilizando-se de sistemas vetoriais e estratégias booleanas, ou combinações de métodos de pesquisa. A linguagem natural vem sendo o meio de organização do universo da web, pois, considera-se impossível em um âmbito mais amplo, um processamento dos objetos digitais utilizando-se de linguagens documentárias estruturadas, usadas para o um tratamento da informação a posteriori, nas quais termos, simbolizando conceitos, não unidades lexicais (palavras), seriam usados como chaves de recuperação. O campo da modelagem de dados, arquitetura de dados, especialmente no que se refere à recuperação semântica ou por conceitos, acha-se fartamente referenciado na literatura de muitas disciplinas.

O desafio de organizar a internet extrapola hoje fronteiras disciplinares e torna-se a busca de uma solução para que realmente o valor do meio possa efetivamente ser usufruído em toda a sua potencialidade. Dentre os profissionais que se atêm a trabalhar nessa área destacam-se filósofos, lingüistas, profissionais da ciência da computação, da psicologia cognitiva, da lingüística e da semiologia, todos focalizando as dificuldades atuais e procurando "uma luz no fundo do túnel".

A descrição do projeto CATRIONA, Cataloguing and Retrieval of Information Over Networks Applications, citado por Woodward (1996), inicia-se com a assertiva de que a biblioteca sem paredes está se tornando realidade e de que "o problema de como o usuário encontra as fontes e serviços apropriados às suas necessidades devem ser primeiramente resolvidos". A resposta para essa pergunta poderia ser: "combinando uma adaptação de novas tecnologias de software com uma adaptação de métodos e práticas biblioteconómicas já estabelecidas".

A literatura tem sido vasta na ciência da informação sobre pesquisas e experiências de organização da internet, utilizando-se ferramentas tradicionais da biblioteconomia. Por outro lado, muitos são os argumentos contrários ao uso dos métodos da biblioteconomia para o tratamento de documentos na internet.

Na comunidade da computação e da eletrônica, muitos autores argumentam que, mesmo antes da internet,

especialistas no campo da biblioteconomia e da ciência da informação já admitiam que os esquemas de classificação, então usados, já eram antiquados e impróprios para classificar os temas novos presentes na literatura atual.

Outros recorrentes argumentos são evocados, tais como o que advoga que as classificações tradicionais seriam úteis apenas para arranjar estantes, mas consideradas impróprias para o tratamento dos documentos digitais dos nossos dias. Esses argumentos, entretanto, denotam não possuírem seus defensores o conhecimento da existência do trabalho de pesquisa que subjaz à construção e atualização dos sistemas de classificação tradicionais, e parecem desconhecer a classificação facetada, preconizada por Ranganathan, que permite uma grande multiplicidade de combinações de conceitos, possibilitando uma estruturação semântica dos conceitos de um documento.

É certo que, classicamente, a formação do bibliotecário convencional garante-lhe pouca fundamentação matemática e computacional, o que de certa forma limita uma ação mais intensa e segura por parte desse. A solução do estabelecimento de parcerias entre profissionais de diversas áreas, há muito vem sendo praticada por alguns grupos e a consequência da inexistência, especialmente no Brasil, de um trabalho mais integrado e contínuo de pesquisa do pessoal da ciência da informação com a ciência da computação e eventualmente da lingüística, faz com que se veja atrasado o desenvolvimento de bibliotecas digitais.

Enquanto a internet se apresenta como novas possibilidades de disponibilização e processamento de informações eletrônicas, bibliotecários e especialistas da informação se esforçam para desenvolver métodos para a descrição, organização e recuperação de objetos digitais acessados remotamente. Eles não estão sós nesse desafio, porque criadores, provedores e usuários de fontes eletrônicas nos meios acadêmicos, públicos e comerciais também se acham dedicados à pesquisa sobre esse vasto corpo de informações. Cada grupo tem abordado o problema de organização e acesso, a partir de seus próprios esquemas de referência. Vellucci adverte para a importância do reconhecimento das contribuições de cada um dos membros do grupo, para que o esforço cooperativo possa prover uma estrutura flexível para a organização e o acesso das fontes da internet.

O ponto positivo da estrutura da web é a sua habilidade para organizar, dispor e trocar dados descritivos originados de uma vasta variedade de criadores, em um sistema integrado de organização, pautado pela liberdade que o meio propicia.

A informação eletrônica em rede é, na maioria das vezes, transitória e carente de desejáveis controles de qualidade e estabilidade, providos pelos familiares sistemas de peer-review, praticados nos processos tradicionais de edição de publicações.

## **Classificação**

No processo de interação do ser humano com o universo, existem possibilidades infinitas de se perceber objetos concretos e abstratos e desse processo origina-se o conhecimento.

Uma das grandes dificuldades que vem sendo ressaltada no processo de recepção, e de o tratamento do conhecimento, para fins de preservação e acesso, constitui-se no fato de que as tentativas de se classificar seres, coisas e textos que sobre estes são produzidos, revestem-se a priori da constatação de que as coisas, e portanto os conhecimentos, não se reduzem ao que deles pode ser visto explicitamente, mas, apenas poderão vir a ser devidamente compreendidos, a partir da atitude filosófica do compreender que, segundo Paul Ricoeur, significa o ato da interpretação criadora de sentido. A esse problema pode-se acrescentar a situação delicada e complexa de intermediação em que se colocam os profissionais da informação entre esta e seus usuários.

Assim sendo, ao se tentar classificar objetos, seres ou idéias, não é suficiente que se observe um objeto captando sua essência e características, ou se considere, apenas, o nível lexical que define essas realidades. É absolutamente necessário que se aprofunde no conhecimento das relações entre as similitudes e diferenças entre esses objetos, extrapolando uma visão individual daquele objeto e, a partir dos termos que os simbolizam, fazendo o mesmo aprofundamento nos seus conteúdos semânticos ou conceituais. O processo compreende ainda uma atitude hermenêutica ou interpretativa, igualmente indispensável à análise e à classificação de um universo de coisas. O ato de interpretar é inerente a qualquer abordagem ao conhecimento, qualquer que seja o grau de suposta clareza com que este tenha sido produzido.

Entretanto, a despeito dessa constatação o homem não deixa de classificar, perfeita ou superficialmente as coisas e pensamentos, até porque este é o processo natural através do qual se torna possível o entendimento dos

pensamentos expressos pelas pessoas, sejam de forma oral ou escrita.

Nesse processo de comunicação amplo, estabelecido entre os homens em suas tentativas de compreensão e descrição do universo, a mediação da linguagem ocupa um papel primordial e, no contexto da biblioteconomia e da ciência da informação, o tratamento da informação e conhecimento registrado se depara com uma tarefa complexa que assim poderia ser resumida: a despeito de todas as fragilidades dos atos de conhecer e comunicar, envolvendo coisas, seres, palavras, imagens e sons, torna-se imperativo que se encontre uma forma de se construir interfaces entre os acervos de documentos e informações e seus usuários.

Considerando-se o aspecto da representação simbólica do conceito, emerge o papel da terminologia, no contexto amplo da produção e organização da ciência e da tecnologia. Sabe-se que nenhuma disciplina pode progredir sem a socialização da linguagem entre os membros da comunidade que a sustenta.

Satija (2000), pesquisador e bibliotecário indiano, detentor de vasta produção científica no campo da ciência da informação e estudioso da contribuição de Ranganathan, afirma que este considerava a terminologia científica duplamente importante para os profissionais e pesquisadores da área da biblioteconomia e da ciência da informação: primeiramente, pela necessidade de que os bibliotecários e cientistas da informação têm que entender a terminologia de sua própria disciplina como professores e pesquisadores da área; em segundo lugar, Ranganathan destaca a importância da compreensão, pelo mesmo grupo de especialistas, para o entendimento da terminologia técnica das demais disciplinas, com a finalidade de organização do conhecimento dos diversos campos de conhecimento (p. 221).

Deduz-se da complexidade do ato de classificar as implicações nas tentativas de uma classificação automática de documentos, independente da intervenção humana. Alguns dizem que a matemática é capaz de representar todas as manifestações do pensamento humano, embora hoje ainda estejamos órfãos dessa linguagem universal aplicada a um entendimento amplo entre os homens.

## **Ontologias e metadados**

Atualmente, em contextos externos à ciência da informação, a palavra ontologia, muitas vezes usada no plural, ontologias, designa a faceta semântica da representação dos seres, dos entes, aquilo que se convenciona chamar de assuntos, conteúdos temáticos dos registros sobre a realidade.

Segundo o dicionário Aurélio (Ferreira, 1986) a ontologia consiste "na parte da filosofia que trata do ser enquanto ser, isto é, do ser concebido como tendo uma natureza comum que é inerente a todos e a cada um dos seres".

No comentário sobre o verbete, recorrendo a uma frase do escritor Alexandre Herculano, o dicionário aponta para as dificuldades imbricadas no problema ontológico da existência e essência dos seres, ao evocar dois ilustres filósofos que no ocidente se dedicaram ao estudo da teoria do conhecimento: "Com Kant, o universo é uma dúvida; com Locke é dúvida nosso espírito; e num destes abismos vêm precipitar-se todas as ontologias" (Herculano, Lendas Narrativas II, p.107, apud Ferreira, 1986). Complementando o que já foi discutido nos itens referentes à teoria do conceito e à classificação, o tratamento da informação dependente de uma abordagem ontológica sofre, portanto, da eterna e intransponível dúvida metafísica que povoa as certezas da essência do conhecimento humano...

Segundo a teoria do conceito, esboçada anteriormente, aos seres concretos e abstratos correspondem os conceitos compostos de enunciados sobre esses mesmos seres e a esses conceitos são atribuídos seus respectivos nomes ou símbolos que, nos documentos textuais mais convencionais acham-se representados por termos. Daí as relações entre ontologias, conceitos e terminologia.

O trabalho de Guarino [guarino@ladseb.pd.cnr.it](mailto:guarino@ladseb.pd.cnr.it) enfoca as ontologias do ponto de vista do desenvolvimento de sistemas baseados em conhecimento [Knowledge-Based Systems, KBS] e discute como os princípios da ontologia formal e engenharia ontológica podem ser usados, na prática da engenharia do conhecimento.

Na realidade, os estudos hoje denominados ontológicos vêm sendo retomados como substrato para as pesquisa de modelagem cognitiva de conhecimentos. O citado autor aponta em seu trabalho uma forte conexão entre ontologia formal e análise conceitual; e afirma que a engenharia do conhecimento pode contribuir para estabelecer as fundações de um campo emergente, a engenharia ontológica.

Da leitura de trabalhos dessa natureza, pode-se constatar que novos campos estão surgindo, dedicados a tratar de realidades essencialmente presentes nos domínios da ciência da informação. Uma interpretação desse fenômeno

resultaria em conclusão que aponta para a impossibilidade de hegemonia e manutenção dos domínios de velhos campos do conhecimento, desde que estes não apresentem resultados que atendam ao interesse do mercado ao qual a ciência contemporânea também se acha a serviço. Daí a urgência de desenvolvimento da pesquisa envolvendo os conhecimentos da área da ciência da informação.

Voltando à linha de raciocínio do presente trabalho, verifica-se que metadado etimologicamente, quer dizer "dado sobre dado"; dado que descreve, a essência, atributos e contexto de emergência de um recurso (documento, fonte, etc.) e caracteriza suas relações, visando-se seu acesso e uso potencial. O prefixo grego meta significa mudança, posterioridade, além, trancendência... (Ferreira, 1986).

Metadado pode ser considerado sinônimo de ponto de acesso, termo da área da catalogação, e parece ter sido cunhado em contextos externos à ciência da informação.

Os denominados padrões de metadados, atualmente presentes no instrumental de software para criação de objetos digitais, são considerados as linguagens de marca que, segundo Bax:

*"identificam, de forma descritiva, cada 'entidade informacional' digna de significado presente nos documentos, como por exemplo parágrafos, títulos, tabelas, gráficos. A partir dessas descrições os programas de computador podem melhor tratar ou processar a informação contida em documentos eletrônicos" (Bax, 2001, p. 32-3).*

Linguagens de marca diversas foram desenvolvidos por comunidades responsáveis pela criação ou provimento de acesso a fontes eletrônicas interligadas, desenvolvidas utilizando-se de métodos e padrões diversos para a organização de recursos digitais presentes na web.

Deve estar claro que, tal como na catalogação, os metadados não se restringem a dados temáticos, constantes no texto mas, podem também incluir outros tipos de pontos de acesso, relacionados à descrição física e ao contexto de produção, dependentes do tipo de objeto em questão.

A atribuição de metadados a priori, no momento de produção do objeto, por seus próprios autores ou a posteriori, após a inclusão do objeto no espaço digital, parece ser o ponto focal da maior importância na discussão sobre os metadados. Considerando que a organização da www é basicamente assunto de seus próprios organizadores, CRAIGMILE, citado por Woodward, acredita ser a internet uma fonte pouco estável para permitir uma organização a posteriori; essa constatação também se ancora no uso intensivo das novas tecnologias e no fato de serem as fontes da web, prioritariamente, produtos resultantes da competição do mundo dos negócios.

Outra dificuldade na criação de sistemas automáticos de recuperação da informação consiste na identificação pela máquina de conceitos presentes nas imagens e sons, que prescindem de uma escrita utilizando-se de termos convencionais da linguagem. Uma ressalva pode ser feita para os documentos sonoros, registrados através da música escrita, que é um símbolo lingüístico bastante concreto e possível de leitura pela máquina.

Sintetizando, os metadados incluem pontos de acesso expressos igualmente nos próprios textos, nas imagens e registros sonoros, presentes no meio digital, muitos desses não sendo passíveis de marcação do texto como nos padrões baseados nas linguagens de marca mais tradicionais HTML e SGML.

O trabalho de Bax (2001) publicado na revista Ciência da Informação trata das "markup languages", dentre as quais se incluem o padrão Dublin Core desenvolvido a partir do padrão MARC, criado originalmente para registros catalográficos.

A atribuição de metadados a partir da marcação dos textos em sua materialidade digital já se constituía em processo previsto, há décadas, por bibliotecários e pesquisadores. Lancaster afirma que a literatura do final da década de 1980, já fazia previsões relativas ao futuro dos serviços de recuperação de informações, sintetizando o autor que um dos pontos de convergência dessas previsões seria a integração entre as editorações primária e secundária. As publicações secundárias e catálogos isolados tenderiam a desaparecer, com o advento das bibliotecas inteiramente digitais.

Lancaster afirma, retomando os pensamentos dos autores Goldhar e Burchinal que:

*"Na medida que se tornam disponíveis em linha textos completos de periódicos, as fronteiras tradicionais entre editoração primária e secundária se tornam cada vez mais imprecisas. Burchinal (1977) predisse, na década de 1970, que por volta da década de 1990 os serviços secundários (serviços de acesso) seriam*

*obtidos automaticamente, a partir de materiais primários de base eletrônica e estariam disponíveis simultaneamente com a edição da literatura primária. Os usuários poderiam mover-se entre essas ferramentas de acesso e os textos completos dos documentos, a partir do mesmo terminal em linha". (Goldhar, 1979; Burchinal, 1977, apud Lancaster, 1991).*

Vellucci revê uma grande variedade de esquemas de metadados usados pelas comunidades da biblioteconomia e da ciência da informação, e examina seus próprios backgrounds, desenvolvimento e implementação. Alguns dos esquemas foram desenvolvidos dentro de uma comunidade específica ou domínio, como por exemplo, IAFA templates, MARC, Text Encoding Iniciative (TEI) Headers, Encodea Archival Description (EAD), Computer Interchange of Museun Information (CIMI), Governmente Information Locator Service (GILS) e Content Standards fo Digital Geoespacial Metadata (CDSGM). O conteúdo e sintaxe de cada tipo de metadado varia, dependendo das necessidades específicas de seus usuários. O trabalho de Milstead e Feldman (2001), dentre outros, discorre sobre sistemas de padrões de metadados em desenvolvimento sob os auspícios de diferentes grupos de pesquisa.

### **As pesquisas no campo da classificação e da recuperação semântica na web**

A filosofia e a psicologia do conhecimento, denominação esta usada por Dahlberg, foram campos específicos, onde a gênese e a organização do conhecimento constituíram-se em objetos privilegiados de estudo do conceito. A biblioteconomia e a ciência da informação, que desde seus primórdios contou com a atuação de pensadores eruditos e com ampla visão do campo científico, incorporou esses conhecimentos em seu corpo teórico, levantando as questões específicas da área, ou seja, utilizando métodos de pesquisas e princípios de reflexão, a partir de seus objetos específicos voltados aos processos de organização e recuperação de informações.

Em nossos dias constata-se uma evolução nas pesquisas nesse campo. Com a facilitação do tratamento de dados por computador, novos desafios vêm sendo colocados aos pesquisadores que, mais do que nunca, vêm formando equipes interdisciplinares com o objetivo de resolver problemas na área da formação e representação de conceitos.

Aos cientistas da informação se juntam pesquisadores do campo da informática, da neurologia e da psicologia, no intuito de continuarem as pesquisas que em última análise podem ser rotuladas de pesquisas no campo da representação do conhecimento e das idéias, tendo como objeto componente do triângulo do conceito, a linguagem e o termo como instrumentos de comunicação humana.

Se nas décadas de 1950, 1960 estávamos preocupados com a linguagem de recuperação de informação, na construção de classificações, de índices para recuperação de informações, antes mesmo de chegarmos à solução desejável dessa problemática, chegamos a pensar que a linguagem utilizada para designar os conceitos parece ter chegado ao seu limite, não dando mais conta de atender às necessidades de registro e recuperação de informações, dado ao meio rápido e potente hoje disponível - as redes eletrônicas de computadores.

Essa tendência aponta para um novo paradigma em que a linguagem, a representação do pensamento parece estar em busca de uma nova descoberta que venha transformar a maneira de se produzir e organizar conhecimentos. Os padrões de registros que já privilegiam a abordagem de imagens, o uso de ícones na comunicação e desenvolvimento dos sistemas especialistas e outras linguagens artificiais buscam intensamente encontrar superação do status quo, com a criação de novos processos que possam ser compatíveis com a agilidade e capacidade de armazenagem e processamento de informações das máquinas atuais.

A teoria do conceito, na qual se fundamentam as categorias de Aristóteles, retomadas com ligeiras alterações por Ranganathan, e que propiciam a análise dos conceitos em ângulos diversos de abordagem, as facetas, tem sido considerada por muitos a "luz no fundo do túnel" vislumbrada como insumo potencial para a criação de novas ferramentas destinadas à resolução do grande problema atual da ciência que pode ser identificado na confluência dos processos de produção, representação e organização de conhecimentos.

Outra pista para uma proposta de sistematização estrutural do conhecimento pode estar numa proposição lógica emanada do pensamento do filósofo Leibniz. Peter Jaenecke em artigo no periódico Knowledge Organization apresenta os fundamentos de uma pesquisa baseada no pensamento de Leibniz. Nesse projeto o resgate do pensamento de pensador, deve-se às suas reflexões sobre a sistematização do conhecimento.

De acordo com Leibniz:

*"A infinita variedade de pensamentos é somente aparente; ela se origina do número infinito de*

*possibilidades por meio das quais os pensamentos básicos podem ser combinados uns com os outros [...] todos os pensamentos humanos podem ser reduzidos a alguns poucos pensamentos originais que, quando identificados podem servir como um alfabeto dos pensamentos humanos. Similarmente ele insiste que a maioria dos conceitos podem ser subdivididos em subconceitos e esses em outros até que se atinja conceitos tão elementares que não sejam mais possíveis de qualquer decomposição"*

O Departamento de Ciência da Computação da UFMG vêm desenvolvendo pesquisas na área das bibliotecas digitais, desenvolvimento de ferramentas de modelagem de dados para documentos já existentes na web, ou em bases de dados referenciais on-line. São utilizados modelos matemáticos vetoriais, tentando classificar documentos para fins de recuperação. Dissertações e teses de doutorado têm sido defendidas.

O Departamento de Ciência da Informação da mesma universidade mantém por sua vez, na linha de pesquisa tratamento da informação, projetos em andamento, relacionados especialmente à criação de bibliotecas e arquivos digitais na UFMG e à modelagem de dados, utilizando-se das categorias de análise de Ranganathan e do uso de sintagmas nominais para classificação e recuperação de informações na web.

O projeto de modelagem de dados, a partir das categorias de Ranganathan, pretende propor um modelo de tratamento apriorístico, aplicado a um campo específico de conhecimento, ou seja, um método para disponibilizar para os criadores de hipertextos um instrumental que possibilite construir estruturas conceituais que orientem as relações entre os conceitos e elementos formadores de conceitos expressos nos links dos hipertextos.

As dificuldades encontradas são inúmeras, especialmente considerando-se a inexistência de arquivos para implementação das pesquisas. No Brasil as bibliotecas e arquivos digitais devem ser construídos, também com a finalidade de se constituir como universos de dados para possibilitar o desenvolvimento de pesquisas. O que se pretende na UFMG é que as bibliotecas digitais dos acervos de teses e produção científica se constituam também em universos para testes de modelagem de dados, num trabalho em parceria entre os departamentos de computação e ciência da informação. Essa parceria já está de certa forma instituída por meio de um Grupo de Estudos sobre Bibliotecas Digitais, que mantém um fórum virtual e vem compartilhando conhecimentos, através de encontros, participação em bancas de teses e dissertações e projetos em comum apresentados a agências financeiras da universidade e do país. O site do grupo de estudos é : <http://gebd.eci.ufmg.br/fórum>

## Conclusão

O conceito pode ser visto como fator invariante envolvido na concepção e tratamento de documentos, quando se fala da evolução das bibliotecas tradicionais para as digitais. Daí a complexidade das propostas de tratamento automático de objetos digitais, estando envolvidos nesse processo as chamadas ontologias, os metadados, cabeçalhos de assuntos, tesouros, pontos de acesso, dentre outros componentes.

Mudam-se os meios, sofisticam-se os instrumentos e surgem nomes novos para designar coisas velhas. Entretanto, a essência das coisas permanece. Os tempos digitais acenam para a concretização de sonhos: da grande biblioteca universal, do conhecimento sem fronteiras, da extensão infinita dos limites da memória, do processamento eletrônico de dados complexos, coisas que embora profetizadas, há anos, décadas, séculos, somente agora encontram as condições para se tornarem realidade.

As pesquisas em andamento indicam que a teoria de Ranganathan, composta basicamente de um esquema categorial universal, possa vir a ser usada de muitas formas, contribuindo para o processo de modelagem de dados a priori.

Oxalá se possa presenciar o encontro de novos sonhos com velhas práticas e teorias que se encontravam à espera de meios que as pudessem fazer frutificar. A biblioteconomia estaria pronta para esse desafio? E se não estiver, que outra disciplina melhor que ela estaria? O que não pode ser ignorado é que a mudança radical requerida parece se achar em nível da linguagem, consistindo na descoberta de uma linguagem universal de domínio público, capaz de simbolizar textos sons e imagens e que esteja à altura do novo meio digital, uma realidade já disponível com toda a sua potencialidade e poder. Achamo-nos no limiar de uma grande ruptura epistemológica?

## NOTA

[\*] Complementação da frase de Vickery, entre parênteses, pela autora do presente artigo.

## REFERÊNCIAS BIBLIOGRÁFICAS

- BAEZA-YATES, Ricardo, RIBEIRO-NETO, Berthier. Modern information retrieval. New York : ACM Press, 1999.
- BAX, Marcelo Peixoto. Introdução às linguagens de marcas. Ciência da Informação, v.30, n.1, p.32-38, jan./abr. 2001.
- DAHLBERG, Ingetraut. Knowledge organization and terminology; philosophical and linguistic bases. International classification, v.19, n.2, 1992, p.65.
- DIAS, Eduardo José Wense. Contexto digital e tratamento da informação. DataGramZero: Revista de Ciência da Informação, v.2, n.5, out. 2001. p. 1-10.
- FERREIRA, Aurélio Buarque de Holanda. Novo dicionário de língua portuguesa.2ed. Rio de janeiro : Nova Fronteira, 1986.
- JAENECKE, Peter. Elementary principles for representing knowledge. Knowledge organization, v.23, n.2, 1996, p. 88-102.
- LANCASTER, F.W. Indexação e resumos: teoria e prática. Brasília : Briquet de Lemos Livros, 1991. p.268-9.
- MILSTEAD, Jéssica. Cataloguing by any other name. [http://www.findarticles.com/cf\\_1/m1388/1\\_23/53457500/p1/article](http://www.findarticles.com/cf_1/m1388/1_23/53457500/p1/article)
- MILSTEAD, Jessica; Feldman, Susan. Metadata projects and standards. <http://www.onlineinc.com/onlinemag/OL1999/milstead1.html>
- MORENTE, Manuel Garcia. Fundamentos de filosofia; noções preliminares. São Paulo : Mestre Jou, 1964. p. 114.
- OGDEN, C.K; RICHARDS, I.A. O significado de significado; um estudo da influência da linguagem sobre o pensamento e sobre a ciência do simbolismo. Tradução de Álvaro Cabral. Rio : Zahar, 1972. 349p.
- RASMUSSEN, Edie M. Libraries and bibliographical systems. In: BAEZA-YATES, Ricardo, RIBEIRO-NETO, Berthier. Modern information retrieval. New York : ACM Press, 1999. 513p. p. 397-413.
- SATIJA, Mohinder P. Library classification: an essay in terminology. Knowledge Organization, 27, n.4, 2000.
- SHERA, Jesse H. Pattern, structure and conceptualization in classification. In: International Study Conference on Classification for Information Retrieval, Proceedings, London : ASLIB, 1957. p. 3-13
- VICKERY, D.C. Knowledge representation: a brief review. Journal of documentation, v.42, n.3, sept. 1986. p. 145-59.
- ZIMAN, John. Conhecimento público. Belo Horizonte : Itatiaia ; São Paulo : Editora da USP, 1979. 164p.

### Sobre a autora / About the Author:

Lidia Alvarenga

[lidiaalvarenga@eci.ufmg.br](mailto:lidiaalvarenga@eci.ufmg.br)

Professora da Escola de Ciência da Informação da Universidade Federal de Minas Gerais,  
Área de Tratamento da Informação e Bibliometria.