



INFORMAÇÃO, FERRAMENTAS ONTOLÓGICAS E REDES SOCIAIS: A INTEROPERABILIDADE NA CONSTRUÇÃO DE TESAUROS E ONTOLOGIAS

Maria Aparecida MOURA

Escola de Ciência da Informação da Universidade Federal de Minas Gerais (ECI/UFMG)- Brasil
mamoura@eci.ufmg.br

RESUMO

As mudanças tecnológicas decorrentes da digitalização da informação provocaram inúmeras alterações na produção teórica e nos experimentos empíricos da Ciência da Informação. Nesse contexto, a compreensão acerca do desempenho dos usuários e de suas redes sociais tornou-se fundamental na estruturação de instrumentos de organização da informação. A digitalização das informações, a redução dos rituais sincrônicos e a produção da informação sob demanda provocaram mudanças radicais na disseminação da informação e, conseqüentemente, exigem que sejam realizadas alterações nas metodologias de elaboração dos instrumentos verbais de representação da informação. Analisa-se, nesse trabalho, o surgimento dos novos instrumentos de representação da informação em ambientes digitais. Apresenta-se o conceito de ferramentas ontológicas. Analisam-se os impactos decorrentes da interoperabilidade entre as ferramentas ontológicas disponíveis na *Web* e as redes sociais a elas incorporadas. Discute-se a nova configuração das garantias (literária, estrutural e de uso) no processo de construção de linguagens de indexação em ambientes digitais. Sugere-se, em termos metodológicos, a incorporação das novas mediações informacionais no processo de construção de linguagens de indexação.

ABSTRACT

The technological changes resulting from information digitalization have triggered innumerable alterations in the theoretical production and in the empiric experiments of the Information Science. In this context, the comprehension of the users performance and their social nets has become fundamental in structuring instruments of information organization. The digitalization of the information, the reduction of synchronic rituals and the production of information on demand have set off radical changes in the dissemination of information and consequently, require that alterations in the methodologies of elaboration of verbal instruments of information representation be carried out. In this work we analyze the appearance of new instruments of information



representation in digital environments. We present the concept of ontological tools. The impacts resulting from the interoperability among the ontological tools available in the web and social nets incorporated to them, are also analyzed. We discuss the new configuration of warranties (literary, structural and of use) in the process of constructing indexing languages in digital environments. In methodological terms, the incorporation of the new informational means in the process of construction of the indexing languages was suggested.

PALAVRAS-CHAVE

Ferramentas ontológicas, redes sociais, tesouro-metodologia, linguagens de indexação.

KEY WORDS

ontological tools, social nets, thesaurus – methodology, indexing language



1 INTRODUÇÃO

A Ciência da Informação é uma área de conhecimento que privilegia o estudo dos fenômenos informacionais em seus diversos aspectos, buscando compreender e acompanhar os seus desdobramentos na futura disponibilização em sistemas de informação. Trata-se de um campo, visivelmente atravessado pelo imperativo tecnológico. O campo tem por objetivo compreender a ação manifesta nas relações humanas mediadas pela informação e os desdobramentos dessa ação. Busca para tanto compreender, do ponto de vista do sujeito, os aspectos intelectuais, sociais e técnicos envolvidos na ação de produzir, sistematizar, organizar, disseminar e recuperar informação. Tais informações são sustentadas organicamente por ferramentas, objetos, processos organizacionais e manifestações sociais e culturais.

Nos últimos anos as mudanças tecnológicas decorrentes da digitalização da informação provocaram inúmeras alterações na produção teórica e nos experimentos empíricos da Ciência da Informação. O campo opera em estreita sintonia com a idéia de relativização da noção de lugar para a ocorrência dos eventos sociais e oferece a base, através do aperfeiçoamento da fluidez informacional, para a implementação da chamada Sociedade da Informação, em que o tempo é também relativizado.

Nesse sentido, a CI é responsável pelo aperfeiçoamento técnico da fluidez da informação, através de procedimentos, técnicas e tecnologias que buscam promover a interação entre usuários e informação, considerando a infra-estrutura digital e seus impactos na disseminação informacional.

No contexto de inovações a informação é compreendida como um signo que se atualiza na interface com o sujeito, “*algo que se forma sem cessar*”.

Atualmente, a compreensão acerca do desempenho dos usuários e de suas redes sociais tornou-se fundamental na estruturação de instrumentos de organização da informação. A digitalização das informações, a redução dos rituais sincrônicos e a produção da informação sob demanda provocaram mudanças radicais na disseminação da informação e, conseqüentemente, exigem que sejam realizadas alterações nas metodologias de elaboração dos instrumentos verbais de representação da informação. Hoje, verificam-se grandes transformações nas formas de agregação e de arbitragem em torno da informação e do conhecimento.

Nesse trabalho, apresenta-se inicialmente uma discussão teórica sobre a articulação entre a Web 2.0, a interoperabilidade e as ferramentas ontológicas de representação e recuperação da informação; em seguida apresenta-se um estudo da ferramenta Google Trends e sua potencialidade nos processos de elaboração de



linguagens de indexação e bases terminológicas. Posteriormente realiza-se uma discussão de carácter teórico sobre as comunidades de práticas e as redes sociais considerando o papel das mesmas na dinamização e validação da terminologia adotada nos domínios de conhecimento e formalizada em repositórios on-line de percursos individuais e coletivos de informação. Por fim, relata-se nesse trabalho, como exemplo da viabilidade da abordagem, o desenvolvimento simultâneo de uma ontologia e um de tesouro voltados à representação de discursos académicos na área relacionada à mundialização e ao desenvolvimento sustentável no espaço de conhecimento digital denominado *Mondialisation et Développement Durable* (http://semioweb.msh-paris.fr/corpus/mdc/FR/Prog_missions.asp) e apresentam-se as considerações finais.

Para o desenvolvimento do trabalho foram privilegiadas as representações dos atores sociais envolvidos com a temática nos inúmeros dispositivos digitais disponíveis na Web, com vistas a compreender como as novas dinâmicas de produção e difusão do conhecimento impactam as garantias canonicamente estabelecidas no domínio da construção de linguagens de indexação, a saber; garantia de uso, garantia estrutural e garantia literária. Que atualizações podem ser incorporadas? É possível falar que, nos ambientes digitais, tais categorias encontra-se justapostas?

2 CIÊNCIAS DA INFORMAÇÃO E AS MUDANÇAS NA MATERIALIDADE INFORMACIONAL

A explosão digital e a conseqüente disponibilização de informação em rede passou a exigir soluções mais sofisticadas, ágeis e específicas no trato dessa documentação. Muitas iniciativas foram tomadas, mas ainda se verifica um desequilíbrio nas articulações teóricas e nos conseqüentes experimentos. Ora privilegia-se o tecnológico, ora privilegia-se a experimentação decorrente da experiência e em outros momentos reitera-se a distinção teórica envolvida na questão.

Assim, tem surgido uma série de estudos em que parte das soluções representa a o retorno aos esforços de pioneiros como Peter Luhn (1896-1964) que contribuiu significativamente para a compreensão do papel da computadorização da informação nos processos de recuperação da informação.

Na atualidade, o debate acerca da noção de documento expandiu e, de acordo com a rede pluridisciplinar de pesquisadores internacionais reunidos em torno do tema (<http://rtpc-doc.enssib.fr>), abriga três dimensões: o documento como forma, como signo e como meio.



O documento como forma incorpora a idéia de um objeto, material ou imaterial, que precisa ser manipulado, utilizado e conservado. Essa abordagem considera a legibilidade do documento como elemento central.

A abordagem relativa ao significado, enfatiza a interpretação e se orienta para o documento como portador de sentido, dotado de uma intencionalidade indissociável do sujeito que o produz e utiliza. Considera ainda que o documento é integrante de um sistema documentário ou de conhecimentos.

A abordagem focada na mediação enfatiza a sociabilidade e analisa o estatuto do documento nas relações sociais. Nessa perspectiva, o documento é um traço de comunicação construído ou encontrado e representa um elemento do sistema identitário e um vetor de poder.

De acordo, com o coletivo Pedauque (2003: p. 26) as transformações radicais pelas quais passou o documento, da tecnologia impressa ao digital, envolve sempre a necessidade de se refletir sobre o processo evolutivo que o acompanha sendo necessário pensá-lo em termos dos contratos de leitura a ele vinculado. Segundo os autores o documento é um contrato entre os homens que funda sua humanidade e a sua capacidade de viver juntos, seja pelas qualidades antropológicas (legibilidade - percepção), seja pelo caráter intelectual (compreensão-assimilação) ou pelo aspecto social envolvido (sociabilidade-integração). Nesse contexto, o aspecto digital se apresenta como uma modalidade de multiplicação desse contrato.

O impacto sofrido pela informação em virtude do crescente processo de digitalização também é parte integrante desse debate. Verifica-se que a transformação da informação analógica em digital, permitiu a ampliação da geração, estocagem e processamento dos dados em tempo real. E, como consequência, significou ampliar a noção de informação com a qual o campo trabalhava até então. Se antes as ações teóricas e experimentos práticos tinham como foco o aspecto material da informação, hoje, devido aos intercâmbios entre os papéis sociais de autores e receptores de informação, é possível afirmar que o foco migrou para a recuperação de trechos informacionais. E devido à contigüidade presente nos acervos digitalmente dispostos em rede, o usuário de informação está interessado em encontrar um trecho, ou um segmento de um documento, não necessariamente, um documento em sua manifestação física.



3 WEB 2.0 E A INTEROPERABILIDADE ENTRE AS FERRAMENTAS ONTOLÓGICAS

O aprofundamento da discussão do carácter interativo da WEB e da efetiva presença dos usuários de informação no ambiente digital criou os elementos para o surgimento da Web 2.0. O termo foi proposto entre 2003 e 2004 e designa a segunda geração de produtos e serviços desenvolvidos no contexto da Internet. Pensada inicialmente como uma estratégia de *marketing* segmentado, a Web 2.0 privilegiou a efetiva colaboração do usuário no processo de organização e personalização da informação e proporcionou, desse modo, a alteração das lógicas de organização da informação, agora não mais centralizadas no papel do mediador da informação.

Assim, as tecnologias construídas no contexto da Web 2.0, como o wiki, os blogs, o Consumer-Generated Media (CGM), o Really Simple Syndication (RSS) e as folksonomias apresentam a oportunidade de compartilhar a informação e ampliar o processo interativo entre usuários e informações.

A estrutura da Web semântica envolvida na constituição da Web 2.0 requer três camadas distintas: a camada lógica, a camada ontológica e a camada esquema. A camada lógica envolve a interface de conversação, os motores de busca e o motor heurístico. A camada ontológica envolve as representações semânticas presentes em estruturas taxonômicas. E a camada esquema estabelece o provimento das definições do documento e dos significados a eles associados.

Nesse contexto, a interoperabilidade é considerada como a capacidade de sistemas autônomos (informatizados ou não) comunicarem de modo transparente entre si, devido à adoção de padrões comuns e protocolos que permitem o uso compartilhado de informações.

A interoperabilidade é desenvolvida no contexto digital tomando como referência o estabelecimento da representação estrutural, sintática, semântica e lógica, expressas, de acordo com Ribeiro (2008) nas seguintes camadas:

- **Unicode / URI** – responsável pelo fornecimento da interoperabilidade no nível da codificação de caracteres, do endereçamento e da nomeação de recursos;
- **XML / Namespace / XML Schema** – responsável pela interoperabilidade em relação à sintaxe de descrição de recursos;
- **RDF / RDF Schema** responsável pelo fornecimento de um *framework* para representar a informação (metadados) sobre os recursos;
- **Ontologia** - responsável pela integração, com um nível mínimo de conflitos, da terminologia utilizada e as informações disponíveis.



- **Lógica** – é associada ao fornecimento de suportes para a descrição de regras que expressam as relações entre os conceitos em uma ontologia;
- As camadas de **Prova e Confiança** – Disponibilizam o suporte para a execução das regras e para a avaliação da correção e da confiabilidade da execução.

4 AS FERRAMENTAS ONTOLÓGICAS

O ambiente de inovações composto por fluxo informacional, trabalho colaborativo em rede e a interoperabilidade entre sistemas, abriu maior possibilidade para a participação dos usuários na descrição da informação utilizada em seus percursos informacionais. Além disso, tornou viável o câmbio entre as terminologias sugeridas pelos diversos atores através de antigos e novos arranjos terminológicos que compõem as contemporâneas ferramentas ontológicas.

As ferramentas ontológicas são mediações de linguagem adotadas em sistemas de informação compostas por estruturas informacionais contextualizadas, derivadas de esquemas intelectuais mais complexos. Tais ferramentas são desenvolvidas sob pontos de vista e propósitos específicos e têm por objetivo orientar os sujeitos na compreensão acerca do conhecimento em áreas específicas, bem como apoiar a adoção consciente desses esquemas representacionais em sistemas de organização e recuperação da informação.

Inspirados no conceito filosófico de ontologia - categoria aristotélica que se refere ao estudo das propriedades mais gerais do ser em que se enfatiza sua natureza plena, integral e abrangente, em contraposição a uma visão parcial e particular do mesmo - os campos relacionados à organização e formalização da informação em ambientes digitais passaram a adotar essa terminologia. Desse modo, a denominação “ferramentas ontológicas” tem sido utilizada para designar os estudos conceituais específicos que visam caracterizar dada área de conhecimento a partir do mapeamento das suas categorias mais gerais. De posse desse amplo mapeamento conceitual têm sido desenvolvidos programas computacionais específicos, envolvendo inteligência artificial, com vista a oferecer um instrumento dinâmico de produção, organização e disseminação de conhecimento. Integram a categoria das ferramentas ontológicas as linguagens de indexação (verbais e simbólicas), as taxonomias, os mapas conceituais, as ontologias e mais contemporaneamente as *folksonomias*.

As linguagens de indexação são instrumentos auxiliares no processo de representação da informação e visam estabelecer uma equivalência aproximativa entre as linguagens do autor do documento, a linguagem de abordagem adotada pelo usuário da informação e aquela adotada pelo sistema de informação. Nesse sentido, busca



representar o assunto de maneira consistente permitindo a compatibilidade e o diálogo entre a linguagem do autor, do indexador e a do pesquisador.

A taxonomia é a ciência natural que se dedica à descrição, identificação e classificação dos organismos. Nos últimos anos, a Ciência da Computação tem nomeado as ações classificação de informação como taxonomias. Contudo, não há uma uniformidade nessa adoção. Por vezes, utiliza-se o método das Ciências Naturais para a determinação e construção das classificações em outros domínios do conhecimento e em outros momentos, incorpora-se simplesmente a denominação sem necessariamente, incorporar o princípio ordenador.

Os mapas conceituais ou (derivação conceitual) são esboços totalizadores de um determinado tópico temático que visam estabelecer a formalização computacional e o arranjo visual em termos da interface. No caso dos mapas conceituais está simultaneamente embutida uma visão de mundo classificadora e um esquema de visualização. No contexto da organização da informação esses mapas têm sido muito utilizados no processo de recuperação informação contextualizada. É possível encontrar softwares para a organização de mapas conceituais fixos hiperbólicos (mapas que, por serem mais amplos ou complexos e com infinitas relações, contam com a animação computacional que permite enfatizar, na consulta, um determinado estágio da categoria ou classe escolhida).

A ontologia informacional é a denominação dos softwares e bases terminológicas verbais, articuladas por infinitas relações, que visam representar contextos informacionais específicos. Dentre as principais vantagens das ontologias está a possibilidade de formalizar, no nível computacional, as relações entre os termos em função dos agentes ou grupos de agentes que dela fazem uso. Os principais elementos que uma ontologia descreve são os indivíduos, as classes, os atributos e os relacionamentos.

Atualmente as ontologias têm sido adotadas na identificação e formalização do conhecimento explícito e/ou tácito presentes nas organizações.

As *folksonomias* resultam do neologismo formado pela associação do termo *folk* (povo) e taxonomia. Essa ferramenta tem por objetivo gerar taxonomias que são utilizadas na categorização e recuperação de conteúdos na web, decorrentes da experiência dos usuários e de sua rede social na descrição dos recursos usando rótulos chamados *tags* (etiquetas). Os principais recursos descritos pelos usuários são os *bookmarks*, as fotografias on-line, os vídeos, os blogs, as referências bibliográficas, dentre outros. Consiste, portanto em uma forma de os próprios usuários disponibilizarem os rótulos adotados intuitivamente no compartilhamento, organização e na recuperação futura da informação.

As *folksonomias* permitiram, em certa medida, a realização de algo já assinalado por Lancaster nos 80, quando o mesmo enfatizou que o uso de linguagens naturais em sistemas de representação da informação deveria incluir as questões dos usuários. Naquela ocasião e ainda hoje, os principais problemas referentes à organização da informação em contextos automatizados eram a rapidez que as informações sofrem obsolescência, a escassez de tempo nas buscas retrospectivas, a falta de entendimento sobre o funcionamento das rotinas dos sistemas de informação (falta de padrão comum), a incompatibilidade entre o vocabulário de abordagem (usuário), o vocabulário utilizado no sistema de informação e a morosidade no processo de tratamento de novos documentos.

De posse das possibilidades tecnológicas atuais e em vista dos desafios anteriormente assinalados, sites como o Delicious (<http://www.delicious.com>), o Scivee (<http://www.scivee.tv>), o Twine (<http://www.twine.com>), dentre outros, representam a iniciativa de promover a formação de comunidades de cooperação segmentadas pelo uso das informações, com vistas a aprimorar tanto, o uso compartilhado da informação quanto o *marketing* dirigido às características específicas dos usuários, que se pode deprender inclusive pelo vocabulário de abordagem que os mesmos compartilham.

5 O GOOGLE TRENDS



Figura. 1 - Interface de consulta do Google Trends
(<http://www.google.com/Trends>)



O Google Trends é um aplicativo vinculado ao Google Labs e tem por objetivo avaliar, por meio de algoritmos, o interesse mundial em um determinado assunto. Além disso, o software mostra a frequência com que os temas têm surgido na WEB em escala temporal e espacial.

O algoritmo do Google opera através da análise de pesquisas feitas no motor de busca. O seu principal ganho é adotar metodologias que permitem maior segurança em relação ao dado, na medida em que oferece as médias para além do histórico de tráfego padrão.

O Google Trends analisa uma mostra de buscas na Web e calcula o volume de pesquisas que têm sido feitas no site com o tópico proposto em relação ao número total de pesquisas feitas no Google ao longo do tempo.

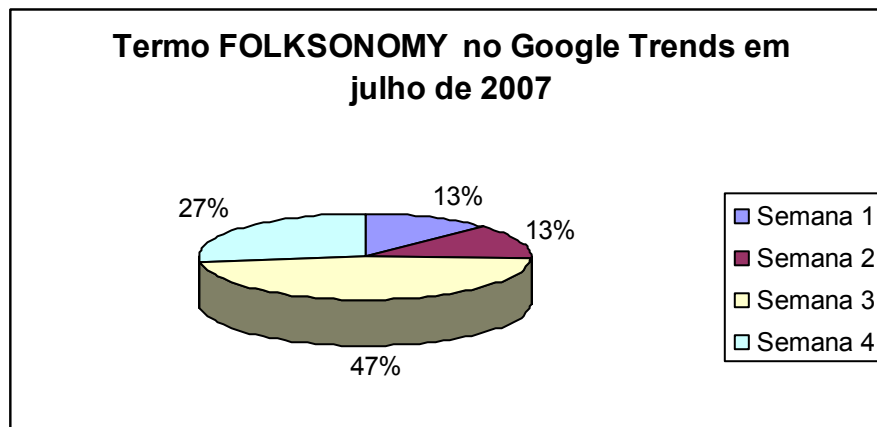
A partir da geração do gráfico no Google Trends é possível exportar os dados CVC para uma tabela eletrônica. Nas tabelas é possível acessar os percentuais de aparição do termo por semana e inferir a origem geográfica das buscas.

O arquivo inclui, juntamente com o índice de pesquisa, a margem de erro relativo para os dados obtidos. Desse modo, na planilha eletrônica é exibida uma coluna com o volume semanal e outra correspondente às margens de erro. Esta informação pode ser usada para calcular o intervalo de confiabilidade para um conjunto de dados.

Os volumes de dados são calculados considerando o modo relativo e o fixo. No modo relativo, os dados são escalonados considerando o tráfego do tópico pesquisado em médio prazo (representado como 1.0) durante o período determinado pelo pesquisador. Já no modo fixo, os dados são dimensionados pela média de tráfego para o tópico pesquisado no decurso de um ponto fixo (janeiro de 2004).

Nesse caso, ao realizarmos uma pesquisa com o termo *folksonomy* em julho de 2007, verifica-se uma tendência com volume total de 7,44, distribuída em 0,96 na primeira semana, 0,94 na segunda, 3,54 na terceira e 2,0 na última.

Tabela. 1 – Termo *Folksonomy* no Google Trends



Ao realizarmos uma pesquisa no mesmo campo temático para *social bookmarks*, verifica o início das manifestações de interesse no tema em agosto de 2007 na terceira e quarta semanas de agosto com um volume de 5,65 e 5,35 respectivamente. Entre outubro de 2007 e julho de 2008 o termo ganha estabilidade e presença no contexto digital.

Tabela . 2 - Estabilização do termo *Social bookmarks* captado pelo Google Trends (outubro 2007 – julho 2008)

Estabilização do termo <i>Social bookmarks</i> captado pelo Google Trends (outubro 2007 – julho 2008)				
Meses	Semanas			
	1	2	3	4
outubro	4,75	4,8	4,35	4,35
novembro	4,35	5	5,6	4,9
dezembro	4,4	5	5,3	5,3
Janeiro	5,15	6,35	6,15	5,65
fevereiro	5,3	6,45	6,1	5,15
março	3,85	5,5	5,3	4,8
abril	6,1	4,9	5,8	5,65
maio	5,2	4,55	4,35	5,3
junho	4,4	4,7	4,45	5,45
julho	6,7	5,9	5,2	6,35

Tomando esses dados como referência, acredita-se que o movimento, capturado pelos algoritmos evidencia a configuração sobreposta das garantias de uso, literária e estrutural manifestas no contexto digital para o termo pesquisado. A sobreposição se



deve ao fato de no contexto digital estarem em ação múltiplos atores sociais, dentre os quais usuários, autores e gestores de informação.

Evidentemente, ao apresentarmos essa ponderação, não se trata de afirmar que o *Google Trends* poderá sozinho resolver o problema das garantias preconizadas no processo de composição das linguagens de indexação. Entretanto, acredita-se que ele pode ser um instrumento auxiliar importante na medida em que tem a capacidade de evidenciar o nascedouro de um termo, a representação das necessidades informacionais dos sujeitos no espaço e no tempo e ainda, fornecer os indícios da obsolescência de um tópico no contexto digital.

De posse desse monitoramento terminológico de grande escala proporcionado pelo Google Trends, o elaborador da linguagem poderá, a partir dessas informações, obter os balizadores para a tomada de decisão sobre a inclusão de um termo em um tesouro ou ontologia e o nível de relação proposto na sintaxe para o mesmo, por exemplo.

6 COMUNIDADES DE PRÁTICA E AS REDES SOCIAIS

O funcionamento da Internet e os diversos produtos e serviços a ela vinculados alteraram significativamente os padrões de agregação social na contemporaneidade. A relativização das noções de tempo e espaço e a redução dos rituais sincrônicos abriram espaço para a mobilidade e o estabelecimento de comunidades não constrangidas pela dimensão geográfica e ocasionou também a implementação de novos padrões de cooperação. A difusão global de informações permitiu uma série de agregações que se constituem em torno do interesse informacional, tornado fluxo. É nesse contexto que surgem as comunidades virtuais, uma modalidade de agregação de sujeitos dispersos geograficamente em torno de interesses comuns.

De acordo com Bauman (2005, p. 17), podemos identificar dois tipos de comunidades, as vida e as de destino. Nelas os membros são marcados por uma ligação absoluta, ou “fundidos unicamente por idéias ou por uma variedade de princípios”.

Nesse ambiente de cooperação abrem-se também espaço para a articulação das comunidades de prática. As comunidades de prática são, de acordo com McDermott (1999),

Agrupamentos de pessoas que compartilham e aprendem uns com os outros por contato físico ou virtual, com um objetivo de resolver problemas, trocar experiências, desvelamentos, a construção de modelos padrões, técnicas ou



metodologias, tudo isso com previsão de considerar as melhores práticas.

Essas comunidades, consideradas territórios neutros das pressões sociais e da demanda por produtividade, devem possuir um domínio de atuação partilhado de forma colaborativa ou comunitária e compartilhar práticas comuns (experiências, problemas e soluções, ferramentas, vocabulários e metodologias).

Em sua origem, o termo foi proposto nos estudos de Jean Lave e de Etienne Wenger baseado em trabalhos desenvolvidos no final dos anos oitenta, quando investigaram a aprendizagem em diversos tipos de comunidades e o papel da participação periférica no fortalecimento das práticas partilhadas. De acordo com Lueg (2008), em comunidades de prática a integração a partir da periferia dos processos não implica em subalternização da participação, mas pode auxiliar ao novo membro na compreensão das formas e as dinâmicas de funcionamento do grupo.

As ações dessas comunidades envolvem simultaneamente o compartilhamento dos conhecimentos explícitos e tácitos. O conhecimento explícito constitui-se em um saber que pode ser formalizado em procedimentos comuns tais como vocabulários, conceitos e bases de conhecimento. O conhecimento tácito, por outro lado, é um conhecimento pessoal, de mais difícil formalização na medida em que agrega crenças, *savoir-faire*, histórias, anedotas e linguagens corporais, elementos esses que passam por sistemas de codificação e explicitação.

No caso específico das comunidades científicas, há em suas formações maior facilidade em trabalhar com o conhecimento explícito, contudo, o desafio contemporâneo é a criação de procedimentos que permitam a rápida incorporação e ou formalização do conhecimento tácito em explícito.

Nesse universo podem se identificar dois tipos de comunidades: as auto-reguladas e as patrocinadas.

As **comunidades auto-reguladas** são aquelas com alto grau de informalidade e descentralização, cujos principais focos de atenção são os interesses comuns de seus membros. Em geral, as trocas realizadas nesses contextos podem agregar valor às ações profissionais e intelectuais de seus participantes. De caráter resiliente, essas comunidades se conduzem pela dinâmica de interesse de seus membros e podem no curso de suas ações alterar seus objetivos iniciais ou mesmo assumirem a forma de uma comunidade patrocinada.



As **comunidades patrocinadas** são iniciativas mantidas ou apoiadas por organizações. Por receberem apoios institucionais o curso das ações é orientado a resultados mensuráveis, as responsabilidades são partilhadas e seus membros ocupam papéis específicos. Apesar disso, a possibilidade de auto-regulação, próprio às comunidades de práticas, as distingue de uma equipe de projetos comum.

Os principais dispositivos adotados na condução e formalização das comunidades de prática são os *groupewares*, os repositórios de dados e conhecimento, a gestão de documentos, os motores de busca, os agentes inteligentes, as *intranets* e os *websites*.

Parte integrante desse mesmo universo, as comunidades virtuais de prática (CoPS) são movimentos colaborativos na WEB que através de tecnologias síncronas e assíncronas ampliaram a comunicação, interação e a produção e o compartilhamento de informações entre sujeitos dispersos geograficamente em escala internacional. As principais dificuldades para o funcionamento de uma comunidade virtual de prática são: o nível de disciplina exigido, a natureza do conhecimento partilhado, a fluidez das relações que se estabelecem pelo meio digital e o estabelecimento da confiança mútua e o respeito aos acordos pactuados na comunidade.

Dado que a intensificação da comunicação é o principal elemento agregador desse tipo de comunidade, a linguagem exerce uma função central nesses contextos. A linguagem no âmbito desse trabalho é compreendida como mediações compartilhadas e tacitamente construídas com vistas a efetivar a expressão do pensamento humano.

As redes sociais, por seu turno, são agregações sociais organizadas em torno de temáticas e interesses específicos que partilham, produzem e disseminam conhecimento e informações. Essas redes atuam aos moldes dos colégios invisíveis e utilizam intensamente as tecnologias digitais em rede como um mecanismo de agregação e produção coletivas.

As redes possuem uma lógica de agregação que se pauta pela horizontalidade das relações sociais e, de acordo com Dias e Silveira (2007), caracterizam-se por processos produtivos e de experiência; poder e cultura; dinamismo e descentralização na tomada de decisão; autonomia dos membros, horizontalidade das relações e desconcentração do poder.

Além disso, elas incorporam, de acordo com Marteleto (2000) três dimensões fundamentais: a dimensão sócio-comunicacional, a lingüístico -discursiva e a de produção de sentidos.



A dimensão sócio-comunicacional envolve os elos, motivações e interações entre os atores sociais. A dimensão lingüística e discursiva incorpora os aspectos cognitivos e informacionais envolvidos no compartilhamento social. Finalmente, a dimensão de produção de sentidos explicita o fluxo e a dinâmica da ação colaborativa partilhada.

No ambiente digital os provedores de soluções colaborativas em rede logo trataram de programar funcionalidades aos softwares para que eles pudessem facilitar o nível de cooperação já, a partir do pacto de linguagem. Em decorrência disso, os membros das comunidades virtuais puderam constituir e instanciar no nível computacional os principais eixos temáticos desenvolvidos e expressos em linguagem.

Desnecessário dizer que tal movimento, tem causado enorme impacto sobre a organização da informação em ambientes digitais, sobretudo, porque se criou uma grande expectativa de que as redes e seus membros seriam eles próprios capazes de coordenar os fluxos informacionais e o aprovisionamento de soluções no âmbito da gestão dos sistemas de informação.

Nesse debate, as *folksonomias* ganharam status de instrumento de representação da informação apesar de elevarem a complexidade do tratamento da linguagem verbal em contextos de representação da informação. E, em um ambiente em que é crescente a presença dos usuários que conduzem suas pesquisas e de autores que disponibilizam conteúdos on-line sem a intervenção de mediadores, tornou-se evidente que o campo da organização da informação precisaria rever seus métodos com vistas a compreender esse movimento de transformação intelectual inevitável.

Atualmente essas redes têm funcionado em torno das trocas informacionais que se realizam em ambientes dotados de *folksonomias* que permitem, através da etiquetagem dos interesses informacionais, o compartilhamento e a validação de uma linguagem de referência. Evidentemente que o controle desse léxico ainda encontra-se atravessado pela intensificação das redes de significação que se estruturam no uso social da linguagem, mas acredita-se que a dinâmica estabelecida pode oferecer pistas, rumo à atualização da linguagem de representação informacional de modo mais ágil, conforme já salientado.



7 A EXPERIÊNCIA DE ESTRUTURAÇÃO DO TESAURO SOBRE MUNDIALIZAÇÃO, DIVERSIDADE VIVA E DESENVOLVIMENTO SUSTENTÁVEL: INTEROPERABILIDADE ENTRE FERRAMENTAS ONTOLÓGICAS E REDES SOCIAIS

De acordo com Svenonius (2000) existem atualmente inúmeras linguagens de indexação que exibem um variável grau de refinamento. Essas linguagens quando utilizadas nos sistemas de recuperação da informação podem provê-lo de qualidade adicional de modo que por seu intermédio possamos transformar informação em conhecimento. Segundo a autora, quando isso acontece a linguagem de indexação torna-se análoga ao próprio conhecimento. Nesse sentido, as linguagens de indexação são fundamentais não apenas na recuperação da informação, mas também em outros contextos que referenciam o conhecimento estruturado tais como, o processamento automático da linguagem e a engenharia de conhecimento.

Em virtude da afirmativa da pesquisadora buscou-se nesse trabalho compreender, do ponto de vista teórico e prático, as tendências de representação da informação em curso no contexto digital e experimentar novas abordagens na construção de instrumentos de mediação no processo de representação e recuperação da informação.

A experiência de construção do tesauro sobre mundialização, diversidade viva e desenvolvimento sustentável nasceu no âmbito das atividades do laboratório de pesquisas ESCoM – *Equipe Sémiotique Cognitive et Nouveaux Médias, sediado na Maison des Sciences de L’Homme*, Paris, França.

O Laboratório desenvolve pesquisas e experimentações em semiótica cognitiva com vistas a permitir o acesso público às unidades discursivas e às representações sociais presentes nos estudos desenvolvidos por pesquisadores no âmbito das Ciências Sociais e Humanas, em nível mundial.

O Laboratório ESCoM disponibiliza atualmente os AAR (*Archives Audiovisuelles de la Recherche*) um site especializado com um patrimônio vivo composto por cerca de 300 entrevistas, 100 colóquios e 55 seminários, além de 1500 entrevistas com pesquisadores franceses e estrangeiros.

Criado em 2001, os AAR têm como missão a produção, digitalização, montagem, indexação e publicação de registros audiovisuais dos patrimônios culturais e científicos sob a forma de arquivos digitais on-line; a concepção e o desenvolvimento de tecnologias e ferramentas apropriadas para a produção e a gestão à distância de arquivos multimídia, a descrição e exploração prática de registros audiovisuais; o estabelecimento de cooperação entre instituições produtoras de conhecimento



científicos e culturais, com vistas a facilitar a transferência de conhecimento científicos e culturais em audiovisuais; a realização e difusão editorial de conhecimentos científicos em meios digitais, além de valorizar, através de parcerias, o conhecimento científico de alto nível.

Parte integrante dos AAR, o espaço de conhecimento MDD (*Mondialisation et Développement Durable*) é dedicado ao debate atual em torno do tema mundialização, diversidade viva e desenvolvimento sustentável e tem por objetivo contribuir para a preservação e valorização do patrimônio mundial das culturas materiais e imateriais; construir e acompanhar a evolução da identidade sócio-cultural de cada povo; e apoiar as relações do homem e seu ambiente no nível global e local.

O acervo audiovisual do site foi organizado nos seguintes grandes temas: mundialização e preservação das identidades sociais e culturais, gestão de recursos naturais e do clima mundial, futuro do planeta, gestão do território e urbanismo, governança e atores sociais, acesso ao conhecimento, novas tecnologias de informação, educação e difusão das especialidades culturais locais, dinâmicas das populações e migrações humanas, desigualdades, crises e catástrofes, segurança local e paz mundial, saúde mundial e acesso local aos cuidados, e abordagem do desenvolvimento por regiões geopolíticas.

8 O DESENVOLVIMENTO DO TESAURO E DA ONTOLOGIA NO ESPAÇO DE CONHECIMENTO MDD

O desenvolvimento do tesouro e da ontologia partiu simultaneamente da compreensão das garantias literária, estrutural e de uso percebidas e identificadas na bibliografia da área, nos documentos digitais disponíveis na *Web* e nos discursos científicos realizados e disponibilizados no acervo do AAR e no MDD.

Por se tratar de um tema muito discutido, polêmico, transdisciplinar e com grande instabilidade conceitual, na primeira etapa utilizou-se o *Google Trends* para identificar a relevância e a rede semântica associada ao tema proposto. Esse procedimento nos deu a possibilidade de compreender em termos geopolíticos como estava organizada a produção científica referente ao tema e as suas manifestações no ambiente digital.

Na segunda etapa, identificamos, através do sistema baseado em *folksonomias* Delicious, as redes sociais que se organizam em torno do tema e o modo como as mesmas categorizam os assuntos tratados, os principais documentos eletrônicos disponíveis citados e as sobreposições semânticas operadas no campo.



Na terceira etapa, chegou-se a um léxico composto por cerca de duas mil expressões envolvendo os três focos temáticos (mundialização, diversidade viva e desenvolvimento sustentável). Realizou-se a categorização e descrição dos termos com a redução do léxico inicial que resultou em uma base terminológica composta por 650 termos e dispostos em 39 facetas. As facetas foram organizadas em: atividades humanas e seus produtos e efeitos sobre o ambiente, atividades, bens e serviços culturais, agricultura, análise ambiental, aspectos interdisciplinares da pesquisa ambiental, aspectos sociais e ações da política ambiental, aspectos teóricos da cultura, bens públicos, comunidade, conflitos sociais e coesão social, consumo, cooperação internacional, diversidade cultural, desregulamentação, ecologia, economia, educação ambiental, ética, geografia física, humana econômica e territorial, identidade social, migração, mundialização, movimentos sociais, ordem social, organismos, pobreza, políticas sociais, qualidade de vida, saúde, ciência e tecnologia, sociedade e populações, sistemas políticos, tecnologias da informação e da comunicação, turismo sustentável.

Na quarta etapa, cotejaram-se os termos identificados e categorizados com base no discurso audiovisual dos pesquisadores disponíveis na base da ESCoM e nas redes sociais identificadas a partir do Delicious.

Na quinta etapa formalizou-se o tesouro no software TCS 10 – Thesaurus Construction Systems e realizou-se um pré-teste da representatividade do tesouro em uma mostra de quarenta vídeos do acervo dos AAR, através do software Interview, desenvolvido pela Equipe ESCoM.

Na sexta etapa, identificaram-se as categorias necessárias ao desenvolvimento da ontologia que requeria a formalização no nível computacional. Desse modo, criaram-se dez ontologias gerais que permitem a combinação com a base terminológica central nomeada taxema. São elas: grafos, línguas, gêneros retóricos, tipos de documento, áreas de conhecimento, pessoas físicas, traduções, exploração de conteúdos informacionais, recursos e autoria.

Na etapa seguinte formalizaram-se as ontologias no software *Ontoeditor* e geraram-se os arquivos XML que depois foram incorporados ao software Interview dedicado à indexação dos acervos audiovisuais disponíveis nos espaços de conhecimento mantidos pela Equipe ESCoM.



9 CONSIDERAÇÕES FINAIS

Pensar a organização da informação na contemporaneidade parece um desafio intransponível, sobretudo em face dos múltiplos interesses e abordagens postos em ação.

Verifica-se que as pesquisas no campo da Ciência da Informação estão sempre desafiadas pela tendência tecnológica do momento. Dispensa-se tempo e esforço em perscrutar e descrever, do ponto de vista informacional, o desconforto, que tais instrumentos e modismos sinalizam, sem necessariamente, retomar a base da discussão estabelecida pelo próprio campo.

De acordo com o que Whittaker citado por Swanson (1988) denominou “postulado da impotência”, há sempre uma margem de indeterminação que os sistemas informacionais são “incapazes” de alcançar no nível da organização da informação. O autor convida à reflexão acerca do fato de os sistemas de informação construírem suas ações no limiar dos processos cognitivos e de significação estabelecidos pelos usuários. Torna-se necessário, nesse contexto, o estabelecimento de novos e distintos diálogos teóricos e metodológicos, pois, se considerarmos que o contexto de busca informacional prepondera no estabelecimento da relevância atribuída a uma informação, somos levados, implícita e explicitamente, a incorporar a colaboração expressa entre os usuários e suas redes sociais como um elemento indispensável na efetivação dos modernos sistemas de informação.

Em nosso experimento, a experiência de apropriação e utilização dos instrumentos disponíveis na Web para fins de construção de ferramentas ontológicas mostrou-se válido dado à agilidade com que os recursos informacionais são disponibilizados, exigindo uma redução drástica no *gap* estabelecido entre produção de conhecimento e a representação da informação.

As garantias, literária, de uso e estrutural, caras à construção de linguagens de indexação no âmbito da Ciência da Informação, se mostraram sobrepostas no contexto digital em virtude da ampliação da instância autoria na produção dos documentos eletrônicos. Isso significa dizer que os documentos gerados e/ou referendados pelas as redes sociais forneceram nova dinâmica aos processos de constituição das linguagens de indexação. Desse modo, é necessário incorporá-los aos métodos de estruturação de tais linguagens, com vistas a ampliar, com apoio dos usuários efetivos, a legitimidade e o poder de representação dos instrumentos de indexação.

As ferramentas verbais de representação da informação permanecem como bons instrumentos de mediação no acesso à informação em contextos digitais, devido à perenidade do conjunto de regras adotadas no seu estabelecimento. Assim, acredita-se



que ao incorporarmos criticamente o discurso das redes sociais, poderemos aprimorar as formas de acesso à informação e à representação das informações disponíveis em contextos digitais.

Finalmente, vale lembrar que, dado ao caráter onipresente do documento eletrônico em nossa vida cotidiana e devido ao seu processo de construção orientado pelas demandas, é necessário transformar radicalmente o nosso modo de pensar o aspecto material, a cognição e a percepção empregadas pelos usuários num contexto de interações hipertextuais.

Acredita-se, pois, que na consideração, explicitação e formalização dos mecanismos de significação, presentes nos percursos e contratos de leituras realizadas pelas redes sociais, repousam as possibilidades de alterarmos as lógicas de organização da informação e privilegiar, como foco, a circulação social do saber.

10 AGRADECIMENTOS

Agradecimentos são devidos ao CNPq, a CAPES, à Equipe Sémiotique Cognitive et Nouveaux Médias (ESCoM- Paris - França) e à Fondation Maison des Sciences de l'Homme (FMSH- Paris - França) pelo apoio concedido nas diferentes etapas do desenvolvimento desse trabalho.

11 REFERÊNCIAS

AGAMBEN, Giorgio (2007). *Qu'est-ce qu'un dispositif?*. Paris: Payot e Rivages.

BERNERS-LEE, T.; et al (2008). *Semantic Web Development Proposal*. Disponível em: <http://www.w3c.org/2001/sw> Acesso em: 17 de jul. 2008.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. (2001) *The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities*. Scientific American, New York, may. Disponível em: <http://www.sciam.com/2001/0501issue/0501berners-lee.html>.

BORBA, F. S.(1996) *Uma gramática de valências para o português*. São Paulo : Ática, 199p.



BRÄSCHER, M. (1999) *Tratamento automático de ambigüidades na recuperação da informação*, 286p. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília.

CAMPOS, M.L.de A. (2001). *Linguagem documentária: teorias que fundamentam sua elaboração*. 5.ed.Rio de Janeiro:Ed.UFF, 133 p.

CAMPOS, M.L.de A.,GOMES,H.E. (2004) Tesouro e normalização terminológica:o termo como base para o intercâmbio de informações.*Datagramazero*,v.5,n.6, dez.2004.Disponível em <http://www.dgzero.org/dez.2004>.

CHU, Heting (2005) *Information representation and retrieval in the digital age*. New Jersey: ASIS&T,

DIAS, Leila Christina, SILVEIRA, Rogério Leandro da (orgs.) (2007). *Redes, sociedades e territórios*. Santa Cruz do Sul.

FRANCIS, Élie, QUESNEL, Odile (2007). Indexation collaborative et folksonomies. Paris : *Documentaliste Science de l'information*. v.44, n. 1. février.

HOLAND, Seth van (2007). From spectator to annotator: possibilities by user-generated metadata for digital cultural heritage collections. Disponível em: <http://homepages.ulb.ac.be/~svhoolan/Usergeneratedmetadata.pdf>

LANCASTER, Frederick Wilfrid (2004). *Indexação e resumo: teoria e prática*. Brasília: Briquet de Lemos/Livros.

LANCASTER, Frederick Wilfrid (1986). *Vocabulary control for information retrieval*. 2.ed. Virginia, 270p.

LUEG, Christopher (2008). Where is the Action in Virtual Communities of Practice? Disponível em: <http://www-staff.it.uts.edu.au/~lueg/papers/commdcscw00.pdf>. Acesso em 20.nov. 2008.

MANIEZ, Jacques (2002). *Actualité des langages documentaires : Fondements théoriques de la recherche d'information*. Paris; ADBS.



MARTELETO, Regina Maria. (2000). Análise de redes sociais: aplicação nos estudos de transferência da informação. *Ciência da Informação*, Brasília, v.30, n.1, p.71-81, jan./abr.

MARTELETO, Regina Maria, TOMAÉL, Maria Inês (2005). A metodologia de análise de redes sociais. In: VALETIM, Marta Lígia Pomim (org). *Métodos qualitativos de pesquisa em Ciência da Informação*. São Paulo: Polis.

MARTINS, Rosane Maria (2008). *Web Semântica: uma visão geral*. Disponível em: http://www.eng.uerj.br/~rosane/survey_generico.pdf> Acesso em: 17 de jul. 2008.

McDERMOTT, R (2000). Knowing in Community: 10 Critical Success Factors in Building Communities of Practice. Disponível em: <http://www.co-il.com/coil/knowledgegarden/cop/knowning.shtml>

MORA, José Ferrater (1993). *Dicionário de filosofia*. São Paulo: Martins Fontes.

MORAIS, Erikson Freitas; SOARES, Marcelo Borghetti (2008). *Web Semântica para máquinas de busca*. Disponível em: homepages.dcc.ufmg.br/~nivio/cursos/pa03/seminarios/seminario7/seminario7.pdf
Acesso em: 17 jul. 2008.

NOVELLINO, Maria Salet Ferreira (1996). Instrumentos e metodologias de representação da informação. *Informação & Informação*, Londrina, v.1, n. 2, p. 37-45, jul./dez.

PEDAUQUE, Roger (2004). *Document: forme, signe et médium, les re-formulations du numérique*. Disponível em : <http://rtpc-doc.enssib.fr>.

RIBEIRO, Adagenor Lobato. As camadas da arquitetura da Web semântica (2008). Disponível em: <http://adagenor.blogspot.com/2008/03/as-camadas-da-arquitetura-da-web.html>. Acesso: 30. Jul.2008.

SVENONIOUS, E (2000). *The intellectual foundation of information organization*. Cambridge: MIT Press, 2000.



SWANSON, Don R (1988). Historical note; information retrieval and the future of an illusion. *Journal of American Society for Information Science*, v.49, n. 2, p. 92-98.

VOSS, Jakob (2007). Collaborative thesaurus tagging the wikipedia way. Disponible: <http://arxiv.org/trackback/cs/0604046>

ZUCCALA, Alesia (2006). Modeling the Invisible College. Zuccala A. *Journal of the American Society for Information Science and Technology* v. 57, n. 2. p.152-168, 2006.

ZINGLÉ, H (1999). *La modelisation des langues naturelles*: aspects théoriques et pratiques. Travaux du LILLA, numéro spécial, 1999. 151p.