

Requisitos mínimos para gerenciamento e recuperação de textos e imagens

Regina de Barros Cianconi

TIPOS DE SISTEMAS DE TEXTO E IMAGEM

A diversidade e despadronização das tecnologias voltadas ao gerenciamento e recuperação de textos e imagens dificultam sua seleção adequada. Ainda não existe uma categorização formal para os diferentes produtos existentes no mercado, mas pode-se estabelecer critérios de diferenciação, baseados em características e facilidades, de forma a compreender as funções de cada tipo de *software*.

A classificação a seguir foi elaborada apenas para fins de entendimento, pois existem produtos bastante flexíveis com características de hipertexto e recuperação no texto, armazenamento livre e estruturas de campos, gerenciamento de imagem e recuperação no texto etc.

SOFTWARE DE RECUPERAÇÃO DE TEXTOS

Sistema que permite criar bases de dados com conteúdo predominantemente textual, possuindo recursos de recuperação adequados à consulta de textos, seja na integra ou ementas, resumos, referências bibliográficas. Há quem os denomine *softwares bibliográficos*, embora este seja um nome restritivo. São exemplos desta categoria Lightbase, SRI, Microlsis, MQuestel.

SOFTWARE DE GERENCIAMENTO DE TEXTOS

Sistemas que, além das funções de recuperação peculiares aos sistemas de recuperação de texto, têm algumas facilidades que agilizam o tratamento das informações.

Alguns exemplos de facilidades possíveis são os seguintes: flexibilidade de armazenamento, tratando igualmente textos e bases estruturadas, sem necessitar definir registros e tipos de documentos; acesso ao conteúdo dos arquivos em tem-

po real - sem a geração de índices (listas invertidas); ou ainda a geração automática de índices que apontam para qualquer palavra do texto; busca tipo hipertexto (por meio de nós e *links* criados automaticamente e que possibilitam prosseguir a busca a partir de palavras assinaladas na tela de resposta); integração de arquivos com diferentes estruturas (bases de dados relacionais ou não, textos gerados por editores de texto).

Geralmente tais produtos efetuam importação de arquivos e aceitam imagens digitalizadas. Caracterizam-se principalmente pelas facilidades proporcionadas na etapa de indexação, pelos recursos sofisticados de busca da informação e pela possibilidade de integração de diferentes tipos de documentos. Nesta categoria, enquadram-se produtos como o Asksam, que trabalha com textos livres e estruturados ao mesmo tempo¹.

Os produtos voltados à recuperação e gerenciamento de texto podem ter diferentes formas de estruturação das informações:

- Formato livre — Os dados são armazenados em registros, porém de modo livre (semelhante a um editor de textos), sem a estrutura de campos de um Sistema de Gerenciamento de Banco de Dados — SGDBS (ou DBMS). A recuperação se dá tal como nos SGDBS, com a diferença de que se pode também efetuar buscas por assunto no contexto (textos completos).
- Formato estruturado — Os dados são organizados em campos e em registros seqüenciais — lineares, portanto. Quase todo *software* de recuperação de texto tem esta estrutura de armazenamento.

A diferença entre os produtos de recuperação de texto e os SGDBS, tipo Dbase, é basicamente o fato de possuírem linguagem de busca voltada para

Resumo

Categorização de produtos voltados à recuperação e gerenciamento de textos e imagens, desde os já tradicionais softwares de recuperação textual, aos mais modernos produtos voltados ao gerenciamento de textos, aos sistemas para desenvolvimento de aplicativos hipertexto e/ou hipermídia, ou de processamento e gerenciamento da imagem. Os critérios para definição dos recursos desejáveis devem levar em conta a missão, as características e o ambiente do sistema ou serviço, o conteúdo informacional existente, o meio atual de armazenamento, além dos objetivos de uso. São sugeridos diversos parâmetros para avaliação de produtos.

Palavras-chave

Armazenamento e recuperação da informação; Recuperação de textos; Gerenciamento de textos; Gerenciamento de imagem; Hipertexto; Hipermídia; Avaliação de software.

Este artigo é uma adaptação da palestra com o mesmo título apresentada na Comdex 94, no Rio de Janeiro, promovido pela Sucesu, em abril de 1994.

textos, efetuando busca de assuntos no contexto, utilizarem operadores de proximidade, permitirem campos repetitivos e campos de tamanho variável.

Armazenamento hipertexto - O texto é armazenado de forma não-linear, em rede, com ligação entre trechos de texto (nós).

Assim como há produtos que trabalham com textos em formato livre e em formato estruturado, diversos produtos de hipertexto permitem, além da navegação, a busca por meio da recuperação de texto e a tendência é ter as duas funções interligadas^{2,3}.

SOFTWARE DE GERENCIAMENTO DA IMAGEM

Caracterizado como um sistema que permite o arquivamento das imagens dos documentos digitalizados, o que significa armazenar e recuperar imagens de documentos em um formato digital (usando código binário). Devido ao grande espaço em disco necessário ao armazenamento de uma imagem, geralmente se utilizam os discos ópticos para seu armazenamento.

Um sistema de processamento da imagem é composto de vários equipamentos, como *scanners, drivers, jukeboxes*, controladoras de computador, monitores de vídeo gráficos e impressoras laser^{4,5}. O tratamento que a imagem do documento recebe varia conforme o uso. Para editoração eletrônica, esta é submetida a softwares de computação gráfica, *desktop publishing* etc. Para arquivamento eletrônico, a imagem é armazenada em sistemas de arquivamento e em geral possui funções chamadas de *workflow*, que permitem automatizar o fluxo de trabalho, circulando documentos via computador, alguns, inclusive, integrando micros ao grande porte. Tem recursos de indexação e de busca, nem sempre muito completos ou sofisticados, pois o objetivo é o armazenamento e recuperação da imagem do documento. Exemplos destes produtos são File Power, Clarity, Keyfile, Metaview, BD&I, Image Plus. Uns rodam em PC sob Windows, ou OS/2 ou DOS, outros em plataformas como SUN, AS 400 etc.

SOFTWARE HIPERTEXTO

Hipertexto consiste em uma técnica de organização da informação que incorpora uma filosofia de armazenamento diferente da tradicionalmente usada nos sistemas de informação. A idéia em si é bem antiga. Foi concebida em 1945 por Vannevar Bush em um trabalho intitulado *As we may think*, que visava a organizar documentos em tre-

chos e combiná-los conforme a necessidade de compreensão e abordagem do leitor. Mais de 20 anos depois, em 1968, Douglas Engelbert apresentou o sistema Augment, no qual introduziu os conceitos hoje conhecidos como hipertexto.

O termo é empregado para designar a organização não-linear da informação, que é armazenada em trechos de texto (e/ou som e/ou imagem), ligados entre si por meio de elos ou *links*. A tendência é o hipertexto ser sempre utilizado com recursos de imagem e som (multimídia), recebendo a denominação de hiper­mídia^{6,7,8,9}.

O armazenamento com nós e *links* se presta à construção de diferentes ferramentas baseadas em hipertexto, tais como sistemas para desenvolvimento de aplicações multimídia, sistemas para desenvolvimento de aplicações de autotreinamento, sistemas para desenvolvimento de hiperdocumentos: trechos de informação com vários "caminhos" possíveis a seguir (navegação).

Software de hipertexto, portanto, é aquele que permite gerar documentos hipertexto, com organização não-linear, pela função autoria. Deve ser possível conectar trechos de texto e imagens, som; impor mais de uma organização ao documento ou informação, isto é, o mesmo conteúdo pode ser lido: a) em superfície - somente a introdução ou os tópicos de cada assunto; b) em profundidade - apenas os detalhes de um pequeno subtópico; reutilizar um mesmo trecho em mais de um documento. Alguns exemplos de software para autoria de hipertexto são Guide, Folioviews, HTIES, Hypersoft, cada qual com diferentes características.

SOFTWARE MULTIMÍDIA

À apresentação da informação de maneira multissensorial, integrada e interativa, dá-se o nome de multimídia.

A tecnologia básica da multimídia existe há anos no mercado como entidades separadas: Windows, CD-ROM, placas de som, videodiscos. Os primeiros computadores lidavam somente com informações alfanuméricas. Em seguida, os computadores pessoais passaram a trabalhar de modo gráfico, mas em baixa resolução, insuficiente para texto e gráficos simultaneamente. As primeiras interfaces gráficas poderosas foram o Machintosh, 1984; depois veio o PC, com o primeiro Windows em 1986.

Posteriormente, vários ambientes e aplicativos começaram a fazer apresentações do tipo *WYSIWYG* (What You See Is What You Get, ou seja, você vê o que se

obtem), com múltiplos tipos de letras, estilos e tamanhos. Surgiram, ao mesmo tempo, periféricos que facilitam a leitura e impressão de gráficos e de texto, tais como *mouse, scanner*, impressora a laser. Nos sistemas multimídia, mais de um sentido humano está envolvido no processo de recuperação da informação, que se apresenta simultaneamente sob a forma texto, imagem em movimento, ou não, e som.

Alguns exemplos de sistemas para desenvolvimento de aplicações multimídia são Toolbook, AVC, Iconauthor, Authority. Estes produtos permitem que se construam nós e *Links* de hipertexto, desenvolvendo aplicações hiper­mídia, portanto.

De modo geral, os sistemas aqui classificados como de "gerenciamento de texto" e de "gerenciamento da imagem" permitem a incorporação de som — característica de multimídia. Contudo, como sua vocação primeira é permitir o gerenciamento de textos e de imagens de documentos, assim foram categorizados.

Uma observação a ser feita é quanto à digitalização de documentos: se os documentos forem armazenados como imagem, ocuparão grande espaço de arquivamento, exigindo discos ópticos e terão de ser indexados manualmente, após ou antes da digitalização. No entanto, caso se deseje fazer a busca pelo conteúdo do texto, é necessário submeter o texto digitalizado a um *software* de OCR, para que os caracteres possam ser interpretados e transformados em ASCII, sendo "lidos" por um editor de texto ou por um software de recuperação ou gerenciamento de texto. Os produtos de OCR estão evoluindo rapidamente graças ao emprego de inteligência artificial.

Entre os produtos de OCR disponíveis no mercado, podem ser citados Recognita, Galera, Omni Page, Onion Multireader, View Graber, Forformix, cada um com suas peculiaridades, exigindo também critérios de seleção voltados às necessidades específicas. São diferentes as exigências para reconhecimento de textos em formulários de coleta de dados, em impressos, em textos de jornal, em livros, em fichas com assinatura, em documentos manuscritos etc.

CRITÉRIOS PARA DEFINIR OS RECURSOS DESEJÁVEIS

A escolha de um software de recuperação de bases textuais, de gerenciamento de textos, hipertexto e/ou gerenciamento de imagem deve ser baseada em características e padrões mínimos desejáveis¹⁰, que precisam levar em conta;

DO PONTO DE VISTA DO SISTEMA OU SERVIÇO DE INFORMAÇÕES

- Critérios de compatibilidade e integração com o ambiente de *hardware* e *software* existente; capacidade de conectividade e interoperabilidade; facilidade e flexibilidade de uso; capacidade de customização; funções e características do *software*; serviço de treinamento e atendimento ao usuário.

DO PONTO DE VISTA DA INFORMAÇÃO

- Natureza da informação e necessidade de integração de informações em diferentes estruturas - texto não estruturado; texto estruturado em bases de dados (bibliográficas ou não); dados numéricos (em planilhas, quadros, tabelas); gráficos, gravuras, fotos.
- Uso dessa informação (necessidade a atender, perguntas mais freqüentes, tipo de usuário e nível de familiaridade com a informática; freqüência de consulta; benefício obtido etc.
- Atualização, volume de informações a armazenar e estimativa de crescimento do acervo.
- Podem ser combinados diferentes critérios: informações para recuperação por itens objetivos e conhecidos (códigos, data, nome etc.); informações para recuperação por itens aleatórios e subjetivos, como assunto, classificações; informações que estejam indexadas por vocabulário controlado (tesauro); informações que possam ter indexação automática das palavras do texto; informações para consulta sob a forma de navegação (hipertexto).

DO PONTO DE VISTA DOS OBJETIVOS DE USO

- a) Quando o objetivo é arquivamento eletrônico da imagem de documentos, especialmente em discos ópticos em substituição aos arquivos em papel, visando essencialmente à economia do espaço físico e a agilização na localização, manuseio e fluxo de documentos —> Sistema de gerenciamento da imagem.

Nesse tipo de sistema, geralmente não se necessita submeter os documentos a um *software* de OCR. A recuperação é feita por meio de índices definidos pelo usuário - externos ao documento e criados a partir de campos estruturados, como em uma base de dados convencional, por exemplo, data, nome, empresa, origem, destino.

Certos sistemas de gerenciamento da imagem possuem recursos de recuperação voltados a textos, outros não. Aí é preciso cuidado. Alguns destes últimos chamam de indexação a simples atribuição de um código e um nome ao documento - meras chaves de acesso. Não necessariamente trabalham com busca por assuntos (operadores booleanos), menos ainda trabalham com a busca no texto (operadores de proximidade etc.).

Convém observar que, caso a empresa queira trabalhar com documentos administrativos e técnicos, é essencial que o sistema de gerenciamento da imagem permita também atribuir assuntos (indexar por conceitos) e possua pelo menos os recursos de álgebra booleana.

- b) Quando o objetivo é fazer buscas em arquivos com dados textuais, usando palavras-chave extraídas da linguagem natural ou vocabulário controlado (tesauro) —> Sistema de recuperação de textos.

Convém observar que, se os documentos estiverem em papel, a entrada de dados pode ser feita mediante digitação ou um *scanner*, mas, neste último caso, é imprescindível que o documento seja submetido ao OCR para reconhecimento dos caracteres ópticos, de forma a possibilitar que se faça recuperação a partir das palavras contidas no próprio texto do documento.

- c) Quando a necessidade é trabalhar com texto, recuperando informações, quer por palavras-chave extraídas da linguagem natural dos documentos, quer por vocabulário controlado e, além disso, informatizar o ambiente de trabalho com bases textuais estruturadas, & não estruturadas —> Sistema de gerenciamento de textos.

Interessante pelo aproveitamento da mesma tecnologia para diferentes tipos de informação e pela possibilidade de padronizar todo o ciclo da informação: desde a geração ao arquivamento e recuperação, armazenando a correspondência, fax, bancos de dados, simplificando e otimizando os recursos de informação.

- d) Quando a necessidade básica é efetuar buscas não lineares, isto é, sem estrutura hierárquica pré-definida, associando trechos de textos e imagens (desenhos, gravuras, fotos etc.) —> Sistema de hipertexto.

Neste caso, a recuperação é primordialmente através da navegação no texto completo do hiperdocumento. É útil em

documentos que visam à leitura não seqüencial, tal como enciclopédias, dicionários, catálogos, qualquer material de referência, manuais de treinamento, tutoriais visando ao ensino por computador, elaboração de ajuda (*help*) etc. É interessante quando se deseja que o usuário tome conhecimento gradativo do que existe disponível na base. Por exemplo, visita a um museu, a uma biblioteca, a uma exposição, onde se pode ir exibindo o mapa geral do local e ir abrindo telas ou janelas que mostrem o que existe sala a sala, arquivo a arquivo, item a item.

- e) Quando se deseja navegar exibindo essencialmente elementos de imagem e som, ou seja, mais de um sentido humano (visão e audição) —> Sistemas de multimídia.

Utilizado em apresentações, *marketing* e treinamento. A ligação entre textos de texto também pode fazer parte dos produtos de multimídia, cuja navegação é a mesma do hipertexto - pressionando-se botões da tela do micro, "pula-se" para outros trechos de texto ou imagem ou som a ele associados.

A recuperação no texto através de buscas booleanas, proximidade de termos etc., ou seja, busca pelo contexto — não costuma ser contemplada nos sistemas ditos de multimídia. Há, porém, o inverso: *software* de gerenciamento de textos e de hipertexto possuindo capacidade de fazer multimídia. O *software* de hipertexto com capacidade multimídia é denominado hipermedida.

QUANTO ÀS NECESSIDADES DE BUSCA

Quanto às necessidades de efetuar buscas ou seja, fazer perguntas, devemos considerar os seguintes aspectos no momento de escolher o produto:

- Busca por comandos — Permite o uso de operadores booleanos, operadores de proximidade no texto com indicação de intervalo de palavras, operadores de comparação etc.

- Busca orientada por *menus*—Através de opções do menu. Convém observar que o menu é útil para o iniciante, porém à medida que o usuário se torna experiente, acha cansativo ter que ir e vir em telas de *menu apenas*. Combinar opções de *menu* e comandos tem sido uma estratégia adotada com sucesso.

- Busca baseada em hipertexto — Os sistemas de hipertexto podem ser implementados de diferentes modos. É

preciso observar se corresponde às expectativas. Por exemplo: se é possível voltar ao ponto de partida, refazendo o caminho inverso; se permitem associar textos, imagens, arquivos e programas entre si; se existe, além da navegação, a possibilidade de busca por comandos.

DO PONTO DE VISTA DO EQUIPAMENTO E DO SUPORTE ATUAL DE ARMAZENAMENTO

Meio de armazenamento da maioria das informações: em computadores de grande porte, em microcomputadores; em microfimes; em papel. A situação atual de armazenamento precisa ser levada em consideração na decisão do *software* mais adequado, pois implicará necessidades distintas.

Quando a maior parte dos documentos está em papel é importante que o *software* aceite importação de arquivos digitalizados. Geração automática de índices e integração de diversas estruturas e formatos de arquivo são igualmente relevantes. Se houver documentos armazenados em computador, porém em distintas arquiteturas, é necessário um *software* que possibilite integrar os diferentes ambientes, talvez micro e grande porte até. Ou talvez seja recomendável uma conversão. Estas características precisam ser cruzadas com os parâmetros de avaliação a seguir, de forma a se estipular a pontuação de cada item.

É fundamental a priorização das necessidades, pois cada *software* existente no mercado possui um determinado conjunto de atributos que permite atender melhor a uma ou outra necessidade. Como a potência, o equipamento necessário e o preço também variam bastante em função do maior ou menor número de recursos oferecidos, o correto dimensionamento da necessidade a ser atendida é fator fundamental para a seleção adequada.

PARÂMETROS PARA AVALIAÇÃO DOS PRODUTOS

Alguns parâmetros básicos, aos quais deverão ser atribuídos pesos de acordo com a aplicação desejada, são descritos a seguir^{11,12,13,14,15}.

- Ambiente — Plataformas em que o sistema roda. Portabilidade, conectividade e compatibilidade com o ambiente de *software* e *hardware* pré-existente (DOS, UNIX, OS/2; PC, MAC, RISC etc.); Arquitetura aberta ou fechada. Redes, etc.;
- Equipamento — *Hardware* necessário e compatível. Configuração mínima necessária;

- Modulariedade — Possibilidade de crescimento à medida das necessidades;
- Armazenamento — Tipos de discos ópticos (worm, regravável), especificação do disco rígido, da memória etc.
- Características gerais — Dados de desempenho e características que facilitam e/ou limitam o desenvolvimento de aplicativos, tais como facilidades do *software*. Capacidade de acentuar, facilidade de desenvolvimento, facilidade de atualização, limites e restrições, tempo de resposta, tamanho do código objeto, *help* contextual.
- Automação do fluxo de trabalho — Possibilidade de envio e circulação de documentos via correio eletrônico e/ou fax, a partir do micro, sem precisar recorrer ao papel.
- Gerenciamento de documentos — Capacidade de armazenamento, indexação e recuperação de documentos com diferentes estruturas de dados (textos não estruturados, bases de dados estruturadas, planilhas etc.) e em diferentes ambientes; capacidade de recuperar textos, imagem, voz (multimídia) etc.;
- Recursos de indexação — Capacidade de criar índices de acesso às informações através de arquivos de palavras-chave ou ainda, de possibilitar acesso direto a todas as palavras do texto. Entre os recursos possíveis, temos a considerar:
 - acesso direto ao registro — Busca em tempo real ao conteúdo dos registros, sem uso de listas de termos.
 - Indexação pelas palavras significativas do registro — Índices gerados automaticamente, com todas as palavras do registro, excetuando-se as constantes em listas de palavras não-significativas (*stop-words*), como artigos, conjunções, preposições.
 - Indexação usando a filosofia de hipertexto — Busca por palavras destacadas na própria tela, com ligação entre os conceitos através de elos e nós.
- Consumo na indexação — Espaço de disco utilizado pelo índice, em termos do percentual do espaço utilizado pelo arquivo indexado. Obs.: Este item não é necessário, se o sistema efetua acesso direto, sem gerar listas invertidas;
- Localizarão do índice da base de dados — os sistemas de gerenciamento da imagem (arquivamento da imagem digitalizada, sem reconhecimento óptico) podem ter o índice junto ou separado da base.
- Recursos de recuperação da linguagem de busca (*Query language*). — Os recursos desejáveis para efetuar a busca são os seguintes:
 - por lógica booleana — Capacidade de efetuar a busca utilizando os operadores lógicos E, OU, NÃO, de modo a precisar melhor o que se deseja recuperar;
 - Busca no contexto — Capacidade de especificar o contexto no qual as palavras aparecem, identificando termos próximos no registro. Ou seja, indicar a vizinhança ou intervalo entre termos no campo, na frase, no parágrafo, no documento etc., recuperando palavras precedidas ou seguidas de outras.
 - Por comparação — Capacidade de permitir comparações entre conjuntos, geralmente utilizando operadores como > < = (aplicável principalmente a datas, campos numéricos).
 - Truncamento ou mascaramento — Aumento da capacidade de recuperação através do mascaramento de caracteres no início, meio ou final de uma palavra (geralmente com símbolos como #, *, ?).
 - Busca por listas de sinônimos e/ou tesouro — Facilidade de utilizar vocabulários controlados para auxiliar a consulta.
 - Capacidade de recuperar informações em diferentes formatos de arquivo — Capacidade de acesso a informações em bases de dados textuais (não estruturadas) e em bases de dados estruturadas — relacionais ou não —, em planilhas etc.
 - Capacidade de salvar buscas — Permite guardar as buscas efetuadas para reutilização posterior.
 - Sort — Possibilidade de reorganizar as saídas na tela ou em relatórios impressos.
- Navegação em nós de hipertexto — São diversas características desejáveis: capacidade de navegação na estrutura dos nós da rede de hipertexto, saltando em qualquer direção, pulando nós; percorrer de volta os nós consultados; exibir

- mapa (gráfico) de localização do usuário na rede de hipertexto; capacidade de fazer *links* e recuperar imagem de documentos digitalizados, gráficos, desenhos, som etc. (recursos multimídia); possibilidade de chamar outros programas a partir de um nó; capacidade idêntica a dos recuperadores de texto, para atender à necessidade de o usuário fazer perguntas, além de navegar.
- Digitalização — Variedade de *scanners* habilitadas (compatibilidade com *scanners* do mercado); padrão utilizado nas imagens digitalizadas de documentos (GIF, TIFF etc.).
- Reconhecimento óptico — Existência ou não de OCR integrado ao sistema. Em caso negativo, compatibilidade com *software* de OCR encontrado no mercado.
- Compressão de dados e imagens — Capacidade de compressão e descompressão da imagem digitalizada. Quanto maior a taxa de compressão, mais reduzido o custo de armazenamento e aumento de desempenho. Padrão e nível de compactação de dados e imagens. Há duas categorias de compressão: proprietária (que não permite permuta entre diferentes sistemas) e padrão, tipo CCITT 3 e 4 (que oferece menor taxa de compressão, porém suporta transferência entre sistemas).

- Facilidades de manipulação — Facilidade visualizar e ampliar a imagem; rotação; janelas (múltiplos documentos acessados simultaneamente); retro-que, etc.
- Recursos de imagem e de som — Facilidades de animação, movimento de objetos, imagens de vídeo, desenhos, fundos, *boxes* etc. Recursos de sonorização. Efeitos especiais e recursos de apresentação.
- Segurança dos dados — Níveis de acesso e autorização para alterar ou acessar as informações (senhas, criptografia).
- Preço — Preços dos produtos, por configuração; preço do *run time*; manutenção; novas versões.

Estas categorias terão peso variado, conforme o objetivo de uso. Alguns requisitos podem mesmo ser dispensáveis. Por exemplo, todos os referentes à imagem em *software* voltado exclusivamente à recuperação de bases textuais, com entrada de dados baseada em digitação. Outros terão de ser acrescentados de acordo com as especificidades. Como exemplo, em um *software* voltado a aplicações de *tutorial*, ou treinamento baseado em computador, que podem ter estrutura de armazenamento baseada em hipertexto ou hierárquica, a interação com usuário e facilidades para correção de erros, bem como pontuação final, são requisitos desejáveis. É, portanto, a análise de cada necessidade que determina os parâmetros e seus pesos, podendo estes variar conforme o tipo de aplicação.

Requests for text and image management and retrieval

Abstract

Presents a categorization for products designed for text and image management and retrieval, including classical information retrieval software and modern text management, hypertext/hypermedia, document image management and workflow software. Criterion to specify the necessary resources must consider the goals, the environment of the system or service, the information content, the current storage media, and the present and potential uses. Several criterion for evaluation of these kind of software are suggested.

Keywords

Information storage and retrieval; Text retrieval; Text management; Document image system (dip); Hypertext; Hypermedia; Software evaluation.

SCHWABE, Daniel, BERNSTEIN, Mark. Tools for designing hyperdocuments. In: BERK, Emily Devlin, Joseph. (Ed.) *Hypertext/Hypermedia handbook*. McGraw-Hill, Chapter 13, 1991. p.179-207

10. WHITE, Ron, BONNER, Paul. Information overload. *PC Computing*: p. 70-90, Aug. 1990
11. DOMINICK, Wayne D. A performance measurement and evaluation environment for information systems. *Inf. Processing and Management*, v. 23, n.1 p. 7-15, 1987.
12. BUSH, Earlene. Search and Retrieval; How to evaluate large text-retrieval systems. *Byte*: p. 271 -276, June 1992.
13. CIN/CNEN, CNPq/IBICT. *Proposta para a linguagem comum de recuperação de informações*. Brasília, mar. 1989.
14. BELKIN, Nicholas J., CROFT, W. Bruce. Retrieval techniques. *Annual Review of Information Science and Technology*, n. 22, p. 109-143, 1987.
15. SALTON, Gerard. Thoughts about modern retrieval technologies. *Information Services & Use*, n. 8, p. 107-113, 1988.
16. BENDER, Avi. Application of a full-text retrieval system for records management. *Journal of Information & Image Management*: p.19-24, p.36, April 1986.
17. EARL, Lois L. Experiments in automatic extracting and indexing. *Inform. Storage and Retrieval*, n. 6, p. 313-334, 1970.
18. HARVEY, David A. Cath the wave of DIP. *Byte*: p. 173-182, April, 1991.
19. HARVEY, David. A., BOB. Ryan. Practically paperless. *Byte*: p. 185-190, April 1991.
20. KAUFMANN, Todd., MONARCH, Ira. *Knowledge-based indexing and hypertext*. Carnegie-Mellon University. Computer Science Department, July 1988.

21. LOCKE, Christopher. The dark side of DIP. *Byte*: p. 193-204, April, 1991.
22. MEGHINI, Carlo, RABITTI, Fusto THANOS, Constantino. Conceptual modeling of multimedia documents. *IEEE Computer*: p. 23-30, Oct 1991.
23. REYNOLDS, Louis R., DEROSE, Steven J. Electronic Books. *Byte*: p. 263-268, June 1992.
24. SWANSON, Rowena Weiss, MAYER, Joseph. Performing evaluation studies in Information Science. *Journal of the American Society for Information Science*: p. 58-73, May-June, 1975.

Artigo aceito para publicação em 15 de setembro de 1994

Regina de Barros Cianconi

Bacharel em biblioteconomia e documentação pela UFF. Mestre em Ciência da Informação pela UFRJ/ CNPq/IBICT. Analista de informática do Serpro-Filial Rio de Janeiro, especialista em sistemas de gerenciamento e recuperação da informação e hipertexto/hipermídia. Atua na área de novas tecnologias, na avaliação, teste e propostas de solução para gerenciamento de textos e imagens.

REFERÊNCIAS BIBLIOGRÁFICAS

1. CÔTÉ, Raymond Ga & DIEHL, Stanford. *Searching for common threads*. *Byte*, v.17, n.6, p.229-305, 1992.
2. CROFT, Bruce. TURTLE, Howard. A retrieval model for incorporating hypertext links. In: *Hypertext'89 Proceedings*, p. 213-224, Nov. 1989.
3. SALTON, Gerard, ALLAN, James. Selective utilization and text traversal. In *Hypertext'89 Proceedings*, p. 131-144, Nov. 1989.
4. CONWAY, Paul. Digital imaging and optical storage over the long run. *INFORM*; p. 31-36, Nov./Dec. 1992.
5. AVEDON, DON M. *Discos ópticos e imagens eletrônicas*. São Paulo: CENADEM, 1991.
6. HARPOLD, Terence. *Hypertext and hypermedia: a selected bibliography*. Brown Univ. Serv., Dennis Short, Purdue University, 1988.
7. D'IPOLITTO, Claudio. *Hipertexto: uma visão geral*. Rio de Janeiro, COPPE/ UFRJ, jul. 1989.
8. MARTIN, James. *Hiper documentos e como criá-los*. Rio de Janeiro; Campus, 1992.
9. GARZOTTO, Franca, PAOLINI, Paolo,