

## 5 Sistemas de busca e recuperação de informação

Sistemas para recuperação de informação baseados na *web* são sistemas complexos e exigem a soma de conhecimentos de profissionais com domínio de conhecimento em diversas áreas. O profissional de ciência da informação pode contribuir com conhecimentos específicos para a criação de bancos de dados e oferecer uma perspectiva técnica para indexação e recuperação de informação (Dzekaniak, 2004). Por sua vez, o especialista em Ergonomia e usabilidade poderá conduzir testes com usuários por meio da análise da tarefa e de outros métodos de projeto. O programador e o designer de interfaces trabalharão para materializar o produto com o qual os usuários irão interagir. Conhecer a forma como o usuário interage em um ambiente físico é a melhor maneira de se iniciar um projeto como o de um *website* para biblioteca.

Essa crescente preocupação da área de Ergonomia e usabilidade em propor recomendações para o projeto de interfaces para sistemas de recuperação de informação *on-line* tem como foco fatores como: orientação para a realização da busca, ajudas, alternativas para o uso de operadores lógicos e interfaces para uso de linguagem natural. Para que o usuário localize a informação mais adequada às suas necessidades, é importante que seja realizado um trabalho de estruturação da informação, estruturação essa que pode ser entendida como “sua organização lógica para posterior recuperação e linguagem de busca como os comandos que permitem a recuperação da informação através de palavras contidas nos títulos, resumos ou outros campos de dados” (Branski, 2004).

Ferreira *et al* (2003) apontam que diversos estudos revelam problemas básicos relativos à falta de consistência entre a organização da informação e sua expressão no sistema. Outros problemas referem-se à expectativa e ao interesse do usuário, assim como dificuldades determinadas pelo comportamento não sistemático do ambiente de interação. Tão mais graves são estes problemas quando verificados em *websites* de bibliotecas virtuais, *on-line* ou digitais.

Os sistemas de busca são, usualmente, categorizados em diretórios, motores de busca e metaferramentas. Os diretórios de busca, também conhecidos como catálogos, são formas de organizar hierarquicamente a

informação para minimizar o esforço do usuário. No entanto, organizar um diretório demanda trabalho intenso, pois exige que um operador humano pesquise por endereços na internet, faça a classificação e a inclusão em categorias distintas. Além disso, é necessário certificar-se a respeito da atualidade dos *websites* para onde apontam os *links*. Todo esse trabalho é feito de forma manual e o conteúdo é sempre submetido à crítica da pessoa responsável pela manutenção do catálogo.

A figura 43 apresenta uma tela do catálogo da Biblioteca Virtual de Política Científica e Tecnológica.

The screenshot shows the homepage of the Prossiga Virtual Library of Scientific and Technological Policy. At the top, there are logos for Prossiga, UNICAMP, and CGEE. A search bar is prominently displayed with the text 'Os acentos devem ser usados, quando for o caso.' Below the search bar are buttons for 'Consultar', 'Limpar', and 'Busca Avançada'. A navigation menu titled 'navegue pela biblioteca' lists various categories such as 'Planos, Programas, Projetos e Fundos', 'Legislação', 'Instrumentos de Política', 'Órgãos de Política e Fomento', 'Gestores, Especialistas e Pesquisadores', 'Organizações Multilaterais e Internacionais', 'Empresas, Instituições, Movimentos e Organizações Não-Governamentais', 'Artigos, Entrevistas, Relatórios e Outros Textos', 'Avaliações de Políticas, Programas e Projetos', 'Ensino e Pesquisa em C&T', 'Periódicos e Outras Publicações', 'Finanças e Estatísticas de C&T', and 'Eventos', 'Listas de Discussão e Newsgroups'. On the left side, there are sections for 'Web-curriculo' and 'Deixe seu e-mail'. At the bottom, there are buttons for 'Sugira um Site', 'Livro de Visitas', and 'Informações Gerais'. A footer section includes 'Serviços do Prossiga' and a list of links for various services like 'Bases de Dados Brasileiras', 'Bibliotecas Virtuais Temáticas', etc.

Figura 43 – Biblioteca Virtual de Política Científica e Tecnológica<sup>30</sup>

Os diretórios e motores possuem uma base de dados que contém representações das páginas indexadas. Nos diretórios o trabalho de indexação é feito manualmente.

<sup>30</sup> Pode ser acessado em <http://www.prossiga.br/politica-ct/>



Figura 44 – Tela do diretório de busca do Yahoo! Brasil<sup>31</sup>

Vale salientar que os motores de busca varrem os *websites* da internet por meio de robôs que indexam, automaticamente, a informação coletada. Metaferramentas são softwares que realizam serviço de busca em outras bases de dados, por exemplo, nos motores de busca (Alencar, 2001).

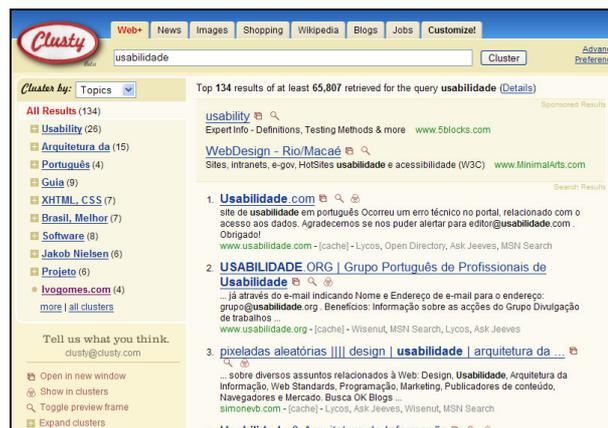


Figura 45 – Tela da metaferramenta Clusty<sup>32</sup>

Os mecanismos de busca se fizeram necessários a partir do momento em que houve exponencial aumento do conteúdo disponível na internet. Tal aumento tornou inviável a coleta manual de informações sobre *websites* e outras fontes. Conhecidos como robôs, essa tecnologia executa, de modo automático, varreduras periódicas e armazena os dados em grandes bancos de dados. Desta forma, se ganha em abrangência e em volume a ser recuperado. No entanto, decai a possibilidade de localizar itens relevantes, uma vez que tais robôs armazenam tudo o que encontram em suas varreduras. Percebe-se aí a

<sup>31</sup> Pode ser acessado em <http://br.yahoo.com/info/diretorio.html>

<sup>32</sup> Pode ser acessado em <http://www.clusty.com>

importância de haver um profissional de informação capacitado para avaliar os dados indexados.

### 5.1.

#### **Definições e perspectivas de sistemas de recuperação de informação**

Recuperação de informação é um termo amplo e ainda não plenamente definido, assim como o próprio termo "informação". Um sistema de recuperação de informação não informa o usuário sobre o assunto de seu questionamento, ou seja, não muda seu estado de conhecimento. Um sistema de recuperação de informação apenas informa sobre a existência ou não, e a localização de documentos relacionados à sua requisição. Isso exclui os sistemas de perguntas e respostas, além de os sistemas do tipo utilizados para acompanhamento de movimento em bolsas de valores.

Cabe, nesse momento, tratar das definições dos termos "dados" e "informação". Stair (1998) define dados como fatos em sua forma primária, por exemplo, nome de autor ou ano de publicação. Os dados representam as coisas no mundo real. Para Shedroff (2001), dado é matéria-prima, geralmente abundante no ambiente. Mesmo que tenha significado para o especialista, são, em sua maioria, os blocos sobre os quais se constrói a relevância.

Já informação é um conjunto de fatos organizados de tal forma que adquirem valor adicional, além do valor dos fatos em si. É necessário estabelecer regras e relações para transformar dados em informação útil e valiosa. Desta forma, entende-se que informação é um conjunto de dados dotados com relevância e propósito. O tipo de informação criada depende da relação definida entre os dados existentes. Transformar dados em informações é um processo, ou uma série de tarefas logicamente relacionadas e executadas para atingir um resultado definido. Para desenvolver esse processo é necessário conhecimento, ou seja, é necessário ter noção do corpo ou regras, diretrizes e procedimentos usados para selecionar, organizar e manipular os dados.

O diagrama a seguir (figura 46) apresenta as relações entre os conceitos de dado, informação, conhecimento e sabedoria como etapas sucessivas.



Figura 46 – Cadeia de conceitos de dados até sabedoria

Van Rijsbergen (1979) elaborou o quadro abaixo para distinguir as diferenças entre as propriedades da recuperação de dados e da recuperação de informação.

Quadro 2 – Propriedades da recuperação de dados e da recuperação de informação

	<b>Recuperação de dados</b>	<b>Recuperação de informação</b>
Adequação	Exata	Adequação parcial, melhor adequação
Inferência	Dedução	Indução
Modelo	Determinística	Probabilística
Classificação	Monotética	Politética
Linguagem de consulta <sup>33</sup>	Artificial	Natural
Especificação de questão	Completa	Incompleta
Itens desejados	Adequação	Relevante
Resposta ao erro	Sensível	Insensível

Na recuperação de dados, normalmente, a preocupação se concentra na adequação exata, ou seja, verifica-se se um item está, ou não está, presente em um arquivo. Na recuperação de informação, algumas vezes, isso pode ser interessante, mas de maneira mais geral o que se deseja é localizar itens que sejam parcialmente adequados à requisição e, a partir daí, selecionar dentre esses alguns poucos dos mais adequados.

<sup>33</sup> Traduziu-se por “consulta” o termo *query*, que se refere a um argumento de entrada no sistema, ou uma requisição precisa por informação, geralmente composta por palavras-chaves, operadores booleanos ou algum outro modificador.

A inferência utilizada em recuperação de dados é do tipo dedutivo. Na recuperação da informação é muito mais comum o uso da inferência indutiva e as relações somente são especificadas com um grau de certeza ou de incerteza e, por essa razão, nossa confiança nessa inferência é variável. Essa distinção conduz alguém a descrever a recuperação de dados como determinística e a recuperação de informação como probabilística. Na recuperação de dados se está interessado em classificação monotética, ou seja, uma classificação com classes definidas por objetos que possuem atributos tanto necessários quanto suficientes para pertencer a uma classe. Na recuperação de informação, tal classificação não é de todo útil. Na verdade, o que se deseja é uma classificação politética. Nesse tipo de classificação, cada indivíduo em uma classe possuirá somente uma proporção de todos os atributos característicos de todos os membros de determinada classe. Por isso, nenhum atributo é necessário nem suficiente para agregar um objeto a uma classe.

A linguagem de consulta para recuperação de dados, via de regra, será do tipo artificial, com sintaxe e vocabulário restrito. Na recuperação de informação, prefere-se utilizar linguagem natural, apesar de haver notáveis exceções. Na recuperação de dados a questão geralmente é uma especificação completa do que se busca. Na recuperação de informação, tal especificação é invariavelmente incompleta. Essa última diferença surge, em parte, do fato de que na recuperação da informação busca-se por documentos relevantes, em oposição a itens exatamente adequados. A extensão da adequação em recuperação de informação é assumida para indicar a probabilidade de relevância de determinado item. Uma consequência simples dessa diferença é que recuperação de dados é mais sensível ao erro no sentido de que, um erro na adequação não recuperará o item desejado, o que implica total falha do sistema. Na recuperação da informação pequenos erros na adequação, geralmente, não afetarão de maneira significativa a *performance* do sistema.

Wives (2004) destaca que informação é aquilo que um agente receptor (uma pessoa) recebe de um agente emissor em um processo de comunicação (em uma mensagem). Porém, a mensagem necessita ser, primeiramente, compreendida e, em seguida, identificada como contendo alguma coisa nova (para o receptor). Caso o receptor receba uma mensagem contendo algo que já conhece ou que não compreende, esta mensagem não lhe transmite nenhuma informação.

Com base em Stair (1998), é possível reconhecer as características da boa informação:

- Precisão: Não conter erros;
- Completude: Conter todos os dados importantes;
- Economia: Ser de produção econômica e oferecer boa relação entre a informação e o custo de sua produção;
- Flexibilidade: pode ser utilizada para diferentes finalidades;
- Confiabilidade: Ser obtida de fontes confiáveis e por meio de métodos de coleta igualmente confiáveis;
- Simplicidade: Ser facilmente compreensível e sem detalhes desnecessários;
- Atualidade: Ser enviada quando necessária e
- Verificabilidade: Poder ser verificada a fim de aferir sua correção.

Sistemas de informação podem ser entendidos, segundo a ótica de Lancaster e Fayen (1973), como interfaces entre um conjunto específico de usuários e o universo de recursos informacionais disponíveis com o objetivo de atender a certas necessidades predefinidas de um público preestabelecido, o que delimita de maneira razoavelmente clara e precisa o perfil de usuário. De acordo com os autores, os sistemas *on-line* para recuperação de informação permitem ao usuário acessar, de modo direto, uma base de dados de documentos por meio de um computador. Nesses casos, surgem como características a operação em tempo real e o compartilhamento entre duas ou mais atividades, o que permite que diferentes usuários acessem o sistema simultaneamente, sem conflito.

Alencar (2001) destaca que essas definições baseiam-se, principalmente, em sistemas de recuperação de informação fechados, em ambientes que permitem o controle do que é introduzido no sistema. Apesar de os sistemas estudados na presente investigação serem acessados *on-line* no ambiente *web*, não se trata de uma busca ampla, em toda rede, à maneira dos sistemas de busca que utilizam robôs para vasculhar grandes volumes em incontáveis *websites*. A informação que pode ser recuperada nos sistemas de bibliotecas *on-line* das universidades federais é proveniente de um banco de dados fechado alimentados pelo pessoal da própria instituição e o conteúdo, por maior que seja, está limitado ao acervo da biblioteca da universidade.

Para uma definição de sistema, Moraes e Mont'alvão (2003) ressaltam que esse é um conceito crítico sem o qual a Ergonomia não pode ser entendida. As autoras observam que um sistema é um conjunto de objetos juntos com as relações entre os objetos e entre os atributos relacionados uns com os outros e

com seu ambiente, de modo a formar um todo. A visão de Stair (1998) completa essa definição ao afirmar que esse conjunto possui objetos (elementos ou componentes) que interagem para se atingir objetivos. O que determina a forma de trabalho de um sistema são os objetos e as relações entre eles. Um sistema de informação é um tipo especializado de sistema, mas, como todos os sistemas, possui entradas, mecanismos de processamento, saídas e oferece respostas.

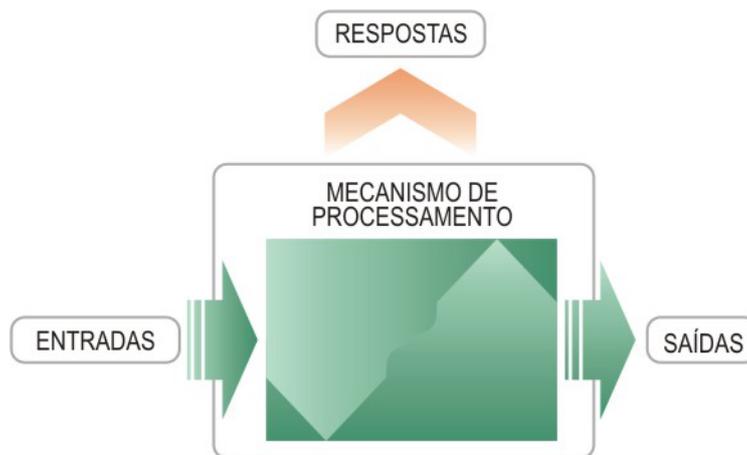


Figura 47 – Modelo de funcionamento de um sistema

Entradas: são os insumos fornecidos para que o sistema possa funcionar. Por entradas, deve-se entender somente os objetos que são processados pelo sistema para gerar uma saída em sintonia com a meta do sistema. Desta forma, os elementos que entram na produção, horas de trabalho, por exemplo, não podem ser computados como entrada. Nos sistemas de informação, as entradas envolvem atividades de captar e juntar dados primários.

Mecanismo de processamento: o mecanismo de processamento consiste de todas as etapas e recursos necessários para gerar as saídas. Nesse momento é que ocorrem as interações entre os diversos objetos ou elementos independentes que compõem o sistema. Também se pode falar em função do sistema, que é o objetivo de uma ação e implica na consecução de um requisito do sistema. A função do sistema é desempenhada por um subsistema ou componente. Em sistemas de informação, o processamento envolve a conversão dos dados e, informação útil (Moraes e Mont'alvão, 2003; Stair, 1998).

Saídas: são resultados do processamento das entradas, coerentes com a meta do sistema, e são lançadas no meio externo. No caso de sistemas de informação, a saída será composta por informações úteis na forma de relatórios, documentos etc.

Respostas: as respostas são saídas que servem para fazer ajustes ou modificações nas atividades de entradas ou de processamento. Desta forma, erros ou problemas podem fazer com que os dados de entrada sejam corrigidos ou que um processo seja modificado. As respostas são fundamentais para a tomada de decisão.

## 5.2. Arquitetura computacional para sistemas de busca

Outro aspecto importante nos sistemas de busca de bibliotecas diz respeito à arquitetura computacional. Considerando sua arquitetura, os sistemas de informação podem assumir, de maneira geral, uma das seguintes configurações (Senac, 1999):

Tempo compartilhado: esse modelo é baseado na arquitetura de *mainframe* e possui basicamente dois *hardwares*: a CPU<sup>34</sup> e os terminais. O processamento era centralizado no *mainframe* e todos os usuários utilizavam o mesmo tempo de CPU.

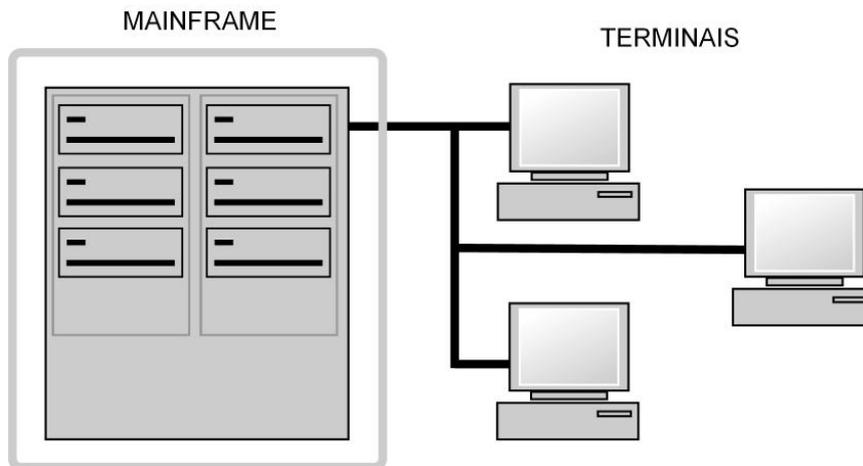


Figura 48 – Modelo de arquitetura de tempo compartilhado

Recurso compartilhado: o modelo de recurso compartilhado surgiu com a introdução das redes locais de microcomputadores, com capacidade própria de processamento e de armazenamento. Essa capacidade fez com que aumentasse a velocidade de execução das atividades, pois não havia o compartilhamento de tempo de CPU, já que o processamento se dava em cada estação.

<sup>34</sup> CPU: *central processing unit*, ou unidade central de processamento

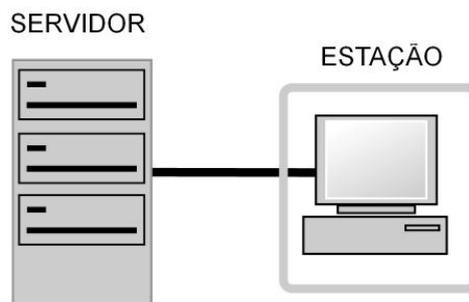


Figura 49 – Modelo de arquitetura de recurso compartilhado

Cliente-servidor: nesse modelo os dados ficam centralizados, o que garante sua integridade, e as tarefas são distribuídas entre cliente e servidores. Uma operação de consulta a um banco de dados cliente-servidor envolve as seguintes etapas:

1. o cliente faz uma consulta (*query*);
2. a consulta é enviada ao servidor (que faz parte da estrutura de retaguarda ou, às vezes, chamada de *back end*);
3. o servidor processa o pedido;
4. o servidor devolve o pedido;
5. a aplicação do cliente exibe o resultado.

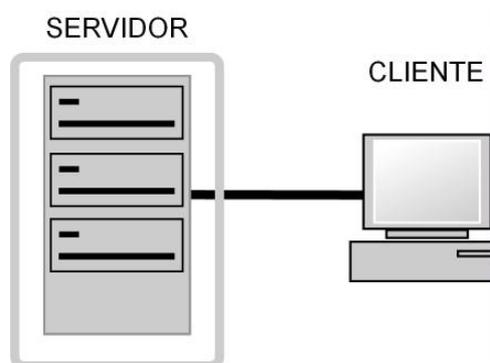


Figura 50 - Modelo de arquitetura cliente-servidor

De acordo com esse modelo, parte do processamento se dá no cliente e parte se dá no servidor. No cliente são processadas as telas e a carga de programas, enquanto no servidor é realizado o processamento dos dados, o que evita a replicação de dados em cada estação. O fato de vários clientes realizarem consultas simultaneamente pode aumentar, de forma considerável, o tráfego de dados pela rede. Tal situação pôde ser solucionada por meio do desenvolvimento do modelo em três camadas de *software*: camada de apresentação, camada de aplicação e camada do banco de dados. Por se tratar de camadas de *software*, podem residir em várias máquinas.

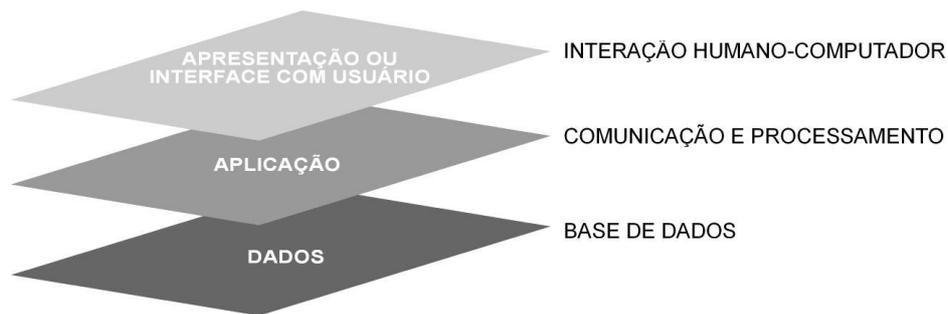


Figura 51 – Modelo em três camadas de software

Entende-se que a interface roda nas estações-clientes, com a qual o usuário entra em contato. A comunicação e processamento operam em um servidor de aplicação e os dados residem em um servidor de banco de dados. Desta forma, torna-se mais claro o escopo desta pesquisa, que é avaliar o nível de satisfação do usuário ao acessar o banco de dados por intermédio da camada de apresentação, onde ocorre a interação humano-computador.

### 5.3. Avaliação de sistema de informação

Lancaster e Fayen (1973, *apud* Alencar, 2001) afirmam que se pode avaliar o desempenho de um sistema de informação a partir do nível de satisfação do usuário no atendimento de sua necessidade. Na área de ciência da informação, os critérios mais utilizados para avaliar a satisfação do usuário com um sistema de informação são: revocação, precisão, cobertura, esforço do usuário, formato de saída, tempo de resposta e atualidade.

Revocação: a taxa de revocação se refere à relação entre o número de documentos relevantes recuperados e o número total de documentos relevantes contidos no sistema. Para medi-la é necessário conhecer o número total de documentos relevantes contidos no sistema.

Precisão: pode ser expressa como a taxa entre o número de documentos relevantes recuperados e o número total de documentos recuperados, sendo necessário avaliar a relevância dos documentos recuperados. Esse é um parâmetro fundamental para a avaliação de sistemas de busca. Para avaliar a precisão é importante que o sistema informe o número total de documentos recuperados.

Cobertura: é a capacidade de recuperar todo o volume de informações disponíveis sobre determinado tópico. No entanto, a satisfação do usuário será

delimitada pelos parâmetros de revocação e de precisão. A abrangência é um fator de avaliação da cobertura, pois está relacionada ao volume de informações indexadas pelo sistema. Conhecer o escopo do sistema também é importante, pois isso influencia a cobertura do sistema. Por exemplo, saber se o sistema de recuperação de uma biblioteca *on-line* cobre um acervo especializado em uma área específica de conhecimento ou o acervo de uma biblioteca universitária.

Esforço do usuário: este parâmetro, originalmente, teve foco no treinamento dos usuários e no projeto do sistema, de maneira a minimizar erros do usuário. No entanto, sua abrangência pode ser ampliada para todos os fatores que servem para facilitar a operação do sistema. Neste parâmetro, se enquadram os aspectos de Ergonomia e de usabilidade da interface, pois essa é a região do sistema com que o usuário entra em contato durante a operação, além das opções oferecidas ao usuário. Este critério pode ser avaliado segundo os fatores documentação do sistema, interface com o usuário e capacidades de busca.

1. Documentação do sistema: para sistemas baseados na *web*, a documentação pode estar disponível na forma de ajuda *on-line*. Essa ajuda deve apresentar instruções para que o usuário consiga por si só entender o funcionamento e solucionar possíveis problemas.
2. Interface com o usuário: a interface é o ponto de contato entre o sistema e os usuários, que certamente possuem diferentes níveis de experiências. Por isso, deve oferecer recursos para a realização da tarefa de maneira eficiente e sem erros. Também é fundamental seguir os critérios de Ergonomia e de usabilidade a fim de facilitar o uso e auxiliar a recuperação e armazenagem da informação encontrada.
3. Capacidade de busca: refere-se ao relacionamento da necessidade do usuário e a base de dados do sistema. Pode ser dividida nos seguintes fatores:
  - 3.1. Relacionamento entre termos: é a facilidade do sistema em fazer conexões entre as palavras-chave especificadas pelo usuário. Isso pode ser feito por meio de operadores booleanos, pela proximidade entre termos, por meio da linguagem natural ou de vocabulário controlado.
  - 3.2. Interpretação de uma única palavra: são os recursos disponíveis pelo sistema para auxiliar o usuário no uso de uma palavra-chave para busca. Tais recursos dividem-se em: truncagem, distinção entre maiúsculas e minúsculas, delimitação de campo, eliminação de palavras

não significativas, atribuição de pesos a termos, incorporação automática de sinônimos.

**Formato de saída:** esse aspecto tem fundamental importância para o julgamento da precisão, pois é o ponto inicial de contato entre o usuário e a informação recuperada. Quanto mais informação oferecida sobre a representação do documento, mais fácil se torna prever sua relevância. Para o estudo das interfaces de sistemas de recuperação de informações, este é um parâmetro crucial para avaliar a satisfação do usuário. O formato de saída tem crucial importância para a facilitação ou não do esforço do usuário na obtenção dos itens buscados. Este critério pode ser dividido em dois fatores: critério de ordenação e formato de exibição.

**Critério de ordenação:** está relacionado com a previsão de relevância que o sistema adota. No entanto, pode ser altamente positivo possibilitar que o usuário possa escolher qual critério de ordenação adotar, de acordo com seu interesse.

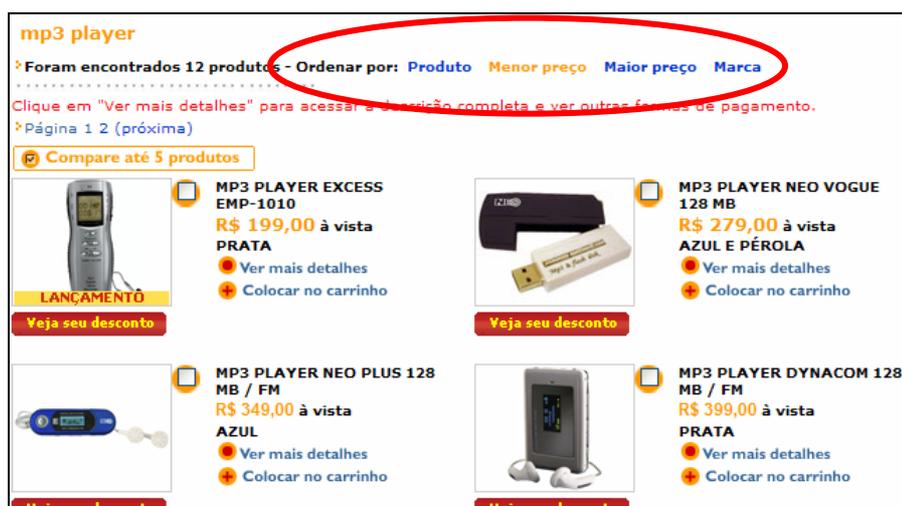


Figura 52 – Recurso de ordenação de produtos (www.pontofrio.com.br)

**Formato de exibição:** em sistema de bibliotecas *on-line*, os resultados apresentados podem incluir a bibliografia ou ainda um resumo do documento recuperado. O formato de exibição também tem influência direta sobre a relevância, pois auxilia a perceber a adequação dos resultados da busca a uma necessidade específica. É importante apresentar a ocorrência do termo pesquisado entre os resultados recuperados, o que, geralmente, é feito por meio de destaques como negrito, cor diferente de texto e outros. Alencar (2001) destaca, por sua vez, outras opções para o formato de exibição:

- número total de registros recuperados;

- possibilidade de determinar a quantidade de registros exibidos por página;
- possibilidade de refinar o resultado obtido, por exemplo, por meio de uma busca avançada e
- armazenagem da busca para consultas posteriores.

Tempo de resposta: apesar de ser um fator importante para avaliar sistemas de informação, para sistemas acessados pela internet, torna-se sujeito a inúmeras variáveis, como velocidade de acesso, tamanho da banda, equipamento utilizado pelo usuário, entre outros fatores. No entanto, é importante oferecer respostas sobre o processamento das entradas para manter o usuário informado a respeito dos resultados de suas ações.

Atualidade: define a proporção de novos documentos recuperados para o requisitante. Esse critério pode incorporar a taxa de atualidade entre o surgimento de um novo documento e sua inclusão no banco de dados do sistema. Para uma biblioteca *on-line*, a taxa de atualidade é outro fator crucial para a satisfação

Desta forma, Alencar (2001) propõe o seguinte modelo para avaliar a recuperação de informações na *web*.

- **Precisão**
- **Cobertura**
- **Esforço do usuário**
  - Documentação do sistema
  - Interface com o usuário
  - Capacidade de busca
    - *Relacionamento entre termos*
      - Operadores booleanos
      - Proximidade entre termos
      - Linguagem natural
      - Vocabulário controlado
    - *Interpretação de uma única palavra*
      - Truncagem
      - Distinção entre maiúsculas e minúsculas
      - Delimitação de campo
      - Eliminação de palavras não significativas
      - Atribuição de pesos a termos
      - Incorporação automática de sinônimos

- **Formato de saída**
  - Critério de ordenação
  - Formato de exibição
- **Atualidade**

#### **5.4. Estratégias de busca**

Lopes (2002) define uma estratégia de busca “como uma técnica ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada em uma base de dados” (p.41). A autora destaca que mesmo após diversos esforços realizados por projetistas de interfaces, “o processo de busca continua sendo um fator de dificuldade que ainda não foi minimizado pelas novas tecnologias disponíveis” (p. 42). No entanto, a estratégia de busca precisa refletir a necessidade de informação do usuário. Todas as estratégias de busca são baseadas na comparação entre a pergunta e o documento armazenado. Às vezes, essa comparação é alcançada, indiretamente, quando a pergunta é comparada com blocos<sup>35</sup>, ou, mais precisamente, com os perfis que representam os blocos. A distinção feita entre diferentes tipos de estratégias de busca, por vezes, pode ser entendida por meio da linguagem de consulta, que é a linguagem pela qual a necessidade de informação é expressa. Desta forma, a natureza da linguagem de consulta dita a natureza da estratégia de busca (van Rijsbergen, 1979).

O trabalho de indexação é parte fundamental dos sistemas de recuperação de informação, pois é a partir daí que serão estabelecidos os pontos de acesso do usuário ao conteúdo das bases de dados. Segundo Lancaster (1993), “o propósito principal da elaboração de índices e resumos é construir representações de documentos publicados numa forma que se preste a sua inclusão algum tipo de base de dados”. Os termos utilizados na indexação serão, com frequência, extraídos de algum tipo de vocabulário controlado, como um tesouro, mas também podem ser termos livres, extraídos do próprio documento. Lancaster representa, por meio da figura 53, a função da elaboração de índices e resumos no quadro mais amplo da recuperação da informação:

---

<sup>35</sup> *clusters* (Sawaya, 1999)

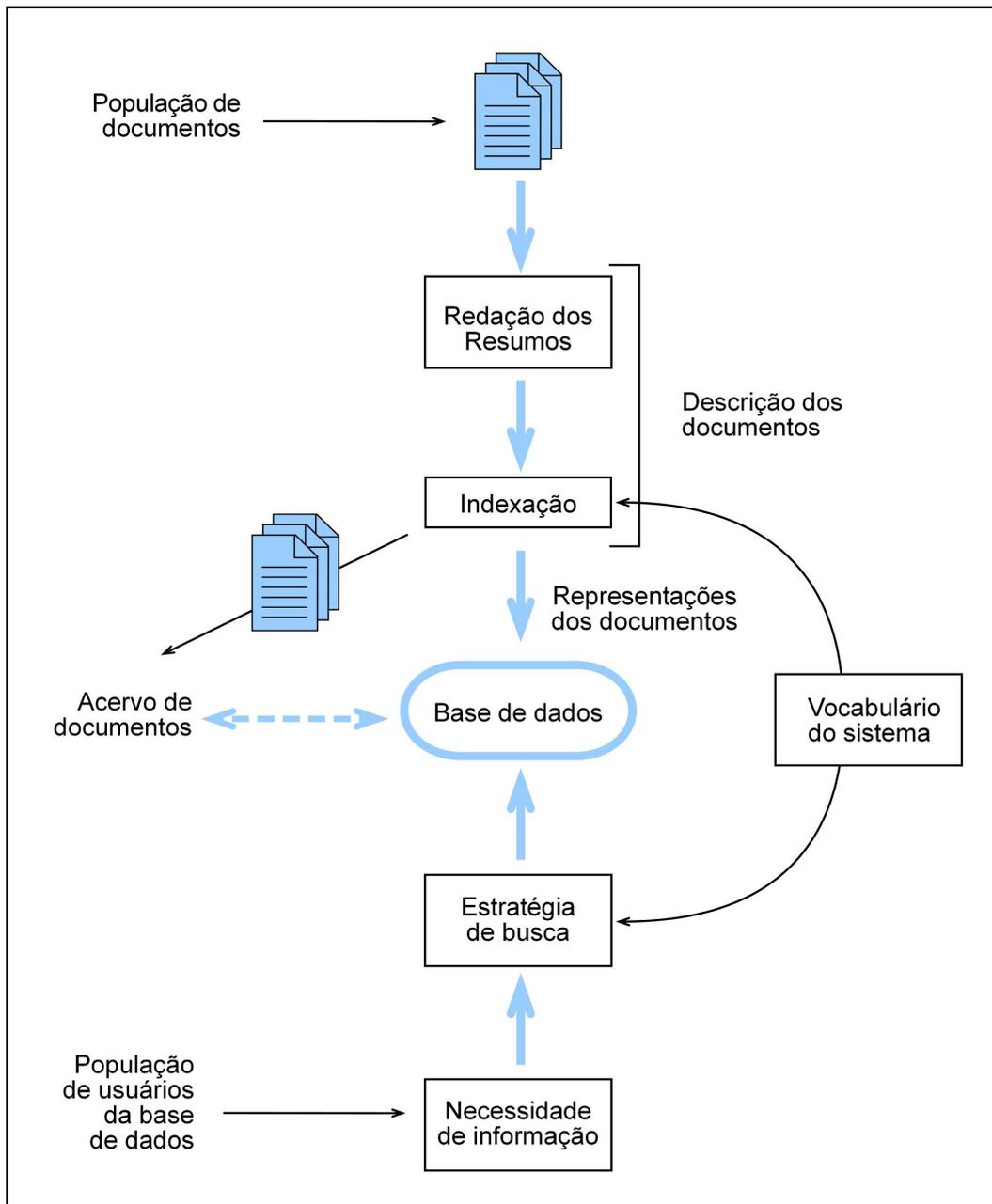


Figura 53 – Índices e resumos na recuperação da informação

Lopes (2002) define linguagem natural como sendo a linguagem do discurso técnico-científico e cita Lancaster (1993) para destacar que “a expressão normalmente se refere às palavras que ocorrem em textos impressos, considerando-se como seu sinônimo a expressão ‘texto livre’” (p. 42). Lancaster (1993) considera que linguagem natural pode ser entendida como discurso comum, que é a linguagem utilizada, de modo habitual, na escrita e na fala, configurando o contrário de vocabulário controlado. Um texto livre pode ser um título, um resumo, um extrato ou mesmo o texto integral da publicação. Lancaster observa, no entanto, que mesmo a expressão texto livre é empregada

para designar palavras ou expressões extraídas do texto por um indexador e incluídas em um registro bibliográfico que representa o texto. As consultas em linguagem natural permitem uma declaração longa ou uma questão que descreva a informação que se deseja encontrar.

Em contraponto, há a chamada "linguagem controlada" ou "vocabulário controlado", que pode ser definida como um conjunto limitado de termos autorizados para uso na indexação e busca de documentos.

Entende-se como representação da informação a substituição de uma entidade lingüística longa e complexa, como todo o texto de um documento, por sua descrição abreviada, com ênfase ao que é essencial no documento, com vistas a sua recuperação em um sistema de informação. No âmbito da Ciência da Informação, comumente são utilizadas as linguagens formalizadas. Tais linguagens facilitam a comunicação entre produtores de informação (autores), organizadores da informação (bibliotecários/arquivistas) e entre usuários da informação (pesquisadores). Dentre essas linguagens encontra-se o vocabulário controlado que é uma lista de termos, conhecidos como "descritores" ou "palavras-chave". Um vocabulário serve como fonte básica da linguagem documentária para indexação e recuperação de documentos que integram um determinado sistema de informações bibliográficas. Tem por finalidade ser um instrumento responsável pela eficácia das ações de transferência da informação e ser um dispositivo para criação de uma linguagem de indexação e recuperação única e padronizada para todos os sistemas de informações bibliográficas de uma entidade.

Um vocabulário controlado inclui uma forma de estrutura semântica que se destina, especialmente, ao controle de sinônimos, à diferenciação de homógrafos e também à reunião ou ligação de termos cujos significados apresentam uma relação estreita entre si. Os vocabulários controlados são construídos a partir de termos que ocorrem em determinados recortes da literatura científica e tecnológica. Um vocabulário controlado é um organismo vivo, em permanente mutação, dada a evolução das áreas de conhecimento. É um sistema aberto e em constante evolução, na medida em que novos assuntos de interesse da entidade forem sendo tratados. A figura 54 representa a tela inicial do *website* UK Archive Thesaurus, vocabulário controlado para consulta *on-line* com as relações entre termos de língua inglesa aprovados pela comissão da Unesco.

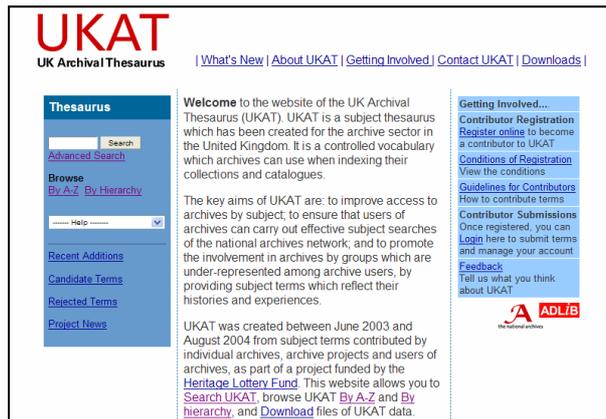


Figura 54 – Tela inicial do UKAT<sup>36</sup>



Figura 55 – Tela de pesquisa do UKAT

Em sua revisão de literatura, Lopes (2002) constata que a recomendação geral por parte dos autores pesquisados é de que sejam implementadas estratégias de busca que conjuguem os dois tipos de linguagens: a controlada e a natural. Esta abordagem permitirá maior efetividade por parte dos sistemas de busca em linha. A autora afirma, ainda, que é necessário investir no tratamento técnico dos recursos informacionais, assim como na sua organização, de maneira mais conveniente, visando a racionalização de sua estocagem e, naturalmente, a busca e a recuperação de informação de maneira eficaz e eficiente. Tal preocupação se dá pelo fato de que nada adianta arquivar um documento que não se sabe como encontrar, por não ter sido indexado ou, ainda, ter sido indexado de maneira incorreta. Para Gardin (1974, *apud* Pinto, 2001) a indexação é um conjunto de atividades que consiste em identificar, nos documentos, os seus traços descritivos (TDs) ou macroproposições e, em seguida, extrair os elementos/descriptores (sintagmas) indicadores do seu

<sup>36</sup> Pode ser acessado em <http://www.ukat.org.uk/>

conteúdo, visando a sua posterior recuperação. Esses descritores vão se constituir na representação, pois esta só pode ser pelo próprio documento.

Pinto (2001) destaca que a indexação - seja manual (feita por seres humanos), automática (também chamada de "mecânica", realizada pelas ferramentas de informática) ou semi-automática (misto de humana e mecânica) - deve ser calcada tanto em um objetivo teórico – estabelecer mecanismos para a elaboração de índices – quanto em um objetivo operacional – possibilitar a busca e a recuperação da informação.

Rosenfeld e Morville (1998) observam que os cientistas da informação e bibliotecários, há muito tempo, têm estudado os hábitos de localização de informação dos usuários. Até recentemente, tais estudos pertenciam aos sistemas de informação tradicionais. Na atualidade, os estudos relacionados aos hábitos de busca dos usuários também interessam aos projetistas de interfaces para mídia digital. O fato principal a ser considerado é que usuários são diversos entre si. Cada pessoa desejará buscar uma informação de modo diferente de uma outra e essas variações se referirão à quantidade de informação necessária, ao nível de profundidade e extensão do assunto coberto pelo conteúdo a ser recuperado, além do tempo disponível para realizar a busca, e, fundamentalmente, a confiabilidade tanto da fonte de consulta e quanto do resultado atingido com a busca.

Uma página de informação é fácil de explorar, mas quando a representação da informação tem o tamanho de um livro ou de uma biblioteca, se torna difícil localizar itens conhecidos ou folhear a fim de ter uma noção geral (Shneiderman, 1996). Para o autor, a busca de informação pode assumir características distintas. Com relação às suas ações, o usuário pode procurar por um fato específico; procurar por fatos relacionados; esquadrinhamento<sup>37</sup> ilimitado; explorar disponibilidade.

Assim, as necessidades dos usuários variam, enormemente e os sistemas de informação que os atendem, devem reconhecer, diferenciar e atender a essas diferentes necessidades. As principais estratégias de busca podem ser sumarizadas em: busca por item conhecido, busca pela existência, busca exploratória e pesquisa.

A busca por item conhecido é mais freqüente quando o usuário tem, de forma clara, suas necessidades definidas e estas possuem uma resposta única e correta. Como exemplo, toma-se a consulta à cotação de uma moeda

---

<sup>37</sup> *browsing*.

estrangeira em um periódico sobre economia. Neste caso, sabe-se o que se busca e onde pode ser encontrado. A busca pela existência, por sua vez, acontece quando o usuário sabe o que deseja, porém não sabe como descrevê-lo. Ao invés de uma questão clara que possui uma resposta certa, se tem um conceito ou uma idéia abstrata. Já na busca exploratória, alguns usuários sabem como redigir suas questões, mas não sabem exatamente o que esperam encontrar, estão apenas explorando e tentando aprender mais. Em qualquer caso, não se tem clareza do que se vai descobrir, mas se deseja gastar tempo para aprender mais.

De maneira geral, as formas como o usuário realiza buscas são influenciadas pelos recursos disponíveis no sistema para realização da tarefa. As capacidades de busca têm por objetivo adequar a necessidade de informação de um usuário aos itens de uma base de dados que a responderão (Hahn, 1998). As capacidades de busca podem ser de dois tipos: as que ajudam a especificar a relação entre termos em uma declaração de busca e as que facilitam a interpretação de uma palavra em particular.

As capacidades, que ajudam a especificar a relação entre termos, são os operadores booleanos (AND, OR, NOT), que permitem configurar intersecção, união ou diferenças. Uma estratégia booleana de busca recupera os documentos que sejam verdadeiros para a consulta. Essa formulação somente faz sentido se as consultas forem expressas em termos de índice, de palavras-chaves combinadas pelos conectores lógicos mais comuns: AND, OR ou NOT. Também, os operadores de proximidade (NEAR) permitem que duas ou mais palavras sejam buscadas como uma unidade singular ou uma expressão. Assume-se que, quanto mais próximos, estiverem dois termos em um texto, tanto mais estarão relacionados ao mesmo conceito de busca.

Por sua vez, foram desenvolvidas diversas capacidades que facilitam a interpretação de uma palavra em particular. O truncamento - que é a busca em fragmentos de palavras, tais como: sufixos ou prefixos - permite busca e recuperação com base em parte de uma palavra. Lancaster (1993) afirma que o truncamento talvez seja o recurso mais poderoso das buscas em linguagem natural. Programas de computadores permitem a busca em quaisquer fragmentos de palavras, eliminando-se prefixos, sufixos ou mesmo infixos. No entanto, o autor destaca que este tipo de busca se mostra mais adequado para buscas em bases de dados de ciência e tecnologia. O truncamento pode ser representado com o uso do cifrão após a raiz de um termo. Deste modo, todos os registros que contenham qualquer palavra com a raiz fornecida serão

recuperados. Na busca com a expressão "ECONO\$" poderão ser recuperados os registros que contenham as palavras: ECONOMIA, ECONOMIAS, ECONÔMICO, ECONÔMICOS, ECONOMICAMENTE.

A eliminação de palavras aumenta a velocidade da busca ao tornar algumas palavras "invisíveis" para o sistema. A incorporação automática de sinônimos na formulação da busca aumenta a revocação, sem que o usuário tenha que pensar em todos os sinônimos. A distinção entre caixa alta e caixa baixa aumenta a precisão da busca ao diferenciar, por exemplo, o sobrenome Carvalho do substantivo carvalho. A habilidade de requisitar que o sistema encontre termos semelhantes permite ao usuário que recuperou, pelo menos, um registro relevante, identifique documentos similares sem ter que entrar com outra estratégia de busca. A lógica difusa<sup>38</sup> oferece a habilidade de localizar palavras e conceitos similares ao termo de busca e é usado para compensar erros ou variações na escrita das palavras.

Ao se considerar que as buscas, usualmente, recuperam itens irrelevantes, as capacidades de navegação e esquadramento colaboram para dar foco aos itens que tenham maiores possibilidades de ir ao encontro das necessidades de informação. A capacidade de ordenar e de indicar relevância embasa a ordem de saída em valores preditos de relevância, ou seja um resultado que pretende refletir a proximidade de um item recuperado ser relevante para a requisição do usuário. O zoneamento é a capacidade de apresentar uma porção-chave de cada registro recuperado, que caiba em uma tela, e que seja suficiente para que se possa tomar uma rápida decisão de relevância. O uso de realces é a capacidade de apresentar nos registros recuperados em destaques as palavras iguais ao termo buscado. O realce dá ao usuário uma indicação do motivo um item foi selecionado e em qual contexto os registros recuperados são adequados às necessidades do usuário. A possibilidade de escolher quantos resultados devem ser exibidos é uma característica que tenta acomodar a preferência pessoal de cada usuário para a quantidade de informação desejada sobre um assunto.

---

<sup>38</sup> *fuzzy*

## 5.5.

### Estudo sobre interfaces para sistemas de busca

Peterson (1998) destaca que, dada à imensa base de dados de documentos em que se configura a *web*, é essencial examinar o projeto de interfaces de sistemas de busca a partir de um ponto de vista da usabilidade a fim de torná-los efetivos para o usuário típico. De acordo com Shneiderman (1998), os projetistas de interfaces possuem cinco tipos de estilos de interação para escolher ao projetar uma interface com usuário: linguagem de comandos, seleção de menus, preenchimento de formulários, linguagem natural e manipulação direta.

Cada um desses têm suas respectivas vantagens e desvantagens que devem ser consideradas ao se projetar interfaces para recuperação de informação com usabilidade. Como Shneiderman aponta: quando os usuários utilizam entrada de dados por linha de comando estão ativamente iniciando a operação, pelo fato de que devem lembrar a notação e dar entrada ao comando diretamente. Interfaces baseadas em menus exigem que os usuários sejam mais passivos, uma vez que escolhem dentre um conjunto de opções circunscritas. Preenchimento de formulários, da mesma forma, apresenta opções passivas ao usuário, mas talvez seja ainda mais constrangedora do que as interfaces baseadas em menus, já que cada caixa de entrada do formulário possui somente um tipo de dado possível e não oferece alternativas ao usuário. Por outro lado, alguns mecanismos de busca têm implementado interfaces de linguagem natural, à medida que os estudos sobre o interesse do usuário por esse tipo de solução avançam.

Uma das grandes desvantagens da entrada de dados baseada em comandos é que o usuário novato ou casual não tem consciência das opções de comandos disponíveis, seja porque nunca as viu ou por tê-las esquecido devido ao uso pouco freqüente. Essa situação não é amenizada quando o mecanismo de busca requer que o usuário vá uma outra página para buscar ajuda e muitas dessas páginas são difíceis de navegar e pobremente organizadas.

Em geral, um aumento da possibilidade de aprendizagem de uma interface aumentará, concomitantemente, a capacidade de ser lembrada, apesar de serem conceitos diferentes. Contudo, ao se tratar adequação a tarefas simples, interfaces facilmente aprendidas também podem tornar-se bastante fáceis de lembrar.

Desta forma, o objetivo do projetista de interfaces para mecanismos de busca deve ser criar interfaces que possam ser, instantaneamente, aprendidas e que requeiram um uso mínimo de telas de ajuda, minimize os erros e auxilie o usuário a formular questões complexas de busca ao mesmo tempo em que aumentará a precisão da busca.

Peterson (1998) realizou um experimento em que foram testadas quatro variações de tela de mecanismo de busca com a intenção de verificar qual tipo de interface de busca era preferido pelos usuários. Os tipos de interfaces avaliadas foram:

Interface A: Menu-conector – Composta por quatro campos de entrada de texto que são ligadas por um menu cascata com as opções AND, OR, NOT ou NEXT TO. A interface não foi baseada em nenhum mecanismo de busca existente conhecido do autor.

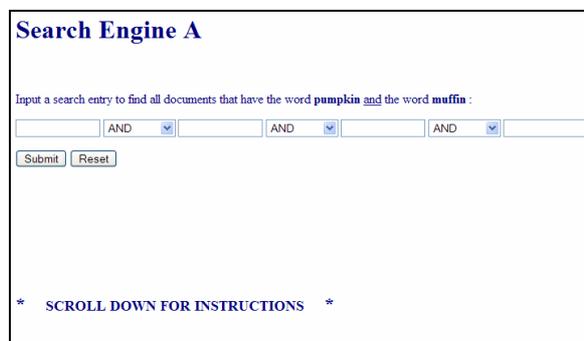


Figura 56 – Tela de busca da interface A<sup>39</sup>

Interface B: Preenchimento de formulário – Possui quatro caixas de texto que são rotuladas de acordo com um operador booleano específico.

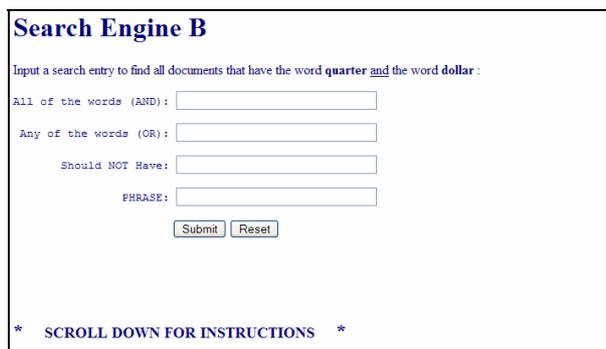
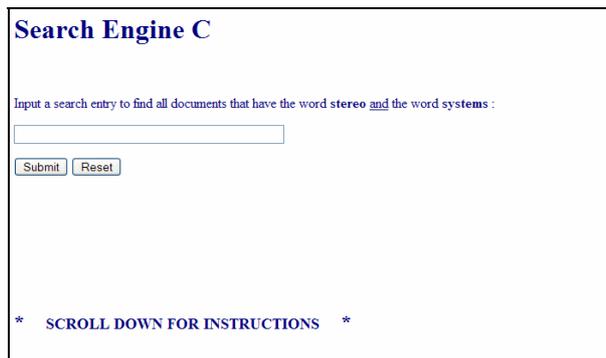


Figura 57 – Tela de busca da interface B<sup>40</sup>

<sup>39</sup> Disponível em <http://ella.slis.indiana.edu/g/bpm//scholarship/peterson/engineA.html>)

<sup>40</sup> Disponível em <http://ella.slis.indiana.edu/g/bpm//scholarship/peterson/engineB.html>

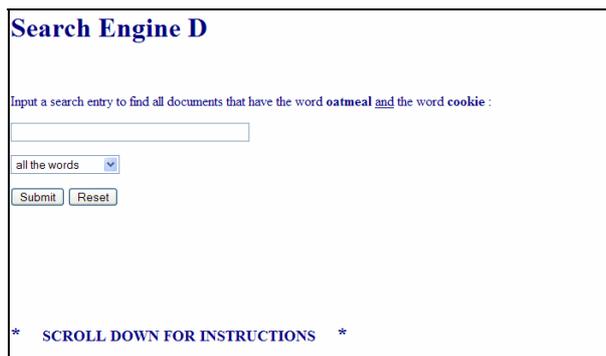
Interface C: Linha de comando – Possui somente um campo simples de texto e é a referência óbvia das interfaces típicas de sistemas de busca.



The screenshot shows a web page titled "Search Engine C". Below the title, there is a line of text: "Input a search entry to find all documents that have the word stereo and the word systems :". Below this text is a single text input field. Underneath the input field are two buttons: "Submit" and "Reset". At the bottom of the page, there is a line of text: "\* SCROLL DOWN FOR INSTRUCTIONS \*".

Figura 58 – Tela de busca da interface C<sup>41</sup>

Interface D: Menu simples – Possui uma caixa de texto em conjunção com um menu em cascata que permite que o usuário selecione uma das quatro opções de operadores booleanos.



The screenshot shows a web page titled "Search Engine D". Below the title, there is a line of text: "Input a search entry to find all documents that have the word oatmeal and the word cookie :". Below this text is a single text input field. Underneath the input field is a dropdown menu with the text "all the words" and a small downward arrow. Below the dropdown menu are two buttons: "Submit" and "Reset". At the bottom of the page, there is a line of text: "\* SCROLL DOWN FOR INSTRUCTIONS \*".

Figura 59 – Tela de busca da interface D<sup>42</sup>

O estudo de Peterson foi realizado com estudantes universitários dos cursos de Biblioteconomia e de Biologia. Os resultados dos questionários aplicados apontaram, no geral, que estudantes de Biologia utilizam menos mecanismos de busca na *web* do que os de Biblioteconomia. Além disso, o uso geral estimado de operadores booleanos para estudantes de Biblioteconomia se confirmava em cerca de 50% das vezes, enquanto para estudantes de Biologia esse percentual era um pouco menos que isso. Um estudante de Biologia respondeu que nunca utilizara operadores booleanos em suas buscas, ao passo que um de Biblioteconomia afirmou utilizar em todas as buscas realizadas em

<sup>41</sup> Disponível em <http://ella.slis.indiana.edu/g/bpm//scholarship/peterson/engineC.html>

<sup>42</sup> Disponível em <http://ella.slis.indiana.edu/g/bpm//scholarship/peterson/engineD.html>

mecanismos de busca na *web*. É claro que existe uma diferença entre estudantes das duas áreas, apesar da distribuição geral das respostas terem sido próximas para os dois grupos.

Peterson (1998) obteve quatro importantes conclusões em seu estudo:

1. os critérios definidos pela literatura (Nielsen, 1993) para tornar um sistema fácil de lembrar para uso casual se aplicam a uma grande proporção de usuários de mecanismos de busca;
2. muitos usuários sem treinamento em sistemas de recuperação de informação não sabem como construir questões simples com operadores booleanos nos atuais sistemas de busca na *web*;
3. usuários que têm experiência com uso de operadores booleanos adotam prioritariamente o uso da notação “NOT” em detrimento da “-” e
4. a área de estudo e de formação implica diretamente na experiência de uso de mecanismos de busca (por exemplo, entre estudantes de Biblioteconomia e de Biologia).

Os resultados de performance apresentaram a interface A como a mais adequada para realização de buscas nas *web*, além de ser também a preferida pelos participantes do teste. A interface A foi a de uso mais eficiente para todos os participantes, além apresentar menor taxa de erros. Também foi a mais fácil de aprender e, certamente, a mais fácil de recordar. A interface B, apesar de não ter obtido tão grande destaque nos resultados, também apareceu como uma solução com boa usabilidade.

Todos os erros associados à interface C envolviam o uso de operadores e os sujeitos que cometeram tais erros não consultaram às instruções. Cada pessoa que utilizou a notação correta consultou as instruções. É um achado importante, pois implica que o uso de notações simbólicas não é intuitivo para os usuários e que sistemas de busca projetados como interfaces de linhas de comando devem evitar as notações “+” e “-”, ou possibilitar o uso de palavras e notações simbólicas.

Empresas que desenvolvem e comercializam mecanismos de busca, geralmente, lançam foco em questões técnicas. Por exemplo, procuraram melhorar a relevância e a atualidade de artigos recuperados e o número dos artigos posicionados (Xie *et al.*, 1998). Do mesmo modo, as avaliações dos sistemas de busca tendem a se centrar nestes aspectos. À medida que aumenta a qualidade técnica dos sistemas de busca, diminuem as diferenças entre os

diversos produtos, do mesmo modo que a experiência passa a dominar as preferências do usuário (Norman, 1999).

A maioria dos desenvolvedores de sistemas de busca tem se esforçado para melhorar a experiência do usuário, ordenando os resultados recuperados, categorizando os resultados em grupos, oferecendo opções de mudança de idioma e apresentando interfaces de busca avançada mais explícitas. Por outro lado, a maioria dos sistemas de busca oferece uma interface de linha de comando em seu modo de busca simples ou padrão.

Algumas pesquisas têm mostrado que usuários do *website* de busca entrevistados encaram a "facilidade de uso" como uma determinante muito importante da satisfação com as interfaces. Recentemente, a maioria dos sistemas de busca tem se esforçado em melhorar a experiência do usuário, por meio da ordenação dos resultados recuperados, da categorização de resultados em grupos, pela possibilidade de construir questões em linguagem natural e pela apresentação de interfaces de busca avançada mais explícitas.

## **5.6. Processo humano de busca por informação**

Marchionini (1995) afirma que muito da existência humana é caracterizada pela noção de busca, seja por objetos ou por sensações. Informação, de maneira geral, é qualquer coisa que pode mudar o conhecimento de alguém. Busca de informação, segundo o autor, é o processo no qual as pessoas, propositadamente, se envolvem a fim de mudar seu estado de conhecimento. Lopes (2002) cita Spink e Saracevic (1993) para dizer que a busca em sistemas de recuperação de informação é um processo de alta complexidade que envolve numerosos fatores e variáveis, além de decisões e do entrelaçamento dos subprocessos inter-relacionados com a busca. Kuhlthau (1991, *apud* Lima, 2001) destaca que o processo de busca pode ter foco em três aspectos: o afetivo, o cognitivo e o físico. Tais aspectos estão presentes em todos os estágios do processo de busca que, segundo Lima (2001), podem ser enumerados em:

1. inicialização;
2. seleção;
3. exploração;
4. formulação;
5. coleta e
6. apresentação.

A busca de informação começa com o reconhecimento e aceitação do problema e continua até que o problema seja resolvido ou abandonado. O processo de busca de informação é composto por um conjunto de subprocessos. Marchionini (1995) os representa com a figura 60.

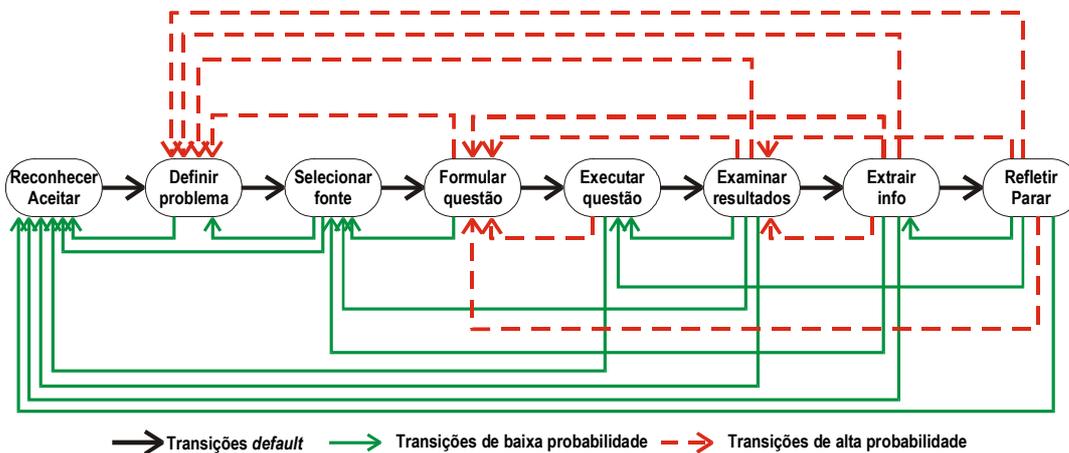


Figura 60– Subprocessos de busca de informação

Esses subprocessos podem tornar-se padrão para fases ou passos num algoritmo seqüencial, mas são melhor considerados como funções ou módulos de atividade que podem ser chamados à ação recursivamente a qualquer momento, podem estar ativos, de forma contínua; temporariamente paralisados; enquanto outros são processados e que podem chamar para outros subprocessos. Dessa forma, o processo de busca de informação pode ser efetuado ao longo de linhas paralelas de progresso e aproveitar as oportunidades que surjam dos resultados aleatórios ou intermediários.

O processo de busca de informação é dinâmico e orientado à ação. A figura 61 (Marchionini, 1995, *apud* Santos, 2000) ilustra alguns dos paralelismos e descreve três classes de subprocessos de busca de informação: entender; planejar e executar; avaliar e usar.

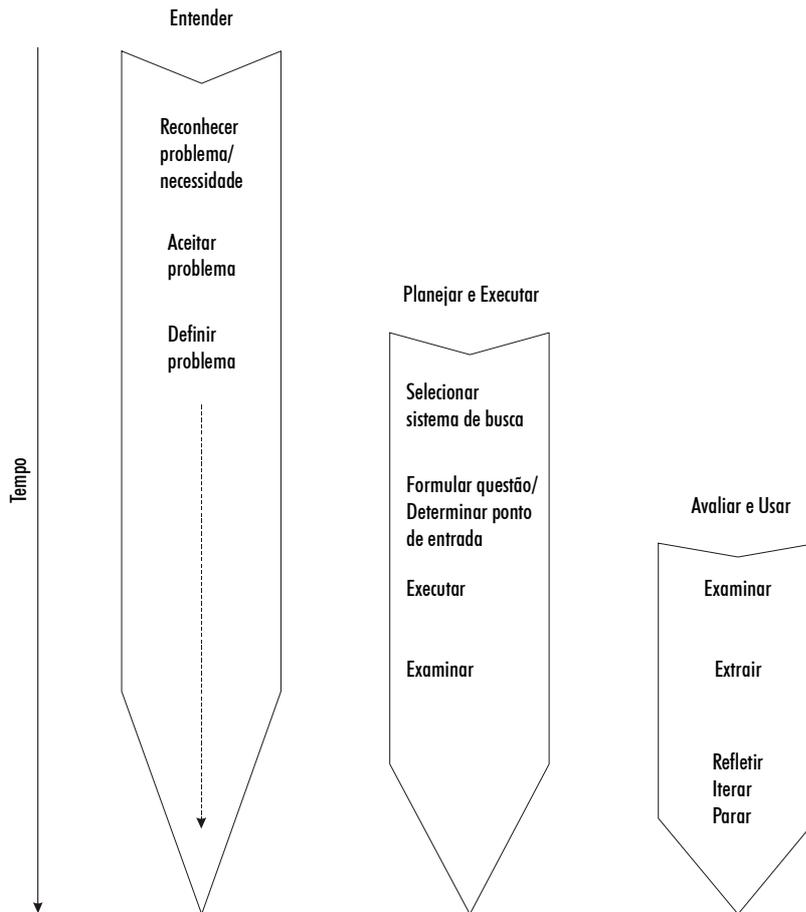


Figura 61 – Paralelismos e subprocessos da busca de informação

Pelo fato de esses subprocessos serem controlados pelo usuário, é bem maior a probabilidade de tomarem caminhos gerados pelo momento, de acordo com suas habilidades e experiência. Esses caminhos dependem de julgamentos sobre a relação custo/benefícios do progresso alcançado, redefinições dos objetivos da tarefa, e julgamentos de relevância da informação recuperada. Os sistemas de busca eletrônicos têm substanciais impactos em muitos dos subprocessos, especialmente na formulação da questão e no exame dos resultados. Os sistemas de busca interativos em base de dados de texto têm contribuído para fazer desaparecer os limites de separação entre os subprocessos.

Uma página de informação é fácil de explorar, mas quando a representação da informação tem o tamanho de um livro ou de uma biblioteca, ou ainda mais, torna-se difícil localizar itens conhecidos ou folhear a fim de ter uma noção geral (Shneiderman, 1998).

Para Shneiderman (1998), a busca de informação pode assumir características distintas. Com relação às suas ações, o usuário pode procurar

por um fato específico (o telefone do Presidente da República); procurar por fatos relacionados (quais outros livros do autor de *Policarpo Quaresma*); folheamento ilimitado (há relação entre monóxido de carbono e níveis de desertificação?); explorar disponibilidade (quais informações estão disponíveis sobre Aids, diabetes, doenças mentais?). O autor afirma que, quando a tarefa envolve a busca de informação, os focos são os objetos de informação e suas estruturas, além das ações de informação para navegação.

## 5.7. Relevância

O usuário deve julgar a relevância das unidades individuais recuperadas relacionando-as à tarefa de busca de informação em andamento. Relevância é um tema central da ciência da informação e tem sido considerada tanto da perspectiva teórica quanto da prática. A relevância atribui um caráter subjetivo à recuperação de informação, uma vez que está relacionada às necessidades do usuário e, principalmente, à sua capacidade de reconhecer os resultados apresentados. Uma mesma informação pode ser bastante relevante para uma pessoa, enquanto para outra possui baixo nível de relevância, de acordo com o ponto de vista da pesquisa e de seu conhecimento prévio do assunto. Saracevic (1975) discute as bases para entender relevância na recuperação de informação e formulou seguinte expressão:

Relevância é o (a) A de um(a) B entre um(a) C e um(a) D conforme determinado por um(a) por E.

Nesta expressão, cada incógnita pode ser substituída por um dos termos presentes no quadro a seguir:

Quadro 3 – Variáveis da expressão de relevância

A	B	C	D	E
medida grau dimensão estimativa avaliação relação	correspondência utilidade conexão satisfação ajuste cruzamento	documento artigo forma textual referência informação oferecida fato	pergunta requisição necessidade do usuário ponto de vista	pessoa julgador usuário solicitante especialista de informação

Cooper (1971, *apud* Marchionini, 1995) definiu "relevância lógica" como a base formal para avaliação de sistema de recuperação. Já Wilson (1973, *apud* Marchionini, 1995) descreveu "relevância situacional" como dependente do

problema de informação particular em mãos. A relevância situacional provê uma forma mais específica para o usuário julgar a relevância à medida que examinam os resultados intermediários da busca. De uma perspectiva prática, relevância serve como critério principal para computar medidas de *performance*, como retorno e precisão.

Do ponto de vista do usuário, relevância pode ser considerada como decisão a respeito das ações a serem executadas a cada passo no processo de busca de informação. Alternativas incluem:

- terminar a busca porque o objetivo foi alcançado;
- examinar o documento mais detalhadamente;
- registrar a existência e localização do documento e continuar examinando outros resultados para depois voltar ao anterior para exame mais detalhado;
- examinar mais detalhadamente outras implicações do documento para a continuação da busca (identificar termos para usar em questões subseqüentes);
- continuar examinando outros resultados nessa interação;
- formular nova questão ou redefinir o problema;
- rejeitar o documento por completo e continuar examinando os resultados e
- rejeitar o documento e parar a busca de informação sem terminar a tarefa.

Medidas para avaliar recuperação de informação como precisão – a precisão dos resultados da busca – e revocação – a completude dos resultados de busca – funcionam bem para avaliar tarefas de localização de item conhecido, especialmente para busca. A relevância, por outro lado, é muito mais difícil de medir, pois envolve tomada de decisão por parte do usuário (Toub, 2000).

Interfaces de bibliotecas *on-line* são dedicadas à recuperação de informações bibliográficas e de outros itens pertencentes ao acervo. Como tais, são orientadas à tarefa, com forte suporte na navegação, o que as torna produtos bastante adequados para estudar a Ergonomia da interação humano-computador. Conhecer as necessidades e expectativas do usuário é um passo importante e fundamental para se iniciar um projeto de interface para esse tipo de sistema. Também é fundamental conhecer conceitos da biblioteconomia e da ciência da informação, áreas tradicionalmente envolvidas com atividades similares.

No estágio atual da web, a possibilidade de localizar objetos de informação é de fundamental importância para praticamente todas as áreas de atividades, desde a acadêmica até o mundo dos negócios. Qualquer que seja o contexto, o usuário sempre desejará encontrar um item relevante de maneira precisa e com o mínimo de esforço para obtê-lo. O ergonomista, ao se debruçar sobre esse tipo de produto basicamente deve ter como pontos de interesse o esforço do usuário e verificar o quão fácil de usar e de aprender é a interface, suas capacidades de busca, além dos formatos de saída e de exibição, que são os elementos que informam ao usuário o nível de sucesso de sua tarefa.

Os próximos capítulos abrangem a parte empírica desta pesquisa, apresentam os métodos e técnicas utilizados para investigar o problema. Apresentam, também, os procedimentos de diagnóstico e de projeto executados, bem como os resultados da pesquisa e suas conclusões.