

Espaço Discente

# Indexação Colaborativa na Web<sup>1</sup>

## Iara Aparecida Oliveira da Silva

Bacharel em Biblioteconomia pela Faculdade de Biblioteconomia e Ciência da Informação da Fundação Escola de Sociologia e Política.

## Cibele Araújo Camargo Marques dos Santos

Doutora em Ciência da Informação pela Escola de Comunicações e Artes da USP. Professora da Faculdade de Biblioteconomia e Ciência da Informação da Fundação Escola de Sociologia e Política.

**Resumo:** O artigo apresenta o procedimento de indexação na Web, essencialmente realizado por usuários de maneira colaborativa. Aborda o procedimento de etiquetagem em diversas ferramentas da Web 2.0 e sugere a utilização desse procedimento colaborativo, em conjunto com a indexação tradicional nas instituições, empresas e bibliotecas do Brasil.

**Palavras-chave:** Indexação; Folksonomia; Colaboração; Internet; Web 2.0

**Abstract:** The article presents the indexing procedures on the Web, performed essentially by users in a collaborative way. It presents the procedures of tagging in different tools on the Web 2.0 and suggests the use of this collaborative procedures along with traditional tagging on institutions, companies and libraries in Brazil.

**Keywords:** Indexing; Folksonomy; Collaboration; Internet; Web 2.0

<sup>1</sup> Trabalho de Conclusão de Curso apresentado à Faculdade Biblioteconomia e Ciência da Informação, da Fundação Escola de Sociologia e Política de São Paulo.

## INTRODUÇÃO

Com o grande fluxo de informações e registros produzidos pelo homem em diversos suportes durante a História, criou-se a necessidade do desenvolvimento de ferramentas e sistemas, que possibilitassem sua organização e recuperação. Ao longo do tempo o homem desenvolveu formas de “reencontrar” os registros feitos por ele para a posterior divulgação de informações, controle e entretenimento. A organização de tais registros tornou-se uma necessidade quando a quantidade de documentos passou a exceder a capacidade de controle e de recuperação.

A democratização da informação, o desejo de que a informação seja compartilhada por uma quantidade cada vez maior de pessoas, atravessa séculos na História e faz surgir o desenvolvimento de ferramentas específicas e/ou tecnologias próprias. No contexto da organização e recuperação da informação, o surgimento da linguagem controlada nas décadas de 1950/60 em decorrência do avanço científico e tecnológico motivou o desenvolvimento de um novo enfoque conceitual baseado no controle de linguagem.

Com o surgimento da internet, inicialmente sem a interferência direta do usuário, a Web 1.0 tinha fluxo unidirecional. Ela possibilitava acesso, mas não oferecia ainda as possibilidades de diálogo e interação que conhecemos hoje. Contudo, indiscutivelmente, promoveu o acesso a informações e aumentou a oferta sobre temas específicos e gerais,

tornando o usuário mais seletivo e exigente, criando uma demanda por novos “produtos” informacionais, dirigidos para públicos com perfis específicos. Esses públicos exigiam que seus desejos fossem ouvidos, fazendo com que os serviços passassem a ter um foco maior no relacionamento com o usuário (consumidor), ou seja, uma espécie de SAC (Serviço de Atendimento ao Consumidor), onde o mesmo poderia deixar o seu recado e ser prontamente atendido para sanar a sua dúvida.

Dessa primeira forma de interação surgiu a Web 2.0, onde o usuário pode produzir conteúdo e democratizar a informação o que revolucionou o acesso e a interação com a mesma. É desse trabalho compartilhado, dessa cooperação, que este artigo vai tratar, dessa ação conjunta em busca de um resultado para muitos. Da troca e da soma de conhecimentos e de como eles podem ser utilizados na organização e recuperação da informação em tempos de Web 2.0. O foco será a indexação da informação, o modo tradicional feito por especialistas em organização da informação (bibliotecários), o modo colaborativo feito pelo usuário da informação e a possibilidade de interação entre eles.

## **1 INDEXAÇÃO E LINGUAGENS: FERRAMENTAS PARA A RECUPERAÇÃO DA INFORMAÇÃO**

O bibliotecário reconhece a importância do processamento da informação tendo em vista o desenvolvimento de estratégias eficientes para a recuperação da informação. Um sistema de recuperação precisa justificar e comprovar a qualidade de seus serviços. Os estudos tradicionais de desempenho dessas instituições bem como os esforços para seu aperfeiçoamento, têm se orientado para o desenvolvimento das técnicas de processamento da informação: os meios sendo considerados mais importantes que os fins. Entretanto eles são sistemas sociais que devem ser avaliados à luz do processo de comunicação, de um lado, a informação; de outro o usuário; como canal, o sistema de recuperação da informação (CESARINO, 1985, p.166).

Independente da finalidade do sistema, a eficiência deste depende do conhecimento sobre o usuário deste conteúdo e de uma análise conceitual dos documentos feita com qualidade, pois grande parte dos ruídos obtidos nesse processo se deve a erros ou omissões na interpretação dos documentos e no desconhecimento do público-alvo. Analisar assuntos contidos em documentos, visando a sua recuperação quando necessário é uma das tarefas mais difíceis para o bibliotecário. Talvez isso se explique pela subjetividade inerente ao processo. É desejável o conhecimento do potencial interesse dos usuários do sistema, podendo-se dizer que o conhecimento da técnica de análise de assunto é imprescindível. (PINTO, 1985, p.185).

Segundo Cesarino (1985), existe um número grande de linguagens de indexação sendo utilizadas em sistemas de recuperação da informação. São elas linguagens naturais e controladas; linguagens pré e pós-coordenadas, entre outras. Cabe ao profissional da informação escolher, adaptar ou criar a linguagem que se ajuste a necessidade de seu sistema e de seu usuário.

Os membros da comunidade a ser atendida utilizarão base de dados ou catálogos, para satisfazer a diferentes necessidades de informação. Para lograr isso, devem converter uma necessidade de informação em algum tipo de 'estratégia de busca', a qual pode ser tão simples quanto a escolha de um único termo para consultar um índice impresso ou um catálogo em fichas, ou exigir a combinação de muitos termos numa estratégia mais elaborada e complexa [...] (LANCASTER, 2004, p. 2).

Indexação é o processo intelectual que envolve atividades cognitivas na compreensão do texto e a composição da representação do documento. Por ser uma atividade intelectual, utiliza especialmente teorias e métodos da psicologia cognitiva e da teoria de soluções de problemas (LIMA, 2006, p.104). O conceito de indexação surgiu a partir da elaboração de índices e atualmente está mais vinculado a análise de assunto. Com a evolução da prática, em decorrência da necessidade de recuperação das informações cada vez mais rápido, por parte de todos os que trabalham com a informação, os profissionais passaram a contar com um aparato lógico e metodológico diversificado e voltado para o contexto de cada documento, conforme citam Silva e Fujita (2004).

Para Lancaster (2004), uma indexação de assuntos eficiente implica que se tome uma decisão não somente quanto ao assunto tratado em um documento, mas pelo que também, através dele, possa despertar o interesse de determinado grupo de usuários. Para o autor não há receita nem conjunto de termos que sejam corretos para a indexação de documento algum. O mesmo documento em diferentes contextos e indexado por diferentes pessoas trará consigo um conjunto diferente de palavras e um diferente valor para cada leitor, usuário ou pesquisador.

A linguagem natural apresenta os dados da experiência de cada indivíduo, seguindo padrões da cultura e dos modelos sociais de cada povo que a fala. De acordo com Cintra (2002, p.17) "podemos dizer que cada linguagem natural é uma análise da sociedade, do homem participante de um grupo e sua cultura". Esta linguagem natural possui um conjunto de regras e um vocabulário, que se bem utilizados permitem a comunicação de informações e o registro das mesmas. O uso da linguagem natural para a indexação implica utilizar os termos que estão descritos no próprio documento sem compará-los com um instrumento estruturado de indexação. Como na linguagem natural, a linguagem documentária também é feita de sistemas simbólicos que visam facilitar a comunicação,

mas essa comunicação fica restrita a contextos documentários, facilitando a relação do usuário com o sistema e a posterior recuperação da informação.

De acordo com Cintra (2002), as linguagens documentárias mais consistentes dispõem de um vocabulário que integra, de um lado, a linguagem de especialidade e, de outro, a linguagem natural, que é a linguagem dos usuários e ainda, a linguagem dos autores dos documentos. Compete a estas linguagens transformar estoques de conhecimento em informações adequadas aos diferentes segmentos sociais. É esse compartilhamento que está na base do caráter público da informação e que não pode ser obtido na ausência de uma linguagem documentária. De fato, durante muito tempo acreditou-se que a disponibilização dos estoques seria suficiente para sua socialização. Mas atualmente, o fundamental é a existência de uma forma de organização que garanta compartilhamento. Apesar de todos os cuidados a serem tomados e persistentes defeitos ainda não sanados, não se pode deixar de lado o caráter social do trabalho de indexação, a instrumentalização do acesso, a disponibilização de diferentes tipos de informação a quem lhe é de interesse e a “redescoberta” do conteúdo que se encontra sob guarda, mas sem caminho de acesso. A indexação pode ser encarada como um dos procedimentos fundamentais para a disseminação e democratização da informação.

## **2 DA WEB 1.0 À WEB 2.0 – A COLABORAÇÃO DO USUÁRIO**

Os primeiros serviços da Web 1.0 ou Web clássica como alguns autores a denominam, se concentravam em fluxos unidirecionais, as empresas, universidades e instituições, simplesmente mostravam suas informações e produtos para apreciação, sem permitir ao consumidor/internauta participar de maneira interativa do seu conteúdo. Os papéis eram bem definidos, de um lado, os produtores de conteúdo (as instituições, empresas e Estado) ou pessoas com conhecimentos sobre a linguagem de programação, e de outro os consumidores, o restante da população (com acesso a rede).

O dinamismo ocasionado pela constante mudança da tecnologia empregada na internet pode estar diretamente vinculado à demanda por interatividade e este fato também decorre da necessidade de geração de demandas na dinâmica de mercado (MEIRELLES; MOURA, 2007).

A Web 1.0 chegou às casas, envolta de expectativas quanto ao seu potencial, porém logo foram apontados diversos pontos fracos quanto à pouca interação com os usuários. Essa participação ainda não havia sido desenvolvida pelos criadores e pesquisadores da área, os conteúdos eram pensados para o público, mas ainda não se pensava no público como parte do processo, produzindo e gerando valor para os conteúdos (LEVY, 1993).

O termo Web 2.0 surgiu em uma sessão de “brainstorming”, de Tim O’Reilly, em uma série de conferências feitas pela O’Reilly Media que tiveram início em outubro de 2004 (PRIMO, 2006, p.1). Alguns críticos ainda acreditam que se trata simplesmente de um modismo ou estratégia de marketing, outros já aceitam o termo como um novo paradigma segundo afirmou o próprio O’Reilly (2005).

A segunda geração de serviços on-line caracteriza-se por potencializar as formas de publicação, compartilhamento e organização de informações, além de ampliar os espaços para interação entre os participantes do processo. A Web 2.0 refere-se não apenas a uma combinação de técnicas informáticas, mas também a um determinado período tecnológico, a um conjunto de novas estratégias mercadológicas e processos de comunicação mediados pelo computador (PRIMO, 2006, p.1).

Como visão de mercado, podemos destacar que a Web 2.0, dá ouvidos e mãos à seus usuários. Segundo O’Reilly (2005), o princípio fundamental que se esconde por trás do êxito dos gigantes (empresas) nascidos na era da Web 1.0, que sobreviveram para liderar a era da Web 2.0 parece ser que estão abraçando o poder da Web para explorar a inteligência coletiva. De acordo com Lévy (1999, p.29), quanto mais os processos de inteligência coletiva se desenvolvem, melhor é a apropriação, por indivíduos e por grupos, das alterações técnicas, e menores são os efeitos da exclusão ou destruição humana resultantes da aceleração do movimento tecno-social.

### **3 INDEXAÇÃO COLABORATIVA**

A organização, classificação e recuperação da informação na Web 2.0 também se dá por meio da ação colaborativa. A participação e colaboração do usuário/internauta acontece também por meio da folksonomia, que se tornou ferramenta fundamental para essa fase de criação do conhecimento coletivo. O termo folksonomia é a tradução de um neologismo criado por Thomas Vander Wal, a partir da junção de folk (povo, pessoas) com taxonomy (taxonomia), (CATARINO; BAPTISTA, 2007). Folksonomia pode ser definida como a classificação livre de documentos, feita por etiquetagem, as chamadas “tags”, que utiliza a linguagem natural, nem sempre a utilizada nos documentos, e sim a linguagem comum do usuário, presente na memória coletiva. Através da folksonomia um novo link se forma com prática um novo tipo de hipertexto: o hipertexto 2.0 (AQUINO, 2007).

Uma etiqueta (tag) criada por um usuário pode ser utilizada por todos os outros, assim um item classificado por um usuário também é livre para ser classificado por quem quiser, criando um ambiente favorável ao desenvolvimento do conhecimento coletivo.

Esse conhecimento emerge de um sistema governado por forças bottom-up, isto é, da periferia para o centro (GUEDES; DIAS, 2010).

Esse sistema de etiquetagem, classificação, indexação coletiva ou colaborativa na Web 2.0 recebeu várias outras denominações na literatura, além de folksonomia. Catarino e Baptista (2007) destacaram doze termos possíveis para nomear etiquetagem na Web como: Tagging; Tagging Systems; Social Tagging; Social Tagging Systems; Collaborative Tagging; Collaborative Tagging Systems; Social Classification; Bookmarking; Social Bookmarking; Social Bookmarks Manager; Social Ontologies e Taxonomia dinâmica. As autoras esclarecem que existem dois aspectos diferentes para etiquetagem: atribuição de etiquetas aos recursos da Web como classificação e a etiquetagem das urls (ou favoritos), Bookmarking.

A grande quantidade de termos para explicar o mesmo conceito é o espelho do que ocorre com a etiquetagem dos documentos na Web, várias palavras podem ter o mesmo significado e chegar ao mesmo resultado. Aumenta-se o “leque” de opções de busca, porém, a variedade pode trazer o efeito contrário, caso se conheça apenas um termo e esse não seja utilizado. Vários serviços e sites estão utilizando a indexação colaborativa para auxiliar na organização, representação e recuperação de seus documentos. A inclusão de etiquetas ou “tags” pode ser feita por profissionais da área de informação ou criação do próprio site, por usuários, ou pelos dois grupos.

A indexação colaborativa se apresenta como solução paliativa de recuperação da informação, tendo em vista o excesso de informação na Web. É feita por todos instintivamente, sendo improvável que cada usuário da Web tenha consciência da importância de seu trabalho em escala global, o resultado desse trabalho ainda não é quantificável, apesar dos inúmeros indícios da bem feitoria dessa prática em ambientes digitais.

A biblioteca sempre foi o lugar onde é possível encontrar conhecimento. Sendo considerada como “detentora” da produção intelectual dos diversos povos de todas as épocas. Com o passar do tempo, diversos tipos de materiais e suportes foram surgindo e foram utilizados para disseminar e facilitar o acesso às informações produzidas e registradas. Durante muitos anos a biblioteca exerceu um papel de “guardiã” de todo o conhecimento, possuindo o acervo como foco dos seus serviços. Com o passar dos séculos “revoluções” sociais e intelectuais ocorreram no mundo e não seria diferente no ambiente das bibliotecas. Com o amadurecimento das teorias biblioteconômicas, os profissionais bibliotecários identificaram que o foco das bibliotecas deveria mudar dos acervos para os usuários.

A mudança consolidada e o pesado fluxo de informações que transitam pela internet, nos fazem pensar em novas formas de atender o público. Deve-se considerar de que modo estes usuários em potencial podem ser atendidos. A função da biblioteca é otimizar o tempo do usuário, informando-o de maneira assertiva e de acordo com os assuntos de seu interesse, atraindo esse público que tem a internet como ferramenta principal de trabalho, estudo, pesquisa e entretenimento.

Com as ferramentas que a Web 2.0 disponibiliza para o acesso e distribuição da informação, cabe às bibliotecas atualizarem seus serviços para aproximar-se do usuário e, por meio das tecnologias disponíveis chamar esse usuário, antigo receptor de informações, para ser agente de transformação, criação e colaboração na biblioteca. A indexação colaborativa e a utilização do chamado Social Bookmarking podem ser muito interessantes no ambiente da biblioteca, agregando valor aos livros e documentos. Disponibilizando os favoritos dos usuários na internet, pode-se criar demanda por livros ou assuntos, além de possibilitar para outros usuários mais uma possibilidade de acesso e recuperação de documentos, atualização dos assuntos mais procurados entre outras coisas.

#### **4 EM BUSCA DA INDEXAÇÃO IDEAL!**

Após investigar os dois modelos de indexação tradicional e colaborativo observou-se que tanto um, quanto outro apresentam ruídos com relação à recuperação da informação. Portanto, sugere-se a utilização das duas formas de indexação, a tradicional organizadora e classificadora de documentos por hierarquia, baseada em vocabulário controlado, a ser estudado de acordo com a instituição que dela irá fazer uso. E a colaborativa que não segue nenhum padrão de classificação, a ser utilizada como base de pesquisa para a atualização de vocabulário e suporte da primeira na tarefa de facilitar a recuperação da informação por ser representativa da instância de recepção da informação.

A idéia é utilizar as duas formas de indexação complementarmente, dando origem a uma espécie de indexação híbrida, aumentando as possibilidades de recuperação da informação por diversos públicos. Familiarizado ou não com um dos tipos de indexação, o usuário poderá escolher qual forma de busca mais lhe agrada. Além de posteriormente poder fazer parte do procedimento de indexação, sugerindo etiquetas para os materiais que mais gosta ou utiliza.

Assim, poderia-se incentivar a participação do usuário para atualizar, validar e complementar os recursos das ferramentas de busca das unidades de informação, empresas e instituições. Uma nova forma de pensar tem início, não é somente sobre

buscar, mas sobre encontrar o que se precisa que estamos falando, não é só sobre acesso a informação, mas também e essencialmente sobre compartilhamento.

## 5 CONSIDERAÇÕES FINAIS

Ao entrar em contato com tema indexação colaborativa na Web, buscou-se conhecer como o trabalho colaborativo de etiquetagem era utilizado na Web 2.0, e se esse trabalho poderia ser aproveitado como ferramenta auxiliar no procedimento de indexação tradicional já realizado pelos profissionais da informação no Brasil.

A maioria das empresas, instituições e bibliotecas no Brasil ainda não utilizam a indexação colaborativa para auxiliar na recuperação da informação. Nota-se certa resistência a esse vocabulário “aberto”, feito livremente pelos usuários.

Existe um desconhecimento das ferramentas por parte dos profissionais da informação e também por parte dos usuários. Outro fator a ser discutido é a falta de softwares de sistemas de bibliotecas que acompanhem esse procedimento de indexação colaborativa. O fluxo de informação na maioria das vezes de mão-única, pode ser repensado com o uso da indexação colaborativa, gerando assim uma retro-alimentação de informações, vindas também do usuário representante da instância da recepção, que pode oferecer parâmetros para atualização dos termos do vocabulário controlado. As redes sociais mostraram o sucesso do trabalho colaborativo em várias áreas, são exemplo de como a indexação colaborativa pode ser tornar algo natural na vida dos usuários e ser utilizada para auxiliar na recuperação da informação.

O ideal é complementar o procedimento de indexação tradicional, com a indexação colaborativa e não substituir uma pela outra. Benefícios além da dinâmica de vocabulário podem ser observados com a disseminação da informação e com a divulgação do ambiente que se utiliza do trabalho coletivo em busca do conhecimento.

## REFERÊNCIAS

AQUINO, M. C. Hipertexto 2.0, folksonomia e memória coletiva: um estudo das tags na organização da Web. **E-Compós**, Brasília, v.9, ago. 2007. Disponível em: <[http://www.compos.org.br/files/15ecompos09\\_MariaClaraAquino.pdf](http://www.compos.org.br/files/15ecompos09_MariaClaraAquino.pdf)>. Acesso em: 20 mai. 2010.

CATARINO, M.E.; BAPTISTA, A. A. Folksonomia: um novo conceito para a organização de recursos digitais na Web. In: **Revista Data Grama Zero**, v.8, n.3, jun. 2007. Disponível em: <[http://www.dgz.org.br/jun07/Art\\_04.htm](http://www.dgz.org.br/jun07/Art_04.htm)>. Acesso em: 05 mai. 2010.

CESARINO, M. A. da N. Sistemas de recuperação da informação. In: **R. Esc. Bibliotecon. UFMG**, Belo Horizonte, v.14, n.2, p.157-168, set. 1985.

- CINTRA, Ana Maria et al. **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002.
- GUEDES, Roger de Miranda; DIAS, Eduardo José Wense. Indexação Social: abordagem conceitual. In: **Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis**, v.15, n.1, p. 39-53, jan./jun. 2010. Disponível em: <[http://revista.acbsc.org.br/index.php/racb/article/view/686/pdf\\_17](http://revista.acbsc.org.br/index.php/racb/article/view/686/pdf_17)>. Acesso em: 20 out. 2010.
- LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Brinquet de Lemos, 2004.
- LÉVY, Pierre. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. Rio de Janeiro: Editora 34, 1993.
- \_\_\_\_\_. **Cibercultura**. Rio de Janeiro: Editora 34, 1999.
- LIMA, Gercina Ângela Borém. **Organização da informação para sistemas de hipertexto**. In: NAVES, Madalena Martins Lopes; KURAMOTO, Hélio (Org.). **Organização da informação: princípios e tendências**. Brasília: Brinquet de Lemos, 2006. p. 99-116.
- O'REILLY, Tim. What is Web 2.0? design patterns and business models for the next generation in software. **O'Reilly Media**, Califórnia, 2005. Disponível em: <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>. Acesso em: 20 mai. 2010.
- PINTO, Maria Cristina Mello Ferreira. Análise e representação de assuntos em sistemas de recuperação de informação. In: **R. Esc. Bibliotecon. UFMG**, Belo Horizonte, v.14, n.2, p. 169-186, set. 1985.
- PRIMO, Alex. O aspecto relacional das interações na Web 2.0. In: XXIX Congresso Brasileiro de Ciências da Comunicação, 2006, Brasília. Anais, 2006. Disponível em: <<http://www6.ufrgs.br/limc/PDFs/web2.pdf>>. Acesso em: 20 out. 2010.
- SILVA, Maria dos Remédios da; FUJITA, Mariângela Spotti Lopes. A prática de indexação: análise da evolução de tendências teóricas e metodológicas, **Transinformação**, Campinas, v.16, n.2, p.133-161, mai./ago. 2004.