

MESA-REDONDA

A DIMENSÃO INTERDISCIPLINAR NAS PESQUISAS EM ORGANIZAÇÃO DO CONHECIMENTO E RECUPERAÇÃO

Moderadora: Cíntia de Azevedo Lourenço (ECI/UFMG)

Participantes:

Prof. Dr. Marcel Ferrante Silva (UFG/GO)

Prof. Dr. Maurício B. Almeida (ECI/UFMG)

Prof. Dr. Renato R. Souza (EMAP/FGV-RJ)

Prof. Mestre Eduardo R. Felipe (BDMG/MG)

Representação de informação, metadados, interoperabilidade e recuperação da informação na atualidade

INFORMATION REPRESENTATION METADATA, INTEROPERABILITY AND INFORMATION RETRIEVAL TODAY

Cíntia de Azevedo Lourenço (ECI/UFMG)

cal@eci.ufmg.br

Resumo: A representação da informação para a organização e recuperação da informação teve sua época áurea no final do século XIX e início do século XX. Atualmente, as pesquisas se intensificaram com o uso das novas tecnologias de informação, agregando à área além dos pesquisadores oriundos da Biblioteconomia, outros das áreas de Computação, Linguística, Engenharias e Administração. O grande desafio é garantir a interoperabilidade entre os diversos repositórios informacionais disponíveis na internet. A atuação conjunta desses pesquisadores e a integração entre todas as partes envolvidas na representação da informação garantirá o sucesso para uma organização e recuperação da informação do século XXI.

Palavras-chave: Representação da informação. Descrição bibliográfica. Indexação de assuntos. Metadados.

Abstract: The representation of information for the organization and retrieval of information had its heyday in the late 19th century and early 20th century. Today, research has intensified with the use of new information technologies, adding to the area beyond the coming of librarianship researchers other areas of Computer Science, linguistics, Engineering and Administration. The challenge is to ensure interoperability between the various informational repositories available on the internet. The joint work of these researchers and integration between all parties involved in the representation of information will ensure success for an organization and recovery of the 21st century information.

Keywords: Representation of information. Bibliographic description. Indexing issues. Metadata.

1 INTRODUÇÃO

Representar a informação foi desde o final do século XIX uma das formas mais eficientes que a humanidade pensou o controle bibliográfico e a recuperação da informação, através do registro de informações descritivas e temáticas dos documentos.

Dessa forma, a Biblioteconomia conheceu seu grande auge exatamente no final do século XIX e início do século XX, quando os profissionais da área, impulsionados pelo início da explosão bibliográfica, delinearão normas para a descrição e linguagens para a classificação e indexação dos documentos.

Grandes nomes como Dewey, Jewett, Outlet, Lafontaine, Cutter, Ranganathan, Bliss, entre outros foram grandes inovadores das pesquisas em representação e organização da informação, sempre pensando a organização da informação em seus dois aspectos centrais: a representação descritiva e a temática (BARBOSA, 1978).

É nesse contexto de desafios e inovações que surgem os catálogos de fichas, mecanismo eficiente de índice invertido manual, desenvolvido para a recuperação da informação em bibliotecas, os sistemas de classificação bibliográfica que permitem a organização dos documentos nas estantes por assuntos de uma forma sistemática, evitando que os bibliotecários precisassem re-etiquetar os documentos a cada nova remessa de livros recebida pela biblioteca, além dos primeiros vocabulários controlados para uma indexação de assuntos padronizada para uso em bibliotecas.

Apesar de alguns desses grandes pesquisadores da época terem se especializado ou em representação descritiva ou temática, alguns deles, como Cutter, se dedicaram à essas duas vertentes da representação bibliográfica.

Infelizmente, após esse período áureo da representação da informação, tivemos grandes avanços na área da representação temática, evoluindo das listas de cabeçalhos de assuntos, para os tesouros, as taxionomia e ontologias, numa busca incessante pelo aperfeiçoamento da recuperação da informação por assunto, além das atualizações constantes dos sistemas de classificação bibliográficas.

Já no tocante a representação descritiva, a área ficou por muito tempo esquecida no pragmatismo da prática profissional, sofrendo apenas algumas atualizações das normas de descrição bibliográfica, quanto às inovações dos suportes de informação, como o Código de Catalogação Anglo-Americano (AACR2) e o Padrão Internacional para Descrição Bibliográfica (ISBD).

Apenas no final da década de 1990, a área de representação descritiva retomou suas pesquisas efetivamente, originando em estudos de modelagem conceitual das normas de descrição

bibliográficas vigentes, padrões de metadados para a representação de informação em ambiente digital, culminando com a nova norma de Catalogação que substituirá as anteriores: o Descrição e Acesso de Recursos (RDA).

2 A REPRESENTAÇÃO DA INFORMAÇÃO NA ATUALIDADE

Neste novo século a sociedade mundial tem se dinamizado e o saber do homem adquiriu proporções gigantescas. Neste cenário surge a Sociedade da Informação, onde a informação e o conhecimento humano adquirem valor econômico.

Atualmente a grande discussão dentro das disciplinas fronteiriças ao campo da Ciência da Informação refere-se ao tratamento de informação no ambiente da *web*. E esta discussão extrapola os limites da Biblioteconomia, ciência que há muitos milênios vem estudando e desenvolvendo técnicas de organização, tratamento e recuperação da informação.

Assim, a Ciência da Informação tem tentado definir diversos conceitos como informação, conhecimento, dado, competência, disseminação da informação, recuperação da informação, sistemas de informação, internet, linguagens de marcação e outros tantos, necessários para sua consolidação enquanto ciência. E é neste oceano conceitual dessa nova ciência, que surge mais um conceito: o de metadados, que longe de ser um conceito novo, tem sido atualmente reconhecido como a chave para a resolução dos problemas de produção e recuperação de informações da *web*.

Contudo, na literatura sobre o tema os metadados tem sido analisado e estudado sob a luz das normas e padrões da catalogação e da indexação bibliotecária, sendo desenvolvidos e aperfeiçoados, diversos padrões de metadados tanto para descrição de objetos digitais, como para estruturação semântica das bibliotecas digitais.

No entanto, a Biblioteconomia é uma ciência antiga, que detêm técnicas de organização, recuperação e tratamento da informação milenares e bem consolidadas. E entre as técnicas biblioteconômicas pouco lembradas no estudo dos metadados, está a teoria da classificação, onde descobriremos os princípios de classificação de Ranganathan, bibliotecário indiano que estabeleceu o alicerce teórico inicial para o desenvolvimento das mais variadas de tabelas de classificação, conhecidas até nossos dias. Estes princípios, que ele chamava de cânones, estabelecem normas para a elaboração de uma classificação conceitualmente consistente e organizada.

Este novo enfoque se justifica nesta época de globalização e da sociedade da informação, que simboliza uma nova “reforma” um novo “renascimento”, no sentido de que antigos paradigmas são abandonados, novos paradigmas surgem e principalmente, velhos paradigmas

esquecidos são resgatados. Na Ciência da Informação não é diferente, pois tendo uma de suas raízes na prática biblioteconômica e documentalista, também tem resgatado antigas teorias de organização de informação na tentativa de atender a demanda e as necessidades informacionais desta nova era.

Neste contexto, a afirmação de Souza (1996), de que o novo paradigma da Biblioteconomia é o modelo Ranganathiano de organização da informação, mostra que a Ciência da Informação tem buscado modelos mais adequados entre as ideias dos antigos classificadores, que foram esquecidos e “sufocados” pelo capital, que impôs o modelo Deweyano de organização da informação.

Neste caos informacional, de uma Internet em constante crescimento, onde as informações devem ser usadas, pois informação é poder, cresce mais a necessidade de se direcionar cada informação ao seu usuário e cada usuário à sua informação, pois informações irrelevantes levam o usuário à uma perda do tempo que lhe é crucial para suas tomadas de decisão, assim, as antigas leis de Ranganathan são cada vez mais relevantes.

E é neste ambiente “vivo” da Internet que os modelos de organização da informação da Biblioteconomia têm se mostrado insuficientes, surgindo a necessidade de aperfeiçoá-los e até mesmo mudá-los. Ranganathan, parece estar mais atual do que nunca, tanto com suas leis básicas, como em seus princípios de classificação e organização do conhecimento, através de sua classificação facetada.

No contexto atual, a área de representação e organização da informação tem sido pesquisada e desenvolvida, não só pela Biblioteconomia, mas também por estudiosos oriundos das áreas de Informática e Tecnologia e Linguística Computacional.

Assim, com o uso de metadados tanto descritivos quanto estruturais e de conteúdo, a área tem buscado a interoperabilidade entre os grandes repositórios de informação especializados tanto em formato digital quanto em catálogos eletrônicos e bases de dados referenciais.

Apesar deste termo ter emergido apenas há alguns anos, ele não é um termo novo. Segundo Lange e Winkler citados por Vellucci (1998, p. 191), o termo metadados existe desde a década de sessenta, mas somente nos anos oitenta é que ele começa a aparecer na literatura de sistemas e administração de bancos de dados. Porém, Domenico (2001, p. 61) nos lembra que a importância do termo no ambiente de apoio à tomada de decisões certamente é nova. Neste contexto, metadados é a descrição das informações contidas nos bancos de dados, ou simplesmente, metadados é o dado sobre o dado (VELLUCCI, 1998; TAKAHASHI, 2000; DOMENICO, 2001).

O termo metadado vem sendo utilizado atualmente em larga escala principalmente pelas áreas da Ciência da Informação e da Informática, para o desenvolvimento de Bibliotecas Digitais no ambiente *Web*, além de taxionomias e ontologias especializadas. Sua definição na maioria dos trabalhos geralmente se limita a defini-lo como dado do dado. Foram identificadas outras definições mais amplas, as quais citamos neste trabalho para um melhor embasamento de nossa discussão posterior.

Vários autores entendem metadados como:

Uma melhor definição seria a de que metadado é o significado dos dados, ou seja, de que é uma abstração dos dados, ou ainda, dados de mais alto nível que descrevem dados em um nível inferior (DOMENICO, 2001, p. 62).

Metadados são [...] a definição dos dados, feita numa linguagem comum que permita que todos os intervenientes se comuniquem e se entendam, permitindo a partilha de informação (MURRAY, 2000, p. 1).

Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics (DEMPSEY; HEERY, 1997 citado por VELLUCCI, 1998, p. 194).

Metadata is defined here as data that describe attributes of a resource, characterize its relationships, support its discovery and effective use, and exist in an electronic environment. Metadata usually consist of a set of data elements where each element describes an attribute of the resource, its management, or use (VELLUCCI, 1998, p. 192).

El término metadatos describe varios atributos de los objetos de información y les otorga significado, contexto y organización” (KENNEY et al., 2001, p. 5).

O dado estruturado sobre o dado que é interpretado de forma automática (W3C citado por BARRETO, 1999, p. 41).

Assim, os modelos de metadados descritivos, têm se constituído nos novos padrões de “catalogação” da Biblioteconomia para documentos eletrônicos, podendo-se entender padrões de metadados como:

Um padrão de metadados provê um conjunto de elementos descritores organizados segundo um modelo de dados que é empregado para descrever objetos do domínio de informação do usuário. Um modelo de dados, por sua vez, pode ser definido como um conjunto de conceitos e regras de composição associadas que são usadas para descrever tipos e relacionamentos entre as instâncias desses tipos. No contexto relacional, o modelo de dados básico consiste de tabelas, registros e atributos. No contexto da orientação a objetos, o modelo de dados básico consiste de classes e objetos. Em ambos os modelos, regras de composição distintas se aplicam. No modelo orientado a objetos, essas regras se referem à herança entre classes, polimorfismo, domínio de atributos e agregação. No modelo relacional, essas regras especificam, por exemplo, que um registro deve conter um número qualquer de atributos e que cada atributo deve ser definido em termos de um tipo primitivo de dado (BARRETO, 1999, p. 52).

Os trabalhos apresentados na mesa redonda “A dimensão interdisciplinar nas pesquisas em organização do conhecimento e recuperação da informação”, no II Seminário do Grupo MHTX demonstra os temas desenvolvidos pelos pesquisadores em organização e tratamento da informação acerca da questão da representação da informação para recuperação da informação: metadados e ontologias.

Estes tem sido os temas mais presentes nas pesquisas em Ciência da Informação, por pesquisadores tanto da Biblioteconomia quanto da Ciência da Computação, como se pode perceber pela própria composição da mesa, que traz pesquisadores de renome de ambas as áreas.

3 CONCLUSÕES

Sem pretender esgotar o assunto, este trabalho foi uma tentativa teórica de clarear o conceito de metadados, principalmente para utilização na Ciência da Informação. Deste modo, verificamos que este termo não é atual, ele existe deste a década de sessenta, mas estava como que “guardado no fundo do baú”, pois principalmente a Biblioteconomia sempre lançou mão de metadados para a definição de seus padrões de descrição bibliográfica, mas não sabia que o estava fazendo.

Mesmo assim, notou-se aqui que já existem muitas iniciativas em adequar os formatos existentes para sua utilização em ambiente *Web* e também, que existem várias outras iniciativas para um formato menos técnico para utilização dos autores de documentos eletrônicos. Portanto, é de suma importância que os estudos sobre metadados sejam intensificados, principalmente pela Biblioteconomia e pela Ciência da Informação como um todo, pois historicamente estas áreas já o faziam por muitas décadas.

Apesar dos cânones de Ranganathan terem sido idealizados para a elaboração de esquemas de classificação do conhecimento, alguns deles são pertinentes e observáveis às normas de descrição de recursos informacionais da Internet: os metadados descritivos.

Estes cânones, como princípios básicos de organização da informação, têm sua aplicação estendida a todo ambiente *Web*, podendo ser considerados uma contribuição teórica significativa para o desenvolvimento:

- a) das Bibliotecas Digitais;
- b) dos mecanismos de busca, coleta e indexação automática da *Web*;
- c) dos estudos semânticos e ontológicos das arquiteturas de metadados como o RDF e o OIL.

Portanto, o resgate desses princípios é necessário e obrigatório não só aos classificadores e indexadores modernos, como também aos estudiosos da catalogação e das normas para

descrição e identificação dos recursos informacionais tanto em meio físico quanto em meio eletrônico, para que os metadados descritivos e as arquiteturas semânticas de *Web* sejam aperfeiçoadas de maneira a aumentar a eficiência e a eficácia da recuperação dos recursos informacionais da *Web*.

Além desse resgate de princípios, será também imprescindível para o desenvolvimento qualitativo e equilibrado da área da representação de informações, um resgate de pesquisas conjuntas entre a representação descritiva e a representação temática, pois a recuperação de informações não pode se limitar apenas aos elementos extrínsecos dos documentos, os assuntos ou sua temática, é importante que os elementos intrínsecos ou descritivos também sejam bem trabalhados para uma perfeita representação de informações que muito contribuirá para os avanços da organização e tratamento da informação.

REFERÊNCIAS

BARBOSA, A. P. **Novos rumos da catalogação**. Rio de Janeiro: BNG: Brasilart, 1978.

BARRETO, C. M. **Modelo de metadados para a descrição de documentos eletrônicos na web**. 190f. Dissertação (Mestrado em Ciências em Sistemas de Computação) – Instituto Militar de Engenharia, Rio de Janeiro, 1999.

DOMENICO, J. A. D. **Definição de um ambiente Data Warehouse em uma instituição de ensino superior**. 2001. 137f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2001.

MURRAY, D. G. O que é... um portal de informação empresarial?. **InfoIMAGEM: Jornal da Gestão Eletrônica de Imagens, Documentos e Processos**, n. 29, set./out. 2000. Disponível em: <<http://www.dotecome.com/infoimagem/infoimagem/infoimagem.htm>>. Acesso em: 31 maio 2001.

SOUZA, F. C. Os paradigmas da Biblioteconomia e suas implicações no ensino desta ciência. **Encontros Bibli: Revista de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 1, n. 2, set. 1996. Disponível em: <<http://www.ced.ufsc.br/bibliote/encontro/bibli2.html>>. Acesso em: 12 jun. 2001.

TAKAHASHI, T. (Org.). **Sociedade da informação no Brasil**: livro verde. Brasília, DF: Ministério da Ciência e Tecnologia, 2000. 195p.

VELLUCI, S. L. Metadata. **Annual Review of Information Science and Technology (ARIST)**, v. 33, p. 189-222, 1998.