

Representação e economia da informação

Carlos Henrique Marcondes

Professor, Departamento de Documentação, Universidade Federal Fluminense
marcondes@ax.apc.org

Resumo

Este trabalho discute o papel da representação na atual economia da informação/conhecimento que se delineia a partir da emergência da Internet. Neste contexto, identificação e avaliação se tornam atividades essenciais para a utilização de recursos informacionais disponíveis na rede. A identificação/avaliação de informações é apresentada como um processo semiótico, no qual a representação substitui a informação original e serve a um usuário como um dispositivo inferencial com relação à sua relevância. O trabalho informacional de desenvolvimento e sistematização de representações das informações disponíveis na Internet joga um papel decisivo, viabilizando processos de identificação, avaliação e uso, atuando como intermediária entre usuários e suas necessidades de informação e recursos informacionais potencialmente relevantes.

Palavras-chave

Representação da informação; Organização do conhecimento; Recuperação da informação; Internet.

Information representation and economy

Abstract

This paper discusses the role of information representation in the present information/knowledge economy outlined by the emergence of Internet. In this context, information discovery and evaluation are essential activities for information/knowledge use and re-use. Information discovery is presented as a semiotic process, where representation substitutes the original information and serves to a user as an inferential device concerning its relevance. Information work of developing and systematizing representations of information available in Internet plays an essential role, allowing processes of information discovery, evaluation and use, acting as an intermediate between users with information needs and potentially relevant resources.

Keywords

Information retrieval; Knowledge organization; Information representation; Internet.

INTRODUÇÃO

“Todo o símbolo implica a morte da coisa por ele representada. Simbolizar é perder, para ganhar.” (Peregrino¹, 1986, p. 17)

Este trabalho discute o papel da representação de informações na atual economia da informação/conhecimento que se configura a partir do surgimento da Internet. Na atualidade, constata-se que a informação se tornou um recurso cada vez mais valorizado como viabilizador de decisões e de processos de conhecimento/inteligência nos mais diferentes campos. Um aspecto problemático da cultura de nosso tempo relacionado à questão informacional é o assim chamado fenômeno da explosão informacional, a grande quantidade de informações produzidas e disponibilizadas por diferentes atividades sociais, dificultando sua identificação, acesso e utilização. Na emergência da sociedade da informação, o valor desta como insumo para qualquer atividade, seja ela uma decisão econômica, um processo cultural ou de ensino/aprendizagem, uma pesquisa científica ou tecnológica, está relacionado diretamente ao seu potencial de orientar de forma econômica o dispêndio de energia para a realização desta atividade. Para que possa realizar todo este potencial, a informação relevante para um dado problema precisa estar disponível no tempo certo. De nada adianta a informação existir, se quem dela necessita não sabe da sua existência, ou se ela não puder ser encontrada.

Representar objetos e conceitos na mente, manipular estas representações antes de agir na realidade de acordo com os resultados da manipulação mental prévia e socializar estas experiências, sob a forma de representações orais ou escritas, para outros membros da espécie, é uma das mais fundamentais características do gênero humano. Constitui o que chamamos de cultura, uma segunda natureza humana, intersubjetiva, capaz de transcender os limites individuais de espaço e tempo.

O trabalho informacional de desenvolvimento e sistematização de representações dos recursos informacionais hoje disponíveis na Internet joga um papel essencial como facilitador da identificação, avaliação e acesso a estes recursos, intermediando um usuário com suas necessidades e informações potencialmente relevantes. *Bibliotecários e indexadores têm produzido e padronizado metadados por séculos (Milstead², 1999).* A representação em atividades de informação, tais como

catalogação, classificação e indexação, é tão comum e trivial entre outras desenvolvidas pelos profissionais de informação que sua importância e valor são frequentemente menosprezados. Esquemas tradicionais de representação como a CDD, a CDU, tesouros, indexação pré e pós-coordenada e mesmo propostas recentes como os metadados, em função de suas origens empíricas e práticas, se ressentem de sólidas bases teóricas com relação a seus aspectos representacionais e os processos cognitivos por eles viabilizados.

Em seus primeiros estágios de uso, a tecnologia da informação era utilizada em um ciclo de identificação, localização e acesso a documentos em papel, armazenados em bibliotecas. Com o surgimento da Internet, os estágios de localização e acesso se tornaram triviais: um número crescente de registros da cultura humana estão migrando para formatos digitais e uma quantidade grande dos mesmos são criados já diretamente em formato digital; estes podem agora ser facilmente acessados, uma vez conhecidas suas URLs.

Em contraste, a enorme quantidade de informação armazenada e disponibilizada via Internet torna crucial o problema da *identificação* de informação relevante. Diversas estratégias para fazer frente à explosão informacional trazida pela Internet podem hoje ser divisadas, como os mecanismos de busca gerais (AltaVista, Excite, Lycos, Infoseek, Yahoo e outros), os localizadores de informações especializados, como o GILS ou o PROSSIGA (www.prossiga.br) ou LIS (www.bireme.br) e as propostas de indexação pelo próprio autor do documento com *tags* embutidas no código das páginas HTML, como o Dublin Core e a linguagem XML (Medeiros, 2000). O desenvolvimento de esquemas de metadados como o Dublin Core (Weibel³, 1995) e sua utilização pelo próprio autor do documento eletrônico para descrevê-lo e indexá-lo é uma possível solução; juntamente com *tags* HTML especiais, as *tags* META, metadados constituem uma possível solução para a explosão informacional trazida pela Internet, tornando possível que páginas Web sejam vasculhadas por programas especiais, chamadas “aranhas” ou agentes inteligentes, indexando-as com muito mais qualidade.

Uma novidade em termos de mecanismos de busca que parece alentadora são os projetos CLEVER⁴ e GOOGLE⁵, com suas propostas de ordenamento e priorização *ranking* dos resultados de uma busca tendo por base *sites* mais referenciados por *link* a partir de outros (Clever⁴, 1999). Em todas estas propostas, representações consistentes das informações jogam um papel fundamental em atividades como busca de informações, avaliação e uso.

Intimamente ligado às atividades de recuperação de informações, principalmente utilizando tecnologia da informação, está o conceito de Relevância (Greisdorf⁶, 2000). Este é um conceito central para a ciência da informação (Saracevic⁷, 1970). Relevância pode ser definida como “*medindo a correspondência entre um documento e uma consulta*” (Le Coadic⁸, 1996). Conforme colocado por Robins⁹ (2000), o conceito de relevância na recuperação de informações evoluiu desde uma perspectiva centrada no sistema, onde um documento, em um sistema de recuperação de informações *é ou não é relevante*, para modelos mais sofisticados, baseados em aspectos cognitivos de um usuário em sua interação com um sistema de informações – os chamados modelos cognitivos ou modelos baseados no usuário (Belkin¹⁰, 1982), (Saracevic¹¹, 1988), (Ingwersen¹², 1996).

Na perspectiva cognitivista, a recuperação de informações está associada com imprevisibilidade (Ingwersen¹², 1996). Esta situação implica, da perspectiva de um usuário, um processo de busca de informações hierarquizado, desde necessidades mais gerais por informações – *information behavior* –, passando pela interação do usuário com fontes e sistemas de recuperação de informações – *information seeking behavior*, *information searching behavior* –, até o uso final da informação pelo usuário – *information use behavior* – (Wilson¹³, 2000, p. 49). Este processo se mostra assim um processo contínuo de tomada de decisões. Wang¹⁴ (1998) propõe um modelo para estas decisões, baseado em critérios como topicalidade, orientação, qualidade, novidade, disponibilidade, autoridade, relação e em valores como epistêmico, funcional e emocional. Aspectos do modelo proposto por Wang relacionados com o problema da representação são seu estudo do impacto do que ele chama DIEs (*Document Information Elements*) – componentes de informação do documento, os campos de uma referência bibliográfica, por exemplo – nas decisões relacionadas à relevância dos documentos recuperados de um sistema de recuperação de informações.

A referência bibliográfica pode ser vista como um texto sobre outro texto – o documento que ele representa. Diversas disciplinas se voltam para o problema da interpretação de textos. A interpretação como atividade cognitiva de um leitor (no caso de sistemas de informação, um usuário), é preocupação da hermenêutica (Demo¹⁵, 1995), (Reale¹⁶, 1991). A teoria literária de base semiótica (Eco¹⁷, 1995), (Violi¹⁸, 1998), a Análise do Discurso (Orlandi¹⁹, 1993) e a ciência da informação (Dervin²⁰, 1999) destacam o papel ativo do leitor, do seu contexto, da sua história, em dar sentido ao texto.

O referencial teórico da semiótica, conforme estabelecida por Charles Sanders Peirce, é usado neste trabalho para elucidar o papel de representações de documentos no contexto de processos de busca de informações: principalmente conceitos como *abdução* – processo inferencial baseado em representações – e o de interpretação (Hobbs²¹, 1993). Peirce associa a abdução ao *insight* criativo na ciência, quando uma nova descoberta científica é inferida, baseado em poucos e incompletos dados. Assim também considera Ingwersen¹² (1996), p. 20) quando diz que “*Em vários aspectos a recuperação de informações tem paralelo com o processo científico de formular uma hipótese, testá-la e obter um novo conhecimento*”. A formulação de Peirce para o raciocínio abduutivo esclarece o papel de representações de informações – uma referência bibliográfica, por exemplo – na economia da informação: servir como um dispositivo de inferência – um artefato cognitivo, na concepção de Hutchins²² (1999) – sobre a relevância do recurso informacional para as necessidades de informação do usuário que a interpreta. No espaço da Internet, por exemplo, representações de documentos ou de outros recursos informacionais devem apoiar decisões de seus usuários potenciais relativas à busca, identificação, avaliação de relevância, custo-benefício de acessar a informação e lê-la.

É discutido também o trabalho informacional que envolve o processo de produção de representações de recursos informacionais. Quando comparado com a leitura, comunicação e interpretação de textos, já bastante discutidos na literatura, as atividades de representação se mostram mais complexas, pois envolvem um terceiro papel além do autor e do leitor, o do profissional da informação. Este interpreta o texto, infere as necessidades de informação de possíveis leitores e prepara um (meta)texto que vai intermediar o acesso, a identificação e a avaliação de relevância de um usuário com relação ao texto original.

Todas estas colocações mostram a complexidade cognitiva envolvida em processos de busca de informações e reforçam a importância de se desenvolver e aperfeiçoar modelos destes processos. A busca de informações por um usuário é vista aqui como um processo contínuo e interativo entre o usuário e fontes e recursos informacionais, onde, por meio de processos de avaliação-decisão sucessivos, o usuário vai descartando conjuntos de recursos informacionais e assim estreitando o espaço de busca, até que eventualmente informações relevantes para suas necessidades sejam encontradas. A questão que o trabalho se propõe a discutir é qual o papel da representação de informações nas etapas do ciclo de

identificação, avaliação, localização e acesso a informações; mais especificamente, qual o papel das representações na etapa de *avaliação* da relevância de um recurso informacional.

O artigo está organizado nos seguintes tópicos: depois desta introdução, a parte 2 discute a representação desde um ponto de vista antropológico, cultural, cognitivo e filosófico. A parte 3 discute o papel da representação na economia da informação/conhecimento que emerge da sociedade da informação e do surgimento da Internet. Então, a parte 4 discute a descoberta de informações como um processo semiótico; vários aspectos teóricos da representação, como modelagem, relação com o objeto sendo representado, propósito, modelos semióticos para a atividade de representação, qualidade da representação, aspectos cognitivos suportados pela representação, são discutidos também. Finalmente, são apresentadas conclusões e algumas linhas de pesquisa relacionadas a estas questões.

O PROBLEMA DA REPRESENTAÇÃO – ASPECTOS ANTROPOLÓGICOS, CULTURAIS, COGNITIVOS E FILOSÓFICOS

“*O fato de que o Homem pode manipular signos (símbolos que se referem a fenômenos naturais) na ausência dos próprios fenômenos constitui uma diferença vital nas características dos animais*” (Foskett²³, 1980, p. 24).

A representação de objetos e conceitos na mente, o processamento destas representações antes de agir praticamente no mundo, de acordo com o resultado deste processamento, é uma das características mais fundamentais da humanidade. Rosemberg, citado por Santaella²⁴ (1998, p.17), diz que “*a atividade humana essencial e característica é a representação*”. Esta atividade mental e cognitiva foi mais útil ao homem em sua sobrevivência que seus atributos de força física. No amanhecer da humanidade, elas jogaram um papel fundamental na sobrevivência do homem.

A representação de experiências da vida diária e seu processamento mental permitem ao homem transcender as limitações de tempo e espaço. Ela podia fazer muito mais, acreditava o homem primitivo: as impressionantes pinturas de caçadas, com animais feridos, nas cavernas paleolíticas de Altamira na Espanha e Lascaux na França (Mauduit²⁵, 1964) e em muitos outros lugares, diante das quais eram celebradas cerimônias antecedendo as caçadas, como forma de trazer boa sorte, testemunham a crença do homem primitivo no poder da representação, até para influenciar os resultados da caçada.

A expressão de representações mentais através de uma crescentemente sofisticada linguagem falada trouxe um mecanismo para comunicação intersubjetiva a transferência social de conhecimentos e experiências entre os membros do grupo e também entre gerações. Assim é a cultura, uma natureza adicional para a humanidade, que transcende e é independente de qualquer indivíduo. Mais que isso, a cultura humana é incremental e tem sido transmitida e ampliada de geração em geração.

Etimologicamente, representar significa *fazer alguma coisa presente novamente*. A semiótica, conforme formulada por Pierce, descreve a representação como um processo envolvendo um objeto, alguma coisa que o *representa* e o efeito da representação, na ausência do objeto, na mente de um usuário. Representação é, desta maneira, um processo ocorrendo na mente de alguém, produzindo nesta mente algo distinto do objeto a que se refere. A representação então relaciona o objeto que ela representa com a mente que o percebe.

A representação é parte de um complexo processo cognitivo. Pierce afirma que “*a lógica é a ciência das leis necessárias gerais dos signos, especialmente dos símbolos*” (Pierce²⁶, 1977, p. 29) e, ainda, que “todo o pensamento, portanto, deve necessariamente, estar nos signos” (Pierce, 1977, p. 253). Este processo cognitivo, em seu primeiro estágio, envolve um confronto com a realidade, aprendizado a partir desta experiência, conceitualização e abstração. Estes processos são atalhos ou auxílios para operar com a *complexidade* da realidade. A vida real é uma realidade demasiadamente vasta e bastante complexa para ser captada em sua totalidade pela mente humana. Aspectos desta realidade, para serem armazenados na mente de alguém para uso futuro, devem ser filtrados, sintetizados. A literatura oferece um *insight* notável para este processo: Epstein²⁷ (1986, p. 10), citando o famoso escritor argentino Jorge Luis Borges²⁸ e seu personagem Funes, diz que Funes “*lembrava cada folha de cada árvore e também de cada momento que ele a tinha percebido ou imaginado na sua vida. Mas Funes era incapaz de pensar. Pensar é esquecer diferenças, é generalizar, é abstrair*”.

O mesmo processo cognitivo é amplamente desenvolvido pela ciência. A construção de modelos é uma atividade fundamental em ciência; modelos permitem simular um fenômeno e fazer inferências baseadas neles (Dupuy²⁹, 1996); são abstrações de algum aspecto da realidade que apresentam características homomórficas com respeito a esta realidade que eles representam (Epstein²⁷, 1986).

O processo de conceitualização consiste em se abstraírem aspectos particulares das características de um objeto e reter uma representação mental do mesmo, contendo somente aspectos genéricos. Escolhas, consciente ou não, devem ser feitas acerca de que características são relevantes e devem ser retidas e de quais não devem, produzindo uma representação mental do mesmo. O processo continua com a memorização desta representação e com a capacidade de evocar experiências passadas, relacioná-las com a situação atual e agir de acordo.

Qual seria a vantagem (cognitiva) de um processo em que alguma coisa faz com que outra coisa ausente torne-se presente na mente de alguém? De um ponto de vista cognitivo, trazer objetos ausentes como representação permite *processá-los na mente*. Esta é uma característica excepcional, uma vez que o processamento de representações mentais, relacionando-as com um acervo de experiências e conhecimento de alguém, permite antever e *inferir* conseqüências de processamento de uma maneira econômica e agir na realidade de uma maneira mais segura, de acordo com estas conclusões.

O LUGAR DO TRABALHO INFORMACIONAL NA ECONOMIA DA INFORMAÇÃO E DO CONHECIMENTO

Localizar/identificar informações é uma etapa fundamental na economia da informação e do conhecimento. Algumas estratégias podem ser divisadas para lidar com o problema da localização/identificação de informações no contexto da Internet. Em primeiro lugar, os mecanismos de busca, como Lycos, AltaVista, Infoseek, HotBot etc. Estes mecanismos de busca mantêm uma base de dados contendo URLs de páginas Internet juntamente com palavras-chave extraídas destas páginas automaticamente. A base de dados é mantida por programas especiais chamados aranhas (*spiders*) ou robôs, que, ao gerarem automaticamente valores numéricos de URLs*, acessam diferentes páginas Web, vasculham seu conteúdo, extraindo informações do texto de cada página HTML visitada e armazenando na base de dados. Regularmente, programas-aranha vasculham a Internet ou partes selecionadas dela, mantendo bases de dados do mecanismo de busca atualizadas.

* Universal Resource Locator – Localizador Universal de Recursos, o endereço Internet de um recurso.

Sneiderman³⁰ (1997), discutindo a eficácia dos mecanismos de busca de uso geral disponíveis na Internet, diz que

“Embora mecanismos de busca como as Infoseek, AltaVista, Lycos, WebCrawler e Open Text sejam largamente usados, existe um consenso público e geral e entre profissionais acerca das grandes dificuldades de buscar informações”.

Um relatório do instituto de Pesquisas NEC, publicado pelo boletim *Edupage* em português, em 3 de abril de 1998, relata que os melhores mecanismos de busca cobrem não mais de 30% de todas as páginas *Web*. Esta afirmação reforça a dimensão do problema localização/identificação colocado pela Internet.

Outra estratégia é o desenvolvimento de serviços de localizadores de informações especializados como os serviços de localizadores de informações governamentais (GILS) de diferentes países, os portais especializados como o SOSIG (www.sosig.ac.uk), a experiência das bibliotecas virtuais especializadas do programa Prossiga, o serviço LIS – Localizadores em Ciência da Saúde. A catalogação e indexação de recursos Internet nestes serviços é feita por profissionais de informação; eles indexam recursos informacionais e não páginas isoladas, como é feito automaticamente pelos mecanismos de buscas (Kirriemuir³¹, 1998). Apesar do fato de que a qualidade da informação é boa e os localizadores de informações especializados são realmente eficazes, eles se destinam a comunidades específicas.

Uma terceira estratégia consiste no uso descentralizado de esquemas de metadados por diferentes comunidades para descrever e indexar recursos de informação Internet à medida que eles são criados/publicados. Metadados embutidos no texto HTML de páginas podem tornar mais fácil e correto o exame e indexação das páginas *Web* pelas “aranhas” e robôs, favorecendo a localização/identificação de informações relevantes.

Todas estas experiências destacam o papel dos metadados, isto é, de representações de recursos informacionais disponíveis na *Web*, na localização/identificação de informação/conhecimento que emerge com o surgimento da Internet e da sociedade da informação. No trabalho informacional de criação de representações, atores e papéis são bastante distintos dos de processos cognitivos individuais e devem ser identificados de maneira precisa. Primeiro, um ator, o profissional de informação, representa um documento mediante estratégias cognitivas de indexação e resumo (Endres-Niggemeyer³², 1998). Ele

interpreta o documento para alguém mais, o usuário, com o objetivo de desenvolver uma representação do documento. Esta representação é preparada para ser utilizada por outro ator, o usuário. Este é um processo ativo, à medida que o profissional de informação não somente extrai a informação contida no documento original, mas, freqüentemente, *adiciona* informações novas baseado nas necessidades de informação e no quadro conceitual que o profissional de informação supõe que usuários possuam ou que ele próprio possua. Ao construir um registro bibliográfico para inclusão em uma base de dados, por exemplo, o profissional de informação assinala campos como *tipo de documento, classificação* etc., informação que não está explícita no documento original.

REPRESENTAÇÃO E A ECONOMIA DO CONHECIMENTO

Processos semióticos, como estabelecidos por Eco¹⁷ (1995), diferentemente de processos determinísticos de estímulo-resposta, que são binários (A implica B), são *triádicos*, envolvendo um objeto A ausente, um signo B e uma entre várias possibilidades de *interpretação* C que signo traz ou representa, na mente de alguém. Colocada desta maneira, a semiótica é um processo humano e social fundamental, que implica a comunicação entre seres humanos e, desta forma, em decisões e escolha de interpretações de mensagens.

Dois aspectos devem ser destacados, como fazem vários autores, nas tarefas cognitivas de interpretação e compreensão de mensagens: *a importância da informação contextual e o papel ativo desempenhado pelo receptor ou leitor da mensagem*. A escolha entre diferentes opções da interpretação de um signo depende fortemente do contexto do processo de comunicação, de conhecimento implícito, do quadro conceitual prévio do receptor ou leitor, do cenário do diálogo, das expectativas do receptor, do seu nível cultural etc. Estes fatores favorecem algumas interpretações em relação a outras, ou fazem com que algumas interpretações tenham mais sentido que outras. O receptor tem um papel ativo na interpretação de uma mensagem: Eco afirma que qualquer texto pode ser visto como um mero esquema indicativo, solicitando do leitor que preencha os vazios. Eco, citado por Violi¹⁸, diz que *“a text can be seen as a lazy machine asking the reader to fill in a whole series of gaps”*. Pierre Lévy também destaca o papel criativo da interpretação:

“Toda criação é equivalente ao uso, de uma maneira original, de elementos preexistentes. Toda criação é uso, na medida que traz novas possibilidades, como no plano

da criação... Criação e uso são, realmente, dimensões complementares das mesmas operações de conexão, com seus efeitos de reinterpretação e construção de novos significados “(Lévy,³³ 1993, p. 58).

A busca de informações, principalmente no contexto de uma busca na Internet ou em uma base de dados, pode ser vista como um processo semiótico, uma vez que envolve o uso e interpretação de representações, as referências bibliográficas. Discutindo este processo, Belkin diz que, em situações de busca de informações, o usuário se encontra em um “estado anômalo de conhecimento” que deve ser preenchido com informação/conhecimento procurado para se tornar coerente. Esta situação é a que cria motivação para a busca de informações e fornece os parâmetros para qualquer avaliação de relevância de representações como referências bibliográficas.

Quando interpretada pelo usuário, a representação deve suportar/apoiar vários tipos de operações cognitivas de inferência acerca do conteúdo do documento original, de modo a permitir que o usuário, no limite, *possa decidir despendendo energia para obter e ler o documento completo*. O valor da representação consiste no fato de que ela pode economizar energia para o usuário, tornando o processo de busca de informações mais econômico de um ponto de vista do seu dispêndio de energia; ler um resumo despende muito menos energia que a leitura do documento completo.

Pierce, discutindo o processo de aquisição de conhecimento, afirma que existem três modos de raciocínio na ciência: *dedução, indução e abdução*. Dedução, diz Pierce, prova que algum aspecto não claramente conhecido de alguma coisa já conhecida é *verdadeiro*; indução prova que alguma coisa é *provavelmente verdadeira*; somente o último, *abdução*, segundo Pierce, é um processo criativo, à medida que somente sugere que alguma coisa pode ser um caso particular de uma regra ou princípio; desta maneira, a abdução permite a criação de um conhecimento novo; baseado em algumas características de um fenômeno, é possível inferir outras ou inferir princípios gerais. Abdução é associada por Pierce com o *insight* criativo em ciência. Permitir inferências é a razão de ser do processo de representação no contexto da busca de informações. Quais são os elementos e os limites deste processo? Quais são as condições de otimizá-lo no trabalho de informação?

Pierce conceitua signo como “*um signo ou representâmen, é aquilo que, sob certo aspecto ou modo, representa algo para alguém*” (Pierce²⁶, 1977, p. 46); representar seria “*estar no*

lugar de, isto é, estar numa tal relação com um outro que, para certos propósitos, é considerado por alguma mente como se fosse este outro” (Pierce²⁶, 1977, p. 61). Na representação de um documento por meio de uma referência bibliográfica, o objeto seria o documento original, o signo seria a referência bibliográfica e o interpretante seria a idéia do documento criada pela referência na mente de um usuário. Este signo seria, mais especificamente, um ícone. Um ícone é um tipo de signo que pode representar um objeto, porque seus atributos têm relações análogas com os atributos do objeto sendo representado. Uma representação é construída colecionando-se atributos ou predicados acerca do objeto, diz Dahlberg³⁴ (1992, p. 66), na teoria do conceito. Predicados acerca de um documento poderiam ser do tipo: *um documento tem um título <TÍTULO>, tem um autor <AUTOR>, diz respeito a um assunto <ASSUNTO> etc.* Assim, uma referência bibliográfica é um ícone com relação ao documento que ela representa. Um ícone é definido como um “*signo que tem alguma semelhança com o objeto representado*” (Coelho Neto³⁵, 1980, p.58). Como o documento original, uma referência é um (meta) texto que aparece ao usuário com o mesmo autor, com o mesmo título que o documento original. Geralmente, a referência indica também a fonte do documento original, palavras-chave associadas ao assunto do documento em questão e um resumo descrevendo seu conteúdo. Todos estes componentes são interpretados pelo usuário e, assim, constituem um dispositivo de inferência ou uma *tecnologia cognitiva* conforme estabelecido por Pierre Lévy, que permite inferir o *aboutness* e a relevância do documento.

Sob outro aspecto, indexação e resumo são processos fortemente baseados em inferências por parte do profissional de informação acerca das necessidades do usuário, suas atividades e sobre a maneira pela qual ele usa informação. Pesquisas recentes têm como foco a modelagem de usuários, suas necessidades expressas durante o processo de busca de informações etc. Indexação e resumo consistem em um processo de construção, pelo profissional de informação, de um metatexto fortemente relacionado com o texto original. Em que consistem estas relações? Na construção deste metatexto, o profissional de informação faz escolhas acerca de que aspectos do texto original devem ser mantidos no metatexto e quais não devem ser mantidas, quais, mesmo que não estejam no texto original, devem ser adicionadas ao metatexto.

Outro conceito semiótico importante para a compreensão do papel de representações de documentos é o conceito de *denotação* e *conotação* (Coelho Netto³⁵, 1980, p. 24). Todo o signo é composto de *significante* (o objeto do signo, por exemplo, o conjunto das letras MESA) e o *significado*

(o interpretante do signo, ou seja, o sentido atribuído normalmente ao objeto MESA). A primeira função de um signo, isto é, o signo como notativo (ou denotativo), é representar, através do significante o conceito contido no significado. Mas muitas vezes, o uso social do signo faz com que (significante + significado) de um signo constituam o significante, para um outro signo, com um outro significado. É o que normalmente é conhecido em literatura como o fenômeno lingüístico da metáfora, um outro significado carregado ou *conotado* pelo signo. No campo da ciência da informação, Langridge³⁶ (1977, p. 23) também chama a atenção para este fenômeno, no estudo das propriedades que definem uma classe.

Denotação e conotação ocorrem também quando um usuário interpreta uma referência bibliográfica. A construção desta também faz uso intenso deste fenômeno. O trabalho de Wang¹⁴ (1998) indica a importância dos DIES (Document information elements) – componentes de informação do documento, os campos de uma referência bibliográfica, para o julgamento de relevância de um usuário, e destaca que os campos mais importantes são Título, Resumo e Autor. O fenômeno da denotação e conotação amplia a compreensão deste processo; o campo Autor, além de simplesmente informar quem é o autor do texto, informação irrelevante para um usuário que não o conheça, torna-se altamente significativo como fator de relevância para um usuário que conheça o autor. Além de sua conotação como o responsável pela autoria do trabalho, o autor denota também toda sua trajetória acadêmica, seus trabalhos anteriores etc.

Se tomarmos alguns campos típicos de uma referência bibliográfica, descritos na norma ISO8777³⁷, constatamos que isto se repete. Vários campos carregam um forte potencial *conotativo* adicional ao seu valor denotativo, dependendo do conhecimento prévio do usuário; além do campos AU - Autor, já comentado, tem-se também:

AF – Afiliação do Autor – denota o renome da instituição a que o autor está filiado; as linhas de pesquisa da instituição; a contribuição da instituição para a ciência etc;

DT – Tipo de documento – denota uma forma e um tipo de conteúdo esperado; são bastante conhecidos por profissionais de informação a importância, o uso e manuseio de diferentes fontes de informação (Campelo³⁸, 2000); em termos de ciclo de comunicação científica, um livro pressupõe um conhecimento já aceito, enquanto um *pre-print*, ou uma tese pressupõe um conhecimento mais recente, ainda objeto de discussão e polêmica; um trabalho

em congresso e um artigo de periódico colocam-se em níveis intermediários de consolidação do conhecimento comunicado;

PD – Data de publicação – indica a atualidade da informação contida no documento;

JT – Título do periódico – o título do periódico também indica o renome do mesmo, seu destaque na área do conhecimento, trabalhos anteriores aí publicados, a respeitabilidade do seu corpo editorial.

Os fenômenos de denotação e conotação destacam a importância do usuário, do seu quadro de referência, na implementação de inferências abduativas acerca do potencial de relevância suprido pela representação de um documento.

Quais os limites deste processo? A representação mais completa do ponto de vista informacional de um texto é o próprio texto. Diferentemente, no entanto, do processo semiótico, a construção deste metatexto, a referência bibliográfica, envolve um aspecto que é a raiz de sua utilidade para a economia da informação/conhecimento. A representação deve situar-se entre dois extremos para economizar energia e assim realizar seu papel: ser suficientemente rica sob o aspecto cognitivo e, ao mesmo tempo, sintética para economizar a energia do usuário de uma maneira significativa. Este compromisso deve ser claramente compreendido pelo profissional de informação. Na economia da informação/conhecimento, a representação deve ser tão rica cognitivamente de modo a permitir a um usuário inferir o conteúdo do documento, de maneira a permitir-lhe decidir acerca da validade de obter e realizar a leitura do documento completo. Simultaneamente, este processo deve implicar uma significativa economia de energia em comparação com a obtenção e leitura do documento original.

CONCLUSÕES

O crescimento da Internet como meio de disponibilização de informação/conhecimento, sob as mais diferentes formas – documentos, imagens, hiperdocumentos, – reforça ainda mais a importância da representação para suportar atividades de descoberta/identificação de informações, avaliação e uso. Estas atividades são essenciais para que o conhecimento, tornado assim disponível, possa ser acessado e se torne operacional. Nestas atividades, o raciocínio abduutivo é o mecanismo cognitivo utilizado por um usuário na avaliação de relevância de um recurso de informação/conhecimento na Internet.

No espaço da Internet, estas atividades são desenvolvidas não só por usuários humanos – a avaliação de relevância já tão estudada na ciência da informação –, mas também por meios automáticos, por robôs, “aranhas”, programas agentes inteligentes. É necessário, portanto, que descrições/representações de recursos informacionais possam ser intelegíveis não só por seres humanos, mas também por máquinas. É cada vez maior a tendência em publicar documentos na Internet contendo não só seu conteúdo, mas descrições/representações processáveis diretamente por computador.

É nesta direção que se desenvolvem iniciativas como a linguagem XML (Medeiros³⁹, 2000) - Extensible Markup Language, um padrão do W3C. Trata-se de publicar na Internet documentos escritos (e descritos simultaneamente) em uma linguagem onde, em um todo autocontido, o conteúdo do documento e sua descrição compõem um único arquivo, de tipo texto puro – portáteis, portanto sem problemas para qualquer tipo de sistema computador, já que o formato texto puro é lido por todas as plataformas computacionais. A descrição é feita através de *tags* – marcações ou etiquetas, que envolvem o conteúdo do documento. Este “todo autocontido” tem também a possibilidade de fazer referência para dois elementos: o primeiro é a assim chamada *Document Type Description* (DTD), uma descrição do documento mais completa que a contida nas marcações que o acompanham, que informa que marcações e que valores são válidos, permitindo validar um documento; o segundo, chamado *name space*, que indica a semântica dos campos que descrevem o documento, evitando assim qualquer ambigüidade, já que diferentes publicadores podem utilizar campos de mesmo nome para descrever documentos distintos. O documento é assim “descrito na fonte”, desde sua publicação, tornando-se passível de ser processável por computador para as mais diferentes finalidades.

Outra iniciativa importante no sentido de facilitar e tornar mais rápido o acesso à comunicação científica entre pares é o OpenArchive (www.openarchives.org). Esta iniciativa tem origem nos bancos de *pre-prints* eletrônicos. Seu objetivo é criar sistemas acessíveis via Internet, onde documentos eletrônicos possam ser submetidos e imediatamente disponibilizados para os pares. O OpenArchives vem tendo um impacto muito grande sobre os mecanismos tradicionais de comunicação científica, principalmente o periódico eletrônico tradicional, com relação a seu esquema de avaliação por *referees*, a questão do *copyright* e principalmente ao tempo de publicação.

No bojo desta iniciativa, surgem propostas como a *Research Documents Information Format* (ReDIF), de padronização da descrição de documentos de pesquisa eletrônicos à medida que eles forem submetidos ao banco de *pre-prints*, através de um conjunto padronizado de metadados. Pode-se esperar uma convergência e integração cada vez maior destas iniciativas, suportadas por padrões abertos.

Neste contexto, uma possibilidade interessante seria que a representação de um documento de pesquisa, por exemplo, um *paper*, possa incorporar também algum tipo de representação de conhecimento processável por computador, algo parecido com as cláusulas da linguagem PROLOG (TurboPROLOG⁴¹, 1990). Um documento poderia então ter em meio legível por computador, além de seus elementos descritivos, suas premissas, os dados e fatos levantados e suas conclusões. Isso praticamente transformaria um documento em uma base de conhecimentos, que poderia ser combinada com outras e utilizada para a transferência e geração de novos conhecimentos.

À medida que mais e mais documentos forem publicados dentro do padrão XML, contendo implicitamente, além do conteúdo, seus elementos de representação processáveis por máquina, as atividades de descoberta/identificação de informações, avaliação e uso se tornarão muito mais fáceis.

A Internet se parece cada vez mais com o sonho de Otlet⁴² (1934) de uma biblioteca universal. Para que ela não evolua no sentido de um caos informacional, tem de ser trabalhada cooperativamente pelos profissionais de informação de todo o mundo. Conceitos como ecologia da informação se aplicam a este espaço gigantesco, diversificado, compartilhado e inter-relacionado: qualquer nova informação disponibilizada na *Web* afeta todas as demais. Geri-la eficientemente é dever de todos.

O autor agradece as contribuições valiosas das professoras Lena Vânia Ribeiro Pinheiro e Rosali Fernandez de Souza, do DEP-IBICT/RJ

REFERÊNCIAS BIBLIOGRÁFICAS

1. PEREGRINO, Hélio. Édipo e a paixão. In: Os sentidos da paixão. Rio de Janeiro, São Paulo : FUNARTE, Ministério da Cultura, 1986. P. 17. (Ciclo de debates)
2. MILSTED, Jessica, FELDMAN, Susan. Metadata: cataloging by any other name. ONLINE, jan. 1999. (<http://www.online.com/onlinemag/OI1999/milstead1.html>, visitado em 28/08/99).
3. WEIBEL, Stuart. Metadata: the foundations of resource description. D-Lib Magazine, July 1995. (<http://www.dlib.org/dlib/July95/O7weibel.html>).
4. CLEVER project. Hypersearching the Web. Scientific American, n. 6, 1999. <http://www.sciam.com/1999/0699issue/0699raghavan.html>, consultado em 31/05/99.
5. GOOGLE: como usar o Google. http://www.google.com/mu/pt/why_use.htm
6. GREISDORF, Howard. Relevance: an interdisciplinary and information science perspective. Informing Science, v. 3, n. 2, 2000. (Special issue on Information Science Research).
7. SARACEVIC. T. Introduction to Information Science. New York : Bowker, 1970.
8. LE COADIC, Yves-François. A Ciência da Informação. Brasília : Briquet de Lemos/Livros, 1996. 119 p.
9. ROBINS, David. Interactive information retrieval: context and basic notions. Informing Science, v. 3, n. 2, 2000. (Special issue on Information Science Research).
10. BELKIN, N. J., ODDY, R. N., BROOKS, H. M. ASK for information retrieval. Journal Documentation, 38, n. 2, p.61-71, Part I, n.3, p.145-65, 1982. Part II.
11. SARACEVIC, T., Kantor, P. A study of information seeking and retrieving. Journal of American Society for Information Science, v.3 9, n.3, part I, p.161-176, part II, p.177-196, part III, p.197-216, may 1988.
12. INGENSEN, Peter. Cognitive perspectives of information science retrieval interaction: elements of a cognitive IR theory. Journal of Documentation, v. 52, n. 1, p. 3-50, mar. 1996.
13. WILSON, T. D. Human information behavior. Informing science, v.3, n.2, p.49-55, 2000.
14. WANG, Peiling, SORDEL, Dagobert. A cognitive model of document use during a research projet. Study I. Document selection. Journal of American Society for Information Science, v.49, n. 2, p.115-135, 1998.
15. DEMO, Pedro. Metodologia científica em ciências sociais. São Paulo : Atlas, 1995. 293 p.
16. REALE, Giovanni, ALTISSERI, Dario. História da filosofia: do Romantismo aos nossos dias. São Paulo : Paulus, 1991. (Coleção Filosofia).
17. ECO, Umberto. Os limites da interpretação. São Paulo : Perspectiva, 1995. 315 p. (Coleção Estudos).
18. VIOLI, Patrícia. Semiotics and cognition. (http://mitpress.mit.edu/MITECS/work/violi_r.html, visitado em 15/11/1998).
19. ORLANDI, Eni P. O Integível, o interpretável e o compreensível. In: O Orlandi, Eni P. Discurso e leitura. Campinas : Ed. UNICAMP, 1993. P.101-119.
20. DERVIN, Brenda. Chaos, order and sense-making. 1999. <http://www.mcc.ufc.br/etagi/projetobb/zendervinmit99.html>, consultado em 18/05/2000.
21. HOBBS, Jerry R., STICKEL, Mark E. Interpretation as abduction. Artificial Intelligence, v. 63, p. 69-142, 1993.
22. HUTCHINS, Edward. Cognitive artifacts. (http://www.mitpress.mit.edu/MITECS/work/htchins_r.html, visitado em 12/02/1999).
23. FOSKETT, D. J. Informática. In: Ciência da informação ou informática? Rio de Janeiro : Calunga, 1980. p. 9-51.
24. SANTAELLA, Lucia, NOTH, Winfried. Imagem, cognição, semiótica, mídia. São Paulo : Iluminuras, 1998. 224 p.
25. MAUDUIT, J. A. Quarenta mil anos de arte moderna. Belo Horizonte : Itatiaia, 1964. 271 p.
26. PIERCE, Charles S. Semiótica. São Paulo : Perspectiva, 1977. 337 p.
27. EPSTEIN, Isaac. Cibernética. São Paulo : Ed. Ática, 1986. 88 p.
28. BORGES, Jorge Luis. Obras Completas. Buenos Ayres : Emecé, 1969.
29. DUPUY, Jean-Pierre. Nas origens das ciências cognitivas. São Paulo : Ed. UNESP, 1996. 228 p. Tradução de "Aux origines des sciences cognitives".
30. SHNEIDERMAN, Ben. Clarifying search: a user-interface framework for text searches. Dlib Magazine, January 1997. (<http://www.dlib.org/dlib/january97/retrieval/01sneiderman.html>).
31. KIRRIEMUIR, John, BRICKLEY, Dan, WELSH, Susan, KNIGHT, Jon, HAMILTON, Martin. Cross-searching subject gateways: the query routine and forward knowledge approach. D-Lib Magazine, January, 1998 (<http://www.dlib.org/january98/01kirriemuir1/html>)
32. ENDRES-NIGGEMEYER, Brigitte, NEUGEBAUER, Elisabeth. Professional Summarizing: no cognitive simulation without observation. Journal of American Society for Information Science, v. 49, n. 6, p. 486-506, 1998.
33. LÉVY, Pierre. As tecnologias da inteligência: o futuro do pensamento na era da informática. Rio de Janeiro, Ed. 34, 1993. (Coleção Trans). 205 p.
34. DAHLBERG, Ingeltraut. Knowledge, organization and terminology: philosophical and linguistic bases. International Classification, v. 19, n. 2. p. 65-71. 1992.
35. COELHO Netto, J. Teixeira. Semiótica, informação e comunicação. São Paulo : Ed. Perspectiva, 1980. 222 p. (Coleção Debates).
36. LANGRIDGE, Derek. Classificação: uma abordagem para estudantes de biblioteconomia. Rio de Janeiro : Interciência, 1977.
37. ISO8777 – Commands for interactive text searching. In: European Commission's Open Information Interchange – OII standards and specification list – Library information Interchange Standards – <http://158.169.50.95:10080/oii/en/library.html>, visitado em 15/09/2000).
38. CAMPELLO, Bernardete Santos, CENDON, Beatriz Valadares, KREMER, Jeannette Marguerite, org. Fontes de informação para pesquisadores e profissionais. Belo Horizonte, Ed. UFMG, 2000. 319 p. (Coleção Aprender).

39. MEDEIROS, Norm. XML and the resource description framework: the great web hope. ONLINE, set. 2000. (<http://www.onlinemc.com/onlinemag/OL2000/medeiros89.html>, visitado em 09/01/2001).
40. REDIF: www.openarchives.org. (visitado em 01/10/2000).
41. TURBOPROLOG: the natural language of artificial intelligence. California, EUA : Borland International, 1990.
42. OUTLET, Paul. Traité de documentation; le livre sur le livre, theorie et pratique. Bruselles, Belgiun : Ed. Mundaneun, 1934.
- CHEN, Heisung, HOUSTON, Andreas L., SEWELL, Robin R.,
- SCHATZ, Bruce. Internet browsing and searching: user evaluations of category maps and concept space techniques. Journ. American Society for Information Science, v. 49, n. 7, p. 582-603, 1998.
- LANCASTER, F. W. Indexação e resumos: teoria e prática. Brasília : Briquet de Lemos, 1993. 347 p. Translation of Indexing and abstraction in theory and practice.
- SPERBER, Dan, Hirschfield, Lawrence. Introduction: culture, cognition and evolution. 1998. (http://mitpress.mit.edu/MITECS/cultureintro_r.html, visitado em 13/11/1998).