

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



**UMA PROPOSTA DE INTERFACE DE CONSULTA PARA  
RECUPERAÇÃO DE INFORMAÇÃO EM DOCUMENTOS  
SEMI-ESTRUTURADOS**

MIRELLA SILVA JUNQUEIRA

Uberlândia - Minas Gerais

2009

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



MIRELLA SILVA JUNQUEIRA  
Orientador: PROF. DR. ILMÉRIO REIS DA SILVA

**UMA PROPOSTA DE INTERFACE DE CONSULTA PARA  
RECUPERAÇÃO DE INFORMAÇÃO EM DOCUMENTOS  
SEMI-ESTRUTURADOS**

Dissertação de Mestrado apresentada à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como parte dos requisitos exigidos para obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Banco de Dados.

Uberlândia, Minas Gerais  
2009

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada “**Uma proposta de Interface de Consulta para Recuperação de Informação em Documentos Semi-Estruturados**” por **Mirella Silva Junqueira** como parte dos requisitos exigidos para a obtenção do título de **Mestre em Ciência da Computação**.

Uberlândia, 19 de Fevereiro de 2009

Orientador:

---

Prof. Dr. Ilmério Reis da Silva  
Universidade Federal de Uberlândia/Minas Gerais

Banca Examinadora:

---

Prof. Dr. Ricardo da Silva Torres  
Universidade Estadual de Campinas UNICAMP/São Paulo

---

Prof. Dr. João Nunes de Souza  
Universidade Federal de Uberlândia UFU/Minas Gerais

UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Data: Fevereiro de 2009

Autor: **Mirella Silva Junqueira**  
Título: **Uma proposta de Interface de Consulta para Recuperação de  
Informação em Documentos Semi-Estruturados**  
Faculdade: **Faculdade de Computação**  
Grau: **Mestrado**

Fica garantido à Universidade Federal de Uberlândia o direito de circulação e impressão de cópias deste documento para propósitos exclusivamente acadêmicos, desde que o autor seja devidamente informado.

---

Autor

O AUTOR RESERVA PARA SI QUALQUER OUTRO DIREITO DE PUBLICAÇÃO DESTE DOCUMENTO, NÃO PODENDO O MESMO SER IMPRESSO OU REPRODUZIDO, SEJA NA TOTALIDADE OU EM PARTES, SEM A PERMISSÃO ESCRITA DO AUTOR.

# Dedicatória

*Aos meus pais....*

# Agradecimentos

Agradeço...

A Deus, por minha vida.

Aos meus pais pela dedicação, confiança, apoio, carinho e amor incondicional em todos os momentos.

À minha mãe pelo incentivo neste trajeto, sempre me aconselhando, torcendo e fazendo de tudo para dar certo.

Aos meus amigos do mestrado, Tauller, Klicia, Núbia, Vinícius, Marcos, Crícia, Juliana, Felipe, Klérison, Rodrigo, Robson, Valquíria, Liliane, Éverton, Ítalo, Lúcio e todos os demais colegas do mestrado, companheiros e ajuda sempre disponível, boas lembranças desses anos de estudo...

Ao CNPQ, pelo apoio financeiro.

Ao meu orientador pela oportunidade.

Ao Victor pelo apoio, ajuda, força, incentivo e motivação. Sem a sua presença, desde o início, este trabalho não existiria, você sabe o quão foi importante e essencial nesses anos.

*"E ainda se vier noites traiçoeiras  
Se a cruz pesada for, Cristo estará contigo  
E o mundo pode até fazer você chorar  
Mas Deus te quer sorrindo  
Seja qual for o seu problema  
Fale com Deus, Ele vai ajudar você  
Após a dor vem a alegria  
Pois Deus é amor e não te deixará sofrer"*

# Resumo

A Recuperação Semi-Estruturada é uma forma de recuperação de informação intermediária entre a Recuperação Textual e a Recuperação Estruturada (típica em sistemas de banco de dados relacionais). Em sistemas de recuperação estruturada, o usuário geralmente conhece a estrutura dos dados e as linguagens de consulta disponíveis, conseguindo assim formular consultas que produzem resultados mais precisos. Na Recuperação Textual o usuário não conhece a estrutura dos dados e formula as consultas apenas com palavras-chaves, as quais geram resultados não tão precisos. Na Recuperação Semi-Estruturada, o usuário geralmente desconhece a estrutura dos dados e formula consultas que mesclam buscas textuais e mecanismos de recuperação estruturada. Neste contexto, surge o problema de como melhorar a precisão dos resultados aproveitando a estrutura contida nos documentos semi-estruturados. Dados semi-estruturados são comumente armazenados como documentos XML, os quais podem ser vistos como árvores. Nós internos dessas árvores contêm a estrutura do documento enquanto os nós folhas contêm os dados. O projeto de interfaces para usuários neste contexto é um dos grandes desafios na recuperação semi-estruturada. Isso ocorre especialmente porque os usuários não conhecem a estrutura do documento e têm dificuldade na formulação de consultas estruturadas. Este trabalho apresenta uma proposta e um protótipo de interface desenvolvido para auxiliar os usuários no processo de formulação de consultas estruturadas. Pretende-se com isso aumentar a precisão nos resultados das consultas. A proposta é validada por meio de experimentos envolvendo usuários voluntários e pela comparação de resultados obtidos com consultas textuais e consultas estruturadas formuladas com o auxílio da ferramenta. A melhoria atinge 440% para consultas bem estruturadas, realizadas por usuário que conhece bem a interface, e 179,75% para consultas razoavelmente estruturadas, realizadas por usuários sem experiência no uso da interface.

**Palavras chave:** Banco de Dados, Recuperação de Informação, Recuperação semi-estruturada, Interfaces Homem-Máquina, XML

# Abstract

Semi-Structured Information Retrieval is an intermediate way to retrieve information between Textual Retrieval and Structured Retrieval (typical in relational database systems). In structured retrieval systems, users generally know the available data structure and query languages, so they can formulate queries that produce more accurate results. In textual retrieval users don't know the data structure and formulate queries with keywords only, which produces not so accurate results. In Semi-Structured Retrieval, users generally don't know the data structure and formulate queries that mix textual search and structured retrieval mechanisms. In this context, the problem of how to improve the results accuracy using the structure inside semi-structured documents appears. Semi-structured data is usually stored as XML documents and can be seen as trees. Internal nodes of these trees have the structure of documents, while leaf nodes contain text. The design of interfaces for users in this context is one of the biggest challenges in semi-structured information retrieval. This occurs especially because the users don't know the document structure and have problems in formulating structured queries. This dissertation presents a proposal and a prototype interface developed to help users in the process of formulation of structured queries. The aim is to increase the precision in the results of the queries. The proposal is validated by experiments involving volunteer users and by comparing the results of textual queries and structured queries made with the help of the interface. The improvement reaches 440% for well structured queries, with a user who knows the interface, and 179.75% for reasonably structured queries, by users without experience to use the interface.

**Keywords:** Database, Information retrieval, Semi-Structured Information Retrieval, Human-Machine Interfaces, XML

# Sumário

|                                                                      |             |
|----------------------------------------------------------------------|-------------|
| <b>Lista de Figuras</b>                                              | <b>xi</b>   |
| <b>Lista de Tabelas</b>                                              | <b>xiii</b> |
| <b>Lista de Abreviaturas e Siglas</b>                                | <b>xiv</b>  |
| <b>1 Introdução</b>                                                  | <b>1</b>    |
| <b>2 Recuperação de Informação para Documentos Semi-estruturados</b> | <b>4</b>    |
| 2.1 Recuperação de Informação . . . . .                              | 4           |
| 2.2 Extensible Markup Language (XML) . . . . .                       | 5           |
| 2.2.1 Principais Benefícios da Linguagem XML . . . . .               | 5           |
| 2.2.2 Estrutura de um Documento XML . . . . .                        | 7           |
| 2.2.3 Document Type Definitions (DTD) . . . . .                      | 8           |
| 2.2.4 Características do Conjunto de Documentos . . . . .            | 9           |
| 2.3 Recuperação Semi-estruturada (Arquivos XML) . . . . .            | 9           |
| <b>3 Sistemas de Recuperação de Informação para Documentos XML</b>   | <b>12</b>   |
| 3.1 Modelo Vetorial . . . . .                                        | 12          |
| 3.2 JuruXML . . . . .                                                | 15          |
| 3.3 sTerm . . . . .                                                  | 17          |
| <b>4 Avaliação de RI para Documentos XML</b>                         | <b>20</b>   |
| 4.1 Avaliação de Sistemas de Recuperação de Informação . . . . .     | 20          |
| 4.2 Avaliação para Documentos XML . . . . .                          | 22          |
| 4.3 INEX . . . . .                                                   | 23          |
| 4.3.1 A Coleção INEX . . . . .                                       | 24          |
| 4.3.2 Tópicos na INEX . . . . .                                      | 24          |
| 4.3.3 Sistema TopX Search . . . . .                                  | 25          |
| 4.4 Consulta Estruturada e o Impacto na Precisão . . . . .           | 25          |
| 4.4.1 Precisão no sTerm . . . . .                                    | 25          |
| 4.4.2 Verificação do sTerm . . . . .                                 | 28          |

|          |                                                                                                 |            |
|----------|-------------------------------------------------------------------------------------------------|------------|
| <b>5</b> | <b>Interface Gráfica de Usuário</b>                                                             | <b>30</b>  |
| 5.1      | Ergonomia . . . . .                                                                             | 31         |
| 5.2      | Usabilidade . . . . .                                                                           | 34         |
| 5.2.1    | A Importância do Usuário . . . . .                                                              | 37         |
| 5.2.2    | Importância da Avaliação . . . . .                                                              | 39         |
| 5.3      | Projeto de Interface Humano-Computador . . . . .                                                | 39         |
| 5.4      | Interfaces para Sistemas de Busca . . . . .                                                     | 41         |
| 5.5      | Padrões para Interface . . . . .                                                                | 45         |
| 5.6      | Sugestões para Interfaces de Sistemas de Recuperação de Informação . . .                        | 54         |
| 5.6.1    | Requisitos Propostos . . . . .                                                                  | 55         |
| 5.6.2    | Requisitos Ergonômicos . . . . .                                                                | 56         |
| <b>6</b> | <b>Uma Proposta de Interface para Especificação de Consultas Semi - estruturadas</b>            | <b>63</b>  |
| 6.1      | Prototipação . . . . .                                                                          | 63         |
| 6.2      | Exemplo de Aplicação da Interface . . . . .                                                     | 71         |
| 6.3      | Resultados da aplicação da interface . . . . .                                                  | 77         |
| 6.3.1    | Cálculo da Precisão-0 em Consultas Realizadas com a Interface -<br>Usuário Experiente . . . . . | 77         |
| 6.3.2    | Testes Realizados com Usuários sem Experiência . . . . .                                        | 82         |
| 6.4      | Considerações Finais . . . . .                                                                  | 85         |
| <b>7</b> | <b>Trabalhos Relacionados</b>                                                                   | <b>87</b>  |
| 7.1      | Interface de Usuário para Recuperação de Documentos XML . . . . .                               | 87         |
| 7.2      | Interface de Consulta para Dados Semi-estruturados . . . . .                                    | 89         |
| 7.3      | XQBE: Interface para Consultas Semi-estruturadas usando XQuery . . . .                          | 91         |
| 7.4      | XQuery Gráfica . . . . .                                                                        | 93         |
| 7.5      | BBQ: Interface para navegação e consulta de documentos XML . . . . .                            | 94         |
| 7.6      | Considerações Finais . . . . .                                                                  | 95         |
| <b>8</b> | <b>Conclusão e Perspectivas Futuras</b>                                                         | <b>96</b>  |
|          | <b>Referências Bibliográficas</b>                                                               | <b>98</b>  |
| <b>A</b> | <b>Tópicos INEX Triviais</b>                                                                    | <b>102</b> |

# Lista de Figuras

|     |                                                                                                                                                 |    |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 | Estrutura de uma árvore XML (nós externos não são representados em árvores XML.) . . . . .                                                      | 8  |
| 2.2 | Consultas XML . . . . .                                                                                                                         | 11 |
| 3.1 | Exemplo de vetores documento e consulta no espaço vetorial tridimensional ( $t = 3$ ) . . . . .                                                 | 14 |
| 3.2 | Um mapeamento de um documento XML (esquerda) para um conjunto de termos (direita) . . . . .                                                     | 16 |
| 3.3 | Consultas no JuruXML . . . . .                                                                                                                  | 16 |
| 3.4 | Mapeamento de um documento XML para um árvore rotulada. . . . .                                                                                 | 18 |
| 4.1 | Tópico 431 da INEX . . . . .                                                                                                                    | 25 |
| 4.2 | Interface de consulta TopX . . . . .                                                                                                            | 26 |
| 4.3 | <i>Ranking</i> com a consulta não-estruturada (linha tracejada) e a consulta com estrutura (linha contínua) no sTerm . . . . .                  | 27 |
| 4.4 | <i>Ranking</i> com a consulta não-estruturada (linha tracejada) e a consulta bem estruturada (linha contínua) no sTerm . . . . .                | 27 |
| 4.5 | <i>Ranking</i> com a consulta não-estruturada (linha tracejada) e a consulta bem estruturada (linha contínua) na verificação do sTerm . . . . . | 29 |
| 5.1 | Estrutura da Usabilidade . . . . .                                                                                                              | 37 |
| 5.2 | Estruturas Físicas . . . . .                                                                                                                    | 49 |
| 5.3 | Lista de Relevância ressaltada pelo tamanho da fonte [Submarino, 2009, Globo.com, 2009] . . . . .                                               | 57 |
| 5.4 | Preferências Google . . . . .                                                                                                                   | 59 |
| 5.5 | Mensagem de ajuda na linguagem do usuário . . . . .                                                                                             | 60 |
| 6.1 | Interface-Protótipo . . . . .                                                                                                                   | 65 |
| 6.2 | Início de uma consulta sem estrutura. . . . .                                                                                                   | 65 |
| 6.3 | Ajuda e preenchimento da consulta. . . . .                                                                                                      | 66 |
| 6.4 | Árvore representando a consulta $A[B[B1[D[D1 D2[H J[J1]] E F G]] C]$ . 6.5.                                                                     | 67 |

|      |                                                                                                                        |    |
|------|------------------------------------------------------------------------------------------------------------------------|----|
| 6.5  | Representação da consulta estruturada: A[B[B1[D[D1 D2[H J[J1]] E F G]] C]. . . . .                                     | 68 |
| 6.6  | Ajuda inicial - Raiz da coleção. . . . .                                                                               | 69 |
| 6.7  | Ajuda na construção da árvore de consulta. . . . .                                                                     | 69 |
| 6.8  | Mensagem quando não são encontrados resultados para a consulta. . . . .                                                | 69 |
| 6.9  | Painel D: Resultado de uma consulta, representando o Ranking de similaridade . . . . .                                 | 70 |
| 6.10 | Mensagens de processamento das consultas. . . . .                                                                      | 70 |
| 6.11 | Raiz da coleção. . . . .                                                                                               | 71 |
| 6.12 | Tela inicial da interface. . . . .                                                                                     | 72 |
| 6.13 | Filhos da tag article. . . . .                                                                                         | 72 |
| 6.14 | Ajuda para a tag "name". . . . .                                                                                       | 73 |
| 6.15 | Escolha do termo "weather" para a consulta. . . . .                                                                    | 73 |
| 6.16 | Ajuda para um termo (nó-folha). . . . .                                                                                | 74 |
| 6.17 | Seleção da tag "body", filha de "article". . . . .                                                                     | 74 |
| 6.18 | Ajuda para a tag "body". . . . .                                                                                       | 75 |
| 6.19 | Ajuda para a tag "section". . . . .                                                                                    | 75 |
| 6.20 | Ajuda para a tag "title". . . . .                                                                                      | 76 |
| 6.21 | Interface final: Consulta estruturada: article[name[weather] body[ section[ title[ prediction]]]] e resultado. . . . . | 76 |
| 7.1  | Interface de Consulta, baseada na geração de consultas XIRQL [Großjohann et al., 2002] . . . . .                       | 88 |
| 7.2  | Interface de Apresentação de Resultados [Großjohann et al., 2002]. . . . .                                             | 89 |
| 7.3  | Interface de Consulta Pathway Explorer - Parte A [Newman and Ozsoyoglu, 2004]. . . . .                                 | 90 |
| 7.4  | Interface de Consulta Pathway Explorer - Parte B [Newman and Ozsoyoglu, 2004]. . . . .                                 | 91 |
| 7.5  | Interface de Consulta XQBE. . . . .                                                                                    | 92 |
| 7.6  | Interface de consulta GXQL. . . . .                                                                                    | 93 |
| 7.7  | Interface de consulta BBQ. . . . .                                                                                     | 95 |

# Lista de Tabelas

|      |                                                                                                                                                                                                                                               |    |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1  | Principais características de banco de dados, recuperação de informação e recuperação semi-estruturada. Não existe consenso ainda sobre modelo formal, linguagem de consultas e estrutura de dados para recuperação semi-estruturada. . . . . | 2  |
| 5.1  | Recomendações para projeto de interface (1ª parte). . . . .                                                                                                                                                                                   | 45 |
| 5.2  | Recomendações para projeto de interface (2ª parte). . . . .                                                                                                                                                                                   | 46 |
| 6.1  | Consultas realizadas utilizando a interface-protótipo (1ª parte). . . . .                                                                                                                                                                     | 79 |
| 6.2  | Consultas realizadas utilizando a interface-protótipo (2ª parte). . . . .                                                                                                                                                                     | 80 |
| 6.3  | Resultados obtidos. . . . .                                                                                                                                                                                                                   | 81 |
| 6.4  | Precisão-0 nas consultas sem estrutura. . . . .                                                                                                                                                                                               | 82 |
| 6.5  | Precisão-0 nas consultas com estrutura. . . . .                                                                                                                                                                                               | 82 |
| 6.6  | Média de melhoria. . . . .                                                                                                                                                                                                                    | 82 |
| 6.7  | Resultados obtidos dos tópicos utilizados para os testes com usuários. . .                                                                                                                                                                    | 83 |
| 6.8  | Resultados obtidos usuário 1. . . . .                                                                                                                                                                                                         | 84 |
| 6.9  | Resultados obtidos usuário 2. . . . .                                                                                                                                                                                                         | 84 |
| 6.10 | Resultados obtidos usuário 3. . . . .                                                                                                                                                                                                         | 85 |
| 6.11 | Resultados obtidos usuário 4. . . . .                                                                                                                                                                                                         | 85 |
| 6.12 | Resultados obtidos usuário 5. . . . .                                                                                                                                                                                                         | 86 |
| 6.13 | Resultados obtidos usuário 6. . . . .                                                                                                                                                                                                         | 86 |

# Lista de Abreviaturas e Siglas

|      |                                                |
|------|------------------------------------------------|
| CAS  | content-and-structure                          |
| CO   | content-only                                   |
| INEX | INitiative for the evaluation of XML retrieval |
| RI   | Recuperação de Informação                      |
| SRI  | Sistema de Recuperação de Informação           |
| XML  | eXtensible Markup Language                     |
| IHC  | Interface Humano Computador                    |
| DTD  | Document Type Definition                       |

# Capítulo 1

## Introdução

Sistemas de Recuperação de Informação (RI) são frequentemente contrastados com banco de dados relacionais. Sistemas de R.I. recuperam informações à partir de texto não estruturado - forma pela qual frequentemente denominamos um texto sem marcadores. Bancos de dados são projetados para pesquisas em dados estruturados: conjunto de registros que tem valores para atributos pré-definidos, tais como número de empregado, título e salário.

Existem dois tipos de problemas na recuperação de informação que são intermediários entre a recuperação de texto e a busca em bancos de dados relacionais. Um destes é a busca paramétrica, o outro, foco deste trabalho, é a recuperação semi-estruturada, especificamente em documentos XML (Extensible Markup Language). Documentos XML podem ser vistos como árvores cujos nós folhas contêm texto e os nós internos rotulados definem os papéis dos nós folhas no documento. Este tipo de documento é chamado de semi-estruturado e a recuperação de informação contida nestes documentos é a recuperação semi-estruturada.

Existem diferenças fundamentais entre recuperação de informação e sistemas de banco de dados em relação ao modelo de recuperação, estruturas dos dados e linguagem de consulta, como mostrado na Tabela 1.1 [Manning et al., 2007].

Recuperação semi-estruturada tem se tornado cada vez mais importante por causa do crescimento do uso de XML. Hoje em dia, a maioria dos dados semi-estruturados é codificada em XML [Manning et al., 2007].

Os dois grandes problemas na recuperação semi-estruturada são o tamanho do índice e a formulação de consultas. Na RI tradicional o índice tem tamanho proporcional ao número de termos da coleção. Já na recuperação semi-estruturada as relações hierárquicas entre os termos também devem ser indexadas, provocando um crescimento significativo do índice. A formulação da consulta encontra uma barreira na necessidade de se formular a consulta com estrutura, o que na maioria das vezes não é de conhecimento do usuário.

Neste trabalho trataremos deste segundo problema: Como auxiliar os usuários na construção de uma consulta estruturada? Que tipo de interface para estes sistemas que

|                                     | <b>Banco de Dados</b> | <b>Recuperação de Informação</b> | <b>Recuperação semi-estruturada</b> |
|-------------------------------------|-----------------------|----------------------------------|-------------------------------------|
| <b>Objetos</b>                      | Registros             | Documento não estruturado        | Árvore com texto nas folhas         |
| <b>Modelos</b>                      | Relacional            | Espaço vetorial e outros         | ?                                   |
| <b>Estrutura de Dados Principal</b> | Tabela                | Índice invertido                 | ?                                   |
| <b>Consultas</b>                    | SQL                   | Consultas textuais               | ?                                   |

Tabela 1.1: Principais características de banco de dados, recuperação de informação e recuperação semi-estruturada. Não existe consenso ainda sobre modelo formal, linguagem de consultas e estrutura de dados para recuperação semi-estruturada.

auxiliariam o usuário no processo de consulta? Qual o impacto da estruturação da consulta na qualidade da recuperação?

Hoje há um consenso entre os desenvolvedores de software de que a qualidade do desempenho do usuário está ligada à qualidade de sua interface com o sistema. Assim, o desenvolvimento de sistemas interativos de qualidade exige a utilização conjunta e integrada de conceitos e técnicas específicas do domínio conhecido como Interação Humano-Computador (IHC) e de conceitos e métodos de desenvolvimento de sistemas considerados integrantes do domínio da Engenharia de Software.

Tradicionalmente a Engenharia de Software tem se ocupado do desenvolvimento do componente funcional do sistema, também conhecido como aplicação, enquanto que é do domínio de IHC, o componente que interage com o usuário, também conhecido como Interface com o Usuário (IU). Esta separação de componentes, no ponto de vista da arquitetura de sistemas interativos, é denominada independência de diálogo e possibilita dois benefícios essenciais:

- a aplicação e a IU podem ser modificados e evoluir separadamente.
- A IU pode ser projetada por um especialista da área, o projetista de interfaces, que pode usar conhecimentos de design visual, psicologia cognitiva, ergonomia, semiótica, etc.

O conceito de independência do diálogo, no entanto, tem sido entendido e praticado como sinônimo de isolamento de diálogo, uma postura na qual o desenvolvimento do sistema interativo é feito à partir do desenvolvimento dos dois componentes em separado: tipicamente, primeiro a aplicação, depois a IU.

Assim, os engenheiros de software (analistas e projetistas da aplicação) têm uma perspectiva interna destes sistemas, priorizando aspectos essencialmente funcionais, como por exemplo, eficiência, manutenibilidade e portabilidade, conferindo ao desenvolvimento uma

orientação funcional em detrimento da operacional. Assim, colocando em um plano secundário o desempenho do usuário face aos objetivos de suas tarefas, tem-se como resultado sistemas que efetivamente funcionam, mas que muitas vezes, não são adaptados aos usuários e inadequados às suas tarefas.

Com a crescente complexidade atribuída aos produtos - principalmente os produtos relacionados com tecnologia da informação e de comunicação, como aparelhos celulares ou microcomputadores - a Ergonomia e a usabilidade passaram a ser peças fundamentais no desenvolvimento de projetos mais adequados aos usuários. Entender as dimensões da usabilidade e seu impacto no produto final colabora para o aumento da efetividade e da eficiência dos artefatos gerados.

Este trabalho utiliza conceitos de recuperação de informação e recuperação de informação semi-estruturada. Exemplos de sistemas que provêm a melhora na qualidade dos resultados se a consulta incluir alguma informação de estrutura que melhor especifique a necessidade de informação do usuário. Estudos e padrões para o desenvolvimento de uma boa interface com exemplos e sugestões.

Neste trabalho, além dos conceitos é apresentado um protótipo de interface para ajudar o usuário na formulação de consultas, que será aplicado em um sistema existente para a verificação da melhora da qualidade dos resultados da consulta quando esta é realizada com estrutura. A principal contribuição diz respeito à proposta de uma interface, por meio de um protótipo e a análise qualitativa do impacto da estrutura das consultas na qualidades dos resultados, medidas por meio da precisão. Na avaliação do protótipo temos uma melhoria de mais de 440% na precisão dos resultados, quando realizadas consultas bem estruturadas e de 156,25% quando foram realizadas consultas com alguma estrutura por usuários iniciantes na utilização da interface.

Esta dissertação, além desta introdução, está dividida em conceitos de recuperação de informação e recuperação semi-estruturada, apresentados no capítulo 2, sistemas de recuperação semi-estruturada no capítulo 3 e formas de avaliação destes sistemas são apresentadas no capítulo 4. O Capítulo 5 discute conceitos sobre Ergonomia e Usabilidade para o projeto de interfaces. No capítulo 6 apresenta-se o protótipo da interface e a avaliação dos resultados deste protótipo. O Capítulo 7 mostra alguns trabalhos relacionados; e a conclusão e os trabalhos futuros são apresentados no Capítulo 8.

## Capítulo 2

# Recuperação de Informação para Documentos Semi-estruturados

Neste capítulo serão apresentadas as definições de recuperação de informação, documentos XML e por fim a recuperação de informação em documentos XML.

### 2.1 Recuperação de Informação

Recuperação de Informação(RI) é uma área da computação que estuda o armazenamento, classificação, agrupamento e recuperação automática de documentos. A abundância de informações disponíveis na *Web* torna necessário o uso de ferramentas de busca e é uma das principais razões para a crescente popularidade de RI. A *Web* é uma ferramenta importante, não somente para comunicação, mas também para armazenamento e compartilhamento de informação. Do ponto de vista de RI, a *Web* pode ser vista como grande repositório de dados contendo documentos, ou páginas *Web*, que são interconectados.

O problema central em recuperação de informação é encontrar informações de interesse dos usuários e a principal ferramenta usada para resolver este problema é o emprego de sistemas de recuperação de informação (SRI). A área de recuperação de informação apresentou importantes resultados, desde o seu início, nas tarefas de localizar e classificar documentos em sistemas bibliográficos, até então restritos a bibliotecas e redes de menor escala. Mais recentemente, o emprego de RI para busca de informações na *Web* contribuiu enormemente para a criação de máquinas de buscas.

O usuário de um SRI busca obter documentos que atendam à sua necessidade de informação. Para isto, ele, geralmente, expressa sua necessidade por meio de uma consulta. O sistema então realiza o casamento da consulta com seus documentos por meio de um modelo de recuperação de informação. Um modelo para RI propõe uma representação para documentos e consultas e define como calcular a similaridade entre os mesmos. O modelo, então, é capaz de quantificar a relevância de cada documento da coleção em re-

lação à consulta do usuário e, com base nesta relevância, apresentar-lhe os documentos que melhor atendam à sua necessidade de informação i.e., os mais similares à consulta. Por fim, o modelo apresenta um *ranking* de documentos que aparecem em ordem decrescente de similaridade. Para realizar a tarefa de recuperar documentos, um SRI trabalha com modelos de representação de documentos e consultas.

Os tipos de modelos de RI desenvolvidos podem ser divididos em: modelos clássicos como booleano, vetorial e probabilístico; modelos alternativos: por exemplo, redes de crença, vetorial generalizado, booleano generalizado, entre outros [Baeza-Yates and Ribeiro-Neto, 1999]; modelos estruturados: como o modelo *proximal nodes*, sistema *sTerm* [Baeza-Yates and Ribeiro-Neto, 1999, Schlieder and Meuss, 2002]; modelos lógicos de RI [Chen, 1994, Chiaramella, 1992]. Existem ainda modelos baseados em Inteligência Artificial, como modelos baseados em conhecimento, em redes neurais artificiais, em algoritmos genéticos e em linguagem de processamento natural [Croft, 1993, Croft, 1995, Kang and Choi, 1997].

Os diferentes tipos de modelos de RI refletem a complexidade de RI em geral e, em especial da tarefa de modelagem. Os tipos de modelos de RI diferem, em sua maioria, na forma que os objetos (documentos, imagens, etc) são representados e como a recuperação é definida.

Para que estes conceitos sejam estendidos para a recuperação semi-estruturada, devemos entender o que é um documento XML para verificar as diferenças existentes entre recuperação de informação não estruturada e semi-estruturada. Tais conceitos de XML são apresentados na próxima seção.

## 2.2 Extensible Markup Language (XML)

Extensible Markup Language (XML) é uma linguagem de marcação de dados (*meta-markup language*), desenvolvido pela W3C [W3C, 2009], que provê um formato para descrever dados estruturados. Isso facilita declarações mais precisas do conteúdo e resultados mais significativos de busca através de múltiplas plataformas. O XML também permitirá o surgimento de uma nova geração de aplicações de manipulação e visualização de dados via internet.

### 2.2.1 Principais Benefícios da Linguagem XML

O XML tem por objetivo trazer flexibilidade e potencializar o desenvolvimento de aplicações para Web. Dentre os benefícios para desenvolvedores e usuários temos: Buscas mais eficientes; Desenvolvimento de aplicações Web mais flexíveis; Distribuição dos dados via rede de forma mais comprimida e escalável e Padrões abertos. Esses e outros benefícios do XML são descritos a seguir [Júnior, 2008]:

### **Buscas mais eficientes**

Os dados em XML podem ser unicamente "etiquetados", o que permite que, por exemplo, uma busca por livros seja feita em função do nome do autor. Atualmente, usando R.I. não estruturada, uma busca com o nome do autor poderia levar a qualquer site que tivesse referência a tal nome, não importando se fosse o autor do livro ou simplesmente um livro sobre o autor.

### **Desenvolvimento de aplicações flexíveis para a Web**

Isso inclui integração de dados de fontes completamente diferentes, de múltiplas aplicações; computação e manipulação local dos dados; múltiplas formas de visualização e atualização granulares do conteúdo.

### **Integração de dados de fontes diferentes**

Atualmente é praticamente impossível a procura em múltiplos bancos de dados heterogêneos. A XML permite que tais dados possam ser facilmente combinados. Essa combinação seria feita via software em um servidor intermediário, estando os bancos de dados na extremidade da rede. Os dados poderiam ser distribuídos para outros servidores ou clientes para que fizessem o processamento, a agregação e a distribuição.

### **Computação e manipulação locais**

Os dados XML recebidos por um cliente são analisados e podem ser editados e manipulados de acordo com o interesse do usuário. Ao contrário de somente visualizar os dados, os usuários podem manipulá-los de várias formas. A separação da interface visual dos dados propriamente ditos permite a criação de aplicações mais poderosas, simples e flexíveis.

### **Múltiplas formas de visualizar os dados**

Os dados recebidos por um usuário podem ser visualizados de diferentes formas uma vez que o XML define somente os dados e não a forma de visualização. A interpretação visual poderia ser dada de várias maneiras diferentes, de acordo com as aplicações. Os recursos de CSS e XSL [XSL, 2009b, XSL, 2009a] permitem essas formas particulares de visualização.

### **Atualizações granulares dos documentos**

Os dados podem ser atualizados de forma granular, evitando que uma pequena modificação no conjunto de dados implique a busca do documento inteiro novamente. Dessa forma, somente os elementos modificados seriam enviados pelo servidor para o cliente. Atualmente, usando HTML (*HyperText Markup Language*) [HTML, 2009], uma modificação em um item de dados acarreta na necessidade de atualização da página inteira. O

XML também permite que novos dados sejam adicionados aos já existentes, sem a necessidade de reconstrução da página.

### **Fácil distribuição na Web**

Assim como o HTML, o XML, por ser um formato baseado em texto aberto, pode ser distribuído via HTTP sem necessidade de modificações nas redes existentes.

### **Escalabilidade**

Devido ao fato dos documentos XML separarem completamente os dados da forma com a qual são visualizados, autores de aplicações de visualização de dados podem torná-las muito poderosas e interativas, permitindo ao usuário visualizar os dados da forma que lhe agrade.

### **Compressão**

A compressão de documentos XML é fácil devido à natureza repetitiva das tags usadas para definir a estrutura dos dados. A necessidade de compressão é dependente da aplicação e da quantidade de dados a serem movidos entre clientes e servidores. Os padrões de compressão do HTTP 1.1 podem ser usados para o XML.

## **2.2.2 Estrutura de um Documento XML**

Um documento XML é uma árvore rotulada, contendo um nó externo e nós internos, onde um nó externo consiste de:

- dados de caracteres (uma sequência de texto);
- instruções de processamento (anotações para os processadores), tipicamente no cabeçalho do documento;
- um comentário;
- uma declaração de entidade;

Um exemplo de nó externo é:

```
<?xmlversion = "1.0" encoding = "UTF - 8"? >
```

Um nó interno é um elemento, o qual é rotulado com:

- um nome ou
- um conjunto de atributos, cada qual consistindo de um nome e um valor.

Um exemplo de nó interno é:

```
< nameid = "233488" » Machinelearning < /name >
```

Cujo nome é *name*, o valor é *Machine learning* e possui atributo *id* com o valor 233488

Normalmente, comentários e informações DTD (ver seção 2.2.3) não são explicitamente representadas na árvore. Um exemplo de estrutura de um documento XML é mostrada na Figura 2.1, representando nós internos, um nó externo não é representado na árvore:

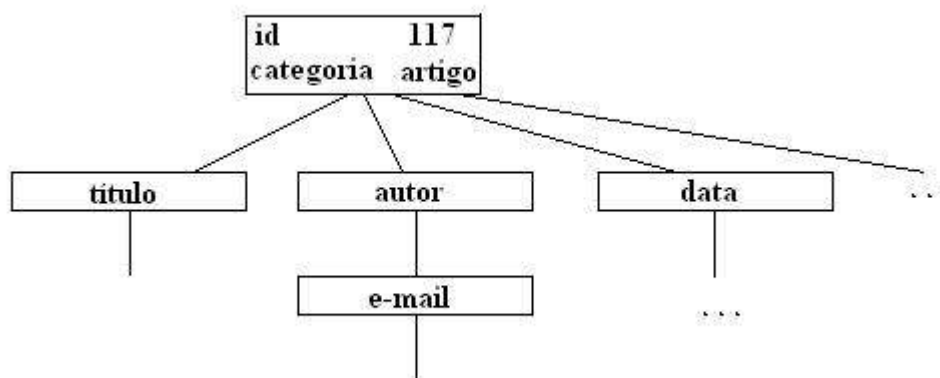


Figura 2.1: Estrutura de uma árvore XML (nós externos não são representados em árvores XML.)

### 2.2.3 Document Type Definitions (DTD)

No XML as regras que definem um documento são ditadas por DTDs (*Document Type Definitions*), as quais ajudam a validar os dados quando a aplicação que os recebe não possui internamente uma descrição do dado que está recebendo. Mas os DTDs são opcionais e os dados enviados com um DTD são conhecidos como dados XML válidos. Um analisador de documentos pode checar os dados que chegam analisando as regras contidas no DTD para ter certeza de que o dado foi estruturado corretamente. Os dados enviados sem DTD são conhecidos como dados bem formatados. Nesse caso, o documento pode ser usado para implicitamente se auto-descrever.

Com os dados XML válidos e com os bem-formatados, o documento XML se torna auto-descritivo porque as tags dão idéia de conteúdo e estão misturadas com os dados. Devido ao formato do documento ser aberto e flexível, ele pode ser usado em qualquer lugar onde a troca ou transferência de informação é necessária. Desta forma, podemos usar o XML para descrever informações sobre páginas HTML, ou descrever dados contidos em objetos ou regras de negócios, ou transações eletrônicas comerciais.

O que define formalmente quais elementos e quais combinações possíveis são permitidas dentro de um documento XML é o "*schema*", ou seja, esquema. Existem novos esquemas

propostos ao W3C, dentre eles estando o DCD (Document Content Description), que provêm a mesma funcionalidade dos DTDs, e que, pelo fato de linguagens esquema serem extensíveis, os desenvolvedores podem aumentá-los com informações adicionais, tais como regras de apresentação, tornando essas novas linguagens esquema mais poderosas que os DTDs.

## 2.2.4 Características do Conjunto de Documentos

Um conjunto de documentos XML pode ser constituído por documentos provenientes de uma única fonte, por exemplo, uma revista, jornal ou biblioteca digital. Pode tratar de uma única área do conhecimento ou pode ser constituído de documentos de origens variadas e discorrerem sobre diversas áreas do conhecimento.

Os documentos podem ser classificados como *text-centric* quando possuem poucos marcadores intercalados em grandes quantidades de texto, ou podem ser *data-centric* quando são criados e interpretados por alguma aplicação, possuindo em geral uma estrutura bem definida. Sob o ponto de vista da estrutura XML, os documentos podem ter estruturas semelhantes ou diversificadas, dando origem a dois tipos de coleção:

- Coleção homogênea - Constituída por documentos que utilizam a mesma DTD, ou seja, constituída por documentos que possuem os mesmos elementos XML.
- Coleção heterogênea - Constituída por documentos com DTDs variadas, com publicações de diversas fontes, ou sem DTD.

A heterogeneidade da coleção é um fator importante na determinação da estratégia de recuperação a ser utilizada. Em coleções homogêneas a indexação de elementos específicos torna-se viável, uma vez que todos os documentos possuem os mesmos elementos. Para coleções heterogêneas este tipo de indexação só será possível a partir do mapeamento entre as diversas DTDs. Por outro lado, estratégias que independem da DTD podem ser implementadas em ambas as coleções. A interface-protótipo que será apresentada trabalha com a utilização de coleções heterogêneas.

## 2.3 Recuperação Semi-estruturada (Arquivos XML)

Documentos XML podem ser vistos como árvores que têm os nós folhas contendo texto e os nós internos rotulados que definem os papéis dos nós folhas no documento. Este tipo é chamado de texto semi-estruturado e a recuperação de informação nestes documentos é a recuperação semi-estruturada. Recuperação semi-estruturada tem se tornado cada vez mais importante por causa do crescimento do uso de Extensible Markup Language ou XML. XML é usada para conteúdo web, para muitos documentos, para importação

e exportação de textos com conteúdo geral, e muitas outras aplicações. Hoje em dia, a maioria dos dados semi-estruturados é codificada em XML [Manning et al., 2007].

Conforme seção anterior, um documento XML é uma árvore ordenada e rotulada. Os nós da árvore são elementos XML e são escritos com tags de abertura e fechamento. Um elemento pode ter um ou mais atributos XML.

Um grande desafio para a recuperação semi-estruturada é o escolha da unidade de indexação, então um critério de decisão tem sido proposto para seleção da parte mais apropriada de um documento. É o princípio da recuperação de documentos estruturados: *Um sistema deveria sempre retornar a parte mais específica de um documento em resposta à consulta*. A decisão de qual nível da árvore é a resposta correta para a consulta é um problema difícil [Manning et al., 2007].

Existem três diferentes abordagens pra definição da unidade de indexação na recuperação XML:

1. Indexar todos os componentes que são elegíveis para serem retornados no resultado da consulta - a desvantagem é que este esquema pode conter sobreposição de unidade que têm que ser filtradas em uma etapa de pós-processamento para reduzir a redundância.
2. Agrupar nós em pseudo-documentos não sobrepostos. Isto evita o problema da sobreposição, mas pseudo-documentos podem não ser intuitivos para o usuário. E eles devem ser fixados em tempo de indexação, levando a nenhuma flexibilidade para responder às consultas em um nível mais específico ou mais genérico.
3. Designar um elemento XML como um substituto para a unidade. Muitas abordagens escolhem um artigo/documento como unidade de indexação.

Um outro desafio existente é que é preciso distinguir diferentes contextos de um termo durante o cálculo de estatísticas para um *ranking*.

Em relação à especificação de consultas, os esquemas de um documento XML são um grande desafio para interfaces de usuário, principalmente se os usuários não são familiarizados com a estrutura e as convenções de nomenclatura dos documentos da coleção.

As consultas XML podem ser representadas como árvores. Todos os documentos que "casam" com a árvore-consulta são considerados. A figura 2.2 mostra consultas para livros com as palavras Romeu e Julieta no título (a); livros com título contendo Minha Vida e autor contendo Romeu (b); e livros contendo Hamlet em algum lugar qualquer (c).

A interface do usuário deveria expor a estrutura da coleção e permitir ao usuário especificar os nós que serão consultados. Como consequência, a interface da consulta deverá ser mais complexa que uma caixa de pesquisa como na recuperação não-estruturada ("textos livres").

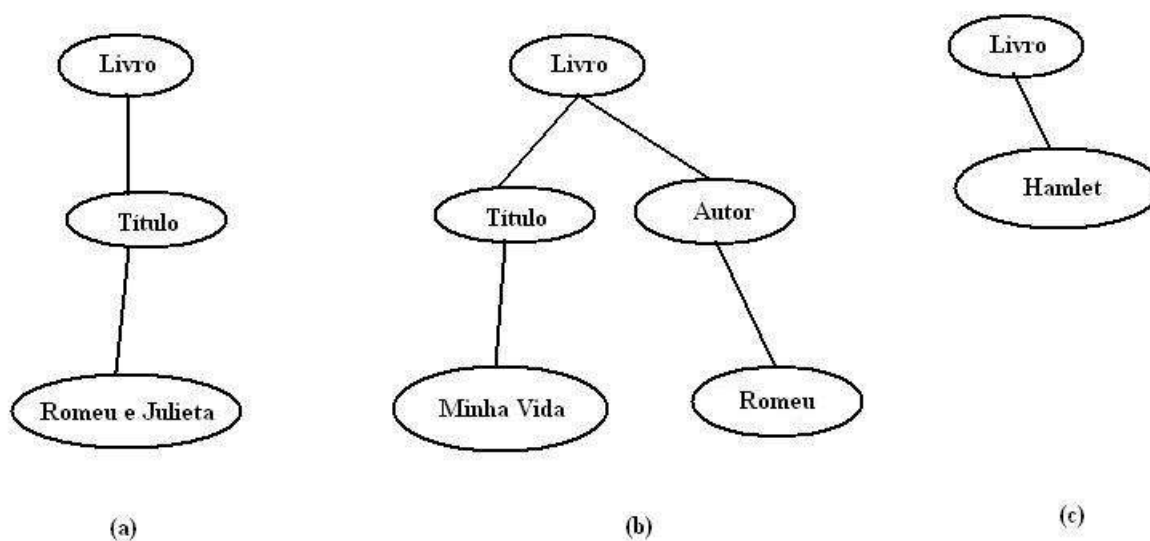


Figura 2.2: Consultas XML

A aplicação dos conceitos de recuperação de informação apresentados neste capítulo é feita pela criação de modelos implementados em sistemas que permitem a realização da busca da informação. No próximo capítulo serão apresentados alguns exemplos de sistemas de recuperação de informação semi-estruturada.

## Capítulo 3

# Sistemas de Recuperação de Informação para Documentos XML

Os sistemas de recuperação de informação são implementações dos modelos de recuperação. Neste capítulo apresentam-se dois sistemas de recuperação semi-estruturada que são baseados no modelo vetorial, apresentado na seção 3.1.

### 3.1 Modelo Vetorial

O modelo vetorial foi inicialmente proposto por Gerard Salton [Salton, 1974]. A apresentação deste modelo segue [Baeza-Yates and Ribeiro-Neto, 1999]. Neste modelo, os objetos relevantes para o sistema de recuperação de informação, documentos e consultas, são representados como vetores de termos. Uma medida entre vetores é utilizada para ordenar os documentos recuperados para uma consulta.

Um termo, para [Baeza-Yates and Ribeiro-Neto, 1999] é o resultado do processamento de todas as palavras de um documento. Este processamento é feito a partir da eliminação de stopwords (tais como artigos e conectivos), de redução das palavras à sua raiz (*stemming*) e a identificação de grupos de substantivos (o que elimina adjetivos, advérbios e verbos).

Cada termo  $k_i$  é representado como um vetor  $t$ -dimensional, em que  $t$  é o número de termos distintos da coleção. No modelo vetorial, o vetor  $\vec{k}_i$  representa o termo  $k_i$ . Se  $c_s$  é o  $s$ -ésimo elemento do vetor  $\vec{k}_i$ , então  $\vec{k}_i = (c_1, c_2, \dots, c_t)$  onde

$$\begin{cases} c_s = 0 \Leftrightarrow s \neq i \\ c_s = 1 \Leftrightarrow s = i \end{cases}$$

ou seja,

$$\begin{aligned}\vec{k}_1 &= (1, 0, 0, \dots, 0) \\ \vec{k}_2 &= (0, 1, 0, \dots, 0) \\ &\vdots \\ \vec{k}_t &= (0, 0, 0, \dots, 1)\end{aligned}$$

O conjunto de todos os vetores de termos  $\mathbf{K} = \{\vec{k}_1, \vec{k}_2, \dots, \vec{k}_t\}$  é linearmente independente e forma a base canônica para o espaço  $\mathbb{R}^t$  do modelo vetorial. Os vetores de termos são todos ortogonais entre si e, como consequência, os termos correspondentes são considerados independentes. A presença de um termo não indica a presença ou ausência de outro termo. A presença de um termo também não aponta sua relação com os outros termos.

Vetores de documentos e consultas são representados utilizando o conjunto de vetores de termos  $\mathbf{K}$ . Estes vetores são construídos como uma combinação linear dos vetores de termos. O vetor  $\vec{d}_j$  associado ao documento  $d_j$  é definido por:

$$\vec{d}_j = \sum_{i=1}^t w_{i,j} \vec{k}_i \quad \text{ou} \quad \vec{d}_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j})$$

Analogamente, o vetor para a consulta  $q$  é definido por:

$$\vec{q} = \sum_{i=1}^t w_{i,q} \vec{k}_i \quad \text{ou} \quad \vec{q} = (w_{1,q}, w_{2,q}, \dots, w_{t,q}).$$

Nas igualdades acima,  $w_{i,j}$  e  $w_{i,q}$  são pesos do termo  $k_i$  no documento  $d_j$  e na consulta  $q$ , respectivamente. A representação gráfica dos vetores de documentos e consultas é exemplificada na Figura 3.1

A definição de peso dos termos no modelo vetorial é baseada em estatísticas de ocorrência no documento e na coleção.

Um esquema comum para o primeiro fator é empregar a frequência do termo  $k_i$  no documento  $d_j$ . O fator de frequência do termo é usualmente referenciado por  $tf$  e fornece uma medida de quão bom é o termo para descrever o conteúdo do documento [Salton, 1974].

A definição dos pesos de termos mais eficiente para a recuperação de informação balanceia esses fatores [Baeza-Yates and Ribeiro-Neto, 1999], conforme é mostrado a seguir.

Sejam  $N$  o número total de documentos no sistema e  $n_i$  o número de documentos no qual o termo  $k_i$  aparece. Sejam  $freq_{i,j}$  a frequência do termo  $k_i$  no documento  $d_j$  e  $max(s)$  uma função que retorna a frequência do termo  $s$  com maior frequência na coleção de documentos. A frequência normalizada  $f_{i,j}$  do termo  $k_i$  no documento  $d_j$  é dada por

$$f_{i,j} = \frac{freq_{i,j}}{max(s)freq_{s,j}} \quad (3.1)$$

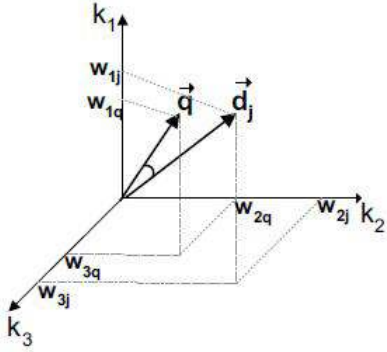


Figura 3.1: Exemplo de vetores documento e consulta no espaço vetorial tridimensional ( $t = 3$ )

em que o máximo é computado sob todos os termos mencionados no texto do documento  $d_j$ . Se o termo  $k_i$  não aparece no documento  $d_j$ , então  $freq_{i,j} = 0$ . A frequência inversa  $idf_i$  para  $k_i$  é dada por

$$idf_i = \log \frac{N}{n_i} \quad (3.2)$$

ou por variações desta fórmula. O peso do termo  $k_i$  no documento  $d_j$  é dado por

$$w_{i,j} = f_{i,j} \cdot idf_i \quad (3.3)$$

Essa estratégia de peso dos termos é denominada *tf-idf* [Baeza-Yates and Ribeiro-Neto, 1999].

O modelo vetorial avalia o grau de similaridade do documento  $d_j$  em relação à consulta  $q$  como a correlação entre os vetores  $\vec{d}_j$  e  $\vec{q}$ . A relevância de um documento para uma consulta é proporcional à distância entre os respectivos vetores. Usualmente, essa correlação é quantificada pelo co-seno do ângulo entre esses dois vetores. Isto é,

$$sim(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} = \frac{\sum_{i=1}^t w_{i,j} \cdot w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad (3.4)$$

Observa-se que a similaridade do documento em relação à consulta baseada no co-seno utiliza um fator de normalização no tamanho do documento. Documentos grandes são mais propensos a serem recuperados que documentos pequenos uma vez que os maiores,

comumente têm um conjunto de termos maior que os menores [Salton, 1974].

Como  $w_{i,j} \geq 0$  e  $w_{i,q} \geq 0$ , então  $0 \leq \text{sim}(d_j, q) \leq 1$ . Os documentos mais similares (mais próximo no espaço) à consulta são considerados relevantes para o usuário e retornados como resposta para a consulta. Após o cálculo dos graus de similaridade, é possível montar uma lista ordenada (*ranking*) de todos os documentos e seus respectivos graus de relevância à consulta, em ordem decrescente de relevância.

O Modelo Vetorial apresentado trabalha com documentos não estruturados.

Ao contrário da recuperação não-estruturada, a recuperação XML considera o contexto estrutural dos termos. Um documento cujo autor seja Romeu deverá "casar" com a segunda consulta da Figura 2.2, mas não com a primeira. O objetivo é usar o modelo vetorial para representar este contexto estrutural. Na recuperação não estruturada, haveria um único termo para Romeu. Na recuperação XML, deve-se separar a palavra título Romeu do autor Romeu. Assim, os termos devem representar não somente palavras, mas também sua posição dentro da árvore XML.

Considerando que no modelo vetorial cada termo é apresentado por um eixo no espaço vetorial, ao lidar com o contexto, o número de eixos do espaço tende a aumentar significativamente.

As seções seguintes apresentam dois exemplos de SRI semi-estruturada.

## 3.2 JuruXML

JuruXML [Carmel et al., 2002, Mass et al., 2003] foi um dos primeiros sistemas que fizeram a modificação no modelo vetorial para a recuperação XML. A representação básica do documento adotada no JuruXML é mostrada na figura 3.2. As dimensões no espaço vetorial são definidas como as subárvores dos documentos que contém ao menos um termo.

No JuruXML a indexação é feita com a criação de um vetor (t,c), onde t indica o termo e c é o contexto (caminho) a partir da raiz da árvore-documento. O índice é armazenado na forma t#c. Para a recuperação de documentos, a consulta não deve necessariamente conter todo o caminho correspondente ao documento, alguns nós internos na hierarquia podem ser omitidos, mas são preferíveis aqueles cuja árvore de consulta seja o mais próximo possível da árvore-documento. Assim, quanto melhor for formulada uma consulta como árvore, melhor será o resultado da consulta.

A figura 3.3 mostra duas consultas. Nesta podemos ver claramente que o processo de construção da consulta requer conhecimento sobre a estrutura, o que limita a quantidade de usuários que poderiam fazê-la.

Neste sistema, a similaridade entre a consulta e um documento, para a ordenação do



$$cr(\vec{q}, \vec{d}) = \frac{1 + |\vec{q}|}{1 + |\vec{d}|} \quad (3.6)$$

### 3.3 sTerm

O sistema apresentado em [Schlieder and Meuss, 2002], que recebeu o nome *sTerm* em [dos Santos, 2006a] é outro exemplo de sistema de recuperação de informação para documentos semi-estruturados que segue o modelo vetorial clássico. Os conceitos tais como documento, consulta e termos são estendidos para uma interpretação estruturada.

Uma consulta no modelo vetorial clássico é uma lista de palavras-chaves. No modelo *sTerm*, é adicionada estrutura nas palavras chaves de tal forma que as consultas possam ser interpretadas como árvores rotuladas. Documentos XML são interpretados como árvores rotuladas, também.

Uma coleção de documentos é modelada como uma única árvore, e cada subárvore como um documento lógico. A raiz da árvore da consulta determina a noção de documentos admissíveis: todo documento lógico, cujo nó raiz é igual ao nó raiz da consulta, é um candidato potencial a ser retornado como resultado. Este documento é comparado com a consulta, e atribui-se um grau de similaridade que determina a sua posição no ranking. O grau de similaridade é calculado utilizando a distribuição dos termos estruturados (s-terms).

Termos estruturados são, essencialmente, subárvores da consulta e dos documentos. O número de ocorrências de um termo estrutural dentro de um documento lógico e o número de documentos lógicos que contém o termo estrutural são contados, normalizados e utilizados para computar o peso de um termo, de forma análoga ao que ocorre no modelo vetorial. Os pesos são utilizados para construir os vetores documentos. Por outro lado, os pesos do vetor consulta podem ser definidos pelo usuário. Os vetores consulta e documento são comparados utilizando critérios próprios do modelo.

Cada documento XML físico é mapeado para uma árvore rotulada. Os elementos são representados por um nó que tem o nome do elemento como rótulo. Sequências de textos são decompostas em palavras. Cada palavra é mapeada para um nó folha rotulado com a respectiva palavra. Atributos são mapeados para dois nós que são pai e filho um do outro: o nome do atributo é o rótulo do nó pai, e o valor do atributo é o rótulo do filho. A Figura 3.4 mostra o mapeamento de um exemplo de documento XML para uma árvore com seus respectivos rótulos.

Os cálculos de peso e similaridade seguem as equações 3.3 e 3.4:

Frequência do termo: Sejam  $T$  um termo estrutural e  $D$  um documento lógico. Seja  $freq_T(D)$  o número de ocorrências de  $T$  em  $D$ , e  $maxfreq(D)$  o número máximo de ocorrências de um termo qualquer em  $D$ . A frequência de um termo  $tf_{T,D}$  de  $T$  em  $D$  é definida

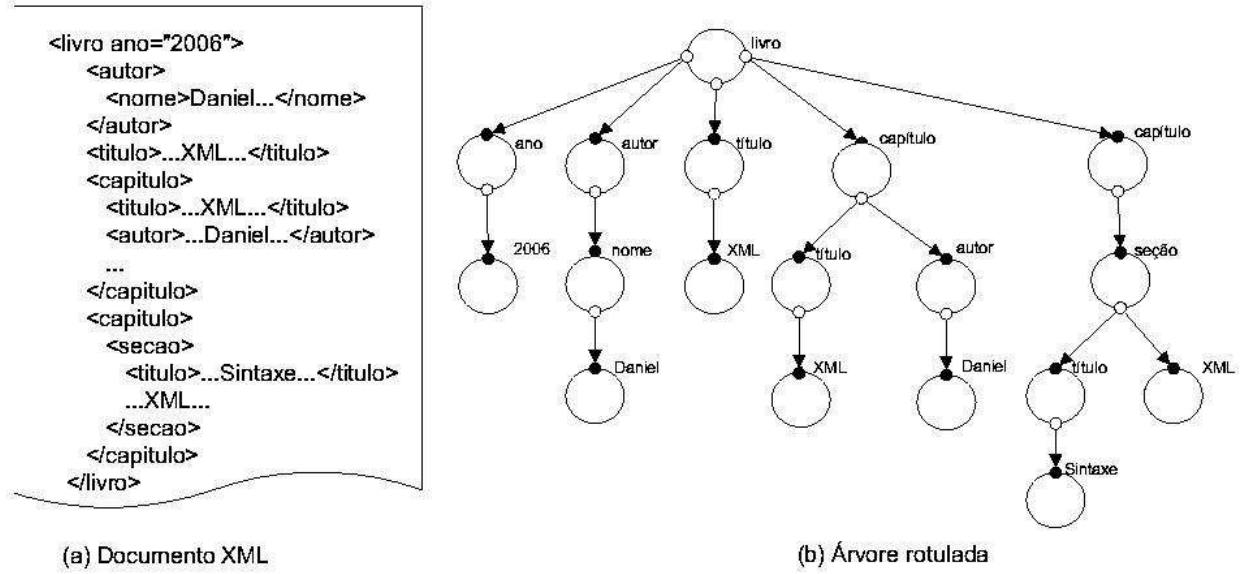


Figura 3.4: Mapeamento de um documento XML para um árvore rotulada.

por:

$$tf_{T,D} = \frac{freq_T(D)}{maxfreq(D)} \quad (3.7)$$

Frequência inversa de documentos: Seja  $T$  um termo estrutural, e  $t$  um tipo. Seja  $|D^t|$  o número de documentos do tipo  $t$ , e  $n_T$  o número de documentos em  $D^t$  casados por  $T$ . A frequência inversa  $idf_T^t$  de  $T$  é definida por:

$$idf_T^t = \log\left(\frac{|D^t|}{n_T}\right) + 1 \quad (3.8)$$

Peso do documento: Seja  $D^t$  um documento de tipo  $t$ . O peso  $w_{T,D}^t$  do termo estrutural  $T$  em  $D$  é definido como:

$$w_{T,D}^t = tf_{T,D} \times idf_T^t \quad (3.9)$$

Similaridade: Seja  $Q$  uma consulta com o vetor consulta  $v_Q$ , e  $D$  um documento com o vetor documento  $v_D$ . A similaridade  $sim(Q,D)$  é obtida por meio de uma operação entre os vetores  $v_Q$  e  $v_D$  é definida por:

$$sim(Q,D) = \cos(v_Q, v_D), \text{ se } D \in D^{tipo(Q)} \text{ e}$$

$$sim(Q,D) = 0, \text{ caso contrário,}$$

onde o  $\cos$  denota o cosseno entre os vetores, calculada como na equação 3.4 e  $D^{tipo(Q)}$  é o conjunto de documentos lógicos cujo tipo é o mesmo tipo da raiz da consulta ( $Q$ ).

Podemos perceber que a consulta, para que se obtenha resultados melhores, deve ser feita na forma de uma árvore, tornando assim o processo de especificação de consulta mais trabalhoso. Percebe-se aqui também o problema que é tratado neste trabalho.

Abaixo mostram-se dois exemplos de consultas realizadas neste sistema [Schlieder and Meuss, 2002]:

Consulta não estruturada (apenas texto plano):

$q_1$ : [*law*, *minister*, *enact*, *regulation*].

Consultas estruturada:

$q_2$ : [*preamble*[*law*], *minister*, *enact*, *regulation*].

$q_3$ : [*preamble*[*law*], *paragraph*[*minister*, *enact*, *regulation*]].

A consulta  $q_1$  é especificada apenas como uma lista de palavras, sem especificação de estrutura. A consulta  $q_2$  tem como especificação de estrutura apenas o termo *law* designado para um local específico: *preamble*. Na consulta  $q_3$  vemos que o contexto dos demais termos estão também designados.

## Capítulo 4

# Avaliação de RI para Documentos XML

Sistemas de R.I. devem ser avaliados para a verificação de suas qualidades, existem várias formas de se avaliar um SRI. Neste capítulo citamos algumas delas e apresentamos uma iniciativa mundial para avaliação de SRI semi-estruturado e ainda uma avaliação do sTerm, apresentado no capítulo anterior com relação à precisão dos resultados da consulta.

### 4.1 Avaliação de Sistemas de Recuperação de Informação

Um SRI classifica os documentos recuperados para cada consulta, de acordo com uma ordem de relevância gerando um resultado. Avalia-se o SRI através da comparação das respostas geradas por este sistema com um conjunto ideal de respostas. O conjunto ideal faz parte de uma coleção de teste que é utilizada para avaliar e comparar SRIs. Para isso, os dois conjuntos são comparados, obtendo-se dois índices de avaliação: precisão e revocação.

**Definição de Precisão:** é a fração dos documentos já examinados (recuperados) que são relevantes, calculado pela divisão do número de documentos relevantes recuperados pelo número total de documentos recuperados na busca.

**Definição de Revocação:** é a fração dos documentos relevantes observada dentre os documentos examinados (recuperados), calculado pela divisão do número de documentos relevantes recuperados pelo número total de documentos relevantes existentes (os quais deveriam ser recuperados).

**Definição de Precisão média interpolada:** é a média de precisão calculada para 11 pontos de revocação (0 a 1, variando de 0.1). A Precisão em 0% de revocação, aqui chamada de Precisão-0, é a precisão quando o primeiro documento relevante é encontrado no topo do *ranking*. É calculada pelo inverso da posição do primeiro documento relevante recuperado. Esta será a métrica utilizada nos testes deste trabalho.

Lancaster e Fayen [Lancaster and Fayen, 1973] afirmam que se pode avaliar o de-

sempenho de um SRI a partir do nível de satisfação do usuário no atendimento de sua necessidade. Eles citam, além das métricas acima, critérios como: cobertura, tempo de resposta, atualidade, formato de saída e esforço do usuário [dos Santos, 2006b]:

**Definição de 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.

**Definição de tempo de resposta:** um fator importante para avaliar sistemas de informação. Em sistemas acessados pela internet, torna-se sujeito a inúmeras variáveis. É importante oferecer respostas sobre o processamento das entradas para manter o usuário informado a respeito dos resultados de suas ações.

**Definição de atualidade:** define a proporção de novos documentos recuperados para o requisitante.

**Definição de formato de saída:** 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.

**Definição de 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. Se enquadram os aspectos de Ergonomia e de usabilidade da interface, 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: a documentação pode estar disponível na forma de ajuda. 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. 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:
  - (a) 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.

- (b) 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.

Tais métricas medem a interseção entre o conjunto resposta apontado por especialistas e aquele obtido por cada algoritmo de recuperação de informação.

## 4.2 Avaliação para Documentos XML

A avaliação da RI tradicional é efetuada a partir da utilização de uma coleção de testes constituída de três partes: um conjunto de documentos, um conjunto de necessidades de informação, chamados de tópicos e um conjunto de julgamentos de relevância listando, para cada tópico, um conjunto de documentos relevantes. Com base na coleção de testes e no *ranking* obtido pelo SRI, medidas de precisão e revocação são calculadas.

Uma coleção de testes para recuperação de informação em documento XML difere das coleções de teste tradicional usada em RI em muitos aspectos. Embora ainda seja constituída das mesmas três partes, a natureza dessas três partes é fundamentalmente diferente. Em coleções de testes para RI, documentos são considerados unidades de texto não estruturados, consultas são geralmente tratadas como coleções de termos e/ou frases, e o julgamento de relevância avalia se um documento como um todo é ou não relevante para uma consulta.

Os documentos XML, por outro lado, organizam seus conteúdos em elementos menores e estruturalmente aninhados. Cada um dos elementos dentro da hierarquia do documento, bem como o documento como um todo, é uma unidade de recuperação. Com relação aos tópicos, com o uso de linguagens de consulta para XML, os usuários de um sistema de recuperação de informação em documentos XML são capazes de combinar condições de estrutura e conteúdo em suas necessidades de informação e restringir sua consulta a elementos específicos dentro da coleção XML. Finalmente, o julgamento de relevância para coleções XML deve também considerar a natureza estruturada dos documentos e prover julgamentos em diferentes níveis da estrutura.

Uma consulta realizada com estrutura retorna resultados com maior precisão para o usuário em um sistema de recuperação semi-estruturado, como podemos observar na seções seguintes.

## 4.3 INEX

Uma grande parte da pesquisa acadêmica em recuperação XML é conduzida dentro do programa INEX (INiciative for the Evaluation of XML retrieval) [INEX, 2009], um esforço colaborativo que inclui coleções de referência, conjunto de consultas, julgamentos de relevância e um encontro anual para apresentar e discutir os resultados da pesquisa.

Em sua primeira edição, ocorrida em 2002, foram criados uma coleção de teste com documentos XML do mundo real, em conjunto com tópicos padronizados e os respectivos julgamentos de relevância. Grupos de pesquisas de 36 organizações participaram, apresentando as primeiras ferramentas de recuperação de informação, com seus resultados e as respectivas avaliações.

A partir de então, edições anuais vêm dando prosseguimento à criação da coleção de testes, estudos de métricas de avaliação e apresentação das ferramentas para recuperação de informação em documentos XML.

Existem dois tipos de consultas, chamadas tópicos na INEX: somente conteúdo (CO) e conteúdo e estrutura (CAS). Tópicos CO são consultas regulares como na recuperação não-estruturada. Tópicos CAS têm conteúdo estrutural, além das palavras-chaves.

A INEX é dividida em algumas frentes de estudo, dependendo da coleção e formas de recuperação e avaliação, são elas:

- Adhoc Track: os participantes comparam a eficácia de recuperação dos seus sistemas de recuperação XML a outros. Na realização, portanto, eles contribuirão para a construção da coleção de teste de XML. A coleção de teste possibilitará a realização de futuros experimentos comparativos por parte dos participantes.
- Multimedia track (MM): considera outros tipos de meios de comunicação além do texto. O seu objetivo é explorar a estrutura de XML, que fornece um nível lógico ao qual objetos multimedia são conectados, para melhorar a performance de recuperação.
- Document mining track (XDM): foi criado para explorar duas idéias principais: primeiro identificar problemas-chave para mineração de documentos semi-estruturados e novos desafios deste campo emergente e segundo estudar e avaliar o potencial de técnicas de máquinas de aprendizagem para tratar com tarefas genéricas em uma Máquina de Aprendizagem (*Machine Learning* - *ML*) em um domínio estruturado, isto é, classificação e agrupamento de documentos semi-estruturados.
- Entity Ranking track (XER): a necessidade para devolver entidades em vez de só documentos ou elementos está ficando cada vez mais importante. Esta *track* tenta fornecer um fórum de pesquisadores para comparar e avaliar técnicas de sistemas de busca para devolver listas de entidades.

- Link the wiki track (LTW: usando o Wikipedia como coleção, a *Link the wiki* tenta produzir um procedimento padrão e a métrica da avaliação da descoberta de *links* entre documentos.
- Book search track (BS): o objetivo desta frente é estudar estratégias de *ranking* de relevância específica para livros em tarefas de busca de informação, investiga questões de interface de usuário e comportamento de usuário, explorando características específicas de livro, como os índices fornecidos por autores, e associado metadata como biblioteca para catalogar a informação.

### 4.3.1 A Coleção INEX

A coleção de documentos usada na INEX é a Coleção Wikipedia XML, contendo 659.338 artigos. Em média um artigo contém 161 nós XML. A profundidade média de um nó na árvore XML é 6,72.

### 4.3.2 Tópicos na INEX

A criação e descrição dos tópicos utilizados na INEX seguem um formato padrão, consistindo das seguintes partes:

<title> Onde são dadas as consultas CO

<castitle> Onde são dadas as consultas CAS (Esta parte é opcional)

<description> Uma ou duas sentenças em linguagem natural que definam a informação buscada

<narrative> Onde são dadas definições do que é relevante ou irrelevante.

Em 2007 foram selecionados 130 tópicos para a tarefa de recuperação Ad-Hoc (a tarefa de recuperação para consulta específica).

Todos os tópicos que não foram criados considerando a parte <castitle> foram completados com esta parte de uma forma artificial, gerando consultas triviais:

```
// * [about(., keyword title)]
```

Por exemplo, o tópico 414 foi criado com consulta CO:

*hip hop beat*

e foi gerada a consulta CAS trivial:

```
// * [about(., hip hop beat)]
```

Foram utilizados somente 72 tópicos criados com uma consulta CAS não-trivial para verificar o relevância destas consultas nos sistemas avaliados.

A figura 4.1 mostra um tópico criado com uma consulta CAS não-trivial. Neste exemplo é possível perceber que a consulta CO é bem mais simples que a consulta CAS.

```
<inex_topic topic_id="431" ct_no="37">
<title>OpenGL in computer graphics</title>
<castitle>//article[about(.,computer graphics)]//section[about(.,OpenGL)]</castitle>
<description>find information about OpenGL used in computer graphics</description>
<narrative>I am supposed to use OpenGL in my computer graphics course. An element speaking
about OpenGL in computer graphics should be considered as relevant whereas an element
speaking about computer graphics without mentioning OpenGL should be considered as
irrelevant.</narrative>
</inex_topic>
```

Figura 4.1: Tópico 431 da INEX

### 4.3.3 Sistema TopX Search

TopK [Theobald et al., 2005, TopX, 2009] é uma máquina de busca para arquivos XML, disponibilizado pela INEX. Através dele é possível a realização de consultas na coleção INEX.

Os tópicos desenvolvidos na INEX podem ser utilizados/testados por meio do TopX. O sistema realiza consultas CO e CAS, mas a construção da consulta não é facilitada. O usuário deve ter conhecimento da estrutura do documento para a realização da consulta e esse é o problema principal deste trabalho.

A Figura 4.2 mostra a tela de consulta. Pode-se verificar que, independente do tipo de consulta (CO ou CAS) ela deve ser descrita no campo *Query*.

## 4.4 Consulta Estruturada e o Impacto na Precisão

Em um sistema de recuperação de informação semi-estruturada que permita a realização de consulta com estrutura, temos um aumento na precisão dos resultados das consulta, conforme mostrado a seguir.

### 4.4.1 Precisão no sTerm

Em [Schlieder and Meuss, 2002], após o desenvolvimento do sistema, foram realizados testes que comprovaram que consultas realizadas com estrutura obtiveram resultados mais precisos, como mostrado a seguir:

Supondo que a informação desejada seja:

*"Get all laws that describe how ministers enact regulations."*

# TopX Search

*logged in as "uberlandia"*



The screenshot shows a web interface titled "TopX Search". Below the title, it says "logged in as 'uberlandia'". The main section is a box labeled "Search options". Inside this box, there are several controls: a "Collection" dropdown menu currently showing "INEX Wikipedia [Ad-Hoc]"; a "Query" text input field; a "Granularity" section with two radio buttons, "Fetch&Browse" (which is selected) and "Thorough"; a "Parameters" section with a text input for "Results/page" set to "10" and a checkbox for "Conjunctive" which is unchecked. At the bottom of the "Search options" box are two buttons: "Submit" and "Previous settings".

Figura 4.2: Interface de consulta TopX

Seja a consulta não estruturada:

*document : [law, minister, enact, regulation].*

O ranking gerado nesta consulta, usando os cálculos de similaridade do modelo vetorial clássico, é apresentado na Figura 4.3 com a linha tracejada.

A segunda consulta, com alguma estrutura:

*document[preamble[law], minister, enact, regulation].*

requer que **law** apareça no pre-âmbulo do documento, e é apresentada na Figura 4.3 como a linha contínua.

Os documentos 9, 18, 19, 15, 7, 4, nos quais *law* está especificada no pre-âmbulo estão agora nas primeiras posições do *ranking*, como apresentado na Figura 4.3 que mostra que a adição de estrutura aumenta a precisão dos resultados da consulta.

Uma terceira consulta estendendo a estrutura da consulta e esperando que as palavras-

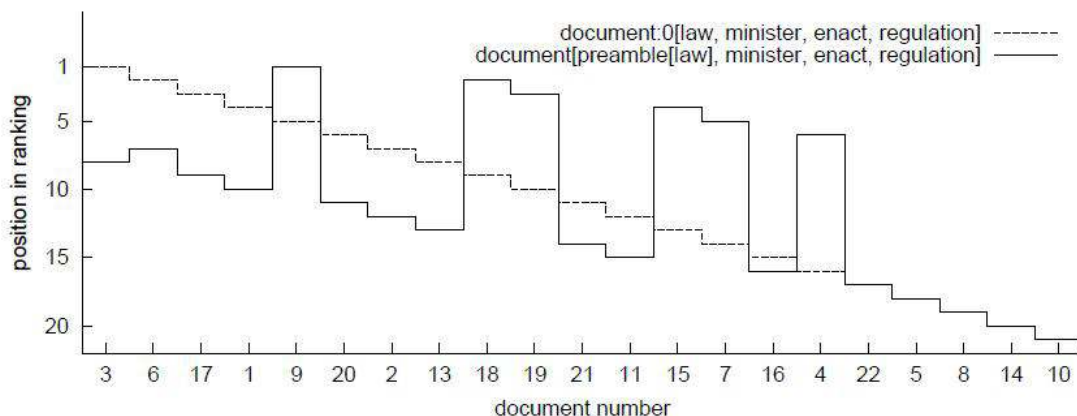


Figura 4.3: *Ranking* com a consulta não-estruturada (linha tracejada) e a consulta com estrutura (linha contínua) no sTerm

chave *minister*, *enact* e *regulation* estejam no mesmo parágrafo:

`document[preamble[law],paragraph[minister,enact,regulation]]`

Esta extensão modifica novamente a ordem do *ranking* acima. Os documentos que contêm as palavras no mesmo parágrafo (documentos 18, 15 e 7 na Figura 4.4).

A Precisão-O obtida na consulta não estruturada foi de 11,1%. Na consulta com estrutura a Precisão-0 foi de 50%, alcançando um aumento de 400%. E na consulta bem estruturada a Precisão-0 foi de 100%, alcançando um aumento de 800%.

## 4.4.2 Verificação do sTerm

O sistema sTerm foi implementado como parte deste trabalho, para a verificação de alguns resultados apresentados em [Schlieder and Meuss, 2002]. Foram mantidas as características no sistema original. A coleção exemplo utilizada agora era composta de alguns documentos da coleção INEX. É o sistema utilizado na interface-protótipo que será apresentada.

Podemos verificar novamente, nesta implementação, que a inserção de estrutura nas consultas gera o aumento da precisão dos resultados.

Para o entendimento de algumas tags XML que aparecem nas consultas estruturadas, apresentamos sua semântica:

Um parágrafo `<p>` pode ter palavras, elementos ligados, listas, etc. Dentre esses elementos destacamos:

*Stylistic emphasis* `<emph>`: este elemento é usado para acentuar palavras ou frases se-

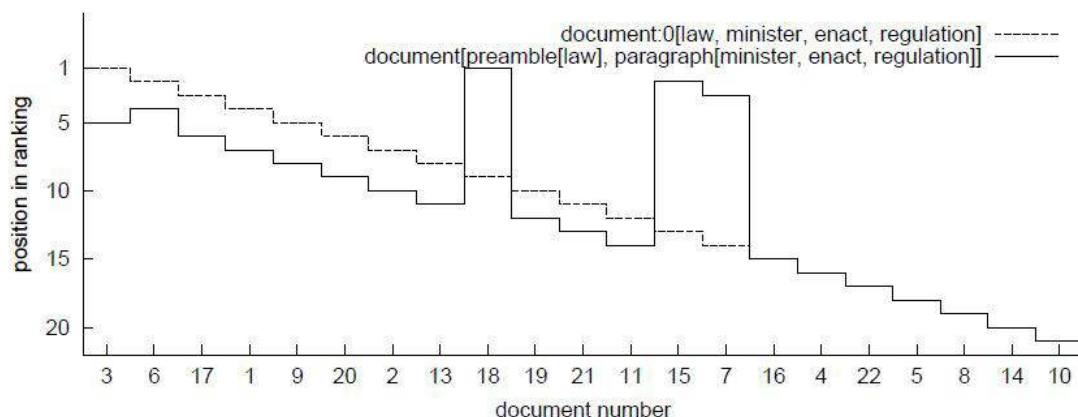


Figura 4.4: *Ranking* com a consulta não-estruturada (linha tracejada) e a consulta bem estruturada (linha contínua) no sTerm

gundo a sua aparência desejada. O formato da ênfase (negrito, itálico, etc.) é determinado pelo atributo "de tipo" do elemento <emph>.

*Structural emphasis* <emph2>: este elemento é usado para acentuar palavras ou frases segundo o seu significado. O tipo da ênfase (cuidado, aviso, etc.) é determinado pelo atributo "de tipo" do elemento <emph2>. O elemento <emph2> é tipicamente aplicado a parágrafos inteiros.

*Cross-reference* <xref>: este elemento é usado para estabelecer um *link* entre uma palavra ou frase e outro elemento dentro do documento (isto é, *hyper link*). Para estabelecer este *link*, o valor do atributo de "refid" do elemento <xref> deve combinar com o valor de atributo de "id" do elemento de objetivo.

*Foot node* <ftn>: este elemento é usado para fornecer anotações curtas ao texto. A presença do elemento <ftn> gerará uma referência à nota ao pé de uma página no ponto no texto no qual a nota ocorre.

A primeira consulta realizada foi na forma não estruturada, como se segue:

`article[notebook, academic, prediction, companies].`

Em seguida, foi inserido o uso de estruturas na consulta:

`article[emph2[notebook], academic, prediction, companies].`

As linhas tracejadas na Figura 4.5 mostram o ranking gerado pela primeira consulta (sem estrutura) e a linha contínua mostra o ranking na segunda consulta (com estrutura). Documentos que não continham os termos consultados estão nas últimas posições do ranking, ordenados pelo nome já que todos têm similaridade 0 com a consulta.

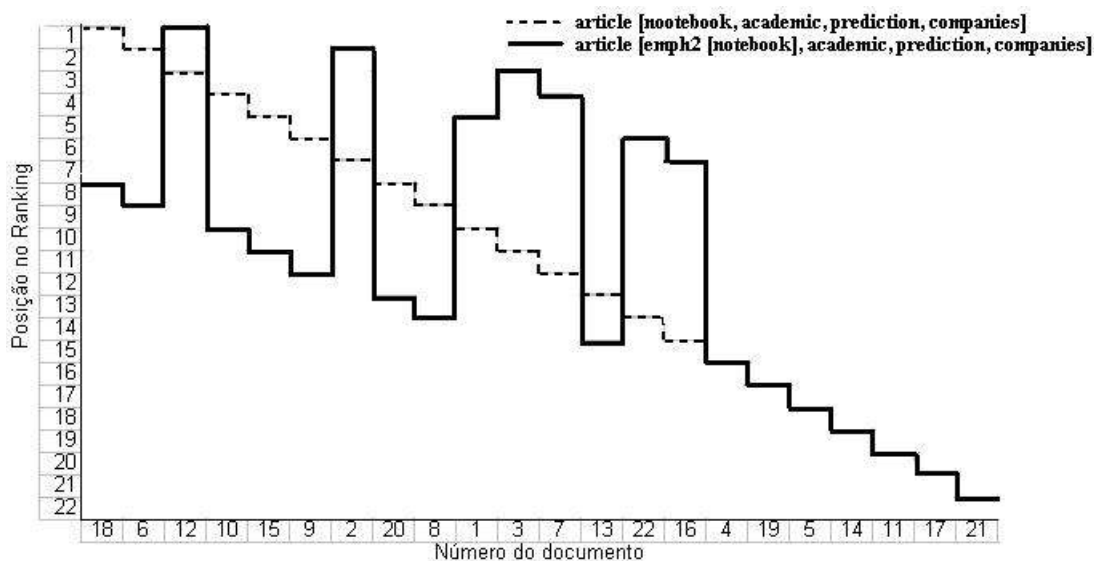


Figura 4.5: *Ranking* com a consulta não-estruturada (linha tracejada) e a consulta bem estruturada (linha contínua) na verificação do sTerm

Os documentos 12, 2, 1, 3, 7, 22 e 16 que contém o termo *notebook* em *emph2* agora alcançam as primeiras posições no *ranking*.

A Precisão-0 para a consulta não estruturada foi de 33,3% e de 100% na consulta estruturada, gerando um aumento de 200 %

Com base nos resultados do INEX e das propostas do JuruXML e sTerm, confirmando que a estrutura presente na consulta ajuda no aumento da precisão inicial, e na ausência de uma interface que auxilie na construção destas consultas, uma interface para estes sistemas será proposta neste trabalho. O trabalho inclui ainda uma avaliação quanto à capacidade da interface de auxiliar o usuário a elaborar consultas bem estruturadas.

Muitos cuidados devem ser tomados no desenvolvimento de uma interface. Alguns serão apresentados no próximo capítulo. E, à partir destes, um protótipo de interface será desenvolvido e mostrado no Capítulo 6, como principal contribuição deste trabalho.

## Capítulo 5

# Interface Gráfica de Usuário

O termo Interface tem uma série de significados diferentes, ampliados ainda mais com o advento das novas tecnologias, desde os computadores pessoais até os mais recentes telefones móveis que integram a tecnologia de armazenamento de dados, mesclando o celular e os PDAs (Personal Digital Assistant, como o Palm-OS) em um único aparelho. Os usuários dessas novas tecnologias controlam todas essas funções através da utilização das Interfaces Gráficas de Usuário (G.U.I.), ou simplesmente Interface. De acordo com o dicionário, interface significa um ponto comum, uma divisa ou fronteira entre duas coisas. Assim as G.U.I.s são o ponto de interconexão entre o usuário e a tecnologia, para um sistema de interação entre o homem e a máquina.

Para existir um sistema homem-máquina é necessário que existam ações ligando os elementos homem e máquina na busca por um objetivo, ou seja, uma tarefa comum que leve esses elementos a interagir. As máquinas podem ser consideradas extensões do ser humano, sendo subordinadas às ações do homem, agindo com eficiência, rapidez, força e velocidade em situações nas quais o corpo humano não alcança um determinado potencial. Com isso, possibilitam ao homem o cumprimento de tarefas que ele não poderia desempenhar sozinho.

A interface é o ponto no qual usuário, tarefa e ferramenta comunicam-se, interagem e se tornam um único elemento. Isso exige do homem apenas que ele conheça o funcionamento da ferramenta, pois ela já está adaptada ao seu manuseio.

A máquina pode estar onde o homem não pode, como o robô que foi ao planeta Marte, mas todas as ações do robô, desde a exploração do solo do planeta até o momento de seu desligamento são totalmente decididas pelo homem.

O software centrado no usuário envolve uma equipe multidisciplinar na sua elaboração e testes, capaz de conciliar na sua arquitetura as habilidades, objetivos e preferência deste usuário. Além disso, deve ser relevante, consistente, fácil de aprender, gerenciar adequadamente os erros, e fornecer *feedback* para as ações do usuário. Dessa forma poderá alcançar um ótimo padrão de usabilidade (ver seção 5.2).

A seção 5.1 traz o conceito de ergonomia, apresentando os oito critérios ergonômicos

e características que descrevem softwares ergonômicos. A seção 5.2 apresenta o conceito de usabilidade, com critérios para avaliação, incluindo uma norma ISO, assim como a importância do usuário e da avaliação do sistema desenvolvido. A importância da realização de um projeto baseado na interação homem máquina é apresentada na seção 5.3. Já a seção 5.4 mostra características úteis para a concepção de uma interface para sistemas de busca e a 5.5 alguns padrões que auxiliarão o desenvolvedor na construção da interface.

À partir dos conceitos apresentados nas primeiras 5 seções, a seção 5.6 traz sugestões que poderiam para a aplicação destes conceitos no desenvolvimento de uma interface auxílio à construção de consultas semi-estruturadas para a recuperação semi-estruturada. Algumas das sugestões apresentadas são aplicadas no protótipo desenvolvido e apresentado no capítulo 6.

## 5.1 Ergonomia

A ergonomia [Bastien et al., 1992] no âmbito da computação busca enfrentar os problemas e propor soluções lógicas para o desenvolvimento de software interativo que sejam adaptados a seus usuários e adequados a suas tarefas. Atualmente, face ao acelerado processo de informatização da sociedade, verifica-se a importância que a ergonomia da informática passa a ter para a qualidade de vida em geral, quando pessoas sem um treinamento específico passam a ter necessidades de acesso aos dispositivos informatizados existentes no seu trabalho, em sua casa ou mesmo em locais públicos [Walter de Abreu Cybis, 1999].

A ergonomia se vale dos conhecimentos sobre o homem no trabalho, com o objetivo de conceber e avaliar produtos e ferramentas que possam ser utilizados com o máximo de conforto, segurança e eficiência.

Trabalhos de pesquisa convergem para uma idéia central: a concepção de sistemas interativos mais úteis e usáveis depende do conhecimento das tarefas que as pessoas realizam e da aplicação deste conhecimento no processo de concepção de sistemas [Benyon, 1992, Johnson, 1992].

A ênfase da ergonomia não está em evitar a demanda excessiva sobre o operador humano, mas em explorar as capacidades humanas, por exemplo, por meio do aumento do prazer na execução da tarefa.

Com o objetivo de construir sistemas que suportem de maneira efetiva e satisfatória as tarefas do(s) usuário(s), existem alguns princípios ergonômicos para o processo de construção de sistemas. A formulação destes princípios é o resultado de uma experiência coletiva acumulada pela comunidade de ergonomia e de IHC (Interface Humano-Computador), examinando as práticas bem (ou mal) sucedidas. Os princípios são apresentados na forma de critérios, recomendações ou normas, e raramente são associados a algum método de concepção. São por isto, altamente dependentes de quem os interpreta e aplica. Bastien & Scapin, desenvolveram, a partir do exame exaustivo de uma base de recomendações

ergonômicas, um conjunto de critérios ergonômicos (C.E.) [Bastien and Scapin, 1993]:

1. **condução** se define na presteza do sistema, na legibilidade das informações e telas, no feedback imediato das ações do usuário e no agrupamento e distinção entre itens nas telas. Esse último sub-critério refere-se tanto aos formatos (agrupamento e distinção por formato) quanto à localização (agrupamento e distinção por localização) dos itens.
2. **Carga de trabalho** se define na brevidade das apresentações e entradas (concisão), no comprimento dos diálogos (ações mínimas) e na densidade informacional das telas como um todo.
3. **Controle explícito** se define no caráter explícito das ações do usuário (ações explícitas) e no controle que ele tem sobre os processamentos (controle do usuário).
4. **Adaptabilidade** refere-se tanto às possibilidades de personalização do sistema que são oferecidas ao usuário (flexibilidade) quanto ao fato da estrutura do sistema estar adaptada a usuários de diferentes níveis de experiência (consideração da experiência do usuário).
5. **Gestão de erros** refere-se tanto aos dispositivos de prevenção que possam ser definidos nas interfaces (proteção contra erros) quanto à qualidade das mensagens de erro fornecidas e às condições oferecidas para que o usuário recupere a normalidade do sistema ou da tarefa (correção dos erros).
6. **Consistência** refere-se à homogeneidade e à coerência das decisões de projeto quanto as apresentações e aos diálogos.
7. **Significado dos códigos e denominações** refere-se à relação conteúdo-expressão das unidades de significado das interfaces.
8. **Compatibilidade** se define no acordo que possa existir entre as características do sistema e as características, expectativas e anseios dos usuários e suas tarefas.

O software ergonômico aconselha, orienta, informa, e conduz o usuário na interação com o computador. Além disso proporciona aprendizado rápido e fácil utilização permitindo que o usuário melhore seu desempenho e diminua o número de erros na operação do sistema. Baseado nos critérios ergonômicoa acima, o software deve apresentar as características:

- Software prestativo: é aquele que engloba os meios utilizados para levar o usuário a realizar determinadas ações.
- Software Fácil: é aquele que segue características que sejam semelhantes a softwares já utilizados.

- Software Claro: é aquele que não deixa o usuário com dúvidas, quando se depara com a interface do aplicativo. Deve apresentar e dispor claramente todas as informações na tela.

- Software Confortável: é aquele que proporciona uma diminuição da carga de trabalho perceptiva e cognitiva do usuário, durante a interação. O software confortável deve evitar ações desnecessárias, para assim proporcionar interações mais rápidas.

- Software Obediente: é aquele que permite o controle explícito sobre o sistema, deixando o usuário definir as entradas, podendo ainda interromper, cancelar, postergar e retomar os processamentos a qualquer instante. As ambigüidades são limitadas e, em consequência, o usuário comete menos erros.

- Software Adaptável: a adaptabilidade de um sistema diz respeito à sua capacidade de reagir conforme o contexto e conforme as necessidades e preferências do usuário. Uma interface não pode atender ao mesmo tempo a todos os seus usuários, por isso o Software Adaptável permite a personalização da interface, a fim de levar em conta as exigências da tarefa, as estratégias ou os hábitos de trabalho dos usuários.

- Software Flexível: é aquele que respeita o nível de experiência do usuário, desde o especialista até o leigo. A interface de um software deve ser concebida para lidar com as variações de nível de experiência. A previsão de atalhos torna um software mais atraente para o usuário experiente, pois lhe permite acesso rápido às funções do sistema.

- Software Seguro: é aquele que emprega mecanismos que permitem evitar ou reduzir a ocorrência de erros e que favorecem a sua correção. As interrupções provocadas pelos erros têm consequências negativas sobre a atividade do usuário. Elas prolongam as transações e perturbam a planificação. Quanto menor é a possibilidade de erros, menos interrupções ocorrem e melhor é o desempenho. Os erros são bem menos perturbadores para o usuário, quando eles são fáceis de corrigir.

- Software Coerente: é aquele no qual procedimentos, rótulos, comandos, etc. são melhor reconhecidos, localizados e utilizados, quando seu formato, localização ou sintaxe são estáveis de uma tela para outra ou de uma seção para outra. Nestas condições, o sistema é mais previsível e a aprendizagem mais generalizada e os erros são diminuídos.

- Software Expressivo: é aquele que favorece a correta relação entre o objeto, a informação apresentada ou pedida e sua referência. Códigos e denominações significativas possuem uma forte relação semântica com seu referente. Termos pouco expressivos para o usuário ocasionam problemas de condução, levando o usuário a selecionar opções erradas.

- Software Compatível: é aquele que coloca em acordo as características do usuário e de sua tarefa com a organização das saídas, das entradas e do diálogo de uma dada aplicação. Esta qualidade diz respeito também ao grau de similaridade entre diferentes ambientes e aplicações.

Esse conjunto de critérios tem sido, durante os últimos anos, alvo de uma série de experimentos que visam validá-los em suas pretensões adicionais de organizar os conheci-

mentos do domínio da ergonomia, contribuir para a normalização e de aperfeiçoar o diagnóstico ergonômico [Bastien et al., 1992]. Assim, os autores consideram que atualmente, os critérios são válidos, confiáveis e úteis para, em particular, aumentar a performance da ergonomia [Bastien and Sacpin, 1995].

A adoção de uma abordagem baseada nestes princípios e práticas permite não só conceber (incluindo analisar, especificar e construir) sistemas que considerem adequadamente os usuários, suas reais necessidades e suas práticas de trabalho, mas também construir e testar a usabilidade do sistema desenvolvido.

O papel da Ergonomia na interação humano-computador é melhorar a qualidade da interação do usuário com os sistemas informatizados.

A crescente preocupação da área de Ergonomia e usabilidade em propor recomendações para o projeto de interfaces para sistemas de recuperação de informação 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 a partir de palavras contidas nos títulos, resumos ou outros campos de dados" [Branski, 2004].

## 5.2 Usabilidade

A usabilidade é uma abordagem importante na análise ergonômica. Não deve ser confundida com a ergonomia, pois a ergonomia é uma disciplina que contém ferramentas cognitivas (técnicas analíticas, projetivas e abordagens) e utiliza-se da usabilidade como conjunto de conhecimento, regras e normas para análise e validação de requisitos relacionados ao uso de uma interface. Usabilidade pode ser compreendida como a capacidade, em termos funcionais humanos, de um sistema ser usado com facilidade e com eficiência pelo usuário [dos Santos, 2006b]. Requisitos estéticos não são necessariamente abordados no contexto da usabilidade, mas, em boa parte, pela ergonomia e principalmente pelo projeto gráfico [Hoelzel, 2004].

Conceitualmente, a usabilidade tem sido um termo de compreensão comum e de concordância entre designers gráficos, ergonomistas e desenvolvedores, quanto a ser uma meta indispensável a um sistema interativo. É vista de modo geral como a propriedade que os produtos têm de serem fáceis de usar e aprender. [Bastien and Scapin, 1993] consideram que a usabilidade está diretamente ligada ao diálogo na interface e é a capacidade do aplicativo em permitir que o usuário alcance suas metas de interação. Ela é aplicada aos produtos, por isso é uma característica passível de medida e formalização quanto a requisitos e formas de análise. Desta forma, a engenharia de usabilidade tem sido largamente

utilizada para determinar qualidades essenciais dos sistemas interativos nas áreas de informática, ergonomia e design gráfico. As definições e elementos da usabilidade esclarecem para o design os pontos-chave a serem observados e auxiliam no entendimento comum entre design e engenharia a respeito das orientações da qualidade dos projetos orientados ao usuário.

[Moraes, 2004] relaciona os principais fatores pertinentes à abrangência do termo Usabilidade. São eles:

- Facilidade de aprendizagem: o sistema deve permitir que os usuários alcancem níveis de desempenho aceitáveis dentro de um tempo especificado.
- Efetividade: um desempenho aceitável deve ser alcançado por uma proporção definida da população-usuária, em relação a um limite de variação de tarefas e a um limite de variação de ambientes.
- Atitude: um desempenho aceitável deve ser atingido considerando custos humanos aceitáveis, em termo de fadiga, estresse, frustração, desconforto e satisfação.
- Flexibilidade: o produto deve ser capaz de lidar com um limite de variação de tarefas, além daquelas inicialmente especificadas.
- Utilidade percebida do produto: o maior indicador da usabilidade de um produto é se ele é usado.
- Adequação à tarefa: um produto "usável" deve apresentar uma adequação aceitável entre as funções oferecidas pelo sistema e as necessidades e requisitos dos usuários.
- Características da tarefa: a frequência com que uma tarefa pode ser desempenhada e o grau no qual a tarefa pode ser modificada, em termos da variabilidade dos requisitos de informação.
- Características dos usuários: conhecimento, habilidade e motivação da população usuária.

Em termos gerais, durante a prática cotidiana de projeto, a preocupação com a usabilidade só ocorre no final do ciclo de projeto, quando se faz a avaliação do produto já finalizado. Por esta razão, poucas modificações são implementadas, o que é agravado pelo custo elevado de modificações substanciais nessa etapa. Portanto, desde o início da atividade de projeto, a observação da usabilidade deve estar presente.

### **Critérios para avaliação de usabilidade**

As medidas de usabilidade mais freqüentemente consideradas, inclusive homologadas pela International Standard Organization (ISO), são: efetividade, eficiência e satisfação. a ISO publicou a norma ISO DIS 9241-11 [de Normas Técnicas. NBR 9241-11, 2002], que destaca os três aspectos separados da usabilidade. Essa estabelece os processos, critérios e os conceitos envolvidos na medida desta característica em sistemas informatizados:

"A ISO 9241 define usabilidade e explica como identificar a informação necessária a ser considerada na especificação ou avaliação de usabilidade de um computador em termos de medidas de desempenho e satisfação do usuário. E traz orientações sobre como descrever explicitamente o contexto de uso do produto (hardware, software ou serviços) e as medidas relevantes de usabilidade."(NBR 9241-11:2002, p.2) [de Normas Técnicas. NBR 9241-11, 2002]

Abaixo as definições de termos a partir da NBR (op. cit., p.3) de referência:

- **Usabilidade:** medida na qual um produto pode ser usado por usuários específicos para alcançar objetivos específicos com eficiência, eficácia e satisfação em um contexto específico de uso.
- **Eficácia:** acurácia e completude com as quais usuários alcançam objetivos específicos.
- **Eficiência:** recursos gastos em relação à acurácia e abrangência com as quais usuários atingem objetivos.
- **Satisfação:** ausência de desconforto e presença de atitudes positivas para com o uso de um produto.
- **Contexto de uso:** usuários, tarefas e equipamentos (hardware, software e materiais), e ambiente físico e social no qual o produto é usado.
- **Sistema de trabalho:** sistema composto de usuários, equipamentos, tarefas e o ambiente físico e social com o propósito de alcançar objetivos específicos.
- **Usuário:** pessoa que interage com o produto.
- **Objetivo:** resultado pretendido.
- **Tarefa:** conjunto de ações necessárias para alcançar um objetivo.
- **Produto:** parte do equipamento (hardware, software e materiais) para o qual a usabilidade é especificada ou avaliada.
- **Medida:** valor resultante da medição e o processo usado para obter tal valor.

Estes conceitos são componentes de uma estrutura que permite medir e especificar a usabilidade a partir da identificação dos objetivos e decomposição da eficácia, eficiência e satisfação, e os componentes do contexto de uso em subcomponentes com atributos mensuráveis e verificáveis. A Figura 5.1 está representando o relacionamento entre os componentes da usabilidade:

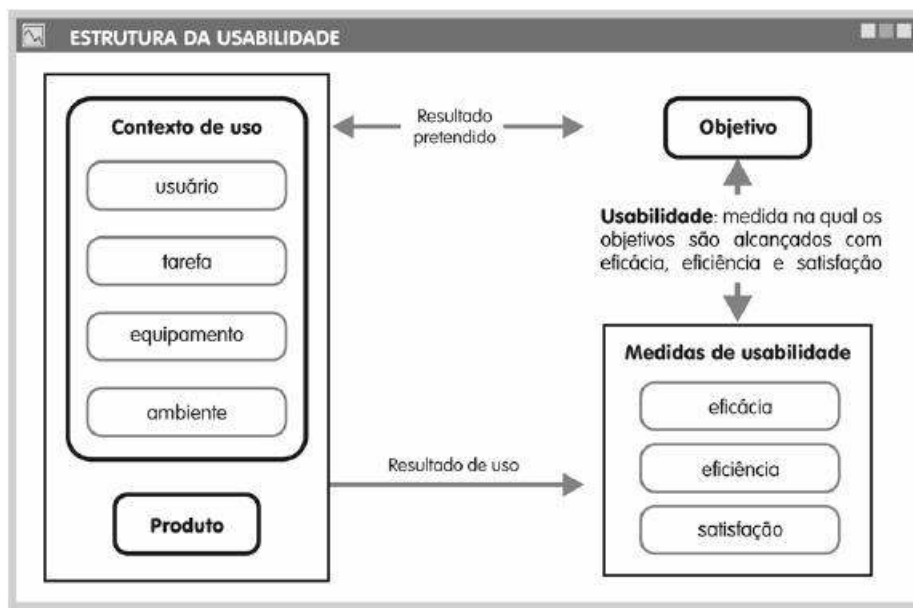


Figura 5.1: Estrutura da Usabilidade

Observa-se que para especificar ou medir a usabilidade é necessário saber:

- quais os objetivos pretendidos;
- uma descrição dos componentes do contexto de uso com detalhes suficientes para representar os aspectos relevantes do contexto;
- identificar quais os valores reais ou desejados de medidas de eficácia, eficiência e satisfação no uso do produto. Esses valores podem ser objetivos ou subjetivos de acordo com o contexto, o produto e a possibilidade de resposta do usuário em ensaio de usabilidade.

Deve-se medir a aderência do sistema aos requisitos da norma (na opinião dos usuários), avaliando-o quanto ao desempenho em tolerar erros, dar suporte à individualização, oferecer recursos de controlabilidade das tarefas dos usuários, conformidade com as expectativas, auto-descrição, adequação ao aprendizado e adequação à tarefa.

### 5.2.1 A Importância do Usuário

Meister e Enderwick [Meister and Enderwick, 2002] reforçam que o usuário pode ser encarado sob diversos aspectos, tais como:

- o usuário como sujeito em testes de usabilidade e prototipagem. Neste caso, o foco é avaliar sua performance com relação a um equipamento particular ou a uma característica deste equipamento;
- o usuário como alguém que tem preferências, particularmente com produtos comerciais;
- o usuário como especialista em assuntos específicos, como provedor de informações.

Diversos autores incluem em suas recomendações para projeto de sistemas e de interfaces a preocupação com o usuário como requisito inicial para o desenvolvimento.

Conforme [Constantine, 1999], o modelo centrado no usuário apresenta um grupo de questões sucessivas necessárias para considerar-se um projeto centrado no usuário:

- Quem são os usuários e como eles querem relacionar-se com o sistema?
- Que tarefas os usuários tentarão realizar no sistema a ser projetado?
- O que eles necessitam do sistema e em que ordem para realizar suas tarefas, e como isto deve ser organizado?
- Quais são as condições operacionais sob as quais o sistema será usado?
- Que aspecto a IU deve apresentar e como isso se deve comportar?

Ao se adotar o enfoque centrado no usuário a tecnologia deve servir para atender as necessidades e características humanas.

O usuário, consciente ou inconscientemente, irá calcular a razão de custo-benefício em termos de quantidade de esforço físico ou esforço mental que despenderá para atingir o resultado desejado com o uso de determinado produto de tecnologia. Se o benefício não for significativamente maior que o custo, o usuário rejeitará a nova tecnologia.

A abordagem de projeto para usuário é a abordagem dominante em estudos e projetos que envolvam Ergonomia. Esta abordagem pressupõe que o projetista detém conhecimento suficiente para agir em nome do usuário, por meio da aplicação de conceitos e princípios gerais desenvolvidos e validados por meio de pesquisas. Esse tipo de abordagem é aplicado a projetos que são caracterizados por sua generalidade, que possuam um público-usuário muito amplo, como os grandes portais de informação. Nesse caso é o projetista quem decide o que é melhor para o usuário, de acordo com a ciência experimental de métodos empíricos tradicionais, que junto com dados, testam hipóteses sobre

comportamento humano. Teorias gerais, leis e princípios são levantados para que possam ser aplicados em nome dos indivíduos.

A abordagem de projeto com usuário nasceu exatamente como crítica à prática descrita anteriormente e se enquadra dentro de uma visão de projeto participativo, que vê o usuário como colaborador no processo de projeto. Esta abordagem leva em consideração questões inerentes ao ser humano, como ambições, crenças, emoções e outras. O papel do projetista, dentro desta ótica, é o de agente de mudança, ao invés de alguém que possui conhecimento total. Sua aplicação se dá mais eficazmente em projetos localizados e bem identificados, como em intranets ou sistemas especialistas, pois os usuários conhecem sua realidade e podem decidir por si só.

Percebe-se que a aplicação pura de qualquer uma das duas abordagens levaria a projetos deficientes. Sabe-se, inclusive, que é praticamente impossível haver uma ou outra abordagem de forma pura, pois, em alguma instância, usuários representativos foram considerados na geração das recomendações de projeto e, por outro lado, o conhecimento do projetista também influencia decisões finais, mesmo que tenham sido elaboradas pelos próprios usuários.

Disso se conclui que uma abordagem mais eficiente será aquela que considerar o conhecimento de regras, normas e recomendações para projeto que o projetista possui, e que também busque informações reais e contextualizadas extraídas diretamente dos usuários. Essa visão, de projeto para usuário com usuário, tem condições de conduzir a projetos mais adequados em atender as reais necessidades do usuário e da tarefa.

### 5.2.2 Importância da Avaliação

O sucesso do desenvolvimento, implantação ou compra de um *software* interativo está intimamente ligado ao projeto da atividade de verificação e validação de sua usabilidade. Pode-se gerar desperdícios, desenvolvendo idéias ou soluções equivocadas, ou ao contrário, concentrar recursos nas idéias que poderão satisfazer os usuários. Pode-se liberar um produto de *software* para o mercado ou ao contrário, solicitar alterações de última hora em sua interface com o usuário. Pode-se suspender uma transação de compra e venda de *software* ou assinalá-la, com base nos resultados de uma avaliação de usabilidade.

No processo de avaliação verifica-se se os requisitos esperados para a atividade foram efetivamente alcançados. A retroalimentação sobre o que deu e sobre o que não deu certo durante a realização dos testes é fundamental para garantir o sucesso de novos projetos de avaliação.

## 5.3 Projeto de Interface Humano-Computador

Nas últimas décadas do século XX, o computador era de domínio de especialistas ou pessoas treinadas. Porém sua presença nas mais variadas atividades do ser humano é cada vez mais intensa (aparelho celular, caixa eletrônico de sistemas bancários, computadores de bordo, aparelhos domésticos programáveis e outros equipamentos interativos). Esse avanço é proporcionado principalmente pela preocupação dos grandes fabricantes de software de disponibilizarem produtos que facilitam a vida por estarem mais adequados às necessidades dos usuários. É esta interação humano-computador que vem sendo aprimorada e tornando-se uma nova fronteira de geração de conhecimento [da Silva, 2006].

### Concepção de uma interface

"Um computador pensa (se pensar é a palavra correta no caso) através de minúsculos pulsos de eletricidade, que representam um estado "ligado" ou um estado "desligado", um 0 ou um 1. Os seres humanos pensam através de palavras, conceitos, imagens, sons, associações. Um computador que nada faça além de manipular seqüências de zeros e uns não passa de uma máquina de somar excepcionalmente ineficiente. Para que a mágica da revolução digital ocorra, um computador deve também representar-se a si mesmo ao usuário, numa linguagem que este compreenda" [Johnson, 2001].

O computador se auto representa para o homem através de imagens, sons, associações hipertextuais. Esse meio, no qual a linguagem é a metáfora do pensamento humano e que os projetistas de programas de computador utilizam para alcançar a compreensão do usuário, é a interface.

Com a interface, o ponto de vista do homem é priorizado quando o projetista de interface faz uso da metáfora de uma tecnologia ou de um ambiente que o usuário do computador já dominava antes de ter essa experiência digital. Assim, o usuário tem uma curva de aprendizagem em contato com a interface, que diminui de acordo com o conhecimento anterior para o cumprimento da sua tarefa. Por essa razão que é comum encontrar uma série de termos para identificar determinados programas que funcionam através de interfaces com nomes de ambientes e tecnologias conhecidas por todos, até mesmo por quem nunca utilizou um computador: "salas de bate-papo", "correio eletrônico", "loja virtual"; sempre proporcionando uma familiarização com o sistema.

A interface precisa ser concebida de maneira a garantir uma boa relação do usuário-máquina e, ao mesmo tempo, aproveitar as potencialidades e funcionalidades da tecnologia de maneira a favorecer também a relação custo x benefício. Uma interface com estas características é certamente uma boa interface. Esta visão define a Interface Humano-Computador como um projeto que visa melhorar a usabilidade produzindo aplicações

seguras, úteis e funcionais. O projetista deve ter como regra básica o conhecimento do usuário antes de qualquer outra coisa. A usabilidade da aplicação produzida será identificada pelas seguintes proposições:

- Fácil de aprender: não requer treinamento formal;
- Fácil de memorizar: lembrar como usar mesmo após algum tempo;
- Maximizar a produtividade: a tarefa é feita rápida e eficientemente;
- Minimizar taxas de erros: permitir que o usuário tome conhecimento dos erros;
- Maximizar a satisfação do usuário: dar-lhe confiança e segurança.

Existem várias etapas no ciclo de desenvolvimento de um projeto de IHC. Dentre elas temos:

- a constituição da equipe envolvida (que deve conter usuários), busca de informações e requisitos;
- a prototipação que visa definir o nível de detalhes, detectar problemas de usabilidade e corrigi-los. Nesta fase tanto pode ser utilizado simulações simples como papel e tesoura, quanto ferramentas de projetos de interface como Visual Basic, Delphi, etc. ;
- uma etapa de avaliação, antes da implementação do protótipo, usando modelos que permitem prever desempenho com a desvantagem de não ter a participação do usuário.
- avaliação no pós-protótipo, usa o protótipo completo e técnicas heurísticas de inspeção;
- Finalmente tem-se os testes com o usuário que tem como vantagem detectar problemas não previstos pelos especialistas.

## 5.4 Interfaces para Sistemas de Busca

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 tem 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 relembrar 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 a 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 relembrar [dos Santos, 2006b].

[Moraes, 2004] apresenta um experimento realizado por Michael Peterson 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.:

- 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.
- Interface B Preenchimento de formulário: Possui quatro caixas de texto que são rotuladas de acordo com um operador booleano específico.
- Interface C Linha de comando: Possui somente um campo simples de texto e é a referência óbvia das interfaces típicas de sistemas de busca.
- 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.

O estudo de Peterson foi realizado com estudantes universitários dos cursos de Biblioteconomia e de Biologia. E obteve quatro importantes conclusões:

1. Os critérios definidos 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.

Os resultados de performance apresentaram a Interface A (menu-conector) como a mais adequada para realização de buscas na 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.

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 uma conclusão 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.

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 [dos Santos, 2006b]. 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 [dos Santos, 2006b].

Considerando que a própria experiência de busca pode ser uma experiência muitas vezes demorada e frustrante. O acesso às telas de consultas deve se dar da maneira mais imediata possível e com o mínimo de desvios [dos Santos, 2006b].

Dentre as regras para a navegação eficaz, [Fleming, 1998] inclui a economia. Por "economia" entende-se que quanto menores forem o esforço, o tempo gasto e a quantidade de ações requeridas, menos cansativo será o processo de interação do usuário com o sistema. Já [Bastien and Scapin, 1993] incluíram, em seus critérios ergonômicos, o item "brevidade", especificado nos subcritérios "ações mínimas" e "concisão", a fim de reduzir a carga de trabalho do usuário.

A interface deve deixar claro para o usuário qual o conteúdo principal. Deve apresentar-se como confiável, seja pela apresentação dos créditos das informações contidas, seja pela referência direta às fontes do conteúdo. O mais importante é disponibilizar para usuários

com diferentes perfis, recursos de busca e de navegação que permitam tanto a busca por um item conhecido, quanto a busca por exploração.

[George, 2005] descreve o processo de realização de teste de usabilidade e projeto de um website de biblioteca. Foi realizada uma pesquisa na web para determinar necessidades, procedimentos para o protótipo e completar o processo com o projeto final e teste de usuário. Os resultados indicaram muitos pontos fracos referentes à navegação, ao design de telas e à nomenclatura usada, o que orientou para outras revisões e o produto final. Os testes indicaram que: o uso de cores e de gráficos colabora para atrair atenção do usuário; fontes e localização aumentam a visibilidade; dividir em grupos e identificar com palavras-chaves aumenta a capacidade de leitura e a consistência aumenta a usabilidade.

Uma avaliação das interfaces da Scientific Library Online (SciELO), realizada em [Oliveira, 2001], buscou identificar os pontos de melhoria para o uso do sistema, especificamente em seus aspectos de apresentação, de navegação e funcionalidades. Com a aplicação de métodos qualitativos, a autora obteve a visão de usuários novatos e de usuários experientes sobre as qualidades e deficiências vividas durante a interação com a interface do SciELO. Dentre as sugestões de melhoria, pode-se citar "um maior cuidado com a condução do usuário durante o diálogo. A interface deveria utilizar mensagens convidativas para os campos de preenchimento".

[Oliveira, 2001] afirma ainda que os usuários tiveram dificuldades em recordar os pontos fracos da interface. No entanto, durante a observação da interação, as verbalizações trouxeram à tona as dificuldades experimentadas. A pesquisadora associa tal fato a uma predisposição do usuário em aceitar os problemas existentes na interface.

Realizado a partir das reclamações de usuários sobre dificuldades de busca e recuperação de informação, [Lima, 2001] desenvolveu estudo que teve como objeto a informação recuperada. A partir de uma abordagem de projeto centrado no usuário, foram elaborados requisitos de usabilidade e "sugeridos por meio de questionário e submetidos aos futuros usuários, para que priorizassem os mais relevantes para a concepção de uma interface ergonômica de biblioteca virtual". Neste estudo, foram apontadas ainda algumas recomendações relevantes para projeto de interfaces para recuperação de informação. Dentre as quais se destacaram:

- a interface deve auxiliar o usuário a formular suas consultas, a selecionar entre as fontes de informação disponíveis, a entender os resultados obtidos e a manter um histórico sobre os progressos da pesquisa;
- a interface deve ser capaz de salvar e recuperar as configurações feitas pelo usuário no próximo acesso à biblioteca;
- as interfaces devem ser configuráveis, como uma forma de ajudar o usuário a alterar propriedades de acordo com suas preferências.

Jasek [da Silva, 2004], líder do Elsevier User Centered Design Group, publicou uma lista de recomendações para o projeto de interface. As tabelas 5.1 e 5.2 apresentam as recomendações apresentadas pelo autor.

|                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>CONSISTÊNCIA</b>           | <b>Oferecer link para Ajuda em todas as páginas</b><br>Colocar um link para ajuda no canto superior direito de todas as páginas.<br><b>Usar elementos de design consistentes</b><br>Usar fontes e cores de forma consistente para dar uma aparência uniforme e profissional. Também deve-se esforçar para ser consistente em outras áreas tais como <i>layout</i> das páginas e na terminologia, bem como na maneira que os usuários interagem com o sistema. |
| <b>ORIENTAÇÃO E NAVEGAÇÃO</b> | <b>Usar uma barra de navegação</b><br>Para orientar os usuários no sistema, deve-se usar uma barra de navegação de maneira consistente e bem aplicada. De maneira geral, deve aparecer no topo das páginas, com destaque para a seção que o usuário visita no momento.                                                                                                                                                                                        |
| <b>HIPERTEXTO E LINKS</b>     | <b>Tratar <i>links</i> de acordo com as convenções</b><br>Deve-se sublinhar os <i>links</i> e usar uma cor diferente para indicar <i>links</i> que os usuários já visitaram.                                                                                                                                                                                                                                                                                  |
| <b>LAYOUT DA PÁGINA</b>       | <b>Usar a área útil da página sabiamente</b><br>Certificar-se que o conteúdo principal de cada página herdará maior espaço possível. Deve-se tentar minimizar a quantidade de espaço do logotipo e da barra de navegação que ocupam o topo da página, desta forma os usuários poderão ver o conteúdo principal sem rolagem.                                                                                                                                   |

Tabela 5.1: Recomendações para projeto de interface (1ª parte).

O uso de sistemas para recuperação de informação é feito em situações específicas de necessidades informacionais, o que exige não só que a interface tenha sido projetada para facilitar o uso e a aprendizagem, como também exige que o conteúdo esteja organizado, rotulado e possa ser acessado por meio de um sistema de navegação ou de busca.

## 5.5 Padrões para Interface

Padrões podem ser descritos como as melhores práticas dentro do domínio do projeto, entretanto, não se pode medir a qualidade de um objeto contando os padrões que existem nele. Cada projeto tem um contexto e precisa-se resolver um problema. Um dado padrão pode ser uma solução pobre neste contexto. Nada pode substituir um bom julgamento. Nem pode ser substituto para um bom processo de design, o qual ajuda a encontrar e corrigir erros de design.

[Tidwell, 2005] apresenta padrões para a criação de interfaces e indica quais perguntas

|                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ESTÉTICA E GRÁFICOS</b>                          | <p><b>Usar poucas cores e mínimos gráficos</b></p> <p>Usar somente três ou quatro cores combinadas, a fim de evitar uma aparência circense.</p> <p>Evitar ou minimizar o uso de texto animado, rolante ou com piscamento.</p>                                                                                                                                                                                                                                                                                                                                             |
| <b>FLEXIBILIDADE E EFICIÊNCIA DE USO</b>            | <p><b>Organizar a informação de múltiplas formas</b></p> <p>Organizar informação por tipo de material</p> <p>Oferecer cruzamentos de links quando for possível.</p> <p><b>Minimizar o número de cliques que o usuário deve dar.</b></p> <p>Coloque <i>links</i> para diferentes tipos de materiais e fontes logo na primeira página.</p>                                                                                                                                                                                                                                  |
| <b>ADEQUAR O SISTEMA ÀS ATIVIDADES DOS USUÁRIOS</b> | <p><b>Organizar a interface com base nas tarefas do usuário e em sua freqüência</b></p> <p>A organização deve atender às necessidades de pesquisa dos usuários</p> <p><b>Tornar a busca clara</b></p> <p>Certificar-se que os usuários podem facilmente ver e entender que materiais ou conteúdos podem ser buscados pelo sistema.</p> <p><b>Assegurar bom desempenho</b></p> <p>Para aumentar a aceitação, o sistema não deve sofrer de desempenho fraco, apesar da potência ou confiabilidade do equipamento, deve-se atentar para um bom <i>layout</i> de páginas.</p> |
| <b>TESTE DE USABILIDADE</b>                         | <p><b>Testar a usabilidade da interface</b></p> <p>Conduzir um estudo de usabilidade para certificar-se que o sistema atende às necessidades do usuário.</p>                                                                                                                                                                                                                                                                                                                                                                                                              |

Tabela 5.2: Recomendações para projeto de interface (2ª parte).

devemos fazer aos usuários para entender qual o objetivo dele. Uma boa interface não inicia com figuras, inicia com o entendimento das pessoas: o que eles gostam, porque eles usam aquele software e como eles podem interagir com ele. Quanto mais você sabe sobre os usuários, e mais você compartilha os sentimentos com eles, mais efetivamente você pode projetar para eles. O software, no fim de tudo, é simplesmente um meio para o fim, objetivo do seu uso para as pessoas. Quanto melhor se satisfaz aquele fim, mais feliz aqueles usuários serão.

Se há o conhecimento do que os usuários esperam da aplicação, e se é feita uma escolha cuidadosa de ferramentas de idiomas, controles e padrões então o usuário achará aquela interface "familiar".

O primeiro passo no projeto de uma interface é compreender o que os seus usuários realmente estão tentando realizar. Fazer as perguntas certas pode ajudar na conexão entre os objetivos e o processo de desenho. Encontre o que os usuários realmente gostam e o que eles pensam e sentem.

Especificamente, o que se deve aprender:

- Os objetivos no uso do sistema;

- As tarefas específicas que eles buscam;
- A linguagem e as palavras que eles usam para descrever o que deve ser feito;
- Habilidade em usar software similares ao que está sendo desenvolvido;
- As suas atitudes em relação ao que está sendo projetado, e como os desenhos diferentes poderiam afetar aquelas atitudes.

Infelizmente essa fase de descobrimento do usuário consome muito tempo no ciclo de desenho. Se não há tempo para métodos formais, é melhor apenas fazer uma reunião informal com poucos usuários do que não fazer qualquer descoberta.

Uma interface que suporta padrões que descrevem comportamentos humanos ajudará o usuário a atingir melhor seu objetivo que uma interface que não tem esse suporte. Abaixo apresenta-se alguns desses padrões:

- Exploração Segura (*Safe Exploration*) - Permite exploração na interface sem ter perdas ou ocorrerem problemas
- Gratificação Instantânea (*Instant Gratification*) - Resultados imediatos para as ações.
- Satisfação (*Satisficing*) - O usuário não quer gastar tempo no aprendizado.
- Mudanças de Percurso (*Changes in Midstream*) - O usuário pode mudar o que está fazendo durante a execução.
- Escolhas Diferentes (*Deferred Choices*) - usuários não querem despendar tempo em respostas.
- Construção Incremental (*Incremental construction*) - O usuário vai montando, construindo, buscando o melhor.
- Hábitos (*Habituation*) - O usuário já tem alguns hábitos em relação ao uso de interface, como por exemplo, teclas de atalho.
- Memória Espacial (*Spatial Memory*) - Onde os elementos estão??
- Memória Prospectiva (*Prospective memory*) - Usada quando planeja-se fazer algo no futuro.
- Repetição (*Streamlined Repetition*) - Encontrar uma forma de se reduzir repetições de ações.
- Teclado Apenas (*Keyboard only*) - Minimizar/ evitar o uso do mouse.

- Conselhos de Outros (*other people's advice*) - A opinião de outros usuários influencia.

A partir destes padrões de comportamento/expectativas dos usuários e do conhecimento das suas necessidades inicia-se a construção da interface.

Deve-se sempre lembrar que usuários normalmente ficam mais satisfeitos quando encontram algo próximo ao que ele já conhece, ao que já existe.

A maioria das aplicações são organizadas de acordo com uma ou mais aproximações seguintes:

- Lista de objetos, por exemplo, uma caixa de e-mails.
- Lista de ações, por exemplo, navegar, comprar, registrar.
- Lista de categorias, por exemplo, saúde, tecnologia.
- Lista de ferramentas, por exemplo, calendários, agenda de endereços.

A seguir uma análise mais detalhada sobre como essas quatro categorias são realizadas.

### **Lista de Objetos**

Útil para coleções, resultados de pesquisas. Estes objetos serão apresentados em listas selecionáveis, tabelas, árvores, entre outros que sejam apropriados. Algumas características que podemos citar de forma a organizar tais objetos:

- Dados apresentados de forma linear, usualmente ordenada.
- Tabelas 2D, que permite ao usuário ordenar ou realizar filtros a partir de vários critérios.
- Uma hierarquia que agrupa itens em categorias.
- Uma hierarquia que revela relacionamentos: pai/filho, etc.

### **Lista de ações**

Uma vez tendo selecionado um objeto, o que deve ser feito com ele?

Esta é a pergunta chave para esta categorização. Ela é tratada em termos de verbos. É mostrada muitas vezes em menus. Pode ser apresentada diretamente na interface com as opções de ações e uma explicação em texto plano sobre o significado de tal ação.

### **Lista de categorias**

Muito comuns em websites e referências online. O sucesso depende também de quão bem o projetista antecipa o que os usuários estão procurando quando vem pela primeira vez a interface. Existem muitas formas de categorização, como alfabética, cronológica e geográfica.

### **Lista de Ferramentas**

Funciona bem quando os usuários tem uma clara expectativa do que as ferramentas devem ser ou fazer.

Um modelo organizacional à base de ferramentas falha quando os nomes das ferramentas se misturam com nomes de ações, tarefas, ou objetos. Essa mistura tende a quebrar as expectativas dos usuários do que aqueles itens se apresentam como escolhas de menu, itens de lista, botões, ou links, especialmente para novos usuários que não sabem o que os nomes das ferramentas significam ainda. Os usuários observados muitas vezes não articulam bem esta confusão. O projetista deve olhar os usuários enquanto este vê a interface para perceber os sinais sutis da confusão, mas não deve perguntar ao usuário sobre eles diretamente.

Organizações lineares (disposição linear dos objetos) são as formas mais comuns de apresentação das ferramentas.

Deve-se traduzir os conceitos de projeto em uma estrutura física de janelas, painéis e controles. A aplicação usará múltiplas janelas, uma simples janela com muitos painéis, ou uma janela como página da web? Ou alguma combinação entre elas?

Tem-se que tomar cuidado com o objetivo do uso da interface, das limitações e desejos dos usuários. Uma aplicação com múltiplos painéis é difícil de se usar sem um mouse.

Softwares de desktop e aplicações web dão maiores possibilidades de escolhas. Não existem regras para determinar o que é melhor, mais existem algumas sugestões para guiar o desenho da interface. Antes da decisão devem-se analisar as tarefas e os seus usuários e como irão trabalhar. A compreensão obtida deve direcionar as decisões.

A figura 5.2 mostra três diferentes estruturas físicas, que são explicadas a seguir:

**Múltiplas Janelas:** Trabalham melhor em aplicações sofisticadas quando os usuários querem customizar o layout. Alguns usuários acham múltiplas janelas irritante ou confuso.

**Uma-Janela:** Aplicações web simples trabalham melhor com este modelo. É apresentada uma única página a cada vez.

**Painéis:** Ótima para usuários que querem ver muita coisa ao mesmo tempo e dispende pouco esforço no gerenciamento de janela. As pessoas aceitam a idéia de interagir em um painel e ver o resultado em outro.

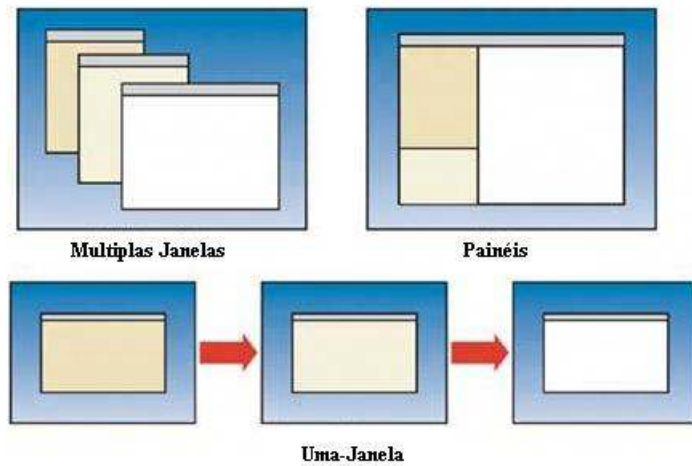


Figura 5.2: Estruturas Físicas

### Padrões de Projeto

Além dos padrões de comportamento/ expectativa do usuário apresentados, temos também padrões para o desenho da interface. A junção de conteúdo com estrutura física fornece alguns padrões para a realização de um bom projeto de interface. Algumas combinações trabalham muito bem:

- Seletor de dois painéis (*Two-Panel Selector*)
- Tela mais Paleta (*Canvas Plus Palette*)
- Navegação em Uma Janela (*One-Window Drilldown*)
- Visões alternativas (*Alternative Views*)

Os seguintes padrões tratam com o conteúdo abstrato. O Wizard pode ser implementado com qualquer número de apresentações físicas. Os Extras sob Demanda e os Desvios Intrigantes descrevem modos adicionais de dividir o conteúdo.

- Wizard (*Wizard*)
- Extras sob demanda (*Extras on Demand*)
- Desvios Intrigantes (*Intriguing Branches*)

Muitos padrões contribuem em vários níveis para a aprendizagem de uma interface. O conjunto Ajuda Multi-Nível traz formas de integrar a ajuda diretamente na aplicação:

- Ajuda Multi-Nível (*Multi-level Help*)

A seguir apresenta-se o que são estes padrões, quando, porque e como usá-los:

1. Seletor de dois painéis:

- O que é: Coloque dois painéis lado a lado na interface. No primeiro, mostra o conjunto de itens que o usuário pode selecionar, no outro, mostre o conteúdo do item selecionado.
- Quando usar: Para apresentação de listas de objetos, categorias ou ações. Fisicamente, o local de exibição da interface deve ser grande o suficiente para mostrar dois painéis separados de cada vez.
- Por que: É comum e poderoso. Os usuários aprendem a usá-lo rapidamente. Vantagens: Reduz esforço físico, reduz carga visual cognitiva e reduz carga na memória do usuário.
- Como: Coloque a lista selecionável no painel de cima ou da esquerda e o painel de detalhes abaixo ou à direita. Quando o usuário seleciona um item, imediatamente é mostrado o conteúdo, ou detalhes no segundo painel.
- Exemplo: Microsoft Windows Explorer.

2. Tela mais Paleta:

- O que é: Coloque um ícone da paleta próxima à tela em branco. O usuário clica nos botões da paleta para criar objetos na tela.
- Quando usar: Para desenhar em algum tipo de editor gráfico.
- Por que: É um mapeamento natural a partir de objetos físicos familiares para o mundo virtual
- Como: Apresente uma área grande e vazia como a tela. A paleta pode ser um *grid* de botões. A paleta deve ficar à esquerda ou acima da tela.
- Exemplo: Paint.

3. Navegação em Uma Janela:

- O que é: Mostre cada uma das páginas da aplicação dentro de uma janela. Quando o usuário navega através de um menu de opções ou detalhes de um objeto, a janela é completamente substituído com uma nova página.
- Quando usar: A aplicação consiste em muitas páginas ou painéis de conteúdo. Pode ser aplicado para construção em dispositivos com restrições de espaço ou com complexidade limitada.

- Por que: Mantenha simples. As opções são claras em cada estágio e o usuário não precisa focar a atenção em nada mais.
- Como: Estruture o conteúdo em painéis que cabem bem na janela: não muito grande e não muito pequeno. Estruturas lineares simples ou hierárquicas trabalham melhor. Em testes de usabilidades, tenha certeza de que os usuários podem usar tudo sem perder nada.
- Exemplo: Menus de iPod.

#### 4. Visões alternativas:

- O que é: Deixe o usuário escolher entre visões alternativas que são estruturalmente diferentes.
- Quando usar: Web page, editor, ou qualquer aplicação que mostre o conteúdo formatado de alguma forma.
- Por que: Não podemos sempre acomodar todas as possibilidades de cenários em um desenho simples. Existem diferentes tecnologias para o usuário final. As preferências dos usuários variam. Pode existir necessidade de uma visão diferente temporariamente.
- Como: Escolha alguns cenários que o modo normal da aplicação não pode facilmente servir. Desenhe visões especializadas para eles, e apresente essas visões como alternativas.
- Exemplo: Visões alternativas de arquivos em um sistema de arquivo.

#### 5. Wizard:

- O que é: Conduza o usuário através da interface passo a passo, executando tarefas em uma ordem prescrita.
  - Quando usar: Para uma tarefa longa ou complicada, que seja nova para o usuário e realizada com pouca frequência.
  - Por que: Dividir para conquistar. Por dividir as tarefas em uma sequência de pedaços, cada um dos quais o usuário pode enfrentar com um discreto "espaço mental", a tarefa é efetivamente simplificada.
  - Como: Dividindo a tarefa: Quebre as operações da tarefa em uma série de pedaços ou grupos. Estes grupos devem ser apresentados em uma sequência. Colocar escolhas relacionadas juntas é mais simples para o usuário.
- Estrutura física: Wizards que apresentam cada passo em uma página separada, navegação com botões de retrocesso e avanço são a implementação mais óbvias e conhecidas deste padrão.

## 6. Extras sob demanda:

- O que é: Mostrar o conteúdo mais importante mas esconder o resto. Deixe o usuário alcançá-lo por uma via única, com um simples gesto.
- Quando usar: Há muita coisa para ser mostrada na página, mas algumas não são muito importantes.
- Por que: Uma interface simples é frequentemente melhor que uma mais complexa, especialmente para novos usuários.
- Como: Corte impiedosamente a interface abaixo dos itens mais importantes. Coloque o restante em sua própria página ou seção. Ocultar esse ponto, por *default*; na recém-simplificada interface, colocar um botão ou *link* claramente marcados para o restante, como "Mais Opções". Muitas interfaces usam setas ou divisas, " > >", como parte do *link* ou botão. Outros utilizam "...", especialmente se o botão abre uma nova janela ou página.
- Exemplo: Caixa de diálogo de cores no Windows

## 7. Desvios Intrigantes:

- O que é: Colocar links para conteúdos interessantes em lugares inesperados, e nomeá-los de uma forma que atraia os usuários curiosos.
- Quando usar: O usuário se move ao longo de um percurso linear como um texto narrativo, uma tarefa bem-definida, uma apresentação, um filme, etc. Pretende-se apresentar conteúdo adicional que não é o principal foco de atenção, no entanto.
- Por que: Pessoas são curiosas.
- Como: Crie "portas" para este conteúdo que sejam um apelo ao usuário. Há uma arte ao fazê-lo. Quando em dúvida, faça um teste de usabilidade, com uma amostra representativa de seu usuário base.

## 8. Ajuda Multi-Nível:

- O que é: Use uma mistura de técnicas de ajuda leve e pesada para dar apoio aos usuários com diferentes necessidades.
- Quando usar: Sua aplicação é complexa. Alguns usuários provavelmente necessitarão de um verdadeiro sistema de ajuda, mas a maioria dos usuários não terão tempo para usá-lo. Pode-se estender a ajuda para dar apoio aos impacientes e/ou usuários ocasionais.
- Por que: Usuários de praticamente todos artefatos de software têm necessidades diferentes de níveis de apoio para as tarefas que eles estão tentando realizar.

- Como: Criar ajuda em vários níveis.
- Exemplo: O Excel possui várias técnicas de ajuda, integradas na interface.

A partir das definições dos padrões acima, inicia-se a criação dos protótipos da interface. Tidwell [Tidwell, 2005] traz em seu livro também os padrões utilizados no projeto, tais como padrões para:

- Formas de movimentação - os problemas da navegação;
- Layout de organização dos elementos na página;
- Ações e comandos;
- Como mostrar dados complexos: árvores, tabelas e outras informações gráficas;
- Formulários e controles para entrada de dados dos usuários;
- Editores;
- Estética e um bom visual.

## 5.6 Sugestões para Interfaces de Sistemas de Recuperação de Informação

Com base em uma avaliação preliminar de como uma interface pode melhorar a busca, e baseado nos conceitos apresentados nas seções anteriores, a seguir são mostrados critérios para o projeto de uma boa interface, com foco especial na contribuição deste trabalho, a interface para RI semi-estruturada. Deve-se lembrar que características apresentadas neste capítulo são úteis para o desenvolvimento de interfaces, independente de qual seja o sistema na qual esta será aplicada. No Capítulo 6 será apresentado um protótipo de interface que contém algumas das sugestões apresentadas nesta seção. Deve-se lembrar que o protótipo é uma versão não pronta do sistema e que interações com o usuário e continuações no desenvolvimento irão complementá-lo antes que ele se torne o interface pronta.

Para apoiar a busca e recuperação de informação o sistema deverá contar com uma interface ergonômica através da aplicação de requisitos ergonômicos para atender às necessidades dos usuários e garantir a eficácia do processo de interação entre o humano e o computador. O objetivo do protótipo é auxiliar o usuário na construção de consultas semi-estruturadas com informações do contexto para alcançar alta precisão nas respostas. Isto será avaliado posteriormente.

Em [Pollock and Hockley, 1997] foi realizado um estudo com o objetivo de desenvolver um protótipo de SRI para Internet que oferecesse qualidade e aceitabilidade. Neste

estudo foi verificado que uma das grandes dificuldade dos usuários é formular os termos da pesquisa. [Cardoso, 1999] também acredita que para a maioria dos usuários o problema para encontrar a resposta exata de suas buscas é a dificuldade em formular perguntas.

Pollock e Hockley também destacam que usuários novatos acham complicado realizar as buscas e consideram que devam ser melhor orientados. Cita ainda que a forma de apresentação das opções de busca tais como, pesquisa avançada ou especialista, levou à interpretação equivocada por parte destes usuários e que os sistemas deveriam apoiá-los com instruções durante o processo de pesquisa.

[Nielsen, 1999] recomenda que antes de se projetar a interface de um site é bom observar os mais visitados/ usados. Se a grande maioria (90%) fizerem de uma forma similar isto pode ser considerado um padrão e deve ser seguido, além disso deve-se lembrar que os usuários visitam sites esperando que estes disponham de elementos para conduzir a navegação do mesmo modo com o qual estão acostumados.

A interface deve ser centrada nos usuários e projetada para facilitar a navegação, ser simples e elegante, sem abusar de gráficos e figuras pesadas. Deve haver um equilíbrio entre usabilidade e funcionalidade de modo a não subjugar o usuário com beleza de projeto e lentidão nas respostas.

O usuário não espera que uma interface contenha longos tópicos e textos ininterruptos, muitas caixas de diálogo (podem parecer confusas) e a exibição automática de gráficos e mídias. É fundamental que a interface ofereça *feedback* informativo, permita fácil reversão de ações e reduza a carga de memória do usuário, assim como, oferecer uma interface alternativa para iniciantes e experientes. Enfim, uma interface deve auxiliar o usuário a formular suas consultas.

### 5.6.1 Requisitos Propostos

Para a proposição de requisitos, deve-se lembrar dos objetivos do sistema e da interface a ser desenvolvida, e dos usuários e responder a algumas perguntas apresentadas no decorrer do texto:

- Quais os objetivos pretendidos? Criação de uma interface que auxilie um usuário que não tenha conhecimento da estrutura de um documento XML a realizar uma consulta estruturada. Que esta interface seja prática e de fácil utilização.
- Quais os valores desejados de medidas de eficácia, eficiência e satisfação? Deseja-se que os usuários consigam realizar as consultas de uma maneira fácil, tranquila, consigam proceder com a construção de uma árvore mesmo sem ter o conhecimento prévio sobre esta estrutura e sobre os meios para fazê-la.
- Quem são os usuários e como eles querem relacionar-se com o sistema? Podemos ter dois tipos de usuários, aqueles com conhecimento na área de computação, que

entendam bem o significado e formato de estruturas em arquivos XML e similares. E aqueles usuários inexperientes, que devem ser bem auxiliados pela interface no processo de criação da consulta. O segundo grupo é aquele que necessita de mais auxílio na elaboração da árvore de consulta.

- Quais tarefas serão realizadas? Consulta (com e sem estrutura).
- O que os usuários necessitam do sistema? Informações e auxílio de como construir uma consulta estruturada.

A partir do estudo baseado nas necessidades do usuário, procedemos a busca por padrões que atendam aos requisitos ergonômicos. Os padrões, se corretamente utilizados, facilitam a criação de uma boa interface.

Como sugestão, podemos utilizar como base os padrões de desejos/attitudes dos usuários citados em [Tidwell, 2005]. Dentre os apresentados devemos tomar cuidado em tratar: a Exploração Segura, pois usuários não querem perder informações e deve-se evitar ao máximo a ocorrência de problemas; A Gratificação Instantânea, talvez um dos mais importantes já que os usuários em um processo de busca de informações muitas vezes desejam encontrar o resultado buscado na primeira busca.

A satisfação deve ser pensada, já que existem sistemas de recuperação de informação que são muito bons e satisfazem o usuário com os resultados retornados, mesmo sem o uso de estrutura. A interface deve ser de fácil aprendizado.

Além deste padrões, sempre implementar o uso do teclado, deixar que todas as atividades possam ser acessadas via teclado.

## 5.6.2 Requisitos Ergonômicos

Sobre os 8 critérios ergonômicos descritos na seção 5.1, podemos enfatizar as características que são extremamente úteis para o desenvolvimento de interfaces em geral, e que devem estar presentes inclusive em uma interface de consulta para a recuperação de informação semi-estruturada, conforme mostrado a seguir:

- Interface prestativa: apresentar as informações a respeito de como as consultas devem ser realizadas. Ter sempre um ícone de ajuda na tela para esclarecer possíveis dúvidas.
- Interface fácil: seguir características que sejam semelhantes a softwares já utilizados. Por exemplo, ter como base inicial a interface de um sistema de busca muito utilizado, tais como Google, Yahoo!.
- Interface clara: não sobrecarregar a interface com informações, figuras que poluam a tela. O foco da interface é a consulta com estrutura. Manter na interface apenas informações que auxiliem o usuário na formulação de consulta.
- Interface confortável: diminuir o esforço do usuário utilizando-se de mensagens passadas de forma visual e não cansativa. Um bom exemplo para a visualização de termos

que representam algum grau de relevância ou de ocorrência é uma visualização em tamanhos de fontes diferentes, quanto maior a fonte maior o significado do texto. Tal significado pode representar, por exemplo, relevância, mais procurados ou frequência, que é o significado aplicado no protótipo. Podemos ver exemplos desta aplicação na figura 5.3.



Figura 5.3: Lista de Relevância ressaltada pelo tamanho da fonte [Submarino, 2009, Globo.com, 2009]

- Interface obediente: a interface deve apenas responder aos estímulos do usuário; é ele quem tem o controle. As dicas serão apresentadas conforme a vontade do usuário; é ele quem irá montar a consulta, solicitar ajuda para a construção e executar a consulta.

- Interface adaptável/ flexível: deixar disponíveis opções de busca simples e busca avançada. A busca simples realizará a consulta sem estrutura. A Busca avançada será com ajuda para criação da consulta estruturada e especificação de relevância.

- Interface segura: a interface deve deixar claro as instruções sobre o funcionamento do sistema, já que ocorre a diminuição das possibilidades de erros. Se a orientação na especificação da consulta for bem sistematizada evita-se erros. Esta é uma análise que também deve ser exaustivamente testada antes da interface ser colocada à disposição do usuário.

- Interface coerente: a interface deve ser padronizada, mesmo havendo mudança de contexto, de busca simples para a avançada.

- Interface expressiva: a interface deve ser simples, clara e com as informações essenciais. Evitar termos que não são de conhecimento do usuário. Se estes foram necessários, como algumas tags de documentos XML, ter a explicação do significado na ajuda do sistema.

- Interface compatível: deve-se sempre lembrar do objetivo da interface: ajudar a formulação de consultas semi-estruturada. Para se obter este tipo de interface é primordial considerar os anseios dos usuário no processo de concepção.

Quanto mais destas características citadas a interface tiver, melhor será. Uma interface ideal, pronta para uso, é aquela que engloba todas estas características. Um protótipo

pode não conter todas as características, como é o caso do protótipo apresentado no Capítulo 6.

Temos também que analisar e ter cuidado com os requisitos de diálogo, que vão garantir a qualidade dos sistemas de recuperação:

1. Apoio ao diálogo na realização da tarefa:

- (a) Recuperar dados perdidos: usar o comando "desfazer". Atentar para a padronização de teclas de atalho. Como convenção usar "Ctrl+Z".
- (b) Mostrar diálogos curtos, objetivos: os diálogos devem ser bastante didáticos e auto-explicativos.
- (c) Escolha de estilo de diálogo adequado ao usuário: Sempre que possível inserir possibilidades de alteração de estilos, diálogos, fontes que atendam às preferências do usuário. O site da Google oferece preferências de escolha, que vão desde o idioma à preferência por sugestões de pesquisa. A figura 5.4 mostra esta interface.
- (d) Diálogos e ações devem ser passo-a-passo, com instruções detalhadas: Deve-se tornar o processo tranquilo e fluido.

2. Descrição da operação:

- (a) Mensagem de Erro: utilizar a mesma linguagem do usuário. Não usar mensagens que serão reconhecidas apenas por conhecedores de informática. Sempre deixar disponível a opção da visualização breve ou detalhada do erro.
- (b) Prevenção de erros: sempre dar *feedback* de confirmação para o usuário quando realizar uma tarefa que pode ser permanente.
- (c) Entrada de dados de forma clara: deve ser bem posicionada e de fácil visualização e localização.
- (d) Informações sobre o sistema: Quanto mais próximo à linguagem do usuário, melhor. Nesta proposta devemos lembrar que exemplos de como realizar uma busca estruturada serão sempre bem vindos no sistema de ajuda. O assistente de ajuda do pacote Office da Microsoft é um bom exemplo de ajuda explicativa, clara e com linguagem de fácil entendimento. Ela direciona as dúvidas do usuário sem que ele tenha que ler todo o tópico. A figura 5.5 nos mostra este exemplo. O sistema de ajuda não será apresentado no protótipo, mais é um componente essencial da interface pronta.
- (e) Fornecer o tempo de processamento: sabemos que os usuários gostam de respostas rápidas (*feedback imediato*). Em um processo de recuperação de informação



The image shows two screenshots of the Google interface. The top screenshot is the Google Brasil homepage, featuring the Google logo, a search bar, and navigation links. The 'Preferências' link is highlighted with a red box. The bottom screenshot is the Google Preferences page, which allows users to customize their search experience. It includes sections for interface language, search language, SafeSearch filtering, number of results, results window, and search suggestions.

**Google Brasil**

Pesquisa Google | Estou com sorte

Pesquisar: ☒ a web ☐ páginas em português ☐ páginas do Brasil

[Pesquisa avançada](#)  
[Preferências](#)  
[Ferramentas de idiomas](#)

[Soluções de publicidade](#) - [Tudo sobre o Google](#) - [Google.com in English](#)

©2009 - [Privacidade](#)

a) Tela do Google com opção para Preferências destacado.

---

**Google Preferências**

Salvar preferências no final e retornar à pesquisa.

**Preferências Globais** (as mudanças aplicam-se à todos os serviços Google)

**Idioma da interface** Exibir dicas e mensagens Google em: português (Brasil)

**Idioma de pesquisa**

☒ Pesquisar páginas da Web escritas em qualquer idioma (Recomendado).

☐ Prefiro páginas escritas neste(s) idioma(s):

|                                                |                                    |                                    |                                     |
|------------------------------------------------|------------------------------------|------------------------------------|-------------------------------------|
| <input type="checkbox"/> alemão                | <input type="checkbox"/> eslovaco  | <input type="checkbox"/> húngaro   | <input type="checkbox"/> português  |
| <input type="checkbox"/> árabe                 | <input type="checkbox"/> esloveno  | <input type="checkbox"/> indonésio | <input type="checkbox"/> romeno     |
| <input type="checkbox"/> Armênio               | <input type="checkbox"/> espanhol  | <input type="checkbox"/> inglês    | <input type="checkbox"/> russo      |
| <input type="checkbox"/> Bielorrusso           | <input type="checkbox"/> Esperanto | <input type="checkbox"/> islandês  | <input type="checkbox"/> sérvio     |
| <input type="checkbox"/> búlgaro               | <input type="checkbox"/> estoniano | <input type="checkbox"/> italiano  | <input type="checkbox"/> sueco      |
| <input type="checkbox"/> catalão               | <input type="checkbox"/> Filipino  | <input type="checkbox"/> japonês   | <input type="checkbox"/> tailandês  |
| <input type="checkbox"/> chinês (simplificado) | <input type="checkbox"/> finlandês | <input type="checkbox"/> letão     | <input type="checkbox"/> tcheco     |
| <input type="checkbox"/> chinês (tradicional)  | <input type="checkbox"/> francês   | <input type="checkbox"/> lituano   | <input type="checkbox"/> turco      |
| <input type="checkbox"/> coreano               | <input type="checkbox"/> grego     | <input type="checkbox"/> norueguês | <input type="checkbox"/> Ucraniano  |
| <input type="checkbox"/> croata                | <input type="checkbox"/> hebraico  | <input type="checkbox"/> Persa     | <input type="checkbox"/> Vietnamita |
| <input type="checkbox"/> dinamarquês           | <input type="checkbox"/> holandês  | <input type="checkbox"/> polonês   |                                     |

**Filtragem SafeSearch** O SafeSearch do Google bloqueia páginas da Web que contenham conteúdo sexual explícito

☐ Usar filtro restrito (Filtrar textos e imagens explícitos.)

☒ Usar filtro moderado (Filtrar apenas imagens de conteúdo explícito. - comportamento padrão)

☐ Não filtrar os resultados de minha pesquisa.

**Número de resultados** O padrão do Google (10 resultados) proporciona os resultados mais rápidos.

Exibir 10 resultados por página.

**Janela de Resultados** ☐ Abrir uma nova janela para os resultados da pesquisa.

**Sugestões de pesquisa**

☒ Fornecer sugestões na caixa de pesquisa.

☐ Não fornecer sugestões na caixa de pesquisa.

b) Tela para a seleção de Preferências do Usuário

Figura 5.4: Preferências Google

esta resposta pode ser demorada, dependendo da consulta e do tipo de informação buscada, então deve-se sempre fornecer o tempo de processamento ou

alguma informação que mostre ao usuário como está o andamento da operação.

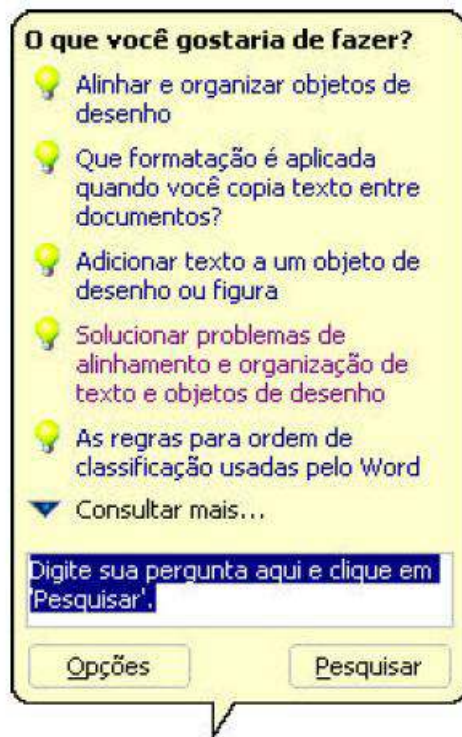


Figura 5.5: Mensagem de ajuda na linguagem do usuário

3. Usuário com controle das ações:

- (a) Usuário deve ter sempre a opção de desfazer suas ações.
- (b) Possibilitar que o usuário configure preferência da quantidade de registros que ele deseja ver. O usuário deve ter total controle sobre a visualização dos resultados.
- (c) O usuário deve ter o direito de escolher qual a próxima ação, qual o próximo campo a ser preenchido: A interface deve mostrar todas as possibilidades que o usuário tem de preenchimento e deixar livre para que ele escolha qual quer preencher e em qual ordem. Usuários são diferentes e têm necessidades diferentes em momentos diferentes.

4. Atendendo às expectativas do usuário:

- (a) Linguagem familiar ao usuário: interface com clareza e objetividade na comunicação.

- (b) Padronização de teclas e funções: manter sempre os padrões conhecidos nas execuções de tarefas. Exemplos: Para ajuda, tecla F1; para desfazer, "Ctrl+Z"; botão direito do *mouse* para obter detalhes e funções.
- (c) Se for inevitável o uso de termos técnicos, deve-se oferecer ajuda. Deixa bem visível para o usuário que esta ajuda existe.

5. Tolerância a erros:

- (a) Garantir a prevenção a erros (explicado no item 2, tópicos a e b da página 65).
- (b) Oferecer ajuda na correção dos erros: além da mensagem explicativa do erro citada acima, deve-se apresentar como corrigir tal erro.
- (c) Dar possibilidade de correção para erros comuns, ortográficos, entre outros. Se o sistema de recuperação de informação der suporte, a interface deverá trazer sugestão para correção de certos erros. O Google novamente disponibiliza esta ajuda: Quando há a digitação de algum termo de um modo diferente ele apresenta um *link* para uma outra consulta, com a frase: "Você quis dizer..." ou "Sua intenção era pesquisar..."
- (d) Robustez do sistema: o sistema não deve travar na ocorrência de um erro de entrada de dados. O usuário não deverá perder o trabalho já executado.

6. Individualização:

- (a) Recurso para deficientes: a interface deve auxiliar o sistema na questão de deixá-lo disponível para o mais variado público. Todas as possíveis limitações devem ser pensadas e geradas possibilidades de supri-las.
- (b) Opções de escolha de idioma.
- (c) Possibilitar que o próprio usuário defina teclas de atalho para facilitar o processo.

7. Aprendizado:

- (a) Opções de nível de pesquisa de acordo com a experiências do usuário: Definir a pesquisa avançada com explicação clara de como utilizá-la para alcançar resultados mais precisos.
- (b) Lições passo-a-passo para aprendizagem com exercícios. Na parte de ajuda, ensinar o usuário a usar o sistema e colocar exercícios para melhor o aprendizado. Este recurso ajudará inclusive usuários experientes quando utilizarem o sistema pela primeira vez.

Os requisitos apresentados acima devem ser analisados e inseridos na interface sempre que possível. O protótipo apresentado no Capítulo 6 contém alguns destes requisitos, a interface final deverá conter os requisitos restantes, como por exemplo, o sistema de ajuda e a tratativa de erros. 8

Ainda dentre as verificações dos requisitos necessários e proposições de formas de aplicá-los, devemos lembrar da importância que as cores têm em uma interface. Algumas possuem efeito subliminar muito forte. Vermelhos e negros são cores que devem ser usadas com parcimônia, pois seu efeito psicológico está associado a situações de tensão. Verdes e azuis podem ser utilizados em larga escala, pois são cores repousantes para os olhos, já que as lentes do olho as focam diretamente sobre a retina. O marrom pode ser reabilitado: embora o uso não seja freqüente, esta cor elimina a fadiga e estabiliza o humor. Por esta razão, é a cor mais recomendada para ambientes de trabalho [Teodoro, 2001].

Diante de todas as características de projeto apresentadas neste capítulo percebe-se que a melhor forma de concepção de uma interface é realizar o trabalho desde o princípio como o auxílio e o foco no usuário. A interface não deve ser pensada somente no fim do desenvolvimento do sistema. A aplicação dos requisitos, padrões e características apresentados gerará uma boa interface. Questionários de avaliação verificam se tais requisitos estão presentes e permitem melhorias antes que a interface esteja totalmente pronta.

No próximo capítulo apresenta-se um protótipo de interface. O protótipo respeita alguns requisitos e padrões. Sendo possível existir protótipos diferentes com a aplicação de diversos padrões, como os apresentados neste capítulo.

## Capítulo 6

# Uma Proposta de Interface para Especificação de Consultas Semi - estruturadas

Como principal contribuição deste trabalho, neste capítulo será apresentado um protótipo de interface para consulta a documentos semi-estruturados. A validação deste protótipo será feita à partir da execução de testes, mostrando que com a ajuda da interface podemos obter resultados mais precisos na RI semi-estruturada.

As características visualizadas no protótipo, apresentado na seção 6.1 são algumas aplicações das sugestões apresentadas na seção 5.6 do Capítulo 5. Na seção 6.2 tem-se um exemplo da realização de uma consulta com estrutura realizada com o auxílio da interface-protótipo, e finalizando, a seção 6.3 traz a avaliação do protótipo com relação à Precisão-0 dos resultados.

### 6.1 Prototipação

Um protótipo é uma aplicação, normalmente experimental e incompleta, que permite aos projetistas avaliarem suas idéias de desenvolvimento durante o processo de criação da aplicação pretendida. Ele deve ser construído rapidamente, com baixo custo e seu tempo de vida não é definido. Constitui uma das etapas do ciclo de desenvolvimento de um projeto de IHC.

Dentre as informações extraídas de um protótipo, podemos destacar a funcionalidade necessária ao sistema, seqüências de operação, necessidades de suporte ao usuário, representações necessárias e comunicabilidade da aplicação.

Um protótipo pode ser classificado em relação à sua função no processo de desenvolvimento, o seu objetivo de avaliação e a técnica de sua construção. Esta facilidade faz com que esta etapa do teste possa ser feita com um orçamento menor.

Na elaboração de uma interface pode-se ter mais de um protótipo. Cada um com características diferentes para a verificação de quais são melhores, ou protótipos que surjam sucessivamente a partir de correções e melhorias nos anteriores.

Como protótipo apresentaremos uma interface contendo algumas das características citadas. A interface será simples e elegante. Concebida à partir das respostas às perguntas para a verificação dos objetivos (seção 5.6.1). Tendo como foco a necessidade do usuário. Foram utilizados padrões como Exploração Segura, Gratificação e atenção ao Uso do Teclado. A tela da interface é composta por uma única janela com vários painéis. A verificação da Usabilidade é feita através de testes com usuários de diferentes níveis para a avaliação da precisão dos resultados, auxílio na criação das consultas e características visuais da interface.

Para a tela inicial, de consultas simples (consultas sem estrutura), sugere-se seguir o modelo visto no site Google, considerando que este tem se tornado o site de busca mais visitado. A motivação desta escolha é que o usuário gosta e tem conhecimento do formato da interface.

Esta tela inicial é representada no protótipo com a primeira linha (Painel A da Figura 6.1). No protótipo as duas formas de consulta estão na mesma tela para facilitar a visualização das diferentes consultas para efeito de teste. Esta visualização deverá ser também apresentada aos usuários quando forem realizados testes e questionários para a avaliação da usabilidade. Os usuários deverão escolher qual das duas formas de apresentação mais lhe agrada.

A figura 6.1 apresenta a interface-protótipo. O campo marcado com A é o local onde deve ser montada uma consulta sem estrutura, devendo o usuário apenas digitar os termos da consulta.

Os painéis B e C representam a consulta avançada, ou seja a consulta com estrutura. O painel B deve conter as palavras (termos ou estrutura) que mais ocorrem nos documentos da coleção e são filhas da palavra que se pede ajuda, do decorrer do texto será usada a palavra ajuda no sentido de auxílio na construção da árvore de consulta, e não a ajuda como é mais conhecida, sobre o sistema.

No painel C será montada a consulta estruturada. Serão inseridos as tags e os termos em caixas de texto (C1). À medida que o usuário clica no sinal de "+" (C2) surgem caixas de texto para que seja inseridos filhos do termo que foi digitado na caixa de texto correspondente, gerando assim uma árvore.

O usuário tem a possibilidade de digitar a consulta diretamente, ou começar a montá-la a partir da escolha de uma tag que está no painel D da Figura 6.1.

O painel marcado com D apresenta o resultado das consultas. Usuários devem também opinar sobre a preferência desta visualização. Da forma apresentada, ou da forma mais aplicada em sites de busca, onde os resultados aparecem abaixo do campo de pesquisa.

Na Figura 6.1, a área destacada por D1 mostra a quantidade de páginas de resposta.

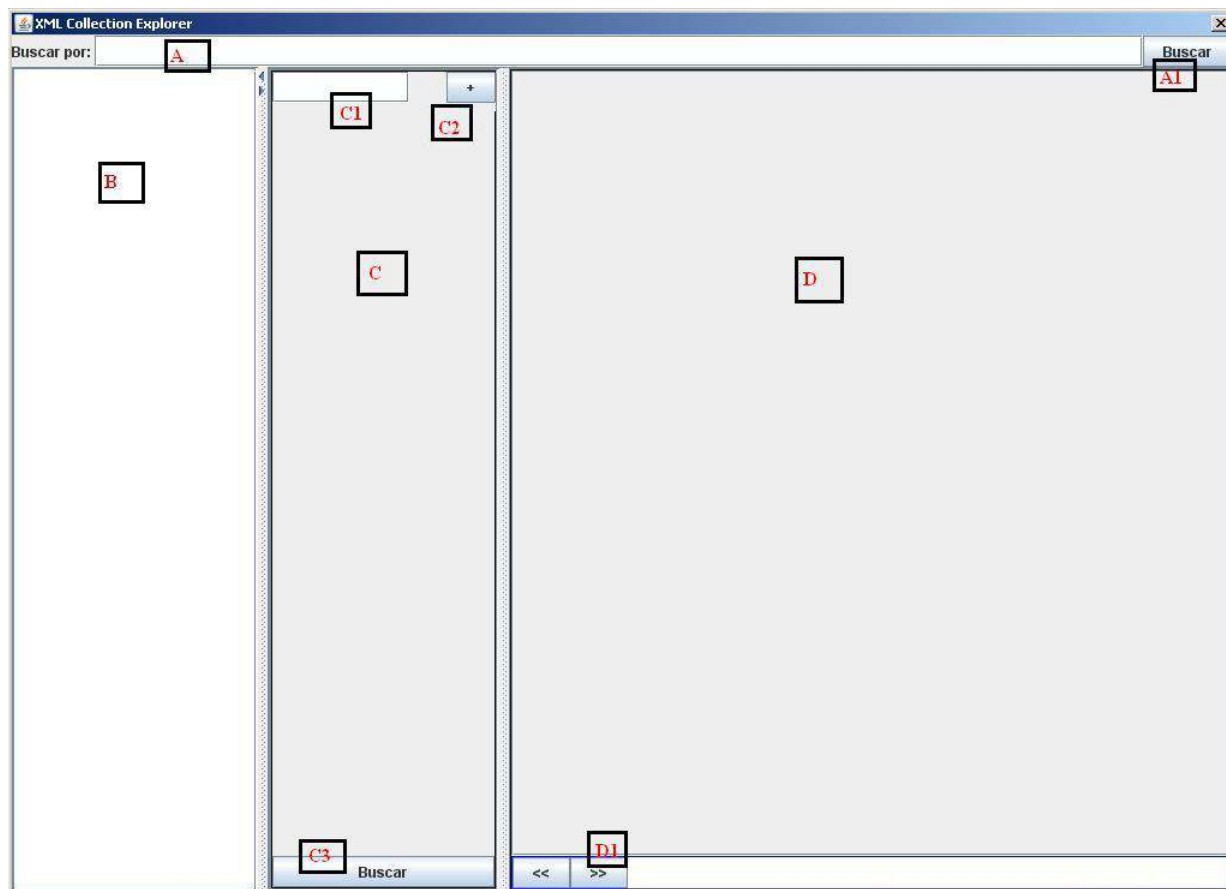


Figura 6.1: Interface-Protótipo

Os painéis B, C e D podem ter a largura alterada, e quando a sua largura é menor do que a largura do conteúdo surge uma barra de rolagem para que o usuário possa visualizar todo o conteúdo. O mesmo ocorre com a altura nos painéis C e D.

Abaixo apresentamos cada um dos painéis separadamente:

A Figura 6.2 mostra uma consulta sem estrutura: consulta dos termos "java language". Após a digitação dos termos o usuário solicita a realização da consulta pelo uso do botão (marcação A1 da Figura 6.1) ou apertando a tecla Enter.



Figura 6.2: Início de uma consulta sem estrutura.

A Figura 6.3 mostra o painel B preenchido, com tags que são filhos do termo já

digitado, no caso do exemplo, `article`. Podemos visualizar nesta figura que a tag marcada está agora na caixa de texto do painel da direita, que representa o painel C da Figura 6.1.

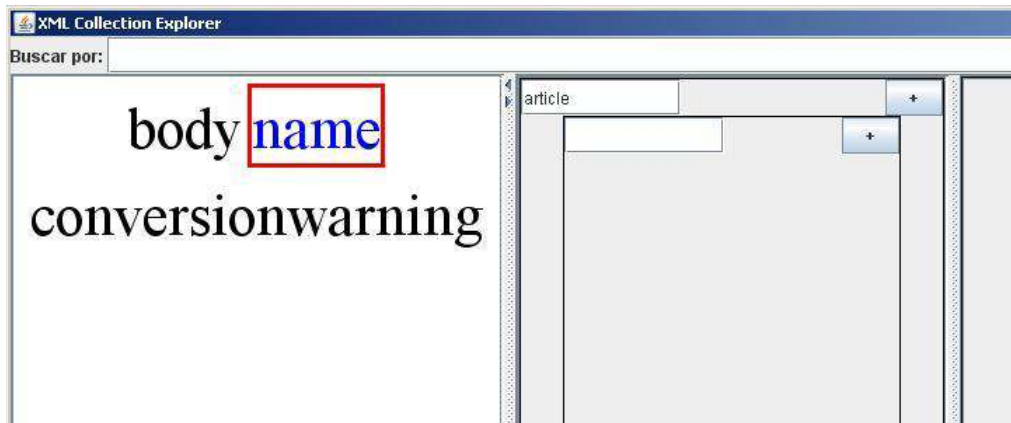


Figura 6.3: Ajuda e preenchimento da consulta.

A Figura 6.5 mostra o painel C da figura 6.1. No início existe apenas uma caixa de texto (C1). Para a criação da árvore de consulta deve ser utilizada o sinal "+" (C2), este sinal representa a vontade de criação de um filho para o termo que está na caixa de texto a ele associada. Visualmente, será gerada uma árvore similar a uma árvore de diretórios, como no *Windows Explorer*. Cada nível da árvore está um pouco afastado para a direita. Irmãos estão no mesmo nível. Quando houver a necessidade de se excluir uma caixa de texto devemos clicar no botão "-". Considerando os conceitos de ergonomia e padrões de interface, o ato de clicar nos botões de "+" e "-" pode ser substituído pelo uso do teclado, com as teclas de setas para direita + shift e esquerda + shift, respectivamente. A seta para baixo + shift gera um irmão para o termo da caixa de texto. O usuário poderá escolher mais de um filho para cada tag. À medida que ele vai clicando e criando as tags e termos ele estará vendo a geração da árvore. A consulta é processada clicando no botão C3 ou Enter no teclado.

A Figura 6.4 mostra a representação de uma consulta fictícia `A[B[B1[D[D1 D2[H J[J1]] E F G]] C]` por meio de uma árvore. Observe que na figura fica clara a hierarquia dos itens da consulta, o que fica difícil de visualizar através da consulta se esta fosse apresentada de maneira plana, como acima. A Figura 6.5 representa a consulta se esta fosse realizada na interface-protótipo.

Quando o usuário desejar auxílio para a escolha de tag/termos filhos, o sistema irá auxiliá-lo com sugestões. Para que estas sugestões apareçam, é necessário que o usuário selecione shift+enter ao fim da digitação do termo.

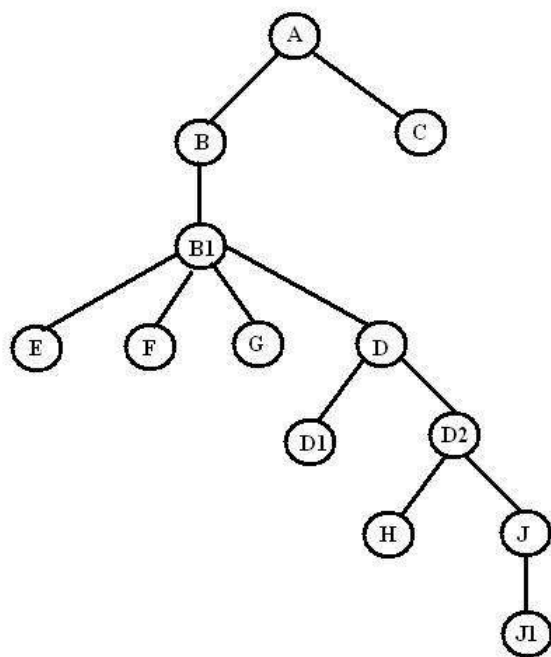


Figura 6.4: Árvore representando a consulta  $A[B[B1[D[D1\ D2[H\ J[J1]]\ E\ F\ G]]\ C]$ . 6.5.

A figura 6.6 mostra a ajuda inicial. Quando não temos nenhum termo digitado e pressionamos shift+enter, o sistema apresenta a tag raiz da coleção, ou as tags se for mais de uma coleção.

O usuário escolhe então a tag raiz para iniciar sua consulta, mas ele ainda não tem conhecimento de quais filhos esta raiz tem, então, ele novamente pressiona shift+enter ao fim do termo e aparecerá a lista de tags-filhas da raiz. A figura 6.3 mostra um exemplo de ajuda pra uma tag filha da raiz (article). Neste exemplo, podemos perceber que aparecem tags e termos para a consulta, o usuário pode optar por escolher o(s) termo(s) ou continuar a árvore de consulta.

Se o usuário optar pela escolha do termo, e pressionar shift+enter, o painel de ajuda (B) ficará vazio, pois um termo é representado na árvore como um nó folha, não tendo filhos. Assim que a consulta estiver pronta basta o usuário pressionar enter ou o botão buscar para executar a consulta.

A Figura 6.8 apresenta uma das possibilidades de preenchimento do painel D quando não são encontrados resultados.

A Figura 6.9 mostra como são apresentados os resultados de uma consulta no Painel D. A quantidade de documentos retornados é apresentado logo abaixo no painel.

Podemos perceber que o formato da apresentação dos documentos recuperados é feita para que o usuário possa verificar a localização dos termos consultados no arquivo de

resposta. A partir do nome do documento (com o caminho) é fácil abrir o documento e visualizá-lo.

Enquanto a consulta é executada, a interface informa ao usuário o andamento. Esta mensagem fica abaixo do Painel B. A Figura 6.10 mostra as mensagens que são exibidas.

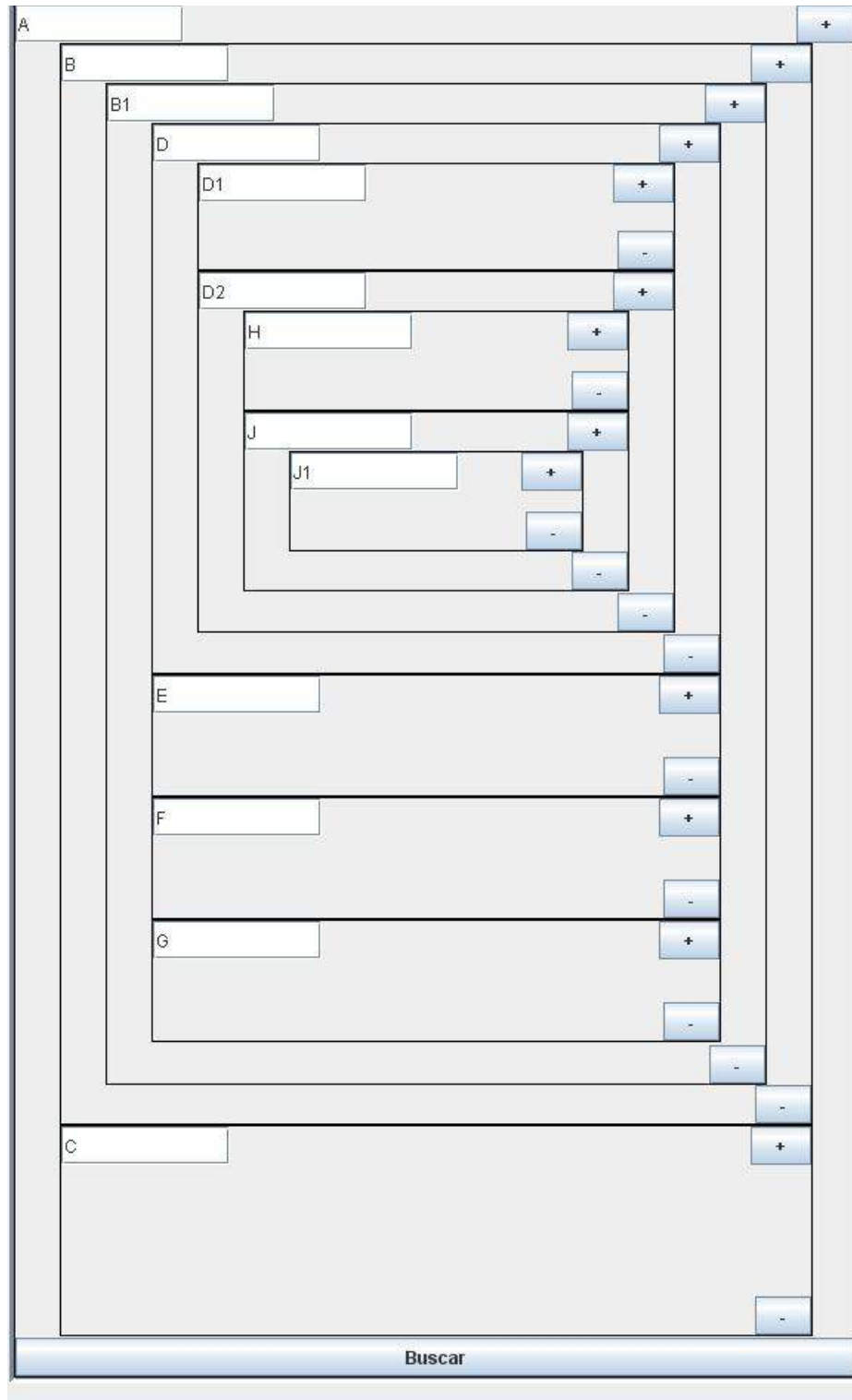


Figura 6.5: Representação da consulta estruturada: A[B[B1[D[D1 D2[H J[J1]] E F G]] C].

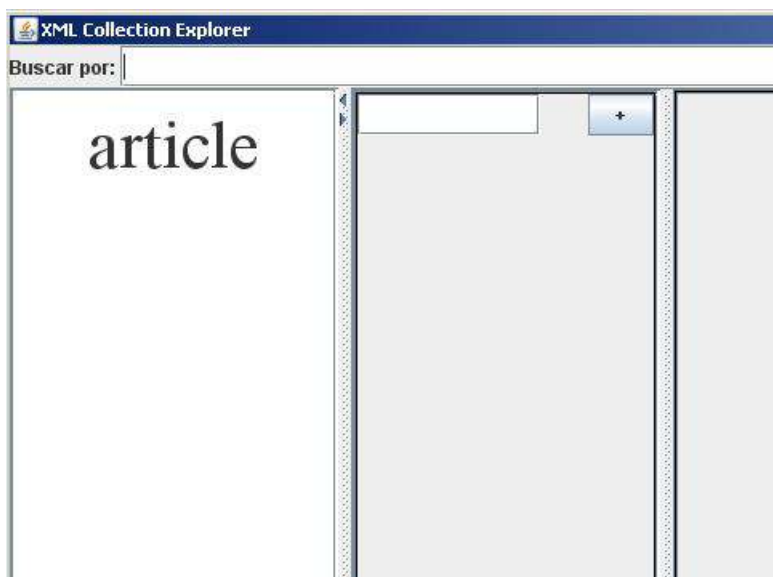


Figura 6.6: Ajuda inicial - Raiz da coleção.

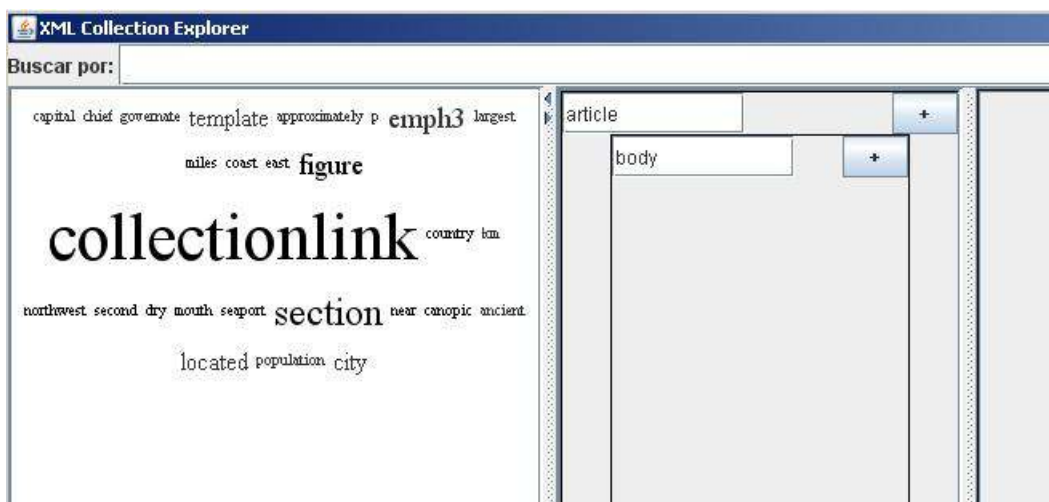


Figura 6.7: Ajuda na construção da árvore de consulta.



Figura 6.8: Mensagem quando não são encontrados resultados para a consulta.

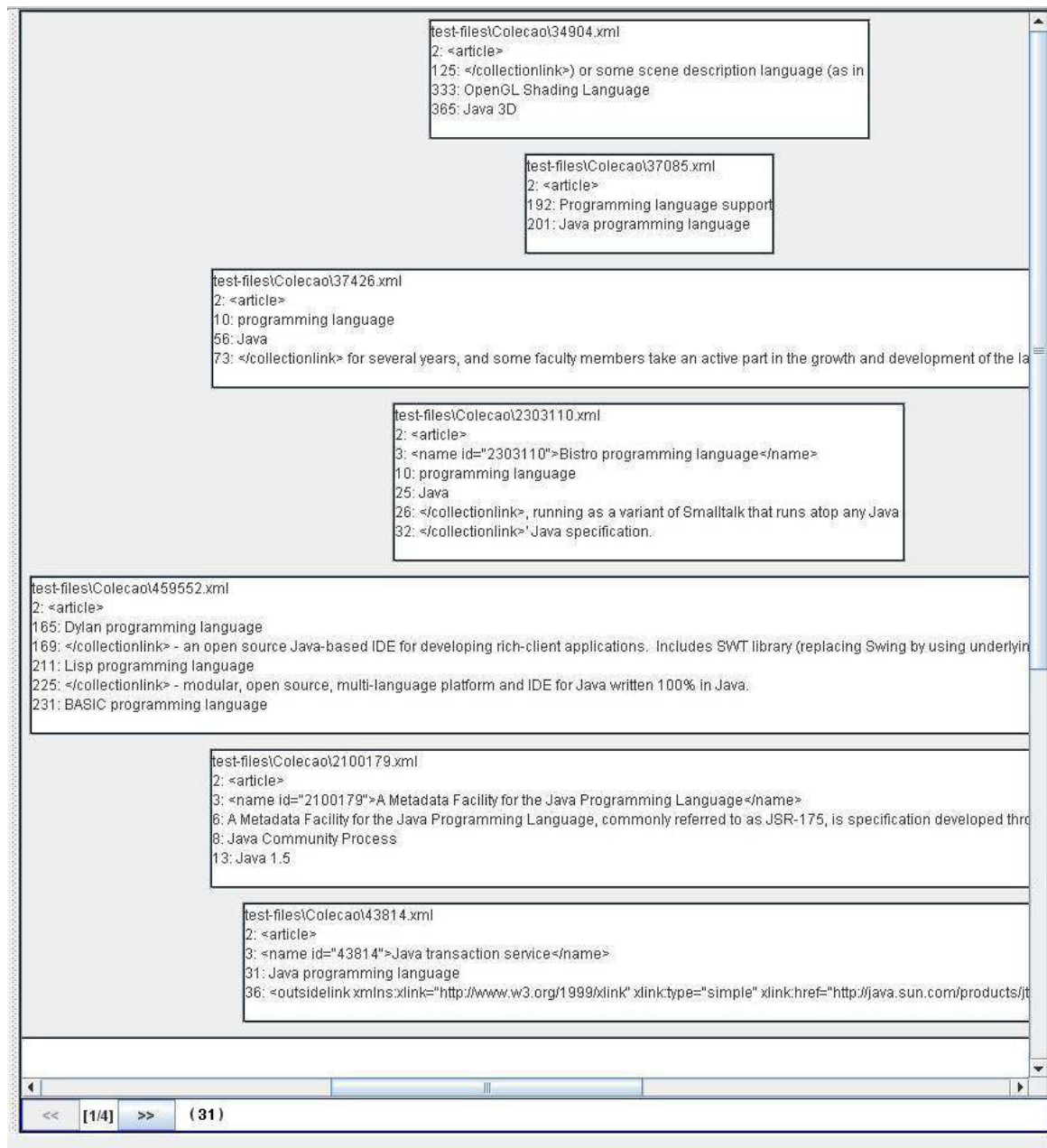


Figura 6.9: Painel D: Resultado de uma consulta, representando o Ranking de similaridade



Figura 6.10: Mensagens de processamento das consultas.

## 6.2 Exemplo de Aplicação da Interface

A seguir mostram-se todos os passos da realização de uma consulta com a utilização da interface:

O exemplo representa a construção da consulta estruturada para atender à necessidade de informação mostrada no tópico 449, descrito no Apêndice 1. Os termos a serem consultados são: "weather" e "prediction". Este exemplo foi escolhido por mostrar que, com o auxílio da interface podemos seguir a construção da árvore de consulta, da raiz até uma folha(termo) e voltar À raiz (ou outra tag dependendo da consulta) e realizar a construção de mais um caminho para um outro termo a ser consultado.

Todas as consultas apresentadas nas tabelas 6.1 e 6.2 foram realizadas com o auxílio da interface.

A Figura 6.12 mostra a tela inicial da interface. Se nesta tela o usuário já solicitar ajuda pressionando shift+enter a interface irá apresentar ao usuário a(s) tag(s) que representam a(s) raiz(es) das coleção, como mostrado na Figura 6.11.



Figura 6.11: Raiz da coleção.

O usuário insere a tag "article" na caixa de consulta e novamente pressiona shift+enter para obter ajuda a respeito dos filhos de "article", gerando a ajuda apresentada na Figura 6.13, mostrando que a tag "article" possui três filhos diretos: "body", "name" e "conversionwarning". Deve-se lembrar que, a qualquer momento da construção da árvore de consulta, o usuário pode inserir os termos da consulta, mesmo que não se tenha chegado ao fim da árvore ou que estes termos não apareçam na lista de filhos diretos da tag.

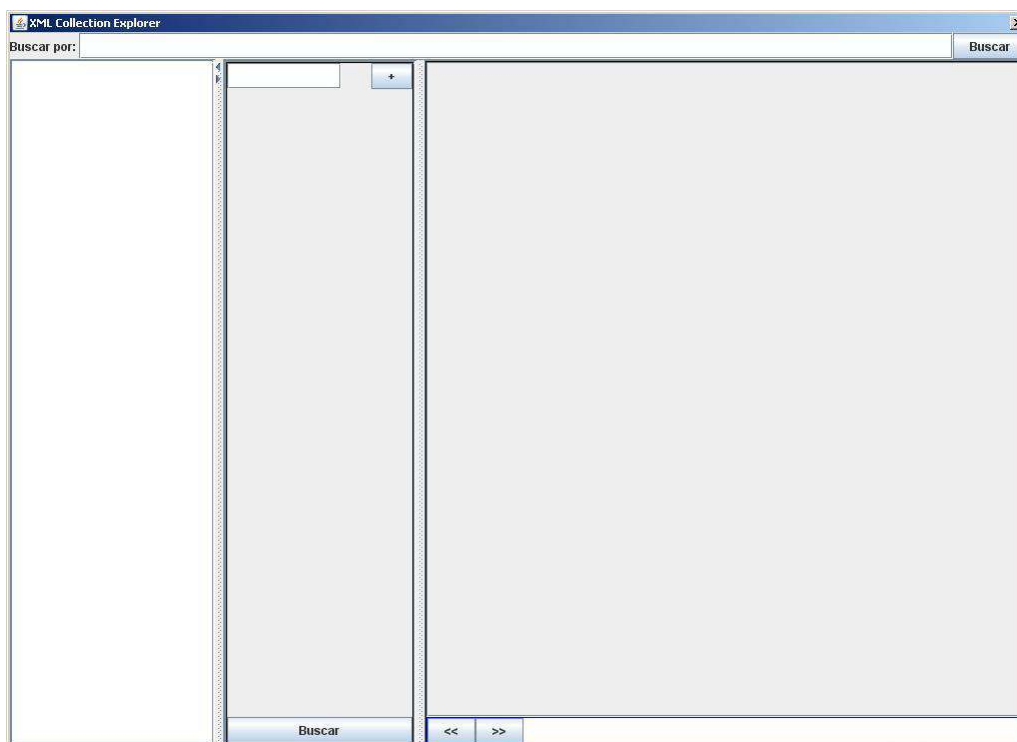


Figura 6.12: Tela inicial da interface.

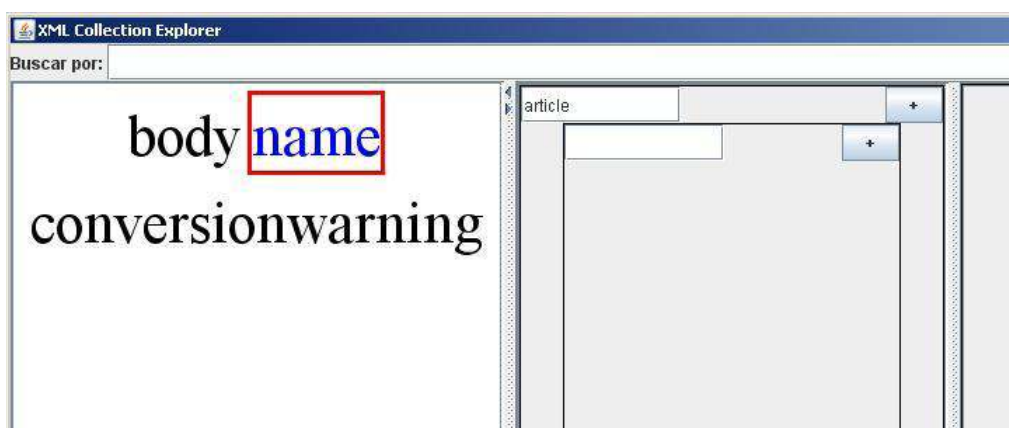


Figura 6.13: Filhos da tag article.

Para a inserção de um filho na tag "article", clica-se no botão "+" (ou pressiona-se shift+seta para direita). Uma nova caixa de texto aparecerá, onde deve-se inserir um filho para "article". A Figura 6.13 mostra que a tag-filho selecionada foi "name". Ao pressionar

shift+enter quando o cursor do mouse esta à frente de "name" teremos a ajuda para esta tag, como mostrado na Figura 6.14.

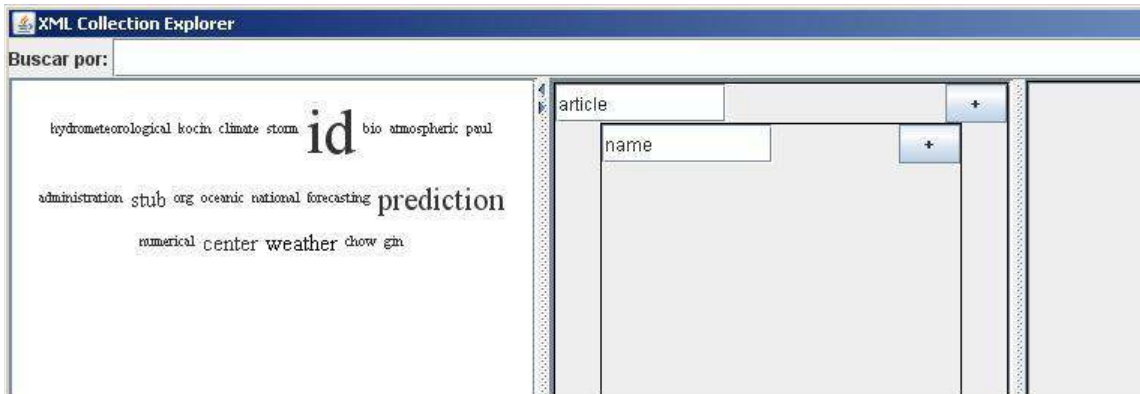


Figura 6.14: Ajuda para a tag "name".

De acordo com a ajuda mostrada, escolhe-se o termo "weather", que é um dos termos a serem consultados (Figura 6.15). Ao se pressionar shift+enter para a obtenção de ajuda para um termo, a tela ficará vazia, já que termos são nós-folha, não tendo portanto, filhos, como mostrado na Figura 6.16.

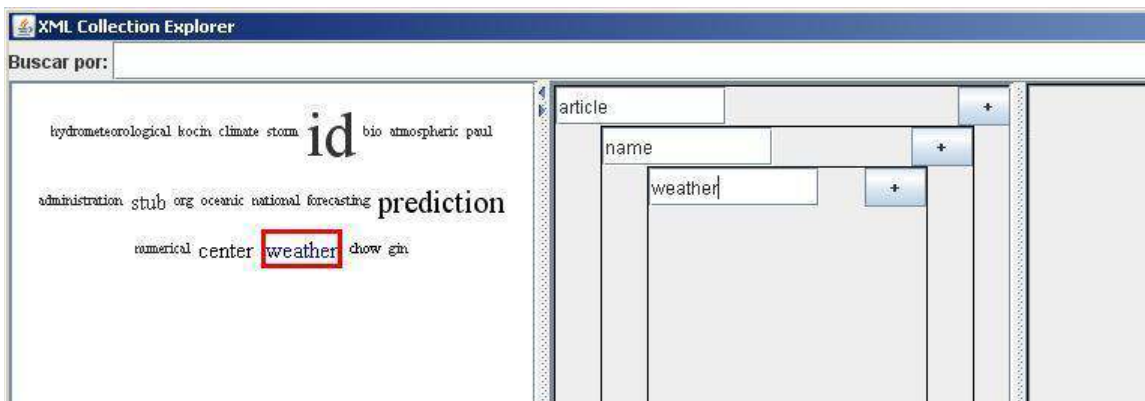


Figura 6.15: Escolha do termo "weather" para a consulta.

Deseja-se ainda consultar o termo "prediction". Para isso volta-se na construção da árvore, na tag article e seleciona-se a tag "body" (Figura 6.17), ou seja, o termo "prediction" deve estar em alguma parte do corpo do documento, não necessariamente no título, como foi com a tag "weather". A Figura 6.18 mostra a ajuda pra a tag "body".

Dentre os possíveis filhos diretos de "body" escolhe-se a tag "section", demonstrando que o termo "prediction" deve estar em alguma seção do documento. Solicita-se ajuda



Figura 6.16: Ajuda para um termo (nó-folha).

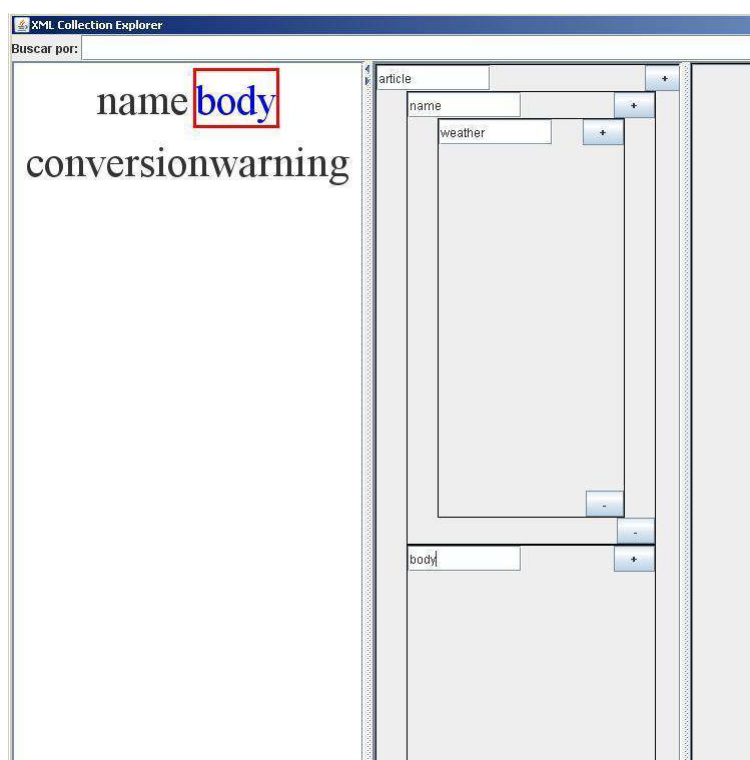


Figura 6.17: Seleção da tag "body", filha de "article".

para a tag "section", obtendo os dados apresentados na Figura 6.19.

Seleciona-se então a tag "title", significando que o termo "prediction" deve estar no título de uma seção. Solicita-se a ajuda para "title" e obtêm-se os dados apresentados na Figura 6.20.

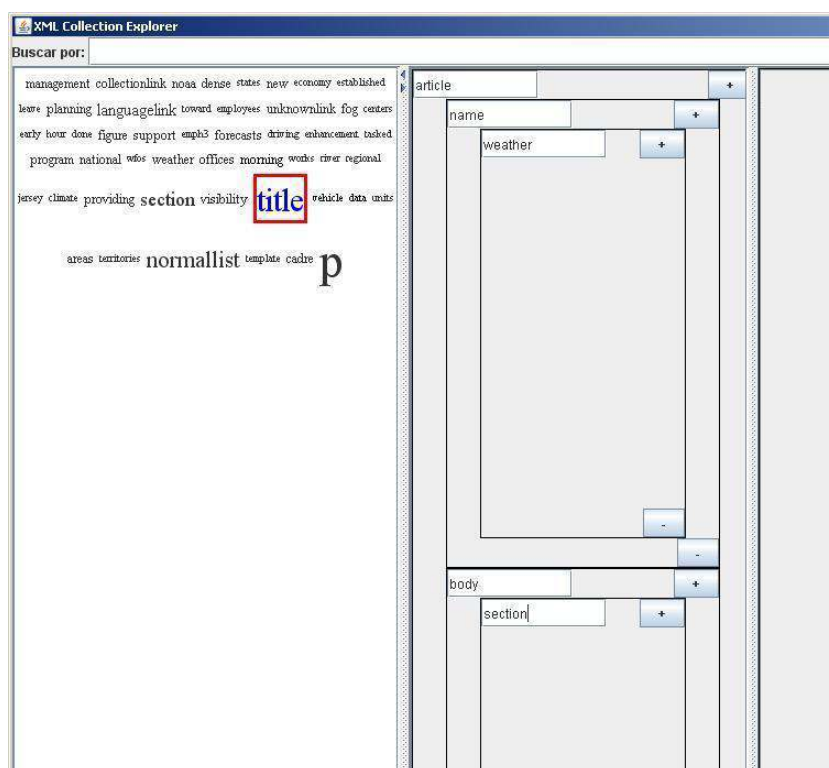
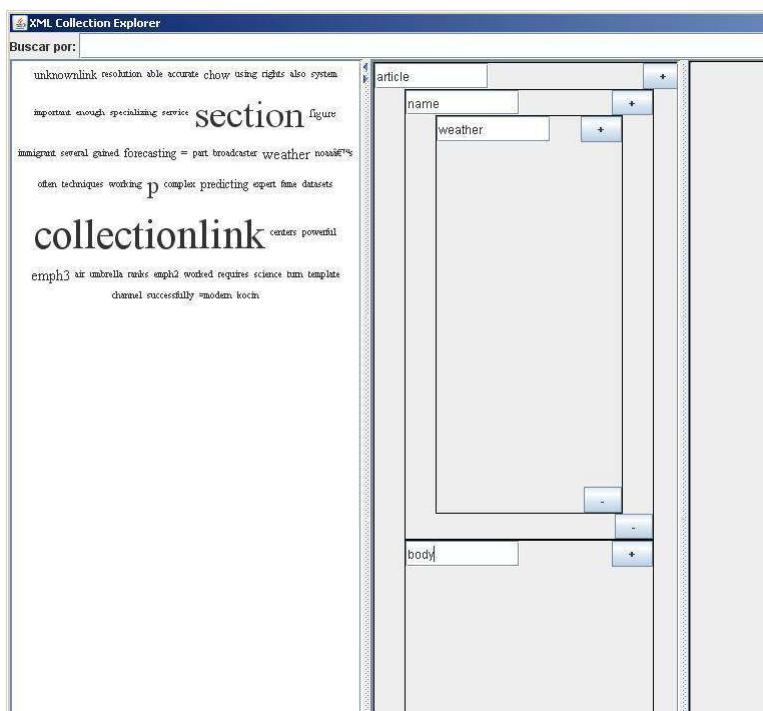


Figura 6.19: Ajuda para a tag "section".

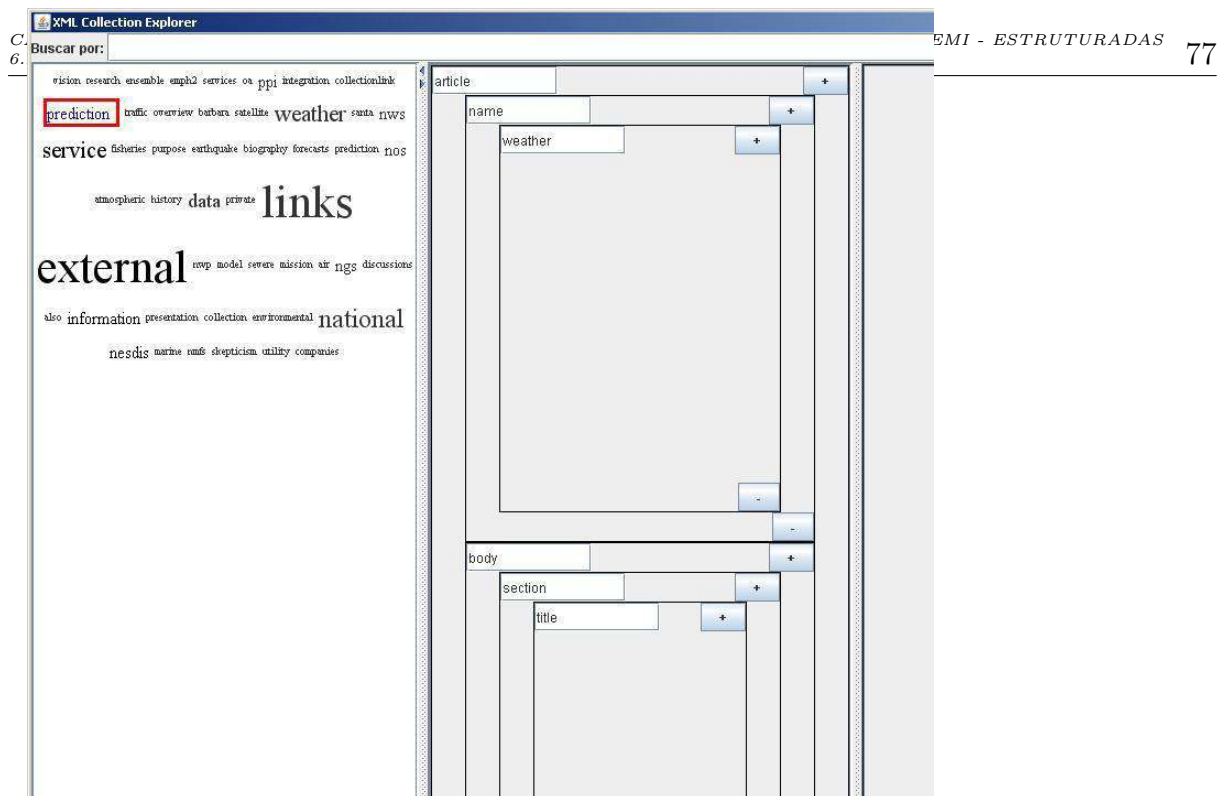


Figura 6.20: Ajuda para a tag "title".

Escolhe-se então o termo a ser consultado : "prediction". A consulta então está pronta, com os dois termos que representam a necessidade do usuário. Clica-se então no botão "Buscar" para que a consulta seja processada. A Figura 6.21 mostra a interface final da consulta: `article[name[weather] body[section[title[prediction]]]]` com os resultados obtidos.

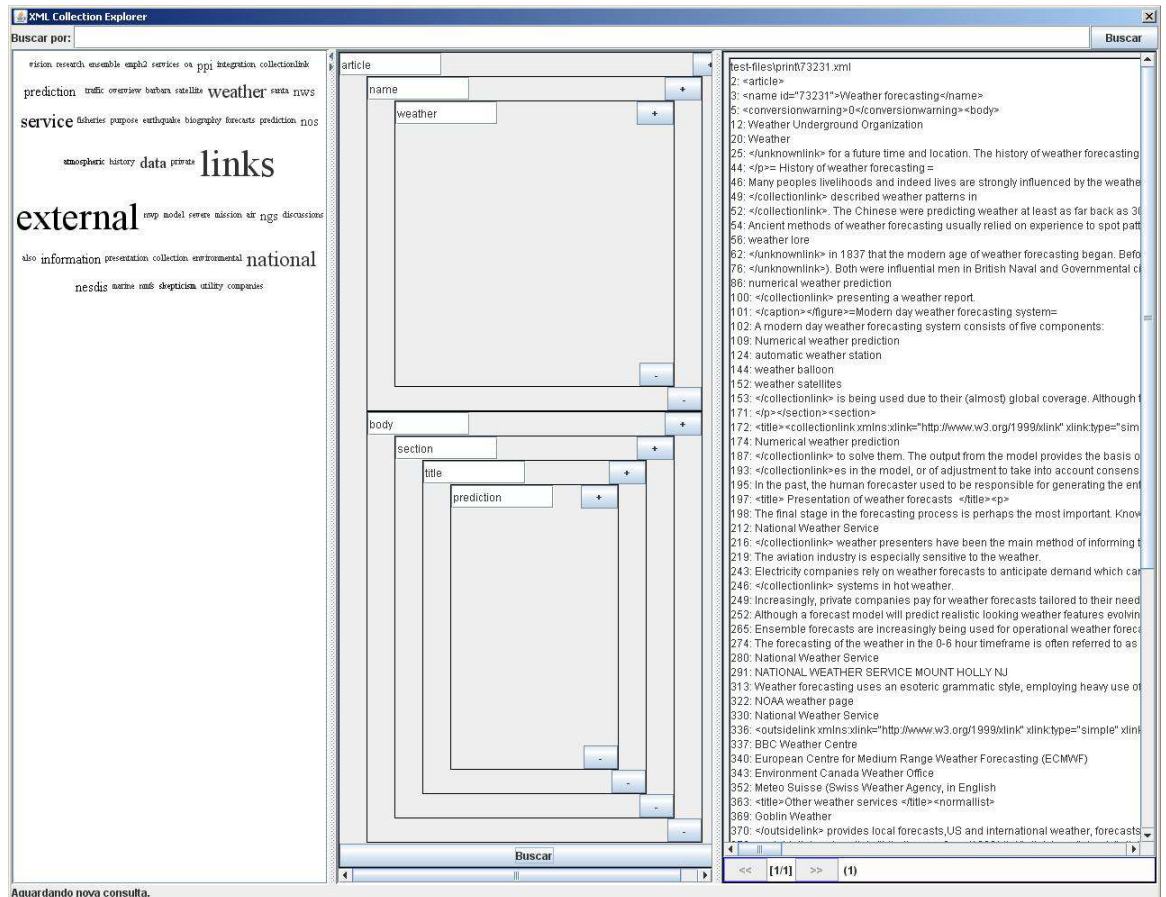


Figura 6.21: Interface final: Consulta estruturada: article[name[weather] body[ section[ title[ prediction]]]] e resultado.

## 6.3 Resultados da aplicação da interface

Para a verificação da melhora dos resultados da consulta, com a realização de consulta com estrutura auxiliada pela interface apresentada, foram realizadas consultas existentes na INEX. Foram selecionados 20 tópicos triviais.

A escolha de tópicos triviais foi feita para que a inserção de estrutura ocorresse apenas pela ajuda da interface. Os tópicos com consultas CAS não-triviais são úteis na verificação do sistema, se ele está ou não preparado para a melhora dos resultados com uma consulta estruturada, o que não é o foco deste trabalho.

O sistema escolhido para a realização dos testes foi uma implementação própria do modelo sTerm. Todas as consultas foram realizadas sob as mesmas condições.

A coleção utilizada para testes foi um conjunto de documentos da INEX composto por 467 documentos. O principal problema para grandes coleções é o grande aumento de termos para a indexação, que é um dos grandes problemas para a recuperação de informação semi-estruturada, mas que não é o foco deste trabalho. Esse número foi suficiente para mostrar a eficácia da interface proposta.

Como mostrado na Figura 6.3, são apresentados alguns termos tag (estrutura) para auxiliar o usuário que não tem conhecimento sobre estas estruturas no documento.

Os testes foram realizados com dois tipos de usuários: 1º - usuário com experiência e conhecimento da interface e da forma de construção da consulta, este usuário foi a autora deste trabalho; 2º - usuários com conhecimento em computação, mas que tiveram o primeiro contato com a interface no momento da realização dos testes e com uma breve explicação do funcionamento do protótipo.

### 6.3.1 Cálculo da Precisão-0 em Consultas Realizadas com a Interface - Usuário Experiente

Como algumas das *tags* apresentadas não são de conhecimento da maioria dos usuários, as consultas estruturadas foram formuladas sempre tentando utilizar tags mais comuns do vocabulário, como por exemplo: *body*, *title*, *name*, *section*. Esta escolha se deve ao fato de ser possivelmente claro para um usuário que *title* seja referente ao título do documento ou de algum elemento dentro do documento; *name* seja o nome do documento; *body* seja o corpo do documento e *section* seja uma seção em especial dentro deste documento.

A seguir são apresentados os dados dos testes realizados. Nas tabelas 6.1 e 6.2 têm-se as consultas realizadas (com e sem estrutura). As consultas estruturadas foram geradas por meio da interface proposta. Nas tabelas 6.4 e 6.5 temos as médias de Precisão-0 para consultas sem estrutura e com estrutura, respectivamente. Na tabela 6.3 temos a precisão-0 para cada consulta, usada para a avaliação da melhoria da precisão dos resultados destas consultas aplicados no sistema. E na tabela 6.6 a média da precisão-0 calculada sobre

todas as consultas.

Para efeito de padronização da coleção sempre colocaremos a palavra "article" no início de todas as consultas, como raiz para todos os documentos. Esta inserção é necessária devido à características do sistema de indexação do sTerm.

| ID CONSULTA | TÓPICO | CONSULTA NÃO-ESTRUTURADA                 | CONSULTA ESTRUTURADA                                                                     |
|-------------|--------|------------------------------------------|------------------------------------------------------------------------------------------|
| 1           |        |                                          |                                                                                          |
| 2           | 417    | therapeutic breathing                    | [body[section[title[therapeutic]]]breathing]                                             |
| 3           |        |                                          |                                                                                          |
| 4           | 449    | weather prediction                       | [name[weather]body[section[title [prediction]]]]                                         |
| 5           |        |                                          |                                                                                          |
| 6           | 452    | seven ancient wonders world              | [name[seven wonders]ancient world]<br>[name[wonders world]seven ancient]                 |
| 7           |        |                                          |                                                                                          |
| 8           |        |                                          |                                                                                          |
| 9           | 458    | cellphone technology                     | [section[cellphone] technology]                                                          |
| 10          |        | mobile phone technology                  |                                                                                          |
| 11          |        |                                          | [body[section[title[mobile phone]]]technology]                                           |
| 12          |        |                                          |                                                                                          |
| 13          | 460    | java mvc framework model-view-controller |                                                                                          |
| 14          |        | java mvc framework                       |                                                                                          |
| 15          |        | java model-view-controller               |                                                                                          |
| 16          |        | mvc                                      | [ name[model-view-controller]java]                                                       |
| 17          |        |                                          |                                                                                          |
| 18          | 462    | bea weblogic                             | [name [bea weblogic]]                                                                    |
| 19          |        |                                          | [body[section[bea weblogic]]]                                                            |
| 20          |        |                                          | [body[p[bea weblogic]server]]                                                            |
| 21          |        |                                          |                                                                                          |
| 22          | 465    | difference American British English      | [name[difference American British English]<br>[name[American British]difference English] |
| 23          |        |                                          |                                                                                          |
| 24          |        |                                          |                                                                                          |
| 25          | 467    | Definition ressources text mining        | [name[text mining]]                                                                      |
| 26          |        |                                          |                                                                                          |
| 27          | 473    | Kernel Machines Learning                 | [name[kernel] machine learning]                                                          |
| 28          |        |                                          | [body[section[title[machine] kernel learning]]]                                          |

Tabela 6.1: Consultas realizadas utilizando a interface-protótipo (1ª parte).

|    |     |                                                          |                                                                                                                            |
|----|-----|----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| 29 | 499 | Grammy Award                                             | [name[Grammy Awards]]                                                                                                      |
| 30 |     |                                                          |                                                                                                                            |
| 31 | 502 | Britney Spears                                           | [name[Britney Spears]]                                                                                                     |
| 32 |     |                                                          |                                                                                                                            |
| 33 | 503 | conjectures unsolved problems mathematics                | [name[conjectures] problems unsolved mathematics<br>[name[mathematics] conjectures unsolved problems ]                     |
| 34 |     |                                                          |                                                                                                                            |
| 35 |     |                                                          |                                                                                                                            |
| 36 | 505 | software patent europe                                   | [name[software patent] europe]<br>[name[ patent] software europe]<br>[name[patent] body [section[title[software]] europe]] |
| 37 |     |                                                          |                                                                                                                            |
| 38 |     |                                                          |                                                                                                                            |
| 39 |     |                                                          |                                                                                                                            |
| 40 | 506 | image processing                                         | [name[image processing]]<br>[name[image] processing]]                                                                      |
| 41 |     |                                                          |                                                                                                                            |
| 42 |     |                                                          |                                                                                                                            |
| 43 | 509 | music Bob Dylan                                          | [name[Bob Dylan] music]                                                                                                    |
| 44 |     |                                                          |                                                                                                                            |
| 45 | 514 | food additives physical health risk grocery store labels |                                                                                                                            |
| 46 |     | food additives                                           |                                                                                                                            |
| 47 |     |                                                          | [name[food]additives]                                                                                                      |
| 48 | 518 | tidal power wind                                         |                                                                                                                            |
| 49 |     |                                                          | [body[ section[title[title]] section[title[wind]] power]]                                                                  |
| 50 | 519 | types of bridges vehicles water ice                      |                                                                                                                            |
| 51 |     |                                                          | name[bridge] water vehicles type]                                                                                          |
| 52 | 523 | fuel efficient cars                                      |                                                                                                                            |
| 53 |     |                                                          | [name[car] fuel efficient]                                                                                                 |
| 54 | 524 | home heating solar panels                                |                                                                                                                            |
| 55 |     |                                                          | [nome [solar panel] home heating]<br>[name[solar] panels home heating]                                                     |
| 56 |     |                                                          |                                                                                                                            |

Tabela 6.2: Consultas realizadas utilizando a interface-protótipo (2ª parte).

| Tópico | ID Consulta com Estrutura | PRECISÃO-O da Consulta sem Estrutura | PRECISÃO-O da Consulta com Estrutura | Melhoria |
|--------|---------------------------|--------------------------------------|--------------------------------------|----------|
| 417    | 2                         | 25%                                  | 100%                                 | 300%     |
| 449    | 4                         | 11,1%                                | 100%                                 | 800%     |
| 452    | 6                         | 12,5%                                | 50%                                  | 300%     |
| 452    | 7                         | 12,5%                                | 100%                                 | 700%     |
| 458    | 9                         | 0% (consulta id=8)                   | 0                                    | 0%       |
| 458    | 11                        | 6,6%                                 | 100%                                 | 1415%    |
| 460    | 16                        | 25% (consulta id=14)                 | 100%                                 | 300%     |
| 462    | 18                        | 7,1%                                 | 100%                                 | 1308%    |
| 462    | 19                        | 7,1%                                 | 10%                                  | 40%      |
| 462    | 20                        | 7,1%                                 | 25%                                  | 252%     |
| 465    | 22                        | 14,2%                                | 50%                                  | 252%     |
| 465    | 23                        | 14,2%                                | 25%                                  | 76%      |
| 467    | 25                        | 0                                    | 100%                                 | 1000%    |
| 473    | 27                        | 20%                                  | 100%                                 | 400%     |
| 473    | 28                        | 20%                                  | 33,3%                                | 66,5%    |
| 499    | 30                        | 25%                                  | 50%                                  | 100%     |
| 502    | 32                        | 9,1%                                 | 33,3%                                | 266%     |
| 503    | 34                        | 25%                                  | 100%                                 | 300%     |
| 503    | 35                        | 25%                                  | 100%                                 | 300%     |
| 505    | 37                        | 14,2%                                | 50%                                  | 252%     |
| 505    | 38                        | 14,2%                                | 50%                                  | 252%     |
| 505    | 39                        | 14,2%                                | 100%                                 | 604%     |
| 506    | 41                        | 5,2%                                 | 100%                                 | 1823%    |
| 506    | 42                        | 5,2%                                 | 100%                                 | 1823%    |
| 509    | 44                        | 11,1%                                | 20%                                  | 80%      |
| 514    | 47                        | 25% (consulta id=46)                 | 100%                                 | 300%     |
| 518    | 49                        | 50%                                  | 100%                                 | 100%     |
| 519    | 51                        | 50%                                  | 100%                                 | 100%     |
| 523    | 53                        | 100%                                 | 100%                                 | 0        |
| 524    | 55                        | 33,3%                                | 100%                                 | 200%     |
| 524    | 56                        | 33,3%                                | 50%                                  | 50%      |

Tabela 6.3: Resultados obtidos.

Observações à respeito de algumas consultas:

Consultas 10 e 11 (sem estrutura): Diante da ocorrência de resultados ruins ou da não existência de resultados relevantes, baseado na descrição da necessidade de informação, as consultas foram alteradas. No lugar de "cellphone", foram realizadas consultas com: mobile phone, já que significam a mesma coisa. Esta modificação nos termos da consulta foi feita para mostrar que muitas vezes a consulta não tem resultados satisfatórios por problemas na formulação da consulta.

Consultas 13 a 16: Considerando que "mvc" é a abreviação de "model-view-controller"

foram realizadas consultas diferentes, com os dois termos ou com apenas um deles.

Consulta 20: A inserção do termo "server" foi realizada à partir da descrição na necessidade da informação.

|                                                                  |               |
|------------------------------------------------------------------|---------------|
| <b>Média de Precisão-0 alcançada nas consultas sem estrutura</b> | <b>21,38%</b> |
|------------------------------------------------------------------|---------------|

Tabela 6.4: Precisão-0 nas consultas sem estrutura.

|                                                                  |              |
|------------------------------------------------------------------|--------------|
| <b>Média de Precisão-0 alcançada nas consultas com estrutura</b> | <b>80,3%</b> |
|------------------------------------------------------------------|--------------|

Tabela 6.5: Precisão-0 nas consultas com estrutura.

|                                               |                |
|-----------------------------------------------|----------------|
| <b>MÉDIA GERAL DA MELHORIA DOS RESULTADOS</b> | <b>443,85%</b> |
|-----------------------------------------------|----------------|

Tabela 6.6: Média de melhoria.

Observação à respeito de algumas execuções:

Quanto mais próximo da posição número 1 do *ranking* o documento relevante estiver maior é a precisão do resultado da consulta. A tabela 6.5 mostra o valor médio alcançado de Precisão-0 para consultas com estrutura, lembrando que o objetivo quando avalia-se a Precisão dos resultados é que esta alcance 100%.

Consulta 2: como resultado tem-se apenas um documento, que possui *links* para diversos documentos que contêm informações importantes para a consulta. A não recuperação de tais arquivos ocorre por eles não fazerem parte da coleção utilizada.

Nesta recuperação temos um fato que deve ser levado em consideração na recuperação de documentos XML, que é a possibilidade de navegação dentro do documento.

Consulta 24: o segundo documento retornado foi considerado irrelevante por tratar de *data minig*, que foi citado como não sendo de interesse na narrativa do tópico.

### 6.3.2 Testes Realizados com Usuários sem Experiência

A Interface-protótipo foi apresentada para alguns usuários, para que estes pudessem utilizá-la e avaliá-la com relação às características físicas e com o resultado das consultas com estrutura. Foi dada uma explicação sucinta sobre o funcionamento da interface a estes usuários e eles realizaram as consultas usando a interface.

Foram realizadas consultas com 6 usuários com algum conhecimento na área de computação. Estes realizaram consultas com estrutura para 6 tópicos apresentados. Estes tópicos foram escolhidos porque quando a interface foi utilizada pelo usuário experiente obteve-se consultas bem estruturadas que alcançaram Precisão-0 de 100% (Tabelas 6.1 e 6.2), mostrando que podem ser formuladas consultas ideais.

Isolando das tabelas 6.1 e 6.2 os tópicos utilizados nos testes teremos a tabela 6.7.

Consultas 09 não considerada no cálculo das médias, já que não foi encontrado nenhum resultado para ela.

| Tópico | ID<br>Consulta<br>com<br>Estrutura | PRECISÃO-0<br>da Consulta<br>sem<br>Estrutura | PRECISÃO-0<br>da Consulta<br>com<br>Estrutura | Melhoria |
|--------|------------------------------------|-----------------------------------------------|-----------------------------------------------|----------|
| 452    | 6                                  | 12,5%                                         | 50%                                           | 300%     |
| 452    | 7                                  | 12,5%                                         | 100%                                          | 700%     |
| 458    | 9                                  | 0 (consulta id=8)                             | 0                                             | 0        |
| 458    | 11                                 | 6,6%                                          | 100%                                          | 1415%    |
| 473    | 27                                 | 20%                                           | 100%                                          | 400%     |
| 473    | 28                                 | 20%                                           | 33,3%                                         | 66,5%    |
| 503    | 34                                 | 25%                                           | 100%                                          | 300%     |
| 503    | 35                                 | 25%                                           | 100%                                          | 300%     |
| 505    | 37                                 | 14,2%                                         | 50%                                           | 252%     |
| 505    | 38                                 | 14,2%                                         | 50%                                           | 252%     |
| 505    | 39                                 | 14,2%                                         | 100%                                          | 604%     |
| 518    | 49                                 | 50%                                           | 100%                                          | 100%     |
| Média  |                                    | 19,47%                                        | 80,3%                                         | 426,31%  |

Tabela 6.7: Resultados obtidos dos tópicos utilizados para os testes com usuários.

O valor médio de Precisão-0 alcançada nos testes apresentados foi de 19,47% para consultas sem estrutura e de 80,3% para consultas com estrutura, gerando um aumento médio de 426,31% na melhoria destes resultados, média próxima a encontrada com todos os tópicos avaliados.

### Precisão-0 das Consultas Realizadas com Usuários sem Experiência

As consultas realizadas pelos usuários alcançaram um aumento médio na Precisão-0 de 179,75%. Este resultado mostra que, com pouco conhecimento sobre a interface já é possível a realização de consultas com estrutura suficiente para a melhoria na precisão dos resultados.

Os usuários alcançaram em média, 46,72% de Precisão-0 nas consultas realizadas com estrutura, valor próximo a 50% que representa que o documento mais relevante está, em média, na segunda posição do ranking. Este valor já é satisfatório, já que usuários em geral, verificam os resultados nas primeiras posições do ranking, e não somente na primeira.

As tabelas 6.8 a 6.13 sintetizam os valores de Precisão-0 alcançados pelos seis usuários, cada tabela representa um usuário.

Percebe-se que os usuários iniciantes na utilização desta interface realizam consultas com alguma estrutura. A ocorrência de consultas bem estruturadas são bons indicadores de que a interface ajuda os usuários na formulação de boas consultas estruturadas. Tais consultas bem estruturadas dependem de dois outros fatores, como a melhoria no conhecimento da interface e melhoria no processo de realização e consultas com estrutura, já

| <b>Tópico</b> | <b>PRECISÃO-O da Consulta sem Estrutura</b> | <b>PRECISÃO-O da Consulta com Estrutura</b> | <b>Melhoria</b> |
|---------------|---------------------------------------------|---------------------------------------------|-----------------|
| 452           | 12,5%                                       | 25%                                         | 100%            |
| 458           | 6,6%                                        | 0                                           | 0               |
| 473           | 20%                                         | 100%                                        | 400%            |
| 503           | 25%                                         | 100%                                        | 300%            |
| 505           | 14,2%                                       | 100%                                        | 604%            |
| 518           | 50%                                         | 100%                                        | 100%            |
| Média         | 21,38%                                      | 70,83%                                      | 250,66%         |

Tabela 6.8: Resultados obtidos usuário 1.

| <b>Tópico</b> | <b>PRECISÃO-O da Consulta sem Estrutura</b> | <b>PRECISÃO-O da Consulta com Estrutura</b> | <b>Melhoria</b> |
|---------------|---------------------------------------------|---------------------------------------------|-----------------|
| 452           | 12,5%                                       | 0                                           | 0               |
| 458           | 6,6%                                        | 25%                                         | 279%            |
| 473           | 20%                                         | 33,3%                                       | 66,5%           |
| 503           | 25%                                         | 0                                           | 0               |
| 505           | 14,2%                                       | 100%                                        | 604%            |
| 518           | 50%                                         | 100%                                        | 100%            |
| Média         | 21,38%                                      | 43,05%                                      | 174,91%         |

Tabela 6.9: Resultados obtidos usuário 2.

que muitos usuários, independente da interface não são familiarizados com a realização de consultas com estrutura.

Uma característica interessante observada nestas consultas é que os usuários tendem a direcionar as palavras-chaves a serem consultadas no nome do documento.

### Percepção das Características da Interface pelos Usuários

As primeiras observações sobre a interface trouxeram comentários que foram aceitos na modificação do protótipo que foi apresentado: Disposição dos resultados, antes centralizados, agora alinhados à esquerda; Inclusão de um botão de "?" ao lado do botão "+" para que a solicitação de ajuda na construção da árvore fosse feita também pelo uso do mouse, e não somente pelo teclado; Os termos apresentados como ajuda agora são apresentados em ordem alfabética.

Os usuários acharam que a interface foi de fácil utilização, aprovando o objetivo da interface, salientando que esta alcança os objetivos propostos, sendo uma idéia muito promissora.

As cores, botões, teclas de atalho e fonte foram aprovados pelos usuários, sendo de

| <b>Tópico</b> | <b>PRECISÃO-O<br/>da Consulta<br/>sem<br/>Estrutura</b> | <b>PRECISÃO-O<br/>da Consulta<br/>com<br/>Estrutura</b> | <b>Melhoria</b> |
|---------------|---------------------------------------------------------|---------------------------------------------------------|-----------------|
| 452           | 12,5%                                                   | 50%                                                     | 200%            |
| 458           | 6,6%                                                    | 50%                                                     | 657,57%         |
| 473           | 20%                                                     | 25%                                                     | 25%             |
| 503           | 25%                                                     | 100%                                                    | 300%            |
| 505           | 14,2%                                                   | 33,3%                                                   | 134,5%          |
| 518           | 50%                                                     | 100%                                                    | 100%            |
| Média         | 21,38%                                                  | 59,71%                                                  | 236,17%         |

Tabela 6.10: Resultados obtidos usuário 3.

| <b>Tópico</b> | <b>PRECISÃO-O<br/>da Consulta<br/>sem<br/>Estrutura</b> | <b>PRECISÃO-O<br/>da Consulta<br/>com<br/>Estrutura</b> | <b>Melhoria</b> |
|---------------|---------------------------------------------------------|---------------------------------------------------------|-----------------|
| 452           | 12,5%                                                   | 33%                                                     | 164%            |
| 458           | 6,6%                                                    | 20                                                      | 203%            |
| 473           | 20%                                                     | 25%                                                     | 25%             |
| 503           | 25%                                                     | 25%                                                     | 0               |
| 505           | 14,2%                                                   | 0                                                       | 0               |
| 518           | 50%                                                     | 100%                                                    | 100%            |
| Média         | 21,38%                                                  | 33,83%                                                  | 82%             |

Tabela 6.11: Resultados obtidos usuário 4.

boa percepção e fácil utilização.

Algumas sugestões apresentadas foram: inserir mensagens explicativas no momento do preenchimento dos termos da consulta para facilitar o uso por usuários menos experientes; redimensionar melhor o painel de formação de consulta, pois consultas maiores ficam difíceis de serem visualizadas; apresentar árvore de consulta de uma maneira mais familiar, como por exemplo a utilizada nos sistemas de arquivos do Windows; diferenciar a apresentação de termos-tag e os termos-folha, quando da apresentação destes na painel de ajuda;

Algumas dificuldades visualizadas: alguns termos-tag não são de conhecimento do usuário, gerando dúvida em sua utilização.

## 6.4 Considerações Finais

Diante destas consultas realizadas com o auxílio da interface sobre os tópicos da INEX podemos ver que com poucas modificações na consulta, com a inserção de estruturas que seriam de fácil entendimento para vários usuários, gerando consultas bem estruturada temos um grande aumento na precisão dos resultados, com uma melhoria média de

| <b>Tópico</b> | <b>PRECISÃO-0<br/>da Consulta<br/>sem<br/>Estrutura</b> | <b>PRECISÃO-0<br/>da Consulta<br/>com<br/>Estrutura</b> | <b>Melhoria</b> |
|---------------|---------------------------------------------------------|---------------------------------------------------------|-----------------|
| 452           | 12,5%                                                   | 50%                                                     | 200%            |
| 458           | 6,6%                                                    | 50%                                                     | 657,57%         |
| 473           | 20%                                                     | 50%                                                     | 150%            |
| 503           | 25%                                                     | 0                                                       | 0               |
| 505           | 14,2%                                                   | 100%                                                    | 604%            |
| 518           | 50%                                                     | 100%                                                    | 100%            |
| Média         | 21,38%                                                  | 58,33%                                                  | 285,26%         |

Tabela 6.12: Resultados obtidos usuário 5.

| <b>Tópico</b> | <b>PRECISÃO-0<br/>da Consulta<br/>sem<br/>Estrutura</b> | <b>PRECISÃO-0<br/>da Consulta<br/>com<br/>Estrutura</b> | <b>Melhoria</b> |
|---------------|---------------------------------------------------------|---------------------------------------------------------|-----------------|
| 452           | 12,5%                                                   | 0                                                       | 0               |
| 458           | 6,6%                                                    | 11,1%                                                   | 66,5%           |
| 473           | 20%                                                     | 33,3%                                                   | 400%            |
| 503           | 25%                                                     | 50%                                                     | 200%            |
| 505           | 14,2%                                                   | 50%                                                     | 252%            |
| 518           | 50%                                                     | 100%                                                    | 100%            |
| Média         | 21,38%                                                  | 40,73%                                                  | 169,75%         |

Tabela 6.13: Resultados obtidos usuário 6.

443,85%.

Quando ocorre a realização de consultas por usuários iniciantes no uso da interface estes conseguem, sem grandes dificuldades realizarem consultas com alguma estrutura, trazendo uma melhoria média de 156,25% da Precisão-0 dos resultados.

Os testes foram realizados sempre na tentativa de se evitar qualquer influência externa. Todos os termos pesquisados, em todas as consultas, estruturadas ou não tinham a mesma relevância para a pesquisa.

# Capítulo 7

## Trabalhos Relacionados

Neste capítulo são apresentados alguns exemplos de interfaces projetadas para recuperação semi-estruturada. Estas propostas são comparadas com a interface porposta neste trabalho.

### 7.1 Interface de Usuário para Recuperação de Documentos XML

O trabalho [Großjohann et al., 2002] combina conceitos relacionados a RI com a busca em documentos semi-estruturados. Os autores propõe dois sistemas: 1. uma interface visual para formulação de consultas XIRQL; 2. uma interface para apresentação dos resultados.

A interface de consulta é baseada na linguagem XIRQL e permite que o usuário construa consultas nessa linguagem de forma mais amigável e sem a necessidade de conhecer sua sintaxe. A vantagem no uso de XIRQL como base é que ela foi criada para consulta de coleções XML e para suplantar as limitações de outras linguagens de consulta em relação ao contexto de RI. A linguagem estende XPath com operadores necessários em RI, tais como "sounds\_like" e "contains" que permitem a busca em diferentes tipos de dados usando predicados vagos. O objetivo da interface visual é tornar a consulta em documentos estruturados mais acessível e amenizar a necessidade de se conhecer a sintaxe de linguagem de consulta, neste caso, XIRQL. A figura 7.1 ilustra a interface de consulta que é dividida em três áreas. A área A, *structure condition area*, permite ao usuário formular condições estruturais "isoladas" para as consultas. A área B, *condition list*, permite editar essas condições e definir como combiná-las para formar a consulta. A área C, *paraphrase area*, permite visualizar a versão em XIRQL para a consulta realizada pelo usuário. As condições estruturais podem ser formuladas com base em exemplos (*Query by Example*) ou com base na DTD. Nas formulações por exemplos, o usuário seleciona elementos de interesse em um documento da coleção e, à partir daí, generalizações são derivadas do

documento e apresentadas ao usuário para definir sua consulta. Na opção por DTD, a consulta não é definida com base em exemplos de um documento específico, mas sim na estrutura implícita das DTDs.

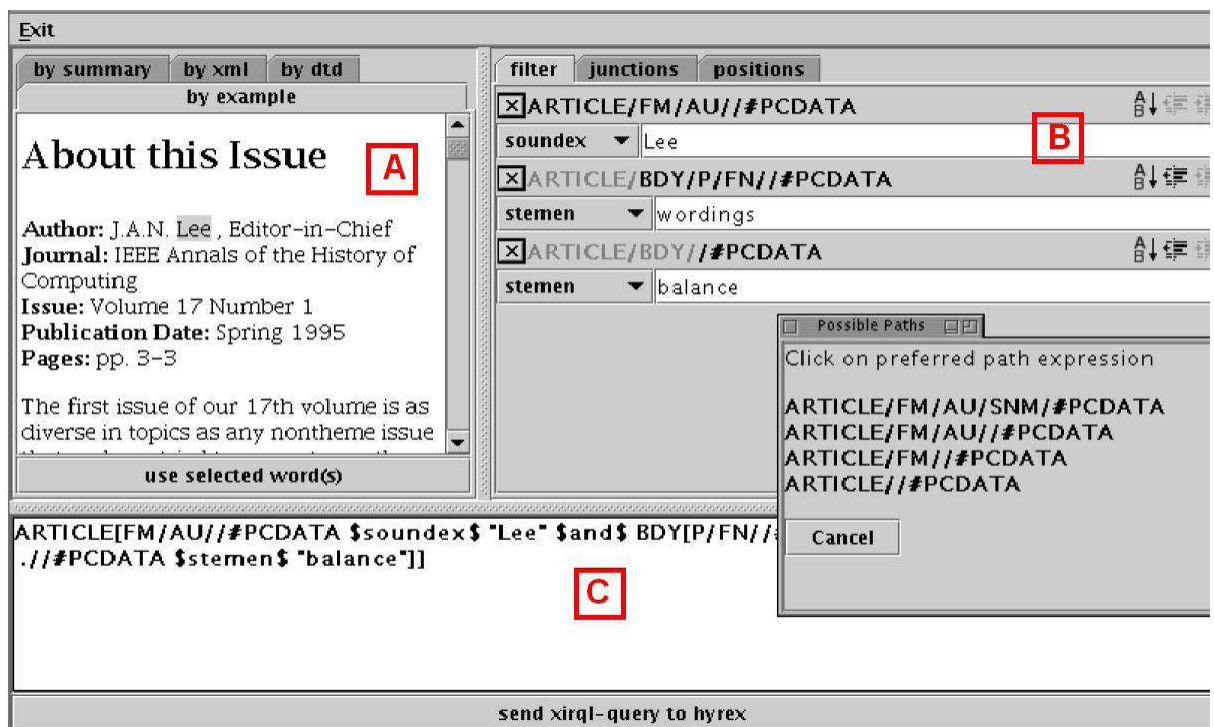


Figura 7.1: Interface de Consulta, baseada na geração de consultas XIRQL [Großjohann et al., 2002]

A interface de apresentação visa adaptar a realidade RI a coleções semi-estruturadas. Nesse âmbito, conceitos de relevância (Peso e Ranking) devem ser projetados sobre as noções de estruturas hierárquicas e interrelacionadas presentes em documentos XML. Com isso, os resultados relevantes podem ser provenientes de trechos de vários documentos. Os resultados são expressos com o objetivo de sintetizar e facilitar a visualização simultânea de vários documentos recuperados. Por sua vez, o grau de relevância de um documento (ou parte dele) é representada utilizando escalas de cores. A figura 7.2 mostra os principais elementos da interface. A área A mostra várias janelas representando os documentos relevantes para consulta na forma de TreeMaps Parciais. A coloração baseada em escalas de cores é usada para destacar as partes mais relevantes nesses documentos. Adicionalmente, a interface faz uso de ToolTips (área B) para detalhar informações sobre um documento e tabelas ou índices de conteúdo (Table of contents) para dar informações de alto nível (área C). O documento original também é apresentado no resultado (área D).

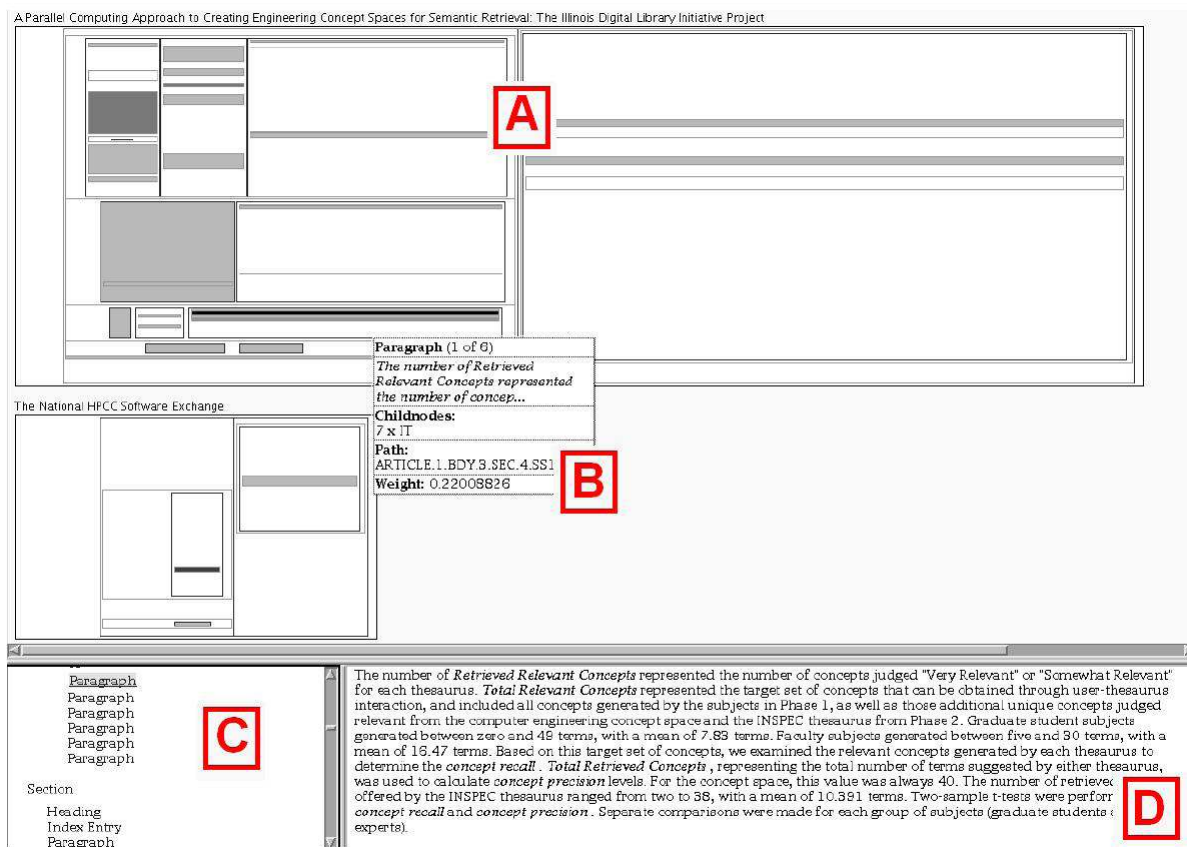


Figura 7.2: Interface de Apresentação de Resultados [Großjohann et al., 2002].

Pode-se perceber que o objetivo deste trabalho e a interface-protótipo proposta é tornar a consulta mais acessível e amenizar a necessidade de se conhecer a sintaxe da linguagem de consulta, entretanto esta interface foi projetada para a consulta em uma linguagem específica, a XIRQL, usando a estrutura contida na DTDs. A interface-protótipo apresentada nesta dissertação tem como finalidade a consulta a qualquer coleção de documentos, com e sem DTD, uma coleção de documentos heterogênea. A ajuda é retirada dos documentos e não das DTDs.

## 7.2 Interface de Consulta para Dados Semi-estruturados

A interface para consulta em repositórios XML proposta em [Newman and Ozsoyoglu, 2004] é baseada em uma representação em árvore e em mecanismos de filtragem de atributos e elementos na forma de caixas de seleção e listagem (*drop-down box*). O trabalho visa fornecer uma interface genérica, simples, não ambígua e de fácil uso para usuários inexperientes. O trabalho segue o paradigma "Query By Example".

A figura 7.3(a) mostra a forma como um documento é mapeado para a interface de consulta. Cada elemento presente no documento é representado como um nó selecionável. A cada atributo é associado uma caixa de listagem, também selecionável, contendo os

possíveis valores para o atributo presentes nos documentos. A medida que o usuário marca os elementos e atributos, a árvore de consulta é filtrada dinamicamente de forma a refletir apenas documentos que atendem as condições selecionadas pelos usuários. Após finalizar a consulta, os documentos recuperados podem ser visualizados de duas formas: 1. como uma hierarquia que se expande e contrai, similar a uma arquivo XML; 2. uma tabela plana, onde cada coluna representa um atributo. A árvore de consulta não está limitada a estrutura dos documentos na coleção, podendo o usuário "customizar" a consulta, da maneira que preferir. Desde que haja relacionamentos entre os nós inseridos pelo usuário, a árvore de consulta será apresentada.

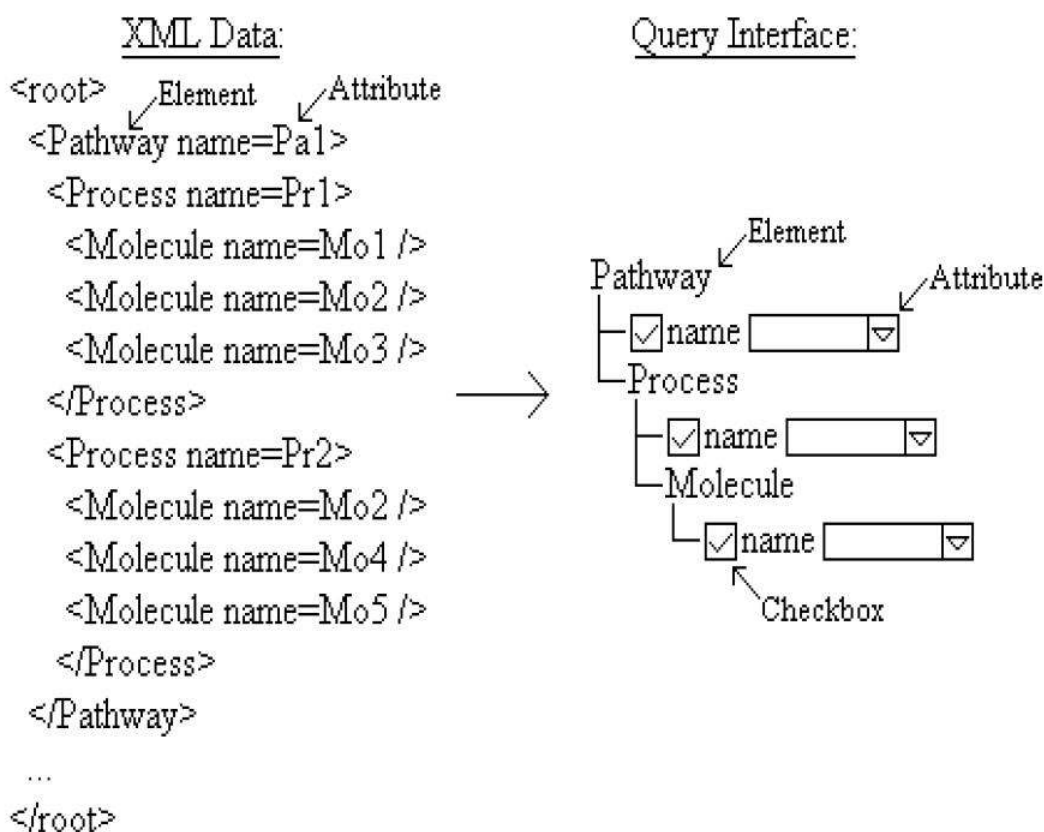


Figura 7.3: Interface de Consulta Pathway Explorer - Parte A [Newman and Ozsoyoglu, 2004].

A Figura 7.4 (b) mostra esse esquema de consulta implementado em uma aplicação no domínio da biologia, chamada Pathway Explorer. À esquerda é mostrado o componente para a formulação das consultas. À direita, na parte superior, são listados os resultados da consulta. Na parte inferior, os resultados são apresentados na forma de grafos. O sistema foi criado para facilitar a extração de informações em um banco de dados relacional neste domínio. Após a formulação da consulta pelo usuário na interface gráfica, consultas SQL são geradas. O desempenho é comparável ao que seria obtido em um repositório XML acessado através de Xquery.

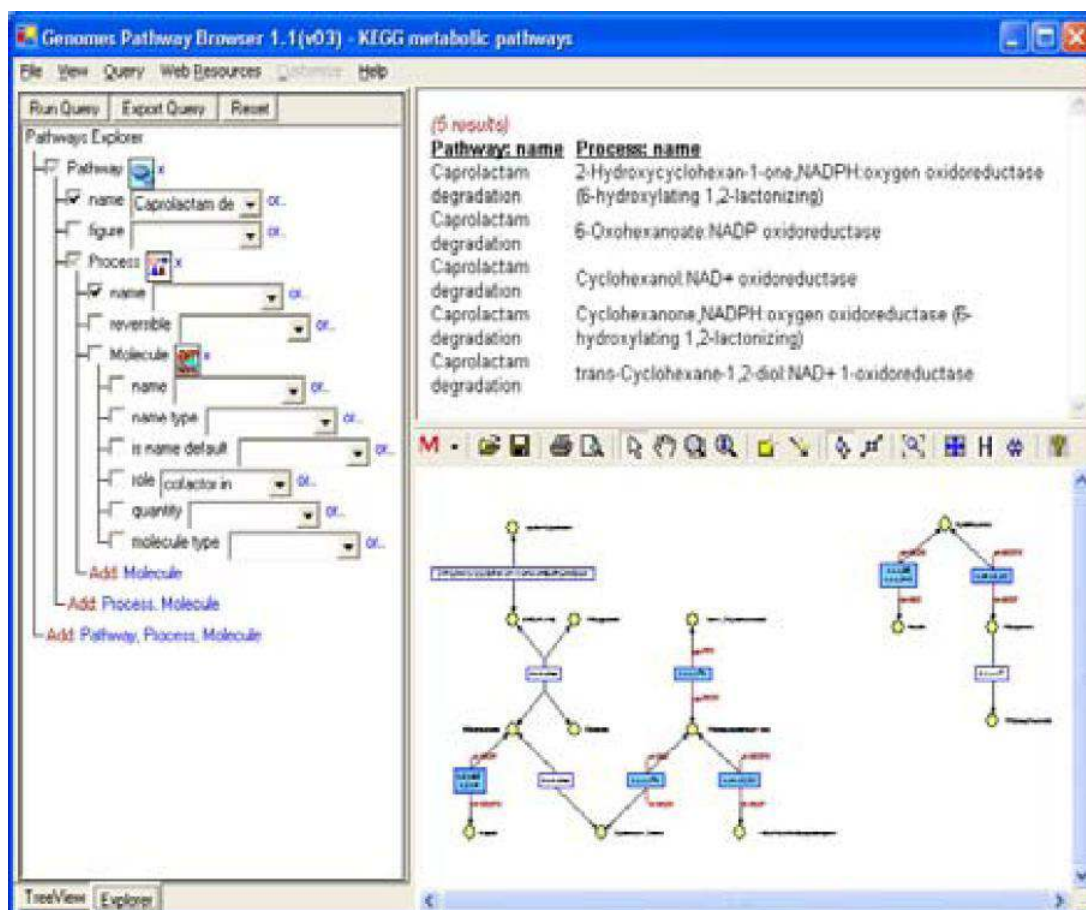


Figura 7.4: Interface de Consulta Pathway Explorer - Parte B [Newman and Ozsoyoglu, 2004].

Da mesma forma que o trabalho de [Großjohann et al., 2002], esta dissertação é voltada para uma interface mais específica, para facilitar a extração de informações em um banco de dados. Isto permite a apresentação da estrutura da árvore, já que documentos XML representando um banco de dados possuem a mesma estrutura, o que não ocorre com documentos de coleções variadas, que não permitem a apresentação direta das tags que representam tal estrutura. A interface-protótipo foi formulada para atender documentos de coleções diversas.

## 7.3 XQBE: Interface para Consultas Semi-estruturadas usando XQuery

Em [Augurusa et al., 2003, Braga et al., 2005, Braga and Campi, 2003], os autores criaram a XQBE, uma linguagem gráfica de fácil manipulação para a realização de consultas em documentos XML e que tira proveito da linguagem XQuery (linguagem oficial, segundo a W3C - World Wide Web Consortium, para manipulação de documentos XML). XQBE foi criada com dois objetivos principais: 1. ser amigável e acessível para usuários

sem conhecimentos avançados de programação; 2. ser traduzida diretamente em XQuery para reutilizar os recursos já desenvolvidos para essa linguagem. Dentre suas aplicações estão a consulta, transformação, modificação e validação de documentos XML. Algumas de suas características são o poder expressivo cobrindo a maior parte da sintaxe de XPath, a possibilidade de realizar consultas do tipo FLWOR (for, let, where, order-by, return), construção de novos elementos XML e a reestruturação de documentos existentes. Com isso, estima-se que a maioria das necessidades de consulta dos usuários em geral seja suprida pela linguagem. XQBE não inclui alguns recursos como if-then-else, o que aliado a outros fatores, a torna não tão expressiva em relação a XQuery (de fato, isto não fazia parte dos objetivos, já que a complexidade adicionada poderia inviabilizar a linguagem para os usuários em geral). O paradigma Query by Example serve de base também para XQBE.

A interface criada para a linguagem é mostrada na figura 7.5. A figura 7.5 (a) resume a idéia mostrando alguns elementos gráficos de XQBE. Na área da fonte de dados (painel à esquerda), o usuário manipula os elementos presentes no documento a ser consultado. Na área de construção (painel à direita), é possível definir o formato de saída para resultados e define condições para a consulta. Cada elemento gráfico tem uma semântica bem definida. Os elementos entre cada área devem ser interligados pelo usuário para definir a sua correspondência. Ao final do processo de construção da consulta e formatação do resultado desejado, é gerada uma consulta correspondente em XQuery. Esta é passada a um motor para a extração das informações necessárias. A figura 7.5 mostra a aplicação criada com base nesses conceitos. Há uma barra de ferramentas que permite inclusão de elementos gráficos. O usuário pode ainda alternar entre o modo gráfico e a consulta correspondente em XQuery (abas na parte inferior).

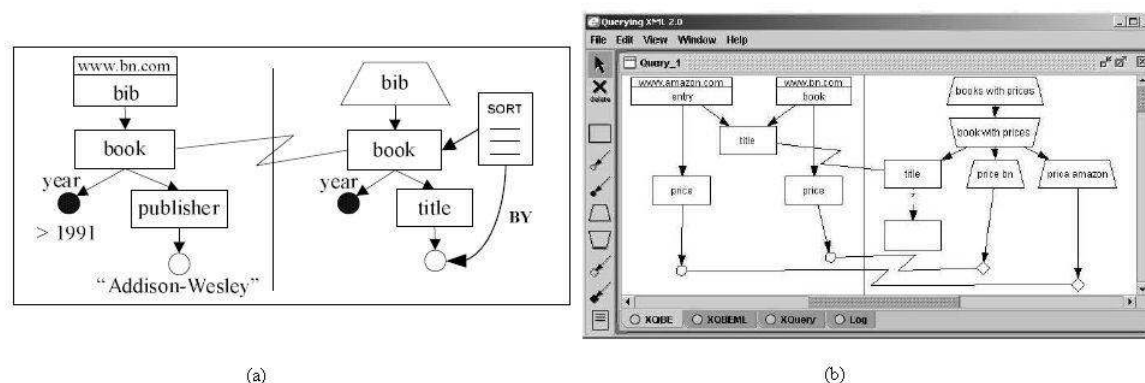


Figura 7.5: Interface de Consulta XQBE.

Dos dois objetivos principais deste trabalho tem-se que o primeiro é semelhante ao objetivo da interface-protótipo. Diferem-se, no entanto, no uso da XQuery. A interface-protótipo foi projetada para ser de uso de sistemas em geral. Ela auxilia na construção de uma árvore, resultando em uma linguagem simples, sem o uso de operadores e conectivos.

## 7.4 XQuery Gráfica

Em [Qin et al., 2004], novamente a linguagem XQuery serve de referência para a criação de GXQL, uma interface gráfica de consulta em documentos XML. Neste caso, o conceito de janelas aninhadas (Nested-Windows) é utilizado para representar a estrutura dos documentos XML. Elementos, atributos e valores são distinguidos na interface de forma clara. A estrutura do documento pode ser explorada de maneira gradativa, começando pelos elementos raízes e expandindo seus elementos filhos. A interface traduz a consulta gráfica realizada pelo usuário em consultas XQuery, sendo suportado consultas com FLWOR, predicados, quantificadores e operadores booleanos. Como em outras linguagens gráficas algumas cláusulas XQuery não são suportadas, tais como expressões condicionais, *type casting* e funções.

A figura 7.6 exemplifica a idéia da interface. As janelas aninhadas são utilizadas para expressar as relações de parentesco entre os elementos XML. Quando comparadas a outras linguagens visuais como XQBE, a quantidade de símbolos em GXQL é propositalmente menor. A idéia é utilizar elementos mais intuitivos e reduzir a carga do usuário quanto ao número de elementos. A idéia de se ter uma área de formulação da consulta com base nos documentos a serem consultados (painel à esquerda) e uma área para definir a estrutura do resultado desejado é mantida (painel à direita). Uma barra de ferramentas dá acesso aos elementos utilizados para formular a consulta.

Percebe-se que este trabalho assemelha-se bastante com o proposto na interface-protótipo, a principal diferença está na forma de ajudar o usuário na formulação da árvore para a consulta, feita passo-a-passo com a apresentação dos filhos de cada tag.

## 7.5 BBQ: Interface para navegação e consulta de documentos XML

Em [Munroe and Papakonstantinou, 2000] a ferramenta BBQ foi desenvolvida para permitir a consulta e navegação em repositórios XML. BBQ é voltada para pessoas que tenham conhecimento sobre a estrutura desse tipo de documento e consigam entender conceitos de cruzamento e agregação de dados.

Novamente, as consultas são feitas usando o paradigma Query by Example. A representação dos dados segue uma estrutura em árvore, sendo possível alternar entre a

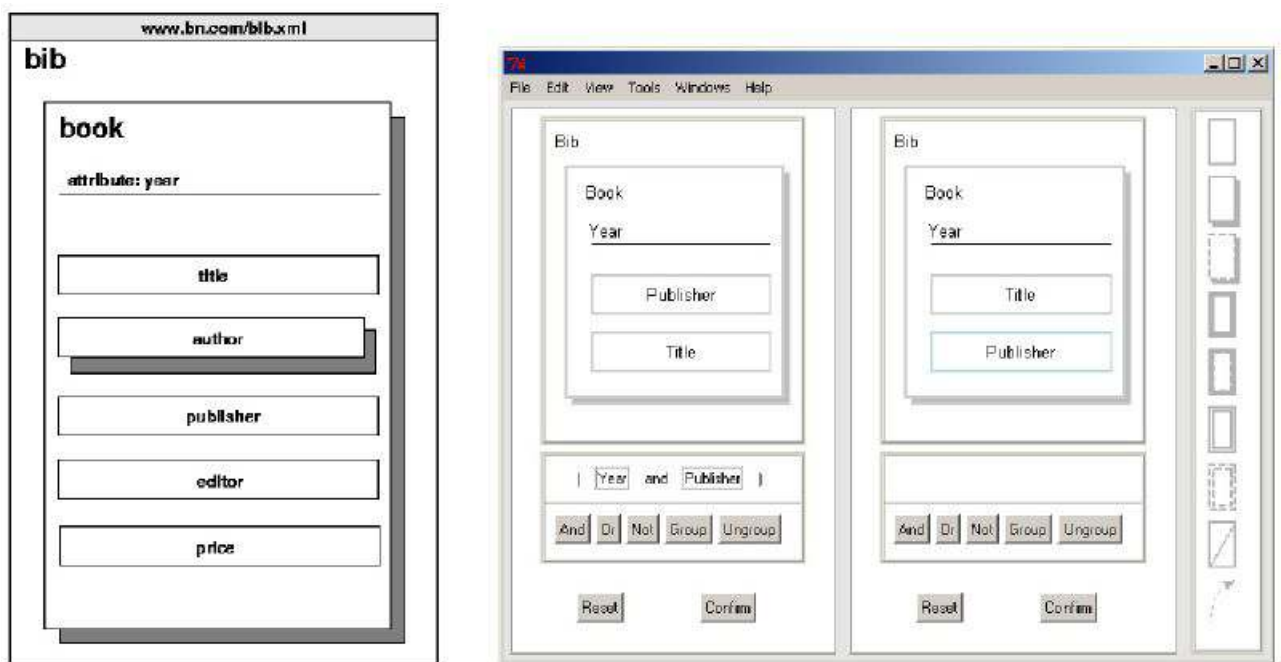


Figura 7.6: Interface de consulta GXQL.

estrutura da DTD e os dados do documento. Além disso, é possível navegar por múltiplas fontes de dado XML simultaneamente. Essas várias fontes podem ser combinadas na construção das consultas e formatação dos resultados. Outro ponto interessante é a possibilidade de reaproveitar consultas em diferentes fontes de dados. O sistema trata tanto documentos com DTD bem definida, quanto documentos sem uma especificação precisa. Para isso um sistema de inferência de DTD é utilizado na geração dos resultados. As consultas especificadas visualmente são traduzidas na linguagem XMAS, a qual tem a forma *CONSTRUCT head WHERE body*, onde *head* refere-se ao formato de resultado desejado e *body* às restrições selecionadas pelo usuário na formulação da consulta. A formulação de consultas e geração dos resultados ocorre em um ciclo iterativo formado pelos seguintes passos: 1. Definição de restrições sobre as fontes de dados (*body*); 2. Criação da "árvore de resultados" (*head*); 3. Geração da consulta em XMAS e encaminhamento ao mediador; 4. Geração de DTD para a árvore de resultados e visualização do esquema e dados resultantes.

A figura 7.7 ilustra parte da interface de BBQ. As janelas à esquerda e à direita representam duas fontes de dado XML, sobre as quais o usuário aplica restrições (neste caso sobre a DTD) para a formulação da consulta. A estrutura do resultado desejado é visualizado no painel central (*head*). O sistema dá suporte a várias operações para a inclusão de restrições sobre as DTD e dados XML tais como: comparação ('=' , '<' , '>' , etc), cruzamento e ligação de campos (join e binding) e filtros.

Percebe-se que, mesmo que a interface desenvolvida neste trabalho permita o trabalho

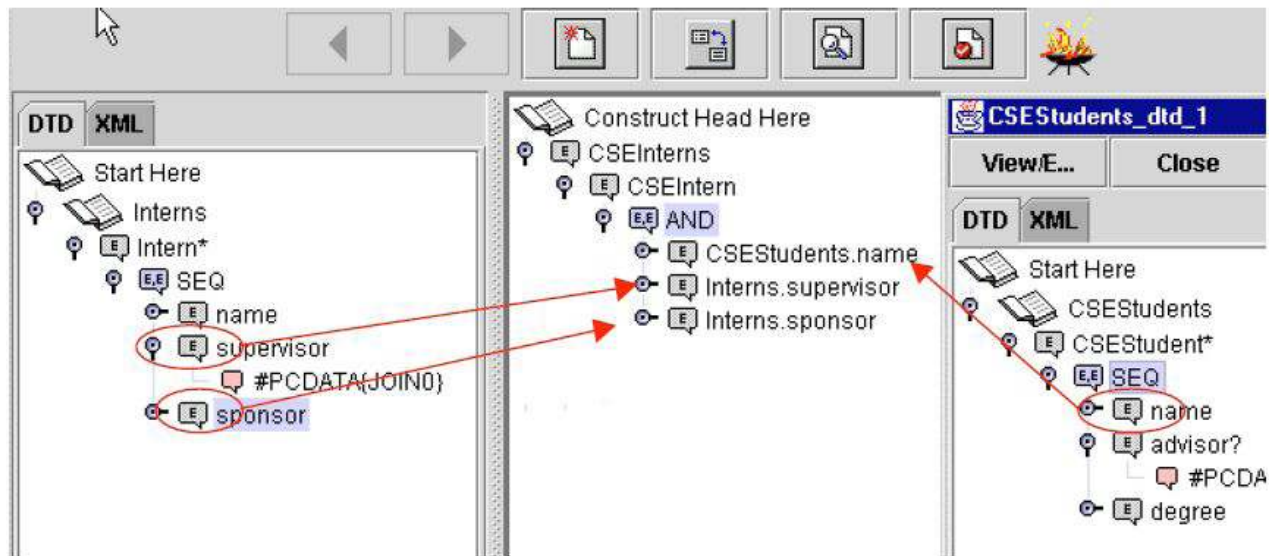


Figura 7.7: Interface de consulta BBQ.

com documentos contendo ou não o DTD, característica que se assemelha ao objetivo da interface-protótipo, este trabalho é voltado para aqueles que já tenham conhecimento sobre os documentos XML e entendam conceitos de cruzamento e agregação de dados, característica oposta à interface-protótipo, desenvolvida para usuários que não tenham conhecimento sobre os documentos e sua estrutura.

## 7.6 Considerações Finais

Muitos autores trabalham com o desenvolvimento de interface para linguagens, sistemas específicos, enquanto outros, procuram desenvolver uma interface que atenda a diversos sistemas, como a interface-protótipo proposta neste trabalho.

O protótipo apresentado nos fornece uma interface de fácil utilização que independe de características inerentes ao sistema, a ajuda nela apresentada é feita a partir da coleção, sem necessidade do DTD, podendo ser aplicada a coleções heterogêneas ou documentos XML representando dados estruturados.

## Capítulo 8

# Conclusão e Perspectivas Futuras

É fato que as consultas estruturadas melhoram a precisão dos resultados e que montar tais consultas não é fácil. A interface deve ajudar os usuários em tal tarefa.

Seguir padrões de desenvolvimento de interface ajuda o usuário em sua utilização e na obtenção do objetivo final no uso do sistema.

Neste trabalho foi apresentado um protótipo de interface para a consulta a documentos semi-estruturados. Tal interface auxilia o usuário no processo de construção de consulta estruturada, sem que o usuário tenha grandes conhecimentos sobre a estrutura do documento.

Além deste protótipo, apresentamos algumas características que, se inseridas no desenho da interface auxiliariam o usuário ainda mais, tornando a interface agradável de se utilizar.

Para a validação da eficiência deste protótipo, foram realizadas algumas consultas, sem estrutura e com estrutura, estas totalmente construídas com o auxílio da interface-protótipo. O aumento da precisão dos resultados foi realmente grande e válido, obtendo uma melhoria média de 443,85% nos resultados. Isso mostra que a interface tem condições de auxiliar o usuário a formular consultas bem estruturadas, alcançando ganhos significativos na precisão. Entretanto uma questão importante é a dependência do conhecimento prévio da interface para que esses resultados fossem alcançados.

Foram realizados alguns testes para a verificação da percepção de usuários sem conhecimento prévio da interface com relação a este protótipo. O protótipo foi aprovado pelos usuários, que realizaram consultas com estrutura que alcançaram uma melhoria média de 156,25% na precisão das respostas. Essa melhoria mostra que mesmo usuários sem experiência e conhecimento da interface alcançam bons resultados de precisão quando utilizam a interface protótipo para a formulação de consultas semi-estruturadas, mostrando que a interface-protótipo atinge o objetivo de auxiliar o usuário na formulação de consultas com estrutura.

Como trabalhos futuros, temos a necessidade de uma verificação dos critérios de usabilidade e ergonomia realizada a partir de testes na interface. Usuários de grupos distintos

devem testá-la e responder questionários de percepção sobre a interface. A Usabilidade deverá ser avaliada seguindo os critérios da norma ISO 9241-11.

Em [Tidwell, 2005] cita-se que um dos elementos no processo do desenho da interface é ter tempo para haver interação entre várias versões da interface já que pode não ser alcançada a melhor opção na primeira vez. Neste sentido, consideramos de extrema importância que, após a interação do usuário com a interface protótipo, que esta seja melhorada de acordo com as percepções destes, e sejam novamente testadas, até que se alcance uma interface ótima.

Como primeiros resultados da avaliação da interface por usuários, sugere-se melhorias como:

- para cada termo a ser pesquisado, independente de qual tag ele esteja ligado, o usuário poderá clicar sobre ele (clique duplo) e aparecerá um seletor de peso do termo na consulta, o peso estabelece a relevância daquele termo para o usuário em relação à consulta como um todo. O usuário deslizará neste seletor escolhendo entre 0 e 100% de relevância. Se não houver esta seleção todos os termos serão consultados com o mesmo peso.
- modificação no tamanho dos painéis de construção da consulta e de apresentação dos resultados, para que consultas maiores sejam melhor visualizadas.
- inserção de mensagens de auxílio quando os usuários estão realizando as consultas.
- diferenciação entre termos-tag e termos-folha quando apresentados na ajuda para facilitar a visualização do usuário.
- existência de um link de explicações para a utilização da interface, com exemplos, que deve ficar bem visível ao usuário e com fácil acesso.

Na implementação da interface deve-se tomar cuidado com relação à escalabilidade e tempo de resposta. Deve-se lembrar que um dos grandes problemas na recuperação semi-estruturada é o aumento na quantidade de termos que devem ser indexados.

Para a realização das ajudas, a interface-protótipo foi implementada com o auxílio de um gerenciador de banco de dados, além do índice gerado pelo próprio sistema sTerm. Alternativas para a melhoria no tempo de resposta do sistema não foram exploradas, sugere-se que sejam realizados experimentos com outros sistemas gerenciadores de banco de dados para verificar a melhora na velocidade do processamento.

# Referências Bibliográficas

- [Augurusa et al., 2003] Augurusa, E., Braga, D., Campi, A., and Ceri, S. (2003). Design and implementation of a graphical interface to xquery. In *SAC '03: Proceedings of the 2003 ACM symposium on Applied Computing*, pages 1163–1167, New York, NY, USA. ACM.
- [Baeza-Yates and Ribeiro-Neto, 1999] Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison Wesley, Cambridge UP.
- [Bastien et al., 1992] Bastien, J. M. C., , and Scapin, D. L. (1992). A validation of ergonomic criteria for the evaluation of human-computer interfaces. *International Journal of Human-Computer Interaction*.
- [Bastien and Sacpin, 1995] Bastien, J. M. C. and Sacpin, D. L. (1995). Evaluating a user interface with ergonomic criteria. *International Journal of Human-Computer Interaction*.
- [Bastien and Scapin, 1993] Bastien, J. M. C. and Scapin, D. L. (1993). Ergonomic criteria for the evaluation of human-computer interfaces. *INRIA -Institut National de Recherche en Informatique et en Automatique*.
- [Benyon, 1992] Benyon, D. (1992). The role of task analysis and systems design.
- [Braga and Campi, 2003] Braga, D. and Campi, A. (2003). A graphical environment to query xml data with xquery. In *WISE '03: Proceedings of the Fourth International Conference on Web Information Systems Engineering*, page 31.
- [Braga et al., 2005] Braga, D., Campi, A., Martinenghi, D., Raffio, A., and Salvi, D. (2005). Xqbe: the swiss army knife for semi-structured data. In *SEBD*, pages 284–291.
- [Branski, 2004] Branski, R. M. (2004). *Recuperação de Informações na Web*, volume 9. Perspectivas em Ciência da Informação - Belo Horizonte.
- [Cardoso, 1999] Cardoso, J. C. (1999). Considerações ao projeto de interfaces de bibliotecas digitais. Trabalho de curso - pós-graduação em ciência da computação, Pontifícia Universidade Católica do Rio Grande do Sul, PUC-RS.
- [Carmel et al., 2002] Carmel, D., Efraty, N., Landau, G. M., Maarek, Y. S., and Mass, Y. (2002). An extension of the vector space model for querying xml documents via xml fragments. *XML and Information Retrieval workshop of SIGIR 2002*.
- [Chen, 1994] Chen, S. (1994). On inference rules of logic-based information retrieval systems. In *Information Processing and Management*, pages 43–59.

- [Chiaramella, 1992] Chiaramella, Y. (1992). About retrieval models and logic. In *The Computer Journal*, pages 233–241.
- [Constantine, 1999] Constantine, Larry L. and Lockwood, L. A. D. (1999). *Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design*.
- [Croft, 1993] Croft, B. (1993). Knowledge-based and statistical approaches to text retrieval. *IEEE Expert: Intelligent Systems and Their Applications*, pages 8–12.
- [Croft, 1995] Croft, B. (1995). Effective text retrieval based on combining evidence from the corpus and users. *IEEE Expert: Intelligent Systems and Their Applications*, pages 59–63.
- [da Silva, 2004] da Silva, F. C. M. (2004). How to design library web sites to maximize usability. [http : //www.elsevier.com/framework\\_librarians/LibraryConnect/lcpamphlet5.pdf](http://www.elsevier.com/framework_librarians/LibraryConnect/lcpamphlet5.pdf).
- [da Silva, 2006] da Silva, F. C. M. (2006). Interface humano-computador: Uma visão principiante. *Governo do Estado do Piauí - Fundação de Amparo à Pesquisa do Estado do Piauí*.
- [de Normas Técnicas. NBR 9241-11, 2002] de Normas Técnicas. NBR 9241-11, A. B. (2002). Requisitos ergonômicos para trabalho de escritório com computadores: Parte 11: Orientação sobre usabilidade. Technical report.
- [dos Santos, 2006a] dos Santos, D. G. (2006a). Meta-modelo funcional para recuperação de informação baseado em  $\lambda$  cálculo. Dissertação de mestrado, Faculdade de Computação, UFU.
- [dos Santos, 2006b] dos Santos, R. L. G. (2006b). Usabilidade de interfaces para sistemas de recuperação de informação na web-estudo de caso de bibliotecas on-line de universidades federais brasileiras. Tese de doutorado, Pontifícia Universidade Católica - Rio De janeiro.
- [Fleming, 1998] Fleming, J. (1998). *Web navigation: designing the user experience*. O'Reilly, Sebastopol, CA.
- [George, 2005] George, C. (2005). Usability testing and design of library website: An iterative approach. *OCLC systems and services*, 21(3):167–180.
- [Globo.com, 2009] Globo.com (2009). [Online; accessed 4-fevereiro-2009].
- [Großjohann et al., 2002] Großjohann, K., Fuhr, N., Effing, D., and Kriewel, S. (2002). A user interface for xml document retrieval. *Informatik bewegt: Informatik 2002 - 32. Jahrestagung der Gesellschaft für Informatik e.v.*, 19:166–170.
- [Hoelzel, 2004] Hoelzel, C. G. M. (2004). Design ergonômico de interfaces gráficas humano-computador: Um modelo de processo. Tese de doutorado, Universidade Federal de Santa Catarina.
- [HTML, 2009] HTML (2009). Html 4.01 specification. <http://www.w3.org/TR/html4/>.
- [INEX, 2009] INEX (2009). Inex - initiative for the evaluation of xml retrieval. <http://qmir.dcs.qmul.ac.uk/INEX/>.

- [Júnior, 2008] Júnior, M. B. F. (2008). Tutorial xml. [Online; Último acesso: 15-setembro-2008].
- [Johnson, 1992] Johnson, P. (1992). *Human-Computer Interaction: Psychology, Task Analysis and Software Engineering*. Mc-Graw Hill, Maidenhead, UK.
- [Johnson, 2001] Johnson, S. (2001). *Cultura da Interface*. Jorge Zahar Editor, Rio de Janeiro.
- [Kang and Choi, 1997] Kang, H. and Choi, K. (1997). Two-level document ranking using mutual information in natural language information retrieval. *Inf. Process. Manage.*, pages 289–306.
- [Lancaster and Fayen, 1973] Lancaster, F. W. and Fayen, E. G. (1973). Information retrieval on-line. In *Los Angeles, CA: Melville*.
- [Lima, 2001] Lima, M. E. R. (2001). Análise e especificação de requisitos ergonômicos para sistemas de recuperação de informação (sri) na web. Dissertação de mestrado, Universidade Federal de Santa Catarina, UFSC.
- [Manning et al., 2007] Manning, C. D., Raghavan, P., and Schütze, H. (2007). *Introduction to Information Retrieval*. Preliminary draft (c), Cambridge UP.
- [Mass et al., 2003] Mass, Y., Mandelbrod, M., Amitay, E., Maarek, Y. S., and Soffer, A. (2003). Juruxml - an xml retrieval system at inex'02. *Proceedings of the 1st Workshop of the Initiative for the Evaluation of XML Retrieval (INEX)*.
- [Meister and Enderwick, 2002] Meister, D. and Enderwick, T. P. (2002). *The User - Human Factors in Systems Design*. Lawrence Erlbaum, London.
- [Moraes, 2004] Moraes, A. d. (2004). Ergonomia, ergodesign e usabilidade: Algumas histórias, precursores, divergências e convergências. *Ergodesign/usich coletânea de palestras de convidados internacionais e nacionais*.
- [Munroe and Papakonstantinou, 2000] Munroe, K. D. and Papakonstantinou, Y. (2000). Bbq: A visual interface for integrated browsing and querying of xml. pages 277–296.
- [Newman and Ozsoyoglu, 2004] Newman, S. and Ozsoyoglu, Z. M. (2004). A tree-structured query interface for querying semi-structured data. In *SSDBM '04: Proceedings of the 16th International Conference on Scientific and Statistical Database Management*, page 127.
- [Nielsen, 1999] Nielsen, J. (1999). When bad design elements become the standard. [Online; Último acesso: 12-dezembro-2008].
- [Oliveira, 2001] Oliveira, E. R. d. (2001). Avaliação ergonômica de interfaces da scielo - scientific electronic library online. Dissertação de mestrado, Engenharia de Produção, Universidade Federal de Santa Catarina.
- [Pollock and Hockley, 1997] Pollock, A. and Hockley, A. (1997). What's wrong with internet searching. In *D-Lib Magazine*. <<http://www.dlib.org/dlib/march97/bt/03pollock.html>>. Acesso em 12 jan. 2009.

- [Qin et al., 2004] Qin, Z., Yao, B. B., Liu, Y., and McCool, M. D. (2004). A graphical xquery language using nested windows. In *WISE*, pages 681–687.
- [Salton, 1974] Salton, G. and Wong, A. Y. C. (1974). A vector space model for automatic indexing. Technical report tr74-218, Cornell University, Computer Science Department.
- [Schlieder and Meuss, 2002] Schlieder, T. and Meuss, H. (2002). Querying and Ranking XML Documents. *Journal of the American Society for Information Systems*, 53(6):489–503.
- [Shneiderman, 1998] Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, 3. ed. Reading, MA.
- [Submarino, 2009] Submarino (2009). Submarino. [Online; accessed 4-fevereiro-2009].
- [Teodoro, 2001] Teodoro, V. (2001). Uma análise da usabilidade de diferentes interfaces para máquinas de busca. Monografia de graduação, Universidade Federal de Lavras.
- [Theobald et al., 2005] Theobald, M., Schenkel, R., and Weikum, G. (2005). An efficient and versatile query engine for topx search. In *Proceedings of the 31st VLDB Conference*.
- [Tidwell, 2005] Tidwell, J. (2005). *Designing Interfaces*. O'Reilly, Sebastopol, CA.
- [TopX, 2009] TopX (2009). Topx search. <http://infao5501.ag5.mpi-sb.mpg.de:8080/topx/?login>.
- [W3C, 2009] W3C (2009). Extensible markup language (xml). <http://www.w3.org/XML/>.
- [Walter de Abreu Cybis, 1999] Walter de Abreu Cybis, Marcelo Soares Pimenta, M. C. S. L. G. (1999). Uma abordagem ergonômica para o desenvolvimento de sistemas interativos. *Laboratório de Utilizabilidade da Informática - LabIUtil/Inf - Universidade Federal de Santa Catarina*.
- [XSL, 2009a] XSL (2009a). Extensible stylesheet language (xsl) family. <http://www.w3.org/Style/XSL/>.
- [XSL, 2009b] XSL (2009b). Extensible stylesheet language(xsl) version 1.1. <http://www.w3.org/TR/xsl/>.

# Apêndice A

## Tópicos INEX Triviais

A Seguir são apresentados os tópicos INEX utilizados para a execução dos testes na interface protótipo:

```
<inex_topic topic_id="417" ct_no="9" >
<title>therapeutic breathing</title>
<castitle>// * [about(., therapeutic breathing)]</castitle> // <description>Find the use
of special breathing techniques for therapeutic effects</description>
<narrative>The relevant documents should talk about use of special breathing techniques
used for therapeutic purposes, either preventive or curative ones. These techniques must
not use neither drugs nor special devices. Retrieved techniques should include: rebirthing,
holotropic breathwork, yoga.</narrative>
</inex_topic>
```

```
<inex_topic topic_id="449" ct_no="62" >
<title>weather prediction</title>
<castitle>// * [about(., weather prediction)]</castitle>
<description>I want informations about weather prediction.</description>
<narrative>A friend of mine will defend its PhD thesis next week. He works on weather
prediction. In order to understand his work a little bit, I'd like to find basic informations
on weather forecast, like for example how weather is predicted, the history of meteorology,
which phenomena scientific methods are able to predict,...</narrative>
</inex_topic>
```

```
<inex_topic topic_id="452" ct_no="65" >
<title>seven ancient wonders world</title>
<castitle>// * [about(., seven ancient wonders world)]</castitle>
<description>Find the name and description of the seven ancient wonders of the world
</description>
<narrative>I am interested in knowing more about the seven wonders of the world, but
```

only the ancient ones. I am already aware of the Great Pyramid of Giza and the lighthouse of Alexandra, but I am not sure of the others. Also I would like a description of each of them. To be relevant a document (part) must mention the names of one or several ancient wonders of the world, preferentially with a description.</narrative>

</inex\_topic>

```
<inex_topic topic_id="458" ct_no="75" >
<title>cellphone technology networks</title>
<castitle>// * [about(., cellphone technology – networks)]</castitle>
<description>Find information about different cellphone technologies, but not about specific cellphone networks</description>
<narrative>Recently in New Zealand, a major telecommunications company shut down one of its older cellphone networks, stating that the technology was outdated. This perked my interest in the different technologies that have been used with cellphones, both historically and in the present day. I want to learn more about these different technologies.
```

To be relevant the element should talk about a technology that has been used in cellphones, either past or present. This could talk about the different generations of phones (like the current '3G' phones being short for third generation), or the actual network technology used.

It is not relevant if the element discusses an actual cellphone network, such as Vodafone, with out mentioning some form of technology as well. </narrative>

</inex\_topic>

```
<inex_topic topic_id="460" ct_no="78" >
<title>java MVC framework model-view-controller</title>
<castitle>// * [about(., java mvc framework model – view – controller)]</castitle>
<description>To find information about Model-View-Controller(MVC) Framework implemented in Java.</description>
<narrative>As a significant framework when building various applications(b/s or c/s whatever), Model-View-Controller(MVC) Framework is widely used in many mature resolutions of prevailing Programming. I want to get information about the implementations of Model-View-Controller(MVC) Framework especially in Java programming language. An acceptable elements should contain the basic concept of Model-View-Controller(MVC) Framework, the tutorial of how to build applications with the framework and the comparison of efficiency and performance between different implementations.</narrative>
```

</inex\_topic>

```
<inex_topic topic_id="462" ct_no="82" >
<title>bea weblog</title>
```

```
<castitle>// * [about(., bea weblogic)]</castitle>
<description>To find information of BEA Weblogic server, such as how to deploy ap-
plications, the system requirement, the compatibility and comparison to other preveiling
application servers.</description>
<narrative>BEA Weblogic is one of the preveiling Java Application servers. As a J2EE
developer, I would like to get some information on Weblogic to help me make a decision
whether to use it as a default environment in our recent projects. The relevant elements
should include information about the deployment of applications in weblogic, the system
requirement to run the server, the compatibility of the server container to various ver-
sions of jsp and java servlet. Furthermore, elements about the comparison of the cost of
enterprise deploying applications on weblogic to that of deploying on other application
servers, such as IBM websphere and Apache, will be appreciated.</narrative>
</inex_topic>
```

```
<inex_topic topic_id="465" ct_no="87" >
<title>difference American British English</title>
<castitle>// * [about(., difference American British English)]</castitle>
<description>I'm looking for differences between American English and British English.
</description>
<narrative>I'm looking for differences between British English and American English.
I'm interested in examples and explanations of differences of any kind, including spelling,
pronunciation, grammar, vocabulary and meaning. Larger documents discussing the
differences between these variants are higly relevant, but also smaller sections mentioning
just one or a few example are appreciated. If an item only mentions a difference without
giving examples or explaining (e.g. 'BE and AE differ in spelling'), it is not considered
relevant. Also, I'm not interested in learning where the different variants are used, or
what their respective influences were on other variants like Malaysian or Indian English.
Finally, pointers to information on differences are not good enough, I want direct access
to examples or explanations.</narrative>
</inex_topic>
```

```
<inex_topic topic_id="467" ct_no="91" >
<title>Definition and ressources on text mining</title>
<castitle>// * [about(., Definition and ressources on text mining)]</castitle>
<description>I would like to know what is text mining</description>
<narrative>I would like to know what is text mining and to find ressources : links, tools,
corpora. I am also interested in biomedical text mining. I am not interested by data
mining or mining.</narrative>
</inex_topic>
```

```
<inex_topic topic_id="473" ct_no="101" >
<title>Kernel Machines Machine Learning</title>
<castitle>// * [about(., Kernel Machines Machine Learning)]</castitle>
<description>I want to know the theory of Kernel Machines.</description>
<narrative>Kernel Machines are a set of related supervised learning methods using Kernel Methods for classification and regression. And the most famous Kernel Machine is Support Vector Machine(SVM), but SVM is unfit for fast additional learning. Then my aim is simply to know any other Kernel Machines. However, 'Kernel' means some different senses. So I hope that irrelevant documents excepted, which talk about the central component of most computer operating systems, a Unix-like operating system kernel like the Linux kernel, and also mathematics means.</narrative>
</inex_topic>
```

```
<inex_topic topic_id="499" ct_no="153" >
<title>Grammy Awards</title>
<castitle>// * [about(., Grammy Awards)]</castitle>
<description>Find information on 'Grammy awards' about its history and award-winners for different categories in different years.</description>
<narrative>Grammy award is one of the great milestones for every music performer. I am interested in the history of Grammy awards and the name of the award-winners. Sections or paragraphs mentioning the name of audio or video album or cassettes, records which were either nominated for or awarded are relevant. But paragraph or element on a Grammy award-winner's life and works is not important.</narrative>
</inex_topic>
```

```
<inex_topic topic_id="502" ct_no="156" >
<title>Britney Spears</title>
<castitle>// * [about(., Britney Spears)]</castitle>
<description>I want to know facts about Britney Spears. I'm interested in both her career as a professional musician and her personal life including her friends, family, hobbies, and the curious incidents that involve her.</description>
<narrative>Relevant documents describe Britney Spears or people closely related to her. What happens outside her professional career is as interesting as her musical achievements. Description of fan products such as celebrity dolls is not relevant. Neither is technical information about her albums, dvds, movies, or the chart performance of her singles.</narrative>
</inex_topic>
```

```

<inex_topic topic_id="503" ct_no="160" >
<title>conjectures and unsolved problems in mathematics</title>
<castitle>//*[about(., conjectures and unsolved problems in mathematics)]</castitle>
<description>I'm looking for descriptions of mathematical unsolved problems or conjectures</description>
<narrative>I want descriptions of mathematical unsolved problems. I am a student in mathematics and I want to see the difficulty of unsolved problems. I am more interested in proper formulations of these problems than just a list of them. Recently solved conjectures can also be accepted since they have been unsolved for a long time.</narrative>
</inex_topic>

```

```

<inex_topic topic_id="505" ct_no="163" >
<title>software patent europe</title>
<castitle>//*[about(., software patent europe)]</castitle>
<description>informations about patentability of software in Europe. </description>
<narrative>I would like to write a report concerning software patent in Europe. First of all, I want to know the definition of a software patent, its advantages and disadvantages. Then, I would like collect the reasons supporting and against patentability and also the position of Europe in the debate. At the end, if there is a comparison between copyright and patent I will be a plus. I do not require details such as patent cooperation treaty, open patent or open software licence and I am not interested in the general definition of patent.</narrative>
</inex_topic>

```

```

<inex_topic topic_id="506" ct_no="164" >
<title>image processing</title>
<castitle>//*[about(., image processing)]</castitle>
<description>informations about image processings. </description>
<narrative>I would like to make an introductory course on the image processing. I want some general information such as the definition, history and uses. In addition, I would like seek information on image processing techniques like Morphological image processing or segmentation... </narrative>
</inex_topic>

```

```

<inex_topic topic_id="509" ct_no="172" >
<title>Life and music of Bob Dylan</title>
<castitle>//*[about(., Bob Dylan)]</castitle>
<description>Find information about Bob Dylan's life and music albums and songs which make him great.</description>

```

<narrative>I am fond of Bob Dylan's songs. I am interested in his life, times and music. Why his songs contain political messages, which scenario triggered him making those compositions are of my interest. Any element discussing his musical achievement or biographical sketch or anecdotes will be relevant.</narrative>

</inex\_topic>

<inex\_topic topic\_id="514" ct\_no="180" >

<title>food additives physical health risk grocery store labels</title>

<castitle> // \*[*about(., food additives physical health risk grocery store labels)*] </castitle>

<description>Find information about food additives and health risks.</description>

<narrative>You recently heard about the book 'Fast Food Nation,' and it has really influenced the way you think about your diet. You note in particular the amount and types of food additives contained in the things that you eat every day. Now you want to understand which food additives pose a risk to your physical health, and are likely to be listed on grocery store labels.</narrative>

</inex\_topic>

<inex\_topic topic\_id="518" ct\_no="185" >

<title>tidal power wind power</title>

<castitle> // \*[*about(., tidal power wind power)*] </castitle>

<description>Find information about the pros and cons of tidal power and wind power.</description>

<narrative>You just joined the citizen's advisory committee for the city of St. John's, Newfoundland. With the increase in fuel costs, the city council is contemplating supplementing its power with alternative energy. Tidal power and wind power are being discussed among your fellow committee members. As you want to be fully informed when you attend the next meeting, you research the pros and cons of each type.</narrative>

</inex\_topic>

<inex\_topic topic\_id="519" ct\_no="187" >

<title>types of bridges vehicles water ice</title>

<castitle> // \*[*about(., types of bridges vehicles water ice)*] </castitle>

<description>Find information about different types of bridges suitable for vehicles and able to withstand water and ice.</description>

<narrative>Your community is contemplating building a bridge across a span of water measuring 1000 M in order to ease traffic congestion. There will be a presentation this evening about the type of bridge proposed for the project. To date, many types of bridges have been discussed: 'folding bridge,' 'suspension bridge', 'retractable bridge', and 'basculer bridge'. In order to be well informed when you attend the meeting, you need

information on what type of bridge would best suit the community's needs, bearing in mind that the solution must accommodate vehicles and be sturdy enough to withstand a body of water that can be rough and frozen over with ice in winter.</narrative>

</inex\_topic>

<inex\_topic topic\_id="523" ct\_no="191" >

<title>fuel efficient cars</title>

<castitle>// \* [about(., *fuel efficient cars*)] </castitle>

<description>Find information about fuel efficient and environmentally friendly cars.

</description>

<narrative>A close friend is planning to buy a car for the first time, but is worried about fuel costs and the impact on the environment. The friend has asked for help in learning about options for vehicles that are more fuel efficient and environmentally friendly. What types of different types of engines, manufacturers and models of cars might be of interest to your friend? What would be the benefits of using such vehicles?</narrative>

</inex\_topic>

<inex\_topic topic\_id="524" ct\_no="192" >

<title>home heating solar panels</title>

<castitle>// \* [about(., *home heating solar panels*)] </castitle>

<description>Find information about solar panels in relation to home heating. </description>

<narrative>Friends are planning to build a new house and have heard that using solar energy panels for heating can save a lot of money. Since they do not know anything about home heating and the issues involved, they have asked for your help. You are uncertain as well, and do some research to identify some issues that need to be considered in deciding between more conventional methods of home heating and solar panels.</narrative>

</inex\_topic>