



Universidade de Brasília - UnB
Faculdade de Ciências da Informação – FCI
Curso de Graduação em Biblioteconomia

**A PRÁTICA DA INDEXAÇÃO: DAS ANOTAÇÕES DE CONTEÚDO E
RESUMOS À ETIQUETAGEM NAS MÍDIAS SOCIAIS E AMBIENTE *WEB***

Paulo Gustavo Halfeld

Brasília – DF

2015

PAULO GUSTAVO HALFELD

**A PRÁTICA DA INDEXAÇÃO: DAS ANOTAÇÕES DE CONTEÚDO E RESUMOS À
ETIQUETAGEM NAS MÍDIAS SOCIAIS E AMBIENTE *WEB***

Projeto de Conclusão de Curso apresentado na Faculdade de Ciência da Informação da Universidade de Brasília como requisito básico para a elaboração da monografia de conclusão do curso de Biblioteconomia

Orientadora: Prof. Dra. Fernanda Monteiro

Brasília

2015

H1699

Halfeld, Paulo Gustavo.

Indexação: das anotações de conteúdo e resumos à sua etiquetagem nas mídias sociais e ambiente web, Paulo Gustavo Halfeld. – Brasília, 2015.

72 f.: il.

Orientadora: Fernanda Monteiro

Monografia (Graduação) – Universidade de Brasília, Faculdade de Ciência da Informação, Curso de Graduação em Biblioteconomia, 2015.

1. Indexação. 2 Representação da Informação.
3. Web 2.0. I. Título.



Título: A prática da indexação: das anotações de conteúdo e resumos à etiquetagem nas mídias sociais e ambiente web.

Aluno: Paulo Gustavo Halfeld da Silva.

Monografia apresentada à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Brasília, 23 de março de 2015.

Fernanda de Souza Monteiro - Orientadora
Professora da Faculdade de Ciência da Informação (UnB)
Doutora em Ciência da Informação

Dulce Maria Baptista – Membro
Professora da Faculdade de Ciência da Informação (UnB)
Doutora em Ciência da Informação

Simone Bastos Vieira – Membro
Professora da Faculdade de Ciência da Informação (UnB)
Doutora em Ciência da informação

Dedico o presente trabalho a meus pais, irmãos, e amigos que me apoiaram em todo processo de construção deste estudo.

AGRADECIMENTOS

Agradeço a aos meus pais pelo incentivo aos estudos e a leitura, compreensão e carinho dedicados a mim em todos os momentos da vida.

Obrigado meus irmãos, Luis e Rafael pelo apoio mesmo que distante.

Obrigado minha companheira e melhor amiga Thaís, por estar sempre presente em todas as horas, pela compreensão e ajuda, e por me mostrar que não devemos virar as costas para o novo.

Agradeço aos amigos que estão sempre por perto e me apoiando, em especial a Ingrid Schiessl pelas sucessivas leituras e discussões acerca do tema.

Agradeço a professora Fernanda Monteiro pelo auxílio na construção do trabalho, pela paciência, pela cobrança, e por estar sempre presente no decorrer de todo o processo.

“Nós não nos relacionamos, mas nos conectamos, não pela facilidade da conexão, mas pela facilidade da desconexão.”

ZygmuntBauman

RESUMO

O presente trabalho é uma revisão de literatura na área da indexação com enfoque na sua prática, e se propõe a entender o que propiciou a indexação se adaptar ao contexto da *Web 2.0*. Para tanto, é necessário traçar um panorama histórico que evidencie a evolução da prática da indexação até o atual contexto, que se dá de forma colaborativa pelos próprios usuários, no ambiente da *Web 2.0*. A metodologia compreende: Identificar, por meio da revisão de literatura os temas: Indexação e Contexto atual, que abarca *Web 2.0* e folksonomia, além de apontar os marcos históricos que definiram mudanças na prática da indexação; Caracterizar o ambiente *Web 2.0* e sua diferenciada de indexação. Foi possível traçar um panorama histórico que demonstrou a adaptação da maneira de indexar ao longo do tempo e ao analisar a evolução da área, percebeu-se que a prática sempre esteve ligada a necessidades e particularidades do seu tempo, e acompanhou as mudanças da sociedade na forma de lidar com a informação. Alcançou-se os objetivos de identificar as evoluções na prática da indexação e sua adaptação aos diferentes contextos, até os dias atuais. As principais contribuições desta pesquisa são a formulação de um panorama histórico da prática da indexação até os dias atuais, acréscimo de literatura em português para a área da etiquetagem, indexação como é desempenhada na *Web*, e por inserir a temática da folksonomia em um contexto de mudanças sociológicas e tendências aceleradas pelo rápido desenvolvimento e popularização das novas Tecnologias de Informação e Comunicação.

Palavras-chave: Indexação; Representação da Informação; *Web 2.0*; Folksonomia; Mídias Sociais.

ABSTRACT

This study is a literature review in the area of indexing with focus in its practice, and propose to understand what led indexing to adapt to Web 2.0 context. Therefore, it is necessary to have a historical overview showing the evolution of the practice of indexing until the current context, which happens collaboratively by the users in the Web 2.0 environment. The methodology comprises: Identify, through the themes literature review: Indexing and current context, which includes Web 2.0 and folksonomy, while pointing out the landmarks that defined changes in the practice of indexing; Characterize the Web 2.0 environment and its differentiated indexing. It was possible to build a historical view that demonstrated the adaptation of the way to index over time and to analyze the evolution of the index, it was noticed that the practice has always been linked to the needs and particularities of their time, and followed the changes in society in how to handle the information. It achieved the goal of identifying developments in the practice of indexing and their adaptation to different contexts until the present time. The main contributions of this research are the development of a historical overview of the indexing practice to the present day, in Portuguese literature addition to the area of tagging, as indexing is performed on the Web, and enter the theme of folksonomy in a context of sociological changes and trends accelerated by the rapid development and popularization of the new Information and Communication Technologies.

Keywords: Indexing; Representation of Information; Web 2.0; Folksonomy; Social Media.

Lista de figuras

Figura 1 - Ciclo informacional.....	53
Figura 2 - Desambiguação na Wikipedia.....	59
Figura 3 - Postagem de vídeo no Instagram.....	61
Figura 4 - Etiquetagem do vídeo com hashtags.....	62
Figura 5 - Recuperação da informação via pesquisa de hashtags	62
Figura 6 - Acesso à hashtag #handmade.....	63
Figura 7 - Etiquetagem no Delicious	64

Lista de Quadros

Quadro 1 - Evolução da prática da representação e da indexação até os dias atuais 49

Sumário

1. INTRODUÇÃO.....	14
1.1. Contextualização.....	16
1.2. Objetivos.....	17
1.2.1. <i>Objetivo geral</i>	17
1.2.2. <i>Objetivos específicos</i>	17
1.3. Justificativa.....	17
2. REVISÃO DE LITERATURA.....	19
2.1 Indexação.....	19
2.1.1 <i>Definições e conceitos</i>	19
2.1.2 <i>As diferentes concepções</i>	24
2.2 Histórico.....	27
2.3 Indexação na Web.....	35
2.3.1 <i>Ambiente Web</i>	35
2.3.2 <i>Etiquetagem de Recursos na Web</i>	37
3. PROCEDIMENTOS METODOLÓGICOS.....	43
3.1. Etapas da pesquisa.....	43
4. DESENVOLVIMENTO.....	46
5. RESULTADO.....	49
6. CONSIDERAÇÕES FINAIS.....	66
7. CONCLUSAO.....	68
REFERÊNCIAS.....	69

1. INTRODUÇÃO

Este estudo se propõe a entender de que forma a prática da indexação se adaptou aos diversos contextos até sua forma atual na *Web 2.0*, evoluindo dos processos manuais e completamente realizados por profissionais, passando pela informatização do método ou parte dele, até a forma mais atual, em que o próprio usuário realiza a representação da informação por ele propagada.

Sabe-se que constituir índices é prática bastante antiga, mas a indexação como processo, desenvolveu-se em conjunto com a necessidade de criação de mecanismos de controle bibliográfico em centros de documentação especializados. Dessa forma, o aparecimento destes mecanismos aconteceram“(...) fora do âmbito das bibliotecas tradicionais, representando uma evolução no processo de tratamento da informação e dando origem teórico-prática, naquela ocasião, a uma nova área, a Documentação.” (SILVA; FUJITA, 2004, p. 134). Embora seja importante entender a origem da Indexação como área da Documentação, o presente trabalho não objetiva estudar esta área geral, e sim a prática da indexação, dos primórdios da representação da informação, até sua forma atual e colaborativa em ambiente *Web*.

A indexação é uma área da representação da informação, e tem papel fundamental para que um documento ou seu conteúdo seja recuperado, através do levantamento de conceitos ou palavras que representem as informações originais, que serão localizados pelos mecanismos de busca. Segundo Robredo (1994, apud SILVA; FUJITA, 2004, p. 141), “o índice é o mais significativo instrumento para recuperação da informação, sendo definido como uma ‘chave’ condensada que dá acesso à informação contida nos documentos, ou como uma ponte entre o conteúdo de um acervo de informação e usuários”.Kuramoto(1995, p.3) define a finalidade da indexação

O processo de indexação produzindo uma lista de descritores visa à representação dos conteúdos dos documentos. Ou seja, este processo tem como objetivo extrair as informações contidas nos documentos, organizando-as para permitir a recuperação destes últimos.

Segundo Lancaster (2004), representar um documento facilita sua recuperação, de forma que os termos atribuídos servem como pontos de acesso para que os suportes sejam localizados. De acordo com Feitosa (2006), a representação tem como principal função a elaboração dos índices, que são

instrumentos utilizados para a representação do conteúdo de documentos primários. Estes índices, ainda segundo o autor, tem a finalidade de simplificar o processo de recuperação de informações e conteúdos contidos no documento representado ou resumido. Nesta linha, Fujita (1989) destaca a importância dos índices para a recuperação da informação, ao levantar a questão de quão cruciais são os índices de assunto enquanto ferramentas que objetivam possibilitar ao usuário a localização das informações desejadas. Desta forma, através de um conjunto de palavras, torna-se possível o acesso às ideias e conhecimentos contidos em um acervo documentário.

Com a popularização dos computadores pessoais, das Tecnologias de Informação e Comunicação (TICs) e o advento da internet, a produção e o acesso a informação tomaram proporções jamais vistas anteriormente. O aumento da quantidade de materiais produzidos diariamente tornou inviável a exigência de que um profissional analisasse, resumisse e indexasse todo o volume de informação gerado. Este ambiente *online* em que a produção de informação tem crescimento exponencial e o usuário contribui diretamente para a criação, disseminação e organização dos dados é denominado *Web 2.0*, “plataforma onde são compartilhados todos os dispositivos conectados” (O’REILLY, 2005 apud CATARINO; BAPTISTA, 2007, p. 1). O termo, usado para designar a segunda geração de comunidades e serviços na *Web*, denota uma mudança de paradigma na concepção de serviços *online*, que eleva o cliente ou usuário ao patamar de criador e organizador, tendo grande importância para que a informação e os aplicativos da rede sejam gerenciados. Uma vez que cada usuário tem o poder de criar e organizar a informação, o ambiente virtual torna-se muito mais rápido e dinâmico, apto a comportar a crescente geração de conteúdo, mas também passível de falhas.

Neste contexto, os processos de representação de informação deixaram de ser uma tarefa exclusiva dos profissionais envolvidos no tratamento dos materiais. Com a explosão informacional, após a II guerra mundial, a prática da indexação passou a contar também com a ajuda de computadores, tornando-se total ou parcialmente um processo automatizado (GIL LEIVA, 1999). Posteriormente, como explicaram Catarino e Baptista (2007), a publicação e o acesso a informação tornaram-se ações de fácil execução, e as pessoas passaram a ter a possibilidade de participar ativamente destes processos de construção e organização de

conteúdos na *Web*. Nos ambientes virtuais, a representação da informação cada vez mais tornou-se uma atividade colaborativa, passando a ser desempenhada pelo próprio usuário, e é nomeada por Thomas Vander Wal (2004) como Folksonomia.

No desenvolvimento deste estudo, a metodologia adotada constitui-se de revisão de literatura acerca da indexação e sua evolução histórica, *Web 2.0* e etiquetagem na *Web*, buscando definições e conceitos cunhados por autores de referência na área, com foco na prática e experiência no processo de representação da informação. Após delimitar o tema, os textos e artigos foram resumidos e fichados, e então separados por ordem cronológica quanto ao assunto abarcado. Posteriormente, o conteúdo foi analisado e resumido a fim de extrair citações e passagens importantes que contribuíssem para a elucidação dos objetivos propostos.

1.1. Contextualização

O presente trabalho aborda a temática da indexação, área da representação da informação, e a evolução de sua prática ao longo do tempo, até o contexto digital. Trata-se de uma importante área da Ciência da Informação, pois está diretamente ligada ao fechamento do ciclo informacional e sua eficiência, relacionando-se a recuperação da informação.

Este estudo tem como objetivo traçar um panorama histórico em relação ao tema indexação, começando nos primórdios da representação da informação até o contexto da *Web 2.0*. É importante ressaltar que o trabalho não pretende ser exaustivo e abarcar áreas além da prática da indexação, por se tratar de um projeto de conclusão de graduação com tempo limitado para conclusão.

Considerando a escassez de literatura, principalmente nacional, acerca da representação temática da informação nas mídias sociais, que tem como fundamento o processo de indexação tradicionalmente realizado para constituição de índices de bibliotecas, é pertinente desenvolver um trabalho traçando o processo de evolução da indexação de materiais impressos, até sua forma colaborativa nos meios virtuais, em poder dos usuários. Como a prática da indexação evoluiu até o contexto da *Web 2.0*?

1.2. Objetivos

1.2.1. Objetivo geral

Traçar um panorama histórico que evidencie a evolução da prática da indexação, demonstrando como esta se adaptou ao contexto da *Web 2.0*.

1.2.2. Objetivos específicos

1. Identificar marcos históricos que definiram mudanças na prática da indexação;
2. Levantar aspectos que influenciaram a prática da indexação, da forma como é realizada atualmente, na *Web 2.0*;

1.3. Justificativa

O estudo de indexação e representação da informação é muito importante, pois tem papel determinante sobre a eficiência na recuperação dos documentos desejados pelo usuário. Nas palavras de Pinto (2000, p. 63), “a indexação tem por objetivo teórico expressar de maneira o mais fiel possível, a representação dos elementos que pertencem ao conteúdo de um documento – seja ele real ou virtual – a fim de que o mesmo possa ser recuperado posteriormente”. Sendo assim, uma indexação criteriosa e coerente com as políticas da instituição e o perfil do usuário impactam de maneira significativa, possibilitando às estratégias de busca sucesso ao recuperar a informação.

Embora a representação da informação e indexação tenham sido amplamente abordadas por teóricos conceituados, por ser temática sempre atual devido a sua possibilidade de adaptação a novos contextos, as repercussões e implicações destas práticas no ambiente virtual da *Web 2.0* ainda são recentes e pouco abordadas, principalmente na literatura nacional. A folksonomia e as práticas de etiquetagem¹ na internet, bem como suas consequências, ainda não são claras, sendo importante compreender este contexto para auxiliar os usuários na indexação

¹Tradução do termo *tagging*, é uma forma de indexação realizada através de palavras-chave na web.

e recuperação da informação *online*, uma vez que, conforme mencionado, uma indexação bem feita, em qualquer ambiente, é fator preponderante para que a informação primária seja acessada.

O presente estudo visa traçar um panorama histórico da evolução da prática da indexação, analisar e identificar que contextos propiciaram a adaptação dos processos de representação da informação e indexação à plataforma *Web 2.0*. Trata-se de um tema atual e relevante para a área de Ciência da Informação, uma vez que os produtos deste contexto, como as *wikis*² e as mídias sociais, tornam-se cada vez mais populares e demandadas como fontes de informação pelos usuários. A possibilidade e difusão da criação de *tags* nas plataformas *Web*, segundo Catarino e Baptista (2007), podem ser vistas como um novo paradigma para a organização dos conteúdos digitais, mas ao mesmo tempo, pouco se sabe sobre seus impactos sobre a recuperação da informação. A fim de compreender melhor este cenário, torna-se imprescindível conhecer a prática da indexação, sua evolução, progresso e adaptação aos novos ambientes virtuais e colaborativos.

²Coleções de documentos na *Web* criados de forma colaborativa

2. REVISÃO DE LITERATURA

A presente revisão de literatura foi realizada através de leitura, análise e fichamento de materiais para os temas propostos. Posteriormente os temas foram confrontados, permitindo que as análises fossem dispostas de forma a colaborar com elucidação da questão do estudo.

Objetivou-se levantar conceitos e definições acerca da indexação, bem como suas visões distintas entre as correntes inglesa e francesa, e a evolução da sua prática ao longo do tempo, até sua forma atual e utilizada nas mídias sociais inseridas no ambiente *Web 2.0*: a folksonomia.

2.1 Indexação

A documentação do modo que é hoje, segundo Kobashi (1994), nasceu no século XVII com a edição de *Le Journal des Sçavans* publicado em Paris no ano de 1665. Era um periódico semanal que trazia resumos dos trabalhos científicos, filosóficos e artísticos.

Segundo Silva e Fujita (2004, p. 134) a indexação surge, com a atividade de elaboração de índices. As autoras afirmam que:

[...] a atividade de indexação, como processo, é realizada mais intensamente desde o aumento das publicações periódicas e da literatura técnico-científica, surgindo a necessidade de criação de mecanismos de controle bibliográfico em centros de documentação especializados.

Na subseção a seguir são apresentadas definições e conceitos relativos ao tema.

2.1.1 Definições e conceitos

A atividade de representar a informação, mesmo que de forma rudimentar, sempre esteve presente de alguma forma nos processos de tratamento da informação, desde as primeiras formas de registro. A representação da informação tem como fim tornar possível o futuro uso dos conteúdos existentes nos documentos e consiste em sistematizar conceitos, descritores e palavras chave com

a finalidade de representar as informações e possibilitar sua recuperação. Neet(1989apud GUIMARÃES, 2009, p. 107) afirma que seu objetivo é “facilitar a pesquisa de documentos ou de informações contidas em documentos”.

Segundo Lancaster (2004), representar um documento facilita sua recuperação, de forma que os termos atribuídos servem como pontos de acesso para que os suportes sejam localizados. Em um breve resumo histórico da atividade, Ribeiro (2005, p. 1) traça o processo evolutivo da representação da informação:

[...] desde sempre que, associados ao processo de produção e conservação da informação, existiram procedimentos de organização e representação, requisitos indispensáveis (por isso mesmo, sempre ocorreram de forma natural ou espontânea) para potencializar o uso recorrente dessa mesma informação. As práticas empíricas milenares associadas à produção, tratamento, uso e armazenamento da informação foram se tornando mais complexas e tornaram-se objeto de reflexão e de estudo, estando na origem de um processo evolutivo, iniciado ainda no século XIX, que acabaria por desembocar na construção da moderna Ciência da Informação.

Esta atividade possui a característica de desmembrar o documento original em documentos secundários, que representam imagens resumidas dos documentos analisados, ou indicadores que resumem seu conteúdo, e isto propiciará a recuperação da informação. Ribeiro (2005, p. 3) disserta sobre os índices, os quais inclui na categoria dos instrumentos de acesso à informação:

[...]os catálogos, os repertórios, os inventários, as listas de referências ordenadas, enfim, todos os tipos de índices constituem aquilo a que modernamente chamamos ‘instrumentos de acesso à informação’ e não são mais do que representações (imagens) da informação, que servem de intermediárias entre os pesquisadores e o produto informacional que é procurado.

Considerando que uma das funções da indexação é a de gerar índices, Robredo (1994 apud SILVA; FUJITA, 2004, p. 141) destaca sua importância enquanto ferramenta de busca, e o define como:

[...] um conjunto ordenado de códigos representativos de assuntos, tópicos ou conceitos (por exemplo, códigos de classificação, grafismos diversos, incluindo palavras ou frases), os quais podem servir como critérios de busca relacionado com alguma chave de acesso que permita localizar os documentos – ou suas partes ou representações – relativos a cada assunto.

Silva e Fujita (2004) classificam a indexação como um processo de tratamento temático, abarcando análise, síntese e representação, na perspectiva do tratamento da informação. Observa-se que a indexação é uma das formas de representação de conteúdo que surgiu com a finalidade de elaborar índices ou listas

de termos significativos. Pinto (2000, p. 63) faz considerações importantes acerca destas finalidades:

A indexação tem por objetivo teórico expressar de maneira mais fiel possível, a representação dos elementos que pertencem ao conteúdo de um documento – seja ele real ou virtual – a fim de que o mesmo possa ser recuperado posteriormente.

Fica claro então que a indexação possui um objetivo teórico, que estabelece técnicas e mecanismos para a representação dos documentos e posterior elaboração dos índices, e outro operacional, que visa possibilitar a recuperação da informação.

Gardin(1974,apud PINTO, 2000, p. 65), complementa a definição de indexação, levantando aspectos relacionados a sua prática e objetivos teóricos, ao mencionar o uso de descritores:

Consideramos a indexação documentária como um conjunto de atividades que consiste em identificar , nos documentos , os seus 'Traços Descritivos' (TD's) ou macroproposições e em seguida extrair os elementos/descriptores (sintagmas) indicadores do seu conteúdo visando à sua recuperação posterior.

É importante ressaltar, sobre a extração dos elementos que irão indexar o documento, que as palavras-chave ou conceitos podem ser extraídos do próprio texto ou retirados de outras bases de dados, a exemplo das Linguagens Documentárias (LD's). Já os sintagmas, que representam frases completas, somente podem ser retirados do texto (PINTO, 2000).

Existem várias maneiras e técnicas para realizar a indexação, e o século XX comportou a inserção dos computadores e máquinas na atividade de representar a informação. A indexação pode ser realizada de forma manual ou intelectual, feita exclusiva e inteiramente pelo homem, de forma mecânica e automática, através unicamente de softwares e algoritmos, ou de forma semi-automática, em que o processo é misto: parte informatizado e parte humano (GIL LEIVA, 1999). Neste caso, uma das técnicas utilizadas faz uso primeiramente de um software que realiza a triagem automática dos conceitos mais mencionados no texto, e em seguida um profissional da indexação completa o trabalho, levando em conta a relevância e os aspectos semânticos necessários.

O processo manual de indexação, completamente realizado pelo ser humano, inicia-se com a análise do conteúdo do documento. Este é lido através de suas estruturas lógicas, e não de forma completa. Trata-se da leitura técnica,

processo intrínseco a identificação do assunto pelo indexador. Parte-se ou das estruturas propostas pelo autor/editor, ou a partir do julgamento do indexador, de forma a compreender seu conteúdo para extrair conceitos que representem o documento. A etapa seguinte, chamada tradução, compreende o casamento entre a linguagem dos descritores retirados do documento, e os contidos nas Linguagens Documentárias da instituição, e pode ser responsável pela ocorrência de silêncio ou ruído no momento da recuperação da informação, de acordo com Fidel (1994, apud PINTO, 2000). Na visão deste autor, a indexação manual ainda é uma atividade muito utilizada e que apresenta a resposta mais eficiente na recuperação da informação, mas que tem custo elevado, pela exigência de profissionais especializados. Por ser realizada pelo homem, pode apresentar baixa coerência entre indexadores diferentes.

Pesquisas realizadas por Fujita (1999; 2003 apud SILVA; FUJITA, 2004) concluíram que quando a indexação é realizada por profissionais especializados, ela ocorre de forma mais ágil, o que foi atribuído ao fato dos indexadores dominarem a área de assunto dos documentos. Concluíram também que os não especialistas, apesar de indexarem de forma menos ágil, fizeram um uso maior de estratégias, em que a mais usada foi a associação com a linguagem do sistema de indexação.

É importante observar que a indexação é atividade primordial para a existência e sobrevivência de uma unidade de informação e para o fechamento do ciclo informacional, conforme observa Booth (2001 apud GUIMARÃES, 2009, p. 107):

[...] usuários de bibliotecas esperam dispor de ferramentas de pesquisa que lhes permitam recuperar documentos específicos sobre um determinado assunto, enquanto pesquisadores dependem da existência de índices de periódicos, ou de coleções de documentos de diferentes ordens para produzir literatura científica.

Neet (1989, apud GUIMARÃES, 2009) complementa o papel da indexação nas instituições, ao afirmar que o seu objetivo primordial é o de facilitar a pesquisa de documentos ou de informações contidas em documentos. Assim, torna-se necessário observar questões como a dimensão documento/informação e documento/usuário, considerando sempre as demandas informacionais as quais os documentos se propõem a suprir. É necessário ao profissional da área não só deter noções técnicas e bom nível de conhecimento, mas manter-se coerente e ter uma

compreensão afeiada acerca das necessidades do usuário, bem como do contexto da instituição (GUIMARÃES, 2009).

Apesar de ser desejável que a indexação seja coerente com as Linguagens Documentárias e políticas da instituição, é importante observar que existe autonomia por parte do profissional para manter o termo retirado do documento, e não o correspondente contido nas listas. Nas palavras de Lancaster (1993, apud SILVA; FUJITA, 2004, p. 135), “a terminologia usada pelo autor pode não corresponder exatamente aos termos do vocabulário controlado” e que “devem ser indexadas as ideias do autor do texto e não as palavras”. Mesmo contando com essa liberdade, deve-se manter um grau de consistência e foco na recuperação da informação, pois, segundo afirmam Cleveland & Cleveland (1990 apud GUIMARÃES, 2009, p. 108), “índices e resumos não farão qualquer sentido se não se prestarem a busca informacional”.

A atividade de indexação deve sempre levar em consideração a natureza do documento, de forma que uma monografia não deve ser indexada, necessariamente, da mesma forma que um livro. Desta forma, Pinto (2000) salienta que não existe uma regra única para realização da indexação, que deve considerar três elementos: o autor, o indexador e o usuário. Além disto, um outro elemento importantíssimo que deve sempre ser levado em conta são as políticas da unidade de informação, de forma que o processo de indexação seja coerente também dentro do acervo ao qual os documentos estão inseridos. É recomendável que as instituições possuam um manual que contenha políticas, missão e objetivos da unidade de documentação, a fim de nortear a representação da informação de forma a reduzir a subjetividade inerente a tarefa. Fungmann (2000 apud GUIMARÃES, 2009, p. 108) aborda a questão da falta de objetividade na indexação, e observa que “se um mesmo documento é indexado de forma completamente diferente por distintos indexadores ou mesmo pelo mesmo indexador, ao longo do tempo, significa que efetivamente alguma coisa está errada”.

Mesmo com todos os avanços referentes aos processos de indexação e ao desenvolvimento de técnicas e padrões para a realização da atividade e elaboração e utilização das linguagens documentárias, algumas vezes conceitos e procedimentos não são tão bem delineados, conforme observam Cleveland & Cleveland (1990):

Como em qualquer processo de indexação de assunto, o indexador começa com o reconhecimento das próprias palavras do texto, 'escaneando' cada sentença e grifando as palavras-chave utilizadas pelo autor. O indexador, então, avalia as referidas palavras face à estrutura geral do parágrafo de modo a determinar os assuntos que estão sendo discutidos. Certamente, nem toda palavra grifada em um parágrafo constitui assunto significativo. Muitas palavras são de menor importância e não seriam utilizadas em um índice. Que tópico ou tópicos importantes são discutidos no parágrafo? Que palavras são simplesmente modificadores e não efetivamente indicadores de assunto? Por exemplo, em os mosquitos atacaram com a ferocidade de um tigre, apenas mosquitos são um indicador de assunto.

Este exemplo reitera a importância de que o indexador, mesmo se atendo às regras e normas, precisa ter bom senso e coerência na hora de indexar, utilizando sua liberdade para a escolha dos descritores e tendo em mente as opções de busca de um usuário. Em sua natureza, a Indexação, nas palavras de Pinto (2000, p. 66), “desmonta o discurso montado pelo autor do documento, à medida que faz recortes neste discurso.” Desta forma, torna-se compreensível que, em alguns casos, os descritores utilizados pelo indexador não sejam exatamente palavras utilizadas pelo autor ou as contidas nas linguagens documentárias da instituição.

2.1.2 As diferentes concepções

Diferentes correntes entendem o termo indexação de maneiras distintas, pois o mesmo tem significado abrangente, o que torna difícil delimitar sua extensão dentre as tarefas do tratamento da informação. Nas palavras de Silva e Fujita (2004, p. 136):

A partir da evolução que determinou a importância do contexto do documento para a recuperação da informação, a área de indexação passa a incorporar os estudos dirigidos à compreensão do conteúdo dos textos a serem analisados. Esses estudos, porém, estão claramente inseridos em correntes teóricas e é fácil confundir, na literatura, a função da indexação perante a necessidade de análise de conteúdo. Na literatura observa-se a existência de duas correntes teóricas: a francesa e a inglesa.

A definição de Análise Documentária precede a discussão acerca das diferentes visões, mais especificamente francesa e inglesa, sobre a indexação, por ser ponto central de discordância entre os autores. Gardin (1981 apud SILVA; FUJITA, 2004, p. 136) define Análise Documentária como “um conjunto de

procedimentos efetuados com a finalidade de expressar o conteúdo de documentos científicos, sob formas destinadas a facilitar a recuperação da informação”. Serres (s.d.apud GUIMARÃES, 2009, p. 109) aponta o objetivo da análise documentária, ao ressaltar que ela “visa a representar o conteúdo intelectual do documento, a fim de facilitar a consulta ou a posterior recuperação”, dotada de natureza de operação intelectual “que conduz, por outro lado, ao resumo do documento e, por outro, à indexação do mesmo, revelando seus conceitos-chave de conteúdo”.

A Análise Documentária é composta por um processo comunicativo, conforme observa Fox (2005apud GUIMARÃES, 2009, p. 110), em que “um conjunto de operações cognitivas de natureza analítico-sintética transforma, por meio do reconhecimento e da representação do conteúdo” o documento fonte em outro mais resumido. Este novo documento, de natureza secundária, facilita a recuperação e subsidia a pesquisa científica. Esta abordagem, segundo Guimarães (2003, p. 100), “pressupõe o resgate do ciclo informacional enquanto base para o fazer documental, em cujo âmbito se verificam, como operações fundamentais e interdependentes, a produção, o tratamento ou organização, a recuperação, a disseminação e o uso da informação que, por sua vez, poderá gerar nova produção, completando o ciclo”. De forma objetiva, é possível afirmar que a análise documentária, para fins de tratamento da informação, traduz-se em um conjunto de processos de análise e síntese, por meio do emprego de representação em linguagens documentárias, visando assegurar eficiência e precisão na recuperação da informação (GUIMARÃES, 2003).

É importante observar que diversos campos do conhecimento oferecem aporte teórico à área, que se configura como multidisciplinar, ao empregar subsídios da Administração, por ser parte do contexto organizacional das unidades de informação, da Psicologia, por abordar questões como conhecimentos prévios do usuário e seu vocabulário, bem como noções cognitivas, também abordadas pela Linguística (GUIMARÃES, 2003).

Desta forma, observa-se que os conceitos de Análise Documentária e Indexação são próximos e em muitos aspectos, convergentes. Para os teóricos da corrente inglesa e norte-americana, representada por Foskett, Lancaster e Campos, indexação e análise documentária são sinônimos, e abrangem a mesma atividade e

conceituação, com objetivo de representação do conteúdo informacional de documentos para a elaboração de índices.

Nesta linha, para Borko e Bernier(1978apud SILVA; FUJITA, 2004, p. 137), a indexação se define como o “processo de analisar o conteúdo informacional dos registros do conhecimento e sua expressão na linguagem do sistema de indexação”, e Van Slype(1991apud SILVA; FUJITA,2004, p. 137), define como “a operação que consiste em enumerar os conceitos sobre os quais trata um documento e representa-los por meio de uma linguagem combinatória: lista de descritores livres, lista de autoridades e o *thesaurus* de descritores”. Observa-se, que na visão destes autores, os termos indexação e análise documentaria constituem processos idênticos, não fazendo nenhuma diferenciação entre eles.

Teóricos da corrente francesa entendem a indexação como sendo uma sub-atividade da Análise Documentária. Sendo assim, a indexação seria a etapa final deste processo, vista como um conceito mais amplo. Seguindo esta visão, Chaumier(1982apud GUIMARÃES, 2009, p. 109) diferencia indexação de análise documental, ao afirmar que a análise documentária abrange dois tipos de tratamentos diferentes: a condensação, que se vale da extração de conceitos para servir de apoio à difusão da informação, e a indexação, que se vale da extração de conceitos para servir de apoio à recuperação”, e Gardin (1981 apud GUIMARÃES, p. 111) ressalta que a indexação é tida como uma das operações da análise documentaria, com a finalidade prática de propiciar a recuperação da informação.

Desta forma, Silva e Fujita (2004) se posicionam de acordo com esta visão, ao entender que a Análise Documentária é um macro universo no qual a indexação está inserida, e Chaumier (1980 apud SILVA; FUJITA, 2004, p. 138) eleva a indexação à posição mais importante dentro deste processo de representação documentária, sendo fator determinante para o valor de um sistema documentário. Isto porque, na visão do autor, a mesma “condiciona o valor do sistema documental”. Adverte ainda que “uma indexação insuficiente representa 90% das causas essenciais para aparição de ruídos ou de silêncios”, e que a indexação constitui “operação primordial sem a qual não é possível o efetivo uso da informação” (CHAUMIER, 1982 apud SILVA; FUJITA, 2004, p. 137).

Guimarães (2009, p. 111) define bem esta divergência de concepções, ao destacar que:

[...] enquanto alguns autores, mais voltados para a escola inglesa (Foskett, 1973; Cavalcanti, 1982; Amaro, 1991 e Lancaster, 1993, dentre outros), vêem identidade entre o tratamento temático da informação e a indexação, outros, mais ligados a escola francesa (Gardin, 1981; Ruiz Perez, 1992; Pinto Molina, 1993 e Guimarães, 2003) encaram a análise documental enquanto uma área (todo) na qual se insere a indexação propriamente dita (parte), o que se reflete na representação documental, fase final do tratamento, em que se utilizam os instrumentos documentais (linguagens) para a geração de produtos documentais (índices, notações classificatórias, etc.).

Apesar de existirem visões distintas a respeito do tema, Silva e Fujita (2004, p. 137) ponderam:

[...] é preciso considerar que, dentro de uma perspectiva histórica, a Indexação tinha uma finalidade específica de construção de índices e o termo 'Indexação' se ajustava perfeitamente à atividade, porém, com o uso de tecnologias de recuperação da informação a necessidade de elaboração de índices foi sendo substituída pela necessidade de representação do conteúdo documentário por termos de indexação em decorrência da análise de assunto.

Desta forma, torna-se necessário compreender a evolução da representação documental e da indexação ao longo do tempo, a fim de perceber como a prática levou o conceito a se adequar aos momentos e ambientes aos quais a atividade estava inserida.

2.2 Histórico

A atividade de representação da informação, enquanto processo de tratamento de documentos, é muito antiga. Segundo Ribeiro (2005, p. 2),

Embora as práticas empíricas de organização da informação tenham estado sempre associadas ao respectivo processo de produção e armazenamento, porque só assim se tornava possível a recuperação e o uso recorrente da própria informação, a reflexão sobre essas práticas e o desenvolvimento disciplinar em torno delas são criações relativamente recentes, quando comparadas com a história milenar do processo informacional.

Indexar e constituir índices, instrumentos de armazenagem e recuperação da informação, originaram-se a partir do momento em que “o homem passou a se preocupar em tornar acessível a informação registrada em um documento e para isso resolve ordená-la de alguma forma” (GOMES; GUSMÃO, 1983 apud SILVA; FUJITA, 2004, p. 138).

Nos primeiros e mais primitivos sistemas de informação, a principal preocupação no trato dos documentos a fim de possibilitar sua localização física, era

referente a separação entre suportes (tábuas de argila, pergaminho, papiros), e os assuntos ou tipologias informacionais sobre as quais versavam os autores, como leis, registros de contabilidade e tratados de comércio, literatura, etc. Observa-se que a atitude de separar estes materiais, segundo os critérios supracitados, já revela o cuidado com que a informação guardada e ordenada fosse localizada posteriormente, mesmo sem ainda propiciar a localização das informações contidas nestes documentos (RIBEIRO, 2005).

Segundo Collinson (1971), a indexação se iniciou baseada na memória. Nas primeiras civilizações humanas, grandes contos, textos religiosos sobre a origem do homem e epopeias eram transmitidos de forma verbal, e mais tarde observou-se a criação dos primeiros índices rudimentares, que vinham arranjados pela primeira sentença de cada parágrafo.

Na antiga Suméria e Mesopotâmia, cerca de XVII a.C., encontramos a forma mais antiga de armazenamento de informação que se tem conhecimento. Em tábuas de argila, através da escrita cuneiforme, se registravam tratados de comércio, leis e estudos científicos. Nestas tábuas de argila também foi encontrada “uma espécie de resumo dos livros antigos considerada como a forma de representação de informação condensada do conteúdo informacional que dava acesso ao assunto dos livros” (WITTY, 1973apud SILVA; FUJITA, 2004, p. 138).

A Antiga Biblioteca de Alexandria, uma das maiores bibliotecas do mundo antigo, teve seu acervo organizado segundo a classificação de Calímaco (310-240 a.C.), poeta erudito e bibliotecário, e já praticava a condensação de suas principais obras, em forma de resumos. Seu catálogo, a *Pinakes*, era constituído por 120 volumes e era ordenado por temas e autores, com notas bibliográficas sobre estes últimos, tornando-se obra de referência sobre a literatura grega (RIBEIRO, 2005). No século II, o guia *De Libris Propriis Libris* foi compilado por Cláudio Galeno (129-217 d.C.), proeminente médico e filósofo romano. O surgimento deste guia cria um marco, em que compilações de obras isoladas são condensadas em um só documento, que continham cabeçalhos de capítulos na margem dos parágrafos. A obra anônima *Apothegmata*, datada do século V, consistia de uma listagem de provérbios gregos sobre assuntos teológicos, e foi a que mais se aproximou do atual índice alfabético de assuntos (GARRIDO ARILLA, 1999). A partir deste momento histórico, as obras começam a se configurar de forma a facilitar a recuperação de

suas partes, através de capítulos, seções numeradas, cabeçalhos de capítulos, sumário e tábuas de matéria.

Por volta do século XIV, a concepção de índice evoluiu para a de listagem de conteúdo ou lista de resumos e notas, ainda distante da noção atual de índice (SILVA; FUJITA, 2004). Os copistas, de forma subjetiva e pessoal, escreviam às margens dos livros palavras ou frases que representassem o conteúdo do mesmo, tentando desta forma criar uma espécie de indicação de relevância, conforme observa Collinson (1971). Nos mosteiros, a elaboração de catálogos de livros e listas já era uma das metas estabelecidas pelos monges. Pela primeira vez, observou-se que a indexação realizada por pessoas distintas ocasionaria diferença de qualidade, uma vez que substituindo o copista, dificilmente se manteria a coerência, devido à falta de parâmetros para a realização destas tarefas (SILVA; FUJITA, 2004).

Em 1440, Johannes Gutenberg aperfeiçoa a prensa móvel, criação atribuída ao chinês Bi Shengainda no século XI, criando uma máquina capaz de reproduzir documentos em uma escala nunca vista anteriormente, através de caracteres gravados em blocos de chumbo ou madeira (TOSSERI, Oliver. Gutenberg não inventou a imprensa)³. Da prensa móvel deriva a palavra imprensa, que designa os meios de comunicação e informação, que tiveram um crescimento significativo a partir do momento em que o processo de reprodução de documentos se tornou mecanizado. A obra *Bibliotheca Universalis*, de autoria de Konrad Gesner, repertório geral europeu que remetia a aproximadamente 12 mil títulos, e seu posterior índice alfabético *Pandectarumsivepartitonumuniversalium, libri XXI*, representam o ápice do surgimento de listas, no século XVI (RIBEIRO, 2003).

A reforma protestante, no século XVII, propiciou a tradução da bíblia e seu acesso ao público. Este fato demandou um esforço de indexação em grande escala, realizado em 1737 por Alexandre Cruden, que lança a primeira concordância completa da Bíblia. Esta concordância relacionava citações com sua localização no texto, aspecto importante que influenciou os próximos índices (SILVA; FUJITA, 2004). No século XVIII Samuel Johnson indexa a língua inglesa em seu Dicionário da Língua Inglesa, estabelecendo passagens a serem representadas e quais termos deveriam tornar-se as entradas (RIBEIRO, 2005). Pela primeira vez foram estabelecidos padrões e diretrizes para a realização do trabalho de indexação,

³Disponível em: <http://www2.uol.com.br/historiaviva/artigos/gutenberg_ao_inventou_a_imprensa.html>. Acesso em: 15 set. 2014.

segundo Silva e Fujita (2004), que afirmam que a partir das orientações propostas por Cruden e Johnson, os procedimentos de indexação foram se disseminando, até que surge pela primeira vez na Alemanha a concepção de *schlagwort* ou palavra-chave, que constituiu um marco importante para a indexação e recuperação da informação.

Com o aperfeiçoamento da prensa móvel e o advento da imprensa industrial, foi possibilitado o crescimento da circulação dos periódicos referenciais no século XVIII, assim como outras formas de acessar os livros contidos nas bibliotecas dos mosteiros, que até então só ocorria através dos índices, segundo Silva e Fujita (2004). As autoras afirmam que a indexação em si, como processo, foi realizada mais intensamente desde o aumento destas publicações periódicas e da literatura técnico-científica, que tiveram um crescimento exponencial observado ao longo dos séculos XVIII e XIX.

Neste momento, as publicações periódicas passaram a incluir listas e índices que remetiam para os documentos originais, e em outros casos, resumos dos trabalhos científicos, filosóficos e artísticos. Ainda no século XVIII, Jean-Baptiste Massieu cria o primeiro código de catalogação nacional em 1791, que embora sem grandes avanços observados em relação a códigos anteriores, aborda aspectos referentes a indexação de documentos (RIBEIRO, 2005).

Ao longo do século XIX, diversos trabalhos contribuíram para a área da indexação e o aprimoramento das técnicas deste processo, como o periódico *Punch, the Illustrated London News*, e *An Alphabetical index to Subjects treated in the Review and other Periodicals*, publicado por Fredrick Poole, que em 1882, criou o *Poole's Index*. Esta obra apresentou um grande avanço para a área ao introduzir índices cumulativos de periódicos, usando palavras-chave retiradas dos títulos dos artigos para representar as publicações indexadas (RIBEIRO, 2005).

Nos Estados Unidos, pouco antes, em 1876, Melvil Dewey publica a Classificação Decimal de Dewey, também conhecida como CDD, mencionada por Silva e Fujita (2004) como a mais expressiva contribuição à indexação. Nesta obra, também inserida na ótica do positivismo, todo o conhecimento é dividido em dez classes principais, subdivididas em números decimais, o que permite uma hierarquia infinita de classes.

Observa-se então que o século XIX compreende um momento em que a Indexação passou por avanços técnicos significativos, e sua execução foi bastante aprimorada. Isto ocorreu em sintonia com a demanda por estes serviços de representação. Não obstante, é possível notar que a prática da indexação, antes executada por profissionais de outras áreas, começa a ser desempenhada por profissionais e especialistas no assunto. Uma vez que uma massa documental crescia em excesso, o público sentia a necessidade de controlar este volume de informação (RIBEIRO, 2005). Desta forma, ocorreu o aumento de instituições dedicadas exclusivamente a indexar livros, refazer índices antigos, elaborar índices retrospectivos, índices cumulativos, índices cooperativos, entre outros (SILVA; FUJITA, 2004). Ribeiro (2005, p. 4) complementa a questão, observando que:

A criação de instituições/serviços destinados a conservar e a gerir as massas documentais nacionalizadas seguiu, em toda a Europa, o modelo institucionalizado na França com os *Archives Nationales*. Mas, as incorporações feitas em bibliotecas implicaram também a criação de espaços para depósitos, uma vez que as estantes das salas de leitura eram manifestamente insuficientes para instalar a documentação incorporada.

Observa-se que neste contexto, ocorreu um movimento a favor da concentração dos serviços de documentação, incluindo a indexação, nas mãos do Estado. Ribeiro (2005, p. 4) contextualiza bem esta questão e levanta suas consequências, ao observar que a Revolução Francesa teve papel preponderante nesta tendência:

A afirmação disciplinar da Arquivologia e da Biblioteconomia e o conseqüente reforço da sua vertente mais tecnicista são conseqüências inegáveis da Revolução Francesa, dos ideais do Liberalismo e de todo o processo de nacionalizações que foi desencadeado, transferindo para a posse do Estado a documentação imprescindível à gestão patrimonial dos bens usurpados às classes dominantes no Antigo Regime e à redação da História da Nação, ideal muito caro aos revolucionários que impuseram a nova ordem liberal. Com a concentração de tão grande quantidade de documentação nas mãos do Estado, vão surgir problemas novos, como, por exemplo, a necessidade de espaços para instalar convenientemente os documentos e a exigência de profissionais qualificados para organizar, catalogar e difundir os acervos reunidos.

O século XX se apresenta como marco do surgimento de diversos bons índices, tanto gerais como específicos. Se no século XIX, um aumento da massa documental propiciou um avanço qualitativo e quantitativo no processo de

indexação, no século XX observamos um refinamento das obras de representação, que evoluíram para índices em vários volumes e esforços para criar índices cooperativos em nível internacional (SILVA; FUJITA, 2004).

H.W. Wilson, em 1901, lança a obra *Reader's Guide to Periodical Literature*, em que artigos são indexados por autor e assunto específico, além de apresentar inúmeras remissivas coordenando assuntos correlatos, o que se configura como avanço significativo em grau de aprimoramento da indexação. Conforme observou Collinson (1971), o trabalho mantém, do início ao fim, um padrão altíssimo de exatidão e uniformidade. Outros índices, como o da *Encyclopaedia Britânica*, considerada por diversos autores como maior representante do aperfeiçoamento da indexação no século XIX, foram aprimorados e passaram a contar com uma clareza maior e apresentação gráfica (SILVA; FUJITA, 2004).

Com o avanço científico e tecnológico propiciado pela II guerra mundial, diversos processos passaram a contar com algum grau de automatização. O surgimento dos computadores e sua utilização em diversos serviços, apresentaram-se como resposta para o tratamento de um volume documental que crescia de forma exponencial, fenômeno conhecido como explosão bibliográfica (RIBEIRO, 2005). Hans Peter Luhn, cientista computacional da IBM, por volta de 1950, propõe uma análise automática da frequência da repetição das palavras em um texto, e cria um modelo de indexação baseado em um ranking de palavras retiradas dos documentos (GIL LEIVA, 1999). Assim, propôs o primeiro método de indexação automática, o KWIC (*Keyword in Context*), analisando as palavras contidas no título dos documentos (VIEIRA, 1988). Robredo (1982, p. 238) define bem a questão, e afirma que “a primeira aplicação generalizada da indexação automática de documentos técnicos, a partir de palavras significativas dos títulos, se deu com KWIC, produzindo índices de títulos permutados”.

Embora simples, o método de Luhn serviu de base para diversas outras proposições de automatização do processo de indexação, que se tornaram cada vez mais refinadas, ao passar a contar com metodologias estatísticas e ponderação de termos. Posteriormente, foram realizados estudos incluindo análises linguísticas quanto a relevância das palavras extraídas, e embora sejam mencionados neste trabalho, não cabe aqui realizar uma análise detalhada destes métodos.

Neste sentido, vale ressaltar que diversos avanços na área da indexação automática no Brasil se devem ao trabalho do professor Jaime Robredo, que compara o processo de indexação automática ao processo de leitura-memorização humano. O autor afirma que o princípio desta indexação é baseado no confronto das palavras do texto com uma relação de palavras vazias de significado previamente estabelecidas, e com base na eliminação, corre a seleção das palavras significativas (ROBREDO, 1982).

Nos anos 60 inicia-se a execução de indexação automática baseada nas técnicas de Processamento de Linguagem Natural (PLN), que elevam a análise, segundo Gil Leiva (1999), a uma componente morfológica, uma sintática e uma semântica. É importante observar que neste momento, profissionais de diversas áreas trabalhavam em conjunto para elaborar estas técnicas, que configuram-se como um processo semiautomático, ou automatizado, ao contar parte com a ajuda de artifícios tecnológicos, parte com análise humana para realização do trabalho de indexação.

Cabe aqui distinguir os termos indexação automática, que na visão de Robredo (1986 apud SILVA; FUJITA, 2004, p. 145), “abarca qualquer procedimento que permita identificar e selecionar os termos que representem o conteúdo dos documentos, sem a intervenção do documentalista”, e indexação automatizada, que envolve o uso de computadores para análise e levantamento de palavras, mas conta também com o trabalho intelectual de um profissional para validar e checar este levantamento (SILVA; FUJITA, 2004).

A partir da década de 80, diversos esforços internacionais culminaram na elaboração de normas, na tentativa de oferecer uma maior padronização da atividade de indexação, como o caso da UNISIST, de 1981, ISO 5696 (*International Organization For Standardization*) de 1985, posteriormente traduzida pela ABNT (Associação Brasileira de Normas Técnicas) na norma ABNT 12676, de 1992. Esta norma apresenta a Indexação como exercício de identificação e descrição do conteúdo de um documento, a partir de descritores ou termos representativos, extraídos de seu assunto. É importante observar que a partir deste momento, começa a ser mencionada também a preocupação de indexar tendo em vista os objetivos para os quais as informações serão selecionadas (SILVA; FUJITA, 2004).

Os anos 90 marcam a popularização dos computadores pessoais, a difusão dos serviços de internet, além das TICs (Tecnologias de Informação e Comunicação). Dada a evolução dos *mainframes*, máquinas enormes utilizadas basicamente pelos governos, até as unidades domésticas, difundidas principalmente pela IBM e pela Apple, a sociedade vivenciou um fluxo de informação em proporções gigantescas, o que impactou no processo de indexação e determinou uma demanda de rapidez na sua execução, a fim de acompanhar este contexto.

A partir deste momento, qualquer pessoa se tornou apta a gerar informações e conteúdo na rede em quantidades muito maiores, proporcionado pela velocidade de processamento crescente dos computadores e da comunicação *online* (GIL LEIVA, 1999). Esta década assiste também o surgimento do *Google* e outros motores de busca, como *Yahoo*, que no início baseavam-se na indexação de páginas através da sua categorização, e posteriormente passaram a possibilitar as meta-buscas através de palavras-chave.

Uma vez propiciado um ambiente global e colaborativo, a *Web* se tornou uma plataforma, e seu conceito ampliado para o de *Web 2.0* (O'REILLY, 2004). A participação ativa dos usuários e da comunidade *online* na criação, desenvolvimento e aprimoramento de *softwares* e recursos *Web* marca esta etapa, que registra também o surgimento das *wikis*, os programas abertos e as mídias sociais, todos estes, alimentados pelos usuários e dependentes de sua participação. A organização da informação na *Web*, uma forma de indexação realizada pelos usuários em linguagem natural, passa a ganhar popularidade (SANTOS, 2013). O termo folksonomia abarca esta maneira de indexar os dados da rede, que se dá através de etiquetagem e o emprego de marcadores, caracterizado como um processo descentralizado (VANDER WAL, 2004).

Neste momento, nota-se que a indexação dos recursos gerados e disponibilizados na rede deixa de ser tarefa exclusiva de profissionais e especialistas, e passa a ser delegada também ao usuário, através de linguagem natural utilizada pelas comunidades. O contexto *Web 2.0* e a etiquetagem de informação na internet são a face mais recente da indexação, neste caso aplicada a recursos digitais, e por isso demandam uma seção dedicada a seu estudo.

2.3 Indexação na Web

2.3.1 Ambiente Web

Posterior ao surgimento da internet, ainda no século XX, a *World Wide Web*(WWW) foi criada por um projeto iniciado em 1989 pela Organização Europeia para Investigação Nuclear⁴(CATARINO; BAPTISTA, 2007).A criação da *Web* se deu, notadamente, devido aos esforços do físico britânico Tim Berners-Lee, com a intenção de criar um ambiente global utilizando-se de protocolos e hipertextos, documentos virtuais não lineares e não dotados de hierarquia (LANDOW, 1992; LÉVY, 2003; LEÃO, 2005; NEGROPONTE; TELLAROLI, 1995) para navegação *online*, de modo a facilitar o compartilhamento de informações e documentos.

Este ambiente *Web* proporcionou o compartilhamento de informações e interação entre usuários sem precedentes, e as pessoas passaram a interferir no processo de geração e disseminação da informação ativamente, de forma fácil e rápida.Conseqüentemente, passaram também a ser agente da organização dos conteúdos gerados. Por se tratar de um ambiente flexível e adaptável, este contexto está em constante evolução, com adição permanente de serviços e funcionalidades que cada vez mais, "permitem que os seus usuários participem de forma ativa na construção e organização dos conteúdos lá disponíveis" (CATARINO; BAPTISTA, 2007, p. 1).

O conceito de *Web 2.0*, cunhado em 2004 por Tim O'Reilly, deriva da *World Wide Web*(WWW), está imerso em um contexto de profundas transformações sociológicas e parte da premissa de que os próprios usuários participam de maneira ativa e colaborativa da geração e organização de conteúdo *online*, na visão de Catarino e Baptista (2007). O'Reilly (2005) afirma que o termo *Web 2.0* não possui uma delimitação claramente definida, mas um núcleo de definições e atitudes que permeiam este ambiente social e colaborativo. Esta seria a segunda geração da *Web*, e neste momento histórico, mais do que nunca, o usuário se torna protagonista na geração não só de informação e conteúdo *online*, mas na criação e propagação dos serviços na rede.Além disto,o usuário torna-se agente de colaboração e promove a evolução desta plataforma, tendência nomeada por O'Reilly (2005) como

⁴Link: <http://info.cern.ch>

“arquitetura da participação”, o que, na sua visão, não estava presente na primeira geração dos serviços *online*.

Neste momento, o usuário passa a fornecer seus “dados e serviços de forma a que sejam reorganizados por outros” (CATARINO; BAPTISTA, 2007, p. 1) e as funcionalidades, plataformas e *softwares* derivadas deste contexto passam a ter atualização constante. Na visão de Santos (2013), essas aplicações agora configuram-se como uma versão *beta* constante, em alusão à versão de *softwares* considerada ainda inacabada. Se antes esperava-se que uma grande corporação gerasse um *software* ou aplicativo para suprir uma demanda do usuário, neste momento o próprio usuário contribui ativamente na geração destas funcionalidades, como no caso do sistema Linux, em que o código do sistema operacional é aberto para personalização e infinitas versões podem ser criadas e difundidas, de acordo com a experiência dos usuários e seu perfil de uso. Percebe-se, neste momento, uma inclinação dos usuários a desempenhar suas atividades em aplicativos *online*, via navegadores ou *browsers*, e a alocação dos recursos na nuvem ou rede, e não mais em *softwares offline* em memória física.

A *Web 2.0* passa a ser tratada como uma atitude, e não uma tecnologia (DAVIS, 2005 apud CATARINO; BAPTISTA, 2007), deixa de ser uma rede para tornar-se uma plataforma, em que o que mudou foi nossa forma de entendê-la (OLIVEIRA, 2006 apud CATARINO; BAPTISTA, 2007). Exemplo disto são a atualização constante de *softwares* abertos, passando pela construção coletiva de conteúdos por meio das *wikis* e mídias sociais. É importante notar que neste momento, a coleta, tratamento e disseminação de dados e informações para a comunidade, que antes dependia de um serviço especializado, agora podem ser realizados pelo usuário de maneira colaborativa, que passam a alimentar bancos de dados disponíveis virtualmente, que na visão de Strehl (2011, apud SANTOS, 2013, p. 95), “tão melhores ficam quanto mais pessoas os utiliza”.

A partir das possibilidades propiciadas pela plataforma *Web 2.0*, popularizou-se uma forma de criação e compartilhamento em que o conteúdo na *Web* é gerado pelo próprio usuário de forma ativa. Desta forma, os *softwares*, mídias sociais e *wikis* passam a ser uma espécie de repositório, que fornece interface e funcionalidades para possibilitar esta geração de conteúdo e a edição das informações compartilhadas, além da interação entre os usuários, na visão de

Santos (2013). Segundo o autor, as mídias digitais passaram a permitir que “o usuário reinterprete, parodie e comente categorias e enunciados, como videocliques, novelas, filmes, da mesma forma que permite a esses usuários compartilharem conteúdos de autoria própria”.

É importante mencionar que existem correntes de estudiosos que discordam do termo 2.0 por entenderem que a *Web* sempre foi aberta a participação dos usuários, e nesta visão, esta plataforma não se trata de uma revolução, mas uma evolução. Davis (2005) observa que a *Web* sempre teve como objetivo ser colaborativa, e sem essa base não teria chegado a sua forma atual, e acredita que a *Web 2.0* seja um conjunto de esforços para que se retome sua forma original. De toda forma, fica patente que hoje a participação dos usuários é estimulada como nunca foi, e que este modelo colaborativo é essencial para a manutenção e organização do fluxo atual de informação.

Consoante aos tempos atuais, em que a explosão e oferta de informação atingem um patamar inimaginável antes do advento da internet, o modelo *Web 2.0* propicia maior velocidade e praticidade no tratamento e organização dos recursos digitais. Estas evoluções, na visão de Catarino e Baptista (2007) podem ser consideradas como um novo paradigma na organização e indexação dos conteúdos *online* e recursos da *Web*, em que o usuário passa a deter o poder não só da geração e propagação das informações na *Web*, mas da sua organização e etiquetagem colaborativa, a fim de tornar possível que estas informações sejam recuperadas.

2.3.2 Etiketagem de Recursos na Web

Etiketagem de recursos na *Web* é um termo genérico que abrange a prática da indexação dos documentos gerados e disponibilizados *online*, de forma natural e não hierárquica. É realizada pelo próprio usuário ou grupos de usuários a fim de organizar as informações *online* no contexto da *Web 2.0* de forma colaborativa (SANTOS, 2013).

Não existe consenso na área de representação da informação sobre a melhor nomenclatura a ser adotada para designar este tipo de indexação na *Web*, apesar de diversos termos terem sido criados para definir esta prática. Desta forma,

pretende-se nesta seção, apresentar de que forma a etiquetagem vem sendo praticada nas aplicações *online* da *Web 2.0*, bem como algumas visões sobre seu uso.

Uma vez que no contexto da etiquetagem os próprios usuários se utilizam de ferramentas cooperativas para organizar a informação de forma que possam recuperá-la por meio de uma busca apoiada em conexões e significados, na visão de Santos (2013), constata-se uma mudança nos padrões de organização e representação da informação. Essa quebra de paradigma trouxe vantagens e desvantagens para a indexação e recuperação da informação, que por um lado torna ágil e em tempo real as atividades de representação, mas por outro pode se tornar pouco precisa e prejudicar o processo de busca dos dados.

As informações e objetos alvos da etiquetagem são definidos por Vander Wal (2004) como qualquer dado dotado de *URL*⁵, ou seja, localização e endereço virtual. Miller (1998) adota a nomenclatura recursos para estas mesmas informações em espaço *online*. Nesta seção serão adotadas ambas as definições para se referir à informação disponibilizada na *Web*.

Segundo Brandt (2009), etiquetagem é uma maneira de as próprias pessoas ou usuários da informação realizarem a indexação, de modo a classificar recursos e objetos informacionais. Vander Wal (2005) aborda o fato de esta ser uma indexação não hierárquica, em que a informação é localizada por meio de correlações associativas e inferências, ou seja, as palavras que serão usadas para recuperar a informação não seguem nenhum padrão, sendo basicamente resultado de instintos de demanda de informação *online*. O autor em questão reconhece que as hierarquias foram uma maneira eficiente de organizar informações antes deste momento histórico, marcado pela popularização dos computadores e da *Web*.

Diferentemente da indexação e representação do conhecimento, em que, em linhas gerais, um profissional realiza um trabalho técnico de leitura, análise de assunto e extração de termos com base em listas pré-definidas de conceitos, a etiquetagem é uma indexação livre criada pelos próprios usuários, muitas vezes os próprios geradores da informação *online*, com base nas suas próprias percepções sobre os documentos (Santos 2013).

⁵ Abreviação de *Uniform Resource Locators*, que se refere a localização de informação na rede

Conforme explanam Catarino e Baptista (2007, p. 4), etiquetagem significa atribuir etiquetas aos recursos da *Web*, por meio de linguagem natural, em que não são adotadas regras e/ou políticas de indexação nem o controle dos vocabulários, e afirma que “o ato de etiquetar é do próprio usuário da informação (...) não é o autor nem o profissional de indexação que indexam o recurso *Web*; é o seu próprio usuário”. Santos (2013) defende que a etiquetagem trata-se de uma indexação colaborativa e democrática, centrada no usuário, que colabora com seus próprios termos advindos de seu contexto e experiência prévia. Afirma ainda que esta prática flexibiliza formas de representação da informação anteriores, por ser permeada pelo ideal de cooperação inserido na plataforma *Web 2.0*.

Os termos “classificação social”, “classificação distribuída”, “etnoclassificação”, “*bookmarking* social”, e “ontologia social” também são termos utilizados para designar a prática de etiquetagem dos recursos *Web* (CATARINO; BAPTISTA, 2007, p. 4). Algumas definições dão mais ênfase ao cunho social e colaborativo (PETERSON apud CATARINO; BAPTISTA, 2007), outras, como no caso da “etnoclassificação”, defendida por Merholz (2004, apud MATHES, 2004), contém no nome o prefixo que indica popular, em alusão ao fato de ser uma classificação realizada pelas pessoas. A palavra distribuída, referente à “classificação distribuída”, tem sido empregada para designar o caráter múltiplo da atividade, que é realizada por diversas pessoas e alterada com o tempo (Hammond, 2005). As definições relacionadas a *bookmarkings* são mais restritas à etiquetagem relacionada a um *URL*, sendo a representação do endereço *online* indexado através de marcadores ou favoritos.

Todos estes termos, porém, se relacionam a esta mesma atividade de representação da informação *online* e de etiquetagem de recursos na *Web*. Folksonomia, termo cunhado por Thomas Vander Wal em 2004, é o mais difundido para nomear esta prática (CATARINO; BAPTISTA, 2007). Este termo possui o prefixo *folk* relacionado à pessoas, e deriva da palavra taxonomia, método pelo qual os sistemas de representação, organização e recuperação de informações funcionam, agrupando informações semelhantes (SANTOS, 2013).

A palavra taxonomia está relacionada a organização das informações por meio de classes e subclasses, através de listagens construídas por pessoal especializado. No caso da Folksonomia, não se utilizam listagens de termos

previamente elaborados ou vocabulário controlado para sua realização, e por isso este termo tem sido criticado por alguns autores. Guy e Tonkin(2006) observam que a folksonomia é diferente das taxonomias, por designar um sistema de classificação distribuído, realizado pelas próprias pessoas e usuários individuais em linguagem natural, diferente do que ocorre nas taxonomias.

Não existe consenso quanto ao termo mais adequado para designar etiquetagem na *Web*,mas Folksonomia é um termo muito empregado pelos estudiosos das áreas de informação, que o utilizam para designar a indexação de recursos digitais na *Web*.Apesar desta prática também ser chamada de classificação social ou *socialtagging*, segundo Catarino e Baptista (2007), Folksonomia vem sendo a forma mais usada para denominar a etiquetagem da informação. Vander Wal (2006 apud CATARINO; BAPTISTA, 2007) define a folksonomia como o resultado da atribuição livre, natural e pessoal de etiquetas a informações ou objetos virtuais, visando sua recuperação em um ambiente social aberto.

Morrison (2008) defende que o termo Folksonomia deve ser estritamente utilizado para se referir à utilização de marcadores sem outros recursos, e define esta prática como uma coleção de termos constituída por contribuições de usuários, em que os mesmos participam da classificação e avaliação dos itens a serem avaliados, em um ambiente social virtual. Catarino e Baptista (2007, p. 5) acrescentam que este ambiente social no qual a Folksonomia se desenvolve, necessariamente é compartilhado e aberto à comunidade, e que tem como finalidade a recuperação da informação, assim como a indexação. Além disso, as autoras destacam três fatores essenciais da Folksonomia: “1) é resultado de uma indexação livre do próprio usuário do recurso; 2) objetiva a recuperação a posteriori da informação e 3) é desenvolvida num ambiente aberto que possibilita o compartilhamento e, até, em alguns casos, a sua construção conjunta”.

Existem diferentes visões a respeito da natureza da Folksonomia. Alguns autores defendem ser o resultado de um processo ou produto, no qual se encaixam o próprio criador do termo, Thomas Vander Wal,e outros, como Peterson (2006), Russel (2005) e Guy; Tonkin(2006) que acreditam que a Folksonomia é uma abordagem ou metodologia, considerando-a como um novo paradigma ou metodologia na área da indexação. A definição de Folksonomia como “resultado da etiquetagem na *Web*” (VANDER WAL, 2005) e como resultado de um sistema de

classificação desenvolvido de forma cooperativa (TRANT, 2006) mostram a inclinação destes autores para designar Folksonomia como parte final do processo de indexação na *Web*, ou seja, a prática de etiquetar informações *online*.

Catarino e Baptista (2007, p. 3) resumem bem as aplicações da Folksonomia na *Web*, evidenciando sua forma de funcionamento na prática:

Sucintamente, pode-se dizer que as ferramentas de folksonomia permitem que usuários da *Web* indexem os recursos a partir da atribuição de etiquetas para seu armazenamento, organização e recuperação. Além disto, estas ferramentas permitem que as etiquetas fiquem disponíveis em rede (na *Web*), de forma que outros usuários que tenham os mesmos interesses possam aceder aos recursos, bem como mostram as várias formas pelas quais um mesmo recurso foi indexado por outros. É uma maneira colaborativa e livre de indexar que geralmente não se pauta em nenhum vocabulário controlado ou qualquer outro sistema predefinido de classificação tradicional.

A Folksonomia pode ser percebida claramente nas mídias sociais, da etiquetagem pelos próprios usuários com *tags* no Flickr⁶, a criação de páginas para agrupamento de informações semelhantes no Facebook⁷, até o uso de *hashtags* no Twitter⁸ e Instagram⁹, de forma que todas as vezes que uma expressão é mencionada após o sinal de #, o conteúdo é integrado a um banco de dados comum, sendo indexado por esta etiqueta (SANTOS, 2013). Outras mídias sociais, como o caso do Delicious¹⁰, permitem o agrupamento de listas de favoritos, mais relacionadas ao conceito de *bookmarking social*, nada mais do que outra nomenclatura para etiquetagem na *Web*.

Na prática, a Folksonomia se dá quando um usuário, ao publicar uma informação na *Web*, seja ela um texto de blog, foto, vídeo ou pensamento, atribui palavras-chave, etiquetas ou *hashtags* a fim de representar esta informação *online*. Uma vez que etiquetas são atribuídas a uma postagem, esta passa a pertencer a um banco de dados que contém todas as informações etiquetadas com as mesmas palavras-chave. Isto possibilita, além do agrupamento destes conteúdos, a recuperação da informação por parte de qualquer usuário que busque pelas etiquetas utilizadas, ou seja, compartilhe dos mesmos interesses.

⁶<http://flickr.com>

⁷<http://facebook.com>

⁸<http://twitter.com>

⁹<http://instagram.com>

¹⁰<http://delicious.com>

Seja por motivações sociais ou pessoais, as pessoas estão aderindo à Folksonomia, e independente dos fatores que incentivam seu uso, a etiquetagem na *Web* tem se revelado fator importante para que a informação disponibilizada seja indexada e organizada, mesmo que isso muitas vezes ocorra de forma desordenada (SANTOS, 2013). Por isso, torna-se importante pensar maneiras de melhorar os mecanismos de indexação na *Web*, como as sugestões de etiquetas com base em vocabulários controlados, modelo já utilizado pelo Delicious.

3. PROCEDIMENTOS METODOLÓGICOS

A metodologia, disciplina relacionada a epistemologia, está ligada a origem e a validade do conhecimento, e trata de estudar e avaliar os vários métodos disponíveis, identificando suas limitações ou não em nível das implicações de suas utilizações, como um conjunto de procedimentos utilizados na obtenção de determinado conhecimento (BARROS; LEHFELD, 2000).

De modo geral, a metodologia é utilizada com o intuito de avaliar e examinar as técnicas de pesquisa, além da geração e verificação de novos métodos que conduzem à captação e ao processamento de informações para que se chegue à resolução de problemas de investigação. A metodologia surge então, como um conjunto de procedimentos utilizados na obtenção de determinado conhecimento (BARROS, 2000).

Para que este trabalho fosse desenvolvido, foi realizada uma revisão de literatura, apresentada na seção 2, que abrange o conteúdo desta pesquisa. A revisão de literatura, segundo Santos (2006), tem papel fundamental no trabalho acadêmico, pois por meio dela é possível situar a área de estudo em que o trabalho se encontra, contextualizando-o.

3.1. Etapas da pesquisa

O estudo consistiu em uma pesquisa histórico-descritiva, com base em levantamento documental e revisão de literatura, e abarcou análise de livros, teses, dissertações, monografias, ou seja, diversos tipos de materiais bibliográficos relacionados ao assunto do trabalho.

Após ter sido delimitado o tema e objeto da pesquisa como a prática da indexação até os dias atuais, a coleta de dados foi feita a partir do levantamento documental. Foram escolhidos textos e artigos de autores referência nas áreas de indexação e representação da informação, com foco na prática, *Web 2.0* e etiquetagem, pesquisados em bases de dados, repositórios institucionais e Google acadêmico. Posteriormente, estes artigos foram resumidos, fichados e organizados por ordem cronológica referente ao assunto abarcado, sintetizados e tiveram

passagens importantes listadas para servirem de arcabouço teórico para elaboração do estudo.

Estes passos são detalhados a seguir:

1. Delimitar o tema: a fim de entender de que forma a indexação se adaptou às mudanças na sociedade, optou-se por abordar a prática da representação da informação e indexação até os dias atuais, bem como a plataforma *Web 2.0*, ambiente no qual a indexação em sua forma atual, a etiquetagem, ocorre na internet;
2. Buscar documentos bibliográficos que atendam aos assuntos da pesquisa: pesquisar diversas fontes, como o Repositório Institucional da UnB, a Scielo (Scientific Electronic Library Online), Portal de Periódicos da UFMG (Perspectivas em Ciência da Informação), Portal CRUESP, que unifica as bases da USP, UNESP e UNICAMP, Pesquisa Brasileira em Ciência da Informação e Biblioteconomia, Transinformação (Portal de Periódicos da PUC-Campinas), Google Acadêmico, Por Dentro da Pesquisa (Google), outros periódicos. Os termos utilizados para realizar a busca, em português e inglês, são “prática da indexação”, “história da indexação”, “evolução da indexação”, “folksonomia”, posteriormente combinados com “*Web 2.0*” e “recuperação da informação”, realizados em cada fonte nos campos “Título”, “Resumo”, “Palavras-chave”, e no próprio texto, quando o artigo estiver inteiramente disponível.
3. Resumir e fichar os textos encontrados: Após realizar a leitura crítica, identificar as partes principais nos materiais pesquisados, considerando a delimitação da pesquisa, e realizar o resumo;
4. Organizar cronologicamente: Uma vez que um dos objetivos do estudo foi traçar um panorama histórico acerca da prática da indexação, tornou-se necessário separar os materiais levantados por ordem cronológica referente aos assuntos abordados;
5. Sintetizar as informações: articular os conteúdos destacando o contexto e como se dava a prática da indexação no momento histórico;

6. Discutir os marcos históricos e práticas da indexação identificadas.

4. DESENVOLVIMENTO

O presente trabalho trata-se de uma monografia para conclusão de curso de graduação de Biblioteconomia da Universidade de Brasília, que se compõe de revisão de literatura concentrada na prática da representação da informação e indexação, na qual foram contextualizadas a área de pesquisa e apresentadas as visões dos principais autores que abordam o tema, além de uma discussão acerca da aplicação da indexação em ambiente *Web* nas suas formas atuais.

Sob a ótica da delimitação do tema, concentrado na prática da indexação até os dias atuais, diversas fontes foram pesquisadas, incluindo o Repositório Institucional da UnB, a Scielo (Scientific Electronic Library Online), Portal de Periódicos da UFMG (Perspectivas em Ciência da Informação), Portal CRUESP, que unifica as bases da USP, UNESP e UNICAMP, Pesquisa Brasileira em Ciência da Informação e Biblioteconomia, Transinformação (Portal de Periódicos da PUC-Campinas), Google Acadêmico, Por Dentro da Pesquisa (Google), e outros periódicos. Os materiais utilizados como referência para elaboração do pré-projeto também foram referenciados no trabalho final.

Os termos utilizados para realizar a busca, em português e inglês, foram “prática da indexação”, “história da indexação”, “evolução da indexação”, “folksonomia”, posteriormente combinados com “*Web 2.0*” e “recuperação da informação”, realizados em cada fonte nos campos “Título”, “Resumo”, “Palavras-chave”, e no próprio texto, quando o artigo estivesse inteiramente disponível. Cabe ressaltar que como o presente trabalho objetivou chegar às conclusões a respeito de como a indexação se adaptou ao contexto atual, as pesquisas foram focadas em artigos publicados após a década de 1990. Para a revisão de literatura relativa à área de etiquetagem e folksonomia, foram levantadas publicações posteriores a 2004, ano em que o termo foi cunhado por Thomas Vander Wal.

Foram levantados textos e artigos de autores referência nas áreas abordadas, posteriormente realizada leitura crítica, e os textos resumidos, a fim de extrair citações e passagens importantes que contribuíssem para a elucidação dos objetivos propostos, de modo a traçar um panorama histórico da prática da indexação. Considerando que não existe consenso sobre os métodos de classificação da informação, pois cada autor defende um critério, por vezes muito

semelhantes. (APOLINÁRIO, 2006), optou-se por organizar o conteúdo em ordem cronológica quanto ao assunto abarcado.

Seguindo estes passos, foi possível levantar diversas considerações a respeito da representação da informação e indexação, identificando seus objetivos de facilitar a recuperação da informação (LANCASTER, 2004). Notou-se o fato de que desde sempre, associados a produção e organização do conhecimento, existiram formas e esforços para representar a informação (RIBEIRO, 2005). Foi possível também identificar vertentes distintas que versam sobre ao pertencimento ou não da indexação à Análise Documentária, apontadas por Silva e Fujita (2004) como a corrente inglesa, que considera Análise Documentária e Indexação sinônimos, e a corrente francesa, que compreende Indexação como parte final do processo de Análise Documentária.

Ao levantar marcos históricos relativos a prática da indexação, foi possível entender que a indexação se adaptou aos novos contextos conforme a sociedade evoluiu. Nota-se esta adaptação ao perceber que inicialmente, as experiências de indexação foram calcadas em uma prática manual e simples, de forma intuitiva (PINTO, 2000), que tomou formas mais especializadas, envolvendo técnicas e o uso do computador posteriormente (GIL LEIVA, 1999).

Com o surgimento da internet e da *Web 2.0*, rede de compartilhamentos *online*, as informações geradas na internet passaram a ser indexadas no momento de sua criação ou compartilhamento, pelo próprio usuário da informação, com linguagem natural (WAL, 2006), que nomeou a prática como Folksonomia.

O trabalho se desenvolveu seguindo a ordem abaixo descrita:

1. Introdução: apresenta uma breve discussão sobre os temas que serão abordados, os objetivos do trabalho, justificativas e o resultado esperado.
2. Revisão de Literatura: abarca todos os subtemas relacionados às disciplinas desta monografia.
3. Indexação: definições, concepções e histórico da prática da indexação.
4. Definições e Conceitos: apresenta a literatura na área de Ciência da Informação abrangendo a conceituação da indexação, com foco na sua prática.
5. Diferentes Concepções: uma breve discussão acerca das visões das correntes inglesa e francesa, e a concepção de indexação como processo.

6. Histórico: versa sobre os primeiros métodos de representação da informação e como a prática da indexação evoluiu ao longo do tempo, até sua aplicação no contexto atual *Web*.
7. Indexação na *Web*: levantamento bibliográfico sobre o ambiente *Web* e como se dá a indexação neste contexto.
8. Ambiente *Web*: trata do ambiente colaborativo e participativo da *Web 2.0*
9. Etiquetagem de Recursos na *Web*: aborda a indexação na *Web 2.0*, realizada pelo próprio usuário com linguagem natural.
10. Desenvolvimento, Resultado e Considerações finais: são discutidos os temas abordados, forma de desenvolvimento do trabalho, apresentado um quadro relacionando o período histórico aos marcos e a contribuição e/ou prática da indexação no período, e feitas considerações a respeito da evolução do processo de indexação e sua adaptação ao ambiente virtual.

5. RESULTADO

Através da revisão de literatura, foi possível identificar marcos históricos que definiram mudanças na prática da indexação e levantar aspectos que influenciaram esta tarefa, até sua caracterização atual, na *Web 2.0*.

Esta seção dedica-se a apresentar a síntese do trabalho na forma de um Quadro (Quadro 1), bem como uma discussão acerca da adaptação da indexação ao longo dos diferentes contextos pelos quais a sociedade avançou e o que influenciou estas mudanças, com enfoque maior no momento atual, por ser um tema menos explorado e ainda escasso de bibliografia especializada. Buscou-se avaliar a prática dos conceitos abordados na revisão de literatura, coordenando-os com contexto atual.

Quadro1 - Evolução da prática da representação e da indexação até os dias atuais

PERÍODO	MARCO/CONTEXTO	CONTRIBUIÇÃO/PRÁTICA
XVII a.C. aproximadamente	Antiga Suméria/Mesopotâmia.	Resumo do conteúdo dos livros antigos, como acesso aos textos originais (WITTY, 1973 apud SILVA; FUJITA, 2004).
III a.C. a IV d.C.	Biblioteca de Alexandria.	Seguindo a classificação de Calímaco, foi criado um catálogo chamado Pinakes, que condensava as principais obras em forma de resumos, ordenado por tema e autor (RIBEIRO, 2005).
Séc. II	Roma. Guia <i>De Libris Propriis Liber</i> foi compilado por Cláudio Galeno.	Criação de um único documento que condensava compilações de obras isoladas, que continham cabeçalhos de capítulos na margem dos parágrafos (GARRIDO ARILLA, 1999).
Séc. V	Grécia. Obra anônima <i>Apothegmata</i> .	Apresentava uma listagem de provérbios gregos sobre assuntos teológicos, e foi a obra que mais se aproximou do atual índice alfabético de assuntos (GARRIDO ARILLA, 1999). A partir deste momento as obras aparecem arranjadas em capítulos numerados, permitindo a localização de suas partes (SILVA; FUJITA, 2004).
Séc. XIV	Europa. Idade	Elaboração de inventários/catálogos

	Média/Mosteiros.	dos livros constantes nos mosteiros (COLLINSON, 1971).
Séc. XVI	<i>Bibliotheca Universalis e Pandectarumsivepartitonumu niversalium, libri XXI</i> , obras de Konrad Gesner.	Elaboração de um repertório geral europeu que remetia a aproximadamente 12 mil títulos e seu índice alfabético, considerado até então o ápice do surgimento de índices (RIBEIRO, 2003).
Séc. XVIII	Reforma Protestante/Concordância completa da Bíblia, por Alexandra Cruden.	Publicada em 1737, consistia em uma obra com uma seção de citações relacionada a sua localização no texto (SILVA; FUJITA, 2004).
Séc. XVIII	Dicionário da Língua Inglesa, por Samuel Johnson.	Publicada em 1755, a obra indexou a língua inglesa, determinando quais termos deveriam ser as entradas e partes a serem representadas (COLLINSON, 1971).
Séc. XVIII	Surgimento na Alemanha da concepção de palavra-chave.	Passou-se a utilizar as <i>schlawortna</i> representação de itens com sistema de escolha das palavras-chave (SILVA; FUJITA, 2004).
Séc. XVIII	Advento da imprensa industrial.	Aumento da circulação de periódicos referenciais, que apresentavam de resumos das novas obras científicas, a listas e índices que remetiam às obras originais (SILVA; FUJITA, 2004).
Séc. XVIII	Criação do primeiro código de catalogação nacional, por Jean-Baptiste Massieu.	Embora sem grandes avanços, a publicação de 1791 foi a primeira grande obra a apresentar diretrizes para a indexação de documentos (RIBEIRO, 2005).
Séc. XIX	Publicação da Classificação Decimal de Dewey (CDD), por Melvil Dewey.	Publicada em 1876, passou a utilizar classificação analítica, considerada a maior contribuição a indexação até então, ao dividir todos o conhecimento em dez classes principais e oferecer subsídios para que as obras fossem melhor indexadas segundo sua categoria (SILVA; FUJITA, 2004).
Séc. XIX	Publicação do <i>Poole's Index</i> , por Fredrick Poole.	Publicada em 1882, introduziu índices cumulativos de periódicos, utilizando palavras-chave retiradas dos títulos das publicações (RIBEIRO, 2005).
Séc. XIX	Êxito dos <i>Archives Nationales</i> da França.	Após a experiência francesa, os países europeus começam criar

		instituições dedicadas exclusivamente a conservar e gerenciar massas documentais (SILVA; FUJITA, 2004), contando com o trabalho de profissionais especialistas no assunto (RIBEIRO, 2005).
Séc. XX	Publicação da obra <i>Reader's Guide to Periodical Literature</i> , por H.W. Wilson.	Publicada em 1901, indexou os artigos por autor e assunto específico, além de apresentar inúmeras remissivas coordenando assuntos correlatos (COLLINSON, 1971).
Déc. 1950	Método <i>KWIC</i> de indexação automática.	Proposto por Hans Peter Luhn, este foi o primeiro método a propor uma análise automática e computadorizada de documentos, em que a indexação se daria com base na frequência das palavras (VIEIRA, 1988).
Déc. 1960	Técnicas de PLN de indexação semi-automática.	Integrando profissionais de várias áreas, passam a considerar variáveis morfológicas, sintáticas e semânticas na indexação, que passa a ser automatizada, ou seja, conta com análise humana, além do auxílio de computadores (GIL LEIVA (1999).
Déc. 1980	UNISIST e ISO 5696.	Esforços internacionais com objetivo de padronizar a atividade de indexação, que a define como exercício de identificação e descrição do conteúdo de um documento, a partir de descritores ou termos representativos, extraídos de seu assunto. Concepção de indexação levando em consideração o objetivo da informação (SILVA; FUJITA, 2004).
Déc. 1990	Popularização dos computadores pessoais, TICs e Internet.	Este contexto proporcionou um crescente fluxo de informação digital, que passou a contar com a possibilidade de ser gerado por qualquer usuário conectado (GIL LEIVA, 1999). Surgimento dos motores de busca como <i>Google</i> e <i>Yahoo</i> , que realizam buscas com base no texto completo ou em marcações de conteúdo.

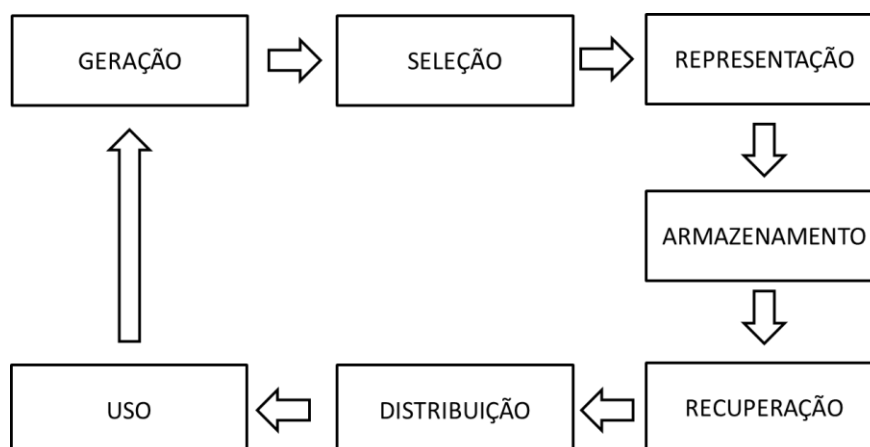
Déc. 2000 em diante	Etiquetagem na <i>Web 2.0</i> .	Imersos em um contexto colaborativo e digital, os próprios usuários passam a indexar a informação compartilhada na rede com linguagem natural, através de etiquetas. Esta prática é chamada Folksonomia (WANDER WAL, 2004).
---------------------	---------------------------------	---

Fonte: Elaborado pelo autor

A importância histórica da elaboração de índices e resumos, produtos da indexação, é inegável. Através dos dados levantados, foi possível compreender que desde os primórdios da prática do tratamento e organização da informação o homem percebeu que sem representar o conteúdo dos documentos gerados pela sociedade, seria impossível fazer com que as informações contidas nos suportes chegassem até o usuário. Gomes e Gusmão (1983 apud SILVA; FUJITA, 2004) abordaram esta questão, ao defender que as mais primitivas formas de ordenação dos documentos foram resposta a demandas relacionadas a tornar acessível a informação. Sem ordenar e representar, seria impossível que a informação gerada fornecesse subsídios ao avanço das comunidades, sendo a matéria prima para a criação de mais e mais informação.

A teoria do ciclo informacional (Figura 1), mencionada na revisão de literatura, corrobora esta tese, ao pressupor as etapas de Geração, Seleção, Representação, Armazenamento, Recuperação, Distribuição, e Uso. Desta forma, a etapa de Uso realimenta o processo de Geração, formando um ciclo que propicia o desenvolvimento do conhecimento na sociedade, conforme observou Guimarães (2003). Sabe-se que o ciclo inicia quando existe uma dúvida ou lacuna de conhecimento, chamado também de necessidade informacional, e que o usuário, neste momento, recorre à recuperação da informação para sanar sua questão. Percebe-se, desta forma, que uma quebra em qualquer das etapas prejudica o ciclo, incluindo a fase de representação, efaz com que novas informações não sejam geradas.

Figura 1 - Ciclo informacional



Fonte: Adaptado de POJUÁN DUARTE, 1998

No caso específico da Indexação, assunto deste trabalho, percebe-se que um documento não indexado ou mal representado não será encontrado pelo usuário, sendo assim inútil para o desenvolvimento de posteriores pesquisas e impossibilitando a criação de novos documentos a partir de seu conteúdo, conforme observou Booth(2001 apud GUIMARÃES, 2009).

Os primeiros homens a lidar com a informação, mesmo não dotados de arcabouço teórico que permitisse discernir sobre as melhores técnicas de representação, perceberam que todo o documento criado, para ser encontrado e ser útil a quem necessitasse, deveria passar por um processo de tratamento que possibilitasse sua recuperação a quem interessasse. Conforme abordado em sessão anterior, os primeiros métodos de representação foram os resumos dos documentos na antiga Suméria e Mesopotâmia(WITTY, 1973 apud SILVA; FUJITA, 2004), e na Biblioteca de Alexandria, que adotou a classificação de Calímaco e condensou compilações de obras, ordenadas por tema e nome do autor (RIBEIRO, 2005).

Nos primeiros séculos da era atual até a Idade Média, diversas compilações e listagens de obras foram criadas, como o guia *De Libris Propriis Libris*, de Cláudio Galeno (GARRIDO ARILLA, 1999) e a obra anônima *Apothegmata*(SILVA; FUJITA, 2004), que apresentaram evoluções e tentativas de melhor representar os documentos, como anotações de rodapé, cabeçalhos e capítulos numerados que permitiam a localização de suas partes, até a criação de índices. Importantes catálogos e repertórios foram publicados na Europa, do século XIV ao XVI, em um esforço de organizar e tornar possível a localização das obras

publicadas ou simplesmente ordenar melhor a informação, como a *Bibliotheca Universalis* de Konrad Gesner (RIBEIRO, 2003) e o dicionário da Língua Inglesa, de Samuel Johnson (COLLINSON, 1937).

Estas obras apresentam grandes avanços para a área da Indexação: a *Bibliotheca Universalis* por introduzir um índice alfabético que remetia a aproximadamente doze mil títulos, e o Dicionário da Língua Inglesa por ser um grande esforço de indexação trazendo a concepção de entrada para descrever as partes a serem representadas. Percebe-se, neste contexto, um grau maior de refinamento das práticas de indexação, buscando estratégias de organização e melhor disposição das informações representadas, como ordenamento alfabético e uma ideia primitiva do que viria a ser chamado posteriormente de descritor.

A elaboração de catálogos e índices foi amplamente estimulada com o aumento da circulação de informação, proporcionado pela imprensa e pela popularização das publicações periódicas no século XVIII. Novas formas foram sendo desenvolvidas para suprir uma demanda crescente de organização da informação gerada, como as listas de índices e os periódicos referenciais, que apresentavam resumos das novas obras científicas, literárias e artísticas (SILVA; FUJITA, 2004).

Neste contexto, concepções como a de palavra-chave, surgida na Alemanha sob a denominação de *schlawort*, tornam-se importantes para melhor representar o volume crescente de publicações. Esta técnica foi empregada por Fredrick Poole no *Poole's Index*, que apresentava índices cumulativos de periódicos, formulados com base na extração de palavras-chave retiradas dos títulos das publicações (RIBEIRO, 2005).

O século XIX foi marcado pela institucionalização, pelos países europeus, da conservação e gerenciamento das crescentes massas documentais. Seguindo o exemplo da França, que durante a Revolução Francesa criou o *Archives Nationales*, instituição pública responsável pela guarda e tratamento da produção informacional francesa, outros países europeus criam suas próprias bibliotecas nacionais (SILVA; FUJITA, 2004; RIBEIRO, 2005). Analisando este momento histórico, é possível notar duas características importantes da prática da indexação neste século: a apropriação de tarefas de tratamento da informação pelo Estado que, tomado por ideais de nacionalismo e liberdade, passa a reconhecer a importância de gerenciar a

massa documental e produção científica produzida; o emprego de profissionais especializados para desempenhar as tarefas de representação da informação, antes fruto de iniciativas predominantemente individuais.

O século XX, permeado pela II guerra mundial, o avanço tecnológico sem precedentes e explosão informacional por este contexto desencadeada, foi cenário de diversas evoluções na área da Indexação, como a automatização dos processos ou parte deles, constituindo algumas das maneiras encontradas para aumentar a rapidez na representação da informação. Uma maior agilidade no tratamento da informação tornou cada vez mais urgente a medida que se tornava cada vez mais difícil lidar, através dos métodos tradicionais, com o crescimento da massa documental. Outra necessidade identificada foi a de coordenar esforços internacionais para a padronização dos processos de indexação (SILVA; FUJITA, 2004), o que é reflexo de uma sociedade cada vez mais globalizada e da flexibilização das fronteiras. O UNISIST e o ISO 5696 foram importantes iniciativas para oferecer diretrizes internacionais com a finalidade de padronizar a indexação.

Observou-se que o aumento da produção de materiais, incluindo livros, artigos, publicações impressas, entre outros, culminou em uma quantidade enorme de informação, e a informação é elemento chave na atual sociedade. A informação configura-se atualmente como fator de inclusão e exclusão social dos indivíduos, que precisam estar informados e atualizados para exercerem sua cidadania, além de ser um elemento de alto valor agregado para a produção científica, industrial e empresarial.

Sabemos que o desenvolvimento tecnológico propicia o incremento da produção de conhecimento, bem como sua fragmentação, uma vez que novos campos do saber são criados e integrados de forma interdisciplinar. Ao passo que a produção de conhecimento aumenta, faz-se necessário torná-la acessível ao usuário de maneira rápida e precisa para acompanhar o ritmo de produção, de forma que os documentos gerados sejam recuperados e supram as necessidades de informação de quem as demanda.

Pudemos notar a evolução do conceito "Sociedade da Informação", conceito cunhado por Nora & Minc em 1980 (CASTELLS, 1999), para o de Sociedade do Conhecimento. Aspectos cognitivos relacionados ao processamento das informações passam a ser considerados mais relevantes, e Drucker (1994) eleva

o conhecimento a recurso econômico mais valioso. Em parte, isto é atribuído ao maior fluxo de informação e seu compartilhamento na rede, pois a quantidade de informação passa a não ter valor se não for apreendida pelo homem, e despreende-se desta visão que sem indexação e a conseqüente possibilidade de recuperação por parte do usuário, também passa a não ter valor.

A mudança de paradigmas relacionados à explosão informacional, proporcionada pelo advento da *Web* e as novas Tecnologias da Informação e Comunicação (TICs), tecnologias integradas que mediam processos informacionais, proporcionaram ambiente favorável para o surgimento de novas formas de representação da informação, conforme observou Gil Leiva (1999). Percebe-se que, neste contexto, especialistas ou profissionais não são mais capazes de gerenciar todo o volume de informação produzido pela sociedade, posto que rompeu-se o domínio da geração e veiculação de informação, e este processo saiu da mão de um número limitado de pessoas.

Se em outros tempos era necessária a existência de jornais, emissoras de TV e rádio, editores e revisores para que a informação fosse gerada e propagada, de forma gradual todo e qualquer usuário passou a ser capaz de desempenhar tarefas similares, mesmo que de forma não tão apurada como a praticada por veículos de informação. Desta forma, não é mais obrigatória a existência de uma estrutura organizacional complexa para que a informação seja gerada e manipulada, e aos poucos percebe-se que este poder foi sendo transmitido a cada cidadão. Mais adiante, conectados à *Web* e dotados de conhecimento mínimo para operar dispositivos eletrônicos, são capazes de desempenhar os processos de compartilhamento e organização da informação de forma muito ágil e em grande quantidade.

Nos anos 50 a expressão *do it yourself* surgiu para designar um conjunto de mudanças que colocaram o sujeito como centro da força de criação, em um contexto em que se tornou cada vez mais caro contratar pessoas ou empresas especializadas para desempenhar serviços. Observa-se uma correlação entre este princípio e o que vivemos atualmente na *Web*: da mesma forma que empresas passaram a vender produtos em quantidades menores e com uso simplificado para que toda pessoa fosse capaz de desempenhar quaisquer tipos de tarefas, as TICs foram sendo adaptadas e sua interface simplificada de modo que qualquer pessoa

pudesse operá-las, e com isso, se comunicar e divulgar suas criações e produções. A partir dos anos 90, presenciou-se um contexto de facilidade para divulgação de textos, pensamentos, vídeos, fotos, áudio e qualquer tipo de mídia gerada, e esta informação precisa ser representada.

Concomitante a este cenário, observa-se o surgimento de uma cultura de colaboração e sociabilidade entre os usuários da *Web*, o que foi chamado por O'Reilly de arquitetura da participação. Nota-se o agrupamento de pessoas em torno de círculos em comum na *Web*, que além de compartilhar experiências e socializar, contribuem ativamente para a criação e aperfeiçoamento dos serviços informacionais em todo mundo. Se todos estão conectados, rompe-se a barreira física antes existente, que limitava os círculos de criação e organização de informação. Neste momento, nota-se que a expressão *do it yourself* já não representa mais o contexto de criações conjuntas e cooperação estabelecido na *Web*, e o termo *do it together* passa a ser mais adequado para designar este conjunto de mudanças, que se estabelece na *Web 2.0*, a face colaborativa da rede. Desta forma, o século XX assistiu o empoderamento da atividade de criação e disseminação de informação digital pelo usuário, e posteriormente a formação de grupos agindo em conjunto para desempenhar estas tarefas de forma mais ágil.

Um exemplo que corrobora esta tese é o aumento da criação e a popularização dos *softwares* abertos ou livres, programas com o código editável e personalizável, de modo que os próprios usuários criam funcionalidades e adaptam estes *softwares* a diversos contextos. O Sistema Operacional Linux foi desenvolvido pelo finlandês Linus Torvalds em parceria com grupos de programadores voluntários que participavam do *Usenet*, uma espécie de fórum agrupado por assuntos existente nos primórdios da internet, e teve sua primeira versão lançada em 1991. Apresenta código fonte aberto, o que significa que sua licença permite a qualquer usuário visualizar, editar, reprogramar e divulgar uma versão nova do software. Inúmeras versões adaptadas e adaptáveis a ambientes bancários, de jogos, comércio ou mesmo uso diário, com funcionalidades e interface mais simples foram lançadas desde então, com a colaboração da comunidade *online* e de grandes empresas como IBM, Google e HP. Além de editar e aprimorar o programa, que desta maneira está em constante aperfeiçoamento pelas comunidades virtuais, as novas versões

são compartilhadas em fóruns, disponíveis para que qualquer pessoa conectada faça o *download*, utilize e compartilhe experiências acerca do Sistema Operacional.

Inseridos neste contexto, mesmo os editores de softwares fechados, que não apresentam a possibilidade de ter seu código editável, abrem espaço para que funcionalidades e programas adicionais sejam inseridos em suas plataformas, como no caso da Apple e da Microsoft, que permitem a instalação de *Apps* ou Extensões, que embora não modifiquem os *softwares* em sua essência, adicionam funcionalidades.

Outros mecanismos, como a revisão entre pares, em que a revisão de artigos científicos, jornalísticos, ou mesmo textos de blogs passam a contar com a possibilidade de serem realizados por colegas, e não necessariamente por profissionais editores ou revisores, e o agrupamento e atuação de grupos em prol de objetivos comuns, como as comunidades criadas com a finalidade de desenvolver *softwares* abertos para determinados fins, demonstram que se desenvolveu um contexto de cooperação virtual com o propósito de gerar e gerenciar informação *online* de forma ágil e não necessariamente presa a organizações formais.

As mídias sociais, sistemas projetados para possibilitar a interação social a partir do compartilhamento e da criação colaborativa de informação nos mais diversos formatos e as *wikis* são plataformas emblemáticas deste contexto de cooperação. Elas constituem-se de base de dados dotadas de funcionalidades que permitem o compartilhamento e edição de conteúdo, mas este conteúdo, de fato, é gerado e organizado pelos usuários. A Wikipedia, uma enciclopédia *online* e gratuita para consulta, foi desenvolvida por Jimmy Wales e Larry Sanger e lançada em 2001 sobre plataforma *wiki*, trata-se de um banco de dados global sobre temas em geral, em que seus artigos são escritos e revisados pelos próprios usuários. Atualmente a base de dados apresenta artigos em 287 idiomas, e artigos divulgados na Reuters¹¹, Alexa¹² e na Time¹³ a apontam como a obra de referência mais popular na *Web*, tendo aproximadamente 365 milhões de leitores.

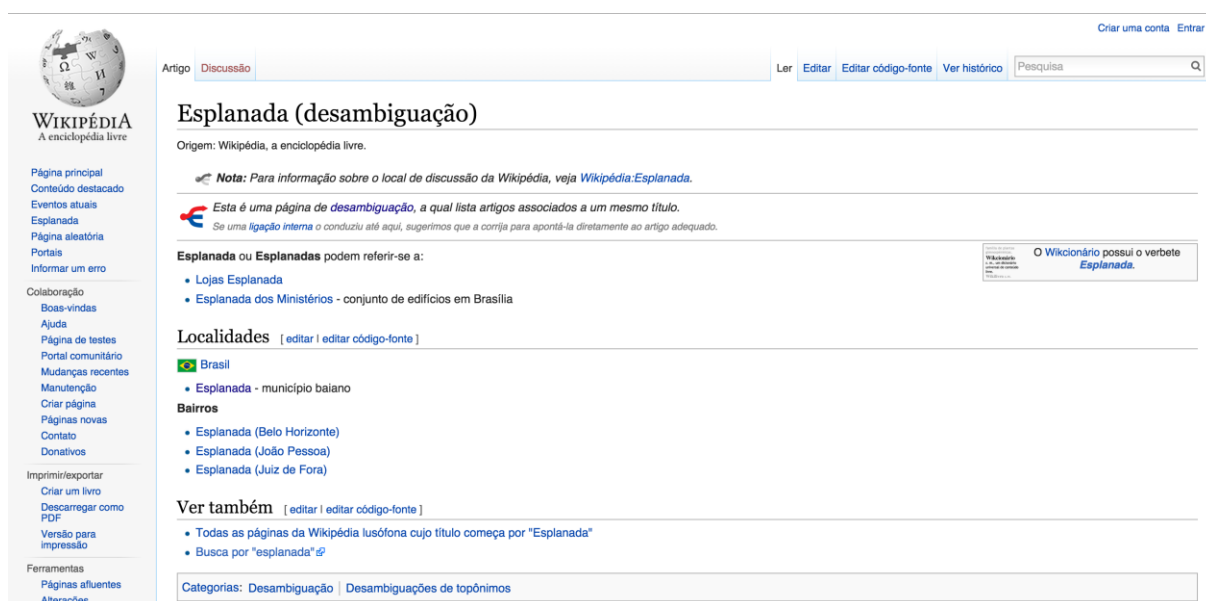
¹¹<http://www.reuters.com/article/2007/07/08/us-media-wikipedia-idUSN0819429120070708>

¹²<http://www.alexa.com/siteinfo/wikipedia.org?range=5y&size=large&y=t>

¹³<http://content.time.com/time/business/article/0,8599,1595184,00.html>

Os artigos da Wikipedia possuem um indicador de credibilidade, que levam em consideração a necessidade de revisão, checagem de fontes e referências e avaliação dos usuários, e a própria plataforma oferece orientações para que os usuários atuem como colaboradores. Estes artigos são indexados pelas palavras constantes no título dos textos, e caso haja alguma palavra ambígua, ou seja, com mais de um significado, é possível escolher entre os resultados. Esta ferramenta existente para refino dos resultados é a desambiguação (ver Figura 2), que permite a escolha de resultados distintos para um mesmo título ou entrada de busca. Ao realizar a busca através da palavra “Esplanada”, o usuário é direcionado ao resultado julgado mais relevante ou o que mais corresponde a busca, que se trata de um município brasileiro situado na Bahia. Através deste resultado, ao acessar a desambiguação, o usuário tem acesso aos demais resultados homônimos, como “Esplanada”, referente a bairros de Belo Horizonte, João Pessoa e Juiz de Fora e outros como “Esplanada dos Ministérios” e “Lojas Esplanada”.

Figura 2 - Desambiguação na Wikipedia



Fonte: Elaborada pelo autor.wikipedia.org

Apesar de ser alvo de críticas quanto a confiabilidade e precisão das informações contidas na base, pesquisas realizadas em 2005 pela revista *Nature*¹⁴, uma das mais antigas revistas científicas do mundo, e outra pelo jornal britânico *The*

¹⁴<http://www.nature.com/nature/journal/v438/n7070/full/438900a.html>

*Guardian*¹⁵, demonstraram que a própria comunidade é eficiente em inibir a incidência do que é conhecido como vandalismo *online*: a inclusão de dados e informações inverossímeis. Utilizando-se de métodos estatísticos, a revista *Nature* selecionou 42 artigos científicos e os submeteu à análise de analistas acadêmicos, que os avaliaram segundo o critério de acurácia e precisão. O resultado apontou que, dentro desta amostra, a Wikipédia continha o mesmo número de erros graves que os constantes na *Enciclopedia Britannica*, uma das obras de referência mais respeitadas no mundo, e 4 erros considerados leves, contra 3 da *Britannica*. No estudo realizado pelo *The Guardian*, sete verbetes foram analisados por especialistas, que conferiram, em sua maioria, notas de 5 a 8 em uma escala de 0 a 10 aos artigos da Wikipédia, destacando opiniões favoráveis em relação à confiabilidade e utilidade das informações, bem como a possibilidade de remeter, via links, a outras informações de interesse da busca. Estes estudos apontam que o usuário é capaz de gerar e gerenciar conteúdo na *Web*, por mais frágil que os ambientes editáveis e cooperativos pareçam ser, e representam bem o contexto de colaboração em que o ambiente *Web* está inserido atualmente.

Inseridas neste mesmo contexto, diversas mídias sociais, além de permitir a interação, formação de comunidades de interesse e o compartilhamento de textos, áudio e fotografias, oferecem funcionalidades para a representação dos conteúdos gerados pelos usuários. Em geral, a representação da informação é realizada pelos próprios membros, com linguagem natural e familiar aos grupos em que estão inseridos no mesmo momento em que o conteúdo é disponibilizado, constituindo um estágio que a primeira vista parece um retrocesso em relação aos avanços da área da indexação, por não estar baseado em técnicas ou vocabulários controlados, mas que confere agilidade a adaptabilidade ao processo.

Este processo, chamado folksonomia, conceituado em sessão anterior neste trabalho, no contexto das mídias sociais é em geral derivado das concepções de palavra-chave, que no contexto *Web* se popularizou pela inserção da entrada, ou palavra a descrever a informação com o prefixo #, popularmente chamada de *hashtag*. Esta forma de representação, também chamada de etiquetagem, por atribuir etiquetas aos conteúdos, possibilita sua recuperação mediante acesso à essas *tags*. Desta forma, quando um usuário compartilha uma foto capturada em

¹⁵<http://technology.guardian.co.uk/opinion/story/0,16541,1599325,00.html>

câmera analógica e a etiqueta com #film, automaticamente esta imagem passará a constar em um banco de dados compreendido por todas as figuras etiquetadas com esta mesma *hashtag*, possibilitando seu acesso por outros usuários que acessem esta etiqueta. Algumas mídias sociais permitem que outros usuários da comunidade adicionem outras etiquetas, ou repostem o conteúdo atribuindo suas próprias palavras-chave, tornando o processo de representação na *Web* dinâmico e cooperativo, reflexo de um ambiente adaptável por natureza.

Nas mídias sociais Twitter, Instagram e Facebook, a Folksonomia se dá pelo uso das *hashtags*. Durante a publicação, via edição ou comentários, é possível incluir palavras que representem o conteúdo seguidas do prefixo # para realizar a indexação (Figuras 3 e 4). Neste exemplo específico em que uma marca divulga seus produtos, várias palavras foram utilizadas para designar o vídeo compartilhado, como #handmade, por se tratar de algo feito à mão, #queromelancia, para facilitar a recuperação de informações relacionadas ao nome da marca, dentre outras.

Figura 3 - Postagem de vídeo no Instagram



Fonte: Elaborada pelo autor. <http://instagram.com/queromelancia>

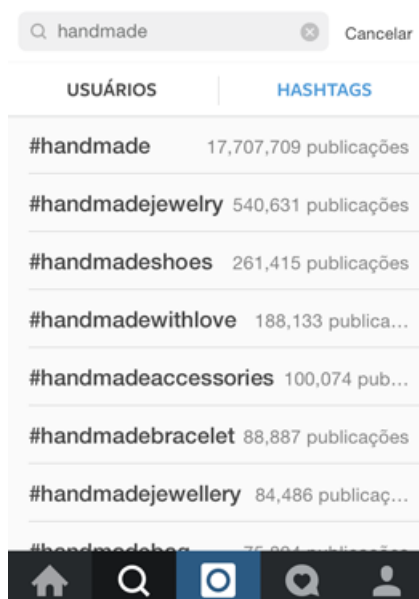
Figura 4 - Etiquetação do vídeo com hashtags



Fonte: Elaborada pelo autor. <http://instagram.com/queromelancia>

É importante mencionar que, com exceção do Twitter, em que as postagens não podem ser editadas, no Facebook e Instagram as *hashtags* podem ser incluídas após a publicação do material por qualquer usuário da rede, e o processo de indexação se dará da mesma forma. Estas mídias sociais também possibilitam a recuperação da informação, via pesquisa de *hashtags* (Figura 5) ou por um clique na própria etiqueta, que remeterá ao conjunto de publicações que foram indexadas com a mesma palavra-chave.

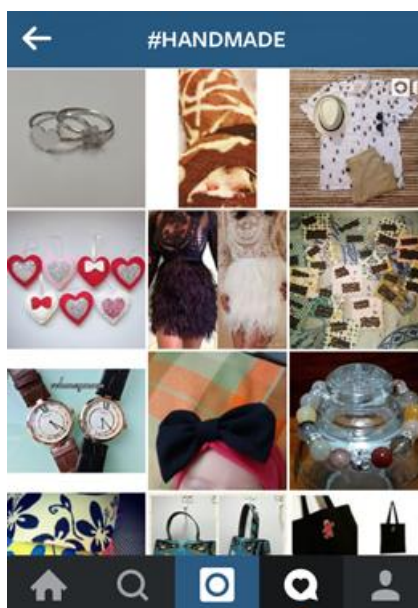
Figura 5 - Recuperação da informação via pesquisa de hashtags



Fonte: Elaborada pelo autor. <http://instagram.com/queromelancia>

Desta forma, quando um usuário pesquisar nas mídias sociais pela *hashtag* *#handmade* ou acessá-la diretamente via clique, um conjunto de publicações etiquetadas com esta *hashtag* será disposto, em uma página que reúne as postagens indexadas com a palavra-chave *handmade*, em ordem de publicação, inclusive o vídeo exemplificado nas figuras (Figura 6).

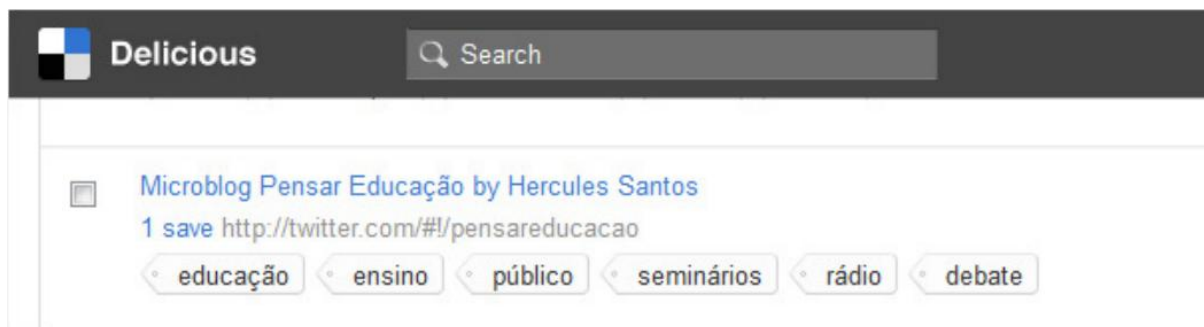
Figura 6 - Acesso à hashtag *#handmade*



Fonte: Elaborada pelo autor. <http://instagram.com/queromelancia>

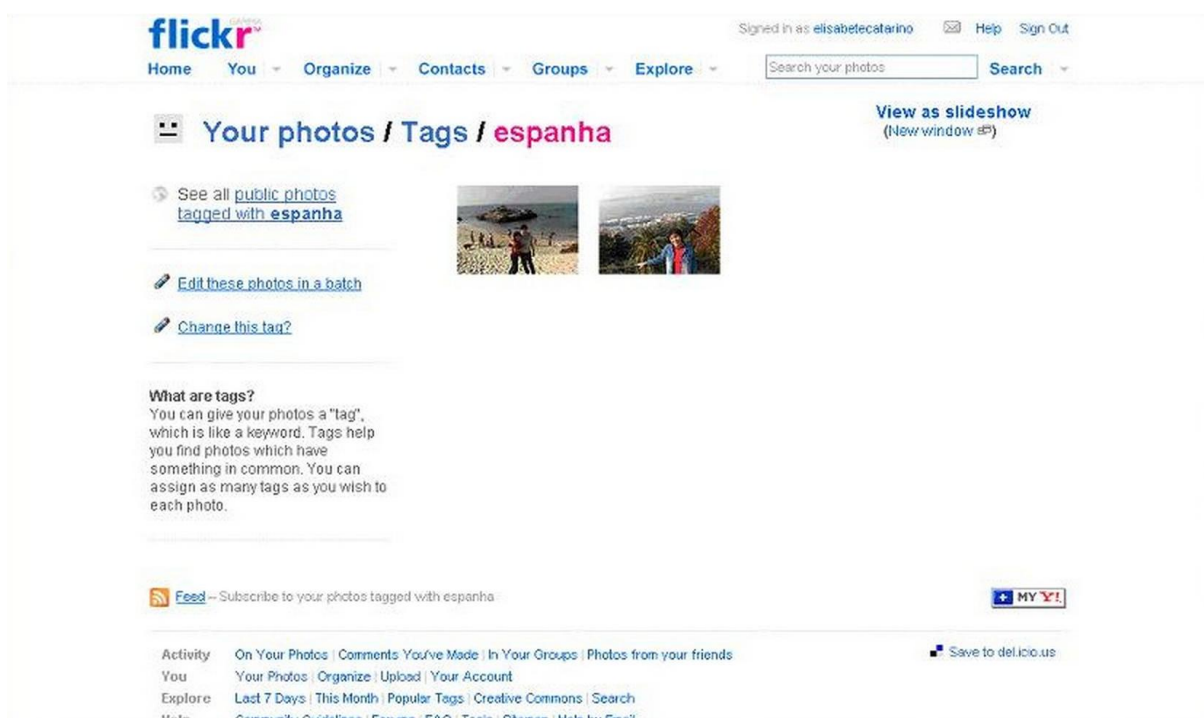
Em outras mídias sociais, a exemplo do Flickr, base de armazenamento e compartilhamento de fotos, e do Delicious, que permite aos usuários armazenarem marcadores ou favoritos e terem acesso aos favoritos de toda a comunidade, a folksonomia não se dá pelo uso de *hashtags*, mas de etiquetas (SANTOS, 2013). Neste caso, elas funcionam de forma semelhante ao das mídias sociais anteriormente exemplificadas, de forma que o usuário atribui etiquetas ao material publicado e é possível recuperar informação buscando por elas (Figuras 7 e 8).

Figura 7 - Etiketagem no Delicious



Fonte: Santos, 2013, p. 96

Figura 8 - Etiketagem no Flickr



Fonte: Catarino; Baptista, 2007, p. 18

O Twitter é uma mídia social que possibilita o compartilhamento de pequenos textos e links, limitados a 140 caracteres, e popularizou o uso das *hashtags*. Segundo o próprio site, seus usuários criaram essa forma de representação de conteúdos de maneira orgânica, como forma de categorizar as mensagens. Ao possibilitar o agrupamento dos *tweets* contendo a mesma etiqueta, a mídia social passou a divulgar as *hashtags* mais utilizadas pelos usuários em determinados momentos, favorecendo o compartilhamento das mensagens e a difusão dos temas por elas abarcados. O fato de milhões de usuários estarem conectados ao mesmo tempo e com acesso instantâneo ao conteúdo postado

possibilitou a difusão de mensagens, etiquetas e conteúdos em evidência no momento, conhecidos como *trendingtopics*, ou tópicos em tendência. Trata-se de mais um mecanismo de medição de popularidade de termos e assuntos, e representação de conteúdos gerenciado exclusivamente pelos usuários em tempo real.

Seguindo a tendência de outras mídias sociais como Twitter, Instagram, Tumblr e Pinterest de dar autonomia aos usuários para indexação de suas postagens, em junho de 2013 o Facebook liberou o uso das *hashtags*, que passaram a ser clicáveis e as postagens etiquetadas, parte de um contexto maior. Além disso, foi permitido recuperar assuntos e verificar quais os temas mais comentados. Esta ação reconhece que os usuários desejam participar da criação e gerenciamento dos recursos da *Web*, e foi implementado pouco tempo depois do *Wall Street Journal* ter publicado um artigo que afirmava que a incorporação das *hashtags* poderiam manter os usuários mais tempo na rede.

Percebe-se que a sociedade atual e a *Web* permeiam um contexto de cooperação e criações conjuntas, citado neste estudo como *do it together*, bem como o de uma cultura de personalização, em que o sujeito, mesmo que de formas direcionadas, valoriza a possibilidade de opinar ou fazer parte das criações. Assim como prestadores de serviços e estabelecimentos passaram a oferecer um Menu de opções passíveis de incrementação por parte dos clientes, as aplicações virtuais se adaptaram a esta realidade e passaram a realimentá-la, influenciando a forma de interação entre pessoas e serviços, que encontraram na *Web* possibilidade de divulgação e rompimento de barreiras.

A folksonomia é ponto fundamental nesta relação, ao possibilitar que conteúdos individuais façam parte de um contexto maior e acessível a mais pessoas, potencializando a recuperação dos conteúdos gerados por qualquer usuário.

6. CONSIDERAÇÕES FINAIS

De acordo com o estudo realizado pode-se afirmar que a indexação evoluiu para sua forma *online* na *Web* não só pelo avanço tecnológico, difusão do acesso a internet e a popularização do uso dos dispositivos eletrônicos, mas por um conjunto de mudanças sociológicas que elevaram o indivíduo ao papel de ente criador e organizador, e principalmente pelo surgimento de um contexto colaborativo. Esta quebra de paradigmas, em que a indexação deixou de ser realizada por profissionais ou especialistas e passou a ser realizada em tempo real em linguagem natural, foi exemplificado pelo movimento *do it yourself*, e posteriormente por sua evolução para *do it together*, um contexto de cooperação e criações conjuntas, em que se inserem os softwares abertos, *wikis* e mídias sociais.

Foram apresentados estudos que apontam que o usuário é eficaz em controlar acessos e alterações mal-intencionadas na rede, e que a confiabilidade de trabalhos realizados pelos próprios usuários é comparável aos realizados por instituições conceituadas. Embora tenham sido apresentados estes estudos, não foi objeto deste trabalho fazer juízo de valor acerca das novas formas de compartilhamento e indexação da informação, como mídias sociais e *wikis*, mas demonstrar que estas fazem parte de tendências e de um contexto complexo, cabendo estudar e apresentar maneiras para aprimorar o tratamento da informação nestes ambientes.

As principais dificuldades encontradas para realização deste trabalho foram relacionadas a contemporaneidade dos temas relacionados à *Web 2.0* e folksonomia, a conseqüente escassez de estudos avaliando a indexação na *Web* em comparação aos métodos atuais, e a quase ausência de literatura na área da sociologia acerca do contexto colaborativo no qual a sociedade está inserida. Desta forma, recomenda-se que outros estudos sejam realizados em conjunto com profissionais de outras áreas como sociologia, psicologia e administração, por se tratar de uma área inter e multi-disciplinar, que se beneficiará com contribuições de outras disciplinas.

No âmbito acadêmico, o presente trabalho contribui para a área da Ciência da Informação por acrescentar referência em português para um tema de literatura escassa no Brasil, principalmente quando se trata de assuntos

relacionados a *Web 2.0*, por oferecer um panorama histórico acerca da evolução da indexação ao longo do tempo, e por inserir a temática da folksonomia em um contexto de mudanças sociológicas e tendências aceleradas pelo rápido desenvolvimento e popularização das novas Tecnologias de Informação e Comunicação.

Finalmente, seja por necessidade de pertencimento ou popularidade, a indexação se adaptou a este contexto, e na rede, assume faces paralelas e atualizadas de princípios fundamentais da Biblioteconomia cunhados por Ranganathan, uma vez que a folksonomia possibilita a informação a ser acessada pelos seus usuários de forma mais rápida, poupando seu tempo, em um meio vivo, em crescimento e mutação - a *Web*. Se os tempos atuais configuram-se como pós-modernidade ou modernidade líquida, como define Zygmunt Bauman (2003), tempos em que a informação é gerada e descartada em velocidades sem precedentes, e sabe-se que informação é matéria-prima para o conhecimento, sinônimo de poder, representar este conteúdo de forma rápida torna-se necessidade fundamental para que a informação se torne acessível antes que sua vida útil se esgote.

7. CONCLUSÃO

Por meio da apresentação de revisão de literatura acerca da conceituação de indexação e levantamento de marcos históricos que evidenciaram a evolução da prática da sua prática, foi possível traçar um panorama histórico que demonstrou a adaptação e evolução da maneira de indexar ao longo do tempo, e suas características. Foram apresentadas também definições sobre *Web 2.0*, ambiente *online* no qual a sociedade atual está inserida e sobre folksonomia, a forma de indexar em linguagem natural realizada pelos próprios usuários na rede.

Ao analisar a evolução da indexação, foi possível perceber que a prática sempre esteve ligada a necessidades e particularidades do seu tempo, e acompanhou as mudanças da sociedade na forma de lidar com a informação. Desta forma, alcançou-se o objetivo de compreender de que forma a indexação se adaptou ao contexto *Web*.

Através da revisão bibliográfica, foi possível entender o que propiciou e de que forma a indexação se deu em ambiente *Web*. Neste contexto mais recente, com o advento da internet, a popularização das novas Tecnologias de Informação e Comunicação e o surgimento de um contexto colaborativo, um volume maior ainda de informação passa a ser gerado e compartilhado *online*, e esta massa informacional precisa ser indexada em velocidade recorde. Imersa em um contexto colaborativo, a tarefa passou a ser desempenhada pelo próprio usuário em linguagem natural, o que tornou possível o gerenciamento e indexação dos conteúdos fossem realizados na mesma velocidade com que fossem gerados.

Por se tratar de um contexto atual, é importante que o profissional da informação conheça e estude as novas formas de indexação colaborativas e a *Web*, e dotado de conhecimentos mais profundos sobre a prática da representação da informação, possa oferecer subsídios para que as aplicações da *Web 2.0* auxiliem o usuário na escolha das etiquetas, propiciando melhoria no processo de recuperação da informação.

REFERÊNCIAS

APOLINÁRIO, F. **Metodologia da Ciência: filosofia e prática da pesquisa**. São Paulo: Thomson, 2006.

BARROS, A.J.S.; LEHFELD, N.A.S. **Fundamentos de metodologia científica**. 2.ed. São Paulo: Makron, 2000. 122p.

BAUMAN, Z. **Modernidade Líquida**. Tradução: Plínio Dentzien. Rio de Janeiro: Zahar, 2003. 258p.

BRANDT, M. B. **Etiquetagem e folksonomia: uma análise sob a óptica dos processos de organização e recuperação da informação na web**. 142 f. Dissertação (Mestrado em Ciências da Informação) – Universidade de Brasília, Faculdade de Economia, Administração, Contabilidade e Ciência da Informação, 2009.

CAMPOS, M. L. A. Perspectivas para o estudo da área de representação da informação. **Revista Ciência da Informação**, v. 25, n. 2, 1995.

CASTELLS, M. **A sociedade em rede**. 4. ed. São Paulo: Paz e Terra, 1999. v.1.

CATARINO, M. E.; BAPTISTA, A. A. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. **DataGramZero**, v. 8, n. 3, jun/2007. Disponível em: <http://www.dgz.org.br/jun07/Art_04.htm>. Acesso em: 10 ago. 2014.

CHAUMIER, J. **Analyse et langages documentaires**: Le traitement linguistique de l'information documentaire. Paris: Enterprise Moderne d'Édition, 1982.

CLEVELAND, D. B.; CLEVELAND, A. D. **Introduction to indexing and abstracting**. Englewood: Libraries Unlimited, 1990.

COLLINSON, R.L. **Índices e indexação: guia para indexação de livros, e coleções de livros, periódicos, e coleções de livros, periódicos, partituras musicais, com uma seção de referência e sugestões para leitura adicional**. Trad. Antônio Agenor Briquet de Lemos. São Paulo: Polígono, 1971.

DIAS, E. W. Análise de Assunto: percepção do usuário quanto ao conteúdo de documentos. **Perspectivas Ciência da Informação**, v.9, n. 2, p. 148-157, jul/dez/2014.

DRUCKER, P. F. **Sociedade pós-capitalista**. São Paulo: Pioneira, 1994.

FEITOSA, A. **Organização da informação na Web**: das tags à Web semântica. Brasília, DF: Thesaurus, 2006. 131 p. (Estudos avançados em Ciência da Informação, v. 2).

FUJITA, M. S. L. A avaliação da eficácia de recuperação do sistema de indexação PRECIS. **Ciência da Informação**, Brasília, DF, v. 18, n. 2, p. 120-134, jul./dez. 1989. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1361/987>> . Acesso em: 10 set. 2014.

GARRIDO ARILLA, M. R. **Teoria e historia de lacatalogación de documentos**. Madrid : Editorial Síntesis, 1999. ISBN 84-7738-344-8. p. 63.

GIL LEIVA, I. **La automatización de laindexación de documentos**. Gijón (Astúrias): EcicionesTrea, 1999. 220 p.

GUIMARÃES, J. A. C. A análise documentária no âmbito do tratamento da informação: elementos históricos conceituais. In: RODRIGUES, G. M.; LOPES, I. L. (Org.). **Organização e representação do conhecimento na perspectiva da Ciência da Informação**. Brasília: Thesaurus, 2003, v. 2, p.100-117. Disponível em: <<http://pt.scribd.com/doc/60189087/Analise-documentaria-no-ambito-do-tratamento-da-informacao>> Acesso em: 10 set. 2014.

GUIMARÃES, J. A. C. **Abordagens teóricas de tratamento temático da informação**: catalogação de assunto, indexação e análise documental. Ibersid.,2009 105-117. ISSN 1888-0967. Disponível em:<<http://ibersid.eu/ojs/index.php/ibersid/article/viewFile/3730/3491>> Acesso em: 10 set. 2014.

GUY, M.; TONKIN, E. Folksonomies: tidyinguptags? **D-Lib Magazine**, v. 2, n. 1, 2006. Disponível em: <<http://www.dlib.org/dlib/january06/guy/01guy.html>>. Acesso em: 5 out 2014.

HAMMOND, Tony et al. Social Bookmarking Tools (I): a general review. **D-Lib Magazine**, v.11, n.4, apr. 2005. Disponível em: <<http://www.dlib.org/dlib/april05/hammond/04hammond.html>>. Acesso em: 08out. 2014.

KOBASHI, N. Y. **A elaboração de informações documentárias**: em busca de uma metodologia. 1994. 267 f. Tese (Doutorado em Comunicações e Artes) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1994.

KURAMOTO, H. Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais. **Revista Ciência da Informação**, v. 25, n. 2, 1995. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/435/393>>. Acesso em: 16 ago. 2014.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília, DF: Briquet de Lemos, 2004.

LANDOW, G. P. **Hipertexto: laconvergencia de lateoría crítica contemporánea y latecnología**. Tradução de Patrick Ducher. Barcelona: Paidós, 1992.

LEÃO, L. **O labirinto da hipermidia: arquitetura e navegação no ciberespaço**. 3. ed. São Paulo: FAPESP: Iluminuras, 2005.

LÉVY, P. **O que é o virtual?** Tradução de Paulo Neves. São Paulo: Editora 34, 2003.

MATHES, A. **Folksonomies**: cooperative Classification and Communication through shared metadata. Computer Mediated Communication – LIS590CMC, Urbana : University of Illinois, 2004. Disponível em: <<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>>. Acesso em: 25 out. 2014

MILLER, E. **An Introduction to the Resource Description Framework**. **D-Lib Magazine**, v.4, n.5, may 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 29 out. 2014.

MORRISON, P. J. **Tagging and searching: search retrieval effectiveness of folksonomies on the World Wide Web**. Information Processing and Management, v. 44, p. 1562-1579, 2008.

NEGROPONTE, N.; TELLAROLI, S. **A vida digital**. 2. ed. São Paulo: Companhia das Letras, 1995.

O'REILLY, T. Not 2.0? 2005. Disponível em: <<http://radar.oreilly.com/2005/08/not-20.html>> Acesso em: 20 out. 2014.

O'REILLY, T. What is Web 2.0? 2005. Disponível em: <<http://oreilly.com/web2/archive/what-is-web-20.html>>. Acesso em: 20 out. 2014.

PETERSON, E. **Beneath the Metadata: some philosophical problems with Folksonomy**. D-Lib Magazine, v.12, n.11, nov. 2006.

Disponível em: <<http://www.dlib.org/dlib/november06/peterson/11peterson.html>>. Acesso em: 01 out. de 2014.

PINTO, V. B. Indexação documental: uma forma de representação do conhecimento registrado. **Rev. De Letras**, v. 1/2, n. 22, 2000. Disponível em: <<http://www.revistadeletras.ufc.br/rl22Art09.pdf>>. Acesso em: 15 ago. 2014.

PONJUÁN DANTE, G. **Gestión de informacionenlasorganizaciones: principios, conceptos y aplicaciones**. Santiago, Chile: Centro de CapacitaciónenInformación, Universidad de Chile,1998.

RANGANATHAN, S. R. **As cinco leis da Biblioteconomia**. Brasília: Briquet de Lemos Livros, 2009. 336 p.

RIBEIRO, F. – **O Acesso à informação nos arquivos**. Lisboa : FundaçãoCalousteGulbenkian; Fundação para a Ciência e a Tecnologia, Ministério da Ciência e do Ensino Superior, 2003. ISBN 972-31-1017-2. vol. 2, p. 633-694. (Dissertação de doutoramento. Disponível em: <<http://ler.letras.up.pt/revistas/index.htm>> Acesso em: 25 set. 2014.

RIBEIRO, F. Organizar e representar informação : apenas um meio para viabilizar o acesso?.**Revista da Faculdade de Letras: Ciências e Técnicas do Património**, v. 4 , p. 83-100, 2005. Disponível em:<<http://hdl.handle.net/10216/9019>>Acesso em: 15 set. 2014.

RUSSELL, T. **Contextual authoritytagging: cognitiveauthoritythroughfolksonomy**. Disponível em: <<http://www.terrellrussell.com/projects/contextualauthoritytagging/conanthtag200505.pdf>>. Acesso em: 2 out. de 2014.

ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. In: Machado, U. D. (Org.). **Estudos Avançados em Ciência da Informação**. Brasília, DF.:Associação dos Bibliotecários do Distrito Federal, 1982. v. 1, p. 235-274.

SANTOS, H. P. Etiqueta e folksonomia: o usuário e sua motivação para organizar e compartilhar informação na web 2.0.**Perspectivas em Ciência da Informação**, v. 18, n. 2, p. 91-104, abr./jun. 2013. Disponível em: <http://www.scielo.br/pdf/pci/v18n2/07.pdf> Acesso em: 10 out. 2014.

SILVA, M. dos R. da; FUJITA, M. S. L. A prática da indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, Campinas, 16(2): 133-161, mai/ago., 2004. Disponível

em:<<http://www.brapci.ufpr.br/documento.php?dd0=0000000388&dd1=0fc70>>Acesso em 17 set. 2014.

TOSSERI, O. **Gutenberg não inventou a imprensa** [Internet]. São Paulo, Segmento, s.d. Disponível em: http://www2.uol.com.br/historiaviva/artigos/gutenberg_nao_inventou_a_imprensa.html. Acesso em: 15 set. 2014.

TRANT, J. **Exploring the potential for social tagging and folksonomy in art museums: proff of concept**. New Review of Hypermedia and Multimedia, v.12, n.1, p.63-81, jun. 2006a.

VIEIRA, S. B. Indexação automática e manual: revisão de literatura. **Ci. Inf.**, Brasília, 17 (1): 43-57, jan./jun. 1988. Disponível em: <http://revista.ibict.br/index.php/ciinf/article/download/1391/1017> Acesso em: 15 ago. 2014.

WAL, T. V. **Folksonomy definition and wikipedia**. 2005. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>. Acesso em: 28 out. 2014.

WAL, T. V. **From tag to the future**. 2005. Disponível em: <<http://www.vanderwal.net/random/category.php?cat=153>>. Acesso em: 17 set. 2014.

WAL, T. V. **Understanding the personal info cloud: using the model of attraction**. Presented at University Maryland., 2004. Disponível em:<<http://www.vanderwal.net/essays/moa/040608/040608.pdf>> Acesso em 17 set. 2014.