

LUCIANE FÁTIMA GABARDO FERREIRA PROSTT

**PROBLEMÁTICAS NA RECUPERAÇÃO DA INFORMAÇÃO EM SISTEMAS
VIRTUAIS: UM ESTUDO DE CASO DO PORTAL DA INFORMAÇÃO DA
UFPR**

**Trabalho de conclusão da disciplina
Pesquisa em Informação II, Curso de
Gestão da Informação, Setor de Ciências
Sociais Aplicadas, Universidade Federal
do Paraná.**

**Orientadora: Prof.^a Ligia Leindorf
Bartz Kraemer**

**CURITIBA
2001**

Prostt, Luciane Fátima Gabardo Ferreira

Problemáticas na recuperação da informação em sistemas virtuais:
um estudo de caso do Portal da Informação da UFPR / Luciane Fátima
Gabardo Ferreira Prostt. – Curitiba, 2001.

77 p. : il.

Trabalho de conclusão da disciplina Pesquisa em Informação II, Curso
de Gestão da Informação, Setor de Ciências Sociais Aplicadas,
Universidade Federal do Paraná.

1. Sistema de informação. 2. *World Wide Web* (Sistema de recuperação
da informação). 3. Universidade Federal do Paraná – Portal da Informação.
I. Título.

CDD 004.6

SUMÁRIO

LISTA DE SIGLAS	iv
LISTA DE FIGURAS	v
LISTA DE QUADROS	vii
RESUMO	viii
1 INTRODUÇÃO	1
2 SISTEMAS DE INFORMAÇÃO	5
3 RECUPERAÇÃO DA INFORMAÇÃO	14
3.1 A AMBIGÜIDADE DAS PALAVRAS.....	15
3.2 TÉCNICAS DE BUSCA.....	16
3.2.1 Operadores Lógicos (ou Booleanos).....	17
3.2.2 Operadores de Existência (ou Matemáticos).....	20
3.2.3 Operadores de Proximidade.....	20
3.2.4 Truncagem.....	22
4 ÍNDICE DE REVOCAÇÃO E PRECISÃO	24
5 SISTEMAS DE BUSCA NA INTERNET	24
5.1 DIRETÓRIOS E FERRAMENTAS DE BUSCA.....	25
5.1.2 Diretórios.....	25
5.1.3 Ferramentas de Busca.....	27
5.2 META-FERRAMENTAS DE BUSCA.....	29
5.3 PORTAIS.....	30
5.4 SERVIÇOS PERSONALIZADOS.....	34
6 PROCEDIMENTOS METODOLÓGICOS DO ESTUDO DE CASO	35
7 APRESENTAÇÃO E INTERPRETAÇÃO DOS RESULTADOS	36
7.1 ABRANGÊNCIA E PROPÓSITO.....	37
7.2 CONTEÚDO.....	39
7.3 PLANEJAMENTO VISUAL/GRÁFICO (<i>WEBDESIGN</i>).....	40
7.4 FUNCIONALIDADE (INTERFACE E NAVEGABILIDADE).....	41
8 CONSIDERAÇÕES FINAIS	63

9 SUGESTÕES.....	64
GLOSSÁRIO.....	67
REFERÊNCIAS.....	74

LISTA DE SIGLAS

- HTML** - *Hypertext Markup Language*
- HTTP** - *Hypertext Transfer Protocol*
- IA** - Inteligência Artificial.
- SGML** - *Standard Generalized Markup Language*
- TCP/IP** - *Transmission Control Protocol/Internet Protocol*
- TGS** - Teoria Geral de Sistemas
- TIC** - Tecnologias de Informação e Comunicação
- UFF** - Universidade Federal Fluminense
- UFPR** - Universidade Federal do Paraná
- URL** - *Uniforme Resource Locator*
- XML** - *Extensible Markup Language*
- WWW** - *World Wide Web*

LISTA DE FIGURAS

FIGURA 1 – PORTAL DA INFORMAÇÃO DO SISTEMA DE BIBLIOTECAS DA UFPR.....	36
FIGURA 2 – <i>HOMEPAGE</i> DO PORTAL DA INFORMAÇÃO DA UFPR.....	37
FIGURA 3 – <i>HOMEPAGE</i> DO PORTAL DE REFÊNCIA EM ARQUIVOLOGIA, BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO DA UNIVERSIDADE FEDERAL FLUMINENSE (UFF).....	42
FIGURA 4 – INTERFACE DA PÁGINA BASES <i>ON-LINE</i> DO PORTAL DA INFORMAÇÃO DA UFPR.....	43
FIGURA 5 – INTERFACE DA PÁGINA NOVIDADE DO PORTAL DA INFORMAÇÃO DA UFPR.....	44
FIGURA 6 – INTERFACE DA PÁGINA BASES <i>ON-LINE</i> DE ACESSO REGULAMENTADO DO PORTAL DA INFORMAÇÃO DA UFPR.....	45
FIGURA 7 – INTERFACE DA PÁGINA BASES <i>ON-LINE</i> DE ACESSO PÚBLICO DO PORTAL DA INFORMAÇÃO DA UFPR.....	46
FIGURA 8 – INTERFACE DA PÁGINA BASES REFERENCIAIS DO PORTAL DA INFORMAÇÃO DA UFPR.....	47
FIGURA 9 – INTERFACE DA PÁGINA BASES TEMÁTICAS DO PORTAL DA INFORMAÇÃO DA UFPR.....	48
FIGURA 10 – MAPA DO <i>SITE</i> DO PORTAL DA INFORMAÇÃO DA UFPR.....	49
FIGURA 11 – CÓDIGO FONTE DA <i>HOMEPAGE</i> DO PORTAL DA INFORMAÇÃO DA UFPR.....	50
FIGURA 12 – INTERFACE DA PÁGINA ACERVO DO PORTAL DA INFORMAÇÃO DA UFPR.....	52
FIGURA 13 – DETALHE DE OBRA RECUPERADA NA INTERFACE LIVROS.....	54
FIGURA 14 – DETALHE DE OBRA RECUPERADA NA INTERFACE FONTES BIBLIOGRÁFICAS.....	54

FIGURA 15 – DETALHE DE OBRA (LOCALIZADA NA BIBLIOTECA DE HUMANAS E EDUCAÇÃO DA UFPR) RECUPERADA NA INTERFACE PERIÓDICOS.....	55
FIGURA 16 – DETALHE DE OBRA (LOCALIZADA NA BIBLIOTECA DE CIÊNCIAS FLORESTAIS E DA MADEIRA DA UFPR) RECUPERADA NA INTERFACE PERIÓDICOS.....	56
FIGURA 17 – INTERFACE DE BUSCA LIVROS DO PORTAL DA INFORMAÇÃO DA UFPR.....	58
FIGURA 18 – INTERFACE DE BUSCA FONTES BIBLIOGRÁFICAS DO PORTAL DA INFORMAÇÃO DA UFPR.....	59
FIGURA 19 – INTERFACE DE BUSCA DOS PERIÓDICOS DO PORTAL DA INFORMAÇÃO DA UFPR.....	60
FIGURA 20 – INTERFACE DE APRESENTAÇÃO DO RESULTADO DE BUSCA.....	61

LISTA DE QUADROS

QUADRO 1 – COMPARAÇÃO ENTRE AS CARACTERÍSTICAS DAS INTERFACES DE BUSCA DO SISTEMA DE BIBLIOTECAS DA UFPR.....	53
QUADRO 2 – RESULTADO DE BUSCA COM PROBLEMA DE POLISSEMIA.....	62
QUADRO 3 – RESULTADO DE BUSCA COM PROBLEMA DE SINONÍMIA.....	62
QUADRO 4 - RESULTADO DE BUSCA COM PROBLEMA DE HOMOGRAFIA.....	63

RESUMO

Pesquisa exploratória, descritiva e bibliográfica sobre sistemas de informação desde o início de sua concepção, com a Teoria Geral de Sistemas (TGS), do estudioso BERTALANFFY, até a “Era dos Portais”, com o advento da Internet. Aborda o agravamento dos problemas na recuperação da informação, na passagem dos sistemas de informação centralizados para os virtuais, descrevendo o problema da ambigüidade das palavras (polissemia, sinonímia e homonímia), as diversas técnicas de busca (operadores lógicos, de existência, de proximidade e truncagem), os sistemas de busca (diretórios, ferramentas de busca, meta-ferramentas de busca, portais e serviços personalizados). Define e exemplifica índice de revocação e índice de precisão em um sistema de informação. Efetua um estudo de caso sobre o Portal da Informação da Universidade Federal do Paraná (UFPR), utilizando-o como exemplo de um sistema de informação virtual, avaliado mediante aplicação de um roteiro de critérios de avaliação de *sites* de informação, tais como: abrangência e propósito; conteúdo; planejamento visual/gráfico (*webdesign*) e funcionalidade (interface e quesitos de navegabilidade). Delimita a interface de busca do Acervo, do Sistema de Bibliotecas da UFPR, pertencente ao referido portal, como elemento de amostra para a verificação da reação do mesmo à aplicação de palavras ambíguas e das referidas técnicas de pesquisa na execução de buscas, assim como traça os índices de revocação e precisão por meio dos resultados obtidos na recuperação da informação. Sugere, como resultado, propostas para a melhora do portal, de acordo com os critérios de avaliação.

1 INTRODUÇÃO

Ao procurar definir informação, no contexto da explosão e do caos documentários, o autor clássico BRIET¹, *apud* BRAGA (1995), enfocou-a como sinônimo de documento, ou seja, com uma “base de conhecimento, fixada materialmente, suscetível de estudo, prova ou confronto”. Ou seja, a informação é colocada pelo autor como algo tangível, pois encontra-se alocada em um suporte físico (papel, filme, disquete), sendo passível de manipulação.

Em contrapartida, SHANNON & WEAVER², *apud* BRAGA (1995), ao retirarem a informação do suporte físico imposto por BRIET e transportarem-na para um suporte qualquer, como o ar, por exemplo, colocam-na como a essência intangível de um processo de comunicação, no qual a mesma independe de suporte físico, mas necessita de um emissor, um receptor e um canal (processo de comunicação) para difundir-se.

Unificando-se as diferentes visões de BRIET; SHANNON & WEAVER, verifica-se a existência de informações tangíveis e intangíveis, que tornam ainda mais complexa e caótica a explosão da informação. Para BRAGA³, *apud* BRAGA (1995), “o caos parece ser o princípio da informação e um atrator para sistemas de recuperação da informação”. Nesse contexto, desde a década de 50 vêm surgindo os sistemas automatizados de informação, advindos da necessidade de armazenar e acessar rápida e relevantemente milhares de documentos que apresentavam crescimento exponencial desde o século XVII.

¹ BRIET, S. **Qu'est-ce que la documentation?** Paris: Presses Universitaires de France, 1953.

² SHANNON, C.; WEAVER, W. **The mathematical theory of communication.** [Urbana]: [s.n.], 1949.

³ BRAGA, G. M. Caos, princípio da informação. **Cadernos de Ciência da Informação**, Belo Horizonte, 1995.

Inicialmente esses sistemas eram centralizados, com política definida, planejados a longo prazo e estáveis. Posteriormente, com o advento das novas tecnologias de informação e comunicação (TIC), em especial a Internet, o caos informacional tornou-se ainda mais complexo e trouxe consigo novos formatos e canais de comunicação que não conhecem barreiras geográficas, modificando o perfil dos sistemas de informação, extrapolando na quantidade de informações disponíveis e ultrapassando as possibilidades do acervo local para o virtual.

A virtualização dos sistemas de informação agravou as dificuldades de recuperação da informação. Um exemplo são os diversos sistemas de busca⁴ que, ao executarem uma determinada pesquisa, oferecem centenas e mesmo, não raramente, milhares de referências das quais somente poucas são relevantes e atendem às necessidades de informação do usuário, encontrando-se dispersas no decorrer da lista de resultados. Portanto, intui-se que a política de métodos de busca empregados pelos sistemas são insuficientes para recuperar informações com precisão. Sendo assim, verifica-se a necessidade de estudos sobre os problemas de recuperação da informação que apontem para soluções do ponto de vista do tratamento da informação.

Dentre os diversos problemas existentes na recuperação dos sistemas de informação, KURAMOTO (2002) cita tanto o problema da ambigüidade das palavras: polissemia, sinonímia, homografia, quanto a diversidade das técnicas de busca, cada qual com suas diferenças no modo de recuperação: operadores lógicos (ou booleanos), de existência (ou matemáticos), de proximidade, truncagem e outros. Esses problemas foram delimitados como objetos de estudo desta pesquisa para poder atingir o seu objetivo geral que é descrever os elementos negativos existentes na recuperação da informação em sistemas de informação virtual, contribuindo para o desenvolvimento e o aprimoramento dos sistemas de informação.

⁴ Neste trabalho, “sistema de busca” é adotado como um termo genérico que engloba os conceitos: diretórios, ferramentas, meta-ferramentas, portais e os serviços personalizados.

Para que o objetivo geral desta pesquisa fosse atingido, foram relacionados os seguintes objetivos específicos:

- a) focar conceitual, tipológica e caracteristicamente os sistemas de informação desde a concepção da Teoria Geral de Sistemas (TGS), do estudioso BERTALANFFY, até o advento da Internet;
- b) abordar teórico-conceitualmente os problemas da recuperação da informação relacionados a: ambigüidade das palavras; técnicas de busca; índices de revocação e precisão; sistemas de busca na Internet;
- c) analisar um sistema de informação virtual, utilizando-se como modelo o Portal da Informação da Universidade Federal do Paraná (UFPR), aplicando um roteiro de critérios de avaliação de *sites* de informação, tais como abrangência e propósito; conteúdo ; planejamento visual/gráfico (*webdesign*) e funcionalidade (interface e quesitos de navegabilidade), assim como para a investigação dos índices de revocação e precisão do mesmo.

Trata-se de uma pesquisa do tipo bibliográfica caracterizada pelos procedimentos de coleta de dados em busca de materiais já elaborados e publicados. Pode também ser considerada como de cunho exploratório por seus objetivos e descritiva pela inclusão de estudo de caso (SANTOS, 1999, p. 26-29).

Assim sendo, inicia-se no capítulo 2 a literatura pertinente sobre sistemas de informação, citando-se suas várias definições, assim como caracterizando historicamente a evolução dos mesmos.

A recuperação da informação é abordada no capítulo 3, enfocando-se tanto o problema da ambigüidade das palavras quanto a diversidade de técnicas de busca existentes, mediante descrições e exemplos.

O capítulo 4 define os índices de revocação e precisão, exemplificando-os por meio da execução de uma busca em base de dados.

Os sistemas de busca na Internet são estudados no capítulo 5, sendo descritos os diretórios e ferramentas de busca; as meta-ferramentas de busca, os portais e os serviços personalizados.

O capítulo 6 arrola os procedimentos para o estudo de caso do Portal da Informação da UFPR que, posteriormente, tem a interpretação e análise de seu estudo apresentando no capítulo 7.

As considerações finais para o Portal da Informação são apresentadas no capítulo 8 e, as sugestões para o aprimoramento do mesmo, no capítulo 9.

Em uma seção pós-textual, apresenta-se um glossário de termos, nomes e expressões utilizados nesta pesquisa, complementando-a e, posteriormente, arrola-se as referências para a sua execução.

2 SISTEMAS DE INFORMAÇÃO

Em meio às mudanças de paradigmas impostas pela II Guerra Mundial, em 1945, equipes trabalhavam de forma interdisciplinar, com especialistas das mais diversificadas áreas, para resolver as complexidades advindas do momento. O resultado foi a concepção da Teoria Geral dos Sistemas (TGS), como “instrumento organizador de um universo complexo” e das várias idéias comuns geradas pelas mais diferentes disciplinas ou ciências. Seu idealizador foi o biólogo austríaco Ludwig von Bertalanffy (1901-1972), radicado no Canadá, professor da Universidade de Alberta, em Edmonton (Canadá), que explica a idéia em seu livro *Teoria Geral de Sistemas*: “um sistema pode ser definido como um conjunto de elementos em interrelação entre si e com o ambiente” (BERTALANFFY, 1975, p. ix).

Especificamente, para CHURCHMAN (1972, p. 17-18), sistema é um “conjunto de elementos com funções próprias que interagem, agem em conjunto para atingir um ou mais objetivos e pode ser simples ou formado por vários subsistemas que interagem entre si e se complementam para alcançar objetivos”.

Historicamente, para TOFFLER (1999, p. 28-30) existiram ao longo dos anos “três ondas” ou fases, que podem ser definidos como avanços tecnológicos, que revolucionaram e continuarão revolucionando a sociedade. A primeira onda foi a agrícola, a segunda industrial e a terceira dos sistemas de informação que, atualmente, segundo POLLONI (2001, p. 17), é a onda na qual encontramos-nos sob efeito.

Trata-se do início da era pós-industrial, como bem ilustra BARBOSA⁵, *apud* CARVALHO & KANISKI (2000, p. 34), na qual predominam:

... os esforços (científicos, tecnológicos e políticos) no sentido de informatizar a sociedade. Se, por um lado, o avanço e a cotidianização da tecnologia informática já nos impõem sérias reflexões sobre questões [éticas, deontológicas, jurídico-políticas, de soberania, culturais e político-sociais], por outro lado, seu impacto sobre a ciência vem se revelando considerável [pois essa deixa de ser] vista como atividade ‘nobre’, ‘desinteressada’, sem finalidade

⁵ BARBOSA, W. do V. Tempos pós-modernos. In: LYOTARD, J. F. **O pós-moderno**. Rio de Janeiro: J. Olympio, 1986.

preestabelecida [para se fortalecer como recurso gerador de riqueza, pois] descobriu-se que a fonte de todas as fontes chama-se **informação** e que a ciência – assim como qualquer modalidade de conhecimento – nada mais é que um certo modo de organizar, estocar e distribuir certas informações ..., o que representa entendê-la como fonte de mensagens na lógica computacional do 0 e do 1, exigindo um **sistema de informação** ágil .

As novas TIC, aceleraram a evolução dos sistemas de informação, assim como os processos econômicos e socioculturais, a partir do instante em que passaram a converter conhecimento em vantagem competitiva, pois os ativos do capital intelectual aumentaram com o uso, ou seja o conhecimento compartilhado gerou mais conhecimento, permanecendo este com o seu doador, mas ao mesmo tempo enriquecendo o seu receptor, gerando indefinidamente novas idéias.

Assim sendo e considerando-se o fator humano, um sistema de informações pode ser caracterizado como uma tecnologia intelectual porque afeta a organização das funções cognitivas do homem: a coleta, o armazenamento e a análise de informações assim como atividades de previsão, concepção, escolha e decisão.

Para CASSARRO (1999, p. 27-28), a informação é o ativo mais importante de uma organização, tornando-a mais dinâmica, agressiva e atuante que as outras a partir do momento em que dispõe de um sistema de informação qualificado, ou seja, um conjunto de funções logicamente estruturadas, que possibilitem planejamento, coordenação e controle da organização, com a finalidade de subsidiar a tomada de decisões.

Tanto STAIR (1998, p. 11) quanto POLLONI (2001, p. 30-31) acordam que os sistemas de informação são sistemas específicos que objetivam a realização de processos de comunicação, mediante uma série de elementos ou componentes inter-relacionados que coletam (entrada), manipulam e armazenam (processo), disseminam (saída) os dados e informações e fornecem um mecanismo de *feedback*:

- a) entrada: captura e junta dados primários caracterizados como itens em forma primária, dispersos e incoerentes (ex.: total de unidades vendidas, datas, descrição de produtos e clientes do mês);
- b) processamento: transforma os dados em saídas úteis, ou seja em

informações relevantes (ex.: relatório mensal de vendas, com a data de cada venda, descrição e quantidade de produtos vendidos por cliente);

c) saída: produz informações úteis (ex.: documentos relatórios, dados de transações);

d) *feedback* (realimentação do sistema): saída utilizada para fazer ajustes ou modificações nas atividades de entrada ou processamento (ex.: erros ou problemas podem ser corrigidos na entrada do sistema para que expresse corretamente as informação na saída).

Nos sistemas de informação estão envolvidos tanto o fator humano quanto o fator físico, fazendo-se necessária a utilização de equipamentos que possam suprir necessidades, que colaborem no suporte às atividades desempenhadas pelos seus profissionais, assim como no aprimoramento de seus colaboradores.

Encontra-se tal colocação em FREITAS & LESCA (1992, p. 95) ao definirem sistema de informação como um conjunto interdependente das pessoas, das TIC, dos procedimentos e métodos que deveriam permitir à organização dispor, no tempo desejado, das informações que necessita (ou necessitará). Acrescentam os referidos autores que um sistema de informação possibilita a ligação entre três componentes organizacionais: as pessoas que participam no processo de informação da organização, as estruturas organizacionais e as TIC. Nessa implantação, é necessário que todo o esforço e investimentos estejam voltados para o negócio da organização. O impacto mais significativo na implantação de tecnologias da informação é na produtividade: as operações antes realizadas manualmente, passam a ser realizadas por meios mais rápidos e seguros, agilizando todos os processos encadeados.

Neste sentido, FREITAS *et al.* (1997, p. 77) colocam os sistemas de informação como mecanismos (de interação entre homem e tecnologia), cuja função é coletar, guardar e distribuir informações para suportar as funções gerenciais e operacionais das organizações.

Em uma conceituação bem abrangente, SOARES NETO (1993, p. 22)

compreende o sistema de informação como um conjunto de elementos interrelacionados, processos, dados e tecnologia, cuja finalidade é alimentar os centros de decisão com as informações necessárias à escolha de diretrizes de ação que permitam a consecução dos objetivos da organização.

Sendo assim, os sistemas de informação possibilitam aos gerentes tomarem decisões apropriadas, seja pelo número de informações disponíveis, seja pela possibilidade de organização e estruturação de suas funções.

Dentro deste contexto, BIO (1996, p. 34-35) classifica os sistemas de informação em dois grupos principais:

- a) sistema de apoio às operações: típico sistema de processamento de transações que trata e processa trocas diárias de negócios para auxiliar nas decisões voltadas para a operação como, por exemplo, controle de estoque, folha de pagamento, entre outros;
- b) sistema de apoio à gestão: sistema de tomada de decisão, voltado para as decisões referentes às operações e muitas outras operações complexas e transacionais tais como planejamento e controle de produção, custos, contabilidade, entre outras.

Conceitualmente, BIO (1996, p. 120) afirma que se faz necessário distinguir as informações operacionais das gerenciais. O referido autor destaca que uma informação operacional, gerada por um sistema qualquer, tem por finalidade simplesmente permitir que determinadas operações continuem ocorrendo dentro do ciclo operacional da organização. Já as informações de natureza gerencial destinam-se a alimentar os processos de tomada de decisão.

Para ROSARIO GIRARDI (2001, p. 12) os sistemas de informação possuem uma classificação mais complexa que a de BIO e estão divididos segundo dois critérios principais:

- a) hierárquico: ligado aos níveis organizacionais onde o sistema de informação é utilizado;

b) funcional: relacionado às atividades desenvolvidas por seus usuários.

Conforme a autora, as informações das organizações dividem-se em três níveis:

- a) nível de controle operacional (básico): contém informações obtidas no interior das organizações, utilizadas com alta frequência e detalhadas;
- b) nível de controle gerencial (intermediário): possui informações de baixa frequência de utilização e referentes ao futuro;
- c) nível de planejamento estratégico (alto): encerra informações obtidas no ambiente externo e de valor agregado.

E, mediante esta divisão, a hierarquia classifica os sistemas de informação em:

- a) sistema de processamento de transações: armazena e processa dados sobre as atividades de rotina dos negócios. Suas operações baseiam-se em procedimentos padronizados que estabelecem exatamente como dar entrada em dados, processá-los e apresentar resultados na saída como, por exemplo, as transações bancárias automáticas (aplicações, transferências entre contas e saques);
- b) sistema de apoio ao planejamento e controle operacional: processo pelo qual os gerentes asseguram que os recursos são obtidos e utilizados de forma eficiente e eficaz para alcançar os objetivos organizacionais como, por exemplo, em um processo de fabricação, esse sistema apoia o processo de desenvolvimento de produção diária, realizando um *feedback* constante, até decidir-se por uma solução satisfatória;
- c) sistema de apoio ao planejamento e controle gerencial: sistema de apoio à tomada de decisões em ambientes complexos e não-estruturados (intuitivos). Trabalha com dados agregados, oriundos tanto da organização como de fontes externas;
- d) sistema de apoio ao planejamento estratégico: fornece informações para apoiar os gerentes de alto nível a tomar decisões focalizando as questões

que afetam a capacidade da organização em competir, a longo prazo, nos mercados de produtos e serviços em que opera.

De acordo com ROSARIO GIRARDI (2001, p. 15) o critério funcional classifica os sistemas em:

- a) sistema de processamento de transações, descrito acima;
- b) sistema de informação gerencial;
- c) sistema de apoio à decisão;
- d) sistema de informação para executivos.

Nos anos de 1950 e 1960, começaram a surgir os primeiros sistemas de informação computadorizados, os quais focavam o nível operacional da organização. Com o passar do tempo, outros tipos de sistemas de informações vieram agregar-se aos anteriores, atendendo às diferentes necessidades das organizações. O primeiro a ser desenvolvido foi o **sistema de processamento de transações**. Após a organização das atividades operacionais, surgiu a necessidade de gerar informações consolidadas, seguras e rápidas, para que os gerentes pudessem ter maior certeza de que rumo tomar e como a organização se encontra no mercado, o que auxiliaria no processo decisório da mesma. Com isso, foi desenvolvido o sistema de informação gerencial, que auxiliaria os gerentes nessas questões (O'BRIEN, 2001, p. 27).

Segundo STAIR (1998), CASSARO (1999), LAUDON & LAUDON (1999), OLIVEIRA (1998), O'BRIEN (2001), POLLONI (2001) e ROSARIO GIRARDI (2001), o **sistema de informação gerencial** é um processo de transformação de dados em informações que são utilizadas na estrutura decisória da empresa, através do provimento ao gerente de informações, passadas e presentes, sobre as operações, reunindo informações pertinentes a cada uma delas e, proporcionando, ainda, a sustentação administrativa para otimizar os resultados esperados.

A demanda por diferentes tipos de sistemas de informação começou a crescer no início dos anos 70. Concomitantemente, acarretou a necessidade de se obter um sistema de informação que ao apoiar a tomada de decisão, aumentasse a qualidade da

mesma. Assim surgiu o sistema de apoio à decisão (O'BRIEN, 2001, p. 27).

O **sistema de apoio à decisão**, de acordo com FREITAS (1997), OLIVEIRA (1998), STAIR (1998), LAUDON & LAUDON (1999), O'BRIEN (2001), POLLONI (2001) e ROSARIO GIRARDI (2001), auxilia o executivo nas etapas de tomada de decisão, utilizando modelos e dados de diferentes fontes para simulações, considerando-se o estilo do decisor. Pode ser classificado em *ad hoc*, quando desenvolvido para uma situação única, e institucional, quando utilizado em situações que ocorrem periodicamente.

Este tipo de sistema de informação computadorizado fornece, normalmente, suporte às decisões estruturadas, semi-estruturadas e não-estruturadas, abaixo definidas:

- a) estruturadas: ligadas diretamente à gerência de controle operacional, seguem normas e/ou regras bem definidas e seus procedimentos são documentados e automatizados;
- b) semi-estruturadas: envolvem a combinação de soluções e procedimentos padrões, que não mudam e julgamento individual baseado na experiência. Não são automatizadas por completo, requisitando informações não-estruturadas, de forma conjunta, resumidas e com projeções;
- c) não-estruturadas: são decisões tomadas pelo alto escalão da organização. Tratam de processos vagos e problemas complexos, onde a intuição humana é frequentemente utilizada para tomar tais decisões.

Tendo quase todas as necessidades das organizações e principalmente dos gerentes atendidas, ainda faltava um sistema que fornecesse dados e informações com mais rapidez para os executivos. Após análises de certas características que envolviam tais decisões, foi desenvolvido o sistema de informação para executivos, que possibilitaria maior aproximação entre o executivo e o sistema de informação para a obtenção das informações em menor tempo, melhorando, assim, as decisões estratégicas, e possibilitando autonomia de pesquisa ao decisor. O sistema de

informação para executivos surgiu nos anos 80, tendo como público alvo o nível estratégico das organizações, ou seja, os altos executivos das mesmas (O'BRIEN, 2001, p. 27).

FURLAN *et al.* (1994), OLIVEIRA (1998), STAIR (1998), LAUDON & LAUDON (1999), O'BRIEN (2001), POLLONI (2001) e ROSARIO GIRARDI (2001), apontam o **sistema de informação para executivos** como sistema de apoio às atividades de planejamento estratégico, cujo objetivo principal é a filtragem dos dados mais relevantes para os executivos, reduzindo o tempo de obtenção e gerando informações de real interesse, as quais permitam o acompanhamento e controle da organização. As principais características deste tipo de sistema são eliminar o intermediário entre o executivo e o computador, adaptar-se ao estilo de decisão do executivo, ser de fácil utilização, fornecer uma visão global e precisa da organização, possuir recursos gráficos de alta qualidade para que as informações possam ser apresentadas de várias formas e destaquem exceções e variações, ter a capacidade de *drill down* (visualização das informações em vários níveis de detalhe).

Além destes sistemas, verifica-se a existência de:

- a) sistemas especialistas;
- b) sistemas documentais.

Ainda no início dos anos 80, constatou-se a necessidade de pessoas especializadas em determinadas funções, o que tornava os custos maiores e a necessidade constante dessa pessoa na organização. Criou-se então o conceito de inteligência artificial que originaria o sistema especialista, cujo objetivo é a incorporação do conhecimento humano sobre atividades especializadas (O'BRIEN, 2001, p. 28).

O **sistema especialista**, baseia-se em uma das técnicas de inteligência artificial (outros exemplos seriam a linguagem natural e a robótica) que modela o raciocínio de especialistas na resolução de problemas. Assim sendo, todo o conhecimento do especialista, deve ser transferido para a máquina, possibilitando uma

maior agilidade na conclusão das operações, conforme STAIR (1998), LAUDON & LAUDON (1999), O'BRIEN (2001) e POLLONI (2001).

E, finalmente, pode-se citar o sistema documental, considerado como “registro da memória humana”, ou seja, comum a todos nós. Antes, porém faz-se necessário enfatizar que no contexto de um sistema documental, dados e informações muitas vezes são tratados como sinônimos, pois alguns autores consideram dados, referindo-se ao que está realmente armazenado na base de dados e, informações, para referenciar o significado desses dados na recuperação da informação de um sistema documental (DATE, 2000, p. 4).

Um **sistema documental**, segundo FREIRE & ARAÚJO (1999), é um sistema de informação que manipula informações potencialmente contidas em documentos impressos e/ou eletrônicos.

Tais informações, são categorizadas por BELKIN & ROBERTSON⁶, *apud* ARAÚJO (1995), como “cognitivos sociais”, pois tratam-se de estruturas conceituais sociais referentes ao conhecimento, ou seja, as estruturas de conhecimento compartilhadas pelos membros de um grupo social (manuscritos, livros, periódicos, mapas, filmes, vídeos, quadros, partituras, CD-ROMS, disquetes, base de dados e outros).

Para ARAÚJO (1995), o sistema documental remonta à origem do suporte mais rudimentar de armazenamento da informação, a terracota: um tipo de argila, que após modelada e cozida, recebia de nossos antepassados primatas a inserção de seus sinais e desenhos, com o objetivo de registrar ali suas memórias para a posteridade.

Ainda segundo a autora, o sistema documental passou por inúmeras transformações até culminar em “grandes sistemas de bases de dados em registros magnéticos capazes de mandar, de um canto ao outro do mundo, grandes volumes de mensagens, a velocidades fantásticas e de armazenar milhões de informações em *microchips*” (ARAÚJO, 1995).

⁶ BELKIN, N. J.; ROBERTSON, S. E. Information science and the phenomenon of information. *Jasis*, v. 27, n. 4, p. 197-204, 1976.

Assim sendo, se faz necessária a visão de abrangência de sistemas de para o entendimento da inserção da problemática da representação da informação, para a sua eficácia e reuso.

3 RECUPERAÇÃO DA INFORMAÇÃO

Segundo LANCASTER (1993, p. 15-17), GUINCHAT & MENOUE (1994, p. 75-76) e MARCONDES (2001), a indexação é o melhor meio pelo qual procura-se representar (descrever e identificar) o conteúdo de um documento, esteja ele em formato impresso ou eletrônico, para que o mesmo possa depois ser recuperado.

O termo representar um documento, significa:

representar objetos e conceitos na mente, manipular estas representações antes de agir na realidade de acordo com os resultados da manipulação mental prévia e socializar estas experiências sob a forma de representações orais ou escritas, para outros membros da espécie, é uma das mais fundamentais características do gênero humano. Constitui o que chamamos de cultura, uma segunda natureza humana intersubjetiva, capaz de transcender os limites individuais de espaço e tempo (MARCONDES, 2001, p. 61).

Antes do surgimento das novas tecnologias de informação, o processo da representação de documento nas antigas bases de dados era realizado por profissionais da área de informação especializados em indexação, que utilizavam instrumentos (vocabulários controlados, tesouros, tabelas, listas e outros) padronizados que forneciam descritores adequados à elaboração da representação de cada documento.

Com o advento da Internet, MARCONDES (2001) coloca que os registros da cultura humana estão migrando para formatos digitais, ou mesmo, em grande parte, já são criados em formato digital. Diante disso, uma inesgotável quantidade de informações torna caótico o problema da identificação da informação relevante na rede. Assim sendo, a utilização dos descritores das antigas bases de dados, utilizados por indexadores e denominados por KURAMOTO (2002) como “termos portadores de informação, que fazem referência a um objeto ou fato do mundo real”, tornou-se

inviável, necessitando-se criar mecanismos de indexação automática, que se baseiam na extração de palavras.

LANCASTER (1993, p. 229-250) explica que “na indexação por extração, palavras ou expressões que aparecem num texto são extraídas e utilizadas para representar o conteúdo de um texto como um todo”. Esse tipo de indexação é realizado por um *software* que, como se fosse um ser humano, dentro de um documento procura extrair as palavras que aparecem com mais frequência dentro do texto, resumo, título, figuras e outros.

Segundo BRÄSCHER (2002), a problemática está inserida também no contexto das técnicas de busca que, sejam elas lógicas (ou booleanas), de existência (ou matemáticas), truncadas ou outras, ao utilizarem as palavras, que são ambíguas e possuem problemas de polissemia, sinonímia e homografia, como ponto de acesso ao conteúdo, fazem com que os sistemas de busca recuperem informações irrelevantes, trazendo consigo baixos índices de revocação e precisão.

3.1 A AMBIGÜIDADE DAS PALAVRAS

Segundo FUCHS⁷ (1987) e FUCHS⁸ (1996), *apud* BRÄSCHER (2002), por ambigüidade entende-se uma expressão da língua (palavra ou frase), que possui vários significados distintos, podendo, conseqüentemente, ser compreendida de diferentes maneiras por um receptor, ocorrendo quando palavras ou frases geram mais de uma interpretação de seu significado, como nos exemplos abaixo:

- a) na frase: *o arquivo está precisando de manutenção*, a ambigüidade latente

⁷ FUCHS, C. **L’ambigüité et la paraphrase en linguistique**. In: FUCHS, C., ed. *L’ambigüité et la paraphrase: opérations linguistiques, processus cognitifs, traitements automatisés*. Caen: Centre de Publications de L’Université de Caen, 1987, p. 9-20.

⁸ FUCHS, C. **Les ambigüites du français**. Paris: Orphys, 1996. 183 p.

da palavra *arquivo* pode culminar em várias interpretações de um *arquivo* como *móvel*, como *conjunto de documentos* ou como *instituição*;

b) na frase: *neutralização de contaminação do leite*, a ambigüidade permite interpretar que *a neutralização é feita com leite* ou que *a contaminação é causada pelo leite*.

Mediante esses exemplos, pode-se verificar que a ambigüidade das palavras possui características inerentes às mesmas, conforme KURAMOTO (2002) e BRÄSCHER (2002):

- a) polissemia: ocorre quando uma mesma palavra pode ter vários significados, numa única frase, como no exemplo da palavra *arquivo*, que pode ser um *móvel*, uma *instituição* ou um *conjunto de documentos*;
- b) sinonímia: ocorre quando duas palavras diferentes possuem o mesmo significado, como no exemplo do fruto *abóbora* que tem como sinônimo, no Nordeste do Brasil, o *jerimum*;
- c) homografia: ocorre quando uma palavra tem a mesma grafia (escrita) de outra, mas seus significados são diferentes, como no exemplo do *cobre (metal)* e *cobre (do verbo cobrir)*.

Verificando-se os exemplos citados, conclui-se com BRÄSCHER (2002) que a ambigüidade causa interferência na recuperação da informação.

Além disso, KURAMOTO (2002) aponta as técnicas de busca como outro fator de pouca precisão na recuperação da informação.

3.2 TÉCNICAS DE BUSCA

Técnicas de busca são métodos utilizados para recuperar a informação e podem ser classificados, segundo DENECA (2000, p. 21), como:

- a) operadores lógicos (ou booleanos);

- b) operadores de existência (ou matemáticos);
- c) operadores de proximidade;
- d) truncagem.

3.2.1 Operadores Lógicos (ou Booleanos)

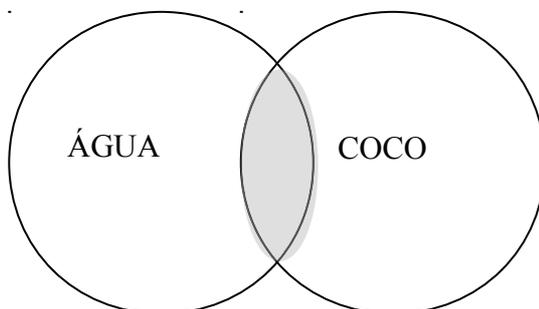
GUINCHAT & MENOUE (1994, p. 306), citam a pesquisa booleana como a mais utilizada em sistemas de informação tanto manuais como automatizados, por sua simplicidade e clareza. Como em qualquer outro sistema de informação, a lógica de Boole também é utilizada em sistemas de informação virtuais, como a Internet.

Como descrevem os autores, essa lógica de pesquisa resulta da operação da aplicação da lógica de Boole e permite estabelecer três tipos de relação entre termos/descriptores, cujo resultado é representado pelo diagrama de Venn:

- a) relação de intersecção;
- b) relação de união;
- c) relação de exclusão.

A **relação de intersecção**, que utiliza o operador “E” (AND), permite relacionar dois descritores, que devem ser encontrados na indexação de um mesmo documento para que a busca seja considerada precisa. Exemplo: ao executar uma pesquisa sobre *água de coco*, os itens recuperados devem, necessariamente, conter o descritor *água* e descritor *coco* ao mesmo tempo, como na representação da intersecção do conjunto unitário *água* com o conjunto unitário *coco*.

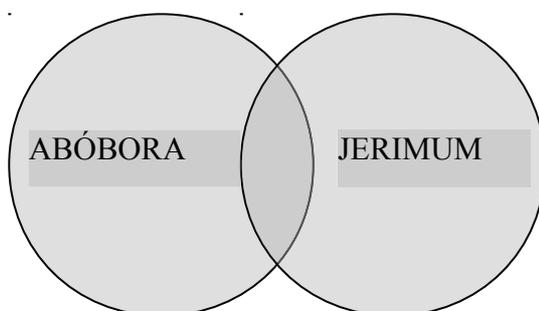
Diagrama de Venn do conjunto intersecção *água* AND *coco*



DENEGA (2000, p. 22) acrescenta que poderá ser utilizado o caracter & em vez de AND.

A **relação de união**, que utiliza o operador “OU” (OR), permite relacionar dois descritores, onde um ou outro, ou os dois devem ser encontrados. Geralmente, segundo MARCHIORI (2001), esse operador é mais utilizado em buscas de descritores com problemas de sinonímia, pois permite encontrar maior número de itens, mediante a utilização dos vários sinônimos de um termo, como é o caso de *abóbora* e *jerimum*, na representação do conjunto união entre o conjunto unitário *abóbora* e o conjunto unitário *jerimum*.

Diagrama de Venn do conjunto união *abóbora* OR *jerimum*



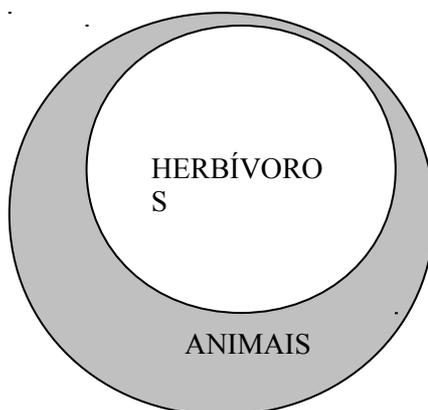
A autora destaca que os termos não precisam necessariamente serem sinônimos e, quando utilizada a operação OU o resultado será a união do primeiro com o segundo conjunto, cujo resultado conterà documentos com um ou outro ou ambos os termos.

Para a autora, o operador OU é útil também quando não se tem certeza da grafia da palavra, como no exemplo de busca *cusco* OU *cuzco*.

DENEGA (2000, p. 24) acrescenta que o caracter | poderá ser utilizado em vez de OR e que, quando não digitamos nenhum operador entre as palavras, os sistemas de busca interpretam como se tivéssemos colocado OR.

A **relação de exclusão**, que utiliza o operador “NÃO” (NOT), permite relacionar dois descritores onde o primeiro deve estar presente e o segundo ausente. Exemplo: na busca por *animais NÃO herbívoros*, todos os itens contendo *animais* serão recuperados à exceção daqueles que contém a palavra *herbívoros*.

Diagrama de Venn do conjunto união *animais* NOT *herbívoros*



DENEGA (2000, p. 23) acrescenta que o caracter ! poderá ser utilizado em vez de NOT.

3.2.2 Operadores de Existência (ou Matemáticos)

Algumas ferramentas de busca permitem o uso da matemática, na qual os operadores booleanos são representados por símbolos de adição (+), ou subtração (-), como descritos abaixo:

- a) presença: especifica as palavras que devem aparecer no resultado. Utiliza o símbolo de “+” e funciona como o operador booleano AND, já descrito;
- b) ausência: especifica palavras que não devem aparecer no resultado. Utiliza símbolo de “-” e funciona como o operador booleano NOT, já descrito.

MARCHIORI (2001) alerta para o fato de não se poder deixar um espaço entre o descritor anterior e o sinal gráfico que vem antes do descritor. Como, por exemplo, em uma busca na qual a pessoa queira um roteiro de viagem para a Malásia, a busca fica assim: *+roteiro +viagem +malásia*.

3.2.3 Operadores de Proximidade

Esses operadores permitem que os resultados de busca sejam mais precisos, pois permitem que a relação entre os termos seja especificado e classificam-se, segundo DENEGA (2000, p. 25-27), em:

- a) próximo (NEAR);
- b) junto (ADJ);
- c) frases.

O **operador próximo (NEAR)**, é a intersecção de dois conjuntos da pesquisa, sendo parecido com o E lógico e “indica que o termo de pesquisa obrigatoriamente deve aparecer entre ‘x’ palavras de outro termo. Os sistemas de busca que permitem tal operador, definem tal número de palavras. No Altavista, por exemplo, os termos se separam por 10 outros termos, enquanto que no Open Text, estão separadas por 80

palavras” (MARCHIORI, 2001). Tomando-se como exemplo *estudantes NEAR brasileiros*, significa que as palavras *estudantes* e *brasileiros* apareçam relativamente juntas, como na frase o “candidato dos *estudantes* permitiu que somente entrassem na sala aqueles que fossem *brasileiros*”.

O **operador junto (ADJ)**, é a intersecção de dois conjuntos de pesquisa, semelhante ao E lógico, mas exige, diferentemente do NEAR, que entre as palavras não haja nenhuma outra. Tendo-se como exemplo *estudantes ADJ brasileiros*, as palavras *estudantes* e *brasileiros* deverão aparecer juntas, como na frase “o candidato dos *estudantes brasileiros...*”.

DENEGA (2000, p. 26) alerta para o fato de se poder utilizar aspas (“ ”), caso o sistema de busca não permita o ADJ (exemplo: “*estudantes brasileiros*”) e que também poderá ser utilizado o hífen (-) entre as palavras que se queira agrupar (exemplo: *estudantes-brasileiros*).

O **operador frases** é o que mais se identifica com a linguagem natural utilizada pelo homem. Trata-se da “intersecção das palavras pesquisadas que estão exatamente próximas uma das outras e de acordo com a ordem em que foram digitadas; é parecido com o E lógico, mas exige que a seqüência das palavras seja idêntica”, segundo o autor (DENEGA, 2000, p. 27), assim como na frase “*estudantes brasileiros conectados à Internet*”, que deverá aparecer idêntica no resultado da pesquisa. Ainda segundo o autor, pode-se utilizar hífen (-) entre as palavras da frase e obter o mesmo resultado (exemplo: *estudantes-brasileiros-conectados-à-Internet*).

Entretanto, MARCHIORI (2001) sinaliza para o fato de a frase limitar demasiadamente a pesquisa, pois somente serão obtidas as referências que contenham a frase exata que foi digitada entre aspas. Além disso, caso a mesma seja mal formulada e contenha plurais ou variações de palavras que não constam nos índices, o retorno será nulo ou inconsistente, por mais que a base de dados utilizada pela ferramenta contenha, efetivamente, a resposta para a pesquisa.

3.2.4 Truncagem

“Alguns sistemas de busca permitem a busca pela ‘raiz’ de uma palavra – o que permite a recuperação por distintas variações no final desta. Em alguns instrumentos isto é feito automaticamente, em outros, como no Altavista, por exemplo, deve-se utilizar o caracter asterisco (*), como na busca: auto* (permite recuperar ‘automático’, ‘automóvel’, ‘automóveis’)” (MARCHIORI, 2001). A truncagem é um recurso de pesquisa onde a busca é feita utilizando-se fragmentos da palavra, como nos tipos e exemplos abaixo arrolados por WILLIAMS⁹, *apud* LANCASTER (1993, p. 204):

- a) truncamento à direita: busca todas as palavras iniciadas por ‘condens’, por exemplo;
- b) truncamento à esquerda: busca todas as palavras que terminam com ‘micina’, por exemplo;
- c) truncamento com ‘infixos’ (especificam-se o começo e o fim, mas não meio de cada palavra), ou qualquer combinação possível desses recursos: busca todas as palavras que incluam a cadeia de caracteres ‘managet’, independentemente de onde a mesma apareça, por exemplo.

As técnicas de pesquisa são utilizadas para o refinamento das buscas, contribuindo para uma pesquisa precisa e com alta revocação.

⁹ WILLIAMS, M. E. Experiences of IIT Research Institute in operating a computerized retrieval system for searching a variety of data bases. **Information Storage and Retrieval**, n. 8, 1972, p. 57-75.

4 ÍNDICE DE REVOCAÇÃO E ÍNDICE DE PRECISÃO

LANCASTER (1993, p. 4), emprega o termo revocação para designar a capacidade de recuperar documentos úteis. O índice empregado, habitualmente, para expressar a extensão com que todos os itens úteis são encontrados é o coeficiente de revocação.

O termo precisão, conforme o autor, designa a capacidade de evitar documentos inúteis. O índice da relação entre os itens úteis e o total de itens recuperados é denominado de coeficiente de precisão.

Exemplificando com o autor, tem-se em uma busca em bases de dados 57 itens recuperados, na qual supõe-se que poderiam ser encontrados 11 itens úteis (caso a busca fosse perfeita) que responderiam à sua necessidade de informação, mas somente 6 itens úteis foram recuperados e 51 inúteis. Sendo assim, o coeficiente de precisão resultou em 10% (6/57) e o coeficiente de revocação em 54% (6/11).

Segundo o autor, um aumento no índice de revocação implica em diminuição do índice de precisão que, trata de uma característica lamentável, inerente à recuperação da informação (LANCASTER, 1993, p. 4).

Para KURAMOTO (2002), a falta de recuperação relevante da informação deve-se à palavra, que não possui *status* de descritor, para representar a unidade básica de acesso à informação, ou seja, “ao atribuir um descritor, isto é, um termo de indexação a um documento, o indexador declara que esse descritor possui um alto grau de relevância para o conteúdo do documento, ou seja, o descritor está fortemente associado a um conceito incorporado ao documento, e que é adequado à área temática do documento” (TINKER¹⁰, *apud* LANCASTER, 1993, p. 11).

Tanto KURAMOTO (2002) quanto BRÄSCHER (2002) colocam que a palavra possui problemas de ambigüidade (polissemia, sinonímia e homografia), inerentes à mesma, que a consolidam como fator de ruído no processo de

¹⁰ TINKER, J. F. Imprecision in indexing. *American Documentation*, n. 17, 1966. p. 93-102.

comunicação entre um sistema de informação e seu usuário.

5 SISTEMAS DE BUSCA NA INTERNET

A palavra Internet é uma abreviação de *International Network* (Rede Internacional de Informações) por isso, não se deve confundir a mesma, que é uma rede (física) de computadores, com a *Wide World Web* (WWW), ou simplesmente *web*, que é uma rede (virtual) de documentos interconectados e disponibilizados para serem acessados (LIMA, 2002).

A *web* conta hoje com centenas de milhares de *sites* espalhados ao redor do mundo que crescem de forma exponencial. Desde a sua concepção, a rede passou a ser uma preocupação constante dos especialistas no sentido de poderem ser maximizados os seus recursos informacionais e para que as informações verdadeiramente relevantes não ficassem “escondidas”.

Ao longo do tempo, conforme MARCHIORI (2001), “a busca por informações na *web* já passou por quatro fases (na ordem abaixo) que convivem, atualmente, de maneira harmoniosa e interdependente”.

- a) diretórios e ferramentas de busca;
- b) meta-ferramentas de busca;
- c) portais;
- d) serviços personalizados.

DENEGA (2000, p. 1) ressalta que, na língua inglesa, os sistemas de busca são denominados de *search engine* e que, no Brasil, existe o uso indistinto dos termos mecanismo de busca, sistema de busca, ferramenta de busca, ferramenta de procura, motor de procura e *search engine*. MOURA (2002) corrobora com DENEGA e acrescenta ainda outros termos à lista de denominações: diretório, catálogo, *site* de busca, programa de busca, serviço de busca e outros.

5.1 DIRETÓRIOS E FERRAMENTAS DE BUSCA

Os diretórios e as ferramentas foram os primeiros sistemas de busca básicos que culminaram nos atuais e tornaram a recuperação da informação ainda mais complexa e volátil.

5.1.2 Diretórios

Os diretórios surgiram como a primeira proposta para se tentar organizar e localizar informações na *web*, quando o conteúdo da mesma ainda era pequeno o bastante para que se pudesse coletar informações de forma não automática, ou seja, utilizando-se pessoas.

Para VALADARES CEDÓN (2001, p. 39-40) os diretórios são sistemas que organizam *sites* em uma base de dados pequena, mediante a utilização de cabeçalhos de assunto para que as palavras possam ser armazenadas e depois recuperadas pelos usuários com mais relevância. Tratam-se de sistemas genéricos, ou seja, que atendem a um público variado, pois contêm informações relacionadas com educação, esportes, entretenimento, viagens, compras, informática e outros.

Como descreve a autora, os *sites* coletados são selecionados, na maioria das vezes, pelos editores dos diretórios (que tomam conhecimento desses novos recursos por meio de sugestões que a comunidade usuária envia-lhes) ou pelos robôs (hospedeiros das ferramentas de busca) que, atualmente, em parceria com os diretórios, coletam os novos endereços na *web*. Os critérios de seleção utilizados pelos editores não são divulgados mas, segundo a autora, somente os melhores recursos são utilizados para a inclusão. Todavia, esse processo manual pode demorar até um mês.

Atualmente, um dos diretórios mais conhecidos é o Yahoo. Trata-se de um *site* em inglês, muito bem organizado, com catálogos de *sites* em diversas línguas, inclusive português e que, quando não encontra a palavra pesquisada, busca-a automaticamente no Alta Vista (uma ferramenta de busca) (BOANERGES, 2002).

VALADARES CEDÓN (2001, p. 40-41) ressalta que, apesar de todos os diretórios seguirem os mesmos princípios genéricos, acima descritos, podem ainda classificarem-se em:

- a) **diretórios avaliativos ou acadêmicos:** utilizam-se de estudantes de mestrado ou mestres na área de informação, ou ainda especialistas em assuntos específicos para que, nas análises de *sites* a serem incorporados no diretório, hajam descrições criteriosas e detalhadas dos recursos, assim como críticas e avaliações construtivas dos mesmos. Geralmente estão associados a bibliotecas ou instituições de ensino;
- b) **diretórios temáticos ou especializados:** cobrem áreas específicas, tais como: diretórios especializados em imagens, jornais, *softwares*, listas de discussão, saúde, ciências, legislação, informática, ferramentas de busca em países específicos, listas de diretórios temáticos e outras;
- c) **diretórios de sistemas de busca:** foram criados com a função de listar sistemas de busca, para facilitar a identificação dos mesmos, devido ao fato de freqüentemente surgirem na *web* novos sistemas de busca. Um diretório importantíssimo desse tipo é o *Search Engine Watch*.

Na atualidade, os diretórios trabalham em parceria com as ferramentas de busca, para que, eventualmente, caso o usuário não encontre o que deseja, o diretório acione a ferramenta de busca e, instantaneamente, recupere na *web* os *sites* que contenham a pesquisa realizada pelo cliente. Para possibilitar o entendimento de como essas ferramentas atuam, se faz necessário discorrer sobre as mesmas.

5.1.3 Ferramentas de Busca

Com o aumento incessante da *web*, tornou-se complexa a indexação manual da informação contida na mesma. Assim sendo, surgiram as ferramentas de busca que, diferentemente dos diretórios, não organizam os *sites* que coletam em suas bases de dados.

Conforme VALADARES CEDÓN (2001, p. 41), as ferramentas de busca compõem-se de quatro componentes, que as auxiliam na recuperação da informação na *web*:

- a) o robô;
- b) o indexador;
- c) a ferramenta propriamente dita e;
- d) a interface.

Os **robôs** (*robots*), que também são denominados de aranhas (*spiders*), agentes viajantes (*wanderes*), rastejadores (*crawlers*), vermes (*worms*), conforme a autora, tratam-se de *softwares* que o computador hospedeiro da ferramenta de busca lança regularmente na *web*, para monitorarem a mesma, tendo como intuito capturar mais informações para serem armazenadas em suas bases de dados. Esse procedimento é executado estrategicamente, como bem esboça a autora:

Existem várias estratégias que os robôs podem utilizar para se locomoverem de um documento a outro, utilizando-se dos *links* existentes nas páginas da *web*. Geralmente, eles iniciam a busca a partir de *sites* conhecidos, especialmente daqueles que possuem muitos *links*, recuperam a sua *homepage* e, sistematicamente, seguem os *links* encontrados nesta página inicial. Usam algoritmos próprios para determinar que *links* devem seguir. Por exemplo, alguns recuperam os documentos da hierarquia superior de um grande número de servidores (abordagem *breadth-first*), enquanto outros capturam todos os documentos em *links* de um mesmo servidor (abordagem *depth-first*) (VALADARES CEDÓN, 2001, p. 41).

Depois de capturados pelos robôs, os documentos são encaminhados aos **indexadores** que extraem informações julgadas importantes, tais como páginas *Hipertext Markup Language* (HTML), títulos, resumos e palavras, armazenando-as em uma base de dados. Os critérios para essa indexação, encontram-se aqui descritos pela

autora:

As ferramentas de busca criam índices ... utilizados para dinamizar a busca de informações na sua base de dados ... esses índices contêm termos que podem ser utilizados na busca de informações, pelo usuário e o *Uniforme Resource Locator* (URL) das páginas que os contêm ... podendo ainda serem armazenados dados sobre a posição das palavras na página e sobre os *tags* HTML associados ao texto. Se um termo não estiver incluído no índice, ele não será encontrado, portanto os critérios utilizados para a indexação influenciam os resultados da busca ... algumas ferramentas indexam cada palavra do texto visível das páginas ... outras, entretanto, indexam apenas o URL, as palavras que ocorrem com frequência ou as palavras e frases mais importantes contidas no título ou nos cabeçalhos e nas primeiras linhas ... outras ferramentas indexam outros termos, que não fazem parte do texto visível, mas que contêm informações importante e úteis, cujo exemplo são os textos incluídos nos *metatags* para classificação, descrição e palavras-chave e texto ALT do *tag Image*, ou seja, texto associado com imagens. Os *metatags* de classificação fornecem uma palavra-chave que define o conteúdo da página. Os de descrição retornam à descrição da página feita pelo seu autor no lugar do resumo que o robô criaria automaticamente. Os de palavras-chave fornecem as palavras-chave designadas pelo autor para descrever seu conteúdo ou assunto. Por exemplo, no *metatag* `<META name="keyword" content="Brasil, informação para negócios">`, as palavras Brasil e informação para negócios podem não fazer parte do texto visível da página, entretanto foram indicadas pelo seu autor como indicadores do assunto sobre os quais a página versa. Algumas ferramentas não incluem no seu índice algumas palavras do texto, chamadas de palavras proibidas (*stop words*) ... e outras as incluem, mas as ignoram ao executar a busca, para torná-la mais rápida ... para o usuário torna-se um problema, pois as ferramentas não fornecem explicações sobre quais são essas palavras proibidas utilizadas (VALADARES CEDÓN, 2001, p. 42-43).

Após indexadas, as informações são disponibilizadas mediante uma **interface web**, na qual os usuários realizam suas consultas, utilizando-se pesquisas simples ou as técnicas de pesquisa, já citadas em capítulo anterior, que depois de transmitidas às ferramentas de busca, recuperam as informações pertinentes ao pedido. Além disso, segundo a autora, a ferramenta também ordena os resultados, de modo que os mais relevantes apareçam em primeiro lugar na lista de resultados. Esses resultados são ilustrados como uma lista de descrições de *sites* e seus *links* (VALADARES CEDÓN, 2001, p. 41).

Depois da concepção das ferramentas, surgiu o conceito de meta-ferramentas de busca.

5.2 META-FERRAMENTAS DE BUSCA

Com o intuito de facilitar a obtenção de melhores resultados em pesquisas na *web*, foram criadas as meta-ferramentas que tem por princípio executar, em uma única busca, uma pesquisa em vários diretórios e ferramentas concomitantemente. Assim como na busca, os resultados da mesma também são simultâneos, ou seja, apresentam-se ao mesmo tempo e em uma única lista, contendo todos os diretórios e ferramentas com suas respectivas respostas (COOKE 1999, p. 42).

VALADARES CEDÓN (2001, p. 47) ressalta que as meta-ferramentas não possuem nenhuma base de dados e utilizam-se exclusivamente dos dados das ferramentas e diretórios que monitoram para a recuperação da informação. A autora ainda alerta para o fato de existirem “pseudometas-ferramentas”, ou seja, falsas meta-ferramentas, que apenas fornecem uma interface onde várias ferramentas encontram-se listadas, sem que haja um mecanismo de busca integrado, ou seja, existe uma caixa de pesquisa para cada ferramenta ou diretório, mas as buscas são efetuadas isoladamente e, conseqüentemente, as respostas também.

Segundo a autora, o resultado dessas pesquisas, em algumas meta-ferramentas, são parte de um subconjunto de respostas das quais somente as primeiras e, supostamente, as mais relevantes são recuperadas. Esse tipo de busca possui como inconveniente o fator tempo, pois existe uma demora adicional para a compilação dos dados, além do fato de os buscadores possuírem características intrínsecas que os tornam uns mais lentos que os outros e, o tempo de resposta somente findará quando o buscador mais lento der a sua resposta (VALADARES CEDÓN, 2001, p. 49).

5.3 PORTAIS

O'BRIEN (2001) afirma que, atualmente, o perfil dos sistemas de informação caracteriza-se pelo rápido crescimento da Internet (rede mundial de computadores), juntamente com a Intranet (rede quando se refere ao meio ambiente de um local), Extranet (quando se conectam redes distintas em longas distâncias) e outras redes.

Dentre estas modalidades de rede, a Internet vem se destacando no que se refere ao acesso a sistemas de armazenamento, tratamento e disseminação de informações.

BLATTMANN *et al.* (1999), citam como vantajoso o uso da Internet, em relação aos métodos tradicionais, como instrumento de busca de informação, pois possui características especiais, como as abaixo arroladas:

- a) facilidade e rapidez no acesso às informações;
- b) visão de outras realidades culturais;
- c) aumento da comunicação interpessoal;
- d) interação e integração com outros alunos e professores, enriquecendo seus conhecimentos de forma individual e grupal;
- e) ganhos no trabalho cooperativo, com a promoção de um maior intercâmbio;
- f) melhoria na comunicação escrita;
- g) produção de materiais, como a criação e construção de documentos hipertexto;
- h) aprendizado em diferentes conteúdos, visando um trabalho interdisciplinar.

Uma autora que retrata com clareza esta mudança de perfil, em extensa revisão de literatura, é Cláudia Augusto Dias, em seu artigo *Portal corporativo: conceitos e características*, publicado na revista *Ciência da Informação*, v. 30, n. 1, p. 50-60, jan./abr. 2001. O referido artigo será utilizado neste capítulo, assim como os autores citados por DIAS (2001), aqui abordados, terão suas obras citadas em nota de rodapé.

Segundo a autora, o que hoje denomina-se de portal, há três ou quatro anos tratava-se de ferramenta de busca, cujo intuito era facilitar o acesso às informações pela *web*. Conforme REYNOLDS & KOULOPOULOS¹¹, os portais *web* iniciaram-se pela utilização de operadores lógicos (ou booleanos) para restringir ou ampliar o universo da pesquisa; passaram pela navegação por categorias na qual vários *sites* de busca incluem categorias, ou seja, filtram *sites* e documentos em grupos pré-configurados de acordo com o seu conteúdo – esportes, meteorologia, turismo, finanças, cultura e outro; integraram-se, às funções anteriores, as comunidades virtuais e suas listas de discussão, *chats* em tempo real, possibilidades de personalização dos *sites* de busca (Meu Yahoo!, Meu Excite!) e, finalmente, tiveram suas funções expandidas para outras áreas dos mundos informacionais e comerciais:

Esses portais, segundo os autores, podem ser classificados de duas formas:

- a) em relação ao contexto de sua utilização: público ou corporativo;
- b) em relação às suas funções: suporte à decisão e/ou processamento corporativo.

O **portal público**, também denominado de portal Internet, portal *web* ou portal de consumidores, tem como função, mediante uma única interface, colocar à disposição dos usuários todos os servidores que compõem a rede. Seu intuito é atrair, para o seu *site*, todas as pessoas que navegam pela *web*, pois quanto maior for o número de visitantes no mesmo, maior será a probabilidade do estabelecimento de comunidades virtuais que, potencialmente, comprarão os produtos anunciados naquele *site*.

O **portal corporativo**, segundo REYNOLDS & KOULOPOULOS¹², “trata-se de uma revolução das Intranets, incorporado a ferramentas que possibilitam a identificação, captura, armazenamento, recuperação e distribuição de grandes

¹¹ REYNOLDS, H.; KOULOPOULOS, T. **Enterprise knowledge has a face**. Disponível em: <<http://www.intelligententerprise.com/993003/feat1.shtml>>.

¹² REYNOLDS & KOULOPOULOS, *op cit*.

quantidades de informações de múltiplas fontes, internas e externas, para os indivíduos de uma instituição”.

Os **portais com ênfase em suporte à decisão**, para MURRAY¹³, ECKERSON¹⁴ e WHITE¹⁵, auxiliam executivos, gerentes e analistas de negócios a acessarem as informações corporativas para a tomada de decisão em negócios. Tais portais subdividem-se em:

- a) portal de informações ou de conteúdo: é aquele que organiza grandes acervos de conteúdo, a partir de temas ou assuntos neles contidos e conecta pessoas às informações (MURRAY¹⁶). Entretanto, pode ser denominado portal Intranet, pois o mesmo inclui *links* para informações e *sites web* dentro e fora da organização (WHITE¹⁷);
- b) portal de negócios: equivale aos portais comerciais da *web* (Yahoo, Excite e outros), pois torna possível aos usuários das organizações informações (estruturadas e não-estruturadas) para a tomada de decisões em negócios, tais como relatórios, pesquisas, negócios e outros (ECKERSON¹⁸);
- c) portal de suporte à decisão: utiliza informações estruturadas, capturadas por ferramentas inteligentes, tanto de sistemas internos quanto externos da instituição, para servir de apoio à tomada de decisão.

Os **portais com ênfase em processamento corporativo**, caracterizam-se por manipular informações tanto dos sistemas de informação da instituição, como de

¹³ MURRAY, G. **The portal is the desktop**. Disponível em: <http://www.groupcomputing.com/Back_Issues/1999/MayJune1999/mayjune1999.html>.

¹⁴ ECKERSON, W. **Business portals: drivers, definitions, and rules**. Disponível em: <<http://www.viador.com/pdfs/SeyboldWhitePaper.pdf>>.

¹⁵ WHITE, C. **The enterprise information portal requirements**. Disponível em: <<http://www.decisionprocessing.com/papers/eip2.doc>>.

¹⁶ MURRAY, *op cit*.

¹⁷ WHITE, C. **Decision threshold**. Disponível em: <<http://www.intelligententerprise.com/991611/feat1.shtml>>

¹⁸ ECKERSON, *op. cit*.

grupos ou indivíduos de fora da mesma e dividem-se em:

- a) portal cooperativo ou para processamento cooperativo: utiliza ferramentas de trabalho em grupo e de fluxo de tarefas/documentos, para prover o acesso às informações geradas por indivíduos ou grupos. As informações encontram-se personalizadas em textos, memorandos e outros;
- b) portal de especialistas: relaciona e une pessoas com base em suas experiências, ou seja, serve de meio de comunicação para vários especialistas de diferentes áreas.

Os **portais de suporte à decisão e processamento cooperativo** são os mais abrangentes, pois “aliam as funções de suporte à decisão e processamento cooperativo, conectam os usuários não só a todas as informações, mas também a todas as pessoas necessárias para a realização dos negócios da empresa” e dividem-se em:

- a) portal do conhecimento: “é um ponto de convergência entre os portais de informações, cooperativos e especialistas, que é capaz de implementar tudo o que os outros portais implementam, além de fornecer conteúdo personalizado a cada usuário” (MURRAY¹⁹);
- b) portal de informações empresarias: “utiliza metadados e a *Extensive Markup Language* (XML) para integrar dados não-estruturados, mantidos e arquivos textuais, relatórios e outros, aos dados estruturados do *data warehouse*, fornecendo acesso às informações institucionais a partir de uma interface individualizada, disponível via Intranet” (DIAS, 2001).

Nesse mesmo contexto, tanto KURAMOTO (2002) quanto BRÄSCHER (2002) colocam que, apesar dos benefícios, a Internet continua enfrentando os mesmos problemas de recuperação da informação como qualquer outro tipo de sistema de informação.

Como já visto anteriormente, os portais são importantes sistemas, pois atraem uma enorme gama de usuários, utilizando-se da qualidade e diversidade de

¹⁹ MURRAY, *op cit.*

informações, com múltiplos recursos, contidos em seus *mega-sites*. Os portais são a evolução e união de diretórios, ferramentas e meta-ferramentas que, acoplados ao mesmo, transitam pelo universo da *web* monitorando-a.

De acordo com SABBATINI (2002), os portais surgiram como sistemas de busca mais atrativos para os usuários, como sonhos de consumo, pois contêm e organizam, em uma única página, todo o “caótico mundo da informação” da *web*.

Assim como os diretórios avaliativos ou acadêmicos citados por VALADARES CEDÓN (2001), os portais, segundo COOKE (1999, p. 34), utilizam-se de profissionais da área de informação, assim como de especialistas para analisar, indexar e incorporar informações qualitativas aos mesmos, mediante descrições criteriosas e detalhadas dos recursos, para que as mesmas possam ser bem recuperadas. Geralmente estão associados a bibliotecas ou instituições de ensino.

Segundo MARCHIORI (2001), esses sistemas voltam-se para a “compilação de conteúdos e *links* de áreas temáticas determinadas”, tais como Ciências Sociais Aplicadas, Ciências da Terra, Ciências Humanas, Ciências Sociais e outras. Dentro deste contexto de portal temático, verifica-se como exemplo o Portal da Informação da UFPR, inaugurado em sete de agosto do ano dois mil, pelo Sistema de Bibliotecas da referenciada instituição de ensino, que visa suprir as necessidades informacionais contidas em todas as áreas dos cursos que compõem a universidade.

5.4 SERVIÇOS PERSONALIZADOS

Como o próprio nome já diz, segundo MARCHIORI (2001,.), “tratam-se de serviços que são executados de acordo com o usuário, ou seja, o mesmo escolhe o tipo de informação que quer acessar e adapta o visual de suas páginas”. Exemplos são os diversos portais *web* que oferecem o modo de personalizar a página, ou mesmo os provedores de acesso à Internet que, durante o processo de instalação do *software*,

pede ao cliente que o mesmo escolha qual o nome que ele quer que apareça na tela, ou selecione um desenho que mais se adequa à sua personalidade, entre outros.

6 PROCEDIMENTOS METODOLÓGICOS DO ESTUDO DE CASO

Para complementar a pesquisa com a retratação de uma realidade realizou-se um estudo de caso.

O estudo de caso é uma técnica de pesquisa cujo objetivo é o estudo de uma unidade que se analisa profunda e intensamente, dentro de um contexto real, procurando reunir o maior número de informações detalhadas, com o objetivo de apreender a totalidade de uma situação e descrevê-la (MARTINS & LINTZ, 2000, p. 36).

Inicialmente, determinou-se o Portal da Informação da UFPR, como exemplo de sistema virtual de informação, para ser o objeto do estudo de caso. A escolha deveu-se ao fato de, como usuária do sistema e estudante do curso de Gestão da Informação, da instituição de ensino referenciada, poder contribuir com sugestões para a melhoria da busca e recuperação das de informações disponíveis no Portal.

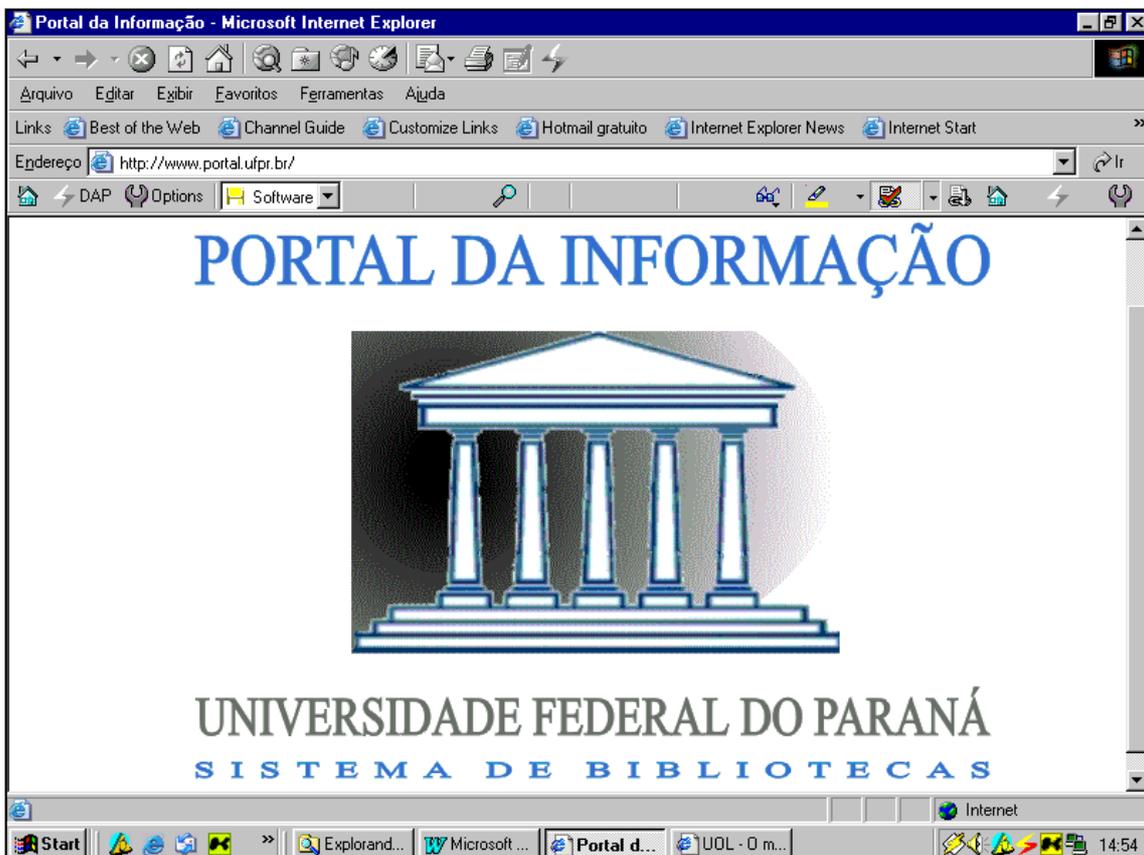
Para a avaliação do Portal, foram adotados critérios estabelecidos por BARBOZA *et al.* (2000), tais como abrangência e propósito; conteúdo; planejamento visual/gráfico (*webdesign*) e funcionalidade (interface e quesitos de navegabilidade). Dentre esses critérios de avaliação, foi dada especial atenção ao de funcionalidade, uma vez que mesmo trata essencialmente da forma de busca e recuperação da informação em *sites*.

Desta análise resultou a apresentação e interpretação dos resultados, as considerações finais e as sugestões.

7 APRESENTAÇÃO E INTERPRETAÇÃO DOS RESULTADOS

Nesta seção, apresenta-se o relato dos resultados obtidos, mediante a análise do Portal da Informação da UFPR (<http://www.portal.ufpr.br>) (figura 1), embasados no roteiro de critérios de avaliação, anteriormente mencionados, estabelecidos por BARBOZA *et al.* (2000).

FIGURA 1 – PORTAL DA INFORMAÇÃO DO SISTEMA DE BIBLIOTECAS DA UFPR



FONTE: UNIVERSIDADE (2002b)

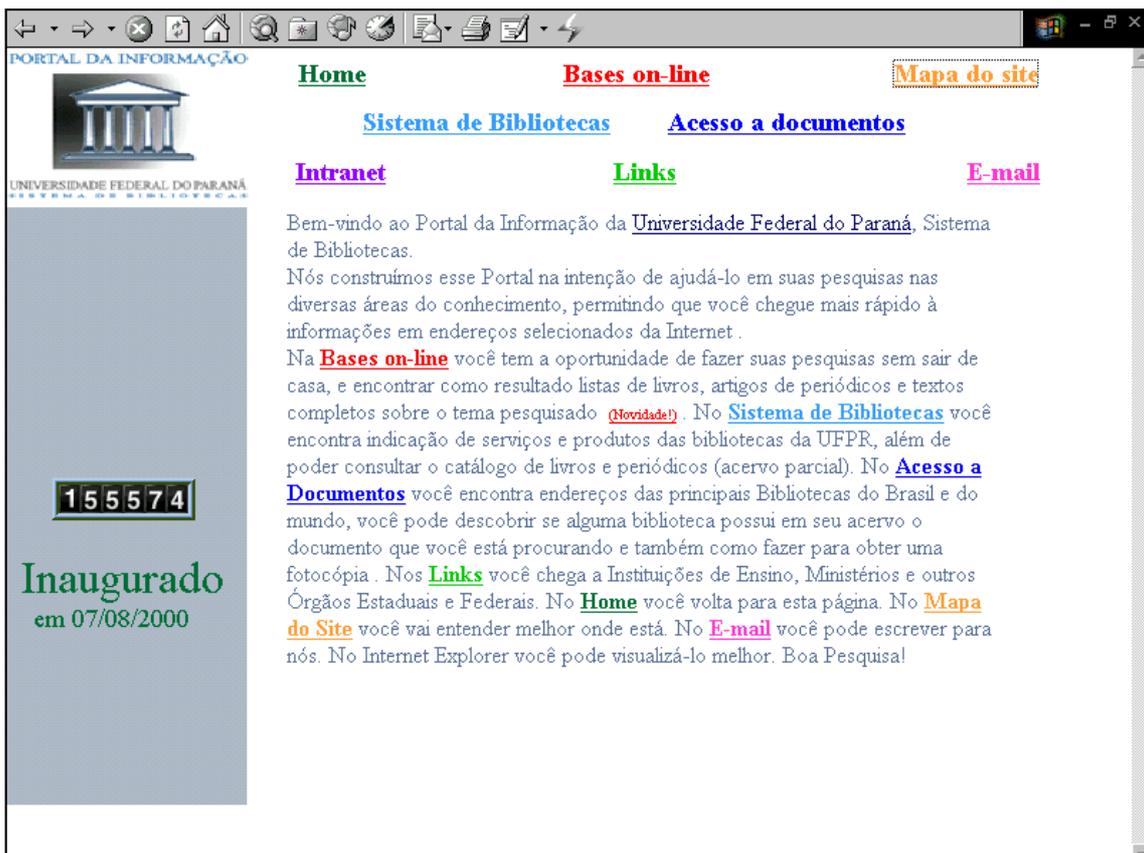
7.1 ABRANGÊNCIA E PROPÓSITO

A abrangência e o propósito são os critérios pelos quais pode-se verificar “a

amplitude ou limitação da fonte de informação, sua profundidade e nível de detalhe, alguma restrição com relação à informação durante certos períodos; se a página atingiu seus objetivos ao expor tópicos escolhidos, não deixando de mencionar informações significativas” (BARBOZA *et al.*, 2000, p. 121).

O propósito do portal vem claramente enunciado em sua *homepage* (figura 2): “Nós construímos esse Portal na intenção de ajudá-lo em suas pesquisas nas diversas áreas do conhecimento, permitindo que você chegue mais rápido à informações em endereços selecionados da Internet” (UNIVERSIDADE, 2002b).

FIGURA 2 – *HOMEPAGE* DO PORTAL DA INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

Quanto à abrangência, a *homepage* (figura 2) traz, em destaques diversificadamente coloridos e com descrições simplificadas, os serviços e produtos de informação que oferece (UNIVERSIDADE, 2002b):

- a) “na **Bases on-line** você tem a oportunidade de fazer as suas pesquisas sem sair de casa, e encontrar como resultado listas de livros, artigos de periódicos e textos completos sobre o tema pesquisado”;
- b) “no **Sistema de Bibliotecas** você encontra indicação de serviços e produtos das bibliotecas da UFPR, além de poder consultar o catálogo de livros e periódicos (acervo parcial)”;
- c) “no **Acesso a Documentos** você encontra os endereços das principais bibliotecas do Brasil, e do mundo, você pode descobrir se alguma biblioteca possui em seu acervo o documento que você está procurando e também como fazer para obter uma fotocópia”;
- d) “nos **Links** você chega a Instituições de Ensino, Ministérios e outros Órgãos Estaduais e Federais”;
- e) “no **Home** você volta para esta página”;
- f) “no **Mapa do Site** você vai entender melhor onde está”;
- g) “no **E-mail** você pode escrever para nós”.

Com relação ao *link* **Intranet** e ao *hyperlink* **Universidade Federal do Paraná** porém, não há indicação alguma de qual é a função dos mesmos na página. Entretanto, pressupõe-se que remetam a um local condizente com o termo utilizado, ou seja, **Intranet** para os serviços do portal que encontram-se disponíveis somente na rede interna da instituição e **Universidade Federal do Paraná** para a página principal da instituição de ensino. Contudo, a página deveria conter a descrição dos mesmos para facilitar a navegabilidade do usuário, que talvez não esteja familiarizado com os termos utilizados.

7.2 CONTEÚDO

O conteúdo é o critério pelo qual pode-se “avaliar a apresentação da

informação com relação à sua correção gramatical e ortográfica, aos erros de digitação, à qualidade de sua escrita e à clareza da comunicação de seu conteúdo facilitando a leitura; se está explícita a organização que produziu a página; se quem a produziu tem formação e/ou experiência na área; se consta no *site* a data de atualização, se a sua informação é exclusiva em termos de disponibilidade em outros formatos ou em outros *sites*; se existem *links* para outras fontes de informação” (BARBOZA *et al.*, 2000, p. 121).

A escrita utilizada para descrever o conteúdo da *homepage* (figura 2) é extremamente simplória, tornando-se muito convidativa à comunidade mas, ao mesmo tempo, incompatível com o perfil da maior parte de seus usuários, haja vista que tratam-se de acadêmicos. Porém, o hipertexto foi bem elaborado e seus *hyperlinks* descritos de forma condizente à informação para a qual remetem.

O nome organização (UFPR), assim como a equipe de profissionais da informação que produziram a página (Sistema de Bibliotecas) encontram-se explícitos ao ler-se “Bem-vindo ao Portal da Informação da Universidade Federal do Paraná, Sistema de Bibliotecas” (figura 2). Entretanto, a página não referencia quantos, quais e de que áreas são os profissionais que fizeram parte da criação do portal. Essencialmente, o portal deve ser um trabalho elaborado por diversos profissionais, não somente da área de informação, como também das outras áreas (humanas, biológicas, tecnológicas), pois os especialistas no tratamento da informação precisam trabalhar de forma interdisciplinar com especialistas em outras áreas do conhecimento, para que possam gerar produtos informacionais de alta qualidade, como é o modelo ideal de portal.

Com relação à data de atualização do *site*, verificou-se a inexistência da mesma, acarretando um problema de falta confiabilidade e obsolescência nas informações prestadas pelo portal.

O conteúdo das páginas do *site* apresenta parcerias entre a UFPR e outras instituições tanto nacionais como internacionais, como se pode verificar nos *Links* e

no **Acesso a Documentos** (figura 2).

O único meio de ajuda do portal é o **E-mail**, pelo qual pode-se escrever para a equipe responsável pelo suporte do *site*.

7.3 PLANEJAMENTO VISUAL/GRÁFICO (*WEBDESIGN*)

O planejamento visual/gráfico (*webdesign*) é o critério pelo qual pode-se “verificar se as letras e os tipos estão claros, simples e facilmente legíveis; seu tamanho ou corpo está relacionado com a superfície que ocupa; sua disposição dá o destaque de acordo com o interesse dos títulos, e em contraste e harmonia com os espaços destinados a brancos e imagens; se os ícones representam claramente o que pretendem e se algum conteúdo de multimídia está incorporado de forma apropriada” (BARBOZA *et al.*, 2000, p. 121).

Na página do portal verifica-se a implementação de letras de fonte e tamanho visíveis que, porém, tornam-se pouco legíveis nos tons mais claros (figura 2) utilizados nos *links*: **Sistema de Bibliotecas** (azul-claro); **Links** (verde-claro); **E-mail** (rosa-claro) e, extremamente chamativas nos *links*: **Intranet** (violeta); **Bases on-line** (vermelho vivo); **Mapa do Site** (amarelo escuro). Já o *link* **Acesso a documentos** (azul) apresenta-se em um bom tom de cor por sua sobriedade e legibilidade. Outro problema é a cor do hipertexto, cujo tom encontra-se num cinza extremamente claro, tornando-se pouco legível, apesar de o tamanho da fonte ser bom. Caso o hipertexto fosse escrito em preto e os *hyperlinks* em negrito e sublinhados, tornariam-se mais legíveis e condizentes com o perfil dos usuários que acessam o portal, por tratarem-se de jovens e adultos. Além disso, a utilização das cores verde e vermelho não faz diferença para os daltônicos, pois os mesmos não conseguem distinguir uma cor da outra por problemas oftalmológicos e, no caso do portal, verifica-se claramente que a diversidade de cores foi utilizada para que cada uma simboliza-se um *link* de acesso a

uma determinada informação, destacando-se uns dos outros.

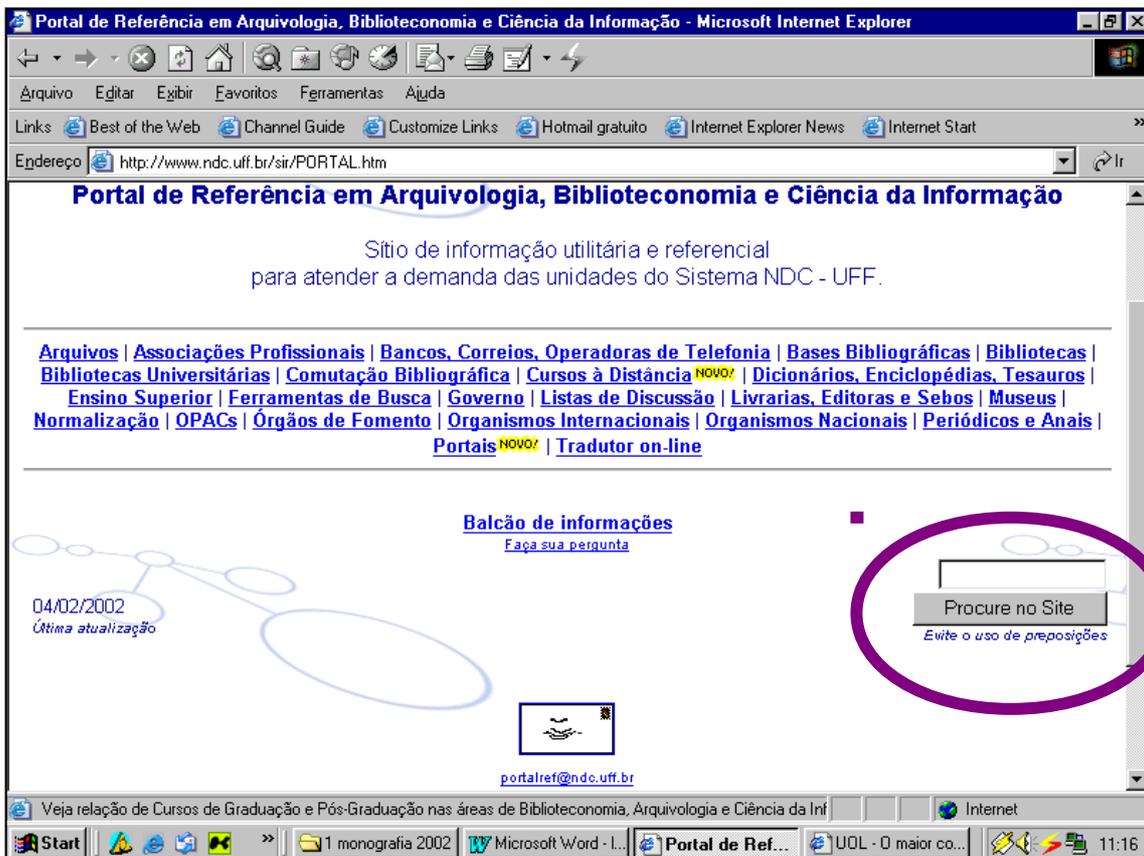
Quanto aos conteúdos multimídia, a página não incorpora nenhum.

7.4 FUNCIONALIDADE (INTERFACE E NAVEGABILIDADE)

A funcionalidade (interface e quesitos de navegabilidade) é o critério pelo qual pode-se verificar “se os aspectos relativos à navegação pelo *site* são bem objetivos, tais como o *design* do menu e a legibilidade da tela facilitando a navegação; se a ferramenta de busca para a recuperação da informação contém interfaces amigáveis e indexação da fonte como um todo; se as características de seus formatos são sempre funcionais, têm resultados e agregam valor ao *site*, se suas página carregam rapidamente ou se existe um espelho de *site* alternativo” (BARBOZA *et al.*, 2000, p. 122).

A navegação pelos diversos *links* do portal (figura 2), possui os problemas referentes ao *design* do menu, já mencionados, além de não possuir um elemento básico em sua *homepage*, a caixa de busca, ou seja, um retângulo no qual digita-se um termo que se queira buscar como o “Procure no *Site*”, da *homepage* do Portal de Referência em Arquivologia, Biblioteconomia e Ciência da Informação da Universidade Federal Fluminense (UFF) (figura 3).

FIGURA 3 – *HOMEPAGE* DO PORTAL DE REFÊNCIA EM ARQUIVOLOGIA, BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO DA UNIVERSIDADE FEDERAL FLUMINENSE (UFF)

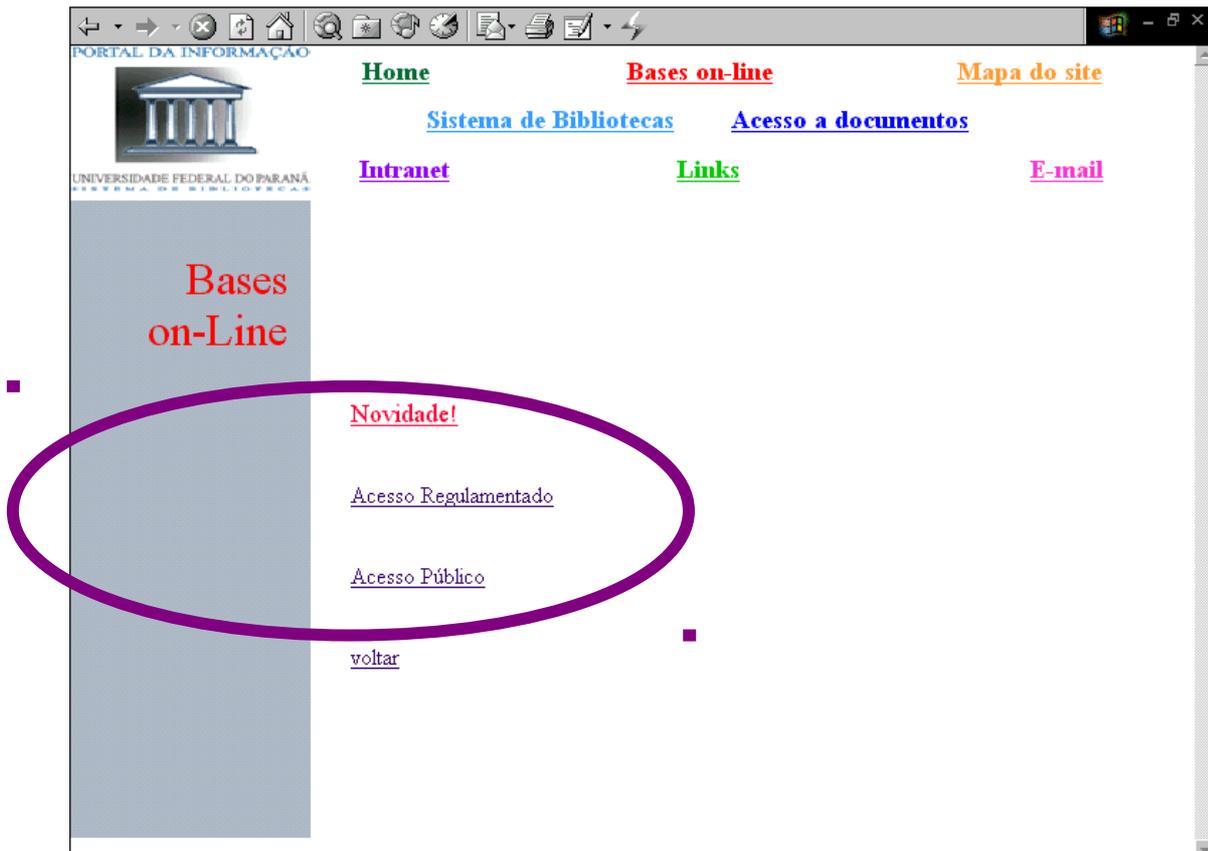


FONTE: UNIVERSIDADE (2002a)

Como não há uma caixa de busca no portal (figura 2), o usuário obriga-se a clicar em vários links do menu até chegar ao local desejado, como no exemplo de procura por uma base *on-line*, em que o mesmo terá que executar todos os passos abaixo, contando sempre com um *link* de retorno (**voltar**) à página anterior, para chegar ao que pretende:

- a) passo 1 - clicar em **Bases on-line** (figura 2), para obter como resultado os *links* abaixo circutados (figura 4):

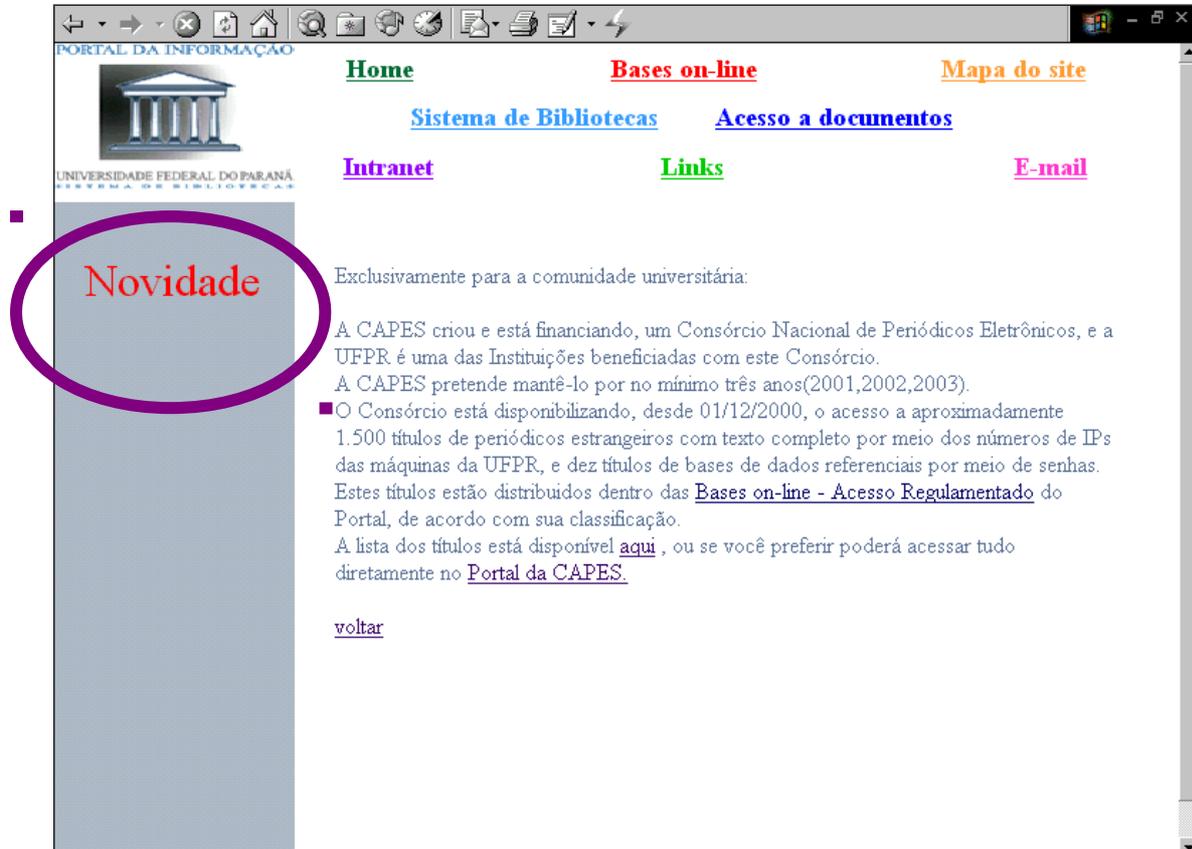
FIGURA 4 – INTEFACE DA PÁGINA BASES *ON-LINE* DO PORTAL DA INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

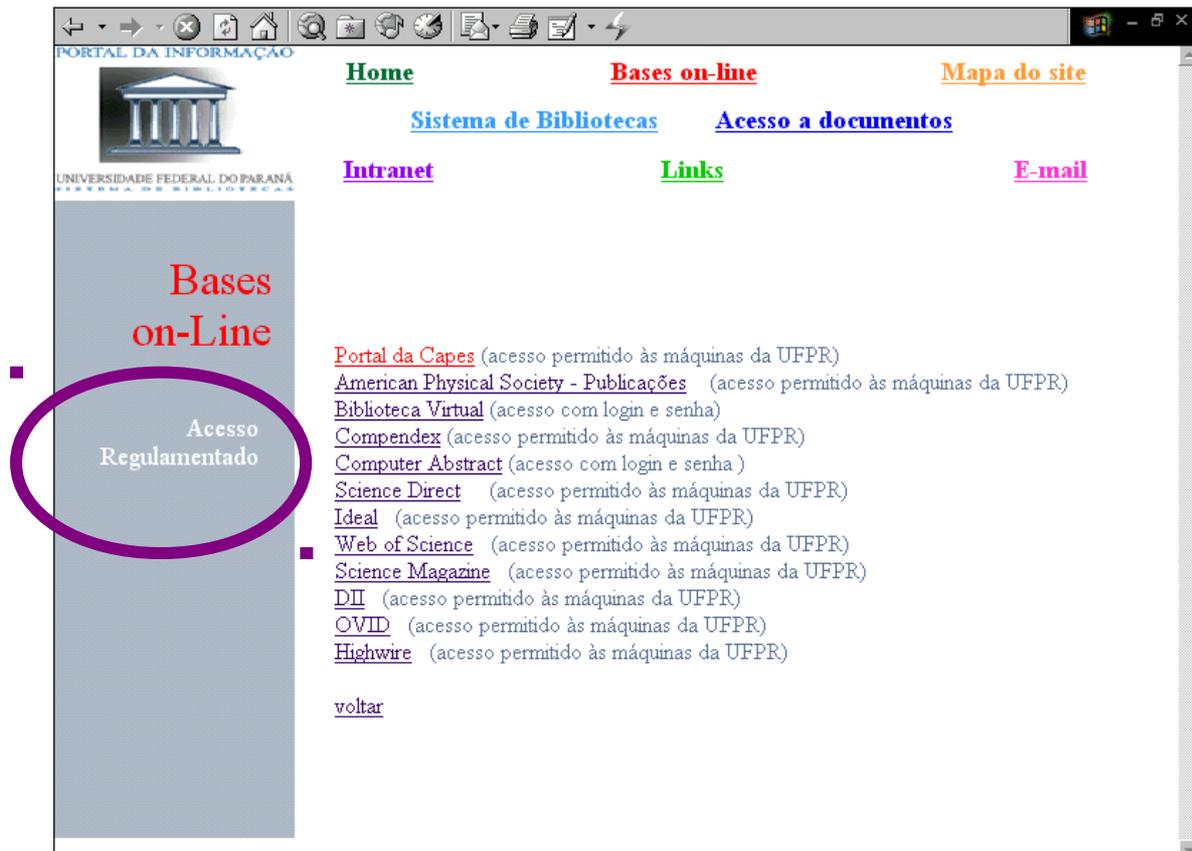
- b) passo 2 - clicar em cada um desses *links* para verificar se algum deles responde à sua busca e descobrir que aparecem outros *links* (figuras 5, 6 e 7):

FIGURA 5 – INTEFACE DA PÁGINA NOVIDADE DO PORTAL DA INFORMAÇÃO DA UFPR



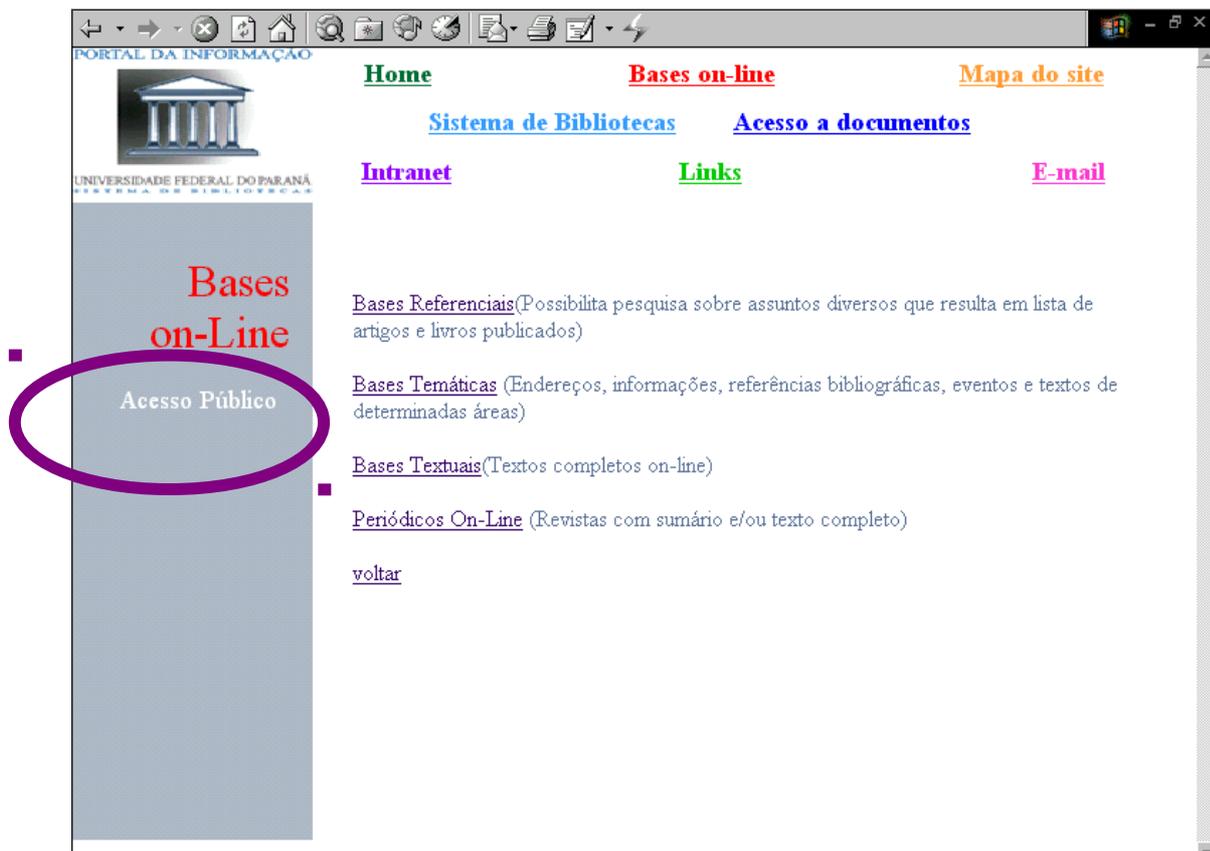
FONTE: UNIVERSIDADE (2002b)

FIGURA 6 – INTEFACE DA PÁGINA BASES *ON-LINE* DE ACESSO REGULAMENTADO DO PORTAL DA INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

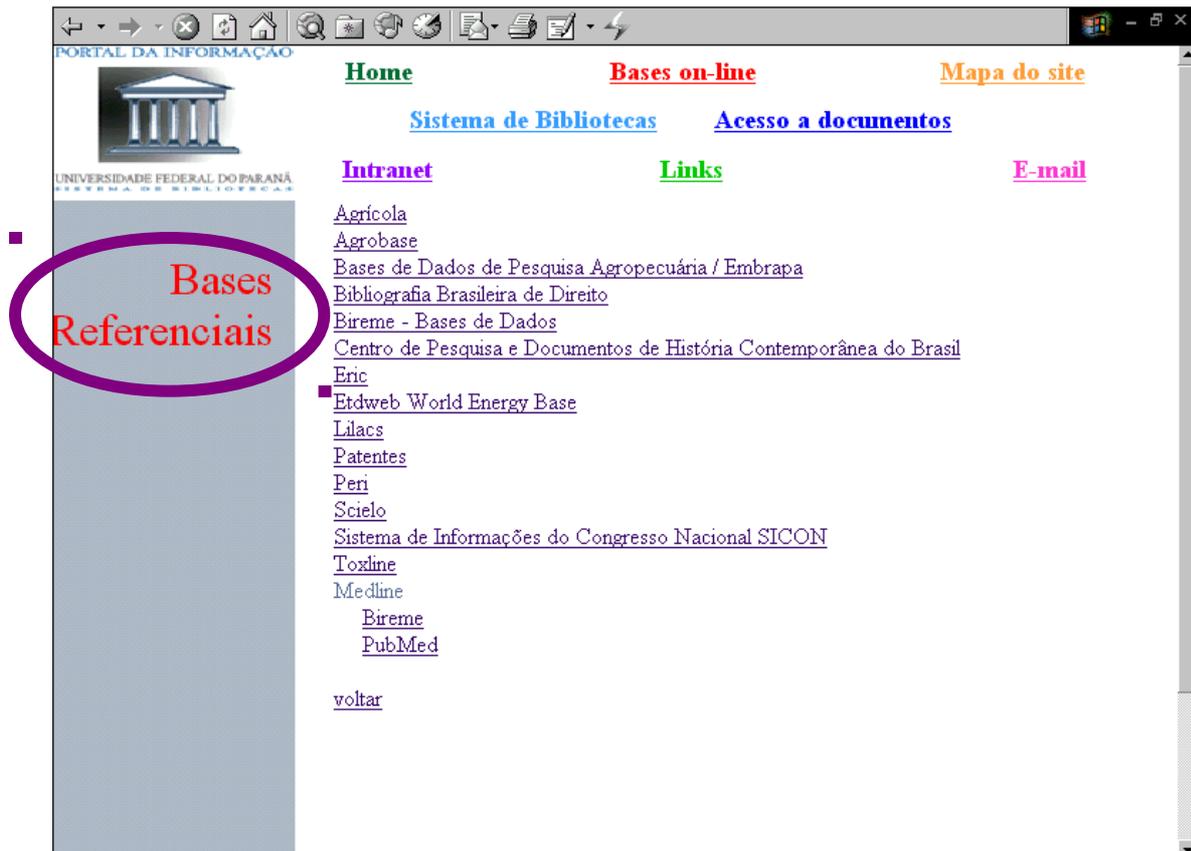
FIGURA 7 – INTEFACE DA PÁGINA BASES *ON-LINE* DE ACESSO PÚBLICO DO PORTAL DA INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

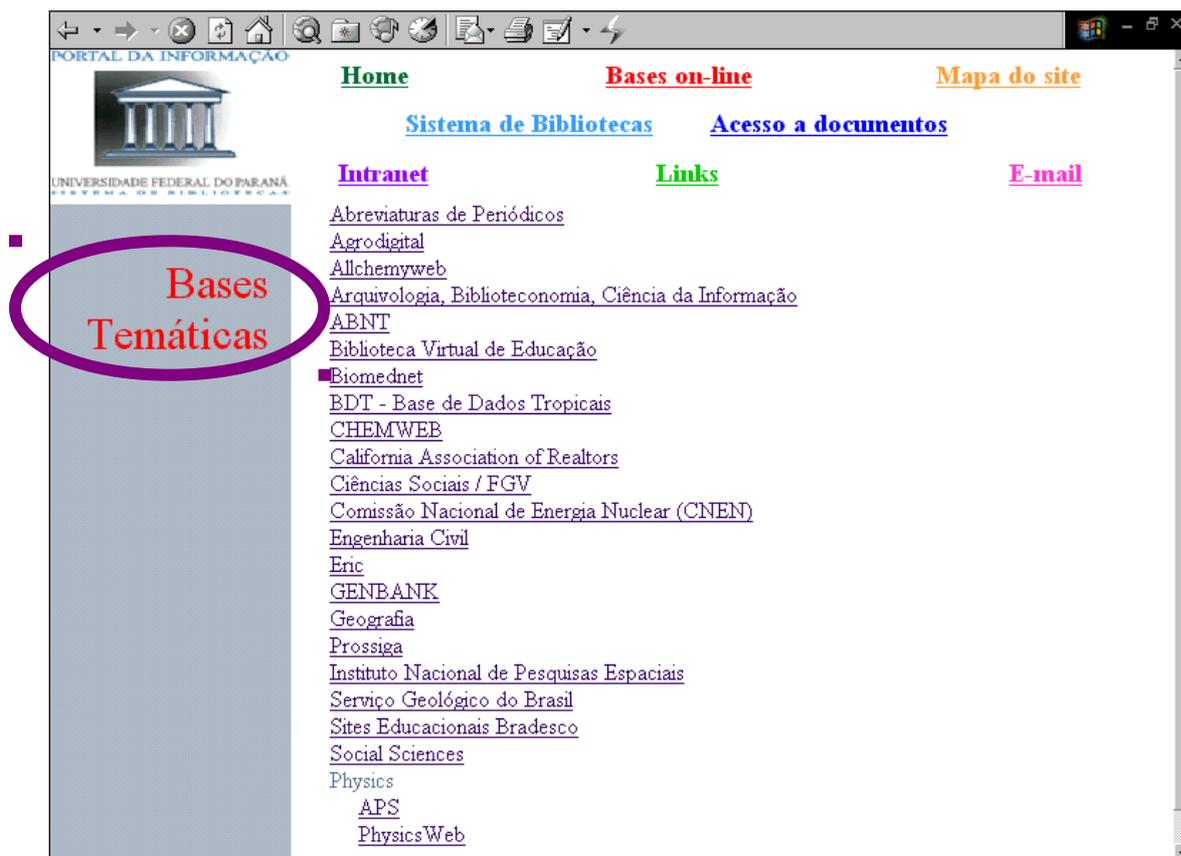
- c) passo 3 – clicar em cada um dos tipos de base para verificar em qual deles encontra-se a base procurada (figuras 8 e 9):

FIGURA 8 - INTEFACE DA PÁGINA BASES REFERENCIAIS DO PORTAL DA INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

FIGURA 9 – INTEFACE DA PÁGINA BASES TEMÁTICAS DO PORTAL DA INFORMAÇÃO DA UFPR



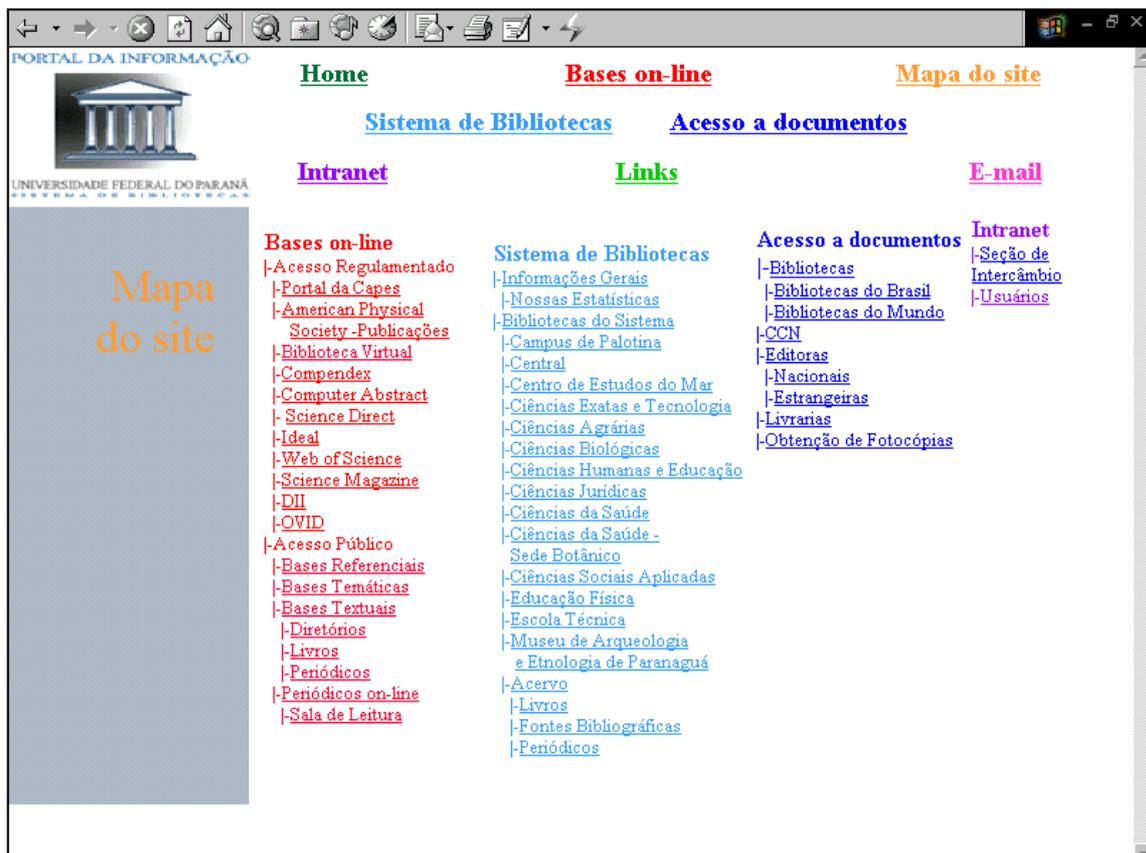
FONTE: UNIVERSIDADE (2002b)

- d) passo 4 – continuar abrindo sucessivamente os *links* até encontrar o que deseja ou visualizar “todas” as hierarquias existentes portal, de uma única vez, no **Mapa do Site** (figura 10). A palavra “todas” foi colocada entre aspas (“ ”), pois a estrutura do mapa encontra-se incompleta, como no exemplo das **Bases Textuais** que contêm *links* para Diretórios, Livros e Periódicos e esses três elementos ainda contêm dentro de si outros *links*

que, finalmente remetem o usuário ao destino da página para poder efetuar a pesquisa. O mesmo acontece com a hierarquia do *link* **Acesso a Documentos**, no qual ainda faltam itens dentro do mapa do *site*.

As hierarquias dos *links* do **Sistema de Bibliotecas**, assim como da **Intranet**, encontra-se em conformidade com o exposto no mapa do *site*.

FIGURA 10 – MAPA DO *SITE* DO PORTAL DA INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

Os problemas relativos à navegação não acabam por aqui, pois além de não haver uma caixa de busca e de o mapa do *site* encontrar-se incompleto, os *links* com os nomes das bases de dados *on-line* trazem poucas explicações sobre o seu tema e não contêm informações sobre as bases inseridas em sua hierarquia, ou seja, o usuário não sabe do que tratam e nem que técnicas de pesquisa (simples, operadores lógicos, de proximidade e outros) devem ser utilizadas nas mesmas para a recuperação da

informação, a menos que navegue pela base e a mesma contenha informações sobre que recursos de pesquisa deverão ser utilizados ou não. Além disso, é necessário que o usuário tenha um bom domínio do idioma inglês, pois a maior parte das bases encontram-se neste idioma e não possuem o recurso de mudar para o idioma português.

Com relação à indexação do *site* como um todo, o portal não possui *metatags* em seu código fonte escrito em *Hipertext Markup Language* (HTML) (figura 11) - linguagem caracterizada por possuir um alto nível de exaustividade para a recuperação da informação; essa falha no sistema acaba dificultando a pesquisa pela ausência desse recurso importante. Entretanto, a URL é textual (<http://www.portal.ufpr.br>) tem relação entre as palavras ou siglas usadas e o nome da instituição, no caso portal e ufpr, que auxiliam os robôs de busca a encontrarem-no com facilidade.

FIGURA 11 – CÓDIGO FONTE DA *HOMEPAGE* DO PORTAL DA INFORMAÇÃO DA UFPR

```

<html>
<head>
<TITLE>Portal da Informa&ccedil;&atilde;o da UFPR</TITLE>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1"></head>
<body bgcolor="#FFFFFF" text="#52678d" link="#330066" vlink="#000066" alink="#003333">
<div id="Layer1" style="position:absolute; left:2px; top:4px; width:111px; height:58px; z-index:1"

<tr>
<td width="15%">
<div align="left"><font face="Times New Roman, Times, serif" size="4">
<a href="frontal1.html"><b><font size="4" font color="#006633">Home</font></b></a></fon
</td>
<td width="5%">
<div align="center"></div>
</td>
<td width="51%">
<div align="center"> <a href="bases_on_line.htm"><b><font size="4" font color="#FF0000">B.
on-line</font></b></a></div>
</td>
<td width="29%">
<div align="right"> <a href="mapa.html"><b><font size="4" font color="#FF9933">Mapa
do site</font></b></a></div>
</td>
</tr>
</table>
</div>
<div id="Layer4" style="position:absolute; left:204px; top:41px; width:523px; height:29px; z-inde:
<table width="98%" border="0">
<tr>
<td width="2%">&nbsp;&nbsp;&nbsp;</td>
<td width="41%">
<div align="right">

```

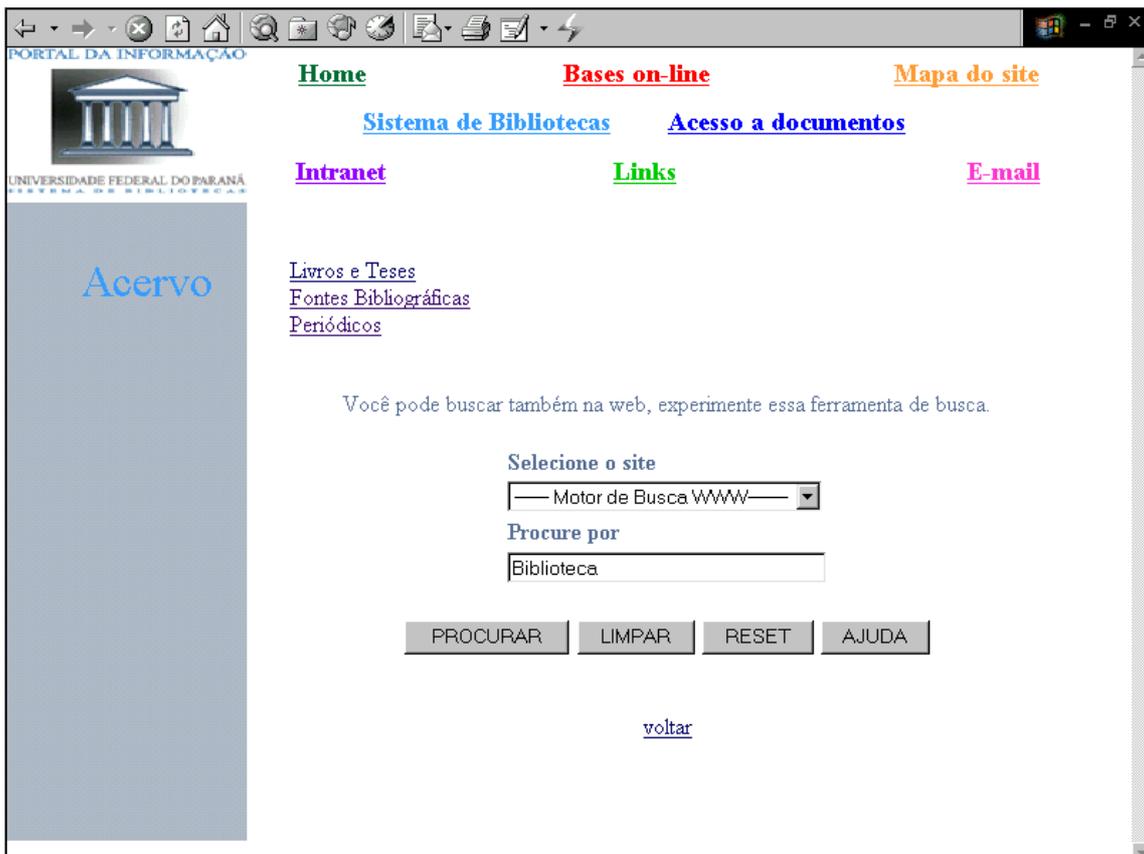
FONTE: UNIVERSIDADE (2002b)

Com relação à interface de busca e recuperação da informação dentro do portal, ou seja, dentro de suas principais hierarquias utilizadas para pesquisa e seus respectivos *links* (figura 2) - **Bases on-line**, **Sistema de Bibliotecas** e **Acesso a Documentos**, foi delimitado a página do Acervo (figura 12), assim como os *links* contidos em sua hierarquia (**Livros e Teses**, **Fontes Bibliográficas** e **Periódicos**), dentro do **Sistema de Bibliotecas**, para se poder verificar os índices de revocação e precisão do Portal da Informação da UFPR, mediante a execução de buscas utilizando-se palavras com problemas de ambigüidade (polissemia, sinonímia e homografia) e variadas técnicas de busca (pesquisa simples e pesquisa refinada: operadores lógicos, matemáticos, de existência, de proximidade e truncagem). Para essa verificação foi utilizada a técnica da amostragem que, segundo MARCONI²⁰, *apud* OLIVEIRA (1997,

²⁰ MARCONI, M. de A; LAKATOS, E. M. **Técnicas de pesquisa**. São Paulo: Atlas, 1985.

p. 159), “quando se deseja coletar informações sobre um ou mais aspectos de um grupo grande ou numeroso (diga-se neste caso um universo numeroso), verifica-se, muitas vezes, ser praticamente impossível fazer um levantamento do todo ... necessitando-se escolher uma amostra do universo para poder averiguá-lo”.

FIGURA 12 – INTERFACE DA PÁGINA ACERVO DO PORTAL DA
INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

Procurando-se visualizar melhor as características de cada uma das interfaces de busca do Sistema de Bibliotecas da UFPR, com base na literatura, construiu inicialmente um quadro comparativo (quadro 1), através do qual pode-se verificar que as interfaces não foram padronizadas dentro do portal e essa falta de padrão resultará tanto em formas de busca diferentes quanto em formatos visualmente de recuperação diferentes.

QUADRO 1 – COMPARAÇÃO ENTRE AS CARACTERÍSTICAS DAS

INTERFACES DE BUSCA DO SISTEMA DE BIBLIOTECAS DA
UFPR

INTERFACE CARACTERÍSTICAS	LIVROS*	FONTES BIBLIOGRÁFICAS	PERIÓDICOS
possui nota explicativa de como pesquisar	não	sim (nota simples)	não
possui dispositivo de ajuda total de caracteres dos campos de busca	não	não	não
	30	35	35
aceita letras maiúsculas	sim	sim	sim
aceita letras minúsculas	sim	sim	sim
aceita sinais diacríticos e acentos	não	sim	não
aceita operadores lógicos	não	não	não
aceita operadores de existência	não	não	não
aceita operadores de proximidade	não	não	não
aceita truncagem	sim	sim	sim
mostra número de itens encontrados	não	não	não
mostra data de atualização	não	não	não
possui <i>link</i> à biblioteca depositária da obra	não	sim	sim
possui <i>link</i> de retorno à página anterior	sim	sim	sim
possui <i>link</i> para alguma URL	sim	não	sim

FONTE: Elaboração do autor

*a interface de busca da página [livros](#) possui três campos de busca: assunto, autor, título, que estão sendo caracterizados ao mesmo tempo na tabela, pois possuem características em comum.

Visualizando-se o quadro de comparação entre as características das interfaces de busca do Sistema de Bibliotecas do Portal (quadro 1), verifica-se a falta de uma política de padronização em suas características, que pode ser melhor observada na ficha de detalhes das obras, que contém os campos dispostos em diferentes formatos,

como se pode constatar nos exemplos abaixo (figuras 13,14 e 15):

FIGURA 13 – DETALHE DE OBRA RECUPERADA NA INTERFACE LIVROS

TÍTULO	Tecnologia do cobre na arquitetura /
AUTOR	Simoes, Joao Roberto Leme.
ASSUNTO	///
BIBLIOTECA	BIBLIOTECA DE CIÊNCIA E TECNOLOGIA
PATRIMÔNIO	301592
NR CHAMADA	721.5/S593t
EDIÇÃO	
IMPRESSÃO	Sao Paulo : Pini, 1998.

FONTE: UNIVERSIDADE (2002b)

FIGURA 14 – DETALHE DE OBRA RECUPERADA NA INTERFACE FONTES BIBLIOGRÁFICAS

Título:	CAB Abstracts on CD
Editor:	CAB International, Oxon
Propósito:	Arrolar referências de artigos de periódicos, relatórios, conferências e livros nas áreas de Agronomia, Medicina veterinária, Engenharia florestal, Zootecnia, Reprodução animal, Ecologia, Agricultura – Economia, Pastagens, Gramíneas
Cobertura	Internacional
Suporte:	CD-ROM
Idioma	Inglês
Periodicidade:	Mensal
Período:	1972-
Disponibilidade:	Consulta realizada pelo funcionário, com hora marcada, mediante pagamento de taxa
Assuntos:	Agronomia; Medicina veterinária; Engenharia Florestal; Zootecnia; Reprodução animal; Economia Agrícola; Ecologia; Pastagens; Gramíneas; Forragicultura; Fisiologia Vegetal; Animais
Biblioteca:	BIBLIOTECA DE CIÊNCIAS AGRARIAS

FONTE: UNIVERSIDADE (2002b)

FIGURA 15 – DETALHE DE OBRA (LOCALIZADA NA BIBLIOTECA DE HUMANAS E EDUCAÇÃO DA UFPR) RECUPERADA NA INTERFACE PERIÓDICOS

TÍTULO	BIB-REVISTA BRASILEIRA DE INFORMACAO BIBLIOGRAFICA EM CIENCIAS SOCIAIS
Exemplares Disponíveis:	(29-30); (31-32); (33); (35); (36); (43-44); (45-46)
ISSN	0100-199X
Localização	HE
Url	

FONTE: UNIVERSIDADE (2002b)

As figuras 13, 14 e 15 foram inseridas para visualizar-se que cada uma das interfaces possui um padrão diferente do outro. Porém, observou-se que o formato tanto da de Livros quanto da de Fontes Bibliográficas, apesar das diferenças, foram adotadas por todas as bibliotecas do sistema. Já a de Periódicos, como se pode observar nas figuras 15 e 16, encontra-se com formatos diferentes em cada uma das bibliotecas. Verifica-se que cada uma das bases possui campos obrigatórios diferentes e adequados ao perfil de cada uma, porém seria adequado padronizar pelo menos os itens em comum na visualização de detalhes das obras, tais como: tamanho e tipo da fonte; se negritada ou não; se em letras maiúsculas e/ou minúsculas; *link* à biblioteca depositária; *link* à URL e outros.

FIGURA 16 – DETALHE DE OBRA (LOCALIZADA NA BIBLIOTECA DE CIÊNCIAS FLORESTAIS E DA MADEIRA DA UFPR) RECUPERADA NA INTERFACE PERIÓDICOS

TÍTULO	BIT – BOLETIM DE INFORMACAO TECNOLOGICA
Exemplares Disponíveis:	ANO:1996/1996 VOL:0001/0001 NUM:0002/0002 MES:03/04 ANO:1996/1996 VOL:0001/0001 NUM:0003/0003 MES:05/06 ANO:1996/1996 VOL:0001/0001 NUM:0004/0004 MES:07/08 ANO:1996/1996 VOL:0001/0001 NUM:0005/0005 MES:09/10 ANO:1996/1996 VOL:0001/0001 NUM:0006/0006 MES:11/12 ANO:1997/1997 VOL:0002/0002 NUM:0002/0002 MES:03/04 ANO:1997/1997 VOL:0002/0002 NUM:0004/0004 MES:07/08 ANO:1997/1997 VOL:0002/0002 NUM:0006/0006 MES:11/12 ANO:1998/1998 VOL:0003/0003 NUM:0001/0001 MES:01/02 ANO:1998/1998 VOL:0003/0003 NUM:0002/0002 MES:03/04 ANO:1998/1998 VOL:0003/0003 NUM:0003/0003 MES:05/06 ANO:1998/1998 VOL:0003/0003 NUM:0005/0005 MES:09/10 ANO:1998/1998 VOL:0003/0003 NUM:0006/0006 MES:11/12 ANO:1999/1999 VOL:0004/0004 NUM:0001/0001 MES:01/02
ISSN	
Localização	CE
Url	

FONTE: UNIVERSIDADE (2002b)

Alguns itens específicos importantes foram observados na análise da busca e recuperação da informação nas interface do sistema de bibliotecas do portal:

- a) **interface de busca Livros** (figura 17): os três campos de busca (assunto, autor, título) não podem ser preenchidos ao mesmo tempo, ou seja, não se pode refinar a busca. Caso os três campos pudessem ser preenchidos simultaneamente, o usuário perderia menos tempo. Alguns dos campos da interface possuem particularidades que deveriam estar explícitas para que o usuário não perdesse tempo:

- **campo autor:** pelo campo autor não é possível recuperar um livro colocando-se o nome e sobrenome do autor ao mesmo tempo como normalmente se escreve, como no exemplo: *Gareth Morgan*. Neste sentido, foram encontradas duas soluções: ou coloca-se uma palavra de cada vez (primeiro faz-se a busca por *Gareth* e depois faz-se uma nova busca por *Morgan*), causando-se assim uma quantidade excessiva de respostas, pois o sistema não sabe distinguir qual dos dois é nome e sobrenome, ou coloca-se o sobrenome antes do nome e

os dois separados por uma vírgula, como se fosse o campo de autoria de uma referência bibliográfica, como no exemplo: MORGAN, Gareth (pode-se colocar em letras maiúsculas e/ou minúsculas). Dessa última maneira o sistema recupera exatamente pelo autor que o usuário procura (a pergunta é: como o usuário iria adivinhar?), bastando apenas verificar qual dos títulos de livros é o desejado (entre os vários recuperados);

- **campo título:** o total de caracteres comportados pelos campos, em geral, é pequeno, havendo casos em que o usuário sabe exatamente o título completo da obra mas não consegue digitá-lo até o final por falta de espaço. No caso das buscas realizadas na interface da página livros (figura 17), um maior número de caracteres no campo título auxiliaria o usuário em sua busca, pois o campo não recupera as obras por partes de seu título, ou seja, o sistema retorna e responde que não foi localizado nenhum livro;

FIGURA 17 – INTERFACE DE BUSCA LIVROS DO PORTAL DA INFORMAÇÃO

PORTAL DA INFORMAÇÃO

[Home](#) [Bases on-line](#) [Mapa do site](#)

[Sistema de Bibliotecas](#) [Acesso a documentos](#)

[Intranet](#) [Links](#) [E-mail](#)

Livros

O serviço de busca do Sistema de Bibliotecas, está disponível ainda em fase de implementação, por isso é necessário ressaltar que a base de dados encontra-se incompleta.

O processo de busca pode ser feito de várias formas. Escolha a que melhor lhe convier e faça sua pesquisa:

Assunto

Autor

Título

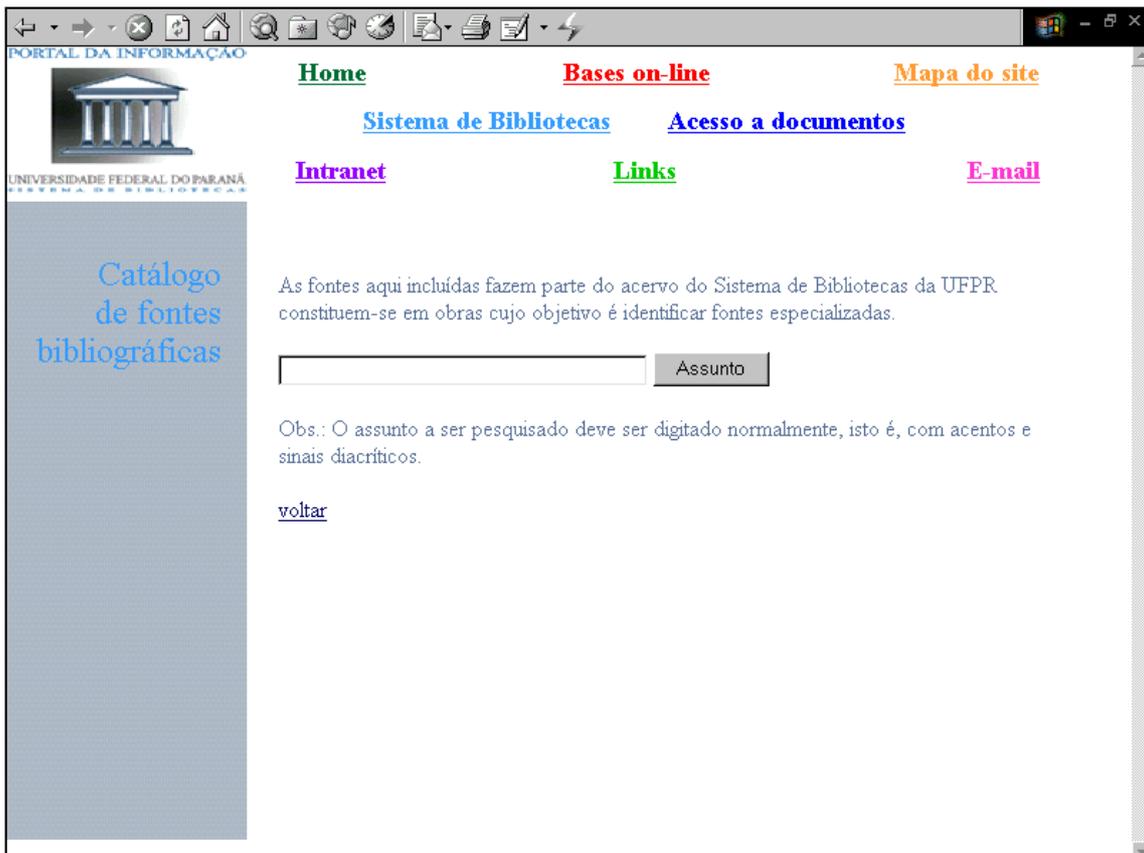
[voltar](#)

DA UFPR

FONTE: UNIVERSIDADE (2002b)

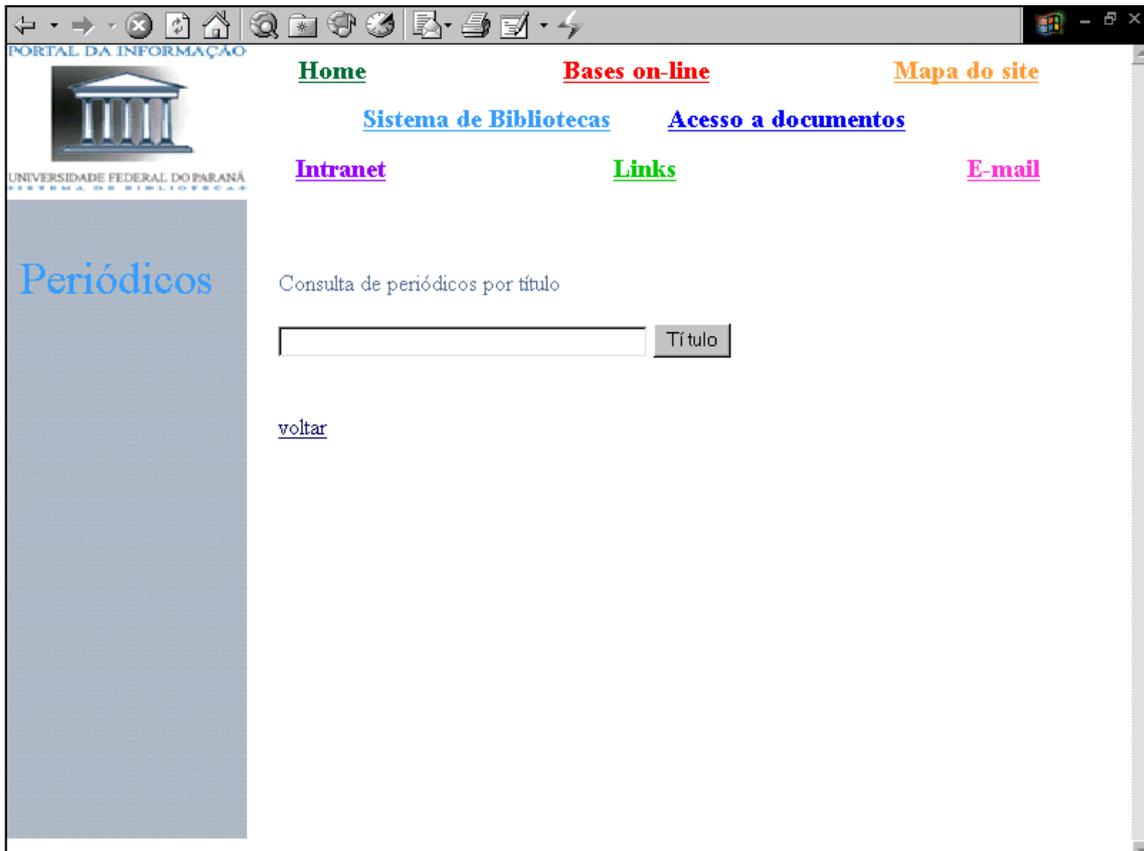
- b) **interface de busca Fontes Bibliográficas** (figura 18): é a única que apresenta uma nota explicativa precisa, mesmo que simples, de como preencher adequadamente o campo de busca;
- c) **interface de busca Periódicos** (figura 19): na verificação de detalhes dos periódicos recuperados, observou-se que apesar de muitos já possuírem versão *on-line*, a URL dos mesmos não foi colocada na ficha de detalhes, demonstrando falta de atualização de dados, que já foi mencionada inúmeras vezes. O problema maior está no conflito de padronização entre as informações repetitivas dentro do portal, pois de um lado encontra-se o *link* de Periódicos com informações defasadas e do outro encontra-se o *link* Periódicos on-line com informações atualizadas. Entretanto, verifica-se que ao menos o *link* de Periódicos on-line acaba contribuindo para a “atualização” de dados do acervo de periódicos do Sistema de Bibliotecas, restando apenas uma pergunta: será que o usuário descobrirá que existe esta possível saída, sem que verifique o mapa do *site*, apenas entrando diretamente para executar buscas no acervo do Sistema de Bibliotecas? Para sanar essa deficiência, se faz necessária a inserção de recursos de ajuda ou de notas explicativas expondo que a seção na qual o usuário se encontra está defasada, podendo o mesmo tentar o *link* tal para recuperar informações atuais.

FIGURA 18 – INTERFACE DE BUSCA FONTES BIBLIOGRÁFICAS DO PORTAL DA INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

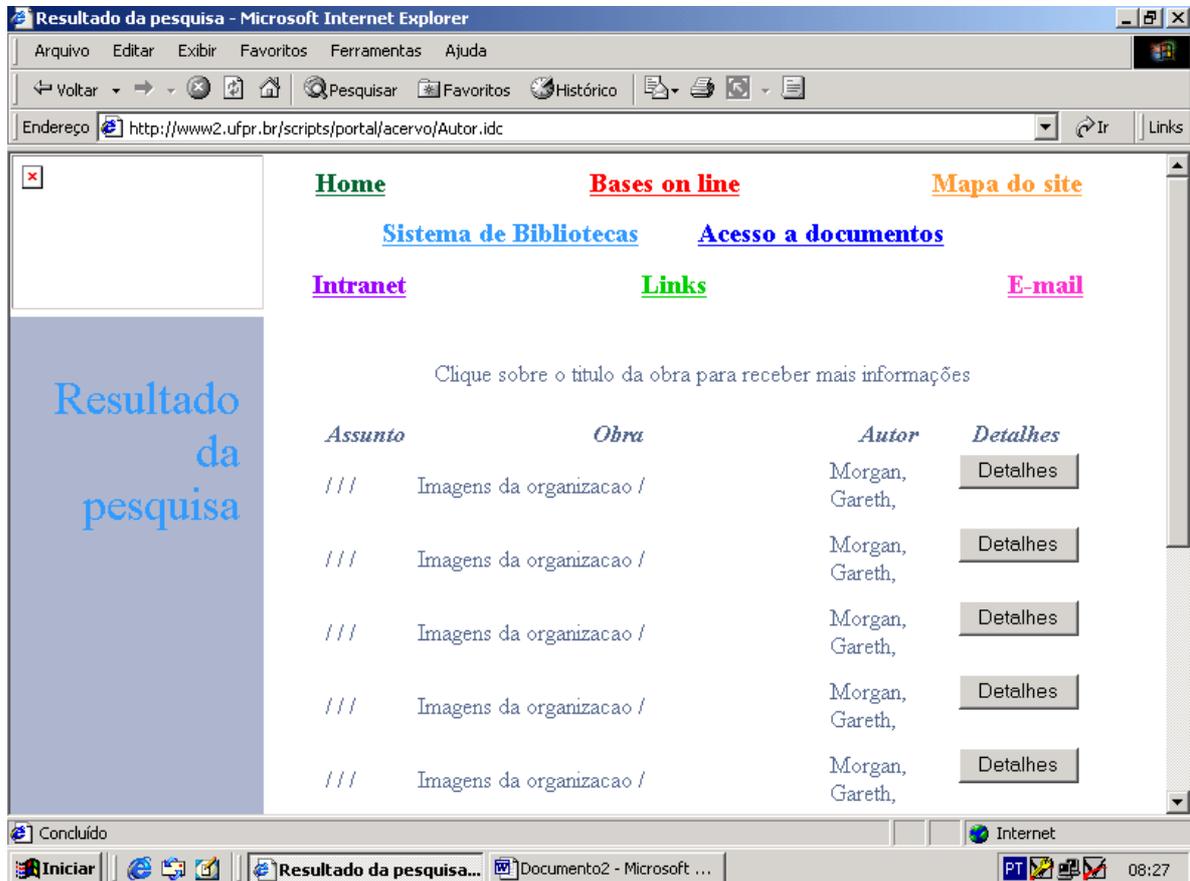
FIGURA 19 – INTERFACE DE BUSCA DOS PERIÓDICOS DO PORTAL DA INFORMAÇÃO DA UFPR



FONTE: UNIVERSIDADE (2002b)

Após a verificação das principais características das interfaces de busca do Sistema de Bibliotecas do Portal da Informação da UFPR, partiu-se para a execução de buscas utilizando-se algumas palavras encontradas na literatura que apresentam os problemas de ambigüidade (polissemia, sinonímia e homografia) já descritos, para poder verificar os índices de precisão e revocação do portal, tendo como interface de busca somente o campo título da interface de busca Livros, pois o resultado mais visível da busca na tela é o campo do meio denominado de **Obra**, sem que seja necessário abrir a parte de **Detalhes**, no canto direito (figura 21). Visualizando-se a figura, vê-se que na parte denominada **Assunto** ou não há nenhuma informação, ou apresenta alguns caracteres barra (///) sem denominação alguma e, mesmo a ficha de detalhes não traz o assunto da obra, somente as barras (///) aparecem.

FIGURA 20 – INTERFACE DE APRESENTAÇÃO DO RESULTADO DE BUSCA



FONTE: UNIVERSIDADE (2002b)

Em um primeiro exemplo, colocou-se a palavra *gato* como polissêmica na frase “tenho um *gato* e preciso de informações específicas sobre o seu comportamento”. Entretanto, a palavra *gato* nesta frase é ambígua e pode significar tanto *gato (animal)*, como *gato (gíria - homem bonito)*, como *gato (ladrão)*, entre outros, causando problemas na recuperação da informação (quadro 2). O resultado obtido foi um **total de 123 itens recuperados**, dos quais **71 itens prováveis** para *gato (animal)* e outros 52 inúteis (48 + 4 = 52), sendo que somente **51 itens úteis**, dos 71, eram sobre *comportamento do gato (animal)* e foram aproveitados, gerando **72 itens inúteis** (52 + 20 = 72). Assim sendo, conclui-se que o índice de precisão (relação entre itens úteis e o total de itens recuperados, 51/123) é 41,5% e o índice de revocação (relação entre os

itens úteis e os itens prováveis, 51/71) é 71,8%.

QUADRO 2 – RESULTADO DE BUSCA COM PROBLEMA DE POLISSEMIA

INTERFACE DE BUSCA	TOTAL	GATO (ANIMAL)	GATO (GÍRIA – HOMEM BONITO)	OUTROS
<u>Livros</u> (campo título)	123	71	04	48

FONTE: Elaboração do autor

Posteriormente, efetuou-se buscas com as palavras sinônimas *aipim*, *macaxeira* e *mandioca* (quadro 3). O resultado obtido revela que somente a palavra *mandioca* é recuperada, ou seja, caso o usuário execute a busca pelos termos *aipim* ou *macaxeira*, o resultado será nulo. Esta pesquisa foi somente para averiguar a recuperação de palavras sinônimas, na qual concluiu-se que somente uma das palavras é escolhida na indexação e as outras são rejeitadas. Quanto aos índices de revocação e precisão, torna-se impossível calculá-los pela insuficiência de dados, haja vista que neste exemplo procurou-se visualizar a aceitação ou não de sinônimos.

QUADRO 3 – RESULTADO DE BUSCA COM PROBLEMA DE SINONÍMIA

INTERFACE DE BUSCA	TOTAL	AIPIM	MACAXEIRA	MANDIOCA
<u>Livros</u> (campo título)	18	00	00	18

FONTE: Elaboração do autor

Finalmente, colocou-se a palavra *cobre* na execução de busca, procurando-se pelo termo como “*cobre (metal)* e seu uso na arquitetura”. Entretanto, o termo é ambíguo podendo significar tanto *cobre (metal)*, como *cobre (verbo cobrir)*, causando problemas na recuperação da informação (quadro 4). O resultado obtido foi um **total de 24 itens recuperados**, dos quais **19 itens prováveis** para *cobre (metal)* e outros 05 inúteis (01 + 04 = 05), sendo que somente **4 itens úteis**, dos 19, eram sobre *cobre (metal)* e seu uso na arquitetura e foram aproveitados, gerando **22 itens inúteis** (05 +

17 = 22) . Assim sendo, conclui-se que o índice de precisão (relação entre itens úteis e o total de itens recuperados, 4/24) é 16,6% e o índice de revocação (relação entre os itens úteis e os itens prováveis, 4/19) é 21,5%.

QUADRO 4 - RESULTADO DE BUSCA COM PROBLEMA DE HOMOGRAFIA

INTERFACE DE BUSCA	TOTAL	COBRE (METAL)	COBRE (VERBO COBRIR)	OUTROS
<u>Livros</u> (campo título)	24	19	01	04

FONTE: Elaboração do autor

Estas verificações podem apresentar um certo grau de imprecisão, pois como a ficha de detalhe da obra traz o campo assunto, mas o mesmo não é preenchido, é difícil precisar somente pelo título se o termo é exatamente compatível com a necessidade de informação. Pode-se tentar verificar o assunto pela número de classificação da obra. Entretanto, a classificação é um produto subjetivo que depende da política da instituição em que a obra se encontra. Além disso, somente pessoas com conhecimentos específicos conseguem desvendar essa numeração. Conclui-se então, que o usuário terá que anotar os itens das prováveis obras pertinentes às suas resposta, ir até o local de depósito para verificá-las, uma a uma, pois a falta de colocação de termos e de seu real significado no campo assunto, pode causar tanto ambigüidades quanto nulidades na recuperação da informação.

8 CONSIDERAÇÕES FINAIS

As tecnologias da informação e comunicação influenciam a área de informação, gerando o aparecimento de novas técnicas de representação e recuperação de conteúdo que visam desenvolver sistemas inteligentes de recuperação da

informação, baseadas no processamento da linguagem natural, visando a disponibilização de textos completos, mediante a utilização de interfaces amigáveis, com o objetivo de suprir as necessidades de informação de seu usuário final.

Todavia, como se pode observar no estudo de caso do Portal por mais inteligentes e amigáveis que sejam as interfaces de busca, estas tornam-se falhas quando conduzem à recuperação de informações irrelevantes, ocasionadas por problemas relativos ao tratamento indevido do conteúdo das informações, que requer habilidades e aplicação de conhecimentos lingüístico-cognitivos para uma recuperação relevante. Porém, estas implementações requerem um trabalho complexo e de manutenção rotineira, que envolve a definição de políticas de seletividade de informações, assim como tempo, dinheiro e, principalmente, pessoas especializadas em tratamento da informação pois, devido ao excesso de informações disponibilizadas virtualmente, os sistemas adotaram a extração automática de palavras e não conseguiram solucionar os problemas lingüísticos de ambigüidade das palavras (polissemia, sinonímia e homografia) que nos sistemas tradicionais eram tratados mediante a utilização de linguagens documentárias.

Assim sendo, cabe aos profissionais da área de informação monitorar as TIC, assim como o tratamento da informação executado nas mesmas, visando aprimorar e/ou criar métodos adequados às técnicas de recuperação da informação.

9 SUGESTÕES

As sugestões arroladas neste capítulo referem-se, preliminarmente, a melhorias que podem ser implementadas no Portal da Informação da UFPR, deixando-se aqui a proposta de que, posteriormente, mais graduandos da área de informação e afins, executem análises de *sites* de informação pois, atualmente, encontramos na era da economia baseada na informação como mola propulsora do desenvolvimento de

países em vias de desenvolvimento. Como encaixamo-nos no perfil de cidadãos das nações, temos o dever de contribuir, de alguma forma, para melhorar o acesso às informações e disponibilizá-las no meio virtual, como propõe o livro *Sociedade da Informação no Brasil: livro verde*, do Ministério de Ciência e Tecnologia (MCT).

Para uma melhora do *site* do portal da informação da UFPR, encontram-se aqui listados os problemas e suas sugestões de aprimoramento do mesmo:

a) abrangência: indicar a abrangência do *link* **Intranet**, pontuando que o mesmo refere-se às bases que só podem ser acessadas pela rede interna da instituição, e do *hyperlink* **Universidade Federal do Paraná**, colocando que o mesmo remete para a página da UFPR, já na *homepage*, para facilitar a navegabilidade do usuário, que talvez não esteja familiarizado com os termos acima utilizados;

b) conteúdo:

- adequar a linguagem do texto utilizado para descrever o conteúdo da *homepage* ao perfil de seus usuários acadêmicos, cientistas, entre outros;
- referenciar quantos, quais e de que áreas são os profissionais que fizeram parte da criação do portal pois, essencialmente, o portal deve ser um trabalho elaborado por diversos profissionais, não somente da área de informação, como também das outras áreas (humanas, biológicas, tecnológicas), pois os especialistas no tratamento da informação precisam trabalhar de forma interdisciplinar com especialistas em outras áreas do conhecimento, para que possam gerar produtos informacionais de alta qualidade;
- colocar data de atualização do *site*, valorizando-o;
- acrescentar dispositivo de ajuda na *homepage* e em cada um dos *links* de acesso;

c) planejamento visual/gráfico (*webdesign*):

- implementar uma caixa de busca, ou seja, um retângulo no qual digita-se um termo que se queira buscar como o “Procure no *Site*”, da *homepage* do Portal de Referência em Arquivologia, Biblioteconomia e Ciência da Informação da Universidade Federal Fluminense (UFF) (figura 3);
- completar a hierarquia da estrutura do mapa do *site*;

d) funcionalidade (interface e quesitos de navegabilidade):

- explicitar o conteúdo existente nas bases de dados, assim como em outros *links*;
- no *link* do Acervo e suas interfaces de busca, do **Sistema de Bibliotecas**: referenciar, exemplificadamente, como fazer as buscas; implementar a alternativa de busca refinada; expandir o número de caracteres que compõem os campos; apresentar o número de itens recuperados; padronizar a apresentação dos resultados, assim como evidenciar todos os detalhes de seus campos.

Mediante essas sugestões, pretende-se contribuir para a melhora dos principais critérios na implementação e avaliação de *sites*, em especial do Portal da Informação da UFPR, que necessita ser analisado cuidadosa e detalhadamente, para poder condizer, efetivamente, com as características teóricas existentes na literatura pertinente sobre o tema portal da informação, abordadas ao longo desta pesquisa, para que o mesmo, na prática, possa denominar-se uma ferramenta chamada portal pois, a princípio, o perfil do mesmo qualifica-o como *site*, pelo fato de possuir inúmeras de páginas da *web*, sobre variados assuntos, instituições, entre outros. Essa característica acaba tornado-o um simples repositório de informações sem tratamento que, conseqüentemente, pode culminar em problemas ou nulidades na recuperação da informação, deixando assim de cumprir a grande meta de um portal, que é fornecer aos seus usuários informações relevantes, de modo descomplicado e rápido.

GLOSSÁRIO

Caracter

Unidade básica de informação em uma hierarquia de dados. Um conjunto de caracteres consiste em letras maiúsculas e minúsculas, dígitos numéricos e caracteres especiais, como sinais de pontuação (STAIR, 1998, p. 409).

Chat

Bate-papo eletrônico pela Internet (DIAS, 2001, p. 59).

Correio eletrônico

Modalidade de transmissão de mensagens por meio de rede eletrônica, em particular a Internet. Usa-se também a forma inglesa *e-mail* (BRASIL, 2000, p. 167).

Data warehouse

Grande repositório de dados, elaborado com a finalidade de dar suporte ao processo decisório estratégico da empresa (DIAS, 2001, p. 59).

E-mail

ver correio eletrônico.

Extensible Markup Language (XML)

Linguagem de Marcação Extensível. Metalinguagem, subconjunto da SGML, que permite a criação de formatos de informação compartilháveis por um conjunto de usuários, no que se refere tanto aos conteúdos quanto aos próprios formatos. Usa símbolos de marcação para indicar tipos de informação em conteúdos, a forma como os dados vão ser interpretados e como se vai interagir com eles (BRASIL, 2000, p. 168).

Feedback

Saída utilizada para ajustar e controlar um sistema. É também o processo pelo qual a saída de controle é coletada e reintroduzida no sistema (STAIR, 1998, p. 416).

Hyperlink

Palavra destacada ou figura de um hipertexto que produz um salto para a outra página da *web* ou para outra parte da página exibida (MOURA, 2002).

Hipermídia

Mecanismo de ligação que inclui elementos de texto a outras formas de representação, tais como imagens e sons (DENECA, 2000, p. 205).

Hipertexto

Texto que contém *links* (apontadores, elos) para outros documentos ou outras partes do mesmo documento. Os elos estão associados a palavras ou expressões que permitem ao leitor se deslocar automaticamente para as partes por eles apontadas (BRASIL, 2000, p. 169).

Homepage

Principal e primeira página de um *site*. Serve como índice ou sumário de documentos armazenados naquele ou em outro *site* (BRASIL, 2000, p. 169).

Host

No contexto da Internet um *host* é um computador ou dispositivo que possua um endereço Internet e possa se comunicar com outros *hosts*. É um nó da rede (BRASIL, 2000, p. 169).

HTML

ver Hypertext Markup Language (Linguagem de Marcadores para Hipertexto).

HTTP

ver Hypertext Transfer Protocol (Protocolo de Transferência de Hipertexto).

***Hypertext Markup Language* (HTML)**

Linguagem de Marcação de Hipertexto. Linguagem padrão usada para escrever páginas de documentos para WWW, variante (subconjunto) da SGML. Possibilita preparar documentos com gráficos e *links*, para visualização em sistemas compatíveis com o WWW (BRASIL, 2000, p. 169).

***Hypertext Transfer Protocol* (HTTP)**

Protocolo de Transferência de Hipertexto. Conjunto de regras para intercâmbio de informações (texto, imagens, som, vídeo, e outros arquivos multimídia) na WWW. É um protocolo de aplicação (BRASIL, 2000, p. 170).

IA

ver Inteligência Artificial.

Índice

Catálogo dos principais termos de busca elaborados e criados nos sistemas de busca. Os mecanismos de busca utilizam a página de índice para construir e apresentar os resultados de uma pesquisa (DENECA, 2000, p.205).

Inteligência Artificial (IA)

Capacidade de imitar ou reproduzir as funções do cérebro humano (STAIR, 1998, p. 417).

Interface

Qualquer dispositivo através do qual um usuário pode se comunicar com um sistema (STAIR, 1998, p. 417).

Internet

Sistema mundial de redes de computadores – uma rede de redes – que pode ser utilizado por qualquer pessoa, em qualquer parte do mundo, onde haja ponto de acesso, e que oferece um amplo leque de serviços básicos, tais como correio eletrônico, acesso livre ou autorizado a informações em diversos formatos digitais, transferência de arquivos. Os protocolos básicos para o transporte dos dados são do TCP/IP (BRASIL, 2000, p. 171).

Intranet

Aplicação da tecnologia Internet no âmbito interno da empresa (DIAS, 2001, p. 59).

Link

Elo ou apontador para outros documentos e/ou partes do documento, em hipertexto (BRASIL, 2000, p. 171).

Metadados

ver metatags.

Metatags (Metadados)

Dados a respeito de outros dados, ou seja, qualquer dado usado para auxiliar na identificação, descrição e localização de informações. Trata-se, em outras palavras, de dados estruturados que descrevem as características de um recurso de informação (BRASIL, 2000, p. 172).

Standard Generalized Markup Language (SGML)

Linguagem Padrão de Marcação. Sistema de organização e marcadores dos elementos de um documento (BRASIL, 200, p. 176).

SGML

ver Standard Generalized Markup Language (Linguagem Padrão de Marcação generalizada).

Site

Coleção de páginas da *web* referentes a um assunto, instituição, empresa, pessoa etc. Diz-se também *website*. A forma portuguesa *sítio* é pouco usada (BRASIL, 2000, p. 175).

Sítio

ver site.

Standard Generalized Markup Language (SGML)

Linguagem padrão de marcação generalizada. Sistema de organização e marcadores dos elementos de um documento (BRASIL, 2000, p. 176).

Tag

É uma seqüência de caracteres na HTML e na SGML que codifica a correlação estrutural de uma seqüência de texto num documento. Uma *tag* é, normalmente, mostrada na forma: `<tag>`, que indica o início de uma seqüência de um texto de tipo específico, e `</tag>`, que indica o fim da seqüência (GLOSSÁRIO, 2002).

TCP/IP

ver Transmission Control Protocol/Internet Protocol (Protocolo de Controle de Transmissão/Protocolo da Internet).

Tecnologias de Informação e Comunicação (TIC)

Tecnologias utilizadas para tratamento, organização e disseminação de informações. Pode-se utilizar a abreviação TIC (BRASIL, 2000, p. 176).

TIC

ver tecnologias de informação e comunicação .

Transmission Control Protocol/Internet Protocol (TCP/IP)

Protocolo de Controle de Transmissão/Protocolo da Internet. Conjunto de protocolos básicos da Internet, que permitem a entrega de pacotes de um *host* para outro (BRASIL, 2000, p. 177).

Uniforme Resource Locator (URL)

Localizador de Páginas da Internet. Forma padronizada de endereçar recursos na Internet (BRASIL, 2000, p. 177).

URL

ver Uniforme Resource Locator (Localizador de Páginas da Internet).

XML

ver Extensible Markup Language (Linguagem de Marcação Extensível).

Web

ver WWW.

Website

ver site.

WWW

ver World Wide Web (Teia Global).

World Wide Web (WWW)

Teia Global. Conjunto de documentos e serviços, que faz parte da Internet, organizados em forma de páginas de hipertexto, em que cada página é identificada por um URL. Também é chamada de *web* (BRASIL, 2000, p. 178).

REFERÊNCIAS

- ARAÚJO, V. M. R. H. de. Sistemas de informação: nova abordagem teórico-conceitual. **Ciência da Informação *On line***, v. 24, n. 1, 1995. Disponível em: <<http://www.ibict.br>> Acesso em: 15 ago. 2001.
- BARBOZA, E. M. F. *et al.* Web sites governamentais, uma esplanada à parte. **Ciência da Informação *On line***, v. 29, n. 1, p. 118-125, jan./abr. 2000. Disponível em: <<http://www.ibict.br>> Acesso em: 15 ago. 2001.
- BERTALANFFY, L. von. **Teoria geral dos sistemas**. Petrópolis: Vozes, 1975. 351 p.
- BIO, S. R. **Sistemas de informação: um enfoque gerencial**. São Paulo: Atlas, 1996. 183 p.
- BLATTMANN, U. *et al.* Recuperar a informação eletrônica pela Internet. **Revista da ACB: Biblioteconomia em Santa Catarina**, v. 4, n. 4, p. 9-27, 1999. Disponível em: <<http://www.geocities.com/ublattmann/papers/buscanet.html>> Acesso em: 24 fev. 2002.
- BOANERGES, S. **Como achar algo (útil) na Internet**. Disponível em: <http://www.bhnet.com.br/~tiao/como_achar.html> Acesso em 20 fev. 2002.
- BRAGA, G. M. Informação, ciência da informação: breves reflexões em três tempos. **Ciência da Informação *On line***, v. 24, n. 1, 1995. Disponível em: <<http://www.ibict.br>> Acesso em 15 ago. 2001.
- BRÄSCHER, M. A ambigüidade na recuperação da informação. **DataGramaZero** – revista de ciência da informação, v. 3, n. 1, fev. 2002. Disponível em: <http://www.dgzero.org/Atual/F_I_art.htm> Acesso em 20 mar. 2002.
- BRASIL. Ministério da Ciência e Tecnologia. **Sociedade da informação no Brasil: livro verde**. Brasília: Ministério da Ciência e Tecnologia, 2000. 195p. Disponível em: <http://www.socinfo.org.br/livro_verde> Acesso em 20 ago. 2001.
- CARVALHO, I. C. L. ; KANISKI, A. L. A sociedade do conhecimento e o acesso à informação. **Ciência da Informação *On line***, v. 29, n. 3, p. 33-39, set./dez. 2000. Disponível em: <<http://www.ibict.br>> Acesso em 15 ago. 2001.
- CASSARRO, A. C. **Sistemas de informações para a tomada de decisões**. 3. ed. rev. e ampl. São Paulo: Pioneira, 1999. 129 p.
- CHURCHMAN, C. W. **Introdução à teoria dos sistemas**. Rio de Janeiro: Vozes, 1972. 309 p.

COOKE, A. **A guide to finding quality information on the Internet**: selection and evaluation strategies. 5. ed. London: Library Association Publishing, 1999. 169 p.

DATE, C. J. **Introdução a sistemas de banco de dados**. Rio de Janeiro: Campus, 2000. 803 p.

DENEGA, M. A. **Como pesquisar na Internet**. São Paulo: Berkley Brasil, 2000. 217 p.

FREIRE, I. M. ; ARAÚJO, V. M. R. H. de. A responsabilidade social da ciência da informação. **Transinformação**, Campinas, v.11, p. jan./abr. 1999. Disponível em: <<http://www.puccamp.br/~biblio/transinformacao/pag11.html>> Acesso em 15 ago. 2001.

FREITAS, H. *et al.* **Informação e decisão**: sistemas de apoio e seu impacto. Porto Alegre: Ortiz, 1997. 214 p.

FREITAS, H. ; LESCA, H. Competitividade empresarial na era da informação. **Revista de Administração da USP**, São Paulo, v. 27, n. 3, p. 92-102, jul./set. 1992.

FURLAN, J. D. *et al.* **Sistemas de informação executiva = EIS**: como integrar os executivos ao sistema informacional das empresas. São Paulo: Makron Books, 1994. 157 p.

GLOSSÁRIO de termos sobre Internet . Disponível em: <<http://www.startpoint.com.br/glossa0.htm>> Acesso em: 08 mar. 2002.

GUINCHAT, C. ; MENO, M. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. Brasília: IBICT, 1994. 540 p.

KURAMOTO, H. Sintagmas nominais: uma nova proposta para a recuperação de informação. **DataGramaZero** – revista de ciência da informação, v. 3, n. 1, fev. 2002. Disponível em: <http://www.dgzero.org/Atual/F_I_art.htm> Acesso em: 20 mar. 2002.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília: Brinquet de Lemos/Livros, 1993. 347 p.

LAUDON, K. C. ; LAUDON, J. P. **Sistemas de informação**. 4. ed. Rio de Janeiro: Livros Técnicos e Científicos, 1999. 389 p.

LIMA, C. M. **O que é a Internet e como utilizá-la para pesquisa**. Disponível em: <<http://www.ufma.br/filosofia/#TUTOR>>. Acesso em: 20 fev. 2002.

MARCHIORI, P. **Navegar é preciso**: como entender a estrutura de busca na Web. In: PESQUISA AVANÇADA NA INTERNET, 2001, Curitiba: UFPR, 2001. [mensagem pessoal]. Mensagem recebida por: <prostt@starmedia.com.br> em: 16 abr. 2001.

MARCONDES, C. H. A representação e economia da informação. **Ciência da Informação *On line***, v. 30, n. 1, p. 61-70, jan./abr. 2001. Disponível em: <<http://www.ibict.br>> Acesso em 15 ago. 2001.

MARTINS, G. de A. ; LINTZ, A. **Guia para a elaboração de monografias e trabalhos de conclusão de curso**. São Paulo: Atlas, 2000. 108 p.

MOURA, G. A. C. **Sistemas de busca da web**: diretórios e mecanismos de busca Disponível em: <http://www.quatrocantos.com/tec_web/sist_busca/11introd.htm>. Acesso em: 20 fev. 2002.

O'BRIEN, J. A. **Sistemas de informação**: as decisões gerenciais na era da Internet. São Paulo: Saraiva, 2001. 436 p.

OLIVEIRA, D. P. R. **Sistemas de informações gerenciais**: estratégicas, táticas, operacionais. 5. ed. São Paulo: Atlas, 1998. 275 p.

OLIVEIRA, S. L. de. **Tratado de metodologia científica**: projetos de pesquisa, TGI, TCC, monografias, dissertações e teses. São Paulo: Pioneira, 1997. 320 p.

POLLONI, E. G. F. **Administrando sistemas de informação**: estudo de viabilidade. 2. ed. São Paulo: Futura, 2001. 272 p.

ROSARIO GIRARDI, María del. **Tópicos em sistema de informação**. Maranhão: UFMA, 2001. 72 p.

SABBATINI, R. **A era dos portais**. Disponível em: <<http://www.epub.org.br/correio/cp991022.htm>>. Acesso em: 20 fev. 2002.

SANTOS, A. R. dos. **Metodologia científica**: a construção do conhecimento. Rio de Janeiro: DP&A, 1999. p. 26, p. 29. 139 p.

SOARES NETO, H. O. **Análise vital de sistemas**. Rio de Janeiro: Datamec, 1993. 199 p.

STAIR, R. M. **Princípios de sistemas de informação**: uma abordagem gerencial. Rio de Janeiro: Livros Técnicos e Científicos, 1998. 451 p.

TOFFLER, A. **A terceira onda**. 24. tir. Rio de Janeiro: Record, 1999. 491 p.

UNIVERSIDADE FEDERAL FLUMINENSE. Sistema NDC. **Portal de referência em Arquivologia, Biblioteconomia e Ciência da Informação**. Disponível em: <<http://www.ndc.uff.br/sir/PORTAL.htm>>. Acesso em: 08 mar. 2002a.

UNIVERSIDADE FEDERAL DO PARANÁ. Sistema de Bibliotecas. **Portal da informação da Universidade Federal do Paraná**. Disponível em: <<http://www.portal.ufpr.br>>. Acesso em: 08 mar. 2002b.

VALADARES CEDÓN, B. Ferramentas de busca na *Web*. **Ciência da Informação *On line***, v. 30, n. 1, p. 39-49, jan./abr. 2001. Disponível em: <<http://www.ibict.br>>
Acesso em: 15 fev. 2002.