

XVI ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO

Declaração de Direito Autoral

Autores que submetem a esta conferência concordam com os seguintes termos:

a) Autores mantêm os direitos autorais sobre o trabalho, permitindo à conferência colocá-lo sob uma licença Licença Creative Commons Attribution, que permite livremente a outros acessar, usar e compartilhar o trabalho com o crédito de autoria e apresentação inicial nesta conferência.

b) Autores podem abrir mão dos termos da licença CC e definir contratos adicionais para a distribuição não-exclusiva e subsequente publicação deste trabalho (ex.: publicar uma versão atualizada em um periódico, publicar e compartilhar disponibilizar em repositório institucional, ou publicá-lo em livro), com o crédito de autoria e apresentação inicial nesta conferência.

c) Além disso, autores são incentivados a seus trabalhos online (ex.: em repositório institucional ou em sua página pessoal) a qualquer momento antes e depois da conferência.

FONTE:

<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/paper/viewFile/2939/1015>. Acesso em: 22 nov. 2015.

REFERÊNCIA:

BRITO, Marcilio de; CARIBÉ, Rita de Cássia do Vale. Princípios da indexação por imagens. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Anais...** João Pessoa: ANCIB, 2015. Disponível em:<<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/paper/viewFile/2939/1015>>. Acesso em: 22 nov. 2015.



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 2 – Organização e Representação do Conhecimento
Comunicação Oral

PRINCÍPIOS DA INDEXAÇÃO POR IMAGENS¹

PRINCIPALS OF INDEXING BY IMAGES

Marcilio de Brito, UnB
mdebrito@unb.br

Rita de Cássia do Vale Caribé, UnB
rita.caribe@gmail.com

Resumo: Apresenta fundamentação teórica pela linguística da semiologia de Saussure e da semiótica de Peirce para uma abordagem metodológica da indexação por imagens. A proposta explica os mecanismos da indexação tradicional e suas relações com as palavras da língua, do léxico, do discurso e em terminologia abrindo o espaço para a proposta de elaboração de signos imagéticos para representar o conteúdo dos documentos. O processo de indexação é analisado sob o ponto de vista do signo-palavra e do signo-imagem estabelecendo-se a ponte entre os dois universos. Utiliza-se da lógica peirceana sobre a descrição da imagem na tríade ícone, índice e símbolo para a elaboração de uma nova mensagem. Nesse contexto de compreensão constroem-se os exemplos de representação dos documentos por imagens-chave, apoiando-se nos estudos de Jacques Bertin, sobre a construção intencional de imagens. Estabelece-se o procedimento reverso (construção de imagens) que constitui a base da metodologia de indexação por imagens. O uso da imagem como elo de conexão entre documentos e usuários apresenta-se como um vasto campo de verificação científica em pesquisa como em desenvolvimento.

Palavras-chave: Indexação por imagens. Semiótica. Representação do conhecimento. Representações sintéticas de documentos.

Abstract: Presents the methodology of indexing by images based on Saussure's semiology and Peirce's semiotics theory. So, it is explained how the traditional indexing mechanisms are related to the words of the language, lexicon, discourse and in terminology to represent the documents' contents. Indexing processes are though analysed under the perspective of the sign both as a word and as image to show the links between literal and imagetive universes. Presents the Peirce's logic of image description in its simplified triad form of icon, index and symbol, which is used to elaborate new messages. In that context some examples are built to show the resulting indexing card meaning to establish the reversed procedure

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

of the indexing by images' methodology, based on Jacques Bertin's studies about the principals of intentional image construction. Using images to tie relationships between users and documents is a vast field for scientific research and development.

Keywords: Indexing by images. Semiotic. Knowledge representation. Document surrogates.

1 INTRODUÇÃO

Indexar por imagens consiste em usar imagens em lugar de palavras-chave ou descritores. Se a profundidade do discurso é uma característica da linguagem escrita, difícil de se representar por qualquer outro artifício que não seja a cognição, mais complexo ainda é usar palavras isoladas para representar este conteúdo. Até o momento, representa-se o conteúdo de um texto por elementos do próprio texto, o propósito deste trabalho propõe que a dimensão da imagem seja analisada para ser utilizada para este fim devido às suas características multidimensionais na transmissão da mensagem no processo de comunicação. Trata-se de um salto dimensional nas ferramentas de representação do conteúdo dos documentos, e particularmente da informação registrada.

Em comunicação precedente para ISKO-BRAZIL CONFERENCE 2015 discutiu-se o conceito de catálogo imagético (iOPACS) como instrumento de acesso diferenciado aos acervos das bibliotecas. Esta proposta de acessibilidade, inovadora em seu princípio, está sustentada por uma camada mais profunda de teoria da indexação cujos princípios fundamentais serão discutidos para anunciar as especificidades da proposta de indexação por imagens.

Inicialmente, importa compreender o contexto da imagem na transmissão da mensagem do emissor e na representação do conteúdo dos documentos, meio em que se desenvolvem os alicerces desta proposta. O texto estabelece uma dinâmica de revisitação dos conceitos tradicionais acrescentando, em cada passagem, uma camada de entendimento, conduzindo o leitor ao conceito de “imagem-chave”.

As vantagens sobre o uso de imagens para transmitir e receber informações há muito são objeto de pesquisas em semiologia e semiótica. Na atualidade das tecnologias da informação esses aspectos levantam considerações importantes sobre a mídia informacional (ASHTON, 2015). Saber ler um texto é essencial, mas nem todo mundo lê da mesma forma. A informação registrada não é apenas a informação escrita, a imagem conquistou um espaço nesse processo de comunicação.

Em semiologia aprende-se que a imagem é uma linguagem de comunicação completa. Nossa sociedade é amplamente consumidora de imagens. As leis e códigos de comunicação que

regem a linguagem imagética são, por conseguinte, mais compreensíveis para um número mais significativo de indivíduos do que os tradicionais códigos escritos.

A leitura de imagens supera a *performance* da mensagem escrita com maior precisão e por um público mais amplo, evidenciados pelos comportamentos individuais e sociais:

- O cérebro consome 1/10 segundo para perceber uma imagem de paisagem em oposição aos 60 segundos necessários à leitura da descrição dessa mesma cena (200-250 palavras).

- Imagine-se viajando a 80 Km/h e lendo o texto a seguir,

“Adverte-se ao condutor do veículo da existência adiante, de área sujeita a desmoronamento ocasionado por instabilidade no talude ou por obras”, em lugar desta imagem



- A inclusão de vídeo em mensagens no Twitter aumenta os *Retweets* de 28%, e *tweets* que incluem fotos atingem 35% mais *Retweets* (ROGERS, 2014);

- Um estudo publicado por Social Bakers (2014) mostrou que fotos representam 87% dos *posts* compartilhados em 30.000 páginas, enquanto qualquer outro tipo de conteúdo não atinge mais do que 4% dos compartilhamentos.

2 O CENÁRIO DO PROCESSO DE COMUNICAÇÃO

Com o aperfeiçoamento das técnicas geométricas aplicadas à arquitetura do movimento renascentista de Florença, no século XV, o desenho em perspectiva tomou forma e dimensão na representação dos objetos e em sua tridimensionalidade. A imagem se aproxima dos objetos da realidade identificando-os não somente pela aparência, mas também por sua distribuição e localização espacial. Comparativamente, enquanto a imagem mostra uma realidade percebida visualmente permitindo múltiplas inferências (interpretação do contexto), a informação escrita limita sua percepção visual ao reconhecimento dos sinais linguísticos transferindo o restante de suas competências para o nível intelectual, aonde se processa a decodificação dos signos pela competência linguística do decodificador e pela habilidade literária do autor da mensagem em descrever sua realidade percebida ou imaginária. A primeira operação envolvendo a imagem põe em funcionamento um mecanismo intuitivo de comparação visual e de verificação desta correspondência. A lógica põe-se da seguinte forma: se o que é visto na realidade também se encontra na imagem, então a proposição objeto-representação visual é verdadeira. Em prática de recuperação de informações, isto equivale à operação mental - esse documento contém o que estou procurando. O observador reconhece na imagem os objetos que procura no contexto

adequado. Em condições ideais, o processamento da mensagem seria o mesmo, todavia, no caso da imagem, tudo acontece muito mais rapidamente do que no caso da informação escrita.

Entende-se que o processo de registro codificado do conhecimento ou representação da informação, seja ela imagética ou literária, obedece a princípios diferentes. Quando se produz a informação (codificação) põe-se em aplicação determinadas regras que diferem daquelas da leitura-interpretação (decodificação). A gramática de produção é distinta daquela de interpretação da mesma informação (BERRENDONNER, 1983). Como se não bastasse esta assimetria de gramáticas, ambos os processos veem-se acrescidos do diferencial de bagagem de conhecimentos (referências) que cada um dos agentes, codificador e decodificador, possui em suas experiências individuais (LOPES, 2005; 2008). Assim, anuncia-se o cenário do processo de comunicação, um falante e um ouvinte, no mesmo contexto do processo produção, tratamento e recuperação da informação documentária. A mensagem é produzida e registrada, em seguida é tratada (indexada) por um intermediário, e por fim procurada, lida e interpretada por um usuário.

De posse desses elementos de argumentação, é possível perceber a complexidade do processo de comunicação: o falante, produtor da informação restringe-se às regras de uma gramática de expressão (linguagem) para registrar sua mensagem, posteriormente um agente externo (indexador), não necessariamente competente na área de conhecimento, interpreta a mensagem (frequentemente por meios de uma leitura transversal) e atribui etiquetas conceituais (descritores/palavras-chave), que um terceiro agente, o ouvinte (público alvo) vai utilizar-se como elemento representativo da informação procurada, para localizar o documento que contém os objetos de sua busca. Um exame racional simples permite deduzir que trata-se ali de um exercício aproximativo de correspondência.

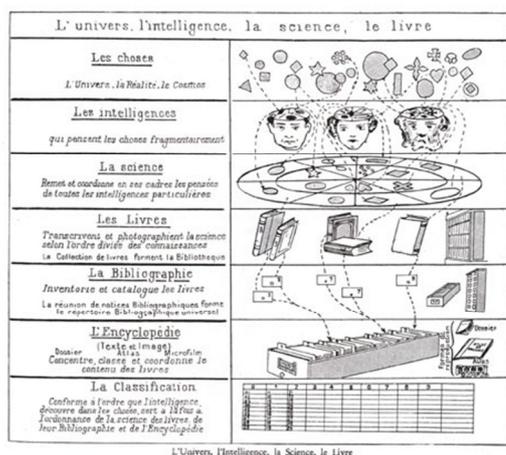
3 INDEXAÇÃO

Indexar ou organizar em índice significa comumente estabelecer uma lista metódica de nomes de autores, de livros consultados, de assuntos etc., de tal forma que cada elemento da lista aponta para uma ocorrência deste elemento no corpo do documento. O elemento, ou forma de entrada na lista, aponta para o objeto (ocorrência) por meio (frequentemente) de seu número de página. Assim, a forma de entrada assume o papel de “representante” do objeto (coisa) ocorrência no documento, referência no discurso. Enquanto essas formas (palavras) são idênticas, a correspondência biunívoca que se estabelece é uma relação sem ambiguidades. Quando, todavia esta operação introduz elementos de lógica (intensional e extensional) convencionadas pelas linguagens documentárias, a leitura dessas relações de correspondência

assume graus de subjetividade, e, por conseguinte, riscos de interpretação, que são simultaneamente, aproximações no apontamento das referências linguísticas (objetos), e ambiguidades na decodificação da mensagem.

Formalmente, a operação denominada indexação, é o “ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação” (ABNT, 1992, p. 2). O UNISIST (1981) complementa especificando que a indexação torna possível recuperar os conceitos contidos nos documentos que podem ser representados por termos selecionados da linguagem natural (palavras-chave), de um tesauro (descritores) ou por símbolos (número de classificação retirado de um sistema de classificação bibliográfica como a CDD ou CDU).

Figura 1 – O Universo, a Inteligência, a Ciência, o livro



Paul Otlet em seu clássico *Traité de Documentation* (OTLET, 1934, p.41), ilustra seu entendimento de que os livros são apenas meios (canais) nos quais encontram-se registrados os objetos (*les choses*) da realidade, por atos de linguagem como expressão da inteligência e do saber.

Fonte: OTLET, 1934, p. 41

3.1 REFERENCIAIS LINGUÍSTICOS

Na perspectiva da semiótica de Peirce, fazendo-se intervir a tríade, ícone, índice e símbolo, um ícone representa uma experiência adquirida no passado, uma lembrança daquela experiência, um retrato do momento concluído. Um índice realiza sua experiência no presente, apontando para o objeto, sem nada mais ensinar sobre ele. O símbolo é o fato real que será experimentado no futuro, quando a mensagem chegar ao seu receptor e reconstituir o conjunto das intenções postas na mensagem pelo emissor.

O processo de busca de informações leva à constatação de uma relação certamente complexa entre escrever, comunicar e descrever a língua, para expressar o pensamento. De fato, o que se deseja é poder representar os conhecimentos materializados, por atos de linguagem, em um texto. Esta representação faz uso de métodos capazes de canalizar as propriedades

intrínsecas da língua e de sua gramática. A pesquisa que se empreende aqui consiste em transpor os conteúdos documentários em uma metalinguagem utilizando-se de imagens como representantes dessa (meta) linguagem controlada. Para tanto importa conhecer os princípios regentes desta formulação. O eixo condutor usado para este desenvolvimento, em termos linguísticos, orienta-se das pesquisas realizadas sobre línguas latinas, pela especificidade da língua portuguesa. Destaca-se que a complexidade das estruturas linguísticas latinas não permite a adoção direta das soluções propostas para as línguas anglo-saxônicas, apesar da superioridade quantitativa na produção científica.

Na França dos anos 1960, os linguistas se aplicaram em pesquisas descritivas. Os problemas até então limitados, foram tratados sob o ponto de vista dos “campos morfo-semânticos” ou sob a perspectiva lexicológica dissociada da morfologia, trabalhando-se sobre *corpus* circunscritos: descrição semântica do vocabulário da habitação, animais domésticos da moda entre outros. Pouco a pouco adere-se às propriedades físicas dos objetos denotados pelas palavras estudadas, certo progresso se dá a partir do exame dos códigos ou linguagens documentárias destinadas à recuperação dos objetos em textos de origens arqueológicas. Na pesquisa em teoria semântica destacam-se, principalmente, as pesquisas em linguística sobre a representação da informação (BARTHES, 1967; GREIMAS, 1966; 1970; 1973; 1976; GREIMAS, s.d.; GUIRAUD, 1966; MOUNIN, 1965) que deram origem aos trabalhos sobre o tratamento da linguagem natural e em particular sobre seu processamento automático, levando-se em conta os componentes morfológico, lexical, sintático, semântico e lógico, assim como as relações e dependências entre os elementos simples e complexos do discurso.

Intuitivamente, sabe-se que em algum lugar do texto existe uma significação (objetos que inferem em conhecimentos). Constata-se, não obstante, que esses objetos podem ser localizados por operações naturais, tais como a leitura, mas cuja representação artificial não se dá convenientemente. O critério essencial de modernidade nos métodos de tratamento da informação não reside no emprego de equipamentos sofisticados e modernos, mas na adoção de formas inovadoras de se apresentar o problema. A análise de um documento prevê que este seja tratado convenientemente por um sistema documentário. Essa operação consiste em indexar o documento, descrevendo seu conteúdo informacional, respeitando-se as limitações impostas pelas linguagens documentárias. A indexação, tal como é conhecida, é uma tradução lexical de unidades da língua, ou ainda uma tradução sintática para expressar as relações entre as diferentes partes do discurso, aquelas que descrevem o conteúdo, os descritores.

Segundo o uso que se pode fazer das palavras, elas podem ser de quatro tipos distintos:

- Palavras da língua;

- Palavras do léxico;
- Palavras do discurso, e
- Palavras em terminologia (Termos).

A indexação é uma operação que visa a representar os objetos de que fala o documento, o que diz o falante na mensagem. Sendo o discurso composto de objetos e suas relações, esses são também chamados de referentes. Assim, as palavras da língua ou do dicionário (morfemas) remetem unicamente aos seus significados, e não se pode dizer que elas apontam para referentes.

As palavras do léxico, lista de palavras de um sistema documentário, não possuem o status de palavras da língua, nem tampouco de palavras do discurso. As palavras do léxico não designam objetos, elas remetem a um conjunto aberto de objetos que possuem características comuns. É impossível delimitar precisamente as fronteiras deste conjunto e, por conseguinte designar um referente, para se estabelecer uma conexão direta entre as palavras do léxico e um objeto da realidade (realidade extra-linguística ou realidade imaginária). As sutilezas que separam as palavras do léxico das palavras em terminologia explicam e justificam a tendência a confundir estes dois termos.

As palavras em terminologia introduzem a questão de fronteira. Como explica o linguista Michel Le Guern (1989), no léxico como em terminologia encontram-se “palavras”, mas de um lado como de outro, não são as mesmas palavras. O objeto “palavra” do léxico é uma realidade distinta, o léxico trata as palavras independentemente das coisas, enquanto em terminologia, as palavras estão ligadas às coisas. As palavras, em lexicografia designadas como substantivos são em realidade predicados. Elas falam das qualidades e não das substâncias. As palavras do léxico remetem a propriedades e não a substâncias, a qualidades e não à objetos. A existência de um objeto demanda um termo (palavra em terminologia) para entrar no discurso.

É somente pela aplicação no discurso e pela sintaxe que as palavras do léxico podem participar na construção dos termos, i.e. entrar em relação direta com as coisas (objetos, referentes).

Esta é uma operação da lógica, e em lógica as palavras do léxico não possuem extensão, elas possuem apenas intensão (compreensão). Ao discurso corresponde uma lógica extensional e às palavras da língua uma lógica intensional, e uma se opõe à outra. A lógica extensional constrói segmentos que são portadores de referência a uma realidade, que designa objetos do universo. Fala-se, portanto de classes de objetos, deve-se considerar um universo constituído de objetos. A lógica intensional, das palavras da língua, aplica-se às noções (predicados) que fazem referência aos conceitos.

O fato de se ter no léxico, predicados livres, e em terminologia predicados vinculados, implica que pela lógica combinatória é possível se ter predicados complexos de ambos os tipos. Também pelas operações da lógica, basta que um dos predicados seja vinculado para que o predicado complexo também o seja.

A situação que consiste em descrever o conteúdo dos documentos com palavras-chave do léxico parece assim insuficiente, mesmo admitindo-se a existência de um código (a linguagem documentária) pretensamente universal, para assegurar a transição léxico-discurso. É preciso buscar no discurso os elementos necessários para a representação dos temas do discurso.

4 DESCRIÇÃO TRADICIONAL DE CONTEÚDO

Tradicionalmente, no processo de análise de conteúdo identificam-se os assuntos e a forma sob a qual foram tratadas suas contiguidade temporal e espacial, dentre outros aspectos relevantes para facilitar a localização do documento quando procurado.

No processo de indexação os conceitos são extraídos do documento por meio da análise, portanto, em linguagem natural, rica em significados e figuras, como metáforas, licenças poéticas, entre outros. A linguagem natural é caracterizada pela conotação.

A operação que permite utilizar as palavras-chave da indexação consiste em transformar conceitos em termos, os mesmos encontrados nos tesouros, listas de cabeçalhos de assunto, sistemas de classificação etc.

As abordagens linguísticas deram seus frutos (BRITO, 1992) e comprovaram, com os progressos da tecnologia, a pertinência deste método. Apesar disso, procedimentos simplificados de metadados, como os *tags* das folksonomias, se propagam consideravelmente nos sistemas de gestão da informação. A razão deste sucesso é a simplicidade da implementação, embora conhecidos os efeitos de revocação e pertinência na recuperação.

Nem tudo, portanto é ruim e nem tudo é eficiente por completo. Mesmo os modelos linguísticos mais sofisticados, dos analisadores (*parsers*) de texto integral, confrontam-se a desvios de performance devido à grande flexibilidade das leis que regem o discurso produzido pelas comunidades, em contraste com a rigidez das gramáticas de suas línguas.

Com a evolução das ferramentas de tratamento da informação escrita, verificou-se relativa evolução no reconhecimento da forma (morfologia) e das relações sintáticas inter e entre unidades do discurso. Os sistemas baseados principalmente em estatísticas tomaram proveito deste avanço atribuído aos métodos linguísticos e viu-se o aparecimento dos sistemas ontológicos e da web semântica.

O uso de métodos simplificados, híbridos e estatísticos, em oposição ao proposto pelos

métodos linguísticos, devolve ao usuário a competência para identificar a pertinência dos resultados. Aceita-se por conveniência, a imprecisão dos sistemas em localizar os documentos pertinentes, e a incerteza na exaustividade da busca.

A problemática posta é aquela em que o momento tecnológico pede novas soluções sem alterar o balanço desta relação custo-benefício. Até que ponto é possível modificar a cadeia dos processos técnicos tradicionais (indexação da informação), sem abalar as atuais instituições documentárias, e concomitantemente aumentar a satisfação do usuário, é uma questão que merece ser discutida não somente no plano teórico. A indústria da informação desempenha também um papel considerável nesta relação.

5 CONTRIBUIÇÕES DA SEMIOLOGIA E SEMIÓTICA

Partindo-se do referencial, em que se esclarece a natureza do signo, representante das referências da realidade extra-linguística, i.e. dos objetos do discurso, entende-se que a qualidade dos sistemas de recuperação está intimamente ligada às propriedades de vinculação entre unidades do discurso e seus objetos. Esta é a essência dos mecanismos de busca documentária. A eficiência dos sistemas de recuperação de informação está intimamente relacionada à identificação das coisas de que tratam os documentos.

As duas ciências Semiologia, desenvolvida na Europa por Ferdinand Saussure linguista francês, e Semiótica, desenvolvida nos Estados Unidos pelo filósofo americano Charles Sanders Peirce, ambas estudam os signos na vida social, embora sejam escolas diferentes. A semiótica provê conceitos e ideias que permitem uma análise sistemática dos signos, que podem ser palavras, imagens, textos completos, artefatos, gestos, posturas físicas etc., estuda ainda o meio em que os signos são utilizados e como os significados são construídos.

Figura 2 – O signo e a linguagem segundo Saussure: representação esquemática.



Fonte: Elaboração própria

Paralelamente e na mesma época (1900), aparecem em semiótica, as teorias da indexicalidade de conteúdos de/por imagens. A revisão dessa abordagem converge para a teoria do signo de Charles Sandes Peirce, que defende uma semiótica fundamentada na relação entre o símbolo, o ícone e o índice. Linguagem e imagem compõem-se assim de signos que possuem uma função icônica, indicial e simbólica. Para maior objetividade do assunto proposto, o nível de simplificação adotado para a teoria de Peirce resume-se à tríade índice, ícone e símbolo, da seguinte forma:

- O ícone é um signo em que o significante corresponde ao significado. A imagem é uma representação fiel da realidade, relações aonde estabelece-se uma completa analogia com as partes do objeto imaginário.

- O índice é um signo representado pela resultante que o significado causa no significante. Índices relacionam signos já experimentados. Assim, pessoas com guarda chuva na rua relacionam-se com a condição de que está chovendo. A representação da chuva que não se vê na imagem passa pela expressão do(s) ícone(s) relacionado(s).

- Um signo simbólico existe quando a relação entre significante e significado é puramente convencional, tal como as placas de trânsito, tal como a pomba que representa a paz.

- Finalmente, Peirce considera que a completude da mensagem está presente num signo “quando os papéis icônicos, indicativos e simbólicos são misturados tão irmãmente quanto possível” (PEIRCE apud NÖTH, 2013a, p. 94).

A teoria semiótica de Peirce é amplamente aplicada à descrição de imagens (BOULANGER, 1996). Nas coleções imagéticas e nas práticas arquivísticas, trata-se de um processo de descrição literal de um documento imagético para efeito de recuperação.

Vê-se, assim, que a representação de conteúdos através de imagens pressupõe uma operação diferenciada em indexação. As vantagens que se anunciam são múltiplas: atração visual do usuário pelo sistema, compreensão intuitiva do código de indexação, elevado potencial didático, ampliação da portabilidade dos indexadores, interoperabilidade entre as leis que regem os universos do discurso e do especialista indexador, e em consequência, também com o universo de compreensão do usuário.

Relembrando, no processo de organização da informação, especificamente a etapa de descrição do conteúdo do documento, a linguagem é utilizada para representar os conceitos que nele são tratados. Nesse procedimento, em que o usuário realiza uma busca aleatória com imagens, o resultado final deve ser um registro do documento, ou conjunto de registros de documentos semelhantes em imagens e em texto, os quais darão acesso à localização física dos documentos no acervo.

A esta representação esquemática de navegação pressupõe uma argumentação teórica subjacente. Segundo a teoria pragmática de Peirce, cada objeto do universo se relaciona com um indeterminado número de outros objetos, tanto direta como indiretamente, de tal forma que cada elemento transporta um potencial indicial indeterminado (LEFEBVRE, 2007, p. 220). Como então o signo representa seu objeto, e como ele revela seu objeto? A representação icônica aparece como uma qualidade que possui o objeto. Se essas qualidades emergem, então o objeto pode ser identificado como tal e através do seu ícone. O caráter indicial está intimamente ligado ao objeto: o frio ou temperaturas baixas estão intimamente ligados a imagens de neve ou de gelo. Simbolicamente, o signo precisa ser interpretado. É pela ação dos signos que se infere que uma realidade é verdadeira e neste contexto pode-se participar da indexação de imagens. Pelo conhecimento, o signo pode ser interpretado, ampliando-se os limites da relação signo-objeto.

6 DA DESCRIÇÃO À CRIAÇÃO DO SIGNO IMAGEM-CHAVE

Considerando-se a tarefa de leitura (descrição) da imagem, as teorias saussuriana e peirciana seriam suficientes para que, com a prática adequada, um profissional da indexação pudesse indexar imagens atribuindo-lhe palavras-chave segundo seus níveis de percepção: significante e significado, conotação e denotação, ícone, índice ou símbolo. O vetor resultante desses indicadores serviriam para armazenar e recuperar a imagem adequadamente. Mas, a inovação que se deseja agregar ao processo é justamente substituir o papel do signo palavra-chave tradicional, por um signo imagem, representante da mensagem do conteúdo dos documentos. Sobre isso, a contribuição de Jacques Bertin (1967 apud DANTIER, 2008; BERTIN, 1970) é particularmente enriquecedora.

Sobre a análise da informação, Bertin esclarece que todo pensamento se expressa por um sistema de signos, imitando uma codificação natural. A linguagem verbal é um código de signos sonoros, a escrita de uma língua é outro tipo de código, e a representação gráfica também. Dado que a representação gráfica é a transcrição da informação em um sistema gráfico de signos, a representação gráfica faz, portanto parte da semiologia.

Transcrever significa separar os elementos do pensamento qualquer que seja o sistema de signos disponível. Independente da qualidade dessa transcrição, somente interessa a qualidade e a eficácia da transcrição gráfica, pois é esta que representa o conteúdo da informação. Cada sistema de signos dispõe de seus próprios meios, estilo e estética para representar o pensamento. Trata-se de uma relação entre diversos conceitos reconhecidos e isolados, escolhidos dentre todos os conceitos imagináveis. Assim, diz Bertin (apud DANTIER,

2008, p. 19, tradução nossa, grifo do autor) que

[...] na representação gráfica, chamar-se-á INFORMAÇÃO o conteúdo traduzível de um pensamento. Ele [o conteúdo] constitui-se essencialmente de uma ou várias CORRESPONDÊNCIAS ORIGINAIS dentre um conjunto finito de conceitos de variação, e um conjunto invariante.

Bertin introduz também em sua teoria da imagem o conceito de eficácia definida em função do tempo de percepção da mensagem. Para se obter uma resposta correta e completa a uma dada questão, a “construção” que requer o tempo mais curto de percepção da resposta é a construção mais eficaz, e corresponde à imagem. Esta noção de “custo mental” foi posta em evidência pela lei de Zipf (1949), bem conhecida em Biblioteconomia.

Bertin evidencia assim três operações sucessivas para a leitura de uma imagem que vão guiar, posteriormente, o processo inverso de construção da imagem (mensagem):

1. Identificação do externo – concretizar no pensamento os conceitos propostos, os componentes.
2. Identificação interna – por meio de que variáveis os componentes se expressam.
3. Percepção das correspondências originais. Esta percepção é o resultado de uma questão, consciente ou não. A questão que se pode colocar diante de uma informação.

Significa que o olho humano isola as informações visuais pelo reconhecimento das correspondências definidas e percebe a mensagem em seu contexto. Esse exercício de percepção conduz à definição de imagem como sendo “a forma visual significativa perceptível no instante mínimo de visão”.

Estabelecendo-se um paralelo entre a percepção da mensagem pela imagem, Winfried Nöth explica sobre a importância que Peirce atribui aos signos no processo de aprendizagem, ressalta, assim, que a combinação de índices com símbolos e ícones, “faz os índices serem instrumentos didáticos muito poderosos” (NÖTH, 2013a, 2013b). Esta noção de aprendizagem há muito vem sendo ressaltada pelos pesquisadores em semiótica (TUFTE, 1998; SOUSANIS, 2012; 2015). Em Sousanis (2015) o autor traz uma contribuição importante sobre o pensamento e a narrativa visual. Sua tese de doutorado, totalmente escrita em forma de história em quadrinhos, permite modos de entendimento mais profundos do que normalmente se apreende de uma mensagem. O suporte imagético assessoria o texto como descrito por Peirce na completude do signo perfeito.

Finalmente, retém-se como justificativas para a composição de uma representação gráfica, as três funções básicas: registrar, comunicar e tratar a informação.

O processo de indexação por imagens prevê assim um momento de leitura e outro de

concepção. Indexar por imagens consiste em construir uma imagem representante da mensagem do conteúdo do documento.

Nesse processo de concepção, o especialista indexador dispõe de ferramentas semelhantes?

Às linguagens documentárias:

- Deve-se constituir um banco de imagens, preferencialmente padronizadas, segundo uma política de programação visual estabelecida. Entende-se que o corpus de “imagens-chave” deve seguir padrões de coerência tais como o estilo, a técnica etc. de forma a guardar a homogeneidade da interpretação da mensagem contida nos signos. Observa-se também que, quanto mais universais forem os signos imagéticos, mais fácil será a compreensão da mensagem pelo grande público.

- As mesmas remissivas dos instrumentos tradicionais (Imagem genérica, Imagem específica, Use, Usado para, Ver etc.) devem ser previstas para as imagens, estabelecendo-se padrões para o uso das imagens (GHEORGHITA, 2011).

- Os princípios anunciados anteriormente por Bertin devem ser observados para se dar ao corpus de imagens uma identidade e um estilo de linguagem.

- Um banco de imagens pode, tal como um tesouro, ser construído por profissionais capacitados, a fim de se obter uma coleção homogênea e coerente com a área de especialização.

- A recuperação das imagens, no bando de imagens, pode adotar métodos específicos ou híbridos de busca: uso de imagem-chave ou palavras-chave para se recuperar imagens. Nesse caso, às imagens desse banco de imagens serão associados metadados para facilitar a recuperação dos registros e assim transmitir suas características temáticas às colagens (imagens compostas).

O resultado da indexação:

- Contrariamente às palavras-chaves, o produto não será uma seleção de imagens adjacentes umas às outras, mas uma colagem de imagens significativas, criando-se assim uma “carta gráfica”, “carta de indexação” ou “imagem-chave” da obra. Esta imagem de indexação será o signo indicial, icônico e simbólico da obra, herdando todas as características referenciais das imagens que a compõem e criando novas por associação.

- A composição desta “imagem-chave” segue a orientação do “signo completo” de Peirce, em cuja composição mínima encontram-se signos icônicos, indiciais, e simbólicos completando-se a tríade peirciana.

- A escolha do signo icônico deve ser, tanto quanto possível, a imagem da capa da obra, ou aquela que melhor a representa (há livros que não possuem imagens na capa). Entende-

se que a imagem-capa aponta igualmente para o objeto documento, preenchendo uma dupla função. Não obstante é desejável que a colagem seja composta por no mínimo três imagens, mesmo que redundantes em sua função, mas que permitam detalhar o conteúdo da obra.

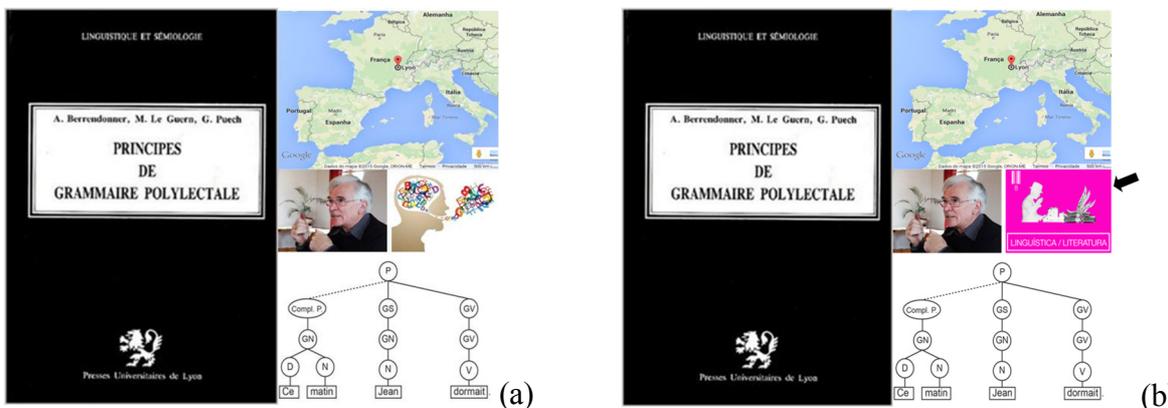
- Como, pelas dimensões reduzidas, o resultado é uma composição com poucas imagens, deve-se dar prioridade a uma imagem de localização geográfica e outra temporal ou temática, situando a obra nesses eixos lógicos.

O experimento de indexação por imagens apresentado a seguir, é apenas ilustrativo do que se pode obter como resultado. Pelo fato de não se ter ainda constituído um banco de imagens padronizadas, equivalentes a uma linguagem documentária tradicional, usou-se para este fim imagens de domínio público recuperadas na web.

Escolheu-se aleatoriamente obras reais referenciadas neste trabalho ou recuperadas no catálogo da Biblioteca Universitária (BU), para se mostrar o mecanismo de coerência no processo indexação por imagens.

As informações do registro catalográfico são de autoria da BU.

Exemplo 1 (a,b) - BERRENDONNER, Alain; LE GUERN, Michel; PUECH, Gilbert. Principes de grammaire polylectale. Presses universitaires de Lyon, 1983.



Fonte : Elaboração própria (imagens de domínio público)

As colagens acima seguem os princípios anunciados na fundamentação teórica. A título demonstrativo introduziu-se uma imagem na colagem (b) representando a classe CDU (8) à qual pertence a obra. Este artifício quando usado sistematicamente permite uma leitura mais fácil da colagem, uma vez que todas as obras da mesma classe poderão ser identificadas visualmente por um índice gráfico que situa a obra em um contexto facetário.

Figura 3 - Representação gráfica das 10 classes da CDU



Fonte: Biblioteca pública. Governo dos Açores

Exemplo 2 : FAULSTICH, E. L. de J. Como ler, entender e redigir um texto. 1999.



Fonte: Composição artificial sobre a tela de resposta do sistema documental da BU. Imagens de domínio público

A imagem-chave de indexação desta obra adota o mesmo o procedimento descrito anteriormente: Imagem (icônica e indicial) da obra, imagem indicial geográfica, e duas imagens assessórias para situar o tema (recorrência da classe CDU). Acrescentou-se à imagem geográfica tanto um marcador para a cidade de Brasília como o logotipo institucional do autor, acessórios que agregam informação e mostram a flexibilidade do sistema, aumentando-se a precisão da descrição. O resultado é o conjunto “imagem-chave”.

Figura 4 – Tela de resposta (real) do sistema da BU com os indexadores de assunto



Fonte: Sistema de gestão documentária da BU.

Observa-se a expressividade proporcionada pelo uso da imagem-chave em lugar das palavras-chave usadas para descrever o assunto do documento. Nesse mesmo sentido, ressalta-se que, no atual sistema documentário tradicional, não se admite a busca por semelhança ortográfica ou fonética, o descritor usado deve corresponder exatamente ao registrado no sistema, daí a dificuldade em se encontrar determinados documentos aonde, por exemplo, a busca pelo nome do autor (ou nomes próprios) estabelece uma relação de correspondência crítica quando não se sabe exatamente a ortografia da palavra.

Observa-se no exemplo seguinte a introdução de uma imagem que situa o usuário na dimensão da autoridade da obra: o Instituto de Antropologia e História, e outra para o assunto na Classe 3 CDU, que concorrem para a identificação da obra. Nota-se ainda a translação do elemento geográfico mostrado no mapa, quando se lê na obra “Noueva España”, que hoje é o México.

Exemplo 3 - CARRILLO Y GARRIEL, Abelardo. El traje en la nueva españa. 1959.

CATALOGAÇÃO PRÉ-MARC

Número de Chamada

391(72) C317t

Autor Principal

[Carrillo Y Garriel, Abelardo](#)

Título Principal

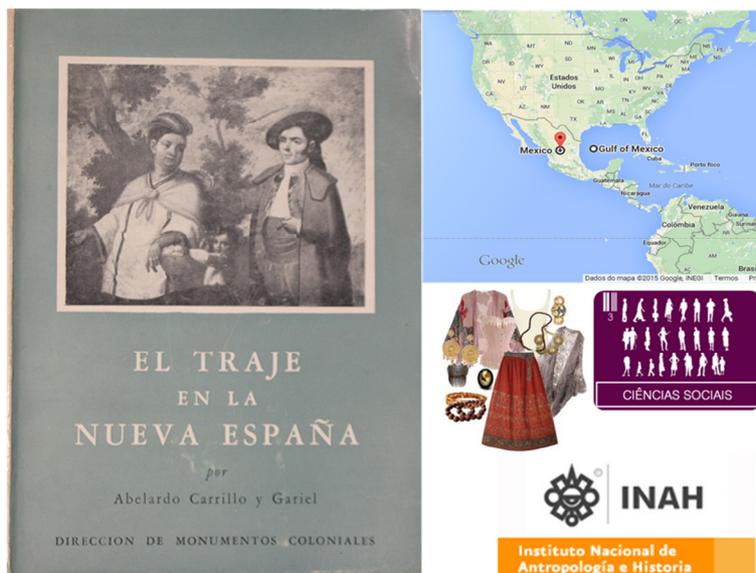
Traje en la Nueva Espana(el)

Publicação

México, DF : Inst Nac Antropol E Hist, 1959.

Descrição Física

207 p. : il.



Assuntos

[Adornos:antropologia cultural](#)

[Elegancia:antropologia cultural](#)

[Mexico:america do norte](#)

[Moda:antropologia cultural](#)

[Trajes:antropologia cultural](#)

[Vestuario:antropologia cultural](#)

Fonte: Composição artificial com a resposta do sistema documentário da BU. (Imagens de domínio público)

7 CONSIDERAÇÕES FINAIS

Indexar por imagens reconsidera os paradigmas da primazia das palavras sobre as imagens e anuncia uma legitimidade com perspectivas múltiplas para a indexação de documentos, distinguindo-se o que seria um mundo reservado aos praticantes de uma determinada língua escrita, por um sistema de comunicação mais universal baseado em imagens. A representação do conteúdo proposta por meio de imagens-chave, sustentada pelas teorias da semiologia saussuriana ou da semiótica peirciana, introduz uma nova proposta no cenário da documentação que merece ser testada mais amplamente. A descrição das imagens através de suas representações textuais comportam os fundamentos para os sistemas de gestão de bases de dados imagéticas, e constituem as ferramentas para a composição do princípio de

indexação por imagens anunciado neste trabalho. As diretrizes revisitadas de Bertin mostram como se pode chegar à composição de uma imagem em concordância com as teorias da indexação textual e, por conseguinte com os mecanismos da indexação por imagem.

Os ensaios de composição das fichas descritivas dos documentos demonstram a aplicação das bases teóricas citadas e esclarecem sobre a viabilidade de uma indexação mais eficaz. Nesses termos, os resultados abrem possibilidades múltiplas de acessibilidade aos acervos documentários, aos acervos imagéticos, como à indústria editorial da imagem, e em particular às estórias em quadrinhos, que encontram nesta proposta outras formas de expansão. Os especialistas da indexação descobrem potencialidades para novas habilidades, com repercussões diretas na estrutura interna dos sistemas de gestão e recuperação da informação. Consequentemente, os motores de busca na web dotam-se dessas estratégias de indexação para uma recuperação otimizada da informação. Os próximos passos consistem em testar a aplicabilidade deste processo de indexação sobre conteúdos abstratos, aonde se fará intervir a composição de imagens-chave mais complexas.

A metodologia proposta abre, oportunamente, campo para o desenvolvimento de interfaces mais adaptadas aos usuários com necessidades especiais, deficientes e iletrados funcionais, com propostas de realização de estudos mais aprofundados sobre os mecanismos de concepção de imagens. Competências transdisciplinares em Neurociência, Linguística e Ciências da Informação podem, juntas, explorar questões tais como a natureza dos processos cognitivos em organização e classificação de informações de deficientes auditivos, o funcionamento das gramáticas de comunicação dos autistas, os mecanismos de ação da biblioterapia para idosos, pacientes com Alzheimer e outras patologias. Enfim, o uso da imagem para o estabelecimento de elos entre usuários e documentos parece encontrar nos sistemas de recuperação da informação por imagens um vasto campo de verificação científica tanto em pesquisa como em desenvolvimento.

REFERÊNCIAS

ABNT. **Métodos para análise de documentos**: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992. 4 p.

ASHTON, D. **10 reasons why you should care about visual content marketing**. Blog. Publicado em 19 de fevereiro, 2015. Disponível em: <<http://neomam.com/blog/13reasons/>>. Acesso em: 5 maio 2015.

BARTHES, R. *Le système de la mode*. Paris: Seuil, 1967.

BERRENDONNER, A.; LE GUERN, M.; PUECH, G. **Principes de grammaire polylectale**.

Presses Universitaires de Lyon, 1983.

BERTIN, J. La graphique. **Communications**, v. 15, n. 1, p. 169–185, 1970.

BIBLIOTECA pública. Governo dos Açores. Disponível em:

<<http://www.bparah.azores.gov.pt/html/bparah-seccao+infanto-juvenil+cdu.html>>. Acesso em: 3 maio 2015.

BOULANGER, C. La surimage. **Applied Semiotics/Sémiotique Appliquée**, v. 1, n. 2, p. 159–172, 1996.

BRITO, M. de. Sistemas de informação em linguagem natural: em busca de uma indexação automática. **Ciência da Informação**, v. 21, n. 3, p. 223–232, 1992.

DANTIER, B. La représentation et l'étude visuelles des informations. In.: BERTIN, J.

Semiologie graphique, les diagrammes, les reseaux, les cartes. 4. ed. Paris. Ehes, 2008. 32 p.

GHEORGHITA, I. Methodologie de construction automatique du thesaurus pour l'indexation et la recherche des images. **RECITAL**, n. 2, p. 221–228, jun. 2011. Disponível em:

<<https://hal.archives-ouvertes.fr/hal-00695722>>. Acesso em: 2 maio 2015.

GREIMAS, A. J. **La mode en 1830**: essai de description du vocabulaire vestimentaire d'après les journaux de mode de l'époque. Paris, Sorbonne (thèse dactylographiée). s.d.

GREIMAS, A. J. **Sémantique Structurale recherche de méthode**. Paris: Larousse, Langue et Langage, 1966.

_____. Du Sens. **Essais Sémiotiques**. Paris: Seuil, 1970.

_____. Les actants, les acteurs et les figures, In : SEMIOTIQUE narrative et textuelle, Claude Chabrol (Dir.), Paris, Larousse, 1973. p. 161-176.

_____. **Sémiotique et sciences sociales**. Paris : Seuil, 1976. 219 p.

GUIRAUD, P. De la grive au maquereau: le champ morpho-syntaxique des noms de l'animal tacheté. **Le Français Moderne**. v. 34, 1966.

LEFEBVRE, M. The art of pointing: on Peirce, indexicality, and photographic images.

Photography Theory, p. 220-244, 2007.

LE GUERN, M. Sur les relations entre terminologie e lexique. **META Journal de traducteurs**. Montréal, 1989.

LOPES, A. C. M. Texto e coerência. **Revista Portuguesa de Humanidades**, v. 9, n. 1/2, p. 13-33, 2005.

_____. Texto, gramática e processamento. In: COLÓQUIO DE HOMENAGEM A FERNANDA IRENE FONSECA, 2008. **Actas...** 2008. p. 57-68.

MOUNIN, G. Essai sur la structuration du lexique de l'habitation. **Cahiers de Lexicologie**, v. 5,

1965.

_____. Un champ sémantique: la dénotation des animaux domestiques. **La Linguistique**. v. 1, 1965.

NÖTH, W. Os signos como educadores: Insights peircianos. **TECCOGS: Revista digital de tecnologias cognitivas**, n. 7, p. 75-97, jan./jun. 2013a. Disponível em: <www.pucsp.br/pos/tidd/teccogs/index.html>. Acesso em: ago. 2014.

_____. Semiótica visual. **Triáde: Comunicação, Cultura e Mídia**, v. 1, n. 1, p. 13-40, jun. 2013b. Disponível em: <periodicos.uniso.br/ojs/index.php>. Acesso em: ago. 2014.

OTLET, P. **Traité de documentation** : le livre sur le livre - théorie et pratique. Bruxelles, Editions Mundaneum, 1934. 451 p.

ROGERS. S, **What fuels a tweet's engagement?** (@smfrogers). Data editor 03/10/2014. Disponível em: <<https://blog.twitter.com/2014/what-fuels-a-tweets-engagement>>. Acesso em: abr. 2015.

SOCIAL BAKERS. Photos cluttering your facebook feed? here's why pictures account for majority of brand facebook posts. Abril 21, 2014, **Advertising & Marketing**. Disponível em: <<http://www.emarketer.com/Article/Photos-Cluttering-Your-Facebook-Feed-Here's-Why/1010777/1>> Acesso em: maio 2015.

SOUSANIS, N. The shape of our thoughts: a meditation on & in comics. **Visual Arts Research**, v. 38, n. 1, p. 1-10, 2012.

_____. **Unflattening**. [S.l.]: Harvard University Press, 2015. 208 p.

TUFTE, E. R. **Envisioning information**. Connecticut, USA: Graphics Press, 1998, 126 p.

UNISIST. Princípios de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 10, n. , p. 83-94, mar. 1981.

ZIPF, G. K. **Human behavior and the principle of least effort**. Cambridge, (Mass.): Addison-Wesley, 1949.