

# II Seminário do Grupo de Pesquisa

## <MHTX>



**Pesquisa Integrada em  
Organização do Conhecimento  
e Recuperação da Informação**



**02 e 03 de junho de 2016  
Escola de Ciência da Informação/UFMG**

Gercina A. B. O. Lima, Ivo Pierozzi Junior,  
Benildes C. M. S. Maculan, Célia C. Dias,  
Cíntia A. Lourenço  
(Org.)

**ANAIS DO II SEMIÁRIO  
DO GRUPO DE PESQUISA <MHTX>:**

Pesquisa Integrada em Organização do Conhecimento e  
Recuperação da Informação

ISBN 978-85-65609-03-6

BELO HORIZONTE

ECI/UFMG

2016

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Reitor: Jaime Arturo Ramírez  
Vice-Reitora: Sandra Regina Goulart Almeida

ESCOLA DE CIÊNCIA INFORMAÇÃO - ECI/UFMG  
Diretor: Carlos Alberto Ávila Araújo  
Vice-Diretor: Adalson de Oliveira Nascimento

GRUPO DE PESQUISA MHTX  
Coordenadora: Gercina Ângela Borém de Oliveira Lima

EMBRAPA INFORMÁTICA AGROPECUÁRIA, CAMPINAS – SP  
Representante: Ivo Pierozzi Junior

PROJETO GRÁFICO  
Décio Wey Berti Junior  
Eduardo R. Felipe  
Graciane Bruzinga Borges  
Lucinéia Maia  
Webert J. Araújo

A532a 2016 Anais do II Seminário do Grupo de Pesquisa MHTX : pesquisa integrada em organização do conhecimento e recuperação da informação (2. : 2016 : Belo Horizonte, MG, ECI/UFMG, EMBRAPA).

Anais [recurso eletrônico] / II Seminário do Grupo de Pesquisa MHTX : pesquisa integrada em organização do conhecimento e recuperação da informação, 2-3 de junho em Belo Horizonte, MG. / Organizadores: Gercina A. B. O. Lima, Ivo Pierozzi Junior, Benildes C. M. S. Maculan, Célia C. Dias, Cíntia A. Lourenço. – Belo Horizonte, ECI/UFMG, 2016.

ISBN 978-85-65609

1. Evento – Ciência da Informação. 2. Evento – Pesquisa em Ciência da Informação. I. Título. II. Embrapa Informática Agropecuária, Campinas – SP.

CDU: 02(063)(81)

#### DIREITO AUTORAL E DE REPRODUÇÃO

Direitos de autor ©2016 para os artigos individuais dos autores. São permitidas cópias para fins privados e acadêmicos, desde que citada a fonte e autoria. A republicação desse material requer permissão dos detentores dos direitos autorais. Os editores deste volume são responsáveis pela publicação e detentores dos direitos autorais.

## SUMÁRIO

<b>APRESENTAÇÃO.....</b>	<b>2</b>
<i>Gercina A. B. O. Lima</i>	
<b>PALESTRA DE ABERTURA.....</b>	<b>4</b>
KNOWLEDGE ORGANIZATION AND INFORMATION RETRIEVAL: A RESEARCH AGENDA .....	4
<i>Dagobert Soergel (University at Buffalo, USA)</i>	
<b>MESA-REDONDA.....</b>	<b>25</b>
DEBATES SOBRE AS PRÁTICAS DE ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO NO MERCADO DE TRABALHO SOB A ÓTICA DA BIBLIOTECONOMIA	
<i>Moderadora: Célia da Consolação Dias (ECI/UFMG)</i>	
A organização e a representação da informação no mercado de trabalho sob a ótica da Biblioteconomia .....	25
<i>Célia da Consolação Dias (ECI/UFMG)</i>	
Consultoria em gestão de informação: desafios e oportunidades.....	32
<i>Elisângela Cristina Aganette (ECI/UFMG)</i>	
Bases de dados de informações científicas e tecnológicas e a organização e representação da informação: relato de experiência .....	39
<i>Alessandra Rodrigues (Embrapa Informação Tecnológica, EMBRAPA)</i>	
Biblioteconomia: relato de experiência e perspectivas para o profissional da informação .....	45
<i>Fernanda Pereira (IFMG-Sudeste / Campus Barbacena)</i>	
Consultoria técnica em gestão de documentos e serviços de informação no contexto corporativo .....	51
<i>Graciane B. Borges (ECI/UFMG)</i>	
Organização e representação em bibliotecas escolares .....	57
<i>Kelly Cristiane Santos Moraes (Colégio Franciscano Sagrada Família)</i>	
<b>MESA-REDONDA.....</b>	<b>61</b>
AS TEORIAS DA PESQUISA INTEGRADA EM ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO	
<i>Moderadora: Benildes Coura M. S. Maculan (ECI/UFMG)</i>	
A noção de integração nas pesquisas em organização e recuperação da informação	61
<i>Benildes C. M. S. Maculan (ECI/UFMG)</i>	
Da organização do conhecimento à recuperação da informação: teorias e técnicas .	67
<i>Gercina A. B. O. Lima (ECI/UFMG)</i>	

Emprego da abordagem onomasiológica na elaboração de definições em ontologias de domínio: estudo preliminar .....	78
<i>Maria Luiza de Almeida Campos (Universidade Federal Fluminense)</i>	
Organização do conhecimento na perspectiva da abordagem cognitiva e sociocognitiva .....	83
<i>Mariângela Spotti Lopes Fujita (UNESP)</i>	
<i>Paula Regina Dal ´Evedove (UFSCar)</i>	
<b>PALESTRA.....</b>	<b>90</b>
A PESQUISA NA EMPRAPA: O CONVÊNIO COM O GRUPO DE PESQUISA MHTX	
<i>Ivo Pierozzi Júnior (Embrapa Informática Agropecuária, Campinas – SP)</i>	
<b>MESA-REDONDA.....</b>	<b>95</b>
A DIMENSÃO INTERDISCIPLINAR NAS PESQUISAS EM ORGANIZAÇÃO DO CONHECIMENTO E RECUPERAÇÃO	
<i>Moderadora: Cíntia de Azevedo Lourenço (ECI/UFMG)</i>	
Representação de informação, metadados, interoperabilidade e recuperação da informação na atualidade .....	95
<i>Cíntia de Azevedo Lourenço (ECI/UFMG)</i>	
<i>Tainacan</i> : sistema de gestão de biblioteca digital social .....	102
<i>Marcel Ferrante Silva (UFG/GO)</i>	
Provocações acadêmicas: ontologias, tesouros, documentos, conteúdo de documentos, e... Unicórnios.....	117
<i>Maurício B. Almeida (ECI/UFMG)</i>	
A dimensão interdisciplinar nas pesquisas em organização do conhecimento e recuperação da informação.....	124
<i>Renato R. Souza (EMAP/FGV-RJ)</i>	
<i>Flávio Codeço Coelho (EMAP, FGV)</i>	
<i>Matthew Connelly (Department of History, Columbia University)</i>	
Metadados: princípios e práticas na organização informacional .....	134
<i>Eduardo R. Felipe (BDMG/MG)</i>	

## APRESENTAÇÃO

Gercina A. B. O. Lima

As tecnologias de informação e comunicação são, hoje, constitutivas dos espaços e práticas interdisciplinares de geração de conhecimento e de inscrição de memórias, condicionando todos os contextos e pontos de partida da inovação, da deliberação e da tomada de decisão. O ser humano adquire, processa e armazena o conhecimento, por meio de um processo que vai concatenando, buscando aportes teóricos em outras disciplinas, permitindo a elaboração de um conhecimento novo, completando, assim, o círculo informacional quando a informação é convertida em conhecimento, capaz de promover desenvolvimento e de ser comunicada, gerando, assim, novos estoques de informação e retroalimentando o ciclo informacional. Diante desse panorama, quanto maior for a integração entre os profissionais da área de Organização do Conhecimento (OC) com outros profissionais das áreas interdisciplinares, maior será o avanço nas pesquisas.

O Grupo de Pesquisa Modelo Hipertextual de Organização de Documentos – MHTX, se insere neste contexto, como um espaço de ensino, pesquisa e extensão, que prepara alunos do curso de Biblioteconomia, e orienta alunos oriundos de diversas áreas do conhecimento, contribuindo para suas inserções na pesquisa acadêmica. Assim, inicia-se a formação desde a graduação, por meio de bolsas dos programas de iniciação científica, monitoria e do programa Pro-Noturno, visando ao ingresso desses alunos no Programa de Pós-Graduação em Ciência da Informação. Portanto, as pesquisas e produções intelectuais do Grupo MHTX têm um papel significativo na constituição ou composição de um conhecimento coletivo, que dissemina a informação e agrega estudantes, iniciantes na pesquisa acadêmica.

Desde 2004, as atividades do Grupo MHTX vêm contribuindo para intermediação, acesso e distribuição da informação, ou, ainda, nas interações comunicativas em meio digital, por meio dos resultados das pesquisas realizadas pelos membros do Grupo. Além disso, o Grupo MHTX tem sido um ambiente interativo, no qual as ações cruzam fronteiras disciplinares e culturais, construindo uma rede de divulgação do conhecimento.

O I SEMINÁRIO DO GRUPO DE PESQUISA MHTX ocorreu em maio de 2013. Nesse Seminário, pesquisadores e orientandos do PPGCI, participantes do Grupo MHTX, tiveram a oportunidade de apresentar e discutir os resultados de suas pesquisas junto à comunidade científica da UFMG, o que contribuiu para que os alunos da graduação se sentissem incentivados em participar e contribuir para essas novas interseções entre as formas sociais de informar e ser informado e as utilizações das novas tecnologias da informação.

O II SEMINÁRIO DO GRUPO DE PESQUISA MHTX tem como objetivo dar continuidade aos trabalhos iniciados no I Seminário, realizado na Escola de Ciência da Informação da UFMG, no ano de 2013. O presente Seminário conta desta forma, com três mesas-redondas e três palestras. O II Seminário teve como temática principal “A pesquisa integrada sobre organização do

conhecimento e recuperação da informação”. O objetivo foi de propiciar uma reflexão acerca dos fundamentos teóricos metodológicos das pesquisas realizadas pelos membros do Grupo de Pesquisa MHTX, discutindo o diálogo interdisciplinar que complementam esses fundamentos, bem como as aplicações tecnológicas e considerações sobre o perfil do profissional bibliotecário que o mercado de trabalho têm exigido atualmente.

As expectativas desses resultados, em conjunto, caracterizadas pela integração das pesquisas, permitem que os estudos possam avançar, para além dos anos que a academia permite aos alunos de Pós-Graduação. Para a realização dessas pesquisas, também é necessário buscar aportes teóricos e técnicos em outras áreas do conhecimento, tais como a Ciência da Computação, a Linguística e a Psicologia Cognitiva.

Espero que este trabalho seja útil para os alunos de Graduação e de Pós-Graduação, bem como para professores e pesquisadores da área. Espera-se que a presente obra possa contribuir para a interlocução científica entre a comunidade de OIC e que ela seja um incentivo para a integração entre pesquisadores/professores e alunos dos cursos de Pós-graduação e Graduação em Biblioteconomia e Ciência da Informação (BCI).

Gercina Ângela de Lima  
Universidade Federal de Minas Gerais  
*glima@eci.ufmg.br*

## PALESTRA DE ABERTURA

### KNOWLEDGE ORGANIZATION AND INFORMATION RETRIEVAL: A RESEARCH AGENDA

#### *ORGANIZAÇÃO DO CONHECIMENTO E RECUPERAÇÃO DA INFORMAÇÃO: UMA AGENDA DE PESQUISA*

Dagobert Soergel (University at Buffalo, USA)

With the assistance of Catherine Dillon

**Abstract:** Research in information retrieval is design research; it produces designs and implementation of solutions that are intended to improve information problems and then studies the effectiveness of these solutions. This paper shows many ways in which Knowledge Organization Systems could be used in solutions to improve information seeking, retrieval, and sensemaking and learning, some well-known and others suggested here, and it discusses ways to test such solutions and implications for KOS suggesting many interesting research problems. The paper considers the following component processes of information seeking, retrieval, and sensemaking and learning: 1) Assistance to users in clarifying the information need and formulating a good query for the system being searched; 2) Search and post-search processing. Sensemaking. Includes support for reading. See also #3; 3) Creating structure in whole collection or in search results (also part of sensemaking); 4) Interface design. This is a wide-ranging concept paper focusing on ideas. It is not a paper with empirical results.

**Resumo:** A pesquisa em recuperação da informação é *design* de pesquisa; ela produz concepções e implantações de soluções que pretendem aprimorar os problemas de informação e a partir daí, estuda a eficácia dessas soluções. Este artigo mostra várias formas nas quais os sistemas de organização do conhecimento (SOCs) poderiam ser usados em soluções para melhorar a busca, recuperação, *sensemaking* (geração de sentido) e aprendizagem da informação. Alguns são muito conhecidos e outros sugeridos aqui e este trabalho discute formas de testar essas soluções e implicações de SOCs, sugerindo muitos problemas de pesquisa interessantes. Este artigo considera os seguintes componentes dos processos de busca, recuperação, *sensemaking* e aprendizagem: 1) assistência aos usuários em esclarecer a necessidade da informação e de formular uma boa *query* para os sistemas que estão sendo buscados; 2) processo de busca e pós-busca. *Sensemaking*. Inclui suporte para leitura (ver #3); 3) criar estrutura na coleção toda ou nos resultados da busca por (isso também é parte do *sensemaking*); 4) *design* de Interface. Este é um artigo conceitualmente amplo, focado em ideias. Este não é um artigo com resultados empíricos.

## 1 INTRODUCTION

### 1.1 Information seeking, retrieval, and use processes

Information retrieval is a component of a broader process of information seeking, retrieval, and use (including sensemaking and learning). Sensemaking is a process of finding data and arranging them into structures to better support problem solving and decision making; it involves fitting new data into knowledge structures. The sensemaker may already have a knowledge



structure, or she may have to search for one or build one bottom-up from data. Information retrieval tools will be more useful when they assist users in building knowledge structures and if they show how isolated pieces of data fit into knowledge structures. This is where knowledge organization comes in. As this paper will show, knowledge organization and Knowledge Organization Systems have a pervasive and useful role to play in all phases of the information seeking process.

Knowledge structures – in the mind, in text, in collections – can be viewed as

- (1) neural networks and exploited through machine learning or
- (2) as explicit, understandable, meaningful structures and exploited through knowledge-based methods.

Both approaches have been used successfully; the statistical / machine learning approach has become prevalent. This paper discusses the knowledge-based approach.

Creating and using information systems has many component processes. This paper discusses only some of these as shown in Table 1, starting on the user side with information

Table 1. Overview: Information seeking, retrieval, and use processes

<b>1</b>	<b>Assistance to users in clarifying the information need and formulating a good query for the system being searched automatic query enhancement</b>
	1.1 Assistance to users with clarifying their problem and information need
	1.2 Query formulation and query enhancement
	1.2.1 Assistance with manual query formulation. Manual query term expansion
	1.2.2 Automatic query enhancement. Query term expansion
<b>2</b>	<b>Search and post-search processing. Sensemaking</b>
	2.1 Executing a query
	2.2 User-system interaction in search, exploration, processing search results, and sense making
	2.3 Supporting reading and quick assimilation of information
<b>3</b>	<b>Detecting underlying meaning and structure and creating new structure in whole collection or in search results</b>
	3.1 From words and images to concepts: phrase recognition, word sense disambiguation, term to concept mapping, coreference resolution
	3.2 Indexing and metadata creation.
	3.2.1 Human indexing or tagging, including social tagging
	3.2.2 Computer-assisted and automatic indexing. Semantic enrichment, entity recognition
	3.3 Document segmentation. Detecting meaningful segments

- 3.4 Creating hypertext. Multi-document hypertext
- 3.5 Abstracting and summarization, human or automatic. Multi-document summarization
- 3.6 Classification, grouping, clustering
- 3.7 Information extraction

#### **4 Interface design. Visualization**

needs clarification, query formulation, search, and information use moving on to processes of detecting and creating structure that can be applied to whole collections to enable search or to the processing of search results, and closing with brief observations on interface design. These processes are not executed in a fixed order or in a linear sequence; rather the order of execution is dynamic, ever changing, and iterative.

These processes can be executed by people, by computer programs, or by a collaboration of both. Human-computer collaboration may happen in several ways, for example:

- (1) The program simply selects knowledge from a KOS that is helpful for the current task and presents it to the user;
- (2) the user enters an item and the program extends it or just repeats #1;
- (3) the program executes a task (such as synonym expansion) and presents the results to the user for review.

For each process the paper discusses the application of KOS across this whole spectrum.

All these processes can be improved by application of Knowledge Organization Systems that give information on concepts and terms with concept-concept and concept-term relationships and possibly much other information.

### **1.2 Many types of KOS and many sources for accessing KOS**

This paper takes an expansive view of KOS to include, among others, scheme known under any of the labels listed below. For example, there is a wide range of schemes, even just simple lists of terms, called "ontology" by their creators because ontology has become a fashionable term

- (1) systems that specify entity types and relationship types – ontologies in the proper meaning, data models, metadata schemes (called "vocabularies" in the Semantic Web context), metadata inventories (such as schema.org), data dictionaries.
- (2) systems that give entity values and relationship instances – traditional classification schemes, taxonomies, typologies, subject heading lists, thesauri (many available in multiple languages), concept maps / causal maps / other node-link diagrams and collections thereof, dictionaries (mono-, bi-, or multi-lingual), glossaries, linguistic lexicons, gazetteers, term lists giving statistical associations between pairs of terms in

some collection of documents; for example, a list that give co-occurrence information for tags assigned in social tagging, folksonomies enriched with hierarchical links and mapping to other KOS, integrated systems relating multiple KOS, such as UMLS and CISMef, other mapping tables (cross-concordances) (not a complete list).

In the context of business information, the kind of data given by KOS, broadened to other entity types (such as persons, organizations, currencies, products, which pose exactly the same problems) are referred to as *reference data*.

KOS provide much information:

- concepts with hierarchical and other relationships and indication of their entity type or broad category (such as sentient being), links to concept maps / causal maps / other relationship diagrams in which the concept occurs
- terms with relationships to concepts and relationships between terms, senses of a term and their definitions, images, usage data for each sense, audience level (register), etymology, linguistic information such as morphology and part of speech or how a word is properly used (for example, *laugh* needs as a subject a sentient being, or in *this software runs on X*, *X* must be some kind of computing device), pronunciation (useful for automatic speech recognition),
- co-occurrence information, and much more.

KOS information can be produced by any combination of automatic methods and intellectual work.

KOs vary widely in quality, and that must be considered in their use and the evaluation of the effects of such use.

This paper suggests that applying knowledge from such KOS can make information seeking, retrieval, and sensemaking and learning more successful. But to apply such knowledge from KOS, the user or a computer program must first find it. A user or a computer program may obtain information

- only from the KOS used in a particular system (if the system has a KOS that is accessible; many retrieval systems, or search engines, including Google, do not) or
- from any KOS that can be accessed by a user in print or through or by a computer program through an API or Web service or as linked data.

### **1.3 The complexity of research on the use of KOS in information seeking, retrieval, and use**

In spite of its potential benefits the application of KOS is often neglected. We need more evidence to demonstrate the effectiveness of KOS in these information processes. This paper sets out a research agenda for this purpose, presenting many design ideas that can and should be studied for their

- feasibility: how well can different systems implement the idea and
- effectiveness: how much does a design idea, if implemented, help users and/or improve retrieval.

In assessing the results of such research, one needs to be aware that any such results are influenced by many factors and are valid only for unique constellation of factors. Beware of generalizations that are not supported by the data. For example, a researcher may study the effects of synonym expansion using WordNet, find that retrieval performance actually decreases, and pronounce to the world that synonym expansion lowers retrieval performance, while the actual result is that synonym expansion using WordNet when searching for questions in question set X in a collection in subject field Y lowers retrieval performance. The researcher clearly is unaware that WordNet, an excellent and very rich online dictionary, is a poor choice for synonym expansion because it gives for each word all its senses (both frequent and rare), and for each sense all words that express this sense (the synonyms). In the query, a word is used in a particular sense, but if not used carefully WordNet delivers synonyms for all senses. Leading to the retrieval of many irrelevant documents. Furthermore, WordNet does not cover much technical vocabulary, and it focuses on single words, not phrases. Using a more appropriate source for synonyms may well improve retrieval results compared with no synonym expansion.

## **2 ASSISTANCE TO USERS IN CLARIFYING THE INFORMATION NEED AND FORMULATING A GOOD QUERY FOR THE SYSTEM BEING SEARCHED. QUERY ENHANCEMENT**

The discussion in this section makes a conceptual distinction between 1.1 *Assistance with information needs clarification* and 1.2 *Assistance with query formulation*. However, these two are closely linked; in practice they often occur together in one and the same process.

Clearly, the usefulness of the processes described in this section, particularly as they involve work by the user, depends on the complexity of the problem to be solved and the information need and the desired quality of retrieval results. A quick question, such as finding the definition of a simple word or the symptoms of a common disease do not require all this elaboration, exploring a doctoral dissertation topic or finding the most efficient algorithm for a complex computation do.

### **2.1 Assistance to users with clarifying their problem and information need**

Perhaps the most important factor in the success of a search is a good understanding of the information need. For problems of some complexity, KOS can assist the user with thinking

through the problem to be solved and the information needed to solve the problem. A KOS can map out a concept space, provide classification/typology, put concepts and terms into context, relate concepts to terms, and give definitions, thus provide orientation and serve as a reference tool; all of this helps the user clarify her information need, be it information for solving a problem or information for learning about a topic or writing a term paper.

Showing facet frames and concept hierarchies will be useful to the user. One can hypothesize that even more useful are concept maps that show connections between different areas of a user's problem or variables under study, such a causal maps that show how the factors or variables in a research problem, such as obesity, influence each other. A causal map is highly structured, other concept maps are less formal just showing general associations between concepts. This kind of assistance requires a very rich KOS with links from concepts to concept maps in which they appear. Such a system should support highly flexible viewing and editing of the concept map. For example, a user might expand a concept to show the hierarchy around it, change the links between concepts, add concepts, and highlight concepts she deems important to her search. The user should be able to view multiple concept maps in separate windows, for example, the user may select concept A in one concept map and ask the system to show other concept maps that include concept A. Concept maps may be especially useful when several users work jointly (collaborative information retrieval) or in a reference interview, where the user and an information intermediary (reference librarian) explore the search topic together.

Showing a causal map might remind the user of additional factors to consider in solving a problem. Another way to expand a user's thinking is through a KOS that is organized around problems, such as *buying a house, deciding on a mortgage application, planning a new town, planning a campaign to reduce obesity, decide on a university to attend*. Such a KOS would give for each problem the steps to perform to arrive at a solution, sub-problems, and the factors to consider in addressing the problem. In the absence of such a KOS, the system might suggest to the user to conduct a subsidiary search to explore the problem space, finding, for example, documents that discuss how to solve the problem at hand or documents that discuss how other people went about solving the problem.

The system and the user can collaborate to come up with an initial concept map in several ways:

- The user types an initial query (possibly just one word). This query signals one or more concepts. The system then
  - shows one or more ready-made concept maps around the user concept(s)
  - places the user concept(s) in a concept map drawing window and encourages the user to think of other concepts and relates these concepts to each other, expressing the

relationships through links. The system can provide assistance in this process by showing other concepts and relationships drawn from a traditional KOS, such as a thesaurus, or from a richer KOS that has access to many concept maps.

- The user enters a paragraph-long statement of the information need. The system identifies concepts and relationships by analyzing this paragraph and displays them in a concept map; this approach was first described in the pioneering and much cited paper Belkin et al. 1982. (The "system" could be a human information intermediary or computer program capable of sophisticated natural language understanding, see Section 3.6 Information extraction). The user can then view and edit this concept map as described above.

There are many research issues here, one group revolving around the structure of the KOS required and the capabilities of the software for interacting with the KOS and another around cognitive processes of users and how effective different kinds of displays and interactions are for users, depending on characteristics of users and the problems they try to solve. User characteristics of interest include how much the user knows about the subject and its vocabulary, how skilled the user is at interpreting concept maps and at creating concept maps (which also depends on the software used).

## 2.2 Query formulation and query enhancement

From a good statement of the information need or *query statement*, a good query formulation can be constructed through the following steps. This is an ideal process; some or all of these steps may be omitted, or the steps may be executed in a different sequence. If the query statement is clearly written out, a sophisticated computer program can execute all steps, if the query statement is only in the user's head, the user must execute Steps 1 and 2 or otherwise enter some concepts or terms for a computer program to do Steps 3 and/or 4. The first three deal with conceptual issues the user needs to address for searching any type of retrieval system, the fourth is necessary for free-text and folksonomy search, which in many areas is prevalent today. As the following discussion shows, *query concept expansion* would be a better term.

The steps are:

- 1 Isolating the concepts that need to be combined using Boolean AND. Ideally this is accomplished through facet analysis of the query.
- 2 Within each facet, find the key concept at the appropriate level of specificity.
- 3 For each of these concepts find all the narrower concepts for hierarchic query expansion.
- 4 For all concepts, find synonyms for query term expansion for free-text and folksonomy search.

Query term expansion results in a query that ANDs one or more conceptual components, where each conceptual component is represented by an OR-combination of terms. This structure is

very useful in systems that compute a relevance score, such as Google: The relevance score is based on the number of conceptual components occurring in a document; a conceptual component occurs in a document if the document contains any of the ORed terms. In this context, such a query has been called a *structured query* as opposed to a query that is just an unstructured bag of words.

### **2.2.1 Assistance with manual query formulation. Manual query term expansion**

For *Step 1*, the system can assist through guided facet analysis (provided a faceted KOS is available), either suggesting facets based on the analysis of the users natural language query statement or displaying the entire facet frame (all facets in the KOS) so the user can think about whether or not a facet would provide a useful criterion in specifying the search and check the facets found promising. Faceted search has become very popular, especially on ecommerce sites. On such a site, specifying values in several facets may be the final query formulation; the query-based search is then often followed by browsing the results. Faceted search is predicated on the items to be searched being indexed explicitly with facet values. However, facet analysis of the query is useful as a step towards a good query formulation for any systems, even a free-text search as on Google.

For *Step 2*, the system can assist with clear and meaningful displays of the concept hierarchy within each facet and of other concept relationships.

*Step 3* is best executed by the search program (see next section) but . But can be executed manually. For example, shelf arrangement by a classification such as the Dewey Decimal Classification supports hierarchical expanded search: When the user goes to the shelves looking for a given class number (which users often do, having obtained a class number from a bibliographic record found in the library catalog), he finds books not only in this class but also in narrower classes following it in the arrangement. In an online search, the user can manually enter an OR combination of the key concept in a facet and all the narrower concepts under it; these narrower concepts should be visible in the concept hierarchy of a KOS; if not, the user must think of the narrower concepts herself.

For *Step 4* the system should suggest synonyms and term variants (spelling variants, morphological variants) found in a KOS, possibly also terms in different languages; otherwise the user needs to think of synonyms herself. In manual query term expansion, the user is in complete control. The system makes suggestions, the user selects. This brings to the task much intelligence that is often missing in automatic query term expansion, to be discussed next.

Throughout this process the system can make different kinds of concept/ term suggestions. For example, Google suggests phrases which begin with or otherwise contain the word the user

typed in the search box. These suggestions come from a very large KOS of such phrases that Google assembled for vast quantities of text and user queries. Google clearly has an algorithm to select from all possible such phrases, considering, I suspect, the individual user's search history and the frequency of a phrase in the collection.

Research issues here include

- What kind of displays are most helpful to users, for example, a simple outline form hierarchy as used in Explorer or a graphic display of the hierarchy as a tree or radiating out from a center?
- How successful are users in accomplishing these tasks?
- How does the quality of the underlying KOS influence the results?
- What are the effects on retrieval performance?

Keep in mind that the answers depend on user characteristics, problem and subject field characteristics, and characteristics of the database searched.

### **2.2.2 Automatic query enhancement. automatic query term expansion**

This requires accomplishing the same four tasks automatically. This involves using the methods described in Section 3, in particular phrase recognition and word sense disambiguation.

Step 3 *Hierarchic query expansion* (hierarchically expanded search, inclusive search) is used in many systems where search is based on a controlled vocabulary. The user specifies one or more concepts to be searched, using the appropriate term(s) or code(s). If the KOS used has a hierarchy, it is then a simple process to search for all narrower concepts as well. MEDLARS and its online incarnation MEDLINE may have been the first computer system to support hierarchically expanded search using the hierarchy of the Medical Subject Headings (MeSH). MeSH is now used by many systems for hierarchic expansion.

Much has been written under the label *automatic query expansion*. Most of this literature reports on methods that do just Step 4, often just based on single words.

Research issues include:

- To what extent can Steps 1 – 2 be automated?
- What are the effects of automatic query expansion on retrieval performance?

There are many factors that can influence the results, so any general statements are misleading at best. The factors to be considered include

- the subject field, in particular its conceptual structure and how tight or diffuse its terminology is;
- the nature of the query;
- the source(s) used for hierarchic and synonym expansion;



- the nature and quality of the program performing the expansion, for example, how good is word sense disambiguation.

### 2.2.3 Concluding remarks on query processing

In all steps of query processing, the techniques discussed in Section 3.1 From words to concepts are important. In particular, misspellings should be detected and corrected (informing the user of such action). This is particularly important for young children and users who search in a language with which they are not intimately familiar.

## 3 SEARCH AND POST-SEARCH PROCESSING. SENSEMAKING

### 3.1 Executing the search

In straight Boolean retrieval this is simple. Query expansion, especially hierarchic expansion, is sometimes thought of a part of search execution; some authors refer to this as using semantic networks (or Bayesian networks if the relationships between terms are statistical) to enhance search. In search based on computing for each document a score indicating expected relevance, this becomes more complicated. Many systems of this nature use the entire text of a document, comparing it to the query, in the simplest case considering the document just as a bag of words. With more sophistication there are many more refinements, for example, considering phrases.

Instead of computing relevance scores based on the raw text a system can use text pre-processed by the word-to-concept mapping and other methods discussed in Section 3.

Relevance scores are generally computed by giving a document "credit" for each search criterion it fulfils. This credit is adjusted by weighting, where the weight of a criterion (in the simplest case a word) is often determined as product or other function of two factors:

- (1) The strength with which the criterion is represented in the document, often computed as a weighted frequency of occurrence where, for example, an occurrence in the title is weighted more heavily than an occurrence somewhere in the text. Known in its simplest form as term frequency (TF).
- (2) The importance of the criterion, of the measured by its rarity in the collection or inverse document frequency (IDF).

These weighting factors can be adjusted in several ways using KOS information. First, when a criterion is expressed as an OR-combination of terms (query term expansion), the frequency of the criterion is the sum of the frequencies of the ORed terms, so it should be called *concept frequency*. Of course, document frequency should likewise be computed for the OR-

combination of terms. Next, one can introduce nuance to query term expansion as illustrated by this example: In a search for *lawyer* a document gets full credit for *attorney* and half a credit for *judge*. This "semantic nearness" weights could be determined for each search, a rather cumbersome process, or they could be given in a KOS for all suitable term pairs.

The importance weight of a concept or term could be given in a KOS rather than being computed as a function of document frequency. For example, a term representing a document type could be given a low importance even if the term for particular document type is rare in the collection. Or in a search conducted for a medical expert could give high importance to terms from medical expert terminology and low importance or health consumer terms (as seen from the audience level indicator given in a KOS).

There are two further issues relating to query term expansion and search;

- (1) Searching across multiple databases using different controlled index languages, for example, across multiple databases of electronic health records.
- (2) Searching across multiple natural languages (cross-language search)

Both require mappings from one language to another; such mappings can be found in suitable KOS. The mappings may be imprecise (the problem of semantic interoperability).

Finally, there is a search method known as "more like this". It works by starting from a known relevant document (sometimes a document found in an earlier stage of a search). The system then searches for similar documents, often using terms as the document features for computing similarity. This method might work better if one first applies the words-to-concepts methods discussed in Section 3.1 and the uses the resulting concepts as features.

### **3.2 User-system interaction in search, exploration, processing search results, and sense making**

These processes are closely intertwined, so they are discussed here together. There are large overlapping bodies of literature on each of these; this section gives just a few examples how KOS can support these processes.

Users can profit much from a meaningful well-structured display of search results (the first 100 or 500 or whatever cutoff is chosen) as opposed to a simple ranked list. Here are some examples:

- Arrange search results by a classification, preferably one that is known to user. Usually this requires assignment to classes "on the fly" (see Section 3.2.2) (fitting results into an existing structure).
- Arrange search results in clusters arrived at by knowledge-based clustering (creating new structure bottom-up).

- Show search results on a map, using entity recognition of places (see Section 3.2.2) in conjunction with a gazetteer that gives coordinates
- Arrange search results in a structure that supports solving a problem. This would first list documents that discuss the problems as a whole and the remaining results by one or more of the following: steps in the problem-solving process, problem components or sub-problems, factors to be considered in the solution. This arrangement would use a KOS of problems as described in Section 1.1.
- An overview of the major topics, places, organizations, people (distinguished by whether they are authors or document subjects), and perhaps other entities occurring in the search results and their relationships, both relationships drawn from co-occurrence analysis in the search results and from one or more KOS. This requires KOS support to make sure that all occurrences of a given entity, say an organization, are brought together regardless of the various labels used in documents for that entity. It also requires a topic hierarchy since topic occurrences need to be aggregated at fairly broad topics. From any element in the overview one could drill down to individual documents or to more information on any node found in some KOS or another knowledge source, such as a biographical data base. Such an overview facilitates topic exploration and query modification. It is particularly useful in collaborative information retrieval.

For reviewing the results it is useful to have summaries of the documents, especially multi document summaries. The system can support the user in working with and making sense of the information in the document by dividing the documents into meaningful segments, selecting the relevant segments, and arranging them by topic. Even better, the system could put together a hypertext consisting of all relevant segments extracted from the documents found.

To support sensemaking even further, the system should use information extraction (see Section 3.7) to represent the combined relevant knowledge contained in the documents found as a potentially large set of statements that can then be organized in various ways to see patterns. For example, assume a US intelligence analyst wants to track the movements and potential meetings of suspected leaders and members of terrorist groups. From a large set of news stories, broadcasts, intercepted communications, and field reports, the system can extract many statements about terror group activities, among them statements of the form

Person P <*wasInLocation*> (Location L, PointInTime, PointInTime)

A difficulty is that both people and locations can be referred to by many different names, some public and some secret code names that US intelligence has obtained. All references to the same person or the same location must be unified so that each person or location is always referred to by the same identifier. Now the system can group all these statements on the whereabouts of people together and then figure out when two people of interest were at the same location at the

same time. For another example, assume that a system extracted many statements from a large set of documents dealing with all aspects of obesity. Assume the system has used several KOS to ensure that in the extracted statements each factor or variable has one unique identifier. The system could then group all cause-effect statements and piece them together in one large causal map. There may well be variables that differ only slightly in their definition; in a KOS such variables could be grouped under one broader variable in a hierarchy. To make a causal map simpler and perhaps more useful, one might use that hierarchy to transform extracted statement to map one or both of the connected variables to a variable higher up in the hierarchy. Two original statement may transform into the same statement, resulting in a simpler causal map.

All of this requires that each of the entities mentioned in one or more documents is referred to by the same label or identifier throughout no matter how many different labels were used in different documents. This, in turn, requires access to several comprehensive KOS and other knowledge bases.

### 3.3 Supporting reading and quick assimilation of information

Users can be assisted in understanding and quickly processing text and extracting information from text by one or more of the methods listed below. Users can be similarly assisted in extracting needed information from images, but we will not discuss this here. Most of the methods listed below rely on or can improved by first processing a document by the word-to-concept mapping and other methods discussed in Section 3. (Note: There is much work in education on tools to assist struggling readers. Our focus here is on supporting all kinds of users in assimilating information. The issues and methods are related.)

- Translate.
- Highlight all occurrences of search concepts (or a concept specified by the user) in the text, possibly using colors to distinguish concepts.
- Highlight entities of certain types by different colors (for example people = red, organizations/companies = blue, products = green, places = yellow, dates = brown).
- Give the user information on words and phrases on demand: definition, synonyms, translation into the user's preferred language, conceptual context (linear hierarchy, related concepts, concept map), images.
- Provide tools for navigating the text: Outline, mark meaningful segments in the text below the smallest section level (see 3.3), hypertext map (either based on the native hypertext structure of the document or on an automatically created hypertext version, see Section 3.4), concept map showing the major concepts covered in the text and their relationships. In each: drill down to appropriate location(s) in the text.
- Provide tools for searching within the text in order to limit reading to just the relevant parts. If the text is in a hypertext structure, the system can lay out a reading path that

includes all the pieces of the text that are required to understand the part that gives the information needed to solve the problem at hand.

Giving the user information on words and phrase uses KOS directly. The others are based on preprocessing the text with methods that use KOS support.

Research issues

- Effect on the speed of assimilating the important information from the text
- Effect on the quality of comprehension
- Effect on the appropriate application of information in the text to the problem at hand

#### **4 DETECTING UNDERLYING MEANING AND STRUCTURE AND CREATING NEW STRUCTURE IN WHOLE COLLECTION OR IN SEARCH RESULTS**

##### **4.1 From words and images to concepts: phrase recognition, word sense disambiguation, term to concept mapping, coreference resolution**

Much information seeking, retrieval, and use is based on text (mostly printed, but also spoken), with still and moving images also taking on increasing importance. But to understand, process, and reason with information we need a more formal representation using concepts (understood here very broadly as people, organizations, things, abstract concepts – any entity value). This change in representation can be applied in all processes discussed in this paper, especially in the functions discussed in Section 3.2 – 3.7 and might lead to improvements.

This issue warrants further discussion. Most processes discussed in this paper, especially the functions discussed in Section 3.2 – 3.7, rely on document features. Often words in the text are used as the document features; in some cases there is also phrase recognition so that phrases can be used as document features. An alternative is to first apply the words-to-concepts methods discussed in this Section 3.1 and then use the resulting concepts as features. Whether and under what circumstances this works better for what processes is an empirical question providing many opportunities for research. The change in representation loses information. *myocardial infarction* and *heart attack* will both be mapped to the same concept (perhaps using the concept URI). This mapping loses the information that in one document the medical expert term is used and in the other a lay term; sometimes this is important. For another example, it might be important for an intelligence whether in some intercepted message a suspected terrorist is mentioned by his public name or by a supposedly secret code name. Some terms may have a derogatory connotation that gets lost in mapping to a standard concept. To have the best of both worlds, one could map from words to concepts but carry special markers for audience level, connotation, and other characteristics of interest.

The techniques discussed here are very important in getting from words to concepts. Brief remarks on each must suffice here, emphasizing KOS support.

#### 4.1.1 Phrase recognition (phrase detection)

This is very important, at least in English, because meaning is often conveyed by a phrase, not by individual words. Phrase recognition is best accomplished through a combination of syntactic and/or statistical analysis and matching with a large phrase dictionary. Phrases are found in many technical KOS (dictionaries, thesauri, glossaries, ...), in acronym dictionaries (as most acronyms are brief forms of phrases), and in directories of organizations.

#### 4.1.2 Getting the base form of a term or phrase

Getting the *base form* of a term or phrase is a prerequisite for term-to-concept mapping. This includes morphological normalization (including stemming) and unifying spelling variants, including misspellings.

#### 4.1.3 Word sense disambiguation (WSD)

*Word sense disambiguation (WSD)* is as important as it is difficult. On closer inspection, most words, in English anyway, have several meanings; to go from words (or, more accurately, phrases) to concepts requires selecting the meaning that applies in the context. Methods to resolve the ambiguity of a word include

- Recognize the word as part of a phrase, such as *fire station*. Or consider the highly ambiguous word *pool* in the phrases  
*swimming pool, gene pool, betting pool, office pool, funding pool, pool table.*
- Determine the part of speech: consider the word *fire*, which has at least five meanings as seen from KOS that gives, for example, the meanings for *fire [verb]* and for *fire [noun]*
  - *fire* as a verb (narrows meanings down from 5 to 2: *fire (dismiss from a job)* and *fire (shoot)*)
  - *fire* as a noun (has a literal meaning and two metaphorical meanings)
- Use information that is useful for disambiguation found in a KOS. For example, a KOS may give for each sense of a word
  - the prevalence in a given context (in a subject field or specialty, in an organization); for example, WordNet gives for each sense the frequency with which the word is used in that sense in the Brown corpus of standard English.
  - surrounding words that indicate that the word is likely used in that sense.

#### 4.1.4 Term-to-concept mapping

This is one of the major functions of classical thesauri.

#### 4.1.5 Coreference resolution

Coreference resolution means to identify mention of the same person, place, action, event, etc. (the same entity value in a text or across several texts, for example, across several email messages).

One form of coreference resolution is finding all instances where the same entity is referred to by different terms or name. This is accomplished simply by term-to-concept mapping, replacing the term or name by a unique identifier (such as the preferred term or name or the URI (Universal Resource Identifier) for the entity. The second form of co-reference is reference to an entity by a pronoun (more generally: proform), an anaphoric or cataphoric reference. It is this second form we will focus on in this section, and we simply call it anaphora.

Coreference resolution is important for many functions, for example

- information extraction and extracting sentences for an automatic summary;
- proximity searching (one of the two concepts required to appear near each other in a text may be referred to by an anaphoric reference such as *it* or *this disease*);
- counting the mentions of a concept in a text;
- disambiguation.

For information extraction consider these two examples:

**Example 1.** From the text

*Plaque in the arteries can be caused by high levels of cholesterol. This condition can be treated with simvastatin*

a good system would extract

*High levels of cholesterol can be treated with simvastatin*

**Example 2.** From the text

*Cranial arteritis is the most common cause of sudden blindness in the elderly. Usually only one eye is involved but sometimes it occurs in both. This condition is successfully treated with corticosteroids.*

a good system would extract

*Cranial arteritis can be treated with corticosteroids.*

Anaphora resolution can be supported by a KOS of words and phrases used for anaphoric reference, such as pronouns (more generally proforms), demonstratives, and general terms such as *the disease*, *this disease*, *the system*, and give for each of these the type(s) of entity it may refer to, for example (simplified)

*she* – female person, female animal, ship

*it* – any object, place, ...

The system must then find an antecedent of the proper entity type.

This information can assist with word sense disambiguation, as in this example: *Lee* is a male or female first name or surname, thus refer to a male or female person; it is also the name of many places. Anaphora resolution can contribute to disambiguation:

Lee is beautiful. She is shown here at the Met gala (*Lee* is a female person)

Lee is beautiful. It has many bed and breakfasts (*Lee* is a thing or place)

#### 4.1.6 From images to concepts

With the present state-of-the-art of content-based image retrieval (CBIR), it is possible to map from images and image detail to concepts (in the broad definition given above). Given a large KOS that includes one more images for many of its concepts, a CBIR system can start from an image, possibly divide it into meaningful segments, then find similar images in an image-enriched KOS, and assign a corresponding descriptor, such as *female person* or *Hillary Clinton* or *Greek vase*. A program could also analyze a causal map and extract causal relationships from it.

### 4.2 Indexing and metadata creation

#### 4.2.1 Human indexing or tagging, including social tagging

Retrieval performance is increased if a document or document record includes terms that express concepts users want to search for. The purpose of request-oriented (or user-centered) indexing is to increase the chance that this is the case. User concepts may not be explicitly mentioned in a document even though the document is relevant. In that case it takes an intelligent indexer familiar with user needs to assign the appropriate descriptor. In request-oriented indexing it is the function of the index language to crystallize the concepts of interest to the intended audience in a systematic structure that communicates user interests to the indexers. This structure must be presented in a meaningful arrangement of facets and meaningful hierarchy within each facet so that the indexer can use this presentation as a checklist and check the descriptors for which the document is relevant.

The KOS must also support finding the right descriptor starting from a term in the document or a term the indexer thinks of, and make further suggestions using concept relationships (RT). For example, if an indexer enters *cancer* as a descriptor, the KOS should remind the indexer to also index with a body part by showing the top level of the body part hierarchy.

There are attempts at improving social tagging by showing the user a facet frame to solicit tags from several perspectives, making spelling corrections, suggesting a preferred label from an



organized KOS to reduce harmful variety, and suggest additional tags based on KOS structure. One system presents a social tagging interface that encourages users to collaboratively build a hierarchy of tags to be used in tagging, but whether this is the best use of tagging volunteers' time is questionable.

#### **4.2.2 Computer-assisted and automatic indexing and classification. Semantic enrichment, entity recognition**

There are two approaches to automatic indexing.

Approach 1 selects terms from the document based on some weighting scheme, often just term frequency in the document but sometimes also considering a term's intrinsic importance, similar to TF-IDF weighting schemes in search. This approach is document-oriented rather than request-oriented. One can hypothesize that this approach would be improved by first applying the methods discussed in Section 3.1 to the document text. The terms or concepts finally selected could be mapped to the descriptors of a KOS, perhaps with some loss of information.

Approach 2 uses rules that use document features, sometimes including certain patterns in the text, to predict descriptors from an index language (including classes from a classification). The totality of these rules is often called a *classifier*. The rules can be crafted manually or derived through machine learning from a training set. They could be as simple as mapping from a text string to a concept in the target index language. A KOS can assist in manually crafting rules. Any classifier, whether constructed manually or through machine learning, is a KOS. Statistical procedures, to which machine learning belongs, require a certain number of cases to work well. Therefore, one might suspect that classifiers produced by machine learning could work well for assigning broad descriptors, such as genre or general subject discipline but not for specific descriptors that occur only a few times in million document collection (and are very valuable for a user interested in such a descriptor precisely because it is rare).

*Semantic enrichment* is augmenting existing metadata for a document (which may be none), either manually or (the default case if nothing else is specified) automatically. The added metadata may include any of the following:

- Concepts for which the document is believed to be relevant, generally concepts from a standard or quasi-standard indexing language (such as the Gene Ontology) so as to create a link from the document to the standard index language. These concepts can be assigned through automatic indexing as discussed above
- Entities (such as people, organizations, projects, places, products, ..., often called *named entities*) that have been recognized in the text and that – here is the key – have been assigned an
  - an entity type from a standard scheme and

- an entity identifier from a standard or quasi-standard scheme for that entity type

This information is often given through XML tags around the text string that represents the entity in the text. Adding this information relies heavily on KOS (in the broadest sense, including people, organization, and other reference data) makes it possible to reliably retrieve all documents that mention a certain entity and, as importantly, to link to places where more information about this entity can be found, both inside an organization's information systems / intranet and on the Web at large, the essence of linked data.

#### **4.3 Document segmentation. Detecting meaningful segments**

Segmentation applies to text, still images, moving images (film, video). It is merely mentioned here as one of the processes using document features, offering the opportunity to test the hypothesis that using concepts as document features improves segmentation.

#### **4.4 Creating hypertext, more generally hypermedia. Multi-document hypermedia documents**

Creating a hypermedia document can work just on the document sections defined by the author, but a more fine-grained approach using also automatically identified meaningful segments at the level below the section might be more useful. Obviously hyperlinks can be derived from the formal document structure. Other links could be identified based on concept relationships found in a KOS. For example, Segment 15 might deal with concept A and Segment 37 with concept B, and a KOS shows a relationship A <causes> B; one might create a hyperlink between Segment 15 and Segment 37.

A multi-document hypermedia document may contain segments of multiple texts and segments of multiple images (still and/or moving), all interlinked. For this to work, the terminologies among all the texts must be aligned with each other and with the terminology used for the features detected in images. For example, if the same person is mentioned in several texts and depicted in several images, the system needs to make sure that the same identifier is used for that person throughout the hypermedia document and connect all occurrences with links. This is where KOS and other "reference data" systems can help.

#### **4.5 Abstracting and summarization, human or automatic. Multi-document summarization**

Abstracting and summarization can work with text at the surface (original text or text transformed through word-to-concept mapping) or it can work through a deep understanding of the text, which can be operationalized as extracting a set of statements that among them capture the

information the text conveys. If this is done well, information overlap among several documents can be detected and eliminated from a multi-document summary. Likewise, contradictory or deviating statements can be detected and differences between documents can be highlighted.

#### 4.6 Classification, grouping, clustering

This is another example where the effectiveness of transforming text through word-to-concept mapping can be studied. The result of clustering is a newly created structure, possibly a multi-level hierarchical structure, that might help a user in analyzing a problem. Clustering at the same time creates classes and assigns documents to these classes. In that way, it contributes to semantic enrichment.

#### 4.7 Information extraction

Entity recognition identifies the entities (people, organizations, places, abstract concepts, etc.) mentioned in a document, text or image, a process highly dependent on KOS as we saw above. Information extraction identifies statements about relationships between these entities, which is the actual information the document conveys. This can be supported by a KOS that gives extraction templates and cue words that indicate relationships.

Here are two examples:

Text	Pattern	Extracted assertion
Stress <i>due to</i> {the effect of} wind	Effect <i>due to</i> Cause	Wind <causes> Stress
Adding 10% aluminum <i>doubles</i> the strength of copper	Cause <i>doubles</i> Effect (Also <i>reduce, increase, ...</i> )	Adding 10% aluminum <causes> 2x strength of copper

It can be seen that a general KOS of this nature would be very large.

## 5 INTERFACE DESIGN. VISUALIZATION

A good KOS can assist interface design in two major ways:

- A KOS can suggest good labels for items, such as concepts, places, products, or people named in the interface, including images (as in a catalog for very young children), internationally standardized icons for language-independent communication, or terms in other languages for interface localization
- A KOS can suggest meaningful arrangement of items in a list or of the overall screen layout. Meaningful arrangement is easier to grasp and memorize, collocates items that the user might want to consider as likely alternatives, and might even support the user in learning something.

Good visualizations of KOS structure can promote use of KOS. A major principle in visualization is to keep it simple and do not deviate too much from the realm of the intended users' experience. Unfortunately, many visualizations deviate from this principle. Often a hierarchy display in simple outline form, with levels indicated by indention and major subdivisions signaled by larger / bold type is much easier for users than complex circular diagrams. On the other hand, a well-designed concept map or other relationship diagram that uses box shape to distinguish entity types and colors to distinguish relationship types may be quite useful.

## **6 CONCLUSION**

This concludes a wide-ranging tour of the pervasive applications of KOS that have the potential of vastly improving information seeking, retrieval, and use, including sensemaking and learning. I hope that this presentation will inspire many projects of designing and evaluating innovative systems that will improve service to users. And perhaps recognition of the importance of KOS will stimulate the emergence of systems that facilitate unified and easy access to the rich information available in many different Knowledge Organization Systems to promote the application of this knowledge.

## **REFERENCES**

Note: An extensive bibliography for this paper, organized by section, will be available by the author after June 30, 2016 at MHTX's blog.

## MESA-REDONDA

### DEBATES SOBRE AS PRÁTICAS DE ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO NO MERCADO DE TRABALHO SOB A ÓTICA DA BIBLIOTECONOMIA

Moderadora: Célia da Consolação Dias (ECI/UFMG)

Participantes:

Profa. Dra. Elisângela Cristina Aganette (ECI/UFMG)

Mestre Alessandra Rodrigues (EMBRAPA)

Mestre Fernanda Pereira (IFMG/Barbacena)

Mestre Graciane B. Borges (ECI/UFMG)

Mestre Kelly Cristiane Santos Moraes (Colégio Franciscano Sagrada Família)

#### A organização e a representação da informação no mercado de trabalho sob a ótica da Biblioteconomia

*THE ORGANIZATION AND REPRESENTATION OF INFORMATION IN THE LABOR MARKET FROM THE PERSPECTIVE OF LIBRARIANSHIP*

Célia da Consolação Dias (ECI/UFMG)

*celiadias@gmail.com*

**Resumo:** Este trabalho apresenta algumas das discussões apresentadas na Mesa 1 do segundo Seminário MHTX que foi realizado na Escola de Ciência da Informação em junho de 2016. A temática da Mesa 1 compreende o relato da experiência de cinco bibliotecárias sobre suas práticas profissionais de organização e representação da informação em diferentes cenários informacionais. Em termos de conteúdo foram tratadas temáticas como organização do conhecimento e da informação, bem como as estratégias adotadas pelas bibliotecárias no desenvolvimento de projetos em contextos tão diversos como as bibliotecas escolares que exige a organização da informação para usuários em plena formação de sua competência informacional, em empresas abordando a organização da informação tecnológica para usuários especializados como o caso da Embrapa, os desafios e as oportunidades dos profissionais que trabalham como consultores em gestão de informação e gestão dos documentos no contexto corporativo e por último a análise das perspectivas para o profissional da informação.

**Palavras-chave:** Organização da informação. Mercado de trabalho. Biblioteconomia.

**Abstract:** This paper presents some of the discussions presented in Table 1 of the second MHTX Seminar which was held at the Information Science School in June 2016. The theme of Table 1 comprises the story of five librarians experience on their organization of professional practices and representation of information in different informational scenarios. In terms of content they were treated topics such as organization of knowledge and information, as well as the strategies adopted by librarians in developing projects in such diverse contexts as school libraries that requires the organization of information to users in full training their information literacy, companies dealing with the organization of technological information for specialized users as the case of Embrapa, the challenges and opportunities of professionals working as consultants in

information management and document management in the corporate context and finally the analysis of the prospects for professional information.

**Keywords:** Information organization. Job market. Librarianship.

## 1 INTRODUÇÃO

A organização da informação é um processo de arranjo de acervos tradicionais ou em formato eletrônico realizado por meio da descrição física e do seu assunto. As atividades relacionadas à organização da informação e do conhecimento estão no *core* das atividades realizadas pelo bibliotecário para preparar um determinado acervo informacional e disponibilizá-lo para recuperar informação.

Este trabalho apresenta algumas das discussões apresentadas na Mesa 1 do segundo Seminário MHTX que foi realizado na Escola de Ciência da Informação em junho de 2016. A temática da Mesa 1 compreende o relato da experiência de cinco bibliotecárias sobre suas práticas profissionais de organização e representação da informação em diferentes cenários informacionais. Em termos de conteúdo foram tratadas temáticas como organização do conhecimento e da informação, bem como as estratégias adotadas pelas bibliotecárias no desenvolvimento de projetos em contextos tão diversos como as bibliotecas escolares que exige a organização da informação para usuários em plena formação de sua competência informacional, em empresas abordando a organização da informação tecnológica para usuários especializados como o caso da Embrapa, os desafios e as oportunidades dos profissionais que trabalham como consultores em gestão de informação e gestão dos documentos no contexto corporativo e por último a análise das perspectivas para o profissional da informação.

O texto está organizado da seguinte forma: uma síntese dos aspectos teóricos sobre a organização do conhecimento e da informação, as práticas de organização e representação da informação no mercado de trabalho, as considerações finais e as referências usadas neste texto.

## 2 A ORGANIZAÇÃO DO CONHECIMENTO E DA INFORMAÇÃO

Diversas palavras-chave estão relacionadas à organização e representação da informação e, mais especificamente quando se pretende relacionar tais assuntos às práticas associadas ao fazer do bibliotecário. Ainda pensando nas palavras-chave relacionadas ao assunto em questão, pode se destacar algumas tais como: organização da informação, representação da informação, biblioteconomia. De acordo com Saracevic (1996, p. 48) “a biblioteconomia tem uma longa e orgulhosa história, remontando a três mil anos, devotada à organização, à preservação e ao uso dos registros gráficos humanos”.

Para começar o entendimento sobre o assunto ressalta-se que a organização da informação compreende a descrição dos documentos de acordo com seus aspectos físicos e temáticos. Vale a pena dizer que, deve-se pensar em documentos de forma ampliada e não só nos documentos que fazem parte do universo bibliográfico como livros ou periódicos em formato impresso/ eletrônico e que são tão comuns nas bibliotecas. Mas, também pensar que há um universo composto por objetos diversos, inclusive os objetos digitais presentes nos domínios de forma geral.

Para a ciência da informação a representação da informação sempre fez parte do trabalho de permitir o acesso às informações das diversas coleções presentes nas bibliotecas, arquivos e museus e que, estão disponíveis em mídia impressa, eletrônica ou outro suporte do contexto digital que as organizações possuem e precisam manusear no dia-a-dia para implementar as suas atividades. Permitir o acesso significa criar recursos para a coleta dos dados extraídos do documento e que possibilita ao usuário identificar os seus elementos físicos e, conseqüentemente levar as pessoas até o material desejado para satisfazer a sua consulta.

O resultado de tal preocupação pode ser verificado na construção dos instrumentos das linguagens documentárias que, em função da evolução da tecnologia e dos sistemas de informação tornam evidentes a necessidade de um processo cada vez mais refinado para a busca e a recuperação de informações. A análise documentária se preocupa com o conteúdo dos documentos e ao mesmo tempo com a recuperação da informação no sentido de desenvolver os procedimentos e os instrumentos para apresentar este conteúdo. Assim, a análise documentária pode ser definida como “um conjunto de procedimentos efetuados com o fim de expressar o conteúdo de documentos, sob formas destinadas a facilitar a recuperação da informação” (CUNHA, 1987, p. 38).

Como exemplos de linguagens documentárias podem ser citados as listas de cabeçalho de assunto, os tesouros, as taxonomias e as ontologias.

### **3 AS PRÁTICAS DE ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO NO MERCADO DE TRABALHO**

Sempre que se discute sobre as práticas de organização e representação da informação para a atuação do profissional bibliotecário, pergunta-se: que mercado é esse que tal profissional pode atuar? Existe mercado para atuação do bibliotecário relacionada à organização e representação da informação? Para responder tais perguntas é necessário pensar que os cenários informacionais que demandam organização não se restringem apenas às bibliotecas. Mas, a todas as organizações que possuem informação e que, por esse motivo elas produzem, recebem, processam, disseminam,

precisam recuperar e usar informação diariamente em suas atividades, inclusive para gerar novos conhecimentos.

Ressalta-se nesse sentido que, as metodologias, os instrumentos, os processos, os padrões usados para organizar informação e conhecimento são usados, também, para organizar livros e materiais bibliográficos em geral. Entretanto, podem ser usados para o desenvolvimento de representações de informação em suas dimensões descritivas e temáticas sobre os inúmeros objetos informacionais físicos e digitais que atualmente precisam ser gerenciados por todas as instituições públicas ou privadas e otimizam a recuperação da informação.

Nesse aspecto os instrumentos de organização e representação da informação e do conhecimento utilizados pelos bibliotecários oferecem inúmeros benefícios tais como:

- 1) Mapeamento conceitual conectados com os objetos informacionais de um domínio qualquer.
- 2) Estruturação e organização dos conceitos presentes nos objetos do domínio;
- 3) Uso de linguagens específicas, normas e formatos que padronizam a descrição dos objetos informacionais.
- 4) Criação de produtos de informação com objetivo específico de sintetizar informações sobre os objetos e que facilitam a recuperação da informação.
- 5) Permite a localização de informações em bases de dados a partir do uso de termos e relacionamentos entre esses termos organizados em um vocabulário controlado.
- 6) Permite a localização de objetos relacionados com outros objetos a partir do controle dos relacionamentos presentes no vocabulário organizado.
- 7) Permite a criação de novo conhecimento a partir do existente.

Com o objetivo de ajudar na realização de todas as tarefas indicadas acima, Bräscher; Carlan (2010) ressaltam o papel dos sistemas de organização do conhecimento (SOC). Os SOCs são tipos de sistemas conceituais, ou tipos de representações do conhecimento, frutos do processo de organização do conhecimento. Dahlberg (2006) citado por Bräscher; Carlan (2010) afirma que tais sistemas são aplicados para mapear objetos informacionais, ou seja, para representar os assuntos dos documentos num sistema de informação. Ainda, em relação a esses tipos de sistemas de organização do conhecimento Soergel (2008) citado por Bräscher; Carlan (2010) amplia o seu papel ao afirmar que tais sistemas fornecem esquemas para armazenar e organizar dados, informação, conhecimento sobre o mundo e seus pensamentos. Nesse sentido, os elementos que fazem parte deste sistema são os conceitos, as categorias, classes, os relacionamentos, os termos e seus relacionamentos.



Apenas para exemplificar, tais metodologias, instrumentos, esquemas e padrões podem ser usados pelos bibliotecários para organizar de maneira mais geral ou específica todos os recursos em todos os tipos de bibliotecas. Para além das bibliotecas, entretanto, todo esse ferramental ou parte dele pode ser usado para a organização de informação de objetos informacionais, tais como imagens de objetos no domínio da história da arte, objetos de um laboratório do tipo biológico, objetos de um cenário artístico, informações de uma instituição financeira, dentre outros.

É importante lembrar, ainda, que organizar informação e conhecimento em ambientes digitais traz novos desafios relacionados tanto à gestão dos acervos, definição de tecnologias de informação, questões relacionados aos direitos autorais dos recursos e também quanto à representação descritiva ou da catalogação dos dados que descrevem os objetos. Nesse contexto digital, os dados dos catálogos dos sistemas e serviços de informação “encontram-se representados por novos conjuntos de informações, dando origem a produtos que poderiam ser considerados metainformações ou metadocumentos” (ALVARENGA, 2006, p. 77).

Os metadados possuem diversas funções entre elas podem ser destacadas a de descrever, localizar, representar, recuperar e descobrir recursos presentes nos objetos informacionais, por exemplo. Lourenço (2005) em sua pesquisa de doutorado sobre o Padrão Brasileiro de Metadados de Teses e Dissertações apontou a descrição dos objetos informacionais como uma das funções dos metadados. Ainda nesta linha de pensar sobre o que são os metadados Cleveland (1998) citado por Lourenço (2005, p. 51) nos apresenta uma definição. Segundo ele

Metadado é o dado que descreve o conteúdo e os atributos de qualquer item em uma biblioteca digital. É um conceito familiar para bibliotecários porque é uma das primeiras coisas que eles fazem – criam registros catalográficos que descrevem documentos. Metadado é importante em bibliotecas digitais porque é a chave para a descoberta de recursos e usos para qualquer documento. (CLEVELAND, 1998).

Tais recursos facilitam gerenciar e recuperar informação, esteja na forma eletrônica ou não. Nesse contexto, Bräscher e Café (2008) ressaltam que a descrição da informação é fundamental na recuperação da informação, pois cumpre a função de possibilitar o acesso a itens específicos que se encontram armazenados. Ainda em relação aos metadados Gilliland-Swetland (2002, p.3) citado por Lourenço (2005) discute a importância dos metadados. Estes recursos são importantes para recuperar informações analíticas e resumidas dos objetos de informação que possam ter sido tratados por profissionais da informação; identificar e explorar as relações estruturais externas e internas dos objetos de informação com outros objetos de informação e para fornecer pontos de acesso para os usuários recuperarem estes objetos informacionais.

#### 4 CONSIDERAÇÕES FINAIS

Tratar do assunto organização e representação da informação conectada ao mercado significa um desafio duplo. Primeiro para os professores que trabalham com a formação dos alunos de biblioteconomia e têm a missão de apresentar aos discentes e futuros bibliotecários o universo teórico da organização, representação da recuperação da informação. Tais discussões giram em torno da compreensão dos fundamentos teóricos da área, bem como o conhecimento sobre o uso dos instrumentos, metodologias, esquemas, padrões e processos relacionados à área de organização e a possibilidade de aplicação de todo esse arsenal às situações da realidade que, demandam pela intervenção do bibliotecário.

Do outro lado observa-se um desafio para a o bibliotecário que desempenha o papel fundamental de implementar as atividades de organização. Nesse sentido, as atividades de organização e representação se revelam complexas não apenas do ponto de vista técnico, mas também acerca do perfil do bibliotecário que demandam entre outras qualidades a de síntese, a de negociador, a de investigador em busca da compreensão do domínio, entre outras. Todo esse esforço tem como objetivo segundo Dahlberg (1993), estruturar e organizar sistematicamente unidades do conhecimento, os conceitos, e ligar esses conceitos aos objetos e assuntos que, sua vez serão ligados a outras bases de dados que vão conversar com outros sistemas de informação. Além disso, tal profissional tem também a preocupação com o processo de categorização das coisas, do nome dado às coisas, da organização e configuração de um vocabulário, com o uso dos padrões que permitem a interoperabilidade entre os sistemas.

Por fim, ressalta-se que onde existir conhecimento para ser organizado existe a oportunidade de aplicações das práticas de organização pelo bibliotecário.

#### REFERÊNCIAS

ALVARENGA, L. Organização da informação nas bibliotecas digitais. In: NAVES, M. M. L.; KURAMOTO, H. (Org.). **Organização da informação: princípios e tendências**. Brasília, DF: Briquet de Lemos, 2006. p. 76-98.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 9., 2008, São Paulo. **Anais...** São Paulo: ECA/USP, ENANCIB, 2008. Disponível em: <[http://saladeaula01.eci.ufmg.br:8080/xmlui/bitstream/handle/123456789/462/organizacao\\_\\_conhecimento\\_3.pdf?sequence=1](http://saladeaula01.eci.ufmg.br:8080/xmlui/bitstream/handle/123456789/462/organizacao__conhecimento_3.pdf?sequence=1)>. Acesso em: 27 maio 2016.

BRASCHER, M.; CARLAN, E. Sistemas de organização do conhecimento: antigas e novas linguagens. In: ROBREDO, J.; BRASCHER, M. (Org.). **Passeios pelos bosques da**

**informação:** estudos sobre representação e organização da informação e do conhecimento. Brasília: IBICT, 2010. p. 147-176.

CAFÉ, L. M. A.; SALES, R. de. Organização da informação: conceitos básicos e breve fundamentação teórica. In: ROBREDO, J.; BRÄSCHER, M. (Org.). **Passeios pelo bosque da informação:** estudos sobre representação e organização da informação e do conhecimento. Brasília, DF: IBICT, 2010. 335 p. ISBN: 978-85-7013-072-3. Disponível em: <<http://repositorio.ibict.br/bitstream/123456789/36/1/eroic.pdf>>. Acesso em: 20 abr. 2016.

CUNHA, I. M. R. F. Análise documentária. In: SMIT, Johanna W. (Org.). **Análise documentária:** a análise da síntese. Brasília: IBICT, 1987. p. 38-60.

DAHLBERG, I. Knowledge organization: its scope and possibilities. **Knowledge Organization**, v. 2, n.4, 211-222, 1993.

LOURENÇO, C. A. **Análise do padrão brasileiro de metadados de teses e dissertações segundo o modelo entidade relacionamento.** 2005. 165 f. Tese (Doutorado em Ciência da Informação)– Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte, 2005.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

## Consultoria em gestão de informação: desafios e oportunidades

*CONSULTANTS MANAGEMENT INFORMATION: CHALLENGES AND OPPORTUNITIES*

Elisângela Cristina Aganette (ECI/UFMG)

*elisangelaaganette@hotmail.com*

**Resumo:** O presente estudo é um relato de experiência sobre a atuação da autora no mercado de trabalho na área de Biblioteconomia e Ciência da Informação, por meio da atuação consultiva em Projetos de Gestão de Informação, desenvolvidos em várias organizações brasileiras. Projetos esses cujo escopo consiste no desenvolvimento de serviços e produtos informacionais, tais como: diagnósticos, políticas, procedimentos operacionais, taxonomias, ontologias, planos de classificação, mapeamento de processos, tabelas de temporalidade, organização física e eletrônica de documentos, apoio na modelagem e implantação de sistemas, entre outros. A gestão documental significa a manutenção e a competitividade de uma organização, conduz a transparência dos processos e atividades. A ausência de uma Gestão Documental causa prejuízos e problemas de toda de ordem: administrativa, financeira e jurídica. Independente do setor, percebe-se que há um cenário de muitas oportunidades de atuação para o “profissional de informação”, uma vez que onde há problemas informacionais há necessidade de solução, e após resolvidos e solucionados tais problemas, as oportunidades continuam a existirem, uma vez que a gestão documental necessita de manutenção. Há oportunidades, mas não significa ser fácil, trata-se de um mercado cada vez mais exigente que demanda profissionais de alto nível e qualificados. A qualificação deve-se ser contínua para atendermos com excelência as exigências do almejado mercado. O relato desta experiência visa colaborar com a motivação e o desenvolvimento profissional relatando cenários, práticas e oportunidades de atuação, fora do contexto de biblioteca.

**Palavras-chave:** Gestão de Informação. Consultoria. Projeto de Gestão de Informação.

**Abstract:** This study is an experience report on the work of the author in the labor market in the field of Library and Information Science, through advisory role in Project Management Information developed in several Brazilian organizations. These projects whose purpose is the development of services and information products, such as diagnoses, policies, operating procedures, taxonomies, ontologies, classification schemes, process mapping, temporality tables, physical organization and electronic documents, support in modeling and implementation of systems, among others. The document management means maintaining and competitiveness of an organization leads to transparency of processes and activities. The absence of a document management causes losses and problems of all order: administrative, financial and legal. Regardless of industry, it is clear that there is a scene of many opportunities for action for the "professional information", since where there are information problems is needed solution, and after solved and solved such problems, opportunities still exist, once the document management needs maintenance. There are opportunities, but by no means be easy, it is an increasingly demanding market that a high level of professional and qualified demand. The qualification must be continuous in order to comply with the requirements of excellence-targeted market. The account of this experiment aims to collaborate with the motivation and professional development reporting scenarios, practices and opportunities for action outside the library context.

**Keywords:** *Information Management. Consulting. Information Management Project.*

## 1 INTRODUÇÃO

O presente estudo é um relato de experiência sobre a atuação da autora no mercado de trabalho na área de Biblioteconomia e Ciência da Informação. Utilizando-se o dialeto do mercado corporativo, a gestão da informação se faz cada vez mais necessária e urgente. As empresas se vêm, cada vez mais, em meio a um crescente volume de informação, informações essas, necessárias à continuidade dos seus processos de negócio. Dessa forma, gerir, administrar e organizar são ações que se destacam como fundamentais, para o sucesso operacional e estratégico das organizações. Trata-se da necessidade de se:

Implementar um conjunto de medidas que visam a racionalização e a eficácia no uso e circulação de dados e informação e a aplicação das teorias e técnicas da ciência da informação nos sistemas de informação (ALVES, 1993, p. 52).

O objetivo deste relato de experiência consiste em colaborar com a formação acadêmica e o desenvolvimento profissional, por meio da discussão de algumas possibilidades, oportunidades e exigências do referido mercado de trabalho. Além disso, evidenciar a importância da atuação do profissional, por meio de atitudes inovadoras, uma postura criativa e diferenciada; e finalmente, possibilitar discussão sobre a necessária interação entre a academia e o mercado de trabalho, e a consequente troca de conhecimento.

As experiências e rotinas profissionais vivenciadas pela autora e relacionadas ao contexto auxiliam na interpretação dos fenômenos observados. Saracevic (1996) lembra que:

Um campo é definido pelos problemas que são propostos e a Ciência da Informação (CI) é definida como um campo englobando, tanto a pesquisa científica quanto a prática profissional, pelos problemas que propõe e pelos métodos que escolheu, ao longo do tempo para solucioná-los (SARACEVIC, 1996, p. 41).

A Escola de Ciência da Informação, tem se preocupado com as diferentes possibilidades de atuação nos procedimentos de ensino e de aprendizagem em suas áreas de conhecimento. Trata-se de questões, em que não se é possível ausentar-se ou não se posicionar. Dessa forma, a Escola de Ciência da Informação, tem se posicionado favorável a esse ambiente de teoria e prática, por meio de eventos, projetos de pesquisa e desenvolvimento, grupos de estudos, seminários e encontros com profissionais da área. Busca-se com isso, discussões que extrapolam a teoria e convergem para um ambiente real de aprendizagem. Faz-se necessário que os envolvidos nestes temas relacionados à escola estejam em alerta e utilizem na prática, as diferentes experiências mostradas em palestras, *workshops* e grupos de trabalho.

## 2 PERCURSO PROFISSIONAL

### 2.1 Atividades consultivas

Os desafios e dificuldades são muitos na atividade consultiva, mas o aprendizado e o retorno financeiro vêm na medida que se desenvolve um trabalho de qualidade. Como “recordes”, tive períodos na vida profissional que atuei em 4 (quatro) projetos em paralelo, em estados diferentes. Felizmente, todas estas atuações foram no contexto da Ciência da Informação.

Por falar em desafio, este tem início, na relação com o personagem mais importante desse contexto: o Cliente. Cliente esse que, na maioria das vezes, desconhece na sua totalidade, as características do serviço contratado. Contratam as consultorias, motivados por problemas que vivenciaram, ou que estejam vivenciando em relação ao contexto informacional da organização, e que de alguma forma impactam em questões de ordem financeira. Dessa forma, buscam por soluções, que de fato garantam e auxiliem no processo de mapeamento, padronização, organização e recuperação de informação. Soluções essas, que só conhecem por algum meio de divulgação, ou simplesmente por terem “ouvido” os benefícios de algum fornecedor, parceiro ou mesmo concorrente.

Mas os desafios não param na fase de fechamento do negócio. A execução do projeto de informação, demanda uma atividade contínua que busca o envolvimento e comprometimento dos colaboradores da organização. Culturalmente, há uma resistência a mudanças, e ainda, infelizmente, a máxima que “informação é poder”, ainda persiste. Diante disso, durante todas as etapas de execução do projeto, que incluem: levantamento, elaboração, validação dos serviços e ou produtos informacionais e capacitação, faz-se necessário adotar técnicas de conscientização das pessoas envolvidas, assim como esclarecimento de dúvidas, discussão e alinhamento quanto aos benefícios do projeto, detalhamento de etapas etc.

A cada projeto realizado, a cada consultoria prestada, faz-se necessário adequar a forma de atuação nos mesmos. O desenvolvimento dos serviços e produtos informacionais, tais como: diagnósticos, políticas, procedimentos operacionais, taxonomias, ontologias, planos de classificação, mapeamento de processos, tabelas de temporalidade, organização física e ou eletrônica de documentos, apoio na modelagem e implantação de sistemas, entre outros, sempre se dão de forma diferenciada. As especificidades e mudanças, vão desde a forma de elaboração (método de levantamento de insumos, complexidade dos processos, técnicas de validação) até a fase de validação e implantação dos serviços. Isso se dá, uma vez que, cada cliente é único, seus processos de negócio possuem especificidades únicas, e por isso, devem ser tratados e analisados de forma a desenvolver serviços e produtos informacionais de acordo com suas necessidades.

As especificidades de cada projeto de gestão de informação, conforme mencionado, variam de acordo com cada cliente. No entanto, algumas situações referentes ao desenvolvimento e a manutenção dos serviços e produtos informacionais merecem destaque:

- **Metodologia de construção e manutenção:** i) na maioria dos casos não existem padrões e métodos a serem seguidos; ii) apesar de orientado às boas práticas e de estar alinhado aos processos de negócio da organização, ainda demandam reestruturação da forma de operacionalização e implementação dos mesmos; iii) quanto à manutenção, quase sempre não se prevê os princípios básicos de auditoria e retroalimentação.
- **Desenvolvimento dos serviços e produtos informacionais:** essa etapa é de crucial importância para a eficiência e eficácia dos mesmos no momento de implementação e uso. Na maioria dos casos, essa etapa é feita de forma aleatória, por profissionais não qualificados para a atividade; sem orientação metodológica e desvinculado da tecnologia de gestão de conteúdos adotada.
- **Manutenção dos serviços e produtos informacionais:** essa etapa prevê o acompanhamento das mudanças, das necessidades do negócio e da evolução do uso da solução. As ações de manutenção devem prever todos os acontecimentos que possam gerar modificações nos serviços e produtos informacionais, tais como: surgimento de novos processos de negócio; ineficiência na recuperação de informações; tratamento de exceções; fusões ou separações na organização. Na maioria dos casos, as ações de manutenção ou não existem ou não são periódicas e focadas em auditar a rotina de gestão documental e em avaliar a consistência do conteúdo e os resultados efetivos de sua utilização.

Tais desafios e situações, proporcionaram-me experiências, que totalizam 12 (doze) anos de prestação de serviços especializados em gestão de informação. Mas também, trouxeram-me várias inquietações teóricas e acadêmicas, dentre as quais se destacam a ausência de padrões metodológicos para a sua construção, implementação e manutenção da maioria dos serviços e produtos informacionais desenvolvidos e implementados.

## 2.2 Atividades acadêmicas

Paralelamente às atividades consultivas, atuei como professora acadêmica na rede privada de educação, por três anos. Como professora, tive oportunidade de ministrar conteúdos na área de Tecnologia da Informação; Segurança da Informação; Empreendedorismo e Inovação e Ciência da Informação.

Conforme mencionei, os trabalhos consultivos continuaram em paralelo, o que me proporcionou a oportunidade de ministrar aulas com abordagens teórica e prática. Aulas baseadas na realidade de mercado, sem, no entanto, prejudicar o ensino fundamentado em uma base teórica necessária para a aprendizagem e formação acadêmica.

Em algum momento dessa corrida e prazerosa vida profissional, dividida entre mercado e academia, deparei-me com uma situação: incompatibilidade de agendas: doutorado, consultoria e docência. Planejei-me a curto, médio e longo prazo. Dessa forma, deixei de atuar provisoriamente como consultora e como docente, e dediquei-me ao doutorado. A oportunidade de lecionar novamente, veio do setor público federal, por meio de um concurso, na área de Ciência da Informação.

### **3 OPORTUNIDADES DE ATUAÇÃO**

Não é novidade o fato de que o dilema das organizações, continua a ser o grande e crescente volume documental, além de um legado informacional acumulado, seja em meio físico e ou eletrônico. Além disso, as organizações continuam com problemas informacionais e por isso continuam sendo penalizadas administrativamente e financeiramente. A gestão de informação é essencial e significa a manutenção e a competitividade de uma organização; conduz e facilita a transparência dos processos e atividades, por meio da governança das informações e do registro e documentação dos processos de negócio, que assegura de forma eficiente, a geração, controle, manutenção e destinação dos documentos.

A ausência da gestão informacional, de forma corporativa, causa prejuízos e problemas para as organizações, problemas de ordem administrativa, financeira e jurídica; dificulta que a mesma possa contar sua história e apresentar sua identidade. Essa ausência da gestão de informação, atualmente é comum na maioria das organizações, afetando diretamente o dia a dia e a memória corporativa da organização.

Há vários outros problemas advindos da referida ausência de uma gestão de informação, comum nas organizações, tais como: documentos armazenados sem as mínimas condições de segurança; ausência de uma política corporativa e de procedimentos definidos para gestão documental, independente do formato; indefinição de um fluxo de informação e de protocolos para segurança do trâmite documental; inexistência de padrões preestabelecidos para organização dos documentos; ausência de pessoal especializado para o desenvolvimento de produtos e serviços de informação, ou ainda, utilização de recursos humanos sem devida qualificação para desenvolvimento das atividades; ausência de softwares adequados para apoio e implementação da



gestão documental; espaço físico inadequado para guarda dos documentos; inexistência de planos de contingência, entre outros.

De forma sucinta, independente da organização, se advinda do setor público ou privado, percebe-se que há um cenário de muito problemas informacionais, e conseqüentemente, muitas oportunidades de atuação para o “profissional de informação”, uma vez que onde há problemas, há necessidade de solução dos mesmos. E as oportunidades não findam após solução e resolução dos problemas, uma vez que a gestão informacional necessita de manutenção contínua.

## 5 CONCLUSÕES

Essa foi uma tentativa de abordar algumas experiências profissionais vivenciadas por mim, com intuito de relatar as experiências adquiridas, os desafios e medos enfrentados, diante de um mercado desprovido de soluções de gestão de informação, cheio de oportunidades, carente de profissionais inovadores, pró ativos e desbravadores.

Há oportunidades, é fato, mas isso não significa facilidade de atuação, trata-se de um mercado exigente, que demanda profissionais de alto nível. Demanda essa que serviria de justificativa para meu retorno à academia em busca do mestrado e do doutorado, uma vez que a qualificação deve ser contínua, para que se atenda com excelência as exigências do almejado e concorrido mercado.

É importante ressaltar, que a universidade capacita, qualifica e “abre portas”, no entanto, cabe a cada profissional, se desenvolver profissionalmente, e buscar qualificações extras, que julgue necessárias, de acordo com sua atuação e necessidade de aprendizado. O desafio é grande.

Conforme já mencionado, para cada projeto, há um escopo novo a ser devolvido, com suas especificidades e necessidades de acordo com o Cliente e com seus processos de negócio, o que demanda constante atualização técnica e capacidade de adaptação. Ao se iniciar um projeto de gestão informacional, é de fundamental importância conhecer bem o Cliente e entender exaustivamente o escopo do mesmo. Deve-se ainda definir três principais atores do projeto: pessoas, processos e tecnologia.

Envolver os colaboradores, essa atividade é imprescindível, pois poderá comprometer a fidelidade do projeto à perspectiva a que ela serve, além de dificultar o consenso e aplicação do mesmo. É desafiante e motivador desenvolver um projeto de gestão de informação, seja qual for seu escopo, de forma que atenda aos objetivos dos principais gestores e usuários.

**REFERÊNCIAS**

ALVES, I. et al. **Dicionário de terminologia arquivística**. Lisboa: Instituto da Biblioteca Nacional e do Livro/Organismo de Normalização Setorial para a Informação e Documentação, 1993, p. 52-53.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jul. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 11 nov. 2015.

## **Bases de dados de informações científicas e tecnológicas e a organização e representação da informação: relato de experiência**

*SCIENTIFIC AND TECHNOLOGICAL INFORMATION DATABASES AND ORGANIZATION AND REPRESENTATION OF INFORMATION: EXPERIENCE REPORT*

Alessandra Rodrigues (Embrapa Informação Tecnológica, EMBRAPA)

*rodriguesal@gmail.com, alessandra.silva@embrapa.br*

**Resumo:** A organização e a representação da informação (ORI) compreendem um amplo universo de estudos e práticas, sobretudo ao serem observadas sob a ótica do mercado de trabalho no qual os(as) profissionais da Biblioteconomia atuam. Este trabalho objetiva apresentar um breve relato sobre a relevância das representações realizadas pelos profissionais bibliotecários em bases de dados de informações científicas e tecnológicas para as práticas contemporâneas de integração de sistemas. Destaca-se que essa atuação modifica não só o alcance das práticas de ORI, mas também as responsabilidades dos profissionais de Biblioteconomia no ciclo de organização e disponibilização das informações, uma vez que a informação inserida em uma base dialoga com outras bases e serve de insumo para a geração de novas informações.

**Palavras-chave:** Organização da informação. Representação descritiva. Representação temática. Bases de dados de informações científicas e tecnológicas.

**Abstract:** Organization and representation of information comprise a broad universe of study and practice, especially to be observed from the perspective of the labor market in which the professional librarianship act. This study aims to present a brief report on the relevance of the representations made by professional librarians in scientific and technological information databases for contemporary practice systems integration. This action modifies not only the scope of ORI practices, but also the responsibilities of librarianship professionals in organizing cycle and availability of information, as the information entered into a base dialogue with other bases and serves as input to generate new information.

**Keywords:** Information organization. Descriptive representation. Thematic representation. Scientific and technological information databases.

### **1 INTRODUÇÃO**

A organização e a representação da informação (ORI) tornaram-se fundamentais em um momento de grandes mudanças sociais e materiais, no qual “a informação se tornou o principal elemento de produção. [...] A informação como elemento de poder, pode ser vista, claramente, a questão da informação para dominação e conscientização social” (ARAÚJO; DIAS, 2005, p. 111-117). Essa relevância repercute no mercado de trabalho no qual os(as) profissionais da Biblioteconomia atuam – mercado este bastante amplo e diversificado, abrangendo de instituições públicas – como bibliotecas escolares, universitárias e especializadas – a empresas privadas – do ramo editorial, da comunicação social e de provedores de internet. A forma como um universo informacional é organizado, representado e disponibilizado está fortemente relacionada ao alcance

que lhe será concedido, isto é, a ORI é antes de tudo um processo de selecionar aquilo que potencialmente será recuperado em um sistema de informação, processo esse mediado pela atuação dos profissionais da Biblioteconomia.

Este trabalho, como parte do II Seminário do Grupo de Pesquisa MHTX, objetiva apresentar um breve relato sobre a relevância das representações descritivas e temáticas realizadas pelos profissionais bibliotecários em bases de dados de informações científicas e tecnológicas para as práticas contemporâneas de integração de sistemas – será apresentada parcela da experiência do Sistema Embrapa de Bibliotecas (SEB) da Empresa Brasileira de Pesquisa Agropecuária (Embrapa). O lugar de fala do qual se parte é o de profissional bibliotecária que procura elucidar como as práticas de ORI, reconhecidas há anos na Biblioteconomia, têm permitido a expansão das possibilidades de atuação de toda a classe.

Para tanto, além desta introdução, o texto está organizado em duas seções, nas quais estão apresentadas noções sobre a ORI, sobre bases de dados de informações científicas e tecnológicas, a integração entre sistemas de informação corporativos e algumas considerações finais.

## **2 ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO**

A ação de organizar informações no âmbito da Biblioteconomia envolve processos específicos da área, como catalogação, classificação, recuperação e preservação de recursos informacionais. Esses processos são realizados no intuito de possibilitar a recuperação das informações contidas nesses recursos, de maneira que são empregadas lógicas e padrões para que se alcance excelência nos procedimentos e para que haja coerência entre os registros de bases de dados. Brascher e Café (2009) esclarecem que:

A organização da informação é, portanto, um processo que envolve a descrição física e de conteúdo dos objetos informacionais. O produto desse processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico. Alguns tipos de representação da informação são construídos por meio de linguagens elaboradas especificamente para os objetivos da OI (BRASCHER; CAFÉ, 2009, p. 5).

Esses processos diferenciam-se de acordo com o objeto com o qual trabalham, isto é, estão relacionados ao tipo (ou aos tipos) de recursos informacionais enfatizados. Assiste-se, na contemporaneidade, à ampliação do conceito de recursos informacionais, uma vez que diversas possibilidades estão contempladas como, por exemplo, os documentos digitais que, por sua vez, podem ser constituídos por textos em diferentes formatos como e-pub, PDF, Word; por dados de

pesquisa; áudios; imagens; arquivos multimídia; bem como documentos impressos (como livros, periódicos, mapas, entre outros) e documentos em suportes diversos.<sup>1</sup>

O escopo das práticas relacionadas à ORI estende-se, assim, em conformidade com a demanda que é apresentada pela sociedade ao profissional, demanda essa fortemente influenciada por um ambiente dinâmico, instável e no qual as práticas informacionais também estão relacionadas com as práticas sociais emergentes. Pensar a ORI é uma das formas de pensar como uma prática social se desenvolve e como esta repercute no âmbito da comunidade que uma unidade de informação busca atender. Dessa forma, amplia-se a relação que cada item de um sistema de informação possui com os outros que o compõem; e, também, entre a maneira como se ordena e aquela como se recupera uma informação, de forma que não é possível pensar na descrição de um item sem considerar um todo complexo a ele relacionado.

## 2.1 Bases de dados de informações bibliográficas

A Embrapa é uma empresa pública, instituída em 1972, vinculada ao Ministério da Agricultura, Pecuária e Abastecimento e “dedicada à geração de conhecimentos e soluções aplicáveis à modernização agrícola” (EMBRAPA, 2015, p. 6). Desde sua criação, a Embrapa apresenta departamentos, setores ou unidades relacionados à gestão da informação científica e tecnológica, o que possibilitou o desenvolvimento na Empresa de um sistema de bibliotecas com bases de dados integradas e serviços compartilhados – Sistema Embrapa de Bibliotecas (SEB).

No início da década de 1990, houve um grande empenho por parte dos(das) profissionais da área de Biblioteconomia em promover a automatização dos catálogos das bibliotecas, seguido, nos anos 2000, pela perspectiva de disponibilizar esses catálogos on-line para toda a sociedade – conhecidos em inglês como *Online Public Access Catalog* (OPAC). Essa tendência também se manifestou na Embrapa que, em 1991, lançou o *software* Ainfo, desenvolvido pela equipe da Embrapa Informática Agropecuária (localizada em Campinas), para a gestão dos acervos impressos e digitais das bibliotecas da Empresa.

Em seu desenvolvimento inicial, o Ainfo atendia aos requisitos de um catálogo, já que possibilitava o acesso aos metadados das publicações que estavam nele indexadas. Esse foi um grande avanço para o SEB, uma vez que otimizou a gestão dos acervos impressos das bibliotecas. Com o passar dos anos, foram implementados outros módulos no *software*, dos quais se destaca o módulo digital que possibilitou o *upload* e a disponibilização dos arquivos digitais das publicações,

---

<sup>1</sup> Existem, por exemplo, coleções de aromas em bibliotecas especializadas em culinária, coleções de amostras de solos, coleções de amostras de pedras, e assim por diante, de maneira que o profissional da área de Biblioteconomia possui um universo vasto de recursos informacionais com os quais trabalhar.

inicialmente, em nível interno e, posteriormente, a toda a sociedade, por meio dos repositórios digitais e da versão de consulta pública do Ainfo intitulada como Bases de Dados da Pesquisa Agropecuária (BDPA).

Esse foi um grande avanço para os processos de ORI na Empresa, em especial tendo-se como referência o escopo das informações científicas e tecnológicas. Cabe destacar que todo o processo de implantação e atualização do *software* Ainfo foi realizado por profissionais da área de Tecnologia da Informação, em parceria com profissionais da área de Biblioteconomia. Além disso, foram criados dois grupos, para abordar: a) questões voltadas à evolução e ao desempenho do *software* e b) questões relacionadas à padronização da descrição dos recursos informacionais no sistema.

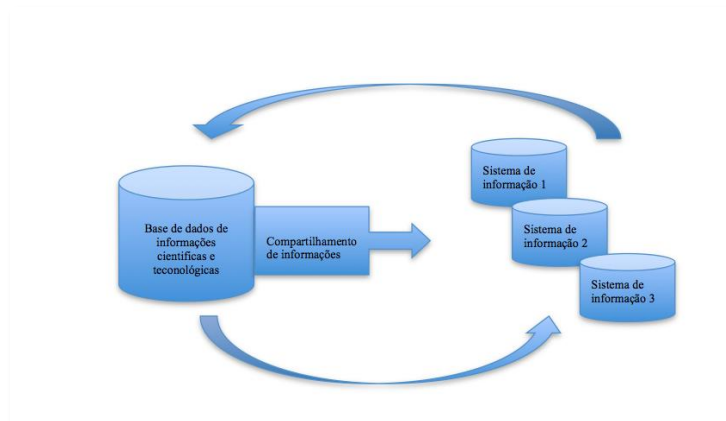
## **2.2 Integração entre sistemas de informação**

A organização das informações científicas e tecnológicas disponível no Ainfo, que inicialmente apresentou benefícios explícitos para a área de Gestão da Informação, foi também reconhecida por outras esferas institucionais. A partir desse reconhecimento, foram propostas ações para integração de sistemas de informação, de maneira que as informações fossem reaproveitadas por outras instâncias corporativas.

A catalogação na qual é feita a representação descritiva de aspectos extrínsecos dos recursos informacionais está fortemente relacionada ao motivo da existência de um sistema de recuperação da informação, ao tipo de usuário(a) que irá consultá-lo, à interface de busca e a outros fatores. Se, em tempos mais remotos, as bases de dados eram concebidas com a definição precisa dos elementos citados, hoje, cada dia mais, bases de dados são reutilizadas nas instituições para a criação e a alimentação de outros sistemas; não basta ser interoperável em nível tecnológico, é necessário também que haja integração prévia no nível das diretrizes de descrição dos recursos a serem adotados.

A imagem a seguir exemplifica uma base de dados de informações bibliográficas que inicialmente foi criada com o objetivo de organizar e recuperar informações científicas e tecnológicas publicadas e adquiridas por uma instituição de pesquisa e gerenciada pelas bibliotecas dessa instituição. Com o passar dos anos, essas informações tornaram-se de fundamental importância para a comprovação de indicadores de desempenho do corpo de empregados(as) da instituição, bem como para comprovação da finalização e entrega de projetos de pesquisa, de maneira que as informações inicialmente de cunho científico e tecnológico também vieram a se constituir como informações estratégicas para a gestão da empresa.

FIGURA 1 – Proposta de integração de bases de dados e sistemas de informação



Fonte: Desenvolvido pela autora.

Sob essa ótica, a ORI na base de dados original foi contemplada sob uma estrutura que não obrigatória e necessariamente atendia à estrutura demandada pelas outras bases com as quais veio a se integrar. Entretanto, a partir do momento em que se promoveu a integração entre esses diferentes sistemas de informação, ações relativas às diretrizes de descrição da representação dos registros passaram a ser demandadas, já que a base de dados de informações científicas e tecnológicas também passou a repercutir naquelas com informações sobre metas e projetos da instituição. Isso não significa que houve alteração nos padrões e códigos seguidos pelos profissionais da Biblioteconomia, significa que esses padrões puderam ser validados em um ambiente de informações on-line e em texto completo.

### 3 CONSIDERAÇÕES FINAIS

Conforme inicialmente apresentado, este trabalho teve como objetivo expor um breve relato de experiência sobre o processo de representação realizado por bibliotecários em bases de dados de informações científicas e tecnológicas para as práticas contemporâneas de integração de sistemas. Destaca-se que essa é uma tendência mundial, uma vez que bases de dados e sistemas de informações implicam um alto custo orçamentário de manutenção, seja em nível de recursos tecnológicos, seja em nível de recursos humanos.

Dessa forma, o trabalho de estruturação, alimentação e acompanhamento da base de dados de informações científicas e tecnológicas recebeu visibilidade, já que antes não necessariamente se reconhecia a participação ativa e crítica de profissionais da Biblioteconomia nessa ação. Essa atuação modifica não só o alcance das práticas de ORI, mas também as responsabilidades dos profissionais de Biblioteconomia no ciclo de organização e disponibilização das informações, uma

vez que a informação inserida em uma base dialoga com outras bases e serve de insumo para a geração de novas informações.

## REFERÊNCIAS

ARAÚJO, E. A.; DIAS, G. A. A atuação profissional do bibliotecário no contexto da sociedade da informação: os novos espaços de informação. In: OLIVEIRA, M. (Coord.). **Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação**. Belo Horizonte: UFMG, 2005. p. 111-122.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., São Paulo, 2009. **Anais...** São Paulo, 2009. Disponível em: <[http://saladeaula01.eci.ufmg.br:8080/xmlui/bitstream/handle/123456789/462/organizacao\\_\\_conhecimento\\_3.pdf?sequence=1](http://saladeaula01.eci.ufmg.br:8080/xmlui/bitstream/handle/123456789/462/organizacao__conhecimento_3.pdf?sequence=1)>. Acesso em: 12 maio 2016.

EMBRAPA. **VI plano diretor da Embrapa: 2014-2034**. Brasília, 2015.



## **Biblioteconomia: relato de experiência e perspectivas para o profissional da informação**

### *LIBRARIANSHIP: REPORT ABOUT EXPERIENCE AND PROSPECTS FOR INFORMATION PROFESSIONAL*

Fernanda Pereira (IFMG-Sudeste / Campus Barbacena)

*fernanda.ufmg@gmail.com*

**Resumo:** Relato sobre a experiência profissional, desde a escolha do curso de Biblioteconomia até a atual prática de mercado. Expõe pontos de vista sobre o que o mercado de trabalho espera do profissional. Também aponta o que pode ser aprimorado durante a formação dos bibliotecários. Por fim, propõe uma reflexão sobre a postura dos bibliotecários em redes sociais.

**Palavras-chave:** Bibliotecário. Mercado de trabalho do bibliotecário. Ética na profissão.

**Abstract:** Report about the professional experience, from the choice of Librarianship course to the current market practice. Exposes views on what the labor market expects from professional. It also points out what can be improved for the education of librarians. Finally, it proposes a reflection on the attitude of librarians in social networks.

**Keywords:** Librarian. Librarian labor market. Ethics in the profession.

## **1 CONSIDERAÇÕES INICIAIS**

Talvez este relato de experiência devesse começar com considerações acerca da escolha do curso, uma escolha que é muito questionada e feita com base em diferentes aspectos, como: o reconhecimento do valor do profissional para a transformação da sociedade, a possibilidade de trabalho em ambientes especializados em diversas temáticas, o amor aos livros, a curiosidade, a baixa concorrência para ingresso na universidade, oferta de vagas de emprego. Fiz a escolha pela possibilidade de poder trabalhar com pesquisas dentro de diversas temáticas e também pela grande oferta de concursos para a área.

Cursei Biblioteconomia entre 2005 e 2008, período em que pude repensar todos os porquês de ter feito essa opção e perceber que muitos dos meus colegas compartilhavam dos mesmos receios de permanecer e motivos para prosseguir até concluir a graduação. Dentre os receios de permanecer, acredito que o maior deles foi e, provavelmente, ainda é, sobre como o mercado de trabalho reconhece o Bibliotecário. Durante as aulas, eu tentava imaginar como as empresas trabalhavam com tantas teorias, como seria o dia-a-dia do profissional em uma empresa, como ele empregaria tantas regras e teorias, sendo que, para algumas delas, possivelmente, já haveria alguma tecnologia em desenvolvimento que fizesse o trabalho.

Busquei, então, entender a aplicabilidade de todos os estudos com os quais passei a ter contato. Comecei a fazer estágios, pois além de entender a aplicabilidade das teorias estudadas, poderia perceber qual era a demanda do mercado para este grupo de profissionais. Fazendo os estágios e estando em contato com pessoas que já atuavam na área, também foi possível entender o que era, de fato, esperado do profissional em Biblioteconomia.

Particpei da Empresa Júnior da ECI, fiz estágio na CEMIG, Práxis, ICEX, Biblioteca da Escola de Farmácia da UFMG, fui bolsista de iniciação à pesquisa e monitoria, grupo de pesquisa. Mesmo com toda essa vivência e contato com profissionais da área, eu não conseguia perceber como as teorias seriam aplicadas no meio profissional. Além disso, a conclusão do curso estava próxima, e o mercado não oferecia boas perspectivas. A sensação era como se tivesse cheia de teorias que não serviriam para a prática.

FIGURA 1 – Obesidade cerebral: saber muito e praticar pouco



Fonte: SILVA, 2006.

O mercado parecia desanimador: poucas vagas, salários baixos. Eu, porém, estava um pouco mais encantada com a Biblioteconomia e decidi insistir um pouco mais nas teorias e contato com a Escola, e então, prossegui nos estudos, fazendo o Mestrado. Acredito ter sido esse o melhor investimento para minha carreira, pois, durante o curso, foi possível consolidar e entender melhor tudo o que eu havia estudado durante a Graduação e finalmente perceber o sentido da grade curricular. Além disso, com a convivência com alunos mais experientes e de outras áreas, que demonstravam, em todas as discussões, a aplicação de nossas teorias em projetos e desenvolvimentos, finalmente me convenci de que fiz a escolha certa!

Finalizei o Mestrado e, em menos de dois meses, comecei a trabalhar na área de engenharia de uma grande empresa de logística do setor ferroviário, onde atuei por aproximadamente cinco anos e meio, entre 2011 e 2016. Foi um emprego bastante promissor, cercado de desafios e projetos a serem desenvolvidos. A empresa anunciou a vaga em busca de um bibliotecário, porém ela própria não conhecia bem as atribuições desse profissional, mas tinha clara sua necessidade de organizar suas informações e prover meios mais acessíveis de pesquisa, visto que se trata de um setor que trabalha constantemente com inovação e pesquisa. Mesmo com uma demanda tão latente, eu era a única bibliotecária, e o projeto para o qual fui alocada cuidava de saneamento de dados para migração de sistemas. Na prática, eram várias planilhas em Excel para conferência de dados. A insatisfação chegou logo, e imediatamente recorri à liderança da empresa e expus a situação. Nessa ocasião, esclareci minha vontade de permanecer na empresa, apresentei as demandas que percebia ao conversar com os pares e demonstrei como minha formação poderia auxiliar a equipe em vários aspectos. Este foi o momento em que, de fato, consegui um espaço para trabalhar como bibliotecária. Nos primeiros dois anos, organizei uma pequena biblioteca especializada, desde o projeto do espaço até a aquisição do acervo, conciliando o tempo com o desenvolvimento de algumas atividades de analista de gestão. À medida que meus pares foram experimentando nossa pequena biblioteca e seus produtos, a demanda cresceu rapidamente: eles buscavam referências para construção de relatórios técnicos, revisão de formatação de trabalhos, construção de modelos de documentos e orientação para construção de relatórios técnicos, aquisição de normas e periódicos, apoio geral nas pesquisas.

Os documentos produzidos começaram a ganhar maior visibilidade, por terem maior embasamento teórico e, principalmente, por serem disponibilizados de forma organizada para consulta em uma biblioteca digital acessível a todos os funcionários.

O sucesso dessa primeira etapa do projeto foi reconhecido por outros setores além da engenharia e eles também começaram a demandar pelos mesmos produtos e reconhecer outras necessidades de organização da informação.

Ao ver os documentos da engenharia organizados, as áreas começaram a se perguntar quantos documentos produziam e quão dispersos e recuperáveis esses estavam. Nesse momento, a empresa reconheceu a necessidade de um projeto de gestão de documentos mais amplo, criando uma arquitetura da informação que espelhasse a visão de negócio de cada uma de suas áreas, assim como instrumentos terminológicos para suportar a indexação de toda sua documentação. À medida que as taxonomias foram construídas e eram apresentadas as áreas refletiam sobre a estrutura de seus processos e se conheciam mais profundamente, pois os instrumentos construídos proporcionavam uma nova perspectiva de representação de suas conexões com outras áreas dentro

da estrutura da empresa. Diante desse cenário, ficou claro o potencial que o bibliotecário possui dentro de uma organização e, com isso percebi alguns *gaps* na nossa formação.

De um pequeno projeto de uma biblioteca, passei a coordenar um projeto de gestão de documentos para toda área de engenharia em que precisava gerir aspectos técnicos e administrativos. Foi mapeado as melhores práticas de mercado, recursos, iniciativas que a empresa já havia tomado para então buscar formas de canalizar todas para o mesmo fim. Todas as atividades e coordenação fizeram-me repensar alguns aspectos do curso, pois somos altamente capacitados tecnicamente, no entanto, ao assumir projetos, nós nos comprometemos com atividades de gestão e precisamos entender sobre metodologia de projeto, conhecer ferramentas, desenvolver planejamentos, apresentações executivas, pensar estrategicamente para vender internamente a ideia do projeto e defender orçamentos. É preciso ser empreendedor e ter espírito de liderança.

Nosso trabalho exige do profissional: flexibilidade, habilidade para negociação e outras competências e conhecimentos que vão além da Biblioteconomia. Acredito que o desenvolvimento dessas habilidades é tão essencial quanto a formação técnica, pois através do uso delas que conseguimos demarcar nosso espaço em projetos e apresentar todo potencial do nosso trabalho ao mercado profissional. Temos contato direto com usuários, nossos “clientes finais”, que, além de serem as pessoas a quem prestamos atendimento, muitas vezes, são nossas fontes de coletas de dados e gestores, para os quais devemos apresentar resultados e defender orçamentos para execução dos projetos.

Durante a formação, somos alertados diversas vezes para a falta de recursos das instituições, o que não traz a ideia de que seremos gestores de grandes projetos. Aprendemos muito sobre planejamento do acervo, demanda do usuário e tratamento técnico dos documentos, ou seja, ao longo do curso, ouvimos muito sobre a parte técnica da profissão. No entanto, acredito que sejam necessários também outros estímulos, especialmente os que se referem a empreendedorismo e liderança, posições que o bibliotecário precisa assumir para desenvolver projetos e gerir unidades de informação. Além disso, o profissional precisa conhecer também ferramentas e metodologias de desenvolvimento de projetos, desde Excel, *Project*, até PMBOK, PMI, para que consiga trabalhar em seus projetos em paridade com as demais áreas das instituições em que atua. Além de toda parte de gestão, não posso deixar de ressaltar a importância de conhecer um pouco a área de sistemas, pois ao propormos uma arquitetura de organização de documentos entramos em contato direto com as áreas de tecnologia para execução de desenvolvimentos, parametrizações de sistemas, suporte e utilização de *softwares* específicos, e esse conhecimento será fundamental para que seja possível utilizar as ferramentas em todo seu potencial.

Deixei a empresa no início deste ano, para assumir uma vaga como bibliotecária no Instituto Federal Minas Gerais – IFMG, *campus* Barbacena. Apesar de minha recente entrada, observo claramente algumas diferenças na dinâmica de trabalho, principalmente com relação ao atendimento direto ao público que a biblioteca oferece.

Acredito que as considerações feitas acima são pertinentes tanto para aqueles que pretendem atuar em ambientes corporativos como públicos, pois o domínio de ferramentas para gestão de projetos permitirá maior domínio dos dados e análises que ajudarão a identificar os pontos de melhoria, justificar orçamentos, evidenciar e valorizar o trabalho desenvolvido para, então, justificar e defender a existência da instituição. As competências flexibilidade, habilidade para negociação ajudarão desenvolver ideias e produtos e conseqüentemente levarão a atitudes empreendedoras. Ou seja, a interdisciplinaridade, tão falada durante o curso, é fator determinante de sucesso para o profissional.

### 1.1 Postura profissional

Nossa profissão, além de carregada por estereótipos, ainda tem pouco reconhecimento em empresas. Esse, felizmente, é um cenário que vem mudando lentamente, à medida que profissionais conseguem demonstrar resultados e ganhos nas instituições. No entanto, para que o reconhecimento aconteça de forma mais ágil, acredito que a postura dos profissionais e futuros profissionais seja determinante.

Sendo assim gostaria de chamar atenção para o comportamento dos bibliotecários em redes sociais: é comum ver profissionais debatendo sobre salários e más condições de trabalho em fóruns de grupos. Acredito que isso gera uma imagem negativa, pois vale lembrar que, hoje, as empresas consultam frequentemente as redes sociais, e, ao ver *publicações* em que os profissionais expõem as fragilidades da área cria-se um ambiente favorável para o empregador, que se sente, então, à vontade para oferecer menos e exigir mais. Afinal, se você fosse o empregador, você se interessaria em ver o que tem sido comentado nas redes sociais do candidato para fazer sua proposta, não é mesmo?

Para os futuros profissionais, acredito que devam observar o mercado desde o início do curso e acompanhar quais são as empresas que empregam mais bibliotecários e quais as competências esperadas pelo empregador. Vale observar: Qual é o salário médio e refletir sobre qual diferencial você teria a oferecer para melhorar seus rendimentos.

Até então, considere, nesta minha fala, o caminho padrão a ser percorrido: faculdade - emprego, mas, também, é preciso pensar o que você prefere: ter um emprego ou gerar empregos? E, para aqueles mais empreendedores que gostariam de ter o próprio negócio, que produto

ofereceriam? O que mais deveria aprender? Você sonha com seu diploma ou com o uso que fará dele?

FIGURA 2 – Diploma *versus* sonhos



Fonte: SILVA, 2006.

## 2 CONSIDERAÇÕES FINAIS

Acredito que as questões aqui colocadas para reflexão sejam um bom começo para que os futuros profissionais possam direcionar suas observações para o mercado em que atuarão e buscar conhecimento e desenvolvimento de habilidades que possam vir a ser essenciais em sua atuação profissional.

## REFERÊNCIA

SILVA, Flávio Augusto da. **Geração de valor**: compartilhando inspiração. Rio de Janeiro: Sextante, 2014. 208 p.

## **Consultoria técnica em gestão de documentos e serviços de informação no contexto corporativo**

### *TECHNICAL CONSULTANCY ON DOCUMENT MANAGEMENT AND INFORMATION SERVICES IN CORPORATE ENVIRONMENT*

Graciane B. Borges (ECI/UFMG)

*gracianesb@yahoo.com.br*

**Resumo:** Este trabalho aborda a Gestão de Documentos e a prestação de Serviços de Informação no contexto corporativo. Propõe-se a contribuir com a atuação do bibliotecário em novas perspectivas no mercado no campo da consultoria e da prestação de serviços em Organização e Representação da Informação. Apresenta dois vieses: o primeiro diz respeito à precária realidade das empresas brasileiras quanto à organização e gestão de seus documentos e informação, e como projetos que visam ao desenvolvimento de Arquiteturas da Informação baseadas na visão de negócio do cliente estão criando sistemas de gestão mais inteligentes e integrados, fator que vem contribuindo para a melhoria dessa realidade. Nesse contexto, aponta-se para o mercado de consultoria técnica em gestão de documentos, pouquíssimo explorado por bibliotecários. O segundo viés aborda o indivíduo dentro da empresa, bem como suas necessidades informacionais particulares e o valor real da informação em seu desempenho profissional e pessoal. Relata brevemente a experiência da autora em biblioteca corporativa especializada na indústria do petróleo.

**Palavras-chave:** Gestão de Documentos. Serviços de Informação. Consultoria. Organização da Informação. Representação da Informação.

**Abstract:** This paper approaches Document Management and provision of Information Services in the corporate environment. This paper proposes to be a contribution to the discussion of the role of librarians in new perspectives in the market in the field of consultancy and of provision of services in Organization and Representation of Information. It presents two biases: the first with respect to the precarious condition of Brazilian companies in relation to the organization and management of its documents and its information and as projects that aim at the development of Architectures of Information based on the business perspective of the client creating smarter and more integrated management systems, a factor that is contributing to the improvement of the present condition. In this context, the aim is at the technical consultancy market in documents management, very little exploited by librarians. The second bias approaches individuals within the company, as well as his/her private informational needs together with the real value of information in his/her professional and personal performance. This paper briefly reports its author's experience in corporative librarianship specialized in the oil industry.

**Keywords:** Documents management. Information Services. Consultancy. Organization of Information. Representation of Information.

## **1 CONSIDERAÇÕES INICIAIS**

O bibliotecário é o hoje um Profissional da Informação de alto valor para o mercado corporativo especializado. O poder da informação no mundo é uma realidade, e estamos vivenciando fortes mudanças em relação ao registro do conhecimento e às formas de utilização de

seus recursos. Não estamos diminuindo o volume informacional no meio corporativo, mas, sim, alterando a maneira de tratamento e manipulação. Hoje, uma empresa que não considera investimentos reais na organização e gestão de seu capital intelectual está se distanciando da tendência internacional de fluidez dos processos, integração das atividades e potencialidade na execução da atividade fim da instituição. A informação é, inegavelmente, o combustível das organizações. É o principal elemento que faz girar o motor da produção, prestação de serviços e é a garantia da retenção do conhecimento gerado diariamente pelo capital humano das grandes corporações.

A realidade das empresas brasileiras no que diz respeito à organização e gestão de seus documentos e informação é precária. É constatado um “caos informacional”, caracterizado por ineficiência ou, como na maioria dos casos, inexistência de qualquer sistemática para o fluxo documental e informacional das empresas. O movimento atual para melhoria desse quadro, embora ainda tímido no Brasil, é retratado por iniciativas de projetos que visam ao desenvolvimento de Arquiteturas da Informação que, assim como o próprio nome diz, arquitetam os sistemas de gestão de documentos para que estes apresentem melhor desempenho.

Nesse contexto, sob uma perspectiva do indivíduo dentro da empresa, temos um campo vastíssimo de trabalho na direção do atendimento desse público quanto às suas necessidades informacionais particulares. Aqui, a informação estratégica é o diferencial em seu desempenho profissional e até mesmo pessoal. Vários Serviços de Informação podem ser desenvolvidos e executados dentro das unidades ou centros de informação corporativos.

### **1.1 Consultoria técnica em gestão de documentos**

O mercado de consultoria técnica em Gestão de Documentos é muito pouco explorado por bibliotecários no Brasil. Temos experiência em projetos que visam ao desenvolvimento de sistemas de gestão por meio de Arquiteturas da Informação baseadas na visão de negócio do cliente. Tecnicamente, essas arquiteturas são elaboradas a partir de instrumentos terminológicos, tais como: Vocabulário Controlado; Taxonomia Corporativa; e Tabela de Gestão Informacional – Temporalidade.

Há de se ressaltar aqui, por vivência de mercado, que estes instrumentos, quando entregues ao cliente isoladamente, embora possam trazer algum ganho para a organização interna dos documentos, não irão significar efetivamente avanços no processo de geração, tramitação, arquivamento, destinação, recuperação e utilização da informação. Estes resultados são concretamente obtidos quando o escopo do projeto prevê a parametrização e/ou a customização de



*softwares* com essa modelagem interna. Os instrumentos, por si só, tornam-se obsoletos, sem uso efetivo, e, na maioria das vezes, são “engavetados” pelos clientes ao término do contrato.

O valor agregado do processo de Organização da Informação deve garantir ao cliente benefícios em vários aspectos: (1) reuso da informação para resolução de problemas e geração de novos conhecimentos internos; (2) resolução de problemas; (3) método sistemático de organização e armazenamento dos documentos com garantias fiscais/legais de destinação do acervo; (4) interação entre os profissionais e compartilhamento de experiências; (5) comunicação interna da empresa; (6) tomada de decisão baseada em registros seguros; (7) recuperação da informação; (8) reuso de lições aprendidas de casos de sucesso e de insucesso vivenciados anteriormente na empresa; (9) reutilização do conhecimento gerado para evitar retrabalho e desperdício de tempo e recursos; e (10) segurança da informação.

Apesar de o processo de Organização da Informação oferecer tantos benefícios, o mercado é carente de profissionais qualificados, o que torna mais difícil o atendimento à demanda das empresas. Destacamos algumas habilidades desejáveis ao profissional para agregar positivamente a este cenário: a) especialização *lato e stricto sensu*: mestrado e doutorado em Ciência da Informação – CI e especialização em Arquitetura e Gestão da Informação são aconselhados; b) formação complementar continuada: capacitação em excelência no atendimento ao cliente; idiomas; oratória e escrita são fundamentais; c) habilidades esperadas: pró-atividade; profissionalismo; disciplina; inteligência emocional; flexibilidade e facilidade em trabalho em equipe.

## 1.2 Serviços de informação em bibliotecas corporativas

A tendência internacional das grandes empresas, quanto a serviços de informação, é a expansão de bibliotecas corporativas tradicionais, que, até então, administravam os acervos técnicos e administrativos das corporações, para verdadeiros centros/unidades de informação. O empréstimo de livros e a realização de pesquisas bibliográficas passam a ser os itens básicos dessas unidades. O conceito é brutalmente ampliado para prestação de Soluções em Serviços de Informação, visando à excelência no atendimento ao cliente (antigo “usuário”). Prestar solução em serviços de informações rompe a barreira tradicional da localização de itens documentais, avançando para um Serviço de Referência criativo, especializado e que gera um valor agregado ao cliente. Este novo modelo apresenta uma vasta gama de serviços.

O Profissional da Informação que se interessar por este segmento precisa, primeiramente, ser apaixonado por *servir ao cliente*. Ter o espírito de auxílio e de ação é primordial. Vou além, acredito que seja necessária uma boa dose de contentamento pessoal com a satisfação do cliente

bem atendido. É necessário ter em mente alguns princípios: *o que trouxe meu cliente à unidade de informação da empresa? O que posso fazer para atendê-lo da melhor forma possível? Qual é o meu limite de “todo o possível”?*

Em minha experiência em bibliotecas corporativas, no segmento da indústria do petróleo, tive a rica oportunidade de trabalhar ao lado de uma equipe de Referência robusta e com uma liderança empreendedora e visionária. O espírito de trabalho era o de que “prestar serviços” é cumprir tarefas, é ser tarefeiro, e que “prestar soluções em serviços de informação” é ir além, é entender, captar e se apropriar do problema do cliente e fazer todo o possível na direção de contribuir, efetivamente, para sua resolução.

Esse ambiente me iniciou e me inspirou por toda a trajetória profissional. Hoje visualizo nichos de mercado potenciais para prestação de serviços especializados. A informação precisa, selecionada e trabalhada por profissionais de alto nível é, para mim, o conceito de informação estratégica que deve estar presente no dia a dia do profissional.

Dentre os vários serviços que podem ser oferecidos, estão: (1) Alerta de Assuntos e de Sumários de Periódicos Científicos; (2) Normalização Documentária; (3) Normalização e Organização de *Curriculum Vitae*; (4) Assessoria Técnica para Editoração de Publicações Científicas e Depósito Legal; (5) Levantamento Bibliográfico; (6) Localização de dados, informações, documentos, instituições e profissionais; (7) Monitoramento de atualização de Normas Técnicas e Legislações; (8) Orientação metodológica para Elaboração e Apresentação do Trabalho Científico; (9) Pesquisa Especializada; e destacamos a (10) Disseminação Seletiva da Informação – DSI.

O serviço de DSI pode ser considerado o principal produto de informação em um centro ou unidade de informação corporativa. Através dele, o cliente é atendido com informações específicas de maneira personalizada e contínua. Para Luhn (1961 citado por LONGO, 1978, p. 101) DSI não é um conceito novo. Sendo considerado na literatura como o precursor dos trabalhos nesta área (LONGO, 1978; SAMPAIO e MORESCHI, 1990; EIRÃO e CUNHA, 2012; AMARAL e FREIRE, 2014), Luhn (1961) define DSI como aquele serviço dentro de uma organização que se refere à canalização de novos itens de informação, vindos de quaisquer fontes, para aqueles pontos dentro da organização, onde a probabilidade de utilização, em conexão com interesses ou trabalhos carentes, é grande.

Souto (2004, p. 125) por sua vez, afirma: “a Disseminação Seletiva de Informações é um serviço que fornece ao usuário uma relação periódica de fontes informacionais, relacionadas com sua área, mediante o preenchimento de um perfil de interesse”.

Nessa perspectiva, podemos constatar um papel fundamental e estratégico da informação, que a situa no cerne das grandes organizações e precisa ser fortemente explorada por bibliotecários e profissionais da informação de maneira geral.

## 2 CONSIDERAÇÕES FINAIS

O perfil do Profissional da Informação moderno, ciente de seu próprio valor no mercado, seja ele público ou privado, passa por um movimento pessoal de desenvolvimento e investimento de autoformação e de colocação no mercado como efetivo profissional autônomo.

O trabalho interdisciplinar, comumente mencionado de forma teórica na academia, já é uma realidade no mercado para este segmento. Equipes de projetos reúnem profissionais da Ciência da Informação, da Computação e de outras áreas do conhecimento, como a Linguística e o *Design* Gráfico. Isso se dá porque são necessários um espírito colaborativo e um profundo entendimento do contexto e das necessidades do cliente para o desenvolvimento de uma Arquitetura da Informação potencialmente estruturada e fundamentada para o desenvolvimento de sistemas de gestão de documentos integrados e eficientes.

Com o olhar de minha curta atuação no mercado em um período de dez anos, onde pude atuar, inicialmente, como bibliotecária de pesquisa, bibliotecária de DSI, supervisora técnica de equipe de referência e consultora em gestão de documento e da informação, acredito e vivencio um mercado absolutamente promissor e rentável. Contudo, o caminho não é nada fácil, pelo contrário, pressupõe uma decisão profissional e até mesmo pessoal de viver uma forma de trabalho onde o retorno financeiro não está no cumprimento de uma carga horária, mas, sim, no desenvolvimento de produtos e prestação de serviços efetivos que serão validados pelo cliente final e remunerados se, e somente se, atenderem às expectativas pelas quais foram contratados.

Essa adrenalina muda, por completo, a nossa perspectiva de trabalho. Acredito e torço por uma geração de bibliotecários empreendedores, seguros e capacitados para um mercado fascinante e de possibilidades ilimitadas de lidar com dois fatores inegavelmente fundamentais para o progresso da sociedade: a informação e o conhecimento. As teorias e técnicas oriundas da Biblioteconomia e da Ciência da Informação fazem desse profissional o mais habilitado para o tratamento da informação, sua organização, recuperação e utilização no mundo contemporâneo. Todo esse potencial e inesgotável nicho de mercado pode e precisa ser fortemente explorado por nós, Profissionais da Informação.

**REFERÊNCIAS**

- AMARAL, M. R.; FREIRE, G. H. A. A disseminação seletiva de informações no contexto das organizações aprendentes e a importância do seu desenvolvimento em bibliotecas universitárias. **Revista do Mestrado Profissional Gestão em Organizações Aprendentes**, João Pessoa, v. 3, n. 2, p. 69-93, 2014. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/mpgoa/article/view/22031/12460>>. Acesso em: 10 maio. 2016.
- EIRÃO, T. G.; CUNHA, M. B. Disseminação seletiva da informação: análise da literatura publicada no período de 1958-2012. **Informação & Sociedade: estudos**, João Pessoa, v. 23, n. 1, p. 39-47, jan./abr. 2012. Disponível em: <[www.ies.ufpb.br/ojs/index.php/ies/article/download/15756/9262](http://www.ies.ufpb.br/ojs/index.php/ies/article/download/15756/9262)>. Acesso em: 29 maio. 2016.
- LONGO, R.M.J. Disseminação seletiva da informação (SDI): "estado da arte" e tendências futuras. **Revista de Biblioteconomia de Brasília**, Brasília, v. 6, n. 2, p. 101-120, jul./dez. 1978. Disponível em: <[http://www.brapci.ufpr.br/journal\\_edicao\\_artigos.php?dd0=8&dd1=1978&dd2=jul./dez.%201978&dd3=v.%206&dd4=n.%202](http://www.brapci.ufpr.br/journal_edicao_artigos.php?dd0=8&dd1=1978&dd2=jul./dez.%201978&dd3=v.%206&dd4=n.%202)>. Acesso em: 9 maio. 2016.
- LUHN, H. P. Selective dissemination of new scientific information with the aid of electronic processing equipment. **American Documentation**, v. 12, p. 131-138, Apr. 1961.
- SAMPAIO, M. I. C.; MORESCHI, E. B. P. DSI – Disseminação Seletiva da Informação: uma abordagem teórica. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 23, n. 4, p. 38-57, jan./dez. 1990. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000002801&dd1=45239>>. Acesso em: 10 maio 2015.
- SOUTO, L. F.; PORTELA, P. O. O SDI como instrumento de educação continuada: a responsabilidade das universidades no treinamento dos usuários. **Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis**, v. 8/9, p. 123, 2003/2004. Disponível em: <<http://revista.acbsc.org.br/racb/article/view/411/522>>. Acesso em: 5 maio 2016.

## **Organização e representação em bibliotecas escolares**

### *ORGANIZATION AND REPRESENTATION IN SCHOOL LIBRARY*

Kelly Cristiane Santos Moraes (Colégio Franciscano Sagrada Família)

*kellycristianesantosmoraes@gmail.com*

**Resumo:** As escolas enquanto instituições que geram conhecimento são fundamentais na formação humana e social dos sujeitos. Dentro desse contexto, as bibliotecas escolares são espaços de aprendizagem que apoiam as práticas educativas. Para desempenhar o seu papel, as bibliotecas precisam ser gerenciadas por bibliotecários que dominem as metodologias e ferramentas aprendidas na academia e que as apliquem nos processos de organização e representação dos documentos, a fim de, capacitar os usuários no uso desses espaços. Nesse trabalho, procura-se apontar algumas experiências vivenciadas pela autora no âmbito do mercado de trabalho, a frente da gestão de uma biblioteca escolar.

**Palavras-chave:** Biblioteca escolar. Organização e representação. Profissional da informação.

**Abstract:** The schools as institutions that generate knowledge are fundamental in human and social formation of the subjects. Within this context, school libraries are learning spaces that help in educational practices. To play its role, libraries need to be managed by librarians that dominate methodologies and tools learned in the university for to apply them in the processes of organization and representation of documents, to capacitate users in the use of these spaces. In this work, will be reported some experiences of the author in the labor market, in management a school library.

**Keywords:** School library. Organization and representation. Professional information.

## **1 INTRODUÇÃO**

A biblioteca escolar como um dos setores de apoio pedagógico tem um papel fundamental dentro das escolas, uma vez que seu objetivo é organizar e representar documentos para o uso e a disseminação nas práticas educativas. Além de, contribuir para o fomento das pesquisas escolares, promoção à leitura, guarda da memória da instituição, capacitação dos docentes e alunos na utilização de diferentes fontes de informação armazenados nos variados suportes.

A atuação do bibliotecário nesse tipo de biblioteca deve ser pensada dentro da missão dessas instituições, aplicando as metodologias e ferramentas aprendidas na academia para a melhoria das práticas educativas. É essencial que o profissional conheça o projeto político-pedagógico da escola e que mantenha um diálogo com a direção, coordenadores e professores para entender as necessidades e expectativas da comunidade escolar. Assim, será possível propor um planejamento que insira a biblioteca dentro do contexto da escola para que o espaço seja uma extensão da sala de aula e alcance o objetivo de apoiar as práticas educativas. Em entrevista

concedida a Torres (2015, p. 19) Christine Fontelles<sup>2</sup> diz que, “é preciso pensar a biblioteca dentro dessa missão da escola e em como essa biblioteca atua, de maneira orgânica, no planejamento educacional proposto”. Neste trabalho, procura-se apontar algumas experiências vivenciadas pela autora no âmbito do mercado de trabalho, à frente da gestão de uma biblioteca escolar.

## 2 ORGANIZAÇÃO E REPRESENTAÇÃO DE ACERVOS

O dia-a-dia da biblioteca escolar exige um profissional multifacetado que consiga administrar os recursos humanos, quando houver, e o tempo nas diversas tarefas que compete ao bibliotecário como: gestão, planejamento, organização e representação de documentos, desenvolvimento do acervo, produtos e serviços e atendimento ao usuário. Desse modo, além das habilidades esperadas de um bibliotecário como organização, representação e disseminação, está dentro das expectativas do mercado a capacidade de liderança, criatividade na solução de problemas, gerenciamento de recursos humanos, domínio de ferramentas tecnológicas, e conhecimentos de *marketing*.

Do ponto de vista da organização e representação de documentos ressalta-se a importância da elaboração de uma política de catalogação para a biblioteca que contemple as definições e parâmetros que subsidiarão os processos de catalogação. A política é importante para documentar as decisões tomadas pelo profissional em uma época específica, que poderá ser de notória importância em momentos de mudança do profissional da informação frente à unidade de informação. A ausência desse documento dificulta o trabalho do novo profissional, pois, será necessário resgatar a história na tentativa de entender, na medida do possível, qual a lógica implantada para organização e representação do acervo.

Observa-se também, que ao assumir bibliotecas já implantadas, a inexistência da aplicação de metodologias oriundas da Biblioteconomia nas tarefas de organização e representação de documentos, e a falta de atualização pelos bibliotecários sobre os novos conceitos e práticas geram problemas de recuperação, frustração e distanciamento dos usuários nas bibliotecas.

Assim, é importante que os bibliotecários dominem e apliquem as metodologias de organização e representação para que a biblioteca possa ser um lugar de descobertas, por meio de um caminho lógico que o usuário entenda e consiga localizar com facilidade o que precisa.

O distanciamento da teoria com a prática profissional coloca em risco a consistência dos sistemas documentários. É preciso uma maior interlocução entre a academia e os profissionais,

---

<sup>2</sup> Christine Fontelles é diretora de Educação e Cultura do Instituto Ecofuturo, que foi criado em 1999, como OSCIP (Organização da Sociedade Civil de Interesse Público); é também autora do livro “Espaços de leitura na escola”.

para que aconteça de fato a interação e aplicação das metodologias e ferramentas aprendidas durante a formação, e que culmine com sistemas que alcancem os objetivos a que se propõem.

A aplicação de lógicas próprias para a organização e representação dos documentos por parte dos bibliotecários, deve ser utilizado com cautela, uma vez que no futuro pode gerar problemas com a recuperação de documentos pela falta de estudos e validação que comprovem a eficiência e eficácia dessas práticas.

Desse modo, a definição e utilização de instrumentos como códigos, manuais, vocabulários controlados que orientem a prática e ajudem a padronizar os procedimentos pela equipe da biblioteca são essenciais. Portanto, recomenda-se que os estudantes de biblioteconomia se capacitem para a utilização desses instrumentos para que os procedimentos validados possam ser aplicados nos sistemas documentários.

Ao ser entrevistada por Torres (2015, p. 19), Bernadete Campello<sup>3</sup> afirma que, “a organização do acervo com base em sistemas bibliográficos universais é essencial, em primeiro lugar, para garantir a localização de itens específicos e com rapidez”. Segundo para permitir que os alunos se familiarizem com os padrões utilizados em universidades, facilitando futuramente a utilização de outras bibliotecas.

Além da aplicação de metodologias e parâmetros é importante o treinamento e capacitação dos auxiliares de biblioteca que ajudam a executar alguns processos como: carimbagem dos documentos, colocação de etiquetas, organização dos materiais nas estantes, etc. A falta de orientação a esses profissionais podem culminar em processos realizados a partir do senso comum o que compromete todo o trabalho de organização e representação do acervo.

As escolas que possuem recursos financeiros, como é o caso da rede particular de ensino investem cada vez mais em ferramentas tecnológicas para aprimorar as práticas educativas e melhorar a interlocução com a comunidade de alunos e de seus familiares. Por meio de laboratórios, biblioteca, mídias e plataformas digitais, aplicativos de acesso a informações pessoais do aluno, redes sociais colaborativas, o uso da internet tem expandido para diferentes aplicações voltadas a educação.

Inseridas nesse contexto as bibliotecas escolares também têm buscado informatizar os seus processos. “Muitas têm, atualmente, seus catálogos automatizados, empréstimo on-line, projetos de digitalização de acervos, acesso a bases de dados externas, repositórios digitais e outros serviços baseados em tecnologia”<sup>4</sup>.

---

<sup>3</sup> A profa. Bernadete Campello é uma das maiores referências em Fontes de Informação e como pesquisadora, de reconhecimento internacional, do tema da biblioteca escolar, e coordena grupo de pesquisa na Universidade Federal de Minas Gerais (UFMG).

<sup>4</sup> Trecho da entrevista de Bernadete Campello concedida a Torres (2015, p. 20).

Do ponto de vista do software para gerenciamento dos processos de gestão e organização dos documentos, é necessário que os bibliotecários tenham um mínimo de conhecimento de tecnologias que ajudem a avaliar o *software* a ser adquirido ou em funcionamento. Assim como, para operar a ferramenta para organizar e representar os documentos, e treinar a equipe e os usuários.

As mídias sociais como *Blog*, *Twitter* e *Facebook* são ferramentas interessantes para divulgar a biblioteca e os seus serviços. Porém, necessitam de atualizações diárias para manter os usuários conectados.

Por fim, sugere-se a avaliação de todos os processos para garantir a qualidade e a correção de erros ou problemas quando identificados. Analisar como os usuários comportam-se ao realizar a busca nos sistemas documentários, e a qualidade da recuperação, são elementos que ajudam a compreender e a avaliar como os processos podem ser melhorados.

### 3 CONCLUSÕES

A formação do bibliotecário deve contemplar diferentes áreas do conhecimento como: Administração, Economia, Comunicação Social, Biblioteconomia, Computação, etc. Dentre os conhecimentos dessas áreas, os bibliotecários devem ser especialistas em organizar e representar documentos armazenados em diferentes suportes. Portanto, é importante a capacitação dos profissionais no uso das novas tecnologias, porém é necessário que o uso das ferramentas seja realizado a partir de padrões e metodologias da Biblioteconomia para garantir a qualidade da organização e representação de documentos e o sucesso da recuperação pelos usuários.

### REFERÊNCIA

TORRES, F. A biblioteca do século 21. **Gestão educacional**, v. 11, n. 121, p. 16-21, jun., 2015.



## MESA-REDONDA

### AS TEORIAS DA PESQUISA INTEGRADA EM ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO

Moderadora: Benildes Coura M. S. Maculan (ECI/UFMG)

Participantes:

Profa. Dra. Gercina A. B. O. Lima (ECI/UFMG)

Profa. Dra. Maria Luiza A. Campos (UFF)

Profa. Dra. Mariângela S. L. Fujita (UNESP / Marília/SP)

Profa. Dra. Nair Y. Kobashi (ECA/USP)

### A noção de integração nas pesquisas em organização e recuperação da informação

*NOTION OF INTEGRATION IN RESEARCH ON ORGANIZATION  
AND INFORMATION RETRIEVAL*

Benildes C. M. S. Maculan (ECI/UFMG)

*benildes@gmail.com*

**Resumo:** Discorre sobre os três importantes elementos desta mesa-redonda: as teorias em Organização da Informação e do Conhecimento (OIC), as teorias em Recuperação da Informação (RI), e a integração desses elementos nas pesquisas realizadas no âmbito das áreas da Biblioteconomia e da Ciência da Informação (CI). Considera que essa integração deve levar em consideração as especificidades disciplinares, num esforço para não incorrer na desconsideração das premissas que estabelecem a ciência da área.

**Palavras-chave:** Integração disciplinares. Representação da Informação. Fundamentos e teorias.

**Abstract:** Discusses the three important elements of this round table: the theories in Information and Knowledge Organization (ICO), the theories in Information Retrieval (IR), and the integration of these elements in the research carried out in the areas of Librarianship and Information Science (CI). It considers that this integration should take into account the disciplinary specificities, in an effort not to incur in disregard of the premises that establish the science area.

**Keywords:** Disciplinary integration. Representation of Information. Fundamentals and theories.

## 1 CONSIDERAÇÕES INICIAIS

O tema desta mesa perpassa três importantes elementos: de um lado, há as teorias em Organização da Informação e do Conhecimento (OIC), e, de outro lado, as teorias em Recuperação da Informação (RI). Conectando esses dois lados, existe a integração desses elementos nas pesquisas realizadas no âmbito da Biblioteconomia e da Ciência da Informação (CI).

De maneira mais ampla, a Organização do Conhecimento (OC) tem relação com um conjunto de processos intelectuais que “visa[m] à construção de modelos de mundo que se constituem em abstrações da realidade” para compreensão de diversos fenômenos (BRASCHER; CAFÉ, 2008, p. 6). Esses processos envolvem diferentes níveis de complexidade:

(a) elementos de conhecimento, por meio dos quais entendemos as características de conceitos que podem ser obtidos com a predicação de suas propriedades ou fazendo declarações sobre seus referentes; (b) unidades de conhecimento, que equivalemos aos conceitos. Elas são a síntese das características do conceito, obtidas pelas referidas declarações sobre um referente e são representadas por um signo (palavra, nome, termo, código); (c) unidades de conhecimento ampliado, que são a combinação de conceitos, por exemplo, em declarações ou em definições ou apenas em textos; e (d) sistemas de conhecimento, que são entidades compostas de unidades de conhecimento arranjados a partir de um planejamento adequado, em uma estrutura coesa (DAHLBERG, 2006, p. 12, tradução minha).

Assim, esses quatro elementos são apreendidos pelo modelador na tarefa de criar representações de referentes do mundo, real ou não, para organizá-los de acordo com um propósito e público-alvo específicos (agrupamento, arranjo e denominação verbal), para que seja possível aos usuários “tirar conclusões úteis a partir deles” (DAHLBERG, 2006, p. 12, tradução minha).

Já a Organização da Informação (OI) tem como finalidade, basicamente, a

[...] organização de um conjunto de objetos informacionais para arranjá-los sistematicamente em coleções, [...] [como] [...] a organização da informação em bibliotecas, museus, arquivos, tanto tradicionais quanto eletrônicos. [...] A organização da informação é, portanto, um processo que envolve a descrição física e de conteúdo dos objetos informacionais (BRASCHER; CAFÉ, 2008, p. 5).

Nessa perspectiva, a OI é também uma representação e tem como objeto o documento, em seu sentido mais amplo, que pode ser representado no todo ou em partes, de forma individual ou a partir de um conjunto de documentos. E essa representação poderá englobar tanto os aspectos descritivos quanto os aspectos de conteúdo temático dos documentos (MACULAN, 2015).

Abarcando as duas concepções, entende-se a OIC como uma

[...] disciplina dedicada ao estudo e desenvolvimento dos fundamentos e técnicas do planejamento, construção, gestão, uso e avaliação de sistemas de descrição, catalogação, ordenação, classificação, armazenamento, comunicação e recuperação dos documentos criados pelo homem para testemunhar, conservar e transmitir seu saber e seus atos, a partir de seu conteúdo, com o fim de garantir sua conversão em informação capaz de gerar novo conhecimento (ESTEBAN NAVARRO; GARCIA MARCO, 1995, p. 149).

Pode-se considerar que para obter geração de novos conhecimentos é necessário haver processos adequados de comunicação e, para isso, são desenvolvidos sistemas de recuperação da informação (SRI). Um SRI possui um conjunto de rotinas de processamento de informações e registros informacionais (representações) que permitem a manipulação do conteúdo contido nesses

registros. As representações são modeladas para atendimento às necessidades de informação de uma comunidade de usuários, e o SRI possibilita o acesso e a recuperação de informações que estão armazenados no sistema. Entretanto, deve-se destacar que, segundo Lancaster (1968), o SRI não informa o usuário sobre o objeto de sua questão, mas somente é capaz de retornar (comunicar) a ele a possível existência de recursos informacionais atinentes à sua questão.

Nesse sentido, Lancaster (1968) aponta que a recuperação da informação é dependente da adequação das necessidades informacionais do usuário ao conteúdo dos documentos. Sendo assim, é preciso organizar a informação de forma que ela seja útil e possa satisfazer a uma necessidade do usuário, evitando que ele recupere itens não relevantes à sua questão. Corroborando essa afirmativa, Miranda et al. (2013) asseguram que,

[...] a recuperação da informação se refere aos sistemas e procedimentos adotados para se obter de um determinado acervo elementos informativos que satisfaçam às necessidades de informação dos usuários. Tradicionalmente, os bibliotecários realizam atividades de organização do conhecimento visando à recuperação da informação, mediante o uso de instrumentos de representação e organização do conhecimento e de recuperação da informação, como por exemplo, os esquemas de classificação, as listas de cabeçalhos de assunto, os tesouros e os catálogos de assunto - quer sejam alfabéticos ou sistemáticos. [...] O processo de recuperação da informação depende muito do tratamento e do armazenamento da informação, que em grande medida, determinam as melhores estratégias possíveis para as buscas num sistema de recuperação da informação (MIRANDA et al., 2013, p. 5).

Nota-se, assim, que todas essas questões exibem tal complexidade que torna o processo comunicacional também complexo, porque exige compreender diferentes elementos (ou partes) que, juntos, constituem um todo sistêmico. Então, fica estabelecida a necessidade de ampliar as fronteiras disciplinares visando dar conta de apreender e refletir sobre as diversas dimensões dos fenômenos investigados.

Dessa forma, a proposta dessa mesa-redonda teve como objetivo refletir sobre as questões da expansão dos limites disciplinares nas investigações, sobretudo no que diz respeito às teorias que formam o arcabouço das pesquisas em OIC e RI, sejam sobre fundamentos do campo da Biblioteconomia e da CI, sejam de disciplinas de áreas afins. Isso ficou evidente no conteúdo das apresentações dos convidados, que estão disponíveis no Anais do evento:

(1) A profa. Gercina A. B. O. Lima apresentou o texto “Da organização do conhecimento à recuperação da informação: teorias e técnicas”, que discorreu sobre a integração entre as teorias e técnicas da organização e da recuperação da informação, e sua interface com áreas tais como a Ciência da Computação, Linguística e Ciências Cognitivas, a partir dos resultados das pesquisas realizadas pelo Grupo MHTX, nesses 12 anos de estudos;

(2) A profa. Maria Luiza A. Campos apresentou o texto “Emprego da abordagem onomasiológica na elaboração de definições em ontologias de domínio: estudo preliminar”, que discorreu sobre as questões que envolvem a construção de definições em Ontologias, quando investiga “métodos e teorias capazes de subsidiar procedimentos metodológicos para elaboração de definições, como um elemento fundamental para a elaboração de modelos conceituais consistentes em ontologias”;

(3) A profa. Mariângela S. L. Fujita apresentou o texto “Organização do conhecimento na perspectiva da abordagem Cognitiva e Sociocognitiva”, que discorreu sobre as “perspectivas das abordagens Cognitiva e Sociocognitiva para se pensar a organização do conhecimento frente às exigências de práticas e modelos contemporâneos de recuperação da informação em ambientes multifacetados”;

(4) A profa. Nair Y. Kobashi apresentou o texto “Organização e recuperação da informação: aspectos comunicacionais”<sup>5</sup>.

É possível perceber que as reflexões que nortearam a mesa-redonda abarcaram as questões: Quais são as teorias, metodologias e técnicas que são utilizadas, advindas da própria área e das áreas convergentes, que dão conta das pesquisas em Organização da Informação e do Conhecimento e da Recuperação da Informação? Quais podem ser os possíveis pontos de convergência e/ou divergência entre essas distintas teorias, metodologias e técnicas? Essas questões foram refletidas a partir da noção de integração.

### **1.1 A noção de integração**

Etimologicamente, a palavra “integração” é um substantivo que tem origem no latim “integratio”, significando a ação e efeito de constituir um todo que se torna completo com as partes que lhe faltam ou que podem potencializar a sua capacidade de resultados positivos. Nesse sentido, para aprimorar as pesquisas em Organização da Informação e do Conhecimento e em Recuperação da Informação é interessante buscar teorias advindas de áreas convergentes, por meio de uma integração. Assim, é possível complementar e ampliar o arcabouço teórico que fundamentam as atividades e reflexões nesse campo de estudos.

Contudo, o uso de teorias de outra área do conhecimento implica sempre um esforço no sentido de que a sua apropriação, com o emprego de novos elementos, leve em consideração as especificidades desse campo de investigação. Dessa forma, os saberes de outra área são empregados de maneira suplementar, em uma formulação crítico-reflexiva, que visa superar a

---

<sup>5</sup> Trabalho cujo conteúdo não estava disponível para acesso até o final da elaboração deste artigo.

fragmentação que deu origem às diferentes disciplinas, sobretudo no final do século XIX (positivismo expressivo na fragmentação do saber) (POMBO, 2005). Com isso, espera-se proporcionar um diálogo entre elas, rompendo com os limites das disciplinas e, assim, incluindo conhecimentos advindos de distintas áreas para a compreensão de uma dada realidade. Certamente, a integração entre as teorias de diversas disciplinas não é capaz de “anular o poder que todo saber implica”, uma vez que tem a assimilação de outros poderes irá potencializar os que lhes são próprios (POMBO, 2005, p. 16).

Desse modo, o que está sendo denominado “integração” é a agregação de teorias de duas ou mais áreas do conhecimento em torno de um ou mais temas de pesquisa. Nessa perspectiva, Ventura e Lins (2014) afirmam que essa integração cria uma relação com a qual se percebe avanços nas ciências. Para as autoras, a transferência de teorias e métodos têm permitido a criação de novos conceitos e têm contribuído para a geração de novos conhecimentos. Dessa forma, segundo as autoras, é possível lidar com realidades, fenômenos e problemas cada vez mais complexos, fato que desafia a ciência contemporânea. Assim, preservando suas próprias premissas conceituais e teóricas, um campo do saber agrega diferentes áreas do conhecimento com a integração de metodologias e teorias que possam dar respaldo à solução de um problema ou de uma reflexão (ALVARENGA et al., 2011). A integração exige sair do conforto e da inércia disciplinar, cujo conhecimento somente replica aquilo que foi aprendido.

## **2 CONSIDERAÇÕES FINAIS**

No âmbito da Biblioteconomia e da Ciência da Informação, sob a perspectiva do tratamento dos saberes registrados (em qualquer suporte documental), representar conteúdos de informação para organizá-los significa recortar, segmentar e apresentá-la de outra maneira. São as necessidades pragmáticas, dos usuários, da instituição e do acervo, que irão determinar qual a melhor estratégia de representação, que visará atender ao fluxo informacional que se pretende. A realidade se apresenta com certa complexidade e exige conhecimentos que, muitas vezes, estão além da fronteira disciplinar.

Nesta mesa-redonda, houve a oportunidade de verificar a integração do conhecimento (teorias, métodos e técnicas) advindos de diversas disciplinas para dar suporte e contribuições às pesquisas em Organização da Informação e do Conhecimento e em Recuperação da Informação. Contudo, não se deve perder de vista as especificidades do campo ao fazer apropriações de teorias das áreas convergentes, conforme pode ser analisado a partir do teor dos trabalhos dos convidados.

**REFERÊNCIAS**

ALVARENGA, A. et al. Histórico, fundamentos filosóficos e teórico-metodológicos da interdisciplinaridade. In: PHILIPPI, A.; SILVA NETO, A. (Ed.). **Interdisciplinaridade em Ciência, Tecnologia & Inovação**. Rio de Janeiro: Manole/Capes, 2011.

CAFÉ, L.; BRASCHER, M. Organização da informação ou organização do conhecimento? ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...** São Paulo: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 2008. p. 1-14. Disponível em: <2008.http://saladeaula01.eci.ufmg.br:8080/xmlui/bitstream/handle/123456789/462/organizacao\_\_conhecimento\_3.pdf?sequence=1>. Acesso em: 12 maio 2016.

LANCASTER, F. W. **Information retrieval systems**. New York: John Wiley, 1968.

MACULAN, B. C. M. S. **Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do THESAGRO**. 2015. 345 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

MIRANDA, M. L. C. et al. Ensino, Pesquisa e Extensão em Organização do Conhecimento: uma proposta integrada para a formação de alunos de biblioteconomia por meio da educação tutorial na UNIRIO. In: XXV CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 25., Florianópolis, SC, Brasil, 7 a 10 de julho de 2013. **Anais...** Florianópolis: CBBD, 2013.

ESTEBAN NAVARRO, M. A.; GARCÍA MARCO, F. J. Las primeras jornadas sobre organización del conocimiento: organización del conocimiento e información científica. **Scire**, v. 1, n. 1, p. 149-157, ene./jun. 1995.

POMBO, O. Interdisciplinaridade e integração dos saberes. **Liinc em Revista**, v. 1, n. 1, p. 3-15, 2005.

VENTURA, D. F. L.; LINS, M. A. D. T. A integração entre disciplinas no âmbito da pós-graduação e da pesquisa. **Cadernos de Pesquisa**, v. 44, n. 151, p. 104-131, jan./mar. 2014.

## Da organização do conhecimento à recuperação da informação: teorias e técnicas

### *FROM KNOWLEDGE ORGANIZATION TO INFORMATION RETRIEVAL: THEORIES AND TECHNIQUES*

Gercina A. B. O. Lima (ECI/UFMG)

*limagercina@gmail.com*

**Resumo:** Os novos conhecimentos adquiridos sofrem alterações, e é necessário lapidá-los na sua concepção para que sejam compreendidos e armazenados, e, posteriormente registrados e disseminados na sua concepção. A utilização de teorias e técnicas da área da Organização e do Conhecimento auxiliam nesses processos tornando a recuperação da informação registrada mais eficiência e eficaz. O surgimento de novas tecnologias, principalmente a internet e a Web, demandaram avanços nessa área do conhecimento que as bases teórico-metodológicas da área não conseguiram acompanhar. A área de Organização da Informação e do Conhecimento busca possíveis soluções através de aportes teóricos em áreas correlatas que contemplem a dimensão interdisciplinar desse campo de estudo, sem perder seu foco principal, que é tratar a informação para ter uma recuperação eficiente.

**Palavras-chave:** Organização do Conhecimento. Organização da Informação. Recuperação da Informação. Metodologias. Tecnologia da Informação.

**Abstract:** The new knowledge acquired has to go through changes, and it is required to polish it from its own conception, so it can be understood, stored, and, subsequently registered and disseminated in its original idea. The use of theories and techniques from the Organization and Knowledge area assist in these processes making the registered information retrieval more efficient and effective. The appearance of new technologies, especially the Internet and the Web, have demanded advances in this area of knowledge that the theoretical and methodological bases of the same field have failed following it. The area of Information and knowledge Organization searches for possible solutions through theoretical contributions in related areas that include the interdisciplinary dimension of this field of study without losing its main focus, which is to treat the information, in order to have an efficient retrieval.

**Keywords:** Knowledge Organization. Information Organization. Information Retrieval. Methodologies. Information Technology.

## 1 INTRODUÇÃO: ANTECEDENTES

O Conhecimento é algo dinâmico, que está presente na vida do ser humano desde o seu nascimento e o acompanha até o final de sua vida, e, no percorrer desse caminho, ele adquire novos conhecimentos. Esses novos conhecimentos adquiridos sofrem alterações, e é necessário lapidá-los na sua concepção para que sejam compreendidos e armazenados, e, posteriormente, registrados e disseminados dentro de uma comunidade. Esse processo é realizado pela mente humana e às vezes ocorre a partir do modo como o indivíduo vê seu mundo e o representa, o que se dá por meio de informações sobre objetos pertencentes à sua realidade. Porém, a realidade não

pode ser vista sob um único aspecto, advindo da percepção daquela realidade, geralmente adquirida por meio de percepções e sensações do indivíduo. Surge aí a necessidade de criar mecanismos de padronização desse conhecimento, tornando-o passível de ser posteriormente compartilhado consensualmente entre os indivíduos.

Um dos primeiros estudiosos a pensar o mundo e a categorizá-lo foi o filósofo Aristóteles há mais de dois milênios, com sua inquietude de saber como o homem poderia representar o seu conhecimento. A teoria clássica de Aristóteles é considerada por alguns como a hierarquia perfeita, na qual as categorias são definidas apenas pelas propriedades comuns a todos os seus membros. Destarte, veio da Filosofia, através dos estudos de Aristóteles, a primeira preocupação em sistematizar o conhecimento, com o intuito de torná-lo padronizado por meio da representação das ideias por símbolos que representassem a realidade visível pelo homem. Dessa maneira, tornar-se-ia possível a manifestação de tais ideias, por meio de um significado convencional, unindo as ideias e as coisas, ou seja, estariam integradas entre elas de forma que pudessem materializar o pensamento humano, dando aos objetos características e atributos que os representassem. Assim, Aristóteles em seu *Organón* (ARISTÓTELES, 2010), sistematiza o saber humano propondo um sistema filosófico representado por dez categorias, a partir de proposições que referem à determinação formal do ser, a saber: Substância, Quantidade, Qualidade, Relação, Duração, Lugar, Ação, Paixão ou Sofrimento, Maneira de ser, e Posição. Outras contribuições deixadas por Aristóteles foram as quatro distinções entre o gênero, o próprio, o acidente e a definição (MORUJÃO, 2006, p.304) e o esquema tripartitivo de classificação, no qual o filósofo dividiu o conhecimento em ciências teóricas, ciências práticas e ciências poéticas, representando o pensar, o agir e produzir, respectivamente (PIEIDADE, 1983). Esses estudos de Aristóteles influenciaram o desenvolvimento dos estudos sobre a classificação do conhecimento e são considerados o embasamento teórico para as classificações bibliográficas que começaram a surgir posteriormente, no século IV. Não é objetivo deste trabalho aprofundar neste assunto, mas devem-se mencionar aqui os estudos de Porfírio e Ranganathan. Porfírio, em sua obra *Isogoge* (Introdução às Categorias), incorporou a lógica aristotélica na categoria Substância, considerada como gênero supremo, processo que resultou em cinco predicáveis<sup>6</sup> – gênero, espécie, diferença, propriedade, acidente –, como forma de relacionar com a Substância. Quando interligados, esses predicáveis formam uma lógica sucessiva de subordinação, resultando em uma classificação dicotômica, que se tornou conhecida como “Árvore de Porfírio”. Os trabalhos de Ranganathan, publicados a partir

---

<sup>6</sup> Porfírio faz distinção entre *predicados* e *predicáveis*. Por exemplo: na expressão «todos os homens são mortais», «mortais» é um predicado, mas é atribuído a «todos os homens» de modo diferente do que o seria, por exemplo, o predicado «bípedes sem penas». O predicável será, então, o modo como certo predicado é atribuído a um sujeito (MORUJÃO, p. 304).



dos anos 30, apresentam a proposta analítico-sintética e as cinco categorias fundamentais, o que representou um marco na história da organização e representação do conhecimento através das classificações bibliográficas.

## 2 ORGANIZAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO

A área de Organização do Conhecimento (OC) tem suas raízes históricas na base teórica da Classificação, sendo que, às vezes, a Classificação pode ser vista como uma área de conhecimento, a qual classifica o conhecimento filosófico, que se inicia com o modelo categorial proposto por Aristóteles; e, às vezes, como classificação bibliográfica, que possui processos e instrumentos para representar conteúdo temático de documentos e aumentar a eficácia na recuperação de informações. As duas abordagens da classificação procuram ligar teoria e prática nos estudos e pesquisas realizadas dentro da área da Biblioteconomia e Ciência da Informação (BCI).

Entretanto, a Organização do Conhecimento, enquanto área de estudo, foi estabelecida através de estudos realizados por Dahlberg (1993) e os membros da *International Society of Knowledge Organization* (ISKO), a partir de 1973, quando os membros da dessa sociedade decidem reconhecer a OC enquanto área, sugerindo que o conceito “classificação” passasse a ser interpretado como um método de classificação que agrupa objetos semelhantes.

Porém, o termo “organização do conhecimento” foi utilizado em estudos anteriores de Henry Evelyn Bliss, *The Knowledge Organization and the system of the science*, e *The Organization of Knowledge in Libraries*, publicados, respectivamente, em 1926 e 1936, e na tese de Dagobert Soergel, intitulada *Organization of knowledge and documentation* (1971), defendida na Alemanha, e na tese da própria autora, que foi publicada sob o título *Foundation of Universal Organization of Knowledge*.

Além desse seu estudo, Dahlberg cria o periódico *Knowledge Organization*, em 1973, sob a responsabilidade da ISKO, tornando-se sua primeira presidente. A partir de 1993, este periódico veio a substituir o periódico *International Classification*, que foi publicado de 1974 a 1992, firmando assim, a Organização do Conhecimento como uma área. A partir daí, sob a influência de Dahlberg, o conceito “classificação” passa ser interpretado como um método de classificação, que agrupa objetos semelhantes.

Nota-se, portanto, que as áreas Classificação e Organização do Conhecimento são muito próximas, tanto no embasamento teórico, quanto em suas atividades práticas. Hagar Gomes (2009) salienta que a própria literatura ainda confunde as bases teóricas de cada área, e não tem definido claramente quais atividades estão inseridas somente em uma delas, ou se estão todas relacionadas, e afirma que a área de OC tem múltiplas aplicações. Sendo que as pesquisas apontam tanto para

uma abordagem teórico-filosóficas voltadas para questões epistemológicas ao representar um domínio, quanto para estudos sobre as aplicações dos Sistemas de Organização do Conhecimento (GOMES, 2009, p. 63).

A diversidade conceitual também ocorre em relação à Organização do Conhecimento (OC) e Organização da Informação (OI). Dahlberg define Organização do Conhecimento como “a ciência que estrutura e organiza sistematicamente unidades do conhecimento (conceitos) segundo seus elementos de conhecimento (características) inerentes e a aplicação desses conceitos e classes de conceitos ordenados a objetos/assuntos” (DAHLBERG, 2006, p. 12).

Já para Hjørland (2008), a Organização do Conhecimento, dentro da Biblioteconomia e Ciência da Informação (BCI), é vista, em seu sentido mais restrito, como a organização da informação de itens bibliográficos para sua recuperação. O autor salienta ainda que o campo da BCI nunca prestou muita atenção na Organização do Conhecimento em seus construtos teóricos. Esse fato pode dificultar o avanço desse campo, pois, para se ter princípios de organização, como é vista dentro da BCI, deve-se partir, primeiramente, do entendimento do que é Organização do Conhecimento em seu sentido mais amplo. Indo além, Hjørland (2003, p. 88) questiona sobre o desenvolvimento teórico da área da OC, criticando o fato de, após os estudos realizados por Dewey (1876), terem surgido, apenas alguns padrões e alguns desenvolvimentos teóricos, tais como as Regras de Cutter (1876; 1904), e as contribuições de Ranganathan (1937). Porém, para seu desenvolvimento, a área de OC precisa unir esforços para estudar as teorias que embasam o entendimento mais claro do que é o conceito, os critérios de inclusão de classes, o significado, indexação, as relações semânticas assunto, e os pontos de acesso de assunto.

Por um lado, Dahlberg (2006) apresenta a Organização do Conhecimento como uma ciência que analisa, estrutura e sistematiza os conceitos de acordo com suas características, para representar um determinado domínio, ou seja, a modelagem de um domínio para sua representação através de sistemas de organização do conhecimento, tais como as taxonomias, tesouros e ontologias. Por outro lado, Hjørland (2003) aponta para a organização da informação como um processo de catalogação, classificação e indexação de documentos.

Bräscher e Café (2010) apresentam um estudo sobre a utilização da OC e OI dentro da área da Biblioteconomia e Ciência da Informação (CI). As autoras concluem que tal maneira de se utilizar os termos demonstra a falta de delimitação do conceito em questão, e afirmam que “por vezes, o termo organização do conhecimento é utilizado no sentido de organização da informação, e vice-versa e, em determinadas situações, empregam-se os termos conjuntamente, organização da informação e do conhecimento” (BRÄSCHER; CAFÉ, 2008, p. 1).

Na visão das autoras, existem dois tipos de processos de organização: a organização do conhecimento, que visa à criação de modelos que representam as unidades de conhecimento de um determinado domínio, de acordo com suas características, e a organização da informação, que tem o objetivo de descrever as características físicas e conceituais dos objetos informacionais, atividades que são realizadas em bibliotecas e centros de documentação (BRÄSCHER; CAFÉ, 2010, p. 93).

Assim, pode-se dizer que essas duas abordagens se complementam, pois, enquanto na área de OC organiza-se o conhecimento a partir da modelagem de domínios, representando sua estrutura com base na descrição de conceitos e dos relacionamentos semânticos entre eles, na área de OI ocorre o processo de representação do conteúdo de um documento específico, o qual tem o conhecimento registrado, para que ele seja posteriormente recuperado pelo usuário em um Sistema de Recuperação da Informação (SRI).

Portanto, verifica-se que esses diferentes pontos de vista existentes, a falta de entendimento das bases teóricas e a inexistência de uma terminologia unificada que possam mediar esses diferentes olhares dificultam a consolidação da Organização do Conhecimento como área do conhecimento.

## **2.1 As teorias e técnicas para Organização do Conhecimento e da Informação**

Os principais fundamentos teórico-metodológicos utilizados pelos classificacionistas para modelagem e representação de um domínio do conhecimento são a Teoria da Análise Facetada (TAF) criada por Ranganathan, em 1937, e a Teoria do Conceito, criada por Dahlberg (1978a), que trazem forte influência da lógica aristotélica para a representação do conhecimento.

A teoria da análise facetada (TAF) foi criada por Shiyali Ramamrita Ranganathan (1892-1972), matemático indiano que se tornou bibliotecário depois de constatar que a maioria dos assuntos tratados em cinco periódicos diferentes era formada de assuntos compostos e, também, pela sua insatisfação com os sistemas de Classificação Decimal de Dewey (CDD) e Classificação Decimal Universal (CDU).

A origem da classificação facetada se deu com os estudos de Ranganathan e com a Classificação de Dois Pontos (*Colon Classification*), publicado em 1933, que mostrava que, em uma área do conhecimento, o número de assuntos específicos que podem ser selecionados em uma classificação é infinito. Entre dois pontos, em um sistema, pode ser inserido um número infinito de novos dois pontos. Ranganathan considerou o conhecimento como um elemento multidimensional: as interligações de cada conceito espalham-se em muitas direções e, frequentemente, cada assunto constitui-se de uma síntese de vários conceitos múltiplos ligados, o que caracteriza essa abordagem

como analítico-sintética. Além disso, Ranganathan explicou os caminhos da divisão, abrangendo a dicotomia difundida por Kant e pela Árvore de Porfírio.

Examinando o método da dicotomia (com dez divisões), Ranganathan concluiu que, em vista do crescimento prolífico e multidimensional do conhecimento ao longo da história, esse método não seria satisfatório. Sugeriu, então, um padrão para mapear o conhecimento com a policotomia ilimitada (número ilimitado de divisões das áreas do conhecimento). Ao discutir a Árvore de Porfírio, com seu eixo único, o autor buscou uma alternativa na Árvore Baniana (*Banyan Tree*) para explicar de forma aceitável a sua proposta de que, em adição a um tronco original, vários troncos secundários podem ser formados de tempos em tempos, ideia que associou a uma árvore de assuntos (RANGHANATAN, 1967, p. 368).

Ranganathan descreve *assunto* como “corpo de ideias organizadas ou sistematizadas, cuja extensão e intensão devem ser coerentemente com o domínio de interesse e confortavelmente ajustadas à competência intelectual e campo especializado de um indivíduo” (RANGHANATAN, 1967, p. 82). Todo assunto é oriundo de um assunto básico somente, e seu componente vem de uma ou mais ideias isoladas, formando, assim, o assunto composto. O assunto básico é um assunto sem nenhuma ideia isolada como componente. A ideia isolada (isolado) é alguma ideia, ou complexo de ideias, moldada para formar um componente de assunto. Mas, sem seu contexto, ela não é considerada um assunto. A ordem de citação em que todo e qualquer assunto é dividido em cinco categorias: Personalidade/Entidade (*Personality/Entity*), Matéria (*Matter*), Energia (*Energy*), Espaço (*Space*), Tempo (*Time*), mnemonicamente conhecidas como PMEST, continua sendo considerada uma das grandes contribuições de Ranganathan para esse campo de estudo.

Na década de 70, autora alemã Ingetraut Dahlberg apresenta a teoria do conceito (1978a), que tem por base os princípios aristotélicos, os princípios analítico-sintéticos ranganathianos, além de buscar aportes teóricos na Teoria Geral da Terminologia (TGT) de Wüster (1998).

Para Dahlberg (1978a), a ideia de *conceito* está ligada às características dos objetos. Dessa forma, considera *conceito* como uma série de enunciados (características) verdadeiros sobre um objeto, ou seja, para a autora o conceito é uma unidade do conhecimento. Na elaboração de um conceito são colocados elementos que se articulam numa unidade estruturada. Para que essa unidade seja, de fato, estruturada, a autora coloca a necessidade de análise e síntese dos enunciados verdadeiros de um conceito, que podem apresentar-se sob a forma de característica ou categoria (hierarquia de características) (DAHLBERG, 1978a, p. 102). Para a construção do conceito, Dahlberg (1978b) propõe um modelo, no qual esta unidade do conhecimento é elaborada atendendo estas afirmações verdadeiras: (A) sobre um objeto específico (B) representado por signo linguístico (C) com fins de comunicação. Assim, a autora sugere a forma de um triângulo para

representar estes componentes onde o vértice superior é o objeto, o vértice esquerdo é o significado, e o vértice direito é o signo linguístico (termo).

Enquanto, para Ranganathan, a ideia/conceito é o produto do pensamento, salientando a característica intuitiva e dedutiva da Teoria da Análise facetada (TAF), Dahlberg aponta que o conceito é uma unidade do conhecimento, evidenciando os aportes teóricos da Teoria da Terminologia utilizados na construção do conceito a partir de seus atributos e sua representação dentro de um campo do conhecimento.

Assim como Aristóteles e Ranganathan, Dahlberg apresenta as categorias que permitem representar domínio específico, através da possibilidade de listar todas as características possíveis de um conceito, tendo por base as categorias aristotélicas, a saber: Matéria (substância), Qualidade, Quantidade (extensão), Relação, Processo (atividade), Modo de ser, Passividade, Posição, Localização (lugar), e Tempo.

Nas sociedades orais, não existia nenhum modo sistematizado de armazenar as representações para futura reutilização nesse tipo de comunicação, pois não havia a escrita. A escrita foi um dos mais importantes desenvolvimentos técnicos do ser humano. Assim como a fala foi o principal instrumento utilizado no tempo da oralidade primária, diversos tipos de sistemas de sinais gráficos, incluindo o alfabeto, tornaram-se os instrumentos principais da escrita. Segundo McGarry (1999, p. 72), o alfabeto tem sido considerado por alguns autores como uma das maiores invenções do homem. A escrita veio permitir a atualização do conhecimento acumulado, de fatos presenciados ou relatos de pessoas que viveram em épocas ou lugares diferentes (DIAS, 1999, p. 267). A comunicação escrita e o modo de transmissão dos textos sofreram profundas mudanças com o advento da imprensa. Para McGarry (1999, p. 73), a escrita em suportes artesanais foi a tecnologia de comunicação mais avançada, desde o quarto milênio a.C. até a invenção da imprensa e tipografia na Mogúncia (c.1450), quando Johann Gutenberg compôs, com tipos móveis, o texto da Bíblia, o primeiro livro a passar do manuscrito para o papel impresso. A partir da revolução técnico-científica posterior à Segunda Guerra Mundial, o grande volume de informações gerado no crescente número de áreas do conhecimento passou a demandar um nível maior de organização informacional, pois a informação deve ser ordenada, estruturada ou fixada materialmente, tornando-se um documento, senão permanecerá amorfa e inutilizável.

Pode-se afirmar que os avanços que ocorreram desde a década de 1950 até os dias atuais foram relevantes e marcou o desenvolvimento tanto na forma de armazenamento e representação quanto na recuperação da informação. Concentram-se esforços estudos sobre no processo da análise de assunto buscando aportes teóricos na semântica, na linguística documentária, na lógica e na cognição para entender o processo do *aboutness*, para fazer representação mais fidedigna das

ideias dos autores dos documentos. A indexação automática surge como alternativa, porém encontra obstáculos na capacidade da máquina reproduzir o pensamento humano no momento da representação da informação, dentro de um contexto específico. No caso da catalogação, observam-se mais progressos na pesquisa, por meio de metadados estruturados, descritivos e conceituais. O padrão *Marc* e o *Dublin Core* surgem com possibilidades para auxiliar a interoperabilidade entre dados e sistemas de informação.

O armazenamento da informação, antes feito somente nos discos rígidos dos computadores com grande capacidade de processamento, agora se dá de maneira diversa, de modo que há a possibilidade de se acessarem arquivos, dados e aplicativos em qualquer lugar e a qualquer hora, utilizando-se tanto um computador como dispositivos móveis, desde que haja conexão com a Internet, por meio do armazenamento nas chamadas “nuvens”.

As necessidades dos usuários passam a ser o ponto central de estudos na área, já que a recuperação da informação é o principal foco de toda a área da Organização da Informação e do Conhecimento. Muitas dessas pesquisas têm sido relacionadas com a interação entre o sistema de RI e o usuário, com o intuito de melhorar a eficiência dos sistemas de recuperação de informação e os métodos de catalogação, de classificação, e de indexação de assunto de um documento. Com o crescimento e desenvolvimento da Internet cada vez mais acelerados, surgem novos interesses de pesquisa em tecnologias da *Web Semântica*, da *Web 2.0*, da *Web 3.0* com uma participação cada vez maior dos usuários, que fornecem metadados e contribuem para a marcação social, mas exigem criações de interfaces mais intuitivas.

### **3 REFLEXÕES: INTEGRANDO OS SABERES PARA MELHORAR A RECUPERAÇÃO DA INFORMAÇÃO**

A Organização da Informação e do Conhecimento é uma área interdisciplinar, com aspectos teóricos e práticos, que busca aportes teórico-metodológicos em outras áreas, tais como Filosofia, Ciência da Computação, Linguística, Terminologia, Ciências Cognitivas, Lógica, Comunicação, Sociologia. A necessidade de buscar referencial teórico e metodológico em outras áreas do conhecimento tornou-se evidente na medida em que as bases teórico-metodológicas da área não conseguiram acompanhar os avanços tecnológicos, tampouco a avalanche da produção bibliográfica que ocorreu após a revolução industrial; nem conseguiu resolver problemas na organização do conhecimento/informação para melhorar sua recuperação, tais como a melhoria na qualidade dos processos de indexação e na catalogação descritiva. Além disso, observou-se a necessidade de melhorar a indexação automática, por meio do desenvolvimento de técnicas de extração mais eficientes, utilizando-se de métodos da linguística, como, por exemplo, o

processamento em linguagem natural. Também se verificou a necessidade de melhorar os relacionamentos semânticos nos sistemas de organização do conhecimento, para auxiliar a adaptabilidade e a hospitalidade, principalmente para domínios específicos, e facilitar sua conversão para uma representação mais formal, o que pode auxiliar na interoperabilidade entre diferentes vocabulários e sistemas disponíveis via *Web* (LIMA, 2015, p. 684).

Nota-se que se carece urgentemente de reflexão sobre os aspectos teóricos e metodológicos que contemplem a dimensão interdisciplinar desse campo de estudo, sem perder seu foco principal, que é tratar a informação para ter uma recuperação eficiente. Para isso, a área de Organização da Informação e do Conhecimento tem necessidade de estudos para desenvolver uma base teórica sólida para dar suporte às suas atividades, fazendo a ponte entre a teoria e a prática, com foco na recuperação da informação, que é seu principal objetivo. Verifica-se, portanto, que a área de Organização da Informação e do Conhecimento busca possíveis soluções através de aportes teóricos em áreas correlatas, conforme foi previsto na concepção teórica de Ranganathan sobre o dinamismo inerente das áreas do conhecimento, quando faz uma analogia com a figura da árvore baniana, na qual sugere que o conhecimento é inter-relacionado, e as disciplinas se integram infinitamente.

Finalmente, destaca-se que este trabalho não teve a pretensão de aprofundar na temática. Aqui se buscou discorrer brevemente sobre o pensar o conhecimento, bem como os modelos e as teorias de sua representação. Esta reflexão pode contribuir para a melhor transformação do conhecimento em registros de informação, o que possibilitará, cada vez mais, que possamos recuperá-la e utilizá-la, fazendo circular o conhecimento que o ser humano é capaz de produzir, transformando-o em registros de informação, possibilitando desta forma a sua recuperação.

## REFERÊNCIAS

ARISTÓTELES. **Órganon**. São Paulo: Edipro, 2010.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo, **Anais...** São Paulo: ANCIB, 2008. Disponível em: <<http://www.enancib2008.com.br>>. Acesso em: 17 jun. 2016.

BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento?. In: LARA, Marilda L.G.; SMIT, Johanna. (Org.). *Temas de pesquisa em Ciência da Informação no Brasil*. São Paulo: Escola de Comunicação e Artes/USP, 2010.

CUTTER, C. A. **Rules for a dictionary catalog**. Washington, D.C.: UNT Libraries Digital Collections, 1876. 274p.

- CUTTER, C. A. **Rules for a dictionary catalog**. 4<sup>th</sup> ed. Washington, DC: Government Printing Office, 1904.
- DAHLBERG, I. Teoria do conceito. Tradução Astério Tavares Campos. **Ci. Inf.**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978a.
- DAHLBERG, I. A referent-oriented, analytical concept theory of Interconcept. **International Classification**, [S.l.], v. 5, n. 3, p. 122-151, 1978b.
- DAHLBERG, I. Knowledge organization: its scopes and possibilities. **Knowledge Organization**, Frankfurt, v.20, n. 4, p.211-222, 1993.
- DAHLBERG, I. Knowledge organization: a new science? **Knowledge Organization**, Frankfurt, v. 33, n. 1, p. 11-19, 2006.
- DEWEY, Melvil. **Dewey Decimal Classification and relative index**. 22. ed. Dublin, Ohio: OCLC, 2003. 4 v. Disponível em: <<http://www.oclc.org/dewey.en.html>>. Acesso em: 26 maio 2016.
- DIAS, C. A. Hipertexto: evolução histórica e efeitos sociais. **Ci. Inf.** Brasília, v. 28, n. 3 set./dez. 1999; p. 267-275
- GOMES, H. E. Tendências da pesquisa em organização do conhecimento. **Tendências da Pesquisa Brasileira em Ciência da Informação**, [S.l.], v. 2, n. 1. p.60-88. 2009. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/issue/view/2/showToc>>. Acesso em: 1 maio. 2016.
- HJORLAND, B. Fundamentals of knowledge organization. **Knowledge Organization**, [S.l.], v. 30, n. 2, p. 87-111, 2003.
- HJORLAND, B. What is Knowledge Organization (KO)? **Knowledge Organization**, [S.l.], v. 35, n. 2-3, p. 86-101, 2008.
- LIMA, G. A. Organização do conhecimento: pesquisa e desenvolvimento. In: ISKO, Marília, SP, 17, 18 e 19 de setembro de 2015, Faculdade de Filosofia e Ciências da UNESP. **Anais eletrônicos...** Marília: FUNDEPE, 2015. Disponível em: <<https://www.marilia.unesp.br/Home/Extensao/CEDHUM/livro-isko-brasil-finalizado.pdf>>. Acesso em: 22 maio 2016.
- McGARRY, K. **O contexto dinâmico da informação**: uma análise introdutória. Tradução de Helena Vilar de Lemos. Brasília, DF: Briquet de Lemos/Livros, 1999. 206 p.
- MORUJÃO, C. A lógica modernorum: lógica e filosofia da linguagem na escolástica dos séculos XIII E XIV. **Revista Filosofia de Coimbra**, n.28, p.301-322, 2006. Disponível em <[http://www.uc.pt/fluc/dfci/publicacoes/a\\_logica\\_modernorum](http://www.uc.pt/fluc/dfci/publicacoes/a_logica_modernorum)>. Acesso em: 18 de jun. 2016
- PIEPADE, M.A. **Introdução à teoria da classificação**. 2.ed. Rio de Janeiro: Interciência, 1983.
- RANGANATHAN, S. R. **Prolegomena to library classification**. Bombay: Asia Publishing House, 1967.
- SOERGEL, D. **Dokumentation und organisation des wissens**. Versuch einer methodischen und theoretischen Grundlegung am Beispiel der Sozialwissenschaften. (Documentation and organization of knowledge. An inquiry into the methodological and theoretical foundations with



particular reference to the Social Sciences.). Berlin: Duncker und Humblot; 1971. 380p. (Ordo Politicus 13) (Doctoral dissertation).

**WÜSTER, E. Introducción a la teoría general de la terminología y a la lexicografía terminológica.** Barcelona: IULA, 1998.

## **Emprego da abordagem onomasiológica na elaboração de definições em ontologias de domínio: estudo preliminar**

### *USE OF THE ONOMASIOLOGICAL APPROACH TO THE DEVELOPMENT OF DEFINITIONS TO DOMAIN ONTOLOGIES: PRELIMINARY STUDY*

Maria Luiza de Almeida Campos (Universidade Federal Fluminense)

*maria.almeida@pq.cnpq.br*

**Resumo:** Este trabalho em andamento é uma continuidade de atividades de pesquisa que iniciamos a partir de estudos interdisciplinares junto às áreas de Ciência da Informação, Ciência da Computação, Inteligência Artificial, Terminologia. Investigamos métodos e teorias capazes de subsidiar procedimentos metodológicos para elaboração de definições, como um elemento fundamental para a elaboração de modelos conceituais consistentes em ontologias. Analisamos a adequação do método onomasiológico para a descrição das propriedades dos referentes na elaboração de Ontologias de Domínio, na expectativa de identificar quais elementos ônticos (funcionais) podem ser agregados aos elementos lógicos (relações de gênero-espécie) para a compreensão do posicionamento do conceito no contexto em questão. Pretende-se assim, explorar as definições, não como produto final da fixação do conceito em um domínio, mas como elemento primeiro visando à estruturação do conceito no interior de um domínio a partir de um dado compromisso assumido.

**Palavras-chave:** Definição. Abordagem Onomasiológica. Ontologia de Domínio.

**Abstract:** This work in progress is the sequel of research activities started from interdisciplinary studies in areas such as Information Science, Computer Science, Artificial Intelligence, and Terminology. We investigated methods and theories adequate to support methodological procedures aimed at the development of definitions, as an aid for the development of a consistent conceptual models in ontologies. We analyzed the suitability of onomasiological method to describe the properties of the referents in the elaboration of Domain Ontologies, to identify which ontic elements (functional) can be aggregated to logical elements (genus-species relations) for the comprehension of the concept positioning in the context in question. Thus, definitions are explored, not as a final product for fixing a concept in a domain, but as a first step aiming at their fixing from of a given commitment.

**Keywords:** Domain ontologies; Definitions, Onomasiological approach.

## **1 CONSIDERAÇÕES INICIAIS**

Este estudo pretende investigar o uso de definições como um elemento básico para a elaboração de modelos conceituais consistentes em Ontologias de Domínios.

Para que um Modelo Conceitual possa ser uma descrição adequada da realidade do domínio, ele deve apresentar informações precisas e claras, não permitindo a ocorrência de ambiguidades sobre os aspectos que devem ser modelados. Uma análise mais detalhada das propriedades dos objetos de um domínio, denominada de análise ontológica, pode ser utilizada como fonte de conhecimento para modeladores conceituais, possibilitando minimizar esses erros.

Para tanto, estamos dedicados a estudar a abordagem onomasiológica apresentada por Riggs (1979, 1981, 1989) e verificar quanto esta abordagem pode ser utilizada para a elaboração de Ontologias de Domínios e compará-las com outras iniciativas no âmbito do contexto das Ontologias. No campo dos estudos informacionais esta abordagem vem sendo utilizada por Dahlberg (1978a, 1978b, 1981, 1983) como um método para a elaboração de definições em tesouros.

Consideramos que as definições podem ser o elemento básico para o entendimento do objeto em um determinado domínio: elas evidenciam as propriedades de um dado objeto através de relações que podem ser identificadas nos elementos de seu enunciado definitório. Assim, pretendemos observar quais elementos deverão constituir uma definição que possa revelar, não apenas o posicionamento do conceito no sistema (relações genéricas e específicas), mas também a descrição de fatos que envolvem tal conceito (relações ônticas).

A seguir, apresentamos algumas questões que envolvem o método onomasiológico e a perspectiva da conceptologia de Riggs, assim como a Teoria do Conceito de Dahlberg, visando apontar alguns aspectos teóricos e metodológicos, já identificados até o momento, que envolvem definição no campo das Ontologias de Domínio.

## **2 DEFINIÇÃO E ONTOLOGIAS DE DOMÍNIOS: BUSCA DE BASES TEÓRICAS E METODOLÓGICAS**

Nossos estudos têm investigado discussões sobre um vasto número de modos de definição e como estes modos estão relacionados com o compromisso ontológico (acordo) assumido em cada domínio. Consideramos que as ontologias, entre outros aspectos, necessitam de um padrão para seus enunciados definitórios que possam apoiar o entendimento pelos agentes inteligentes (software). Consideramos também, que partimos do conteúdo conceitual de uma unidade de representação, seja ela uma entidade ou uma instância, para que seja possível inferir sobre estas unidades um dado entendimento que possa ser “compreendido” por máquinas. (CAMPOS, 2010; CAMPOS et al., 2013, CAMPOS, 2014; GOMES, 2015; GOMES et al., 2010)

Entendemos aqui um enunciado definitório como o conjunto de características que descrevem um referente em um dado contexto e que tem por função, além de permitir o posicionamento do conceito em um sistema de conceitos, possibilitar o entendimento de toda a complexidade daquele conceito em um dado contexto, ou seja, que elementos ônticos (relações funcionais e partitivas) possam ser agregados aos elementos lógicos (relações de gênero-espécie) (DAHLBERG, 1981, 1983).

Há vários métodos para a elaboração de dicionários e instrumento semânticos assemelhados como tesouros e terminologias. A Linguística reconhece duas abordagens para a

construção de tais instrumentos semânticos: semasiológica e onomasiológica. A primeira é característica de dicionários gerais que lista alfabeticamente as palavras com seus vários significados. A segunda volta-se para instrumentos especializados: tem como ponto de partida o significado aceito em um domínio, acrescido de uma “definição”, ou seja, de uma descrição de seu conteúdo semântico e reúne tais significados segundo características comuns, ou seja, requer um arranjo sistemático.

Dada sua importância na recuperação de informação – quando da introdução da computação nos serviços especializados de informação – a abordagem onomasiológica se mostrou mais adequada na elaboração de tesouros documentários, nos anos 50 do século passado, quando princípios de classificação foram introduzidos em sua elaboração (CAMPOS et al., 2015).

A abordagem onomasiológica se caracteriza pelo uso de critérios que permitem uma concepção consensual da realidade sobre um objeto social, refletindo a forma como o referente é conceitualmente construído. Esta forma especial de apreender esta dimensão conceitual considera que a linguagem é um instrumento de construção de uma dada realidade e que é na língua em uso e referendada por um grupo de especialidade, que se instaura, sustenta e altera os processos sociais.

Esta concepção baseia-se no postulado da funcionalidade construtivista da linguagem e encontra alguns de seus elementos na “Lexicologia Construcionista”. De acordo com este postulado, a linguagem desempenha, ao lado das outras funções (comunicativa, interacional, etc.), a tarefa fundamental de “construir” o referente do discurso (RIGGS, 1989b).

Segundo Riggs (1989a e b, 1996) dentro do campo da Linguística existe uma variedade de disciplinas conexas, como Semântica, Sociolinguística, Psicolinguística e Planejamento da Linguagem, que podem reivindicar todas elas, como uma de suas associadas, o campo do conhecimento chamado “Lexicologia”. De outro lado, relacionados com a Filosofia, encontramos o campo da Lógica, da Filosofia da Ciência, e a pesquisa em Classificação. Neste contexto emergiu um subcampo que focaliza a análise de conceitos, incluindo suas relações com a ciência como unidades do conhecimento teoricamente significativas, e suas relações com observações empíricas julgadas por critérios “operacionais” ou “denotativos”. Este subcampo foi denominado por Riggs de conceptologia e está aberto a diferentes abordagens filosóficas e não precisa se unir a nenhuma delas.

O ponto de partida na abordagem onomasiológica é um item de referência em um discurso especializado. Isso não significa que o procedimento utilizado seja semasiológico (FAULSTICH; OLIVEIRA, 2007), porque as relações de significação não entram em jogo no sistema conceitual: ao tomar um item de referência em um dado contexto ele já é a designação de um referente e é este que vai ser analisado, resultando daí sua definição.

No campo dos estudos tesaurográficos, a “Teoria do Conceito voltada para o Referente” (DAHLBERG, 1978 a e b), de abordagem onomasiológica, fornece um método para elaboração de definições que consideramos que pode ser investigado visando apropriação em Ontologias de Domínio. Neste sentido, esta Teoria é ‘voltada para o referente’ porque pressupõe que cada conceito se refere a algo (a um objeto concreto ou abstrato) e é chamada ‘analítica’ porque, baseando-se no referente e assim gerando e identificando suas características conceituais, um conceito é construído de modo analítico.

Neste sentido, estamos considerando que Dahlberg avança na elaboração de definições e seu modelo reflete uma concepção consensual da realidade sobre um objeto social, denominada “abordagem onomasiológica”. Nesta, como abordado, estruturas de expressões refletem a forma como o referente é, conceitualmente, construído; vale dizer, pela seleção, realce e iluminação de alguns aspectos do ser e, por conseguinte, atenuação, ou mesmo ocultamento, de outros aspectos. Em outras palavras, ao iluminar uma faceta do ser, a abordagem onomasiológica produz uma “versão” do referente que constitui o modo como determinado grupo de falantes - no caso de Dahlberg, de especialistas -, interagem socialmente com o referente. E isto permite um movimento de visão sistêmica sobre o referente, em que as Categorias são o elemento agregador.

### 3 CONSIDERAÇÕES FINAIS

A partir das abordagens apresentadas consideramos que em Ontologias de Domínios a elaboração de enunciados definitórios para os elementos que compõem uma dada Ontologia, se faz necessário para que seja possível a aplicação de mecanismos automáticos para a descoberta de conhecimento sobre um dado domínio. As estratégias metodológicas utilizadas para se chegar a estes enunciados, dependem das bases teóricas que serão adotadas. Aqui defendemos o método onomasiológico como o mais apropriado para a descrição das propriedades dos referentes na elaboração de Ontologias de Domínio. Este método está apresentado na literatura por teóricos no campo dos estudos terminológicos e informacionais como Riggs e Dahlberg.

### REFERÊNCIAS

CAMPOS, M. L. A. O papel das definições na pesquisa em ontologia. **Perspectivas em Ciência da Informação**, [S.l.], v. 15, p. 220-238, 2010.

CAMPOS, M.L.A. et al. Information Sciences methodological aspects applied to ontology reuse tools: a study based on genomic annotations in the domain of trypanosomatids. **Knowledge Organization**, v. 40, p. 50-61, 2013.

CAMPOS, M. L. A.; GOMES, H. E. Princípios para modelagem de domínios: a posição de Barry Smith e de Ingetraut Dahlberg. **Ciência da Informação**, [S.l.], v. 41, p. 81-94, 2014.

CAMPOS, M. L. A.; GOMES, H. E. O papel das definições como subsídios a elaboração de Ontologias de Domínios: a perspectiva onomasiológica. In: ENCUESTRO IBÉRICO EDICIC, 7., 2015, 2015, Madrid. **Anais...** Madrid: Universidad Complutense de Madrid, 2015.

DAHLBERG, I. **A Referent-oriented analytical concept theory of interconcept**. *International Classification*, v. 5, n. 3, p.142-150,1978a.

DAHLBERG, I. Conceptual definitions for Interconcept. **International Classification**, [S.l.], v. 8, n. 1, 1981.

DAHLBERG, I. **Optical structures and universal classification**. Bangalore: Sarada Ranganthan Endowment, 1978b.

DAHLBERG, I. **Terminological definitions: characteristics and demands**. Problèmes de la définition et de la synonymie en terminologie. Québec: Association internationale de terminologie, 1983. p.13-51

FAULSTICH, E. O.; MICHELLE, M. Para que serve um dicionário analógico? Um estudo de lexicografia comparativa. **Revista Intercâmbio dos Congressos Internacionais de Humanidades**, v. 10, p. 1-16, 2007.

GOMES, H. E.; CAMPOS, M. L. A.; GUIMARAES, L. S. Organização da Informação e Terminologia: a abordagem onomasiológica. **Datagramazero**, Rio de Janeiro, v. 11, 2010.

RIGGS, Fred W. Information science and social science: the need for onomantics. **International Forum for Information and Documentation**, v. 14, n. 1, p. 12-21, 1989a.

RIGGS, Fred W. Onomantics and terminology. **Knowledge Organization**, v. 23, p. 25-3, 1996.

RIGGS, Fred W. Terminology and lexicography: their complementarity. **International Journal of Lexicography**, v. 2, n. 2, p. 90-110, 1989b.

## **Organização do conhecimento na perspectiva da abordagem cognitiva e sociocognitiva**

### *KNOWLEDGE ORGANIZATION IN THE PERSPECTIVE OF COGNITIVE AND SOCIO-COGNITIVE APPROACH*

Mariângela Spotti Lopes Fujita (UNESP)

*fujita@marilia.unesp.br*

Paula Regina Dal'Evedove (UFSCar)

*dalevedove@ufscar.br*

**Resumo:** Apresentam-se as perspectivas das abordagens cognitiva e sociocognitiva para se pensar a organização do conhecimento frente às exigências de práticas e modelos contemporâneos de recuperação da informação em ambientes multifacetados. A abordagem cognitiva foi desenvolvida com uso da técnica introspectiva do protocolo verbal em diferentes modalidades para estudos de representação temática, em especial da leitura documentária do indexador para análise de assunto de diferentes estruturas textuais em diferentes domínios de assunto e a sociocognitiva com uso do protocolo verbal com enfoque no contexto sócio-cultural do profissional e a relação com o sistema de recuperação da informação. Formas alternativas merecem ser examinadas pela área de Organização e Representação do Conhecimento, uma vez que o avanço com as abordagens cognitiva e sociocognitiva demonstra novas possibilidades e enfrentamentos a serem observados em razão das inúmeras capacidades cognitivas, sociais, culturais e educacionais a serem compreendidas no entorno dos profissionais e usuários de sistemas de informação.

**Palavras-chave:** Organização do Conhecimento. Abordagem cognitiva. Abordagem sociocognitiva.

**Abstract:** The perspectives of the cognitive and socio-cognitive approaches are presented in order to think on the knowledge organization to meet the practical demands and contemporary models of information retrieval in multifaceted environments. The cognitive approach was developed using the verbal protocol introspective technique in different modalities to the studies of thematic representation, especially to the indexer documentary reading for the subject analysis of different textual structures in different domains of subject, and the socio-cognitive approach was developed using the verbal protocol focusing on the professional socio-cultural context and the relationship with the information retrieval system. Alternative forms deserved to be examined by the area of Knowledge Organization and Representation, since the advance with the cognitive and socio-cognitive approaches demonstrates new possibilities and confrontations to be observed because of several cognitive, social, cultural and educational skills to be understood by professionals and users of information systems.

**Keywords:** Knowledge Organization. Cognitive approach. Socio-cognitive approach.

## **1 INTRODUÇÃO**

O campo da informação é permeado por dilemas básicos que envolvem a informação na condição de fenômeno “irresoluto”. Na qualidade de objeto de estudo central deste campo

científico e profissional, onde estão entrecruzadas as trajetórias, tradições e particularidades da Documentação, Arquivologia, Biblioteconomia, Museologia e da Ciência da Informação “[...] como (sub/inter)campos simultaneamente parceiros, cooperativos, conflitantes, relativamente comuns e singulares.” (MARQUES, 2011, p.76), a informação é um objeto abrigador de diversas possibilidades e olhares investigativos. Por tais razões, não se pode omitir as relações que as pessoas estabelecem com os conteúdos informativos e os processos de produção, organização, busca e uso da informação. Em especial, possibilidades dirigidas à área de Organização e Representação do Conhecimento que contemplem aspectos práticos e científicos que perpassam o tratamento da informação e seus desdobramentos mediante contributos interdisciplinares advindos de campos correlatos.

Ao lidar com o conhecimento registrado e socializado, a área ampara a atuação do profissional da informação diante das diversas práticas e atividades sociais vinculadas com o uso da informação e o acesso do conhecimento ao empreender esforços para apresentar subsídios teóricos sobre “[...] tudo aquilo que é relativo ao *tratamento da informação*, particularmente com o *tratamento temático da informação*, e de um modo menos específico – mas não menos importante – com a *gestão do uso social da informação*.” (BARITÉ, 2001, p.38, tradução nossa).

Portanto, a representação temática da informação reclama uma atenção para além do conteúdo informativo do documento. Sua natureza subjetiva condiciona o encaminhamento de pesquisas que coloquem em movimento as problemáticas desta complexa atividade, mediante a aproximação de áreas do conhecimento que ofereçam novos pensares e possibilidades investigativas na resolução de aspectos atinentes.

Neste sentido, a proposta desta comunicação é apresentar as perspectivas das abordagens cognitiva e sociocognitiva para se pensar a organização do conhecimento frente às exigências de práticas e modelos contemporâneos de recuperação da informação em ambientes multifacetados. O pensamento aqui apresentado segue uma trajetória de ensino e pesquisa em representação temática da informação sob o viés da abordagem cognitiva e sociocognitiva.

## **2 (RE)PENSAR A REPRESENTAÇÃO TEMÁTICA DA INFORMAÇÃO**

A análise de assunto é uma etapa indispensável e decisiva para o efetivo atendimento das necessidades informacionais dos usuários, uma vez que possui dimensão estratégica ao imprimir significação à informação, em atenção às características da comunidade usuária – preocupação efetiva com a dimensão de uso.

Ao versar sobre a análise de assunto na indexação, grande parte dos esforços empreendidos desde a década de 1990 frente ao Grupo de Pesquisa “Análise Documentária” foram centrados na



leitura documentária como esteio na identificação de termos representativos, em que o indexador faz uso de estratégias cognitivas e metacognitivas, além de ser imprescindível o conhecimento na área de domínio e da estrutura textual dos documentos. Em leitura documentária, “a cognição é compreendida como os processos mentais inconscientes de uma pessoa, enquanto a metacognição refere-se ao gerenciamento consciente sobre um fenômeno cognitivo pelo indivíduo” (NEVES; DIAS; PINHEIRO, 2006, p. 142).

Por ser um ato cognitivo, a abordagem cognitiva ganha espaço nas discussões que perpassam o conhecimento humano na indexação como alternativa investigativa.

[...] para análise e elaboração de teorias e metodologias. O foco é a cognição – o processo de conhecer humano que oferece uma perspectiva de investigação baseada na compreensão, no processamento e na representação (FUJITA; CERVANTES, 2005, p. 30).

A composição de pesquisas nesta linha viabiliza, desde então, que os efeitos desta complexidade sejam amenizados na prática profissional e, por consequência, na recuperação da informação mediante o encontro da Ciência da Informação com campos correlatos. Os processos cognitivos envolvidos na leitura amparam-se em esquemas mentais e no conhecimento prévio de cada sujeito. Por consequência, a representação temática aspira uma relação mais profícua com o a Psicologia Cognitiva ao passo que estas estratégias de leitura devem ser estudadas e conhecidas.

Frente as novas necessidades de se pensar o processamento da informação (SHERA, 1971), as pesquisas na temática foram ampliadas para a perspectiva sociocognitiva (HJØRLAND, 1997) na qual a realidade é socialmente construída, isto é, a informação como uma dimensão da existência humana, admitindo-se a formação coletiva e de sentido social dos sujeitos na apropriação da informação. Trabalha-se, portanto, com o entendimento de que o processo de conhecimento individual é socialmente condicionado, em que estudos de abordagem sociocognitiva ampliam a perspectiva individual para contextos socioculturais.

Dentre os resultados pragmáticos que foram acurados ao longo de pesquisas pensadas e construídas mediante a aceitação de aspectos cognitivos e sociais no tratamento da informação, estão o emprego de técnicas introspectivas oriundas da Psicologia Cognitiva para a observação do processo de indexação. A inserção do Protocolo Verbal como técnica introspectiva na Ciência da Informação, mediante informações sobre os estados de pensamento dos sujeitos, ganhou notoriedade ao ser utilizado para coleta de dados qualitativos em atividades de leitura documentária (FUJITA; NARDI; FAGUNDES, 2003). Desde então, estudos advindos da Linguística (CAVALCANTI, 1989; NARDI, 1999) viabilizaram formas alternativas de se investigar a experiência consciente dos sujeitos no processamento da informação, cujas modalidades de Protocolo Verbal Individual, Protocolo Verbal Interativo ou Protocolo Verbal em

Grupo facultam que questões cognitivas relacionadas à representação temática da informação sejam observadas de diferentes maneiras.

Os avanços advindos com os protocolos verbais são sentidos, dentre outros, na elaboração de subsídios acerca de políticas de indexação para o contexto de bibliotecas universitárias e escolares. Considerada uma filosofia de trabalho que conduz a indexação mediante diretrizes que visam a eficiência dos sistemas de informação (RUBI, 2008), exercem papel fundamental para o aprimoramento da recuperação da informação (GIL LEIVA; FUJITA, 2012).

Outro avanço significativo refere-se à sistematização de métodos destinados à representação temática da informação. A proposta metodológica de Modelos de Leitura Documentária para indexação de artigos científicos (FUJITA; RUBI, 2006) e para a análise de assunto de livros científicos (REIS; FUJITA, 2014) auxilia bibliotecários indexadores aprendizes e experientes na identificação de conceitos do documento na análise de assunto.

Como produtos de pesquisas científicas, apresentam desempenhos proficientes e comprovam a importância de formas alternativas de se investigar o profissional bibliotecário no cumprimento de suas atribuições. Ademais, estão diretamente imbricados no desenvolvimento de avanços teóricos e metodológicos que fundamentam o ensino e a formação continuada de bibliotecários, como base para delinear pesquisas de vertente mais aplicada.

### **3 DESAFIOS E OPORTUNIDADES TRANSVERSAIS**

Algumas problemáticas que permeiam a organização e a recuperação da informação são melhor compreendidas quando conduzidas pela abordagem cognitiva e sociocognitiva. Estudos e reflexões nesta linha, em especial Cintra (1987), Farrow (1991), Frohmann (1992), Pinto Molina (1993), Naves (2000), Hjørland (2002), Ingwersen (2002), Fourie (2003), Neves (2004) e Fujita (1999, 2003, 2004, 2007), contribuem ativamente com avanços teóricos e aplicáveis em representação temática da informação. Considera-se pertinente que as implicações que permeiam a informação sejam observadas por diferentes ângulos, no intuito de desvelá-las em diferentes concepções por meio de um trabalho de análise e maturação científica, pois o campo da informação, suas disciplinas constitutivas e áreas de investigação, tal como a Organização e Representação do Conhecimento, da teoria às aplicações, são permeados por conceitos, noções e ideias que atuam como ponto convergente de vários outros saberes científicos.

Neste sentido, compreende-se que formas alternativas de se debruçar sobre a questão merecem ser examinadas pela área de Organização e Representação do Conhecimento, uma vez que o avanço neste eixo investigativo descortina de modo paulatino novas possibilidades e enfrentamentos a serem observados em razão das inúmeras capacidades cognitivas, sociais,

culturais e educacionais a serem compreendidas no entorno dos profissionais e usuários de sistemas de informação. Dal'Evedove (2014) traz um estudo preliminar da abordagem sociocultural no eixo da representação temática da informação, sendo articulados os modos sociais de pensamento de docentes, profissionais e usuários para a observância de alguns problemas concretos que perpassam a perspectiva humana do processo em bibliotecas universitárias. Este olhar integrado viabilizou que as reflexões e subsídios obtidos fossem fundamentados nos fundamentos teóricos e metodológicos da área de Organização e Representação do Conhecimento, no contexto sociocognitivo do bibliotecário indexador e, sobretudo, nas características e princípios culturais dos usuários.

Diante da necessidade de não deixar alijada questões de identidade cultural, a Ciência da Informação busca nos fundamentos de outros campos científicos, tais como Antropologia, Sociologia e Psicologia, novas interpretações para tratar questões de representação da informação para culturas específicas ou domínios do conhecimento. Entende-se ser esta uma alternativa contributiva e uma postura que vai ao encontro com o perfil dos usuários contemporâneos que exigem informações cada vez mais personificadas. O grande desafio, portanto, é alinhar as facetas da organização e recuperação da informação às possibilidades contributivas das abordagens cognitiva e sociocognitiva para o fortalecimento das dimensões teóricas e aplicáveis do processo de indexação; perspectiva necessária na Organização e Representação do Conhecimento para o desvendamento de questões que perpassam da produção ao consumo da informação, com fins de construção e socialização do conhecimento.

Apesar dos avanços teóricos e metodológicos advindos pela abordagem cognitiva e sociocognitiva, os aspectos que envolvem a representação temática da informação demandam pesquisas contínuas para seu aperfeiçoamento. Dentre as perspectivas investigativas que reclamam por olhares mais verticalizados estão o aprimoramento de normas para a análise de assunto e políticas de indexação na condição de suportes para a realização da indexação nos contextos de formação e prática profissional que exigem metodologia etnográfica e de pesquisa ação integral como proposta de discussão de problemas com resultados negociados em pactos cujos registros sejam concretizados em manuais de política de indexação.

## REFERÊNCIAS

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (Org.). **Educação, universidade e pesquisa**. Marília: UNESP; São Paulo: FAPESP, 2001. p. 35-60.

CINTRA, A. M. M. Estratégias de leitura em documentação. In: SMITT, J.W. **Análise e documentação**: a análise da síntese. Brasília: IBICT, 1987. p. 29-38.

DAL'EVEDOVE, P. R. **O tratamento temático da informação em abordagem sociocultural: diretrizes de política de indexação em bibliotecas universitárias.** 2014. 266 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

FARROW, J. F. A cognitive process model of document indexing. **Journal of Documentation**, [S.l.], v. 47, n. 2, p. 149-166, June 1991.

FOURIE, I. How can we take a socio-cognitive approach in teaching indexing and abstracting?. **The Indexer**, London, v. 23, n. 2, p.83-85, Oct. 2003.

FROHMAN, B. Rules of indexing: a critique of mentalism in information retrieval theory. **Journal of Documentation**, [S.l.], v. 46, n. 2, p.81-101, 1990.

FUJITA, M. S. L. **A leitura documentária do indexador: aspectos cognitivos e lingüísticos influentes na formação do leitor profissional.** 2003a. 321f. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) – Faculdade de Filosofia e Ciências, UNESP, Marília, 2003a.

FUJITA, M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. **DataGramZero** - Revista de Ciência da Informação, Rio de Janeiro, v.5, n.4, ago. 2004.

Disponível em:

<<http://basessibi.c3sl.ufpr.br/brapci/index.php/article/view/0000007547/f439490af520c6a64bbd99c4a74f163>>. Acesso em: 18 maio 2016.

FUJITA, M. S. L. La enseñanza de la lectura documentaria en el abordaje cognitivo y socio-cognitivo: orientaciones a la formación del indizador. **Anales de Documentación**, [S.l.], v. 10, p. 1-16, 2007b.

FUJITA, M. S. L.; CERVANTES, B. M. N. Abordagem cognitiva do protocolo verbal na confirmação de termos para a construção de linguagem documentaria em inteligência competitiva. In: VALENTIM, M. L. P. (Org.). **Métodos qualitativos de pesquisa em Ciência da Informação.** São Paulo: Polis, 2005. p.29-56.

FUJITA, M. S. L.; NARDI, M. I. A.; FAGUNDES, S. A. A observação da leitura documentária por meio do protocolo verbal. In: RODRIGUES, G. M. R.; LOPES, I. L. (Org.). **Organização e representação do conhecimento na perspectiva da Ciência da Informação.** Brasília: Thesaurus, 2003. p. 141-178. (Estudos avançados em Ciência da Informação, v. 2).

FUJITA, M. S. L.; RUBI, M. P. Um modelo de leitura documentária para a indexação de artigos científicos: princípios de elaboração e uso para a formação de indexadores. **Datagramazero**, Rio de Janeiro, [S.l.], v. 7, n. 3, p.1-18, 2006.

FUJITA, M. S. A leitura do indexador: estudo de observação. **Perspect. cienc. inf.**, Belo Horizonte, [S.l.], v. 4, n. 1, p. 101-116, jan./jun. 1999.

GIL LEIVA, I.; FUJITA, M. S. L. (Org.). **Política de indexação.** São Paulo: Cultura Acadêmica; Oficina Universitária, 2012. 260p.

- HJØRLAND, B. Epistemology and the sócio-cognitive perspective in information science. **Journal of the American Society for Information Science and Technology**, [S.l.], v. 53, n. 4, p. 257-70, 2002a.
- HJØRLAND, B. **Information seeking and subject representation: anactivity-theoretical approach to information science**. Westport: Greenwood Press, 1997. 213p.
- INGWERSEN, P. **Information retrieval interaction**. London: Taylor Graham, 1992.
- MARQUES, A. A. I. C. **Interlocuções entre a Arquivologia nacional e a internacional no delineamento da disciplina no Brasil**. 2011. 399f. Tese (Doutorado em Ciência da Informação) – Faculdade de Ciência da Informação, Universidade Nacional de Brasília, Brasília, 2011.
- NARDI, M. I. A. **A metáfora e a prática de leitura como evento social: instrumentos do pensar a Biblioteconomia do futuro**. 1999. 272f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica, São Paulo, 1999.
- NAVES, M. KOCH, I. G. V. **Desvendando os segredos do texto**. São Paulo: Cortez, 2002. 168p.
- NAVES, M. M. L. **Fatores interferentes no processo de análise de assunto: estudo de casos de indexadores**. 2000. 275f. Tese (Doutorado em ciência da informação) – Escola de ciência da informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2000.
- NEVES, D. A. de B.; DIAS, E. W.; PINHEIRO, A. M. V. Uso de estratégias metacognitivas na leitura do indexador. **Ciência da Informação**, Brasília, v. 35, n. 3, p. 141-152, set./dez. 2006. Disponível em: <http://www.scielo.br/pdf/ci/v35n3/v35n3a14.pdf>. Acesso em: 24 abr. 2016.
- NEVES, D. A. de B. **Aspectos metacognitivos na leitura do indexador**. 2004. 131 f. Tese (Doutorado em ciência da informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.
- PINTO MOLINA, M. **Análisis documental: fundamentos y procedimientos**. 2. ed. rev. aum. Madrid: EUEDEMA, 1993. 270p.
- REIS, D. M.; FUJITA, M. N. S. L. Análise de assunto em bibliotecas universitárias: a observação de estruturas textuais em livros das áreas de humanas, exatas e biológicas. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, [S.l.]. **Anais...** [S.l.: s.n.], 2014.
- RUBI, M. P. **Política de indexação para construção de catálogos coletivos em bibliotecas universitárias**. 2008. 169f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2008.
- SHERA, J. The sociological relation shipsof information science. **Journal of the American Society for Information Science**, [S.l.], v. 22, n. 1, p.76-80, mar./apr. 1971.

## PALESTRA

### A PESQUISA NA EMPRAPA: O CONVÊNIO COM O GRUPO DE PESQUISA MHTX

Ivo Pierozzi Júnior (Embrapa Informática Agropecuária, Campinas – SP)

*ivo.pierozzi@embrapa.br*

**Resumo:** O Grupo de Pesquisa MHTX formalizou um convênio de cooperação técnica com a Embrapa Informática Agropecuária, visando a elaboração e o gerenciamento de Sistemas de Organização do Conhecimento (SOC) da Embrapa. A convergência de interesses e motivações de ambas as instituições permitiram o desenvolvimento e a validação de vários recursos, entre os quais, terminologias; taxonomias facetadas e navegacionais; reengenharia semântica para vocabulários controlados; modelos de avaliação de SOC, prevendo alinhamento terminológico baseado em reuso de tesouros; e ferramentas computacionais amigáveis para visualização de SOC com facilidades para cognição e apreensão conceitual da informação e do conhecimento representados. Diante de tantos resultados significativos, vislumbra-se um horizonte amplo e convidativo à proposição e exploração de novos temas de pesquisa conjunta entre ambas as instituições e de suas aplicações decorrentes.

**Palavras-chave:** Cooperação técnica. Gestão da Informação. Gestão do Conhecimento. Sistemas de Organização do Conhecimento. SOC.

**Abstract:** The MHTX research group has formalized a technical cooperation agreement with Embrapa Agricultural Informatics, aiming at the elaboration and management of Embrapa's knowledge organization systems (KOS). The convergence of interests and motivations of both institutions have allowed the development and validation of various resources, including terminologies; faceted and navigational taxonomies; semantic reengineering for controlled vocabularies; models for KOS evaluation including terminological alignment based on reuse of thesauri besides computational tools with friendly facilities for cognition and conceptual apprehension of the represented knowledge. Face to these significant results, there is a broad and inviting horizon to the proposition and exploring of new themes of joint research between the two institutions and their subsequent applications.

**Keywords:** Technical cooperation. Information Management. Knowledge Management. Knowledge Organization Systems. KOS.

## 1 INTRODUÇÃO

A Empresa Brasileira de Pesquisa Agropecuária (Embrapa) é vinculada ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA), do governo brasileiro e tem como missão viabilizar soluções de Pesquisa, Desenvolvimento e Inovação (PD&I) para a sustentabilidade da agricultura, em benefício da sociedade brasileira (EMBRAPA, 2015). A Embrapa é composta de vários tipos de Unidades Descentralizadas (centros de pesquisa, serviços especializados, laboratórios no Exterior e escritórios internacionais) além de Unidades Centrais na sua sede, em

Brasília, DF, que se ocupam dos macroprocessos de desenvolvimento institucional e administrativos<sup>7</sup>.

Assim como inúmeras outras instituições de PD&I, envolvidas com multi, inter e transdisciplinaridade, a Embrapa precisa incorporar em suas abordagens técnico-científicas a organização e representação da complexidade dos sistemas naturais e antrópicos como novo paradigma do pensamento humano frente à compreensão do mundo real e frente ao desafio de direcionar conhecimento como solução de problemas. Todos os grandes desafios da sociedade moderna, em escala global, dos quais vários são diretamente relacionados à atuação da Embrapa (por exemplo: mudanças climáticas, segurança alimentar e sustentabilidade de recursos naturais), requerem variedade e grande volume de dados e informação, que devem ser compartilhados transpondo diferentes culturas, escalas e tecnologias. Essa situação exige uma nova forma e uma nova concepção de processo, infraestrutura e operação para a organização, engenharia, gestão e apropriação do conhecimento (PIEROZZI JUNIOR et al., 2014).

A Embrapa Informática Agropecuária<sup>8</sup> é um dos centros de pesquisa da Embrapa, focado na viabilização e no desenvolvimento de aplicações de tecnologia da informação e de computação para a agropecuária. Para cumprir seus objetivos, atua especialmente nas áreas de agroinformática e bioinformática aplicadas ao setor agropecuário, reconhecendo o caráter transversal de suas competências técnicas como suporte e aplicação a inúmeras questões relacionadas com a agricultura brasileira. A Embrapa Informática Agropecuária propõe soluções para os diversos domínios agrícolas, valorizando tecnologias emergentes e de padrões livres, com desenvolvimento de *software* e utilização de modelos de inteligência e matemática computacionais, como solução a problemas complexos do setor agropecuário (MASSRUHÁ et al., 2014). No seu itinerário de produção técnico-científica, várias temáticas de pesquisa precisam ser complementadas com a expertise de colaboradores externos, oriundos seja da área acadêmica, seja de outras instituições de PD&I públicas ou mesmo privadas. Essa complementaridade de competências é formalizada em instrumentos jurídicos específicos, como convênios ou contratos de cooperação, que desenham os interesses e motivações da colaboração, as responsabilidades envolvidas, as estruturas disponíveis e necessárias, a negociação das questões de propriedade intelectual, além de um plano de ação detalhado contendo descrição e cronograma de atividades e, quando pertinente, previsão orçamentária e desembolso financeiro.

Nesse contexto, a Embrapa Informática Agropecuária vislumbrou a oportunidade de enriquecer suas abordagens na área de organização e representação do conhecimento propondo

---

<sup>7</sup> <https://www.embrapa.br/quem-somos>

<sup>8</sup> <https://www.embrapa.br/informatica-agropecuaria>

uma parceria com o Grupo de Pesquisa MHTX, a qual foi consolidada pelo Convênio de Cooperação Técnica Embrapa Código 23800.14/0023-0, assinado em 27.08.2014. A execução operacional do convênio, se estabeleceu inicialmente pelo projeto intitulado “Elaboração e gerenciamento de sistemas de organização do conhecimento (SOC) na Embrapa”. O relato das atividades e produtos originados desse projeto, nos primeiros quatro anos de sua execução é objeto do presente capítulo.

## **2 RESULTADOS DA PARCERIA EMBRAPA-GRUPO DE PESQUISA MHTX**

A parceria entre a Embrapa e o Grupo de Pesquisa MHTX já viabilizou diversos resultados desenvolvidos em quase quatro anos de colaboração. Nesse período, a Embrapa participou, no âmbito do Programa de Pós-Graduação em Ciência da Informação da UFMG, de três bancas de exame de qualificação de alunos de mestrado e doutorado e duas bancas de defesa de tese de doutorado (ALVES, 2014; MACULAN, 2015). No período, ainda foram promovidas e realizadas, com a promoção conjunta, quatro reuniões técnicas presenciais, para discussões e alinhamento conceitual de temas de pesquisa, proferimento de palestras e para um curso intensivo de utilização do *software* e-Termos no processo de concepção, construção e gestão de SOC e de gestão terminológica. Decorrente dessa profícua e continuada colaboração, em 2015, foram publicados os primeiros trabalhos, apresentados em congressos e encontros técnico-científicos nacionais e internacionais (MACULAN et al., 2015a; 2015b; MENDES NETTO et al., 2015). Ressaltam-se, ainda, a colaboração recíproca dos parceiros na organização e realização de dois eventos de maior envergadura: (a) II Workshop dos SOC – Embrapa, realizado na Embrapa Informática Agropecuária, no período de 26 a 29 de agosto de 2014, em Campinas, SP e reunindo profissionais das áreas de Ciências Naturais e Agrônômicas, Computação, Ciência da Informação e Linguística e (b) o II Seminário do Grupo de Pesquisa MHTX, realizado na ECI/UFMG, no período de 02 a 03 de junho de 2016, em Belo Horizonte, MG, com convite à discussão de temática relacionada à organização do conhecimento e recuperação da informação.

## **3 CONCLUSÕES E PERSPECTIVAS FUTURAS**

A avaliação dos resultados relatados acima revela que, em quatro anos de parceria, ambas as instituições envolvidas nessa proposta de cooperação técnica, beneficiaram-se reciprocamente. Abusando de uma oportuna metáfora, se de um lado o Grupo de Pesquisa MHTX encontrou um terreno fértil para cultivar oportunidades de pesquisa, desenvolvimento e validação de hipóteses, técnicas e ferramentas operacionais para organização e recuperação da informação, do seu lado a



Embrapa já colhe os frutos desse trabalho, trazendo para dentro de seus processos corporativos de Gestão da Informação e do Conhecimento, soluções metodológicas e tecnológicas de facilitação do percurso dos ciclos de vida de dados, da informação e do conhecimento, garantindo que seu principal compromisso social, a produção e entrega de conhecimento do domínio agropecuário, seja alcançado, cada vez mais, de forma efetiva, eficiente e eficaz.

Com os resultados e produtos alcançados no âmbito dessa parceria, a Embrapa já incorporou, em seu arsenal de recursos, terminologias, taxonomias facetadas e navegacionais, reengenharia semântica para vocabulários controlados; modelos de avaliação de SOC, prevendo recursos de alinhamento terminológico baseado em reuso de tesouros e recursos computacionais amigáveis para visualização de SOC com facilidades para cognição e apreensão conceitual da informação e do conhecimento representados. Numa reflexão de mais alto nível, a Embrapa se beneficia dessa parceria, na medida em que traz para dentro de seu contexto organizacional modelos e ferramentas mais atualizados de organização e representação do conhecimento que produz, permitindo sua maior visibilidade, acessibilidade e apropriação.

Insistindo na metáfora, o terreno fértil onde a cooperação entre a Embrapa e o Grupo de Pesquisa MHTX foi cultivada se estende, agora, por um horizonte amplo e cheio de oportunidades, tão amplas quanto as oportunidades que a pesquisa agropecuária sempre ofereceram para o bem-estar e a sobrevivência da humanidade.

## REFERÊNCIAS

- ALVES, L. L. **Projeto de sistemas de recuperação de informação corporativa: uma abordagem de análise de domínio baseada na análise facetada.** 2014. 178f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2014.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **VI Plano Diretor da Embrapa 2014-2034.** Brasília, DF: Embrapa, 2015. 24p.
- MACULAN, B. C. M. S. **Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do THESAGRO.** 339f. 2015. Tese (Doutorado em Ciência da Informação)– Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.
- MACULAN, B. C. M. S. et al. Aplicação de reengenharia de tesouro: modelagem do THESAGRO. In: CONGRESO ISKO ESPAÑA, 12.; CONGRESO ISKO ESPAÑA-PORTUGA, 2., 2015, Murcia. **Actas...** Murcia: Universidad de Murcia, 2015b. p. 1-11.
- MACULAN, B. C. M. S. et al. Refinamento de relações em tesouros: reengenharia do Thesagro. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Anais...** João Pessoa, Ancib, 2015a.

MASSRUHÁ, S. M. F. S. et al. (Ed.). *Tecnologias da informação e comunicação e suas relações com a agricultura*. Brasília, DF: Embrapa, 2014.

MENDES NETTO, C.; LIMA, G. A. B. de O.; PIEROZZI JUNIOR, I. Aplicação da Teoria da Análise Facetada e Mapas Conceituais para navegação facetada em uma ontologia de domínio: resultados preliminares. In: CONGRESO ISKO ESPAÑA, 12.; CONGRESO ISKO ESPAÑA-PORTUGA, 2., 2015, Murcia. **Actas...** Murcia: Universidad de Murcia, 2015. p. 1-13.

PIEROZZI JUNIOR, I. et al. Gestão da informação e do conhecimento. In: MASSRUHÁ, S. M. F. S. et al. (Ed.). **Tecnologias da informação e comunicação e suas relações com a agricultura**. Brasília, DF: Embrapa, 2014. cap. 12. p. 235-258.

## 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.

## **Tainacan: sistema de gestão de biblioteca digital social**

### *TAINACAN: SOCIAL DIGITAL LIBRARY MANAGEMENT SYSTEM*

Marcel Ferrante Silva (UFG/GO)

*marcelf@gmail.com*

**Resumo:** O estudo apresenta modelos de arquitetura de bibliotecas digitais e os princípios da Web 2.0 para investigar como esses princípios podem ser incorporados em uma biblioteca digital e como a comunidade de usuários pode participar na alimentação, curadoria, interação com redes sociais, construção do sistema de organização, definição do esquema de metadados e no enriquecimento dos metadados dos objetos de uma biblioteca digital. O estudo descreve o desenvolvimento de um sistema de gestão de biblioteca digital social com funcionalidades e características que atendem os princípios da Web 2.0.

**Palavras-chave:** Biblioteca digital. Repositório. Web 2.0. Colaboração.

**Abstract:** The study presents models of digital libraries and the principles of Web 2.0 architecture to investigate how these principles can be incorporated into a digital library and how the user community can participate in feeding, curation, interaction with social networks, the organization system construction, metadata schema definition and enrichment of metadata objects from a digital library. The study describes the development of a social digital library management system with features and functionality that meet the principles of Web 2.0.

**Keywords:** Digital library. Repository. Web 2.0. Collaboration.

## **1 INTRODUÇÃO**

É notório que a partir dos anos 2000 com a popularização da internet e banda larga, teve um crescimento expressivo de bibliotecas digitais, repositórios institucionais, exposições virtuais, arquivos abertos e vários outros ambientes de compartilhamento e difusão da informação. Uma biblioteca digital pode ser definida como o “local onde os documentos eletrônicos podem ser acessados através de redes eletrônicas”, principalmente a internet (PINHEIRO; FERREZ, 2014, p. 195). Não necessariamente o documento, físico ou eletrônico está contido na biblioteca digital, mas ela fornece a referência para encontrá-lo. Segundo Sayão:

[...] proporciona acesso a uma grande variedade de conteúdos, e disponibiliza elenco diversificado de serviços voltados para gestão da biblioteca e de seus estoques informacionais e para a **disseminação e o acesso** a esses estoques. Os conteúdos gerenciados pela biblioteca podem incluir virtualmente qualquer tipo de material em formato digital, assim como bases de dados de periódicos, de artigos e de resumos e ainda descrições de objetos e de coleções físicas. (SAYÃO, 2009, p. 174).

Uma outra definição bastante difundida é a da *Digital Library Federation*:

Bibliotecas digitais são organizações, que disponibilizam recursos, incluindo recursos humanos especializados, para a **seleção, estruturação, interpretação, distribuição e disponibilização** de objetos digitais, e que devem zelar por sua

**integridade/autenticidade**, de forma que sejam acessíveis a baixo custo para a comunidade (DIGITAL LIBRARY FEDERATION, 1998, *on-line*).

No contexto deste estudo, é pertinente apresentar também conceito de repositório digital que é similar ao de uma biblioteca digital. Segundo o tesouro brasileiro da ciência da informação sua definição é:

Mecanismo para **administrar, armazenar e preservar** conteúdos informacionais em formato eletrônico, e que podem ter como foco um assunto (repositórios temáticos) ou a produção científica de uma instituição (repositórios institucionais). Muitos permitem o acesso universal e gratuito a seus conteúdos, que variam de acordo com a política de cada instituição. São coleções digitais de documentos de interesse para a pesquisa científica e, no caso dos institucionais, representam a sua memória científica (PINHEIRO; FERREZ, 2014, p. 195).

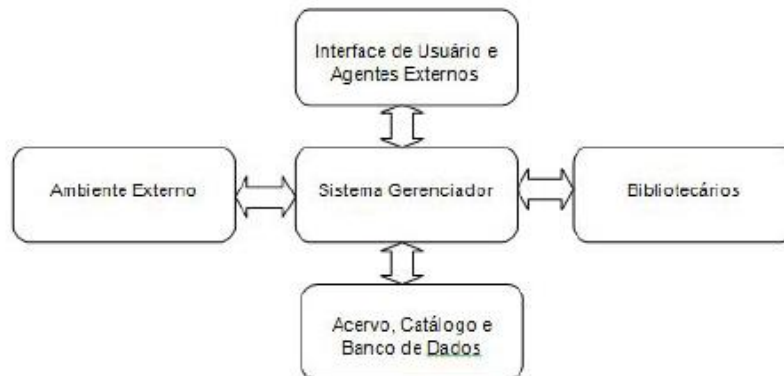
Conceitualizado por Cunha e Cavalcanti (2008, p. 322), um repositório institucional é “a coleção, digital ou em papel, que **capta e preserva** a memória intelectual de uma comunidade ou organização; memória institucional”. Em negrito foram evidenciadas as principais funções que uma biblioteca ou repositório digital deve ter. Percebe-se que a preservação é um requisito mais comum na definição de repositório que na definição de biblioteca digital. De qualquer maneira, é perceptível a similaridade entre esses conceitos, e nesse estudo foi adotado o termo biblioteca digital.

A evolução das bibliotecas digitais também está sendo impulsionada pela Web 2.0, que é caracterizada por ambientes para colaboração, interação e participação comunitária tais como comunidades, redes sociais, *blogs* e *wikis* (BLATTMANN; SILVA, 2007). Alguns estudos abordam como essas mídias sociais da Web 2.0 podem trabalhar em conjunto com as bibliotecas digitais, entretanto, o presente trabalho foca na seguinte questão: Como os princípios da Web 2.0 podem ser incorporados em uma biblioteca digital e como a comunidade de usuários pode ser envolvida na alimentação, curadoria, interação com redes sociais, construção do sistema de organização, definição do esquema de metadados e no enriquecimento dos metadados dos objetos de uma biblioteca digital? Pretende-se obter escalabilidade na construção e manutenção de uma coleção através de uma gestão compartilhada e colaborativa em contraposição a tradicional gestão hierarquizada que é centralizada e onerosa.

## 2 MODELOS DE ARQUITETURA DE BIBLIOTECAS DIGITAIS

Para uma melhor compreensão de uma biblioteca digital são apresentados alguns modelos de arquitetura encontrados na literatura. Um dos primeiros modelos foi sugerido por Frew (FIG.1):

FIGURA 1 - Modelo de uma biblioteca digital

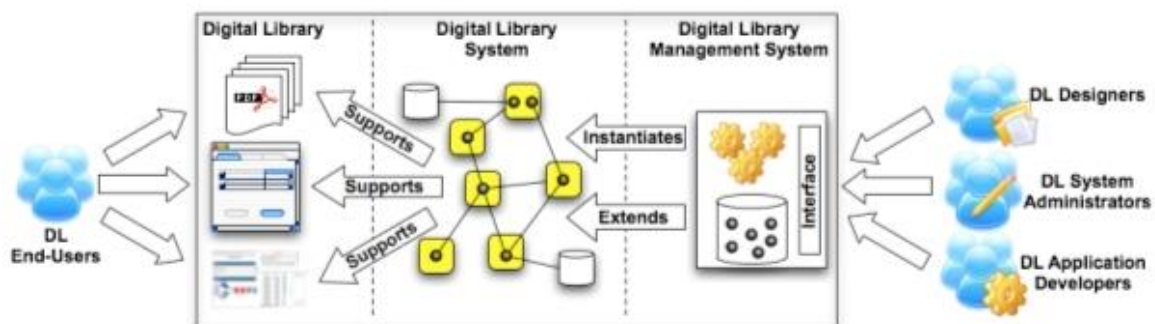


Fonte: Pontes (2013, p. 50 *apud* FREW et al., 1998)

Neste modelo Frew propõe que o núcleo da biblioteca digital seja um sistema gerenciador que faz a interface com o ambiente externo. Ele tem os seguintes componentes: interface de usuário, acervo, catálogo e banco de dados e serve aos usuários, agentes e bibliotecários (PONTES, 2013).

A próxima representação, o modelo DELOS (FIG. 2), se destaca por ser mais completo: neste modelo a biblioteca digital é suportada por um sistema de biblioteca digital que é uma instância do sistema de gestão de biblioteca digital. O modelo também caracteriza os usuários do sistema: os usuários finais (produtores, consumidores e bibliotecários), designers, administradores e desenvolvedores da biblioteca digital.

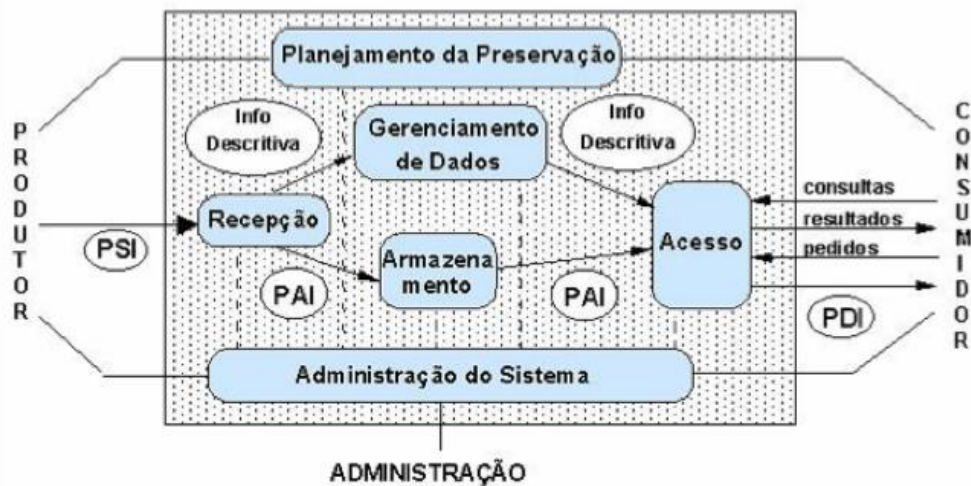
FIGURA 2 - Visão de uma biblioteca digital pelo modelo DELOS



Fonte: CANDELA, 2007.

A seguir apresentamos o modelo OAIS ou (*Open Archival Information System*) traduzido para o português como SAAI (Sistema Aberto de Arquivamento de Informação). Esse modelo foca na função da preservação que uma biblioteca digital deve ter. Neste modelo é nítido a divisão em usuário produtor, usuário consumidor e usuário administrador do SAAI (FIG. 3).

FIGURA 3 – Modelo OAIS



Fonte: THOMAZ; SOARES, 2004.

De uma forma geral, observa-se que os usuários dos modelos tradicionais de biblioteca digital têm perfis bem distintos e claros: usuários finais da biblioteca digital consomem o conteúdo e os produtores de alimentam a biblioteca digital com conteúdo e os ou administradores ou bibliotecários definem o esquema de metadados, vocabulário controlado a ser usado, entre outras funções administrativas.

A Web 2.0 iniciada a partir dos anos 2000 propôs uma revolução nesse aspecto misturando as funções de consumidor e produtor. A seguir são apresentados conceitos da Web 2.0 que se pretende trazer para o mundo das bibliotecas digitais.

### 3 WEB 2.0

Tim O'Reilly, no artigo intitulado "What is Web 2.0", lançado em 2005, consolidou o termo Web 2.0, e definiu seus setes princípios (O'REILLY, 2005).

O primeiro dele aponta a **Web como plataforma**: a Web começou como apenas um imenso repositório de documentos e informações que podiam ser compartilhados e lidos em qualquer parte do mundo a qualquer momento. A Web 2.0 impactou no modelo tradicional de desenvolvimento de software como aplicativo local ou cliente/servidor. Vários programas passam a ser desenvolvidos apenas para Web, que passa a ser adotada como plataforma principal. A visão geral é de que ela é um grande meio para disponibilizar serviços on-line. É importante observar que as informações não são descartadas, pelo contrário, o conteúdo agrega valor aos serviços. Segundo O'Reilly (2005, p. 1), "o valor do software é proporcional à escala e dinamismo dos dados que ajuda a gerir".

O segundo princípio é a questão do **uso da inteligência coletiva**: na primeira fase da Web, muitas empresas apenas armazenavam as informações dos usuários. Na Web 2.0, muitos sites estimulam os usuários a produzir mais informações coletivamente e, quanto mais e em maior escala isso for realizado, mais valor e impacto o site tem. Mais do que apenas produzir informação, um dos objetivos principais desses sites é utilizar da colaboração de cada usuário para a melhoria contínua da qualidade do serviço, o que foi chamado de uso da inteligência coletiva. Por exemplo, na Wikipédia, a inteligência coletiva é aplicada para a construção de uma enciclopédia; no Youtube, para a construção de um portal de canais de vídeo; no Digg, um portal de notícias. Com base nesses exemplos, conclui-se que "os efeitos de rede de contribuições dos usuários são a chave para o domínio do mercado na era da Web 2.0" (O'REILLY, 2005, p. 2).

O terceiro princípio aponta para o **dado como o elemento mais importante**: inúmeras empresas de sucesso na Web têm como principal diferencial competitivo os dados ou o conteúdo que disponibilizam. Seja um banco de dados especializado, tal como o Google Maps, ou um repositório de informações produzidas pelos próprios usuários, tal como a Wikipédia ou Youtube, quanto mais abrangente e confiável for, mais valor o *site* na Web tem. Os direitos autorais sobre esses dados têm sido uma questão cada vez mais importante e, apesar de, em muitos casos, o usuário inserir o conteúdo, tais empresas reclamam para si os direitos autorais para veiculação de determinadas informações, como revisões de produtos e respostas a questões. Além disso, o usuário, ao inserir os dados nessas bases, e concordando com os termos de uso, pode estar sujeito a ter suas informações sendo usadas para marketing e outras atividades comerciais. De uma maneira geral, quanto mais informações forem agregadas, produzidas por especialistas ou usuários, mais valor e importância o site tem. O dado chega a ser superior que a tecnologia empregada na construção da mesma, já que isto pode ser copiado.

O quarto princípio trata do **fim do ciclo de lançamento de software**: tradicionalmente, empresas como a Microsoft ficam anos desenvolvendo produtos, para que eles sejam lançados e consumidos por seus clientes. Esse é o modelo tradicional de desenvolvimento de software desde o seu início. Na Web 2.0, temos um novo paradigma: ao invés de fornecer produtos, as empresas desenvolvedoras de software fornecem serviços. Tais serviços são constantemente aprimorados e lançados em um ritmo muito maior que dos softwares desktops tradicionais. Por exemplo, o Flickr, site de compartilhamento de imagens, já revelou que fazem a publicação de atualizações no sistema a cada meia hora. É comum dizer que essas aplicações acessadas pela Web são um "beta perpétuo", que nunca finalizam seu desenvolvimento ou o ciclo de lançamento do produto.

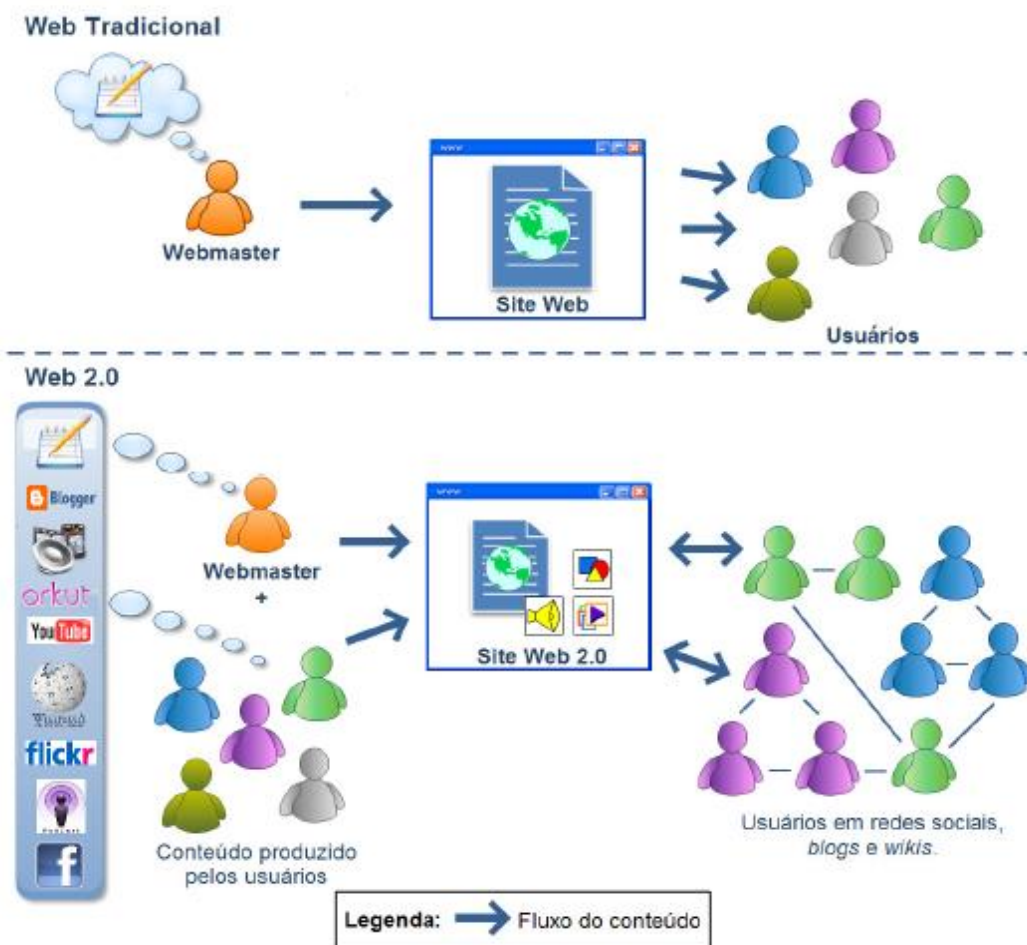
O quinto princípio aborda os **modelos leves de programação**: o "beta perpétuo" foi estimulado também pelas linguagens de programação interpretada (*scripts*), pois não é preciso

compilar o código-fonte para o sistema funcionar. Isso é também uma grande diferença em relação às empresas de desenvolvimento de software e permite a construção de sistemas dinâmicos que estão em constante mudança. As linguagens de programação tais como Perl, PHP, Python e Ruby foram fortemente adotadas. Na Web, um modelo robusto e complexo cede seu lugar a um modelo mais simples, flexível e fracamente acoplado, permitindo uma maior independência dessas aplicações e uma curva de aprendizado no desenvolvimento mais rápida, ou seja, com pouco tempo de estudo, já é possível dar seus primeiros passos, utilizando e adaptando.

O sexto princípio propõe o desenvolvimento e adaptabilidade do software em **qualquer plataforma**: as aplicações da *Web 2.0* são feitas para rodar em qualquer plataforma de computador, independentemente do sistema operacional. Desde que o navegador de Internet esteja de acordo com os padrões da W3C, mesmo que em um dispositivo móvel, isto será transparente para a aplicação *Web*.

O sétimo e último princípio estimula a produção de uma **interface de usuário rica**: no início, a *Web* era um suporte para leitura de documentos estáticos. Com o *Javascript* foi possível adicionar capacidade de processamento no cliente, ou seja, o computador do usuário é responsável pelo processamento de parte da lógica da aplicação. Posteriormente com a utilização do XML, houve o surgimento da metodologia AJAX, acrônimo para *Javascript* Assíncrono e XML, para o desenvolvimento de interface de aplicações *Web 2.0*, que se parecem com aplicações desktop, que são rápidas, interativas e intuitivas. Com o AJAX, é possível a atualização de dados na interface da aplicação *Web* sem a necessidade de o navegador de Internet recarregar toda a página.

A Figura 4 ilustra esse novo paradigma introduzido pela *Web 2.0*, no qual os usuários não são apenas meros leitores do conteúdo disponibilizado, eles passam a exercer o papel de avaliadores ou geradores do conteúdo.

FIGURA 4 - Web tradicional *versus* Web 2.0

Fonte: CIRILO, 2011, p. 23.

Considerando essa ilustração da *Web 2.0*, observa-se que os modelos tradicionais de biblioteca digital se enquadram mais no conceito da *Web 1.0*, no qual os usuários fazem papel de meros leitores. Assim, a proposta desse estudo é evoluir o modelo de biblioteca digital para que os usuários possam atuar ao mesmo tempo como leitores, curadores, produtores do conteúdo e designers do sistema de organização (taxonomias, tesouros ou ontologias) e esquema de metadados.

#### 4 DESENVOLVIMENTO

O *Tainacan* é um projeto de pesquisa com o objetivo de ser uma solução para que usuários não especialistas possam criar bibliotecas digitais. Além disso, objetiva-se uma forte interação com redes sociais. O embrião desse projeto foi iniciado na tese de doutorado em Ciência da Informação defendida na Escola de Ciência da Informação da UFMG em outubro de 2013 e intitulada: Proposta de Modelo de Colaboração para Catálogo Web Facetado (SILVA, 2013). No início de 2015, o projeto foi apoiado pelo Ministério da Cultura como solução para a construção de acervos



digitais culturais e foi completamente reformulado com a refatoração do código. Vai completar 1 ano e meio de desenvolvimento com uma equipe de cerca de 12 pessoas.

Uma etapa importante no início do projeto foi a realização de uma ampla pesquisa na literatura por referências sobre lista de requisitos ou critérios de avaliação de sistemas de gestão de bibliotecas digitais. Esses requisitos foram reunidos em uma grande lista e depois organizados a partir dos componentes da arquitetura da informação (busca, navegação, organização e rotulagem). Posteriormente, novas categorias de requisitos foram acrescentadas ao modelo, como a administração, interoperabilidade, preservação digital e em especial a colaboração/integração com redes sociais. Nessa última dimensão, o objetivo foi desenvolver uma solução alinhada com os princípios da *Web 2.0* como é descrito a seguir.

Ter a **Web como plataforma** (princípio 1) foi uma premissa de partida do projeto que também deveria poder ser acessado de **qualquer plataforma de computador** (princípio 6). Optamos em definir um *framework Web* como plataforma de desenvolvimento. Após uma extensa pesquisa, foi escolhido como plataforma de desenvolvimento, o *software* livre *Wordpress* que utiliza as linguagens PHP, HTML, CSS e *Javascript* e roda independente do sistema operacional (*Windows*, *Linux*, OS X) geralmente utilizando o Apache como servidor *Web* e o *MySQL* ou *MariaDB* como servidor de banco de dados. O *Wordpress* tem uma ampla comunidade de usuários e desenvolvedores, inclusive no Brasil e gerencia de forma nativa (não é necessário implementar): páginas, links, artigos, termos, etiquetas, categorias, usuários e perfis de usuários, o que economiza desenvolvimento no projeto. Além disso, possui no repositório de sua comunidade mais de 40 mil extensões, como o *plugin* de *login* através do usuário do *facebook* ou *backup* com agendamento através de FTP. Essas funcionalidades e muitas outras, podem ser acrescentadas a solução sem nenhum ou com muito pouco esforço de implementação.

Utilizar **modelos leves de programação** (princípio 5) também foi um ponto estruturante no projeto. O *Tainacan* utiliza de forma intensa a técnica AJAX, por meio do *framework* para *Javascript*, *Jquery*, que em conjunto com a biblioteca para CSS, o *Bootstrap*, trabalham integrados com o *backend* do *Wordpress*, propiciando uma alta produtividade no desenvolvimento. O *design patterns* utilizado é o MVC (*Model-View-Controller*), arquitetura amplamente difundida na *Web*, faz uma separação das regras de negócio, modelo de dados e apresentação, e permite o reuso das classes e métodos e facilita a manutenção do código.

Também foi utilizado no *Tainacan*, um esquema de dados inspirado no padrão RDF com triplas que armazenam os metadados. Isto torna a aplicação bastante flexível para suportar novas entidades, sem a necessidade de alteração do esquema de dados. O *Tainacan* em si, é uma extensão do *Wordpress*. Como existe a necessidade de controlar toda a experiência de usuário, ao invés de

construir um *plug-in* foi desenvolvido um tema, que quando ativado acrescenta no *Wordpress*, as funções requeridas para uma biblioteca digital. Na arquitetura desse tema foi previsto a possibilidade de estender as funções do *Tainacan* através de *plug-ins* da mesma forma que ocorre no *Wordpress*.

O princípio 4, ser um **beta perpétuo**, geralmente se aplica aos serviços na *Web*. O *Tainacan* foi projetado para ser tanto um produto, um *software* livre que pode ser baixado e usado localmente, quanto ser um serviço *Web*, para hospedar e estimular a criação de bibliotecas digitais. Desse modo, foi definido que a implementação deveria ser espiral e incremental. O processo de desenvolvimento de *software* foi inspirado no *Scrum*, com ciclos mensais (*sprints*), que resultavam em uma nova versão. Foi utilizado elementos do RUP (*Rational Unified Process*) com a elaboração de casos de uso, prototipação de baixo nível e plano de testes. Também foram realizadas periodicamente oficinas ao longo do projeto, no qual foi possível colher dos usuários um *feedback* sobre o que foi implementado, gerando conhecimento e realimentando o próximo ciclo de desenvolvimento.

O segundo princípio, uso da **inteligência coletiva**, está bastante ligado ao terceiro princípio, no qual os **dados** ou o conteúdo inserido/ofertado aos usuários, é o elemento mais importante. Destaca-se as seguintes funcionalidades implementadas para o atendimento desses princípios:

- Etiquetas e nuvem de *tags*;
- Votar em itens de uma coleção;
- Escolher tipo de votação: curtir, gostei ou não gostei, estrelas;
- Ter múltiplas votações por coleção;
- Redes sociais:
  - Compartilhar item em redes sociais (*Facebook, Twitter, Google++*);
  - Compartilhar coleção em redes sociais (*Facebook, Twitter, Google++*);
- Comentários:
  - Comentar sobre o item;
  - Comentar sobre a coleção;
  - Comentar sobre a categoria;
  - Comentar sobre o repositório;
- Configuração de ação pré-aprovada (*wiki*) para todos os eventos;
- *Workflow* de aprovação pelo administrador (especialista) para todos os eventos;
- *Workflow* de aprovação por votação de usuários para todos os eventos;
- Eventos com *workflow*:
  - Adicionar coleção em comunidade de terceiros;
  - Remover (denunciar) coleção de terceiros;
  - Adicionar item em coleção de terceiros;
  - Remover (denunciar) item de terceiros;

- Adicionar classificação em itens de terceiros;
- Editar classificação em itens de terceiros;
- Excluir classificação em itens de terceiros;
- Adicionar categoria em coleção de terceiros;
- Editar categoria em coleção de terceiros;
- Excluir categoria em coleção de terceiros;
- Adicionar *tag* em coleção de terceiros;
- Editar *tag* em coleção de terceiros;
- Excluir *tag* em coleção de terceiros;
- Adicionar comentário em coleção de terceiros;
- Editar comentário em coleção de terceiros;
- Excluir comentário em coleção de terceiros;
- Adicionar comentário em item de terceiros;
- Editar comentário em item de terceiros;
- Excluir comentário em item de terceiros;
- Adicionar metadado de dado em item de terceiros;
- Editar metadado de dado em item de terceiros;
- Excluir metadado de dado em item de terceiros;
- Adicionar metadado de relacionamento em item de terceiros;
- Editar metadado de relacionamento em item de terceiros;
- Excluir metadado de relacionamento em item de terceiros;
- Adicionar metadado de termo em item de terceiros;
- Editar metadado de termo em item de terceiros;
- Excluir metadado de termo em item de terceiros;
- Editar relacionamento em metadado de relacionamento em item de terceiros;
- Editar valor em metadado de dado em item de terceiros;
- Coleção:
  - Permite homepage da coleção personalizada com miniatura e capa;
  - Moderadores para a coleção;
  - Privacidade da coleção;
  - Filtros para a coleção;
- Conectar com usuário do *facebook* e g+;
- Gestão colaborativa de taxonomias:
  - Categorias privadas;
  - Categorias compartilhadas;
  - Categorias públicas;
- Gestão de licenças e disponibilidade de licenças CC (*Creative Commons*) por padrão;
- Facetas do tipo:
  - Árvore;
  - Menu;
  - Botão rádio;
  - Caixa de seleção;
  - Seleção múltipla;
  - Faixa de valores;

- Definição de esquema de metadados:
  - Suporte ao esquema *Dublin Core* e outros;
  - Relacionamentos;
  - Customização de esquema de metadados;
  - Configurar valores padrão e mensagem de ajuda para campos de metadados;
- Interoperabilidade:
  - Importar/exportar OAI-PMH;
    - Mapeamento dos metadados importados e exportados;
    - Entrada automática de metadados (*harvesting*);
  - Importar/exportar CSV;
    - Importação e exportação de metadados da coleção;
    - Mapeamento dos metadados importados e exportados;
    - Exportação dos resultados de busca;
  - API JSON;
  - Exportar em RDF e gerar Grafo;
  - Importar Recurso *Web URL*:
    - Importar vídeos do *Youtube* e *Vimeo* via URL;
    - Importar canais de vídeos do *Youtube* e *Vimeo* via URL;
    - Importar imagens e arquivos (JPG, PNG, GIF, PDF);
    - Importar uma página Web (ex. notícia);
    - Importar fotos do Instagram, *Facebook* e *Flickr*;
  - Redes sociais:
    - Mapeamento dos metadados dos itens importados das redes sociais (*Youtube*, *Facebook*, *Instagram*, *Vimeo*);
  - Importar automaticamente vídeos, fotos e itens de repositórios OAI-PMH.
  - RSS;

O princípio 7, ter uma **interface de usuário rica**, aumenta a eficiência da interface e a satisfação dos usuários, estimulando a captura da inteligência coletiva e dados dos usuários. Uma interface rica evita recarregamentos da página e deixa a interação mais fluida como o Gmail ou *Google Agenda* realizam por exemplo. Assim, a interface usa muito AJAX, de modo que atender esse princípio foi um investimento significativo no projeto, que passou por um *re-design* completo. Destaca-se as seguintes funcionalidades implementadas para o atendimento desse princípio:

1) Gestão *drag n' drop* de metadados e filtros: essa interface permite ao usuário gerenciar o esquema de metadados adotado na coleção, permitindo-o customizá-lo, acrescentando novos metadados e definindo através do recurso de arrastar e soltar, a ordem desses campos no formulário de adicionar item e também se o metadado será usado como filtro na interface de busca da coleção. Foi de longe a interface mais complexa de idealizar e de implementar. No *Dspace* para alterar o tipo de *widget* no formulário de adicionar item ou colocar um novo filtro na interface de

busca, exige conhecimentos em programação para alterar o código de um arquivo XML. No *Tainacan* isso é feito através de uma interface gráfica, basta arrastar e soltar.

2) Adicionar múltiplos itens: a interface de envio de múltiplos itens visa uma economia de tempo na catalogação de itens no *Tainacan*. Enquanto, no *Dspace* é necessário no mínimo passar por 6 etapas (podendo chegar a 11 etapas) para inserir um único item, no *Tainacan* é possível inserir múltiplos itens com apenas 2 passos (envio e descrição). Na etapa de descrição, o usuário consegue atribuir a múltiplos itens a mesma descrição, como uma *tag* ou a data de criação. Isso pode ser muito útil quando se tem vários objetos com uma parte dos metadados iguais, como fotos de um evento, por exemplo, que compartilharão o mesmo autor, data, assunto, etc. Essa interface foi construída inspirada no site de compartilhamento de imagens *Flickr*.

3) Classificação através de *drag n' drop*: é possível durante a filtragem de itens da coleção, arrastar a categoria ou *tag* que fica no menu esquerdo para dentro de um item, classificando-o.

4) Navegação e busca facetada AJAX: a pesquisa de itens no *Tainacan* ocorre sem dar *refresh* na página. O usuário pode combinar a seleção de filtros no *menu* esquerdo com a digitação de palavra-chave na caixa busca para recuperar os itens da coleção. O usuário pode navegar em uma hierarquia de categorias e adicionar ou editar categorias, atualizando sem interferir no estado da interface (se ele estiver filtrando ou editando um item, essa ação não é interrompida).

5) Autocompletar na busca e na catalogação dos itens: na caixa de busca da coleção tem um autocompletar de categorias, etiquetas e dos títulos dos itens. Assim, se o usuário digitar, o autocompletar mostra a faceta origem e a categoria, e selecionando-a, são exibidos os itens encontrados abaixo. O usuário também utiliza o autocompletar na catalogação, assim, ao preencher um determinado campo de metadado, o sistema exibe uma lista de valores para aquele metadado, baseado no preenchimento dos itens anteriores. Isso pode economizar tempo na catalogação, algo desejável principalmente em coleções com muitos itens para cadastrar.

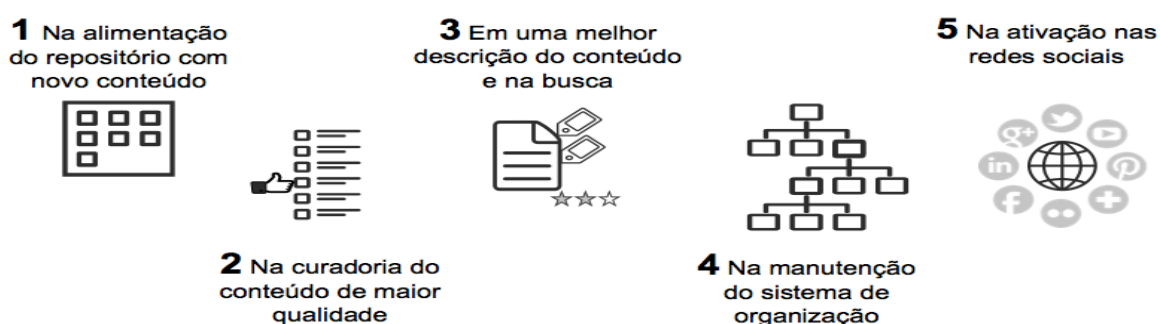
6) *Wizard*: Ao criar uma coleção, o usuário passa por uma sequência de etapas que visam guiá-lo na construção de uma coleção, começando por configurações gerais, depois categorias, esquema de metadados e *layout*.

7) Ter *design* responsivo: a coleção, a listagem de itens, a visualização do item foram tratadas para permitir o acesso por qualquer tipo de dispositivo. A navegação facetada foi adequada para a interface de celulares permitindo a filtragem de itens e seu compartilhamento em redes sociais com poucos cliques na tela.

A colaboração é o grande diferencial do *Tainacan*. Ela foi implementada em 5 dimensões: 1) Na alimentação do repositório com novos conteúdos; 2) Na avaliação dos itens da coleção, evidenciado aquilo que a comunidade de usuários considera ter maior qualidade; 3) Em uma

melhor descrição dos itens, refinando os metadados, editando ou eliminando conteúdo irrelevante ou inapropriado; 4) Na construção do sistema de organização, que são as taxonomias usadas nas coleção, sugerindo novas categorias ou editando as existentes. O esquema de metadados também pode ser aprimorado; 5) Na importação de conteúdo das redes sociais e atribuição de um novo significado a esse objeto dentro da coleção, organizando e descrevendo com uma capacidade não ofertada nas redes sociais. Além disso, é permitido compartilhar em diversas redes sociais, gerando uma maior visibilidade do conteúdo da biblioteca digital. A Figura 5 sintetiza essas dimensões presentes na solução desenvolvida:

FIGURA 5 – Formas de colaboração no *Tainacan*



Fonte: Elaborado pelo autor.

## 5 CONSIDERAÇÕES FINAIS

O sistema de gestão de biblioteca digital, *Tainacan*, construiu seu caminho apoiado em referências internacionais da área de biblioteca digital e nos princípios da *Web 2.0*. Apresenta recursos avançados que o diferencia das demais soluções, com foco na colaboração e integração com redes sociais. O projeto oferece uma alternativa para envolver a comunidade usuários na administração, alimentação, curadoria e organização do conteúdo da biblioteca digital.

Em trabalhos futuros, pretende-se avaliar se o modelo de colaboração para bibliotecas digitais disponível no *Tainacan* é escalável, realizando testes empíricos. Através de projetos pilotos em instituições com demandas nessa área, planeja-se executar testes de usabilidade, questionários de satisfação, grupo focal, e muitos outros experimentos para validar as funcionalidades implementadas e descobrir deficiências, realimentando os próximos ciclos de desenvolvimento.

O impacto dos princípios da *Web 2.0* na preservação digital também é um projeto de pesquisa planejado. Quais são os tipos de interações que devem ser consideradas partes do objeto e também serem preservadas? Se o ideal for preservar todo o contexto do item, isto pode incluir, compartilhamentos, notas, curtidas, comentários, histórico de ações, curtidas nos comentários,

nome dos usuários que interagiram com o objeto, entre outras interações possíveis. Na prática o resultado desse estudo futuro é propor uma nova forma de geração do API (*Archival Information Package*) que inclua os metadados sociais do objeto.

## REFERÊNCIAS

BLATTMANN, Ursula; SILVA, Fabiano Couto Corrêa da. **Colaboração e interação na Web 2.0 e biblioteca 2.0**. 2007.

CANDELA, Leonardo et al. Setting the foundations of digital libraries: The delos manifesto. **D-Lib Magazine**, v. 13, n. 3, p. 4, 2007.

CIRILO, C. E. **Model Driven Richubi**: processo dirigido a modelos para a construção de interfaces ricas de aplicações ubíquas sensíveis ao contexto. 2011. 203f. Dissertação (Mestrado em Ciência da Computação) - Centro de Ciências Exatas e de Tecnologia, Universidade Federal de São Carlos, 2011.

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. **Dicionário de biblioteconomia e arquivologia**. Briquet de Lemos/Livros, 2008.

DIGITAL LIBRARY FEDERATION. **A working definition of digital library**. 1998. Disponível em: <<https://old.diglib.org/about/dldefinition.htm>>. Acesso em: 15 abr. 2016.

FREW, James et al. The Alexandria digital library architecture. In: **Research and Advanced Technology for Digital Libraries**. Springer Berlin Heidelberg, 1998. p. 61-73.

PINHEIRO, Lena Vania ribeiro; FERREZ, Helena Dodd. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia. 2014. Disponível em: <[http://www.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/copy\\_of\\_TESAURO\\_COMPLETOFINALCOMCAPA24102014.pdf](http://www.ibict.br/publicacoes-e-institucionais/tesouro-brasileiro-de-ciencia-da-informacao-1/copy_of_TESAURO_COMPLETOFINALCOMCAPA24102014.pdf)>. Acesso em: 15 jun. 2015.

PONTES, F. V. **Organização do conhecimento em bibliotecas digitais de teses e dissertações**: uma abordagem baseada na classificação facetada e taxonomias dinâmicas. 2013. 233f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

O'REILLY, Tim. **What is Web 2.0**: design Patterns and Business Models for the Next Generation of Software. 30 Sep. 2005. Disponível em: <<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>>. Acesso em: 11 abr. 2016.

SAYÃO, Luis Fernando. Uma arquitetura genérica para sistemas de biblioteca digital como pretexto para a criação de uma agenda de pesquisa. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, Brasília, v. 2, n. 1, p. 173-198, jan./dez. 2009. Disponível em: <[www.brapci.ufpr.br/download.php?dd0=7876](http://www.brapci.ufpr.br/download.php?dd0=7876)>. Acesso em: 6 abr. 2016.

SILVA, Marcel Ferrante. **Proposta de modelo de colaboração para catálogo web facetado**. Belo Horizonte, 2013. 269f. Tese (Doutorado em Ciência da Informação)- Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

THOMAZ, Kátia de Pádua; SOARES, Antonio José. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **DataGramZero**, Rio de Janeiro, v. 5, n. 1, 2004.



## **Provocações acadêmicas: ontologias, tesouros, documentos, conteúdo de documentos, e... Unicórnios**

*AN ACADEMIC GIVE-AND-TAKE: ONTOLOGIES, THESAURI, DOCUMENTS, DOCUMENT CONTENTS AND ... UNICORNS*

Maurício B. Almeida (ECI/UFMG)

*mba@eci.ufmg.br*

**Resumo:** Uma provocação nietzschiana sobre os valores da vida questiona: “qual o valor nos fez escolher os valores que seguimos?”. Na vida acadêmica, tal provocação deveria ser regra: questionar e questionar o que fazemos e o que nos levou a fazer o que fazemos, esse é o papel do pesquisador. É nesse espírito de provocação saudável que se apresenta esse pequeno ensaio no contexto da Organização do Conhecimento. São apresentadas questões que ainda surgem a toda hora no campo de pesquisa. Como provocação assumida que é, não tem pretensão de estabelecer a verdade, mas fomentar o debate acadêmico tão importante para o progresso da pesquisa.

**Palavras-chave:** Ontologia Aplicada. Organização do Conhecimento.

**Abstract:** A Nietzschean provocation about the values of life argues: "which value made us choose the values we follow?" In the academic life, this provocation should be a rule: to question and question again what we do and what led us to do what we do, this should be the role of a researcher. It is within this spirit of a healthy give-and-take that we present this little essay in the context of Knowledge Organization field. We present issues that still arise several times in our research field. As this essay is an assumed provocation, we do not have in mind to establish the truth, but only to foster the academic debate as important tool for the research progress.

**Keywords:** Applied Ontology. Knowledge Organization.

### **1 INTRODUÇÃO**

Tradicionalmente fundamentada em técnicas e teorias milenares de classificação, a Organização do Conhecimento (OC) é um campo vibrante e dinâmico da Ciência da Informação, essencial para a ciência e mesmo para a vida cotidiana.

Esse protagonismo da OC se consubstancia nas habilidades dos profissionais ali formados em representar, organizar e recuperar o conhecimento produzido nas mais diversas áreas do pensamento.

Recentemente, passaram a compor a pesquisa em OC, técnicas provenientes da disciplina da Ontologia Aplicada, a qual traz novos *insights* filosóficos visando a manipulação automática do conhecimento registrado em meio digital. Nesse ensaio, baseado em indagações sobre a relação entre ontologias e conceitos seminais da OC já apresentadas em artigos anteriores, discutem-se questões simples, ainda que interessantes e provocativas.

Por limitações de espaço, não se apresenta o *background* necessário sobre ontologias e OC, considerando que o leitor é minimamente iniciado nesses assuntos. O restante do presente ensaio está organizado em cinco seções: as três primeiras são encabeçadas por perguntas provocativas, a quarta mostrar erros de classificação, e a última traz considerações finais.

## 2 ONTOLOGIAS E LINGUAGENS DOCUMENTÁRIAS

Duas questões relacionadas compõem a pergunta dessa seção, à saber: (a) ontologias são linguagens documentárias? (b) se não são podem ser usadas como tal?

Uma resposta à primeira parte da pergunta, à questão (a), a resposta mais rigorosa é “não”, ontologias não são linguagens documentárias (LDs). Ontologias são criadas como teorias científicas formais, e não como tipos linguagens, apesar de fazerem uso de alguma. Um renomado filósofo explica na primeira frase de seu livro: “*Esse livro é um livro sobre o mundo. Estou preocupado com ontologia, não com linguagem*” (JOHANSSON, 2004, p. 1).

O aviso é claro: existe uma diferença entre as entidades do mundo, por si, e os meios que as pessoas usam para descrever essas mesmas entidades. A ontologia se refere às entidades do mundo, não à convenção usada para se referir as entidades, ou seja, à linguagem.

A resposta para a segunda parte da pergunta, a questão (b), é um “sim”, ou seja, mesmo que ontologias não sejam linguagens documentárias, podem ser usadas como tal. Isso é verdadeiro caso esteja-se referindo ao uso de linguagens de representação adotadas para especificar ontologias, como OWL ou RDFS. O resultado nesse caso será uma LD implementada em um sistema de recuperação da informação da Web, não uma ontologia.

## 3 ONTOLOGIAS E TESAUROS

Assim como na seção 1, duas questões relacionadas compõem a pergunta dessa seção, à saber: (a) ontologias e tesauros exibem a mesma capacidade de representação? (b) se não exibem, onde residem as principais diferenças?

Já é amplamente sabido que os dois instrumentos mencionados nas questões não são diretamente comparáveis, uma vez que servem a objetivos distintos. Ao criar tesauros, empregam-se relações *broader-than* e *narrower-than* para organizar taxonomicamente o vocabulário. Essas relações surgem, por exemplo, em vocabulários controlados como o *Medical Subject Headings* (MeSH)<sup>9</sup>, criado para indexação e catalogação de informação médica. São exemplos do MeSH: i) *FetalBlood narrower-than Blood*; e ii) *Plasma narrower-than Blood*.

Do ponto de vista dos tesauros, essas relações atendem as necessidades de recuperação de documentos: uma consulta usando o termo *blood* retornará tanto artigos sobre *fetal\_blood*, quanto sobre *blood\_plasma*. Entretanto, do ponto de vista ontológico, as duas relações representam tipos

---

<sup>9</sup> <https://www.nlm.nih.gov/mesh/>

diferentes: enquanto na primeira relação *fetal\_blood* é um tipo de *blood*, na segunda *plasma* é parte de *blood*. Essas diferenças são importantes porque enquanto tesouros são criados para uso por pessoas, ontologias são criadas para consumo por máquinas.

#### 4 ONTOLOGIAS, DOCUMENTOS E SEU CONTEÚDO

Nessa seção, a pergunta que norteia a discussão é única: se ontologias são uma descrição de mundo, em que local do mundo se localizam documentos e seu conteúdo?

A parte da pergunta sobre documentos como entidades físicas, não apresenta grandes desafios e, portanto, será omitida aqui por limitações de espaço. A parte da pergunta sobre onde se localiza o conteúdo de documento em uma ontologia não é trivial e nem ontologias como a *Basic Formal Ontology* (BFO) (GRENON; SMITH, 2004) ou a *Information Artifact Ontology* (IAO) (IAO, 2013) se arriscam a definir o conteúdo de documentos.

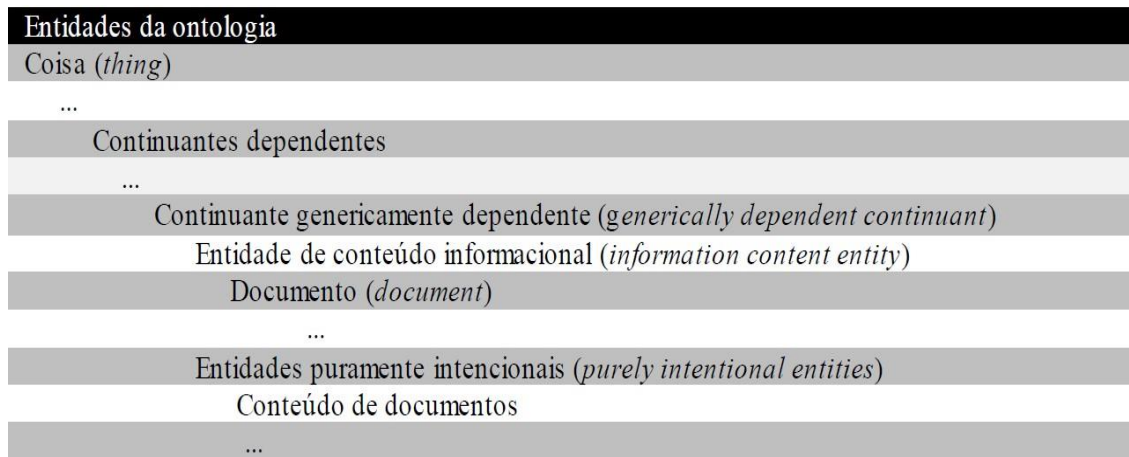
A questão reside no fato de que, no âmbito da IAO e BFO, uma *entidade de conteúdo informacional* (nível 6 da FIG. 1) sempre é sobre (*is\_about*) alguma entidade do mundo, considerado o mundo espaço-temporal. Entretanto, é fácil imaginar livros cujo conteúdo versa sobre unicórnios, coelhos da páscoa, dentre outras ficções, entidades que não existem no mundo espaço-temporal. Para lidar com essa situação, Brochhausen et al. (2013), propõem uma entidade chamada *continuante genericamente dependente*<sup>10</sup>, uma irmã da *entidade de conteúdo informacional*<sup>11</sup>, denominada *entidade puramente intencional*. Abaixo dessa entidade, é possível abrigar a entidade *conteúdo de documentos* que se refere tanto a coisas do mundo espaço-temporal, quanto coisas imaginadas sobre o mundo.

---

<sup>10</sup> Na BFO, continuantes genericamente dependentes são propriedades de outra entidade, mas tal entidade pode variar.

<sup>11</sup> Na IAO, entidades de conteúdo informacional dependem genericamente de outras e mantêm relações de *aboutness*.

FIGURA 1 – Fragmento de hierarquia com entidades “Documento” e “Conteúdo”



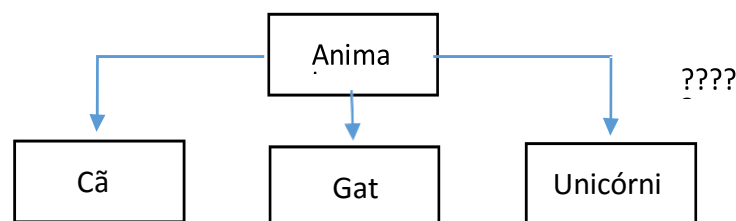
Fonte: ALMEIDA; MENDONÇA; AGANETTE, 2013, p. 15.

## 5 CLASSIFICAÇÃO DE UNICÓRNIOS, CIRURGIAS, PESSOAS, ...

Apresentou-se a localização de entidades ficcionais na ontologia, as quais poderiam ser classificadas abaixo de *entidades puramente intencionais* caso se deseje representar um conteúdo de um livro que tem como assunto unicórnios, por exemplo. Entretanto, ainda cabe questionar onde classificar tais entidades em um sistema de classificação. *Será um unicórnio um animal? Um coelho da páscoa é um coelho? Mudando um pouco o foco dos exemplos: uma cirurgia cancelada é uma cirurgia? Um diretor é uma pessoa?*

Não é difícil mostrar que a classificação desses casos em geral leva a erros básicos e podem levar à problemas em sistemas ontológicos capazes de inferência. A FIG. 2 apresenta uma hierarquia para “*unicórnio é um Animal*”.

FIGURA 2 – Classificação de animais (setas são relação é-um)



Fonte: Elaborada pelo autor.

A ontologia contém entidades – os *continuantes* – as quais existem totalmente durante todo tempo de sua existência, persistem através do tempo mantendo identidade. Exemplos são: um coração, um animal, a cor do tomate, uma orquestra, a disposição da banana em apodrecer, etc. De fato, animais como cães e gatos são entidades chamadas nas ontologias de *continuantes independentes*: entidades portadoras de qualidades, das quais outra entidade pode depender ou ser

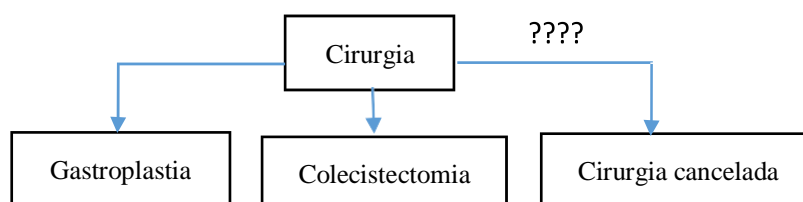
inerente a (a cor vermelha é inerente ao tomate), e que não podem depender de mais nada. Em suma, são entidades independentem da mente humana para existir.

Por outro lado, os unicórnios são entidades nunca vistas, das quais a ciência nada tem a dizer. Dinossauros também são entidades nunca vistas, mas, ao contrário dos unicórnios, a ciência comprova sua existência. O que ocorre é que unicórnios existem apenas na mente das pessoas, não são *continuanes independentes*, são, de fato, *continuanes dependentes*. Um *continuanes dependente* é uma entidade que mantém inerência (ou dependência) de outras entidades, por exemplo: o peso do José não existe sem José; a brancura do queijo depende do queijo; o pensamento de Einstein depende da existência de Einstein. Da mesma forma, unicórnios dependem da mente de alguém que os imagina, e, portanto, não são objetos nem animais, são “propriedades”. Nas ontologias de orientação aristotélica, as propriedades são chamadas *qualidades*. Sendo unicórnios qualidades, a classificação da FIG. 2 é errônea.

O caso do “*coelho da Páscoa é um coelho*” segue um raciocínio similar ao do Unicórnio. O exemplo apenas mostra outra entidade fictícia que pode carecer de classificação, sem que esteja embutida em um assunto ou conteúdo de um livro.

O caso da *Cirurgia Cancelada é um Cirurgia* é diferente, mas ilustra um erro de classificação em que não há preocupação em definir adequadamente a entidade antes localizá-la no sistema de classificação.

FIGURA 3 – Classificação de cirurgias (setas são relação é-um)



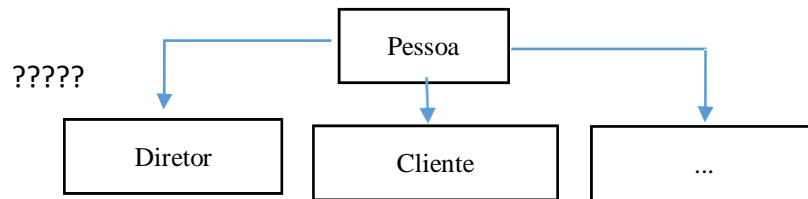
Fonte: Elaborada pelo autor.

Em ontologia existem entidades que “ocorrem” no tempo, e, portanto, denominadas *ocorrentes*. Ocorrentes se desdobram ao longo do tempo e os exemplos prototípicos dessas entidades são os processos: o processo de respiração, o processo de matrícula, etc. Uma cirurgia é um processo e, além de se desenrolar ao longo do tempo, exibe uma característica ontológica básica dos processos: a existência de um processo, requer a existência de um *continuanes independente* que participe do processo. De fato, não existe processo de matrícula sem o aluno; não existe digestão sem a comida; e não existe cirurgia sem paciente. Como uma cirurgia cancelada não possui paciente, ela não é processo e, portanto, uma cirurgia cancelada não é uma

cirurgia. A cirurgia cancelada é, na verdade, um “plano”. Planos, em ontologias, são *entidades de conteúdo informacional*, as quais dependem genericamente de outra entidade, além de manter uma relação de *aboutness* com outras. A classificação apresentada na FIG. 3 é errônea.

O caso “*diretor é-um pessoa*” é um erro comum cometido na modelagem de sistemas de informação e bancos de dados, por profissionais provenientes da informática.

FIGURA 4 – Classificação de pessoas (setas são relação é-um)



Fonte: Elaborada pelo autor.

A FIG. 4 exibe um erro comum: um diretor não é uma pessoa, um diretor é um *papel* que uma pessoa assume durante um período. Papéis são definidos em ontologias como *continuantes dependentes*, os quais não são essenciais para a identidade do continuante.

A consequência desse tipo de erro para sistemas baseados em ontologias é que podem-se obter inferências automáticas totalmente sem sentido. Considere-se, como exemplo, o papel de ser diretor, que é um papel; e considere-se também o papel de um coração artificial em bombear sangue, o qual também é um papel. Sendo esses dois tipos de papéis considerados, e na existência de uma classificação errônea a da FIG. 4, um motor de inferência poderia, hipoteticamente, gerar o fato de que o papel de bombear sangue é papel de uma pessoa. Isso certamente não faz sentido. Esse é um exemplo simples, dentre diversos outros possíveis, que não são apresentados aqui por limitações de espaço.

## 6 CONSIDERAÇÕES FINAIS

O que se pretendeu mostrar ao longo desse ensaio-provocação é a necessidade e entender a origem das entidades sob classificação para classificá-las, independentemente do tipo de sistema de organização do conhecimento. A teoria da Ontologia Aplicada pode ajudar em muito nesse aspecto.

Ainda, cabe destacar que não apenas por se usar um editor de ontologias para construir um sistema de organização de conhecimento, está se garantido que tal sistema seja de fato uma ontologia. Assim como carros, pessoas, árvores, etc., todas as entidades do mundo têm características que as determinam como tal; também as ontologias têm características que precisam

estar presentes para a estrutura seja de fato uma ontologia. É também comum ouvir que inferências não são necessárias, mas parece contraditório se dispender tanto esforço na construção de ontologias e desprezar sua principal utilidade.

Muitos dos enganos cometidos, como os exemplos apresentados aqui, se devem ao desconhecimento ou falta de crença na necessidade de uso de regras de classificação formais. Fica a consideração de que, se as regras de classificação não precisam ser seguidas, então o resultado pode não ser considerado um sistema de classificação.

## REFERÊNCIAS

ALMEIDA, M. B. Revisiting ontologies: a necessary clarification. **Journal of the American Society of Information Science and Technology**, v. 64, n. 8, p. 1682-93, 2013.

ALMEIDA, M. B.; MENDONÇA, F. M.; AGANETTE, E. C. Interfaces entre ontologias e conceitos seminais da Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis: ANCIB, 2013.

BROCHHAUSEN, M. et al. Developing a semantically rich ontology for the biobank administration domain. **Journal of Biomedical Semantics**, v. 4, n. 23. 2013.

GRENON, P.; SMITH, B. **SNAP and SPAN**. 2004. Disponível em: <<http://ontology.buffalo.edu/smith/articles/>>. Acesso em: 10 de jun. de 2013.

IAO: Information Artifact Ontology. 2013. Disponível em: <<http://code.google.com/p/information-artifact-ontology/>>. Acesso em: 16 de jun. 2013.

JOHANSSON, I. **Ontological investigations**. Frankfurt: Ontos Verlag, 2004.

## **A dimensão interdisciplinar nas pesquisas em organização do conhecimento e recuperação da informação**

### *USING ARTIFICIAL INTELLIGENCE TO IDENTIFY STATE SECRETS*

Renato R. Souza (EMAP/FGV-RJ)

*rsouza.fgv@gmail.com*

Flávio Codeço Coelho (EMAP, FGV)

Matthew Connelly (Department of History, Columbia University)

**Abstract:** The challenge of protecting sensitive information in large electronic records systems is a matter of growing public controversy. We report the results of using Natural Language Processing and Machine Learning to identify State Department cables that were originally classified as containing official secrets. By using metadata such as the subject line and sender/recipient information along with the word frequencies in the text we were able to train algorithms to identify 90% of classified cables with <11% false positives. Closer analysis revealed many examples of human error in failing to appropriately classify these communications. The results indicate both the need for research on inter-coder reliability among government officials and the opportunity to develop recommender systems to better manage both classification and declassification.

**Keywords:** Automatic classification. Natural language processing. Machine learning. Sensitive information.

One hundred countries have adopted laws or administrative measures giving citizens a right to information about what governments do in their name. (1) But certain categories are typically excluded, whether for national security or personal privacy. Distinguishing between what citizens are entitled to know and what government officials are obliged to withhold is therefore a matter of growing political importance.

The Freedom of Information Act (FOIA) request that revealed former Secretary of State Hillary Clinton's private e-mail server has put this challenge in stark relief. Whereas critics insist she must have recognized that her communications included sensitive information, Clinton's defenders argue that what is classified is "almost random." (2) It has taken more than eight months for the State Department to review (and redact) the 31,000 emails she turned over to authorities. It is estimated that the State Department generates two billion e-mail every single year, (3) and current declassification methods have been judged completely inadequate to cope with this volume of sensitive information (4).

Is it possible to identify sensitive information automatically, and develop systems that might help officials manage both classification and declassification more rapidly and more



reliably? The release of the first generation of U.S. government electronic records presents new opportunities to analyze this problem using well-developed methods from Natural Language Processing (NLP) and machine-learning. We report on the performance of algorithms to identify diplomatic cables from the 1970s that were originally classified as “secret,” “confidential,” or “limited official use.” Cables were the main form of secure communications in this period. And since the metadata include not only the classification level, but also the sender, recipient, date, and subject, these cables – and research analyzing these cables – are quite relevant to management challenges for newer types of communications, notably email. They can show whether classification has historically been random or predictable, and analyzing the errors machines make in identifying sensitive information can help us understand the errors that humans make.

Human error in managing sensitive information includes the loss of that information. Of the 1,758,279 declassified cable records available for 1973-78, there is no message text for 128,026 (Table 1). Intriguingly, messages classified as “Secret” were more than three times more likely to go missing compared to Unclassified and Limited Official Use messages (22% versus 6.5%). Archivists deleted hundreds of thousands more cables after determining they lacked sufficient historical significance based on relatively crude sampling (5). Our dataset also did not include 410,539 “withdrawn” cables where only the metadata has been declassified, or “Top Secret” cables, which are only available in paper form, and that only for one year (1975). But relatively few cables are Top Secret, and the great majority (~TK%) of the withdrawn cables were never classified, but are likely withheld because they contain personally identifiable information. Finally, we excluded cables that had no text or “n/a” for any of the features we used in building the classifier.

In preparing the remaining 981,083 cables, our goal was to replicate “real world” conditions, i.e. only using data that would be available to a human – or a computer – to identify a cable as containing classified information before it was sent. We carried out many standard NLP operations, such as eliminating hyphenation, tokenizing (separating words and discarding punctuation), and removing common, non-informative words such as “and”, “the”, etc.

These steps are described in the supplementary materials.

Our task was to use features from the metadata and the message text to predict whether a given State Department cable would belong to the classified or unclassified categories. Ironically, in the field of machine learning and statistics this is known as a classification task. In essence, we take a set of observations in which the category membership is known to create a training set. We then use this data to train a statistical model that scores the probability of new observations belonging to the same category.

We have used cross validation to test the performance of different models and avoid over-fitting, i.e. choosing a model that would perform poorly on data outside the sample. This entailed setting aside a random selection of cables and -- after training our classifier with the rest of the data -- testing the accuracy of its predictions for the randomly selected cables.

For the classifiers we have used features vectors constructed under the Bag of Words approach (6). After comparing performance, we found word count vectors worked as well or better than Term Frequency – Inverse Document Frequency (TfIDf) or Word2Vec vectors. In addition to the message text (the body field), we found the following features most useful: 1) sender/recipient, a combination of the “from” and “to” fields, which typically is an embassy or office; 2) concepts, keywords to categorize each document; 3) subject, a short description; 4) office, the part of the State Department with responsibility for either originating or acting on a cable; 5) TAGS, which stands for Traffic Analysis by Geography and Subject. The TAGS field was reserved for one or more acronyms selected from a standard list that best described the content of each cable, whether in terms of country or topic.

The four original classes – (S)ecret; (C)onfidential; (L)imited Official Use; and (U)nclassified – indicate the degree of sensitivity. We grouped them to measure performance depending on how broadly or how narrowly one defines a state secret. Rather than using just one classifier for this task, we built an ensemble of the seven best-performing algorithms (see supplemental material) using different weights for each. This can be likened to assembling experts and giving them more or less weight depending on their comparative reliability.

Of all the features, the relative frequency of different words in the body was the most useful in identifying sensitive information. High recall or precision (but not both) are achievable with some of the other features. For instance, it is possible to identify 95% of the cables that are classified as Secret, Confidential, or Limited Official Use just by using the sender/recipient data. But fully a third of the identified cables would be false positives. The combination produces a relatively low (.57) Area Under Curve score, a common performance metric. Other features produce better overall performance (see Fig. 2) but the best results came from combining all six features.

Table 2 shows the mean scores for precision and recall and several measures of accuracy when we categorize sensitive information more or less broadly. There is a notable improvement in performance when we exclude data that does not clearly belong to either category. The “Limited Official Use” designation has long been a cause for complaint. White House Executive Orders from the period classify as “Secret” information that “could reasonably be expected to cause serious damage to the national security,” and include specific examples. “Confidential” was the

classification for information that could cause “damage” but not “serious damage” (7). Limited Official Use, on the other hand, had no such definition. When the George W. Bush administration attempted such a definition, the Government’s leadign FOIA expert admitted it was “so broad that it is almost harder to think information that does not fall within its scope than information that does” (8). This is consistent with our inability to distinguish these cables despite using many different features and multiple classifiers. It is not clear whether any NLP or machine learning methods would accurately distinguish this category.

But the results also show there is some logic in official secrecy, at least to the extent that State Department cables from the 1970s are indicative. Even when we include Limited Official Use cables, just a few kinds of metadata and the relative frequency of words in the message text are enough to identify 90% of classified cables with relatively few (<11%) false positives. Whether a diplomatic communication should be classified may therefore be relatively predictable -- more so than, say, predicting heart disease by data-mining medical records (9).

If we were able to use all the diplomatic cables from the period, including both those that were lost and those that are still withheld, we would likely achieve higher accuracy. But even with all of the data, these methods will sometimes miss crucial contextual elements, such as whether the subject of a seemingly innocuous travel reservation or visa application was on a sensitive mission. Other errors might instead reflect the intrinsic subjectivity of at least some classification decisions, especially considering that officials often receive inconsistent or inadequate guidance.

When we began to examine the false positives and false negatives – i.e. the cables we predicted would be classified, but were unclassified, and vice versa – we found many instances of human error. For unclassified cables which the algorithm identified as having the highest probability of being classified, this included examples where the National Archives metadata had miscategorized the cable – the message text itself clearly showed it was originally classified. Other errors included cables that were originally secret when received at the State Department but were resent to another post as unclassified (10). And there were many examples of unclassified cables that appear to meet the criteria for secret or confidential communications (Fig. 1) (11).

There is therefore an upper limit to what any algorithm can achieve in machine classification: it cannot be more accurate than the intrinsic consistency of the data allows. To the extent humans misclassify their communications – or simply disagree about how they should be classified – so too will algorithms.

In the debate over the Clinton email, many appear to assume that the post-hoc State Department review created what data scientists would call “ground truth” in identifying classified communications. But one could argue that Clinton and her aides were no less expert in recognizing

what communications deserved to be classified and contained on a secure system. It might instead be seen as a natural experiment in inter-coder reliability in identifying state secrets. If so, this would be the government's first such experiment in 75 years of creating official secrets, or at least the first in which it has published the results.

But this experiment is flawed and incomplete, especially if it is expected to answer the question of whether Clinton and her aides can be faulted for failing to recognize email that should have been classified. These “false negatives” constitute 5.9% of the total reviewed. Even if we accepted this as ground truth, we cannot begin to estimate either precision or recall – and hence accuracy – without reviewing the communications Clinton and her aides correctly classified as sensitive and protected on a secure system (true positives), and those which they overclassified (false positives). Moreover, one would need to compare the error rate with the error rate for the rest of the State Department. Certainly, the idea that government officials make no errors in identifying sensitive national security information is not supported by the data.

Based on what we have learned from machine classification, it should be feasible to develop systems in which classified and unclassified communication streams would continually generate data that would be predictive of the appropriate classification level for new communications. Such a system would automatically default to the predicted classification, requiring manual override if the sender wished to classify at a higher or lower level (such as communications about a new classified program, or a subject that is no longer sensitive).

The same type of system could be used for accelerating the release of electronic records, harnessing data from previous declassification decisions to prioritize records for close scrutiny that are most likely to have sensitive national security or personal information. These systems could therefore “nudge” officials to classify or declassify communications appropriately, and also reveal which officials tend to diverge from the norm when dealing with similar types of subjects, language, etc. Such a system could therefore be self-correcting, and become increasingly accurate as sources of error were continually identified and corrected for.

What is most needed now is government support for applied research on technology to improve management of sensitive information (12). This should include rigorous, double-blind testing of how consistent humans are in both classification and declassification. Without determining the expected error rate among humans, we will never know the baseline against which to test computer-assisted approaches – or evaluate claims that this or that official was negligent in failing to recognize and protect state secrets.

## ACKNOWLEDGEMENTS

This research was made possible through grants from the John D. and Catherine T. MacArthur Foundation, the Columbia Global Policy Initiative, and the Fundação Getulio Vargas. The authors also thank Eric Gade, Thomas Nyberg, and Rohan Shah for expert assistance, and Michael Gill, Daniel Krasner, Aaron Plasek, Owen Rambow, and Arthur Spirling for helpful discussions. Data reported in the paper are described in the supplementary materials and are available from the U.S. National Archives.

## REFERENCES

- 1 MCINTOSH, T. Paraguay is 100th Nation to Pass FOI Law, But Struggle for Openness Goes on. **The Guardian**, [S.l.], 19 Set. 2014.
- 2 LOWELL, A. D. The Broken System of Classifying Government Documents. **New York Times**, New York, NY, 29 Fev. 2016.
- 3 MCALLISTER, W. The Documentary Big Bang, the Digital Records Revolution, and the Future of the Historical Profession. **Passport**, n. 41, p. 12-17 Sept. 2010.
- 4 PUBLIC INTEREST DECLASSIFICATION BOARD. **Transforming the Security Classification System**: Report to the President from the Public Interest Declassification Board, Washington, DC, Nov. 2012. p. 4-5. Disponível em: <[archives.gov/declassification/pidb/recommendations/transforming-classification.pdf](http://archives.gov/declassification/pidb/recommendations/transforming-classification.pdf)>. Acesso em:
- 5 LANGBART, D.; FISCHER, W.; ROBERSON, L. **Appraisal of records covered by N1-59-07-3-P**. 2007.
- 6 HARRIS, Z. Distributional Structure. **Word**, 10: 2, 3, 146-62 (1954).
- 7 FEDERATION OF AMERICAN SCIENTISTS. **Executive Order 11652**. Disponível em: <<http://fas.org/irp/offdocs/eo/eo-11652.htm>>.
- 8 METCALFE, D. The Nature of Government Secrecy. **Government Information Quarterly**, v. 26, p. 305-310 2009.
- 9 CHAURASIA, V.; PAL, S. Data Mining Approach to Detect Heart Diseases (January 9, 2014). **Int. J. Adv. Comp. Sci. Info. Tech.** 2, 56-66 (2013).
- 10 “Philip Di Tommaso - Retirement”, 8/17/1976, Central Foreign Policy Files, Doc. Number-1976ATHENS08481, United States National Archives, date accessed 2/16/2016.
- 11 See, e.g., “South Lebanon Situation - Phalange View”, 7/6/1977, Central Foreign Policy Files, Doc. Number-1977STATE156457, United States National Archives, date accessed 4/6/2016; “Rumored Kidnapping of President Kyprianou’s Son”, 12/15/1977, Central Foreign Policy Files, Doc. Number-1977NICOSI03263, United States National Archives, date accessed 4/6/2016.

12 PUBLIC INTEREST DECLASSIFICATION BOARD. **Transforming the Security**

**Classification System:** Report to the President, Washington, DC: PIDB, Nov. 2012. p. 25-27.

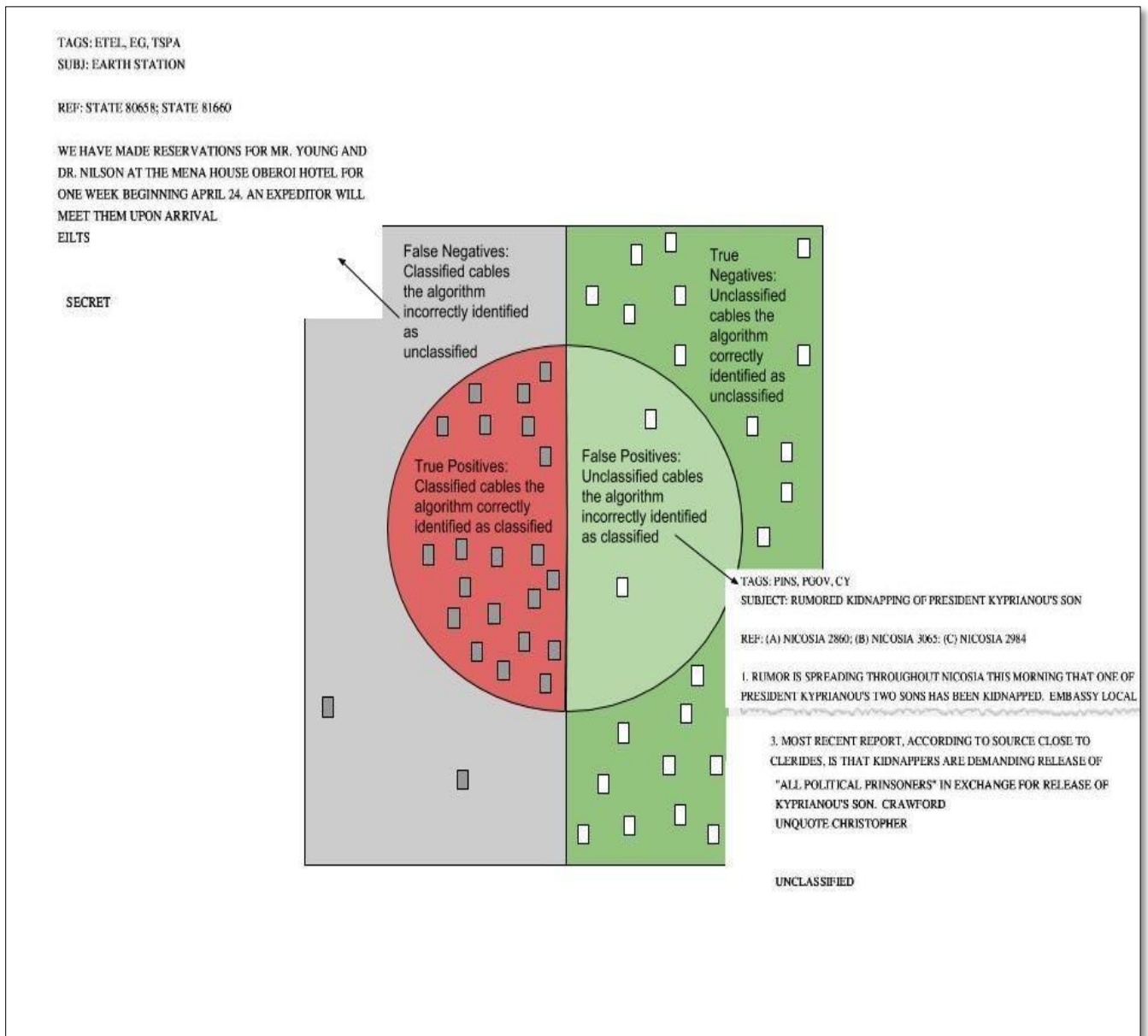
Disponível em:

<[archives.gov/declassification/pidb/recommendations/transforming-classification.pdf](http://archives.gov/declassification/pidb/recommendations/transforming-classification.pdf)>.

**FIGURES**

Classifiers can be optimized for either recall or precision. When the end-user has low tolerance for risk, they will seek to maximize recall, i.e. the percentage of all cables with sensitive information they identify as classified, even if the result is lower precision, i.e. the percentage of all cables identified as classified that actually have sensitive information. But the false positives and false negatives include examples of human error in classification.

**FIGURA 1 – How We Measure Performance in Identifying State Secrets**



To determine which features are most useful, we tested the performance of classifiers for each one individually.

FIGURE 2 – Performance of Classifiers Using Different Features

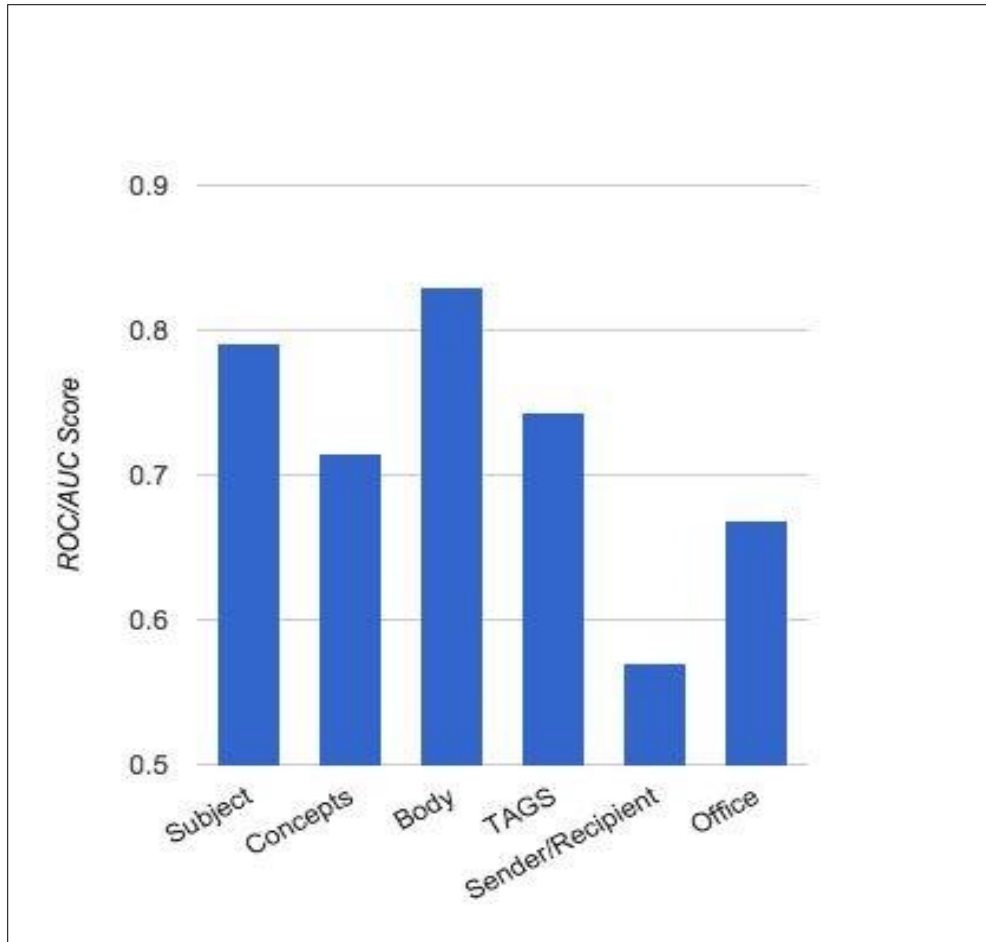




TABLE 1 – Human error in managing sensitive information

Situation	Total in Database	Unclassified	Limited Official Use	Confidential	Secret
<b>declassified cables</b>	<b>1.758.279</b>	<b>876.797</b>	<b>411.973</b>	<b>375.690</b>	<b>93.635</b>
Error messages for <i>body</i>	119.744	53.935	21.744	25.233	18.832
blank <i>body</i>	8282	2.726	1.645	1.924	1.987
blank or n/a <i>concepts</i>	634.967	445.300	114.507	65.502	9.658
blank or n/a <i>subject</i>	26.109	16.490	5.820	2.914	885
blank or n/a <i>from</i>	17	7	6	3	1
blank or n/a <i>to</i>	9.740	6.027	1.572	1.698	443
<b>Used for classifier</b>	<b>981.083</b>	<b>368.043</b>	<b>280.251</b>	<b>270.477</b>	<b>62.312</b>

TABLE 2 – Performance of Classifiers Using All Six Features for Different Tasks

Classification Task	ROC/AUC Score	Accuracy Score	Precision Score for Un-classified	Precision Score for Classified	Recall for Un-classified	Recall for Classified	Average f1-score
(U vs L,C,S)	0.859	0.87	0.834	0.891	0.81	0.9	0.896
(U,L vs C,S)	0.85	0.87	0.884	0.843	0.92	0.78	0.809
(U,L,C vs S)	0.806	0.966	0.974	0.802	0.99	0.62	0.697
(U vs C, S)	0.928	0.929	0.926	0.931	0.94	0.92	0.926

**Supplementary Materials:** a list of the supplementary materials, followed by the actual text of the Supplementary Materials.

## Metadados: princípios e práticas na organização informacional

### *METADATA: PRINCIPLES AND PRACTICES IN INFORMATIONAL ORGANIZATION*

Eduardo R. Felipe (BDMG/MG)

*eduardo@erfelipe.com.br*

**Resumo:** O modelo informacional sofreu uma grande mudança com o advento da computação e suas possibilidades de criação, armazenamento e recuperação. Este trabalho procura investigar as possibilidades dos metadados na contribuição em um ambiente de gravação e recuperação de documentos eletrônicos onde: (1) a definição de descritores para identificação de um item eletrônico seja padronizada; (2) utilizar um mesmo padrão para aproximar a linguagem de usuários indexadores e leitores. Um protótipo de software foi criado a partir da pesquisa, procurando evidenciar a capacidade deste mecanismo em proporcionar uma melhor recuperação do item bibliográfico eletrônico. A fundamentação teórico-metodológica teve como princípios as bibliotecas digitais, metadados e padrões de organização de metadados. Como resultado da pesquisa, a implementação do software obteve sucesso em sua proposta e tornou-se capaz de executar os processos investigados na metodologia. As conclusões direcionam para a valorização destas estruturas de dados alinhadas a uma política informacional, possibilitando um acervo coeso e de fácil recuperação.

**Palavras-chave:** Produção, Organização e Utilização da Informação. Metadados. Biblioteca Digital. Documento Eletrônico. Taxonomia.

**Abstract:** The information model suffered a major change with the advent of computing and the possibilities for creation, storage and retrieval. This work investigates the possibilities of the Dublin Core metadata and their contribution in an environment recording and retrieval of electronic documents where: (1) the definition of descriptors for identification of an electronic item is standardized, (2) use the same pattern to approximate the language users indexers and readers, (3) allow for a faster recovery and correct the object / subject searched. A software prototype has been created from the research, seeking to demonstrate the ability of this mechanism to provide better recovery of electronic bibliographic item. The theoretical and methodological principles had as digital libraries, metadata and metadata standards organization. As a result of the research, the software implementation was successful in its proposal and became capable of performing the processes investigated in the methodology. The findings pointed to the value of these data structures aligned to an informational politics, enabling a collection cohesive and easy retrieval.

**Keywords:** Production, Organization and Use of Information, Metadata, Digital Library, Electronic Document, Taxonomy.

## 1 INTRODUÇÃO

A informação textual é hoje um dos principais mecanismos de comunicação e registro do conhecimento humano. Esta significação que já foi registrada em diversos meios de armazenamento: rochas, argilas, tábuas, peles de animais, papiro, chegando à era do papel, tinha

sua produção controlada, reduzida, tanto pela fragilidade do suporte informacional responsável pelo seu registro, tanto quanto pelo processo de codificação e decodificação dominado por poucos.

A revolução industrial da informação teve um grande avanço com Gutemberg e sua invenção magnífica, a prensa tipográfica. Mais tarde a eletrônica possibilitou a criação dos primeiros componentes eletrônicos que possibilitaram a evolução necessária para que os dispositivos que denominamos computadores pudessem existir. Neste exato momento na história dispomos de recursos incríveis, capazes de manusear a informação de forma cada vez mais rápida e com volume informacional crescente de maneira exponencial; parece não haver limites para a produção e armazenamento informacional. Isso leva a um questionamento: como produzir tanto se não houver formas eficazes de recuperação da informação?

Pode ser mais do que apenas coerente abordar este viés dos Sistemas de Recuperação da Informação, tornou-se fundamental esta abordagem perante a grande produção informacional. Há mais de duas décadas, Salton (1975) iniciou uma de suas obras destacando exatamente o crescente volume informacional e a necessidade das bibliotecas se adequarem a este novo comportamento. Uma demanda crescente de capacidade de processamento documental e maiores custos de armazenamento, aliados à necessidade de novos métodos informacionais visto que esta massa informacional não poderia ser tratada com os métodos convencionais da época. Estamos fazendo referência à década de 70 do século XX, onde os computadores ainda não possuíam a fração de capacidade que nossos aparelhos telefônicos modernos possuem hoje. Salton (1975), assim como Vannevar Busch (1945), viu o futuro, vislumbrou a explosão informacional ainda que o *hardware* da época não evidenciasse tal realidade.

## 2 METADADOS EM BIBLIOTECAS DIGITAIS

Todas essas possibilidades trazem consigo novos desafios para a Ciência da Informação, pois, com a facilidade de inclusão (cadastro) de novos itens no acervo digital, os processos de indexação e categorização precisam ser reavaliados, bem como o processo de recuperação. Ou seja, mesmo no ambiente digital, a correta abordagem dos processos de classificação e indexação é imprescindível para a sua posterior recuperação.

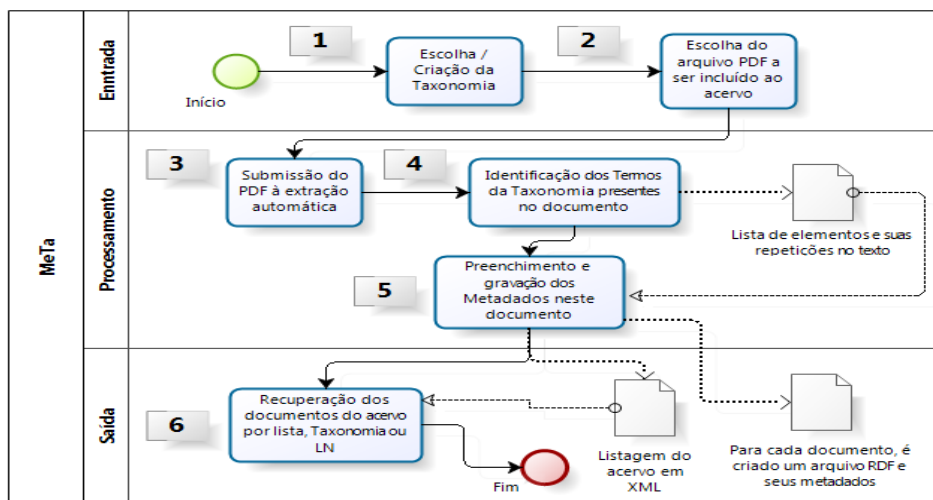
Totalmente incluídas nesse contexto, as Bibliotecas Digitais apresentam-se como resposta a uma demanda eletrônica que se acha em pleno crescimento no momento atual da história. Em países denominados “primeiro mundo”, a versão eletrônica das publicações literárias já é disponibilizada como opção à versão impressa, e um novo negócio pode ser evidenciado com a crescente opção de livrarias virtuais pela Internet. Atrelado a esse momento propício à publicação e utilização do “eletrônico/virtual”, dispositivos eletrônicos denominados *e-readers* e *tablets* estão

sendo lançados para prover uma condição simples e prática na leitura de informação digital e permitir práticas de interação com a Internet.

Tendo em vista a importância do desenvolvimento de mecanismos facilitadores do trabalho do indexador, o software MeTa foi desenvolvido com o enfoque sobre os metadados e sua aplicação em bibliotecas digitais, no intuito de implementar conceitos de gerência e recuperação de informação, em acervos digitais e que funcione como fator norteador do processo de indexação, permitindo ao usuário definir os melhores termos e metadados que propiciem alcançar uma recuperação mais fácil e precisa dos documentos que respondem a determinada requisição.

A partir do embasamento teórico de Lima (2004) e sua proposta no Modelo para Organização Hipertextual de Documentos (MHTX), a construção do software teve início através de um modelo de prototipação<sup>12</sup>. Inicialmente foram definidas algumas funções básicas para a construção de uma sequência de atividades que devem resultar na organização e recuperação dos documentos científicos.

FIGURA 1 – Processos do software MeTa



Fonte: Elaborada pelo autor.

### 3 CONCLUSÕES

O crescente e visível aumento de publicações tem demandado novos mecanismos de gestão e organização da informação. Nesse sentido, este projeto procurou contribuir com os sistemas de recuperação da informação com enfoque em Metadados. Além do papel de descritor, os metadados atuam ainda como tradutores. Embora neste projeto o foco se concentre em documentos científicos (teses e dissertações) e em um formato de arquivos pré-determinados (pdf), ao realizar o mesmo

<sup>12</sup> Modelo em que um *software* possui as características funcionais reduzidas do projeto com a finalidade de testar e avaliar suas funcionalidades básicas.

objetivo de catalogação e indexação em arquivos com informações em áudio, imagens ou vídeo, onde a simbologia textual não se faz presente, os metadados assumem um papel ainda mais importante. Nesse caso, eles intermediam linguagens diferentes, permitindo ao usuário a recuperação da informação através da linguagem textual. Na maioria dos sistemas de pesquisa, esse tipo de linguagem funciona como interface de comunicação entre a origem e a solicitação.

Através desta reflexão, o trabalho afirma a importância dos metadados no ambiente das Bibliotecas Digitais, sendo os principais pontos de destaque:

- Possibilitar a estruturação de informações não estruturadas;
- Definir um padrão para entrada (cadastro) e saída (recuperação) das informações;
- Trabalhar a indexação (catalogação de assunto) e a recuperação através de um mesmo instrumento;
- Permitir a *tradução* da linguagem natural para a linguagem documentária;
- Implementar a interoperabilidade na medida em que outros softwares utilizarem o mesmo padrão através de padrões abertos de codificação;
- Auxiliar o profissional da informação / usuário final no processo de indexação através de um mecanismo semiautomático.

Este trabalho pretendeu oportunizar, no ambiente acadêmico, ensejo para que alunos e professores ampliem suas discussões em temas de grande relevância para a Ciência da Informação, como os vocabulários controlados, linguagens documentárias, processos de indexação (catalogação de assunto), sistemas de recuperação da informação (SRI) e metadados, além do viés tecnológico presente na manipulação do *software* e da representação informacional gravada através das linguagens XML / RDF e recuperada através de interfaces textuais e mapas hiperbólicos. É certo também que este trabalho pode influenciar novas linhas de pesquisa, relacionadas à área de organização e uso da informação, bem como as ligadas à área de tecnologia da informação.

## REFERÊNCIAS

BUSH, Vannervar. As we may think. **The Atlantic Monthly**, July 1945. Disponível em: <<http://web.mit.edu/STS.035/www/PDFs/think.pdf>>. Acesso em: 12 jul. 2012.

LIMA, G. A. B. O. Mapa conceitual como ferramenta para organização do conhecimento em sistemas de hipertextos e seus aspectos cognitivos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n. 2, p. 134-145, 2004.

SALTON, G. **Dynamic Information and Library Processing**. Englewood Cliffs, N.J.: Prentice-Hall, 1975. [ERIC e-Book].