

Leitura documentária para indexação: abordagens interdisciplinares  
6 – A leitura documentária e as perspectivas de aplicação em contexto da  
Web

Gercina Ângela de Lima

SciELO Books / SciELO Livros / SciELO Libros

LIMA, G. A. A leitura documentária e as perspectivas de aplicação em contexto da Web. FUJITA, M. S. L., ALVES, R. C. V., and ALMEIDA, C. C., eds. *Modelos de leitura documentária para indexação: abordagens teóricas interdisciplinares e aplicações em diferentes tipos de documentos* [online]. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2020, pp. 139-171. ISBN: 978-65-8654-607-1. Available from: <http://books.scielo.org/id/96v3r>. <https://doi.org/10.36311/2020.978-65-86546-07-1.p139-172>.



All the contents of this work, except where otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença [Creative Commons Atribuição 4.0](https://creativecommons.org/licenses/by/4.0/).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia [Creative Commons Reconocimiento 4.0](https://creativecommons.org/licenses/by/4.0/).

# 6

## A LEITURA DOCUMENTÁRIA E AS PERSPECTIVAS DE APLICAÇÃO EM CONTEXTO DA *WEB*

*Gercina Ângela de LIMA*

**RESUMO:** A mudança do texto impresso para o texto eletrônico criou uma grande transformação na maneira como se armazena e acessa a informação. O hipertexto, fruto da tecnologia informacional das últimas décadas, é a forma texto que estimula e torna possível o desenvolvimento do universo digital. No ambiente Web, o processo de indexação mostra-se útil e eficiente para criar a estrutura conceitual com *links* para as partes relevantes do texto. Neste capítulo, apresenta-se um estudo sobre a contribuição da leitura documentária na conversão de textos em hipertextos no contexto da Web, do ponto de vista do autor. Os estágios da leitura documentária, no processo de *análise de assunto*, constituem uma abordagem importante a ser considerada na estruturação de hipertextos. Conclui-se que os estágios previstos na leitura documentária realizados no texto linear podem auxiliar o autor na estruturação de hipertextos, desde que sejam adaptados para a representação conteúdo do texto não linear, garantindo a interatividade entre módulos de informação, que garante a compreensão global do texto.

**PALAVRAS-CHAVE:** Leitura documentária. Hipertexto. Web.

**ABSTRACT:** The change from printed text to electronic text has created a major transformation in the way information is stored and accessed. Hypertext, fruit of the information technology of the last decades, is the text form that stimulates and makes possible the development of the digital universe. In the Web environment, the indexing

process is useful and efficient to create the conceptual framework with links to the relevant parts of the text. In this chapter, we present a study about the contribution of documentary reading in the conversion of texts into hypertexts in the context of the Web from the author's point of view. The stages of documentary reading, in the process of subject analysis, constitute an important approach to be considered in the structuring of hypertexts. It is concluded that the stages provided in the documentary reading in the linear text can help the author in the structuring of hypertexts, as long as they are adapted to the nonlinear text content representation, guaranteeing the interactivity between information modules, which guarantees the global comprehension of the text.

**KEYWORD:** Documentary reading. Hypertext. Web.

## 1 INTRODUÇÃO

A necessidade de representação da informação para sua posterior recuperação começou desde quando o papel foi introduzido no mundo ocidental, em meados do século XII, e difundiu-se na Europa, entre os séculos XIII e XV, com a necessidade do aumento da quantidade de novos originais e cópias de livros. À medida que crescia o aumento de exemplares disponíveis, aumentava também o estímulo ao registro do conhecimento, o que influenciou muito o clima intelectual na Idade Média. Assim, a escrita cria a figura do leitor, para o qual a realidade passa por um filtro muito mais refinado. Ampliou-se o número de editores e possíveis leitores. A partir daí, a leitura e a interpretação do texto adquiriram caráter mais individualizado, com particularidades e exigências diversas. As obras começaram a incluir representações gráficas mais precisas e específicas, como tabelas, gráficos e desenhos, incorporando inventos e artifícios anteriores à tipografia. Formas mais sofisticadas de leitura não linear começaram a surgir, como as notas de rodapé, sumários, por exemplo, que direcionavam o leitor a outras partes do mesmo livro ou a outras obras relacionadas, como índices e bibliografias.

Desde então, a comunicação escrita e o modo de transmissão dos textos sofreram profundas mudanças. A escrita em suportes artesanais foi a tecnologia de comunicação mais avançada, desde o quarto milênio a.C. até a invenção da imprensa e da tipografia na Mogúncia (c.1450), quando Johann Gutenberg compôs, com tipos móveis, o texto da Bíblia, o primeiro livro a passar do manuscrito para o papel impresso (McGARRY, 1999).

A informática provocou muito mais do que uma revolução nas formas e nos métodos de geração, armazenamento, processamento e transmissão da informação. Pode-se dizer que o desenvolvimento das tecnologias relacionadas com o computador compara-se à revolução causada com a invenção dos tipos móveis de Gutenberg. A mudança do texto impresso para o texto eletrônico criou uma grande transformação na maneira como armazenamos e acessamos a informação. Os textos eletrônicos, que podem ter as mesmas características dos textos lineares, bem como dos não lineares, são armazenados ou disponibilizados em HDs (*Hard Disk* ou Disco Rígido), disquetes, *CD-ROM*, *LANs*, ou redes digitais. O formato eletrônico pode ser considerado como um formato adicional, em vez de substituição dos formatos existentes. Porém, a possibilidade de explorar o texto de maneira interativa introduziu o conhecimento por simulação.

A internet tem poucos pontos de entrada, uma estrutura pobre e uma forte ênfase na busca por assunto. A teoria e processos da indexação subsidiam diversas atividades em uma teoria comum e diversas técnicas, porém suas aplicabilidades na Web precisam ser adaptadas. Por exemplo: no ambiente Web, o processo de indexação pode ser útil e eficiente para criar estrutura conceitual com termos que podem ser armazenados na base de dados e servir como *links* para as partes relevantes do texto. Somado a isso, o uso de um Sistema de Organização do Conhecimento, do tipo tesouro, pode auxiliar no controle do vocabulário para melhorar a recuperação na Web, reduzindo a frustração e ajudando a assegurar que a informação relevante não se perca.

Lima (2004) aponta que o hipertexto, fruto da tecnologia informacional das últimas décadas, é a forma texto que estimulou e tornou possível o desenvolvimento do universo digital. O hipertexto, como nova tecnologia, também desterritorializa o texto, tornando-o sem fronteiras, o que remete, de certa forma, à transmissão da informação na tradição oral, em que o elemento da aleatoriedade é inerente à comunicação. Porém, o texto eletrônico acelera o tempo e concatena contextos; encadeia e justapõe diferentes documentos, num compartilhamento do mesmo espaço de produção e, às vezes, alterando sua compreensão.

Os hipertextos já eram utilizados, em modo off-line, para representação de informações, imagens, sons, organizados de forma a

promover uma leitura (ou navegação) não linear, baseada em indexações e associações de ideias e conceitos, sob a forma de *links*. No entanto, essa concepção de textualidade possibilitou a surgimento da *World Wide Web* (WWW), nascida em 1991, que é a parte da Internet baseada nos princípios do hipertexto. Uma característica marcante do ambiente Web é o fato de ser uma organização descentralizada, o que a torna um sistema ágil, diferente do ambiente físico (ou real), com seus tradicionais sistemas organizados de forma centralizada. As linguagens de marcação de textos *HTML* (*Hypertext Markup Language* ou Linguagem de Marcação de Hipertexto) e *XML* (*eXtensible Markup Language* ou Linguagem de Marcação Estendida) são consideradas a linguagem básica de representação do hipertexto na Web. Sendo que a linguagem *XML* surgiu como recurso diferenciado na autoria de hipertextos, pois permite aos autores a marcação de conteúdo semântico dentro do próprio texto e o acréscimo de comandos, explicitando decisões de fragmentação e interconexão segundo configurações mais convenientes para aquele documento específico.

De acordo com Mai (2004, p. 93), a maioria das pesquisas da área de organização da informação e do conhecimento concentra-se mais na melhoria da representação do item bibliográfico na WWW do ponto de vista descritivo, por meio de seus metadados, do que a representação de seu conteúdo, e exemplifica esta realidade com uma citação: “Weinberg (1999) calls it “*the stepchild of the metadata literature*.”, que, apesar de ser de 1999, tem essa perspectiva mantida até o momento atual.

No entanto, o processo de indexação, especificamente a leitura documentária na etapa da *análise de assunto*, para modelagem conceitual de um hipertexto, dá-se de forma diferenciada por suas características não lineares, o que torna mais complexo realizar a leitura documentária com perspectivas de sua representação na Web. Agregam-se a este fato as diversas tipologias documentais existentes e suas estruturas, que exigem conhecimentos diferenciados, no que tange à identificação de seus conceitos, tornando, assim, um procedimento ainda mais complexo para sua representação.

O principal papel da leitura documentária é a identificação de conceitos de um determinado documento para que seu conteúdo possa ser representando visando a sua recuperação pelo leitor em uma unidade de informação. Porém, existem variáveis que irão permear esta leitura textual,

que é realizada diferentemente da convencional, que integra o conhecimento profissional do leitor-indexador, ao realizar este procedimento, perpassando pelos aspectos linguísticos, cognitivos e lógicos que são associados a este procedimento, e as variáveis: o texto, o leitor e o contexto.

Neste capítulo, não se pretende esgotar a temática sobre hipertextos como um documento Web tanto na sua construção quanto na sua navegação e recuperação, mas abordar as perspectivas da contribuição da leitura documentária na etapa da *análise de assunto*, para modelagem conceitual de hipertextos, com vistas à sua disponibilização na Web, do ponto de vista do autor conteudista.

## **2 HIPERTEXTO**

A inquietude para o surgimento de uma nova forma textual já se delineava no início da década de 1970. Ironicamente, a necessidade de mudança nos conceitos das funções do autor e do leitor, do livro e do texto, foi trazida à tona por expoentes da teoria crítica literária como Foucault (1979), Derrida (1977) e Barthes (1979), cujas posturas filosóficas tratavam do texto tradicional, e não do hipertexto. Para Foucault (1979, p. 102) “[...] o texto é como uma rede formada por interconexões, na qual as fronteiras não são bem definidas, mas captadas em um sistema de referências e outros livros, outros textos.”. Derrida (1977, p. 173) aponta que “[...] o texto ideal é composto por unidades que, apesar de separadas, podem se complementar na construção do sentido. Cada unidade pode ‘quebrar’, modificar o contexto e abrir para uma infinidade de novos contextos.”. Já Barthes (1979, p. 8) descreve o texto ideal como “[...] um composto de blocos de palavras (ou imagens), perpetuamente ‘não finalizados’, descritos a partir de termos como nós, networks, redes e caminhos.”

Permeando a reflexão desses três filósofos, está a ideia de um rizoma que se materializaria na forma de texto hipertextual, na qual não haveria uma hierarquia de início ou fim, nem de acima ou abaixo, nem de antes ou depois. Foucault fala da interrelação, invisível a muitos, entre obras separadas fisicamente. Derrida reconhece a fragmentação inerente ao texto, mas também sua mobilidade e potencial de transformação como uma entidade viva. Já Barthes aborda o texto como fenômeno em processo, não como matéria acabada.

Segundo Koch (2002), se considerarmos que a linguagem é uma atividade interativa que conduz à concepção processual da construção do sentido e que todo texto é constituído por uma proposta de múltiplos sentidos, como querem Foucault, Derrida e Barthes, pode-se afirmar que todo texto é um hipertexto. Para que o leitor possa construir um sentido, que nem sempre se dá de maneira linear e sequencial, é necessário realizar um constante movimento em variadas direções, recorrendo a diversas fontes de informação textuais e extratextuais. Dessa forma, o hipertexto é uma forma de estruturação textual que permite ao leitor ser uma espécie de coautor do texto, oferecendo-lhe a possibilidade de opções entre caminhos diversificados, de modo a permitir diferentes níveis de desenvolvimento e aprofundamento do tema. Esta variabilidade do grau de aprofundamento é notável em textos impressos. Povoado de referências, citações, notas de rodapé, quadros, tabelas, exemplos etc., sua concentração de estímulos convoca o leitor a pensar sobre outros textos, consultá-los e até mesmo suspender sua leitura, em um processo de alteração de curso para obter outras informações ou fazer anotações. Como o hipertexto oferece uma multiplicidade de caminhos a seguir, compete ao leitor decidir qual o fluxo de informações será incorporado em sua trajetória de leitura.

A definição de texto virtual de Bolter (1991), duas décadas depois de Foucault, Derrida e Barthes, também confirma a inquietude já presente naqueles autores:

[...] texto é um contínuo de parágrafos, dispostos na tela do computador para o leitor ler através de um caminho tradicional. Algumas palavras estão marcadas em negrito; estilo que indica que há uma observação para aquela palavra ou frase, [...] uma segunda janela pode também conter frases em negrito que podem levar o leitor para outros parágrafos. [...] O processo pode continuar indefinidamente [...] formando uma rede que é chamada de hipertexto. (BOLTER, 1991, p. 15).

Bolter (1991) nos mostra, também, que a noção de texto acompanhou o desenvolvimento da tecnologia de transmissão da informação:

Todas as formas de escrita são espaciais [...] Cada tecnologia dá-nos um espaço diferente. Para certas escritas antigas, o espaço [bidimensional] era a superfície interna do rolo contínuo [de pergaminho], que o escriba dividia em colunas. Para a escrita medieval e a impressão na era moderna, o espaço é a superfície branca de uma página, especialmente em volumes encadernados. Para a escrita eletrônica, o espaço é a tela do monitor de um computador, onde o texto é exposto e armazenado em memória eletrônica. (BOLTER, 1991, p. 11).

No contexto histórico, o documentalista e advogado belga Paul Otlet e seus colegas se anteciparam muito da funcionalidade do hipertexto e desenvolveram um complexo de organizações funcionalmente similares a ele, conforme uma série de trabalhos publicados entre 1892 e início deste século. Na obra *Traité de Documentation* (1934), Otlet identificou máquinas intelectuais e processos contendo dispositivos de som e conectados ao telefone, telegrafia e televisão, com o armazenamento baseado principalmente em cartões e folhas de papel de tamanho padrão. A invenção destas máquinas ajudaria a realizar um novo tipo de enciclopédia e a tornaria muito próxima a um anexo do cérebro, um substrato da memória. O autor vislumbrou um sistema internacional de manipulação de informações abrangendo desde a criação de uma entrada em um catálogo até as novas formas de publicação, do gerenciamento de bibliotecas, arquivos e museus como agências de informação inter-relacionadas até o desenvolvimento colaborativo de uma enciclopédia universal codificando todo o conhecimento humano. No centro dessas ideias, estavam o sistema de Classificação Decimal Universal, um novo tipo de agência de informação (*Office of Documentation*), um novo princípio de armazenamento e indexação da informação (Princípio Monográfico) e o microfilme.

Em 1945, Vannevar Bush (1987a) publica o artigo *As we may think*<sup>26</sup>, no qual previu um sistema muito parecido com o hipertexto, com a ideia básica de organizar documentos em trechos e combiná-los conforme as necessidades de compreensão e organização, o qual denominou de *Memex*. Esse sistema teria a capacidade para armazenar um grande volume

---

<sup>26</sup> Publicado originalmente em: *The Atlantic Monthly*, v. 176, n.1, p.101-108, July 1945.



de informações e, por meio de determinado mecanismo, estabeleceria associações entre diferentes pontos da base de dados, propiciando a sua consulta com rapidez e flexibilidade. Assim, o *Memex* serviria como suplemento para a memória dos usuários de um banco de dados de literatura científica. Ele propôs a organização associativa que tentava imitar a estrutura da memória humana.

Além de Bush e Otlet, Theodor Nelson (1965) e Douglas Engelbart (1962) foram outros pesquisadores que compartilharam a mesma visão de que seria estabelecida nova forma de interação homem-computador. Theodor Holm Nelson foi quem criou o termo hipertexto, definindo-o como “[...] uma combinação de texto em linguagem natural com a capacidade do computador de fazer pesquisa interativa e exibição dinâmica de um texto não linear.” (CONKLIN, 1987, p. 17). O sistema *Xanadu*, desenvolvido por ele e sua equipe de pesquisadores, possuía um gerenciador de armazenamento que permitia a criação de ligações entre os tópicos correlatos e mantinha informações sobre a origem, as variações e as interconexões dos textos. Na década de 1960, Nelson começou a estruturar alguns fundamentos do hipertexto e apontou as limitações da informática em face da nova ferramenta. Paralelamente aos estudos de Nelson, Douglas Engelbart, do Instituto de Pesquisa Stanford, EUA, estudou a utilização do computador como instrumento representacional e ampliador do intelecto humano (MARQUES, 1995). Foram apresentados novos conceitos relativos à interação homem-máquina para a construção de sistemas mais dinâmicos e flexíveis. Juntamente com outros pesquisadores, implementou o sistema *On Line System* (OLS), conhecido mundialmente como sendo a primeira implementação que seria conhecida como um hipertexto, estruturado por uma rede de *links* que interligavam blocos pequenos de informação. Esse sistema foi o primeiro a introduzir noções como edição estruturada, uso do *mouse*, correio eletrônico, janelas múltiplas e tela compartilhada. Em 1969, Bush publicou o artigo *Evolution of an information society* (republicado em 1987b),<sup>27</sup> concluindo que o *Memex* podia, finalmente, ser viabilizado, porém a um alto custo, exigindo também manipulação de alta tecnologia e uma enorme dose de paciência dos usuários.

---

<sup>27</sup> BUSH, Vannevar. MEMEX revisited. In: CALKELL, A. E. (ed.). *Evolution of an information society*. London: ASLIB, 1987. p. 179-191. Publicado originalmente em *Science is not enough*. New York: Apollo, 1969. p. 75-101.

Os sistemas de hipertexto são fruto da tecnologia informacional das últimas décadas e causaram uma verdadeira revolução nos conceitos de armazenamento e recuperação de informações. As rupturas com as formas tradicionais de organização da informação vão desde as características de entrada da informação e a criação das estruturas de armazenamento, até o desenvolvimento radical no processo de recuperação da informação, possibilitado pela interação homem–máquina.

Com o surgimento da Web, no início da década de 1990, os sistemas de hipertextos alcançaram maior popularidade e visibilidade mundial. Hoje o sistema de hipertexto mais conhecido é a própria rede Web, na qual o navegar no hipertexto, dado seu alcance mundial, é uma realidade para milhões de pessoas em todo o mundo (NONATO, 2009). A mudança das páginas estáticas disponíveis na internet para páginas interativas com conexões, utilizando o protocolo HTTP (*Hypertext Transfer Protocol*), alterou a forma de acesso da informação na Internet. Gomes (2011, p. 23) afirma que “[...] podemos dizer que a Web, hoje, é um hipertexto aberto, pois quase tudo o que há nela está interconectado, e documentos podem ser acessados e acrescentados a partir de qualquer ponto (mesmo simultaneamente).”.

A maioria das estratégias de estruturação de hiperdocumentos se apoia nas linguagens de marcação de textos como *HTML* (*Hypertext Markup Language* ou Linguagem de Marcação de Hipertexto) e *XML* (*eXtensible Markup Language* ou Linguagem de Marcação Estendida). Segundo o W3C (2010), a linguagem de marcação de hipertexto evoluiu para as versões HTML+, HTML2.0 e HTML3.0, de 1993 a 1995, quando várias mudanças foram propostas para melhoria da linguagem e, somente em 1997, trabalhou na versão 3.2 da linguagem, fazendo com que ela fosse tratada como padrão de linguagem. Também para facilitar a manipulação dos elementos, possibilitando, assim, ao desenvolvedor, modificar as características dos objetos, de forma que sejam transparentes para o usuário final, criando novas *tags* e permitindo modificar a função das demais. Em 2007, a W3C buscou soluções para implementação de uma versão da linguagem HTML moderna e compatível com as versões existentes. Para tanto, a linguagem deve ser única, podendo ser escrita tanto com sintaxe HTML como XML (*eXtensible Markup Language*), denominada HTML5.

Essas linguagens são gerenciadas por uma série de comandos que, uma vez inseridos em um documento tradicional, identificam sua estrutura lógica em títulos, seções e parágrafos. Esses tipos de componentes de um documento servem de guias para o processo de conversão de textos em hipertextos, pois sugerem maneiras de fragmentá-los e de conectar algumas de suas partes entre si. Estas divisões lógicas em títulos, seções e parágrafos podem ser suficientes para determinar uma boa configuração de um hipertexto, no caso de documentos cuja organização interna é bem-definida e rígida, a exemplo de dicionários e catálogos. Todavia, podem ser pouco eficientes em textos livres com subdivisões menos sugestivas, para os quais os algoritmos de conversão gerariam redes não muito adequadas à estruturação de seu conteúdo semântico. Entre as linguagens de marcação, o *XML* surgiu como recurso diferenciado na autoria de hipertextos, pois permite aos autores a marcação de conteúdo semântico dentro do próprio texto e o acréscimo de comandos, explicitando decisões de fragmentação e interconexão segundo configurações mais convenientes para aquele documento específico.

Encontram-se na literatura vários estudos e definições sobre hipertexto advindas de diferentes áreas do conhecimento, tais como: Ciência da Computação, Linguística, Educação, Ciências Cognitivas, Ciência da Informação, Gestão da Informação e Comunicação. Cada uma dessas perspectivas abordam o hipertexto sob um ponto de vista que vai ao encontro de sua aplicabilidade em cada área especificamente. Salienta-se, também, que alguns autores não fazem distinção entre hipertexto e hiper-mídia. Entretanto, a hiper-mídia ultrapassa o hipertexto por considerar qualquer tipo de mídia, não apenas texto, e com o suporte de qualquer tecnologia, e não apenas de computadores (RADA, 1991b).

Uma das definições mais citadas sobre hipertexto é a de CONKLIN (1987, p. 17), que o define como “[...] janelas na tela [que] são associadas a objetos na base de dados e ligações [que] são estabelecidas entre estes objetos, tanto graficamente, na forma de marcas rotuladas, como na base de dados, na forma de ponteiros.”.

Conklin (1987) apresenta, ainda, três elementos que podem ser considerados hipertextos: (1) *um componente de banco de dados*, em que objetos na tela são associados a objetos armazenados em um banco de dados. As conexões entre estes objetos podem ser definidas não apenas graficamente, mas também por ponteiros em uma base de dados; (2)

*um esquema de representação*, no qual a informação textual informal é misturada com processos e operações mais formais e mecanizados. Para ser qualificado como “hipertexto”, o sistema deve proporcionar um suporte eletrônico constante para suas inter-relações; (3) *uma interface com o usuário*, geralmente janelas representando a correspondência “um-para-um” com os nós/nodos na base referencial de dados. A interface gráfica, com o auxílio de *browsers* e diagramas, auxilia o usuário a navegar através de uma grande quantidade de informações, ativando as ligações e acessando o conteúdo dos nodos.

Para Snyder (1997 p. 126), o hipertexto é como “[...] uma estrutura composta por blocos de textos conectados por *links* eletrônicos, os quais oferecem diferentes caminhos para os usuários.”.

Levy (1993, p. 33), por sua vez, define o hipertexto como

[...] um conjunto de nós ligados por conexões. Os nós podem ser palavras, páginas, imagens, gráficos, sequências sonoras, documentos complexos que podem eles mesmos ser hipertextos. Os itens de informação não são ligados linearmente, como uma corda com nós, mas cada um deles, ou a sua maioria, estende suas conexões em estrela de modo reticular.

No Brasil, um dos primeiros autores a estudar sobre o hipertexto foi Vilan Filho (1994), que o definiu como “[...] uma rede de nós e ligações entre documentos, onde documentos são os nós e as ligações são referências cruzadas.” (VILAN FILHO, 1994, p. 297). Esse autor destaca ainda que

Os sistemas de hipertextos são conjuntos de programas de computador (suporte lógico), [com] suas tabelas e dados de controle necessários para a operação de um sistema construído para operar com hiperdocumentos, segundo a filosofia de hipertexto. Em geral, esses sistemas são compostos de: (a) um subsistema de autoria, e (b) um subsistema de navegação (VILAN FILHO, 1992, p. 61).

Lima (2004, p. 23) considera que o hipertexto pode ser

[...] tanto uma ferramenta para o autor quanto para o leitor. Para o autor, funciona como ferramenta que ajuda a organizar, estruturar e armazenar informações. Para o leitor, ele permite, através da navegação, a recuperação da informação organizada.

De acordo com Lima (2004), os sistemas de hipertexto traduzem uma abordagem de estruturação e manipulação de textos caracterizada pela não linearidade, tendo como seus componentes básicos:

a) *Nó* (ou *Nodo*): em cada uma das unidades de informação na base hipertexto, os nós correspondem a uma ou mais exibições de tela. Essas unidades de informação em um hiperdocumento podem conter diferentes tipos de dados, textos, figuras, fotos, sons e são conectadas por *links* a uma variedade de estruturas. Geralmente, o nó descreve um único conceito ou tópico, de modo que pode ser caracterizado como autocontido, não dependendo da leitura prévia de outros nós.

b) *Link*: conceito mais importante do hipertexto, o *link* é também denominado elo, ligação, vínculo, âncora ou botão, de acordo com o sistema de hipertexto. *Links* são marcas que conectam um nodo com outro. A ativação desses *links* implica abertura de nova janela, contendo o documento referenciado. Assim, é possível deslocar-se fácil e rapidamente, de um ponto a outro, no conteúdo das páginas. Esses *links* podem ser representados por palavras ou frases em destaque (negrito, itálico, cores), ou, ainda, por figuras ou ícones.

Frei e Stieger (1995) dividem os *links* em dois tipos: *links* referenciais (ou organizacionais) e *links* semânticos (ou de conteúdo). A finalidade do *link* referencial é tornar a leitura do documento mais confortável. Já a função do *link* semântico é apontar para um detalhamento da informação ou uma informação similar, ou, ainda, uma informação adicional em tópico específico.

Em relação à sua estrutura, o hipertexto determina e descreve um sistema de ligações (*links*) ou relacionamentos entre os nós ou unidades de informação, sendo um fator decisivo na facilidade de criação, uso e atualização do hiperdocumento.

De acordo com os nós conceituais que estão sendo ligados, RADA (1991a) classifica o hipertexto conforme as tipologias microtexto e macrottexto. Microtexto é um neologismo que significa um hipertexto de pequeno volume, com ligações explícitas entre seus componentes. Aparentemente, os microtextos são muito apropriados para navegação. Uma rede de microtextos pode ser vista como uma rede semântica e, por isso, ser tomada como modelos da memória humana. Através dos recursos

computacionais, evidenciam-se o caminho ou os caminhos em que o texto foi internamente construído, e que podem variar de acordo com pontos de vista lógicos. Macrotexto é definido como um sistema hipertextual em larga escala, composto de vários documentos, e cuja ênfase está em *links* que os conectam entre si, em vez de conectar internamente apenas um documento.

Corroborando com essa ideia, Lévy (1993, p. 25) considera o hipertexto como “[...] uma metáfora cálida para todas as esferas da realidade em que significados estejam em jogo.”, e aponta seis princípios básicos de sua lógica, que permitem visão panorâmica, a qual organiza, resume e amplia a ideia de rede que se pretende construir: (1) Princípio de metamorfose: processo de constante construção e renegociação de sentidos, que se dá nos hipertextos. O princípio de metamorfose explicita a ideia de que a rede de significados que constitui o conhecimento está em permanente transformação; (2) Princípio de heterogeneidade: a variedade e diferença entre as informações organizadas em determinada seção do hipertexto, como os nós e as conexões que se estabelecem entre as diversas partes da rede hipertextual. Existem, na memória, informações cognitivas que passam por vários tipos de associações entre estes elementos; (3) Princípio de multiplicidade e de encaixe das escalas: organização “fractal” do hipertexto, ou seja, qualquer nó ou conexão, quando analisado, pode revelar-se como sendo composto por toda a rede, e assim por adiante, indefinidamente, ao longo da escala dos graus de precisão; (4) Princípio de exterioridade: a rede não possui unidade orgânica, nem um “motor interno”. Seu crescimento, diminuição ou recomposição permanente dependem de ação exterior indeterminada. O princípio de exterioridade pretende caracterizar a permanente abertura da rede hipertextual e do conhecimento em construção; (5) Princípio de topologia: nos hipertextos, tudo funciona por proximidade, por vizinhança. Neles, o curso dos acontecimentos obedece a uma topologia, na qual os caminhos seguem a ideia de proximidade entre significados; (6) Princípio de mobilidade dos centros: a rede não tem centro fixo, mas possui diversos centros que são como pontas luminosas perpetuamente móveis, saltando de um nó a outro, trazendo ao redor de si uma ramificação infinita de pequenas raízes, como finas linhas brancas esboçando por um instante um mapa de informações.

Essas seis características da rede hipertextual propostas por Lévy (1993) constituem uma teia básica a partir da qual pode ser tecida a ideia de rede para a representação do conhecimento. Assim, as características próprias do hipertexto e as novas relações entre autor e leitor que este sistema está introduzindo estão transformando radicalmente a comunicação. O hipertexto é capaz de sedimentar uma nova forma de literatura, conforme previsto por alguns autores: Landow (2006), Murray (1997) e Bolter (1991). No cenário atual, vislumbram-se alguns elementos básicos, como: (1) a escrita e a leitura não sequencial; (2) a interatividade possibilitada pelo meio digital; (3) a existência de elos de ligação (*links*) – textuais ou não.

Pode-se dizer, portanto, que o hipertexto não pode ser considerado um gênero textual, mas documentos eletrônicos, organizados em forma de rede, com *links* eletrônicos que possibilitam a interatividade entre as relações inter e intratextuais de diferentes segmentos textuais.

Conclui-se, assim, que o hipertexto é tanto uma ferramenta para o autor (designers/programadores) quanto para o usuário (leitor). Para o autor, funciona como ferramenta que ajuda a organizar, estruturar e armazenar informações. Para o leitor, permite, por meio da navegação, a recuperação da informação organizada. Araújo (2006, p. 38) acrescenta que “[...] se pensarmos que o autor é quem decide como e onde disponibilizar os *links* no hipertexto, não haveria uma liberdade incondicional do leitor e sim uma liberdade, até certo ponto de vista vigiada.” Já Ensslin (2007) registra que existem estudos sobre como os princípios de organização estrutural podem moldar os processos de autoria e leitura do hipertexto. Acrescenta que os papéis de autor e leitor têm, frequentemente, sido revertidos, principalmente no que tange à relativa liberdade de escolha do usuário entre diferentes caminhos de leitura e suposto “desempoderamento” simultâneo do autor.

Para Vilan Filho (1992), a estruturação da informação em sistemas de hipertexto é similar à de uma rede semântica, ou seja, um modelo de representação do conhecimento. Os nós representam conceitos e ideias, e as ligações representam as relações entre eles. As relações devem refletir a estrutura organizacional do assunto relacionado ou, em outras palavras, uma rede semântica de uma área do conhecimento. Se a informação for disposta de maneira organizada, será fácil para o usuário recuperar o que

estiver procurando. Por outro lado, se a estrutura apresentar uma trilha de navegação desorganizada e uma estrutura cognitivamente ineficiente, a busca e o folheio (*browsing*) por parte do usuário será difícil ou ineficiente.

Para este estudo, considerou-se a primeira abordagem, a do autor, apresentada por Pierre Lévy (1993), na qual o hipertexto é utilizado como ferramenta de organização da informação por meio da sua característica de escrita não sequencial e da associação dessas informações através de nós, representados por conceitos, possibilitando assim a construção de uma rede navegacional em documentos completos sem perda significativa do conteúdo semântico após a fragmentação. Como o processo de organização hipertextual é muito parecido com a indexação tradicional, são apresentados, a seguir, os princípios da teoria da indexação e suas possíveis aplicações aos sistemas de hipertexto.

### **3 INDEXAÇÃO**

Indexar é representar um documento por meio de uma descrição abreviada de seu conteúdo essencial. Embora algumas vezes a complexidade desse processo possa ser minimizada, essa é, sem dúvida, umas das atividades mais subjetivas do cotidiano do bibliotecário. A capacidade intelectual de reconhecer o conteúdo de um documento em análise é a questão central do procedimento de indexação. Essa representação é realizada a partir da análise do conteúdo do texto-fonte, com base em metodologias e procedimentos particulares, sempre levando em consideração uma clientela específica.

Para a execução da atividade, faz-se necessário ter uma formação especializada na área de tratamento de conteúdos de documentos. A experiência e os conhecimentos prévios desse profissional diferenciarão o resultado do tratamento informacional realizado por outro indexador menos preparado. Naves (2004, p. 8) afirma que:

O profissional da informação que desenvolve a atividade de indexar assuntos de documentos é chamado de *indexador*, catalogador de assuntos ou classificador. A maioria desses profissionais é graduada em Biblioteconomia, e deve conhecer os fundamentos teóricos e técnicos do tratamento temático da informação.



Por se tratar de uma atividade intelectual, é natural que, no cotidiano dos indexadores, sejam percebidas divergências entre termos atribuídos a um mesmo documento por profissionais de diferentes instituições e contextos. Assim, uma mesma publicação poderá apresentar conjuntos diferentes de termos de indexação, dependendo do grupo de usuários ao qual se destina e dos interesses particulares desse grupo, ou seja, há várias formas ‘corretas’ de se indexar.

De acordo com a ISO 5963, a indexação é vista como “[...] a representação do conteúdo dos documentos por meio de símbolos especiais, quer retirados do texto original, quer escolhidos numa linguagem de informação ou de indexação.” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 1985). Segundo Silva e Fujita (2004, p. 136-137), “[...] o conceito de indexação surgiu a partir da elaboração de índices e atualmente está mais vinculado ao conceito de *análise de assunto*.” De acordo com o UNISIST (1981, p. 8), “[...] indexação é o processo utilizado para descrever e identificar um documento de acordo com seu assunto.”

Na Biblioteconomia e Ciência da Informação (BCI), a *análise de assunto* é parte do processo de indexação que envolve procedimentos cujo número de etapas varia conforme diferentes autores. Jens-Eric Mai (2001) ressalta que o processo de indexação é algumas vezes apontado, na literatura, como tendo duas, três ou mesmo quatro etapas. Conforme as diretrizes da *International Standards Organization-ISO 5963*, do UNISIST (1981) e conforme os autores Foskett (1973), Chaumier, (1988); Fugmann (1993), Taylor (2004) Lancaster (2004), Dias e Naves (2007), esse processo é realizado em duas etapas, sendo a primeira a da *análise de assunto*, ou extração de conceitos; e a segunda etapa a de *tradução* dos conceitos nos termos de uma linguagem de indexação. Já os autores Hutchins (1978), Frohmann (1990) e Blair (1990), Farrow (1991), Taylor (2004) e Petersen (1994) e a ABNT 12.676 (1992) adicionam mais uma etapa, que, apesar de, também, considerarem como primeira a *análise de assunto*, incluem uma etapa intermediária, que reformula o assunto em uma frase de indexação na linguagem natural, para depois fazer a tradução; enquanto Chaumier (1988), Langridge (1989), Van Slype (1991), Chu e O’Brien (1993) e Robredo (2005) consideram quatro etapas. Contudo, todas têm basicamente as mesmas operações. O foco principal deste capítulo é, na primeira etapa, a *análise de assunto*, especificamente, o processo da leitura

documentária e as perspectivas de aplicação em contexto da Web, tendo o hipertexto como objeto de estudo. Sendo assim, na próxima seção, passa-se a descrever mais detalhadamente essa etapa, com ênfase nesse processo da leitura documentária.

### 3.1 ANÁLISE DE ASSUNTO E A LEITURA DOCUMENTÁRIA

A etapa de *análise de assunto* determina de que trata um documento, isto é, qual é o seu assunto. De acordo com Cesarino e Pinto (1980, p. 32, grifo do autor), “[...] a *análise de assunto* é a operação-base para todo o procedimento de recuperação de informação.”. Isso justifica o fato de todo Sistema de Recuperação da Informação (SRI) utilizar essa técnica, tanto na etapa de entrada de dados no sistema quanto no momento de busca da informação pelo usuário final. A *análise de assunto* pode ser considerada a etapa mais importante do processo de indexação.

Segundo Lancaster (1993, p. 20), “[...] ao indexador raramente é dado o luxo de poder ler um documento do começo ao fim.”, realidade que pode ser comprovada no cotidiano de indexadores de instituições nacionais e internacionais. Para tanto, a leitura e a compreensão do texto são primordiais, porém, o tempo restrito do indexador e a quantidade cada vez maior de documentos que demandam tratamento são fatores preocupantes, porque podem comprometer a qualidade da atividade realizada.

Fujita (2003, p. 69) acrescenta que

[...] o processo de análise de assunto reveste-se (*sic*) de uma subjetividade característica, dadas as circunstâncias e elementos envolvidos, pois, a partir da leitura do documento pelo indexador, é realizado um processo de comunicação interativo entre três variáveis: leitor, texto e contexto. Cada uma dessas variáveis estará sujeita a diferentes condições, mas é o indexador como leitor a variável mais influente nessa interação para análise de assunto, porque precisa realizar a compreensão da leitura mediante sua cognição.

Hjørland (1992) agrega que, para a execução dessa primeira etapa, é preciso considerar o domínio no qual o documento está inserido,

identificando as características específicas do campo de conhecimento, sejam elas de ordem cultural, terminológica, histórica ou linguística. Para tanto, o conhecimento do indexador sobre esse domínio é importante para a qualidade da análise. Assim, a atividade poderá ser feita de acordo com o contexto, pois o documento não será considerado algo isolado, mas parte de um todo.

Segundo Kobashi (1994, p. 112), “[...] o texto, em sentido mais amplo, designa uma unidade de comunicação organizada sintagmaticamente e dotada de coesão e coerência.”. O texto pode ser classificado, quanto à sua estrutura interna, como descritivo, narrativo, dissertativo ou, quanto à sua finalidade, como texto técnico, científico, didático, jornalístico, jurídico, político, de lazer e etc. A grande variedade de tipos de documentos pode ser diferenciada pelas características físicas e intelectuais: livros de leitura corrente, teses, dissertações, livros de referência, publicações periódicas e impressos diversos.

A leitura é um processo interativo de compreensão que depende das variáveis do texto (intenção do autor, forma e conteúdo), do leitor indexador (estruturas e processos), e do contexto (social, psicológico e físico) no qual a leitura está sendo realizada (GIASSON, 1993). O conhecimento prévio do indexador é fundamental para a compreensão do texto e identificação de seu conteúdo, para selecionar os conceitos que serão representados com o uso de vocabulário controlado, para que este possa ser recuperado pelos usuários da unidade de informação. Naves (2004, p. 5) considera:

O texto é o ponto de partida para operações analítico-documentárias, sendo constante a presença dos elementos conteúdo e forma, como partes essenciais do mesmo. Para ter uma competência textual, é preciso que, além de conhecer o texto que tem em mãos para análise sob todos os aspectos, o indexador faça dele uma leitura adequada [...].

Assim, pode-se destacar que, para a assimilação e a compreensão do conteúdo do texto, é necessário o processamento de informações na mente do indexador. De acordo com Naves (2004, p. 8), “[...] parece ser consenso entre os especialistas em leitura que o processamento do ato de ler se dá interativamente, dependendo dessa interação a compreensão ou

a não compreensão de um texto.”. No entanto, a leitura documentária difere da leitura comum, porque esta é realizada pelo leitor indexador fazendo a mediação entre as ideias do autor com as do usuário, visando à representação de conteúdo por meio de conceitos. A leitura documentária pode, então, ser considerada a atividade inicial da etapa da *análise de assunto*, que permitirá identificar o conteúdo de um documento a partir de seus aspectos intrínsecos e extrínsecos.

De acordo com os Princípios do Unisist (1981, p. 8), os três estágios para a etapa *análise de assunto* são: (1) compreensão do conteúdo do documento como um todo; (2) identificação dos conceitos que representam esse conteúdo e (3) seleção dos conceitos válidos para recuperação, sendo que se destaca que “na prática, esses três estágios se superpõem». Para Fujita (2003), esta superposição ocorre no momento da leitura do documento. Essa atividade é subjetiva, o que a torna complexa, exigindo, então, do indexador conhecimentos cognitivos, lógicos e linguísticos, e, conseqüentemente, interdisciplinares, para fazer uma análise coerente.

No primeiro estágio, *compreensão do conteúdo do documento como um todo*, o indexador deve fazer a leitura do documento, embora a leitura integral ideal seja impraticável. Destacam-se as partes entendidas como mais importantes para que o indexador possa se orientar durante a leitura; entre elas: título; introdução e as primeiras frases de capítulos e parágrafos; ilustrações; tabelas; diagrama e suas explicações; conclusão; e palavras ou grupos de palavras sublinhadas ou impressas com tipo diferente. Ressalta-se a importância de o indexador não se dedicar exclusivamente às partes preliminares do texto, uma vez que as intenções do autor se encontram no conteúdo principal, enquanto as partes finais indicam se essas intenções foram alcançadas.

No segundo estágio, *identificação dos conceitos que representam o conteúdo do texto*, o indexador deve abordar o texto de maneira lógica, selecionando os conceitos que representem adequadamente o assunto do documento. Para isso, recomenda-se que seja obedecido um esquema de categorias da área coberta pelo documento, como o indicado a seguir: o fenômeno; o processo; as propriedades; as operações; o material; o equipamento, entre outros aspectos (UNISIST, 1981).

Faz-se necessário aqui apresentar uma definição que será considerada sobre o termo *conceito*. Dahlberg (1978, p. 147) considera que

A formação de um conceito é feita através da identificação, no universo de itens (ideias, objetos, fatos, leis, etc.), de um item de referência (entidades, propriedades, atividades e dimensões). Enunciados verdadeiros acerca desse item são formulados e sintetizados sob forma verbal (termo, palavra) que será utilizada no universo do discurso. Os conceitos são, portanto, definidos como unidades do conhecimento, identificadas através de enunciados verdadeiros sobre um item de referência e representados por uma forma verbal.

Para Dahlberg (1978), quando caminhamos rumo a sintetizar, compilar e reunir enunciados verdadeiros sobre um objeto, estamos caminhando para a formação de um conceito. Assim, pode ser definido como sendo a compilação de enunciados verificáveis sobre determinado objeto, fixado por um símbolo linguístico.

Nesse segundo estágio, o processo cognitivo está intrinsecamente ligado ao processo de identificação da ideia principal apontada em um texto:

O processo de análise de assunto reveste-se de uma subjetividade característica, dadas as circunstâncias e elementos envolvidos, pois, a partir da leitura do documento pelo indexador, é realizado um processo de comunicação interativo entre três variáveis: leitor, texto e contexto. Cada uma dessas variáveis estará sujeita a diferentes condições, mas é o indexador como leitor a variável mais influente nessa interação para análise de assunto, porque precisa realizar a compreensão da leitura mediante sua cognição. (FUJITA, 2003, p. 69).

Sendo o indexador um “sujeito”, essa subjetividade é inerente a ele. O indexador, apesar de ser um profissional com ideologias próprias, é o mediador entre as ideias do autor com as necessidades dos usuários, mantendo-se imparcial.

O terceiro estágio da *análise de assunto*, chamado de *seleção dos conceitos válidos para recuperação*, corresponde à seleção daqueles conceitos que são julgados como úteis para a recuperação do documento indexado, tendo em vista os objetivos para os quais as informações são indexadas. Assim, nem todos os conceitos identificados serão necessariamente selecionados (UNISIST, 1981).

Há pesquisadores, como Lancaster (2004) e Naves (2004), que definem a atividade de seleção de conceitos como a *determinação da atinência*. Em tradução do inglês, a expressão corresponde ao termo *aboutness*, porém, encontram-se na literatura outras traduções: concernência, sobrecidade, temacidade.

O final do estágio é indicado com a definição da chamada *frase de indexação*. Essa é elaborada pelo indexador em Linguagem Natural (LN). Após todo o processo intelectual de leitura e compreensão do texto, de identificação e seleção de conceitos representativos do documento em foco, o indexador deve afirmar: “*Este documento trata de...*” A partir dessa definição, o indexador pode passar para a etapa final do processo de indexação, a tradução da *análise de assunto* em termos de indexação.

Tessier (1992) afirma que o processo de desenvolvimento de classes e relações de conceitos, que ocorre no processo de indexação, no qual o indexador pode utilizar diversas estratégias ao mesmo tempo, é um processo semelhante ao de ligação dos segmentos de textos, em um sistema de hipertexto.

Na próxima seção, passa-se a descrever sobre a perspectiva do uso da leitura documentária na estruturação hipertextual de um documento.

#### **4 A LEITURA DOCUMENTÁRIA NA CONVERSÃO DE TEXTOS EM HIPERTEXTOS**

Esclarece-se que o processo de leitura aqui referido estará direcionado ao procedimento da leitura documentária realizada pelo autor, com vistas à estruturação e conversão de hipertextos, e não a leitura que é realizada pelo usuário, no momento da navegação.

Nota-se, na literatura, que os teóricos em hipertexto têm dado maior ênfase a ele enquanto forma de leitura, interessando-se menos pelas suas características de ferramenta de auxílio ao autor e ao texto. No entanto, Landow (2006, p. 4, tradução nossa) ressalta que

Tanto como uma ferramenta para o autor, ou quanto um meio de leitura, um sistema de hipertexto permite que autores ou grupos de autores juntos vinculem informações, criem caminhos por meio de blocos de informações relacionados, anotem nos textos existentes e

criem notas que apontam os leitores para dados bibliográficos, ou para o corpo do texto referenciado... Os leitores podem navegar pelos textos anotados com referência cruzada vinculados de uma maneira ordenada, mas não sequencial.<sup>28</sup>

A importância fundamental do hipertexto como texto eletrônico é ter o potencial de implementar sua característica crucial – a capacidade do leitor/usuário adicionar *links*, comentários ou ambos, realizando o processo de leitura–escrita como um processo interativo comunicacional. No presente estudo, interessa mais este segundo aspecto: os detalhes que permitem criar um texto dinâmico, estruturado semanticamente em seus aspectos cognitivos e seu papel na navegação. No entanto, a função de leitor se funde com a do autor; e a divisão entre os dois, muitas vezes, é indistinta.

De acordo com Lima (2004), a criação de um hiperdocumento bem estruturado a partir do documento em papel é um processo de conversão que ainda exige um trabalho de reelaboração do texto. A passagem do “texto” ao “hipertexto”, como o acréscimo do prefixo gramatical sugere, é a hiperbolização da forma, dos contornos, dos limites da palavra e do texto. Mas se há mudança pela interrupção, há de se ter também uma continuidade. E esse ultrapassar de fronteiras do suporte impresso para o virtual adquire sentido se forem consideradas as características da textualidade precedente, averiguando a preservação de algumas de suas características, quais limitações são superadas e quais novos domínios são potencializados.

Enquanto o texto tradicional se caracteriza pela univocidade, a certeza, a linearidade, a perenidade, no hipertexto, o documento é dividido em pequenos módulos padronizados, com acesso seletivo e descontínuo, conforme ilustrado na Figura 1; porém deve-se assegurar ao usuário coerência e consistência internas, semelhantes à estrutura tradicional do documento. Assim, apesar de a informação hipertextual ser multissequencial e possibilitar uma leitura dinâmica e randômica, ainda

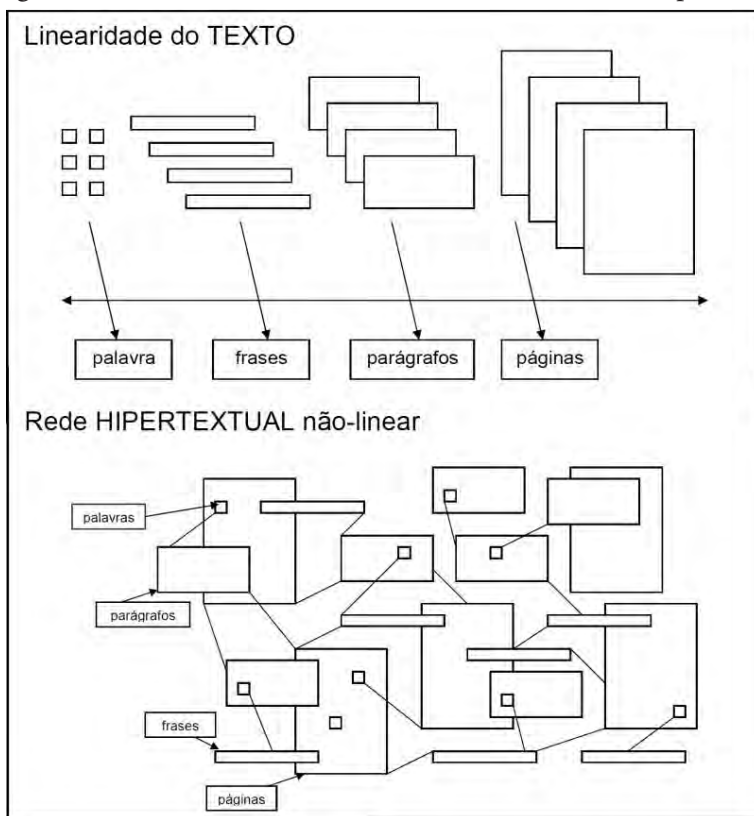
---

<sup>28</sup> *Both an author's tool and a reader's medium, hypertext document system allows authors or groups of authors to link information together, create paths through a corpus of related material, annotate existing texts, and create notes that point readers to either bibliographic data or a body of the referenced text... Readers can browse through linked, cross-referenced, annotated texts in an orderly but nonsequential manner.*

persiste, em sua estrutura, um mínimo de linearidade, de sequência para que se processe a inteligibilidade do material a ser lido.

O primeiro desafio na criação de um hipertexto é estruturar o conhecimento, de maneira que uma visão geral possa ser apresentada ao leitor no nó inicial, ou raiz. Essa visão deve identificar a ideia principal e a amplitude da cobertura, para que os leitores formem um modelo mental dos tópicos cobertos. Dependendo da aplicação e do tipo de informação em foco, um hipertexto estruturado pode ter o formato hierárquico de uma rede ou de filas (no mesmo nível). Os maiores problemas ocorrem nas fases de escolha e estruturação das informações relevantes e criação de *links* pertinentes.

Figura 1- A linearidade do texto e não linearidade do hipertexto



Fonte: Fachinetto (2005, p. 13).



Nessa conversão do texto linear para o texto não linear, faz-se necessária a realização prévia de um projeto para organização deste conteúdo textual, levando em consideração o contexto da obra analisada, a natureza do texto, perspectivas do usuário e o trabalho com a relação das ideais secundárias com a principal, para compreender a estrutura como um todo. Nesse processo, deve-se levar em consideração os *links* referenciais (ou organizacionais) e os *links* semânticos (ou de conteúdo). Os *links* referenciais são criados para a organização da superestrutura textual e o *link* semântico é utilizado para a criação da estrutura semântica hipertextual, direcionando o usuário para o conteúdo do nó. Pode-se utilizar o procedimento *bottom-up*, associando os *links* em estruturas, ou o *top-down*, no qual pegam-se as ideias gerais e as decompõem em detalhes, ou seja, em módulos.

Enquanto, no texto linear, o leitor indexador pode contar com a coerência textual e a estrutura hierárquica da informação, no caso do hipertexto, o autor tem que considerar que nem todos os parágrafos podem ter a mesma importância em todo texto a ser convertido. Na estruturação de um hipertexto, o autor deverá analisar os blocos de informação que terão seus conteúdos representados por unidades de informação/conceitos, que visam a uma representação com coerência e coesão em sua estrutura semântica, que abarque um conteúdo específico, e não dependendo da leitura prévia de outros nós para seu entendimento. Salienta-se que é por meio dos *links* que o autor constrói a coerência semântica do hipertexto, produzindo o significado dentro do contexto.

A questão da fragmentação é um dos maiores problemas que ocorre na estruturação de um hipertexto, pois não se pode comparar o parágrafo com um nó de informação. Segundo Heinonen (1998), a fragmentação é um problema de escolha dos limites dos parágrafos que melhor podem representar o conteúdo semântico fragmentado. Ela se dá na fase de escolha e estruturação das informações relevantes e criação de *links* pertinentes. Esses fragmentos, ou seções, são conectados e, quanto mais *links* existir em um hiperdocumento, mais fragmentado ele será.

Uma possível solução para a questão da fragmentação está no conceito de modularidade, que é a formação de blocos de informação e a estruturação do documento em módulos, permitindo definir os blocos de informação, os seus conteúdos, as ligações entre eles e a sua organização

coerente. O módulo é a unidade básica na construção do sistema. Modularidade, que será definida aqui como a divisão de documentos completos em pequenos módulos lógicos, é vista como importante conceito para as ferramentas biblioteconômicas que representam os conceitos de uma área, proporcionando a divisão mais específica dos esquemas dos documentos.

Cada módulo pode ser composto de um ou mais parágrafos que representam seu conteúdo, mantendo-o independente de outros módulos, sem a necessidade de relação entre eles, podendo ser representado por um único conceito. Isso permite que o autor do hipertexto use o conceito para representar o conteúdo do módulo, em vez de mostrar todas as sentenças e períodos que o compõem. Em cada módulo de informação, três aspectos são considerados essenciais: clareza, foco e completude. Esta interatividade se realiza por meio dos *links* determinados pelo autor do hipertexto, em níveis de modularidade que caracterizam essa navegação.

Dentre os procedimentos realizados na primeira etapa da indexação, *análise de assunto*, a leitura documentária tem como propósito principal a compreensão do texto para sua representação. Para fins de estruturação de um hiperdocumento, este procedimento não é exceção. Porém, o autor do hipertexto tem que ter em mente que o texto será refletido em uma tela, e os recursos de interatividade e multimodalidade devem ser levados em consideração.

De modo geral, o indexador realiza uma leitura rápida em diagonal no texto, analisando título, concentrando-se no início e no fim de capítulos, seções, resumos, índices dos livros e artigos. Analisam-se as partes mais importantes para extrair os conteúdos importantes do texto para, então, traduzir a informação que considera relevante para uma linguagem documentária. A leitura documentária tradicional com objetivo de compreender o conteúdo do documento e representá-lo não segmenta o texto em módulos para representá-los, mas sim, entende o texto como um todo com o intuito de fazer a organização estrutural do hipertexto, em blocos de informação, chamados nós, que serão inteligidos pelos *links* (frase indexatória) escolhidos posteriormente, com vistas à necessidade do usuário para recuperação do documento.

Assim, pode-se realizar o procedimento da leitura documentária para a representação do conteúdo do hipertexto seguindo as mesmas

orientações para sua realização no texto linear, mas deve-se adequá-la de acordo com as características que serão exigidas para representar um texto não linear. Deve-se levar em consideração a macroestrutura e a superestrutura textual, tendo em mente o sentido global do texto e sua estrutura. Aliado a isso, ressalta-se a mesma necessidade do conhecimento prévio do autor do hipertexto, as questões cognitivas, o contexto e o usuário.

O autor do hipertexto pode seguir os mesmos estágios previstos na *análise de assunto* realizada no texto linear. Começando pela leitura documentária, que perpassa todos os estágios, o autor (1) compreenderá o conteúdo do documento como um todo, analisando as partes entendidas como mais importantes, conforme descritas anteriormente, para que possa nortear-se durante a leitura, e entender as intenções do autor. No caso da compreensão do texto para conversão em hipertexto, o autor não precisa ir nas partes tão detalhadas, como sugerida pelo UNISIST (1981), podendo dividir o texto em quatro partes: o título, o resumo, o conteúdo principal do texto e as referências. Diferentemente da *análise de assunto* do texto linear, interessa mais para o autor a macroestrutura do texto, para determinar quais serão os blocos de informação que comporão a estrutura semântica, demarcando os limites textuais e transformando-os em pequenos módulos lógicos e coerentes, que comporão a organização estrutural deste hipertexto, completando, assim, o primeiro estágio da *análise de assunto*.

Na sequência, o autor do hipertexto passa para o segundo estágio, no qual irá (2) identificar os conceitos/*links* que representarão os conteúdos dos módulos selecionados de uma maneira lógica; porém, neste estágio, o autor deve ter em mente que ele não estará selecionando conceitos que irão representar o documento no todo. Como descrito anteriormente, cada módulo selecionado é transformado em um conteúdo independente; portanto deve-se pensar no conteúdo do módulo, porém sem perder a ideia do documento no todo; fazendo a mediação entre as ideias do autor com as necessidades dos usuários, inseridos dentro de um determinado contexto. Esses módulos irão formar a estrutura semântica do documento, na qual o usuário poderá selecionar, fazer associações, que podem ser sequenciais ou não, sendo distribuídos em níveis hierárquicos. Sobrepor diferentes camadas ou vários documentos suplementares sobre um único assunto pode simular um hipertexto estratificado, como se fossem vários textos em um só documento.

No terceiro estágio, o autor irá (3) selecionar os conceitos válidos para representação do conteúdo de cada módulo, transformando-os em *links*, que irão constituir a rede de conexões do hipertexto. Nesse estágio, o autor deve considerar, também, os aspectos conceituais envolvidos na determinação dos *links*, que constituem uma unidade de informação, que levará o usuário ao acesso do conteúdo do nó. Este procedimento de seleção dos conceitos relaciona-se com a escolha dos conceitos julgados relevantes para o acesso ao módulo de informação, tendo em vista os objetivos a serem alcançados. Ressalta-se aqui que, no caso dessa atividade no hipertexto, escolhe-se somente um conceito que representará o módulo de informação, seja por meio de um termo ou uma frase de indexação.

O Quadro 1 apresenta os aspectos das atividades da leitura documentária na conversão do texto em hipertexto, tendo por base os estágios dessa leitura sugeridos pelo Unisist (1981). Ao final desse processo intelectual, passa-se para a implementação da estrutura do hipertexto, combinando os módulos de informação, levando em consideração o processamento cognitivo e lógico do usuário.

Quadro 1 - Atividades da leitura documentária realizadas durante o processo de conversão de texto em hipertexto

ATIVIDADES DA LEITURA DOCUMENTÁRIA- UNISIST (1981).	ATIVIDADES DA LEITURA DOCUMENTÁRIA REALIZADAS DURANTE O PROCESSO DE CONVERSÃO DO TEXTO EM HIPERTEXTO.
Compreensão do conteúdo do documento como um todo.	Compreensão da macroestrutura do texto para determinar quais serão os blocos de informação que comporão a estrutura semântica, demarcando os limites textuais e tornando-os pequenos módulos lógicos e coerentes, que comporão a organização estrutural deste hipertexto.
Identificação dos conceitos que representam esse conteúdo.	Identificação dos conceitos/ <i>links</i> que representarão os conteúdos de cada módulo selecionados, sem perder a ideia do documento no todo.

Seleção dos conceitos válidos para recuperação.	Determinação dos <i>links</i> que representarão os conteúdos dos módulos na estruturação do hipertexto.
---	---

Fonte: Elaborado pela autora.

## 5 CONSIDERAÇÕES FINAIS

A mudança ocorrida no mundo ocidental desde os meados do século XII até a chegada da informática transformou e evoluiu as formas de expressões do conhecimento. A alteração do texto impresso para o texto eletrônico criou uma grande transformação na maneira como armazenamos e acessamos a informação. Surgem os textos eletrônicos, que podem ter as mesmas características dos textos lineares; e podem também ser textos não lineares, representando uma nova forma de escrita, chamados de hipertexto.

A forma não linear do hipertexto traz uma fragmentação textual como um problema, requerendo, assim, uma estrutura de relacionamentos configuradores de uma rede semântica que assegure a organização e recuperação deste conhecimento de acordo com o raciocínio humano. A modularidade apresenta-se como uma solução para a fragmentação, em que se subdividem os textos em unidades menores com conteúdos completos e mais focalizados, facilitando a organização do próprio autor do hipertexto e daqueles encarregados de sua manutenção. Além disso, o fato de serem esquemáticos os torna mais cognitivos, facilitando sua absorção e retenção na memória.

Na construção de um hipertexto, mantém-se a noção de autoria, porém alteram-se as relações entre o livro, o autor e o leitor. Assim, as fronteiras entre ler e escrever se tornam mais tênues. Apesar de o leitor ter a possibilidade de escolher uma trilha para navegar, de uma certa forma, ele continua lendo o texto linearmente, porque a estrutura foi proposta pelo autor, ou seja, as escolhas das trilhas a serem disponíveis são controladas pelo autor do hipertexto. Assim, o autor, o texto e o leitor têm uma interação a partir da característica intertextual do hipertexto.

O conhecimento do autor sobre o conteúdo, o usuário e o contexto é de suma importância, pois ele poderá pressupor quais as trilhas o usuário possa seguir, fazendo com que ele se sinta como se estivesse lendo

um texto tradicional. Por isso, é importante o autor manter a coerência textual, evidente no texto linear, propondo uma representação por meio de nós e *links* que considera relevante para a representação do conteúdo do texto.

Os processos de indexação podem subsidiar a conversão do texto em hipertexto para disponibilizá-lo na Web, porém exigem adaptação. Nesse contexto, o processo de indexação, especificamente a primeira etapa, a *análise de assunto*, pode ser apontado como uma solução para auxiliar nessa estruturação. A leitura documentária, realizada pelo autor do hipertexto durante estes três estágios, possui um papel interativo e mediador entre as ideias do autor, o usuário e o contexto. Ela irá auxiliar na compreensão do texto, na identificação dos módulos a serem representados e na seleção dos conceitos/*links*, contruindo uma estrutura semântica hipertextual com coerência e coesão.

O uso de método estruturado em conceitos, em vez de termos/palavras que são extraídos dos documentos, para representação dos *links* que farão a ligação entre os nós na estrutura hipertextual, possibilita criar uma rede semântica que considera o conteúdo do documento, o contexto e o usuário do hipertexto, evitando problemas semânticos, tais como a polissemia e a sinonímia. A construção de uma rede semântica coerente e o uso de vocabulário controlado podem tornar mais eficiente a recuperação na Web, assegurando que a informação relevante não se perca. Nesse contexto, na primeira etapa da indexação, os procedimentos da *análise de assunto* realizados no momento da leitura documentária se mostram de suma importância, a ser considerada na estruturação de hipertextos. Os *links*, como os conceitos, são unidades de informação, que devem representar o conteúdo do nó, dando sentido na trilha escolhida pelo usuário, formando uma rede de conteúdo que garante a compreensão global do texto. O ideal seria que a organização semântica do texto, que pode ser denominada de estrutura conceitual, refletisse as relações semânticas entre os assuntos que o compõem.

O autor do hipertexto, que, muitas vezes, cria estas estruturas intuitivamente, poderá, a partir dessas reflexões, utilizar as abordagens linguística, cognitiva e lógica, inerentes à leitura documentária para auxiliá-lo na conversão de textos em hipertextos. Espera-se que este estudo

sirva como ponto de partida para futuras reflexões da perspectiva da Biblioteconomia e Ciência da Informação.

## REFERÊNCIAS

- ARAÚJO, M. A. *Compreensão de hipertexto: sob a perspectiva da teoria da mesclagem*. 2006. Dissertação (Mestrado em Estudos Linguísticos) - Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 12676*. Métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.
- BARTHES, R. *From work to text in textual strategies: perspectives in post-structuralist criticism*. Ithaca: Cornell University Press, 1979.
- BLAIR, D. C. *Language and representation in information retrieval*. Amsterdam; New York: Elsevier Science, 1990.
- BOLTER, J. D. *Writing space: the computer, hypertext, and the history of writing*. Hillsdal: Lawrence Erlbaum, 1991.
- CESARINO, M. A. N.; PINTO, M. C. M. F. Análise de assunto. *Revista de Biblioteconomia de Brasília*, Brasília, v. 8, n. 1, p. 32-43, jan./jun. 1980.
- CHAUMIER, J. Indexação: conceito, etapas, instrumentos. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.
- CHU, C. M.; O'BRIEN, A. Subject analysis: the critical first stage in indexing. *Journal of Information Science*, London, v. 19, n. 6, p. 439-54, 1993.
- CONKLIN, J. Hypertext: an introduction and survey. *Computer*, New York, v. 20, n. 9, p. 17-41, 1987.
- DAHLBERG, I. A referent-oriented, analytical concept theory of Interconcept. *International Classification*, Baden-Baden, v. 5, n. 3, p. 122-151, 1978.
- DERRIDA, J. Signature event context. In: MARIN, S.; SUSSMAN, H. *Glyph*. Baltimore, Maryland: Johns Hopkins University, 1977. p. 172-197.
- DIAS, E. W.; NAVES, Madalena M. L. *Análise de assunto: teoria e prática*. Brasília: Thesaurus, 2007.
- ELGELBART, D. *Augmenting human intellect: a conceptual framework*. Chapel Hill: Ibiblio, [1962]. Disponível em: <http://www.ibiblio.org/pioneers/englebart.html>. Acesso em: 07 fev. 2019.
- ENSSLIN, A. *Canonizing Hypertext: explorations and constructions*. London: Continuum, 2007.

- FACHINETTO, E. A. O hipertexto e as práticas de leitura. *Letra Magna: Revista Eletrônica de Divulgação Científica em Língua Portuguesa, Linguística e Literatura*, Brasil, ano 02, n. 3, 2. sem. 2005.
- FARROW, J.F. A cognitive process model of document indexing. *Journal of Documentation*, London, v. 47, n. 2, p. 149-166, 1991.
- FOSKET, A. C. *A abordagem temática da informação*. Tradução de Agenor de Briquet de Lemos. São Paulo: Polígono, 1973.
- FOUCAULT, M. What is an author? In: RABINOW, P. (ed.). *The Foucault Reader*. London: Penguin Books, 1979. p. 101-120.
- FREI, H. P.; STIEGER, D. The use of semantic links in hypertext information retrieval. *Information Processing & Management*, Tarrytown, v. 31, n. 1, p. 1-13, 1995.
- FROHMANN, B. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, London, v. 46, n. 2, p. 81-101, June 1990.
- FUGMANN, R. *Subject analysis and indexing: theoretical foundation and practical advice*. Frankfurt/Main: Indeks Verlag, 1993.
- FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003.
- GIASSON, J. *A compreensão na leitura*. Lisboa: Asa, 1993.
- GOMES, L. F. *Hipertexto no cotidiano escolar*. São Paulo: Cortez, 2011.
- HEINONEN, O. Optimal multi-paragraph text segmentation by dynamic programming. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 36., INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 17., 1998, Montreal. *Proceedings* [...]. Québec: L'Université de Montréal, 1998. Disponível em: <http://citeseer.ist.psu.edu/heinonen98optimal.html>. Acesso em: 13 mar. 2019.
- HJORLAND, Birger. The concept of "subject" in information science. *Journal of Documentation*, London, v. 48, n. 2, p. 172-200, June 1992.
- HUTCHINS, W. J. The concept of "aboutness" in subject indexing. *Aslib Proceedings*, Bingley, v. 30, n. 5, p. 172-181, May 1978.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Documentation: methods for examining documents, determining their subjects, and selecting indexing terms*. Suíça: ISO, 1985. (ISO 5963-1985 (E)).
- KOBASHI, N. Y. *A elaboração de informações documentárias: em busca de uma metodologia*. 1994. 195 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1994.



- KOCH, I. G. V. *Desvendando os segredos do texto*. São Paulo: Cortez, 2002.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 2004.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos, 1993.
- LANDOW, G. P. *Hypertext 3.0: critical and new media in a era of globalization*. Baltimore: The Johns Hopkins University Press, 2006.
- LANGRIDGE, D. W. *Subject analysis: principles and procedures*. London: Bowker-Saur, 1989.
- LIMA, G. Â. B. O. *Mapa hipertextual (MHTX) um modelo para a organização hipertextual de documentos*. 2004. 199 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.
- MAI, J. E. Semiotics and indexing: an analysis of the subject indexing process. *Journal of documentation*, London, v. 57, n. 5, p. 591-622, 2001.
- MARQUES, E. V. Introdução aos sistemas de hipertexto. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v. 24, n. 1, p. 85-111, jan./jun. 1995.
- McGARRY, K. *O contexto dinâmico da informação: uma análise introdutória*. Brasília: Briquet de Lemos, 1999.
- MURRAY, J. H. *Hamlet on the Holodeck: the future of narrative in cyberspace*. New York: Free Press, 1997.
- NAVES, M. M. L. *Curso de indexação: princípios e técnicas de indexação, com vistas à recuperação da informação*. Belo Horizonte: UFMG; Biblioteca Universitária, 2004. Material didático.
- NELSON, T. The hypertext. In: SMITH, J. B.; HALASZ, F.; YANKLOVICH, N. (ed.). *Proceedings International Documentathm Federation: Hypertext '87 Proceedings*. New York: Association for Computing Machinery Presses, 1965. p. vii.
- NONATO, R. S. *Teoria do conceito e hipertextos: uma proposta para determinação de relacionamentos em links conceituais*. 2009. 122 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2009. Disponível em: [http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-7V2GYA/dissertacao\\_nonato\\_2009.pdf?sequence=1](http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-7V2GYA/dissertacao_nonato_2009.pdf?sequence=1). Acesso em: 16 mai. 2016.
- OTLET, P. *Traité de documentation: le livre sur le livre: théorie et pratique*. Bruxelles: Mundaneum, 1934.

- PETERSEN, T. Introduction. In: PETERSEN, T.; BARNETT, P. J. (ed.). *Guide to indexing and cataloging with the Arts and Architecture Thesaurus*. New York: Oxford University Press, 1994. p. xiii-xvi.
- RADA, R. Focus on links: a holistic view of hypertext. *International Classification*, Frankfurt, v. 1, n. 18, p. 13-18, 1991a.
- RADA, R. Trends in... A critical review. *Information Processing & Management*, Tarrytown, v. 27, n. 6, p. 659-617, 1991b.
- ROBREDO, J. *Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas*. 4. ed. Brasília: Editora do Autor, 2005.
- SILVA, M. R; FUJITA, M. S. L. A prática da indexação: análise da evolução de tendências teóricas e metodológicas. *Transinformação*, Campinas, v. 16, n. 2, p. 133-161, mai./ago. 2004.
- SNYDER, I. *Hypertext: the electronic labyrinth*. New York: New York University Press, 1997.
- TAYLOR, A. G. *The organization of information*. 2. ed. Westport, Connecticut: Libraries Unlimited, 2004.
- TESSIER, J. A. Hypertext linking as a model of expert indexing. *Advances in Classification Research*, Medford, p. 1991. Trabalho apresentado no 2º ASIS SIG/CR Classification Research Workshop, 1991. Disponível em: <https://journals.lib.washington.edu/index.php/acro/article/view/12555/11068>. Acesso em: 13 mar. 2019.
- UNISIST. Princípios de indexação. Tradução de Maria Cristina M. F. Pinto. *Revista da Escola de Biblioteconomia*, Belo Horizonte, v. 1, n. 10, p. 83-94, mar. 1981. Título original: Indexing principles.
- VAN SLYPE, G. Languages de indización: concepción, contrucción y utilización en los sistemas documentales. Tradução de Pedro Hípola e Félix de Moya. Madrid: Fundación Germán Sánchez Ruipérez, 1991.
- VILAN FILHO, J. Hipertexto: visão geral de uma nova tecnologia de informação. *Ciência da Informação*, Brasília, v. 23, n. 3, p. 295-308, set./dez. 1994.
- VILAN FILHO, J. *Sistemas hipertexto para microcomputadores: uma aplicação em informação científica e tecnológica*. 105p. Dissertação (Mestrado em Biblioteconomia e Documentação) - Faculdade de Estudos Sociais Aplicados da UNB, Brasília, 1992.
- WEINBERG, B. H. Improved internet access: guidance from research on indexing and classification. *Bulletin of the American Society for Information Science*, Maryland, v. 25, n. 2, p. 26-29, 1999.
- WORD WIDE WEB CONSORTIUM. *Semantic web*. 2010. Disponível em: <http://www.w3.org/standards/semanticWeb/>. Acesso em: 12 dez. 2019.