

ISSN 0103-3786  
VOLUME 24 NÚMERO 2  
MAIO/AGOSTO 2012

# Trans**Informação**

FUNDADA EM 1989

#### **Editora / Editor**

Profa. Dra. Mariângela Pisoni Zanaga (PUC-Campinas)

#### **Editor Adjunto / Adjunct Editor**

Prof. Dr. Rogério Eduardo Rodrigues Bazi (PUC-Campinas)

#### **Editora Associada / Associate Editor**

Profa. Dra. Nair Yumiko Kobashi (USP)

#### **Editora Gerente / Manager Editor**

Maria Cristina Matoso (PUC-Campinas)

#### **Conselho Editorial / Editorial Board**

Prof. Dr. Aldo de Albuquerque Barreto (UNIGRANRIO)

Profa. Dra. Henriette Ferreira Gomes (UFBA)

Profa. Dra. Isa Maria Freire (UFPB)

Profa. Dra. Lígia Café (UFSC)

Prof. Dr. Luís Fernando Sayão (CNEN)

Profa. Dra. Maria de Fátima Gonçalves Moreira Tálamo (USP)

Profa. Dra. Maria Inês Tomaél (UEL)

Profa. Dra. María Néliida González de Gómez (IBICT)

Prof. Dr. Raimundo Nonato Macedo dos Santos (UFPE)

Profa. Dra. Rosali Fernandez de Souza (IBICT)

#### **Conselho Editorial Internacional / International Editorial Board**

Prof. Dr. Antonio García Gutiérrez (Universidade de Sevilla, Espanha)

Profa. Dra. Fernanda Ribeiro (Universidade do Porto, Portugal)

Prof. Dr. Juan Carlos Fernández Molina (Universidade de Granada, Espanha)

Prof. Dr. Pierre Fayard (Universidade de Poitiers, França)

Prof. Dr. Yves-François Le Coadic (Conservatoire National des Arts et Metiers, França)

#### **Normalização / Normalization**

##### **Bibliotecárias / Librarians**

Maria Cristina Matoso (PUC-Campinas)

Maurícia Daniela Pereira Sacchi (PUC-Campinas)

#### **Apoio Administrativo / Administrative Support**

André Gustavo Tomaz dos Santos (PUC-Campinas)

#### **INDEXAÇÃO / INDEXING**

Web of Science, JCR Social Science, Latindex, Clase

FI=0,083

Qualis A1

#### **Copyright** © Transinformação

É permitida a reprodução parcial desde que citada a fonte. A reprodução total depende da autorização da Revista.

*Partial reproduction is permitted if the source is cited. Total reproduction depends on the authorization of the Transinformação*

O Conselho Editorial não se responsabiliza por conceitos emitidos em artigos assinados.

*The Board of Editors does not assume responsibility for concepts emitted in signed articles.*

Transinformação fundada em 1989, é publicada quadrimestralmente e é de responsabilidade da Faculdade de Biblioteconomia, Centro de Ciências Humanas e Sociais Aplicadas, Pontifícia Universidade Católica de Campinas. Publica trabalhos da área da Ciência da Informação realizados na Universidade, bem como de colaboradores externos.

*Transinformação founded in 1989, it is published every four months and it is of responsibility of the Faculdade de Biblioteconomia, Centro de Ciências Humanas e Sociais Aplicadas, Pontifícia Universidade Católica de Campinas. It publishes works carried out in the University in the field of Information Science, as well as external contributors works.*

#### **COLABORAÇÕES / CONTRIBUTIONS**

Os manuscritos devem ser enviados à Secretaria da Revista, <<http://revistas.puc-campinas.edu.br/transinfo>> conforme as "Instruções aos Autores", publicadas no final de cada fascículo.

*All manuscripts should be sent to the Transinformação' Office <<http://revistas.puc-campinas.edu.br/transinfo>> and should comply with the "Instructions for Authors," published in the end of each issue.*

#### **ASSINATURAS / SUBSCRIPTIONS**

Pedidos de assinatura ou permuta devem ser encaminhados a Secretaria.

- Annual:
- Pessoa física: R\$50,00
  - Institucional: R\$140,00

*Subscription or exchange orders should be addressed to the Secretaria.*

- Annual:
- Individual rate: US\$50,00
  - Institucional rate: US\$140,00

*E-mail: [sbi.assinaturane@puc-campinas.edu.br](mailto:sbi.assinaturane@puc-campinas.edu.br)*

#### **CORRESPONDÊNCIA / CORRESPONDENCE**

Toda a correspondência deve ser enviada à Transinformação no endereço abaixo:

*All correspondence should be sent to Transinformação at the address below:*

Núcleo de Editoração SBI

Prédio da Antiga Reitoria Sala 8 - Campus I

Rod. Dom Pedro I, km 136 - Pq. das Universidades

13086-900 - Campinas - SP

Fone/Fax: 55 (19) 3343-7401

*E-mail: [sbi.nucleoeditoracao@puc-campinas.edu.br](mailto:sbi.nucleoeditoracao@puc-campinas.edu.br)*



Centro de Ciências Sociais Aplicadas

**PUC**  
CAMPINAS  
PONTIFÍCIA UNIVERSIDADE CATÓLICA



# Trans**Informação**

## FICHA CATALOGRÁFICA

Elaborada pelo Sistema de Bibliotecas e  
Informação - SBI - PUC-Campinas

Transinformação. Pontifícia Universidade Católica de Campinas. Centro de Ciências Humanas e Sociais Aplicadas. – Campinas, SP, v.1 n.1 (jan./abr. 1989-)

v.24 n.2 maio/ago. 2012

Quadrimestral 1989-1999; Semestral 2000-2002; Quadrimestral 2003-  
Resumo em Português e Inglês.  
ISSN 0103-3786

1. Biblioteconomia – Periódicos. 2. Ciência da Informação – Periódicos.  
I. Pontifícia Universidade Católica de Campinas. Centro de Ciências Humanas  
Sociais Aplicadas.

CDD020

## Artigos | **Articles**

- 77 A representação descritiva no contexto da *web* semântica  
*Descriptive representation in the semantic web context*  
• Maria Elisabete Catarino; Terezinha Batista de Souza
- 91 A reference ontology for digital scientific journals applied to systematic literature review processes  
*Uma ontologia de referência para periódicos científicos digitais aplicada em processos de revisão sistemática da literatura*  
• Fernando Benedet Ghisi; Gleisy Regina Bóries Fachin; Marcos Henrique dos Santos; Denilson Sell; Gregório Jean Varvakis Rados
- 103 PuertoTex: un *software* de minería textual para la creación de resúmenes automáticos en el dominio de ingeniería de puertos y costas basado en ontologías  
*PuertoTex: a data mining software based on ontologies for automatic summarization on port and coastal engineering domain*  
• Amed Leiva-Mederos; Sandor Domínguez-Velasco; Jose Antonio Senso
- 117 Mudança de paradigma e sua ruptura: um estudo de caso na Museologia e a pluralidade paradigmática da Ciência da Informação  
*Paradigm shift and its rupture: a case study and paradigmatic plurality of Museology and Information Sciences*  
• Mário Gouveia Júnior; Raimundo Nonato Macedo dos Santos
- 127 O reflexo da inteligência coletiva nas organizações  
*Effect of collective intelligence in organizations*  
• Ketry Gorete Farias dos Passos; Edna Lúcia da Silva
- 137 Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos  
*Information retrieval systems: a discussion based on enunciative parameters*  
• Vera Maria Araujo Pigozzi de Araujo
- 145 Instruções aos Autores  
*Instructions to the Authors*

# A representação descritiva no contexto da *web* semântica<sup>1</sup>

## *Descriptive representation in the semantic web context*

Maria Elisabete CATARINO<sup>2</sup>

Terezinha Batista de SOUZA<sup>2</sup>

### Resumo

Este artigo contém o resultado de uma pesquisa bibliográfica, cujo objetivo foi verificar como as práticas da representação descritiva estão relacionadas com a proposta de organização dos recursos da *Web* no contexto da *Web* Semântica. A pesquisa foi desenvolvida a partir de duas temáticas: uma que abarcou os conceitos referentes à *Web* Semântica, com o objetivo de compreender os procedimentos e as tecnologias a eles relacionados; outra que focou a representação descritiva, com o intuito de entender como a práxis da catalogação poderá auxiliar na inserção dos dados bibliográficos no contexto da *Web* Semântica, bem como averiguar ações que estão sendo realizadas nesse sentido. O levantamento foi feito nos idiomas português e inglês, em periódicos e livros das áreas de Ciência da Informação, *Internet* e *Web*, bem como em documentos disponíveis no *site* do *Word Wide Web Consortium*. O texto discorre sobre a *Web* Semântica e os conceitos básicos a ela relacionados: dados ligados, vocabulários, busca, inferência e aplicações verticais. Destaca o *Resource Description Framework* como modelo de descrição de recursos, que é o fundamento da *Web* Semântica. Com base na análise dos textos, nos exemplos apontados de ações que estão sendo desenvolvidas por algumas instituições, tais como *Library of Congress*, *Dublin Core Metadata Initiative* e *Joint Steering Committee*, e tendo como foco os objetivos acima explicitados, pôde-se inferir que as práticas da representação descritiva estão relacionadas com a *Web* Semântica na medida em que poderão ser aplicadas para tornar os dados bibliográficos em dados ligados da *Web*.

**Palavras-chave:** *Dublin Core*. Metadados. Representação descritiva. FRBR. RDA. RDF. *Web* semântica.

### Abstract

*This article presents the result of a bibliographic research developed with aim of verifying how the practices of descriptive representation are related to the proposal of Web resources organization in the Semantic Web context. The study was developed based on two main themes. The first involved concepts related to the Semantic Web with objectives of understanding which are the procedures and technology related to Semantic Web. The second concerned descriptive representation in order to understand how cataloging practices can be used to help bibliographic data insertion into the Data Web as well as to verify the actions that are being developed in this direction. The research was developed in Portuguese and English, in papers and books in the information science area as well as in documents available in the site of the Word Wide Web Consortium. This article is based on Semantic Web and its concepts such as: linked data, vocabulary, query, inference and vertical applications. These results highlight the Resource Description Framework that is a resource description model based on Semantic Web. Based on text analysis, in examples of actions pointed out, which are being developed by some institutions, such as: Library of Congress, Dublin Core Metadata Initiative and Joint Steering Committee; and with focus on the objectives already mentioned, it can be inferred that the practices of descriptive representation are related to the Semantic Web, to the extent to which they can be applied to transform the bibliographic data into data linked to the Web.*

**Keywords:** *Dublin Core*. Metadata. Descriptive representation. FRBR. RDA. RDF. Semantic web.

<sup>1</sup> Artigo elaborado a partir da pesquisa em andamento intitulada "Contribuições da Ciência da Informação para *Web* Semântica", coordenada por M.E. CATARINO. Apoio financeiro do MCT/CNPq, Edital Universal 2010, processo 474215/2010-5.

<sup>2</sup> Professoras Doutoras, Universidade Estadual de Londrina, Departamento de Ciência da Informação. Rod. Celso Garcia Cid, Km 380, *Campus* Universitário, Caixa Postal 6001, 86051-990, Londrina, PR, Brasil. Correspondência para/Correspondence to: M.E. CATARINO. E-mail: <beteca@uel.br>.

Recebido em 21/1/2012, reapresentado em 30/3/2012 e aceito para publicação em 25/4/2012.

## Introdução

O estudo apresentado neste artigo inicialmente enfatizava as questões da representação temática. Todavia, com o desenvolvimento da investigação e consequentemente maior compreensão do que é a *Web Semântica* e qual o relacionamento ou contribuições da Ciência da Informação para esse cenário, percebeu-se que também, e inclusive, a representação descritiva deveria ser estudada como contribuição para o desenvolvimento da organização da informação na *Web*.

Desde o início da pesquisa, havia o pressuposto de que a Ciência da Informação, mais especificamente a área de organização da informação, contribuiria com a *Web Semântica* devido à convicção de que a práxis de ambas são compatíveis. Neste texto, o foco está sobre a área da representação descritiva, pois o alicerce da *Web Semântica* está num modelo de descrição dos recursos da *Web*, o *Resource Description Framework* (RDF). Ao abordar a relação entre as práticas da representação descritiva e da *Web Semântica*, vislumbra-se que a atuação do profissional da informação na organização dos recursos da *Web* permitirá o compartilhamento e a preservação dos dados bibliográficos (na *Internet*), conforme prescrevem as regras da catalogação.

O termo *Web Semântica* foi criado para representar o projeto do *Word Wide Web Consortium* (W3C), que pretende estruturar as páginas da *Web*, dando a elas significado, desenvolvendo um ambiente onde agentes de *software* possam criar sofisticadas tarefas para os usuários a partir de conteúdos de várias páginas e bases de dados<sup>3</sup> da *Web* (Berners-Lee *et al.*, 2001). A estruturação desses conteúdos ocorre por meio de normas e regras para a criação de armazéns de dados na *Web*, construção de vocabulários e manipulação de dados (Word Wide Web Consortium, 2010).

Ao se pensar em estruturação dos conteúdos, logo se pode fazer uma analogia com a área de organização da informação, especialmente a representação descritiva e temática. Nesse contexto, entender quais regras e normas estão sendo adotadas pelos desenvolvedores da *Web* e

como a práxis da representação descritiva está se desenvolvendo nessa área é imprescindível. Este artigo apresenta o resultado de um estudo feito a partir de uma revisão bibliográfica sobre os conceitos da *Web Semântica* e da representação descritiva, bem como as ações que já estão em andamento para tornar os dados bibliográficos dados lincados da *Web*.

Com o objetivo de verificar como as práticas da representação descritiva estão relacionadas com a proposta de organização dos recursos da *Web*, no contexto da *Web Semântica*, a pesquisa bibliográfica foi realizada a partir de duas temáticas. O primeiro tema abarcou o conceito *Web Semântica* e outros a ele relacionados (Dados Lincados, RDF e *Web de Dados*) com o objetivo de compreender as tecnologias e os procedimentos necessários para a implantação da *Web Semântica*. O outro tema focou a representação descritiva com o intuito de entender como a sua práxis poderá auxiliar na inserção dos dados bibliográficos na *Web Semântica*, bem como averiguar ações que têm sido realizadas neste sentido.

A pesquisa foi realizada nos termos acima citados, nos idiomas português e inglês, em periódicos e livros das áreas de Ciência da Informação, *Internet* e *Web*. Também foram consultados os documentos disponíveis no site do W3C.

Com base na leitura dos textos e tendo como foco os objetivos explicitados, procedeu-se a uma análise para averiguar o pressuposto de que a prática da representação descritiva está relacionada com a da *Web Semântica* e, mais especificamente, como os dados bibliográficos poderão tornar-se dados lincados.

Para discorrer sobre essa perspectiva, apresentamos, inicialmente, os conceitos relativos à *Web Semântica* e as tecnologias a ela relacionadas, definições da representação descritiva e novas perspectivas da catalogação no cenário da *Web*.

## Web semântica

O conceito da *Web Semântica*<sup>4</sup> foi referido em 2001, na publicação de um artigo na revista *Scientific American*,

<sup>3</sup> Neste texto considera-se a bases de dados no sentido de qualquer conjunto de registros relacionados entre si, conforme define Sawaya (1999, p.112) "Conjunto de informações relacionadas entre si, referentes a um mesmo assunto e organizadas de maneira útil, com o propósito de servir de base para que o usuário recupere informações".

<sup>4</sup> No entanto, Berners-Lee considerou que o título mais adequado para o projeto seria *Data Web ou Web of Data* (*Web de Dados*) (King, 2007; Word Wide Web Consortium, 2010). Optou-se neste texto por padronizar a denominação do projeto, utilizando apenas o termo *Web Semântica* para que haja maior entendimento do conteúdo.

por Tim Berners-Lee, James Hendler e Ora Lassila. Os autores explicam que não é uma *Web* separada, mas uma extensão da atual, na qual a informação é dada com um significado bem definido, permitindo melhor interação entre computadores e pessoas (Berners-Lee *et al.*, 2001). Trata-se de uma nova geração da *Web*, projeto do W3C, cujo intento é desenvolver tecnologias, linguagens, padrões e recomendações que tornem a informação legível pelas máquinas.

Segundo Berners-Lee *et al.* (2001), os computadores necessitam ter acesso a coleções estruturadas de informações (dados e metadados) e a conjunto de regras de inferência que ajudem no processo de dedução automática para que seja aplicado o raciocínio automatizado.

A *Web Semântica* é composta por alguns elementos: uma filosofia, um conjunto de princípios para *design*, grupos de trabalho e tecnologias para a implantação desses princípios e propostas. Alguns elementos são expressos em especificações formais que incluem o modelo RDF, formatos de intercâmbio de dados (como por exemplo: *RDF/XML*, *N3*, *Turtle*, *N-Triples*) e notações, tais como *RDF Schema* (RDFS) e a *Web Ontology Language* (OWL), para prover uma descrição formal dos recursos e dos conceitos, termos e seus relacionamentos, num específico domínio do conhecimento (Word Wide Web Consortium, 2001).

O W3C pretende implementar, divulgar e promover linguagens, normas e recomendações que permitam que os computadores possam, a partir de dados dispersos em *sites* e bases de dados da *Web*, compor interações na rede e gerar serviços e produtos. O termo *Web Semântica* representa uma *Web* de dados ligados (*linked data*). As tecnologias recomendadas pelo W3C permitirão que as pessoas e as instituições criem armazém de dados, construam vocabulários e escrevam regras para a manipulação dos dados (Word Wide Web Consortium, 2010).

Os conceitos básicos que compõem a ideia da *Web Semântica* (dados ligados, vocabulários, busca, inferência e aplicações verticais) serão apresentados a seguir.

## Dados ligados

Conforme o *Word Wide Web Consortium* (2010), para tornar a *Web Semântica* uma realidade, é importante

ter uma grande quantidade de dados disponível em um formato padrão, acessível e gerenciável por tecnologias apropriadas. Além do acesso aos dados da *Web*, é necessário, ainda, criar relações entre os dados. Dados ligados é o nome que se dá às coleções de dados relacionados na *Web*.

Para a publicação e vinculação dos dados, é preciso adotar o modelo recomendado pelo W3C, o RDF, que é o alicerce da *Web Semântica* e um modelo padrão para intercâmbio de dados na *Web* (RDF Working Group, 2004). As tecnologias disponíveis devem ter como fundamento o modelo RDF para possibilitar o acesso aos dados existentes em bases de dados, sejam elas relacionais (estruturadas em tabelas e seus relacionamentos) ou outras, como, por exemplo, as que estão estruturadas na forma de linguagens de marcação (XML ou HTML) (Word Wide Web Consortium, 2010). O W3C recomenda: RDF; *Gleaning Resource Descriptions from Dialects of Languages* (GRDDL) que é uma ferramenta desenvolvida pela W3C para a obtenção de dados RDF a partir de documentos XML e em particular XHTML (Halpin; Davis, 2007); *Protocol for Web Description Resources* (POWDER), um protocolo para descrição e localização de recursos da *Web* que ajuda o usuário a decidir se um determinado recurso é de interesse (Scheppé, 2009); *Resource Description Framework Attribute* (RDFa) que descreve o conteúdo de uma informação utilizando atributos das *tags* de linguagens de marcação como *Extensible Hypertext Markup Language* (XHTML) (Adida; Birbeck, 2008); *Relational Databases to RDF Mapping Language* (R2RML), uma linguagem para mapear dados de bases de dados relacionais para bases de dados RDF (Das *et al.*, 2010); *Rule Interchange Format* (RIF) padrão desenvolvido para facilitar integração de conjunto de regras e é composto por um conjunto interconectado de dialetos representando regras de linguagens com diferentes características (Morgenstern *et al.*, 2010); e *Protocol and RDF Query Language* (SPARQL), uma linguagem de busca padronizada proposta pelo RDF *Data Access Working Group* (DAWG) do W3C, considerada uma chave para a *Web Semântica* (Prud'Hommeaux; Seaborne, 2008; Word Wide Web Consortium, 2010).

Um exemplo de dados ligados, citado pelo W3C, é a *DBPedia* (<http://wiki.dbpedia.org>) cujo objetivo é extrair informação estruturada da *Wikipedia* e torná-la disponível



na *Web*, permitindo executar consultas complexas linkando os dados da *Wikipedia* com outros conjuntos de dados da *Web* (Hahn *et al.*, 2010), essencialmente reformata o conteúdo da *Wikipedia* disponível para RDF. A importância da *DBPedia* não está somente em incluir conteúdos da *Wikipedia*, mas também incorporar *links* de outras bases de dados, como, por exemplo, o *Geonames* (<http://www.geonames.org>) que é uma base de dados de nomes geográficos que contém mais de dez milhões de nomes geográficos, além de informações gerais e estatísticas sobre países, cidades etc. Ao fornecer esses *links* extras (em RDF), as aplicações podem explorar o conhecimento de outras bases de dados e consequentemente fornecer ao usuário um resultado bem mais completo.

## Vocabulários

Nomeadamente, na *Web Semântica*, os vocabulários são constituídos por termos que representam conceitos, possíveis relacionamentos entre eles e as limitações no seu uso. Os vocabulários podem também ser denominados “ontologias”. Segundo o *Word Wide Web Consortium* (2010), na prática, não existe uma recomendação para o uso de um ou outro conceito, mas, comumente, a tendência é usar o conceito “vocabulário” para coleções de termos mais simples e “ontologia” para as coleções mais complexas<sup>5</sup>.

Como exemplo da utilização de vocabulários para a *Web Semântica*, cita-se um destacado pelo *Word Wide Web Consortium* (2010): para a área da saúde - os profissionais médicos podem usar vocabulários para representar informações sobre os sintomas, doenças e tratamentos; já as empresas farmacêuticas, para representar informações sobre medicamentos, dosagens e reações. A combinação dessas informações com os dados sobre os pacientes permitiria uma ampla gama de aplicações inteligentes: ferramentas de apoio à decisão sobre possíveis tratamentos; sistemas de monitoramento da eficácia das drogas nos tratamentos e seus efeitos colaterais; ferramentas de apoio à investigação epidemiológica etc.

O W3C oferece várias possibilidades para definir vocabulários: RDF e RDF *Schema*, *Simple Knowledge*

*Organization System* (SKOS), OWL e RIF. A escolha da tecnologia vai depender da complexidade exigida pela aplicação na qual o vocabulário será utilizado.

## Busca

No contexto da *Web Semântica*, o termo busca (*Query*) é usado genericamente para se referir às tecnologias e aos protocolos desenvolvidos para a recuperação da informação (*Word Wide Web Consortium*, 2010).

Assim como as bases de dados tradicionais necessitam de linguagens específicas de consulta (por exemplo, a SQL para as bases de dados relacionais, e o *XQuery* para as bases de dados em XML), a *Web Semântica*, com seus dados formatados em RDF, necessita de uma linguagem própria, no caso, a SPARQL. Essa linguagem permite fazer buscas e receber os resultados em *Hyper Text Transfer Protocol* (HTTP) ou *Simple Object Access Protocol* (SOAP).

## Inferência

Na lógica, inferência é definida como:

O processo de concluir uma afirmação a partir de outras afirmações. Por exemplo, com base nas afirmações ‘Deus existe’ e ‘Se Deus existe, a felicidade eterna é possível’, pode-se inferir ‘A felicidade eterna é possível’. Um ARGUMENTO é uma inferência, usada para efeitos de persuasão racional (Murcho, 2011, *online*).

Na *Web Semântica*, os dados são modelados como um conjunto de relações entre os recursos da *Web*. Nesse contexto, “inferência” refere-se à descoberta de novas relações entre os recursos *Web* com base nos dados e em algumas informações adicionais sob a forma de um vocabulário ou conjunto de regras (*Word Wide Web Consortium*, 2010).

Ambas as formas de informações adicionais recorrem a técnicas de representação do conhecimento. A primeira forma (os vocabulários) define classes e subclasses de conceitos, e, ainda, como cada recurso pode ser associado a essas classes, caracterizando as relações entre as classes e suas instâncias.

<sup>5</sup> O conceito de complexidade usado aqui é no sentido de grau de formalidade, como, por exemplo, no âmbito das ontologias que, segundo a tipologia de Uschold e Gruninger, vão desde as Altamente Informais (as mais simples), até as Rigorosamente Formais (as mais complexas) (Catarino, 2009).

A segunda forma (as regras) concentra-se na definição de um mecanismo geral de descoberta e geração de novas relações baseadas nas já existentes, bem como em programas lógicos, como, por exemplo, *Prolog*.

As ferramentas recomendadas pelo *Word Wide Web Consortium* (2010), para definir vocabulários são: RDFS, OWL ou SKOS. Para as abordagens baseadas em regras, o W3C recomenda o padrão RIF.

## Aplicações verticais

Aplicações verticais (*vertical applications*) é uma expressão usada pelo *Word Wide Web Consortium* (2010), para se referir às aplicações genéricas desenvolvidas por comunidades específicas que exploram as tecnologias do W3C. Algumas dessas áreas de aplicação podem formar grupos de trabalho no W3C, colaborando com os demais membros do consórcio para explorar novas possibilidades para a Web Semântica. Esses grupos trazem importantíssimas contribuições para o aperfeiçoamento das recomendações, tecnologias e padrões desenvolvidos pelo W3C.

O W3C cita vários exemplos, destacando-se o da área da Saúde e Ciências da Vida: *The Health Care and Life Science Interest Group*, que foi criado, em 2005, para explorar a usabilidade de tecnologias de Web Semântica na área, como a divulgação do conhecimento científico, processos de aprovação de medicamentos, entre outros. O grupo tem produzido várias manifestações e publicações, tem organizado *workshops* e conferências, e conseguiu transformar essa área de aplicação em uma das comunidades mais ativas de usuários das tecnologias da Web Semântica.

## RDF: o fundamento da web semântica

Apresentados os conceitos básicos da Web Semântica, passamos ao RDF, considerado o fundamento da Web Semântica e que está relacionado com a representação descritiva dos recursos da Web. Segundo Allemang e Hendler (2011, p.27), "RDF, RDFS e OWL são as linguagens básicas de representação da Web Semântica".

RDF é um modelo para a descrição de recursos. Allemang e Hendler (2011, p.49) o definem como um "sistema de modelagem de dados". Segundo Manola e Miller (2004), RDF é uma forma de representação dos metadados dos recursos *Web*, tais como título, autor, data de modificação de uma página da *Web*, *copyright* e informações sobre licenças de documentos *Web* ou informações sobre o compartilhamento.

Nesse modelo de descrição de dados, existe uma estrutura bastante simples. Os recursos *Web*, ou seja, todas as coisas na *Web*, podem ser descritos em triplos: *Subject*, *Predicate*, *Object* (Allemang; Hendler, 2011). Ainda nesse sentido, Manola e Miller (2004) dizem que o RDF é baseado na identificação de recursos por meio de identificadores da *Web*, denominados *Uniform Resource Identifiers* (URI), e na descrição dos recursos em termos de propriedades, ou seja, atributos que descrevem um recurso ou as características que podem identificar um recurso e seus valores correspondentes. Para exemplificar os conceitos do modelo RDF (*Subject/Recurso*, *Predicate/Propriedade*, *Object/Valor*), a seguir serão apresentados alguns exemplos de descrição de dados no modelo RDF.

Conforme Manola e Miller (2004), no *site* do W3C, se quisermos especificar o autor (*Creator*) ao descrever um dado recurso, teríamos a seguinte declaração:

```
http://www.example.org/index.html has a
creator whose value is John Smith.
```

No exemplo acima, o triplo seria:

```
Subject: o site - http://www.example.org/
index.html
```

```
Predicate: a propriedade - creator
```

```
Object: o valor correspondente à propriedade:
John Smith
```

Essa estrutura permite uma forma bastante simples de mesclagem de dados (*merging data*). Para Allemang e Hendler (2011), não é necessário estruturar os elementos de metadados em forma de tabelas/colunas para poder mesclar ou recuperar os dados. A mescla dos dados é, então, transformada, a partir dos triplos, numa simples questão de reunir todas as declarações<sup>6</sup> oriundas de diversas fontes num único lugar.

<sup>6</sup> Declaração pode ser definida como sendo uma parte dos metadados que descreve um recurso na *Web*. Composta de um "*Subject*", um "*Predicate*" e um "*Object*", ou seja, de um Recurso, uma Propriedade e seu Valor correspondente: "*Moby Dick/has author/Herman Melville*" é uma declaração. Na proposta da Web Semântica, não há um registro de metadados, mas um grupo de declarações relacionadas. As declarações relacionadas expressam a mesma descrição que um registro de metadados expressaria em outros sistemas (Word Wide Web Consortium, 2011).

Essa simples estrutura dos triplos tem um desafio (que não é apenas do RDF, mas sim de toda a *Web*): um identificador (*identifier*) único para cada recurso disponível. Na *Web*, a solução é o URI, adotada também pelo RDF.

Considerando que o conceito básico da RDF é o de que tudo seja identificado por meio de URI, como uma forma adequada para que as declarações em RDF e RDF/XML sejam processáveis por máquina, nesse sentido, a declaração do exemplo anterior seria da seguinte forma:

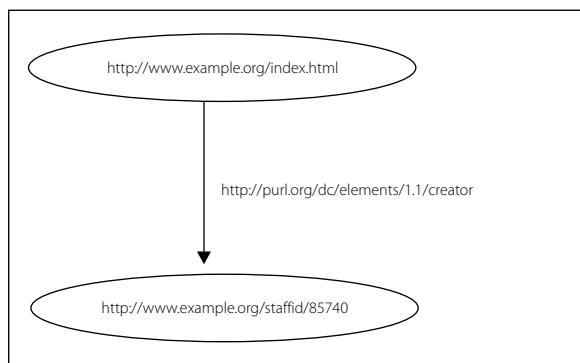
Subject: o site - <http://www.example.org/index.html>

Predicate: a propriedade creator - <http://purl.org/dc/elements/1.1/creator>

Object: o valor correspondente à propriedade - <http://www.example.org/staffid/85740>

Nota-se que tanto a propriedade quanto o seu valor correspondente foram representados em seus respectivos *namespaces* que é uma coleção de nomes (ou termos) usados em documentos XML como tipos de elementos (propriedades) e nomes de atributos (valores) (Word Wide Web Consortium, 2003) (Figura 1). A propriedade *Creator* faz parte do conjunto de termos do *Dublin Core Metadata Elements* (<http://purl.org/dc/elements/1.1/creator>) e o valor correspondente possui um *namespace* específico (identificadores numéricos para o *staff* do referido exemplo: <http://www.example.org/staffid/85740>).

Essa declaração (Figura 1) seria representada, em RDF/XML, da seguinte forma: versão XML usada; indicação dos *namespaces* utilizados (no caso: RDF, DC e EXTERMS) e



**Figura 1.** Uma declaração em RDF.

Fonte: Manola e Miller (2004).

na sequência a descrição propriamente dita (Manola; Miller, 2004).

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:exterm="http://www.example.org/terms/">
<rdf:Description rdf:about="http://www.example.org/index.html">
<dc:creator rdf:resource="http://www.example.org/staffid/85740"/>
</rdf:Description>
</rdf:RDF>
```

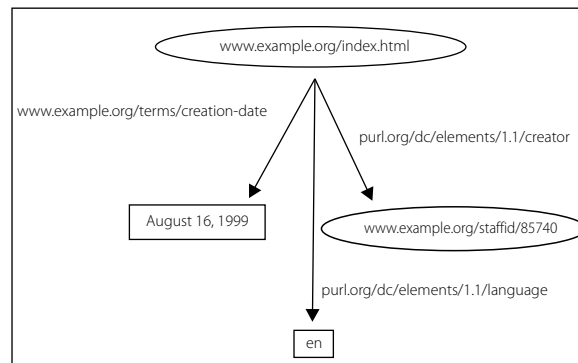
Deve-se ressaltar que a descrição de um recurso poderá ser composta, ou seja, a um *Subject* poderá relacionar-se mais de um *Predicate* e seus correspondentes *Object* (Figura 2). No exemplo utilizado, a declaração é referente ao *Creator* (autor) de uma página da *Web* (exemplo). No entanto, poderíamos ainda complementar a descrição com outras propriedades, tais como a data de criação e o idioma.

Subject: o site - <http://www.example.org/index.html>

Predicate: a propriedade creator - <http://purl.org/dc/elements/1.1/creator>

Object: o valor correspondente à propriedade: <http://www.example.org/staffid/85740>

Predicate: a propriedade creation-date - <http://www.example.org/terms/creation-date>



**Figura 2.** Descrição composta por várias propriedades.

Fonte: Manola e Miller (2004).

Object: o valor correspondente à propriedade:  
August, 16, 1999

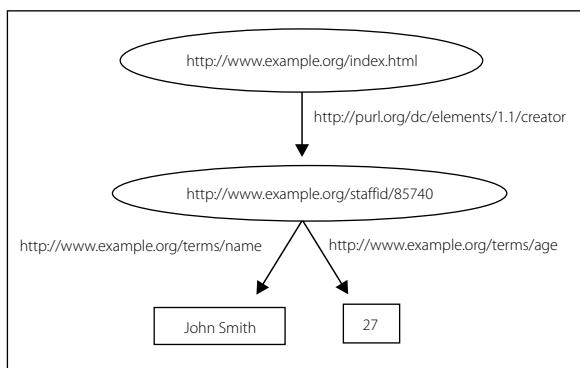
Predicate: a propriedade language - <http://purl.org/dc/elements/1.1/language>

Object: o valor correspondente à propriedade:  
en

A Figura 2 mostra o gráfico dessa descrição.

Em RDF/XML, numa forma adaptada de Manola e Miller (2004):

```
<?xml version="1.0"?>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:exterms="http://www.example.org/terms/">
    <rdf:Description rdf:about="http://www.example.org/index.html">
      <dc:creator rdf:resource="http://www.example.org/staffid/85740"/>
    </rdf:Description>
    <rdf:Description rdf:about="http://www.example.org/index.html">
      <exterms:creation-date>August 16, 1999</exterms:creation-date>
    </rdf:Description>
    <rdf:Description rdf:about="http://www.example.org/index.html">
      <dc:language>en</dc:language>
    </rdf:Description>
  </rdf:RDF>
```



**Figura 3.** Descrição composta.

Fonte: Manola e Miller (2004).

O exemplo representa uma declaração simples, um recurso e suas propriedades. Contudo, as descrições no modelo RDF podem ser mais complexas. Por exemplo, um valor *Object* poderá ser ele próprio um *Subject* com suas propriedades *Predicate* e seus respectivos valores (Figura 3).

O intuito deste texto é apresentar a ideia RDF, com seus princípios básicos, no entanto, ainda existem muitas outras questões importantes que permeiam a descrição por meio do RDF. Uma dessas questões é o fato de um valor poder ser representado tanto por um URI (no exemplo <http://www.example.org/staffid/85740> - URI referente ao *Creator* da página exemplo: John Smith) quanto por um *string*, ou seja, por um conjunto de caracteres alfanuméricos que podem formar dados, sejam eles palavras e/ou frases ou quaisquer outros tipos de dados. Conforme Sawaya (1999, p.450), *string* é "sequência conectada de caracteres ou de *bits* tratados como um só item de dados". No exemplo, os valores "John Smith" e "27" são *strings* que representam os valores das propriedades "*name*" e "*age*", respectivamente.

Um *slogan* que se refere à *Web* é "*Anyone can say Anything about Anytopic (AAA)*" (Allemberg; Hendler, 2011, p.11), ou seja, Qualquer um pode dizer Qualquer coisa sobre Qualquer assunto (QQQ). RDF está de acordo com esse *slogan* porque permite que qualquer fonte se refira a recursos em quaisquer *namespaces*. Mesmo um único triplo (*subject/predicate/object*) pode se referir a recursos em vários *namespaces*.

Como um modelo de dados, RDF fornece uma clara especificação do que tem que ser feito para mesclar informações de múltiplas fontes. RDF por si só não fornece algoritmos ou tecnologia para implementar esses processos. As tecnologias necessárias são: *RDF Parser/Serializer*; *RDF Store*; *RDF Query Engine*; *Application* (Allemberg; Hendler, 2011). Neste texto, não serão abordadas as questões referentes a essas tecnologias; o objetivo é fazer uma análise de como essa proposta de organização dos recursos da *Web*, no contexto da *Web Semântica*, que tem como base o RDF, está relacionada com as práticas da representação descritiva abordada na *Ciência da Informação*.

Nesse sentido, serão abordadas questões a respeito da representação descritiva e, na sequência, sua relação com a *Web Semântica*.

## Representação descritiva

A representação da informação registrada tem como principal objetivo proporcionar a comunicação entre os documentos e os usuários da informação. Pode ser subdividida em representação descritiva e representação temática. A descritiva representa as características específicas do documento que permitem sua individualização e também define e padroniza os pontos de acesso responsáveis pela busca e recuperação da informação e pela reunião de documentos semelhantes. A temática detém-se na atribuição de assuntos aos documentos a partir da classificação bibliográfica, da indexação e da elaboração de resumos, facilitando a recuperação de materiais relevantes que dizem respeito a temas semelhantes.

Embora ambas sejam complementares, neste artigo, serão abordadas apenas as questões relativas à representação descritiva, uma vez que ela:

[...] refere-se aos aspectos da descrição formal dos documentos, o que inclui a descrição física e a descrição dos elementos para identificação dos mesmos; a atividade de representação descritiva é também chamada de catalogação (ou, mais especificamente, catalogação descritiva) em especial entre a comunidade de bibliotecas, e de descrição bibliográfica entre a comunidade dos serviços de informação científica (Ortega, 2011, p.45).

O trabalho de representação descritiva depara-se, muitas vezes, com questões relativas a diferentes documentos que precisam de instrumentos normativos, claros, abrangentes, cujos exemplos são a 2ª edição da *Anglo-American Cataloguing Rules* (AACR2) e o formato de registro bibliográfico *Machine Readable Cataloging* (MARC), conhecidos e aceitos internacionalmente. Além disso, o profissional da área pode recorrer ao compartilhamento ou captura de registros bibliográficos para facilitar seu trabalho por meio da catalogação cooperativa, bem como por meio dos catálogos *online* disponíveis por diferentes bibliotecas.

Os termos representação descritiva e representação temática, provavelmente, tenham sido adotados com o objetivo de ampliar o conceito de catalogação, o

qual remete à produção de catálogos de biblioteca. Entretanto:

[...] o desenvolvimento teórico e metodológico da representação descritiva e da representação temática foi realizado de modo separado, conduzindo a dificuldades conceituais na articulação entre os dois processos. Já a Catalogação, incluindo a catalogação descritiva e a catalogação de assunto, responde pela construção global do sistema, nesse caso, o catálogo (Ortega, 2011, p.45).

A catalogação é o processo por meio do qual se descreve formalmente um documento ou recurso e se estabelece um variado e variável número de pontos de acesso, objetivando proporcionar, ao usuário final, a condição de encontrar, identificar, selecionar e obter o documento ou o recurso descrito, ou a informação nele contida. Portanto, é um processo que gera o registro bibliográfico que, por sua vez, identifica e descreve um documento, tanto na sua forma física, como na de conteúdo, funcionando como sua representação na medida em que permite a um usuário encontrar, identificar, selecionar e obter uma obra por meio de um instrumento de recuperação, nomeadamente o catálogo. Em outras palavras, uma catalogação bem elaborada permite poupar tempo, uma vez que evita a consulta a documentos não pertinentes e permite substituir a consulta a alguns documentos, pelo fato de o usuário já ter encontrado a informação desejada (Souza, 2009).

Naturalmente, bibliotecas cujos objetivos e usuários são diferentes não possuem as mesmas necessidades. Existe um elenco mínimo de dados capaz de identificar e individualizar um documento, a partir do qual qualquer Unidade de Informação pode definir as informações adicionais necessárias aos seus objetivos e aos seus usuários.

Considera-se, a partir do exposto, que a representação descritiva é uma parte fundamental no tratamento dos documentos ou recursos, ao permitir sua identificação dentro da memória documental.

No panorama atual, a representação descritiva vem sofrendo nos modelos convencionais de descrição bibliográfica, mudanças significativas que são provocadas pelas tecnologias de informação e comunicação, o que

obrigatoriamente requer olhares diferenciados no tratamento dos recursos informacionais, inclusive a adoção de suas práticas na organização da informação na *Web*.

### Representação descritiva e a *web* semântica

Como destacado anteriormente, o alicerce da *Web Semântica* é o RDF, que é uma forma de representação dos metadados sobre recursos *Web* (tais como título, autor, data de modificação de uma página da *Web*, *copyright* etc.) (Manola; Miller, 2004).

Lendo essa definição à luz dos conceitos apresentados relativos tanto à *Web Semântica* quanto à representação descritiva, fica clara a relação das atividades desenvolvidas em ambas as áreas.

Concorda-se com Mey (1995, p.6) quando afirma que “a riqueza da catalogação repousa nos relacionamentos entre os itens estabelecidos, de forma a criar alternativas de escolha para os usuários”. Isso quer dizer que, se um usuário busca determinada obra, ele terá a oportunidade de optar por diferentes versões ou suportes, ou seja, as diferentes manifestações do mesmo item, como livro, dvd, CD-ROM, documento *online*, bem como em diferentes edições, traduções, línguas, e isso certamente ampliará o universo de pesquisa e de escolha desse usuário. Os relacionamentos entre os itens também lhe permitem localizar itens por ele desconhecidos, uma vez que a catalogação propicia a reunião de todos os itens sobre um determinado assunto, aumentando e melhorando a busca.

Enquanto Mey (1995) assevera que a riqueza da catalogação está nos relacionamentos entre itens, o fundamento da *Web Semântica* está nos dados linkados. Observa-se, então, que tanto as atividades da catalogação quanto as da *Web Semântica* preocupam-se em propiciar aos usuários formas mais ricas e amplas de recuperação da informação.

O uso de códigos normalizados na representação descritiva é fundamental por regularizar e padronizar procedimentos em sistemas de informação, especialmente de médio e grande porte, possibilitando maior consistência, eficiência e qualidade ao sistema, ao fornecer

registros coerentes entre si; tal uso é fundamental também por viabilizar a cooperação catalográfica entre Unidades de Informação, diminuir a necessidade de tomar decisões individuais para cada documento a ser catalogado e facilitar os processos de informatização.

Mey (1995, p.39) observa que:

À descrição cabe extrair diretamente do item todas as informações, de interesse para o usuário, que individualizem o item tornando-o único entre os demais. Cada item possui apenas uma descrição e cada descrição se aplica a um único item.

As recomendações do W3C também objetivam a padronização de metadados de forma a tornar a descrição de cada item única, mas, ao mesmo tempo, formatada de modo a permitir que todos os dados fiquem acessíveis para que se criem vários produtos ou aplicações.

O W3C propõe que todos os dados e metadados constantes ou citados na *Web* possam ser processados por máquinas e para tanto sugere como fundamento o RDF, que é um modelo de dados para a descrição de recursos da *Web*. Portanto, a base da *Web Semântica* é a representação descritiva dos recursos a partir de um modelo em que os registros de metadados são representados por um conjunto de declarações formadas pelos já mencionados triplos (*Subject/Predicate/Object*).

No contexto tradicional da organização da informação nas Unidades de Informação (bibliotecas, arquivos, museus etc.), como os profissionais da informação poderão proceder para que os dados por eles processados, disponibilizados e recuperados possam ser processáveis por máquinas, conforme propõe o W3C?

Os tradicionais códigos de catalogação, vocabulários, tesouros, índices etc., deverão ser formatados dentro dos padrões propostos pelo W3C para a *Web Semântica*. No entanto, deve-se manter a adoção dos códigos de catalogação e das normas utilizadas para a descrição dos recursos bibliográficos. Contudo, tanto os pontos de acesso quanto os elementos de descrição deverão ser formatados no modelo RDF, ou seja, cada ponto de acesso e cada elemento de descrição devem ser representados em declarações RDF. Desta forma, os dados bibliográficos seriam legíveis por máquinas conforme recomendações do W3C. Esta constatação reforça o pressuposto de que a práxis da representação descritiva não só pode, como

deve, ser uma contribuição para a *Web Semântica*, no sentido de que se mantenham características e funcionalidades das bases bibliográficas que vêm sendo aprimoradas há décadas.

Inicialmente devemos pensar em como estão organizados os bancos de dados bibliográficos atualmente. Estes bancos de dados contêm os registros bibliográficos, produtos da descrição, compostos de conjuntos de metadados, tais como o formato MARC (presente nos OPAC) e o *Dublin Core* adotado pela *Open Archive Initiative* (OAI) como a base da interoperabilidade dos repositórios.

Coyle (2010c, p.15) argumenta que, apesar de o MARC estar estruturado em campos e subcampos que são legíveis por máquinas, tem sua base em linguagem textual, o que dificulta a aplicação das tecnologias da *Web Semântica*. Porém o fato de os dados bibliográficos serem fundamentalmente textuais não significa que sejam livres; estes possuem valores controlados, a exemplo do controle de autoridade e de cabeçalhos de assuntos.

Um outro componente importante para que possamos visualizar a relação entre a representação descritiva e a proposta de organização de recursos da *Web Semântica* são os novos desenvolvimentos da área: o *Functional Requirements for Bibliographic Records* (FRBR) desenvolvido pela *International Federation of Library Associations* (IFLA), é resultado de um estudo sobre os requisitos funcionais para registros bibliográficos empreendido de 1992 a 1997. O objetivo do estudo “era produzir uma estrutura (*framework*) que proporcionasse uma compreensão clara, definida com precisão e comumente compartilhada de como o registro bibliográfico pode proporcionar informação sobre o que se espera que o registro possa arquivar em termos de resposta às necessidades do usuário” (Le Boeuf, 2007, p.1) (um dos principais objetos de estudo, atualmente, na área da representação descritiva) e o *Resource Description and Access* (RDA) que é a proposta do novo padrão que visa fornecer um conjunto de diretrizes e instruções na descrição de recurso e acesso cobrindo todo tipo de conteúdo e mídia.

Supõe-se que para incluir os tradicionais catálogos no ambiente da *Web Semântica*, ou seja, transformar os

dados bibliográficos em dados linkados na *Web* faz-se necessário modelar os dados já existentes que se utilizam de MARC ou *Dublin Core*, por exemplo, nos moldes do RDF. Deve-se, ainda, tornar o modelo conceitual FRBR e o código de catalogação RDA, recentemente publicado como sucessor do AACR, em base RDF; além de escrever os vocabulários controlados existentes em linguagens propostas pelo W3C (RDF e *RDF Schema*, *Simple Knowledge Organization System* (SKOS), OWL e RIF). Neste sentido, várias ações já vêm sendo desenvolvidas. A seguir apresentaremos algumas ações identificadas nesta pesquisa.

Uma importante iniciativa é da *Library of Congress* (LC), que disponibiliza uma lista de cabeçalhos em URI (<http://id.loc.gov/>), ou seja, para cada cabeçalho, foi criado um URI, que é um identificador único (base para as declarações em RDF). Exemplo para o nome Jorge Amado: para Amado, Jorge, 1912-2001 a URI é <<http://id.loc.gov/authorities/names/n50024126>>. Portanto, ao se desenvolverem aplicações de dados bibliográficos em RDF, pode-se utilizar o *namespace* dos URI criados pela LC.

Adaptando o exemplo sugerido por Coyle (2010c), pode-se dizer que para uma declaração em RDF do ponto de acesso “autor” para a obra Gabriela Cravo e Canela, poder-se-iam utilizar alguns identificadores já existentes. A seguir apresenta-se a declaração (*Subject/Predicate/Object*) com seus respectivos identificadores URI:

Subject: o livro (identificador do *WordCat*) - <http://www.worldcat.org/title/gabriela-cravo-e-canela/oclc/11427374>

Predicate: a propriedade (identificador do *RDAVocab*) author - <http://rdvocab.info/roles/author>

Object: o valor correspondente à propriedade (identificador da LC) - <http://id.loc.gov/authorities/names/n50024126>

Esta iniciativa vem ao encontro do principal desafio da *Web Semântica*, que são os identificadores únicos para cada recurso. Segundo Coyle (2010c), para darmos visibilidade aos dados das bibliotecas na *Web*, primeiramente, é necessário que eles tenham identificadores na forma de URI. Neste sentido, vale destacar as atividades do *Dublin Core Metadata Initiative* (DCMI), organização responsável pela manutenção e desenvolvimento do conjunto de metadados *Dublin Core*, que é um

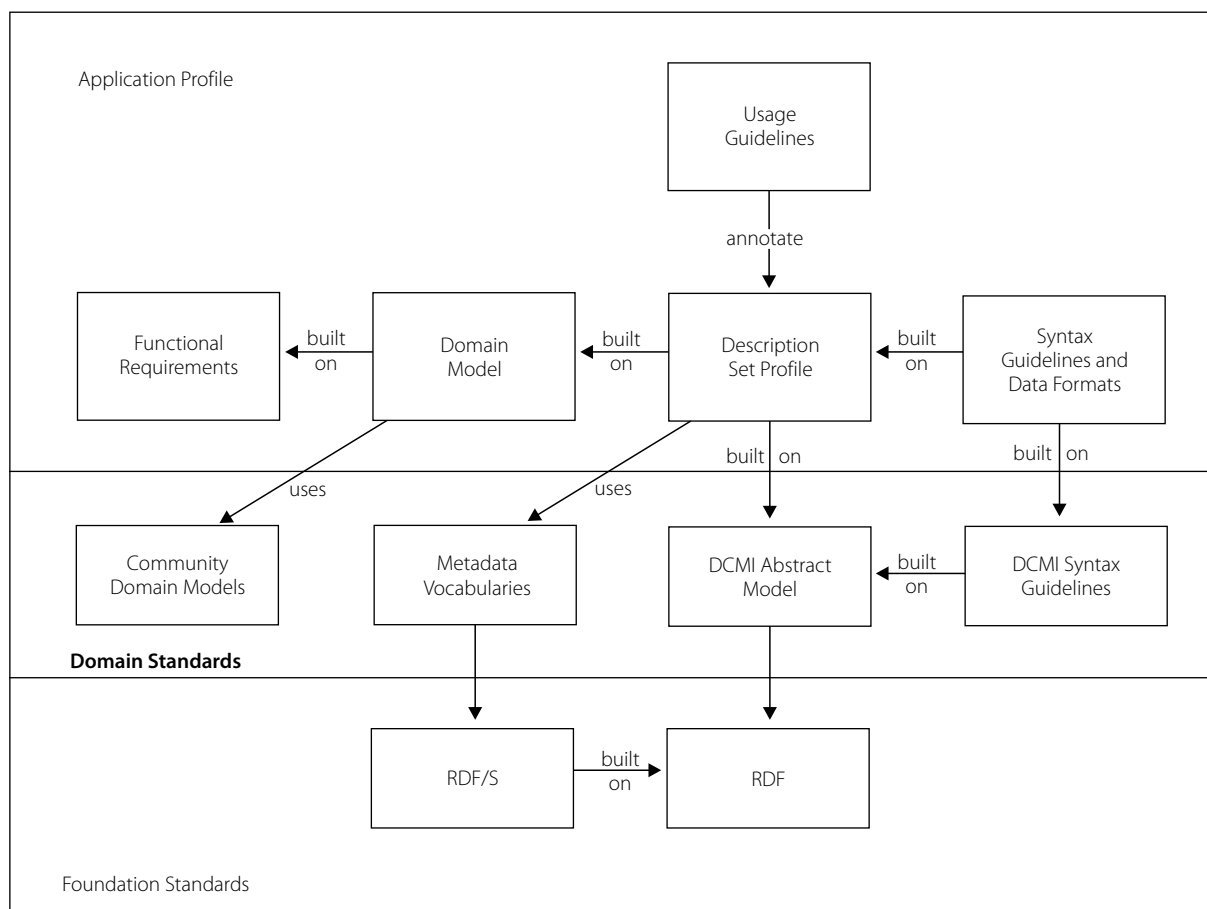
conjunto de metadados criado para a descrição de recursos da *Web* e que é atualmente a base para a interoperabilidade dos repositórios digitais (temáticos ou institucionais).

A *Dublin Core Metadata Initiative* criou um identificador para cada um de seus elementos de metadados, disponíveis em <<http://purl.org/dc/terms>>. Cada um dos termos é definido de forma a especificar detalhadamente o seu uso. Para alguns termos, o DCMI define, inclusive, o tipo de valor que pode ser utilizado na descrição, como, por exemplo, para o metadado *Date*, sugere o uso da norma ISO 8601, que estabelece como descrever uma data (Coyle, 2010c). Os termos DCMI também estão representados em linguagem *RDF Schema* (Dublin Core Metadata Initiative, 2011).

Uma outra iniciativa DCMI é a publicação de um documento contendo recomendações para expressar os

metadados *Dublin Core* em RDF (Nilsson *et al.*, 2008) por meio de uma representação do modelo DCMI, ou seja, o *Dublin Core Abstract Model* (DCAM) em RDF.

O RDF também está presente no documento *Guidelines for Dublin Core Application Profiles*, que apresenta diretrizes para a criação de perfis de aplicação; descreve os componentes-chave; bem como delinea o processo de desenvolvimento de um perfil de aplicação *Dublin Core* (Coyle; Baker, 2008). O documento apresenta, ainda, o *Singapore Framework*, um modelo que relaciona os elementos descritivos de um perfil de aplicação (camada *Application Profile*) com os padrões de domínio (camada *Domain Standards*) e com as normas que fundamentam a *Web Semântica* (camada *Foundation Standards*), ou seja, a base para que um perfil de aplicação seja processável por máquina e, conseqüentemente, inserido na *Web Semântica* é o RDF e o *RDFS* (Figura 4).



**Figura 4.** *Singapore Framework*.

Fonte: Coyle e Baker (2008).



Relativamente ao RDA, vale destacar a iniciativa do DCMI em conjunto com o *Joint Steering Committee* (JSC). Num encontro promovido, em 2007, pela *American Library Association* (ALA) em conjunto com a *British Library*, os membros do DCMI e do JSC se dispuseram a criar uma expressão do RDA em RDF incluindo as associações deste com as entidades e relacionamentos do FRBR. O resultado é um registro *online* de RDA em RDF, que é a primeira definição de dados bibliográficos no formato para *Web Semântica* (Coyle, 2010a). Um artigo foi publicado na *D-Lib Magazine* relatando o processo de criação do RDA *Vocabularies* (Hillmann *et al.*, 2010).

Na Figura 5, pode-se visualizar o detalhamento da propriedade *Author (Work)* apresentado no *Open Metadata Registry*.

Em relação ao FRBR, existem críticas ao modelo, mas também algumas iniciativas para expressá-lo em RDF. Coyle (2010b, p.19, 2010d, p.25) cita: “*expression of Core FRBR Concepts in RDF*” (ver <http://vocab.org/frbr/core.html>) e o FRBRoo, que é parte do “*CIDOC Conceptual Reference*

*Model*” (ver [http://cidoc.ics.forth.gr/frbr\\_inro.html](http://cidoc.ics.forth.gr/frbr_inro.html)). A primeira iniciativa é um vocabulário que inclui as classes de entidades dos grupos 1, 2 e 3 do FRBR, bem como as propriedades correspondentes para os relacionamentos entre essas entidades. A segunda, o FRBRoo, é uma ontologia criada para a captura e representação da semântica subjacente das informações bibliográficas e para facilitar a integração, mediação e intercâmbio de informações bibliográficas e de museus. Ambas consideram necessário incluir mais algumas entidades no modelo FRBR para poder formatá-lo adequadamente em RDF.

Pode-se considerar que as iniciativas acima descritas são importantíssimas e essenciais para inserir as bases bibliográficas na *Web Semântica*. No entanto há, ainda, muito a ser feito. Para a Ciência da Informação, torna-se um campo fértil para desenvolvimento de pesquisas que, inclusive, resultarão no aprimoramento da formação dos profissionais arquivistas, bibliotecários e museólogos, no sentido de estes serem capacitados para

Element Sets: RDA Roles	
Elements: Author (Work)	
Detail	Statements
Metadata +	
<b>Detail</b>	
Label:	Author (Work)
Name:	authorWork
URI:	<a href="http://rdvocab.info/roles/authorWork">http://rdvocab.info/roles/authorWork</a> (RDF)
Description:	A person, family, or corporate body responsible for creating a work that is primarily textual in content, regardless of media type (e.g., printed text, spoken word, electronic text, tactile text) or genre (e.g., poems, novels, screenplays, blogs). Use also for persons, etc., creating a new work by paraphrasing, rewriting, or adapting works by another creator such that the modification has substantially changed the nature and content of the original or changed the medium of expression.
Comment:	
Type:	subproperty
Parent:	Creator
Domain:	<a href="http://rdvocab.info/uri/schema/FRBRentitiesRDA/Work">http://rdvocab.info/uri/schema/FRBRentitiesRDA/Work</a>
Range:	
Status:	New-Proposed
Language:	English
Note:	

**Figura 5.** Propriedade Autor da Obra no *Open Metadata Registry*.

Fonte: <<http://metadataregistry.org/schemaprop/show/id/121.html>>.

atuar na *Web Semântica*, ou seja, a aplicação dos conceitos e práticas da organização da informação na organização da informação da *Web*.

## Considerações Finais

No que se refere ao objetivo deste estudo, pode-se concluir que as práticas da representação descritiva estão relacionadas com a proposta de organização dos recursos da *Web*, no contexto da *Web Semântica*. Esta constatação se deve ao fato de que, para que os dados bibliográficos contidos nos catálogos possam ser transformados em dados linkados da *Web*, é necessário que os catalogadores atuem na implantação do modelo de descrição RDF nos códigos e normas de catalogação vigentes atualmente.

A representação descritiva é uma ação que está presente no projeto da *Web Semântica* na medida em que a base para a existência deste projeto está no uso do RDF (um modelo de descrição). Portanto, a práxis do profissional da informação que cataloga recursos com base nos códigos e modelos que norteiam este fazer (ISBD, AACR, RDA, FRBR, MARC etc.) pode ser aplicada para colaborar com o projeto do W3C. A chave está na conversão dos dados bibliográficos atuais em descrições baseadas no modelo RDF.

Observou-se, ainda, que as ações que já vem sendo realizadas têm caminhado no sentido de permitir que os dados bibliográficos contidos nos catálogos bibliográficos possam ser inseridos em declarações nos moldes do RDF.

A transformação dos catálogos das bibliotecas e de outras bases de dados para dados linkados permitirá uma infinidade de aplicações. Os usuários poderão, ao consultar um catálogo, não somente obter informações a respeito de materiais bibliográficos (impressos ou

digitais) que estão contidos numa coleção específica; mas poderão, por exemplo, ao realizar uma busca, saber quais títulos relacionados existem na biblioteca local, em outras bibliotecas, em livrarias, em coleções particulares; obter acesso aos dados disponíveis no *Google*, *Google Acadêmico*, *Open Library*, *Googlereader*, *Library Thing*, entre outros; ser informados de grupos de discussão existentes sobre o tema de interesse; descobrir pesquisas em andamento; obter o *ranking* dos mais consultados etc.

Os catálogos poderão estar com seus metadados disponíveis por meio dos mecanismos de busca, e este seria o caminho contrário, ou seja, os usuários que estão realizando suas pesquisas nos mecanismos de busca seriam direcionados aos dados linkados das bibliotecas. Esta é uma ação essencial, pois é fato que a maioria dos usuários têm dado prioridade às buscas na *Web*, acessando cada vez menos os tradicionais catálogos para suprir suas necessidades informacionais.

Transformar os dados bibliográficos em dados linkados não enriquecerá apenas os serviços ofertados pelas bibliotecas aos seus usuários, mas, inclusive, a própria *Web*, na medida em que será possível, por meio dos mecanismos de busca, acessar os dados bibliográficos presentes nos catálogos de milhares de bibliotecas do mundo todo.

O resultado desta pesquisa não esgota o assunto, porém aponta para a real colaboração da representação descritiva para a *Web Semântica*, inclusive demonstrando que algumas organizações, tais como LC e DCMI, já estão realizando algumas ações iniciais. Portanto, é imprescindível que outros estudos sejam realizados, tanto por pesquisadores da área de Ciência da Informação (em especial a Organização da Informação) quanto do W3C, de forma a propor ações sistematizadas e globais para que os dados bibliográficos possam estar inseridos no conjunto dos dados linkados.

## Referências

ADIDA, B.; BIRBECK, M. *RDFa primer*. Massachusetts: W3C, 2008. Available from: <<http://www.w3.org/TR/xhtml-rdfa-primer>>. Cited: 27 Apr. 2010.

ALLEMANG, D.; HENDLER, J. *Semantic web for the working ontologist: effective modeling in RDFS and OWL*. 2nd.ed. Amsterdam: Morgan Kaufmann, 2011.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, v.284, n.5, p.34-43, 2001.

CATARINO, M.E. *Integração das folksonomias aos metadados: identificação de novos elementos de metadados como contributo para a descrição de recursos em repositórios*. 2009. Tese (Doutorado em Tecnologias e Sistemas de Infor-

mação) - Departamento de Sistemas de Informação, Escola de Engenharia, Universidade do Minho, Minho, 2009.

COYLE, K. Library data in the web world. *Library Technology Reports*, v.46, n.2, p.5-11, 2010a.

COYLE, K. Metadata models of the world wide web. *Library Technology Reports*, v.46, n.2, p.12-19, 2010b.

COYLE, K. RDA vocabularies for a twenty-first-century data. *Library Technology Reports*, v.46, n.1, p.14-29, 2010c.

COYLE, K. Understanding the semantic web: bibliographic data and metadata. *Library Technology Reports*, v.46, n.2, p.20-25, 2010d.

COYLE, K.; BAKER, T. *Guidelines for Dublin Core application profiles*. 2008. Available from: <<http://dublincore.org/documents/2008/11/03/profile-guidelines/>>. Cited: 11 Oct. 2011.

DAS, S.; SUNDARA, S.; CYGANIAK, R. *R2RML: RDB to RDF mapping language*. Massachusetts: W3C, 2010. Available from: <<http://www.w3.org/TR/2010/WD-r2rml-20101028/>>. Cited: 10 Oct. 2011.

DUBLIN CORE METADATA INITIATIVE. *DCMI terms declarations represented in RDF schema language*. 2011. Available from: <<http://dublincore.org/schemas/rdfs>>. Cited: 18 Oct. 2011.

HAHN, R. *et al.* Faceted wikipedia search. In: INTERNATIONAL CONFERENCE ON BUSINESS INFORMATION SYSTEMS, 13., 2010, Berlin. *Proceedings electronics...* Berlin: Springerlink, 2010. Available from: <<http://www4.wiwi.wiwi.fu-berlin.de/bizer/pub/hahn-et-al-faceted-wikipedia-search-BIS2010.pdf>>. Cited: 19 June 2011.

HALPIN, H.; DAVIS, I. *GRDDL primer*. Massachusetts: W3C, 2007. Available from: <<http://www.w3.org/TR/grddl-primer>>. Cited: 8 July 2010.

HILLMANN, D. *et al.* RDA vocabularies. *D-Lib Magazine*, v.16, n.1-2, 2010. Available from: <<http://www.dlib.org/dlib/january10/hillmann/01hillmann.html>>. Cited: 20 Oct. 2011.

KING, R. Q&A with Tim Berners-Lee. *BusinessWeek*, 2007. Available from: <[http://www.ece.uc.edu/~mazlack/ECE.716.Sp2010/Semantic.Web.Business.Week.2007/QA\\_with\\_Tim\\_Berners\\_Lee.pdf](http://www.ece.uc.edu/~mazlack/ECE.716.Sp2010/Semantic.Web.Business.Week.2007/QA_with_Tim_Berners_Lee.pdf)>. Cited: 8 July 2010.

LE BOEUF, P. *O admirável mundo novo do FRBR: versão 5*. 2007. Disponível em: <[http://www.imeicc5.com/download/portuguese/Presentations2c\\_BraveNewFRBRWorld\(PR\)\\_Port.pdf](http://www.imeicc5.com/download/portuguese/Presentations2c_BraveNewFRBRWorld(PR)_Port.pdf)>. Acesso em: 9 out. 2011.

MANOLA, F.; MILLER, E. *RDF primer*. 2004. Available from: <<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>>. Cited: 1 Sept. 2010.

MEY, E.S.A. *Introdução à catalogação*. Brasília: Briquet de Lemos, 1995.

MORGENSTERN, L.; WELTY, C.; BOLEY, H. *RIF primer*. 2010. Available from: <<http://www.w3.org/2005/rules/wiki/Primer>>. Cited: 14 Oct. 2011.

MURCHO, D. *Dicionário escolar de filosofia*. Lisboa: Plátano, 2011. Disponível em: <<http://www.defnarede.com/i.html>>. Acesso em: 10 out. 2011.

NILSSON, M. *et al.* *Expressing Dublin Core metadata using the resource description framework*. Singapore: DCMI, 2008. Available from: <<http://dublincore.org/documents/dc-rdf>>. Cited: 14 Oct. 2011.

ORTEGA, C.D. Do princípio monográfico à unidade documentária: exploração dos fundamentos da catalogação. *Liinc em Revista*, v.7, n.1, p.43-60, 2011.

PRUD'HOMMEAUX, E.; SEABORNE, A. *SPARQL query language for RDF*. Sverige: W3C, 2008. Available from: <<http://www.w3.org/TR/rdf-sparql-query>>. Cited: 19 Oct. 2011.

RDF WORKING GROUP. *Resource description framework*. Massachusetts: W3C, 2004. Available from: <<http://www.w3.org/RDF/>>. Cited: 10 Oct. 2011.

SAWAYA, M.R. *Dicionário de informática e internet*. São Paulo: Nobel, 1999. Disponível em: <<http://pt.scribd.com/doc/2989340/Informatica-Dicionario-Internet>>. Acesso em: 16 fev. 2012.

SCHEPPE, K. *Protocol for web description resources (powder): primer*. 2009. Available from: <<http://www.w3.org/TR/2009/NOTE-powder-primer-20090901/>>. Cited: 8 July 2010.

SOUZA, T.B. *O ensino de representação descritiva nos cursos da área de ciência da informação no Brasil e em Portugal: estudo comparativo*. 2009. Tese (Doutorado em Ciências Documentais) - Faculdade de Letras, Universidade do Porto, Porto, 2009.

WORD WIDE WEB CONSORTIUM. *W3C semantic web activity*. 2001. Available from: <<http://www.w3.org/2001/sw/>>. Cited: 12 Feb. 2008.

WORD WIDE WEB CONSORTIUM. *W3C glossary dictionary*. 2003. Available from: <<http://www.w3.org/2003/glossary/subglossary/REC-xml-names.rdf>>. Cited: 10 Oct. 2011.

WORD WIDE WEB CONSORTIUM. *Semantic web*. 2010. Available from: <<http://www.w3.org/standards/semanticWeb/>>. Cited: 1 Sept. 2010.

WORD WIDE WEB CONSORTIUM. *Semantic web terminology*. 2011. Available from: <[http://www.w3.org/2001/sw/wiki/Semantic\\_Web\\_terminology](http://www.w3.org/2001/sw/wiki/Semantic_Web_terminology)>. Cited: 10 Oct. 2011.

# A reference ontology for digital scientific journals applied to systematic literature review processes

## *Uma ontologia de referência para periódicos científicos digitais aplicada em processos de revisão sistemática da literatura*

Fernando Benedet GHISI<sup>1</sup>

Gleisy Regina Bóries FACHIN<sup>2</sup>

Marcos Henrique dos SANTOS<sup>1</sup>

Denilson SELL<sup>3</sup>

Gregório Jean Varvakis RADOS<sup>3</sup>

### Abstract

This paper presents possible approaches to the use of a reference ontology for digital journals, supporting bibliographic processes in systematic literature reviews. The benefits are highlighted of using specialized services through "batch" or "on-the-fly" processing of information from different repositories, such as recognized and indexed databases and portals or specific websites. It is concluded that the use of reference ontology enables the creation of services that ensure greater interoperability between different repositories, allowing a more inclusive and accurate retrieval of information, as it standardizes the concepts related to the access points of scientific journals.

**Keywords:** Information retrieval. Journal. Reference ontology. Systematic review of the literature.

### Resumo

*O presente artigo apresenta possíveis abordagens para o uso de uma ontologia de referência para periódicos científicos digitais, apoiando os processos de levantamento bibliográfico em revisão sistemática da literatura. Destacam-se os benefícios da utilização de serviços especializados por meio de processamentos em "batch" ou "on-the-fly" de informações provenientes de diferentes repositórios, como bases de dados reconhecidas e indexadas e portais ou sítios específicos. Conclui-se que a utilização de ontologia de referência possibilita a criação de serviços que asseguram maior interoperabilidade entre diferentes repositórios, permitindo a recuperação da informação de forma mais abrangente e precisa, através da padronização dos conceitos relacionados aos pontos de acesso de periódicos científicos.*

**Palavras-chave:** Recuperação de informação. Periódicos. Ontologia de referência. Revisão sistemática da literatura.

<sup>1</sup> Mestrandos, Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. Campus Universitário Reitor João David Ferreira Lima, Trindade, 88040-970, Florianópolis, SC, Brasil. Correspondência para/Correspondence to: F.B. GHISI. E-mail: <fernandoghisi@egc.ufsc.br>.

<sup>2</sup> Professora Doutora, Universidade Federal de Santa Catarina, Departamento de Ciência da Informação. Florianópolis, SC, Brasil.

<sup>3</sup> Professores Doutores, Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. Florianópolis, SC, Brasil.

Received on 19/12/2011, resubmitted on 4/5/2012 and approved on 23/5/2012.

## Introduction

Since the earliest inscriptions and cave drawings made by man, the need to locate, understand and disseminate information has been present in the evolution of human civilization. During the course of history, information has been represented, stored and recovered in some way. Likewise, the organization of documents and information has accompanied the evolution of society, and has been centered on the field of Librarianship, Information Science and Computing, beginning with the first libraries and advancing to modern databases, directories and repositories (Fachin, 2011).

In a digital context, the knowledge areas are responsible for researching, developing, applying and validating resources as a way to ensure that they are self-sustaining. For as long as they have existed, the knowledge areas have been recording their creations and actions, especially in scientific publications. Considering that the evolution, resources and requirements of the digital world of scientific information are constantly changing, new tools, standards, systems and movements - such as the *Open Access* and *Open Archives Initiative* movements, discussed by many authors, such as Guédon (2010) - have been created. Therefore, irrespective of the field, theme, group, profession, language or country, everyone is faced with a proliferation of technologies related to facilitating communication and availability, where one goes from intensive reading to extensive, expanded, intertextual and hypertextual reading (Miranda, 2010).

Nonetheless, undoubtedly there is constant and increasingly faster proliferation of information and technological resources on the Web every day. Moreover, every field of knowledge invests in its own sources of referential information, in order to meet users' needs, adopting, customizing or developing their own applications, without concern about interaction or interoperability with other systems (databases, digital libraries, repositories, portals or specific sites). The research groups, entities and governments are responsible for developing actions in order to maintain the consolidation of quality, certification and legitimation of information, as well as the efficiency of technological resources for broad and accurate recovery of such information. This is especially important in the scientific context, where the

"efforts of the organizations to consolidate rules, regulations and standards of scientific publishing occur worldwide, with the aim of ensuring its standardization and intercommunicability", according to Miranda (2010, p.10).

Among the initiatives to increase the formality and credibility of researches supported by bibliographical reviews, there are "systematic literature reviewing methods" (Centre for Reviews and Dissemination, 2009; Higgins; Green, 2011). In one of the steps, a systematic search for relevant works is made, according to the study subjects on a pre-selected list of publications. This step is directly related to the mechanisms of information retrieval - and the range, quality and precision of these can have a great impact on the entire process. In this sense, one of the main problems or challenges of the Information Retrieval Systems - and more recently, of the mechanisms of intelligent information retrieval, which use intelligent agents - is only to retrieve the relevant documents looked for by the user.

To achieve proper information retrieval, the information must be processed and organized. Classifications, thesaurus and taxonomies are among the most used techniques that can be used in the treatment and organization of information. Zeng and Qin (2008) and Campos (2009) have stated that the structuring of digital documents may be done through the use of metadata. In this case, it would allow the search to be conducted in different ways: term in the title, term in the body, type of media, type of archive, period of publication, place of publication, language, among others. As a basic function, metadata provides information about the digital information and empowers an accurate retrieval, by the exchange of data among the information systems. More recently, with the emergence of the Semantic Web, ontologies have been increasingly applied to support the retrieval of significant information, according to a specific domain (Gruber, 1996; Guarino, 1998).

According to Bräscher and Carlan (2010, p.153), knowledge organization systems handle representations of "knowledge domains that delimit the meaning of terms in the context of areas, establish conceptual relationships that help position a concept in the conceptual system and are used as tools for organizing and retrieving information". The adoption of standardized

concepts in identifying the resources available on the Web (semantic markup) aims to enable the information collection agents operating in the network (Web crawling, Web spider, Web robot) to become capable of intelligently understanding and processing the contents described and represented by the metadata. Campos (2009) collaborated in this discussion by highlighting the importance of terminological studies, multidisciplinary research and systematic studies in order to improve the standards used by the search agents.

In this context, from the problem related to the recovery of scientific information, the aim of this article was to present possible approaches to better use the vast amount of scientific material available on the Web, contributing to systematic processes for the review of literature. In addition to elucidating the use of Semantic Web resources and techniques (such as the use of reference ontology), some steps of the systematic review process that could be further automated by the use of more integrated and intelligent services, will be identified. Thus, according to Creswell (2010), this study is characterized as a qualitative, exploratory and applied research, because it seeks to model a segment of knowledge in an interdisciplinary way, using literature, operationalizing a reference ontology to standardize the essential elements of scientific journals, and demonstrating its applicability in scenarios of information retrieval that are present in systematic review of the literature.

### Systematic review of literature

In the information and knowledge era, with the Web representing a huge repository of documents and information, researchers and scientists are facing an informational context that includes a miscellany of works without scientific quality or relevance. Databases - such as Scopus, Web of Science, Science Direct and EBSCO - try to mitigate part of this problem, including documents according to some quality of criteria and providing specialized search environments. However, when researchers propose a survey question, and investigate related previous publications on which to base their studies, or to support their hypotheses, they may not have had access to all the studies. They may even tentatively leave out some studies with differing

viewpoints on purpose. To attenuate this problem and increase the formality and credibility of bibliographical reviews, as well as increase their application, the methods of systematic review of literature have emerged.

The systematic review is characterized as an investigation of the evidence of a clearly formulated question using systematic and explicit methods to identify, select and critically evaluate previous relevant researches, extracting and analyzing data from studies that are included in the review (Centre for Reviews and Dissemination, 2009). For Sampaio and Mancini (2007, p.84), systematic review is a form of research that uses the literature on a particular theme as a data source, providing a "summary of evidence related to a specific intervention strategy, by applying systematic and explicit methods of searching, critical appraisal and synthesis of information selected". In addition, several authors argue that this process, when documented, can be repeated and followed by other researchers.

Berwanger *et al.* (2005), conducted a study on the importance of systematic reviews, presenting a history in which they pointed out that since the seventies, psychologists have drawn attention to the need for systematization of the steps necessary to minimize systematic errors (biases) and random errors in scientific research reviews, in particular in the health area. As reported by the authors, in 1987, Yusuf and Peto pointed out the need for systematic reviews of randomized trials as an approach to obtain clinically significant responses. Furthermore, in 1987, Mulrow addressed the problem of the poor quality of clinical study reviews. In 1988, Oxman and Guyatt published guidelines to assist readers in critical evaluations (systematic reviews), concerning the validity, applicability and results. According to Freitas and Vieira (2010), the systematic review is a subject extensively discussed and used in the health area, appearing in the area of Computer Science in 2004, more specifically in the area of Software Engineering, with the work *Procedures for Performing Systematic Reviews*, by Kitchenham (2004).

Another contribution to the understanding and construction of systematic reviews is *Cochrane Handbook for Systematic Reviews of Interventions*, which focuses the health area; however, these instructions could be used in other areas, serving as guide for building more concise

reviews. According to this manual, a systematic review brings together all the empirical evidence, according to predefined criteria, to answer a question, promoting more reliable results. Among others, these features of systematic reviews are highlighted: a set of goals with pre-defined eligibility criteria; an explicit and reproducible methodology; a survey of all studies that meet the criteria adopted, within the investigated issue; the constant evaluation of the risks of biases; and the presentation synthesizing the results of the studies - included in a systematic way (Higgins; Green, 2011).

For Kitchenham (2004) and Centre for Reviews and Dissemination (2009), a systematic review can be conducted in three major steps: planning, conducting and publication of the results. In the planning step, the protocol of the review is defined, i.e., the purpose of the review and the procedures to be adopted for the research are described. In the conducting step, the search for relevant works is made (bibliographic review) - considering the object of study, the selection of publications and the extraction and registration of data from each of the selected sources. And finally, in the publishing step, the considerations about the selected sources and conclusions are described.

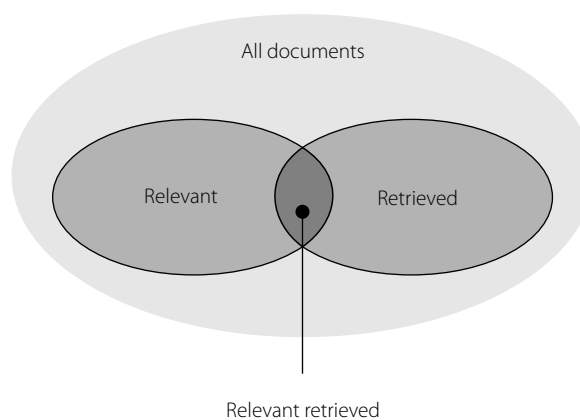
Freitas and Vieira (2010) have stated that even if one adopts a systematic review, it is not possible and nor is it sure that all the works about the subject searched will be obtained. According these authors, some works may not have been found due to the number of selected sources, limitations of the repositories related to the quality of its search system resources, difficulties in defining the search syntax, and invariably, the researcher bias when choosing articles. Thus, it appears necessary to improve the process of systematic literature search, improving the conducting of the information retrieval process as a whole.

## Information retrieval

Information retrieval is an essential tool for research and the evolution of science in every area of knowledge. One of its main issues is to retrieve documents that are relevant to the users. Viewed from a global perspective, information retrieval is about enabling and realizing the meeting of a question with the stored information and providing positive return to the requesting user.

The book *Modern Information Retrieval*, by Baeza-Yates and Ribeiro-Neto (1999), is the one of the preeminent works on information retrieval, providing a broad discussion about this subject. For the authors, information retrieval deals with the representation, storage, organization and access to information. They also make a distinction between data retrieval and information retrieval, stating that to retrieve data consists mainly of identifying documents of a particular collection that have as keywords the same terms as those used by the user in a query. Thus, according to those authors, the recovery process does not satisfy the information needs of users, it only retrieves data. For these authors, information retrieval consists of obtaining relevant information about a particular subject, and not simply acquiring documents that statically answer a query requested.

Grossman and Frieder (2004), presented two categories of documents that corresponded to any given query: the relevant documents and retrieved documents (Figure 1). In a perfect system, the two sets would be equivalent (returning only relevant documents), according to the authors; however, in reality, the systems also return many irrelevant documents. Thus, two metrics are used to evaluate systems, *precision* and *recall*. The first is the ratio between the number of relevant documents returned and the total number of documents returned. The second is the number of relevant documents returned in the total number of documents believed to be relevant.



**Figure 1.** Categories of documents that correspond to any issued query.

Source: Based on Grossman and Frieder (2004).

For Belkin (1996), the process developed by an information retrieval system is complex and should handle: the representation (users stating the problem); the comparison (the representation of user-informed problem with the content); the interaction (user with someone - mediator - or machine - human-computer interaction); the judgment (user analysis and relevance); and the modification (change of the search depending on the results and/or reformulation of the question).

Hoeschl (2006) emphasizes the historical aspects of search engines. The first generation based on directories such as Yahoo; the second based on robots and automated technologies such as *Altavista*; the third generation brought the metasearchers; the fourth led to refinement in the organization of the results (All the Web); the fifth marked by Google, in its way, combined sophistication and a wide coverage, brought its Page Rank algorithm to the Web; the sixth, under development, is likely to be the merging of the various types of media into a single search - text, images, videos and sound; the seventh generation, according to Hoeschl (2006), is marked by quality in the selection of information, through intelligent content analysis, based on ontologies and artificial intelligence techniques. The author reports that many advanced technologies are at present seeking more qualified information retrieval, such as the use of Structured Contextual Search, Dynamically Contextualized Knowledge Representation, Text Mining and Case Based Reasoning (RBC).

With new information retrieval techniques emerging every day, one can envision solutions for recovering documents and scientific information more effectively, as well as optimization in conducting systematic processes of literature review (specifically at the systematic search stage). With the use of ontological resources - such as reference ontologies - supporting and connecting the information retrieval services, promoting interoperability between them and facilitating the processing, retrieval and dissemination of information - is possible to use more intelligent search agents for conducting systematic and replicable surveys, optimizing time and avoiding rework, as well reducing research biases, promoting a greater credibility of the results.

In this interim period, an example is presented of reference ontology for digital scientific journals and its

application is demonstrated in supporting systematic literature review processes based on smart information retrieval techniques.

## Reference ontology

Many resources have been applied by numerous public and private institutions since the ease of access to information and its vast availability on the Web, creating a profusion of resource types, standards and guidelines, hindering the efficient retrieval of relevant information.

The credibility assigned to pages, sites, repositories, portals, and even databases is questionable, pushing people, research groups, and especially institutions and research centers to study ways to provide effective recovery on the Web, as well as to unify and integrate these resources - enabling a greater interoperability - in order to enable a wide and accurate recovery from the databases of the different knowledge areas. In this sense, ontologies have recently been broadly applied.

Fonseca (2007) mentions that ontology can be defined as a tool that helps to describe a specific world and that information systems will only be as good as the associated ontologies. Guarino (2008) collaborates in this context, when he states that reference ontology addresses the aspects of scientific theory, as it seeks to improve the adequacy of the representation of the object under study. According to the author, reference ontology can be renamed as foundational ontology, referring to the ontological theories whose focus is to define or clarify the intended meanings of terms used in very specific areas. Corroborating this context, in the present study, the choice was to use an ontology that "describes a specific world", i.e., describing the conceptual structure of digital scientific journals, standardizing and consolidating its metadata, to support its validation as a vehicle for disseminating knowledge, promoting information exchange and dissemination of knowledge.

Thus, the ontology of digital scientific journals to be used, proposed by Fachin (2011), is characterized as a reference ontology, as it consolidates a set of metadata that standardizes the basic elements related to digital scientific journals, providing understanding and sharing in this area - where there is interaction between people and systems. As described in Fachin *et al.* (2010) and Fachin



(2011), a wide and deep research has been conducted, consolidating these metadata and developing the ontology based on Fernández-López *et al.* (1997) and Noy and McGuiness (2005) methodologies. The tools used were the ontoKEM - Ontology for Knowledge Engineering and Management (<http://ontokem.egc.ufsc.br>) - and the Protégé. The Figure 2, created with the *Jambalaya* plugin for *Protégé*, shows some classes of this ontology, with its relations of hierarchy. Each concept used and their relationships, as well as other concepts that have been suppressed in this figure, are presented in detail by Fachin *et al.* (2010) and Fachin (2011).

In this ontology, the classes and subclasses were created following a hierarchical structure, starting from the broader knowledge - Digital Scientific Journal - to the more specific, trying to group the classes semantically, and thus, showing a structure of all the necessary information for the creation, maintenance and management of digital journals. The following scenarios, shown below, demonstrate the applicability of this ontology in information retrieval contexts present in systematic processes of literature review.

## Ontology applicability

Considering the reference ontology of digital scientific journals included in scientific work information retrieval contexts, and supporting an improvement in processes of systematic review, there is the possibility of two main approaches to the operationalization of this scenario: on one hand with a “batch information processing”, and, on the other hand, with an “on-the-fly information processing”. In the first approach, (Figure 3), the searches would be performed in a centralized database of information previously loaded, following the concepts of the reference ontology. In this context, there are two possibilities for the database loading: providing a service for journals and/or indexers to submit their information (passive mode) or using a service to perform collection, integration and storage of this information (active mode).

In the first case, the journals and indexers would have to adapt themselves to the submission service available (which would follow the concepts of the reference ontology), creating other services for automatic

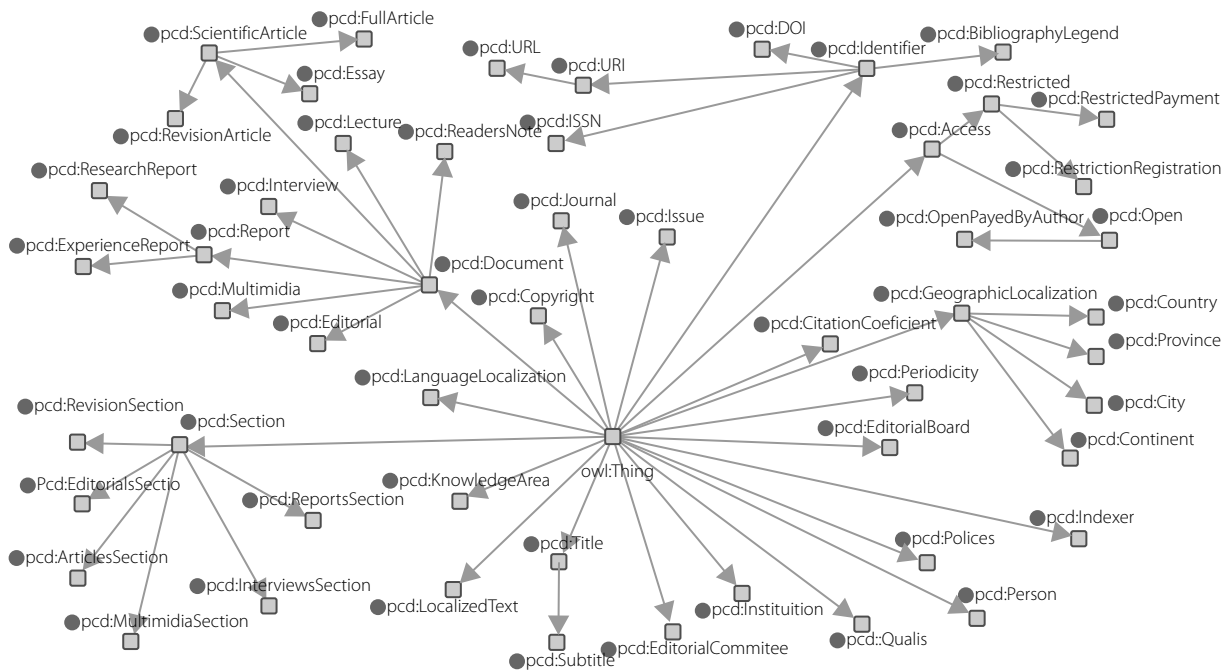


Figure 2. Some classes of the ontology for digital scientific journal.

Source: Based on Fachin (2011).

upload information - in the case of larger repositories - or even making manual entries - in the case of smaller repositories. In the second case, a service would be developed to support the collection, integration and storage of information from different repositories. As journals and indexers do not follow a standard for the publication of information, there is a need to create other services specialized in collecting and processing the data from each repository - dealing with the issues of semantic and syntactic differences. The two possibilities for loading the centralized database (passive mode and active mode) - used in the approach with batch processing of information (Figure 3).

With this approach, involving the use of a centralized repository of previously loaded and integrated information, faster information retrieval is enabled, as well as a greater level of control over the availability and integrity of this information. However, the main disadvantage is the loss of dynamism in adding new content or even in changes or updates of previously published content.

The alternative to this scenario is the "on-the-fly" approach. In this second approach, (Figure 4), the search would be performed in real time and results would be processed dynamically with the aim of integrating and determining the relevance of information, following the concepts of the reference ontology. In this context, two possibilities for operationalization of this service are envisioned: the standardization of services of the different repositories - so that queries would be performed in a uniform way (following the concepts of the reference ontology); or the use of non-standard services - so that specialized operations for querying and processing the information on each service would be required.

In the first case, the repositories that have interest in providing their information would have to create a standardized service to perform queries, according to the system protocols and based on the reference ontology, and the system would simply orchestrate the queries among the services, as well as the information retrieval and integration. In the second case, specialized services would be developed for collecting and transforming the information available in the (non-standard) services of the different repositories - tailoring this information to the concepts available in the reference ontology.

After this stage of "pre-processing", the information could then be integrated and made available by the system. The two possibilities for using the services from the different repositories (using standard or non-standard services) - in the approach with "on-the-fly information processing".

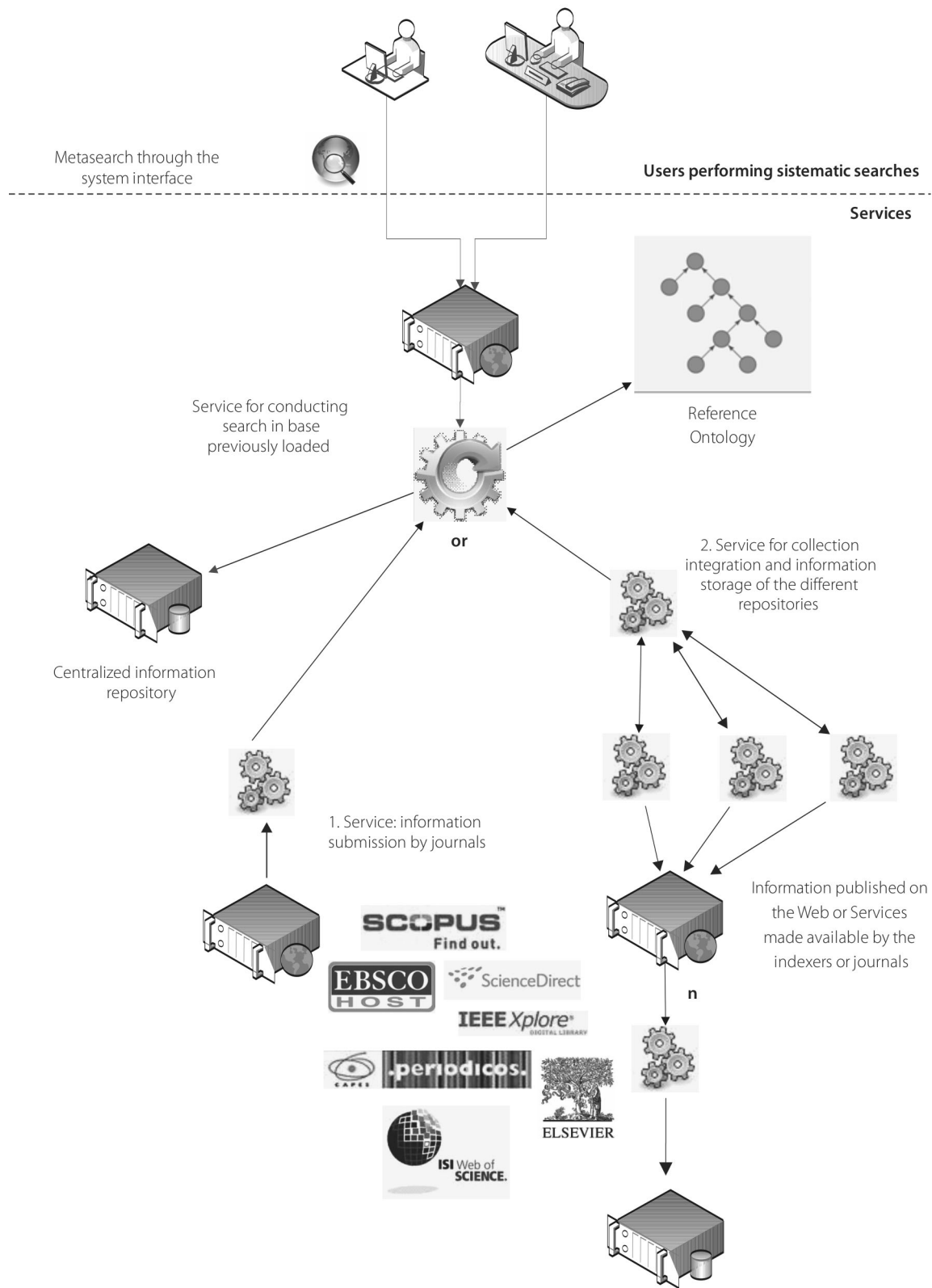
As previously mentioned, by means of this approach, without involving the use of a centralized repository of information (previously loaded and integrated), the searches would be performed in real time and results would be processed dynamically, for integration and relevance determination of the information, following the concepts of the reference ontology - thus obtaining the most up-to-date information available from the different providers. However, there would be a loss in the information recovery speed, as well as less control over the availability and integrity.

Moreover, by consolidating a set of metadata related to digital scientific journals, the reference ontology could provide a standardized representation, availability and access to information in the field of scientific publications, enabling greater interoperability between the different systems and repositories and the development of more accurate and inclusive information retrieval services.

Consequently, improving the mechanisms for systematic search and directly assisting in the process of systematic review of literature (both at the conducting and publishing stages), researchers would not need to deal with different search techniques and tools available in the databases.

In addition, it would be possible to create even more intelligent search agents that would recover the most relevant studies according to different inclusion and exclusion criteria (pre-defined by the researchers), following the ontology concepts and making inferences. These agents could incorporate explicit and formal methods of search, extraction, grouping and summarization of the studies.

Supporting the researchers, even at the stage of critical evaluation of the studies, it would also be possible to generate a "trail" of all procedures performed - making it possible to replicate and update (running over more



**Figure 3.** Approach for realization of meta search using a centralized information repository - "batch information processing".

Source: By authors.

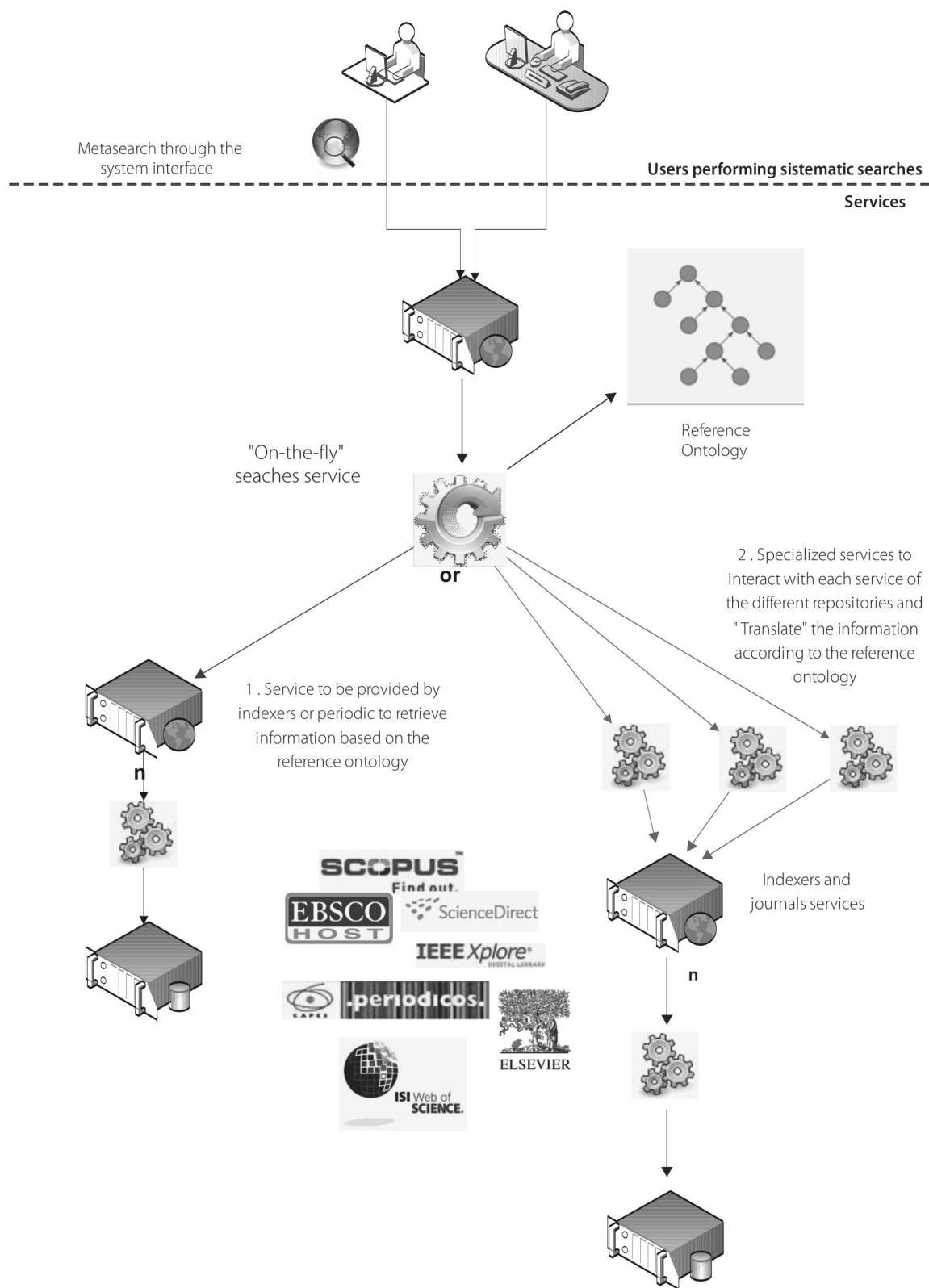


Figure 4. Approach for performing the metasearch “on-the-fly”, without a centralized information repository.

Source: By authors.

recent works) the process automatically, as well as validate the methods performed by the agent. Thus, benefits would also be obtained at the stage of publication of procedures used in the systematic review, without taking into account a possible increase in the credibility of bibliographic reviews - because the automation of some steps might reduce the researcher bias, which may occur in the selection and evaluation of literature, and with the manual use of different tools.

## Conclusion

This article presented possible approaches to the use of a reference ontology for digital scientific journals to enhance information retrieval, supporting the development of systematic search in processes of systematic literature review.

The ontology intends to support the standardization of concepts related to scientific journals, which would ensure greater interoperability between different data repositories and a more accurate and comprehensive

information retrieval, in addition to smart inferences in the systematic review steps.

The proposed ontology could also support intelligent search agents, which would recover the most relevant studies according to different inclusion and exclusion criteria, based on the systematic review goals. These agents could also generate a "trail" of all procedures performed - enabling automatic replication and update of the process, as well as the validation of the process conducted by the agent. The aim is also to benefit the stage of publication of the systematic review processes, and make it possible to increase the credibility of bibliographic reviews - due to the automation of some steps that might reduce researcher bias, which may occur in the selection and evaluation of literature, and in the application of different tools.

The approach discussed in this study is now under evaluation to support scenarios of information integration and inference as well, with regard to the evaluation of courses and journals, based on the Brazilian Graduate Program Evaluation - *Coleta Capes*.

## References

BAEZA-YATES, R.; RIBEIRO-NETO, B. Retrieval evaluation. In: BAEZA-YATES, R.; RIBEIRO-NETO, B. *Modern information retrieval*. New York: Addison-Wesley, 1999. p.73-98.

BELKIN, N.J. *Intelligent information retrieval: whose intelligence?* 1996. Available from: <<http://mariner.rutgers.edu/tipster3/iirs.html>>. Cited: Apr. 2008.

BERWANGER, O.; GUIMARÃES, H.P.; AVEZUM, A. *Revisões sistemáticas em cardiologia*. 2005. Disponível em: <[http://www.moreirajr.com.br/revistas.asp?id\\_materia=3254&fase=imprime](http://www.moreirajr.com.br/revistas.asp?id_materia=3254&fase=imprime)>. Acesso em: abr. 2011.

BRÄSCHER, M.; CARLAN, E. Sistemas de organização do conhecimento: antigas e novas linguagens. In: ROBREDO, J.; BRÄSCHER, M. (Org.). *Passeios no bosque da informação: estudos sobre representação e organização da informação e do conhecimento*. Brasília: Ibict, 2010. cap. 8.

CAMPOS, M.L.A. Aspectos semânticos da compatibilização terminológica entre ontologias no campo da bioinformática. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 10., 2009, João Pessoa. *Anais eletrônicos...* João Pessoa: ANCIB, 2009. Disponível em: <<http://hdl.handle.net/123456789/477>>. Acesso em: nov. 2009.

CENTRE FOR REVIEWS AND DISSEMINATION. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York: University of York, 2009. Available from: <[http://www.york.ac.uk/inst/crd/pdf/Systematic\\_Reviews.pdf](http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf)>. Cited: May 2011.

CRESWELL, J.W. *Projeto de pesquisa: métodos qualitativo, quantitativo e misto*. 3.ed. Porto Alegre: Artmed, 2010.

FACHIN, G.R.B. *Ontologia de referência para periódico científico digital*. 2011. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina, Florianópolis, 2011.

FACHIN, G.R.B. et al. Periódicos científicos digitais e ontologia: uma pesquisa. In: SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL, 3., 2010, Florianópolis. *Anais...* Florianópolis: UFSC, 2010. p.339-349.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. In: SPRING SYMPOSIUM ON ONTOLOGICAL ENGINEERING, 1997, Stanford. *Proceedings...* Stanford: AAAI, 1997. Available from: <[http://oa.upm.es/5484/1/METHONTOLOGY\\_.pdf](http://oa.upm.es/5484/1/METHONTOLOGY_.pdf)>. Cited: May 2012.

FONSECA, F. The double role of ontologies in information science research. *Journal of the American Society for Information Science and Technology*, v.58, n.6, p.786-793, 2007.

FREITAS, L.A.; VIEIRA, R. Revisão sistemática sobre métricas para ontologias. In: SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL, 3., 2010, Florianópolis. *Anais...* Florianópolis: UFSC, 2010. p.331-338.

GROSSMAN, D.A.; FRIEDER, O. *Information retrieval: algorithms and heuristics*. 2nd.ed. Dordrecht: Springer, 2004.

- GRUBER, T. *What is an ontology?* 1996. Available from: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Cited: 21 June 2010.
- GUARINO, N. *Formal ontology and information systems*. 1998. Available from: <<http://www.loa-cnr.it/Papers/FOIS98.pdf>>. Cited: Nov. 2010.
- GUARINO, N. The basic tools of ontological analysis. In: SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL, 2008, Niterói. *Anais eletrônicos...* Niterói: UFF, 2008. 108p. Disponível em: <<http://www.uff.br/ontologia>>. Acesso em: nov. 2010.
- GUÉDON, J.C. Acesso aberto e divisão entre ciência predominante e ciência periférica. In: FERREIRA, S.M.S.P.; TARGINO, M.G. (Org.). *Acessibilidade e visibilidade de revistas científicas eletrônicas*. São Paulo: Senac, 2010. p.21-78.
- HIGGINS, J.P.T.; GREEN, S. (Ed.). *Cochrane handbook for systematic reviews of interventions*: version 5.1.0. 2011. Available from: <<http://www.mrc-bsu.cam.ac.uk/cochrane/handbook/>>. Cited: May 2011.
- HOESCHL, H.C. *Ontoweb: a nova era das ferramentas de busca*. *Consultor Jurídico*, 2006. Disponível em: <[http://www.conjur.com.br/2006-fev-01/ontoweb\\_ferramentas\\_busca](http://www.conjur.com.br/2006-fev-01/ontoweb_ferramentas_busca)>. Acesso em: maio 2011.
- KITCHENHAM, B. *Procedures performing systematic review*. 2004. Available from: <[citeseerx.ist.psu.edu/](http://citeseerx.ist.psu.edu/)>. Cited: May 2011.
- MIRANDA, A. Verbivocovisualidade das revistas no século XXI. In: FERREIRA, S.M.S.P.; TARGINO, M.G. (Org.). *Acessibilidade e visibilidade de revistas científicas eletrônicas*. São Paulo: Senac, 2010. p.9-12.
- NOY, N.F.; MCGUINNESS, D.L. Ontology development 101: a guide to creating your first ontology. 2005. Available from: <[http://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101.pdf)>. Cited: May 2012.
- SAMPAIO, R.F.; MANCINI, M.C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Revista Brasileira de Fisioterapia*, v.11, n.1, p.83-89, 2007. Disponível em: <<http://www.scielo.br>>. Acesso em: maio 2011.
- ZENG, M.L.; QIN, J. *Metadata*. New York: Neal-Schuman Publishers, 2008.



# PuertoTex: un *software* de minería textual para la creación de resúmenes automáticos en el dominio de ingeniería de puertos y costas basado en ontologías

## *PuertoTex: a data mining software based on ontologies for automatic summarization on port and coastal engineering domain*

Amed LEIVA-MEDEROS<sup>1</sup>

Sandor DOMÍNGUEZ-VELASCO<sup>1</sup>

Jose Antonio SENSO<sup>2</sup>

### Resumen

El objeto de este trabajo es elaborar y evaluar los resultados de la implementación de un *software* basado en ontologías, capaz de generar resúmenes automáticos en el campo de la Ingeniería de Puertos y Costas. Para el desarrollo de la herramienta se emplean diversas técnicas emanadas del análisis de discurso así como técnicas cognitivas, que permiten generar reglas para el tratamiento de los textos. También se apela a la construcción de una ontología que facilite los procesos de etiquetado a partir de las potencialidades de *Resource Description Framework* y *Extensible Markup Language*. Se construye un conjunto de agentes que actúa sobre la ontología, de la cual se declaran sus principales elementos. Como producto generado se presenta Puertotex, un *software* para la construcción de resúmenes automáticos basado en ontologías. La evaluación de los resúmenes generados refleja la calidad del sistema, que tiene como única limitación su capacidad para trabajar con el dominio objeto de investigación.

**Palabras-clave:** Autómatas. Desambiguación de textos científicos. Minería de texto. Ontologías. Resumen automático.

### Abstract

*The aim of this work to develop and evaluate the results of the implementation of an ontology-based software capable of generating automatic summaries in the field of Ports and Coastal Engineering. For the development of the tool has been used several techniques emanating from the analysis of discourse and cognitive techniques for generating rules for the treatment of the texts. It also calls for the construction of an ontology that facilitates the processes of labeling from the potential of Resource Description Framework and Extensible Markup Language. It has built a set of agents acting on the ontology, which declared its main elements. Are presented as a product generated, Puertotex, software for the construction of ontology-based automatic summaries. The evaluation of the summaries generated reflects the quality of the system, which is limited only by your ability to work with the domain under investigation.*

**Keywords:** Automata. Disambiguation of scientific texts. Data mining. Ontologies. Automatic summarization.

<sup>1</sup> Profesores Doutores, Universidad Central Marta Abreu de las Villas, Facultad de Ciencias de la Información. Santa Clara, Cuba.

<sup>2</sup> Profesor Titular, Universidad de Granada, Departamento de Biblioteconomía y Documentación. Campus de Cartuja, 18071, Granada, España. Correspondencia a nombre de/Correspondence to: J.A. SENSO. E-mail: <jsenso@ugr.es>.

Recibido el día 19/12/2011, re-apresentado el 4/5/2012 y aceptado para su publicación el 23/5/2012.



## Introducción

A diferencia de la indización y la clasificación, los resúmenes son una reelaboración textual, una propuesta por reducción significativa de los originales que parte de algunas cuestiones previas, como las características del texto a resumir, el destinatario del resumen, los objetivos que se planteen y el proceso que le da lugar, y datan, como servicio organizado, desde mediados del siglo XIX (Hernández, 2007a).

Los resúmenes se erigen como metadatos, pues son la expresión que enuncia el contenido de los documentos textuales (Herrera, 2007). En la literatura de la Ciencia de la Información constantemente se hace referencia a sistemas de metadatos como *Dublin Core*, *Encoded Archival Description* (EAD), Norma Internacional General de Descripción Archivística (ISAD-G), *Machine Readable Cataloging* (MARC 21) etc. Aunque muchos de estos sistemas incluyen etiquetas específicas para incluir un breve resumen del objeto referenciado, en pocos casos se habla del resumen documental como metadato que representa de forma resumida el contenido de los textos, pues es poco su reconocimiento como estructura representacional de datos no estructurados. Esta forma de procesamiento de la información ha sido tratada tradicionalmente como un proceso netamente lingüístico, computacional y semántico.

La producción automática de resúmenes se ha basado en métodos estadísticos desde las investigaciones de Lunh (1958) y continúa hoy con los trabajos del Procesamiento del Lenguaje Natural, las aplicaciones basadas en analizadores automáticos y en modelos socio-pragmáticos del discurso basados en técnicas de "dominio de personalidad propia", subyugados por la extracción de palabras o cadenas significativas, que ya en los años 1990 recortan las "estructuras retóricas" (Mann; Thompson, 1990; Endres-Niggemeyer *et al.*, 1995; Endres-Niggemeyer, 2005), muy relacionadas con el contenido de los documentos, al aportar indicios sobre el esquema conceptual y organizador de las distintas unidades que conforman un texto para reproducir y ordenar determinadas frases (Ono *et al.*, 1994).

Este desarrollo ha sustentado la aparición de múltiples herramientas para la extracción de conocimiento en materias específicas. Sin embargo, existen disciplinas que no poseen *software* que permitan especificar y detallar el conocimiento que poseen, tal es

el caso de La Ingeniería de Puertos y Costas. Para Sánchez-Cuadrado *et al.* (2007) parten de la premisa, compartida en otros trabajos similares, de que la mayor parte del conocimiento está explicitado en los documentos de un dominio mediante términos y relaciones y que sólo el conocimiento que no esté expresado en los documentos tendrá que ser aportado por los expertos del dominio.

La Ingeniería de Puertos y Costas es una disciplina que difiere del resto de las ingenierías. La inexistencia de diccionarios específicos, la ausencia de teorías específicas para este campo multidisciplinar y la necesidad que tienen los profesionales que en ella trabajan de poseer herramientas terminológicas y léxico semánticas para realizar su trabajo le confieren un cierto halo de área discriminada, al menos hablando desde el punto de vista de la documentación y la terminología (Leiva *et al.*, 2009). Esta situación da pie a los objetivos del presente texto. Por un lado, explicar los pasos seguidos para elaborar un *software* basado en ontologías capaz de generar resúmenes automáticos en el campo de la Ingeniería de Puertos y Costas. Por otro, evaluar las cualidades de este *software* dentro del campo de la Minería de Texto.

## Métodos

Para la realización de este trabajo se ha utilizado una muestra aleatoria del corpus del proyecto Puertoterm (<http://lexicon.ugr.es/projects>) de la Universidad de Granada (2 890 documentos, de los cuales se seleccionan 1 390) a la que se aplicó técnicas de análisis de discurso para determinar reglas cohesivas amparadas en la teoría de Mann y Thompson (1990). Se ha utilizado el concepto de discurso asumido por Dijk, quien lo denota como una forma específica de uso del lenguaje, un suceso de comunicación (quién utiliza el lenguaje, cómo lo utiliza, por qué y cuándo lo hace) en el que se transmiten ideas, visto también como una forma específica de interacción social. Así, el discurso se interpreta como un evento comunicativo real (completo) en una situación social dada.

La función fundamental del lenguaje realizado como discurso es la constitución de una realidad semiótica. Ésta se estructura de tal manera que podamos comprenderla y apoderarnos de ella (Metzeltin, 2003). De esta forma, el análisis del discurso, básicamente, se ocupa

de la dimensión interactiva e intersubjetiva del uso del lenguaje mediante la investigación y análisis de datos reales (Pilleux, 2001).

El estudio del discurso encuentra razón en los fenómenos detrás de la oración, aunque no deja de observarlo como un factor dinámico de las interacciones sociales que permiten encontrar modelos para su interpretación y análisis, teniendo en cuenta que los discursos pueden describirse en términos de "acciones sociales que llevan a cabo los usuarios del lenguaje cuando se comunican entre sí en situaciones sociales y dentro de la sociedad y la cultura en general" (Dijk; Kintsch, 1983, p.38).

Unido a estas técnicas se aplican métodos de estudio de usuarios a 23 expertos en el terreno de la Ingeniería de Puertos y Costas conjuntamente con la modelación de Sistema. Para la realización del *software* se utilizó el lenguaje de programación Python.

Para evaluar el sistema se usaron varias metodologías. Por un lado se ha analizado la usabilidad del sistema por medio de un estudio heurístico. Luego se ha analizado la capacidad del sistema para recuperar información determinada. Y para finalizar se analiza la calidad del producto resultante, el resumen automático, empleando dos métodos de evaluación.

### **PuertoTex: componentes del sistema**

PuertoTex es un sistema de Minería de Textos basado en reglas de discurso emanadas de los textos que consumen y generan diversos grupos de investigadores dentro del campo de la Ingeniería de Puertos y Costas. Las prestaciones que facilita este *software* están en consonancia con las necesidades de la comunidad de usuarios a los que va dirigido este tipo de información. Los agentes de *software* que forman este sistema emulan las estrategias resumidoras de 12 expertos en Ingeniería de Puertos y Costas. A continuación analizaremos sus componentes y fases de trabajo.

### **Corpus de trabajo**

Los corpus que han servido de sustento para esta investigación poseen características muy disímiles en lo referente a nivel discursivo, prestigio de sus fuentes y

tipología documental. Esta situación hace que las estrategias de trabajo y análisis sean más complejas, debido a la disparidad en la concepción del corpus (en inglés y español). Si se analiza el corpus desde el punto de vista de estructura discursiva puede llegarse a la conclusión de que el inglés es superior, pues el discurso que genera es eminentemente especializado. Más de 4 000 registros están en esta categoría, de los cuales 818 son semi-especializados y solo 43 son artículos divulgativos. El corpus español es, en esencia, divulgativo y en menor grado especializado.

### **La ontología**

A partir de los documentos de ese corpus se pudo crear Ontosatcol, ontología diseñada para la desambiguación, extracción y búsqueda de información y empleada por la herramienta PuertoTex. Esta ontología (Figura 1) cuenta con 3 005 conceptos, 3 006 subconceptos y 28 900 anotaciones. Posee también 12 propiedades individuales para cada objeto y 8 propiedades especiales para los datos. La herramienta dispone, además, de 890 instancias.

La Ontología del Sistema es generada a partir de los términos que han sido etiquetados en el Corpus, y para ello el sistema incluye un marcador o etiquetador que facilita dicho proceso. Una vez marcados los vocablos dentro del texto, estos comienzan a organizarse teniendo en cuenta la Semántica de Marcos, un modelo de organización del conocimiento que permite que los términos sean descritos de acuerdo a diversas formas de actuación dentro de un contexto, formando un concepto único (Faber *et al.*, 2005). Cada término representa una disciplina, por tanto debe poseer uno o varios eventos (o marcos conceptuales) que describen procesos o acciones que tienen lugar dentro de la disciplina representada por el término. Dichos eventos poseen una estructura que facilita la creación y organización de los diferentes conceptos que los integran. La elaboración del marco conceptual es el culmen del análisis de las definiciones lexicográficas y de corpus extraídas previamente (Senso, 2009).

Una vez finalizada la fase de construcción de la semántica de marcos comienzan a estructurarse los conceptos y se construyen todas las estructuras

lingüísticas de las clases a partir de diversos elementos o relaciones de diversos tipos, como merónimos, hiperónimos o hipónimos. Para estructurar las mencionadas relaciones se contrastan los términos contra obras especializadas en Ingeniería de Puertos y Costas, así como en diccionarios, glosarios, enciclopedias y listas de folksonomías. También se declaran las propiedades de los datos y las de los objetos.

## Etiquetado

Para desarrollar los procesos de resumen se llevó a cabo el etiquetado en *Extensive Markup Language* (XML) de los elementos del texto de cada uno de los documentos seleccionados del corpus, con el fin de organizar las unidades textuales en segmentos. Dichos segmentos añan elementos de diversos órdenes como el discursivo, el sintáctico y el comunicativo que, a su vez, son referentes internos de las cargas semánticas y

estructurales de cada oración. El desarrollo de estas etiquetas permite la obtención de resúmenes de un texto único.

Los textos se han etiquetado tanto a nivel oracional como discursivo. En este punto los autores encontraron que no existe ninguna herramienta para el marcado de relaciones discursivas en español o que facilite la marcación de núcleos, satélites, temas y remas en los textos. Con el objetivo de emplear un sistema semiautomático de etiquetado se ha decidido desarrollar una interfaz de anotación utilizando como herramienta el programa *XMLMarker* (<http://symbolclick.com/>), editor XML que, entre otras cosas, permite administrar bases de datos en XML mediante interfaces *Web*, además de ejecutar sentencias en XML y en Python.

Los pasos que se han seguido para la concepción de las etiquetas obedecen a tres criterios fundamentales: la secuencia en que se realiza el etiquetado, la simbología (el nombre de los elementos XML) utilizada para declarar

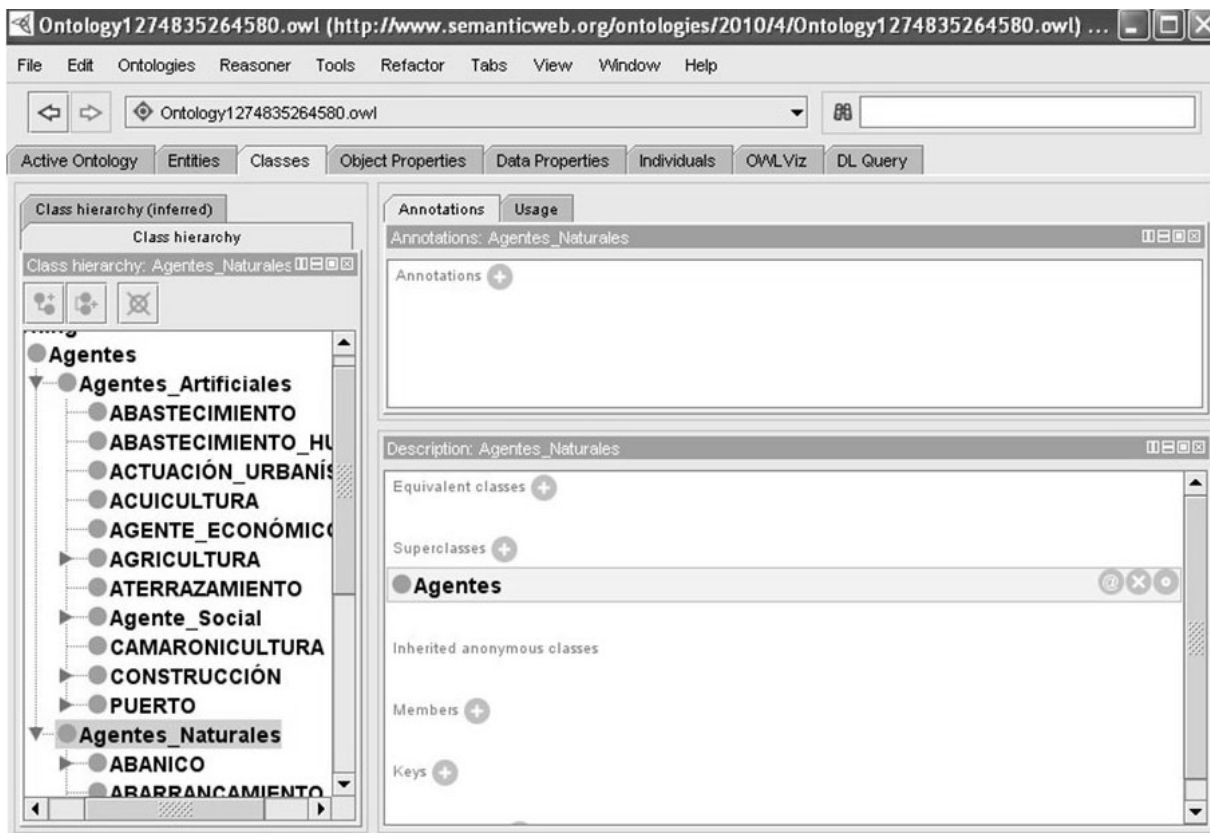


Figura 1. Aspecto de la ontología desde el editor Protégé.

Fuente: Elaboración propia.

los elementos cohesivos del texto y la presencia en el texto de elementos geográficos, verbos, datos estadísticos, fórmulas y procesos.

Lo primero que se marca es la oración que no estará prefijada por un punto, sino por la presencia de formas verbales y sus conectores. A continuación se etiqueta la estructura discursiva de cada sentencia de forma independiente; en otras palabras, se delimitan los núcleos y los satélites.

Como puede intuirse, todas las etiquetas que explican relaciones de tipo núcleo-satélite poseen una estructura similar, a excepción de los casos en que existe un satélite discursivo en relaciones de secuencia, lista o unión. En el caso de la presencia de satélites en la relación de elaboración, se enuncian en las marcas de discurso los elementos tema y rema de su núcleo. A igual que en los trabajos de D'Cunha (2006), esta construcción semántica se ejecuta a través del atributo <tem>, cuyo valor será el tema o el rema de su núcleo (marcados como "t" y "r", respectivamente, junto con el número con el que aparecen en su etiqueta), dependiendo de a cuál de los dos se refiera el satélite (por ejemplo, <item="t">, <item="r">). Es imposible encontrar satélites en relaciones multinucleares, lo cual obliga a declarar en la marcación solo los diversos núcleos, que poseen igual nivel de significado para el resumen.

En este sistema de etiquetado se han declarado elementos de la superestructura del texto, lo que permite que el mismo sea reconocido por estándares como el *Dublin Core*, *Resource Description Framework* (RDF) y *Friend of a Friend* (FOAF).

## Agentes

El objetivo fundamental de esta aplicación es que sea capaz de reproducir e imitar el proceso que sigue un ser humano a la hora de realizar un resumen. Para ello se ha diseñado un conjunto de agentes encargados de realizar diversas tareas, con la meta de conseguir dicho objetivo.

*Agente de Lectura*: se encarga de realizar la primera estrategia cognitiva que hacen los resumidores. Lee el documento y comprueba que esté acorde con el dominio de la ingeniería de puertos y costas. Para ello localiza el texto marcado en XML y contrasta el conocimiento contra

la ontología y la base léxica, determinando que el texto es relevante para el dominio ya que posee términos que son relevantes para él. Envía al agente de resumen el texto para que este lo resuma.

*Agente de Resumen*: este agente toma el texto que le envía el agente de lectura y realiza dos procesos. Primero extrae todas las oraciones candidatas para el resumen de acuerdo a las calificaciones declaradas en el análisis del discurso; o sea, un resumen por extracción. En un segundo paso hace un resumen por abstracción, es decir, reescribe el texto utilizando reglas heurísticas que tiene implementadas. Mediante estas reglas logra la cohesión, la coherencia y el balance textual. En el proceso de extracción, el agente utiliza diversas reglas predefinidas para gestionar la información en dos sentidos. Dichas reglas son de ponderación (permiten elegir las oraciones de acuerdo a la presencia en ellas de términos que hagan referencia a: elementos estadísticos, elementos químicos, nombre geográficos y términos de la lista del discurso), sintáctico-comunicativas (encargadas de la coherencia del texto, manejando criterios que permitan detectar y eliminar los núcleos y satélites en las oraciones y suprimir y eliminar elementos de tema y rema) y de desambiguación (para darle sentido al texto. Se centran en la presencia de pronombres anafóricos para detectar los antecedentes de los grupos oracionales y la asignación de signos de puntuación en las oraciones). El conjunto de autómatas con los que se trabaja en esta fase están basados en una modificación del algoritmo *Multivariate Adaptive Regression Splines* (MARS) (Sidorov; Zazueta, 2006; Domínguez, 2011).

*Agente de Desambiguación*: una vez realizado el resumen, este agente recibe el texto y reconoce en él signos de puntuación erróneos y palabras fuera de contexto utilizando el algoritmo de Aguirre (Aguirre, 1998; Rigau, 2002) y verificando los vocablos por medio de una comparación con las clases de la ontología.

*Agente de Normalización*: se encarga de determinar la estructura del resumen. Recibe el texto del agente de desambiguación y le asigna al texto el segmento de referencia del artículo, así como su *Uniform Resource Locaton* (URL), para que la presentación del resumen permita determinar dónde está el texto fuente y su corpus de referencia.

*Agente de Búsqueda*: es el agente de más trabajo en el sistema, hace búsquedas a través de la ontología y

a través de la base de datos léxica desarrollada. En la ontología se encarga de buscar por autor, título, materia y tipo de documento utilizando las bondades del sistema de metadatos *Dublin Core*. También permite buscar en redes sociales desarrolladas a partir de los datos de los autores que se encuentran en la estructura FOAF. Friend of a Friend facilita la búsqueda y recuperación de los siguientes elementos: institución del autor y título del proyecto, así como el correo electrónico del autor y los autores que colaboran con él.

*Agente de Redundancia*: conecta la solicitud del usuario con los resúmenes que ya tiene almacenados en la base de datos. En el caso de que el documento solicitado ya se encuentre resumido, el agente le proporcionará dicho resumen. Cuando se realice la solicitud de un texto que no está en el corpus, el agente le pedirá al usuario que indique dónde se encuentra, para localizarlo (generalmente por medio de la URI), incluirlo en el corpus e iniciar el proceso con el agente de lectura.

Una vez que un documento entra al sistema se obtiene su contenido semántico y se ubica en la ontología. A continuación se realiza un procesamiento léxico-semántico a partir de la base de datos léxica del sistema (ontología). El siguiente paso consiste en realizar las marcas de discurso en el texto, de acuerdo con la heurística (estrategia cognitiva) tomada de los resumidores humanos que participan en este estudio, generando un documento etiquetado en XML. De ahí el documento se envía al agente de lectura que se encarga de leer el documento emulando el comportamiento de un lector humano. Cuando este agente termina la lectura y comprueba que el texto está etiquetado correctamente, envía el texto al agente de resumen. Éste utiliza varias estrategias de resumen - tal y como se explica en el punto 3.4 - para determinar las oraciones relevantes, asignar pesos a los términos y extraer las oraciones (hace el resumen por abstracción), buscando luego la cohesión mediante el tratamiento de núcleos y satélites de acuerdo con las reglas sintáctico-comunicativas. El resumen resultante de este paso aún no es legible, por tanto el agente de desambiguación tiene que darle sentido a los términos utilizando la ontología. Después de la desambiguación se pasa a la normalización del texto; es decir, se colocan los puntos, las comas y las unidades necesarias para que el texto sea legible. Resumido el texto,

este se almacena y está listo para que el agente de búsqueda localice informaciones (teniendo en cuenta los valores que posibilita la ontología) e interroge al agente de redundancia para cuando se realice una solicitud de información que ya ha sido procesada.

## Las operaciones llevadas a cabo en PuertoTex

La aplicación posee diversas opciones que facilitan el trabajo por medio de un entorno basado en formularios *Web*. Así, es posible el registro, la búsqueda y recuperación de información y la posibilidad de resumir una fuente concreta.

El módulo principal carga la herramienta con las opciones para el ingreso en el sistema, el registro y la ayuda para que llegue a completar todas las acciones que se realizan desde la ventana principal.

El módulo de búsqueda permite una consulta basada en diversos criterios: materia, autor, año y personas. La consulta por materias usa los conceptos expuestos en la ontología, permitiendo el uso de operadores booleanos, y permite visualizar los conceptos y sus dimensiones semánticas de forma dinámica. La búsqueda de autor devuelve el título o lista de títulos que tenga asociados, accediendo al texto completo pinchando dos veces sobre el título escogido.

Como ya se comentó anteriormente, en el proceso de etiquetado se ha empleado FOAF. Así, los nombres de los autores y de las personas o instituciones que se mencionan en los textos están conectados entre sí, formando una red social. De esa manera es fácil obtener información sobre los expertos de un sub-área de conocimiento concreto dentro de la Ingeniería de Puertos y Costas.

Es en este módulo donde también se encuentra la función principal del sistema: resumir una fuente. Una vez que el usuario ha seleccionado un título concreto de la lista de resultados tras su consulta, el agente de búsqueda (apoyado por diversas reglas de inferencia y de las capacidades semánticas de la ontología) filtra la información y localiza el documento en la base de datos, mostrando el resumen del texto, la superestructura del resumen y una URI para que el usuario consulte o

descargue el documento original. Antes de que se resuma la fuente, cada documento pasa por diversos procesos.

Las reglas empleadas para la realización de resúmenes en PuertoTex se han llevado a cabo teniendo en cuenta tres rasgos esenciales, al igual que en los trabajos de D’Cunha (2006) y Paneca (2009): 1) Estructura textual; 2) Unidades léxicas y 3) Estructura discursiva y sintáctico-comunicativa.

Desde el punto de vista de la implementación, se ha intentado hacer que las reglas coincidan con los procesos de realización del resumen, con sus operaciones y la estrategia de construcción del mismo. Estas reglas se unen a las posturas de lectura y son implementadas por agentes de *software* cuyo nivel de especialización en la lectura y construcción del texto ofrece confiabilidad en la calidad del proceso. De esta forma, el agente de lectura y el de resumen asumen diversos niveles textuales para construir un extracto: 1) Nivel textual: consiste en la implementación de reglas basadas en la estructura del texto (denominadas reglas textuales); 2) Nivel léxico: detección, eliminación y selección de unidades léxicas de la base de conocimientos y de la ontología del sistema y 3) Nivel discursivo y sintáctico-comunicativo: formalización de normativas de las reglas discursivo-sintáctico-comunicativas (que hemos denominado reglas PuertoRules).

El análisis de los dos corpus con los que se ha trabajado nos ha mostrado la existencia de una estructura retórica: Introducción, Objetivos, Materiales y Métodos, Resultados y Conclusiones (IOMRC), al igual que lo expuesto en los estudios de Pinto (2001, 2004) y Hernández (2006), y que da sentido a la existencia del agente de lectura de PuertoTex. IOMRC es una extensión de la estructura Introducción, Metodología, Resultados y Conclusiones (IMRC), que facilita la construcción de resúmenes informativos e indicativos. Los primeros se basan exclusivamente en la estructura retórica del texto científico IOMRC y se obtienen mediante métodos de abstracción. Los resúmenes que se realizan sobre textos divulgativos (caso que sucede principalmente con el corpus en español) solo son de categoría indicativa, ya que los textos divulgativos no poseen estructura IOMRC y se construyen exclusivamente a partir de métodos de extracción.

Esta regla, que vamos a denominar *modelo de reconocimiento de texto*, se ha creado con la finalidad de localizar en el texto las unidades retóricas específicas, es decir los apartados introducción, metodología, resultados y conclusiones clásicos de los trabajos científicos.

Partiendo de los criterios de Lunh (1958) y las modificaciones de Mathis *et al.* (1973) se ha constatado la posición estratégica de determinadas oraciones en los segmentos de los textos analizados, las cuales evidencian mayor relevancia que otras en el momento de la selección del resumen.

A modo de táctica operacional, para este *software* se siguen los mismos supuestos teóricos de Lunh (1958) y Salton (1996), donde los modelos de resumen utilizan una aproximación por extracción donde ocurren procesos de transformación de corpus en los cuales se convierte el texto en tokens y se eliminan los signos de puntuación existentes en su volumen. Las acciones propuestas para PuertoTex son las siguientes:

Eliminar oraciones o fragmentos del texto: 1) Asignar un peso ponderado a cada oración del texto según la presencia en ella de vocablos, y su función o posición en el texto y 2) Suprimir oraciones de acuerdo al poco peso en sus niveles semánticos.

A partir de esta forma ya clásica en el tratamiento del resumen se ha decidido desarrollar diversos tipos de reglas para operar con un texto científico: a) Reglas que suprimen secuencias del texto fuente, oraciones y/o pasajes textuales con determinado nivel léxico-semántico; b) Reglas que delimitan aquellas oraciones que en el texto original han de ser eliminadas y almacenadas para otros tipos de construcción; c) Reglas que obligan a buscar en otros niveles de descripción si los candidatos oracionales no son correctos, es decir buscar en la descripción de otros niveles de texto (imágenes y gráficos); d) Reglas que aportan relevancia a las oraciones en dependencia del lugar donde se encuentren y de la presencia en ellas de vocablos o términos de las listas de prioridad y e) Reglas que permiten desarrollar estrategias de desambiguación léxica.

De esta manera, el modelo de extracción de texto asume dos fases con las siguientes particularidades:

Fase I: Constituye la supresión de oraciones y pasajes textuales de bajo nivel de relevancia en el texto (usar reglas del tipo a).

Fase II: Se proponen oraciones o fragmentos de texto candidatos a ser eliminados (mediante las reglas del tipo b); Si no existen oraciones que cumplan este criterio, buscar otros textos asociados al texto que permitan mejorar la descripción. Reglas de tipo c; Descartar aquellas oraciones que poseen menos nivel de puntuación y faciliten la proyección de aquellas que poseen más alto nivel de relevancia; Aplicar reglas de sintaxis y comunicación para lograr la cohesión del texto y Desambiguar el texto para brindar mayor nivel de claridad en la lectura de la información (regla d).

La aplicación exacta de estos elementos en el sistema de desambiguación y extracción quedan de la siguiente forma:

*Nivel 1:* Es utilizado y visualizado como un nivel de segmentación de texto donde, al igual que en varios estudios como los de Endres-Niggemeyer (2005), se delimita el texto en 5 apartados con los contenidos que poseen mayor nivel de preponderancia. En ocasiones estos apartados no están delimitados, lo que obliga a que el agente de lectura reconozca los elementos por los que tradicionalmente se marcan estos segmentos del texto y confronte la base de sinónimos donde aparecen sinónimos de estos apartados. Esto permite asignar la estructura IOMRC.

*Nivel léxico:* utiliza normas basadas en unidades léxicas que facilitan la asignación de relevancias a determinados segmentos textuales y/u oraciones. En este nivel también se aplican reglas basadas en unidades léxicas, que asignan los signos de puntuación a las oraciones.

*Nivel discursivo y sintáctico-comunicativo:* consistente en la aplicación de reglas (PuertoRules) donde, dependiendo de cada segmento textual, se aplicará un determinado elemento de cada relación discursiva. Estas reglas determinan qué elementos discursivos se mantienen o se eliminan, dependiendo de las marcas de cohesión que tenga el texto en función a la relevancia previamente otorgada por los expertos a ciertos grupos textuales y oracionales y almacenadas en la base de datos.

La segunda generación de reglas PuertoRules trabajan en consonancia con los modelos de cohesión de Mann y Thompson (1990). En el modelo de D' Cunha (2006) se tiene en cuenta la longitud de las oraciones deseadas por el usuario; sin embargo, no existe constatación de que ése sea un criterio fundamental para determinar la calidad de un resumen. En el modelo

extractivo que aquí se describe los autores han creído pertinente desarrollar una fase de desambiguación (Nivel 4) a nivel oracional, utilizando mejoras en las oraciones resultantes de los niveles anteriores.

Fase final: constituida por la unión de las oraciones de todos los apartados anteriores. La parte superior (Anexo 1) muestra un documento original almacenado en el repositorio del sistema. En la parte inferior podemos encontrar el resumen generado automáticamente a partir de dicho original.

## Evaluación del sistema

Existen multitud de trabajos centrados en analizar diversos aspectos relacionados con la evaluación de la calidad de los sistemas de recuperación de información. Pero dado que no existe un único método que sea capaz de evaluar este tipo de sistemas de manera integral, los autores se han decidido por realizar tres evaluaciones diferentes del mismo, con la idea de analizar la mayor cantidad posible de vertientes de PuertoTex. Por un lado se ha analizado la usabilidad del sistema por medio de un estudio heurístico. Luego se ha analizado la capacidad del sistema para recuperar información determinada. Para finalizar se analiza la calidad del producto resultante - el resumen automático -, empleando Rouge y un sistema de evaluación por jueces.

La selección de la muestra ha sido de forma aleatoria, escogiéndose 1 200 artículos en inglés y otros tantos en español de entre los almacenados en el sistema. Para las diferentes evaluaciones se han empleado a 94 usuarios en total. Dependiendo de cada fase, estos usuarios eran expertos en un área concreta: documentalistas, ingenieros, biólogos etc.

## Evaluación de la usabilidad

Para esta primera fase se diseñó un experimento donde expertos y usuarios finales intervienen en la evaluación del mismo. En esta fase se han empleado a 59 personas: 5 expertos en Ciencia de la Información, 42 expertos en Medio Ambiente y 12 usuarios de especialidades relacionadas con el tema de la herramienta. En esta investigación se ha determinado emplear un estudio de usabilidad denominado usabilidad de "inspección". Se trata de una forma de evaluación que

tiene nexos directos con la evaluación heurística, actividad que consiste en el concurso de un grupo de evaluadores a los que se encarga la valoración de la interfaz del sistema mediante principios heurísticos de usabilidad. Cada evaluador realiza la valoración de forma individual, como si fuese el usuario del sistema. La revisión se realiza de manera individual y asumiendo el papel de usuario.

En este análisis se han tomado como referencia algunas de las obras más destacadas sobre el diseño centrado en el usuario y, en particular, los principios heurísticos de Nielsen (1994), Machón (2002) y Nielsen (2002a, 2002b). Teniendo en cuenta el tipo de sistema que estamos evaluando, se ha creado una plantilla que recoge indicadores de análisis relativos a varios aspectos. Los parámetros que se evalúan tienen como centro la heurística propuesta por Nielsen (1994), centrada en 10 aspectos claves relacionados con la usabilidad: la navegación, la funcionalidad, el control por parte del usuario, el uso de la lengua, la ayuda en línea y guía del usuario, la información proporcionada por el sistema, la accesibilidad, la coherencia, la prevención y corrección de errores, y la claridad arquitectónica y visual del sistema.

Para lograr una uniformidad en la evaluación del sistema se desarrolló un cuestionario donde se evalúan las variables que se utilizan en la investigación heurística. En el referido cuestionario se les proporcionó a los evaluadores una serie de preguntas con tres opciones de respuesta: "no, nunca", "a veces" y "sí, siempre", a las que se le dio el valor, respectivamente, de 0, 1 y 2, siendo 2 el valor más alto. Con este proceder las respuestas de los evaluadores obtienen cualidades numéricas que se utilizan para valorar los elementos del sistema.

En la Tabla 1 se muestran los resultados de la aplicación del test heurístico de uso del usuario. La navegación y la claridad arquitectónica son los elementos de menor nivel, posiblemente porque los usuarios no están adaptados a trabajar con sistemas que utilizan ontologías como herramientas de consulta y, por lo tanto, no logran identificar a qué lugares del sistema dirigir su búsqueda. El sistema es funcional, pues un 84% de los encuestados identifica la presencia en él de diversas informaciones relativas a idioma, lenguas, esquema de usuarios, algoritmos de trabajo etc.

Por las respuestas obtenidas, se observa que PuertoTex logra que sus usuarios se sientan seguros y con control sobre las acciones, aunque hay que señalar a

**Tabla 1.** Usabilidad del sistema.

Variables de uso	Sí	A veces	No
Navegación	62,0	18,0	10,0
Funcionalidad	81,4	10,6	18,0
Control del usuario	89,0	11,0	0,0
Lengua y contenido	100,0	0,0	0,0
Ayuda en línea	100,0	0,0	0,0
Información del sistema	100,0	0,0	0,0
Accesibilidad	85,0	15,0	0,0
Coherencia	90,0	10,0	0,0
Prevención errores	60,0	20,0	20,0
Claridad arquitectónica	65,5	25,0	10,8

un grupo de usuarios que no poseen un entrenamiento con estos sistemas y que, en algunas ocasiones, dicen sentirse espiados por él anotando observaciones del tipo "este sistema piensa por mí".

### Evaluación de la recuperación de información

Para llevar a cabo esta evaluación se proporcionó un test a los usuarios y los expertos a partir de un conjunto de tareas específicas con PuertoTex. Se preparó una serie de 8 tareas y se realizó la prueba con 10 usuarios con conocimientos en ingeniería y 5 expertos, de manera individual, durante 40 minutos, puntuando cada una de las tareas hasta un máximo de cien. La media de cada uno de los resultados fue: 1) Búsqueda de un término en dos idiomas: 99,55; 2) Búsqueda de información no lingüística: 96,85; 3) Búsqueda de lema: 100; 4) Claridad de presentación de la información (idiomas): 100; 5) Búsqueda de un término: 98; 6) *Feedback*: 95; 7) Claridad de resultados: 86 y 8) Búsqueda de Resúmenes: 100.

La mayoría de los parámetros que se analizan arrojan resultados bastante satisfactorios. La búsqueda de un término en dos idiomas arrojó un 99,5% de efectividad. La recuperación de información no lingüística un 96,5%. La recuperación y búsqueda de información lematizada fue efectiva en un 100%, así como la claridad de la Representación de la Información. La búsqueda de un término es efectiva en un 100%, al igual que la búsqueda de resúmenes.

### Evaluación de los resúmenes automáticos

A pesar de contar con evaluaciones muy favorables en los apartados relativos a la usabilidad del sistema y a



su capacidad para recuperar información, los autores son conscientes de que el punto fuerte de PuertoTex es el de ser un sistema capaz de generar resúmenes automáticos. Por eso es importante evaluar este aspecto de manera independiente.

El primer análisis se realizó por medio de Rouge, herramienta para la evaluación de la sumarización de documentos (Lin; Hovy, 2003). Se ha convertido en una de las pautas esenciales para determinar la calidad de los resúmenes realizados mediante método automáticos. Rouge posee varias métricas para realizar la valoración de los resúmenes; esto facilita la selección de registros adecuados para su adaptación a cada proyecto. En este caso se ha utilizado *Rouge N (N-Gram Co-Occurrence Statistic)*. Esta media estudia la cantidad de n-gramas que ocurren en un resumen candidato y en un resumen de referencia. La ecuación declara que *Rouge N* es el cociente de la suma de las referencias del resumen base entre las referencias de un resumen obtenido mediante medidas automáticas. En ella  $Count_{match}(gram_n)$  es el número máximo de referencias recibidas por un resumen candidato y un resumen de referencia. Los n-gramas son medidas resultantes de n, que significa la cantidad de n-gramas que ocurren en ambos textos.

$$ROUGE-N = \frac{\sum_{S \in \{\text{Reference Summaries}\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{\text{Reference Summaries}\}} \sum_{gram_n \in S} Count(gram_n)}$$

Cuando se trabaja con Rouge es vital poseer resúmenes que sirvan de patrón o referencia para ser comparados con los resúmenes base o candidatos. En este experimento las comparaciones se realizan atendiendo a los siguientes tipos de resumen:

- Resúmenes por extracción generados por el sistema (Puertot\_1a y 1b).
- Resúmenes por abstracción generados por el sistema (Puertot\_2a y 2b).
- Resúmenes por extracción realizados por 10 especialistas de la Comunidad Ciencias Biológicas.
- Resúmenes por abstracción realizados por 10 especialistas de la Comunidad Ciencias Biológicas.

- Resúmenes realizados mediante el sistema *Microsoft Word*.

- Resúmenes Baseline desarrollados por expertos en el tema.

La mayoría de los autores prefieren evaluar con Rouge utilizando como materia de análisis solo los resúmenes de los autores. La experiencia internacional en la evaluación con esta métrica aclara que no es fiable evaluar un resumen automático usando un solo tipo de resumen obtenido bajo un solo modelo, por tanto en esta evaluación se introducen otros resúmenes para ser comparados (Tabla 2).

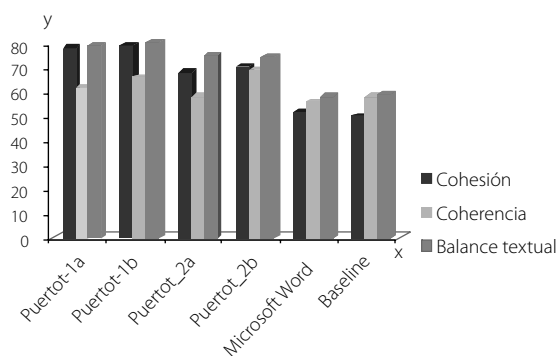
Es evidente que todos los resúmenes hechos mediante PuertoTex son muy similares al corpus baseline; sin embargo, éstos no se asemejan al corpus construido por medio de otros sistemas (para este experimento se optó por usar la opción de resumen automático del procesador de textos *Word*).

Las medidas que se obtienen con Rouge solo son proporciones estadísticas, y es por ello que autores como Pinto (2001) y Hernández (2007b), enfatizan en que los resúmenes no se deben evaluar de acuerdo a métricas de similitud establecidas sobre patrones físicos, en este caso

**Tabla 2.** Resultado de aplicar Rouge N sobre un conjunto de textos generados por PuertoTex.

Rouge-N	Tex 1	Tex 2	Tex 3	Tex 4	Tex 5
Puertot_1a	0.6891	0.6734	0.6791	0.6732	0.6791
Puertot_1b	0.6781	0.6234	0.6451	0.6831	0.6321
Puertot_2a	0.6893	0.6129	0.6234	0.6612	0.6523
Puertot_2b	0.6321	0.6321	0.6247	0.6231	0.6643
Baseline	0.1345	0.1245	0.1891	0.1789	0.6745
Word	0.4567	0.4321	0.4831	0.4890	0.6267
Rouge-N	Tex 6	Tex 7	Tex 8	Tex 9	Tex 10
Puertot_1a	0.6589	0.6128	0.6238	0.6345	0.6538
Puertot_1b	0.6635	0.6238	0.6363	0.6136	0.6156
Puertot_2a	0.6123	0.6341	0.6721	0.6721	0.6825
Puertot_2b	0.6621	0.6458	0.6821	0.6239	0.6523
Baseline	0.1980	0.1213	0.1890	0.1992	0.1389
Word	0.4467	0.4125	0.4712	0.4529	0.4578

Nota: Puertot\_1a y Puertot\_1b son resúmenes hechos por extracción, mientras que Puertot\_2a y Puertot\_2b son resúmenes hechos por abstracción. Para obtener datos más representativos se emplean otros resúmenes realizados sobre el mismo corpus documental. Dichos resúmenes han sido creados automáticamente por el procesador de textos (Word) y un corpus de contraste (baseline) constituido por 10 resúmenes realizados por expertos en la materia.



**Figura 2.** Evaluación de la cohesión y la coherencia textual.

Fuente: Elaboración propia.

Nota: En el eje X se encuentran los diferentes tipos de resúmenes evaluados. En el eje Y la puntuación otorgada por los observadores en base a los criterios de cohesión, coherencia y balance textual.

los bigramas. Por ello, en esta investigación se insiste en la evaluación de la cohesión y la coherencia del texto como única vía para saber si los textos resumidos cumplen con su objetivo pragmático (Pinto, 2001). Es decir, que sean resúmenes legibles y entendibles por los humanos. Para realizar estos análisis sometimos a todos los textos que se analizaron en los experimentos a la valoración de 2 jueces, que determinaron, puntuando de 0 a 100, el nivel de cohesión, coherencia y balance textual de los textos resultantes de la aplicación de los diferentes modelos de resumen que se utilizan en la evaluación (Figura 2).

## Conclusiones

Si bien es cierto que existen otros métodos de desambiguación de probada calidad (entre ellos los que trabajan con resolución de anáforas), tenemos claro que la mayoría de estos métodos no contaban con la posibilidad de emplear una ontología. En el caso de PuertoTex se ha optado por crear y emplear una ontología propia, ya que defendemos la teoría de que los resultados obtenidos con un sistema supervisado nunca serán peores que los de un sistema no supervisado. Aquí se ha partido de la idea de que el texto, como unidad íntegra, debe estar desambiguado. Por eso un agente debe ejecutar la desambiguación empleando para ello la ontología, asignar reglas de puntuación para facilitar la lectura y

ayudar en el etiquetado de las palabras en el texto. La desambiguación culmina cuando el agente desarrolla el proceso de resolución de anáfora pronominal para dar cohesión al texto.

Existen muchos sistemas de resumen desarrollados desde la óptica de la Ciencia de la Computación, sin embargo en PuertoTex se mezclan criterios lingüísticos, comunicacionales, sociocognitivos y matemáticos.

Las prestaciones de PuertoTex son válidas para sus usuarios debido a que en su concepción, además de utilizar los procesos de ingeniería de *software*, se utilizaron técnicas de estudio de usuarios que facilitan que un investigador de la Rama de Ingeniería de Puertos y Costas no solo resuma un documento, sino que también obtenga información con grupos de personas que traban el mismo tema, o localice información en *Internet*, constituyendo una red social que realmente representa información y conocimiento.

Los resultados de la explotación del *software* son halagüeños si se tiene en cuenta que los resultados de los test de usabilidad son de elevada calidad y se corresponden con las prestaciones que ofrece el sistema.

La herramienta facilita la construcción de resúmenes tanto en español como en inglés, siempre y cuando sean textos que pertenezcan al dominio de la Ingeniería de Puertos y Costas. Los textos seleccionados para ambos idiomas facilitan la construcción de resúmenes indicativos e informativos.

La principal limitación de este estudio se centra en las propias estrategias cognitivas y reglas discursivas desarrolladas e implementadas a través de la herramienta, ya que solo operan en el dominio en cuestión y con artículos científicos. Otros textos, como, la poesía o la prosa, serían muy complejos de tratar mediante estas técnicas debido a que la retórica y el estilo del texto suelen variar mucho y no pueden determinarse regularidades para construir reglas de discurso.

## Agradecimientos

Esta investigación se ha realizado dentro del proyecto RECORD: Representación del Conocimiento en Redes Dinámicas FFI2011-22397, financiado por el Ministerio de Ciencia e Innovación de España.

## Referencias

- AGUIRRE, E. *Formalization of concept-relatedness using ontologies: applications in the construction of lexical knowledge bases, word sense disambiguation and automatic spelling correction*. 1998. Tesis (Doctoral) - Universidad de País Vasco, San Sebastián, 1998.
- D'CUNHA, I. *Hacia un modelo lingüístico de resumen automático de artículos médicos en español*. 2006. Tesis (Doctoral) - Universidad Pompeu Fabra, Barcelona, 2006.
- DIJK, T.; KINTSCH, W. *Strategies of discourse comprehension*. Orlando: Academic, 1983.
- DOMÍNGUEZ, S. *Calculus copora*. Santa Clara: Universidad Central Marta Abreu de las Villas, 2011.
- ENDRES-NIGGEMEYER, B. Simsum: an empirically founded simulation of summarizing. *Information Processing and Management*, v.36, n.4, p.659-682, 2005.
- ENDRES-NIGGEMEYER, B.; MAIRE, E.; SIGEL, A. How to implement a naturalistic model of abstracting: four core working steps of an expert abstractor. *Information Processing & Management*, v.31, n.5, p.631-674, 1995.
- FABER, P.; MÁRQUEZ, C.; VEGA, M. Framing terminology: a process-oriented approach. *Meta*, v.50, n.4, p.189-213, 2005.
- HERNÁNDEZ, A. *Indización y resumen*. La Habana: Universidad de la Habana, 2006.
- HERNÁNDEZ, A. *Organización y representación del conocimiento: paradigmas, hipertextos y fundamentación metamodélica*. 2007. Tesis (Doctoral Bibliotecología y Ciencia de la Información) - Universidad de la Habana, La Habana, 2007a.
- HERNÁNDEZ, A. *Paradigmas dominantes de la representación de la información y el conocimiento*. La Habana: Universidad de la Habana, 2007b.
- HERRERA, R. *Formatos de comunicación*. 2007. Disponible en: <<http://fcom.uh.edu.cu>>. Acceso en: 16 jul. 2011.
- LEIVA, A. *et al.* An automat for the semantic processing of structured information. In: INTERNATIONAL CONFERENCE OF DESIGN OF SOFTWARE AND APLICACIÓN, 9., 2009, Pisa, Italy. *Proceedings...* Pisa: IEEE, 2009. p.85-89.
- LIN, C.; HOVY, E. Automatic evaluation of summaries using n-gram co-occurrence statistic. In: HUMAN TECHNOLOGY-CONFERENCE, 2003, Edmonton, Canada. *Proceedings...* Edmonton, Canada: HLT-NAACL, 2003. p.71-78.
- LUNH, H. The automatic creation of literature abstracts. *Journal of Research of Development*, v.2, n.2, p.59-165, 1958.
- MANCHÓN, E. *Evaluación por criterios o heurística*. 2002. Disponible en: <[http://www.ainda.info/evaluacion\\_heuristica.html](http://www.ainda.info/evaluacion_heuristica.html)>. Acceso en: 1 enero 2011.
- MANN, W.; THOMPSON, S. *Rhetorical structure theory: a theory of text organization*. Los Angeles: Information Sciences Institute, 1990.
- MATHIS, B.; RUSH, J.; YOUNG, C. Improvement of automatic abstracts by the use of structural analysis. *Journal of the American Society for Information Science*, v.24, n.2, p.101-109, 1973.
- METZELTIN, M. De la retórica al análisis del discurso. *Revista Electrónica de Estudios Filológicos*, n.6, 2003. Disponible en: <<http://www.tonosdigital.es/ojs/index.php/tonos/article/view/506>>. Acceso en: 1 marzo 2012.
- NIELSEN, J. Heuristic evaluation. In: NIELSEN, J.; MACK, R. (Ed.). *Usability inspection methods*. New York: John Wiley 1994. p.25-52.
- NIELSEN, J. *How to conduct a heuristic evaluation*. 2002a. Available from: <[http://www.useit.com/papers/heuristic/heuristic\\_evaluation.html](http://www.useit.com/papers/heuristic/heuristic_evaluation.html)>. Cited: 26 Jan. 2011.
- NIELSEN, J. *Ten usability heuristics*. 2002b. Available from: <[http://www.useit.com/papers/heuristic/heuristic\\_list.html](http://www.useit.com/papers/heuristic/heuristic_list.html)>. Cited: 21 Jan. 2011.
- ONO, K.; SUMITA, K.; MIKE, S. Abstract generation based on rhetorical structure extraction. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 15., 1994, Kyoto, Japan. *Proceedings...* Kyoto, Japan: DBLP, 1994.
- PANECA, F. *La prensa remediada del siglo XIX: reflejo de la cultura popular tradicional de la localidad*. 2009. Tesis (Doctoral) - Facultad de Humanidades, Universidad Central Marta Abreu de las Villas, Santa Clara, 2009.
- PILLEUX, M. Competencia comunicativa y análisis del discurso. *Estudios Filológicos*, n.36, p.143-152, 2001. Disponible en: <<http://www.scielo.cl>>. Acceso en: 1 marzo 2012.
- PINTO, M. *El resumen documental: principios y métodos*. Madrid: Fundación Germán Sánchez Ruipérez, 2001.
- PINTO, M. Interdisciplinary approaches to the concept and practice of written text documentary content analysis. *Journal of Documentation*, v.50, n.2, p.405-418, 2004.
- RIGAU, G. *Resolución automática de la ambigüedad semántica de palabras*. Soria, España: Fundación Duques de Soria, 2002.
- SALTON, G. On the application of syntactic methodologies in automatic text analysis. *Information Processing and Management*, v.26, p.73-92, 1996.
- SÁNCHEZ-CUADRADO, S. *et al.* Definición de una metodología para la construcción de sistemas de organización del conocimiento a partir de un corpus documental en lenguaje natural. *Procesamiento del Lenguaje Natural*, n.39, p.213-220, 2007.
- SENSO, J. *Representación del conocimiento en la ingeniería de puertos y costas: proyecto investigador*. Granada, España: Universidad de Granada, 2009.
- SIDOROV, G.; ZAZUETA, O.O. Resolución de anáfora pronominal para el español usando el método de conocimiento limitado. In: ENCUENTRO NACIONAL DE CIENCIAS DE LA COMPUTACIÓN, 7., 2006, San Luís Potosi, Mexico. *Anales...* San Luís Potosi, Mexico: SMCC, 2006. p.276-281.

Documento original (parte superior) y resumen automático obtenido a partir de los datos procesados del documento original (parte inferior).

---

Documento original

---

Estudio de corrientes y dispersión en el puerto de Barcelona

M. Díez<sup>1</sup>, J.M. Redondo<sup>1</sup>, J. Vila<sup>2</sup>

1 Universidad Politécnica de Cataluña. Departamento de Física Aplicada, Jordi Girona 1-3 Barcelona 08034

2 Autoridad Portuaria de Barcelona, APB Medio Ambiente C/Circunvalación tramo VI Barcelona 08039

### 1. Introducción

Uno de los problemas medioambientales de los puertos es el vertido de contaminantes, en especial de hidrocarburos. Dado que la dispersión es un mecanismo que hace crecer exponencialmente el tamaño de las manchas y que el coste de su limpieza y daños también es función del tamaño es imprescindible un buen conocimiento del fenómeno que junto con un buen sistema de alerta y acción permita una rápida intervención.

### 2. Experimentos

En el Puerto de Barcelona se están llevando a cabo una serie de experimentos con el fin de estudiar el campo de corrientes y la capacidad dispersiva del medio. Están previstas campañas para cada estación climática, así como campañas especiales para analizar fenómenos hidrodinámicos como la difracción o estratificación. La duración de cada campaña es de dos semanas debido al gran dominio de trabajo.

Los experimentos consisten en el vertido de trazadores lagrangianos y la filmación de su comportamiento. Asimismo con un GPS se registran los puntos de vertido y recogida tomándose como puntos fiduciales en la restitución digital de las imágenes. Los trazadores son de dos tipos lagrangianos para medir corrientes, flotadores adecuadamente lastrados para evitar la interferencia del viento, y manchas de leche con fluoresceína por su elevado contraste y persistencia.

### 3. Resultados

A través del análisis digital de imágenes (Rodríguez *et al* 1997) se obtienen las trayectorias, direcciones y velocidades de la corriente y los coeficientes de dispersión. Por otro lado se analizan los agentes impulsores como el viento, marea, oleaje, corrientes etc.

Los primeros resultados muestran que la influencia de la marea o la pluma del Llobregat pueden desviar la dirección de la corriente respecto del viento significativamente, llegando incluso a oponerse a él. A continuación se muestra un ejemplo en el que se detecta un cambio en el régimen de la corriente de entrante a saliente (figura 2) que no responde estrictamente a una variación del viento. Si se observa la evolución de la marea (figura 3) parece coincidir con un cambio de vaciante a llenante.

Sin embargo, la escasa carrera de marea no hacía esperar un cambio de tendencia.

### 6. Referencias

Rodríguez A., Bahía E., Díez M., Sanchez-Arcilla A., Redondo J.M., Mestres M. Estudio

Experimental de Procesos de Mezcla en Aguas Costeras (Mayo 1997). Proc. IV Jornadas españolas de Ingeniería de costas y Puertos, Ed. J.C. Serra, pp. 643-656. ISBN 84-7721-573-1

### Resumen

Uno de los problemas medioambientales de los puertos es el vertido de contaminantes, en especial de hidrocarburos. Dado que la dispersión es un mecanismo que hace crecer exponencialmente el tamaño de las manchas y que el coste de su limpieza y daños también es función del tamaño es imprescindible un buen conocimiento del fenómeno que junto con un buen sistema de alerta y acción permita una rápida intervención. En el Puerto de Barcelona se están llevando a cabo una serie de experimentos con el fin de estudiar el campo de corrientes y la capacidad dispersiva del medio. Están previstas campañas para cada estación climática, así como campañas especiales para analizar fenómenos hidrodinámicos como la difracción o estratificación. Los primeros resultados muestran que la influencia de la marea o la pluma del Llobregat pueden desviar la dirección de la corriente respecto del viento significativamente, llegando incluso a oponerse a él.

---



# Mudança de paradigma e sua ruptura: um estudo de caso na Museologia e a pluralidade paradigmática da Ciência da Informação

## *Paradigm shift and its rupture: a case study and paradigmatic plurality of Museology and Information Sciences*

Mário GOUVEIA JÚNIOR<sup>1</sup>

Raimundo Nonato Macedo dos SANTOS<sup>2</sup>

### Resumo

Este artigo traça um breve histórico da Ciência da Informação, anotando alguns conceitos ainda em discussão em um campo de estudos dito incipiente em virtude da ausência de paradigmas capazes de atender às inquietações de seus pesquisadores de modo satisfatório. Nesse contexto, busca-se abordar em que medida a configuração de uma crise pode comprometer um modelo de pensamento ao ponto de suscitar um novo paradigma. Analisa-se o conceito de mudança de paradigma em relação à Museologia e à Ciência da Informação, a partir de um breve estudo de caso, por meio do qual se obterá um maior entendimento acerca das negociações que se evidenciam quando se trata de questões inerentes à pesquisa científica e à comunidade acadêmica.

**Palavras-chave:** Informação. Mudança de paradigma. Museologia.

### Abstract

*In this article, brief history of Information Science was traced from a literature review, noting some concepts still under discussion in a field of studies that is allegedly a beginner, due to the absence of paradigms able to meet the concerns of researchers in a satisfactory manner. In this context, an endeavor will be made to address the question of how the emergence of a crisis can compromise a model of thought to the point of raising a new paradigm. The concept of paradigm shift in relation to Museology and Information Science will be analyzed from a brief case study, from which a clearer understanding will be found about the negotiations that are evident when it comes to issues related to scientific research and academic community.*

**Keywords:** Information. Paradigm shift. Museology.

### Introdução

Amparada que está nas novas tecnologias de disseminação de informação e nos fluxos de comunicação

veiculados em tempo real, a Ciência da Informação e toda a literatura por ela produzida encontram-se imersas em constante evolução e resignificação. A dinâmica catalisada pelas mídias digitais não está, todavia, restrita

<sup>1</sup> Mestrando, Universidade Federal de Pernambuco, Programa de Pós-Graduação em Ciência da Informação. Recife, PE, Brasil.

<sup>2</sup> Professor Doutor, Universidade Federal de Pernambuco, Departamento de Ciência da Informação. Av. da Arquitetura, s/n., Campus Universitário, 50740-550, Recife, PE, Brasil. Correspondência para/Correspondence to: R.N.M. SANTOS. E-mail: <rnmacedo@uol.com.br>.

Recebido em 15/2/2012, rerepresentado em 9/4/2012 e aceito para publicação em 25/4/2012.

à academia, uma vez que a sociedade, como um todo, tem recebido o impacto de sua frequência ao mesmo tempo em que ela mesma, individual ou coletivamente, é responsável por sua construção, distribuição e acesso, o que caracteriza a era do acesso ao conteúdo (Capurro; Hjørland, 2007).

Vive-se num tempo de afetividade informacional, em que tanto se pode se corresponder com o vizinho de sala quanto com o colega do outro lado do mundo com a mesma velocidade, o que faz do tempo real, no ciberespaço, o aqui e o agora (Barreto, 2002). De modo análogo, Ortega y Gasset (2006, p.72) considerava, já no começo do século XX, que a vida se mundializara e que cada indivíduo vivia muitas relações de ubiquidade, denotando uma proximidade do distante, uma presença do ausente, ao ponto de sermos capazes de “[...] estar em mais lugares que antes, desfrutar mais idas e vindas, consumir em menos tempo vital mais tempo cósmico”.

Assim, pode-se perceber que a necessidade de informação de que trata Le Coadic (2004) não se restringe à demanda, mas também à sua oferta. Ademais, a transmissão de habilidades práticas, sociais ou intelectuais acaba por se tornar uma espécie de garantia de sobrevivência tanto de seus conteúdos quanto da própria comunidade que os produz (McGarry, 1999). Em uma palavra, “o acesso à informação deixa de ser retórico para tornar-se imprescindível, imperativo, vital, mesmo que de forma fragmentária, como ocorre em períodos de exceção” (Castro, 2007, p.73).

Nessa perspectiva, há a grande quantidade de *blogs* e páginas de provedores da *Internet* que contêm postagens com dicas ou respostas àqueles que têm dúvidas sobre termos e definições ou simplesmente almejam passar de fase em algum jogo de videogame ou de computador. Os autores das postagens dá as respostas ou dá as dicas e truques - uma população indefinida de *internautas* que pode ultrapassar os milhões - na maioria das vezes não se identificam nem parecem atuar ali esperando qualquer reconhecimento que não o da satisfação das necessidades de quem interessar possa.

Tal facilidade de acesso a conteúdos cada vez mais extensos em um espaço de tempo cada vez mais diluído traz a ideia de que se vive numa sociedade cada vez mais tributária da velocidade, que, no entendimento de Lévy (1999b), seria um primeiro grau de virtualização. Em meio

a tanta informação colocada ao dispor do usuário, que parece se encontrar mergulhado num “dilúvio informacional” (Lévy, 1999a), o que requerer do usuário a triagem daquilo que lhe serve em um universo de possibilidades de acordo com suas necessidades.

Da mesma forma se procede no campo científico: na construção de uma ciência, há que se evitar o caos conceitual e a pluralidade de alternativas metodológicas pautadas em definições vagas e circulares, ancoradas em citações acríicas de autoridades (Capurro; Hjørland, 2007), que, por vezes, antagonizam-se e evitam quaisquer meios de se chegar a um consenso mínimo.

Trata-se aqui do que Kuhn (1979) chamou de paradigma. Esse pesquisador, por sinal, trouxe uma noção inovadora acerca do progresso científico. Negando uma evolução linear e contínua da ciência, Kuhn defendeu as rupturas como responsáveis pelas conquistas e descobertas revolucionárias, ocorridas entre determinados espaços de tempo consideráveis. Popper (2007), por seu turno, enxerga que tais transformações se dariam de modo mais constante, a todo momento, ou, ao menos, em intervalos menores.

Vale registrar que “o progresso das certezas científicas, entretanto, não caminha na direção de uma grande certeza” (Morin, 2002, p.23). Não é Buckland (1991) quem diz que a informação-como-conhecimento pode agir tanto no sentido de diminuir quanto de aumentar a incerteza? Aliás, a própria incerteza, que caminha ao lado da dúvida, parece ser a propulsora das atividades científicas e sua busca pelo conhecimento dito verdadeiro, tornado possível, de acordo com Bachelard (1996), quando se toleram as incertezas e se priorizam as perguntas em detrimento das respostas.

Ao longo deste trabalho, serão anotados determinados conceitos ainda em discussão em um campo de estudos dito incipiente em virtude da ausência de paradigmas que sejam capazes de atender às inquietações de seus pesquisadores de modo satisfatório. Nesse contexto, será abordado em que medida a configuração de uma crise pode comprometer um modelo de pensamento ao ponto de suscitar um novo paradigma. Acredita-se que os debates iniciados nos anos de 1970 auxiliam a ilustrar e a entender as negociações que se evidenciam quando se trata de questões inerentes à pesquisa científica e à comunidade acadêmica.

Na primeira parte deste trabalho, será contextualizado o nascimento da Ciência da Informação, na Era da Sociedade Informacional, através de um breve histórico.

O segundo momento traz os primeiros debates, bem como algumas definições e críticas acerca da Ciência da Informação, seu campo de estudos e suas ferramentas. Do mesmo modo, serão evocados alguns conceitos sobre informação, interpretação e aprendizagem.

A terceira e última parte ilustra tanto o advento de um paradigma numa Ciência Social como é a Museologia - com as conferências que se realizaram na América Latina, a partir da década de 1970, com o intuito de traçar novos rumos para aquele campo do conhecimento - quanto a existência de alguns paradigmas na Ciência da Informação, ainda que ela própria não tenha um paradigma definido e amplamente aceito.

### **A Ciência da Informação: um breve histórico**

Apesar de a pólvora, a imprensa e a Reforma serem os três principais agentes causais das transformações tecnológicas, políticas e econômicas que moldaram o mundo atual, Castells (1999) e McGarry (1999) consideram que as principais descobertas tecnológicas em comunicação e eletrônica - microeletrônica, computadores e telecomunicações - do século XX se deram durante e imediatamente após a Segunda Guerra Mundial. É assim que "a guerra, com suas demandas de alta tecnologia, preparou vários processos revolucionários para posterior uso civil" (Hobsbawm, 1995, p.260).

Nessa perspectiva, terminada a referida contenda global:

[...] o incremento da produção documental foi enorme, fruto essencialmente do desenvolvimento tecnológico, científico e industrial. Os anos cinquenta do século XX ficaram mesmo conhecidos como a época da 'explosão da informação', fenômeno que desencadeou, naturalmente, mudanças profundas nos serviços, nos meios de difusão e até mesmo nas técnicas de tratamento da informação (Silva; Ribeiro, 2011, p.108).

Desse modo, a informação até então mantida em sigilo de Estado seria posta à disposição das sociedades em escala global (Barreto, 2002). Todavia essas novas tecnologias da informação só começariam a se difundir

efetivamente a partir de década de 1970. Foi nesse momento em que se conhece o início de uma verdadeira revolução social, científica e técnica, denotada, no Ocidente, pelo uso da informação e do conhecimento nos processos sociais e produtivos (Saracevic, 1996a).

Tais processos, na Era da Sociedade Informacional, ou da Economia Informacional (Castells, 1999), migraram do fabrico de bens materiais para a produção, armazenamento e disseminação de informação e comunicação. Com isso, verifica-se que a técnica produzida pelas ciências e por seu espírito científico, uma vez popularizada, contribuiu significativamente para a transformação da sociedade - apesar de não determiná-la. Esta última, em contrapartida, também modifica e alimenta a própria ciência em suas investigações e descobertas (Morin, 2002). Nesse caso, é forçoso considerar também a ideia de Nagel (1975) de que, por vezes, a valorização em excesso do caráter prático dos avanços científicos - entenda-se a conquista de bens e vantagens - acaba por distorcer os objetivos primeiros da ciência.

Capurro (2003) mostra, na Ciência da Informação, a presença de duas raízes: a Biblioteconomia clássica e a Computação Digital. A primeira dialoga com as origens da sociedade humana, sustentada que esteve e está nas redes de relações mantidas através da linguagem. A segunda, de natureza tecnológica, está ligada ao impacto da computação nos processos de produção, coleta, organização, interpretação, armazenagem, recuperação, disseminação, transformação e uso da informação, sobretudo a científica registrada em documentos - até então apenas - impressos.

De acordo com Saracevic (1996a), ao longo das décadas de 1960 e 1970, a Ciência da Informação se dedicou ao estudo das propriedades dos processos de comunicação e ao comportamento da informação relativos à sua gênese, coleta, organização, estocagem, recuperação, interpretação, transmissão, transformação e uso. Nos anos de 1990, as atenções se voltaram para as questões científicas da área e para a prática profissional direcionadas ao contexto social e suas necessidades de informação.

É exatamente nessa época que Saracevic (1996a) chama a atenção para que estudos sobre usuários e suas demandas e necessidades informacionais estejam acima dos aspectos tecnológicos - funcionalistas -, no *ranking*



de prioridades de estudo entre as ciências e disciplinas que têm na informação o seu objeto. Do mesmo modo, Morin (2002) consideraria de fundamental importância, para toda e qualquer ciência, o ato de se interrogar acerca de suas estruturas ideológicas, contextuais e de seu enraizamento sociocultural.

No momento em que tais preocupações de seleção, armazenamento, recuperação e disseminação da informação se institucionalizavam, a Ciência da Informação dava os seus primeiros passos através de estudos e métodos interdisciplinares, de uma ligação inexorável com a tecnologia da informação e sua participação ativa e deliberada na evolução da sociedade da informação (Saracevic, 1996a).

Essa caminhada começou a ser registrada por inúmeros estudos e debates promovidos em conferências, encontros e seminários; do mesmo modo, muitas revistas e periódicos, nas décadas de 1980 e 1990, foram publicados nos Estados Unidos e na Europa Ocidental. Tais trabalhos e artigos, via de regra compostos por historiadores da informação e bibliotecários, eram direcionados a um público geral, uma vez que se evitava o uso de uma linguagem especializada (Burke, 2007).

Em seus conteúdos estava a exploração de tópicos desde a indexação até o desenvolvimento de bibliotecas e sistemas de informação científica. Do mesmo modo, alguns estudiosos da Ciência da Informação focaram suas atenções nos grandes contextos histórico-explicativos, tais como a modernização do mundo ocidental, a ascensão da chamada sociedade pós-industrial e a contenda entre socialismo e capitalismo (Burke, 2007).

O próprio Burke (2007), entretanto, considera que, apesar de todo o trabalho realizado até agora, não haja um conhecimento suficiente que possa embasar a confecção de uma narrativa histórica abrangente da Ciência da Informação. Alguns pontos essenciais permanecem sem resposta e grande parte da literatura histórica existente aguarda tradução para linguagem e conceitos comuns.

É justamente sobre essa pluralidade de definições e conceitos acerca da informação, do conhecimento, do objeto da Ciência da Informação e da delimitação de alguns paradigmas nesse campo de estudos que os tópicos seguintes tratam.

## Do caos conceitual à organização do conhecimento

De acordo com Capurro e Hjørland (2007), comumente se define a Ciência da Informação como um campo do saber que se ocupa com a gênese, coleta, organização, interpretação, armazenamento, recuperação, disseminação, transformação e uso da informação, aplicando-se, nesse campo, a intervenção das novas tecnologias. Entretanto, ambos ressaltam que nenhuma ciência deveria ser identificada por meio de suas ferramentas - aspectos funcionalistas -, mas pelo seu objeto de estudo e por sua função específica - a informação como fenômeno.

A Sociedade da Informação, brevemente apresentada no tópico anterior, assim como todas as sociedades que passam por transformações históricas e/ou socioeconômicas, demanda as ações de uma ciência no sentido de estudar as propriedades da informação, bem como os processos de sua construção, disseminação e uso, conforme pontua Le Coadic (2004). Esse mesmo autor enxerga na Ciência da Informação - uma ciência social aplicada - a preocupação em esclarecer um problema social evidente: o da informação no que concerne aos aspectos da cognição e da comunicação humanas.

A informação, aliás, desfruta de uma posição privilegiada na atualidade, pois é a partir da informação - como matéria-prima para a tomada de decisões - que a sociedade, em seus vários segmentos, organiza-se, traçando seus planos de ação. Numa palavra, a informação figura como sinônimo de poder - posicionamento estratégico - numa realidade em que seu fluxo é global, contínuo e veloz, e que estratégias de preservação, controle e acesso têm estado nos focos da notícia e na pauta de instituições públicas e privadas, bem como no campo das ciências da tecnologia e da informação.

Le Coadic (2004) define a informação como um conhecimento registrado em um suporte na forma escrita - impresso ou digital -, oral ou audiovisual, compondo um elemento de sentido transmissível a um ser consciente que a interprete através da linguagem.

De modo semelhante, Zins (2007) considera que informação é o significado lógico, porém transitório, de declarações/comunicações emitidas socialmente. Conhecimento, por sua vez, é a capacidade de compreender, explicar e negociar conceitos, ações e intenções. Dito de

outro modo, o conhecimento é a informação que foi apropriada pelo usuário. Quando a informação é adequadamente assimilada para além de sua utilização, mediante análise e reflexão prévia, produz-se conhecimento, que modifica o indivíduo e transforma a sociedade. Apesar de ser uma construção humana, o conhecimento não é transferível, mas, através de informações, pode-se se comunicar sobre determinado assunto.

Capurro e Hjørland (2007) atribuem à informação uma grande variedade de significados e visões teóricas, por vezes, conflitantes. Tais definições dependem, contudo, das funções que lhes são conferidas, visto que a informação ganha significado mediante as interpretações dos seus receptores. Kuhn (1979) considera, nessa perspectiva, que todas as coisas que o homem vê dependem daquilo que sua experiência visual-conceitual prévia o ensinou a ver.

Nesse sentido, negada como matéria e energia por alguns, e entendida como um fenômeno humano que envolve troca de mensagens, a informação não é um elemento observável puro, mas um dado interpretável, uma vez que prescinde do pensamento humano.

Morin (2011) ratifica essa ideia ao anotar que os seres humanos conhecem o mundo graças às mensagens transmitidas, através da percepção, ao cérebro; é lá, na mente, onde se encontra o interior do mundo particular de cada um. De acordo com esse filósofo, a informação existe desde que, entre os homens, haja interação, comunicação e interpretação de signos, cognoscíveis mediante a tradução das realidades do mundo exterior.

De modo semelhante, Silva e Ribeiro (2011) afirmam que o conceito de informação evoca o de comunicação e ambos estão atrelados a um fenômeno humano, e, portanto, social, que dialoga com a capacidade simbólica de cada indivíduo e sua necessidade de interagir com seus pares.

Ainda, segundo Morin (2011), a informação não deve ser considerada como um ingrediente, mas como uma teoria dotada de um conceito indispensável que, apesar de pedir uma problematização, ainda não figura como um conceito elucidativo, apresentando lacunas e incertezas ao mesmo tempo em que se mostra muito mais um ponto de partida do que de chegada.

Zilio e Gonzalez (2010) defendem a ideia de que todos vivem ligados numa teia de informação - não

necessariamente simbólica - em que o conjunto de normas e hábitos partilhados é, possivelmente, o resultado mais contundente desse fato. Não obstante, tal possibilidade seria descartada se o homem fosse incapaz de receber, manipular e difundir informação significativa. O significado, no entanto, ainda de acordo com Zilio e Gonzalez (2010) não está contido diretamente na informação, mas no seu processo de representação, atrelado que está à habilidade de aprendizagem dos indivíduos.

Para Buckland (1991), a informação que, através da percepção, modifica aquilo que se conhece e as opiniões é chamada informação-como-processo. É nesse particular que se pode inferir que por meio de uma nova informação a consciência ou a percepção da realidade são afetadas. Logo, aprendizagem e transformação parecem indissociáveis. O que corrobora também a ideia de que a fonte primária da informação é a observação relevante do fenômeno natural.

McGarry (1999) acrescenta a ideia de que todos possuem um mapa cognitivo que se modifica a cada nova experiência/aprendizagem, para depois se reordenar. Nesse caso, sempre que se interpreta uma nova informação e que ela é acrescentada aos conhecimentos, desorganizam-se os arquivos de nossa mente para depois rearranjá-los mais uma vez. Esta, aliás, parece ser uma atividade dinâmica e inacabada, responsável pela construção do que Barreto (2002) denomina de inteligência - a introdução de um conhecimento assimilado na realidade do receptor, representando um conjunto de atos voluntários através do qual ele reelabora seu mundo e busca modificar seu espaço.

É válido lembrar que este é um acervo subjetivo, ainda que desenvolvido social e culturalmente, e que, no entendimento daquele autor, a informação atua como um instrumento modificador da consciência do homem, trazendo benefícios para o seu desenvolvimento e para o bom funcionamento da sociedade e a harmonização do mundo (Barreto, 2002). Essa concepção harmônica, no entanto, não é convincente devido à inclinação em concordar com a ideia de Buckland (1991) de que a dilatação dos canais de informação tanto pode diminuir quanto aumentar o grau de incerteza. Parece justo salientar ainda que tanto a incerteza como a desordem não necessariamente devem ser afastadas; dúvida e entropia

podem, em muito, contribuir para a pesquisa e a aprendizagem.

E nesse contexto, vale enfatizar que:

A epistemologia não é pontifical nem judiciária; ela é o lugar ao mesmo tempo da incerteza e da dialógica. De fato, todas as incertezas que consideramos relevantes devem ser confrontadas, corrigir umas às outras, entredialogar sem que, no entanto, se imagine possível tapar com esparadrapo ideológico a última brecha (Morin, 2011, p.47).

O próprio Morin (2011) ensina que a capacidade de compreender é inerente à consciência da complexidade.

Tornando a abordar o processo interpretativo, que desarranja para depois rearranjar, precisa-se ainda pontuar que ele advém como uma condição necessária à aquisição de conhecimento, já que interpretar uma informação requer a introdução da perspectiva do receptor, que, por sua vez, confere, ou não, significado àquela mensagem.

Buckland (1991) defende essa natureza subjetiva da informação, ainda nesse contexto, a informação não seria algo idêntico tanto para o emissor quanto para o receptor, mas se apresenta como um elemento a ser constituído, interpretado e partilhado através do processo de comunicação.

Em relação à desordem, é novamente Morin (2011) quem pontua que toda a transformação, assim como a própria vida, está ligada à degradação, à confusão, a qual coopera de certo modo no sentido de organizar o universo. Este, afinal, não teria começado com uma desintegração, e só depois teria se organizado?

Nesse particular, a própria Ciência da Informação já foi - e, por vezes, ainda é - acusada de ser um campo do conhecimento que se encontra num estado de desagregação tamanha e cuja literatura científica é caracterizada pelo caos conceitual, por analogias inadequadas, definições circulares, vagas e contraditórias (Hjorland, 2000).

Todavia, esse mesmo "caos" pode, em sua origem, representar a própria sistematização da Ciência da Informação. De modo análogo, a memória também não existe sem o esquecimento, sendo essa analogia, como aquela, uma relação dialógica na qual forças antagônicas

se complementam através de um equilíbrio dinâmico e mutante. Assim acontece desde a Antiguidade, seja no Oriente - com a filosofia chinesa do *yin-yang* -, seja no Ocidente - com a associação dos contrários de Heráclito (Morin, 2003). No passado, como hoje:

Num universo de pura ordem, não haveria inovação, criação, evolução. Não haveria existência viva nem humana. Do mesmo modo nenhuma existência seria possível na pura desordem, porque não haveria nenhum elemento de estabilidade para se instituir uma organização (Morin, 2011, p.89).

É assim que a chamada complexidade científica parece caminhar entre o consenso e o conflito; entre a racionalidade e o empirismo, a imaginação e a verificação. O não científico está no científico e vice-versa, sem que um anule o outro ou cerceie suas expressões (Morin, 2011).

## Mudança de paradigma e sua ruptura

Apesar de não se refutar de todo os postulados de Popper, há a inclinação em concordar, ao menos em parte, com os pareceres de Thomas Kuhn acerca da relação estreita entre crise e mudança de paradigma.

Um exemplo da aplicação dessa teoria, no campo das Ciências Sociais, e, especificamente, dentro da Museologia, pode ser ilustrado nos novos direcionamentos - entendidos aqui como uma espécie de mudança de paradigma ante uma evidente crise que se anunciava - que tem tomado esta ciência na América Latina, desde os anos de 1970.

Mas o que o museu e seu campo de estudos teriam a ver com a Ciência da Informação? A afirmativa a seguir parece bastante esclarecedora no sentido de ajudar nessa contextualização, ao considerar que:

Dentre as chamadas instituições de memória, o museu tem um papel ímpar na sociedade moderna como mediador entre o público e o acervo, e enquanto comunicador e produtor de discurso. A forma pela qual o público recebe o conteúdo e a mensagem de uma exposição museológica assegura e garante a legitimidade da função social do museu. Talvez seja esta a possibilidade de perceber que a associação entre museu, comunicação e informação configura-se tão estreita e legítima quanto a solidificação das instituições sociais (Castro, 2007, p.105).

No tocante à aproximação entre a Museologia e a Ciência da Informação, ainda se pode lembrar que Le Coadic (2004) teceu algumas reflexões acerca de quatro práticas de organização - entre elas a Museoeconomia - que, segundo ele, atuaram no campo da informação.

Isto posto, o museu, inicialmente, foi pensado e construído para ser visto/consumido por poucos; aliás, "o livro na biblioteca e o objeto no museu foram durante muito tempo recolhidos, armazenados e preservados por um conservador com o fim único da preservação patrimonial" (Le Coadic, 2004, p.12). Suas noções e ações em torno do sagrado e da sacralização de objetos/documentos e monumentos afastavam a muitos. Isso porque se defendia a ideia de que a compreensão e apreensão de determinados conhecimentos não era tangível à coletividade, cientificamente tida como inferior (Schwarcz, 1993).

Essa linha de pensamento, baseada na exclusão do popular e do próprio senso comum e no ranqueamento de competências e capacidades, que vigorou desde o século XIX - época da organização e institucionalização dos primeiros espaços museais (Schwarcz, 1993) - e atravessou o século seguinte, deu os primeiros sinais de crise quando se começou a perceber, para além da ausência de público nos museus, sobretudo na América Latina, profundas crises nos campos da política, economia e cultura desses países (Conselho Internacional de Museus, 1972).

Em virtude de tais constatações, realizou-se, em 1972, uma conferência do Conselho Internacional de Museus, em Santiago, no Chile, onde se deliberou acerca do papel decisivo, que cabe aos museus, no tocante à educação das comunidades. Ficou decidido também que a estes competiria a missão de tornar suas coleções acessíveis ao máximo, tanto a pesquisadores qualificados quanto a entidades públicas, privadas e religiosas. O museu foi definido, ainda, como uma instituição a serviço da sociedade no sentido de participar da formação de sua consciência sem isentar o protagonismo das comunidades nessa ação (Conselho Internacional de Museus, 1972). Doze anos mais tarde, o *The International Council of Museums* (ICOM), reunido no Canadá, através da Declaração de Quebec, ratificaria as deliberações feitas em Santiago (Conselho Internacional de Museus, 1984).

No ano em que se celebraram os vinte anos da confecção da Declaração de Santiago, discutiram-se, em

Caracas, na Venezuela, entre outros aspectos, a confirmação da vigência dos postulados feitos em Santiago; uma reflexão sobre a ação social do museu e a necessidade da tomada de consciência dos poderes públicos acerca da importância da inserção de políticas museológicas no campo da cultura; as crises políticas, econômicas e culturais que se intensificavam por quase toda a América Latina entre fins dos anos de 1980 e começo dos de 1990; e o museu não apenas como fonte de informação ou instrumento de educação, mas, fundamentalmente, como um espaço adequado onde a comunidade possa se expressar e interagir com os processos e produtos culturais, fomentando, assim, um desenvolvimento equilibrado e um maior bem-estar coletivo (Conselho Internacional de Museus, 1992).

Deve-se registrar, por fim, que nas três conferências tratadas, tomaram parte representantes da Argentina, da Bolívia, do Brasil, do Chile, da Colômbia, de Cuba, do Equador, do México, da Nicarágua, do Peru e da Venezuela. Portanto, apenas países latino-americanos, que buscavam, através de novos paradigmas, soluções para problemas e crises que se lhes apresentavam à época. É desse modo que, em fins do século XX, diante de algumas crises e adversidades pelas quais o museu passou em relação à sua funcionalidade e concessão de significado e valor para o usuário, um grupo de especialistas, a partir de 1972, se reuniu para traçar uma mudança de paradigma para a Museologia e para o próprio museu. Este, para se manter vivo e superar suas próprias crises, precisou aceitar-se como um espaço de (re)invenção, de interatividade, de ludicidade, de franco convite à sociedade como um todo. O seu foco, por sinal, transferiu-se do objeto/documento/monumento para o visitante/usuário. Pode-se dizer, com isso, que, tal como nos arquivos e nas bibliotecas, o museu tem agido no sentido de substituir o paradigma custodial, patrimonialista, historicista e tecnicista pelo paradigma pós-custodial, informacional e científico (Silva; Ribeiro, 2011).

É justo e necessário, nesse sentido, enfatizar o ponto de discordância com Thomas Kuhn - acerca da necessidade de ruptura para o advento de novos paradigmas - que, por outro lado, nos aproxima da ideia defendida por Silva e Ribeiro (2011) no que concerne à vigência simultânea de dois paradigmas distintos e até antagônicos, ainda que um deles, o mais antigo, dê sinais de esgotamento ou colapso.

Um clássico exemplo de tal possibilidade é a relação entre a oralidade e a escrita, em que se percebe o entrecruzamento de ruptura e continuidade nesse - apenas aparente - par de contrários (Silva; Ribeiro, 2011), uma vez que a fala e seu registro impresso não de-vem ser percebidos como opostos, mas como “o efeito de distinções recíprocas dentro de configurações históricas sucessivas e imbricadas” (Certeau, 2008, p.223).

Outra ilustração acerca da vigência simultânea de dois paradigmas, que talvez melhor se aplique a essas considerações, é a existência do paradigma custodial, patrimonialista, historicista e tecnicista - cujos traços fundamentais são a supervalorização da custódia ou da guarda e a importância crescente do acesso ao conteúdo - e do emergente paradigma pós-custodial, informacional e científico - marcado pela valorização da informação como um fenômeno social e com prioridade máxima dada ao acesso à informação (Silva; Ribeiro, 2011).

Parece relevante considerar, em relação ao paradigma custodial, o mais antigo, que este ainda parece vigorar na maioria das instituições de memória - a saber: arquivos, bibliotecas e museus - onde a desejada universalização da cordialidade da informação ainda é um sonho distante, posto que o usuário, muitas vezes, é identificado como *persona non grata*, como um inimigo, um agente em potencial da temida entropia no sistema; e o paradigma pós-custodial aparece como um estado de exceção. A busca por essa mudança efetiva depende em muito da biblioteca ou do museu que se quer para o futuro. Se tais instituições foram criadas pelo e para o homem, um ser por natureza dinâmico tal como sua realidade, seu cotidiano e suas relações em sociedade, por que também esses lugares de memória não podem operar transformações em seus métodos e concepções? Um organismo que não se reinventa ante a adversidade está fadado ao desaparecimento.

Concluída a breve análise de aplicação de mudança de paradigma no campo da Museologia, é este o momento que se entende como propício para analisar alguns paradigmas atribuídos à Ciência da Informação. Sabe-se que o conceito de informação é explorado sob diferentes prismas, que estão relacionados a basicamente três paradigmas, e que servem de sustentáculo para diferentes análises do fenômeno da informação. Seriam eles o paradigma físico, o cognitivo e o social (Capurro, 2003).

Para Capurro (2003, p.4):

[...] a Ciência da Informação nasce em meados do século XX com um paradigma físico, questionado por um enfoque cognitivo idealista e individualista, sendo este por sua vez substituído por um paradigma pragmático e social [...] mas agora de corte tecnológico digital.

Em relação ao paradigma físico, este postula que uma mensagem, e não uma informação, é transmitida a partir de um emissor a um receptor. De acordo com essa teoria, defendida por Shannon, não seria a informação, mas a mensagem o que reduz a incerteza. Esse paradigma exclui, assim, o papel ativo do sujeito cognoscente, identificado como usuário, no processo informativo e comunicativo, em geral. Os limites desse paradigma conduziram a um novo paradigma: o cognitivo (Capurro, 2003).

De acordo com Capurro (2003), o paradigma cognitivo, proposto por Brookes, entre outros, respalda-se na ontologia popperiana e sua conhecida distinção de três mundos: o do físico, o da consciência e o do conteúdo intelectual das teorias científicas. Este último é aplicado ao modelo de *Brookes* como agente formador de uma rede existente nos espaços cognitivos, chamada informação objetiva.

E essa informação objetiva seria percebida como uma perspectiva cognitiva a partir do momento em que se busca verificar de que forma os processos informativos são capazes, ou não, de transformar o usuário - sujeito cognoscente dotado de modelos mentais do mundo exterior que são transformados ao longo do processo informacional - que tem necessidade de se informar para corrigir seu estado anômalo de conhecimento (Le Coadic, 2004).

Essa perspectiva cognitiva, no entanto, ignorava os critérios de subjetividade e autonomia no tocante às análises de seleção e relevância conferidas pelo indivíduo. Este, considerado um ser social, está sob influência da sua área de domínio e do seu contexto e, portanto, o ato da busca de informação do usuário depende, inexoravelmente, não só dos seus conhecimentos apriorísticos, mas também do seu estado de conhecimento da área do domínio - contexto -, e de suas necessidades de informação - entropia/grau de ordem do problema - (Ingwersen, 1992). É assim que, de acordo com o próprio Ingwersen

(1992), os processos informativos trazem a possibilidade de transformação - ou não - dos modelos mentais do mundo exterior do usuário.

Nesse contexto, não se pode abordar o paradigma sociocognitivo no campo da Ciência da Informação sem mencionar as considerações registradas em torno da dimensão de relevância. Esta última é percebida por Saracevic (1996b) como um conceito dinâmico, baseado na cognição humana, que demanda um contexto, dentro do qual se alocam determinados atributos de relevância, a saber: relações (comunicações/trocas); intenções (objetivos/expectativas); inferências (percepções); e interações (processo dinâmico de cognição).

Ainda acerca da relevância aplicada ao paradigma sociocognitivo, deve-se considerar que esta se tornou uma área de interesse para a Ciência da Informação, desde as primeiras contribuições de Vannevar Bush, ao ponto de Saracevic (1996b) no que ele mesmo chamou de revisão crítica do pensamento sobre a natureza da relevância em Ciência da Informação - classificar as seguintes manifestações de relevância: relevância algorítmica; relevância tópica; relevância cognitiva (ou pertinência); relevância situacional (ou utilidade); e relevância afetiva ou motivacional.

De acordo com tais postulados, Saracevic (1996b) define a relevância algorítmica como sendo a relação de efetividade entre uma busca e os textos recuperados pelo sistema de informação; já a relevância tópica é pensada como a relação entre os assuntos de uma busca e dos textos recuperados pelo sistema. A relevância cognitiva, também chamada de pertinência, tem a ver com qualidade, novidade e correspondência cognitiva entre uma busca e os textos recuperados pelo sistema, enquanto que a relevância situacional, ou de utilidade - importante na redução de incertezas e na tomada de decisões - é entendida como a relação entre a situação, a tarefa ou o problema e os textos recuperados pelo sistema. Por fim, a relevância motivacional, ou afetiva, estaria relacionada às intenções e metas do usuário e sua satisfação com o produto obtido.

Desse modo, o usuário é percebido como um ser dotado de um papel eminentemente ativo na sua busca por informação, e influenciado pelo contexto social.

Registre-se que, com isso, não se quer dizer que o paradigma social venha sendo aplicado com sucesso inquestionável; há também problematizações em sua práxis, as quais, entretanto, não serão analisadas neste trabalho.

## Considerações Finais

A Ciência da Informação, desde os anos de 1940, vem recebendo valiosas contribuições de várias áreas do conhecimento e com elas vem buscando amadurecer pontos de vista no intuito de consolidar algumas definições. Dessa forma, de modo algum ela parece uma ciência incipiente. A sua maturidade, por outro lado, ainda não foi alcançada, assim como a definição de alguns paradigmas não estão de todo delimitados. Nesse sentido, haveria algum paradigma específico para a Ciência da Informação?

Considerando a complexidade da autonomia humana e sua demanda por condições socioculturais - tais como linguagem e saberes - para consolidar-se em sua independência, devemos destacar a necessidade de que esta cultura demandada seja diversificada ao ponto de criar alternativas de escolhas diante do acervo de ideias existentes, para que se possa refletir de modo autônomo.

Pensa-se que a busca pela autonomia humana, em muitos momentos se aplica à busca por uma autonomia da ciência, onde se pode encontrar muitos e variados caminhos de raciocínio, como também mais dados e mais pontos de vista. E, nesse caso, a Ciência da Informação já dispõe de um estoque de possibilidades minimamente necessário para alcançar a sua autonomia como ciência. Ciência esta que nasceu interdisciplinar e que não parece querer se afastar dessa sábia característica, já que denota a ideia de que os conceitos não são definidos pelas fronteiras.

Pode-se anotar, por fim, que a Ciência da Informação se utiliza de paradigmas de outros campos do conhecimento, mas ainda não possui, ela própria, um paradigma que delimite conceitos e áreas de concentração e atuação. Talvez a emergência de um paradigma consensual neste campo do conhecimento, no contexto atual - de uma sociedade globalizada e interconectada - possa encontrar seu ponto de partida nas perguntas-chave: informação para quem? Informação para quê?

## Agradecimentos

Às agências Conselho Nacional de Desenvolvimento Científico e Tecnológico e Fundação de

Amparo à Ciência e Tecnologia do Estado de Pernambuco, cujas bolsas de produtividade e pesquisa, como também a de mestrado permitiram a realização deste trabalho.

## Referências

- BACHELARD, G. *A formação do espírito científico: contribuição para uma psicanálise do conhecimento*. Rio de Janeiro: Contraponto, 1996.
- BARRETO, A.A. A condição da informação. *São Paulo em Perspectiva*, v.16, n.3, p.67-74, 2002.
- BUCKLAND, M. Information as thing. *Journal of the American Society for Information Science*, v.45, n.5, p.351-360, 1991.
- BURKE, C. *History of information science*. Baltimore: University of Maryland, 2007.
- CAPURRO, R. *Epistemologia e ciência da informação*. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., 2003, Belo Horizonte. *Anais eletrônicos...* Belo Horizonte: ENANCIB, 2003. p.1-21. Disponível em: <<http://www.uff.br/cienciainformacao/Disiplinas/Capurro-Epistemologia%20e%20CI.pdf>>. Acesso em: 1 jul. 2011.
- CAPURRO, R.; HJORLAND, B. O conceito de informação. *Perspectivas em Ciência da Informação*, v.12, n.1, p.148-207, 2007.
- CASTELLS, M. *A sociedade em rede*. 6.ed. São Paulo: Paz e Terra, 1999.
- CASTRO, A.L.S. *Memórias clandestinas e sua museificação*. Rio de Janeiro: Revan, 2007.
- CERTEAU, M. *A invenção do cotidiano*: 1. artes de fazer. 15.ed. Petrópolis: Vozes, 2008.
- CONSELHO INTERNACIONAL DE MUSEUS. *Declaração de Santiago*: mesa-redonda de Santiago do Chile. 1972. Disponível em: <[http://www.museologia-portugal.net/index.php?option=com\\_content&view=article&id=3:declaracao-de-santiago-1972&catid=3:declaracao-de-santiago-do-chile-1072&Itemid=3](http://www.museologia-portugal.net/index.php?option=com_content&view=article&id=3:declaracao-de-santiago-1972&catid=3:declaracao-de-santiago-do-chile-1072&Itemid=3)>. Acesso em: 21 jun. 2011.
- CONSELHO INTERNACIONAL DE MUSEUS. *Declaração de Quebec*: princípios de base de uma nova museologia. 1984. Disponível em: <<http://www.revistamuseu.com.br/legislacao/museologia/quebec.htm>>. Acesso em: 21 jun. 2011.
- CONSELHO INTERNACIONAL DE MUSEUS. *Declaração de Caracas*. 1992. Disponível em: <[http://www.revistamuseu.com.br/legislacao/museologia/decl\\_caracas.asp](http://www.revistamuseu.com.br/legislacao/museologia/decl_caracas.asp)>. Acesso em: 21 jul. 2011.
- HJORLAND, B. Documents, memory institutions and information science. *Journal of Documentation*, v.56, n.1, 2000. Available from: <[http://www.iva.dk/binaries/documents\\_memory%20institutions%20and%20is.pdf](http://www.iva.dk/binaries/documents_memory%20institutions%20and%20is.pdf)>. Cited: 22 June 2011.
- HOBSBAWM, E.J. *Era dos extremos: o breve século XX - 1914-1991*. São Paulo: Companhia das Letras, 1995.
- INGWERSEN, P. *Information retrieval interaction*. London: Taylor Graham, 1992.
- KUHN, T. A função do dogma na investigação científica. In: DEUS, J.D. (Org.). *A crítica da ciência: sociologia e ideologia da ciência*. 2.ed. Rio de Janeiro: Zahar, 1979. p.53-80.
- LE COADIC, Y.F. O objeto: a informação. In: LE COADIC, Y.F. *A ciência da informação*. 2.ed. Brasília: Briquet de Lemos, 2004. p.3-11.
- LÉVY, P. *Cibercultura*. São Paulo: Editora 34, 1999a.
- LÉVY, P. *O que é virtual?*. São Paulo: Editora 34, 1999b.
- McGARRY, K. *O contexto dinâmico da informação: uma análise introdutória*. Brasília: Briquet de Lemos, 1999.
- MORIN, E. *Ciência com consciência*. 6.ed. Rio de Janeiro: Bertrand Brasil, 2002.
- MORIN, E. Da necessidade de um pensamento complexo. In: MARTINS, F.M.; SILVA, J.M. (Org.). *Para navegar no século XXI*. 3.ed. Porto Alegre: Sulina, 2003. p.13-36.
- MORIN, E. *Introdução ao pensamento complexo*. 4.ed. Porto Alegre: Sulina, 2011.
- NAGEL, E. Ciência: natureza e objetivo. In: MORGENBESSER, S. (Org.). *Filosofia da ciência*. São Paulo: Cultrix, 1975. p.11-24.
- ORTEGA Y GASSET, J. *A rebelião das massas*. 2.ed. São Paulo: Martins Fontes, 2006.
- POPPER, K. *A lógica da pesquisa científica*. São Paulo: Cultrix, 2007.
- SARACEVIC, T. Ciência da informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, v.1, n.1, p.41-62, 1996a.
- SARACEVIC, T. Relevance reconsidered. In: INGWERSEN, P.; PORS, N.O. *Information science: integration in perspective*. Copenhagen, Denmark: Royal School of Library and Information Science, 1996b.
- SCHWARCZ, L.M. *O espetáculo das raças: cientistas, instituições e questão racial no Brasil - 1870-1930*. São Paulo: Companhia das Letras, 1993.
- SILVA, A.M.; RIBEIRO, F. *Paradigmas, serviços e mediações em ciência da informação*. Recife: Néctar, 2011.
- ZILIO, D.; GONZALEZ, M. A gênese do significado. *Revista Filosofia, Ciência & Vida*, v.5, n.53, p.14-22, 2010.
- ZINS, C. Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, v.58, n.4, p.479-493, 2007.

# O reflexo da inteligência coletiva nas organizações

## *Effect of collective intelligence in organizations*

Ketry Gorete Farias dos PASSOS<sup>1</sup>

Edna Lúcia da SILVA<sup>2</sup>

### Resumo

As novas tecnologias de informação e comunicação possibilitaram o compartilhamento das informações de maneira rápida e eficaz. Da mesma forma, as organizações têm se apropriado das novas tecnologias para comunicar e disseminar informações. Este artigo tem por objetivo apresentar reflexões sobre como a *Web 2.0* pode empoderar as organizações na sociedade em rede. Para isso, utilizou-se de uma revisão de literatura de caráter exploratório na qual foram abordados conceitos como inteligência coletiva, importância da sociedade em redes para as organizações, *Web 2.0* e Empresa 2.0. Concluiu-se que as organizações estão se conscientizando cada vez mais do potencial da *Web 2.0*, não somente para divulgar produtos, mas também como um espaço para a troca de conhecimento, aprimoramento de serviços, bem como para solução de problemas. Assim as empresas que atuam em redes conseguem superar melhor suas limitações com auxílio de seus colaboradores e clientes por meio da inteligência coletiva.

**Palavras-chave:** Empresa 2.0. Inteligência coletiva. Sociedade em rede. *Web 2.0*.

### Abstract

*The new information and communication technologies have enabled information to be shared quickly and effectively. Similarly, organizations have adopted new technologies to communicate and disseminate information. This article therefore presents reflections on how Web 2.0 can empower organizations in the network society. To do this, concepts of collaborative networking, collective intelligence, the importance of company networks for organizations, Web 2.0 and Enterprise 2.0, were discussed. It was concluded that organizations are increasingly becoming aware of the potential of Web 2.0, not only to advertise products, but also as a space for exchange of knowledge, improvement of services and troubleshooting. Thus, companies that operate in networks can best overcome their limitations with the help of their employees and customers through collective intelligence.*

**Keywords:** Enterprise 2.0. Collective intelligence. Network society. Web 2.0.

### Introdução

A emergência da *Internet* a partir do final da década de 1980 e o surgimento da *World Wide Web* em 1994 prolongaram a precedente evolução da esfera pública, introduzindo ao mesmo tempo elementos

radicalmente novos: a interconexão geral, a desintermediação e a comunicação de todos com todos (Lévy, 2005). Assim, inicia-se um novo ambiente de comunicação, no qual a virtualidade se faz presente e o tempo não é mais o fator que intervém na transferência e no compartilhamento das informações. Estamos na era da informação: a

<sup>1</sup> Mestranda, Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Ciência da Informação. *Campus* Universitário, Trindade, 88045-000, Florianópolis, SC, Brasil. Correspondência para/Correspondence to: K.G.F. PASSOS. E-mail: <ketry2003@gmail.com>.

<sup>2</sup> Professora Doutora, Universidade Federal de Santa Catarina, Departamento de Ciência da Informação e do Programa de Pós-Graduação em Ciência da Informação. Florianópolis, SC, Brasil.

Recebido em 26/7/2011, reapresentado em 5/3/2012 e aceito para publicação em 30/3/2012.



distância entre as pessoas não interfere mais no fluxo entre o emissor e o receptor da informação.

Este artigo tem por objetivo apresentar algumas reflexões sobre como a *Web 2.0* pode empoderar as organizações na sociedade em redes. Portanto, torna-se imperativo apresentar alguns conceitos pertinentes à inteligência coletiva. Para isso, realizou-se uma pesquisa bibliográfica de caráter exploratório. A partir da revisão de literatura, constatou-se que alguns autores se popularizaram em defesa de conceitos, como o de Inteligência Coletiva (Lévy, 1996), Sociedade em Redes (Castells, 1999), Colaboração e Inteligência Coletiva (Lévy, 2005), *Web 2.0* (O'Reilly, 2005) e Empresa 2.0 (Tapscoot; Williams, 2007).

No que diz respeito às mudanças na forma de comunicação, Castells (1999) acredita que estamos numa época de transição de modernos mecanismos para um novo paradigma tecnológico que se organiza em torno da tecnologia da informação. Para Castells (1999), o que caracteriza a atual revolução da informação está relacionado à aplicação da informação para a geração de conhecimentos e de dispositivos de processamento/comunicação da informação em um ciclo de realimentação entre a inovação e seu uso. Com a mudança dos meios de comunicação, o modo de relacionamento entre os indivíduos também se modificou.

Com o advento da *Internet* e de outras tecnologias da informação, como a *Web 2.0*, tornou-se possível ter acesso a muitos dados. Algumas vezes, mais dados sobre clientes, empregados e *stakeholders* do que anteriormente. A princípio, as organizações ganharam mais precisão e compreensão do seu ambiente externo e interno. As recentes tecnologias, incluindo muitas das aplicações da *Web 2.0*, possibilitaram às organizações explorar o "coletivo" em melhor escala. De fato, o aumento do mercado de informação, as *wikis*, *crowdsourcing*<sup>3</sup>, os conceitos de sabedoria das multidões, as redes sociais, os *softwares* colaborativos e outras ferramentas baseadas na *Web* constituem o novo paradigma a partir do qual as organizações tomam decisões (Bonabeau, 2009).

Assim como nas relações sociais, a sociedade em redes e a colaboração em massa podem trazer algumas vantagens competitivas às organizações. Dessa forma,

aborda-se a relação entre a sociedade em rede e a *Web 2.0* como elemento potencializador da inteligência coletiva. As organizações perceberam que a colaboração pode ser uma ferramenta de geração do conhecimento e inovação no desenvolvimento de produtos (bens e/ou serviços) e também um meio de comunicação eficaz e eficiente com clientes e parceiros.

Por isso, quando se aborda o papel da sociedade em redes para as organizações, parte-se do princípio de que, com o advento das novas tecnologias da *Web 2.0*, tornou-se possível para muitas organizações de todo o mundo conhecer melhor os fatores que permeiam seu ambiente externo e interno. As ferramentas da *Web 2.0* possibilitaram a geração de forma contínua e adaptável a novas situações do ambiente organizacional; portanto, se bem utilizadas, podem contribuir para destacar as organizações no mercado cada vez mais competitivo.

De acordo com Lévy (1996), quando refletimos no potencial que a *Web 2.0* pode proporcionar às organizações, estamos pensando no poder da inteligência coletiva. O conceito de inteligência coletiva pode ser definido como um tipo de inteligência que é disseminado em grande escala e em tempo real no ciberespaço.

Em relação a isso, Lévy (2005), defende que um dos aspectos mais transformadores da globalização foi a informatização e a consequente geração da sociedade em rede. Para Lévy (2005), o ciberespaço é muito mais inclusivo do que todos os outros meios de comunicação anteriores, pois ele permite a expressão pública a todos os indivíduos, grupos, instituições e comunidades. Desse modo, as mídias não se ligam mais a um público localizado, mas a uma comunidade virtual distribuída por toda a parte num mundo de ouvintes, expectadores, leitores e contribuintes.

Analogamente, devido à competitividade e crescente mudança no modo de desenvolver produtos, sejam eles bens ou serviços, imperou-se aos gestores a constante preocupação com atualização e qualidade no desenvolvimento de produtos. Nesse contexto, visto que o ciberespaço possibilita a junção de inteligências coletivas em escalas globais e em tempo real, por que as organizações não podem se aproveitar do potencial da inteli-

<sup>3</sup> Modelo de produção que utiliza a inteligência e os conhecimentos coletivos de voluntários da *Internet* para resolver problemas, criar conteúdos ou desenvolver novas tecnologias (Tapscoot; Williams, 2007).

gência coletiva para criar soluções corporativas? Organizações do mundo todo, conscientes do potencial das novas tecnologias da informação, estão buscando cada vez mais ideias que impulsionem o mercado, conectando-se e utilizando-se de ferramentas da *Web 2.0*. Lévy (1996) acredita que na sociedade atual, quanto mais se é virtual, mais se faz dinheiro, quanto mais se penetra no mundo das ideias, mais se é recompensado pelo mercado. Esse potencial está cada vez mais sendo explorado pelas organizações por meio das redes sociais da *Internet*, como, por exemplo, *Orkut*, *Facebook*, *Twitter*, *Myspace*, *Ning*, entre outros.

As organizações se deram conta de que, além de ser um meio de comunicação mais econômico e acessível do que os convencionais, a *Internet* fornece uma variedade de ferramentas, o que possibilitou sanar muitas dúvidas em tempo hábil em relação à percepção do cliente a um determinado produto ou empresa. Também se tornou possível, por intermédio das ferramentas da *Web 2.0*, capturar críticas e/ou sugestões em tempo real sobre o lançamento de produtos ou tendências do mercado. Atualmente, com a *Internet*, é possível perguntar aos clientes quais são suas expectativas em relação a novos produtos (bens ou serviços), avaliar de forma rápida e com custos quase nulos seu processo produtivo e/ou de prestação de serviços, classificar e encontrar demandas de perfis de usuários através da *Web social* (*Orkut*, *Twitter*, *Facebook*, *Ning*, entre outros), descobrir a personalidade e as preferências de determinados clientes, entre outras opções.

Portanto, neste artigo, parte-se do pressuposto de que a *Web 2.0* e a inteligência coletiva podem potencializar as organizações. Para isso, apresentam-se conceitos de comunidade virtual, sociedade em rede e de inteligência coletiva, que são fundamentais para a compreensão dos aspectos que envolvem o conceito de *Web 2.0*.

## A colaboração e a inteligência coletiva

Lévy (2005) explica que a *Internet* criou um novo tipo de comunidade, uma comunidade que não é igual à tradicional. Para compreender como as organizações podem se beneficiar do potencial da inteligência coletiva, é necessário anteriormente explicar o conceito de comu-

nidade virtual. Na visão do autor, as mudanças provocadas pela globalização e pelas novas tecnologias, além da desterritorialização da esfera pública, deixaram antever a ascensão de comunidades desvinculadas dos territórios físicos. Como consequência, as mídias não se ligam mais a um público localizado, mas a uma comunidade virtual distribuída por toda parte num mundo de ouvintes, espectadores, leitores e contribuintes. Para Lévy (2005), as economias virtuais se constituíram de maneira original no ciberespaço, a partir da vontade de comunicação em torno de pontos comuns entre os internautas. O tamanho dessas comunidades envolve questões comerciais, culturais e políticas maiores, na medida em que representam ao mesmo tempo um mercado, uma fração da opinião no novo espaço público e uma potência da inteligência coletiva (coordenação, cooperação, troca de saberes, ajuda mútua).

A transformação da sociedade tradicional para a sociedade virtual está centrada no fato de que, atualmente, com o advento das novas tecnologias de informação, nos tornamos produtores, coprodutores, disseminadores e canais de informação ao mesmo tempo. As tecnologias da *Web 2.0* possibilitaram aos indivíduos a fácil navegação, criação e edição de conteúdos. Desse modo, por exemplo, não precisamos mais contatar uma emissora de televisão para produzir um documentário, basta apenas filmar com uma simples câmera amadora e enviar o conteúdo ao *Youtube* e em questão de segundos algumas centenas ou milhares de pessoas podem assistir à produção, sem custos e muito menos intermediários.

Em relação a isso, dois conceitos pertinentes à colaboração na sociedade em redes podem exemplificar o poder das novas tecnologias de informação e comunicação para a inteligência coletiva: *Crowdsourcing* e *Peering*. Para Tapscoot e Williams (2007), o *Crowdsourcing* consiste no modelo de produção que utiliza a inteligência e os conhecimentos coletivos de voluntários espalhados pela *Internet* para resolver problemas, criar conteúdos ou desenvolver novas tecnologias. E *Peering* é uma forma de produzir produtos que dependem totalmente de comunidades auto-organizadas e igualitárias de indivíduos que se unem voluntariamente para produzir um resultado compartilhado. O *Peering* utiliza as motivações voluntárias de maneira a ajudar a destinar à pessoa certa a tarefa

certa com mais eficácia do que empresas tradicionais. O motivo é a autosseleção. Ao voluntariamente se selecionarem para tarefas criativas e com uso intensivo de conhecimento, as pessoas estão distintamente capacitadas.

O exemplo mais popular de *Peering* é o sistema operacional *Linux*. Em 1991, Linus Torvalds criou uma versão simples do sistema operacional *Unix*. A partir das colaborações de outros programadores, Linus decidiu licenciar o sistema operacional sob uma Licença Pública Geral (GPL) para que qualquer pessoa pudesse usá-lo gratuitamente, contanto que disponibilizasse para os outros as mudanças feitas no programa. Por ser confiável e gratuito, tornou-se um sistema operacional útil para computadores que abrigavam servidores da *Web* e banco de dados. Muitas empresas consideram o *Linux* uma pedra fundamental dos *softwares* empresariais. Apesar de poder ser usado livremente, o *Linux* tem sido incorporado em todos os tipos de produtos lucrativos desenvolvidos por grandes empresas, como BMW, IBM, Motorola, Philips e Sony. A nova promessa de colaboração é que com o *Peering*, pode-se explorar a capacidade, a engenhosidade e a inteligência humana com mais eficiência e eficácia. Essa promessa substituirá as hierarquias empresariais tradicionais como mecanismo essencial para a criação de riqueza na economia (Tapscoot; Williams, 2007).

## O papel da sociedade em redes para as organizações

Castells (1999) entende que a cultura da virtualidade associada às novas tecnologias de informação e comunicação contribui para a transformação do tempo em nossa sociedade de duas formas: na simultaneidade e na intemporalidade. Atualmente, a virtualidade possibilita a comunicação entre as pessoas em tempo real, reunindo indivíduos com os mesmos interesses em conversas interativas, oferecendo instantaneidade, que vem conquistando barreiras temporais. A simultaneidade e a intemporalidade possibilitaram à sociedade organizar-se em torno de redes de comunicação na *Internet*. As redes possibilitaram aos indivíduos compartilhar informações e desenvolver soluções em colaboração, independentemente do local e até mesmo do tempo. Um

exemplo dessa afirmação são os grupos de discussões abertos que recebem diariamente contribuições de milhares de internautas dispostos a ajudar sem receber necessariamente nada em troca, a não ser pelo *status* de pró-atividade e pelo sentimento de colaboração.

Um aspecto importante a ser esclarecido é o papel das redes para as organizações. Na compreensão de Castells (1999), a rede é constituída por nós e conexões dois a dois, que podem interligar-se direta ou indiretamente. Assim, a *Internet* como novo meio de comunicação rompe com a clássica distribuição hierárquica entre emissores e receptores ao possibilitar que cada nó possa produzir e distribuir mensagens simultaneamente. Além disso, a rede é composta por uma estrutura aberta capaz de expandir-se de forma ilimitada, integrando novos nós desde que consigam comunicar-se internamente, o que faz com que o processo de comunicação seja expandido de maneira exponencial. Nesse contexto, Castells (1999) afirma que uma sociedade baseada em redes pode constituir-se num sistema aberto altamente dinâmico e suscetível à inovação sem ameaças ao seu equilíbrio.

Um exemplo do potencial da sociedade em redes é o grupo de discussão *TakingITGlobal*: 110 mil membros registrados em quase 200 países se reúnem com líderes empresariais e governamentais no Fórum Econômico Mundial. O *site* lista mais de 2 mil projetos comunitários iniciados/administrados por jovens (entre 13 e 24 anos) que enfrentam questões desde a eliminação do hiato digital nas áreas rurais da Índia à prevenção do HIV em Uganda. Essas novas colaborações atenderão não apenas a interesses comerciais, mas também ajudarão as pessoas a executarem tarefas como espírito público, desde curar doenças genéticas, prever mudanças climáticas globais e encontrar novos planetas e astros (Tapscoot; Williams, 2007).

Dessa forma, podemos observar que a união da sociedade em redes é um grande passo para a solução de grandes problemas sociais que podem ser resolvidos por meio do compartilhamento de ideias e da inteligência coletiva. Portanto, já que os problemas são da sociedade, por que não deixar que ela possa refletir e encontrar soluções em conjunto, numa junção de redes de pensamentos para solucionar seus próprios problemas?

A esse respeito, Castells (1999) defende que as redes são os instrumentos mais apropriados para a economia capitalista baseada na inovação, globalização e concentração descentralizada. As redes podem beneficiar o trabalho, os trabalhadores e as empresas voltadas para a flexibilidade e a adaptabilidade, bem como para uma cultura de desconstrução e reconstrução contínuas. Portanto, a tendência é que certas organizações desenvolvam produtos e gerem soluções de problemas por meio da colaboração e da exploração do potencial da inteligência coletiva. Para Castells (1999), nem sempre a solução está dentro das organizações, às vezes, é necessário olhar para fora, encontrar outros pontos de vista.

Castells (1999) sustenta a ideia de que as redes são a nova morfologia social da coletividade. A utilização em grande escala das redes pode modificar os resultados de processos produtivos, de experiências, de poder e de cultura. A presença ou a ausência na rede são fontes cruciais de dominação e transformação de nossa sociedade: uma sociedade que, portanto, pode-se denominar de sociedade em rede, caracterizada pela primazia da morfologia social sobre a ação social.

Portanto, o novo modo de se comunicar criado pela sociedade transpôs barreiras temporais e geográficas. Atualmente o poder das redes de comunicação faz com que pessoas de diversas partes do mundo se comuniquem com milhares de outras, utilizando apenas um computador e a *Internet*. Por trás do poder que as redes de comunicação vêm proporcionando, está o poder da inteligência coletiva.

Dispor da colaboração para discussão e apreciação dos fatos na *Web* tornou-se tarefa cotidiana da sociedade em redes, mas esse conjunto de ideias, a mistura do pensamento de diferentes mentes, várias opiniões e diversos pontos de vista não gerariam contradições? Para Lévy (1996), pelo contrário, a sociedade humana atravessou uma barreira de interconexão sem precedentes na história, visto que atingiu o estágio em que ela passa a dispor de uma memória. A sociedade inventou a si mesma como coletivo capaz de aprender em longo prazo, independente da morte dos indivíduos. Por isso, a inteligência é concebida como o conjunto de aptidões cognitivas de saber, de perceber, de lembrar, de aprender, de imaginar e de raciocinar. Na medida em que os indivíduos possuem essas aptidões, podem ser considerados inteligentes.

## Inteligência coletiva

A inteligência coletiva pode ser definida como uma inteligência distribuída em toda parte, continuamente valorizada e cooperada em tempo real. Emergente de processos cooperativos competitivos, a inteligência coletiva se baseia essencialmente na cooperação de agentes concorrentes. Logo, quanto mais se avança na cooperação, mais se abre o leque das formas de competição. Em contrapartida, a cooperação leva à concorrência e à cooperação no interior de um elo auto-criador que engloba uma realidade cada vez mais vasta, diversa e integrada, que é o movimento em si da inteligência coletiva (Lévy, 1996).

Johnson (2003) explica que quando apenas poucos cérebros exploram determinado problema, as células permanecem desconectadas, cada uma seguindo um curso aleatório. No entanto, quando conectados a um maior número de mentes, há uma aglutinação e um novo modo de ver o mundo, compartilhado por milhares de indivíduos. Para Johnson (2003) está havendo uma mudança de pensamento dos indivíduos, estamos mudando regras de pensamento do nível *top-down* para *bottom-up*: não agimos mais a partir de ideias de um líder, as decisões não estão somente sendo tomadas pela alta administração. Existe um imenso potencial para geração de ideias e os gestores devem aproveitar o potencial de todos os membros da organização para fazer uso das ideias coletivas. Atualmente, atos locais geram ações coletivas que podem produzir um comportamento global; como na colônia de formigas, o cérebro da colônia nada mais é do que a soma de milhares e milhares de decisões sendo executadas individualmente. Logo, as formas de comportamento emergentes ficam mais inteligentes com o tempo e reagem melhor às necessidades específicas e mutantes de seu ambiente. Essa nova maneira de pensar é o potencial que a inteligência coletiva possibilita à sociedade conectada.

Esse potencial de exploração da inteligência coletiva pode ser exemplificado no caso da *Goldcorp*. A empresa, em 1999, era uma pequena firma de mineração de ouro, com sede em Toronto, que estava lutando para sobreviver, sitiada por greves, dívidas e um custo de produção excessivamente alto. McEwen, diretor geral, após assistir a uma conferência do *Massachusetts Institute of*

*Technology* (MIT) para jovens presidentes, achou que talvez a solução de seus problemas fosse abrir o processo de exploração da mesma maneira que Torvalds “abriu o código” do *Linux*. Em março de 2000, o “Desafio *Goldcorp*” foi lançado com um prêmio de US\$575 mil em dinheiro para os participantes que tivessem os melhores métodos e estimativas. O objetivo do concurso era encontrar 170 toneladas de ouro. Dessa forma, a ajuda veio tanto de geólogos como de alunos de pós-graduação, consultores, matemáticos, químicos, entre outros. A empresa recebia dos inscritos diversas informações que apresentavam matemática aplicada, física avançada, sistemas inteligentes, computação gráfica e soluções orgânicas para problemas inorgânicos. Mais de 80% dos novos alvos produziram quantidades significativas de ouro. Desde o início do desafio, surpreendentes 230 toneladas de ouro foram encontradas. McEwen estima que o processo colaborativo reduziu dois ou três anos do tempo de exploração. Portanto, a inteligência coletiva pode solucionar problemas de distintas formas (Tapscoot; Williams, 2007).

A partir do conhecimento dos conceitos da sociedade em rede e da inteligência coletiva e do seu potencial para a sociedade em geral bem como para as organizações, pode-se identificar os componentes envolvidos nesse processo de comunicação. Atualmente, com o advento das novas tecnologias de informação e comunicação, criou-se uma variedade de ferramentas na *Internet* com o intuito de facilitar a comunicação no ciberespaço.

Com o desenvolvimento de *softwares* para o compartilhamento de informações, o termo *Web 2.0* foi usado para tentar diferenciar essa nova abordagem de compartilhamento de informação e conhecimento.

## A Web 2.0

A *Web 2.0* não se refere apenas à atualização da *Web* tradicional, mas a uma mudança na forma como ela é encarada por usuários e desenvolvedores, ou seja, o ambiente de interação que engloba inúmeras linguagens e motivações. O conceito “*Web 2.0*” foi criado numa conferência de *brainstorming* entre a O’Reilly e a *MediaLive International* em 2004 (O’Reilly, 2005). A *Web 2.0* ou *Web* social representa um novo padrão de interação: viabiliza e estimula a colaboração, a interação “muitos-muitos” e a

horizontalização da comunicação (Cavalcanti; Nepomuceno, 2007; Antoun, 2008; Primo, 2008; Galdo, 2010).

O sucesso da *Web 2.0* está baseado no fato de que se soube aproveitar o poder da inteligência coletiva. Desse modo, à medida que os usuários adicionam conteúdo e novas páginas da *Web*, eles passam a integrar a estrutura da rede enquanto outros usuários descobrem o conteúdo e se conectam a ele. Do mesmo modo que se formam sinapses no cérebro - com as associações fortalecendo-se em função da repetição ou da intensidade -, a rede de conexões cresce organicamente como resultado da atividade coletiva de todos os usuários da rede (O’Reilly, 2005).

As informações obtidas por meio da *Web 2.0*, além de estratégicas e valiosas para as organizações, podem ser obtidas de forma rápida, segura e mais econômica se comparada aos meios tradicionais de pesquisas. Muitas das tecnologias da *Web 2.0*, principalmente as redes sociais, informam os perfis dos clientes. Comunidades de redes sociais como o *Orkut* podem informar o tipo de atividade do cliente, residência, profissão, livros preferidos, atividades de lazer etc. O *Facebook* é um exemplo clássico de como as empresas podem captar recursos com as ações dos próprios clientes; nele, muitas empresas fazem divulgação de seus produtos e o cliente tem a opção de escolher as propagandas que deseja visualizar.

O diferencial das tecnologias da *Web 2.0* e da *Web 1.0* é que com as ferramentas da *Web 2.0* os indivíduos são consumidores e produtores ao mesmo tempo. Atualmente, muitas ferramentas da *Web 2.0* facilitam a criação do conhecimento no ciberespaço. Um exemplo de como os próprios usuários podem classificar suas informações na *Internet* são os sistemas de *bookmarks* como, por exemplo, o *Del.ici.ous* e outros como o *Flickr*. Nessas ferramentas, é possível adicionar etiquetas (*tags*) e criar suas próprias palavras-chave. Outras tecnologias que reúnem o conceito de *blog* e diferentes formas de interação de usuários, além de facilitarem a troca de informações, estão se tornando muito populares nos últimos anos, entre elas podemos citar o *Flickr* e o *Twitter*. O *Flickr* diferencia-se dos *fotologs* principalmente pela possibilidade de organização das fotos através de *tags* e por disponibilizar opções como inserção de observações dentro da foto, acrescentando informações essenciais para compreensão do contexto no qual a imagem foi retirada (Aquino, 2007).

Autores como Tapscott e Williams (2007) defendem que em vez de se ler um livro é possível escrever um. A Wikipédia é um exemplo do potencial da inteligência coletiva para nossa sociedade. Trata-se de uma enciclopédia criada de maneira colaborativa, escrita por dezenas de milhares de colaboradores, com cinco funcionários em tempo integral. A Wikipédia é dez vezes maior que a enciclopédia britânica e tem aproximadamente a mesma precisão, apesar dos riscos inerentes a uma enciclopédia aberta na qual é possível que qualquer pessoa acrescente suas próprias opiniões. A Wikipédia continua a crescer rapidamente em amplitude, qualidade e tráfego, baseando-se na premissa de que a colaboração entre usuários melhorará o conteúdo ao longo do tempo, da mesma maneira que a comunidade de código aberto aperfeiçoou a primeira versão do *Linux*.

A Wikipédia é o exemplo mais claro do sucesso da colaboração na sociedade em redes. Ela faz tanto sucesso que existem derivações como *wikilivros* (manuais de textos livres), *wikiversidade* (cursos online livres), *wikicionário* (dicionário livre), *wikiquote* (coleção de citações), entre outros. Muitas organizações atentas ao potencial que essa tecnologia oferece estão desenvolvendo programas de incentivo aos funcionários que utilizam *wikis* corporativas para o desenvolvimento da gestão do conhecimento. As *wikis* também podem ser utilizadas para descrição de tarefas complexas de trabalho e registros de eventuais problemas, servindo não apenas como um manual de procedimentos de tarefas, mas também como fonte de informação para as organizações e para futuros membros.

Hoje o fenômeno dos *blogs* aponta para mudanças mais profundas que a nova *Web* causará na economia. Segundo Tapscott e Williams (2007) os *blogs* foram descritos como a maior "cafeteria do mundo". Eles traçam a cada momento um retrato dos pensamentos e sentimentos das pessoas a respeito do que está acontecendo agora, fazendo com que a *Web* deixe de ser uma coleção de documentos estáticos e passe a ser uma conversa em andamento. Os anunciantes já entenderam isso e criaram os seus próprios *blogs* para conversar com clientes interessados em seus produtos. As empresas usam *blogs* como grupos de discussão regularmente, prestando atenção no que as pessoas estão dizendo a respeito de sua própria imagem ou dos seus produtos.

O *blog*, além de ser um meio de comunicação com baixo custo e de simples utilização, necessita de cuidados com sua manutenção, pois não basta apenas criá-los e deixá-los circulando livremente na *Web*. É necessária a constante atualização de informações e constante comunicação com os clientes. Não se trata de uma tarefa muito simples, uma vez que clientes insatisfeitos requerem respostas e esclarecimentos sobre determinados problemas em tempo hábil, porém trata-se de uma nova maneira de se comunicar com os atuais clientes e um espaço para conquistar clientes em potencial. Um exemplo de sucesso com a utilização da ferramenta *blog* é a empresa Claro. O *blog* da Claro (<http://www.claroblog.com.br/>) não serve apenas como meio de divulgação dos produtos da empresa, sua principal função é manter o relacionamento com seus clientes. Para isso, o *blog* da Claro mantém enquetes, promoções, informa os clientes sobre novidades de produtos, bem como constitui-se um canal para sugestões e reclamações.

A partir do conhecimento do potencial que as ferramentas da *Web 2.0* podem proporcionar às organizações, cunhou-se o termo Empresa 2.0. Para Schneckenberg (2009), ferramentas da *Web 2.0*, como plataformas corporativas, possuem o potencial de melhorar o desempenho corporativo ainda que em longo prazo, desde que sejam adotadas como tecnologias de informação pelos funcionários para criar com os pares ambientes de aprendizagem que motivem interações e decisões em suas áreas específicas de interesse.

## Empresa 2.0

McAfee (2006) cunhou o termo *Enterprise 2.0* (Empresa 2.0) para representar as tecnologias da *Internet* e as práticas empresariais que permitem que colaboradores em contextos ricos em informação empresarial produzam e troquem conhecimento em colaboração e interação.

As novas tecnologias da *Web 2.0* garantem às organizações um bom entendimento de eventos e tendências do ambiente externo, interno e no processo de tomada de decisão. As informações advindas das ferramentas da *Web 2.0*, se bem fundamentadas, podem gerar informações e análises de forma contínua e adaptável às novas situações no ambiente organizacional.

Tapscott e Williams (2007) acreditam que empresas inteligentes estão estimulando, em vez de combater, o crescimento vertiginoso de enormes comunidades *online* - muitas das quais surgiram na periferia da *Web*, atraindo dezenas de milhões de participantes da noite para o dia. Com um número maior de empresas descobrindo os benefícios da colaboração em massa, esse novo modo de organização substituirá estruturas empresariais tradicionais como o motor primário de criação de riqueza na economia.

A empresa Fiat pode ilustrar adequadamente como as organizações podem se utilizar da *Web 2.0* para potencializar seu processo produtivo. A Fiat desenvolveu o novo carro Uno pelas sugestões de clientes e internautas por intermédio do *Formspring*. A empresa alocou uma equipe para responder a usuários - predominantemente pertencentes ao público jovem - e captar deles informações e sugestões. Os internautas responderam a cerca de 50 questões relacionadas às características desejadas no novo carro da Fiat, e o resultado foi o novo Uno, desenvolvido pelas sugestões e ideias dos internautas.

A construtora Tecnisa é outro modelo de como as organizações podem se beneficiar das novas tecnologias da *Web 2.0*. Aproveitando a oportunidade de se envolver com os clientes, a construtora criou um apartamento virtual decorado em 3D, semelhante ao apartamento real, e convida os internautas a visitar seu espaço no jogo *Second Life*. O *Second Life* é um *software* que simula os aspectos da vida real e social do ser humano. Dessa forma, os clientes conseguem visualizar o produto antes mesmo de comprá-lo. Isto possibilita que empresas de qualquer parte do mundo divulguem seus produtos de forma rápida e econômica.

Para Bonabeau (2009), se o objetivo é gerar soluções e/ou avaliações de produtos, as organizações devem considerar três tipos de abordagens gerais no uso da *Web 2.0*: solidariedade, agregação aditiva e auto-organização. Esses fatores têm prós e contras e muitas aplicações têm combinado essas diferentes abordagens.

A primeira abordagem - a solidariedade - consiste em coletar ideias para gerar ou avaliar uma solução. Dessa forma, as organizações podem explorar o conhecimento de pessoas ou grupos que não fazem parte do seu quadro funcional, podendo obter ajuda de atores externos. A solidariedade dos usuários da *Web 2.0* amplia o número de indivíduos que estão gerando e/ou avaliando soluções.

O desenvolvimento de *softwares open-source* (código aberto) é o exemplo clássico do poder da coletividade. Um número elevado de pessoas trabalhando com objetivos comuns num projeto pode solucionar e descobrir cada erro existente. A filosofia subjacente é que existem muitas pessoas de fora que podem ajudar, além disso, esses indivíduos não estão necessariamente onde se espera que eles estejam. Já a segunda abordagem envolve a agregação aditiva: as empresas podem coletar informações numa miríade de recursos e realizar uma inferência através delas. O processo pode ser usado para agregar dados de grupos tradicionais de decisão ou tais dados podem ser combinados com a solidariedade, incluindo informações de um conjunto mais amplo de pessoas. A auto-organização possui relação com mecanismos de interação entre membros de grupos que podem potencializar os resultados. Exemplos de uso construtivo de auto-organização que interagem na agregação de valor incluem a *Wikipedia* e a *Intellipedia* (a versão da CIA da *Wikipedia* para comunidades inteligentes) e o *Digg*. Tais aplicações permitem que as pessoas criem valor por adicionar ou apagar contribuições de outros participantes. Mas existe um risco: se os mecanismos de interação não forem projetados corretamente, o todo pode acabar por ser menor que a soma das partes (Bonabeau, 2009).

A *Amazon* pode também ser citada como modelo de sucesso. A empresa desenvolveu uma ciência sobre o envolvimento dos usuários. Ela os convida e os incentiva a participar de avaliações de seus produtos, usa a atividade deles para produzir melhores resultados de busca, incentiva a criação de resenhas dos livros. Enquanto uma busca da *BarnesandNoble.com* vai provavelmente levar a resultados encabeçados pelos produtos da própria companhia ou que sejam patrocinados, os da *Amazon* sempre apresentam em primeiro lugar os "mais populares", uma computação em tempo real baseada não apenas em vendas, mas em outros fatores que os conhecedores da *Amazon* chamam de "fluxo" em torno dos produtos. Contando com a participação do usuário em escala muitíssimo maior, não surpreende que as vendas da *Amazon* excedam à concorrência (O'Reilly, 2005).

Muitos dos problemas enfrentados pelas organizações podem ter solução através de atores externos, que estão muito além dos lugares tradicionais que os gestores podem encontrar fora das organizações. Na verdade, todo sucesso como a *Wikipédia* pode ser olhado

em sua superfície simples, que desmente os complexos mecanismos para o aproveitamento da inteligência coletiva.

Bonabeau (2009) acredita que existe um conjunto de motivações que levam as pessoas a participarem do compromisso coletivo. Incentivos monetários, prêmios e outras promoções podem ser efetivamente estimulantes para participar nas atividades, como, por exemplo, para predições de mercado. Todavia, muitas recompensas explícitas parecem ser importantes, isso significa que as pessoas não procuram ajudar apenas por dinheiro, mas ajudar outras pessoas e colaborar para o desenvolvimento do conhecimento humano parece ser importante para os internautas. Um exemplo de movimento de *open-source*, a Wikipédia e outros similares esforços têm mostrado a participação nas comunidades, o desejo de transferir o conhecimento ou compartilhar experiências e o senso de cidadania como sendo fortes motivadores.

## Considerações Finais

As novas tecnologias da *Web 2.0* possibilitaram a colaboração e o compartilhamento das informações ganhando cada vez mais importância para as organizações visto que por meio destas é possível trocar ideias e experiências, promovendo um ambiente de criatividade, inovação e aprendizagem organizacional.

A busca cada vez maior pelo diferencial competitivo e a nova lógica da economia forçam as organizações a gerar mais soluções e produtos cada vez mais inovadores. Como essa tarefa se torna cada vez mais difícil, nem sempre todas as respostas estão dentro das organizações. Os gerentes devem estar atentos ao poder de comunicação e captação de ideias que as redes sociais e os aplicativos da *Internet* podem proporcionar. A principal vantagem é que essas informações podem ser capturadas em tempo real e a um custo muito baixo.

Deste modo, é a competição que torna as empresas mais inteligentes, o que faz com que utilizem todos os recursos do trabalho cooperativo em rede, da engenha-

ria simultânea, da *Intranet* e da *Internet*, capitalizando a memória da empresa e suas competências, fazendo com que vendam e consumam cada vez mais informações e conhecimentos.

A atual sociedade em redes é tão complexa que nenhuma empresa pode desenvolver inovações necessárias para competir na indústria da tecnologia da informação ou em qualquer outro ramo. As empresas inteligentes devem saber explorar toda inovação gerada utilizando-se da inteligência coletiva e da colaboração para o desenvolvimento de soluções. Portanto, a *Web 2.0* envolve um número de pessoas muito maior do que as organizações jamais poderiam alocar internamente.

Por conseguinte, a *Web 2.0* pode auxiliar as organizações a se tornarem mais inteligentes não apenas compartilhando informações, mas também pela possibilidade da criação e da edição do conhecimento na *Web*. A *Web 2.0* ficou popular pela sua fácil utilização e pela sua atemporalidade, registrando a memória da humanidade. É também popular pela possibilidade de acessar as informações nas nuvens, ou seja, não exige que nenhum *software* seja instalado e pode ser acessada em qualquer dispositivo que tenha acesso à *Internet*. Outra facilidade é que as ferramentas da *Web 2.0* são compatíveis com diversos dispositivos móveis, como celulares, *iphone*, *ipad*, entre outros. Essa mobilidade aumenta ainda mais a frequência de uso e a oportunidade de colaboração.

Da mesma forma que a sociedade em redes, as comunidades virtuais e os fóruns de discussão servem de espaço para o compartilhamento do conhecimento entre os indivíduos. As organizações estão se conscientizando do potencial das novas tecnologias da *Web 2.0*, não somente como um novo espaço de divulgação de produtos, mas também como um espaço para a troca de conhecimento, desenvolvimento de serviços, bem como para a solução de problemas. Assim, as empresas podem ser tornar mais competitivas, explorando o potencial da inteligência coletiva e fazendo com que seus resultados sejam maiores do que a soma das partes.

## Referências

ANTOUN, H. As transformações na sociedade hiperconectada. In: ANTOUN, H. *Web 2.0: participação e vigilância na era da comunicação distribuída*. Rio de Janeiro: Mauad, 2008. p.11-27.

AQUINO, M.C. *O hipertexto como potencializador da memória coletiva: um estudo dos links na web 2.0*. 2007. Dissertação (Mestrado em Comunicação e Informação) - Programa de



Pós-Graduação em Comunicação e Informação, Universidade Federal do Rio Grandes do Sul, Porto Alegre, 2007.

BONABEAU, E. Decisions 2.0: the power of collective intelligence. *MIT Sloan Management Review*, v.50, n.2, p.44-53, 2009.

CASTELLS, M. *A sociedade em rede*. São Paulo: Paz e Terra, 1999.

CAVALCANTI, M.; NEPOMUCENO, C. *O conhecimento em rede*. Rio de Janeiro: Elsevier, 2007.

GALDO, A.M.R. *Web 2.0 e colaboração científica*: análise do uso científico-acadêmico por docentes de pós-graduação *stricto sensu* em ciência da informação no Brasil. 2010. Mestrado (Ciência da Informação) - Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2010.

JOHNSON, S. *Emergência*: a dinâmica da rede em formigas, cérebros, cidades e softwares. Rio de Janeiro: Zahar, 2003.

LÉVY, P. *O que é o virtual?*. São Paulo: Editora 34, 1996.

LÉVY, P. Pela ciberdemocracia. In: MORAES, D. (Org.). *Por uma outra comunicação*: mídia, mundialização cultural e poder. Niterói: Record, 2005. p.367-384.

McAFEE, A. Enterprise 2.0: the dawn of emergent collaboration. *MIT Sloan Management Review*, v.47, n.3, p.9-21, 2006.

O'REILLY, T. *What is web 2.0*: design patterns and business models for the next generation of software. 2005. Available from: <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-Web-20.html>>. Cited: 5 May 2008.

PRIMO, A. *Interação mediada por computador*. Porto Alegre: Sulina, 2008.

SCHNECKENBERG, D. Web 2.0 and the empowerment of the knowledge worker. *Journal of Knowledge Management*, v.13, n.6, p.509-520, 2009.

TAPSCOTT, D.; WILLIAMS, A.D. *Wikinomics*: como a colaboração em massa pode mudar o seu negócio. Rio de Janeiro: Nova Fronteira, 2007.

# Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos<sup>1</sup>

## *Information retrieval systems: a discussion based on enunciative parameters*

Vera Maria Araujo Pigozzi de ARAUJO<sup>2</sup>

### Resumo

A temática deste artigo são os Sistemas de Recuperação da Informação na interface com a teoria da enunciação. O objetivo é discutir a relação que se estabelece entre os sistemas de informação e os usuários. Com o propósito de identificar como se dá essa relação, examinam-se os modelos de comunicação documentária vistos em revisão de literatura, e identificam-se os parâmetros sob os quais são construídos. Por questionar a validade desses modelos, analisam-se as características dessa relação. Identifica-se, no aparelho formal da enunciação formulado por Benveniste, a possibilidade de se contemplarem as peculiaridades que a caracterizam, considerando os seguintes parâmetros: categorias de pessoa (eu-tu), espaço (aqui) e tempo (agora). Conclui com uma proposta alternativa aos modelos apresentados, considerando bibliotecários, usuários e sistemas como interlocutores, e sugerindo ser uma relação de interação e não de comunicação. Esta proposta é fundamentada nas noções de subjetividade, intersubjetividade, categorias de pessoa, de espaço e de tempo.

**Palavras-chave:** Aparelho formal da enunciação. Comunicação documentária. Sistemas de recuperação da informação. Teoria da enunciação.

### Abstract

*The theme of this article is the interface of Information Retrieval Systems with the theory of enunciation. The aim of this study was to discuss the relationship established between information systems and users. In order to identify how this relationship works, it examines the documentary communication models seen in a literature review, and identifies the parameters they are built on. By questioning the validity of these models, it analyzes the characteristics of this relationship. In the formal apparatus of enunciation proposed by Benveniste, it identifies an opportunity to contemplate the peculiarities that characterize the relationship, considering the following parameters: the categories of person, space and time. It concludes with an alternative proposal to the models presented, considering librarians, users and systems as interlocutors, and suggests that it is a relationship of interaction and not one of communication. This proposal is based on notions of subjectivity, intersubjectivity, categories of person, space and time.*

**Keywords:** Formal apparatus of enunciation. Documentary communication. Information retrieval systems. Theory of enunciation.

### Introdução

Este artigo trata dos Sistemas de Recuperação da Informação (SRI), e examina-se nele a comunicação documentária, a partir dos referenciais de Araujo (1995), Lima (1998), Lara (2002), Ortega e Lara (2010) e

Biscalchin e Boccato (2011). Por questionar a validade desses modelos, que identificam sistema e usuário como emissor e receptor desse processo de comunicação, examinam-se os SRI, apresentando sucinta-

<sup>1</sup> Artigo elaborado a partir da pesquisa de doutorado em curso, desenvolvido por V.M.A.P. ARAUJO.

<sup>2</sup> Doutoranda, Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Letras. Av. Bento Gonçalves, 9500, 90040-060, Porto Alegre, RS, Brasil. E-mail: <verapigozzi@hotmail.com>.

Recebido em 29/10/2011, reapresentado em 21/5/2012 e aceito para publicação em 5/6/2012.

mente sua conceituação e função sob a visão de Cendón (2005) e Souza (2006).

No exame da relação dos SRI com os usuários, identificam-se diversas peculiaridades que não estão contempladas nos modelos apresentados. Na busca de um referencial mais satisfatório, percebe-se a possibilidade de desenvolver uma proposta alternativa a esses modelos. É no aparelho formal da enunciação que essa possibilidade é visualizada.

Para discutir a relação que se estabelece entre os SRI e os usuários, o artigo apresenta uma breve revisão de literatura sobre comunicação documentária e sistemas de recuperação da informação. A partir da análise dos modelos de comunicação e dos sistemas na sua relação com os usuários, oferece-se, ainda, uma proposta fundamentada no aparelho formal da enunciação formulado por Benveniste.

## Comunicação documentária

Comunicação documentária, conforme Lima (1998), é o processo de comunicação entre o acervo de um sistema de informação e o seu usuário. Ela é alcançada quando um sistema codifica o conteúdo informacional dos documentos, objetivando sua recuperação pelo usuário. A linguagem documentária é o código utilizado nesse processo.

Os modelos de comunicação documentária têm origem na teoria da informação proposta por Shannon e Weaver (1949), que entende que o processo de comunicação se dá a partir de um sistema (fonte), que influencia outro sistema (receptor), transmitindo signos por intermédio de um canal. O canal é o meio que permite a

comunicação e que viabiliza a transferência da informação. Essa teoria, ainda que tenha recebido alterações no decorrer do tempo, é a que usualmente fundamenta o processo de comunicação documentária.

Lara (2002) entende que as representações documentárias são utilizadas para transmitir informação. No processo de transferência da informação, é possível identificar: 1) o emissor (instituição que formula o sistema documentário); 2) o canal (veículo pelo qual são veiculadas as mensagens documentárias e que seriam os catálogos eletrônicos<sup>3</sup>); 3) o código intermediário (a linguagem documentária usada na indexação e na recuperação da informação e que deve funcionar, simultaneamente, como um sistema de comunicação e de significação); 4) o destinatário (usuário final do sistema). Pelo modelo de comunicação documentária assim proposto, a relação entre SRI e usuários é caracterizada de forma bem simplificada (Quadro 1).

Conforme Ortega e Lara (2010, p.9):

O usuário é o ator que dispara uma situação de comunicação, pois é o ato da recepção que determina o estabelecimento ou não de uma comunicação. [...] para que a comunicação documentária ocorra é necessário que a informação documentária - produto que é transmitido nesse processo - seja portadora de um significado tal que o usuário possa ser capaz de processá-la mentalmente, a fim de obter a informação desejada. A comunicação documentária é então um processo que exige, de um lado, a informação documentária e, de outro, o usuário que se apropria desta informação. Os sistemas documentários lidam com atos de comunicação materialmente explícitos, isto é, o documento, sua representação, a pergunta do usuário, a ação comunicativa do profissional de informação etc.

**Quadro 1.** Modelo de comunicação documentária.

Elementos da comunicação na teoria da informação	Representação dos elementos da comunicação na comunicação documentária
Emissor	Instituição que formula o sistema documentário
Canal	Catálogo eletrônico
Código intermediário	Linguagem documentária
Destinatário	Usuário final do Sistema

Fonte: Elaborado pela autora.

<sup>3</sup> Os catálogos eletrônicos são reconhecidos como SRI por integrarem diferentes atividades: seleção, desenvolvimento de coleções, processamento técnico, busca e recuperação da informação.

Ainda, para que esse processo de comunicação ocorra, é necessário que: 1) a informação documentária - produto que é transmitido nesse processo - seja portadora de um significado, e 2) o usuário - ator que dispara uma situação de comunicação - seja capaz de processar a informação documentária para obter a informação desejada.

Tendo como foco a insatisfação do usuário na relação com os SRI, aborda-se, a seguir, a relação que ocorre entre ambos. Dentre os autores que questionam e procuram respostas para essa insatisfação presente nos processos de busca e recuperação da informação, Araujo (1995, p.1), cuja tese discute essa questão, faz a seguinte afirmação:

Por que vêm falhando os sistemas de informação? E não há dúvidas de que vêm falhando: uma rápida análise da literatura da área evidencia uma série de estudos, projetos etc., visando corrigir, ao menos parcialmente, suas falhas e proporcionar 'maior satisfação aos usuários'. Observa-se, inclusive, que tais estudos incluem 'modernas abordagens' ao planejamento, gestão e, principalmente, *marketing* de seus serviços e produtos, buscando a inserção dos sistemas de informação em uma sociedade pós-industrial, pós-moderna, conhecimento-intensiva.

Em resposta à pergunta *por que vêm falhando os sistemas de informação* e à consequente insatisfação do usuário, Biscalchin e Boccato (2011), em um estudo sobre o uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias, apontaram três aspectos principais: 1) a incompatibilidade existente entre a linguagem de busca e a adotada pelo sistema; 2) a ausência de remissivas; e 3) as inconsistências sintático-semânticas verificadas na linguagem. A pesquisa revelou que a grande maioria dos usuários desconhece a linguagem usada pelo sistema. Acrescenta-se a isso o desconhecimento dos recursos de pesquisa que o catálogo eletrônico oferece. Lima (1998) observa, ainda, que a presença de várias linguagens documentárias em um único SRI prejudica a transferência da informação e, desse modo, a comunicação entre o sistema e o usuário.

Em busca de uma maior satisfação do usuário, são feitas algumas considerações sobre o fato de o bibliotecário, como profissional da informação, ter sob sua responsabilidade o gerenciamento dos SRI. Ele precisa estar atento tanto ao modo de expressar da maneira mais

fiel possível a representação do conteúdo temático de um documento, como ao conhecimento das estratégias de busca oferecidas pelo sistema.

Neste artigo, o aspecto a ser considerado é o da representação fiel do conteúdo temático do documento. Para Lima (1998), a compatibilização das linguagens documentárias com a terminologia oferece condições propícias a uma comunicação documentária mais eficaz.

## Sistemas de recuperação da informação

Os sistemas de informação tratam da representação, do armazenamento, da organização e da localização dos itens de informação. Para organizar e comunicar a informação, eles utilizam linguagens documentárias, que estabelecem um importante elo entre os SRI e os usuários. Conforme Souza (2006), as funções de um SRI são as seguintes: 1) representação das informações contidas nos documentos e expressas pelos processos de indexação e descrição dos documentos; 2) armazenamento e gestão física e ou lógica desses documentos e de suas representações; e 3) recuperação das informações ali contidas e dos próprios documentos armazenados no sistema.

Das vantagens que Cendón (2005, p.62) apresenta a favor dos sistemas no formato eletrônico, há a considerar:

[...] as facilidades que os SRI automatizados oferecem para busca de informação [...] maior número de pontos de acesso que os SRI não-automatizados [...]. Além disso, permitem realizar pesquisas mais complexas, em que vários conceitos necessitam ser relacionados pois pode-se combinar grande número de termos de busca [...], de maneiras que não seriam possíveis nos SRI impressos.

Com a intenção de apresentar uma proposta alternativa que caracterize a relação estabelecida entre SRI e usuários, desenvolve-se a seguir uma reflexão sobre as características apresentadas por esses sistemas. Pela identificação de tais características, verificou-se que os modelos de comunicação até então apresentados não configuram essa relação.

Essa reflexão parte do entendimento de que a função principal de um SRI é dispor de informações contidas nos documentos indexados, a partir de uma descrição sintética, objetiva e representativa de seu con-

teúdo formal e temático. Um SRI integra diferentes processos: seleção, aquisição, indexação e, ainda, busca e recuperação das informações.

Na análise da relação que se estabelece entre os SRI e os usuários, observa-se que ela apresenta peculiaridades que não são explicadas nos modelos de comunicação documentária vistos em revisão de literatura. O que se verifica é que esses sistemas apresentam características bem diferenciadas e muito mais complexas do que aquelas que fundamentam esses modelos, como descrito a seguir:

1) São sistemas em que os locutores (bibliotecário, usuário e SRI) são de natureza heterogênea e não estão presentes quando o alocutário busca a informação;

2) Os códigos utilizados e representados pelas linguagens documentárias são reveladores de diferentes sistemas semânticos; uma mesma Unidade de Informação (cabeçalho de assunto ou descritor) não cobre a mesma noção de um vocabulário para outro, e a gramática das linguagens documentárias é diferente de um sistema para o outro;

3) O processo de indexação faz uso de mais de uma linguagem: linguagem especializada (linguagem do documento) e vocabulário controlado (lista de cabeçalho de assuntos ou tesouro);

4) O processo de busca e recuperação da informação oferece pesquisas por palavras, em catálogos eletrônicos diferenciados quanto à estrutura e estratégias de pesquisa;

5) Os processos que viabilizam a recuperação da informação e a produção do conhecimento se manifestam em lugares e momentos múltiplos;

6) No processo de indexação, bibliotecários e usuários se posicionam como interlocutores, mas no processo de busca e recuperação da informação, usuários e sistemas são os interlocutores;

7) As marcas da subjetividade e intersubjetividade estão presentes e são relevantes nos processos de indexação e de busca e de recuperação da informação;

8) O documento processado é a fonte da informação em ambos os processos.

Visando ao uso da informação e à produção do conhecimento, propõe-se uma aproximação entre os

processos anteriormente mencionados. Entende-se que a relação que ocorre entre SRI e usuários não se limita ao processo de busca e recuperação da informação. A construção de um sistema que se propõe a estabelecer uma relação mais satisfatória com o usuário precisa levar em consideração aquela que ocorre entre o bibliotecário e o usuário durante o processo de indexação. Mas ainda fica a pergunta: que tipo de relação é esta? Trata-se mesmo de uma relação de comunicação? Busca-se esta resposta na teoria de Benveniste.

### **Teoria da enunciação de Benveniste**

Visando compreender o funcionamento das linguagens documentárias, e a relação SRI e usuário, recorre-se à teoria da enunciação. É sob a ótica da semântica, e não da semiótica, que se desenvolve essa reflexão. O semântico toma a seu encargo o conjunto de referentes, e é produzido pelo discurso e se identifica com a enunciação.

Do texto *A forma e o sentido na linguagem* (Benveniste, 1989), apresenta-se a ideia de que a forma é o arranjo formal dos elementos linguísticos ao nível linguístico relevante, enquanto o sentido é o conjunto de procedimentos de comunicação identicamente compreendidos por um grupo de locutores. Para o linguista, a significação é da natureza da linguagem.

Com a noção de semântica, Benveniste entende a língua em emprego e em ação. A frase é a expressão semântica por excelência; com ela nos comunicamos e nos deslocamos para a produção do discurso, para aquilo que o locutor quer dizer. O sentido da frase é sua ideia, que é sempre única, e o das palavras, seu emprego. O sentido da frase se constitui no momento da enunciação, ou seja, quando o locutor exprime sua ideia. Sendo a palavra a unidade semântica, ela mesma guarda suas referências e significações próprias. Entretanto, quando há o agenciamento de palavras para a construção da frase, a referência passa a ser o próprio ato de enunciação; dele decorre o sentido da frase. A noção de referência é essencial no processo de construção do sentido. Sobre as noções de sentido e de referência, o linguista propõe que se estabeleça uma distinção entre elas.

Se o 'sentido' da frase é a ideia que ela exprime, a 'referência' da frase é o estado de coisas que a

provoca, a situação de discurso ou de fato a que ela se reporta e que nós não podemos jamais prever ou fixar [...]. A frase é então cada vez um acontecimento diferente; ela não existe senão no instante em que é proferida e se apaga neste instante [...] (Benveniste, 1989, p.231).

Subjetividade é um conceito fundamental na sua teoria; ela procede da relação 'eu e tu', ou seja, 'pessoa subjetiva e pessoa não-subjetiva'. Mas as noções de intersubjetividade, locutor, referência e categorias de pessoa, espaço e tempo também são importantes. Quanto ao tempo da enunciação, segundo o linguista, ele é sempre presente, uma vez que corresponde ao momento em que 'eu' se diz 'eu'.

Em *A natureza dos pronomes*, Benveniste (1988) aborda a categoria de pessoa. A noção de pessoa está presente apenas nas duas primeiras pessoas, estando ausente na terceira pessoa. O 'ele' é diferente de 'eu e tu' pela sua função e pela sua natureza, e se caracteriza por não remeter a ele mesmo, mas a uma situação objetiva.

No texto *Da subjetividade na linguagem* (Benveniste, 1988), a subjetividade é entendida como a capacidade do locutor para se propor como sujeito. Ao inserir a noção de referência no contexto da comunicação, o linguista questiona a função da linguagem como ferramenta da comunicação. Fundamentado no argumento de que opor o homem à linguagem é colocá-lo em oposição a sua própria natureza, ele nega o aspecto instrumental da linguagem e afirma que a língua não é um instrumento de comunicação, pois isso vai contra o princípio segundo o qual a linguagem está na natureza do homem.

Dentre as teorias revisadas a fim de se formular uma proposta alternativa aos modelos de comunicação documentária, identificados em revisão de literatura, foi na teoria da enunciação de Benveniste, em especial no aparelho formal da enunciação, que foram identificados vários elementos capazes de contemplar as peculiaridades da relação que ocorre entre SRI e usuários.

A opção por essa teoria justifica-se pelas seguintes razões: 1) ela entende que é pela linguagem que o homem se inter-relaciona, recebe e transmite uma mensagem; 2) olha para o estudo da língua do ponto de vista do sentido; 3) permite considerar peculiaridades próprias do tipo de relação que se estabelece entre os SRI e os usuários, e que não são visíveis nos modelos de comunicação apresen-

tados na literatura documentária; 4) caracteriza-se por ser construída a partir de noções como linguagem, subjetividade, intersubjetividade, referência e contexto, que são características presentes no referencial teórico que fundamenta a proposta apresentada.

No tópico a seguir, discute-se o aparelho formal da enunciação com a intenção de propor um novo referencial para a relação que se estabelece entre os SRI e os usuários.

### **Aparelho formal da enunciação e a relação SRI e usuários**

O aparelho formal da enunciação proposto por Benveniste pode ser entendido como um dispositivo que permite que as línguas possam ser enunciadas; ele nada mais é do que a marcação da subjetividade na estrutura da língua. Segundo Benveniste (1989, p.82), "A enunciação é esse colocar em funcionamento a língua por um ato individual de utilização", pois, conforme Flores e Teixeira (2005, p.37), "[...] cada vez que o locutor se apropria do aparelho formal da enunciação - e por ele se apropria da língua toda - produz um uso novo e como tal irrepitível. Essa irrepitibilidade deve-se ao fato de que jamais tempo, espaço e pessoa - categorias fundamentais em enunciação - podem ser perenizados no uso da língua".

Pela compreensão das noções advindas do aparelho formal da enunciação - subjetividade, intersubjetividade, referência e contexto - e pelas categorias fundamentais em enunciação - pessoa (eu e tu), espaço (aqui) e tempo (agora) -, tornou-se possível perceber as peculiaridades e a complexidade da relação que se estabelece entre SRI e usuários, assim como elaborar e construir outro referencial sobre essa relação, que contempla os diversos aspectos que a caracterizam. A partir da configuração registrada (Quadro 2), são feitas algumas considerações.

Ao se inserirem as noções de subjetividade e de intersubjetividade, reconhece-se uma situação de interação entre os diferentes interlocutores envolvidos. A noção de referência aparece como definidora de sentido no nível semântico, sendo a referência ao mundo parte integrante do ato de enunciação: aqui entra a noção de contexto. É em relação ao contexto de situação que o sentido de uma palavra é obtido; fora do contexto, nem

**Quadro 2.** Aparelho formal da enunciação e a relação SRI e usuários.

Aparelho formal da enunciação	Relações SRI e usuários	
	<i>Primeiro momento</i>	<i>Segundo momento</i>
Categoria de pessoa subjetiva (eu)	Bibliotecário (eu)	Usuário (eu)
Categoria de pessoa não-subjetiva (tu)	Usuário (tu)	SRI (tu)
Categoria de não-pessoa (ele)	⇒ Linguagem especializada (ele)	Linguagem especializada (ele)
Categoria de espaço (aqui)	Vocabulário controlado (aqui)	Catálogo eletrônico (aqui)
Categoria de tempo (agora)	Processo de indexação (agora)	Processo de busca e recuperação da informação (agora)

Fonte: Elaborado pela autora.

sempre é possível compreender o sentido que resulta da junção das palavras.

Compreendendo que não se trata de uma situação de comunicação, mas de interação, essa proposta introduz a noção de interlocutores, identificando como interlocutores não apenas o sistema e o usuário, mas também o bibliotecário. São identificadas duas situações de interação, que ocorrem em espaço e tempo diferenciados: 1) uma entre bibliotecários e usuários, interagindo durante o processo de indexação, e 2) outra entre usuários e catálogos eletrônicos, também reconhecidos como SRI, na qual a ação de interlocução acontece durante o processo de busca e recuperação da informação.

Ressalta-se, ainda, que os termos e expressões que identificam o conteúdo informacional de um documento são localizados nos vocabulários controlados e nos catálogos eletrônicos. Os vocabulários controlados são as ferramentas com as quais os bibliotecários reconhecem a terminologia que identifica o conteúdo informacional de um documento. Os catálogos eletrônicos são as ferramentas utilizadas pelos usuários para recuperar a informação que desejam. Nas duas situações, caracterizadas pelos processos de indexação e de busca e recuperação da informação, a linguagem especializada é a fonte de informação.

Assim, entende-se que a relação estabelecida entre os SRI e os usuários não se limita ao processo de busca e recuperação da informação. É preciso que seja considerada também a relação de interação que ocorre durante o processo de indexação entre bibliotecários e usuários. A partir desses pressupostos, conclui-se que essa relação não pode ser vista do modo como tem sido apresentada em modelos de comunicação documentária,

ou seja, a partir de um emissor, canal, código intermediário e destinatário.

## Considerações Finais

Finaliza-se este artigo afirmando que a teoria da enunciação apresenta vantagens teóricas e metodológicas importantes aos estudos desenvolvidos em linguagens documentárias. Esse referencial conduz o bibliotecário a novas reflexões e posicionamentos diante das atividades documentárias, seja na execução das atividades de indexação, seja na elaboração de vocabulários controlados, seja no estabelecimento de uma política de indexação para os sistemas.

O questionamento sobre a relação que se estabelece entre os SRI e os usuários é o seguinte: é essa realmente uma situação de comunicação? Essa discussão, feita a partir de parâmetros enunciativos, possivelmente levará a uma mudança de paradigma no âmbito documentário. Por ora, entende-se que, entre usuários e SRI, estabelece-se uma ação de interação, mas não de comunicação documentária, e como tal não pode ser descrita pelos elementos que usualmente a representam: emissor, canal, código intermediário e usuário.

A insatisfação do usuário na consulta aos catálogos eletrônicos leva à conclusão de que os SRI não estão cumprindo sua função de transferência da informação. Hoje, uma das maiores dificuldades enfrentadas pelo usuário de um sistema não é exatamente a recuperação de documentos, mas a recuperação de informações relevantes para seus propósitos de pesquisa, nos documentos recuperados. Essa realidade leva à seguinte pergunta: o que se espera de um SRI? Seu verdadeiro papel será o da transferência da informação?

## Referências

- ARAUJO, V.M.R.H. Sistemas de informação: nova abordagem teórico conceitual. *Ciência da Informação*, v.24, n.1, p.54-76, 1995.
- BENVENISTE, E. *Problemas de linguística geral I*. 2.ed. Campinas: Pontes, 1988. p.277-293.
- BENVENISTE, E. *Problemas de linguística geral II*. Campinas: Pontes, 1989.
- BISCALCHIN, R.; BOCCATO, V.R.C. Estudo do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: uma abordagem qualitativa-sociocognitiva pela perspectiva do usuário. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 24., 2011, Maceió. *Anais eletrônicos...* Maceió: CBBD, 2011. Disponível em: <<http://febab.org.br/congressos/index.php/cbbd/xxiv/paper/view/39/508>>. Acesso em: 23 mar. 2012.
- CENDÓN, B.V. Sistemas e redes de informação. In: OLIVEIRA, M. (Coord.). *Ciência da informação e biblioteconomia: novos conteúdos e espaços de atuação*. Belo Horizonte: UFMG, 2005. p.45-75.
- FLORES, V.N.; TEIXEIRA, M. *Introdução à lingüística da enunciação*. São Paulo: Contexto, 2005.
- LARA, M.L.G. O processo de construção da informação documentária e o processo de conhecimento. *Perspectivas em Ciência da Informação*, v.7, n.2, p.127-139, 2002.
- LIMA, V.M.A. *Terminologia, comunicação e representação documentária*. 1998. Dissertação (Mestrado em Ciências da Informação) - Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1998. Disponível em: <<http://www.teses.usp.br/teses>>. Acesso em: 20 mar. 2012.
- ORTEGA, C.D.; LARA, M.L.G. A noção de estrutura e os registros de informação dos sistemas documentários. *TransInformação*, v.22, n.1, p.7-17, 2010. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewissue.php?id=24#Artigos>>. Acesso em: 19 maio 2011.
- SHANNON, C.E.; WEAVER, W. *The mathematical theory of communication*. Illinois: The University of Illinois, 1949.
- SOUZA, R.R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em Ciência da Informação*, v.11, n.2, p.161-173, 2006.





# Instruções aos Autores

## Objetivos e política editorial

*Transinformação* é uma revista especializada, com periodicidade quadrimestral, aberta a contribuições da comunidade científica nacional e internacional, distribuída no Brasil e no exterior. Fundada em 1989, é classificada na lista Qualis como A1, publica artigos que contribuem para o estudo e o desenvolvimento científico da Ciência da Informação e Biblioteconomia em suas diversas sub-áreas e interfaces.

Os autores são responsáveis pelas informações contidas nos trabalhos, bem como pela devida permissão ao uso de figuras ou tabelas publicadas em outras fontes.

Serão aceitas contribuições, cujas características:

- a) apresentem enfoque inovador em relação a temas já tratados;
- b) utilizem procedimentos metodológicos inovadores;
- c) no caso de revisão de tema, que seja contribuição nova ao campo;
- d) utilizem metodologia consistente;
- e) apresentem conclusões que decorram de argumentação lógica;
- f) apresentem fontes bibliográficas pertinentes ao tema tratado.

## Tipos de artigos aceitos

- *Originais*: contribuição destinada a divulgar resultados de pesquisa inédita (limite máximo de 12 páginas, preparados em espaço entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Revisão*: síntese crítica de tema de interesse da área, mediante análise e interpretação de bibliografia pertinente (limite máximo de 12 páginas, preparados em entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Ensaio*: reflexão sobre tema que gere questionamentos e permita elaborar hipóteses para futuras pesquisas (apenas sob convite) (limite máximo de 10 a 15 páginas, preparados em entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Debate*: trabalho teórico que se faz acompanhar de cartas críticas assinadas por autores de diferentes instituições (três a cinco especialistas, convidados pelo Editor), seguidas de réplicas do autor do artigo principal. O interessado em submeter um manuscrito para esta seção deve consultar previamente o Editor (limite máximo de vinte páginas, preparados em entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Resenha*: análise de um trabalho frente a um quadro de referência teórica da área proposta pelo autor, o qual deve avaliar e criticar a obra expondo seu ponto de vista pessoal (limite máximo de três páginas, preparados em entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Comunicação*: informações sobre pesquisa em andamento: informações sucintas sobre projetos de pesquisa, dissertações e teses

em andamento (limite máximo de três páginas, preparados em entrelinhas 1,5, com fonte *Arial* tamanho 11 e em folha formato A4).

- *Tradução (reprodução)*: artigos traduzidos, autorizadas pelo detentor dos direitos de reprodução.

## Procedimentos Editoriais

### Processo de julgamento dos manuscritos

São aceitos originais inéditos para serem submetidos à aprovação de avaliadores que sejam especialistas reconhecidos nos temas tratados. Os trabalhos são arbitrados por pelo menos dois revisores pertencentes ao quadro de colaboradores da Revista, em procedimento sigiloso quanto à identidade do(s) autor(es) e revisores.

Opcionalmente, os autores podem indicar três possíveis revisores para o seu manuscrito.

*Manuscritos aceitos*: manuscritos aceitos poderão retornar aos autores para aprovação de eventuais alterações, no processo de editoração e normalização, de acordo com o estilo da Revista.

Manuscritos recusados, mas com a possibilidade de reformulação, poderão retornar como novo trabalho, iniciando outro processo de julgamento.

### Conflito de interesse

No caso de identificação de conflito de interesse da parte dos revisores, o Comitê Editorial encaminhará o manuscrito a outro revisor *ad hoc*.

### Submissão de trabalhos via endereço

<<http://revistas.puc-campinas.edu.br/transinfo>>

São aceitos trabalhos inéditos acompanhados de carta assinada por todos os autores, com descrição do tipo de trabalho, declaração de que o trabalho está sendo submetido apenas à revista *Transinformação* e cessão de direitos autorais.

A carta deve indicar o nome, endereço, números de telefone e *e-mails* dos autores e indicação do autor para o qual a correspondência deve ser enviada.

Caso sejam utilizadas figuras ou tabelas publicadas em outras fontes, deve-se anexar documento que ateste a permissão para seu uso.

### Versão reformulada

Sublinhar as alterações no texto do artigo ou indicá-las com fonte colorida (cor azul), encaminhando-o juntamente com uma carta ao editor, reiterando o interesse em publicar nesta Revista e informando quais alterações foram processadas no manuscrito. Se houver discordância quanto às recomendações dos revisores, o(s) autor(es) deverão apresentar os argumentos que justificam sua posição. O título e o código do manuscrito deverão ser especificados.

*Provas*: serão enviadas provas tipográficas aos autores para a correção de erros de impressão. As provas devem retornar ao Núcleo

de Editoração na data estipulada. Outras mudanças no manuscrito original não serão aceitas nesta fase.

## Os manuscritos deverão apresentar

### Página de título deve conter

a) título completo - deve ser conciso, evitando excesso de palavras, como "avaliação do..."; "considerações acerca de..." "estudo exploratório...".

b) *short title* com até quarenta caracteres (incluindo espaços), em português (ou espanhol) e inglês.

c) nome de todos os autores por extenso, indicando a afiliação institucional de cada um. Será aceita uma única titulação e afiliação por autor. Os autores deverão, portanto, escolher, entre suas titulações e afiliações institucionais.

**Observação:** não havendo vínculo institucional, informar a atividade profissional, cidade e estado.

d) todos os dados da titulação e da afiliação deverão ser apresentados por extenso, sem siglas.

e) indicação dos endereços completos de todas as universidades às quais estão vinculados os autores.

f) indicação de endereço para correspondência com o autor para a tramitação do original, incluindo fax, telefone e endereço eletrônico.

**Observação:** esta deverá ser a única parte do texto com a identificação dos autores.

**Resumo:** todos os artigos submetidos em português, espanhol ou francês deverão ter resumo no idioma original e em inglês, com um mínimo de 150 palavras e máximo de 250 palavras.

Os artigos submetidos em inglês deverão vir acompanhados de resumo, título e palavras-chave em português, além dos respectivos em inglês.

O resumo deve conter o objetivo do trabalho, os procedimentos metodológicos e as conclusões.

**Texto:** com exceção dos manuscritos apresentados como Revisão, Comunicação e Ensaio, os trabalhos deverão seguir a estrutura formal para trabalhos científicos:

**Introdução:** deve conter revisão da literatura atualizada e pertinente ao tema, adequada à apresentação do problema, e que destaque sua relevância. Não deve ser extensa, a não ser em manuscritos submetidos como Artigo de Revisão.

**Métodos:** deve conter descrição clara e sucinta do método empregado, acompanhada da correspondente citação bibliográfica, incluindo: procedimentos adotados; universo e amostra; instrumentos de medida e, se aplicável, método de validação; tratamento estatístico.

Em relação à análise estatística, os autores devem demonstrar que os procedimentos utilizados foram não somente apropriados para testar as hipóteses do estudo, mas também corretamente interpretados.

**Resultados:** sempre que possível, os resultados devem ser apresentados em tabelas ou figuras, elaboradas de forma a serem auto-explicativas e com análise estatística. Evitar repetir dados no texto.

Tabelas, quadros e figuras devem ser limitados a cinco no conjunto e numerados consecutiva e independentemente com algarismos arábicos, de acordo com a ordem de menção dos dados, e devem vir em folhas individuais e separadas, com indicação de sua localização no texto. **É imprescindível a informação do local e ano do estudo.** A cada um se deve atribuir um título breve. Os quadros e tabelas terão as bordas laterais abertas.

O(s) autor(es) se responsabiliza(m) pela qualidade das figuras (desenhos, ilustrações, tabelas, quadros e gráficos), que deverão ser elaboradas em tamanhos de uma ou duas colunas (7 e 15cm, respectivamente); **não é permitido o formato paisagem.** Figuras digitalizadas deverão ter extensão jpeg e resolução mínima de 400 dpi.

Gráficos e desenhos deverão ser gerados em programas de desenho vetorial (*Microsoft Excel, CorelDraw, Adobe Illustrator* etc.), acompanhados de seus parâmetros quantitativos, em forma de tabela e com nome de todas as variáveis.

A publicação de imagens coloridas, após avaliação da viabilidade técnica de sua reprodução, será custeada pelo(s) autor(es). Em caso de manifestação de interesse por parte do(s) autor(es), a revista Transinformação providenciará um orçamento dos custos envolvidos, que poderão variar de acordo com o número de imagens, sua distribuição em páginas diferentes e a publicação concomitante de material em cores por parte de outro(s) autor(es).

**Discussão:** deve explorar, adequada e objetivamente, os resultados, discutidos à luz de outras observações já registradas na literatura.

**Conclusão:** apresentar as conclusões relevantes, considerando os objetivos do trabalho, e indicar formas de continuidade do estudo. **Não serão aceitas citações bibliográficas nesta seção.**

**Abreviaturas e siglas:** deverão ser utilizadas de forma padronizada, restringindo-se apenas àquelas usadas convencionalmente ou sancionadas pelo uso, acompanhadas do significado, por extenso, quando da primeira citação no texto. Não devem ser usadas no título e no resumo.

**Agradecimentos:** podem ser registrados, em parágrafo não superior a três linhas, dirigidos a instituições ou indivíduos que prestaram efetiva colaboração para o trabalho.

**Anexos:** deverão ser incluídos apenas quando imprescindíveis à compreensão do texto. Caberá aos editores julgar a necessidade de sua publicação.

### Citações no texto

Citações bibliográficas no texto: devem constar da lista de referências.

**Não serão aceitas** citações/referências de **monografias** de conclusão de curso de graduação e de **textos não publicados** (aulas, entre outros).

Se um trabalho não publicado, de autoria de um dos autores do manuscrito, for citado (ou seja, um artigo *in press*), será necessário incluir a carta de aceitação da revista que publicará o referido artigo.

Se dados não publicados, obtidos por outros pesquisadores, forem citados pelo manuscrito, será necessário incluir uma carta de autorização, do uso dos mesmos por seus autores.

## Casos específicos

1) Citações literais de até três linhas: entre aspas, sem destaque em itálico e, em seguida, entre parênteses (Sobrenome do autor, data, página, sem espaço entre o ponto e o número). Ponto final depois dos parênteses.

2) Citações literais de mais de três linhas: em parágrafo destacado do texto, com 4cm de recuo à esquerda, em espaço simples, fonte menor que a utilizada no texto, sem aspas, sem itálico, terminando na margem direita do texto. Em seguida, entre parênteses: (Sobrenome do autor, data, página).

3) Vários autores citados em sequência: utilizar ordem cronológica de data de publicação dos documentos, separados por ponto e vírgula: (Crespo, 2005; Costa; Ramalho, 2008; Moresi *et al.*, 2010).

4) Textos com dois autores: Crippa e Bisoffi (2010) (no corpo do texto); (Crippa; Bisoffi, 2010) (dentro dos parênteses).

5) Textos com três ou mais autores: (Griselda *et al.*, 2009) (dentro dos parênteses) e Griselda *et al.* (2009) (fora dos parênteses).

6) Citações do mesmo autor publicados no mesmo ano: acrescenta-se letra minúscula após a data, sem espaçamento. Exemplo: (Morin, 2000a, 2000b).

## Referências

As referências são baseadas na NBR-6023/2002 e recomenda-se limitar a 30 referências para artigos, exceto no caso de artigos de revisão, que poderão apresentar em torno 50. Elas deverão ser ordenadas alfabeticamente pelo sobrenome do primeiro autor.

## Casos específicos

1) Os títulos dos periódicos devem ser referidos por extenso.

2) Referências com autores e datas coincidentes usa-se o título da obra ou artigo para ordenação e acrescenta-se letra minúscula após a data, sem espaçamento.

3) Referências com três ou mais autores, indica-se apenas o primeiro, acrescentando-se a expressão *et al.*

**A exatidão e a adequação das referências a trabalhos que tenham sido consultados e mencionados no texto do artigo são de responsabilidade do autor.**

Para outros exemplos recomendamos consultar as normas da ABNT-NBR-6023/2002.

## Exemplos

### Artigo com um autor

OLIVEIRA, A. Direito à memória das comunidades tradicionais: organização de acervo nos terreiros de candomblé de Salvador, Bahia. *Ciência da Informação*, v.39, n.2, p.84-91, 2011.

### Artigo com dois autores

GRIPPA, G.; BISOFFI, G.C. Memória e hipertexto: uma reflexão sobre o conhecimento relacional. *Transinformação*, v.22, n.3, p.233-246, 2009.

## Artigo em suporte eletrônico

MOURA, M.A. Informação e conhecimento em redes virtuais de cooperação científica: necessidades, ferramentas e usos. *DataGramaZero: Revista de Ciência da Informação*, v.10, n.2, 2009. Disponível em: <[http://www.datagramazero.org.br/abr09/Art\\_02.htm](http://www.datagramazero.org.br/abr09/Art_02.htm)>. Acesso em: 16 maio 2009.

## Livro

GARCÍA GUTIÉRREZ, A. *Epistemología de la documentación*. Barcelona: Stonberg, 2011.

## Livro em suporte eletrônico

BRASIL. Ministério da Saúde. *Parto, aborto e puerpério: assistência humanizada à mulher*. Brasília: Ministério da Saúde, 2001. Disponível em: <<http://bibliotecadigital.puc-campinas.edu.br/services/e-books-MS/01-0420-M.pdf>>. Acesso em: 24 nov. 2010.

## Capítulos de livros

GASQUE, K.C.G.D. Teoria fundamentada: nova perspectiva à pesquisa exploratória. In: MUELLER, S.P.M. (Org.). *Métodos para a pesquisa em ciência da informação*. Brasília: Thesaurus, 2007. p.107-142.

## Capítulo de livro em suporte eletrônico

SABADINI, A.A.Z.P.; SAMPAIO, M. I. C.; NASCIMENTO, M. M. Preparando um periódico científico. In: SABADINI, A.A.Z.P.; SAMPAIO, M. I. C.; KOLLER, S. H. (Org.). *Publicar em psicologia: um enfoque para a revista científica*. São Paulo: Associação Brasileira de Editores Científicos de Psicologia, 2009. p.35-74. Disponível em: <[http://www.ip.usp.br/biblioteca/publicursos/publicar\\_psicologia\\_1edicao\\_2009\\_WEB\\_COR\\_13%20jul%202009.pdf](http://www.ip.usp.br/biblioteca/publicursos/publicar_psicologia_1edicao_2009_WEB_COR_13%20jul%202009.pdf)>. Acesso em: 3 jun. 2011.

## Dissertações e teses

PEREIRA, R. *Espaço Interativo (Eil): o portal de relacionamento como suporte e estímulo à relação universidade-empresa*. 2009. Dissertação (Mestrado em Engenharia de Produção) - Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2009.

## Trabalhos apresentados em congressos, seminários etc.

AMARAL, M.S.; PINHO, J.A.G. Sociedade da informação e democracia: procurando a accountability em portais municipais da Bahia. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 32., 2008, Rio de Janeiro. *Anais... Rio de Janeiro: EnANPAD*, 2008. 1 CD-ROM.

## Trabalhos apresentados em congressos, seminários etc. em formato eletrônico

GAUZ, V.; PINHEIRO, L.V.R. Fluxo da informação entre colecionadores, escribas e cientistas árabes na pré-institucionalização da ciência, séculos IV ao XV. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 11., 2010, Rio de Janeiro. *Anais Eletrônicos ... Rio de Janeiro: Unirio*, 2010. Disponível em: <<http://congresso.ibict.br/index.php/enancib/xienancib/paper/view/394/330>>. Acesso em: 20 abr. 2011.

## Texto em formato eletrônico

HEPWORTH, M. *Information literacy from the perspective of learners: implications for teaching information literacy and skills*. 2002. Available from: <[www.elit-conf.org/itilit2002/papers/ppt/08h1.doc](http://www.elit-conf.org/itilit2002/papers/ppt/08h1.doc)>. Cited: 20 Sept. 2007.

## Lista de checagem

- Declaração de responsabilidade e transferência de direitos autorais assinada por cada autor.
- Verificar se o texto, incluindo resumos, tabelas e referências, está reproduzido com letras fonte *Arial*, corpo 11 e entrelinhas 1,5 e com formatação de margens superior e inferior (no mínimo 2,5cm), esquerda e direita (no mínimo 3cm).
- Verificar se estão completas as informações de legendas das figuras e tabelas.
- Preparar página de rosto com as informações solicitadas.
- Incluir o nome de agências financiadoras e o número do processo.
- Indicar se o artigo é baseado em tese/dissertação, colocando o título, o nome da instituição, o ano de defesa, em nota de rodapé.
- Incluir título do manuscrito, em português, espanhol ou francês e em inglês.
- Incluir título abreviado (*short title*), com quarenta caracteres, para fins de legenda em todas as páginas.
- Verificar se as referências estão citadas no texto.
- Incluir permissão de editores para reprodução de figuras ou tabelas publicadas.

## Documentos

### Declaração de responsabilidade e transferência de direitos autorais

Cada autor deve ler e assinar os documentos (1) Declaração de Responsabilidade e (2) Transferência de Direitos Autorais, nos quais constarão:

- Título do manuscrito:

- Nome por extenso dos autores (na mesma ordem em que aparecem no manuscrito).

- Autor responsável pelas negociações:

1. Declaração de responsabilidade: todas as pessoas relacionadas como autoras devem assinar declarações de responsabilidade nos termos abaixo:

- "Certifico que participei da concepção do trabalho para tornar pública minha responsabilidade pelo seu conteúdo, que não omiti quaisquer ligações ou acordos de financiamento entre os autores e companhias que possam ter interesse na publicação deste artigo";

- "Certifico que o manuscrito é original e que o trabalho, em parte ou na íntegra, ou qualquer outro trabalho com conteúdo substancialmente similar, de minha autoria, não foi enviado a outra Revista e não o será, enquanto sua publicação estiver sendo considerada pela Transinformação, quer seja no formato impresso ou no eletrônico".

2. Transferência de Direitos Autorais: "Declaro que, em caso de aceitação do artigo, a revista Transinformação passa a ter os direitos autorais a ela referentes, que se tornarão propriedade exclusiva da Revista, vedado a qualquer reprodução, total ou parcial, em qualquer outra parte ou de divulgação, impressa ou eletrônica, sem que a prévia e necessária autorização seja solicitada e, se obtida, farei constar o competente agradecimento à Revista".

Assinatura do(s) autor(es)

Data \_\_\_ / \_\_\_ / \_\_\_

## Justificativa do artigo

Destaco que a principal contribuição do estudo para a área em que se insere é a seguinte:

(Escreva um parágrafo justificando porque a revista deve publicar o seu artigo, destacando a sua relevância científica, a sua contribuição para as discussões, na área em que se insere o(s) ponto(s) que caracteriza(m) a sua originalidade e o consequente potencial de ser citado).

Dada a competência na área do estudo, indico o nome dos seguintes pesquisadores (três) que podem atuar como revisores do manuscrito. Declaro igualmente não haver qualquer conflito de interesses para esta indicação.

**Toda correspondência deve ser enviada à revista Transinformação no endereço abaixo**

Núcleo de Editoração SBI - Campus I

Rod. D. Pedro I, km 136 - Sala 8 - Prédio Antiga Reitoria - Pq. das Universidades - 13086-900 - Campinas - SP - Brasil

Fone/Fax: 55 (19) 3343-7401

E-mail: [sbi.nucleoeditoracao@puc-campinas.edu.br](mailto:sbi.nucleoeditoracao@puc-campinas.edu.br)

Home Page: <http://revistas.puc-campinas.edu.br/transinfo>

# Instructions to the Authors

## Objectives and Editorial Policy

*Transinformação* is a triannual specialized journal that accepts contributions from the domestic and international scientific communities and is distributed in Brazil and abroad. Founded in 1989, it is classified in the Qualis list as A1, publishing articles that contribute to the study and scientific development of the information technology and librarianship sciences and its many sub-areas and interfaces.

The authors are responsible for the information contained in the works, as well as for the authorization to use the figures and tables that have been published in other sources.

Contributions will be accepted if they:

- a) Present an innovative approach to themes already addressed;
- b) Use innovative methodological procedures;
- c) In case of review, bring a new contribution to the field;
- d) Use a consistent methodology;
- e) Present conclusions that are based on logical arguments;
- f) Present references that are pertinent to the theme being discussed.

## Types of articles accepted

- *Original*: contribution that seeks to disclose original research results (maximum of 12 pages with 1,5 lines spacing and font Arial size 11 in A4 - size paper).

- *Review*: critical synthesis of an interesting theme for the field, by analyzing and interpreting the pertinent literature (maximum of 12 pages with 1,5 lines spacing and font Arial size 11 in A4 - size paper).

- *Essay*: reflection about the theme that generates questions and allows the creation of hypotheses for future research (only by invitation) (maximum of 10 to 15 pages with 1,5 lines spacing and font Arial size 11 in A4 - size paper).

- *Debate*: technical work accompanied by critical letters signed by authors of different institutions (three to five experts invited by the Editor), followed by replicas from the author of the main article. The person interested in submitting a manuscript to this section must first consult the Editor (maximum of 20 pages with 1,5 lines spacing and font Arial size 11 in A4 - size paper).

- *Digest*: analysis of a work to address a theoretical reference situation of the area proposed by the author, who shall assess and criticize the work disclosing his personal point of view (maximum of 3 pages with 1,5 lines spacing and font Arial size 11 in A4 - size paper).

- *Communication*: information about ongoing research; succinct information on ongoing research projects, dissertations and theses (maximum of 3 pages with 1,5 lines spacing and font Arial size 11 in A4 - size paper).

- *Translation*: translated articles authorized by the owner of the copyrights.

## Editorial procedures

### Manuscript assessment process

Unpublished original articles will be accepted for assessment by referees who are known experts on the theme in question. The works will be assessed by at least two referees from the referees that collaborate with the Journal, in a double-blind process, that is, the authors and referees remain unknown to each other.

The authors may, instead, indicate three referees to assess their manuscript.

Accepted manuscripts may return to the authors for approval of possible changes in the editing and formatting process according to the style of the Journal.

Refused manuscripts but with permission to be reformulated may return as a new work and begin a new assessment process.

### Conflict of interests

If a conflict of interest is identified by the referees, the Editorial Committee will have another *ad hoc* referee assess the manuscript.

### Please submit your works to

<<http://revistas.puc-campinas.edu.br/transinfo>>.

To be accepted, unpublished works must be accompanied by a letter signed by all authors describing the type of work and declaring that the work is being submitted only to the journal *Transinformação*, and by a document transferring the copyrights.

The letter must contain the authors' name, address, telephone numbers and e-mails and indicate which author will be the corresponding author.

If the manuscript uses figures or tables published elsewhere, the authors must attach a permission letter to use them from the copyright owners.

### Reformulated version

Underline the changes in the text or make them in another contrasting color (blue font) and send it with a letter to the editor confirming your interest in publishing your manuscript in this Journal and informing the changes made to the manuscript. If the recommendations of the referees are conflicting, the author(s) must present arguments that justify their position. The title and code of the manuscript must be included.

Proofs: Typographic proofs will be sent to the authors for correction of printing errors. The proofs must return to the Publishing Center within the deadline. Other changes to the original manuscript will not be accepted during this phase.

## Necessary contents of the manuscripts

### The title page should contain

- a) complete title - should be concise, avoiding too many words, such as "assessment of..."; "considerations about..."; "exploratory study...".

b) short title with up to forty characters (including spaces), in Portuguese (or Spanish) and English.

c) full name of all the authors, indicating the institutional affiliation of each author. Only one title and affiliation will be accepted per author. The authors, should, therefore, choose between their titles and institutional affiliations.

**Observation:** if there is no institutional affiliation, please inform your professional activity, city and state.

d) all the data regarding the titles and affiliation must be presented in full, without abbreviations.

e) include full address of all the universities with which the authors are affiliated.

f) indicate corresponding author and inform facsimile, telephone number and e-mail address.

**Observation:** this is the only part of the text where the authors are identified.

**Abstract:** all articles submitted in Portuguese, Spanish or French should have an abstract in the original language and one in English, with at least 150 words and at most 250 words.

Articles submitted in English shall be accompanied by an abstract, title and keywords in Portuguese, in addition to those in English.

The abstract should contain the objective of the work, methods and conclusions.

**Text:** except for the manuscripts presented as Review, Communication and Essay, the works must follow the formal structure for scientific works:

**Introduction:** must contain a review of the current literature pertinent to the theme, appropriate for the presentation of the problem and point out its relevance. It should not be extensive except in Review Articles.

**Methods:** must contain a clear and succinct description of the methods used, followed by the corresponding references, including: procedures used, universe and sample, measurement instruments, and if applicable, validation method and statistical treatment.

In relation to the statistical analysis, the authors should demonstrate that the procedures were not only appropriate to test the hypotheses of the study but also correctly interpreted.

**Results:** whenever possible, the results must be presented in self-explanatory tables and figures and contain statistical analysis. Avoid repeating the data in the text.

Tables, charts and figures should be limited to five in all and given consecutive and independent numbers in Arabic numerals, according to the order the data is mentioned, and should be presented in individual sheets and separated, indicating their location in the text. **It is essential to inform the location and year of the study.** Each one should have a brief title. The charts and tables must be open laterally.

The author(s) are responsible for the quality of the figures (drawings, illustrations, tables and graphs) that should be large enough to fit one or two columns (7 and 15cm, respectively); **the landscape format is not accepted.** Figures should be in jpeg format and have a minimum resolution of 400 dpi.

Graphs and drawings should be made in vector design software (Microsoft Excel, CorelDraw, Adobe Illustrator etc.), followed by their quantitative parameters in a table and the name of all its variables.

The publication of color images will be paid by the author(s) once the technical viability of their reproduction is verified. If the authors are interested, the Journal will provide the costs which will vary according to the number of images, their distribution in different pages, and the concomitant publication of color material by other author(s).

All illustrations must be created in one- or two-column sizes (7.5 and 15 cm, respectively). The landscape format will not be allowed.

Study location and year must be indicated in all illustrations.

**Discussion:** The discussion must properly and objectively explore the results of the study, and compare them with other published data.

**Conclusion:** This section must present the relevant conclusions taking into consideration the objectives of the work, and indicate ways that the study can be continued. **Literature citations will not be accepted in this section.**

**Abbreviations and acronyms:** must be used in a standardized manner, and be restricted to those used conventionally or sanctioned by use, accompanied by their full meaning when they first appear in the text. They cannot be used in the title or the abstract.

**Acknowledgments:** can be included but in a paragraph no longer than three lines and include only the individuals and institutions who effectively collaborated with the study.

**Attachments:** can only be included when they are essential for the understanding of the text. The editors will then decide if they should be published.

## Citations in the text

Literature citations in the text must be included in the references.

Citations and/or references to course completion **monographs** and other **unpublished texts** (lectures etc.) **will not be accepted.**

If an in-press unpublished work from one of the authors of the manuscript is cited, the letter of acceptance by the journal that accepted to publish the study must be included.

If unpublished data from other researchers are cited in the manuscript, a letter from the respective researchers authorizing the use of their data is necessary.

## Specific cases

1) Literal citations of **up to three lines** must be in quotes, not in italics, and followed by, in parenthesis, the last name of the author, date, page, **without space between the period and the number.** A period must follow the parenthesis.

2) Literal citations with more than three lines must be in a separate paragraph, with a left margin of 4cm, using single space between the lines, smaller font than that of the text, without quotes, without italic, ending on the right margin of the text, followed by, in parenthesis, the last name of the author, date and page.

3) Many authors cited in sequence: **use the order in which the documents were published**, separated by a semicolon (Crespo, 2005; Costa; Ramalho, 2008; Moresi *et al.*, 2010).

4) Texts with two authors: Crippa and Bisoffi (2010) (in the body of the text); (Crippa; Bisoffi, 2010) (with in parenthesis).

5) Texts with three or more authors: (Griselda *et al.*; 2009) (in of the parenthesis) and Griselda *et al.* (2009) (out of the parenthesis).

6) Citations of the same author published in the same years: add a small letter to the date, without space Example: (Morin, 2000a, 2000b).

## References

The references are based on the NBR-6023/2002 and should be limited to 30 references per article, except in the case of review articles, which may have as many as 50 references. They should be ordered alphabetically by the last name of the first author.

### Specific cases

1) Journals names cannot be abbreviated.

2) References with coincident authors and dates must be ordered according to the title of the article and a small letter should be added after the date without a space in between.

3) References with three or more authors should be indicated only by the first author, followed by the expression *et al.*

**The preciseness and appropriateness of references to studies that have been consulted and mentioned in the text are of the author's responsibility.**

For other examples, we recommend consulting the ABNT-NBR-6023/2002 norms.

## Examples

### Article with one author

OLIVEIRA, A. Direito à memória das comunidades tradicionais: organização de acervo nos terreiros de candomblé de Salvador, Bahia. *Ciência da Informação*, v.39, n.2, p.84-91, 2011.

### Article with two authors

GRIPPA, G.; BISOFFI, G.C. Memória e hipertexto: uma reflexão sobre o conhecimento relacional. *Transinformação*, v.22, n.3, p.233-246, 2009.

### Electronic article

MOURA, M.A. Informação e conhecimento em redes virtuais de cooperação científica: necessidades, ferramentas e usos. *DataGramZero: Revista de Ciência da Informação*, v.10, n.2, 2009. Disponível em: <[http://www.datagramazero.org.br/abr09/Art\\_02.htm](http://www.datagramazero.org.br/abr09/Art_02.htm)>. Acesso em: 16 maio 2009.

### Book

GARCÍA GUTIÉRREZ, A. *Epistemología de la documentación*. Barcelona: Stonberg, 2011.

## Electronic book

BRASIL. Ministério da Saúde. *Parto, aborto e puerpério: assistência humanizada à mulher*. Brasília: Ministério da Saúde, 2001. Disponível em: <<http://bibliotecadigital.puc-campinas.edu.br/services/e-books-MS/01-0420-M.pdf>>. Acesso em: 24 nov. 2010.

## Book chapters

GASQUE, K.C.G.D. Teoria fundamentada: nova perspectiva à pesquisa exploratória. In: MUELLER, S.P.M. (Org.). *Métodos para a pesquisa em ciência da informação*. Brasília: Thesaurus, 2007. p.107-142.

## Electronic book chapters

SABADINI, A.A.Z.P.; SAMPAIO, M.I.C.; NASCIMENTO, M.M. Preparando um periódico científico. In: SABADINI, A.A.Z.P.; SAMPAIO, M.I.C.; KOLLER, S.H. (Org.). *Publicar em psicologia: um enfoque para a revista científica*. São Paulo: Associação Brasileira de Editores Científicos de Psicologia, 2009. p.35-74. Disponível em: <[http://www.ip.usp.br/biblioteca/publicursos/publicar\\_psicologia\\_1edicao\\_2009\\_WEB\\_COR\\_13%20jul%202009.pdf](http://www.ip.usp.br/biblioteca/publicursos/publicar_psicologia_1edicao_2009_WEB_COR_13%20jul%202009.pdf)>. Acesso em: 3 jun. 2011.

## Dissertations and theses

PEREIRA, R. *Espaço Interativo (Eil): o portal de relacionamento como suporte e estímulo à relação universidade-empresa* 2009. Dissertação (Mestrado em Engenharia de Produção) - Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2009.

## Papers presented in congresses, seminars etc.

AMARAL, M.S.; PINHO, J.A.G. Sociedade da informação e democracia: procurando a accountability em portais municipais da Bahia. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 32., 2008, Rio de Janeiro. *Anais...* Rio de Janeiro: EnANPAD, 2008. 1 CD-ROM.

## Electronic papers presented in congresses, seminars etc.

GAUZ, V.; PINHEIRO, L.V.R. Fluxo da informação entre colecionadores, escribas e cientistas árabes na pré-institucionalização da ciência, séculos IV ao XV. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 11., 2010, Rio de Janeiro. *Anais Eletrônicos ...* Rio de Janeiro: Unirio, 2010. Disponível em: <<http://congresso.ibict.br/index.php/enancib/xienancib/paper/view/394/330>>. Acesso em: 20 abr. 2011.

## Electronic text

HEPWORTH, M. *Information literacy from the perspective of learners: implications for teaching information literacy and skills*. 2002. Available from: <[www.elit-conf.org/itilit2002/papers/pp/08h1.doc](http://www.elit-conf.org/itilit2002/papers/pp/08h1.doc)>. Cited: 20 Sept. 2007.

## Checklist

- Declaration of responsibility and transfer of the copyrights signed by each author.

- Verify if the text, including abstract, tables and references are in Arial font, size 11 and lines with a spacing of 1.5, with upper and lower margins of at least 2.5cm and left and right margins of at least 3cm.



- Verify if the information in the legends of figures and tables is complete.

- Prepare the cover page with the requested information.

- Include the name of the sponsors and process number.

- Indicate if the article is based on a thesis/dissertation, and include the title, name of the institution and the year it was defended in the footnotes.

- Include the title of the manuscript in Portuguese, Spanish or French and in English.

- Include short title with a maximum of forty characters that will be used as legend in all pages.

- Verify if the references are cited in the text.

- Include permission of editors to reproduce figures and tables published elsewhere.

## Documents

### Declaration of responsibility and transfer of copyrights

Each author must read and sign the documents (1) Declaration of Responsibility and (2) Transfer of Copyrights, which must include:

- Title of the manuscript:

- Full name of the authors (in the same order in which they appear in the manuscript).

- Author responsible for the negotiations:

1. Declaration of responsibility: all authors must sign the declaration of responsibility on the terms below:

- "I certify that I have participated in the conception of this manuscript and disclose that I am responsible for its content and that I have not omitted any affiliations or sponsoring agreements between the authors and companies that may have interest in the publication of this article."

- "I certify that the manuscript is original and that the work, in part or in full, or any other work with a substantially similar content, of my authorship, was not submitted to another journal and will not be submitted to another journal while the Journal *Transinformação* is considering publishing it whether in print or electronic format."

2. Transfer of copyrights: "I declare that if the article is accepted for publication, the Journal *Transinformação* will own its copyrights with exclusivity, and any reproduction, partial or full, anywhere else or by any other means, in print or electronic, without the previous and necessary consent of the Journal *Transinformação* is strictly forbidden. If the consent is obtained, I will duly thank the Journal *Transinformação* for it in the paper where the data will be used."

Signature of the author(s)

Date \_\_\_\_ / \_\_\_\_ / \_\_\_\_

### Article justification

I declare that the main contribution of the study for its area is the following:

(Write a paragraph justifying why the Journal should publish your article, pointing out its scientific relevance, its contribution for the discussions, the area of its originality and the consequent potential for being cited).

Given their competence in the area of the study, I indicate the name of the following researchers (three) that may act as referees of the manuscript. I also declare that I have no conflict of interests for this indication.

**All correspondence shall be sent to the journal *Transinformação* at the following address:**

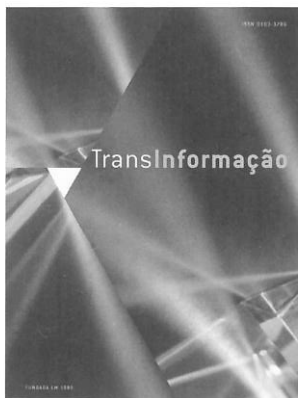
Núcleo de Editoração SBI - *Campus* I

Rod. D. Pedro I, km 136 - Sala 8 - Prédio Antiga Reitoria - Pq. das Universidades - 13086-900 - Campinas - SP - Brasil

Fone/Fax: 55(19) 3343-7401

E-mail: [sbi.nucleodeeditoracao@puc-campinas.edu.br](mailto:sbi.nucleodeeditoracao@puc-campinas.edu.br)

Home Page: <http://revistas.puc-campinas.edu.br/transinfo>



Prezado leitor,

É com satisfação que vimos convidá-lo ASSINAR ou RENOVAR a revista *Transinformação*, a melhor forma de ter contato com os trabalhos desenvolvidos por pesquisadores da área através de uma publicação nacional, indexada nas bases de dados internacionais: Web of Science, JCR Social Science, Latindex e Clase. Qualis A1.

Esperamos contar com sua presença entre nossos assinantes regulares.

Preencha o canhoto abaixo.

Comissão Editorial

<input type="checkbox"/>	ASSINATURA				
<input type="checkbox"/>	RENOVAÇÃO				
<input type="checkbox"/>	Volume 17 (1, 2 e 3) (2005)	<b>Pessoas Físicas</b>	R\$ 40,00	<input type="checkbox"/>	<b>Institucional</b> R\$ 50,00 <input type="checkbox"/>
<input type="checkbox"/>	Volume 18 (1, 2 e 3) (2006)	<b>Pessoas Físicas</b>	R\$ 40,00	<input type="checkbox"/>	<b>Institucional</b> R\$ 60,00 <input type="checkbox"/>
<input type="checkbox"/>	Volume 19 (1, 2 e 3) (2007)	<b>Pessoas Físicas</b>	R\$ 40,00	<input type="checkbox"/>	<b>Institucional</b> R\$ 60,00 <input type="checkbox"/>
<input type="checkbox"/>	Volume 20 (1, 2 e 3) (2008)	<b>Pessoas Físicas</b>	R\$ 40,00	<input type="checkbox"/>	<b>Institucional</b> R\$ 100,00 <input type="checkbox"/>
<input type="checkbox"/>	Volume 21 (1, 2 e 3) (2009)	<b>Pessoas Físicas</b>	R\$ 40,00	<input type="checkbox"/>	<b>Institucional</b> R\$ 100,00 <input type="checkbox"/>
<input type="checkbox"/>	Volume 22 (1, 2 e 3) (2010)	<b>Pessoas Físicas</b>	R\$ 40,00	<input type="checkbox"/>	<b>Institucional</b> R\$ 130,00 <input type="checkbox"/>
<input type="checkbox"/>	Volume 23 (1, 2 e 3) (2011)	<b>Pessoas Físicas</b>	R\$ 50,00	<input type="checkbox"/>	<b>Institucional</b> R\$ 140,00 <input type="checkbox"/>
<input type="checkbox"/>	Volume 24 (1, 2 e 3) (2012)	<b>Pessoas Físicas</b>	R\$ 50,00	<input type="checkbox"/>	<b>Institucional</b> R\$ 140,00 <input type="checkbox"/>

Nome: \_\_\_\_\_

Endereço: \_\_\_\_\_ Bairro: \_\_\_\_\_

CEP: \_\_\_\_\_ Cidade: \_\_\_\_\_ Estado: \_\_\_\_\_ Telefone: \_\_\_\_\_

CNPJ/CPF: \_\_\_\_\_ E-mail: \_\_\_\_\_

Assinatura: \_\_\_\_\_ Data: \_\_\_\_/\_\_\_\_/\_\_\_\_

## FORMA DE PAGAMENTO

### Boleto Bancário

Solicitar via e-mail: [sbi.assinaturane@puc-campinas.edu.br](mailto:sbi.assinaturane@puc-campinas.edu.br)

**Obs.: Favor indicar a melhor data para o pagamento e em nome de quem o boleto deverá ser emitido.**

**Transinformação** - Núcleo de Editoração - Prédio da Antiga Reitoria Sala 8 - Campus I

Rod. Dom Pedro I, km 136 - Pq. das Universidades - 13086-900 - Campinas - SP

Fone: (19) 3343-7351/3343-7640 - Fax: (19) 3343-7271

E-mail: [sbi.assinaturane@puc-campinas.edu.br](mailto:sbi.assinaturane@puc-campinas.edu.br) - Home Page: <http://revistas.puc-campinas.edu.br/transinfo>

**Pontifícia Universidade Católica de Campinas**

(Sociedade Campineira de Educação e Instrução)

**Grão-Chanceler:** Dom Airton José dos Santos

**Reitora:** Profa. Dra. Angela de Mendonça Engelbrecht

**Vice-Reitor:** Prof. Dr. Eduard Pranic

**Pró-Reitor de Graduação:** Prof. Dr. Germano Rigacci Júnior

**Pró-Reitora de Pesquisa e Pós-Graduação:** Profa. Dra. Vera Engler Cury

**Pró-Reitora de Extensão e Assuntos Comunitários:** Profa. Dra. Vera Engler Cury

**Pró-Reitor de Administração:** Prof. Dr. Ricardo Pannain

**Diretora do Centro de Ciências Humanas e Sociais Aplicadas:** Profa. Dra. Luzia Siqueira Vasconcelos

**Diretor-Adjunto do Centro de Ciências Humanas e Sociais Aplicadas:** Prof. Luis Arlindo Feriani Filho

**Transinformação**

Com capa impressa no papel supremo 250g/m<sup>2</sup>  
e miolo no papel couchê fosco 90g/m<sup>2</sup>

**Capa / Cover**

Kátia Harumi Terasaka

**Editoração eletrônica / DTP**

Beccari Propaganda e Marketing

**Impressão / Printing**

Gráfica Editora Modelo Ltda

**Tiragem / Edition**

800

**Distribuição / Distribution**

Sistema de Bibliotecas e Informação da  
PUC-Campinas - Serviço de Publicação,  
Divulgação e Intercâmbio

## Artigos | Articles

- 77 **A representação descritiva no contexto da Web semântica**  
*Descriptive representation in the semantic Web context*  
Maria Elisabete Catarino; Terezinha Batista de Souza
- 91 **A reference ontology for digital scientific journals applied to systematic literature review processes**  
*Uma ontologia de referência para periódicos científicos digitais aplicada em processos de revisão sistemática da literatura*  
Fernando Benedet Ghisi; Gleisy Regina Bóries Fachin; Marcos Henrique dos Santos; Denilson Sell; Gregório Jean Varvakis Rados
- 103 **PuertoTex: un software de minería textual para la creación de resúmenes automáticos en el dominio de ingeniería de puertos y costas basado en ontologías**  
*PuertoTex: a data mining software based on ontologies for automatic summarization on port and coastal engineering domain*  
Amed Leiva-Mederos; Sandor Domínguez-Velasco; Jose Antonio Senso
- 117 **Mudança de paradigma e sua ruptura: um estudo de caso na Museologia e a pluralidade paradigmática da Ciência da Informação**  
*Paradigm shift and its rupture: a case study and paradigmatic plurality of Museology and Information Sciences*  
Mário Gouveia Júnior; Raimundo Nonato Macedo dos Santos
- 127 **O reflexo da inteligência coletiva nas organizações**  
*Effect of collective intelligence in organizations*  
Ketry Gorete Farias dos Passos; Edna Lúcia da Silva
- 137 **Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos**  
*Information retrieval systems: a discussion based on enunciative parameters*  
Vera Maria Araujo Pigozzi de Araujo