

**Título: Contribuições do binômio metadados-ontologias para a Ciência da Informação - uma Análise Sucinta.**

**Resumo:** A Ciência da Informação, como Ciência, é interdisciplinar e como disciplina vem ocupando um espaço deixado pela Biblioteconomia, convivendo harmonicamente entre disciplinas sociais e tecnológicas. O crescente fluxo informacional e o desenvolvimento de novas tecnologias agregaram à Ciência da Informação um sem-número de ferramentas para a construção da comunicação científica. Dentre essas ferramentas destacam-se os metadados e as ontologias que vieram contribuir significativamente para o armazenamento e a recuperação da informação.

**Palavras-chave:** Ciência, Ciência da Informação, integração, metadados, ontologias

## INTRODUÇÃO

A produção do conhecimento cresceu vertiginosamente a partir do último século, graças à sua estreita ligação com as Tecnologias da Informação e Comunicação (NEVES, 2002).

As TIC têm contribuído para a convergência entre os suportes informacionais bem como a criação de novos objetos de informação. Diante de um mundo cada vez mais globalizado, os conteúdos informacionais, particularmente os documentos científicos, são demandados por análise, busca, recuperação e verificação.

A diversidade dos sistemas de informação demonstra que eles acarretam em uma série de problemas, como a falta de integração, a redundância e a inconsistência de dados, perda de conteúdo semântico nas traduções, dispêndio de tempo e dinheiro na transferência de um sistema para outro ou ainda o trabalho de redigitação, conforme o caso.

O desenvolvimento de novos sistemas de informação, baseados em gerência dos dados e a migração ou integração dos sistemas existentes com os novos visa alcançar a tão sonhada interoperabilidade. É a chamada "interconexão efetiva de diferentes sistemas de computador, bancos de dados ou redes com o fim de apoiar a computação distribuída e/ou o intercâmbio de dados", de acordo com PACHECO e KERN (2001).

Diante desse contexto, a interoperabilidade e a integração plena de informações heterogêneas entre sistemas de informação é uma utopia a ser buscada e atingida. Essa utopia passa a ser possível quando os sistemas são desenvolvidos a partir de modelos de dados integrados, fazendo com que uma estrutura comum e integrada permita que vários usuários compartilhem o mesmo banco de dados. A esse modelo de estrutura é dado o nome de metabanco de dados ou metadados.

A definição e a construção de bases de conhecimento interoperáveis, a partir da utilização de metadados, foram facilitadas por meio das ontologias. Essas são ferramentas auxiliares da área da Ciência da Computação, que têm sido amplamente usadas na subárea da representação do conhecimento, como na inserção e recuperação de documentos em sistemas, a extração de

informação em textos de linguagem natural, a aquisição automática do conhecimento e a troca de informações entre softwares agentes inteligentes, dentre outras aplicações.

Na medida em que as fontes de dados crescem de forma assustadora, mais difícil tem se tornado a recuperação da informação, crescendo de importância os estudos relativos à organização da informação.

Na área da Biblioteconomia, a partir dos anos 70, o uso de tecnologias para acesso a bases de dados referenciais passaram a ter maior presença com o advento dos OPACs (*On-line Public Access Catalogs*), tornando as bibliotecas as primeiras instituições a utilizarem os sistemas de informação automatizados para recuperação de informações.

Nos anos 80, a difusão do uso do computador levou, de acordo com VICKERY (1986), para fora dos ambientes das bibliotecas e da própria Biblioteconomia, uma nova visão sobre o tratamento a se dar em relação à estrutura de registros e arquivos de bases de dados. Essa visão incluiu as estruturas específicas dos dados dentro dos programas de computador e suas linguagens, e também uma nova visão da estrutura semântica e sintática da linguagem natural, dos recursos da inteligência artificial e da identificação dos modelos mentais da memória humana.

A partir da década de 90, com o advento e a explosão do uso da Internet, surgiram as bibliotecas digitais como novos meios de comunicação, possibilitando a visualização de uma “biblioteca sem paredes” e de caráter universal.

Para atender a essa mudança de paradigma, da visão da informação catalográfica para uma visão sob a ótica do meio digital, a representação do conhecimento e a recuperação da informação deram espaço a outras mídias além das textuais, como as sonoras e imagéticas, passando a lidar com novos formatos para produzir, organizar e recuperar objetos digitais.

O presente artigo tem por objetivo realizar uma análise sucinta, em face da grande diversidade de aplicações, das contribuições dos metadados e ontologias para a Ciência da Informação.

## 2. REFERENCIAL TEÓRICO

### 2.1 Ciência

O que é Ciência e qual sua origem? A princípio parece fácil responder a tal questionamento, contudo, não o é. A palavra Ciência tem sua origem no latim, *scientia*, e significa conhecimento. Porém, o que é conhecimento e qual a diferença entre conhecimento científico e conhecimento vulgar (senso comum)?

Para se evitar muitos questionamentos, usaremos a definição de FERREIRA (2000) para o significado de Ciência.

[Ciência é um] conjunto de conhecimentos socialmente adquiridos ou produzidos, historicamente acumulados, dotados de universalidade e objetividade que permitem sua transmissão, e estruturados com métodos, teorias e linguagens próprias, que visam compreender e orientar a natureza e as atividades humanas.

Para CHAUI (2000),

historicamente, três têm sido as principais concepções de ciência ou de ideais de cientificidade: o racionalista, cujo modelo de objetividade é a matemática; o empirista, que toma o modelo de objetividade da medicina grega e da história natural do século XVII; e o construtivista, cujo modelo de objetividade advém da idéia de razão como conhecimento aproximativo.

Quanto à sua origem, costuma-se atribuí-la aos antigos pensadores gregos, entretanto, há registro de trabalhos científicos no ocidente por volta do século VI a.C. A busca, de algumas pessoas, por mais explicações de certos fatos e causas, levou ao surgimento de Ciências como a Matemática, Astrologia, Medicina, Filosofia, Biologia, etc.

A partir desses filósofos, o pensamento científico pode evoluir até os dias atuais. Entre os primeiros pensadores sobre o conhecimento científico temos: Pitágoras, Sócrates, Platão, Aristóteles, entre outros. Contudo, até por volta do fim século XVII, período do Iluminismo, o significado de Ciência, ainda, era interpretado como qualquer conhecimento que fosse exato.

Hoje, Ciência tem o significado mais específico de conhecimento científico, ou também, “é usualmente identificada com o conjunto ou sistema organizado de conhecimento científico.” (LUNGARZO, 1989). Ainda segundo LUNGARZO (1989), a Ciência, como conhecimento científico, tem intrínseca

relação com métodos e técnicas de descoberta, com fatores sociais e psicológicos. Essa forte vinculação garante a distinção entre conhecimento científico e senso comum.

Com a evolução do pensamento racional, com a descoberta e criação de novas áreas do conhecimento, foi necessário (re) classificar as áreas e subáreas das Ciências.

Segundo CHAUÍ (2000), “a primeira classificação sistemática das ciências de que temos notícia foi a de Aristóteles (...)”. Ele se utilizou de três critérios para classificar os saberes: a) critério da ausência ou presença da ação humana nos seres investigados; b) critério da imutabilidade ou permanência e da mutabilidade ou movimento dos seres investigados; c) critério da modalidade prática, levando à distinção entre ciências que estudam a práxis e as técnicas.

Atualmente, “das inúmeras classificações propostas, as mais conhecidas e utilizadas foram feitas por filósofos franceses e alemães do século XIX” (CHAUÍ, 2000). Os critérios utilizados, semelhantemente a Aristóteles, foram três: a) tipo de objeto estudado; b) tipo de método empregado; c) tipo de resultado obtido.

Com base nesses critérios e com algumas alterações sobre as várias classificações anteriores, resultou a classificação que é usada até hoje.

- ciências matemáticas ou lógico-matemáticas (aritmética, geometria, álgebra, trigonometria, lógica, física pura, astronomia pura, etc.);
- ciências naturais (física, química, biologia, geologia, astronomia, geografia física, paleontologia, etc.);
- ciências humanas ou sociais (psicologia, sociologia, antropologia, geografia humana, economia, lingüística, psicanálise, arqueologia, história, etc.);
- ciências aplicadas (todas as ciências que conduzem à invenção de tecnologias para intervir na Natureza, na vida humana e nas sociedades, como por exemplo, direito, engenharia, medicina, arquitetura, informática, etc.). (CHAUÍ, 2000)

No entanto, presentemente, cada uma das ciências também se subdivide, assim, novas delimitações de objeto e de métodos de investigação se fazem necessários. Como exemplo, a física subdivide-se em mecânica, acústica, óptica, etc.; a biologia em botânica, zoologia, fisiologia, genética, etc.; a psicologia subdivide-se em psicologia do comportamento, do desenvolvimento, psicologia clínica, psicologia social, etc.

Porém, o uso do termo "Ciência" é, ainda, empregado de forma inusitada, principalmente, quando tratamos de campos novos e interdisciplinares. Alguns desses novos saberes, mesmo em partes, também fazem uso de métodos científicos, como é o caso da Ciência da Computação, da Ciência da Informação, da Ciência Ambiental, etc.

A respeito da Ciência da Informação, sua característica primordial é a sua natureza interdisciplinar, que a faz transitar entre as Ciências Humanas, Matemáticas, Naturais e Aplicadas.

“Constitui-se assim sua interdisciplinaridade, característica cada vez mais presente como componente da Ciência na sociedade atual, em que a magnitude dos problemas enfrentados (ecológicos, étnicos, demográficos) está a exigir soluções inovativas e plurais. A Ciência da Informação vem se consolidando, então, a partir de elementos emprestados da matemática, da física, da biologia, da psicologia, da sociologia, da antropologia, da semiologia e da teoria da comunicação e de quantas ciências puderem contribuir para sua fundamentação e aplicabilidade” (CARDOSO, 1996 apud ARAUJO, 2003).

## **2.2 Ciência da Informação**

A origem da Ciência da Informação, segundo WERSIG (1993), veio da necessidade de se documentar a informação disponível, a partir de uma rede conceitual interdisciplinar que abrange desde a sociologia, a filosofia da ciência, a história até à comunicação no campo científico. Seu objeto de estudo é a informação que, como fenômeno, está relacionado a todos os campos do conhecimento (FREIRE, 2006).

Para MOSTAFA (1995, p. 306),

a Ciência da Informação nasce ao lado de outras configurações, como a do processamento automático de dados, a análise de sistemas, a cibernética, a inteligência artificial, a pesquisa operacional, a psicologia cognitivista, todas ciências novas (disciplinas novas) [...] Das comunicações a Ciência da Informação o absorveu a questão mesma do processo de comunicação; entendeu já no início que a informação flui por um processo de comunicação, concentrando-se no processo de comunicação na ciência. Como esse processo passa por canais, a Ciência da Informação passou a ser a ciência desses canais.

NEVES (2002) afirma ainda que a Ciência da Informação precisa incorporar aos processos de formação novas abordagens que permitam o aprimoramento do diálogo entre os sujeitos que interagem com os sistemas de informação. Desse diálogo com diversas áreas do conhecimento, visando a

comunicação humana, a Ciência da Informação, de acordo com SARACEVIC (1996) é um “campo dedicado às questões científicas e às práticas profissionais voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual, do uso e das necessidades de informação”.

A Ciência da Informação estuda os fenômenos relacionados a todos os campos do conhecimento científico, a partir da essência da informação.

O processo do diálogo entre os diversos sujeitos é operado por meio da linguagem, que, como fenômeno informacional, é dinâmico, complexo e de múltiplas facetas.

Em síntese, a Ciência da Informação visa conduzir os estudos relativos à comunicação humana, particularmente os processos de organização, representação e uso da informação.

### **2.3 Metadados**

Com o crescente aumento na produção da informação torna-se, cada vez mais, necessária a administração desta, de maneira que se possa recuperar o que foi criado.

A localização da informação é um dos desafios principais para a gestão da informação. Para resolver esta questão diversos recursos foram desenvolvidos no sentido de criar ferramentas que auxiliem na administração e recuperação da informação desejada. No mundo de hoje, onde a informação digital é criada com a rapidez com que a eletrônica permite nos comunicar, para a gestão de toda esta informação existem, segundo SENSO (2003), três grandes linhas de ação:

A primeira é a de índices compilados manualmente – estes são base de dados onde os usuários, que são os criadores das páginas Web, sugerem sua publicação dentro de algumas categorias, mediante um formulário.

A segunda são as bases de dados rastreadas por robôs ou *spiders* (aranhas) – partindo de um esquema clássico: uma interface, um motor de busca e uma base de dados, os buscadores utilizam um robô ou *spider* para a alimentação automática da sua base de dados.

Os robôs são programas ordenadores que são feitos para recorrer de forma automática à estrutura de hipertexto de um servidor Web, com o objetivo de alimentar as bases de dados textuais a partir de documentos HTML ou qualquer outro formato eletrônico de edição. Por fim, para completar o trabalho do robô, que é encarregado da extração e indexação de documentos, existe um motor de busca que permite interrogar as bases de dados de diversos clientes na Web.

A terceira linha, segue o método de indexação distribuída – um exemplo clássico deste sistema e o serviço de *Harvest*, que surgiu em 1993 dentro da linha de trabalho da IRTF-RD (*Internet Research Task Force Research Group – Resource Discovery*).

O *Harvest* é uma arquitetura de hardware e software repartida entre diversos servidores distintos por toda Web. Os elementos principais deste modelo de *Harvest* são: *Gatherers*, que consiste em um software instalado em um servidor na Web que periodicamente extrai informação relativa aos arquivos disponíveis para a comunidade de usuários da Internet; *Brokers* que recuperam automaticamente a informação extraída por um ou mais *gatherers* e integram em índices sobre os que se podem lançar equações de busca. A comunicação entre *brokers* e *gatherers* utiliza como protocolo um sistema de metadados denominado SOIF (*Summary Object Interchange Format*). (SENSO, 2003)

No contexto atual, metadado refere-se a alguma estrutura descritiva da informação sobre outra informação ou conhecimento, auxiliando na identificação, descrição, localização e gerenciamento desse recurso.

Dentre as principais funções dos metadados podem ser destacadas:

- Resumir o significado dos dados;
- Permitir a busca;
- Determinar se o dado é o que se necessita;



- Prevenir certos usos (PICS); <sup>1</sup>
- Recuperar e utilizar a cópia de um dado;
- Mostrar instruções de como interpretar um dado;
- Obter informações sobre as condições de uso (direitos do autor);
- Obter informação a respeito do ciclo de vida do dado;
- Oferecer informação relativa ao criador do dado;
- Indicar relações com outros recursos;
- Controlar a gestão da informação. (IANNELA; WAUGH, 1997)

Cada vez mais há a necessidade de melhores formas para encontrar e avaliar informações na Internet e nas Intranets. Para administrar os diversos recursos, nos diversos níveis de gerenciamento do conhecimento, existem algumas formas de administração de dados, assim definidos por IKEMATU (2001).

Estágio	Recurso a ser administrado	Definição de Metadados
Dados	Valores dos dados	Informação necessária para administrar o recurso dos dados
Informação	Valores dos dados e o contexto da informação	Informação necessária para administrar o recurso da informação
Conhecimento	Valores dos dados, contexto da informação e instruções das regras de negócio	Informação necessária para administrar as regras e políticas de negócio da organização
Sabedoria	Valores de dados, contexto da informação, regras de negócio executáveis, monitoração das regras de negócio e regras e métricas de avaliação	Informação necessária para administrar o comportamento da organização de acordo com suas regras e políticas de negócio.

Tabela 1 – Nível de Gerenciamento do Conhecimento

Este quadro demonstra a importância da utilização dos metadados na busca da informação, culminando num sistema de gerenciamento de conhecimento. A importância dos sistemas de metadados é ressaltada, por algumas razões, dentre elas:

<sup>1</sup> Platform for Internet Content Selection. Mecanismo que utiliza metadatos para controlar o acesso a determinado tipo de páginas atendendo a um sistema de classificação previamente estabelecido.

Incrementam a acessibilidade, pois a existência de um conjunto de metadados descreve corretamente um ou vários objetos e aumenta a possibilidade de acesso à informação, conforme definiram GILLILAND e SWETLAND (1998);

- Diminuem o tráfego na rede, porque ao iniciar a representação do objeto, e não o objeto em si, não demanda muito da banda de comunicação para fazer as buscas ou gerar os índices. (ORTIZ; REPISO 1999);
- Expandir o uso da informação, facilitando a difusão de versões digitais em um único objeto;
- Controle de versões não só no que se referir à gestão da vida de um objeto, mas também no que tem haver com sua difusão, gerenciando metadados diferentes a respeito do mesmo objeto.
- Aspectos legais que permitem os metadados estabelecer restrições de exploração e informar sobre os direitos do autor. (SENSO, 2003);
- Preservação do objeto original, assim como afirmam MILSTEAD e FELDMAN (1999), as buscas através da Web são, na atualidade, um processo de equiparação entre os termos da consulta e do documento.

Para maximizar a utilização de metadados, nos dias de hoje, são utilizadas linguagens de controle que permitem aumentar a precisão nas pesquisas feitas na Web. As metalinguagens são linguagens de marcação, ferramentas utilizadas para refinar a recuperação da informação, elevando a pesquisa ao nível semântico e estrutural do documento. As linguagens mais utilizadas são o SGML (*Standard Generalized Markup Language*) e o XML (*eXtensible Markup Language*). A utilização dessas tecnologias permite a construção de ferramentas para tratamento e conversão dos diversos padrões eletrônicos que foram surgindo.

Desta forma, os metadados de uma maneira geral, possibilitam a organização, classificação e descrição da informação, ajudando no gerenciamento da mesma no ambiente Web.

## 2.4 Ontologias

Diversas técnicas buscam dar um tratamento aos dados e objetos digitais, para facilitar a seleção, o processamento, a recuperação e a disseminação da informação, tais como a utilização de glossários, dicionários, categorizações e classificações (taxonomia).

As técnicas utilizadas mais recentemente para lidar com as estruturas de informação a partir de conceitos e seus relacionamentos são as ontologias, os *thesauri* e as redes semânticas.

Algumas definições de ontologias:

*“Ontologia é um ‘catálogo de tipos de coisas’ em que se supõe existir um domínio, na perspectiva de uma pessoa que usa uma determinada linguagem.”* (SOWA,1999).

*“Ontologia é considerada como uma formalização de conceituação, ou seja, a representação da visão de um determinado domínio de atuação. Estes conceitos são normalmente organizados em taxonomias.”* (GRUBER,1995)

*“Uma ontologia é uma especificação explícita dos objetos, conceitos e outras entidades que assumimos existirem em uma área de interesse, além das relações entre esses conceitos e restrições expressados através de axiomas.”* (GRUBER,1995)

*“Estabelece uma terminologia comum, sem ambigüidades, de forma a facilitar a comunicação, a representação e o compartilhamento do conhecimento entre os diferentes profissionais de uma organização.”* (GRUBER,1995)

*“Ontologia é um documento ou arquivo que formalmente define os relacionamentos entre termos.”* (BERNERS-LEE, 1997).

Segundo ALMEIDA e BAX (2003), a utilização de ontologias na organização do conteúdo das fontes de dados é uma abordagem que tem se destacado na organização do conteúdo das fontes de dados.

NOVELLO (2003) indica que com a utilização de ontologias é possível definir uma infra-estrutura para integrar sistemas inteligentes no nível do conhecimento, trazendo grandes vantagens, como:

- Colaboração: possibilitam o compartilhamento do conhecimento entre os membros interdisciplinares de uma equipe;
- Interoperação: facilitam a integração da informação, especialmente em aplicações distribuídas;
- Informação: podem ser usadas como fonte de consulta e de referência do domínio;
- Modelagem: as ontologias são representadas por blocos estruturados que podem ser reusáveis na modelagem de sistemas no nível de conhecimento.
- Busca baseada em ontologia: recuperar recursos desejados em bases de informação estruturadas por meio de ontologias. Desta forma, a busca torna-se mais precisa e mais rápida, pois quando não é encontrada uma resposta exata à consulta, a estrutura semântica da ontologia possibilita, ao sistema, retornar respostas próximas à especificação da consulta.

### **3. CONTRIBUIÇÕES**

#### **3.1 Metadados e Ontologias na construção da informação**

A construção da informação tem origem nas atividades científicas e técnicas de onde fluem os conhecimentos científicos e técnicos que se transformarão, depois de registrados, em informações científicas e técnicas (LE COADIC, 2004, p.26).

O aumento significativo da informação em formato eletrônico, tanto em quantidade como em qualidade, vem transformando a recuperação e a localização da informação em um grande desafio. O advento da Internet e da informação em hipertexto torna este cenário mais catastrófico ainda, pela desorganização da informação e pela demanda por busca, verificação, recuperação e análise de documentos eletrônicos.

Os sistemas de construção e organização do conhecimento estão integrados ao desenvolvimento econômico e social, de forma indissolúvel, onde

a industrialização passa pela ciência e a ciência passa pela industrialização (LE COADIC, 2004, p.27).

As comunidades científicas, como atores da construção do conhecimento, surgem como agentes em busca de identificar formas e modelos para categorizar e localizar a informação como uma ponte entre as necessidades dos usuários e o material disponível.

SILVA (2002) explica que “a produção do conhecimento científico é motivada pelo desejo de alcançar resultados positivos, resultados publicáveis (...) resultados que comprovem as hipóteses levantadas.” Ela acrescenta ainda que “o desejo de alcançar resultados publicáveis é motivado pela necessidade de reconhecimento no meio científico e pela necessidade de obtenção de recursos nas agências estatais de financiamento de pesquisa”.

A necessidade da automação da informação, por meio de técnicas de indexação, visando a classificação, categorização e gerenciamento do conteúdo levantou barreiras na comunidade científica, particularmente entre os profissionais da área de biblioteconomia e cientistas da informação. Eles não chegavam a um consenso quanto às melhores ferramentas, metodologias, estratégias e padrões para construir a informação e facilitar o seu uso diante do constante desafio para organizar e construir o conhecimento em um momento em que o volume de informação cresce geometricamente (HLAVA, 2002).

Em meio a essas discussões entre *indexings*, *nodes*, *keywords*, *descriptors*, *thesauri* e *taxonomies*, o uso da lingüística computacional e da inteligência artificial surgiram para compor ao lado dos metadados e ontologias novas formas de facilitar a indexação e classificação da informação.

Os metadados surgiram como uma necessidade de descrever os dados com um conjunto mínimo de informações para identificar um recurso, incluindo seu contexto, qualidade condições dos dados, conforme IANNELA e WAUGH (1997 *apud* SENSO; PIÑERO, 2003).

Dessa metodologia, aos dados e objetos digitais foi proporcionado tratamento às suas estruturas, considerando-os não apenas destinados à reprodução da informação objetiva, mas surgindo como importante solução para a problemática do crescimento quantitativo da informação e a facilidade de acesso para sua potencial recuperação posterior.

Os princípios das ontologias, sob a ótica do desenvolvimento de sistemas baseados em conhecimento (*Knowledge-Based Systems, KBS*), vêm contribuindo de forma expressiva com as pesquisas de modelagem cognitiva de conhecimentos.

Ao considerar a faceta semântica da representação dos dados, dos seres, dos entes, dos objetos digitais e de tudo aquilo que se representa como informação, contemplando as teorias dos conceitos, as classificações e a modelagem cognitiva de conhecimentos.

Essa modelagem é facilitada pelas inúmeras possibilidades que as ontologias permitem ao especificar explicitamente as áreas de conhecimento de um domínio. A formalização da especificação do domínio reduz inconsistências, ambigüidades e contradições.

Dentre várias técnicas, a ontologia tem se destacado na organização do conteúdo das fontes de dados, onde especialistas de uma área de domínio do conhecimento criam ontologias e definem as regras que regulam a combinação entre termos e relações sobre os objetos desse domínio.

No processo de construção da informação, metadados e ontologias também contribuem para afastar o *amadorismo científico* – esforços isolados de trabalho e auxiliar as *ciências organizadas* e *ciência acadêmica*, lançando fundamentos para que a comunidade científica continue a contribuir para o desenvolvimento econômico e social da sociedade.

### **3.2 Metadados e Ontologias na comunicação da informação**

Primeiramente, para se entender a importância da comunicação da informação, diante do binômio – metadados e ontologia – é necessário saber de que tipo de comunicação está se tratando.

Etimologicamente, a palavra vem do latim *communicatione* (a raiz *communis* significa comum), que é entendida como ato de tornar comum, fazer saber ou partilhar. No contexto atual, comunicação é traduzida como “ato ou efeito de emitir, transmitir e receber mensagens por meio de métodos e/ou processos convencionados”. (FERREIRA, 2000).

Para a Semiótica, ciência que tem por objeto de investigação todas as linguagens possíveis, a comunicação é vista como a transmissão de uma mensagem por meio de signos, dentro de uma linguagem verbal ou não-verbal. (SANTAELLA, 1983).

Na concepção de LE COADIC (2004), a comunicação (leitura, escrita e contato) faz parte de um dos três processos<sup>2</sup> da Ciência da Informação (CI); processos esses que se sucedem e se realimentam formando o “ciclo da informação”. Dentro desse contexto, a comunicação é o meio responsável pelo contato direto e/ou indireto entre cientistas. O objetivo é o compartilhamento de informações a respeito de trabalhos, pesquisas, teses, que estão em fase de desenvolvimento.

O ato comunicacional tem, também, o papel de difundir o conhecimento científico; podendo ser disseminado por meio oral/informal (conferências, seminários, palestras, etc.) ou escrita/formal (revistas, livros, etc.), além de promulgar a ciência (LE COADIC, 2004).

A troca de informação entre pesquisadores, durante a pré/pós-produção de um trabalho científico, torna-se assaz importante porque permite o debate, a crítica e a construção do objeto de modo colaborativo. “Nas ciências físicas, estima-se em 40% o tempo dedicado a atividades de comunicação” (LE COADIC, 2004).

A comunicação, como mostrado, é um dos propósitos da CI para facilitar a transferência de informações entre as pessoas (BELKIN; ROBERTSON, 1976 *apud* SARACEVIC, 1996), ou como afirma SARACEVIC (1996):

A Ciência da Informação é um campo dedicado às questões científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação.

Com o aumento das pesquisas científicas e o avanço tecnológico no período que precedeu a Segunda Guerra Mundial, propiciou o surgimento do primeiro computador, o Enniac. Anos mais tarde, em 1991, Berners-Lee criaria a *World Wild Web* (WWW).

“O advento da WWW trouxe consigo o crescimento exponencial da quantidade de documentos registrados na Internet, o que motivou a

---

<sup>2</sup> Os três processos, segundo Le Coadic (1996), são: construção, comunicação e uso.

implementação de outros serviços de organização e de recuperação de informações.” (FEITOSA, 2006).

Além da necessidade de organização de conteúdos na *Web*, da recuperação de informações, também se manifestou a carência por novas formas de comunicação.

A solução para esses problemas, advindos com a WWW, foi a criação de ferramentas que possam auxiliar as pessoas, instituições e demais usuários da rede de computadores. Entre as soluções estão os metadados e as ontologias.

Os metadados por serem estruturados podem facilitar o processo de comunicação da informação e trazer benefícios, tanto para internautas, como para pesquisadores, empresas, etc., como por exemplo:

a) formar base para o desenvolvimento de sistemas de busca mais avançados; b) podem ser convertidos para outros formatos, de modo que possam interoperar com diferentes protocolos de busca e recuperação; c) [...] pensando-se em recuperação automatizada, torna-se mais fácil a extração de conteúdo de base de dados de descrição de recursos, do que de arquivos textuais em formato HTML (...) (DAY, 2000 apud FEITOSA, 2006)

A questão da escolha de um formato de metadados é assaz importante, pois cada um dos padrões tem formas diferentes de especificação e podem utilizar diferentes linguagens de representação. Hoje, o formato Dublin Core é um dos mais importantes padrões para metadados por sua utilidade, flexibilidade e interoperabilidade com outros sistemas.

Os metadados estão diretamente relacionados com o processo de comunicação da informação, visto que promovem uma padronização da *Web*, facilitam a busca e a recuperação de informações.

Por outro lado, as ontologias surgiram com a proposta de definir “os termos básicos e as relações compreendendo o vocabulário de uma área tópico (...)” (NECHES, 91 *apud* FEITOSA, 2006), ou seja, ela estabelece uma terminologia padrão, evitando as polissemias, as ambigüidades, com o intuito de facilitar a comunicação e o compartilhamento de informação e conhecimento entre pesquisadores, profissionais e cientistas.

As aplicações das ontologias são as mais diversas, como é o caso de uma ontologia elaborada com linguagem neutra, a Ontolíngua, que pode ser traduzida para qualquer linguagem e ter várias utilidades. Ontologias tam-



bém são utilizadas em projetos de domínios como gestão do conhecimento, comércio eletrônico, processamento de linguagens naturais, recuperação da informação na *Web*, de cunho educacional, entre outros.

USCHOLD e GRUNINGER (1996) ressaltam três importantes fatores do uso de ontologias; a) *comunicação* entre indivíduos e organizações por meio da instituição de um vocabulário comum e a integração de diferentes perspectivas dos usuários; b) *interoperabilidade* entre sistemas, utilizando ontologias como uma língua mediadora para troca de informação entre sistemas diversos; c) *engenharia de sistemas*, com a intenção de facilitar o reuso de representações sobre o entendimento básico de entidades, processos, relações ou propriedades.

Ambos os sistemas buscam facilitar a transmissão (emissão e recepção) do conhecimento, por meio do acesso rápido e preciso da informação via Internet.

### **3.3 Metadados e Ontologias no uso da informação**

Uma preocupação constante no processo de utilização da informação diz respeito à sua disponibilidade, facilidade de acesso (universalidade) e uso adequado de ferramentas de recuperação da informação como resultado de pesquisas científicas ou como parte da herança cultural para a sociedade.

Para LE COADIC (2002),

“Usar informação é trabalhar com a matéria informação para obter um efeito que satisfaça a necessidade de informação. Utilizar um produto de informação é empregar tal objeto para obter, igualmente, um efeito que satisfaça a uma necessidade de informação, que esse objeto subsista (fala-se então de utilização), modifique-se (uso) ou desapareça (consumo)”.

Já DAVENPORT (1996 *apud* COHEN, 2002) estrutura a gestão da informação em quatro etapas: a) determinação de exigências; b) obtenção; c) distribuição; d) utilização. Afirmando que "(...) o uso da informação é algo bastante pessoal" e que é "(...) relativamente fácil estimar esse uso". Concluindo que a gestão da informação preocupa-se com o indivíduo, o que este precisa e como utiliza.

Diante deste contexto, as contribuições do binômio metadados-ontologias para a CI podem ser identificadas a partir da construção de modelos e iniciativas que impactam a disseminação e o uso da informação tanto em alcance (profundidade) quanto em qualidade em prol do usuário final.

A utilização dos metadados no âmbito da CI atende às demandas específicas de usuários, de acordo com suas próprias necessidades, variando em conteúdo e sintaxe, conforme os padrões requeridos por grupos de pesquisas.

Na área de representação do conhecimento e reuso da informação, as ontologias contribuem com seu alto nível de abstração, permitindo a mudança de níveis para níveis, conforme o domínio e o propósito a que se destina.

Os sistemas de busca na internet, a análise, a recuperação e verificação de informação têm sido bem sucedidos com a utilização de ferramentas baseadas na utilização de ontologias e metadados.

A conceitualização formal das gramáticas e vocabulários das ontologias tem proporcionado significativo avanço entre a integração de sistemas, reduzindo ou eliminando a incompatibilidade de abstração dos dados.

Os aspectos colaborativos e de interoperação oferecidos pelas ontologias contribuem para derrubar as barreiras que dificultam o acesso pela falta de identificação precisa e exaustiva de fontes ou documentos com atendem as necessidades de informação de usuários.

Propiciar acesso eficiente à informação e ao conhecimento pressupõe que estes sejam organizados através de sistemas de informação e de recuperação da informação, onde os mecanismos de buscas possuam conhecimento especializado e experiência suficiente para interagir com o usuário.

Ontologias e metadados auxiliam o compartilhamento das informações e indicam padrões para o desenvolvimento de sistemas de informação melhores.

Outro aspecto relevante trazido pelo uso de ontologias está na possibilidade de reutilização do conhecimento, trazendo como vantagens a confiabilidade pelo tratamento semântico da informação e pela extração de informações a partir do contexto do domínio.

Os mecanismos de busca, baseados em marcações semânticas, contam com melhores processos de indexação e de recuperação da informação.

O emprego de metadados e ontologias abrem novos caminhos ao desenvolvimento de sistemas de informações alinhados intuitiva e coerentemente ao funcionamento cognitivo do pensamento humano.

Novas metodologias podem ser desenvolvidas e automatizadas com a utilização das ontologias e dos metadados em diversas áreas do conhecimento, para a criação de vocabulários controlados e *thesauri* a partir do uso da semântica genérica e formal da informação. Da mesma maneira, os metadados e ontologias permitem o emprego de indexação automática de documentos (objetos digitais), podendo esta ser obtida pela classificação automática ou semi-automática desses documentos.

As TIC proporcionam vislumbrar novos cenários e possibilidades na recuperação da informação e na interoperabilidade entre sistemas, contribuindo efetivamente com a gestão do conhecimento organizacional, por meio da criação de ontologias próprias e pela padronização dos metadados no mapeamento do conhecimento intelectual da organização.

#### **4. CONCLUSÃO**

Este artigo procurou abordar, de forma sucinta, as contribuições do binômio metadados-ontologias à Ciência da Informação, a partir da análise do referencial teórico do que seja Ciência, Ciência da Informação, metadados e ontologias.

Como ferramentas de apoio à Ciência da Informação, o binômio tem importante parcela na contribuição da produção, organização, classificação, transmissão, utilização e descrição da informação, formando um conjunto de conhecimentos, atendendo as demandas da comunicação humana.

O assunto é vasto e não se esgota, na medida em que a Ciência da Informação, como Ciência, é recente, é marcada pela interdisciplinaridade e possui seu conteúdo permeado pelas ciências matemáticas, sociais e humanas.

As TIC têm contribuído para que a informação transite com mais propriedade do suporte papel para o universo eletrônico, armazenando, transmitindo e disponibilizando todo o tipo de informação no formato eletrônico digital.

O nível de maturidade da pesquisa científica cresce e evolui na medida em que novas técnicas, estratégias e metodologias são agregadas às atividades de criação, armazenamento, comunicação e uso da informação.

No entanto, o principal desafio da Ciência da Informação continua sendo o de tornar acessível ao usuário o conhecimento crescente. As TIC e suas ferramentas são importantes aliados nessa tarefa de oferecer a melhor informação disponível à sociedade.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, Maurício; BAX, Marcelo. **Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção.** In: ENANCIB, 4, 2003, Belo Horizonte. Anais do Enancib. Belo Horizonte ECI/UFMG, 2003b. Disponível em: [http://www.eci.ufmb.br/mba/text/artigo\\_ontologia\\_pesquisa\\_sub2\\_WEB.pdf](http://www.eci.ufmb.br/mba/text/artigo_ontologia_pesquisa_sub2_WEB.pdf). Acesso em 03/12/2006.

ARAUJO, Carlos Alberto Ávila. **A ciência da informação como ciência social.** Ciência da Informação, Brasília, vol.32 no.3, 2003.

BERNERS-LEE, Tim. **Metadata architecture: documents, metadata and link**. 1997. Disponível em: <http://www.w3.org/DesignIssues/Metadata.html>. Acesso em 02/12/2006.

CHAUÍ, Marilena. **Convite à Filosofia**. São Paulo: Ed. Ática, 2000.

COHEN, Max F.. **Some aspects of information use in the information economy**. Ciência da Informação, Brasília, v. 31, n. 3, 2002. Disponível em: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S010019652002000300003&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010019652002000300003&lng=en&nrm=iso). Acesso em 02/12/2006.

COSTA, Sely Maria de Souza. **Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica**. Ciência da Informação, Brasília, Vol. 35, No. 2 (2006). Disponível em: <http://www.ibict.br/cionline/viewarticle.php?id=926>. Acesso em 02/12/2006.

FEITOSA, Ailton. **Organização da informação na web: das tags à web semântica**. Brasília: Thesaurus, 2006.

FERREIRA, Aurélio. **Dicionário Aurélio Século XXI**. Rio de Janeiro: Nova Fronteira, 2000.

FREIRE, Gustavo Henrique de A. **Ciência da Informação: temática, histórias e fundamentos**. [Perspectivas em Ciência da Informação, Vol. 11, N° 1, 2006](#). Disponível em: <http://www.eci.ufmg.br/pcionline/viewarticle.php?id=406>. Acesso em 02/12/2006.

GILLILAND-SWETLAND, A. J. Defining metadata. *In*: \_\_\_\_\_. Introduction to metadata: pathways to digital information. Los Angeles: Getty Information Institute, 1998, p. 1-8.

GÓMEZ PÉREZ, et al., (2004) apud. ABECKER, A., A. BERNARDI, K. HINKELMANN, O. HÜHN, e M. SINTEK. **Toward a Technology for Organizational Memories**. *IEEE Intelligent Systems* v. 13, n. 3, pp. 40-48.

GRUBER, T. R. **Toward principles for the design of ontologies used for knowledge sharing**. International Journal of Human-Computer Studies, v. 43, 1995.

GONTIJO, Silvana. **O livro de ouro da comunicação**. Rio de Janeiro: Ediouro, 2004.

HLAVA, M. M. **Automatic Indexing: A Matter of Degree**. Bulletin of the American Society for Information Science and Technology; Oct/Nov 2002.

IANNELA, R.; WAUGH, A. **Metadata: enabling the Internet**. Disponível em: <http://archive.dstc.edu.au/RDU/reports/CAUSE97>. Acesso em: 10/12/2006.

IKEMATU, Ricardo. **Gestão de Metadados: Sua Evolução na Tecnologia da Informação**. DataGramZero - Revista de Ciência da Informação - v.2 n.6 dez/2001.

LE COADIC, Yves-François. **A ciência da informação**. Brasília: Briquet de Lemos, 2004.

LUNGARZO, Carlos. **O que é ciência**. São Paulo: Ed. Brasiliense, 1989.

MILSTEAD, J.; FELDMAN, S. Metadata: **cataloguing by any other name**. *Online*, n. 1, p. 25-31, 1999.

MIRANDA, Antonio; SIMEÃO, Elmira (Orgs.). **Informação e tecnologia: conceitos e recortes**. Brasília: UnB, Departamento de Ciência da Informação e Documentação, 2005.

MOSTAFA, S. P. **Ciência da Informação: uma ciência, uma revista**. Ciência da Informação, Brasília, v. 25, n. 3, 1995.

NEVES, Dulce Amélia de B. **Aspectos metacognitivos na leitura do indexador**. 2004. Tese (Doutorado em Ciência da Informação)- UFMG, Belo Horizonte, 2004.

NOVELLO, Taisa Carla. **Ontologias, Sistemas baseados em Conhecimento e Modelos de Banco de Dados**. Disponível em [http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo\\_taisa.pdf](http://www.inf.ufrgs.br/~clesio/cmp151/cmp15120021/artigo_taisa.pdf). Acesso em: 6/12/2006.

ORTIZ-REPISO, JIMÉNEZ, V. **Nuevas perspectivas para la catalogación: metadatos versus MARC**. Revista Española de Documentación Científica, v. 22, n. 2, p. 198-219, 1999.

PACHECO, Roberto Carlos dos Santos; KERN, Vinícius Medina. **A common ontology for information and knowledge integration in science and technology**. Ciência da Informação, Brasília, 2001.

SANTAELLA, Lúcia. **O que é Semiótica**. São Paulo: Brasiliense, 1983.

SARACEVIC, T. **Ciência da Informação: origem, evolução e relações**. Perspectivas em Ciência da Informação, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SENSO, José A.; PIÑERO, Antonio de la Rosa. **El concepto de metadato. Algo más que descripción de recursos electrónicos**. Ciência da Informação; Vol. 32, No. 2 (2003). Disponível em: <http://www.ibict.br/cionline/viewarticle.php?id=159>. Acesso em 30/11/2006.

SOWA, J. F. **Building, sharing and merging ontologies**. Tutorial. [S. 1. : s. n.], 1999. Disponível em: <http://users.bestweb.net/~sowa/ontology/ontoshar.htm>. Acesso em: 30/11/2006.

SILVA, L. E. **A construção do conhecimento científico: o processo, a atividade e a comunicação científica em um laboratório de pesquisa.** Perspectivas em Ciência da Informação, Belo Horizonte, v. 7, n. 2, p. 109-125, jul./dez. 2002.

USCHOLD, M.; GRUNINGER, M. **Ontologies: principles, methods, and applications.** Edinburg: The University of Edinburg, 1996. Disponível em:<http://www.aiai.ed.ac.uk/project/pub/documents/1996/96-ker-intro-ontologies.ps>. Acesso em: 09/11/2006.

VICKERY, D.C. **Knowledge representation: a brief review.** *Journal of documentation*, v.42, n.3, sept. 1986. p. 145-59.

WERSIG, G. **Information science: the study of postmodern knowledge usage.** *Information Processing & Management*, v. 29, n. 2, 1993.

**Contribuições do binômio metadados-ontologias para a Ciência da Informação - uma Análise Sucinta.**

Autores: Ly Freitas Filho, Robson do Nascimento, Rodrigo Nunes e Shamash Vicentini

Disciplina: Tecnologia da Informação

Prof<sup>a</sup> Dra Marisa Brascher Basílio Medeiros

Dez/2006