

**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
ESCOLA DE COMUNICAÇÃO**

SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO (SRI) /

SUBSISTEMA DE ENTRADA:

UMA SISTEMATIZAÇÃO DAS

ALTERNATIVAS DE CONSTRUÇÃO

Claudio Ribeiro da Silva

Rio de Janeiro

1996

**SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO (SRI) /
SUBSISTEMA DE ENTRADA:
UMA SISTEMATIZAÇÃO DAS
ALTERNATIVAS DE CONSTRUÇÃO**

Claudio Ribeiro da Silva

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, Convênio CNPq/IBICT-UFRJ-ECO, para obtenção do grau de Mestre em Ciência da Informação.

Orientadora: Gilda Maria Braga, PhD.

Rio de Janeiro

1996

SILVA, Claudio Ribeiro da. Sistemas de Recuperação de Informação (SRI)/ Subsistema de Entrada: uma sistematização das alternativas de construção. Orient.: Gilda Maria Braga. Rio de Janeiro, 1996. 129 p. Dissertação de Mestrado em Ciência da Informação. CNPq/IBICT-UFRJ-ECO.

RESUMO:

Este trabalho sistematiza algumas alternativas, das etapas do subsistema de entrada, necessárias para a construção de um SRI. Avalia o comportamento da literatura em fornecer os conhecimentos necessários para esta construção.

ABSTRACT:

The alternatives of each phase on input's subsystem for an Information Retrieval System Project are described. The literature's behaviour to support this objective is evaluated.

Aos colegas da Superintendência de Informática (SIN), que muito me ajudaram nas diversas etapas deste trabalho.

Aos professores do IBICT, em especial à Gilda Braga, pela atenção dada.

Aos meus pais, por tudo que sou.

À minha esposa e filhas, pela minha ausência.

À Deus, por tudo que tenho.

SUMÁRIO

1 - INTRODUÇÃO	1
2 - OBJETIVO	3
3 - METODOLOGIA.....	4
4 - SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO: Evolução histórica.	5
5 - SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO: Categorias existentes	9
6 - SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO: Etapas de Construção	16
6.1 - PLANEJAMENTO	16
6.2 - COLETA DE DOCUMENTOS	23
6.3 - ENTRADA DE DADOS.....	33
6.4 - CATALOGAÇÃO	39
6.5 - INDEXAÇÃO	47
6.6 - ARMAZENAMENTO	63
7 - CONCLUSÕES: TENDÊNCIAS	81
8 - BIBLIOGRAFIA CONSULTADA	85

1 - INTRODUÇÃO

Os registros do conhecimento humano em documentos surgiram da necessidade do homem em repassar a sua descoberta aos seus semelhantes. Os primeiros registros datam da pré-história onde os nossos ancestrais deixaram as suas atividades (caça, guerra, etc.) e descobertas (fogo, roda, etc.) gravadas nas cavernas como forma de comunicá-las e difundi-las.

Com a evolução da civilização novas formas de registros foram desenvolvidas com o objetivo de organizar os documentos e permitir melhor a sua divulgação. Dentre estas formas destacamos o desenvolvimento do Sistema de Recuperação de Informações (SRI), uma das principais ferramentas utilizadas pelos pesquisadores para consultar e disseminar informações nos dias de hoje. Através dele é possível atender às exigências de disseminação em razão da:

- Quantidade de documentos publicados, com taxas de crescimento cada vez maiores;
- Necessidade de divulgar as informações, em intervalos de tempo cada vez menores entre a publicação e a disseminação;

O planejamento de um SRI deve definir as técnicas e rotinas que serão utilizadas nas etapas de construção e operação do sistema, identificando as entradas, recursos, restrições, limites, meio-ambiente e saídas do sistema, e eliminando prováveis problemas operacionais. As rápidas inovações tecnológicas criam novos mecanismos e alternativas de desenvolvimento e uso dos sistemas e, em alguns casos, rapidamente são aceitas pelos usuários, exercendo forte influência no produto final. Portanto, é necessário conhecer a literatura atualizada, filtrar todos os seus problemas e “garimpar” as respostas dos itens que irão compor o projeto.

Utilizando informações obtidas na literatura, este trabalho procura sistematizar as etapas e alternativas existentes, no subsistema de entrada, para construir um SRI, sendo por este motivo, muito mais do que uma revisão bibliográfica.

Foram considerados os documentos publicados nas literaturas LISOS (Library & Information Science Abstracts), ISA (Information Science Abstracts), ARIST (Annual Review of Information Science and Technology) e no serviço de busca retrospectiva SUPRIR (disponível no Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear) a partir de 1980, sendo estas restrições feitas, período analisado e subsistema de entrada, em razão do grande volume de publicações existentes e utilizadas para a pesquisa.

Embora sejam importantes e necessárias para um projeto de sistema, algumas tarefas não foram analisadas de forma detalhada, tais como: estudo de viabilidade, projeto de arquivos, estudo de usuários, definição de área, etc., por serem específicas de cada sistema não cabendo, portanto, uma avaliação genérica.

A literatura é imprecisa quanto à utilização de alguns termos, sendo eles utilizados como sinônimos e com objetivos diferentes. Para esclarecer eventuais dúvidas no contexto deste trabalho os termos abaixo possuem as seguintes definições:

- Documento - Representa o conteúdo completo do texto.
- Informação - Representa o valor que o texto possui para alterar a estrutura cognitiva de quem o utiliza. Este valor é atribuído pelo usuário que utiliza o texto.
- Etapa - Uma das fases do projeto de construção e elaboração de um sistema contendo um conjunto de processos a serem realizados no sistema (indexação, coleta, formas de armazenamento, etc.).

- Alternativa - Define as técnicas existentes em cada etapa. Exemplo: na etapa de indexação existem as alternativas: indexação automática, indexação utilizando tesauro, etc.
- Base de Dados - Representa o(s) arquivo(s) onde estão armazenados os dados no computador.
- SRI - Representa um conjunto de tarefas, serviços ou procedimentos que coletam, organizam, gerenciam e armazenam documentos, permitindo a sua disseminação.

O presente trabalho está dividido em oito seções. Após esta breve Introdução, seguem-se: Objetivo (2), Metodologia (3) e as etapas relativas ao SRI (4 a 6). As conclusões, apontando as tendências da literatura formam a seção 7, e as referências bibliográficas a seção 8. Ao final de cada seção estão listados os documentos nela citados.

2 - OBJETIVO

Este trabalho tem como objetivo obter na literatura as alternativas existentes em cada etapa da construção de um Sistema de Recuperação de Informações (SRI), no que se refere aos subsistemas de entrada, criando o roteiro necessário para um projetista de SRI.

3 - METODOLOGIA

Objetivando analisar a literatura e sistematizar as alternativas existentes para construção de um SRI, foram realizados os seguintes procedimentos:

- Definição das etapas de construção de um SRI - Identificação das etapas existentes para a criação e operação de um SRI, especificamente do subsistema de entrada.
- Seleção dos documentos na literatura - Identificação dos documentos que abordam as alternativas e etapas existentes no subsistema de entrada, classificando-os, preliminarmente, em cada uma delas. Conforme dito anteriormente, foram considerados para seleção os documentos publicados nas literaturas LISOS (Library & Information Science Abstracts), ISA (Information Science Abstracts), ARIST (Annual Review of Information Science and Technology) e no serviço de busca retrospectiva SUPRIR (disponível no Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear) a partir de 1980.
- Coleta dos documentos selecionados - Obtenção dos documentos selecionados, através do serviço de busca de cópia de documentos SERVIR, disponível no Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear. Foram obtidos, no Brasil e no Exterior, aproximadamente 90% dos documentos selecionados. Parte dos documentos não obtidos se deu, principalmente, pela referência publicada estar incompleta ou errada, não sendo possível a sua localização.
- Leitura dos documentos coletados - Leitura dos textos dos documentos coletados, reclassificando-os, eventualmente, em uma nova etapa ou alternativa.
Durante a leitura foram observados:
 - alteração na classificação de alguns documentos, decorrentes das diferentes abordagens existentes entre o resumo e o texto;
 - avaliação dos documentos eliminando os que tratavam de assuntos fora do escopo deste trabalho ou de forma superficial.
- Análise do conteúdo lido - Sistematização do conteúdo dos documentos criticando-os em relação à sua capacidade de fornecer as informações necessárias para o projeto do sistema.

4 - SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO: Evolução histórica.

Desde a década de 60, os Sistemas de Recuperação de Informações (SRI) vêm sendo mais intensamente objeto de estudos e implementações visando atender às necessidades de conhecimentos das pessoas que o utilizam. Um SRI é um conjunto de tarefas, serviços ou procedimentos que possibilitam a organização e tratamento dos documentos armazenados em computador. Dentre alguns objetivos para o seu uso, podemos relacionar:

- a eliminação da duplicidade de pesquisas e desenvolvimento;
- a promoção e transferência de novas idéias, serviços e tecnologias;
- a identificação das informações básicas necessárias para auxiliar um planejamento de curto, médio e longo prazo;
- auxílio no aprendizado e nas atividades educacionais e de lazer;
- identificação e fornecimento de material documental para a solução de problemas em várias áreas.

A necessidade de repensar a organização tradicional das informações surgiu no final da Segunda Guerra Mundial. O elevado volume de dados científicos gerados e a necessidade de divulgá-los para fins pacíficos, permitiram que alternativas de pesquisas mais simples e rápidas fossem criadas. Surgiram as primeiras bases de dados bibliográficas, que utilizavam as primeiras versões dos índices KWIC e foram incorporados às referências bibliográficas os resumos resultantes dos serviços de resumo e indexação. Inicialmente, estes sistemas permitiam pesquisas não interativas (batch) em bases de dados armazenadas em fitas magnéticas, sendo estes serviços caros e insatisfatórios aos usuários. Anterior a estas bases, existia apenas as bases numéricas, criadas a partir de 1951, conforme afirmação de **NEUFELD**⁴.

No final da década de 50, diversas pesquisas foram feitas com o objetivo de criar sistemas que permitissem acesso on-line. Estes sistemas começaram a surgir já no início da década de 70, juntamente com outras evoluções tecnológicas, sendo a principal delas o armazenamento dos dados em discos magnéticos. Estas evoluções permitiram que os índices KWIC e derivados: KWOC e KWAC fornecessem os resultados das consultas dos usuários no momento da sua realização utilizando diversas formas de acessos tais como: palavras do título ou texto, índice de assuntos, etc., possibilitando o uso mais fácil dos sistemas.

Na década de 70 houve uma grande explosão de criação de SRI's. Diversos sistemas foram construídos, na grande maioria sistemas on-line, envolvendo áreas científicas e sociais, dentre outras. Surgiram os provedores de base de dados, que fornecem todos os serviços de armazenamento e acesso para consulta aos dados por eles gerenciados, e os profissionais especialistas que orientam os usuários na realização de suas consultas. Por serem interativas, estas consultas eram alteradas e refeitas quantas vezes fossem necessárias até atender as necessidades do usuário. Nesta época, as técnicas de recuperação e avaliação dos sistemas se tornaram assuntos importantes na literatura.

Na segunda metade da década de 70 o uso e a importância dos SRI's já estavam consolidados. Novas áreas (comercial, industrial, financeira, etc.) começaram a desenvolver os seus próprios sistemas surgindo outras categorias, notadamente os sistemas numéricos, na área de negócios, e os de texto completo. Novos usuários foram treinados em diversos recursos tais

como: possibilidade de atualização dos dados diretamente na base de dados, serviços de edição, “spelling checking” (verificação ortográfica das palavras), indexação on-line, etc.

Os interesses comerciais começaram a obrigar os produtores de bases de dados comerciais a criarem departamentos de marketing para divulgar os seus sistemas. Os aspectos legais da construção de sistemas para divulgação dos documentos começaram a ser discutidos. Itens como copyright, direitos autorais, controle de cópias, royalties, etc. começaram a ser questionado. A informação passou a ser considerada nos seus aspectos:

- social - como fonte de geração de bem estar;
- econômico - como mercadoria que possui o seu valor comercial;
- psicológico - como recurso que pode alterar a forma de pensar.

Explodiram também as publicações secundárias que consolidavam em um único volume as referências dos documentos publicados em diversas fontes.

No início dos anos 80 ocorreu outro elevado número de criação de SRI's por áreas do conhecimento. Segundo **NEUFELD**⁴ esta quantidade subiu de 300 em 1975 para 600 em 1980 (crescimento de 100% em 5 anos) e 2400 em 1984 (crescimento de 300% em 4 anos). A capacidade de armazenamento e processamento cada vez maiores dos computadores permitiram tornar disponível quantidade maior de dados possibilitando aos pesquisadores aumentar os seus conhecimentos gerando evoluções tecnológicas mais rápidas.

Com o surgimento dos computadores pessoais, diversas pessoas puderam acessar os SRI's em seus próprios locais de trabalho. Este fato obrigou as interfaces dos sistemas a serem mais simples com a implantação de diversos recursos tais como: telas de ajuda, teclas de atalho, consultas orientadas e outros recursos de programação, fazendo com que os usuários utilizassem os sistemas com menos dificuldades.

Na segunda metade da década de 80 surgiram duas tecnologias que viriam a revolucionar todo o uso dos SRI's: a tecnologia ótica e a Internet.

A primeira, iniciada com a criação do equipamento scanner e o desenvolvimento do software de Optical Character Recognition (OCR), era capaz de armazenar no computador o documento no seu formato original com imagens e texto. Estas inovações geraram transformações na construção de SRI's e permitiram surgimento de categorias que manipulavam imagens e outros recursos tornando mais atraente o uso. Com a evolução tecnológica surgiram equipamentos óticos que permitiram capacidade maior de disponibilização dos dados por utilizarem a tecnologia do laser para armazenar os documentos.

Com a Internet, os usuários em seus locais de trabalho ou residências consultam nos SRI's informações armazenadas em qualquer lugar do mundo. Estes sistemas utilizam bases de dados armazenadas em equipamentos interligados de forma que os acessos são transparentes a quem os utiliza. Na rede Internet existem sistemas de várias categorias, tais como: imagens, textos completos, som, etc., sendo os dois primeiros os mais numerosos.

Com as evoluções citadas acima, o uso e a quantidade de dados disponíveis cresceram em percentuais cada vez maiores, conforme quadro abaixo montado por **WILLIAMS**⁹, que mostra os números de formação e uso de base de dados no período de 1975 a 1991:

Base de dados: de 301 a 7637 (crescimento de aproximadamente 2500%)

Produtores: de 200 a 2372 (crescimento de aproximadamente 1200%)

Pesquisas: 800.000 a 34.5 milhões por ano (crescimento de aproximadamente 4300%)

Vendedores: 105 a 933 (crescimento de aproximadamente 900%)

Custo por pesquisa: US\$ 14,80 a US\$ 9,73 (redução de aproximadamente 35%)

Registros: 52 milhões a 4.060 bilhões (crescimento de aproximadamente 7800%)

Horas de conexão: 780.000 a 4.657.000 por ano (crescimento de aproximadamente 600%)

Rendimentos: US\$ 40 milhões a US\$ 690 milhões por ano (crescimento aproximado 1700%)

Todos estes números e, principalmente, percentuais mostram o crescimento no armazenamento e utilização dos dados neste período. Cabe acrescentar que em 1991 ainda não havia explodido de forma significativa o uso da Internet, fazendo acreditar que os percentuais de crescimento sejam atualmente ainda maiores. Informações mais atualizadas sobre estes percentuais não foram encontradas na bibliografia consultada.

Nos dias de hoje, com a evolução tecnológica cada vez mais dinâmica, surgem em intervalos de tempo cada vez menores equipamentos com capacidade de processamento e armazenamento de dados cada vez maiores, aumentando cada vez mais as taxas de crescimento e utilização de SRI's. Estes fatores tornam-os ferramentas indispensáveis para as rotinas de busca de informação.

Referências Bibliográficas:

- 1 CLOUGH, C. R. The Changing Role of Retrieval System and Specialists. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 10, London, 1986. **Proceedings...** London: 1986. p. 175-184.
- 2 FOSTER, Allan. External Databases: An Overview. **Aslib Proceedings**, London, v. 35, n .9, p. 346-353, sep 1983.
- 3 HUGHES, Graeme C. The Information Age. **Information Development**, v. 7, n. 2, p. 72-74, apr 1991.
- 4 NEUFELD, M. Lynne, CORNOG, Martha. Database History: From Dinosaurs to Compact Discs. **Journal of the American Society for Information Science**, New York, v. 37, n. 4, p. 183-190, jul 1986.
- 5 PEREIRA, Vânia Lúcia da Cunha. **Sistema de Redução da Informação: Uma (Ir)recuperação Metodológica**. Rio de Janeiro: UFRJ/IBICT, 1994. 95 p. Dissertação (Mestrado em Ciência da Informação), Escola de Comunicação, Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, 1994.
- 6 SAKSIDA, Marino. Information - A Vision of the Future: The Human Dimension. **Journal of Information Science**, Holborn, v. 18, n. 5, p. 323-327, 1992.
- 7 TRAUE, J. E. Five Thousand years of the Information Society. **New Zealand Libraries**, v. 46, n. 6, p. 17-18, sep 1990.
- 8 WARRO, E. A. Watt have we been signing?: a look at database licensing agreements. **Library Administration and Management.**, v. 8, n. 3, p. 173-177, summer 1994.
- 9 WILLIAMS, Martha E. Highlights of the On-line Database Industry. In: NATIONAL ON-LINE MEETING, 13, 1992. **Proceedings...** Medford: Martha E. Williams, 1992. p. 1-14.

5 - SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO: Categorias existentes

A categoria de um SRI está relacionada ao tipo e conteúdo dos documentos armazenados na base de dados. Em razão de avanços tecnológicos ocorridos, algumas categorias surgiram e outras deixaram de serem utilizadas. Estes avanços permitiram o armazenamento dos documentos de forma cada vez mais fiel ao original.

São poucos os artigos que falam de forma sistematizada sobre o assunto. Dos analisados, **MARTYN**²² e **CUADRA**¹⁰ apenas as relacionam, mas somente **MARTYN**²² tem como objetivo discuti-las. Entretanto, as categorias surgidas mais recentemente (imagem, som, etc.) não são analisadas por este artigo ter sido publicado há mais de dez anos. Os demais comentam sobre projetos ou características específicas de SRI's algumas vezes de forma sucinta.

Algumas categorias são encontradas, tais como em **SUOZZI**³¹ que divide os sistemas em dois grupos:

- Bibliográficos - Fornecem o local onde os documentos podem ser obtidos;
- Não bibliográficos - Fornecem o conteúdo do documento.

Esta divisão deve ser considerada em um primeiro nível de sistematização. Dentre as categorias relacionadas existem as que indicam onde os documentos podem ser obtidos (considerada como Bibliográficas por **SUOZZI**³¹) e as que contêm o documento completo (considerada como Não Bibliográficas). Entretanto, **MARTYN**²² e **CUADRA**¹⁰ categorizam os SRI's pela forma com que armazenam os documentos ou pelo seu conteúdo e não pela possibilidade de torná-lo disponível. Tomando por base estes critérios, encontramos as seguintes categorias:

5.1 - Referenciais

Sistemas que possuem como conteúdo nas suas bases de dados apenas as referências bibliográficas dos documentos (título, autor, forma de publicação, local de publicação, etc.) e o(s) local(is) onde são obtidos.

Estes sistemas são encontrados raramente na literatura atual. Isto se explica por dois motivos:

- serem bastante antigos - pela tecnologia disponível os primeiros SRI's eram Referenciais, tendo ocupado o seu espaço no passado. Hoje se pode considerar que os conhecimentos sobre eles já estão sedimentados.
- pela necessidade de atualidade - a literatura ocupa o espaço abordando as novas categorias que estão sendo utilizadas atualmente.

Estes sistemas são encontrados normalmente em catálogos de bibliotecas que utilizam como forma de acesso as informações contidas na própria referência (palavras do título, autor, local e ano de publicação, etc.).

5.2 - Bibliográficos

Sistemas que armazenam nas suas bases de dados as referências bibliográficas e os locais onde os documentos podem ser obtidos, acrescidos dos serviços de indexação e resumo.

São poucos os artigos que discutem esta categoria. As maiorias dos documentos encontrados estão fora do escopo deste trabalho, por serem anteriores a 1980. Os motivos para esta ausência são os mesmos dos sistemas referenciais. São encontrados com frequência apenas artigos que discutem as formas de indexar documentos tais como: análise de assunto e tesauro (principais tópicos de indexação), sendo objeto de algumas revisões no ARIST.

Estes sistemas são exemplificados por catálogos de bibliotecas (documentos existentes em uma biblioteca) e pelos serviços de referência (indicação do local onde os documentos podem ser localizados), que utilizam os assuntos dos documentos como forma de acesso. Estes catálogos e serviços são encontrados na Internet.

5.3 - Texto Completo ou Textuais

Sistemas que gerenciam o armazenamento e recuperação do texto completo dos documentos nas suas bases de dados. Nas categorias anteriores o processo de entrada de dados mais utilizado é a digitação das referências. Entretanto, para os textos completos os mais utilizados são as entradas via Optical Character Recognition (OCR) e as transferências eletrônicas. Em razão dos recursos tecnológicos necessários estarem mais acessíveis pela redução dos custos de aquisição e manutenção, diversos sistemas surgiram nesta categoria nos últimos anos. Por este motivo, os artigos que os abordam dão ênfase aos aspectos da sua formação. Como exemplo, podemos citar o artigo de **MARTINSEM**²¹ que comenta sobre o processo de formatação de textos para armazenamento na base de dados e posterior disponibilização em serviços on-line.

Os artigos sobre OCR's, que serão vistos no capítulo sobre Entrada de Dados, dão ênfase aos processos de cópia, validação e armazenamento dos documentos na base de dados.

As transferências eletrônicas ocorrem através dos recursos de telecomunicações, tais como: fax, correio eletrônico e, principalmente, a Internet, sendo encontrados artigos sobre estes itens apenas nas literaturas de comunicação de dados e Internet.

5.4 - Texto Completo / Imagens

Armazenam nas suas bases de dados o conteúdo completo dos documentos que contém texto, imagens ou outros caracteres não representados no teclado comum, tais como: alfabeto grego, símbolos químicos, etc., além das diferentes fontes e tamanhos das letras, recursos disponíveis nos editores de textos existentes no mercado.

Os artigos que abordam esta categoria de sistema não são encontrados com frequência. Eles o fazem apenas em aspectos específicos da aplicação, como **BROOKS**⁴, que trata o projeto de representação, armazenamento e visualização dos dados, discutindo o relacionamento entre eles e propondo soluções para a representação de símbolos químicos.

Por armazenarem textos e imagens, os artigos são confundidos com os textos sobre sistemas que armazenam apenas imagens. Entretanto, o tratamento de cada um deles é diferenciado, principalmente no processo de indexação.

Este tipo de sistema é encontrado, normalmente, em aplicações jornalísticas, químicas, etc.

5.5 - Imagens

Esta categoria de sistema gerencia de forma integrada o armazenamento e tratamento de documentos representados por fotos e pelas informações a elas associadas.

Com o surgimento dos equipamentos de OCR's foi possível transformar as características de uma imagem (cor, forma, etc.) em caracteres ASCII (padrão de armazenamento em computadores). Este fato tornou esta a categoria mais utilizada. Por utilizar novas tecnologias, possui grande espaço na literatura com várias revistas especializadas sobre o assunto. Dos artigos analisados sobre categorias de SRT's, aproximadamente 40% falam sobre imagens e exploram desde os critérios para a seleção de um sistema de gerenciamento de imagens, publicado por **O'CONNOR** ²⁵, até projeto de sistemas, publicados por **CAWBELL** ⁷ e **HIBLER** ¹⁵. Entretanto, a ênfase maior é dada aos que revêem conceitos, arquiteturas, recursos, equipamentos, etc., sendo a indexação o item explorado com maior frequência.

Esta grande quantidade de artigos sobre o assunto ocorre pelo fato de não existir uma definição exata do termo, sendo utilizada desde “novos métodos de compressão de dados para armazenamento eficiente de imagens, esquemas de estrutura de dados hierárquico para armazenamento de imagens de grande complexidade, processos de avaliação de recuperação de base de dados de imagens, etc.”, conforme afirmação de **LUNIN** ¹⁹, até discussões sobre alternativas de indexação, ou seja: todos os aspectos de tratamento de imagens são considerados com o título de “Processamento de Imagens”, não existindo uma separação entre os aspectos tecnológicos e de tratamento do documento, sendo encontradas diversas discussões sobre tecnologias óticas, inclusive equipamentos.

Estes sistemas são utilizados em diversas áreas de conhecimento, tais como: educação, turismo, artes, treinamento, etc., ou seja, as maiorias das áreas atualmente fazem uso dos recursos tecnológicos disponíveis, desenvolvendo sistemas que facilitam e tornam mais agradável o uso pelos usuários.

5.6 - Numéricas

Sistemas que gerenciam e armazenam como conteúdo: números, dados numéricos ou estatísticos. Estes conteúdos podem ser contados, medidos ou tratados de diferentes formas de manipulação numérica, através de dimensionamento escalar ou exato. **SOERGEL** ²⁹ afirma que uma base de dados numérica é uma coleção de registros cujos campos definem um ou mais valores que podem ser medidos através da dimensão, escala ou pelo valor exato do número.

São poucos os artigos que abordam este assunto. Após avaliação, observa-se a escassez de conteúdo sobre projetos de construção de sistemas numéricos e estatísticos. Os motivos para esta ausência são os mesmos dos sistemas referenciais, ou seja, literatura antiga com conhecimentos já consolidados.

A utilização destes sistemas está associada às aplicações gerenciais, que mantêm armazenados valores utilizados no dia-a-dia das empresas, tais como: valores de vendas, custos, cotações, etc.

5.7 - Numéricas / Textuais

Sistemas que gerenciam o armazenamento e uso de dados numéricos associados a texto. As características são semelhantes aos sistemas numéricos, acrescidas de informações textuais que as complementam.

Nos artigos pesquisados foram identificadas avaliações e comentários sobre sistemas numéricos e textuais de forma isolada, este último em quantidade maior, não sendo encontrado observações sobre a integração destes em um único sistema.

A sua utilização, assim como a de sistemas numéricos, está incluída no contexto das aplicações gerenciais, por conterem informações atualizadas necessárias às tomadas de decisões.

5.8 - Software

Sistemas que gerenciam o armazenamento de programas de computadores (software) nas bases de dados. Os SRTs possibilitam o acesso, recuperação, utilização e transferência para o computador do usuário dos programas armazenados. Estes programas, considerados de domínio público, também chamado de shareware, não necessitam de qualquer pagamento à título de aquisição ou direito autoral para uso. Segundo **WILLIAMS**³⁴, em 1993 apenas 3% das bases de dados disponíveis na Internet armazenavam programas de computadores, mas atualmente este percentual é muito maior.

Surgidos recentemente na literatura, em razão das disponibilidades de acesso pela Internet às bases de dados, os artigos que tratam deste assunto estão ligados àquela rede. No grande volume de literatura e nas várias revistas especializadas existentes sobre o assunto os itens mais freqüentes são:

- construção de Sistemas de Informações (WWW e Mosaico);
- construções e montagens de redes de comunicações (BBS);
- uso da rede (Manual de Usuário).

Para realização deste trabalho não foram localizados artigos que abordam projetos de sistemas fora do contexto da Internet.

5.9 - Áudio

Sistemas que gerenciam e armazenam bases de dados que possuem conteúdos sonoros.

Sua utilização é restrita, pois nem sempre a informação armazenada é representada apenas por som. Encontramos com freqüência sistemas que apresentam as informações utilizando recursos de multimídia, no qual o som é um dos elementos que os compõem. Por este motivo são encontrados artigos na literatura sobre multimídia, ficando o projeto de sistema onde o som é a informação, pouco explorado.

Dos artigos analisados sobre sistemas de áudio, encontramos **BOSTON**³, que comenta sobre a ausência na literatura deste tipo de assunto e da dificuldade de transferir arquivos com este conteúdo armazenado em dispositivos magnéticos. Além disto, ele relata como seria um sistema sonoro totalmente automatizado. Em outro artigo analisado, **UDARD**²⁷ propõe os critérios que devem ser considerados para a seleção de um sistema de acesso às informações sonoras, armazenadas em LP, Cassetes, CD, etc.

Estes sistemas são encontrados em aplicações musicais e utilizam como forma de acesso informações específicas das músicas, tais como: autor, época, tipo, equipamento, etc.

5.10 - Combinação de tipos

Neste tipo de sistema podemos identificar aqueles que utilizam recursos de multimídia, combinando em um único ambiente diferentes tipos de dados, tais como: som, imagem, texto, etc. Segundo **MALLETT**²⁰, um sistema de multimídia requer a combinação de técnicas e tecnologias com diferentes formas de gerenciamento de bases de dados. Para **ARENTS**¹ um sistema de multimídia pode ser considerado como tendo três faces: multimídia como sistema (armazenando uma base de dados), como interface e como ambiente.

São muitos os artigos que discutem os projetos, problemas e tendências de multimídia, havendo tópicos específicos em publicações secundárias e revistas especializadas. Atualmente, são os sistemas mais utilizados e de crescimento mais acentuado pelos seguintes fatores:

- Interesses financeiros, principalmente os da indústria de equipamentos, conforme afirmação de **FELDMAN**¹³, havendo redução nos preços de equipamentos em razão dos avanços tecnológicos;
- Surgimento de vários modelos de gerenciamento de informação, integrando várias tecnologias, conforme afirmação de **BULICK**⁵.

Dentre os recursos de multimídia existentes atualmente, citamos como tendências o videodisco, ocupando espaço crescente de discussão na literatura, conforme artigos de **CHEN**⁸, **HILL**¹⁶ e **OBERHAUSER**²⁶ e o processamento da voz, conforme **NASH**²⁴, **ARNBAK**² e **PHILIP**²⁸.

Sintetizando as categorias relacionadas acima podemos montar o quadro abaixo:

Quadro A - Comportamento da literatura com relação às categorias de SRI.

CATEGORIA	CONTEÚDO DA BASE DE DADOS	COMPORTAMENTO DA LITERATURA
Referenciais	Referências Bibliográficas e localização	Pouco frequentes na literatura analisada. Conhecimentos consolidados.
Bibliográficas	Referências Bibliográficas com localização, resumo e indexação	Ênfase dada apenas às alternativas de indexação.
Texto Completo	Textos completos dos documentos	Ênfase dada às formas de entrada dos dados.
Texto Completo/ Imagens	Textos completos dos documentos com imagens e caracteres especiais	Incluído com os artigos que tratam de processamento de imagens.
Imagens	Fotos	Abordam todas as técnicas de processamento de imagem, com pouca ênfase à construção de SRI.
Numéricas	Números, Dados Numéricos e Estatísticos	Pouco frequentes na literatura analisada. Conhecimentos consolidados.
Numéricas/Textuais	Números, Dados Numéricos e Estatísticos com textos a eles associados	Pouco frequentes, apesar do uso em sistemas gerenciais.
Software	Programas de computadores	Pouco frequentes. Literatura englobada com a da Internet.
Áudio	Sons	Pouco frequentes. Literatura misturada com multimídia.
Combinação de tipos	Sons, Imagens e Textos	Grande número de artigos que tratam de multimídia.

Referências Bibliográficas:

- 1 ARENTS, Hans C., BOGAERTS, Walter F. L. The Three Faces of Hypermedia In: INTERNATIONAL ON-LINE INFORMATION MEETING, 15, 1991, London. **Proceedings...** London: David I Raitt, Oxford , 1991. p. 253-262.
- 2 ARNBAK, Jens C. Many Voices, One Structure: The Challenge of Telematics. **Information Society**, v. 5, n. 2, p. 101-118, 1987.
- 3 BOSTON, George. The Implications of “New Technology” for Sound Archives. **IASA Journal**. v. 3, p. 68-70, may 94.
- 4 BROOKS, Kristina M., SLADE, Rodney A. Desing and Performance Considerations in the Development of Text/image Data Bases. In: NATIONAL ON-LINE MEETING, 1989, New York. **Proceedings...** New York, Learned Information, 1989. p. 57-61.
- 5 BULICK, Stephen. Future Prospects for Network-based Multimedia Information Retrieval. In: INTERNATIONAL ON-LINE INFORMATION, 15, 1991. London. **Proceedings...** Oxford, David I Raitt, 1991. p. 517-520.
- 6 CAWBELL, A. E. Imaging Systems and Picture Collection Management: a Review. **Information Services & Use**, Amsterdam, v. 12, n. 4, p. 301-325, 1992.
- 7 CAWBELL, A. E. Selected Aspects of Image Processing and Management: Review and Future Prospects. **Journal of Information Science**, London, v. 18, n. 3, p. 179-192, 1992.
- 8 CHEN, Ching-Chih. Analog, Digital and Multimedia: Implications for Information Access. In: ON-LINE INFORMATION, 15, 1991. London. **Proceedings...** Oxford: David I Raitt, 1991. p. 238-291.
- 9 CROW, Linda. Shelf Arrangement Systems for Sound Recordings: Survey of American Academic Music Libraries. **Technical Services Quarterly**, v. 8, n. 4, p. 1-24, 1991.
- 10 CUADRA, Carlos A. Portable Databases: Bridging the Gap Between On-line and Local Databases. **Information Today**, v. 8, n. 1, p. 15-16, jan 1991.
- 11 EAKIN, Diane R., HASSON, Marsha A. Coping with Data Quality in Scientific Data Banks. In: NATIONAL ON-LINE MEETING, 4, 1983, New York. **Proceedings...** Medford: Martha E. Williams and Thomas H. Hogan, 1983. p. 107-117.
- 12 EAST, Harry, VOGEL, Sandra. Indicators and Revenues of Database Production and Hosting in the United Kingdom. **Journal of Information Science**, v. 18, n. 4, p. 241-249, 1992.
- 13 FELDMAN, Tony. Turn of the Tide? A New Age in Multimedia Publishing. **British Book News**, p. 658-659, oct 1991.
- 14 GHOSH, Sakti P. Statistical Databases: Design of Experiment Structures. **Information Systems**, v. 18, n. 4, p. 233-247, jun 1993.
- 15 HIBLER, J. N. D., LEUNG, C. H. C., MANNOCK, K. L., MWARA, N. K. A Database System for the Effective Storage and Retrieval of Images. In: INTERNATIONAL INFORMATION RESEARCH CONFERENCE, 2, 1991, Cambridge. **Proceedings... Multimedia Information**. Bowker-Saur: Feeney and Shirley Day, 1991. p. 83-94.
- 16 HILL, Michael W. Patents on Videodisc: A Future System Almost with us. In: THE FUTURE OF INFORMATION RESOURCES FOR SCIENCE AND TECHNOLOGY AND THE ROLE OF THE LIBRARY. Sweden. **Proceedings...** Sweden: Nancy Fjalbrant, IATUL. p. 151-156.

- 17 LAVIETER, Ir. L. International Databases on Current Research. In: CURRENT RESEARCH INFORMATION IN EUROPE, 1993, Amsterdam. **Proceedings...** Amsterdam: 1993. p. 97-98.
- 18 LEUNG, C. H. C. Architecture of an Image Database System. **Information Services & Use**, v. 10, N. 6, p. 391-397, 1990.
- 19 LUNIN, Lois F. An Overview of Electronic Image Information. **Optical Information Systems**, v. 10, N. 3, p. 114-130, may/jun 1990.
- 20 MALLETT, John., MANNING, Colin. Multimedia and Database Design: a Discussion of Database Technology And its Use in Multimedia. **Journal of Document and Text Management**, v. 1, n. 1, p. 75-84, 1993.
- 21 MARTINSEN, David P., LOVE, Richard A., GARSON, Lorrin R. Multiple Use of Primary Full Text Information: a Publisher's Perspective. **On-line Review**, v. 13, n. 2, p. 121-133, apr 1989.
- 22 MARTYN, John. Factual Databases. **Aslib Proceedings**, London, v. 37, n. 5, p. 231-238, may 1985.
- 23 MASTRODDI, F. A. Managing Image Databases: a European Community Perspective. **Information Media & Technology**, v. 23, n. 2, p. 69-74, mar 1990.
- 24 NASH, Rita. Voice Processing: a Present or Future Technology? **LASIE**, v. 17, n. 3, p. 60-75, sep/oct 1986.
- 25 O'CONNOR, Mary Ann. Implementing Optical Storage: How to Select a Document Image Management System. **Document Image Automation**, v. 11, n. 4, p. 230-233, jul/aug 1991.
- 26 OBERHAUSER, Otto C. Interactive Multimedia in Library and Information Services. **Audiovisual Librarian**, v. 17, n. 1, p. 17-25, feb 1991.
- 27 OUDARD, Denis. Which Technology Should the Sound Archivist Trust? **Phonographic Bulletin**, v. 59, p. 37-40, nov 1991.
- 28 PHILIP, George, YOUNG, Elizabeth S. Man-machine Interaction by Voice: Development in Speech Technology: the Stage-of-the-art. **Journal of Information Science**, London, v.13, n.1, p.3-14, 1987.
- 29 SOERGEL, Dagobert. Structure and Construction of Numeric Databases. **Drexel Library Quartely**, v. 18, n. 3-4, p. 135-146, summer/fall 1982.
- 30 STOKES, John. Image Management. **Microform Review**, v. 21, n. 2, p. 77-81, spring 1992.
- 31 SUOZZI, Patrícia. By the Numbers: An Introduction to Numeric Databases. **Database**, New York, v. 10, n. 1, p. 15-22, feb 1987.
- 32 THIEL, T. J. Automated Indexing of Document Image Management Systems. **Document Image Automation**, v. 12, n. 2, p. 43-49, summer 1992.
- 33 WIGGINS Bob. Document Image Processing - An Overview Part 2: Concepts, Capture and Storage. **Document Image Automation**, v. 12, n. 4, p. 12-20, winter 1992.
- 34 WILLIAMS, Martha E. Highlights of the On-line Database Industry. In: NATIONAL ON-LINE MEETING, 13, 1992. **Proceedings...** Medford: Martha E. Williams, 1992. p. 1-4.

6 - SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO: Etapas de Construção

A fase mais importante a ser realizada para a construção de um SRI é a montagem do projeto onde serão estabelecidos todos os procedimentos de construção e operação do sistema, tais como:

- definição das responsabilidades operacionais de cada etapa;
- planejamento do uso do sistema;
- definição das limitações para futuros crescimentos.

A literatura sobre projeto de SRI não é esclarecedora nem sistematizada. É comum encontrar artigos que abordam projetos, mas eles o fazem apenas sobre as etapas de elaboração de forma isolada. Por este motivo, há dificuldades para elaborar um projeto utilizando apenas os conhecimentos obtidos na literatura. Este problema ocorre por dois motivos:

- o pequeno número de artigos que discutem todas as etapas do projeto;
- a não padronização destas etapas.

Estas situações ocorrem pelo fato de os autores relacionarem as etapas que acham mais importantes, não havendo muitas vezes consenso entre eles. Esta afirmação é comprovada através da análise dos artigos de **GRIFFITH**⁸, **NORMAN**¹³, **NORRIS ET ALL**¹⁴, **CURRAN**⁴, **LEDOUX**⁹, **DURR**⁶ e **LOBO**¹¹, que citam diferentes etapas de projeto.

Sistematizando estes artigos, podemos resumir as etapas abaixo:

6.1 - PLANEJAMENTO

Identificação das reais necessidades do SRI. É subdividido em:

6.1.1 - Levantamento do sistema

Identificação dos principais objetivos e abrangência do sistema, tais como:

- Objetivo - Especificar as necessidades que serão atendidas com o desenvolvimento do sistema.
- Área de interesse - Definir os assuntos ou áreas de interesse que serão cobertos.
- Usuários em potencial - Identificar as pessoas que poderão utilizá-lo.
- Grau de utilização - Identificar a frequência de uso e o período de tempo de utilização (vida útil) previsto do sistema.
- Responsabilidade - Definir as pessoas que serão responsáveis pelo desenvolvimento e operação nas diversas etapas do projeto e do sistema.
- Serviços - Definir os serviços e formas de acesso que deverão haver no sistema.

6.1.2 - Análise de Viabilidade Econômica

Avaliar as viabilidades técnicas e econômicas para o desenvolvimento e operação do sistema, analisando e definindo os seguintes itens:

- Previsão dos custos de desenvolvimento, implantação e operação do sistema em um determinado período de tempo;

- Identificação da existência de sistemas similares que abrangem as áreas de assuntos cobertos pela base, comparando os custos de uso destes sistemas em relação aos custos previstos para o desenvolvimento e operação de um novo;
- Definição de política de preços para utilização do sistema. Esta política poderá ser:
 - por custo - considera o custo de manutenção e operação dos serviços acrescidos do lucro.
 - por valor - considera o quanto que o comprador está disposto a pagar pela informação, ou seja, qual o seu valor agregado.

Nesta etapa do projeto as ferramentas utilizadas mais importantes são:

6.1.2.1 - Estudo de Usuário

Identifica através de entrevistas, questionários, anotações diárias, técnica de Delfos, pesquisas, etc., realizadas por telefone, correio eletrônico, fax, etc., os seguintes itens:

- Áreas de interesses específicas de um grupo de usuários ou os usuários que tenham interesse nas áreas cobertas pelo sistema;
- Ambiente operacional disponível;
- Capacidade de absorção das informações a serem contidas no sistema.

6.1.2.2 - Estudo de Viabilidade

Prevê a viabilidade econômica para o desenvolvimento do sistema na relação custo x benefício. Para a determinação deste custo devem ser consideradas as diversas alternativas de cada etapa do sistema (Entrada de Dados, Coleta de Documentos, Armazenamento, Indexação, etc.), além das despesas com propaganda e marketing para divulgação do mesmo.

6.1.3 - Levantamento dos Dados

Identifica as características dos dados (tipos e formas dos documentos) que serão armazenados na base de dados para definir as estratégias de operação do sistema. Devem ser identificados os seguintes itens:

6.1.3.1 - Características do sistema e dos dados

Definição do tipo da base de dados e as estratégias de operação do sistema. Para identificar estes itens devem ser analisados os seguintes tópicos:

- Tipos de documento - Identificar o suporte físico dos documentos que serão tratados pelo sistema, tais como: papel, microficha, formulário, etc.
- Conteúdo dos dados - Identificar as características dos conteúdos dos documentos a serem armazenados na base de dados, tais como: som, imagem, foto, texto, caracter especial, etc.
- Dispositivos de armazenamento - Identificar os dispositivos ou formas de armazenamento dos documentos, tais como: microfilme, papel, dispositivo magnético, etc.
- Categorização do sistema - Identificar, conforme o conteúdo dos dados, a categorização do sistema que será criado, tais como: textual, referencial, imagem, etc.
- Volume de Dados - Identificar a quantidade dos documentos que serão armazenados no sistema. Esta quantidade deve ser analisada de duas maneiras:

- Quantidade inicial - Definir quantos e quais documentos publicados anteriormente à implantação do sistema serão armazenados.
- Quantidade de atualização - Definir a quantidade média prevista de documentos que serão armazenados em um determinado período de tempo.

Nesta etapa, os limites do sistema devem ser identificados, ou seja, qual a quantidade máxima de documentos que podem ser armazenados sem que haja perda na qualidade do sistema.

6.1.3.2 - Definição das estratégias de operação do sistema

Escolha das estratégias de operação para as seguintes etapas:

- Coleta de documentos - Definir as estratégias de implantação e alternativas de coleta de documentos que podem ser utilizadas de acordo com as características do sistema. Estas alternativas e estratégias serão discutidas em capítulo específico.
- Entrada de Dados - Definir a(s) forma(s) de entrada de dados que será(ão) utilizada(s) de acordo com o(s) tipo(s) dos documentos e dispositivos de armazenamento. Estas alternativas serão discutidas em capítulo específico.
- Política de Indexação - Definir a(s) alternativa(s) de indexação que melhor se adapta(m) às características dos documentos e do sistema. Estas técnicas serão discutidas em capítulo específico.

TENOPIR²¹ em seu artigo analisa a rotina diária de três produtores de base de dados mostrando e comparando as atividades de cada um deles.

6.1.4 - Projeto de Arquivo

Identificação, avaliação e definição dos recursos computacionais necessários ao armazenamento, acesso e recuperação dos dados. Nesta etapa devem ser definidos os seguintes itens:

6.1.4.1 - Local de armazenamento dos dados

Definição do equipamento onde a base de dados será armazenada. A literatura tem mostrado alguns artigos que tratam sobre o assunto, como por exemplo **SWARD**²⁰ que analisa as vantagens e desvantagens entre o armazenamento distribuído e centralizado. Podemos relacionar as seguintes alternativas de armazenamento:

- Armazenamento em equipamento próprio - Armazena a base de dados em computador do próprio produtor. Neste caso ele é o responsável pelos procedimentos de armazenamento, processamento e acesso dos dados, fornecendo a estrutura de telecomunicação necessária para o uso pelos usuários.
- Armazenamento em equipamento de terceiros - Armazena a base de dados em computador alugado de terceiros. Neste caso a responsabilidade do processamento e telecomunicação é de quem fornece o equipamento, ficando o produtor responsável apenas pelo fornecimento dos dados em formato preestabelecido.
- Armazenamento distribuído - Armazena e distribui a base de dados em dispositivos magnéticos, tais como fitas magnéticas, disquetes e CD-ROM. Caso a escolha seja pelo CD-ROM é aconselhável fornecer a base completa em substituição à enviada

anteriormente. Este fato é confirmado por diversos artigos sobre produção de CD-ROM. Não é aconselhável criar mecanismos para atualização dos dados por gerar mais um procedimento operacional no sistema e gerar desperdício de espaço no CD-ROM, a quantidade de informações que pode ser armazenada será sempre superior à quantidade a ser atualizada.

6.1.4.2 - Software de gerenciamento

Avaliação e definição do software bibliográfico que será utilizado para o gerenciamento dos dados armazenados na base de dados. Este software deve possuir recursos de entrada, armazenamento e recuperação dos dados com interface amigável para os usuários. São encontrados diversos artigos que tratam das características de diversos softwares comerciais e de empresas provedoras de base de dados. Dentre as alternativas existentes para gerenciamento dos dados, podemos citar:

- Desenvolvimento de software próprio - Desenvolvimento de um software gerenciador dos dados pelo próprio produtor da base.
- Aquisição de um software bibliográfico - Escolha e aquisição de um software bibliográfico comercial que atenda às características do sistema. Os softwares disponíveis no mercado e suas características serão discutidos em capítulo próprio.
- Utilização de serviços de um provedor de base de dados - Avaliação e escolha de um dos provedores existentes no mercado. Estas empresas têm como objetivo armazenar os dados de um produtor de base de dados, fornecendo as estruturas de armazenamento, distribuição e divulgação dos dados, sendo o produtor responsável apenas pela criação e envio dos dados em formato preestabelecido pelo provedor. Alguns exemplos de bases de dados fornecidas por provedores são Dialog, BRS, etc. **NORRIS ET ALL** ¹⁴ relaciona os fatores que devem ser considerados para esta seleção e **DAEHN** ⁵ mostra os critérios que devem ser avaliados para a escolha entre o desenvolvimento de um software e a utilização dos serviços de um provedor. Este tópico não será analisado neste trabalho, por fugir do escopo do mesmo.

6.1.4.3 - Campos de armazenamento

Definição dos dados que serão armazenados na base de dados de acordo com o tipo de documento. Esta definição irá gerar um “Projeto de Arquivos” que representa as estruturas, organização, métodos de acesso, etc. de todos os arquivos existentes no sistema. Esta etapa é mostrada na literatura por artigos que analisam a modelagem dos dados através de ferramenta utilizada na Análise de Sistemas, chamada Modelo de Entidade e Relacionamento (MER). Esta ferramenta procura automatizar o processo de criação dos arquivos. **ARTZ** ² critica este automatismo e mostra a influência cognitiva do projetista nesta etapa.

Nesta fase serão definidos todos os aspectos computacionais do SRI cuja escolha incorreta pode acarretar diversos prejuízos operacionais. O avaliador deve ter conhecimento técnico suficiente para estas decisões.

6.1.5 - Projeto de Implantação

Definição dos seguintes procedimentos operacionais para implantação e distribuição do sistema:

6.1.5.1 - Atualização

Especificar as formas de atualização da base de dados, periodicidade, responsabilidade, etc.

6.1.5.2 - Segurança e Controle

Definição dos procedimentos necessários para:

- evitar eventuais perdas dos dados por falhas de operação do usuário ou equipamento;
- recuperação, caso haja algum problema operacional.

6.1.5.3 - Avaliação do sistema

Definição de procedimentos periódicos de avaliação do sistema através de testes internos e avaliação por parte do usuário, dentre outras alternativas. Estes procedimentos devem definir o grau de qualidade e atualidade dos documentos e a satisfação dos usuários pelo sistema.

6.1.5.4 - Instalação

Definição dos procedimentos para a instalação do sistema, quando este for de responsabilidade do usuário. Estes procedimentos devem ser simples e auto-explicativos fornecendo sempre alternativas de cancelamento ou interrupção.

6.1.5.5 - Treinamento

Definição da política de treinamento que deverá atender:

- Às pessoas envolvidas na operacionalização do sistema - Estes treinamentos deverão ser feitos no momento da implantação do sistema e periodicamente, através de cursos de reciclagem ou atualização.
- Aos usuários que utilizarão o sistema - Este treinamento deverá ser oferecido pelo produtor do sistema no momento da sua implantação ou por solicitação do usuário, através de cursos periódicos ou previamente marcados.

O projeto de SRI deve gerar um produto útil para o usuário, de uso simplificado e atendendo a requisitos conceituais (facilidade de uso, interface estruturada, possibilidade de retorno e cancelamento, etc.) e tecnológicos (cada nova versão deve ser um complemento da anterior, sem ser necessário aprender tudo novamente), sem que haja mudanças no comportamento do usuário para utilizá-lo.

Nos artigos analisados, estes itens são mostrados isoladamente, ou seja, os autores consideram como planejamento apenas um ou alguns destes itens. São encontradas com maior frequências a identificação do tipo, das características do sistema e do volume dos dados como sendo as informações necessárias para planejar um sistema.

Somente sintetizando todos os artigos foi possível identificar todas as fases necessárias para a etapa de planejamento. Analisando cada uma delas, verifica-se a importância do processo de planejar a construção e operação do sistema, identificando seus limites e restrições.

Referências Bibliográficas:

- 1 AITCHISON, T. M. Codes of Practice - Database Producers. **NEWSIDIC**, n. 52, p. 1-2, mar 1982.
- 2 ARTZ, John. The Database Designer is a Person Too! **Journal of Database Management**, v. 5, n. 3, p. 31-34, summer 1994.
- 3 CUNHA, Murilo Bastos. Metodologia para Estudo dos Usuários de Informação Científica. **Revista de Biblioteconomia de Brasília**, v. 10, n. 2, p. 5-19, jul/dez 1982.
- 4 CURRAN, Charles F. Adverse Event Database Construction and Use: 1. Software Design. **Drug Information Journal**, v. 27, n. 2, p. 561-567, apr/jun 1993.
- 5 DAEHN, Ralph M. Methods and Software for Building Bibliographic Data Bases. **Canadian Library Journal**, v. 42, n. 3, p.147-152, jun 1985.
- 6 DURR, W. Theodore. Inputting Data and Conversion of Data: OCR, Scanning, Utility Source for Merging Files. In: IOLS '90. INTEGRATED ON-LINE LIBRARY SYSTEMS MEETING, 5, 1990, New York. **Proceedings...** Medford: Learned Information, 1990, p. 27-33.
- 7 EDDISON, Elizabeth B. How to Plan and Build your Own Database. **Database**, New York, v. 11, n. 3, p. 15-16, jun 1988.
- 8 GRIFFITH, Cary. The Seven Stages of Database Design. **Information Today**, v. 8, n.6, p. 42-44, jun 1991.
- 9 LEDOUX, Par Marc-Andre. La Creation de Bases de Donnees Bibliographiques. **Argus**, v. 11, n. 5, p. 105-110, sep/oct 1982.
- 10 LEVITIN, A. V., REDMAN, T. C. A Model of the Data (Life) Cycles with Application to Quality. **Information Software Technology**, v. 35, n. 4, p. 217-223, apr 1993.
- 11 LOBO, Maria de Fátima Diniz, BARCELLOS, Silvia de Oliveira. Guias de Fontes de Informação: Metodologia para Geração e Automação. **Ciência da Informação**, Brasília, v. 21, n. 1, p. 75-81, jan/abr 1992.
- 12 McQUEEN, Judy. Creating and Maintaning the Bibliographic Database. **Information Today**, v. 7, n. 3, p. 41-42, mar 1990.
- 13 NORMAN, Margaret. Considerations in Planning an Inhouse Database Retrieval System. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 7, 1982, London. **Proceedings...** Oxford, Learned Information, 1983. p. 213-218.
- 14 NORRIS, Carole L., MARINCOLA, Dian A. Guidelines for Developing an On-line In-house Database Through a Comercial Vendor. In: NATIONAL ON-LINE MEETING, 1982. **Proceedings ...** Medford: Learned Information, 1982. p. 435-444.
- 15 O'LEARY, Mick. Standalone On-line Sources: the Challenge of the Niche Database. **Online**, v. 15, n. 6, p. 56-60, nov 91.
- 16 PAO, Miranda Lee. **Concepts of Information Retrieval**. Colorado, USA: Libraries Unlimited, 1989. 285p.
- 17 PAVELIN, Janet. Software Systems for Managing and Storing Documents. **Information Management and Technology**, v. 25, n. 3, p. 130-131, may 1992.

- 18 SCHLEYER, Judith Rebeca. **Estudos de Usuários: Introdução à Problemática e à Metodologia**. In: Estudos Avançados em Biblioteconomia e Ciência da Informação. Brasília: ABDF, 1982, v. 1, p. 49-71.
- 19 STRICKER, Ulla de. Designing an Electronic Research Tool: Living in a Compromise. **Law Librarian**, v.24, n. 4, p. 176-179, dec 1993.
- 20 SWARD, Robert. Distributed vs. Centralized Information Storage. **International Journal of Micrographics and Optical Technology**, v. 11, n. 3, p. 109-110, sep 1993.
- 21 TENOPIR, Carol. A Day in the Life of a Database Producer. **Database**, New York, v. 15, n. 3, p. 15, 17-20, jun 1992.
- 22 TENOPIR, Carol. Quality Control. **Library Journal**, v. 112, n. 3, 15 Feb 87, p. 124-125.

6.2 - COLETA DE DOCUMENTOS

A etapa de coleta de documentos consiste em identificar, localizar e coletar, nas diversas fontes de informação, os documentos de interesse para o sistema visando a sua inclusão na base de dados.

Utilizando os termos "Coleta" ou "Aquisição" (Survey ou Acquisition em Inglês), em pesquisa no LISA, são encontrados artigos que tratam das políticas de aquisição de livros e coleções em bibliotecas, utilizando diversas técnicas para estudo de uso e de usuário, estando elas fora do escopo de construção de um SRI. Tal fato é confirmado por **DODEBEI**² que afirma: "... a metodologia apresentada nos estudos voltados para a formação de bases de dados bibliográficos concentra-se muito mais na seleção de periódicos para compor coleções básicas do que na determinação de fontes de informação e nas rotinas que compõem o binômio LOCALIZAÇÃO (fonte produtora) / OBTENÇÃO (documento) (grifo do autor)", embora os termos possam ser utilizados para a produção de SRI". Nesta etapa foi identificada a menor quantidade de artigos, dentre as demais. Em pesquisa feita no Dicionário Aurélio, podemos definir estes termos da seguinte maneira:

- Seleção - Processo intelectual de identificar a relevância e utilidade de um item no conjunto de material, através de critérios de escolha.
- Aquisição - Processo de procurar o material identificado.

Tais definições especificam as atividades realizadas para coletar os documentos para um SRI.

A etapa de coleta de documentos pode ser subdividida em duas fases:

6.2.1 - Criação da estrutura de coleta

Criar uma estrutura de coleta visa identificar e criar intercâmbios com as fontes produtoras de informação (autores, instituições e documentos). Uma estrutura bem organizada é um dos principais fatores que possibilitam agilizar a coleta dos documentos. Segundo **QUEIROZ**⁵, "um dos grandes problemas para formação de uma base de dados é a estruturação de uma rede para captação de documentos".

A combinação das fontes produtoras de informação conduz a um encadeamento que **DODEBEI**², afirma que "são ações que encadeiam pontos que se alimentam reciprocamente: localização das fontes primárias e secundárias, identificação das referências bibliográficas e coleta dos originais, em um ciclo que envolve autores, instituições e documento". Tal afirmação é detalhada por **QUEIROZ**⁵ da seguinte forma: "A partir do conhecimento de uma instituição X, solicita-se seu Relatório Anual de Atividades que nos fornece todos os artigos publicados em revistas, relatórios internos, trabalhos apresentados em Congressos, etc. Se a seguir tomarmos um destes trabalhos apresentados em congressos, por exemplo, seremos levados até aos anais. Estes, por sua vez, terão contribuições de interesse apresentadas por outras pessoas, ligadas a novas instituições, às quais serão solicitados os Relatórios Anuais de Atividades, que servirão de ponto de partida para um novo ciclo na coleta."

Para a criação da estrutura de coleta é necessário ter inicialmente o conhecimento prévio sobre o que é produzido na área de interesse, quem produz, para quem e onde. Devem ser conhecidos os documentos convencionais (periódicos, livros, etc.), os não convencionais (relatórios de pesquisas, trabalhos apresentados em congressos, reuniões, teses, etc.) e a literatura

cinzenta (literatura não publicada). As literaturas não convencionais e cinzentas geram dificuldades na sistemática de coleta por possuir circulação restrita, sendo necessário localizar as fontes primárias, secundárias e terciárias de referência.

As primeiras ações para a coleta do documento são realizadas a partir das fontes secundárias e terciárias de referências podendo ser realizadas através de contatos pessoais ou por correspondência. Neles devem ser ressaltados os objetivos da coleta e a importância do sistema que está sendo criado.

6.2.2 - Operacionalização da coleta dos documentos

Execução dos procedimentos operacionais de busca do documento de interesse do sistema, através do intercâmbio feito na fase de estruturação. Esta busca pode ser feita de duas maneiras:

6.2.2.1 - Pelas fontes de informação

As fontes de informação representam a origem, onde os documentos foram gerados. Estas fontes são classificadas de quatro formas:

6.2.2.1.1 - Quanto a origem:

São identificadas três fontes de geração de documentos:

- Pessoas - Representadas pelos autores, editores e orientadores que fornecem seus trabalhos ou informam às instituições as publicações ou pessoas que atuam na área.
- Instituições - Representadas pelos órgãos que editam ou são depositários de documentos, tais como: Bibliotecas, Instituições de Pesquisas, Agências de Fomento, Academias Nacionais, Universidades, etc. Estes órgãos devem fornecer os seus Relatórios de Pesquisa, Relatórios Anuais, Relatórios de Atividades, etc. para identificar os trabalhos publicados.
- Publicações - Representadas pela literatura convencional e não convencional, tais como: teses, periódicos, relatórios, catálogos, etc.

6.2.2.1.2 - Quanto a forma de veiculação e transmissão:

São identificadas quatro formas de transmissão de documentos:

- Verbal - Caracterizada pela comunicação oral entre os profissionais da mesma área, através de encontros informais em Congressos, Associações, Reuniões, etc.
- Impressa - Caracterizada pela informação veiculada em papel, microficha ou microfilme, CD-ROM, etc.
- Audiovisual - Informação veiculada utilizando recursos de sons, imagens, etc. através de filme, rádio, televisão, fita de vídeo, etc.
- Computadorizada - Informação armazenada em Sistemas de Recuperação e divulgada através de arquivos eletrônicos, correio eletrônico, etc.

6.2.2.1.3 - Quanto a localização ou acesso:

São identificadas três formas de publicação dos documentos:

- Fontes convencionais - Informações publicadas nos mecanismos formais de publicação, tais como: periódicos, livros, etc.
- Fontes semi-convencionais - Informações distribuídas apenas por solicitação aos seus produtores, não possuindo finalidades comerciais nesta distribuição. Este tipo de informação é chamada de “literatura cinzenta” .
- Fontes não convencionais - Representam as informações de difícil obtenção, normalmente de caráter informal.

6.2.2.1.4 - Quanto ao conteúdo:

Segundo classificação feita pela **ASSOCIAÇÃO BRASILEIRA DE QUÍMICA**¹ :

- Fontes bibliográficas - Informações que permitem a identificação da literatura impressa.
- Fontes cadastrais - Fornecem perfis descritivos de indivíduos, produtos, equipamentos, etc., através de catálogos, diretórios, etc.
- Fontes textuais - Contém os textos completos dos documentos.
- Fontes numéricas - Armazenam valores ou dados numéricos, temporais ou não.
- Fontes gráficas - Armazenam imagens, desenhos ou outros caracteres não representados no teclado comum.
- Fontes computadorizadas - Armazenam programas de computadores disponíveis para uso (shareware).

6.2.2.2 - *Pelas Fontes de Referência*

As fontes de referências identificam e localizam as fontes de informações, que podem ser:

6.2.2.2.1 - Primárias

São as fontes que contém e divulgam a informação no formato original, ou seja, o próprio documento. São representados pelos livros, periódicos, normas, relatórios, teses, etc.

6.2.2.2.2 - Secundárias

São as fontes que organizam os documentos sob a forma de índice e resumos, facilitando o conhecimento e uso da informação dispersa nas fontes primárias. São representados pelas bibliografias, catálogos coletivos, manuais técnicos, índice de citações, etc.

6.2.2.2.1 - Terciárias

São as fontes que orientam os usuários na utilização das fontes secundárias e primárias, facilitando a localização e, em alguns casos, o acesso à informação dispersa. São representadas por bibliografias de bibliografias, guias, índices de índices, etc.

Sistematizando as fontes de referências em relação às fontes de informação, **DODEBEI**² monta uma matriz com os indicadores das fontes para localização de documentos, conforme descrito no quadro 3. Este quadro contém uma matriz de indicadores de fontes para localização de documentos primários, onde as linhas representam os tipos de documentos primários e as colunas as diversas fontes de referências para cada um deles, ficando assim configurado que uma fonte pode ser secundária para um tipo de documento e primário para outro.

O processo de coletar documentos é iniciado pelas fontes terciárias. Nelas são identificadas as instituições, ou fontes secundárias, de interesse da base. Após a seleção da instituição devem ser realizados os contatos necessários para o conhecimento da sua produção (fontes primárias). Segundo **STIRTON**⁸, estratégias de convênios com instituições devem ser criadas para coletar documentos destacando a importância da intervenção de um Órgão Federal para sistematizar esta coleta.

O quadro abaixo sintetiza a obtenção da fonte de informação em relação às fontes de referência:

Quadro B - Obtenção da informação em relação à fonte de referência.

Fontes de Informação	Primária	Secundária	Terciária
----- Documento Primário			
Pessoas	X		
Instituições	X	X	
Publicações	X	X	X

Quadro C - Indicadores das Fontes de localização dos documentos primários.

6.2.3 - FORMAS DE OBTENÇÃO DE DOCUMENTOS

Localizar um documento de interesse do SRI, na sua fonte primária, requer diferentes esforços para diferentes formas de aquisição, tais como:

Fontes de Informação ----- Documento Primário	Primária	Secundária	Terciária
Relatórios Técnico-Científicos	Listas de Proj. Pesquisas Relatórios Anuais Bolsas de Estudos	Instit. de Apoio a Pesquisa Instit. de Ens. e Pesq. Instit. de Apoio a Pesquisa	Diretórios Órgão Governamental Diret. de Inst. de E & P
Trabalhos de Congressos	Anais Listas pré-prints Relatórios Anuais Agendas Revistas Curriculum Vitae Informativo Assoc.	Catálogo de Bibliotecas Instituição Patrocinadora Instit. de Ens. e Pesq. Revistas Associações de Classe	Diretórios Bibliotecas Diretórios Instituições Catálogos de Editores Quem é Quem/Usuários Listas de Associações
Teses e Dissertações	Catálogo de Tese Relatórios Anuais Artigos Bibliografias Citações Bibliográficas Curriculum Vitae (Autores)	Catálogo de Bibliotecas Instit. de Ens. e Pesq. Revistas Bibliografias Secundárias Artigos	Bases de Dados (outras)
Artigos	Revistas Bibliografias Citações Bibliográficas Curriculum Vitae (Autores)		
Livros	Bibliografia Catálogo de Bibliotecas Catálogo de Editores Agendas (revistas)		
Patentes	Bibliografia Revistas Informativos		
Literatura de Extensão Rural	Programa de Assist. Rural Bibliografias Catálogo de Bibliotecas Catálogo de Editores Agendas (revistas)	Instituições de Assist. Rural	
Legislação	Diário Oficial	Coleções de Leis Bibliografias	
Estatísticas	Publicações Oficiais Artigos	Órgãos Governamentais (Planej., Política, Controle)	
Relatório de Análises Setoriais	Consultores Empresas	Empresas	
Folhetos Comerciais	Feiras/Exposições Anúncios Comerciais Empresas	Calendário de Eventos Revistas	Hotéis Firmas Organizadoras Órgãos Governamentais

6.2.3.1 - Compra ou Assinatura

Representa a aquisição do documento pelo produtor da base de dados. Esta alternativa é utilizada apenas para alguns documentos convencionais, através de compras de livros ou assinaturas de periódicos.

6.2.3.2 - Intercâmbio

Esta alternativa é utilizada quando a instituição que gera o documento realiza sua troca por outros de seu interesse ou pelo uso dos serviços disponíveis no SRI.

6.2.3.3 - Doação

Representa o recebimento do documento desejado gratuitamente.

6.2.3.4 - Depósito

Consiste na obtenção e armazenamento do documento pelo provedor do SRI, passando ele a ser o depositário.

6.2.3.5 - Empréstimo

Esta situação ocorre com o recebimento por empréstimo do documento, com o compromisso de devolvê-lo após o uso.

6.2.3.6 - Análise in situ

Consiste na realização de pesquisas aleatórias em Bibliotecas ou Centros de Documentação com objetivo de localizar artigos de interesse.

6.2.3.7 - Transferencia Eletrônica

Consiste na obtenção do documento através de transferência eletrônica por Fax, Correio Eletrônico, etc.

Marek, sintetizou as alternativas de aquisição no quadro D, abaixo:

Quadro D - Alternativas de aquisição de documento.

<u>Caract.literatura</u> Processos	Tipo de Literatura (mais comum)	Acesso	Observações
Compra	Convencional	Antecipar a obtenção mediante contatos freqüentes com editores. Empréstimo de cópias para indexação	Alto custo
Intercâmbio	Não-convencional	Ampliação do contato com as instituições	Doc. x Doc. Doc. x Produtos
Doação	Convencional Não-convencional	Participação em eventos: feiras, exposições, conferências	Relações informais
Depósito	Não-convencional	Difícil acesso - caráter arquivístico	Doc. gerados internamente

Empréstimo	Convencional Não-convencional	Formação de redes de bibliotecas	Baixo custo uso de cópias
Análise <i>in situ</i>	Convencional	Deslocamento a Bibliotecas e Centros de Documentação	Tarefa efetuada pelo indexador para descobrir artigos em fontes dispersas
Transferência Eletrônica	Convencional	Fontes terciárias disponíveis em redes de comunicação	Acesso à redes de comunicação

6.2.4 - FORMAS DE CONTATO

Representam as alternativas de contato existentes entre o provedor do SRI e a Instituição que possui o documento na sua forma original. Estes contatos devem ser periódicos estabelecendo laços de amizade e cooperação entre as instituições. **DODEBEI**² descreve as seguintes alternativas de contato:

6.2.4.1 - Contato Pessoal

Os contatos pessoais devem ser realizados para conhecer os projetos desenvolvidos, ou em desenvolvimento, e a literatura produzida pela instituição, principalmente a não convencional.

DODEBEI² afirma que "geralmente a literatura não convencional é descoberta somente quando são feitas visitas às instituições". Estas visitas, além de propiciarem o conhecimento dos projetos, servem também para estreitar laços de cooperação entre as instituições. Estes contatos devem ser iniciados na fase de Criação da Estrutura da Coleta, para mostrar às instituições visitadas o processo que se inicia e para conhecer o potencial literário da instituição. Devem ser realizados periodicamente para identificar possíveis publicações.

6.2.4.2 - Contato por correspondência

Deve ser realizado através de cartas ou formulários pré-impresos que, segundo **LAMBERT**³, deverão conter informações sobre a instituição solicitante, tais como: o que faz, qual o papel que desempenha neste processo, qual a finalidade da solicitação e outras informações adicionais. Estas informações deverão ser fornecidas na fase de Criação da Estrutura da Coleta.

Na operacionalização da etapa devem ser enviadas correspondências solicitando publicações específicas ou perguntando sobre a existência de literaturas de interesse do SRI.

Todos estes contatos deverão ser controlados através de formulários de acompanhamento e cadastros das fontes terciárias, secundárias e primárias, conforme propostas de modelo apresentada por **LAMBERT**³, com informações que possam identificar o pedido realizado, evitando duplicidade.

6.2.4.3 - Outras formas de contato

Segundo **LAMBERT**³ outras formas de contato são utilizadas para coleta de documentos, tais como:

- Palestras em Instituições de Ensino e Pesquisa - como forma de divulgar o sistema e coletar documentos gerados pelas instituições participantes.
- Apresentações de trabalhos em eventos - como forma de divulgar os serviços do SRI e conscientizar os pesquisadores sobre a necessidade de enviar os documentos produzidos para inclusão na base de dados e sua conseqüente divulgação.

6.2.5 - Critérios de Seleção

A avaliação de um documento para a sua inclusão na base de dados é um processo subjetivo e nebuloso de caráter individual e pessoal do avaliador. Esta afirmação é confirmada por **PAO**⁴: “o processo de seleção é nebuloso, pois subjetividade e valor não são determinados objetivamente”, e **ROYCE**⁷: “a avaliação da importância do documento é individual”.

Alguns critérios que devem ser utilizados para a seleção de um documento são:

- Relevância - Determina o grau de importância que o documento contém para a área.
- Escopo - Determina se os assuntos tratados pelo documento são de interesse do SRI.
- Unicidade - Representa o nível de novidade das informações contidas no documento.
- Qualidade - Segundo **RICHARDS**⁶, existem dois tipos de qualidade:
 - Qualidade de conteúdo - Avaliada por um conjunto de fatores coletivos, tais como: nível de conhecimento, valor do item, reputação do autor, editor, produtor, etc.
 - Qualidade técnica - Representada pelo tipo de papel e impressão utilizados, clareza das imagens ou figuras, etc.

Para a determinação destes fatores, valores totalmente subjetivos, tais como a experiência e o conhecimento do avaliador determinam a seleção ou não do documento.

Para que não haja esta subjetividade, **TRISKA**⁹ propõe, em sua dissertação de mestrado, um processo de automação de seleção de documentos. Ele avaliou três instituições, identificando processos padrões de seleção em cada uma delas, formando um Diagrama de Fluxo de Dados destes processos viabilizando desta forma sua automação.

No processo de seleção, segundo **DODEBEI**², as ações devem ser permanentes com buscas sistemáticas para a obtenção dos documentos, pois estes representam a matéria-prima para a formação de uma base de dados.

Na etapa de coleta de documentos existe uma grande lacuna na literatura quanto:

- Às formas de coletar documentos;
- À realização de intercâmbios com instituições;
- Às atuais técnicas de coletas eletrônicas.

Na literatura estrangeira são poucos os artigos que abordam aspectos ligados à coleta de documentos. Nos artigos pesquisados para este trabalho apenas os publicados por **DODEBEI**² e **LAMBERT**³ sintetizam algumas alternativas de coleta, e o produzido por **QUEIROZ**⁵ trata das alternativas de criação da estrutura de coleta. Os demais discutem aspectos a serem considerados para a coleta, tais como: qualidade, tipos de documentos, etc.

Sintetizando a etapa de coleta de documentos para a formação de um SRI podemos dividi-la em duas etapas:

- Criação da estrutura de coleta - Nesta fase devem ser realizados os seguintes procedimentos:
 - Contato com as Instituições;
 - Identificação das fontes secundárias e terciárias.
- Operacionalização do sistema - Nesta fase devem ser realizados os seguintes procedimentos:
 - Busca do documento na fonte de referência;
 - Contato com a fonte de informação;
 - Obtenção do documento;
 - Avaliação do documento;

Selecionado o documento é iniciado o processo entrada do seu conteúdo na base de dados, próxima etapa de operação do sistema.

Referências Bibliográficas:

- 1 ASSOCIAÇÃO BRASILEIRA DE QUÍMICA. **Guia de fontes de informação em química e engenharia química no Brasil**. Rio de Janeiro: ABQ, 1995. 410 p.
- 2 DODEBEI, Vera Lúcia Doyle Louzada de Mattos. Metodologia de Coleta de Documentos para Base de Dados Bibliográficos. **Revista de Biblioteconomia de Brasília**, v. 14, n. 2, p. 313-327, jul/dez 1986.
- 3 LAMBERT, Maria Betânia Monte Alto. **Manual de Coleta de Documentos para Alimentação da Base de Dados do INIS**. Rio de Janeiro: Centro de Informações Nucleares, jul 1995. (a ser publicado).
- 4 PAO, Miranda Lee. **Concepts of Information Retrieval**. Colorado: Libraries Unlimited, 1989. 285p.
- 5 QUEIROZ, Gilda Gama de. Fonte. Base de Dados Bibliográficos em Fontes de Energia. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 12, 1983, Camboriú. **Anais...**Camboriú: 1983. 13 p.
- 6 RICHARDS, Daniel T. By your Selection Criteria are ye Known. **Library Acquisitions: Practice and Theory**, New York, v. 15, n. 3, p. 279-285, 1991.
- 7 ROYCE, Bert R. The Place of Objective Measures of Electronic Publishing. In: ASIS ANNUAL MEETING, 45, 1982, Columbus. **Proceedings... Information interaction**, New York: Knowledge Industry Publications, 1982, v. 19, p. 41-42.
- 8 STIRTON, G. BEST Europe: Present and Future. In: CONFERENCE: CURRENT RESEARCH INFORMATION IN EUROPE, 1993, Amsterdam. **Proceedings...** Amsterdam: M. J. de Goede, NBOI, 1993, p. 109-120.
- 9 TRISKA, Ricardo. **Processo de seleção: Rumo a automação**. Rio de Janeiro: Instituto Brasileiro de Informação em Ciência e Tecnologia e Universidade Federal do Rio de Janeiro, 1993. 74 p. Dissertação (Mestrado em Ciência da Informação) Escola de Comunicação, Universidade Federal do Rio de Janeiro, 1993. .

6.3 - ENTRADA DE DADOS

O formato de entrada dos dados que irão compor a base está diretamente relacionado à categoria do sistema (Textual, Referencial, Imagem, etc.), ao tipo de documento (periódico, relatório, etc.) e ao seu conteúdo (texto, foto, imagem, etc.). Esta etapa realiza os procedimentos de cópia, validação e armazenamento dos documentos na base de dados. Nos artigos analisados foi identificado apenas um que sistematiza o processo de entrada de dados. Este fato ocorre por estar o formato da entrada dos dados diretamente relacionado ao conteúdo do documento.

Com o crescimento do número de SRIs disponíveis e do número de registros armazenados, (segundo DANIEL ⁵ passou de 52 milhões em 1975 para aproximadamente 5 bilhões em 1989), houve um crescimento proporcional da importância do processo de validação dos dados armazenados, sendo considerado como aspecto mais importante a qualidade do armazenamento.

A falta de qualidade na informação ocorre, dentre outros fatores, pela obsolescência ou duplicidade dos materiais fornecidos. Estes aspectos devem ser avaliados na etapa de coleta dos documentos, por profissionais especializados, bibliotecários ou profissionais da informação, evitando a seleção de documentos que não atendam à qualidade do sistema. Entretanto, a não nitidez do conteúdo dos artigos (difícil leitura ou visualização) ou a falta de consistência entre este conteúdo e as suas referências (criação de índices), são aspectos que estão diretamente relacionados à entrada de dados e são tratados de forma superficial na literatura.

FURUTA ⁹ destaca a importância da evolução dos sistemas de preparação dos documentos para bases de dados, relacionando vários artigos que discutiram e influenciaram estas evoluções.

As formas para entrada de dados podem ser classificadas em três métodos: Digitação, Optical Character Recognition (OCR) e os Meios Eletrônicos. A escolha de um ou mais métodos está diretamente relacionada ao tipo de material ou a maneira como foi publicado, segundo O'LEARY ¹⁷.

6.3.1 - Digitação

Consiste na transcrição, através do teclado do computador, da informação a ser armazenada na base de dados. Para EAGLE ⁶ esta técnica ficou sobrecarregada com o volume de informações geradas ao longo do tempo. Segundo O'LEARY ¹⁷ para os provedores que recebem os documentos fontes em "listagem de computador" (documento completo), este é o primeiro processo para a entrada dos dados.

O processo de digitação pode ser feito pelo próprio produtor ou terceirizado. Os métodos de validação dos dados digitados devem atender:

- Verificação de ortografia - Desenvolvendo ou utilizando softwares ou outros recursos que avaliam a sintaxe das palavras.
- Detecção de duplicidade - Desenvolvimento de software que avalie a presença do documento digitado na base de dados. Esta verificação pode ser feita comparando-se alguns de seus itens, tais como: título, autor, local, ano de publicação, etc. Segundo FOLLINGSTAD ⁷, mesmo que o programa permita duplicidade, deve ser feita uma atualização do conteúdo substituindo o documento já armazenado pela nova versão. A atualização do conteúdo de um documento na base de dados não é uma situação comum de ocorrer. Ele deve ser recusado na etapa de seleção e coleta evitando a digitação em duplicidade.

FOLLINGSTAD ⁷ define em seu artigo os requisitos necessários e disponíveis no software FILEMAN (software de domínio público para gerenciamento de dados), para a validação dos dados. Para as bases de dados que contem conteúdo formado por fórmulas científicas, **PARLOW** ¹⁸ propõe um programa que aceite os símbolos não disponíveis no teclado.

Mesmo eliminando estes problemas, existe um percentual aceitável de erros que para **O'LEARY** ¹⁷ deve ser de 0,5% e para **DANIEL** ⁵ de 15 caracteres errados para 100.000 digitados, ou seja, 0,015%. Estes percentuais tão diferentes mostram que não existe um padrão ou consenso.

Este método é indicado para os sistemas referenciais e bibliográficos na inclusão dos dados bibliográficos, localização, resumo e índices de acesso dos documentos na base de dados. Na literatura este processo é pouco discutido pela necessidade de abordagens das novas tecnologias e pela sua pouca utilização.

6.3.2 - Optical Character Recognition (OCR)

Segundo **GOODALL** ¹¹, “reconhecimento (recognition) é o ato de perceber se uma pessoa, criatura ou coisa pertence a uma classe previamente conhecida ou é a própria classe.”. No contexto do processamento de imagens este reconhecimento significa identificar padrões de caracteres, palavras ou outras marcas como símbolos especiais, imagens, etc.

Os equipamentos de OCR's surgiram no final da década de 80 com limitados recursos de identificação de fontes e formatos. Com o desenvolvimento dos microcomputadores houve grande aumento destes recursos passando a identificar diversas fontes dos editores de textos, gráficos e manuscritos, conforme afirmação de **EAGLE** ⁶.

A evolução ocorrida nos equipamentos iniciou um processo de substituição do digitador pelo “Scanner”, equipamento utilizado para transferência do conteúdo do documento para seu armazenamento no computador convertendo os seus caracteres ou imagens em códigos ASCII (padrão de reconhecimento do computador). Segundo **EAGLE** ⁶, um Scanner envia ao computador 100 caracteres por segundo, equivalente a 20 digitadores.

Segundo **JARMAN** ¹⁴ e **BRILLIANTINE** ², as vantagens na utilização de OCR são:

- Os documentos fonte podem ser copiados diretamente para o computador;
- Os sistemas OCR's são capazes de lidar com grande volume de trabalho;
- Permite expansão da capacidade de produção com custos fixos;
- Preparação dos dados centralizados ou transferidos via rede local;
- Redução do envolvimento humano, permanecendo este apenas na validação da qualidade dos documentos;
- Permite copiar qualquer tipo de documento;
- Eliminação de digitação;
- Redução de erros;
- Menor custo de equipamento;
- Redução do nível de conhecimento/rotatividade do empregado;
- Redução nos custos de transmissão e custos operacionais;

Conforme descrição feita por **FRUCHTERMAN**⁸ e **SCHEIN**²⁰, o scanner monta para cada caracter uma matriz com conteúdo preenchido com valores 0 ou 1 (notação binária utilizada para linguagem do computador), que representa a existência ou não de conteúdo na posição da matriz. Quanto maior a resolução da scanner maior é a quantidade de elementos na matriz. Para os equipamentos coloridos cada posição representa um conjunto de 0 e 1 que combinados representam a cor copiada. Neste caso, o arquivo com a imagem ou texto gravado ocupa grande espaço para armazenamento, sendo necessário a sua compactação. Vários softwares de OCR's trabalham desta maneira ou com pequenas variações, como por exemplo, o software OMNIFONT citado por **FRUCHTERMAN**⁸.

O uso do OCR como método de entrada de dados atende à maioria das necessidades operacionais e fornece um alto índice de qualidade, fator importante para o processo de formação de qualquer base de dados. Esta qualidade é obtida realizando os seguintes testes:

- Testes de imagens - Consiste em visualizar na tela a imagem do documento que foi "scaneado". Caso a imagem não esteja com boa qualidade o documento deve ser novamente lido pelo scanner. Alguns softwares de OCR possuem recursos que permitem melhorar esta imagem. Entretanto, a boa qualidade da reprodução está diretamente associada à qualidade do documento.
- Teste de leitura - Após o teste de imagem, todos os documentos são armazenados nos dispositivos periféricos do equipamento (disco magnético). O teste de leitura é um processo realizado sem a interferência do operador. O software de OCR realiza a leitura de todos os documentos armazenados mostrando-os na tela. Após a realização deste teste os documentos são indexados.
- Testes de recuperação - Consiste em recuperar, de forma aleatória, os documentos armazenados através dos seus índices de acessos.

Segundo **GROOMS**¹², uma imagem com quantidade de erros superior a 15 caracteres em 100.000 provoca a rejeição do documento. Além disso, são escolhidos aleatoriamente alguns documentos para verificação caracter por caracter.

Existem tipos específicos de scanners e softwares OCR para diferentes finalidades. **CRAWFORD**⁴ relacionou alguns destes softwares em seu artigo. Entretanto, para o tratamento de documentos, devemos reportar a **BROADHURST**³ que relaciona alguns scanners para livros e digitalização de microfichas, identificando os problemas e as características de cada um deles. **JONES**¹⁵ define os tipos de OCR que reconhecem os caracteres no momento em que são escritos.

A tecnologia de reconhecimento automático para entrada de dados, pode ser utilizada em diversas outras aplicações, conforme descrito por **HENDLEY**¹³ e **GOODALL**¹¹, tais como: código de barras, tarja magnética, etc., sendo o OCR uma destas alternativas.

Esta forma de entrada de dados está presente na literatura com bastante frequência havendo várias revistas técnicas publicadas. Nestas revistas, e nas demais publicações convencionais, são discutidos os vários aspectos do OCR's tais como: equipamentos, controle de qualidade, etc. Dos artigos analisados, a maioria trata da qualidade da transferência e de formas de armazenamento dos documentos copiados pelos equipamentos óticos.

6.3.3 - Meios Eletrônicos

Com o avanço da tecnologia de comunicação de dados diversas alternativas eletrônicas podem ser utilizadas para entrada de dados, tais como:

- FAX - Permite o recebimento do documento através da sua transmissão utilizando linhas telefônicas (fax-modem). Este conteúdo pode ser armazenado diretamente em computadores para impressão imediata ou posterior ou transferidos para a base de dados.
- Correio Eletrônico - Transferência do conteúdo de um documento armazenado em um arquivo para o computador do provedor da base de dados. Esta técnica é a tendência atual, sendo a transferência via Internet a alternativa mais utilizada.

Além destas tecnologias eletrônicas, outras começam surgir na literatura como por exemplo as técnicas de reconhecimento de voz. Esta nova tecnologia, ainda em fase de desenvolvimento, avaliação e teste, deverá se constituir em uma nova forma de entrada de dados, que deverá ser utilizada pelos sistemas para realização das consultas pelos usuários (subsistema de saída), não sendo considerada para este trabalho.

6.3.4 - Armazenamento de sons

Para esta tecnologia não foram encontrados artigos que a discutam. Entretanto, considerando os recursos disponíveis em multimídia, deve-se considerar a entrada de dados sonoros através dos equipamentos de CD-ROM e outros recursos de áudio.

Sintetizando as alternativas acima, com as categorias existentes, encontramos o quadro abaixo:

Quadro E - Formas de entrada dos dados conforme a categoria do SRI.

Forma de Entrada de dados ----- Categorias de SRI	Digitação	OCR	Meios Eletrônicos	Sons
Referenciais	X			
Bibliográficas	X			
Texto Completo		X	X	
Texto Completo/Imagens		X	X	
Imagens		X	X	
Numéricas	X		X	
Numéricas/Textuais	X	X	X	
Software			X	
Áudio				X
Combinação de tipos		X	X	X

Referências Bibliográficas:

- 1 APPLIN, Roger. Optical Storage Quality Control. **Document Image Automation**, v. 11, n. 2, p. 71-73, mar/apr 1991.
- 2 BRILLIANTINE, Lance R. OCR Improves Productivity. **Word Processing & Information Systems**, v. 9, n. 3, p. 42-44, 64-65, 67, mar 1982.
- 3 BROADHURST, Roger. The Digitisation of Library Material. **Information Management & Technology**, v. 26, n. 3, p. 128-132, may 1993.
- 4 CRAWFORD, Walt. Catching Pictures, Catching Words: Low-Cost Scanning and Optical Character Recognition. **Library Hi Tech**, v. 9, n. 1, p. 91-102, 1991.
- 5 DANIEL, Evelyn H. Quality Control of Documents. **Library Trends**, v. 41, n. 4, p. 644-664, spring 1993.
- 6 EAGLE, Mark. On Opening the Number one Bottleneck in On-line Service-data Entry. In: NATIONAL ON-LINE MEETING, 1987, New York. **Proceedings...** Medford: Learned Information, 1987, p. 125-131.
- 7 FOLLINGSTAD, Marianne Controlling Data Entry in Fileman-Based System Development or Expansion. **Mumps a World Class Technology**. Orlando, v. 20, n. 1, p. 7-13, jun 1990.
- 8 FRUCHTERMAN, James R. Omnifont Text Recognition: Linking Paper, Electronic and Optical Input. **Inform.**, v. 2, n. 5, p. 16-19, may 1988.
- 9 FURUTA, Richard. Important Papers in the History of Document Preparation Systems: Basic Sources. **Electronic Publishing**, v. 5, n. 1, p. 19-44, mar 1992.
- 10 GAGAN, David. Scanning: a Survival Guide (4). Image Scanning: File Formats **C and L Applications**, v. 6, n. 8, p. 8-10, apr 1993.
- 11 GOODALL, Ian. Applying Recognition Technology for Automatic Data Entry. **Information Management & Technology**, v. 26, n. 2, p. 83-87, mar 1993.
- 12 GROOMS, David W. Quality Assurance of Text and Image Databases at the U.S. Patent and Trademark Office. **Information Services & Use**, Amsterdam, v. 8, n. 2-3-4, p. 161-165, 1988.
- 13 HENDLEY, Tony., PRITCHARD, John. Recognition Technology. **Information Management & Technology**, v. 26, n. 2, p. 70-78, mar 1993.
- 14 JARMAN, Colin. UK Government Has Second Thoughts on benefits of OCR. **Computer Weekly**, v. 759, p. 18-19, 28 may 1981.
- 15 JONES, Keith. Three Data Entry Methods which Eliminate the Need to Straggle with a VDU. **Computer Weekly**, v. 759, p. 17, 28 may 1981.
- 16 KRUZAN, Donald. Reader Scanners, PCs, Fax - New Decade Will See Combined Technologies that Improve Document Management. **International Journal of Micrographics & Optical Technology**, v. 9, n. 4, p. 203-206, 1991.
- 17 O'LEARY, Mick. Producing a Database: Many Choices for Data Entry. **Database**, New York, v. 13, n. 1, p. 38-40, feb 1990.
- 18 PARLOW, A., WEISKE, Ch., HEB-POHL, N., RODEN, G., ROWMELT J., ZIRZ, C. A Computer System for Inputting Generic Reactions and Sequences to Produce Databases and Printed Publications. In: INTERNATIONAL CHEMICAL INFORMATION CONFERENCE, 1992, Montreux. **Proceedings...** Montreux: 1992. p. 225-233.

- 19 SCHANTZ, Herbert F. Forms Automation and Integrated Imaging (OCR) Systems. **Remittance Doc. Process. Today**, v. 13, n. 6, p. 9-11, mar/apr 1991.
- 20 SCHEIN, Alan. Optical Storage and OCR - Key Components of Automated Information Management Systems. **Optical Information Systems**, v. 9, n. 1, p. 9-15, jan/feb 1989.

6.4 - CATALOGAÇÃO

Esta etapa do projeto consiste na escolha do formato com que os registros bibliográficos serão armazenados nas bases de dados. Tem como objetivo:

- Definir regras para a organização dos dados em meio legível por computador, incluindo a estruturação do meio físico de armazenamento;
- Criar códigos e regras para a identificação dos diferentes elementos de dados no registro, ou seja, a identificação de informações, tais como título, autor, data, etc.;
- Definir regras para a formulação dos diferentes elementos de dados, em termos de conteúdo e da forma de registrá-los;

Um registro bibliográfico representa o conjunto de dados (título, autor, data de publicação, etc.) referentes a uma entidade bibliográfica (documento representado por relatório técnico, artigo de periódico, etc.). Estes dados, representados pelos campos da entidade, são armazenados em computadores através de formatos pré-definidos, chamado de lay-out do registro ou registro lógico.

Cada registro lógico pode ser subdividido internamente em vários registros físicos. Esta divisão representa o formato em que eles estão armazenados no disco magnético do computador. O formato físico, também chamado de formato de comunicação, quando padronizado, permite o intercâmbio dos dados entre várias bases.

Um registro lógico contém:

- Descrições bibliográficas do documento;
- Classificação ou Indexação aplicada ao documento;
- Resumo;
- Outras informações.

Um registro físico pode conter:

- Armazenamento de caracteres - Armazenamento de diferentes tipos de caracteres, tais como: letras, números, caracteres subscritos ou superscritos, caracteres gregos, etc.
- Registros com tamanhos variáveis - Para cada campo existente no registro é definido um tamanho suficiente para armazenar o seu conteúdo. Entretanto, por possuírem valores variáveis, e conseqüentemente diferentes tamanhos, devem ser armazenados apenas os caracteres que foram realmente utilizados, desprezando-se os espaços existentes entre o final do conteúdo do campo e o tamanho definido.
- Identificação de níveis bibliográficos - Indicação do tipo de documento e se faz parte de um conjunto de outros documentos, tais como: capítulos de livro, artigo de periódico, etc. Neste caso são armazenados dados sobre o documento (autor, título, etc.) e sobre o conjunto (título do periódico ou livro, data e local da publicação, etc.).
- Identificação de subcampos - Para os campos que possuem mais de um elemento de dados (ex. conferência possui: nome, local, data, etc.). Estes elementos, chamados de subcampos, são necessários para que possam ser recuperados isoladamente (local da conferência) ou no conjunto (conferência, englobando todas as suas informações).

- Identificação do campo - Um campo para ser reconhecido e recuperado pelo computador, deve possuir uma identificação (TAG). Esta identificação pode ser um número, seqüência de caracteres, etc.
- Mecanismos de ligação - Permite interligar um determinado registro a outros registros, tais como: armazenamento de anotações sobre o registro, referências cruzadas entre registros, etc.

Um formato padronizado de armazenamento permite o intercâmbio dos dados entre diversas bases em diferentes tipos de suportes físicos, sem que haja a necessidade de desenvolver programas que realizem as tarefas de transferência e validação dos dados.

Foram identificados artigos sobre dois grupos de formatos de intercâmbio:

- Formato MARC (Machine Readable Catalogue), com suas variações;
- Formato CALCO (Catalogação Legível por Computador), com suas variações.

Entretanto, existem outros formatos de utilização mais restritos e específicos, tais como:

- Formato próprio de um software comercial;
- Formato desenvolvido pela própria instituição.

A literatura reserva espaço para as aplicações, discussões de formato e tendências dos formatos MARC e CALCO com suas variações. Este fato se justifica por serem utilizados a nível internacional há vários anos. Dos artigos analisados, mais de 60% deles discutem o formato MARC padrão, não fazendo referência às suas variações. A maioria deles aborda aspectos ligados a sua estrutura interna; entretanto são encontradas discussões que vão desde o processo de transferência dos documentos deste formato para o MOSAICO (software utilizado para acesso às informações disponíveis na Internet), conforme descrito por **CANTRALL**³, até diversas críticas feitas por **LEAZER**¹¹ sobre a necessidade de realização do processo de "Normalização de Dados", muito comum no desenvolvimento de sistemas automatizados, que visa retirar informações duplicadas. Os quase 40% restantes têm como objetivo abordar aspectos ligados às variações do formato, inclusive a estrutura interna, muitas vezes comparando-as.

Podem ser encontrados outros formatos de catalogação que não estão associados à forma de armazenamento, tais como: OCLC, SGML, AACR2, etc. Eles definem a forma de apresentação dos dados, e não o armazenamento, apresentando algumas regras de catalogação. Para os sistemas que utilizam estes formatos, quando necessitam realizar intercâmbios, é necessário convertê-los para o formato MARC.

6.4.1 - Formato MARC

O formato MARC foi desenvolvido no final da década de 60, para permitir a transferência de dados entre os catálogos existentes, tendo logo recebido o certificado ISO (International Standard Organization) denominado ISO 2709, sendo amplamente utilizado nos Estados Unidos por um grande número de bibliotecas de organismos públicos e de instituições privadas para intercambiar ou difundir informações, nas formas mais variadas.

6.4.1.1 - Estrutura Interna

A sua estrutura interna contém o seguinte formato:

6.4.1.1.1 - Cabeçalho

Contém 24 caracteres, assim distribuídos:

- posição 1 a 5 - Tamanho do registro.
- posição 6 - Status do registro.
- posição 7 a 9 - Tipo do documento.
- posição 13 a 17 - Posição inicial no registro onde estão as informações bibliográficas.

As posições livres existentes no formato e as tabelas com os códigos do status do registro e tipo do documento podem ser criadas livremente para cada implementação.

PAREKH¹⁶ mostra em seu artigo uma comparação no conteúdo dos cabeçalhos dos registros entre as variações existentes no formato.

6.4.1.1.2 - Diretório de campos:

Estrutura que permite a localização e acesso para cada campo (informação bibliográfica) existente no registro. É composto por:

- Tag - Identificação da informação bibliográfica.
- Tamanho do campo - Número de caracteres existentes na informação bibliográfica.
- Posição inicial - Posição na Área de Dados onde se inicia o conteúdo da informação.

Os tags são identificados por 3 dígitos numéricos divididos em:

Tag	Descrição
001	Número do registro ou identificação do documento
008	Nível Bibliográfico (tipo do documento)
010-090	Códigos identificadores (ISBN, Coden, etc.)
1xx	Identificação do responsável pelo documento (nome do autor, conferência, etc.)
2xx	Outras informações do título (edição, imprensa, etc.)
3xx	Descritores físicos
4xx	Informações dos seriados (periódicos, séries, etc.)
5xx	Notas
6xx	Análise de assunto, incluindo resumos, etc.
7xx	Outras informações sobre autores ou títulos
8xx	Outras informações sobre os seriados

Cada símbolo x representa um dígito numérico compreendido de 0 a 9.

A partir de 1981 foi acrescentado mais um elemento à esta estrutura cuja utilização é definida pela própria implementação.

6.4.1.1.3 - Área de Dados - Local onde estão armazenados os conteúdos de todas as informações existentes sobre o documento.

Cada implementação que utiliza este formato, ou sua variação, utilizam tabelas que definem a combinação dos campos necessários para cada tipo de documento.

6.4.1.2 - Variações existentes

A partir deste padrão várias implementações, de diferentes níveis, foram desenvolvidas, tais como:

- A nível nacional - Formato de intercâmbio para um país, tais como: USMARC (USA), UKMARC (Reino Unido), AUSMARC (Austrália), CANMARC (Canadá), IBERMARC (Espanha), etc.
- A nível internacional - Utilizado para o intercâmbio entre vários países, tais como: UNIMARC, CCF, UNISIST-RM, etc.
- A nível Regional - Utilizados por grupos de países de uma região, tal como MARCAL, usado por países da América Latina e INTERMARC, usados pelos países europeus.
- A nível institucional - Utilizados por instituições específicas, ou grupo de instituições, como, por exemplo, os formatos: INIS, desenvolvido pela Agência Internacional de Energia Atômica (AIEA); AGRIS, desenvolvido pela Organização das Nações Unidas para a Agricultura e Alimentação (FAO); DEVSIS, desenvolvido pelo International Development Research Centre (IDRC); REPIDISCA, desenvolvido pelo Centro Panamericano de Engenharia Sanitária e Ciências do Ambiente (CEPIS), com sede em Lima - Peru, etc.

Dentre das principais variações do formato MARC existentes, podemos relacionar:

6.4.1.2.1 - UNISIST-RM (UNISIST Reference Manual)

O padrão UNISIST-RM foi criado a partir da necessidade de existência de um padrão de intercâmbio de dados nos órgãos voltados para os serviços secundários.

Para a sua implementação foram adotadas as seguintes providências:

- Retirada do tratamento de seriadados - Foram criadas tabelas com os campos necessários para cada nível bibliográfico, retirando-se os que tratavam dos seriadados.
- Definição de um carácter alfabético e dois numéricos como "Tag".

Outros formatos criados a partir deste, adotaram codificação de tags diferentes, tais como o DEVSIS (Development Information System) que utilizou no software MiniIsis tag formado por um carácter alfabético e três numéricos, sendo o último carácter utilizado como identificador de subcampos.

6.4.1.2.2 - UNIMARC (Universal MARC)

Uma das principais características deste formato são as alterações sofridas que possibilitaram o armazenamento de novos suportes físicos (mapas, músicas, arquivos de computador, etc.), além de permitir o armazenamento de caracteres não pertencentes ao alfabeto.

Em testes realizados foi constatado que este formato é um dos mais rápidos para recuperação dos dados, sendo utilizado para o intercâmbio de informações em CD-ROM.

A identificação dos campos é feita por tags agrupados de acordo com a característica da informação, conforme descrição abaixo:

Tag	Descrição
1xx	Bloco de identificação (Número do registro, ISSN, ISBN, data de criação, etc.)
2xx	Bloco de descrição da informação
3xx	Bloco de Notas
4xx	Bloco de entrada de ligações
5xx	Bloco de entrada relacionado
6xx	Bloco de análise de assunto

7xx	Bloco de responsabilidade intelectual
8xx	Bloco de uso internacional
9xx	Bloco de uso nacional

Cada símbolo x representa um dígito numérico compreendido de 0 a 9.

Não foi encontrado detalhamento sobre o conteúdo de cada bloco nos artigos analisados.

6.4.1.2.3 - CCF (COMMON COMMUNICATION FORMAT)

Formato criado sob a orientação da UNESCO com o objetivo de integrar as áreas de bibliotecas e de serviços de indexação e resumo. Foi criado a partir de comparações realizadas nos formatos em uso (UNISIST para a área de serviço de indexação e resumo e UNIMARC para a área de biblioteca), retirando-se de cada um deles as características que melhor atendessem a elas, sendo compatível com ambas e obedecendo à estrutura do formato ISO2709.

O registro neste formato é dividido em segmentos, estando as informações agrupadas de acordo com o tipo de documento. Ex. campo título: caso seja monografia, o título é alocado no segmento que contém as informações relacionadas ao nível bibliográfico de monografia, caso seja um artigo de periódico, o título estará alocado no segmento que contém todos os campos referentes ao nível bibliográfico de artigo de periódico.

A estrutura do registro é semelhante ao formato ISO2709 sendo utilizados quatro campos no diretório. Os tags estão agrupados da seguinte maneira:

Tag	Descrição
0xx	Informações para identificação do registro (número do registro, língua, etc.)
1xx	Informações para identificação do documento (ISSN, ISBN, Coden, etc.)
2xx	Título, edição, etc.
3xx	Autor pessoal, autor corporativo, conferência, etc.
4xx	Colação, imprensa, séries, etc.
5xx	Campos de Notas
6xx	Resumo, classificação e definição de assunto.

Cada símbolo x representa um dígito numérico compreendido de 0 a 9.

A necessidade de incorporação de novas características de documentos, tais como: novos suportes físicos, documentos que possuem múltiplos suportes, textos eletrônicos e documentos não textuais (som, vídeo, programa de computador, etc.), vêm fazendo com que o formato sofra atualizações que permitam catalogá-lo. Estas alterações são feitas de forma centralizada em uma coordenação que evite alterar os formatos já existentes, uma vez que sendo eles utilizados para intercâmbio de milhões de documentos por milhares de bases de dados, qualquer alteração causam um custo elevado.

6.4.2 - FORMATO CALCO

O formato CALCO nasceu de uma dissertação de mestrado em Ciência da Informação de Alice Príncipe Barbosa com o objetivo de criar normas de catalogação para as bibliotecas brasileiras. Foi desenvolvido pela Fundação Getúlio Vargas, no Rio de Janeiro.

A partir de 1985, com o apoio da Organização dos Estados Americanos (OEA), foram elaborados estudos para a criação de um formato para o intercâmbio das informações nas línguas espanhol e português, produzidas nos países da América Latina, para a base de dados da Library of Congress. Este formato foi denominado Formato MARCAL (MARC para América Latina).

Sua estrutura interna contém o formato abaixo descrito, que refere-se ao formato IBICT, descrito no item 6.4.2.1. As pequenas variações existentes entre este formato e o CALCO estão ressaltadas nos próprios itens.

a) Cabeçalho:

Contém 24 caracteres, assim distribuídos:

- posição 1 a 5 - tamanho do registro.
- posição 6 - Status do registro (Formato IBICT).
- posição 7 - Tipo de material (Formato IBICT).
- posição 8 - Nível bibliográfico (Formato IBICT).
- posição 11 - Número de indicadores do documento.
- posição 12 - Tamanho do identificador de subcampo.
- posição 13 a 17 - Posição inicial da área de dados na estrutura do registro.
- posição 18 - Nível de catalogação (Formato IBICT).
- posição 21 a 24 - Mapa do diretório.

As demais posições estão disponíveis para uso futuro ou para variações do formato.

b) Diretório:

Estrutura que permite a localização e acesso para campo (informação bibliográfica) existente no registro. É composto por:

- Tag - Armazena o código de identificação do campo.
- Tamanho do campo - Tamanho do campo na área de dados.
- Posição inicial - Posição do primeiro caracter do campo na área de dados.

c) Área de dados:

Local onde estão armazenadas as informações referentes ao documento.

Os tags são identificados por 3 dígitos numéricos divididos em:

Tag	Descrição
001	Número do documento (6 caracteres)
008	Campo de controle, tais como: tipo de acesso, nível bibliográfico, etc. (45 carac.)
010	Agência registradora e/ou depositária (12 caracteres)
0xx	Exceto as relacionadas acima, contém informações de caracter geral, tais como: ISBN, idioma, área geográfica, etc.
1xx	Informações referentes a autoria, tais como: autor pessoal, entidade coletiva, dados da conferência, etc.
2xx	Informações referentes ao título, edição e imprensa.
3xx	Informações referentes a colação.
4xx	Informações sobre a série.
5xx	Informações sobre notas, resumo, etc.
6xx	Informações sobre descritores, cabeçalho de assunto, etc.
7xx	Informações sobre os autores secundários.
8xx	Informações sobre os dados secundários da série.

A partir deste formato foram desenvolvidos os seguintes formatos:

6.4.2.1 - Formato *IBICT (Intercâmbio Bibliográfico e Catalográfico)*

Desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), sendo hoje o formato padrão brasileiro para comutação bibliográfica.

6.4.2.2 - Formato *Bibliodata*

Desenvolvido pela Fundação Getúlio Vargas com o objetivo de realizar a comutação bibliográfica entre diversas bibliotecas. Este formato possibilitou a criação de uma base de dados com o acervo bibliográfico das bibliotecas participantes.

A literatura tem se comportado de forma tendenciosa para o formato MARC e suas variações, mostrando as implementações que os utilizam e realizando comparações entre elas.

Referências Bibliográficas:

- 1 BALES, Kathleen. Format Integration: Coordination the Implementation. **Information Technology and Libraries**, v. 9, n. 2, p. 167-173, jun 1990.
- 2 BRANGER, D. A Computerized Catalogue for Unpublished Recordings. **IASA**, v.1, p.45-53, may 1993.
- 3 CANTRALL, D. From MARC to Mosaic: Progressing Toward Data Interchangeability at the Oregon State Archives. **Archives and Museum Informatics**, v. 8, n. 1, p. 4-12, spring 1994.
- 4 DAVIS, Stephen P. Format Integration: Handling Serials and Mixed Media. **Information Technology and Libraries**, v. 9, n. 2, p. 162-167, jun 1990.
- 5 GAYNOR, E. Cataloguing Eletronic Texts: the University of Virginia Library Experience. **Library Resources & Technical Services**, v. 38, n. 4, p. 403-413, oct 1994.
- 6 GIBBS, George. Applying Format Integration: An Operational Test. **Information Technology and Libraries**, v. 9, n. 2, p. 173-178, jun 1990.
- 7 HOPKINSON, Alan. International Data Exchange: a Question of Formats. **Asian Libraries**, v. 2, n. 1, p. 30-39, mar 1992 .
- 8 INSTITUTO Brasileiro de Informação em Ciência e Tecnologia. **Formato IBICT: Formato de Intercâmbio Bibliográfico e Catalográfico / Instituto Brasileiro de Informação em Ciência e Tecnologia**. Brasília: IBICT, 1987, 400 p.
- 9 JONES, Wayne, QUEINNEC, Young-Hee. Format Integration and Serials Cataloguing. **Serials Librarian**, v. 25, n. 1-2, p. 83-95, 1994.
- 10 LEATHEM, C. A. An Examination of Choice of Formats for Cataloging Nontextual Serials. **Serials Review**, v. 20, n. 3, p. 59-67, 1994.
- 11 LEAZER, Gregory H. An Examination of Data Elements for Bibliographic Description: Toward a Conceptual Scheme for the USMARC Formats. **Library Resources & Technical Services**, v. 36, n. 2, p. 189-208, apr 1992.
- 12 McCALLUM, Sally H. Format Integration: Handling the Additions and Subtractions. **Information Technology and Libraries**, v. 9, n. 2, p. 155-161, jun 1990.
- 13 MORGAN, Eric Lease. The MARC Reader. **Computers in Libraries**, v. 11, n. 11, p. 52-52, dec 1991.
- 14 MOYA, Felix de, HIPOLA, Pedro. The New IBERMARC Format Alongside UNIMARC. **International Cataloguing & Bibliographic Control**, v. 20, n. 3, p. 41-43, jul/sep 1991.
- 15 MURPHY, Catherine. MARC Accountability... Revisiting a Critical Issue. **Emergency Librarian**, v. 19, n. 3, p. 26-28, jan/feb 1992.
- 16 PAREKH, Harsha. Bibliographic Record Structure and Communication Formats. **Lucknow Librarian**, v. 23, n. 1, p. 1-24, jan/mar 1991.
- 17 ROBREDO, Jaime, Cunha, Murilo B. da, colab. **Documentação de hoje e de amanhã: Uma abordagem informatizada da biblioteconomia e dos sistemas de informação**. 2a ed. rev. e ampl. Brasília: 1986, 400 p.
- 18 SIMMONS, Peter. Using CCF: The Common Communication Format. **Information Technology and Libraries**, v. 5, n. 4, p. 285-294, dec 1986.

6.5 - INDEXAÇÃO

O processo de indexação consiste em identificar e atribuir a um documento os conceitos de uma área do conhecimento nele existentes, traduzindo-os em termos, palavras ou frases que serão utilizados para a sua localização.

Desde os controles de livros e periódicos existentes em Bibliotecas ou Centros de Informação, através de listagens de computador (saída dos dados) e cartões perfurados (entrada dos dados), até as tarefas interativas de indexação, através de indexação automática ou inteligência artificial, parte do processo de indexação vem evoluindo em paralelo à computação. Esta evolução permite, até certo ponto, maior eficiência e agilidade na identificação dos conceitos dos documentos a serem indexados. Ela é sentida não apenas na tarefa de indexar documentos, mas também em outras atividades, como as burocráticas (indexação de documentos em editores de textos), controles de qualidade, etc. Este uso cotidiano vem permitindo maior interação e desmistificação dos conceitos. Segundo **HODGE**²³, o avanço da computação gerou uma redução nos custos de indexação atendendo aos baixos orçamentos e às políticas de redução de custos.

Na literatura convencional, a indexação é o assunto onde foi identificada a maior quantidade de artigos, se comparada com as demais etapas do projeto (cerca de 20%). Esta etapa, na literatura é subdividida em diversos assuntos. Alguns destes assuntos ocupam tópicos específicos dentro das publicações secundárias, como é o caso da indexação automática. No **ARIST**^{29,48}, são encontradas algumas revisões sobre estes tópicos. Entretanto, são poucos os artigos que discutem a indexação nos aspectos de projeto e política. Dos analisados apenas **LANCASTER**²⁸ e **GUNCHARUK**²⁰ discutem avaliação de indexação e **FIDEL**¹⁷ e **CARNEIRO**¹³ discutem as políticas de indexação. A grande maioria deles abordam aspectos específicos da indexação e as tecnologias e técnicas emergentes.

Na literatura, diversos autores apresentam diversas definições para a indexação, tais como:

- A indexação deve ser capaz de representar em palavras um documento em detalhes, e ao mesmo tempo ser genérico, conforme **GUNCHARUK**²⁰.
- A indexação é o processo, intelectual ou automatizado, no qual os assuntos tratados pelos documentos são identificados e expressos em palavras, conforme **ALBRECHTSEN**¹.

6.5.1 - Políticas de Indexação

O processo de implantação de uma política de indexação deve considerar as variáveis institucionais, humanas e financeiras pois, segundo **CARNEIRO**¹³, “definir estas variáveis, estabelece os princípios e critérios que servirão de guia na tomada de decisão para otimização do serviço, racionalização dos processos e consistência das operações nele envolvidas”, sendo esta otimização um dos objetivos de uma política de indexação. **CARNEIRO**¹³ relaciona os aspectos a serem considerados para cada uma destas variáveis, que devem ser analisados e definidos relacionando-os ao processo de recuperação da informação.

Estas variáveis são utilizadas para definir o grau de qualidade da indexação e do indexador, e conseqüentemente do sistema, sendo encontradas com freqüência na literatura sob o tópico de “avaliação de sistema”. Estes artigos avaliam diversos sistemas utilizando diversos critérios de avaliação. Lancaster é um dos autores que mais discute o problema, possuindo dois

livros escritos além de diversos artigos publicados em periódicos e anais de conferências. Os principais critérios considerados são:

6.5.1.1 - Exaustividade

Segundo **CARNEIRO**¹³, exaustividade é “uma medida de extensão em que todos os assuntos, discutidos em um certo documento, são reconhecidos na operação de indexação e traduzidos na linguagem do sistema”.

O nível de exaustividade é a quantidade máxima de assuntos, ou termos, utilizados na indexação. Para um documento este número poderá ser inferior ao nível definido. A sua definição depende de decisões políticas que estão relacionadas ao objetivo e aos usuários do sistema (Sistemas mais genéricos, exaustividade menor).

6.5.1.2 - Especificidade

Na literatura diversas definições podem ser encontradas para especificidade, como **GUNCHARUK**²⁰ que afirma como sendo “o uso pelo indexador das palavras que melhor representam o escopo e o conteúdo descrito no documento”.

O grau de especificidade é definido a partir da decisão de indexar um documento com maior ou menor profundidade, estando ele diretamente relacionado ao nível de precisão do sistema. Quanto maior for a profundidade da indexação maior será a precisão. A definição deste grau depende de decisões políticas que estão relacionadas ao ambiente e usuários do sistema (Sistemas mais genéricos, especificidade menor).

6.5.1.3 - Tipo de vocabulário

Conforme **CHANG**¹⁵, um bom esquema de indexação deve fornecer uma recuperação rápida dos documentos desejados e uma manutenção fácil dos dados armazenados no computador. A linguagem de indexação mais apropriada deve ser definida através de decisões políticas, considerando os objetivos e usuários do sistema. **CARNEIRO**¹³, afirma que a linguagem de indexação afeta o desempenho de um sistema de recuperação de informações em dois pontos:

- Na estratégia de busca - Estabelecendo a precisão com que o técnico da busca deve descrever os interesses do usuário.
- Na indexação - Estabelecendo a precisão com que o indexador deve descrever os assuntos dos documentos.

O tipo do vocabulário a ser utilizado pelo sistema deverá ser livre ou controlado (estes tipos serão discutidos adiante).

6.5.1.4 - Índices de revocação e precisão do sistema

Na literatura sobre avaliação de sistemas estes índices são considerados os mais importantes para a avaliação do SRI, por estarem associados a “capacidade do sistema em assegurar a recuperação de um número desejável de documentos relevantes (revocação) e a capacidade de impedir a recuperação de documentos não relevantes (precisão)”, segundo definição de **CARNEIRO**¹³.

Para **LANCASTER**²⁸, as taxas de precisão e de revocação expressam a capacidade de filtragem do sistema, sua habilidade em deixar passar o que é solicitado e impedir o que não interessa.

Os níveis de revocação e precisão estão inversamente relacionados à amplitude da recuperação (quanto maior a revocação, menor é a precisão), podendo ser definidos níveis variáveis de acordo com o objetivo do usuário. Esta variação está condicionada à forma com que a consulta é realizada e aos termos utilizados.

Outros fatores estão associados à qualidade da indexação, tais como: escolha dos termos no tesouro, definição de pesos para cada termo escolhido para o documento (refletindo a importância do tema para o documento), exatidão na escolha dos termos, nível de pré-combinação dos termos, etc., itens citados por **LANCASTER**²⁸ e **FIDEL**¹⁷. **PAO**³⁸ relacionam outros fatores que devem ser considerados no projeto de indexação tais como: seleção dos termos, consistência, uniformidade, especificidade e utilidade, além de definir a forma com que serão utilizados (maiusculo, minúsculo, singular, plural, parênteses, etc.)

6.5.2 - Projeto de Indexação

Para a implantação da estrutura de indexação, diversos aspectos devem ser considerados, tais como:

6.5.2.1 - Escolha do Vocabulário

A escolha do vocabulário afeta os termos utilizados para a indexação e para a recuperação. Estes termos devem ser concisos e auto-explicativos, descrevendo todos os assuntos envolvidos na área de interesse.

A literatura tem mostrado diversos projetos de avaliação de sistemas, tais como SMART, CRAWFORD, etc., avaliando o uso dos seguintes tipos de vocabulários:

- Vocabulário livre - Tipo de vocabulário em que o indexador escolhe os termos que deseja utilizar sem se importar com a sua padronização. A vantagem deste vocabulário é a indexação mais rápida, podendo ser utilizado pessoal menos qualificado. A desvantagem é a dificuldade maior na elaboração da consulta, havendo problemas de recuperação, tais como: grafia, acentuação, etc.
- Vocabulário Controlado - Por ser a base de construção de tesouro, este é o tipo de vocabulário que contém a maior quantidade de estudos na literatura. Consiste em utilizar termos previamente cadastrados em uma lista pré-definida de palavras. Segundo **LEMOS**³¹, o vocabulário controlado é um conjunto de descritores autorizados, ligados entre si por uma estrutura de relações hierárquicas, associativas e/ou de equivalências terminológicas. A indexação neste tipo de vocabulário é realizada de forma lenta pela necessidade de consultas constantes a esta lista devendo ser realizada por pessoal mais qualificado. Permite encontrar os documentos de forma mais precisa.

As regras para a formação de um vocabulário controlado devem considerar os seguintes itens:

- As palavras devem ser simples e consistentes para os objetivos da recuperação eliminando adjetivos, verbos, etc.;
- Devem ser complexas quando formadas por descritores compostos cujas palavras sejam dos grupos: Substantivo-adjetivo, Substantivo-negação-adjetivo,

Substantivo-preposição, etc., porém é aconselhável separá-las em dois ou mais termos distintos;

- Devem ser considerados os sinônimos entre as palavras, associando-os;
- Devem haver padronizações de grafia, optando por palavras no singular;
- Definir o setor ou organização que será responsável pela atualização do vocabulário. Neste tópico, a literatura sugere que o controle seja centralizado, não descartando, porém, as sugestões de forma descentralizada de novos termos.

SVENONIUS⁵¹ avalia em seu artigo alguns estudos comparativos entre os vocabulários controlados e livres, relacionando as vantagens e desvantagens de cada um deles.

Nos artigos analisados, foram encontrados os artigos de **AUSTIN**³, **LEMOS**³¹ e **SVENONIUS**⁵¹ que discutem aspectos ligados ao vocabulário controlado, sendo que apenas os dois primeiros discutem os itens ligados à sua criação.

A partir da necessidade da evolução e estruturação dos termos de um vocabulário, surgiu uma lista que hoje é a mais utilizada e, conseqüentemente, a mais discutida na literatura que é o tesouro.

6.5.2.2 - Construção de tesouros

O sucesso do uso de um tesouro é iniciado com o projeto de construção. Este projeto deve envolver:

- Identificação das fontes de coleta do vocabulário através de dicionário especializado, glossário, índices de outras bases de dados ou outras fontes;
- Definição da largura e extensão do tesouro;
- Seleção dos termos que serão utilizados e montagem do seu encadeamento;
- Definição das regras de validação;
- Definição da estrutura interna de armazenamento dos dados e dos índices de acessos.

Segundo **KROOKS**²⁷ os maiores problemas para a construção de tesouro estão relacionados a:

- Organização dos termos - Criação das ligações entre os termos organizados hierarquicamente e os não organizados.
- Forma dos termos - Definição das classes das palavras (singular/plural, hifenação, etc.).
- Entrada dos vocabulários - Definição das várias formas de apresentação dos termos para, através das ligações, apontar para os termos cadastrados.

A criação do tesouro deve ser feita indicando para cada termo: os seus termos relacionados, os termos mais amplos, os termos mais restritos e outros termos associados que podem ou devem ser utilizados. Estes termos são assim definidos:

- Termos associados - Relaciona as palavras que podem ou devem ser utilizadas para acessar o termo desejado. A sua utilização representa a ligação de termos não estruturados. Podem ser divididos em dois grupos:
 - **USE FOR** - Indica os termos que devem ser utilizados em substituição aos listados no tesouro, podendo ser representados por sinônimos ou outros termos que possuem mesmo significado, podendo ocorrer a referência de dois termos

combinados, utilizando AND. Para estes termos devem existir entradas específicas no tesouro. São representados no tesouro pelo símbolo UF (Used For).

- Termos “Proibidos” - Indica os termos que não podem ser utilizados. São representados no tesouro pelo símbolo USE.
- Termos Relacionados (Related Term) - São termos associados ao termo corrente que podem ser consultados quando necessário. São representados no tesouro pelo símbolo RT.
- Termos Amplos (Broader Term) - Identifica o termo que está hierarquicamente acima do termo corrente (Pai). É representado no tesouro pelo símbolo BT.
- Termos Restritos (Narrower Term) - Relaciona os termos que estão hierarquicamente abaixo do termo corrente (Filhos). É representado no tesouro pelo símbolo NT.

BOOTH⁷ citou o tesouro abaixo para exemplificar a utilização destes termos:

ARCHITECTURE

NT BUILDINGS

BUILDINGS

BT ARCHITECTURE

NT CIVIC BUILDINGS

INDUSTRIAL BUILDINGS

CIVIC BUILDINGS

UF CIVIC CENTRES

TOWN HALLS

BT BUILDINGS

CIVIC CENTRES USE CIVIC BUILDINGS

FACTORIES USE INDUSTRIAL BUILDINGS

INDUSTRIAL BUILDINGS

UF FACTORIES

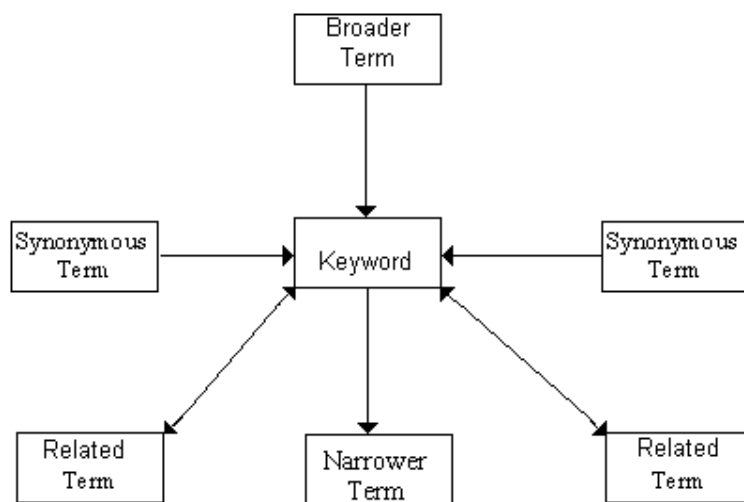
BT BUILDINGS

TOWN HALLS USE CIVIC BUILDINGS

Na literatura outras formas de construção e estruturação de tesouro são encontrados, tais como:

- Montagem representando a estrutura do conhecimento, descrito por **BAUER**⁵;
- Montagem utilizando análise facetada, descrito por **RACKMORE**⁴¹;
- Montagem utilizando software de indexação baseado em sistemas especialistas, descrito por **SCHUEGRAF**⁴⁷.

PAO³⁸ representa a estrutura hierárquica de um tesouro no diagrama abaixo:



Dos artigos analisados, foi observado que 25% tratam da construção e estruturação de tesouros.

6.5.2.3 - Definição das regras de validação de tesouro

Dos artigos analisados que discutem as etapas para a elaboração de tesouro, apenas **FISCHER**¹⁸ e **KENT**²⁵ destacam as regras de validação e consistência como sendo o aspecto mais importante a ser considerado. Segundo eles, para construir um tesouro devem ser considerados as seguintes regras:

- Um termo não pode ficar sem o Pai (BT);
- Um termo não pode apontar para outros que não existam no tesouro;
- No RT os dois termos devem apontar-se simultaneamente.

6.5.2.4 - Armazenamento dos índices

Considerando “índice” como sendo o termo associado ao documento que ele identifica, a escolha do armazenamento dos índices representa a forma com que estão armazenados os termos representativos dos documentos, com suas respectivas ligações hierárquicas e ponteiros para o local onde está armazenado. Para a escolha deste formato diversas decisões devem ser tomadas, conforme relacionadas abaixo.

Em um primeiro nível de decisão deve ser escolhido entre:

- Desenvolvimento do seu próprio software (programa de computador) de armazenamento e recuperação, utilizando as técnicas disponíveis de organização de arquivos e estrutura de dados;
- Escolha de um software comercial que realize todo o processo de indexação, armazenamento e recuperação das informações.

Para a primeira opção, **CHANG**¹⁵ relaciona diversos tipos de organização de arquivo que permitem o armazenamento dos dados. Estas organizações (SAM - Sequential Access Method, DAM - Direct Access Method, ISAM - Indexed System Access Method e VSAM - Virtual System Access Method) possibilitam acessos seqüenciais e diretos, exceto a organização SAM, utilizados na recuperação. Ainda nesta situação **CHANG**¹⁵ e **KWON**²⁶ relacionam os tipos de estruturas de dados hierárquicas (árvore binária, B e B+) indicadas para este armazenamento por possibilitarem pesquisas a qualquer ponto do tesouro com reduzido número de acessos a disco.

Utilizando estes recursos, um modelo de implementação automatizado de tesauro pode ser desenvolvido, conforme sugestão de **CRAVEN**¹⁶.

Para a escolha de um software de indexação a primeira tarefa a ser considerada é a definição dos critérios de avaliação do software. Estes critérios foram definidos por **GANZMANN**¹⁹ e devem considerar:

- A construção e manutenção do tesauro;
- As saídas geradas pelo tesauro;
- Indexação e recuperação de termos com o tesauro;
- Intercâmbio, integração e compatibilidade do vocabulário;
- Compatibilidade técnica do software com equipamento;
- Documentação, suporte, preço, etc.

Com base nestes critérios, **RITZLER**⁴⁴ avaliou os softwares INDEX, PROTERM e TMS, concluindo que o INDEX é o mais eficiente entre eles e o PROTERM o mais indicado na relação preço x performance. Outros autores analisaram outros softwares, tais como: **ROWLEY**⁴⁵ que analisou o software TRACER, **BURGIN**¹⁰ que analisou FASIT e **JONES ET ALL**²⁴ e **CANGANELLI**¹² que analisaram o software INDEX.

Apenas **CHANG**¹⁵, **KWON**²⁶ e **CRAVEN**¹⁶ discutem aspectos de desenvolvimento de software próprio, os demais avaliam e divulgam softwares de indexação. Esta avaliação é fácil de ser entendida por considerar os aspectos comerciais envolvidos.

6.5.3 - Implementação da Indexação

Criada a estrutura, as rotinas necessárias para a operação do SRI, na sua etapa de indexação são:

6.5.3.1 - Análise de Assunto

A fase de análise de assunto, se confunde com a própria etapa de indexação do documento, pois nela são identificados os termos que melhor o representam.

ALBRECHTSEN¹ relaciona os dois procedimentos necessários para a indexação:

- a análise dos assuntos dos documentos expressando as informações percebidas em comandos lingüísticos concretos. Esta análise pode ser feita, de uma forma simplista, considerando os assuntos como palavras do texto. Neste tipo de indexação pode ser utilizada a indexação automática;
- associar ao documento os termos previamente conhecidos no tesauro ou outro esquema de classificação. Esta associação é feita através da interpretação do conteúdo do documento utilizando a capacidade intelectual do indexador.

Na realidade, a análise de assunto define o trabalho intelectual do indexador na identificação dos conceitos e posterior escolha dos termos que representam o documento, através de processos automatizados ou não, utilizando vocabulário controlado ou não. Por este motivo, na literatura são encontrados artigos que discutem a análise de assunto incluídos em outros assuntos, sendo o principal deles a indexação automática.

6.5.3.2 - Utilização de tesouros

A utilização do tesouro no processo de indexação de documento para o SRI é a técnica mais utilizada, independente da categoria do sistema. Esta afirmação é comprovada com a quase totalidade dos artigos sobre indexação que discutem os diferentes aspectos do tesouro, tais como construção, avaliação de softwares, formas de utilização, etc.

O uso do tesouro pode ser feito de três maneiras:

6.5.3.2.1 - Manual

Representa a localização dos termos do documento utilizando livros ou catálogos manualmente. Após a localização, os termos serão informados e anexados às demais informações do documento.

6.5.3.2.2 - Automatizada

Tipo de indexação que utiliza o computador como ferramenta de trabalho. Nela os termos do tesouro estão armazenados em dispositivos magnéticos que ficam disponíveis para a escolha do(s) termo(s) apropriado(s) pelo indexador através de recursos de programação. Com este uso, tem-se as seguintes vantagens:

- Agilização na entrada e escolha do termo através de teclas pré-programadas;
- Realização de procedimentos de verificação e consistência dos termos escolhidos, com os dados do documento. Como exemplo, podemos citar: para um determinado termo escolhido deve haver um conjunto de outros termos ou palavras no documento. Para a verificação deve ser comparada a indexação do documento com outros de temas similares.

6.5.3.2.3 - Automática

Tipo de indexação em que a escolha e associação dos termos ao documento é feita pelo computador, sem a interferência do indexador. Este tipo de indexação, dada a sua importância, será analisada no item abaixo.

6.5.3.3 - Indexação automática

A indexação automática, segundo **MAMFRIM**³⁴, consiste na mecanização do processo de indexação no todo ou em parte, visando estabelecer rotinas que reduzam a interferência da subjetividade do indexador, tanto na análise de assunto dos documentos, identificando os conceitos, quanto na seleção dos termos significativos.

Na literatura diversos artigos abordam a indexação automática, havendo tópicos específicos na literatura secundária. Estes artigos mostram alternativas e propostas de implementação, porém grande parte deles mantém o caráter manual da indexação por realizarem consultas a vocabulários armazenados no computador, caracterizando uma indexação automatizada e não automática. Dentre os artigos analisados a literatura nacional mostrou artigos mais específicos sobre os procedimentos para esta implementação. A literatura estrangeira aborda avaliações ou apresentações de sistemas que implementam a indexação automática.

6.5.3.3.1 - Procedimentos de utilização

Sistematizando os procedimentos de implementação de indexação automáticos, feitos pelo computador, encontramos as seguintes etapas:

- Separação das palavras do texto, resumo ou título;
- Eliminação de palavras não significativas;
- Análise morfológica das palavras - Retirada de sufixos, prefixos, conjugações verbais até atingir a raiz da palavra, conforme algoritmo proposto por **SAVOY** ⁴⁶;
- Análise sintática - Identificação de sujeitos, verbos, advérbios, substantivos, etc.;
- Montagem de tabela estatística com frequência das palavras, considerando aquelas com frequência intermediária;
- Validação das palavras selecionadas com o vocabulário utilizado.

6.5.3.3.2 - Problemas existentes

Alguns dos problemas relacionados à Indexação Automática, foram citados por **AUSTIN** ³, como sendo:

- O computador não é capaz de distinguir palavras de mesma grafia com significados diferentes. Apenas a avaliação do indexador é capaz de associá-la ao documento;
- Identificação de termos associados à palavras chaves mostram diferentes relacionamentos gramaticais;
- Identificação da classe gramatical correta da palavra.

6.5.3.3.3 - Comparação com a indexação manual

Como resultado da comparação entre Indexação automática e manual, **VIEIRA** ⁵³ chegou as seguintes conclusões:

- Indexação Manual - O indexador ao analisar um documento manualmente, tende à utilização de um número maior de descritores, e conseqüentemente reduz a precisão do sistema;
- Indexação Automática - A utilização dos procedimentos de indexação automática de um documento realiza a identificação de três tipos de descritores:
 - Descritores Pontuais - Representado pelas palavras com baixa frequência. A utilização destes descritores como termos de acessos possibilita recuperação mais precisa do documento.
 - Descritores de Facetas - Representado pelas palavras com média frequência. Estes descritores representam os termos mais significativos do documento.
 - Descritores de Escopo - Representado pelas palavras com alta frequência. Caracteriza a indexação macro-categorizada.

Além destas conclusões, **VIEIRA** ⁵³ afirma que “o uso de termos simples e livres apresenta melhor flexibilidade para a formulação da estratégia de busca e o uso da indexação automática, utilizando estes termos, oferece melhor resultado na recuperação”.

Na literatura existem diversas propostas de indexação automática, tais como:

- Indexação automática através de análise morfossintática das palavras, proposto por **BRITO** ⁸;
- **LOCHBAUM** ³³ propõe a técnica chamada de “Indexação Semântica Latente” que, através da criação em uma matriz dos termos dos documentos agiliza o processo de recuperação;

- Aplicação de modelos estatísticos para identificação dos índices dos documentos, proposto por **SRINIVASAN** ⁴⁹.

6.5.4 - A Indexação de acordo com o conteúdo

A utilização das técnicas de indexação para as categorias mais comuns de SRI's, tais como os sistemas textuais, são padronizadas, porém, existem sistemas que necessitam de tratamento específico. As discussões destes tratamentos são encontradas em artigos que tratam de um tipo específico de conteúdo, sendo a imagem o conteúdo discutido em mais da metade deles. Podemos relacionar alguns dos conteúdos encontrados como sendo:

6.5.4.1 - Música

A indexação de Sistemas de Recuperação que têm a música como conteúdo possui características diferentes dos sistemas textuais. Estas características são necessárias pelo fato de que o indexador “deve ser capaz de tratar o assunto com múltiplos tópicos”, conforme opinião de **McKNIGHT** ³⁵.

Segundo ele, a terminologia utilizada no tesauro deve ser diferente de acordo com a área de interesse. Por exemplo, os termos utilizados no campo de etnomusicológico devem ser diferentes de outros campos. A indexação dos documentos musicais deve contemplar os aspectos geográficos e culturais, o gênero musical, os instrumentos utilizados, além dos autores, data, motivo, estilo, etc.

McKNIGHT ³⁵ relaciona alguns trabalhos que avaliam Sistemas de Recuperação musicais e lista as etapas para a construção de um tesauro para estes sistemas.

6.5.4.2 - Hipertexto

O principal motivo de um sistema de hipertexto ou hiperímídia é proporcionar ao usuário mecanismos que permitam consultas livres através de “navegação” no texto ou acessos sequenciais. Para isto, é necessário o desenvolvimento de uma eficiente estrutura interligada da informação que atenda às necessidades do usuário de forma fácil, direta e objetiva. Esta estrutura deve interligar os índices de acessos do documento à sua descrição ou à outros documentos.

Em revisão bibliográfica realizada por **ARENTS** ² são relacionados artigos com diferentes técnicas de indexação para informações hipertextual, avaliando as diferenças existentes entre o espaço do documento (o documento no sistema) e o espaço dos índices utilizados para a navegação no sistema (índices que identificam o documento) em cada uma delas. Ambos os espaços formam o espaço da informação que representa o conteúdo total da informação armazenada no sistema. Neste artigo é proposta a “Técnica de hiperindexação semântica” utilizada em sistemas de hipertexto textual baseada na utilização de tesauro, explicando a sua estrutura interna e forma de implantação.

6.5.4.3 - Imagens

Segundo **LEUNG** ³², nos últimos dez anos a maioria dos trabalhos descritos na literatura sobre indexação de imagens foi feita com propósitos históricos. Os sistemas são desenvolvidos para assuntos específicos utilizando termos e relacionando problemas próprios.

A indexação de um SRI que armazena imagens possui um alto grau de subjetividade pois é feita de acordo com o nível de conhecimentos do indexador.

Na literatura, diversos autores tentam acabar com esta subjetividade desenvolvendo modelagens de dados que permitam automatizar este processo, com por exemplo **CASSI ET ALL**¹⁴, **RABITTI**⁴⁰ e **LEUNG**³². Até o momento essas propostas são de difícil implementação.

Para a avaliação de uma imagem, **LEUNG**³² relaciona três níveis de descrição:

- Identificação dos objetos - Consiste em descrever a imagem segundo os objetos, pessoas, etc.
- Interpretação dos objetos - Consiste em avaliar a imagem de acordo com o ambiente que ela retrata, tais como tradições culturais. Neste nível de descrição fatores subjetivos de conhecimento do avaliador sobre o assunto influenciam no resultado da descrição.
- Entendimento dos significados intrínsecos - Consiste em avaliar a imagem de acordo com os itens que pode representar. Esta avaliação é totalmente subjetiva, e a capacidade de percepção do avaliador é o fator primordial.

Além destas descrições os atributos bibliográficos da imagem devem ser considerados para a sua indexação, tais como: autor, data, local de publicação, etc.

6.5.5 - A Indexação de acordo com o suporte físico

A utilização das técnicas de indexação pode variar de acordo com o suporte físico em que foi publicado o documento. **WELLISCH**⁵⁵ identifica dois tipos de suportes físicos: Documentos com apenas um autor (livros) ou múltiplos autores (periódicos). Poucos artigos foram localizados sobre este tópico, sendo analisado apenas o de **WELLISCH**^{55,56}.

6.5.5.1 - Livros

A indexação de livros ou outros documentos que contenham apenas um autor ou texto, deve ser feita da seguinte maneira:

- Leitura superficial do conteúdo;
- Após a leitura, examinar cada página, marcando as frases, palavras ou nomes candidatos a índices de acessos;
- Inclusão, alteração ou retirada dos índices candidatos, transformando-os em termos do tesouro;
- Avaliação da necessidade de incluir nomes de pessoas, lugares ou eventos como índices de acessos.

Em outro artigo, **WELLISCH**⁵⁶ comenta sobre a padronização de termos utilizados em livros (singular/plural, hifenação. etc.) que facilita a padronização dos índices.

6.5.5.2 - Periódicos

Os periódicos ou outros documentos que contenham mais de um autor ou texto, devem ser indexados da seguinte maneira:

- Devem ser lidos integralmente a introdução, a conclusão e o resumo. O conteúdo deve ser lido superficialmente;

- Devem ser identificados os índices a partir das partes lidas, preferencialmente do resumo;

Obs. A indexação de enciclopédia é feita de forma idêntica a de periódicos, conforme afirmação de **WELLISCH** ⁵⁶.

Com base nas avaliações feitas acima, podemos concluir que:

- Poucas são as alternativas de indexação que não utilizem o tesauro como ferramenta, variando apenas o tipo de vocabulário utilizado;
- As indexações manual e automatizada têm como diferença apenas o uso do computador como ferramenta de agilização do trabalho;
- O uso da indexação automática é possível apenas nos sistemas de textos completo. Para os demais, a presença do indexador é fundamental;
- O recurso do hipertexto deve ser utilizado em sistemas que interligam vários documentos ou associe a eles informações complementares. Para os demais sistemas, este recurso não se aplica.

Mesmo que surjam novas tecnologias para agilizar o processo de indexação, o trabalho intelectual feito pelo indexador continuará sempre presente, ainda que de forma indireta.

Referências Bibliográficas:

- 1 ALBRECHTSEN, Hanne. Subject Analysis and Indexing: From Automated Indexing to Domain Analysis. **Indexer**, v. 18, n. 4, p. 219-224, oct 1993.
- 2 ARENTS, Hans C., BOGAERTS, Walter .F. L. Navigation without Links and Nodes without Contents: Intensional Navigation in a Thirs-order Hypermedia System. **Hypermedia**, v. 5, n. 3, p. 187-204, 1993.
- 3 AUSTIN, Derek. Vocabulary Control and Information Technology. **Aslib Proceedings**, v. 38, n. 1, p. 1-15, jan 1986.
- 4 BATTY, David. Microcomputers in Index Language Design and Development. **Microcomputers for Information Management**, v. 1, n. 4, p. 303-312, dec 1984.
- 5 BAUER, Gerd. Promoting Creative Processes by a Thesaurus-like Representation of Knowlwdge Structure. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 62-69.
- 6 BELL, Hazel K. Vive la Difference! The Survival of the Softest. **Indexer**, v. 18, n. 4, p. 231-236, oct 1993.
- 7 BOOTH, Pat F. Thesauri-their Uses for Indexers. **Indexer**, v. 15, n. 3, p. 141-144, apr 1987.
- 8 BRITO, Marcílio de. Sistemas de Informação em Linguagem Natural: em Busca de uma Indexação Automática. **Ciência da Informação**, v. 21, n. 3, p. 223-232, set/dez 1992.
- 9 BUCHAN, R. L. Intertwining Thesauri and Dictionaries. **Information Services & Use**, Amsterdam, v. 9, n. 3, p. 171-175, 1989.
- 10 BURGIN, Robert, DILLON, Martin. Improving Disambiguation in FASIT. **Journal of the American Society for Information Science**, New York, v. 43, n. 2, p. 101-114, mar 1992.
- 11 BUTCHER, Sharon, PAO, Miranda Lee. Problems of Creating an Index of Piezoelectric Compounds on System 1022. In: NATIONAL ON-LINE MEETING, 1984, New York. **Proceedings... Medford: Learned Information**, 1984, p. 279-285.
- 12 CANGANELLI, Patrick. Sonar Brookends: Index Generation. **Electronic Library**, v. 12, n. 4, p. 253-254, aug 1994.
- 13 CARNEIRO, Marília Vidigal. Diretrizes para uma Política de Indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set 1985.
- 14 CASSI, L., MUSSIO, P., PADULA, M., PROTTI, M., TONOLLI, G. Relational Tools to Manage Pictorial Information. **Interlending and Document Supply**, v. 18, n. 5, p. 375-398, 1992.
- 15 CHANG, R. The Development of Indexing Technology. **Library Software Review**. v. 12, n. 3, p. 30-35, fall 1993.
- 16 CRAVEN, Timothy C. An Experimental Electronic Worksheet for Articulated Index Production. In: ASIS ANNUAL MEETING, 48, 1985, Las Vegas. **Proceedings... ASIS '85**, New York, Knowledge Industry Publications, 1985, v. 22, p. 315-318.
- 17 FIDEL, Raya. User-Centered Indexing. **Journal of the American Society for Information Science**, New York, v. 45(8), Sep 94, p. 572-576.
- 18 FISCHER, Dietrich H. Consistency Rules and Triggers for Thesauri. **International Classification**, v. 18, n. 4, p. 212-225, 1991.

- 19 GANZMANN, Jochen. Criteria for the Evaluation of Thesaurus Software. **International Classification**, v. 17, n. 3-4, p. 148-157, 1990.
- 20 GUNCHARUK, E. D., KONDRATEV, I. A. Quality Control System for an Information Service: Document Indexing Quality Evaluation. **Autom. Doc. Math. Linguist**, v. 22, n. 6, p. 27-32, 1988.
- 21 HALLER, Johan. Indexação Automática de Textos. **Revista de Biblioteconomia de Brasília**, v. 13, n. 1, p. 27-32, jan/jun 85.
- 22 HJERPPE, Roland. The Role of Classification in Hypertext. Issues in Implementing Rogget's Thesaurus as a Hypertext. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmstadt. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 1, p. 206-215.
- 23 HODGE, Gail M. Computer-assisted Database Indexing: the State-of-the-art. **Indexer**, v. 19, n. 1, p.23-27, apr 1994.
- 24 JONES, Leslie P., GASSIE, Edward W., RADHAKRISHNAN, Sridhat. INDEX: the Statistical Basis for an Automatic Conceptual Phrase-indexing System. **Journal of the American Society for Information Science**, New York, v. 41, n. 2, p. 87-97, mar 1990.
- 25 KENT, Robert E. Implications and Rules in Thesauri. In: INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (ISKO) CONFERENCE, 3, 1994, Copenhagen. **Proceedings... Knowledge Organization and Quality Management**. Frankfurt: INDEKS Verlag, 1994, p. 154-160.
- 26 KWON, Yong-Moo, PARK, Yong-Jin. A New Indexing Technique for Nested Queries on Composite Objects. **Ieice Trans. Inf. Syst.**, v. e75-d, n. 6, p. 861-873, nov 1992.
- 27 KROOKS, David A., LANCASTER, F. W. The Evolution of Guidelines for Thesaurus Construction. **Libri**, v. 43, n. 4, p. 326-342, oct/dec 1993.
- 28 LANCASTER. F. W. **Indexação e Resumos: Teoria e Prática**. Brasília, Briquet de Lemos, 1993. 347p.
- 29 LANCASTER. F. W., ELLIKER, Calvin, CONNELL, Tschera Harkness. Subject Analysis. **Annual Review of Information Science and Technology**, v. 24, p. 35-84, 1989.
- 30 LAYNE, Sara Shatford. Some Issues in the Indexing of Images. **Journal of the American Society for Information Science**, New York, v. 45, n. 8, p. 583-588, sep 1994.
- 31 LEMOS, Maria Lúcia Vilar de. Desenvolvimento de um Vocabulário Controlado na Biblioteca do Senado Federal. **Ciência da Informação**, Brasília, v. 15, n. 2, p. 155-158, 1986.
- 32 LEUNG, C. H. C., HIBLER, D., MWARA, N. Picture Retrieval by Content Description. **Journal of Information Science**, London, v. 18, n. 2, p. 111-119, 1992.
- 33 LOCHBAUM, Karen E., STREETER, Lynn A. Comparing and Combining the Effectiveness of Latent Semantic Indexing and the Ordinary Vector Space Model for Information Retrieval. **Information Processing & Management**, v. 25, n. 6, p. 665-676, 1989.
- 34 MAMFRIM, Flávia Pereira Braga. Representação de Conteúdo Via Indexação Automática em Textos Integrais em Língua Portuguesa. **Ciência da Informação**, v. 20, n. 2, p. 191-203, jul/dez 1991.
- 35 McKNIGHT, Mark. Improving Access to Music: a Report of the MLA Music Thesaurus Project Working Group. **Notes**, v. 45, n. 4, p. 714-721, jun 1988.

- 36 MILSTEAD, Jessica L. Methodologies for Subject Analysis in Bibliographic Databases. **Information Processing & Management**, v. 28, n. 3, p. 407-431, 1992.
- 37 MILSTEAD, Jessica L., BERGER, Mary C. The Engineering Information Thesaurus Development Project. **Information Services & Use**, Amsterdam, v. 13(1, p. 71-80), 1993.
- 38 PAO, Miranda Lee. **Concepts of Information Retrieval**. Colorado: Libraries Unlimited, 1989. 285p.
- 39 PARSONS, Julia. Finding your Way Around the Information Maze. **Information Management & Technology**, v.25, n. 3, p. 120-121, may 1992.
- 40 RABITTI, F., SAVINO, P. Automatic Image Indexation to Support Content-based Retrieval. **Information Processing & Management**, v.28, n. 5, p. 547-565, 1992.
- 41 RACKMORE, Marlene. Facet Analysis and Thesauri for Corporate Information Retrieval. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 233-238.
- 42 RADA, Roy. Connecting and Evaluating Thesauri: Issues and Cases. **International Classification**, v. 14, n. 2, p. 63-69, 1987.
- 43 RESS-POTTER, Lorna K. Dynamic Thesauri: the Cognitive Function. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 145-150.
- 44 RITZLER, Claus. Comparative Study of PC-Supported Thesaurus Software. **International Classification**, v.17, n. 3-4, p.138-147, 1990.
- 45 ROWLEY, Jenny. Tracer: Indexing Package. **Information World Review**, n.19, p. 10-11, oct 1987.
- 46 SAVOY, Jacques. Stemming of French Words Based on Grammatical Categories. **Journal of the American Society for Information Science**, v. 44, n. 1, p. 1-9, jan 1993.
- 47 SCHUEGRAF, Ernest J., BOMMEL, Martin F. Van. An Automatic Document Indexing System Based on Cooperating Expert Systems: Design and Development. **Canadian Journal of Information and Library Science**, v. 18, n. 2, p. 32-50, jul 1993.
- 48 SCHWARTZ, Candy, EISENMANN, Laura Malin. Subject Analysis. **Annual Review of Information Science and Technology**. v. 21, p. 37-61 , 1986
- 49 SRINIVASAN, Padmini. On Generalizing the Two-Poisson Model. **Journal of the American Society for Information Science**, New York, v. 41, n. 1, p. 61-66, jan 1990.
- 50 STERN, Alice, RISCHETTE, Norbert. On the Construction of a Super Thesaurus Based on Existing Thesauri. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 134-144.
- 51 SVENONIUS, Elaine. Unanswered Questions in the Design of Controlled Vocabularies. **Journal of the American Society for Information Science**, New York, v. 37, n. 5, p. 331-340, sep 1986.
- 52 TODESCHINI, C. **INIS: Manual for Indexing**. Vienna, International Atomic Energy Agency, jan 1974.

- 53 VIEIRA, Simone Bastos. Análise Comparativa Entre Indexação Automática e Manual da Literatura Brasileira de Ciência da Informação. **Revista de Biblioteconomia de Brasília**, v. 16, n. 1, p. 83-94, jan/jun 88.
- 54 WEINBERG, Bella Hass, CUNNING-HAM, Julie A. The Design of On-line Thesauri In: NATIONAL ON-LINE MEETING, 1988, New York. **Proceedings...** Medford, Learned Information, 1988, p. 411-419.
- 55 WELLISCH, Hans. H. The Art of Indexing and Some Fallacies of its Automation. **Logos**, v. 3, n. 2, p. 69-76, 1992.
- 56 WELLISCH, Hans. H. Book and Periodical Indexing. **Journal of the American Society for Information Science**, New York, v. 45, n. 8, p. 620-627, sep 1994.

6.6 - ARMAZENAMENTO

Esta etapa do projeto define o software de gerenciamento e a forma de armazenamento dos dados que irão compor a base de dados. Principalmente para os sistemas que possuem grande volume de informações ou armazenam textos completos, estas definições devem considerar:

- As consultas que serão realizadas - Estas consultas irão gerar pesquisas aos índices de acessos dos documentos e posterior leitura daqueles de interesse.
- Tempo de acesso aos dados - As operações de leitura e gravação em dispositivos magnéticos são consideradas de elevado tempo de realização se comparadas aos acessos feitos a outros componentes do computador. Por este motivo, elas devem ser minimizadas ao máximo objetivando reduzir o tempo de resposta à consulta do usuário.

Com base nestas necessidades os métodos de armazenamento devem atender, de forma eficiente, às consultas dos usuários, reduzindo ao máximo o tempo de resposta. Na literatura pesquisada foram encontrados dois métodos:

- Desenvolvimento de software próprio - O provedor do sistema deve desenvolver programas que gerenciem o acesso aos dados e atendam às consultas realizadas. Eles devem conter rotinas de compressão e descompressão dos dados, que têm como objetivo armazenar os dados nos dispositivos magnéticos em um espaço físico menor que o inicialmente necessário.
- Utilização de softwares bibliográficos - Deve ser escolhido, dentre os softwares bibliográficos disponíveis no mercado, aquele que melhor atenda às necessidades do sistema considerando suas características, restrições, armazenamento e categoria.

O desenvolvimento do software pelo provedor possibilita-o ser “sob medida” para a aplicação, devendo serem analisadas apenas as alternativas de armazenamento. A alternativa mais comum na literatura é a compressão dos dados, discutindo as características e alternativas de implementação de forma isolada.

Dos 34 artigos analisados para este trabalho, apenas doze discutem aspectos relacionados à compressão dos dados e novas tecnologias, incluindo hipermídia e outros recursos de armazenamento. Os demais discutem as características e utilização de softwares bibliográficos, sendo que deste grupo apenas os artigos de **SIEVERTS ET ALL**²⁹ e **OLIVER**²⁵, com grande ênfase para o primeiro, conseguem relacionar os itens necessários para realizar uma escolha. Os demais discutem características específicas de cada software ou realizam comparações entre dois ou três. Não foram encontrados artigos que discutem desenvolvimento de software por serem estes específicos da aplicação.

6.6.1 - Compressão dos dados

O objetivo da compressão dos dados é utilizar espaço nas áreas do disco magnético menor que o necessário para o armazenamento dos dados, ou seja, o uso de uma das técnicas de compressão objetiva a redução do espaço físico necessário para o armazenamento, gerando grande economia. Esta economia é importante porque o disco magnético possui um limite previamente conhecido e o armazenamento consome grande área, principalmente para os textos completos e imagens. Para os sistemas que utilizam estas modalidades é necessário realizar a compactação dos dados a cada gravação realizada e a descompactação a cada leitura. Esta necessidade é negativamente considerada por **GILLMAN**¹³ que questiona a redução de

aproximadamente 30% na utilização de área em disco, com prejuízo do tempo da recuperação do documento no momento da descompressão. Entretanto, **WITTEN ET ALL**³³ afirma que “em razão da velocidade dos computadores atuais, do tempo de acesso e transferência das unidades de disco serem cada vez mais elevados e da redução da frequência de acesso, o tempo para a descompressão vem se tornando irrelevante”.

Um sistema que utiliza compressão de dados deve ser avaliado pelas operações realizadas, tais como:

- Caso o sistema permita atualização dos dados pelo usuário devem ser considerados os tempos de compressão e descompressão;
- Para os sistemas utilizados apenas para consultas, onde não existe o processo de atualização, o tempo de compressão não é considerado, devendo ser avaliado apenas a descompressão.

6.6.1.1 - Métodos de compressão dos dados

BELTRAMI ET ALL⁴ define dois métodos para a compactação de dados:

- Método Estático - Método no qual o arquivo é analisado integralmente antes de iniciar a compressão. Este método realiza a leitura do arquivo duas vezes: A primeira para realizar as avaliações estatísticas dos caracteres (letras, números e outros caracteres) do texto. Elas são necessárias para definir os códigos a serem utilizadas na compactação para cada um deles. Eles são definidos a partir dos caracteres existentes e identificando, para cada um deles, um código que utiliza uma área menor que a original. A segunda leitura realiza a compactação caracter a caracter, transformando cada caracter lido no código correspondente, definido pela avaliação feita anteriormente.
- Método Dinâmico - Método em que uma tabela de frequência dos caracteres lidos do texto é atualizada cada vez que ele surge, gerando um novo código de acordo com a sua frequência atual. Estes códigos possuem tamanho inversamente proporcional à sua frequência. Neste método o mesmo caracter pode ter diferentes códigos no mesmo texto.

6.6.1.2 - Métodos de descompressão dos dados

Com relação à forma com que os documentos serão restaurados, podemos relacionar dois métodos:

- Com recuperação integral - Onde os textos são restaurados de forma idêntica ao original, ou seja, não existe perda de conteúdo.
- Com recuperação parcial - Quando parte do conteúdo do texto é perdido no momento da compressão. Neste método, quando a descompressão é realizada, este conteúdo não é recuperado integralmente, havendo perda de informação, ou seja, o resultado final é diferente do original. É importante definir quais as informações que podem ser perdidas, pois deverá ser uma perda controlada.

6.6.1.3 - Técnicas de compressão dos dados

Dentre as técnicas de compressão que utilizam os métodos acima, podemos relacionar:

- Codificação de sequência repetida - Substituição de uma sequência de caracteres iguais por apenas um destes acrescido da frequência com que é apresentado.

- Substituição de termos freqüentes - Considerando "termo" como uma palavra ou conjunto de palavras, consiste em substituir os termos mais freqüentes por símbolos menores e os menos freqüentes por símbolos maiores.
- Armazenamento de alfabeto restrito - Considerando como alfabeto restrito apenas as letras minúsculas, consiste em converter as letras maiúsculas em minúsculas e estas por símbolos de tamanho variável. Nesta técnica os 8 bits normalmente utilizados para o armazenamento de cada letra são substituídos por até 5 bits, sendo que as mais freqüentes recebem códigos menores (1 ou 2 bits) e as menos freqüentes códigos maiores (5 bits). Bit é o nome utilizado para representar a menor unidade de armazenamento do computador.
- Compressão de índices - É a técnica utilizada para a compressão dos índices de acessos aos dados. Consiste na substituição de um termo ou palavra completa pela freqüência das suas letras iniciais em relação à anterior acrescidas das suas letras a partir da primeira diferente.
- Armazenamento em binário - Consiste na criação de uma matriz (conjunto de linhas e colunas) onde as linhas representam os registros existentes no arquivo e as colunas os diferentes valores armazenados em cada campo. O conteúdo de cada posição da matriz (linha x coluna) é representado por um bit, cujo conteúdo representa a existência (bit 1) ou não (bit 0) da posição correspondente de conteúdo do campo. Nesta técnica é possível realizar consultas a partir desta matriz, conforme demonstrado por **SPIEGLER**³¹.

Outro aspecto importante no processo de compressão é a geração e o armazenamento dos índices de acesso aos dados. Segundo **BELTRAMI ET ALL**⁴, do espaço total ocupado por uma base de dados, 40% se refere ao armazenamento dos índices. Caso seja uma base numérica, este percentual cai para 20%.

6.6.2 - Softwares Bibliográficos

Segundo **SIEVERTS ET ALL**²⁹ a escolha de um software bibliográfico comercial, (programa de computador utilizado para armazenar e recuperar informações bibliográficas distribuído comercialmente) está diretamente relacionada às necessidades do produtor da base de dados.

Para realizar esta tarefa devem ser consideradas as características dos dados a serem armazenados e as saídas previstas para o sistema. A escolha de um software bibliográfico deve seguir uma metodologia, como **OLIVER**²⁵ propôs em seu artigo, considerando os seguintes aspectos:

- Identificar as necessidades do sistema e dos processos que serão automatizados;
- Relacionar os aplicativos que atendem a estas necessidades;
- Escolher o formato de armazenamento dos dados. Este formato deverá ser compatível com vários outros softwares ou aplicações
- Relacionar os softwares que permitem desenvolver os aplicativos no formato escolhido
- Realizar contato e avaliar propostas dos fornecedores dos softwares relacionados
- Avaliar os softwares em relação aos seus requisitos de equipamentos (hardware), custo, manutenção, treinamento, uso, etc.

Pelas diferentes categorias e recursos existentes nos softwares, a etapa de identificação deve ser realizada com grande atenção. A escolha incorreta pode gerar dificuldades de uso por parte do usuário ou do próprio produtor.

6.6.2.1 - Classificação dos softwares bibliográficos

Analisando os softwares disponíveis no mercado, **SIEVERTS ET ALL**²⁹ relacionou e categorizou cada uma deles, como sendo:

- Programas de Recuperação Clássica - Suporte à lógica booleana; registros de tamanho variável; consulta por palavras do campo; facilidade na entrada dos dados; flexibilidade para geração de relatórios; uso direcionado por menus.
- Programas para Usuário Final - Desenvolvido para o usuário da informação. Deve fornecer mecanismos para facilitar a entrada e saída das referências pessoais; listar as referências de acordo com o tipo do documento; possuir pequena capacidade de pesquisa, dando ênfase à pesquisa seqüencial; facilidade de importação e exportação de dados.
- Programas de Indexação - Programas que geram arquivos de índices independentes do original. A geração deve ser feita após cada atualização a partir dos dados digitados em processadores de texto. Contém operadores de aproximação para consultas.
- Programas de Recuperação de Textos Completos - Possuem características semelhantes aos programas de indexação, porém os arquivos texto e índice não são independentes. Manipula arquivos estruturados e não estruturados e possui, em alguns casos, operadores de aproximação.
- Programas de Gerenciamento de Informações Pessoais - A principal característica deste tipo de software é não possuir a rigidez de uma base de dados. Permite o armazenamento de informações de forma genérica, tais como: agenda pessoal, telefone, etc. Alguns deles possuem recursos de armazenamento e recuperação de dados bibliográficos, índices, pequenas edições, etc.

Legenda quadro F

- A - Programas de Recuperação Clássico
- B - Programas para Usuário Final
- C - Programas de Indexação
- D - Programas de Recuperação de Textos Completos
- E - Gerenciamento de Informações Pessoais
- F - Hipertexto
- G - Pontuação e Ranking

Quadro F - Classificação dos softwares bibliográficos analisados.

Softwares	A	B	C	D	E	F	G
3 by 5		2			1		
Archivist		1					
Ask-It			1				
AskSAM		2			1	1	
Bib/Search	1	1					
CardBox +	1						
CDS/ISIS	1						
Free Base	1					2	
Headfast	1						
IdeaList	1	2				2	
In Magic	1						
KAware	2			1		1	
Note Book-II		1			2		
Nut Shell	1	2					
Papyrus		1					
Personal Librarian						2	1
Polydoc	1						
Pro-Cite		1					
Reference Manager		1					
Ref-Filer		1					
Texplore			1				
Wordcruncher			2	1			

- Programas de Hipertexto - Têm como principal característica o acesso não estruturado às informações. Este acesso é feito através de "Links" (ligações) entre as palavras, ou item de pesquisa, à sua descrição ou explicação, acionada através do texto ou por click de mouse. Com estes softwares é possível desenvolver sistemas com diversos tipos de conteúdo, tais como, texto, som, imagens, etc.
- Programas de Seleção por Pontuação e Ranking - São aqueles que não realizam a combinação lógica dos termos usados para consulta. O usuário fornece um mínimo de termos relevantes para o assunto desejado e o software identificará aqueles que estão relacionados com o documento.

Para melhor avaliação dos softwares discutidos neste trabalho, identificados nos artigos pesquisados, podemos classificá-los segundo as categorias relacionadas acima, conforme mostrado no quadro 6. Este quadro identifica com o código 1 para os software que melhor categoriza o programa e código 2 para os de menor categorização.

6.6.2.2 - Escolha do software bibliográfico

Para avaliar um software bibliográfico, **SIEVERTS ET ALL**²⁹ relacionou a necessidade de se observar os seguintes quesitos:

- Necessidades técnicas - Consiste em verificar se os recursos disponíveis nos equipamentos são compatíveis com os do software. Caso não sejam, devem ser considerados na avaliação os custos de aquisição e provável treinamento. Os principais itens a considerar são: (detalhamento no quadro 7)
 - Sistema Operacional para os quais foram desenvolvidos.
 - Configuração necessária do equipamento - Capacidade de memória, espaço em disco, tipo de processador, etc.
- Versões especiais e segurança - É a identificação no software dos recursos de segurança e da variedade de versões existentes para diferentes ambientes operacionais. Os principais itens a considerar são: (detalhamento no quadro 8)
 - Disponibilidade em versões especiais: Demo, Rede, Monousuário, Multiusuário;
 - Senhas de acessos;
- Utilização do programa - Consiste em relacionar os recursos operacionais disponíveis, que poderão influir na utilização do sistema. Os principais itens a considerar são: (detalhamento no quadro 9)
 - Interface amigável;
 - Qualidade dos manuais e telas de ajuda;
 - Suporte técnico pelo fornecedor;
 - Precauções contra erros de operação;
- Limitações - Consiste em identificar os limites de processamento e armazenamento dos dados do software. Os principais itens a considerar são: (detalhamento no quadro 10)
 - Caracteres por campo;
 - Caracteres por registro;
 - Campos por registro;
 - Caracteres por arquivo;
 - Registros por arquivo;
 - Arquivos por programa.
- Entrada e manutenção dos dados - Identificar os recursos de entrada dos dados disponíveis no software. Os principais itens a considerar são: (detalhamento no quadro 11)
 - Possibilidade de digitação interativa (on-line);
 - Possibilidade de atualização dos dados existentes;
 - Possibilidade de criação e alteração na estrutura dos registros;
 - Validação do conteúdo dos campos no momento da entrada dos dados;
 - Utilização de editor de tela para atualização ou importação dos dados;
 - Entrada de dados por arquivos externos;
 - Importação de dados de bases de dados externos.
- Indexação dos campos - Consiste em avaliar as formas de indexação existentes no software. Os principais itens a considerar são: (detalhamento no quadro 12)

- Flexibilização na definição e atualização dos índices (palavras, termos compostos, subcampos) e dos campos a serem indexados;
- Eliminação de palavras não significativas;
- Definição do número de caracteres por termo indexado (tamanho da chave de acesso);
- Rapidez na indexação;
- Tamanho da área em disco ocupada pelo arquivo de índices em relação ao arquivo de dados;
- Indexação automática e/ou postergada;
- Utilização de estrutura de tesouro e/ou termos controlados.
- Recuperação das informações - Avaliar os recursos de consulta aos dados disponíveis para o usuário. Os principais itens a considerar são: (detalhamento no quadro 13)
 - Pesquisa seqüencial e/ou por índice de acesso;
 - Pesquisa seqüencial de campos não indexados;
 - Impressão do dicionário de índices;
 - Consultas no formato booleano (operadores e, ou e não);
 - Marcação interativa dos campos a serem consultados;
 - Pesquisas com proximidade e truncagens;
 - Pesquisas utilizando resultados de outras pesquisas, realizadas anteriormente;
 - Armazenamento em disco de consultas anteriores.
- Saída dos dados - Identificar as possibilidades de recuperação dos dados selecionados. Os principais itens a considerar são: (detalhamento no quadro 14)
 - Possibilidade de saídas em telas, impressoras ou arquivos;
 - Definição pelo usuário do formato de saída dos dados;
 - Ordenação por vários campos;
 - Saída de valores constantes;
 - Indicação dos termos utilizados na pesquisa;
 - Exportação dos dados;
 - Formato de saída de acordo com o tipo do documento.

Em avaliação feita na literatura foram identificados diversos artigos que relacionam estas características. Alguns artigos foram escritos sobre softwares específicos, (em grande quantidade para os softwares Pro-Cite e CDS/ISIS), outros analisando dois ou mais softwares, muitas vezes comparando-os. Grande ênfase é dada ao desenvolvimento de aplicações utilizando um destes softwares.

Com base nestas avaliações, podem ser montados diversos quadros, mostrados abaixo, que evidenciam a existência ou não de alguns itens relacionados acima nos softwares analisados. (Para os itens disponíveis no software, foi utilizado a letra S (Sim), para os não existentes a letra N (Não). Os itens não encontrados nos artigos foram deixados em branco).

6.6.2.3 - Avaliação dos softwares bibliográficos, segundo as Necessidades Técnicas

Os Sistemas Operacionais identificam o ambiente, a nível de equipamento, necessário para a instalação do software. Os Sistemas Operacionais existentes são:

- DOS - Sistema utilizado em ambiente de microcomputadores PC. Possui pequena capacidade de processamento, embora, quando instalado em ambiente de rede local esta capacidade sofra grande aumento. É o mais comum de ser encontrado e mais fácil de ser utilizado, com grande disponibilidade de aplicativos compatíveis no mercado.
- APPLE - Sistema utilizado em ambiente de microcomputador. Possui pequena capacidade de processamento e pouco uso . É difícil e restrita a localização de aplicativos compatíveis.
- OS/2 - Sistema utilizado em ambiente de microcomputador. Possui pequena capacidade de processamento. Este sistema, de propriedade da empresa IBM, vem sendo bastante difundido atualmente, tendo o seu uso aumentado consideravelmente nos últimos anos, porém com pouca disponibilidade de aplicativos.
- UNIX - Sistema utilizado em ambiente de médio porte. Possui grande capacidade de processamento, e utilização compartilhada simultaneamente por várias pessoas. É o sistema operacional que gerencia a maior parte das atividades e operações feitas na Internet.

Quadro G - Avaliação dos softwares, segundo as necessidades técnicas.

Software	Sistema Operacional	Memória Exigida
1 - AskSAM	DOS	512 Kb
2 - Bib/Search	DOS e UNIX	512 Kb
3 - Bibliography Maker	DOS	256 Kb
4 - Bibliography Writer	APPLE	48 Kb
5 - Bookends Extended	APPLE	128 Kb
6 - Bookends Mac	APPLE	1 Mb
7 - Bookends PC	DOS	256 Kb
8 - Bookends ProDOS	APPLE	64 Kb
9 - Cardbox/Plus	DOS e OS/2	256 Kb
10 - CDS/ISIS	DOS	512 Kb
11 - Document Management System for Citations	DOS	640 Kb
12 - Filemaker Pro	APPLE	4 Mb
13 - Free Base Prof	DOS e UNIX	600 Kb
14 - Headfast	DOS	512 Kb
15 - IdeaList	DOS	450 Kb
16 - In Magic	DOS	384 Kb
17 - Interactive Bibliography	DOS	256 Kb
18 - LITREF		
19 - Note Book II	DOS	512 Kb
20 - Nut Shell-Plus	DOS	384 Kb
21 - Poly Doc	DOS e UNIX	512 Kb
22 - Pro-Cite	DOS	640 Kb
23 - Ref-11	DOS	265 Kb

24 - Reference Manager	DOS APPLE	384 Kb 1 Mb
25 - The Librarian's Helper	DOS APPLE	256 Kb 64 Kb

Memória Exigida - Representa o espaço disponível na memória do computador necessário para a execução do software. Caso o equipamento não possua este espaço, o programa não será executado, caso o espaço disponível seja próximo ao mínimo exigido a performance de execução será reduzida, ficando comprometido o tempo de resposta às operações.

6.6.2.4 - Avaliação dos softwares bibliográficos, segundo as versões especiais e segurança

Quadro H - Avaliação dos softwares, segundo as versões especiais e segurança.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
- Demonstração		S							S			S	S	S	S	S						S				
- Multiusuário		S							S	S		S	S	S		S							S			
- Senha de Acesso	S	S							S	S		S	S	S	S	S						N	S			

Versões especiais - Identifica as versões especiais existentes no software. As versões mais comuns são:

- Versão de demonstração - Representa a disponibilidade no software de parte dos recursos. Esta versão tem como objetivo permitir ao usuário utilizá-lo de forma experimental para avaliação.
- Versão multiusuário - Representa a possibilidade de compartilhamento simultâneo dos dados por mais de um usuário.

Senha de acesso - Item de segurança que exige que apenas as pessoa autorizadas possam acessar os dados.

6.6.2.5 - Avaliação dos softwares bibliográficos, segundo os recursos disponíveis

Representa os recursos operacionais disponíveis no sistema. Os recursos mais comuns são:

Quadro I - Avaliação dos softwares, segundo os recursos disponíveis

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Tecla de Função	S									S		S	S						S		S		S		
Menu	S		S		S					S	S	S	S	S	S			S	S	S		S		S	S
Hipertexto	S																								
Comando Dirigido		S																							
Ajuda Interativa	S	S								S	S	S	S	S							S	S	S		
Recursos de Programação	S									S	S				S	S									

- Teclas de função - Representam um recurso de programação que possibilita a execução de um conjunto de tarefas acionando uma tecla previamente programada. Este recurso impede que o usuário tenha que realizar as tarefas uma a uma.
- Menu - Representa a execução das operações escolhidas pelo usuário, através das opções disponíveis no software.
- Hipertexto - Utilizado na consulta a bases textuais, consiste na “navegação” pelo texto de forma desestruturada, escolhendo os itens de interesse disponíveis para pesquisa.
- Comando Dirigido - Recurso que permite ao usuário ser orientado pelo software na realização da consulta. Este recurso relaciona os campos que podem ser pesquisados e os possíveis valores de pesquisa.
- Ajuda Interativa - Representa a existência de telas explicativas sobre o funcionamento do software no item em que o usuário estiver posicionado.
- Recursos de Programação - É a capacidade que o software tem de permitir ao usuário desenvolver suas próprias aplicações, independente dos recursos disponíveis.

6.6.2.6 - Avaliação dos softwares bibliográficos, segundo as limitações e restrições existentes

Identifica os limites do software, sendo conseqüentemente os limites operacionais do sistema. Os limites mais comuns são:

Quadro J - Avaliação dos softwares, segundo as limitações existentes.

Software	Caracteres por campo	Caracteres por registro	Campos por registro	Registros por arquivo	Formato Registro
1	1.600	infinito	infinito	4 Mb	Variável
2	15.000	15.000	52	infinito	Variável
7	400		1.000		
8		760			
9	8.000	8.000	53	500.000	Variável
10	8.000	8.000	200	16 milhões	Variável
13	65.000	65.000	250	65.000	Variável
14	32.000	32.000	100	32.000	Variável
15	65.000	65.000	infinito	32.750	Variável
16	infinito	infinito	75	infinito	Variável
19	50.000	50.000	50		Variável
20	infinito	infinito	infinito	infinito	Variável
21	10.000	10.000	20	32.000	Variável
22	16.000	32.000	45	2 bilhões	Variável
24	infinito	infinito		2 milhões	Variável

(O termo *infinito* significa que não existe restrições de tamanho, ou seja, podem ser armazenados qualquer quantidade. O termo *caracter* representa uma letra, número ou qualquer outra identificação armazenado na base de dados.)

- Caracteres por campo - Representa a maior quantidade possível de caracteres para um campo.

- Caracteres por registro - Representa a maior quantidade possível de caracteres para um registro. Este tamanho representa o somatório do tamanho de todos os campos armazenados.
- Campos por registro - Representa a quantidade máxima de campos que podem ser definidos para um registro.
- Registros por arquivo - Representa o limite máximo de registros que pode ser armazenado em cada arquivo da base de dados.
- Formato dos registros - Representa a forma de armazenamento dos dados na base de dados. Quando o formato é variável, significa que está havendo economia de espaço, pois os caracteres branco existentes após cada campo não estão sendo armazenados.

6.6.2.7 - Avaliação dos softwares bibliográficos, segundo os recursos de entrada de dados

Identifica os recursos operacionais existentes para a entrada e manutenção dos dados na base de dados. Os recursos mais comuns são:

Quadro K - Avaliação dos softwares, segundo os recursos de entrada de dados.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
Recursos de importação	S	S	N	N	S	S	S	S	S	S	S		S	S	S	S	S	S	S	S	S	S	S	S	S	
Validação dos dados		N							N	N			N	S	N	N				N	N	S				
Interface BD externa	S					S			S	N	S							S	S	S			S	S	S	
Formatos Pré-definidos		S	S	S		S			S				S					S				S	S		S	
Formatos Livre	S	S							S	S	S	S						S		S	S		S			

- Recursos de Importação - Consiste na possibilidade que o software possui de transferir os dados de outros softwares.
- Validação dos dados - Permite avaliar o conteúdo do campo no momento da sua inclusão na base de dados, impedindo que sejam armazenados valores incorretos ou desconhecidos.
- Interface com base de dados (BD) externa - Possibilita a utilização apenas dos recursos de entrada e saída dos dados. O gerenciamento e armazenamento é feito por outro software.
- Formatos Pré-definidos - Consiste na construção automática da planilha de entrada de dados. Esta planilha contém todos os campos que compõem o registro.
- Formatos Livres - Recurso que permite ao usuário desenvolver a sua própria planilha de entrada de dados, possibilitando definir apenas os campos do registro que o interessam.

6.6.2.8 - Avaliação dos softwares bibliográficos, segundo os recursos de indexação dos campos

Identifica as formas de indexação disponíveis nos softwares. Os recursos mais comuns são:

Quadro L - Avaliação dos softwares, segundo os recursos de indexação.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
Automática									S							S									S	
Batch		S							N	S			S	S	N	N					N	S				
Palavras não significativas		S							N	S			S	S	S	S							N	S	S	

- Atualização Automática - Consiste na construção e atualização dos índices de acessos aos dados no momento do seu armazenamento na base de dados.
- Atualização Batch - Este recurso armazena os dados na base sem a atualização dos índices de acesso. Esta atualização é feita posteriormente, quando o usuário desejar.
- Palavras não significativas - Consiste na montagem de uma lista de palavras que não serão índices de acessos, ou seja, não serão permitidas consultas por elas.

6.6.2.9 - Avaliação dos softwares bibliográficos, segundo os recursos para recuperação das informações

Identifica os recursos de consulta e recuperação dos dados armazenados na base de dados. Os recursos mais comuns são:

Quadro M - Avaliação dos softwares, segundo os recursos para recuperação das informações.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Seqüencial	S	S							S	S			S	N	N	S		S	S		N	S		S	
Indexada	S	S							S	S			S	S	S	S		S	S	S	S	S		S	
Truncagens	S	S							S	S			S	S	S	S		S	S	S	S	S		S	
Proximidade	S	N							N	S			S	N	S	N		N	N	N	N	N		N	
Forma Booleana	S	S							S	S	S		S	S	S	S		S	S	S	S	S		S	

- Seqüencial - Consiste na consulta aos dados de forma aleatória, ou seja, o usuário está analisando todos os registros existentes na base de dados. Esta forma de consulta é conhecida como “browse”.
- Indexada - Consiste na consulta aos registros que atendem a uma consulta específica. Neste caso o software irá selecionar, através dos índices de acesso, os registros solicitados pelo usuário.
- Recursos de Truncagens - Representam a consulta, através dos índices de acesso, dos documentos que contém o sufixo ou prefixo do termo pesquisado. Considerando o caracter \$ como sendo de truncagem, podemos citar, como exemplo, a recuperação de todos os termos que iniciam com “cien\$” . Neste caso seriam selecionados todos os documentos que contivessem os termos “ciências, cientista, científico, etc.”
- Recursos de Proximidade - Representam a utilização de um caracter associado a dois ou mais termos. A seleção será feita nos registros que contenham uma ou mais palavras entre os termos utilizados.
- Formato booleano - Forma de consulta mais utilizada. Consiste na seleção de dois ou mais termos associados aos operadores booleanos “e”, “ou” e “não”.

6.6.2.10 - Avaliação dos softwares bibliográficos, segundo os recursos de saída dos dados

Identifica os recursos disponíveis para a saída dos dados da base de dados. Os dados mais comuns são:

Quadro N - Avaliação dos softwares, segundo os recursos de saída dos dados.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Formato Pré - definido					S													S				S			
Formato Livre	S	S	S						S	S		S	S	S	S	S			S	S	S	S	S	S	
Recursos de Ordenação	S	S	S		N		N		S	S		S	S	S	S	S			S	S	S	S		S	
Exportação de dados	S	S	N	N	S	S	S	S	S	N	S		S	S	S	S	S		S	S	N	S	S	S	

- Formato Pré-definido - Consiste na construção automática da planilha de saída de dados pelo software. Esta planilha contém todos os campos que compõem o registro.
- Formato Livre - Recurso que permite ao usuário desenvolver a sua própria planilha de saída de dados, possibilitando definir apenas os campos do registro de interesse.
- Recursos de Ordenação - Permite ao usuário visualizar os registros selecionados ordenados por um ou mais campos existentes no registro. Este recurso está associado as saídas realizadas em papel.
- Exportação de dados - Consiste na possibilidade que o software possui de transferir os dados para outros softwares.

6.6.2.11 - Avaliação dos softwares bibliográficos, segundo outros recursos

Além dos recursos relacionados, outros podem ser avaliados, conforme descrito no quadro 15.

Quadro O - Avaliação dos softwares, segundo recursos adicionais existentes.

Software	Recurso existente
AskSAM	Armazenamento dos dados, utilizando a técnica de compressão com retirada dos espaços em branco.
Bib/Search	Possibilidade de duplicação dos dados no processo de entrada dos dados.
Cardbox/Plus	Projetado para ser uma "caixa de cartão" computadorizada, ou seja, contém um formato padronizado de entrada de dados. Permite realização de consultas por intervalos de valores numéricos.
CDS/ISIS	Sistema com menu em diversos idiomas além de permitir ao usuário personalizar os seus próprios menus.
Filemaker Pro	Contém recursos gráficos.
Headfast	Consultas realizadas pela linguagem estruturada de consulta à base de dados QBE (QUERY BY EXAMPLE).

IdeaList	Interface com Sistema Operacional, permitindo sem sair do software, executar qualquer tarefa no equipamento.
In Magic	Utilizado para gerenciamento de informação textuais e administração de dados.
Nut Shell-Plus	Possibilidade de realização de operações matemáticas.
Poly Doc	Utilizado para gerenciamento de informações textuais.
Pro-Cite	Interface com Sistema Operacional, semelhante ao software IdeaList.

6.6.3 - Novos recursos no armazenamento de base de dados

Com os avanços tecnológicos, novos recursos de armazenamento dos dados nas bases de dados vêm surgindo a cada dia. Estes recursos tornam cada vez mais fácil o acesso às informações.

Termos como hipertexto, multimídia, discos óticos, processamento de imagens, etc., surgiram na literatura a partir da 2ª metade da década de 80 anunciando a evolução (ou revolução?) que ocorreria na década seguinte nos conceitos de utilização de SRI, sendo atualmente assuntos muitos discutidos.

Como ponto de partida para estas tecnologias surgiu o CD-ROM, também chamado disco ótico, dispositivo periférico que permite o armazenamento de grandes volumes de informações em espaço físico menor que os discos magnéticos, utilizando a tecnologia do laser. Segundo **CORAK**⁸, enquanto um disco rígido de um microcomputador armazena de 40 a 150 Mb (cada Mb contém aproximadamente 1.024.000 caracteres) em 5 polegadas de diâmetro, um disco ótico armazena aproximadamente 1 Gb (aproximadamente 1.024.000.000 caracteres) no mesmo espaço. O uso desta tecnologia popularizou a divulgação e uso de grandes bases de dados, permitindo o surgimento de uma nova tecnologia: a imagem eletrônica.

6.6.3.1 - Armazenamento de Imagens Eletrônicas

Segundo **LYNCH**²⁰, a imagem eletrônica é um conjunto complexo de tecnologias interligadas usadas para criar, armazenar, descrever, manipular, enviar e mostrar informações sob a forma de imagens paradas (fotos) ou em movimento (vídeos). Esta utilização só é possível em razão da grande capacidade de armazenamento e compressão dos discos óticos.

A entrada das imagens no computador é feita através do equipamento scanner, que divide-a em pequenos quadros, formando uma matriz. Em cada posição dessa matriz, é usado um bit, que armazena os valores 0 (zero), representando um espaço em branco, ou 1 (um), representando um espaço não branco. Para imagens coloridas um número maior de combinações é feito para identificar as várias tonalidades de cor, sendo necessário até 36 bits para cada posição da matriz. Esta grande quantidade de combinações gera grandes arquivos (25 Mb para uma foto em papel tamanho ofício), sendo fundamental o processo de compressão.

6.6.3.1.1 - Técnicas de compressão de Imagens Eletrônicas

As técnicas mais comuns utilizadas para compressão de imagens são:

- JPEG (Joint Photographic Experts Group) - Técnica utilizada para compressão de fotografias. Ela elimina as redundâncias dos dados, retirando detalhes como as nuances de cor, que pouco contribuem para a aparência da imagem. Esta técnica reduz o tamanho de um arquivo de 25:1.
- MPEG (Moving Pictures Experts Group) - Técnica utilizada para compressão de vídeos. Ela elimina as redundâncias dos dados, armazenando apenas os quadros em que houveram alterações, em relação ao anterior (deve ser considerado que o movimento é feito por uma seqüência de quadros). Esta técnica reduz o tamanho de um arquivo em mais de 50:1.

6.6.3.2 - Outras técnicas de armazenamento

A partir da necessidade de armazenar e manipular grandes bases de dados com diferentes formas de apresentação, outras técnicas foram criadas, tais como:

- Hipertexto - É a técnica que permite acessos não estruturados aos dados. Este acesso é feito através de ligações ("links" ou nós) entre um elemento ou palavra do texto e a sua definição ou explicação. Atualmente alguns softwares bibliográficos oferecem facilidades de desenvolvimento de sistemas utilizando estes recursos. Este fato é confirmado na análise feita por **MORRALL**²⁴.
- Multimídia - É a técnica que combina diferentes tecnologias de dispositivos com recursos de textos, gráficos, som, animação, etc., sendo o seu uso cada vez mais freqüente e com perspectivas cada vez maiores de utilização, conforme **MICHEL**²².
- Hipermídia - É a técnica que combina recursos de multimídia com hipertexto, ou seja, os elementos interligados no hipertexto podem estar armazenados em diferentes dispositivos.

BORNMAN & SONS⁶ apresentam em seu artigo diversas definições para as técnicas acima.

MICHEL²², em seu artigo, relacionou alguns recursos e dispositivos necessários para a utilização destas tecnologias, tais como:

- Grande capacidade de processamento;
- Sistema Operacional com recursos para gerência destes recursos;
- Grande capacidade de armazenamento;
- Elevado fator de compressão dos dados;
- Velocidade de transferência dos dados;
- Dispositivos periféricos de entrada e saída compatíveis;
- Monitor de alta resolução.

Poucas são as alternativas de armazenamento dos dados. Na realidade, a alternativa é entre o desenvolvimento do seu próprio aplicativo, e arcar com todo o custo operacional do desenvolvimento, manutenção e operação do software ou a aquisição de um software comercial. Dentre estas opções observamos que:

- os softwares bibliográficos disponíveis se baseiam nas categorias referenciais e bibliográficas;

- os softwares que acompanham os equipamentos óticos devem ser utilizados nas categorias que utilizam esta tecnologia;
- a opção para o desenvolvimento é viável apenas para as categorias numéricas, referenciais e bibliográficas, cabendo aí possibilidades de escolha.

Portanto, a tarefa de escolher a alternativa é restrita apenas a escolha do software, pois, de acordo com a categoria utilizada, não existem ainda alternativas disponíveis.

Referências Bibliográficas:

- 1 ALLOWAY, James. Software Reviews: Flat-file Database Program. **Library Software Review**, v. 11, n. 6, p. 54-57, nov/dec 1992.
- 2 BEISER, Karl. Hands on Library Computing: AskSAM 5.0. **Computers in Libraries**, v. 12, n. 2, p. 26-29, feb 1992.
- 3 BELLARBY, Liz. Personal Bibliographic Packages: a Comparison of EndNote and Pro-Cite. **C & L Applications**, v. 6, n. 7, p. 3-4, mar 1993.
- 4 BELTRAMI, Giulio, KULKARNI, D. R. Data Compression Scheme for Large Databases on CD-ROM. In: INTERNATIONAL MEETING FOR OPTICAL PUBLISHING AND STORAGE, 1989, Amesterdam. **Proceedings ... Opticalinfo 89**. Oxford: Learned Information, 1989, p. 21-28.
- 5 BONOMI, Mauro. Multimedia and CD-ROM: an Overview of JPEG, MPEG and the Future. **CD-ROM Professional**, v. 4, n. 6, p. 38-40, Nov 1991.
- 6 BORNMAN, Hester, VON SOLMS, S. H. Hypermedia, Multimedia and Hypertext: Definitions and Overview. **Electronic Library**, v. 11, n. 4-5, p. 259-268, aug/oct 1993.
- 7 CLARK, Katie. Biblio-Link To SilverPlatter. **CD-ROM Professional**, v. 4, n. 5, p. 129-132, sep 1991.
- 8 CORAK, Robert. Removing the Mystique Surrounding Optical Disk Technology. **Drug Information Journal**. v. 26, n. 2, p. 163-165, apr/jun 1992.
- 9 COSGRIFF Jr., John C. The Use of AskSAM: A Full-text Database Program as an Added Value Enhancer to On-line Searching. **Library Software Review**, v. 6, n. 6, p. 374-375, nov/dec 1987.
- 10 DELOZIER, Eric P. Bibliography Management with Pro-CiteTM Version 2.0. **Medical Reference Services Quarterly**, v. 12, n. 2, p. 19-34, summer 1993.
- 11 ELLSBURY, Susan H., ROBERTS, LaDonne. Selected List of Software for Organizing Bibliographic References and Creating Bibliographies. **Collection Management**, v. 16, n. 1, p. 55-72, 1992.
- 12 ELECTRONIC Document Imaging: Magnetic Tape: What is its Role in Document Imaging. **Document image automation update**, v. 10, n. 4, p. 1-5, apr. 1991.
- 13 GILLMAN, Peter. Data Handling and Text Compression. **Journal of Information Science**, London, v. 18, n. 2, p. 105-110, 1992.
- 14 HANSON, Terry. LITREF. **Library Micromation News**, n. 29, p. 6-10, sep 1990.
- 15 HANSON, Terry. Pro-Cite and Biblio-Links. **Library Micromation News**, n. 26, p. 10-12, dec 1989.
- 16 HOPKINSON, Alan. CDS/ISIS Information. **Information Development**, v. 8, n. 1, p. 6-8, jan 1992.
- 17 HOYLE, Norman, McNAMARA, Kathleen. Biblio-link and Pro-Cite: The Searcher's Workstation. **Database**, New York, v. 10, n. 1, p. 73-78, feb 1987.
- 18 KAUFMAN, Daufman. Software Reviews: Bibliographic Database Manager. **Library Software Review**, v. 11, n. 6, p. 47-49, nov/dec 1992

- 19 LUNDEEN, Gerald. Software for Managing Personal Files. **Database.**, New York, v. 12, n. 3, p. 36-48, jun 1989.
- 20 LYNCH, Clifford A. The Technologies of Electronic Imaging. **Journal of the American Society for Information Science**, New York, v. 42, n. 8, p. 578-585, sep 1991.
- 21 McCARTHY, Cavan Michael. Pro-Cite: Um Software para Gerenciamento de Bases de Dados Bibliográficos. **Ciência da Informação**, Brasília, v. 18, n. 2, p. 191-198, jul/dez 1989.
- 22 MICHEL, Jacques. Multimedia for the 1990s: Technological Trends and Main Future Applications. In: INTERNATIONAL INFORMATION RESEARCH CONFERENCE, 2, 1991, Cambridge. **Proceedings ... Multimedia Information**. Cambridge: Bowker-Saur, 1991, p. 1-12.
- 23 MORGAN, Paul. Hypertext and the Literary Document. **Journal of Documentation**, v. 47, n. 4, p. 373-388, dec 1991.
- 24 MORRALL, Anne. Evaluation of Hypertext Software to Create a Public Browsing System in a Museum. **Electronic Library**, v. 9, n. 4-5, p. 217-234, aug/oct 1991.
- 25 OLIVER, K. B. Methodology for Selecting Software for an On-line Database. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 9, 1985, London. **Proceedings...** London: 1985. p. 461-465.
- 26 ROADS, Christopher H. Access and Conservation: the Key Elements in Sound Archiving in the 21st Century. **Phonographic Bulletin**, v. 59, p. 41-44, nov 1991.
- 27 ROWLAND, Melanie, SEELEY, Maureen. Image Databases - the Facts of Life. **Audiovisual Librarian**, v. 17, n. 4, p. 217-220, nov 1991.
- 28 SCHLICKE, Priscilla. **CD-ROM: the challenge of the new**. Information Management Report. sep 1991, p. 18-20.
- 29 SIEVERTS, Eric G., HOFSTEDE, Marten. Software for Information Storage and Retrieval Tested, Evaluated, and Compared: Part 1 - General Introduction. **The Electronic Library**, v. 9, n. 3, p. 145-154, jun 1991.
- 30 SIEVERTS, Eric G., HOFSTEDE, Marten, HAAK, Philip H., NIEUWENHUYSEN, Paul, SCHEEPSMA, Gerlinda A. M., VEEGER, Lucas, VIS, Gert C. Software for Information Storage and Retrieval Tested, Evaluated and Compared. Part II - Classical Retrieval Systems. **The Electronic Library**, v. 9, n. 6, p. 301-318, dec 1991.
- 31 SPIEGLER, Israel, MAAYAN, Rafi. Storage and Retrieval Consideration of Binary Data Bases. **Information Processing & Management**, v. 21, n. 3, p. 233-254, 1985.
- 32 STIGLEMAN, Sue. Pro-Cite 2.0: Steady Evolution. **Database**, v. 15, n. 6, p. 68-72, dec 1992.
- 33 WITTEN, Ian H., BELL, Timothy C., NEVILL, Craig G. Indexing and Compressing Full-text Databases for CD-ROM. **Journal of Information Science**, London, v. 17, n. 5, p. 265-271, 1991.
- 34 WOLFF, J. Gerald. Computing, Cognition and Information Compression. **AI Communications**, v. 6, n. 2, p. 107-137, jun 1993.

7 - CONCLUSÕES: TENDÊNCIAS

Analisando-se a literatura convencional publicada a partir de 1980, observam-se as transformações que ocorreram nas etapas de execução e projeto de um sistema. A partir da discussão de como projetá-lo até as tecnologias óticas e distribuição das informações através da Internet, os assuntos surgem, ocupam grandes espaços e, gradativamente, a quantidade de artigos que discutem “execução” e “projeto” diminui ou desaparece.

Pode-se observar que as discussões sobre projetos de sistemas, praticamente se encerram no início da década de 80; as tecnologias óticas de CD-ROM e processamento de imagens surgiram na segunda metade da década passada e permanecem até hoje, indicando a sua atualidade. As discussões sobre tesouros atravessam todo este período discutindo, inicialmente, a forma de utilização, os aspectos da sua construção e atualmente a análise de assunto, indexação automática, etc. Neste último exemplo, apesar das discussões permanecerem, as abordagens mudaram, acompanhando a evolução tecnológica.

Realizando avaliações tópicas, não é difícil identificar as evoluções ocorridas, tanto na área de projeto quanto na tecnológica, esta última influenciando muito a primeira. Mas, como a literatura irá tratar as etapas de um sistema no futuro? Quais os assuntos que são discutidos atualmente e que terá o seu espaço alterado para mais ou menos nos próximos anos? Qual será a evolução desses assuntos no futuro? É muito difícil identificar qual será o rumo da tecnologia. Quem poderia imaginar alguns anos atrás que a Internet seria o que ela é hoje? Na tentativa de identificar as possíveis tendências, pode-se ressaltar:

- 1) Categorias de SRI - Deverão continuar existindo todas as categorias mostradas neste trabalho, sendo as mais utilizadas:
 - Imagens - Nela serão acrescentados todos os recursos da multimídia, tais como: som, áudio, etc., devendo ser incorporada a tecnologia de reconhecimento de voz, já existindo artigos discutindo e especificando algumas aplicações. Estes sistemas utilizarão recursos de armazenamentos ótico cada vez mais sofisticados, como os videodiscos, capazes de apresentarem sons, imagens em movimento, etc., recursos semelhantes ao videocassete, porém, com tecnologia diferente. Os demais equipamentos (scanner, discos óticos, jukebox, etc.) deverão ter seu uso aumentado, com a redução nos seus custos de aquisição e manutenção.
 - Texto Completo - As publicações eletrônicas, periódicos distribuídos através da Internet, serão cada vez mais utilizadas, permitindo ao usuário montar a sua própria base de dados com os textos de seu interesse. O uso de hipertexto estará presente na maioria dos sistemas.
 - Outras categorias - As demais categorias deverão ter seu uso restrito às aplicações específicas de caráter institucional, tais como:
 - as bases numéricas em Sistemas de Apoio à Decisão;
 - as bases referenciais para as Bibliotecas e Centros de Informações;
 - as bases sonoras para os sistemas musicais.
- 2) Coleta de documentos - Por ser um trabalho intelectual, as políticas de seleção devem manter-se inalteradas. A evolução que poderá ocorrer será a seleção automática pelos assuntos em bases de dados ou publicações eletrônicas disponíveis na Internet. As coletas dos dados devem alterar-se em razão das evoluções tecnológicas, devendo ser utilizados

com mais frequência as transferências eletrônicas. Como já foi observada anteriormente, esta etapa do projeto é pouco discutida por que os documentos normalmente fazem parte do acervo da própria instituição, sendo selecionados e coletados por critérios operacionais ou institucionais.

- 3) Catalogação - Não deverão ocorrer grandes alterações nos formatos de catalogação. Este fato se explica pela pequena proporcionalidade de artigos publicados sobre o assunto, pela possibilidade de transferência eletrônica dos documentos e pela utilização cada vez maior de softwares próprios.
- 4) Indexação - As políticas de indexação serão as existentes nos dias de hoje, transpostas para as tecnologias óticas. As evoluções nos softwares por elas utilizadas irão ocorrer em razão das evoluções tecnológicas, fazendo uso cada vez maior da indexação automática. Esta técnica será a tendência também nos sistemas de texto completo, com algumas aplicações implementando, ou pelo menos tentando utilizar, os recursos de inteligência artificial. O tesauro, utilizando vocabulário controlado, continuará a ser o método mais utilizado.
- 5) Entrada de Dados - Para o processo de entrada de dados, a tendência é manter os métodos atuais, utilizando os equipamentos scanner para entrada de imagens e textos completos. O método de digitação deverá ficar restrito a pequenas bases de dados armazenadas nos softwares comerciais que não permitem outro método. A alteração de comportamento que deverá ocorrer será o aumento nas transferências eletrônicas entre as instituições e autores, reduzindo a circulação de papéis. A entrada de dados via reconhecimento de voz será outro método a ser utilizado com mais frequência, embora em percentuais menores que as transferências eletrônicas, por não ser adequado aos SRI's. Esta tecnologia deverá ser utilizada para atender às consultas dos usuários, criando sistemas totalmente interativo ou conversacionais.
- 6) Armazenamento - Em razão das tecnologias óticas, o gerenciamento dos dados será feito pelo próprio software de armazenamento e recuperação das imagens, disponibilizando-as, para os usuários, em redes locais próprias ou na Internet. Os softwares comerciais terão uso restrito às pequenas bases de dados.

Observando os itens acima, conclui-se que os avanços tecnológicos tornarão mais frequentes:

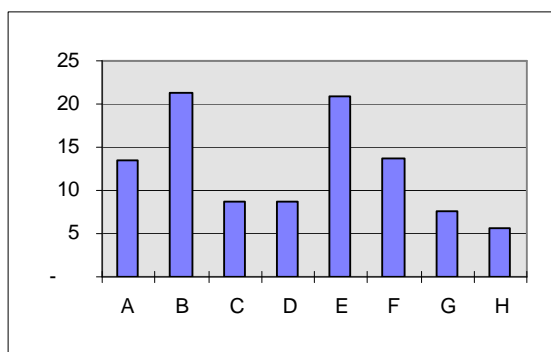
- O uso da Internet no processo de transferência, armazenamento e consultas aos dados;
- A utilização de tecnologias óticas, em razão da redução dos preços de equipamentos, da desmistificação do seu uso e dos conhecimentos adquiridos pelas aplicações já existentes.
- O uso dos recursos da multimídia, que salvo raras exceções, serão aspectos indispensáveis em qualquer sistema.

A sistematização das alternativas existentes para construir um SRI, tendo a literatura como única fonte de conhecimento, cria no projetista a necessidade de avaliar diversos documentos identificando os seus principais assuntos. Esta avaliação deve ser feita inicialmente através do resumo do documento, pois expressa uma visão sucinta do seu conteúdo.

Para a realização deste trabalho, vários documentos foram selecionados na literatura. Estes documentos foram classificados ou reclassificados em cada fase de elaboração da dissertação, nos diversos assuntos que nela são tratados. Após cada classificação foi construído

um gráfico que mostrava o comportamento percentual de cada assunto em relação à quantidade total de documentos, conforme mostrado abaixo:

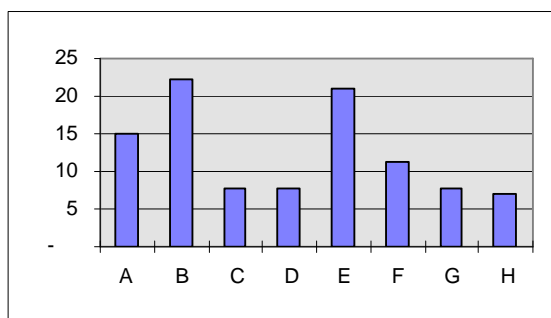
- Na fase de projeto da dissertação - Nesta fase os documentos selecionados foram classificados de acordo com o entendimento do resumo, resultando no gráfico abaixo. Analisando-o observamos o elevado percentual de documentos que tratam sobre indexação e planejamento. Cabe ressaltar que para a fase de Planejamento foram classificados todos os documentos que abordavam algum assunto associado ao tema tendo, por este motivo, um espectro bastante vasto.



LEGENDA

- A - Identif. categorias
- B - Planejamento
- C - Coleta de Dados
- D - Entrada de Dados
- E - Indexação
- F - Catalogação
- G - Armazenamento
- H - Avaliação de SRI

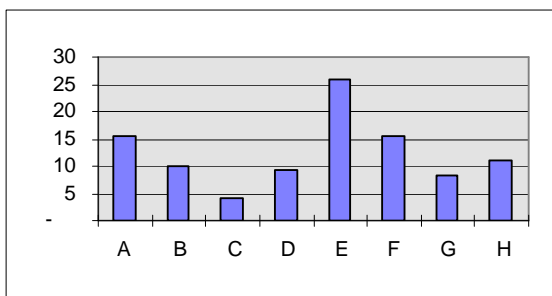
- Na fase de leitura dos documentos - Nesta fase os documentos foram classificados de acordo com o seu conteúdo. Alguns sofreram nova classificação em razão de que este conteúdo não representar exatamente o entendimento que havia sido dado ao resumo. Nesta reclassificação a distribuição percentual dos assuntos ficou distribuída conforme o quadro abaixo.



LEGENDA

- A - Identif. categorias
- B - Planejamento
- C - Coleta de Dados
- D - Entrada de Dados
- E - Indexação
- F - Catalogação
- G - Armazenamento
- H - Avaliação de SRI

- No aproveitamento dos artigos - Nesta etapa foram considerados apenas os documentos que efetivamente foram utilizados no trabalho, resultando no gráfico abaixo. A retirada de um documento ocorreu quando o seu conteúdo estava fora do contexto do trabalho.

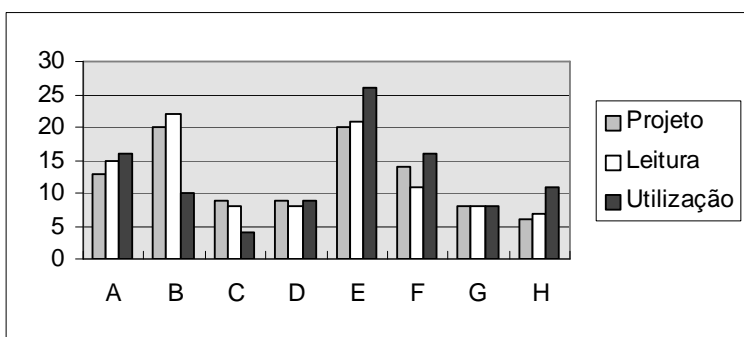


LEGENDA

- A - Identif. categorias
- B - Planejamento
- C - Coleta de Dados
- D - Entrada de Dados
- E - Indexação
- F - Catalogação
- G - Armazenamento
- H - Avaliação de SRI

Comparando os percentuais encontrados, conforme quadro abaixo, destacamos:

- o alto nível de aproveitamento dos itens que abordam aspectos ligados a indexação e catalogação caracterizando a precisão do tema, sendo o primeiro bastante freqüente.
- O baixo percentual de documentos que abordam a Coleta de Dados, sendo que a maioria deles foram coletados na literatura nacional. Muitos dos documentos selecionados para esta etapa tratam de políticas de aquisição e coleta para bibliotecas, estando fora do contexto deste trabalho.
- A grande queda percentual dos documentos que discutem aspectos voltados ao Planejamento. Entretanto, por ser um assunto bastante amplo, vários destes documentos abordam aspectos de SRI que não estão no escopo deste trabalho, tais como: Construção de WWW (World Wide Web), Projeto de criação de SRI em CD-ROM, dando ênfase à elaboração do disco, etc. Esta variedade de documentos e imprecisão do tema leva ao projetista de SRI a consultas desnecessárias.



LEGENDA

- A - Identif. categorias
- B - Planejamento
- C - Coleta de Dados
- D - Entrada de Dados
- E - Indexação
- F - Catalogação
- G - Armazenamento
- H - Avaliação de SRI

O fato da literatura não cumprir o papel de informar, não sistematizando as alternativas existentes nas diversas etapas para a construção de um SRI, no subsistema de entrada, faz com que tal construção seja feita baseada no bom senso do projetista ou nos conhecimentos que este possui, não existindo mecanismos que possam orientá-lo quanto às formas corretas de projeção.

Para a construção de um SRI, não são considerados apenas os subsistemas de entrada, devem ser vistas também as alternativas existentes para os subsistemas de saída, representados pelos produtos e serviços oferecidos pelo sistema. Estes subsistemas de saída, tão importantes quanto os subsistemas de entrada, necessitam ser sistematizados em trabalhos futuros, fornecendo ao projetista as opções necessárias para o projeto.

8 - BIBLIOGRAFIA CONSULTADA

8.1 - SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÕES: Evolução histórica

- CLOUGH, C. R. The Changing Role of Retrieval System and Specialists. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 10, London, 1986. **Proceedings...** London: 1986. p. 175-184.
- FOSTER, Allan. External Databases: An Overview. **Aslib Proceedings**, London, v. 35, n. 9, p. 346-353, sep 1983.
- HUGHES, Graeme C. The Information Age. **Information Development**, v. 7, n. 2, p. 72-74, apr 1991.
- NEUFELD, M. Lynne, CORNOG, Martha. Database History: From Dinosaurs to Compact Discs. **Journal of the American Society for Information Science**, New York, v. 37, n. 4, p. 183-190, jul 1986.
- PEREIRA, Vânia Lúcia da Cunha. **Sistema de Redução da Informação: Uma (Ir)recuperação Metodológica**. Rio de Janeiro: UFRJ/IBICT, 1994. 95 p. Dissertação (Mestrado em Ciência da Informação), Escola de Comunicação, Universidade Federal do Rio de Janeiro, Instituto Brasileiro de Informação em Ciência e Tecnologia, 1994.
- SAKSIDA, Marino. Information - A Vision of the Future: The Human Dimension. **Journal of Information Science**, Holborn, v. 18, n. 5, p. 323-327, 1992.
- TRAUE, J. E. Five Thousand years of the Information Society. **New Zealand Libraries**, v. 46, n. 6, p. 17-18, sep 1990.
- WARRO, E. A. Watt have we been signing?: a look at database licensing agreements. **Library Administration and Management**, v. 8, n. 3, p. 173-177, summer 1994.
- WILLIAMS, Martha E. Highlights of the On-line Database Industry. In: NATIONAL ON-LINE MEETING, 13, 1992. **Proceedings...** Medford: Martha E. Williams, 1992. p. 1-14.

8.2 - SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÕES: Categorias existentes

- AGOSTI, M., GRADENIGO, G., MARCHETTI, P. G. A Hypertext Environment for Interacting with Large Textual Databases. **Information Processing and Management**, v. 28, n. 3, p. 371-387, 1992.
- ARENTS, Hans C., BOGAERTS, Walter F. L. The Three Faces of Hypermedia In: INTERNATIONAL ON-LINE INFORMATION MEETING, 15, 1991, London. **Proceedings...** London: David I Raitt, Oxford, 1991. p. 253-262.
- ARNBAK, Jens C. Many Voices, One Structure: The Challenge of Telematics. **Information Society**, v. 5, n. 2, p. 101-118, 1987.
- BANTZ, David, DAVIES, Nancy, FOELSCHE, Otmar, FENDT, Kurt. Interactive Reading In: INTERNATIONAL INFORMATION RESEARCH CONFERENCE, 2, 1991, United Kingdom. **Proceedings... Multimedia Information**. London: Bowker-Saur, 1991. p. 115-133.
- BOSSEAU, D. **The Era of the CWIS**. Information Management Report, p. 1-5, aug 1992.
- BOSTON, George. The Implications of "New Technology" for Sound Archives. **IASA Journal**. v. 3, p. 68-70, may 94.

- BROOKS, Kristina M., SLADE, Rodney A. Design and Performance Considerations in the Development of Text/image Data Bases. In: NATIONAL ON-LINE MEETING, 1989, New York. **Proceedings...** New York, Learned Information, 1989. p. 57-61.
- BULICK, Stephen. Future Prospects for Network-based Multimedia Information Retrieval. In: INTERNATIONAL ON-LINE INFORMATION, 15, 1991. London. **Proceedings...** Oxford, David I Raitt, 1991. p. 517-520.
- CARLSON, Patricia Ann. The Rhetoric of Hypertext. **Hypermedia**, v. 2, n. 2, p. 109-131, 1990.
- CAWBELL, A. E. Imaging Systems and Picture Collection Management: a Review. **Information Services & Use**, Amsterdam, v. 12, n. 4, p. 301-325, 1992.
- CAWBELL, A. E. Selected Aspects of Image Processing and Management: Review and Future Prospects. **Journal of Information Science**, London, v. 18, n. 3, p. 179-192, 1992.
- CHEN, Ching-Chih. Analog, Digital and Multimedia: Implications for Information Access. In: ON-LINE INFORMATION, 15, 1991. London. **Proceedings...** Oxford: David I Raitt, 1991. p. 238-291.
- CONGER, Lucinda D. The Friendly User. **Database**, New York, v. 7, n. 1, p. 94-95, feb 1984.
- CROW, Linda. Shelf Arrangement Systems for Sound Recordings: Survey of American Academic Music Libraries. **Technical Services Quarterly**, v. 8, n. 4, p. 1-24, 1991.
- CUADRA, Carlos A. Portable Databases: Bridging the Gap Between On-line and Local Databases. **Information Today**, v. 8, n. 1, p. 15-16, jan 1991.
- DUBEY, Yogendra P. Artificial Intelligence Techniques: Prospects for their Application to Information Retrieval System Design. **Journal of Library and Information Science**, India, v. 11, n. 1, p. 1-14, jun 1986.
- EAKIN, Diane R., HASSON, Marsha A. Coping with Data Quality in Scientific Data Banks. In: NATIONAL ON-LINE MEETING, 4, 1983, New York. **Proceedings...** Medford: Martha E. Williams and Thomas H. Hogan, 1983. p. 107-117.
- EAST, Harry, VOGEL, Sandra. Indicators and Revenues of Database Production and Hosting in the United Kingdom. **Journal of Information Science**, v. 18, n. 4, p. 241-249, 1992.
- ESTABLISHMENT of CD-ROM Based Information System: Plans and Prospects. **Library Science with a slant to Documentation**, v. 25, n. 4, p. 244-261, dec 1988.
- FELDMAN, Tony. Turn of the Tide? A New Age in Multimedia Publishing. **British Book News**, p. 658-659, oct 1991.
- FRAPPALO, C. Artificial Intelligence and Text Retrieval: A Current perspective on the State of the Art. In: NATIONAL INFORMATION MEETING, 13, 1991. **Proceedings...** Medford: Martha E. Williams, 1992. p. 113-114.
- GHOSH, Sakti P. Statistical Databases: Design of Experiment Structures. **Information Systems**, v. 18, n. 4, p. 233-247, jun 1993.
- HEMINGWAY, C. Microsoft Modular Windows. **CD-ROM World**. v. 8, n. 3, p. 66-72, apr 1993.
- HIBLER, J. N. D., LEUNG, C. H. C., MANNOCK, K. L., MWARA, N. K. A Database System for the Effective Storage and Retrieval of Images. In: INTERNATIONAL INFORMATION RESEARCH CONFERENCE, 2, 1991, Cambridge. **Proceedings... Multimedia Information**. Bowker-Saur: Feeney and Shirley Day, 1991. p. 83-94.

- HILL, Michael W. Patents on Videodisc: A Future System Almost with us. In: THE FUTURE OF INFORMATION RESOURCES FOR SCIENCE AND TECHNOLOGY AND THE ROLE OF THE LIBRARY. Sweden. **Proceedings...** Sweden: Nancy Fjalbrant, IATUL. p. 151-156.
- HSINCHUN, C. Knowledge-based Document Retrieval Framework and Design. **Journal of Information Science**, v. 18, n.4, p. 293-314, 1992.
- KRANCH, Douglas A. The Development and Impact of a Global Information System. **Information Technology and Libraries**, v. 8, n. 4, p. 384-392, dec 1989.
- LAVIETER, Ir. L. International Databases on Current Research. In: CURRENT RESEARCH INFORMATION IN EUROPE, 1993, Amsterdam. **Proceedings...** Amsterdam: NBOI, M. J. de Goede, 1993. p. 97-98.
- LESCH, A., SZABO, P. Hypermedia Approaches. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings ... Tools for Knowledge Organization and the Human Interface**. Darmstadt: 1990. p. 183-189.
- LEUNG, C. H. C. Architecture of an Image Database System. **Information Services & Use**, v. 10, N. 6, p. 391-397, 1990.
- LUNIN, Lois F. An Overview of Electronic Image Information. **Optical Information Systems**, v. 10, N. 3, p. 114-130, may/jun 1990.
- MALLET, John., MANNING, Colin. Multimedia and Database Design: a Discussion of Database Technology And its Use in Multimedia. **Journal of Document and Text Management**, v. 1, n. 1, p. 75-84, 1993.
- MARTINSEN, David P., LOVE, Richard A., GARSON, Lorrin R. Multiple Use of Primary Full Text Information: a Publisher's Perspective. **On-line Review**, v. 13, n. 2, p. 121-133, apr 1989.
- MARTYN, John. Factual Databases. **Aslib Proceedings**, London, v. 37, n. 5, p. 231-238, may 1985.
- MASTRODDI, F. A. Managing Image Databases: a European Community Perspective. **Information Media & Technology**, v. 23, n. 2, p. 69-74, mar 1990.
- NASH, Rita. Voice Processing: a Present or Future Technology? **LASIE**, v. 17, n. 3, p. 60-75, sep/oct 1986.
- O'CONNOR, Mary Ann. Implementing Optical Storage: How to Select a Document Image Management System. **Document Image Automation**, v. 11, n. 4, p. 230-233, jul/aug 1991.
- O'LEARY, Mick. Computer Databases: a Survey Part I: General and News Databases. **Database**, New York, v. 9, n. 6, p. 15-22, dec 1986.
- OBERHAUSER, Otto C. Interactive Multimedia in Library and Information Services. **Audiovisual Librarian**, v. 17, n. 1, p. 17-25, feb 1991.
- ORMAN, Levent. A Familiar Model of Data Foor a Multilevel Schema Framework. **Information Systems**, v. 7, n. 4, p. 367-376, 1982.
- OSWALD, A. Identification of Electronic Documents: The Diversification of Document Supply and its Consequences for Document Description. In: INTERNATIONAL ESSEN SYMPOSIUM, 14, 1992, Essen. **Proceedings... Libraries and Electronic Publishing: Promises and Challenges for the 90s**. Essen: Ahmed H. Helal and Joachim W. Weiss, 1992. p. 83-95.

- LOUDARD, Denis. Which Technology Should the Sound Archivist Trust? **Phonographic Bulletin**, v. 59, p. 37-40, nov 1991.
- PAIJMANS, H. The Organization of Text Databases. **Microcomputers for Information Management**, v. 5, n. 3, p. 183-195, sep 1988.
- PHILIP, George, YOUNG, Elizabeth S. Man-machine Interaction by Voice: Development in Speech Technology. Part I: the Stage-of-the-art. **Journal of Information Science**, London, v. 13, n. 1, p. 3-14, 1987.
- PRUSSIAN, Michael. Issues for Publishers and Users of CD-ROM on Networks. In: INTERNATIONAL MEETING FOR OPTICAL PUBLISHING AND STORAGE, 1989, Amsterdam. **Proceedings... Opticalinfo 89**. Oxford: Learned Information, 1989, p. 93-97.
- SOERGEL, Dagobert. Structure and Construction of Numeric Databases. **Drexel Library Quarterly**, v. 18, n. 3-4, p. 135-146, summer/fall 1982.
- STOKES, John. Image Management. **Microform Review**, v. 21, n. 2, p. 77-81, spring 1992.
- SUOZZI, Patrícia. By the Numbers: An Introduction to Numeric Databases. **Database**, New York, v. 10, n. 1, p. 15-22, feb 1987.
- THIEL, T. J. Automated Indexing of Document Image Management Systems. **Document Image Automation**, v. 12, n. 2, p. 43-49, summer 1992.
- VACCA, J. R. Telecommunications: Data and Document Image Transmission Systems of the Future. **International Journal of Micrographics and Optical Technology**, v. 12, n. 3, p. 103-109, sep 1994.
- WIGGINS Bob. Document Image Processing - An Overview Part 2: Concepts, Capture and Storage. **Document Image Automation**, v. 12, n. 4, p. 12-20, winter 1992.
- WILLIAMS, Martha E. Highlights of the On-line Database Industry. In: NATIONAL ON-LINE MEETING, 13, 1992. **Proceedings...** Medford: Martha E. Williams, 1992. p. 1-14.
- 8.3 - SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÕES: Etapas de Construção**
- 8.3.1 - Planejamento**
- AITCHISON, T. M. Codes of Practice - Database Producers. **NEWSIDIC**, n. 52, p. 1-2, mar 1982.
- ANDRE, Pamela Q. J. Optical Technology: New Ways of Providing Information for the End User. **Quarterly Bulletin of the International Association of Agricultura Information Specialists**. v. 36, n. 1-2, p. 136-138, 1991.
- APPLEBAUM, David C., CLAYDON, Charles R., HAMPARIAN, Donald L., SMOAK, Keith S. Goals and Structure for Battelle's "People Amplifying" Management. In: NATIONAL ON-LINE MEETING, 1982. **Proceedings...** Medford: Learned Information, 1982. p. 3-14.
- ARTZ, John. The Database Designer is a Person Too! **Journal of Database Management**, v. 5, n. 3, p. 31-34, summer 1994.
- AULD, Dennis B. Databases Production: Multiple Inputs, Multiple Outputs. In: NATIONAL ON-LINE MEETING 90, 11, 1990, New York. **Proceedings...** Medford: Learned Information, 1990. p.25-30.
- BELANI, Rahul. Jane's CD-ROM Project. Step-by-step Through the Development Process. **CD-ROM Professional**, v. 4, n.3, p.69-72, may 1991.

- BIDD, Donald, CHEVIGNY, Louise de, LETOURNEAU, Margo. **FORMAT - Canada's Computerised Information System for Audiovisual Materials. OLAC Newsletter**, v. 5, n. 2, p. 17-26, jun 1985.
- BROOKS, R. M. Principles for Effective Hypermedia Design. **Technical Communication**, v. 40, n. 3, p. 422-428, aug 1993.
- BUCKINGHAM, M. C. S., FRANKLIN, J. Tools for Creating an In-house Database. In: INTERNATIONAL ON-LINE MEETING, 9, 1985, London. **Proceedings...**, p. 243-247.
- CASALE, M. CD-ROM Database Quality. **Online and CD-ROM Review**, v. 17, n. 5, p. 310-312, oct 1993.
- CUNHA, Murilo Bastos. Metodologia para Estudo dos Usuários de Informação Científica. **Revista de Biblioteconomia de Brasília**, v. 10, n. 2, p. 5-19, jul/dez 1982.
- CURRAN, Charles F. Adverse Event Database Construction and Use: 1. Software Design. **Drug Information Journal**, v. 27, n. 2, p. 561-567, apr/jun 1993.
- CUTLER, M., GRAY, M. Construction Manual. Setting up your Own Web Site. **Internet World**, v. 6, n. 2, p. 26-28, feb 1995.
- DAEHN, Ralph M. Methods and Software for Building Bibliographic Data Bases. **Canadian Library Journal**, v. 42, n. 3, p.147-152, jun 1985.
- DANZIGER, Pamela N. Using the PC to Develop In-house Databases. In: NATIONAL ON-LINE MEETING, 7, 1986. New York. **Proceedings...** Medford: Learned Information, 1986. p. 91-96.
- DE, Prabuddha, HASEMAN, William D., SO, Yuk Ho. Four-schema Approach: An Extended Model for Database Architecture. **Information Systems**, v. 6, n. 2, p. 117-124, 1981.
- DURR, W. Theodore. Inputting Data and Conversion of Data: OCR, Scanning, Utility Source for Merging Files. In: IOLS '90. INTEGRATED ON-LINE LIBRARY SYSTEMS MEETING, 5, 1990, New York. **Proceedings...** Medford: Learned Information, 1990, p. 27-33.
- DYKHUIS, Randy. Database Design. **Computers in Libraries**, v. 10, n. 10, p. 50-53, nov 1990.
- EDDISON, Betty. Databases Design: A Checklist for Defining your Needs. **Database**, New York, v. 8, n. 2, p. 78-81, jun 1985.
- EDDISON, Elizabeth B. How to Plan and Build your Own Database. **Database**, New York, v. 11, n. 3, p. 15-16, jun 1988.
- FARMER, Linda. Hypertext: Links, Nodes and Associations. **Canadian Library Journal**, v. 46, n. 4, p. 235-238, aug 1989.
- FINNEGAN, P., FAHY, M. J. Planning for Information Systems Resources? **Journal of Information Technology**, v. 8, n.3, p. 127-138, sep 1993.
- GEYSER, Elsie P., BRAKEL, P.A. Van. Man-machine Interaction as a Factor in the Design of Computerised Information Retrieval Systems. **South African Journal of Library and Information Science**, v. 59, n. 4, p. 256-260, dec 1991.
- GRANICK, Lois. Assuring the Quality of Information Dissemination: Responsibilities of Database Producers. **Information Services & Use**, v. 11, n. 3, p. 117-136, 1991.
- GRICE, R.A., RIDGWAY, L. S. Usability and Hypermedia: Toward a Set of Usability Criteria and Measures. **Technical Communication**, v. 40, n. 3, p. 429-437, aug 1993.

- GRIFFITH, Cary. The Seven Stages of Database Design. **Information Today**, v. 8, n.6, p. 42-44, jun 1991.
- HARRIS, Susan C. The Integration of the Information and Communication Functions, and the Marketing of the Resulting Products. **Information Services & Use**, Amsterdam, v. 5, n. 6, p. 313-322, 1985.
- HENDLEY, T. Digital Mass Storage Options. **Information Management and Technology**, v. 26, n. 1, p. 22-32, jan 1993.
- HUNTER, Eric. CD-ROM and Other Optical Discs: An Overview. **Its News**, v. 18, p. 44-47, feb 1989.
- IGBARIA, M., BANERJEE, S. Computer Capacity Planning Management: Definitions and Methodology. **Journal of Information Technology**, v. 9, n. 3, p. 213-221, sep 1994.
- KITZE, C. Survival Guide for Multimedia Producers - 11 Timely Tips. **CD-ROM Professional**, v. 6, n. 1, p. 58, jan 1993.
- KLING, R. Designing Effective Computing Systems in a Web of Social and Technical Relations. **Human-Computer Interaction**, v. 9, n. 1, p. 86-90, 1994.
- LAWRENCE, Barbara. Database Publishers Challenges for the Future. **Information Services & Use**, Amsterdam, v. 10, n. 5, p. 293-304, 1990.
- LEARN, Larry L., MCGILL, Michael J. The Telecommunication Environment and its Implications for System Design. **Microcomputers for Information Management**, v. 1, n. 2, p. 125-137, jun 1984.
- LEDOUX, Par Marc-Andre. La Creation de Bases de Donnees Bibliographiques. **Argus**, v. 11, n. 5, p. 105-110, sep/oct 1982.
- LEVITIN, A. V., REDMAN, T. C. A Model of the Data (Life) Cycles with Application to Quality. **Information Software Technology**, v. 35, n. 4, p. 217-223, apr 1993.
- LOBO, Maria de Fátima Diniz, BARCELLOS, Silvia de Oliveira. Guias de Fontes de Informação: Metodologia para Geração e Automação. **Ciência da Informação**, Brasília, v. 21, n. 1, p. 75-81, jan/abr 1992.
- LOWRY, Glenn R. A Staffing Profile of U.S. Online Database Producers: A Model and Discussion of Educational Implications. **Online Review**, v. 7, n. 4, p. 329-339, aug 1983.
- LYNCH, Clifford A. Linking Library Automation Systems in the Internet: Functional Requirements, Planning, and Policy Issues. **Library Hi Tech.**, v. 7, n. 4, p. 7-18, 1989.
- MacARTHUR, P. J., CROSSLIN, R. L., WARREN, J. R. A Strategy for Evaluation Alternativa Information System Designes for Business Process Reengineering. **International Journal of Information Management**, v. 14, n. 4, p. 237-251, aug 1994.
- MACAULAY, M. L. A Methodology for Expert Systems Development. **Law Computers and Artificial Intelligence**. v. 2, n. 1, p. 67-76, 1993.
- MACHOVEC, George S. Checklist for Library Database Selection: New Technology Issues. **Online Libraries and Microcomputers**, v. 10, n. 2, p. 1-5, feb 1992.
- MARX, Bernard. Production of Databases in Public Laboratories. In: INTERNATIONAL ON-LINE MEETING, 8, 1984, London. **Proceedings...** London: 1984, p. 413-423.
- MASON, Pamela R. Planning a Multimedia CD-ROM. In: NATIONAL ON-LINE MEETING, 11, 1990, New York. **Proceedings...** Meldford, Learned Information, 1990. p. 265-271.

- McQUEEN, Judy. Creating and Maintaning the Bibliographic Database. **Information Today**, v. 7, n. 3, p. 41-42, mar 1990.
- MEER, Kees Van Der. Online Bibliographic Databases and Grey Literature: a Dutch Approach. In: THE FUTURE OF INFORMATION RESOURCES FOR SCIENCE AND TECHNOLOGY AND THE ROLE OF THE LIBRARY. Sweden: Nancy Fjalbrant, IATUL. p. 157-164.
- NIEUWENHUYSEN, Paul. Criteria for the Evaluation of Text Storage and Retrieval Software. **Electronic Library**, v. 6, n. 3, p. 160-166, jun 1988.
- NILAN, Michael S., PANNEN, Paulina. User-defined Information System Design Specification for Specific Sub-groups: An Example with Resident and Non-resident University Students. In: ANNUAL MEETING OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE, 52, 1989, Washington D.C. **Proceedings... ASIS'89. Managing Information and Technology**. Medford: Learned Information, 1989, v. 26, p. 166-171.
- NORMAN, Margaret. Considerations in Planning an Inhouse Database Retrieval System. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 7, 1982, London. **Proceedings...** Oxford, Learned Information, 1983. p. 213-218.
- NORRIS, Carole L., MARINCOLA, Dian A. Guidelines for Developing an On-line In-house Database Through a Comercial Vendor. In: NATIONAL ON-LINE MEETING, 1982. **Proceedings ...** Medford: Learned Information, 1982. p. 435-444.
- O'CONNOR, Mary. Hints for Designing Effective CD-ROM Databases. **CD-ROM Professional**, v. 4, N. 2, p. 91-93, mar 1991.
- O'CONNOR, Mary Ann. What does it Cost to Develop a CD-ROM? Part 2. **CD-ROM Professional**, v. 4, n. 5, p. 127-128, sep 1991.
- O'LEARY, Mick. Standalone On-line Sources: the Challenge of the Niche Database. **On-line**, v. 15, n. 6, p. 56-60, nov 91.
- PACK, T. CD-ROM Authoring and Retrieval Software: a Look Inside the Leading Packages. **CD-ROM Professional**, v. 7, n. 4, p. 155-156, jul/aug 1994.
- PAO, Miranda Lee. **Concepts of Information Retrieval**. Colorado, USA: Libraries Unlimited, 1989. 285p.
- PAVELIN, Janet. Software Systems for Managing and Storing Documents. **Information Management and Technology**, v. 25, n. 3, p. 130-131, may 1992.
- ROSEBUSH, J. My first CD-ROM. **CD-ROM Professional**, v. 6, n. 6, p. 226-227, nov 1993.
- ROSSITER, B.N., SILLITOE, T. J., HEATHER, M.A. Database Support for very Large Hypertexts. **Electronic Publishing Review**, v. 3, n. 3, p. 141-154, aug 1990.
- SAMBAMURTHY, V., VENKATARAMAN, S., DESANCTIS, G. The Design of Information Technology Planning Systems for Varying Organizational Contexts. **European Journal of Information Systems**, v. 2, n. 1, p. 23-35, jan 1993.
- SCHLEYER, Judith Rebeca. **Estudos de Usuários: Introdução à Problemática e à Metodologia**. In: Estudos Avançados em Biblioteconomia e Ciência da Informação. Brasília: ABDF, 1982, v. 1, p. 49-71.
- SHAW, W. M. Subject and Citation Indexing Part 2: The Optical, Cluster-based Retrieval Performance of Composite Representations. **Journal of the American Society for Informations Science, New York**, v. 42, n. 9, p. 676-684, oct 1991.

- STANFEL, Larry E. Applications of Clustering to Information System Design. **Information Processing & Management**, v. 19, n. 1, p. 37-50, 1983.
- STRICKER, Ulla de. Designing an Electronic Research Tool: Living in a Compromise. **Law Librarian**, v.24, n. 4, p. 176-179, dec 1993.
- SWARD, Robert. Distributed vs. Centralized Information Storage. **International Journal of Micrographics and Optical Technology**, v. 11, n. 3, p. 109-110, sep 1993.
- TENOPIR, Carol. A Day in the Life of a Database Producer. **Database**, New York, v. 15, n. 3, p. 15, 17-20, jun 1992.
- TENOPIR, Carol. Evaluation Criteria for Online. **CD-ROM. Library Journal**, v. 117, n. 4, p. 66,68, 1 mar 1992.
- TENOPIR, Carol. Managing Text: Software Choices for Textual Databases. In: ONLINE INC. CONFERENCE, 1988, New York. **Proceedings... Online '88**. Weston, Online Inc., 1988, p. 143-147.
- TENOPIR, Carol. Quality Control. **Library Journal**, v. 112, n. 3, 15 Feb 87, p. 124-125.
- VANDERGRIFT, Kay E., KEMPER, Marlyn, CHAMPION, Sandra, HANNIGAN, Jane Anne. CD-ROM: An Emerging Technology. Part 2: Planning and Management Strategies. **School Library Journal**, v. 33, n. 11, p. 22-25, aug 1987.
- WICKLIFF, G. A. Special Section on Designing, Testing and Distributing Hypermedia: Introduction. **Technical Communication**, v. 40, n. 3, p. 410-413, aug 1993.
- WIGGINS, R. Building a Campus-Wide Information System. **Campus-Wide Information Systems**. v. 11, n. 1/2, p. 57-66, spring/summer 1994.
- WIGGINS, R. Building a Campus Wide Information System: Part II. **Campus-Wide Information Systems**. v. 11, n. 3, p. 33-43, fall 1994
- WILLIAMS, Martha E. Highlights of the Online CD-ROM Database Industry: Implications of the Internet for Database Procedures. In: ON-LINE MEETING, 15, 1994, New York. **Proceedings...** Medford: Learned Information, 1994, p. 1-6.
- WOLFRAM, Dietmar. Applying Informetric Characteristics of Database to IR System File Desing, Part I: Informetric Models. **Information Processing & Management**, New York, v. 28, n. 1, p. 121-133, 1992.
- WOLFRAM, Dietmar. Applying Informetric Characteristics of Databases to IR System File Design, Part II: Informetric Models. **Information Processing & Management**, New York, v. 28, n. 1, p. 135-151, 1992.
- WRIGHT, Kathy. Bibliographic Database Management in a R & D Organisation: Information Specialists Helping End-users. **Database**, New York, v. 15, n. 3, p. 35-36, 38-40, jun 1992.
- ZEEMAN, J. The Z39.50 Standard: Almost a Reality. **Canadian Library Journal**, v. 49, n. 4, p. 273-276, aug 1992.

8.3.2 - Coleta de Documentos

- ASSOCIAÇÃO BRASILEIRA DE QUÍMICA. **Guia de fontes de informação em química e engenharia química no Brasil**. Rio de Janeiro: ABQ, 1995. 410 p.
- BELANGER, Sandra E., WHITLATCH, Jo Bell, McDERMAND, Robert. Managing Business Periodicals Collections. **Collection Management**, v. 14, n. 1-2, p. 99-131, 1991.

- BRENNAN, M. E. Automation and the Selection Process. In: ALA Annual Conference 1991. **Library Acquisitions: Practice & Theory**, v. 16, n. 3, p. 271-274, 1992.
- BRICK, J. Michael, MORGANSTEIN, David. The Keys to Quality in Survey Products. In: IEEE 1989 NATIONAL AEROSPACE AND ELECTRONICS CONFERENCE, 1989, Dayton, OH. **Proceedings... Naecon 1989**: New York, 1989, p. 1636-1640.
- CASSELL, Kay Ann, FUTAS, Elizabeth. Collection Development Policies. **Collection Building**, v. 11, n. 2, p. 26-29, 1991.
- CLAYTON, Richard, WINTER, Debbie. The Use of Voice Recognition and Voice Response for Survey Data Collection. **Speech Technology**, v. 5, n. 2, p. 52-55, apr/may 1990.
- COLOMBO, R. A. Using Call-backs to Adjust for Non-response Bias. **Survey and Statistical Computing**, p. 269-277, 1992
- DODEBEI, Vera Lúcia Doyle Louzada de Mattos. Metodologia de Coleta de Documentos para Base de Dados Bibliográficos. **Revista de Biblioteconomia de Brasília**, v. 14, n. 2, p. 313-327, jul/dez 1986.
- FARKAS, Doina G., SU, Siew-Phek T. Electronic Order Request Submission at UF Libraries: Three Pilot Programs. **Library Acquisitions: Practice & Theory**, New York, v. 16, n. 3, p. 275-288, 1992.
- IVINS, October. Library Acquisitions: Budget Strategies, Vendor Selection, Vendor Evaluation. **Library Acquisitions: Practice & Theory**, New York, v. 16, n.3, p. 257-263, 1992.
- KATZ, Mark. A Review of the Tools for Secondary Analysis. In: INTERNATIONAL CONFERENCE ON SURVEY AND STATISTICAL COMPUTING. **Proceedings...** 1992, p. 289-296.
- KAYE, G. Roland. The Role of Management Accountants in Information Strategy. **Journal of Information Technology**, v. 3, n. 4, p. 251-264, 1988.
- KELLER, Wouter J. Trends in Survey Data Processing. In: INTERNATIONAL CONFERENCE ON SURVEY AND STATISTICAL COMPUTING. **Proceedings...** 1992, p. 1-10.
- LAMBERT, Maria Betânia Monte Alto. **Manual de Coleta de Documentos para Alimentação da Base de Dados do INIS**. Rio de Janeiro: Centro de Informações Nucleares, jul 1995. (a ser publicado).
- MARINE, Stephen. Highlights of the Alcts Automated Acquisitions/in-process Control Systems in Libraries. **Library Acquisitions: Practice & Theory**, New York, v. 16, n.3, p. 269-270, 1992.
- MEADOR JR, John M., CLINE, Lynn. Displaying and Utilizing Selection Tools in a User-friendly Electronic Environment. **Library Acquisitions: Practice & Theory**, New York, v. 16, n. 3, p. 289-294), 1992.
- MEYER, Paul M. Library Surveys and PC-File Plus. **Computers in Libraries**, v. 9, n. 6, p. 35-36, jun 1989.
- MILLER, Rachel. The Acquisitions Connection: Interfacing Library and Materials Vendor's Systems: Report of the Program Sponsored by the LAMA SASS Acquisitions Systems Committee and the ALCTS RS Acquisitions Committee. **Library Acquisitions: Practice & Theory**, New York, v. 16(3), 1992, p. 299-304.
- O'NEILL, Ann. Evaluation the Success of Acquisitions Departments: a Literatura Review. **Library Acquisitions: Practice & Theory**. New York, v. 16, n. 3, p. 209-219, 1992.

- PAO, Miranda Lee. **Concepts of Information Retrieval**. Colorado: Libraries Unlimited, 1989. 285p.
- PETERS, John. How to Predict the to Retrieve: Selection Documents for Representation in Many Types of Online Databases. **Online**, v. 8, n.5, p. 48-51, sep 1984.
- QUEIROZ, Gilda Gama de. Fonte. Base de Dados Bibliográficos em Fontes de Energia. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 12, 1983, Camboriú. **Anais...**Camboriú: 1983. 13 p.
- RICHARDS, Daniel T. By your Selection Criteria are ye Known. **Library Acquisitions: Practice and Theory**, New York, v. 15, n. 3, p. 279-285, 1991.
- ROYCE, Bert R. The Place of Objctive Measures of Electronic Publishing. In: ASIS ANNUAL MEETING, 45, 1982, Columbus. **Proceedings... Information interaction**, New York: Knowledge Industry Publications, 1982, v. 19, p. 41-42.
- STIRTON, G. BEST Europe: Present and Future. In: CONFERENCE: CURRENT RESEARCH INFORMATION IN EUROPE, 1993, Amsterdam. **Proceedings...** Amsterdam: M. J. de Goede, NBOI, 1993, p. 109-120.
- TRISKA, Ricardo. **Processo de seleção: Rumo a automação**. Rio de Janeiro: Instituto Brasileiro de Informação em Ciência e Tecnologia e Universidade Federal do Rio de Janeiro, 1993. 74 p. Dissertação (Mestrado em Ciência da Informação) Escola de Comunicação, Universidade Federal do Rio de Janeiro, 1993. .
- WHITEHEAD, Derek. Acquisitions in Australia. 1991 Seminars of the National Acquisitions Section. Australian Library and Information Association. **Library Acquisitions: Praticce & Theory**, New York, v. 16, n. 3, p. 243-246, 1992.

8.3.3 - Entrada de Dados

- APPLEN, Roger. Optical Storage Quality Control. **Document Image Automation**, v. 11, n. 2, p. 71-73, mar/apr 1991.
- BRILLIANTINE, Lance R. OCR Improves Productivity. **Word Processing & Information Systems**, v. 9, n. 3, p. 42-44, 64-65, 67, mar 1982.
- BROADHURST, Roger. The Digitisation of Library Material. **Information Management & Technology**, v. 26, n. 3, p. 128-132, may 1993.
- CRAWFORD, Walt. Catching Pictures, Catching Words: Low-Cost Scanning and Optical Character Recognition. **Library Hi Tech**, v. 9, n. 1, p. 91-102, 1991.
- DANIEL, Evelyn H. Quality Control of Documents. **Library Trends**, v. 41, n. 4, p. 644-664, spring 1993.
- EAGLE, Mark. On Opening the Number one Bottleneck in On-line Service-data Entry. In: NATIONAL ON-LINE MEETING, 1987, New York. **Proceedings...** Medford: Learned Information, 1987, p. 125-131.
- FOLLINGSTAD, Marianne Controlling Data Entry in Fileman-Based System Development or Expansion. **Mumps a World Class Technology**. Orlando, v. 20, n. 1, p. 7-13, jun 1990.
- FREDERICK, Janet. Database and Collection Preparation for Automated Circulation. **Technical Services Quarterly**, v. 5, n. 2, p. 35-46, 1987.

- FRUCHTERMAN, James R. Omnifont Text Recognition: Linking Paper, Electronic and Optical Input. **Inform.**, v. 2, n. 5, p. 16-19, may 1988.
- FURUTA, Richard. Important Papers in the History of Document Preparation Systems: Basic Sources. **Electronic Publishing**, v. 5, n. 1, p. 19-44, mar 1992.
- GAGAN, David. Scanning: a Survival Guide (4). Image Scanning: File Formats **C and L Applications**, v. 6, n. 8, p. 8-10, apr 1993.
- GOODALL, Ian. Applying Recognition Technology for Automatic Data Entry. **Information Management & Technology**, v. 26, n. 2, p. 83-87, mar 1993.
- GREENRICH, E. CD-ROM Data Preparation Enhancements. In: NATIONAL ON-LINE MEETING, 14, 1993. New York. **Proceedings ...** Medford: Learned Information, 1993, p. 159-163.
- GROOMS, David W. Quality Assurance of Text and Image Databases at the U.S. Patent and Trademark Office. **Information Services & Use**, Amsterdam, v. 8, n. 2-3-4, p. 161-165, 1988.
- HARVEY, D. A. Catch the Wave of DIP. **BYTE**, v. 16, n. 4, p. 173-174, 176, 178-180, 182, apr 1991.
- HAWKINS, Donald T. Text Retrieval Technologies for Image (and Other) Databases. **Online**, v. 16, n. 1, p. 92-94, jan 1992.
- HENDLEY, Tony., PRITCHARD, John. Recognition Technology. **Information Management & Technology**, v. 26, n. 2, p. 70-78, mar 1993.
- JARMAN, Colin. UK Government Has Second Thoughts on benefits of OCR. **Computer Weekly**, v. 759, p. 18-19, 28 may 1981.
- JONES, Keith. Three Data Entry Methods which Eliminate the Need to Straggle with a VDU. **Computer Weekly**, v. 759, p. 17, 28 may 1981.
- KRUZAN, Donald. Reader Scanners, PCs, Fax - New Decade Will See Combined Technologies that Improve Document Management. **International Journal of Micrographics & Optical Technology**, v. 9, n. 4, p. 203-206, 1991.
- LEE, Jim, O'CONNOR, Mary Ann. So You Want to Produce a CD-ROM? Tips for Successful Data Preparation. **CD-ROM Professional**, v. 4, n. 1, p. 81-84, jan 1991.
- O'LEARY, Mick. Producing a Database: Many Choices for Data Entry. **Database**, New York, v. 13, n. 1, p. 38-40, feb 1990.
- PARLOW, A., WEISKE, Ch., HEB-POHL, N., RODEN, G., ROWMELT J., ZIRZ, C. A Computer System for Inputting Generic Reactions and Sequences to Produce Databases and Printed Publications. In: INTERNATIONAL CHEMICAL INFORMATION CONFERENCE, 1992, Montreux. **Proceedings...** Montreux: 1992. p. 225-233.
- SCHANTZ, Herbert F. Forms Automation and Integrated Imaging (OCR) Systems. **Remittance Doc. Process. Today**, v. 13, n. 6, p. 9-11, mar/apr 1991.
- SCHEIN, Alan. Optical Storage and OCR - Key Components of Automated Information Management Systems. **Optical Information Systems**, v. 9, n. 1, p. 9-15, jan/feb 1989.
- SCHMID, H., NATLACEN, M. Input Control of a Decentralize Information System Including Experiences with a Computer Assisted Translation Program (Spanish to English). **Quarterly**

- Bulletin of the International Association of Agricultural Information Specialists**, v. 39, n. 1-2, p. 98-103, 1994.
- STEIGER, Bettie Alexander, THORNBURG, Bill. Data Preparation and Submission Procedures for RTI's CLASIX DataDrive Series 2000. **Videodisc and Optical Disk**, v. 5, n. 2, p. 125-135, mar/apr 1985.
- 8.3.4 - Catalogação**
- BAKEWELL, K. G. B. Access to Information Via the Index. **Catalogue & Index**, n. 71, p. 1-4, winter 1983.
- BALES, Kathleen. Format Integration: Coordination the Implementation. **Information Technology and Libraries**, v. 9, n. 2, p. 167-173, jun 1990.
- BRANGER, D. A Computerized Catalogue for Unpublished Recordings. **IASA Journal**, v. 1, p. 45-53, may 1993.
- BRYANT, G. Combining Online and Disc. **Online and CD-ROM Review**, v. 17, n. 6, p. 396-398, dec 1993.
- BUCKLAND, Michael K., LYNCH, Clifford, A. The Linked Systems Protocol and the Future of Bibliographic Networks and Systems. **Information Technology and Libraries**, v. 6, n. 2, p. 83-88, jun 1987.
- CANTRALL, D. From MARC to Mosaic: Progressing Toward Data Interchangeability at the Oregon State Archives. **Archives and Museum Informatics**, v. 8, n. 1, p. 4-12, spring 1994.
- DAVIS, Stephen P. Format Integration: Handling Serials and Mixed Media. **Information Technology and Libraries**, v. 9, n. 2, p. 162-167, jun 1990.
- DELFINO, E. The Internet Toolkit: File Compression and Archive Utilities. **Online**, v. 17, n. 6, p. 90-92, nov 1993.
- GAYNOR, E. Cataloguing Eletronic Texts: the University of Virginia Library Experience. **Library Resources & Technical Services**, v. 38, n. 4, p. 403-413, oct 1994.
- GIBBS, George. Applying Format Integration: An Operational Test. **Information Technology and Libraries**, v. 9, n. 2, p. 173-178, jun 1990.
- HOPKINSON, Alan. International Data Exchange: a Question of Formats. **Asian Libraries**, v. 2, n. 1, p. 30-39, mar 1992 .
- INSTITUTO Brasileiro de Informação em Ciência e Tecnologia. **Formato IBICT: Formato de Intercâmbio Bibliográfico e Catalográfico / Instituto Brasileiro de Informação em Ciência e Tecnologia**. Brasília: IBICT, 1987, 400 p.
- JONES, Wayne, QUEINNEC, Young-Hee. Format Integration and Serials Cataloguing. **Serials Librarian**, v. 25, n. 1-2, p. 83-95, 1994.
- KIM, Sook-Hyun. Application of the USMARC Format for Holdings and Locations. **Serials Librarian**, v. 16, n. 3-4, p. 21-31, 1989.
- KHURSHID, Zahiruddin. BiblioFile Versus CDMARC Bibliographic. **CD-ROM Librarian**, v. 6, n. 11, p. 26, 28-31, dec 1991.
- LEATHEM, C. A. An Examination of Choice of Formats for Cataloging Nontextual Serials. **Serials Review**, v. 20, n. 3, p. 59-67, 1994.

- LEAZER, Gregory H. An Examination of Data Elements for Bibliographic Description: Toward a Conceptual Scheme for the USMARC Formats. **Library Resources & Technical Services**, v. 36, n. 2, p. 189-208, apr 1992.
- MACLEOD, Ian A. A Database Management System for Document Retrieval Applications. **Information Systems**, v. 6, n. 2, p. 131-137, 1981.
- McCALLUM, Sally H. Format Integration: Handling the Additions and Subtractions. **Information Technology and Libraries**, v. 9, n. 2, p. 155-161, jun 1990.
- MORGAN, Eric Lease. The MARC Reader. **Computers in Libraries**, v. 11, n. 11, p. 52-52, dec 1991.
- MOTZKIN, Dalia, WILLIAMS Kenneth. A Generalized Database Directory for Nondense Attributes. **Information Processing & Management**, v. 24, n. 2, p. 161-171, 1988.
- MOYA, Felix de, HIPOLA, Pedro. The New IBERMARC Format Alongside UNIMARC. **International Cataloguing & Bibliographic Control**, v. 20, n. 3, p. 41-43, jul/sep 1991.
- MURPHY, Catherine. MARC Accountability... Revisiting a Critical Issue. **Emergency Librarian**, v. 19, n. 3, p. 26-28, jan/feb 1992.
- NICHOLSON, D., STEELE, M. CATRIONA: CATaloguing and Retrieval of Information Over Networks Applications. **Scottish Libraries**, v. 47, p. 11-12, sep/oct 1994.
- PAREKH, Harsha. Bibliographic Record Structure and Communication Formats. **Lucknow Librarian**, v. 23, n. 1, p. 1-24, jan/mar 1991.
- RASMUSSEN, Edie M. Parallel Information Processing. **Annual Review of Information Science and Technology**, v. 27, p. 135-172, 1992.
- ROBREDO, Jaime, Cunha, Murilo B. da, colab. **Documentação de hoje e de amanhã: Uma abordagem informatizada da biblioteconomia e dos sistemas de informação**. 2a ed. rev. e ampl. Brasília: 1986, 400 p.
- SIMMONDS, C. Painless File Extraction: The A(rc) to Z(oo) of Internet Formats. **Online**, v. 17, n. 6, p. 60-62, 64-65, nov 1993.
- SIMMONS, Peter. Using CCF: The Common Communication Format. **Information Technology and Libraries**, v. 5, n. 4, p. 285-294, dec 1986.
- SLOAN, Bernard G. Remote Access: Design Implications for the Online Catalog. **Cataloging and Classification Quarterly**, v. 13, n. 3-4, p. 133-140, 1991.
- WONG, W. Y. P., LEE, D. L. Implementation of Partial Document Ranking using Inverted Files. **Information Processing & Management**, v. 29, n. 5, p. 647-669, sep/oct 93.

8.3.5 - Indexação

- ALBRECHTSEN, Hanne. Subject Analysis and Indexing: From Automated Indexing to Domain Analysis. **Indexer**, v. 18, n. 4, p. 219-224, oct 1993.
- ARENTS, Hans C., BOGAERTS, Walter .F. L. Navigation without Links and Nodes without Contents: Intensional Navigation in a Thirs-order Hypermedia System. **Hypermedia**, v. 5, n. 3, p. 187-204, 1993.
- AUSTIN, Derek. Vocabulary Control and Information Technology. **Aslib Proceedings**, v. 38, n. 1, p. 1-15, jan 1986.

- BAKEWELL, K. G. B. Subject Indexes: Production and Use in the IT age. **Indexer**, v. 13, n. 4, p. 249-251, oct 1983.
- BATTY, David. Microcomputers in Index Language Design and Development. **Microcomputers for Information Management**, v. 1, n. 4, p. 303-312, dec 1984.
- BAUER, Gerd. Promoting Creative Processes by a Thesaurus-like Representation of Knowledge Structure. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 62-69.
- BELL, Hazel K. Vive la Difference! The Survival of the Softest. **Indexer**, v. 18, n. 4, p. 231-236, oct 1993.
- BISWAS, Subal C., SMITH, Fred. Efficiency and Effectiveness of Deep Structure Based Subject Indexing Languages: PRECIS vs. DSIS. **International Forum on Information and Documentation**, v. 16, n. 3, p. 6-21 jul 1991.
- BOOTH, Pat F. Thesauri-their Uses for Indexers. **Indexer**, v. 15, n. 3, p. 141-144, apr 1987.
- BRITO, Marcílio de. Sistemas de Informação em Linguagem Natural: em Busca de uma Indexação Automática. **Ciência da Informação**, v. 21, n. 3, p. 223-232, set/dez 1992.
- BUCHAN, R. L. Intertwining Thesauri and Dictionaries. **Information Services & Use**, Amsterdam, v. 9, n. 3, p. 171-175, 1989.
- BURGIN, Robert, DILLON, Martin. Improving Disambiguation in FASIT. **Journal of the American Society for Information Science**, New York, v. 43, n. 2, p. 101-114, mar 1992.
- BUTCHER, Sharon, PAO, Miranda Lee. Problems of Creating an Index of Piezoelectric Compounds on System 1022. In: NATIONAL ON-LINE MEETING, 1984, New York. **Proceedings...** Medford: Learned Information, 1984, p. 279-285.
- CANGANELLI, Patrick. Sonar Brookends: Index Generation. **Electronic Library**, v. 12, n. 4, p. 253-254, aug 1994.
- CARNEIRO, Marília Vidigal. Diretrizes para uma Política de Indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set 1985.
- CASSI, L., MUSSIO, P., PADULA, M., PROTTI, M., TONOLLI, G. Relational Tools to Manage Pictorial Information. **Interlending and Document Supply**, v. 18, n. 5, p. 375-398, 1992.
- CAWBELL, A. E. An Introduction to Image Processing and Picture Management. **Journal of Documentation Text Management**, v. 1, n. 1, p. 53-63, 1993.
- CHANG, R. The Development of Indexing Technology. **Library Software Review**. v. 12, n. 3, p. 30-35, fall 1993.
- CORCORAN, Maureen. Tracing the Structure of Online Databases Through Their Prini Counterparts: A Review of the Database Search Aids series. **Online**, v. 6, n. 5, p. 77-79, sep 1982.
- CRAVEN, Timothy C. Automatic Structure Modification in the Graphic Display of Thesauri. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 1, p. 146-153.

- CRAVEN, Timothy C. An Experimental Electronic Worksheet for Articulated Index Production. In: ASIS ANNUAL MEETING, 48, 1985, Las Vegas. **Proceedings... ASIS '85**, New York, Knowledge Industry Publications, 1985, v. 22, p. 315-318.
- DAHL, Katherine. The Vertical File Index: a Role to Play a Contribution to Make. Collection Building. v. 11(2), 1991, p.19-22.
- DYKSTRA, Mary. Handling the Stuff Itself: Toward Automated Textual Analysis. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 168-174.
- FIDEL, Raya. User-Centered Indexing. **Journal of the American Society for Information Science**, New York, v. 45(8), Sep 94, p. 572-576.
- FISCHER, Dietrich H. Consistency Rules and Triggers for Thesauri. **International Classification**, v. 18, n. 4, p. 212-225, 1991.
- GANZMANN, Jochen. Criteria for the Evaluation of Thesaurus Software. **International Classification**, v. 17, n. 3-4, p. 148-157, 1990.
- GUNCHARUK, E. D., KONDRATEV, I. A. Quality Control System for an Information Service: Document Indexing Quality Evaluation. **Autom. Doc. Math. Linguist**, v. 22, n. 6, p. 27-32, 1988.
- HALLER, Johan. Indexação Automática de Textos. **Revista de Biblioteconomia de Brasília**, v. 13, n. 1, p. 27-32, jan/jun 85.
- HJERPPE, Roland. The Role of Classification in Hypertext. Issues in Implementing Rogget's Thesaurus as a Hypertext. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 1, p. 206-215.
- HODGE, Gail M. Computer-assisted Database Indexing: the State-of-the-art. **Indexer**, v. 19, n. 1, p.23-27, apr 1994.
- JONES, Leslie P., GASSIE, Edward W., RADHAKRISHNAN, Sridhat. INDEX: the Statistical Basis for an Automatic Conceptual Phrase-indexing System. **Journal of the American Society for Information Science**, New York, v. 41, n. 2, p. 87-97, mar 1990.
- KEEN, Michael E. Aspects of Computer-Based Indexing Languages. In: ANNUAL CONFERENCE ON COMPUTERS IN LIBRARIES, 5, 1991, London. **Proceedings...** London: feb 1991, p. 148-151.
- KENT, Robert E. Implications and Rules in Thesauri. In: INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (ISKO) CONFERENCE, 3, 1994, Copenhagen. **Proceedings... Knowledge Organization and Quality Management**. Frankfurt: INDEKS Verlag, 1994, p. 154-160.
- KROOKS, David A., LANCASTER, F. W. The Evolution of Guidelines for Thesaurus Construction. **Libri**, v. 43, n. 4, p. 326-342, oct/dec 1993.
- KWON, Yong-Moo, PARK, Yong-Jin. A New Indexing Technique for Nested Queries on Composite Objects. **Ieice Trans. Inf. Syst.**, v. e75-d, n. 6, p. 861-873, nov 1992.
- LANCASTER, F. W. **Indexação e Resumos: Teoria e Prática**. Brasília, Briquet de Lemos, 1993. 347p.
- LANCASTER, F. W., ELLIKER, Calvin, CONNELL, Tschera Harkness. Subject Analysis. **Annual Review of Information Science and Technology**, Vol. 24, p. 35-84, 1989.

- LAYNE, Sara Shatford. Some Issues in the Indexing of Images. **Journal of the American Society for Information Science**, New York, v. 45, n. 8, p. 583-588, sep 1994.
- LEMOS, Maria Lúcia Vilar de. Desenvolvimento de um Vocabulário Controlado na Biblioteca do Senado Federal. **Ciência da Informação**, Brasília, v. 15, n. 2, p. 155-158, 1986.
- LEUNG, C. H. C., HIBLER, D., MWARA, N. Picture Retrieval by Content Description. **Journal of Information Science**, London, v. 18, n. 2, p. 111-119, 1992.
- LOCHBAUM, Karen E., STREETER, Lynn A. Comparing and Combining the Effectiveness of Latent Semantic Indexing and the Ordinary Vector Space Model for Information Retrieval. **Information Processing & Management**, v. 25, n. 6, p. 665-676, 1989.
- MAMFRIM, Flávia Pereira Braga. Representação de Conteúdo Via Indexação Automática em Textos Integrais em Língua Portuguesa. **Ciência da Informação**, v. 20, n. 2, p. 191-203, jul/dez 1991.
- McCARN, Davis B., McCARN, Grace H. BEACON: H. W Wilson's Computerized Information System. **Information Technology and Libraries**, v. 2, n. 2, p. 142-150, jun 1983.
- McKNIGHT, Mark. Improving Access to Music: a Report of the MLA Music Thesaurus Project Working Group. **Notes**, v. 45, n. 4, p. 714-721, jun 1988.
- MILSTEAD, Jessica L. Methodologies for Subject Analysis in Bibliographic Databases. **Information Processing & Management**, v. 28, n. 3, p. 407-431, 1992.
- MILSTEAD, Jessica L., BERGER, Mary C. The Engineering Information Thesaurus Development Project. **Information Services & Use**, Amsterdam, v. 13, n.1, p. 71-80, 1993.
- NYAHODA, Philip. Towards an Indigenous Southern African Thesaurus. **African Research & Documentation**, n. 56, p. 43-45, 1991.
- PAO, Miranda Lee. **Concepts of Information Retrieval**. Colorado: Libraries Unlimited, 1989. 285p.
- PARSONS, Julia. Finding your Way Around the Information Maze. **Information Management & Technology**, v.25, n. 3, p. 120-121, may 1992.
- RABITTI, F., SAVINO, P. Automatic Image Indexation to Support Content-based Retrieval. **Information Processing & Management**, v.28, n. 5, p. 547-565, 1992.
- RACKMORE, Marlene. Facet Analysis and Thesauri for Corporate Information Retrieval. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 233-238.
- RADA, Roy. Connecting and Evaluating Thesauri: Issues and Cases. **International Classification**, v. 14, n. 2, p. 63-69, 1987.
- RADA, Ray, BARLOW, Judith, POTHARST, Jan, ZANSTRA, Pieter, BIJISTRA, Djujan. Document Ranking Using an Enriched Thesaurus. **Journal of Documentation**, v. 47, n. 3, p. 240-253, sep 1991.
- RESS-POTTER, Lorna K. Dynamic Thesauri: the Cognitive Function. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 145-150.
- RITZLER, Claus. Comparative Study of PC-Supported Thesaurus Software. **International Classification**, v.17, n. 3-4, p.138-147, 1990.

- ROWLEY, Jenny. Tracer: Indexing Package. **Information World Review**, n.19, p. 10-11, oct 1987.
- SAGER, J.C., SOMERS, H. L, MCNAUGHT, J. Thesaurus Integration in the Sciences. Pt. III. Guidelines for the Integration of Thesauri. **International Classification**, v. 9, n. 2, p. 64-70, 1982.
- SAVOY, Jacques. Stemming of French Words Based on Grammatical Categories. **Journal of the American Society for Information Science**, v. 44, n. 1, p. 1-9, jan 1993.
- SCHMITZ-ESSER, Winifried. New Approaches to Thesaurus Application. **International Classification**, v. 18, n. 3, p. 143-147, 1991.
- SCHUEGRAF, Ernest J., BOMMEL, Martin F. Van. An Automatic Document Indexing System Based on Cooperating Expert Systems: Design and Development. **Canadian Journal of Information and Library Science**, v. 18, n. 2, p. 32-50, jul 1993.
- SCHWARTZ, Candy, EISENMANN, Laura Malin. Subject Analysis. **Annual Review of Information Science and Technology**. v. 21, p. 37-61, 1986
- SHAW, W. M. Subject and Citation Indexing Part 1: The Clustering Structure of Composite Representations in the Cystic Fibrosis Document Collection. **Journal of the American Society for Information Science**, New York, v. 42, n. 9, p. 669-675, oct 1991.
- SRICIVASAN, Padmini. A Comparison of Two-Poisson, Inverse Document Frequency, and Discrimination Value Models of Document Representation. **Information Processing & Management**, v. 26, n. 2, p. 269-278, 1990.
- SRINIVASAN, Padmini. On Generalizing the Two-Poisson Model. **Journal of the American Society for Information Science**, New York, v. 41, n. 1, p. 61-66, jan 1990.
- STERN, Alice, RISCHETTE, Norbert. On the Construction of a Super Thesaurus Based on Existing Thesauri. In: INTERNATIONAL ISKO CONFERENCE, 1, 1990, Darmsdat. **Proceedings... Tools for knowledge organisation and the human interface**. Darmstadt: 1990, part 2, p. 134-144.
- STEWART, Martha M., HATTON Jackie L. Production of a Periodical Index by a Special Library. **Special Libraries**, New York, v. 83, n. 2, p. 92-95, spring 1992.
- SVENONIUS, Elaine. Unanswered Questions in the Design of Controlled Vocabularies. **Journal of the American Society for Information Science**, New York, v. 37, n. 5, p. 331-340, sep 1986.
- TALAMO, M. D. F. G. M., LARA, M. L. G. D., KOBASHI, N. Y. Contribuição da Terminologia para a Elaboração de Tesausos. **Ciência da Informação**, Brasília, v. 21, n. 3, p. 197-200, set/dez 92.
- TODESCHINI, C. **INIS: Manual for Indexing**. Vienna, International Atomic Energy Agency, jan 1974.
- VIEIRA, Simone Bastos. Análise Comparativa Entre Indexação Automática e Manual da Literatura Brasileira de Ciência da Informação. **Revista de Biblioteconomia de Brasília**, v. 16, n. 1, p. 83-94, jan/jun 88.
- WEINBERG, Bella Hass, CUNNING-HAM, Julie A. The Design of On-line Thesauri In: NATIONAL ON-LINE MEETING, 1988, New York. **Proceedings...** Medford, Learned Information, 1988, p. 411-419.

WELLISCH, Hans. H. The Art of Indexing and Some Fallacies of its Automation. **Logos**, v. 3, n. 2, p. 69-76, 1992.

WELLISCH, Hans. H. Book and Periodical Indexing. **Journal of the American Society for Information Science**, New York, v. 45, n. 8, p. 620-627, sep 1994.

8.3.6 - Armazenamento

ALLOWAY, James. Software Reviews: Flat-file Database Program. **Library Software Review**, v. 11, n. 6, p. 54-57, nov/dec 1992.

BEISER, Karl. Hands on Library Computing: AskSAM 5.0. **Computers in Libraries**, v. 12, n. 2, p. 26-29, feb 1992.

BELLARBY, Liz. Personal Bibliographic Packages: a Comparison of EndNote and Pro-Cite. **C & L Applications**, v. 6, n. 7, p. 3-4, mar 1993.

BELTRAMI, Giulio, KULKARNI, D. R. Data Compression Scheme for Large Databases on CD-ROM. In: INTERNATIONAL MEETING FOR OPTICAL PUBLISHING AND STORAGE, 1989, Amsterdam. **Proceedings ... Opticalinfo 89**. Oxford: Learned Information, 1989, p. 21-28.

BONOMI, Mauro. Multimedia and CD-ROM: an Overview of JPEG, MPEG and the Future. **CD-ROM Professional**, v. 4, n. 6, p. 38-40, Nov 1991.

BORNMAN, Hester, VON SOLMS, S. H. Hypermedia, Multimedia and Hypertext: Definitions and Overview. **Electronic Library**, v. 11, n. 4-5, p. 259-268, aug/oct 1993.

CLARK, Katie. Biblio-Link To SilverPlatter. **CD-ROM Professional**, v. 4, n. 5, p. 129-132, sep 1991.

CORAK, Robert. Removing the Mystique Surrounding Optical Disk Technology. **Drug Information Journal**. v. 26, n. 2, p. 163-165, apr/jun 1992.

COSGRIFF Jr., John C. The Use of AskSAM: A Full-text Database Program as an Added Value Enhancer to On-line Searching. **Library Software Review**, v. 6, n. 6, p. 374-375, nov/dec 1987.

DELOZIER, Eric P. Bibliography Management with Pro-CiteTM Version 2.0. **Medical Reference Services Quarterly**, v. 12, n. 2, p. 19-34, summer 1993.

ELECTRONIC Document Imaging: Magnetic Tape: What is its Role in Document Imaging. **Document image automation update**, v. 10, n. 4, p. 1-5, apr. 1991.

ELLSBURY, Susan H., ROBERTS, LaDonne. Selected List of Software for Organizing Bibliographic References and Creating Bibliographies. **Collection Management**, v. 16, n. 1, p. 55-72, 1992.

GILLMAN, Peter. Data Handling and Text Compression. **Journal of Information Science**, London, v. 18, n. 2, p. 105-110, 1992.

HANSON, Terry. LITREF. **Library Micromation News**, n. 29, p. 6-10, sep 1990.

HANSON, Terry. Pro-Cite and Biblio-Links. **Library Micromation News**, n. 26, p. 10-12, dec 1989.

HOPKINSON, Alan. CDS/ISIS Information. **Information Development**, v. 8, n. 1, p. 6-8, jan 1992.

- HOYLE, Norman, McNAMARA, Kathleen. Biblio-link and Pro-Cite: The Searcher's Workstation. **Database**, New York, v. 10, n. 1, p. 73-78, feb 1987.
- KAUFMAN, Daufman. Software Reviews: Bibliographic Database Manager. **Library Software Review**, v. 11, n. 6, p. 47-49, nov/dec 1992
- LOBO, Maria de Fátima Diniz, BARCELLOS, Silvia de Oliveira. Guias de Fontes de Informação: Metodologia para Geração e Automação. **Ciência da Informação**, Brasília, v. 21, n. 1, p. 75-81, jan/mar 1992.
- LUNDEEN, Gerald. Software for Managing Personal Files. **Database.**, New York, v. 12, n. 3, p. 36-48, jun 1989.
- LYNCH, Clifford A. The Technologies of Electronic Imaging. **Journal of the American Society for Information Science**, New York, v. 42, n. 8, p. 578-585, sep 1991.
- McCARTHY, Cavan Michael. Pro-Cite: Um Software para Gerenciamento de Bases de Dados Bibliográficos. **Ciência da Informação**, Brasília, v. 18, n. 2, p. 191-198, jul/dez 1989.
- MICHEL, Jacques. Multimedia for the 1990s: Technological Trends and Main Future Applications. In: INTERNATIONAL INFORMATION RESEARCH CONFERENCE, 2, 1991, Cambridge. **Proceedings ... Multimedia Information**. Cambridge: Bowker-Saur, 1991, p. 1-12.
- MORGAN, Paul. Hypertext and the Literary Document. **Journal of Documentation**, v. 47, n. 4, p. 373-388, dec 1991.
- MORRALL, Anne. Evaluation of Hypertext Software to Create a Public Browsing System in a Museum. **Electronic Library**, v. 9, n. 4-5, p. 217-234, aug/oct 1991.
- OLIVER, K. B. Methodology for Selecting Software for an On-line Database. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 9, 1985, London. **Proceedings...** London: 1985. p. 461-465.
- ROADS, Christopher H. Access and Conservation: the Key Elements in Sound Archiving in the 21st Century. **Phonographic Bulletin**, v. 59, p. 41-44, nov 1991.
- ROWLAND, Melanie, SEELEY, Maureen. Image Databases - the Facts of Life. **Audiovisual Librarian**, v. 17, n. 4, p. 217-220, nov 1991.
- SCHLICKE, Priscilla. **CD-ROM: the challenge of the new**. Information Management Report. sep 1991, p. 18-20.
- SIEVERTS, Eric G., HOFSTEDDE, Marten. Software for Information Storage and Retrieval Tested, Evaluated, and Compared: Part 1 - General Introduction. **The Electronic Library**, v. 9, n. 3, p. 145-154, jun 1991.
- SIEVERTS, Eric G., HOFSTEDDE, Marten, HAAK, Philip H., NIEUWENHUYSEN, Paul, SCHEEPSMA, Gerlinda A. M., VEEGER, Lucas, VIS, Gert C. Software for Information Storage and Retrieval Tested, Evaluated and Compared. Part II - Classical Retrieval Systems. **The Electronic Library**, v. 9, n. 6, p. 301-318, dec 1991.
- SPIEGLER, Israel, MAAYAN, Rafi. Storage and Retrieval Consideration of Binary Data Bases. **Information Processing & Management**, v. 21, n. 3, p. 233-254, 1985.
- STIGLEMAN, Sue. Pro-Cite 2.0: Steady Evolution. **Database**, v. 15, n. 6, p. 68-72, dec 1992.
- WITTEN, Ian H., BELL, Timothy C., NEVILL, Craig G. Indexing and Compressing Full-text Databases for CD-ROM. **Journal of Information Science**, London, v. 17, n. 5, p. 265-271, 1991.

WOLFF, J. Gerald. Computing, Cognition and Information Compression. **AI Communications**, v. 6, n. 2, p. 107-137, jun 1993.

8.4 - AVALIAÇÃO DA LITERATURA SOBRE PROJETOS DE SRI's

AGOSTI, Maristella, SPILOTRO, Franco. Design of Personal Information Retrieval Systems. **Information Services & Use**. v. 6, n. 3, p.161-168, 1986.

ARNOLD, Bruce, ROWELL, Hilary. The Australian Archives Records Information Service System. **Archives and Manuscripts**. v. 17, n. 1, p. 25-35, may 1989.

BAS, M. J. Le. Building a Useful Geochemical Database. **Outlook on Research Libraries**. v. 8, n. 9, p. 1-3, sep 1986.

BATT, Chris., MACDONALD, Sally. SCOTT, Tony. MUSLS - A Multi-media, Multi-discipline Database: Part 1, Defining Requirements, Selecting the System and Initial Development. **Program**. v. 27, n. 1, p. 17-36, jan 1993.

BAXTER, Paula A. Implementing Database Services for Art Research: Planning Considerations. **Art Documentation**. v. 6, n. 1, p. 16-18, spring 1987.

CHIANG, Katherine, et all. Beyond the Data Archive: The Creation of an Interactive Numeric File Retrieval System. **Library Hi Tech**. v. 11, n. 3, p. 57-72, 1993.

CLACK, Mary Elizabeth. The Balance Point: Indexing Policies. **Serials Review**. v. 18, n. 3, p. 49-58, fall 1992.

CRAWFORD, Josephine. Planning, Systems Analysis and Negotiation: Key Elements for a Successful GPO Tapeload. **Cataloging and Classification Quarterly**. v. 18, n. 3/4, p. 1-17, 1994.

CROTTS, Joe, BESNARD, Debbie. Developing an Historical Cartographic Database for Accessing Local History. **Western Association of Map Libraries Information Bulletin**. v. 25, n. 3, p. 129-137, jul 1994.

ERCEGOVAC, Zorana. Design, Implementation, and Evaluation of an Experimental Cataloguing Advisor-mapper. **Geography and Map Division Bulletin**. v. 163, p. 2-29, mar 1991.

FRANKLIN, Carl. An Introduction to Geographic Information Systems: Linking Maps to Databases. **Database**. v. 15, n. 2, p. 12-15, 17-21, apr 1992.

GIBBINS, P. Building an International Information Business - Some Problems and Opportunities. In: INFORMATION AND THE TRANSFORMATION OF SOCIETY; **Proceedings...** Amsterdam: 1982, p. 87-100.

HENNESSEY, Christine. Creating a Database for Art Objects: the National Museum of American Art's Inventory of American Sculpture. **Art Documentation**. v. 6, n. 4, p. 147-149, winter 1987.

HSIANG, Hsiao-Liang, WANG, Jenn-ann, LAI, Ruel-Lung, CHEN, Jin-Bang. The Structure, Creation and Searching of Ideographic Databases on STICNET. In: PACIFIC CONFERENCE ON NEW INFORMATION TECHNOLOGY FOR LIBRARY & INFORMATION PROFESSIONAL, EDUCATIONAL MEDIA SPECIALISTS & TECHNOLOGISTS, 2, Singapore, 1989, **Proceedings...**, Singapore: 1989, p. 165-171.

JAMES, Bonnie B. Designing Bibliographic Databases for Business Searchers. NATIONAL ON-LINE MEETING, 13, New Jersey, 1992. **Proceedings...**, New Jersey: 1992, p. 133-136.

- MENOU, Michel J., GIOVANNETTI, Jean-François, DUPEUBLE, Bernadette. BABINAT and the Development of Microcomputer - Based National Documentation Networks in the Less Developed Countries. **Information Processing & Management**. v. 27, n. 6, p. 699-712, 1991.
- MONEYHUN, Dora H. Building Dynamic Data Bases. In: NATIONAL ON-LINE MEETING, New Jersey, 1982. **Proceedings...** New Jersey: 1982, p. 381-388.
- O'HARA, Michael P., PAGIS, Catherine. The Pharmsearch Database. **Journal Chem. Inf. Comput. SCI**. v. 31, n. 1, p. 59-63, feb 1991.
- PHILIP, George, CROOKES, D., MORROW, P. J., JUHASZ, Z., QUINN. A. Use of a Transputer Network in the Development of an On-line Photographic Database. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 13, London, 1991. **Proceedings...** London: 1991, p. 315-326.
- ROMANOS, Vasso Apostolidis, FITTIPALDI, John J. The Energy Related Construction Information System ERCIS: Response to Need for Information on Army Facilities. **On-line Review**. v. 11, n.3, p. 167-181, jun 1987.
- SMITH, Robert, McSEAN, Tony. Planning and Producing the British National Bibliography on CD-ROM. **Program**. v. 23, n. 4, p. 395-413, oct 1989.
- STORM, William., KALINOSKI, Ronald. TIMS/DADS: A Project to Develop a System of Linking National and International Repositories of Multimedia Information. **IFLA Journal**. v. 18, n. 3, p. 223-227, 1992.
- THIEL, Thomas J., MECONNAHEY, Cathy M., GIELDA, Scott A. Document Indexing for Image-based Optical Information Systems. **Document Image Automation**. v. 11(2), p. 82-88, mar/apr 1991,.
- WARRILOW, Lynda. Design Issues in a Full Text News and Current Database. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 7, London, 1983. **Proceedings...** London: 1983, p. 411-418.