
Um modelo de navegação exploratória para a
infra-estrutura da Web Semântica

Luciano Tadeu Esteves Pansanato

Um modelo de navegação exploratória para a infra-estrutura da Web Semântica

Luciano Tadeu Esteves Pansanato

Orientadora: Profa. Dra. Renata Pontin de Mattos Fortes

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências - Ciências de Computação e Matemática Computacional.

“VERSÃO REVISADA APÓS A DEFESA”

Data da Defesa:	21/11/2007
Visto do Orientador:	

USP – São Carlos
Novembro/2007

Dedicatória

*À minha esposa, Camilla,
pelo amor, pela paciência e
pelas horas roubadas
ao seu convívio.*

*À minha filha, Laissa,
simplesmente por existir.*

Agradecimentos

A Deus.

À minha esposa, Camilla, e à minha filha, Laissa.

Aos meus pais, Hélio e Irene, minha irmã Maria Angélica, minhas “dindas” Victória e Regina, e ao meu primo Germano.

À minha orientadora Renata Pontin de Mattos Fortes.

A todos os professores e funcionários do ICMC/USP.

Aos meus amigos do Intermídia, em especial ao André, Débora e Silvana.

À UTFPR.

À FINEP.

À FAPESP.

Epígrafe

Ensinar é um exercício de imortalidade.

*De alguma forma continuamos a viver naqueles
cujos olhos aprenderam a ver o mundo pela
magia da nossa palavra.*

O professor, assim, não morre jamais...

Rubem Alves

Resumo

Pansanato, L. T. E. Um modelo de navegação exploratória para a infra-estrutura da Web Semântica. 2007. 166 p. Tese (Doutorado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2007.

Esta tese propõe um modelo de navegação exploratória para a infra-estrutura da Web Semântica, denominado *Navigation and Exploration Model* (NAVE). O modelo NAVE foi desenvolvido com base na literatura de *information searching*, nos níveis de atividades de *information seeking* e na estratégia de *orienteering*. O objetivo é facilitar o projeto e desenvolvimento de sistemas de navegação exploratória. O NAVE é descrito por meio de uma representação gráfica dos estágios e decisões do processo de navegação e suas respectivas técnicas de suporte à navegação, além de recomendações. Um sistema, denominado de *Exploratory Navigation System* (ENS), foi desenvolvido para avaliar a viabilidade de utilizar o modelo NAVE em aplicações reais. O sistema ENS é composto de diversas ferramentas de navegação que permitem ao usuário escolher a ferramenta adequada, ou a melhor combinação de ferramentas, provavelmente ajustada ao seu nível de habilidade e conhecimento, à sua preferência e ao tipo de informação que ele está procurando no momento. O sistema permite ao usuário priorizar de maneiras diferentes as suas escolhas de ferramentas em cada passo de uma estratégia de *orienteering*, subjacente ao modelo NAVE. Essas ferramentas podem apresentar vantagens complementares no contexto de uma tarefa de *information searching*. O sistema ENS foi avaliado utilizando uma abordagem tanto qualitativa quanto quantitativa, que serviram para refinar as questões de pesquisa e explorar o modelo NAVE. Primeiro, um estudo de usabilidade foi conduzido que combinou vários métodos, como questionários, *think-aloud*, entrevistas e registro da interação do usuário. Esse estudo forneceu informações com relação às ferramentas e o modelo NAVE subjacente, as quais foram consideradas no seu desenvolvimento. Segundo, um estudo experimental foi conduzido para comparar o ENS com uma abordagem de busca por palavra-chave. Os resultados forneceram indicações estatísticas de que os participantes tiveram desempenho superior utilizando o ENS.

Palavras-chave: Navegação, Metadados, Web Semântica, Recuperação Interativa.

Abstract

Pansanato, L. T. E. A model for exploratory navigation in the Semantic Web infrastructure. 2007. 166 p. Thesis (Doctoral) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2007.

A model for exploratory navigation in the Semantic Web infrastructure called NAVE - Navigation and Exploration Model - is proposed. NAVE is based on literature of information searching, levels of information seeking activities, and an orienteering strategy. This model aims in particular at facilitating the design and development of exploratory navigation systems. It is described by a graphical representation of stages and decisions of the navigation process and their respective navigation support techniques, and recommendations. As a proof of concept and also to evaluate the feasibility of using NAVE in real-life applications, a system called ENS - Exploratory Navigation System - was developed. ENS is composed of a variety of navigation tools, enabling users to choose the appropriate tool or the best combination of tools (that is, the best strategy) in agreement with different levels of users' ability, background, preferences, and kind of information they are looking for at moment. It enables users to prioritize different ways their choices of tools to use at each step in an orienteering strategy embedded on the model NAVE. These tools may present complementary advantages in an information searching task. ENS was evaluated in both qualitative and quantitative approach which served to refine research questions and explore the model NAVE. First, a usability study was conducted which combined a variety of methods, such as questionnaires, think-aloud, interview, and user log recording. This study provided insights regarding the tools and the underlying model which were considered in its further development. Second, an experimental study was conducted in order to compare the ENS with a keyword search approach. The findings provided statistical indications that participants had a better performance using the ENS.

Keywords: Navigation, Metadata, Semantic Web, Interactive Retrieval.

Publicações

Pansanato, L. T. E., e Fortes, R. P. M. Browsing for educational wiki pages: an orienteering strategy. Submetido para publicação no *Journal of Educational Technology & Society*, <<http://www.ifets.info/>>.

Pansanato, L. T. E., e Fortes, R. P. M. Using an Orienteering Strategy to Browse Semantically-Enhanced Educational Wiki Pages. In *Proceedings of the IADIS International Conference e-Learning*, July 6–8, Lisbon, Portugal, 2007. 11–18.

Pansanato, L. T. E., e Fortes, R. P. M. System Description: An Orienteering Strategy to Browse Semantically-Enhanced Educational Wiki Pages. 4th European Semantic Web Conference (ESWC 2007), June 3–7, 2007, Innsbruck, Austria. In *E. Franconi, M. Kifer, and W. May (Eds.): ESWC 2007, Lecture Notes in Computer Science (LNCS), Vol. 4519*. Berlin Heidelberg: Springer-Verlag, 2007. 809–818. DOI=<http://dx.doi.org/10.1007/978-3-540-72667-8_59>.

Pansanato, L. T. E., e Fortes, R. P. M. *MeGen: Um gerador automático de metadados para a CoTeia*. São Carlos: ICMC/USP, 2006. (Relatório técnico, 287). 15 p. Disponível em: <http://www.icmc.usp.br/biblio/BIBLIOTECA/rel_tec/RT_287.pdf>.

Pansanato, L. T. E., e Fortes, R. P. M. Strategies for automatic LOM metadata generating in a Web-based CSCL tool. In *Proceedings of the 11th Brazilian Symposium on Multimedia and the web (WebMedia'05)*, Dec. 5-7, 2005, Poços de Caldas, MG, Brazil, 2005. ACM International Conference Proceeding Series, v. 125. New York, NY, USA: ACM Press, 2005. 1–8. DOI=<<http://doi.acm.org/10.1145/1114223.1114231>>.

Pansanato, L. T. E., e Fortes, R. P. M. Uma pesquisa em documentos Web baseada em estratégia de orientação. In *Anais do XI Simpósio Brasileiro de Multimídia e Web (WebMedia'05)*, 5-7 Dez. 2005, Poços de Caldas, MG, Brazil. Poços de Caldas: PUC Minas, 2005. 238–240.

Pansanato, L. T. E., e Fortes, R. P. M. Strategies for filling out LOM metadata fields in a Web-based CSCL tool. In *Proceedings of the Third Latin American Web Congress (LA-WEB 2005)*, 31 Oct.–2 Nov. 2005, Buenos Aires, Argentina. Los Alamitos, CA, USA: IEEE Computer Society, 2005. 187–190. DOI=<<http://dx.doi.org/10.1109/LAWEB.2005.35>>.

Fortes, R. P. M., Lara, S. M. A., Freire, A. P., e Pansanato, L. T. E. Capítulo 7: Acessibilidade no Projeto de Aplicações Web. In A. C. Teixeira, E. Barrére, I. C. Abrão (Org.). *Web e Multimídia: Desafios e Soluções*. Poços de Caldas: PUC Minas, 2005. 197–225.

Fortes, R. P. M., Pansanato, L. T. E., Lara, S. M. A., e Freire, A. P. Capítulo 2: Universalização do Acesso a Conteúdo Educacional em Aplicações Web. In F. M. Santoro, R. M. E. M. da Costa. (Org.). *Minicursos do XVI Simpósio Brasileiro de Informática na Educação (SBIE 2005)*. Juiz de Fora, MG, Brazil: Editar Editora Associada, 2005. 21–42.

Pansanato, L. T. E., e Fortes, R. P. M. Uma Análise de Metadados para o Acesso Unificado às Informações dos Repositórios de Ferramentas de Software Livre. In *Anais do II Congresso de Tecnologias para Gestão de Dados e Metadados do Cone Sul*, 16-17 Set. 2004, Ponta Grossa, PR, Brazil. Ponta Grossa: Universidade Tuiuti do Paraná, 2004. 12 p. Disponível em: <<http://conged.deinfo.uepg.br/iiconged/Artigos/artigo2.pdf>>

Fumagalli, L. C., Pansanato, L. T. E., e Fortes, R. P. M. Documentation Process in Interactive Systems - A Study Case to Abstract its Structure. In *Proceedings of the 2nd International Information and Telecommunication Technologies Symposium (I2TS 2003)*, Nov. 26-29, 2003, Florianópolis, SC, Brazil. Florianópolis: Fundação Barddal de Educação e Cultura, 2003. 8 p.

Abreviaturas e Siglas

AJAX	<i>Asynchronous Javascript And XML</i>
API	<i>Application Programming Interface</i>
CERN	<i>Conseil Européen pour la Recherche Nucléaire</i>
CGI	<i>Common Gateway Interface</i>
CSCL	<i>Computer Supported Collaborative Learning</i>
CSV	<i>Comma Separated Value</i>
DC	<i>Dublin Core</i>
DTD	<i>Document Type Definition</i>
ENS	<i>Exploratory Navigation System</i>
FRESS	<i>File Retrieval and Editing System</i>
GRDDL	<i>Gleaning Resource Descriptions from Dialects of Languages</i>
GUI	<i>Graphical User Interface</i>
HES	<i>Hypertext Editing System</i>
HTML	<i>HyperText Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
ICMC/USP	<i>Instituto de Ciências Matemáticas e de Computação / Universidade de São Paulo</i>
IETF	<i>Internet Engineering Task Force</i>
JSP	<i>JavaServer Pages</i>
KMS	<i>Knowledge Management System</i>
LOM	<i>Learning Object Metadata</i>

MeGen	<i>Metadata Generator</i>
MVC	<i>Model-View-Controller</i>
NAVE	<i>Navigation and Exploration Model</i>
NE	<i>Navigation Engine</i>
NLS	<i>oN-Line System</i>
OWL	<i>Web Ontology Language</i>
PARC	<i>Palo Alto Research Center</i>
PHP	<i>PHP: Hypertext Preprocessor</i>
RAP	<i>RDF API for PHP</i>
RDF	<i>Resource Description Framework</i>
RIF	<i>Rule Interchange Format</i>
RuleML	<i>Rule Markup Language</i>
SAFE	<i>Software Engineering Available for Everyone</i>
SGML	<i>Standard Generalized Markup Language</i>
SQL	<i>Structured Query Language</i>
SRI	<i>Stanford Research Institute</i>
SWRL	<i>Semantic Web Rule Language</i>
TCP/IP	<i>Transport Control Protocol / Internet Protocol</i>
TIDIA-Ae	<i>Tecnologia da Informação para o Desenvolvimento da Internet Avançada – Aprendizado Eletrônico</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
WWW	<i>World Wide Web</i>
W3C	<i>World Wide Web Consortium</i>
XHTML	<i>Extensible HyperText Markup Language</i>
XML	<i>Extensible Markup Language</i>

Glossário

Application profile: Uma combinação de elementos de metadados selecionados de um ou mais esquemas, ou ainda a extensão de esquemas existentes, com o intuito de se compor um conjunto de elementos adaptado às particularidades e necessidades de um determinado contexto.

Busca exploratória: Uma categoria ou classe de atividades de information searching na qual os usuários geralmente combinam estratégias de consulta e navegação para executar diversas tarefas que envolvem principalmente aprendizagem e investigação.

CoTeia: Uma ferramenta colaborativa assíncrona, baseada em Wiki, utilizada no ICMC/USP no contexto de *Computer Supported Collaborative Learning* (CSCL), para complementar as aulas presenciais com atividades colaborativas.

CoWeb: Uma ferramenta baseada conceitualmente em Wiki, desenvolvida no *Georgia Institute of Technology*, para auxiliar a criação de atividades colaborativas.

Ferramenta: Qualquer programa ou aplicação que fornece suporte para um usuário final executar uma tarefa específica.

Grafo RDF: Um conjunto de triplas (sujeito, predicado e objeto) ou declarações RDF, que pode ser ilustrado como um diagrama de nós e arcos orientados, no qual cada tripla é representada como uma ligação nó-arco-nó.

Informação: O conceito de informação tem de satisfazer duas condições simultaneamente: informação sendo algo que é o resultado de uma transformação das estruturas do conhecimento de um gerador; e informação sendo algo que, quando percebido, afeta e transforma o estado do conhecimento do receptor.

Information behavior: Comportamento humano relacionado à geração, comunicação, uso, e outras atividades que envolvem informação.

Information searching: Comportamento humano relacionado com a interação entre uma pessoa, com uma necessidade de informação, e um sistema de informação.

Information seeking: Comportamento humano relacionado com a pesquisa por informação por meio de fontes de informação formais e informais.

Metadados: Dados utilizados para identificar, descrever ou localizar recursos de in-

formação.

Modelo RDF: Conjunto de declarações armazenadas em uma base de dados que representam um grafo RDF.

Módulo: Uma manifestação genérica de um subprograma. Nas linguagens de programação convencionais, um subprograma pode ser chamado de sub-rotina, procedimento, função, ou outro nome especializado. Nas linguagens de programação orientadas a objeto, a concepção clássica de subprograma é substituída pelo objeto.

Necessidade de informação: Uma lacuna identificada de maneira consciente por uma pessoa no conhecimento disponível a ela.

Precisão: Medida para relevância que consiste na fração dos resultados recuperados que é considerada relevante.

Repositório de páginas Web: Um conjunto de páginas Web dinâmicas que derivam parte ou todo o seu conteúdo a partir de arquivos ou bases de dados.

Revocação: Medida para relevância que consiste na fração dos resultados relevantes que foi recuperada.

Widget: Um termo genérico utilizado para fazer referência a um componente gráfico de uma interface gráfica de usuário ou ao seu programa de controle, ou ainda para fazer referência à combinação de ambos.

Wiki: Um *website* colaborativo composto do trabalho coletivo de diversos autores. O termo refere-se tanto ao website quanto ao software utilizado para a sua criação.

Lista de Figuras

2.1	Linha do tempo para sistemas hipertexto	6
2.2	Linha do tempo para a Web	13
2.3	As camadas da Web Semântica (adaptado de Berners-Lee [2000])	17
2.4	Linha do tempo para a Web Semântica	17
2.5	Modelo Wilson [1997] de <i>information seeking</i>	21
2.6	Um modelo de processo baseado no conjunto de características de Ellis [Wilson 1999]	22
2.7	Modelo “ <i>berrypicking</i> ” proposto por Bates [1989]	25
2.8	Modelo cognitivo proposto por Ingwersen [1996]	26
2.9	Modelo estratificado proposto por Saracevic [1996]	27
2.10	Modelo Spink de <i>information searching</i>	28
2.11	Modelo Navarro-Prieto, Scaife, e Rogers [1999] de <i>information searching</i> na Web para (a) usuários experientes e (b) usuários novatos	29
2.12	Modelo Hölscher e Strube de <i>information searching</i> na Web	31
4.1	Modelo de Navegação Exploratória (NAVE)	49
5.1	Arquitetura do ENS	56
5.2	Exemplo de mensagem de solicitação (<i>request</i>) enviada pelo cliente	58
5.3	Exemplo de mensagem de resposta (<i>response</i>) enviada pelo servidor	58
5.4	Arquitetura dos sistemas de apoio	61
5.5	A hierarquia de classes simplificada do MeGen	64
5.6	Evolução da interface do ENS	66
5.7	Interface gráfica do ENS	67
5.8	Interface da ferramenta Busca por palavra-chave no ENS	69
5.9	Interface da ferramenta Navegação por facetas no ENS	71
5.10	Interface da ferramenta Memória de trabalho no ENS	72
5.11	Interface da ferramenta Classificação no ENS	73
5.12	Interface da ferramenta Agrupamento no ENS	74
5.13	Interface da ferramenta Marcar resultados no ENS	74
5.14	Interface da ferramenta Realce de termos no ENS	75
5.15	Interface da Ferramenta Histórico no ENS	76
5.16	Cenário de uso do ENS com busca por palavra-chave e agrupamento	78
5.17	Cenário de uso do ENS com navegação por facetas e memória de trabalho	79
6.1	Exemplo de análise da interação de um usuário	86

6.2	Porcentagem de vezes que as ferramentas foram utilizadas	87
6.3	Porcentagem de vezes que os metadados foram utilizados	87
6.4	Interface do ENS (Sistema A)	89
6.5	Interface da busca por palavra-chave (Sistema B)	90
6.6	Gráficos <i>boxplot</i> dos dados dos grupos com relação ao tempo total para completar todas as tarefas	98
6.7	Gráficos <i>boxplot</i> dos dados dos grupos com relação ao tempo para com- pletar as tarefas	99
6.8	Gráficos <i>boxplot</i> dos dados dos grupos com relação às medidas de revo- cação (r) e precisão (P) do total de todas as tarefas	100

Lista de Tabelas

2.1	Modelo Kuhlthau [1991] de <i>information seeking</i>	24
2.2	Modelo Choo, Detlor, e Turnbull [2000] de <i>information searching</i> na Web	30
4.1	Matriz dos estágios e decisões em relação às categorias de técnicas de navegação	50
5.1	Um <i>application profile</i> para páginas wiki educacionais	65
5.2	Repositórios utilizados para avaliar o MeGen	65
5.3	Classificação das ferramentas da interface do ENS	68
6.1	Cinco tarefas em quatro tipos de cenários	83
6.2	Resultados estatísticos dos julgamentos com relação a relevância e utilidade	85
6.3	Projeto do experimento	93
6.4	Tarefas para quatro tipos de cenários	94
6.5	Resultados estatísticos dos grupos com relação ao tempo para completar as tarefas; os valores estão em minutos e segundos	98
6.6	Resultados estatísticos dos grupos com relação às medidas de revocação (r) e precisão (P) para as tarefas	99
6.7	Resultado estatístico dos grupos com relação ao julgamento subjetivo de relevância dos resultados das tarefas (questionário pós-tarefa)	100
6.8	Resultados da média harmônica (F) entre as medidas de revocação e precisão	101
6.9	Resumo das questões e dos resultados estatísticos do questionário de fechamento	101

Sumário

Resumo	vii
Abstract	ix
Publicações	xi
Abreviaturas e Siglas	xiii
Glossário	xv
Lista de Figuras	xvii
Lista de Tabelas	xix
1 Introdução	1
1.1 Contextualização	1
1.2 Contribuições	3
1.3 Organização	3
2 Revisão Bibliográfica	5
2.1 Considerações iniciais	5
2.2 Sistemas hipertexto, a Web e a Web Semântica	5
2.2.1 Sistemas hipertexto pré-Web	6
2.2.2 A Web	10
2.2.3 A Web Semântica	16
2.3 Modelos de <i>information seeking</i> e <i>information searching</i>	19
2.3.1 Conceitos	19
2.3.2 Modelos de <i>information seeking</i>	20
2.3.3 Modelos de <i>information searching</i>	24
2.4 Considerações finais	32
3 Metodologia	33
3.1 Considerações iniciais	33
3.2 Motivação e justificativa	33
3.3 Hipóteses	36
3.4 Objetivos	37
3.5 Caracterização da pesquisa	37
3.6 Método de pesquisa	38

3.7	Etapas do trabalho	38
3.7.1	Preparação da infra-estrutura	39
3.7.2	Desenvolvimento de um sistema	39
3.7.3	Avaliação do sistema com usuários	40
3.7.4	Descrição do modelo	40
3.8	Considerações finais	41
4	Um Modelo de Navegação Exploratória para a Web Semântica	43
4.1	Considerações iniciais	43
4.2	Conceitos subjacentes ao modelo	43
4.2.1	Níveis de atividade de <i>information seeking</i>	44
4.2.2	<i>Orienteering</i>	45
4.2.3	Busca exploratória	46
4.3	Desenvolvimento	47
4.4	Descrição	48
4.4.1	Estágios	48
4.4.2	Decisões	51
4.4.3	Entidades	52
4.5	Recomendações	52
4.6	Considerações finais	53
5	Sistema de Navegação Exploratória	55
5.1	Considerações iniciais	55
5.2	Arquitetura e implementação	55
5.2.1	<i>Navigation Engine</i>	59
5.2.2	Independência de domínio	60
5.3	Sistemas de apoio	61
5.3.1	<i>Crawler</i>	61
5.3.2	<i>Indexer</i>	62
5.3.3	<i>Metadata Generator</i>	62
5.4	Visão geral da Interface gráfica	65
5.5	Ferramentas	67
5.5.1	Busca por palavra-chave	69
5.5.2	Navegação por facetas	70
5.5.3	Memória de trabalho	70
5.5.4	Classificação	72
5.5.5	Agrupamento	72
5.5.6	Marcar resultados	73
5.5.7	Realce de termos	75
5.5.8	Histórico (reiniciar e desfazer)	76
5.6	Cenário de uso	77
5.7	Considerações finais	78

6 Avaliação	81
6.1 Considerações iniciais	81
6.2 Estudo de usabilidade	81
6.2.1 Objetivos	82
6.2.2 Metodologia	82
6.2.3 Resultados e observações	84
6.3 Estudo experimental	88
6.3.1 Definição do experimento	90
6.3.2 Planejamento do experimento	91
6.3.3 Operação do experimento	96
6.3.4 Análise e interpretação dos resultados	97
6.3.5 Resumo e conclusões	103
6.4 Considerações finais	103
7 Conclusão	105
7.1 Considerações iniciais	105
7.2 Contribuições	106
7.3 Trabalhos futuros	107
7.4 Considerações finais	109
Referências	111
A Material do Estudo de Usabilidade	123
B Material do Estudo Experimental	145

Introdução

1.1 Contextualização

A Web é a principal fonte de informação para muitas pessoas, ou pelo menos uma das fontes mais consideradas nas suas atividades que envolvem pesquisa por informação. A cada dia, aumenta a quantidade de informações disponíveis por meio da Web, desde notícias, opiniões, tecnologia, até produtos e publicidade. Nesse contexto, os sistemas *on-line* desenvolvidos para facilitar essa atividade se tornaram parte importante do dia-a-dia de muitos usuários de computadores.

A estratégia típica utilizada para encontrar informação na Web consiste em descrever o objetivo por meio de uma ou duas palavras-chave e submeter como uma consulta a uma ferramenta de busca para obter resultados, e depois examinar o conjunto de resultados recuperado e realizar comparações entre esses resultados. Essa estratégia é indicada quando as necessidades de informação do usuário são bem definidas, e as *search engines*, os catálogos ou bases de dados bibliográficas, e as bibliotecas digitais fornecem suporte adequado para os usuários [White et al. 2006a].

No entanto, as pessoas esperam (ou têm a expectativa) que a Web também seja útil em outras situações para as quais os sistemas e ferramentas disponíveis deveriam fornecer suporte mais adequado. São exemplos dessas situações quando: (a) um usuário pode não ter certeza do que ele está procurando até que as opções disponíveis sejam apresentadas; (b) o objetivo não pode ser expresso por palavras-chave ou (c) a terminologia exata utilizada nas páginas é desconhecida; e (d) uma quantidade de informação e contexto deve ser obtida ao longo das páginas (e não somente em uma página final). O tipo das atividades que os usuários se engajam para resolver essas situações está se tornando conhecida como busca exploratória [White et al. 2006a; Marchionini 2006]. Os pesquisadores de diversas comunidades têm trabalhado em tecnologias para fornecer suporte para essas atividades. No entanto, esforços são necessários com o objetivo de elaborar um conjunto de diretrizes para o projeto e desenvolvimento de sistemas para suporte a atividades de busca exploratória [White et al. 2007].

O usuário engajado em atividades de busca exploratória precisa de suporte diferente daquele oferecido, por exemplo, pelos sistemas tradicionais de recuperação de informação. As pessoas precisam de auxílio computacional para descobrir novas associações e tipos de conhecimento, resolver problemas complexos, ou desenvolver a compreensão da terminologia e da estrutura do espaço de informação. No contexto da Web, muitos desses requisitos podem ser alcançados por meio do uso sistemático da infra-estrutura da Web Semântica.

O objetivo da Web Semântica é desenvolver padrões e tecnologias para permitir descoberta (*discovery*), integração de dados, navegação, e automação de tarefas [Berners-Lee et al. 2001]. Nesse contexto, um conjunto de padrões tem sido especificado para a identificação de recursos na Web, assim como para a representação sintática, estrutural, semântica e lógica de informações referentes a esses recursos. No entanto, à medida que aumenta o número de recursos na Web (por exemplo, *wikis* e *blogs*) que disponibilizam os seus dados nesses padrões, novas abordagens para a navegação na Web Semântica são necessárias.

Os modelos de *information seeking* (por exemplo, os modelos propostos por Wilson [1997, 1999], Ellis [1989] e Kuhlthau [1991]) fornecem *frameworks* do comportamento das pessoas quando estão procurando por informação, e podem contribuir com informações sobre como e quando fornecer suporte para auxiliar as atividades de busca exploratória [Wilson e m. c. schraefel 2007]. Segundo Wilson [1999], esses modelos sugerem relacionamentos entre teorias que poderiam ser explorados ou testados. No entanto, a especificação de diretrizes para o projeto e desenvolvimento de sistemas com base nesses modelos pode não ser possível. Por exemplo, a partir do modelo de Wilson [1997, 1999] é possível deduzir (ou levantar a hipótese) de que um sistema deve ser projetado para reduzir o risco de falha do usuário e, portanto, aumentar seu sentimento pessoal de eficácia. Entretanto, a responsabilidade de decidir a maneira pela qual o risco é reduzido é do desenvolvedor do sistema.

O objetivo dos modelos de *information seeking* é fornecer uma visão geral do comportamento de pessoas quando estão procurando por informação, referente ao contexto mais amplo de atividades do que somente o uso de sistemas de informação. Os modelos de *information searching* são mais específicos, pois tentam descrever as atividades relacionadas com a interação entre uma pessoa, com necessidade de informação, e um sistema de informação. No entanto, os modelos de *information searching*, encontrados na literatura (por exemplo, os modelos propostos por Bates [1989], Ingwersen [1996], Saracevic [1996, 1997], Spink [1997], Navarro-Prieto et al. [1999], Choo et al. [2000] e Hölscher e Strube [2000]), também não fornecem diretrizes para o projeto e desenvolvimento de sistemas para navegação (especialmente para a Web Semântica). Portanto, existe a necessidade de se investigar modelos nesse contexto para servir de referência para desenvolvedores de sistemas.

A existência dessa lacuna impede o desenvolvimento de sistemas de navegação que poderiam ser extremamente úteis para a melhoria das atividades das pessoas

relacionadas com a pesquisa por informação na Web Semântica. Nesta tese é investigada a hipótese da viabilidade da criação de um modelo de navegação exploratória para auxiliar o projeto e desenvolvimento de sistemas para pesquisa de informação, no contexto da Web Semântica. O termo “navegação” é utilizado (ao invés de “busca”) para enfatizar a liberdade do usuário de decidir qual estratégia utilizar para satisfazer a sua necessidade de informação.

1.2 Contribuições

As principais contribuições da investigação conduzida neste trabalho podem ser assim resumidas:

- Um modelo de navegação exploratória, denominado de *Navigation and Exploration Model* (NAVE), para servir de referência para o projeto e desenvolvimento de sistemas de navegação exploratória para a infra-estrutura da Web Semântica.
- Um sistema de navegação exploratória, denominado de *Exploratory Navigation System* (ENS), desenvolvido com base no modelo NAVE, para auxiliar a pesquisa por informação na Web Semântica.
- A avaliação do sistema ENS, considerada um estudo de caso de avaliação de sistemas interativos para a pesquisa de informação.
- Um sistema para a geração automática de metadados, denominado de *Metadata Generator* (MeGen), que implementa diversos métodos automáticos de geração de metadados; desenvolvido como sistema de apoio ao ENS.

Estas contribuições são revisitadas na conclusão desta tese, destacando-se o desenvolvimento da pesquisa cuja estrutura é descrita na seção a seguir.

1.3 Organização

Neste capítulo, o contexto e a principal hipótese investigada no desenvolvimento deste trabalho foram apresentados. As contribuições principais que originaram dessa investigação foram descritas de maneira resumida. Os demais capítulos desta tese estão organizados da maneira descrita a seguir.

No Capítulo 2, a revisão bibliográfica é apresentada, dividida em duas partes. Na primeira parte são abordados os primeiros sistemas hipertexto, a Web, e a Web Semântica, com o objetivo de definir a infra-estrutura para a qual o trabalho foi direcionado e apresentar uma classificação das técnicas de auxílio à navegação. Na segunda parte, são descritos os modelos propostos para representar o comportamento das pessoas quando estão procurando por informação. A partir da revisão desses modelos, a existência de uma lacuna no conhecimento científico foi verificada, para a qual foi direcionada a contribuição do trabalho.

A finalidade do Capítulo 3 é apresentar a metodologia empregada no desenvolvimento dos trabalhos e pesquisas realizados. O problema prático que motivou este trabalho é apresentado, assim como o problema de pesquisa identificado. Nesse capítulo também são apresentadas as características da pesquisa realizada de acordo com classificações da literatura e os aspectos referentes ao método de pesquisa utilizado. As etapas do trabalho são apresentadas no final desse capítulo.

No Capítulo 4, um modelo de navegação exploratória, denominado *Navigation and Exploration Model* (NAVE), é proposto para auxiliar o projeto e desenvolvimento de sistemas de apoio a atividades de *information seeking* na Web Semântica. Nesse capítulo são descritos os conceitos subjacentes ao modelo, as características relacionadas com seu desenvolvimento e aplicação, como também a descrição dos elementos que compõem o modelo.

No Capítulo 5, é apresentado o sistema de navegação exploratória, denominado de *Exploratory Navigation System* (ENS). Nesse capítulo são apresentadas as características da arquitetura e implementação do sistema, as decisões de projeto com relação à navegação e independência de domínio, e os sistemas de apoio. Uma visão da interface gráfica também é apresentada e, em seguida, a descrição de cada uma das ferramentas de navegação integradas ao sistema ENS, as técnicas subjacentes às ferramentas e as extensões que foram desenvolvidas. No final desse capítulo é apresentado um cenário de uso do ENS para procurar por informação, visando ilustrar a abordagem de navegação.

No Capítulo 6, são descritos os dois estudos conduzidos para avaliar o sistema ENS, desenvolvido com base no modelo NAVE. Esses estudos consistem em um estudo de usabilidade, realizado com professores, e em um estudo experimental, realizado com alunos de graduação.

No Capítulo 7, as conclusões desta tese são apresentadas, principalmente as contribuições originadas da investigação realizada e as possibilidades no que se refere a extensões do modelo e a outros trabalhos futuros relacionados com os sistemas desenvolvidos.

O material utilizado no estudo de usabilidade e no estudo experimental pode ser encontrado nos Apêndices A e B, respectivamente.

Revisão Bibliográfica

2.1 Considerações iniciais

O objetivo da revisão bibliográfica apresentada neste capítulo é contextualizar o trabalho desenvolvido com relação a sistemas hipertexto/hipermídia, especialmente a Web e a Web Semântica, e também situar o trabalho no contexto dos modelos da literatura que tentam representar o comportamento das pessoas quando estão procurando por informação.

A revisão bibliográfica está dividida, portanto, em duas partes. Na primeira parte, apresentada na Seção 2.2, as tecnologias, descritas no contexto da Web e Web Semântica, definem a infra-estrutura para a qual o trabalho desenvolvido foi direcionado. Além disso, a classificação das técnicas subjacentes às ferramentas propostas para auxiliar a navegação, elaborada a partir dessa revisão e influenciada por outros trabalhos [Bates 1990; Hearst 1999; Teevan et al. 2004], é uma das partes do modelo proposto no Capítulo 4.

Na segunda parte da revisão bibliográfica, apresentada na Seção 2.3, são descritos os modelos que tentam representar o comportamento das pessoas quando estão procurando por informação, desde aqueles modelos mais gerais que consideram características pessoais e do ambiente, passando para modelos utilizados para expressar as atividades relacionadas com a interação entre uma pessoa e um sistema de informação, até os modelos que consideram um sistema de informação específico: a Web. A partir da revisão desses modelos, a existência de uma lacuna no conhecimento científico foi verificada, para a qual foi direcionada a contribuição do trabalho.

2.2 Sistemas hipertexto, a Web e a Web Semântica

O termo hipertexto é freqüentemente utilizado para denominar um conjunto de nós (unidade de informação) que podem conter referências para outros nós; essas referências são denominadas de ligações (ou links) [Nielsen 1995]. Dessa maneira, a estrutura de um hipertexto forma uma rede de nós e ligações. Os leitores se movem por essa rede em uma atividade denominada de navegação, ao invés de leitura, para

ênfatizar que eles podem determinar ativamente a seqüência de leitura dos nós [Nielsen 1995]. O termo hipermídia é utilizado para denominar um hipertexto que não está restrito a texto, mas pode incluir imagem, áudio, vídeo, etc. No entanto, alguns autores (por exemplo, Nielsen [1995]; Lowe e Hall [1999]) usam os termos hipertexto e hipermídia como sinônimos.

As subseções estão organizadas da maneira descrita na seqüência. Na Subseção 2.2.1, os primeiros sistemas hipertexto são descritos. Na Subseção 2.2.2, as tecnologias básicas da Web e os principais padrões para desenvolvimento de aplicações são descritos, além da apresentação de uma classificação das técnicas para suporte à navegação. Na Subseção 2.2.3, as camadas que formam a infra-estrutura da Web Semântica e os principais padrões relacionados são descritos.

2.2.1 Sistemas hipertexto pré-Web

O objetivo desta subseção é descrever, segundo uma perspectiva histórica, alguns dos primeiros sistemas com o conceito de hipertexto até o surgimento da Web. Os sistemas estão distribuídos em uma linha do tempo conforme representada na Figura 2.1 junto com outras referências importantes da história dos computadores pessoais. A revisão é baseada principalmente nos trabalhos de Conklin [1987], Nielsen [1990, 1995], Balasubramanian [1993], Gloor [1997] e de Müller-Prove [2002].

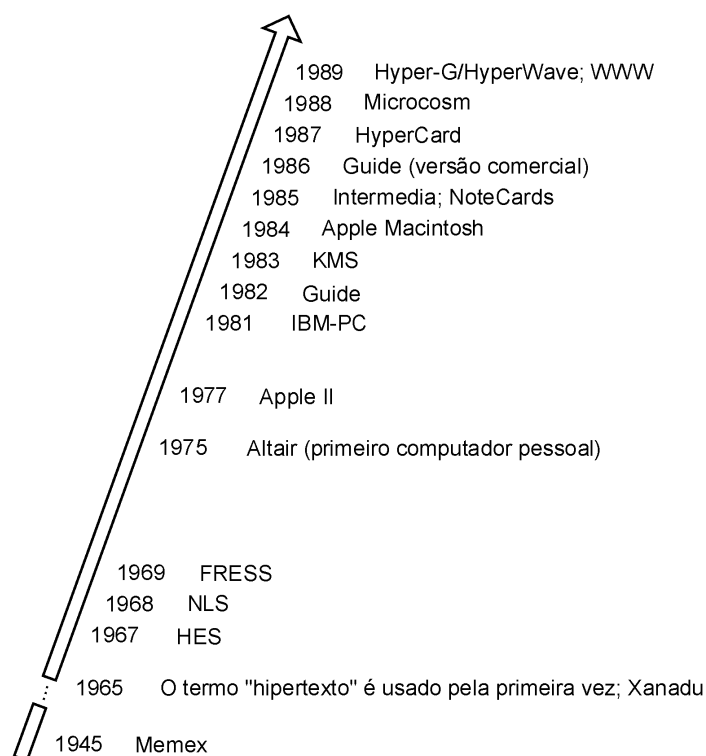


Figura 2.1: Linha do tempo para sistemas hipertexto

A primeira descrição da idéia de hipertexto é atribuída a Vannevar Bush que publicou, em 1945, o artigo clássico intitulado “*As We May Thing*” [Bush 1945]. Nesse

artigo, Bush descreve um equipamento hipotético chamado de Memex (*Memory Extension*) para auxiliar cientistas no seu trabalho diário. Bush era matemático, responsável por uma agência de pesquisa e desenvolvimento, e coordenava o trabalho de mais de seis mil cientistas. Uma das questões confrontadas por Bush era o volume crescente de publicações científicas que deviam ser armazenadas e organizadas de tal forma que permitisse a outros pesquisadores a utilização destas informações de maneira rápida e eficiente.

O Memex deveria permitir a definição de trilhas através dos artigos armazenados, de maneira a organizar seqüências que correspondessem a uma determinada cadeia de pensamento. Bush argumenta que os métodos clássicos de organização, como a classificação por ordem alfabética, são artificiais e não correspondem à maneira de pensar das pessoas. Uma abordagem mais natural seria contextualizar os artigos através de um conjunto de relacionamentos para mapear o estilo associativo da mente humana. As trilhas do Memex são consideradas a primeira concepção do conceito de hiperligações¹ (*hyperlinks*). No entanto, as trilhas do Memex estão mais próximas do conceito de roteiros guiados (*guided tours*), pois trilha é um conceito que está em um nível de abstração mais alto do que hiperligações [Reich et al. 1999].

O termo hipertexto (*hypertext*) foi usado pela primeira vez por Theodor (“Ted”) Nelson, que também propôs (em 1965) um sistema hipertexto denominado Xanadu. A sua idéia central é um “*docuverse*” (*document universe*), um universo de documentos no qual uma nova forma de literatura pode se desenvolver sem as limitações de um meio linear como o livro impresso. O sistema proposto era similar à Web, mas incluía a possibilidade de compor documentos a partir de partes de outros documentos dispersos pela rede sem a necessidade de copiar o conteúdo original (princípio chamado de transclusão), e também um método para gerenciar direitos autorais para proteger a propriedade intelectual de autores e colaboradores. Apesar dos investimentos privados realizados no projeto e dos protótipos com algumas características funcionando, Nelson não conseguiu produzir uma versão completamente funcional do sistema Xanadu.

O sistema *oN-Line System* (NLS) foi desenvolvido por Douglas Engelbart e sua equipe no *Stanford Research Institute* (SRI), e é considerado o primeiro sistema com funcionalidades de hipertexto que se tornou operacional. O objetivo do projeto *Augment*, iniciado em 1962, era construir uma nova ferramenta para “aumentar o intelecto humano”. O sistema NLS facilitava a criação de bibliotecas digitais, além do armazenamento e recuperação de documentos eletrônicos por meio de hipertexto. Em 1968 ocorreu a apresentação pública do NLS e de outras ferramentas novas desenvolvidas no contexto do projeto, por exemplo, *mouse*, ambiente de janelas, e-mail e diversas opções de processamento de texto.

O segundo sistema hipertexto, denominado *Hypertext Editing System* (HES), foi desenvolvido por Andries van Dam e Ted Nelson em 1967 na *Brown University*. Os

¹Os termos hiperligações (*hyperlinks*) e ligações (*links*) são utilizados neste trabalho como sinônimos.

principais objetivos do projeto foram a produção *on-line* de documentos e a exploração do conceito de hipertexto. Após ter assistido a apresentação do NLS, van Dam iniciou o desenvolvimento de um novo sistema, denominado *File Retrieval and Editing System* (FRESS), baseado no HES e com conceitos do NLS. No sistema FRESS as funcionalidades relacionadas às hiperligações apresentavam diversos aprimoramentos. Os links bidirecionais foram implementados pela primeira vez e utilizados para indicar um caminho para outro documento (que aparecia na mesma janela). Os links unidirecionais, por sua vez, eram usados especialmente para indicar a conexão com um único elemento (que aparecia em outra janela), por exemplo, no caso de anotação, definição ou de nota de rodapé. Os links eram armazenados em uma base de dados externa aos documentos. Além disso, os links também eram “tipados”, no sentido de que eles podiam ser rotulados com palavras-chave. Os links com o mesmo rótulo formavam um caminho que poderia ser seguido, de maneira semelhante às trilhas do Memex.

Diversos sistemas hipertexto foram criados e apresentados ao público na década de 1980 e os computadores pessoais (*personal computers*) começaram a ter o seu uso difundido. Em 1981, o IBM-PC foi lançado com o DOS, um sistema operacional baseado em texto. O Macintosh da Apple foi introduzido em 1984, considerado o primeiro computador com uma interface gráfica com o usuário, ou *Graphical User Interface* (GUI).

O *Knowledge Management System* (KMS) foi desenvolvido na empresa *Knowledge Systems*, fundada em 1981 por Donald McCracken e Robert Akscyn. O sistema KMS era baseado em uma unidade básica chamada *frame*. Um *frame* podia conter texto, gráficos, ou imagens, e podiam ser conectados a outros *frames* por meio de links. Existiam dois tipos de links no KMS: árvore, para representar relacionamentos hierárquicos (*hierarchical relationships*); e anotação, para representar relacionamentos de referência (*referential relationships*). O sistema KMS foi comercializado em 1983 e fornecia suporte para outras funcionalidades além de hipertexto, por exemplo, busca por palavra-chave e suporte para colaboração.

O sistema NoteCards foi desenvolvido no Xerox *Palo Alto Research Center* (PARC) em 1985 por Frank Halasz e sua equipe para fornecer um ambiente de hipertexto para texto e gráfico. O sistema utilizava a metáfora de cartões (*notecard*), isto é, cada cartão apresentando seu conteúdo (texto ou gráfico) em uma janela separada. As hiperligações eram referências para outros cartões, mas não faziam referência a uma determinada parte dentro do cartão. O título de um cartão era usado como um marcador de link para aquele cartão. Adicionalmente aos cartões de texto e gráfico, também existiam dois outros tipos de cartões pré-definidos: *filebox* e *browser*. Esses cartões forneciam suporte para o usuário classificar e categorizar os cartões de conteúdo. Um cartão *filebox* era uma maneira simples de coletar cartões que têm aspectos em comum. Um cartão *browser* mostrava um diagrama baseado na estrutura de links entre os cartões. Os cartões eram interconectados por links “tipados”, isto é,

um rótulo podia ser atribuído ao link para descrever o tipo de relacionamento entre dois cartões.

O sistema Guide foi desenvolvido por Peter Brown na *University of Kent*, e comercializado pela companhia britânica *Office Workstations Limited* em 1986 para Apple Macintosh e IBM-PC. O Guide foi considerado o sistema hipertexto comercial mais popular. O Guide foi provavelmente o primeiro sistema hipertexto a utilizar links sublinhados. Ele oferecia três formas diferentes de links (*reference jump*, *pop-up note* e *inline replacement*) que eram diferenciados visualmente através do formato que o ponteiro do *mouse* assumia para cada um deles. Os links do tipo *reference jump* eram referências para outros documentos, mas também podiam fazer referência a localizações dentro do mesmo documento. Os links *pop-up note* apresentavam uma pequena janela com conteúdo. Os links *inline replacement* inseriam um conteúdo na posição do marcador do link.

O sistema HyperCard foi criado por Bill Atkinson em 1987 para o Apple Macintosh. O HyperCard utilizava a metáfora de cartões. No entanto, os cartões eram organizados em pilhas, de maneira diferente do sistema NoteCards no qual uma janela correspondia a cada cartão. Os cartões tinham o mesmo tamanho para preencher a tela original do Macintosh. A pilha fornecia uma ordem padrão para os cartões e mesmo sem quaisquer links o usuário era capaz de navegar através dos cartões das pilhas. Um tipo de cartão especial, chamado de *home card*, funcionava como um marco (*landmark*) de orientação e era acessível a partir de qualquer outro cartão. Além disso, uma galeria com miniaturas (*thumbnails*) de todos os cartões podia ser apresentada para obter uma visão geral da pilha. O HyperCard foi bem sucedido devido à facilidade de uso, disponibilidade (uma cópia gratuita acompanhava o pacote do Macintosh) e a possibilidade de compartilhar cartões em comunidades *on-line*.

Os sistemas hipertextos citados anteriormente nesta seção, com a exceção do Xanadu, são todos considerados sistemas hipertexto fechados. Os hipertextos (ou hiperdocumentos) produzidos nesses sistemas são isolados uns dos outros e não é possível criar uma hiperligação de um hipertexto para outro na mesma máquina, mesmo quando o programa de autoria é o mesmo. Os sistemas Intermedia e Microcosm, descritos a seguir, tentaram preencher a lacuna entre programas e arquivos de formatos diferentes.

O sistema Intermedia foi desenvolvido na *Brown University* em 1985. O Intermedia era um ambiente integrado que fornecia aplicações para a criação de tipos diferentes de conteúdo (por exemplo, textos, gráficos, vídeos, modelos tridimensionais, e outros). A funcionalidade de hipertexto do sistema estava integrada dentro de cada aplicação de modo que a criação e navegação de links podiam ser intercaladas com a criação e edição de documentos. As aplicações compartilhavam o acesso a uma base de dados comum, na qual os links eram armazenados separadamente dos documentos. O Intermedia tinha suporte para o conceito de *webs*. Para o usuário, uma *web* representava uma rede de referências a partir da perspectiva de um contexto específico.

As *webs* podiam ser ativadas individualmente (uma de cada vez), para permitir a apresentação um conjunto diferente de conexões entre os documentos.

O projeto do Microcosm foi iniciado em 1988 na *University of Southampton*. Da mesma forma que o Intermedia, o sistema Microcosm também era baseado na separação entre os links e o conteúdo. Os links eram armazenados em bases de dados específicas, denominadas de *linkbases*, e podiam ser definidos sobre quaisquer tipos de documentos. O Microcosm tinha suporte para três tipos de links com relação a um determinado destino: o *specific link* partia de um objeto em um ponto específico de um documento; o *local link* partia de um objeto em qualquer ponto de um documento; e o *generic link* partia de um objeto em qualquer posição de qualquer documento. Além disso, também existiam funções para *dynamic linking*, isto é, links sem origem ou destino pré-determinado. O destino era calculado com base em algoritmos de recuperação de texto ou na proximidade entre o vocabulário de dois documentos. Com essa estrutura era possível, por exemplo, associar links com trechos genéricos de texto. Em qualquer parte que o texto aparecia, um link era inserido automaticamente. Dessa maneira, os documentos que eram importados pela primeira vez para o Microcosm podiam ter links imediatamente.

O projeto do Hyper-G foi iniciado em 1989 na *Graz University of Technology*. Hermann Maurer e sua equipe tinham como objetivo desenvolver uma versão em rede do Microcosm. A arquitetura do Hyper-G permitia a definição de estruturas de forma independente da distribuição dos arquivos entre os servidores. Os arquivos podiam pertencer a coleções, as quais podiam ser parte de outras coleções. Um tipo especial de coleção, chamado de *cluster*, era utilizado para agrupar um conjunto de arquivos como uma unidade lógica. O *cluster* podia ser utilizado para formar agregações, por exemplo, para fazer referência a vários documentos e objetos multimídia como uma única entidade. As coleções eram integradas ao modelo de dados e podiam também ser o objeto destino das hiperligações. Outra forma de agregação consistia nas seqüências de links, chamadas de *sequences*, que eram ativadas sobre uma série de documentos ou *clusters*. As *sequences* eram coleções com uma ordenação imposta sobre seus membros. A navegação através de *sequences* podia utilizar a ordem dos itens para oferecer comandos como “*next*”, “*previous*”, “*first*” e “*last*”. Quando o Hyper-G foi apresentado em 1995, chamado de HyperWave na versão comercial, o momento não era mais propício para um novo produto, pois a *World Wide Web*, operacional desde o início do projeto do Hyper-G, estava crescendo rapidamente.

2.2.2 A Web

A *World Wide Web* (WWW), ou simplesmente Web, nasceu no laboratório internacional do Conselho Europeu para Pesquisa Nuclear, do francês *Conseil Européen pour la Recherche Nucléaire* (CERN), localizado perto de Genebra, na Suíça. A comunidade científica do CERN é formada por milhares de pessoas. A gerência de uma organização desse porte é difícil e o intercâmbio eletrônico de documentos é ainda mais difícil

devido à existência de muitas incompatibilidades entre os diversos sistemas utilizados (hardware e software). Além disso, muitos dos cientistas não residem na cidade de Genebra, mas trabalham remotamente em muitas partes do mundo.

Berners-Lee e Robert Cailliau propuseram um sistema distribuído de hipertexto para resolver esses problemas [Berners-Lee 1989; Berners-Lee e Cailliau 1990]. Em 1994, os conceitos chave da Web foram apresentados por Berners-Lee et al. [1994]. Basicamente, os autores introduziram um sistema universal de endereço, um protocolo de rede para servidores e uma linguagem de marcação.

O sistema de endereço definido é o *Universal Resource Identifier*² (URI), utilizado para referenciar um recurso de maneira única e global. Uma sintaxe similar estava em processo de definição pela *Internet Engineering Task Force* (IETF), conhecida como *Uniform Resource Locator* (URL). A partir dessas iniciativas foi derivada a sintaxe genérica do *Uniform Resource Identifier* (URI) [Berners-Lee et al. 1998], que incorpora a sintaxe do URL como um subconjunto. Basicamente, um identificador URI é uma seqüência de caracteres para identificar um recurso físico ou abstrato. Nesse contexto, um recurso pode ser qualquer coisa que tem identidade, por exemplo, um documento eletrônico, uma imagem, um serviço e uma coleção de outros recursos. Além disso, os recursos não precisam ser recuperáveis por meio da Internet, por exemplo, pessoas, empresas e livros de uma biblioteca também podem ser considerados recursos.

O protocolo de rede definido para os servidores é o *Hypertext Transfer Protocol* (HTTP), que consiste em um protocolo de transferência de objetos construído sobre os protocolos utilizados na Internet para transporte: *Transport Control Protocol / Internet Protocol* (TCP/IP). O primeiro servidor da Web se tornou operacional no final de 1990. O HTTP/1.1 [Fielding et al. 1999] fornece um conjunto de métodos para a transferência de objetos entre cliente e servidor (OPTIONS, GET, HEAD, POST, PUT, DELETE, TRACE, CONNECT), mas os métodos GET e POST são utilizados com mais frequência para, respectivamente, requisitar dados do servidor e enviar dados para o servidor.

As páginas da Web são codificadas com a linguagem de marcação *HyperText Markup Language* (HTML). A HTML é uma aplicação simples da *Standard Generalized Markup Language* (SGML) [ISO 1986], isto é, uma linguagem em conformidade com um *Document Type Definition* (DTD) da SGML. A idéia de utilizar caracteres de texto simples (*plain text*) facilitou a edição de arquivos HTML com o software existente e a transferência desses arquivos entre todas as plataformas. Além de descrever a estrutura e a apresentação de documentos hipertexto, a HTML [W3C 1999] também tem suporte para vários outros recursos, por exemplo, formulários, scripts, objetos multimídia e aplicações (*applets*³). Para apresentar as páginas codificadas em HTML é utilizado um programa cliente, denominado de navegador (*browser*). O primeiro

²Denominação original do artigo de Berners-Lee et al. [1994].

³Um applet é um programa na linguagem Java que pode ser incorporado em uma página e executado pelo navegador.

navegador, *WorldWideWeb/Nexus*⁴, foi desenvolvido por Tim Berners-Lee em 1990.

As tecnologias básicas da Web (HTTP e HTML) permitem a criação de repositórios de páginas estáticas. O servidor envia ao navegador as páginas previamente construídas e com conteúdo estático (armazenadas em arquivos) de acordo com as solicitações do usuário (por exemplo, a interação do usuário com os links do hipertexto). No entanto, o desenvolvimento da tecnologia conhecida como *Common Gateway Interface* (CGI)⁵ permitiu a evolução da estrutura estática para aplicações mais dinâmicas. O servidor é capaz de gerar páginas sob demanda (dinamicamente), por exemplo, de acordo com os dados fornecidos pelo usuário. Depois foram criadas linguagens específicas para o desenvolvimento de aplicações, como *PHP: Hypertext Preprocessor* (PHP)⁶ e *JavaServer Pages* (JSP)⁷, e JavaScript (padrão ECMAScript [Ecma 1999]). Assim, a interação na Web pode ser proporcionada através de tecnologias tanto no lado cliente quanto no lado servidor.

Padrões Web

O *World Wide Web Consortium* (W3C)⁸ é um consórcio internacional que tem como missão levar a Web ao seu potencial máximo por meio do desenvolvimento de protocolos e recomendações, com o objetivo de promover a sua evolução e assegurar a sua interoperabilidade. Desde a sua fundação por Tim Berners-Lee em 1994, o W3C tem publicado padrões e diretrizes que são utilizadas para a criação e a interpretação de conteúdos baseados na Web. Os principais padrões estão distribuídos em uma linha do tempo conforme representada na Figura 2.2 junto com outras referências importantes relacionadas com a Web.

No contexto de desenvolvimento de aplicações para a Web, os principais padrões relacionados podem ser assim organizados [Johansson 2006; Zeldman 2006]:

- *Linguagens de estrutura*. A estrutura consiste nas partes do conteúdo de um documento e da semântica associada a estas partes.
- *HTML (HyperText Markup Language) 4.01* [W3C 1999]: Essa recomendação desaprova o uso de um conjunto de elementos e atributos que codificam principalmente efeitos de apresentação, adiciona um atributo (para especificar classe) em todos os elementos para poder classificar os elementos em categorias (com o efeito de criar novos elementos) e acrescenta suporte e melhorias para outros recursos, por exemplo, internacionalização, acessibilidade, tabelas e formulários, composição de objetos (multimídia e *applets*), *scripting*.

⁴www.w3.org/People/Berners-Lee/WorldWideWeb.html

⁵As especificações da CGI são mantidas pelo NCSA Software Development Group, <http://hoohoo.ncsa.uiuc.edu/cgi/>.

⁶<http://www.php.net/>

⁷<http://java.sun.com/products/jsp/>

⁸<http://www.w3.org/>

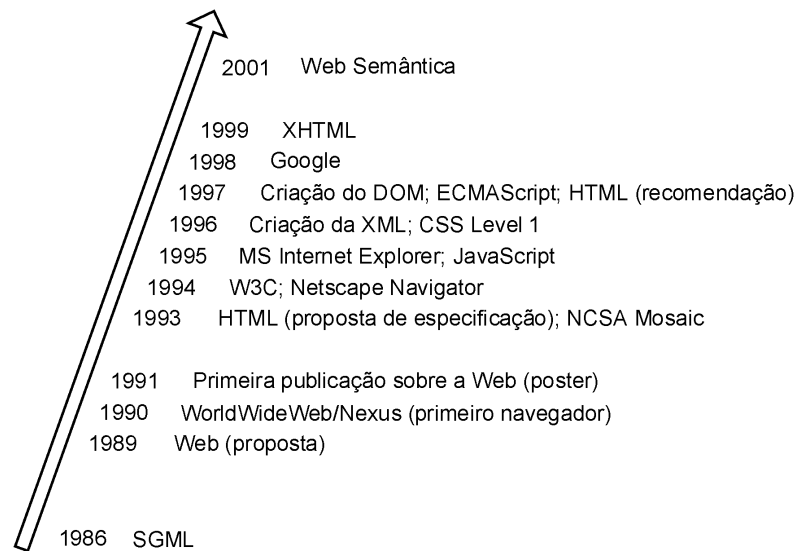


Figura 2.2: Linha do tempo para a Web

- XML (*Extensible Markup Language*) [W3C 2006a]: Essa recomendação define a sintaxe de um subconjunto da SGML para uso na Web. A linguagem XML é uma aplicação da SGML que descreve uma classe de objetos de dados chamados de documentos XML.
- XHTML (*Extensible Hypertext Markup Language*) 1.0 [W3C 2002]: Essa recomendação define uma reformulação da HTML como uma aplicação da XML.
- *Linguagens de apresentação*. A apresentação é o estilo atribuído ao conteúdo de um documento.
 - CSS (*Cascading Style Sheets*) Level 1 [W3C 1999]: Essa recomendação define as propriedades para fontes, margens, cores, etc., necessárias para quase todas as aplicações.
 - CSS Level 2.1 [W3C 2007]: Essa recomendação contém o nível 1 e acrescenta elementos para especificar posições absolutas, numeração automática, quebras de página, alinhamento de texto, entre outros.
 - CSS Level 3 (em desenvolvimento)⁹: Essa recomendação contém o nível 2 e acrescenta novos seletores, bordas e fundos, texto vertical e outros recursos avançados.
- *Modelos de Objeto*. O modelo de objeto consiste em uma interface independente de linguagem e plataforma que define o acesso e a modificação dinâmica do conteúdo, da estrutura e do estilo de documentos.
 - DOM (*Document Object Model*) Level 1 [W3C 2000a]: Essa recomendação define modelos de objeto para documentos estruturados nas linguagens XML e HTML.

⁹<http://www.w3.org/Style/CSS/current-work>

- DOM Level 2 [W3C 2000b]: Essa recomendação adiciona modelos de objeto para visões diferentes associadas a um documento, folhas de estilo, sistema genérico de eventos, e também define interfaces especializadas para percorrer a estrutura do documento e identificar e manipular um trecho de um documento.
- *Linguagens de script*. A linguagem de script é uma linguagem de programação utilizada para manipular, aperfeiçoar e automatizar as facilidades de um sistema existente.
- ECMAScript 262 [Ecma 1999]: Esse padrão define a linguagem de script ECMAScript, baseada principalmente nas linguagens JavaScript (Netscape) e JScript (Microsoft). Em geral, um navegador fornece um ambiente ECMAScript para permitir processamento no lado cliente.

Navegação na Web

A idéia básica da navegação na Web é percorrer um caminho através do espaço de páginas por meio da seleção (voluntária do usuário) dos links disponíveis na página carregada no navegador. Na realidade, o usuário cria o seu caminho a partir de uma página inicial de sua preferência, e para a qual ele sempre pode retornar se ficar perdido (com referência ao problema conhecido como desorientação do usuário ou “*lost in hyperspace*” [Conklin 1987]). O navegador fornece as ferramentas básicas de navegação, tipicamente, um botão para retornar para uma página predefinida pelo usuário (*home*), assim como botões para retroceder (*back*) ou avançar (*forward*) no caminho que o usuário estabelece percorrendo os links. O caminho percorrido pelo usuário é freqüentemente denominado de histórico de navegação.

Além da opção de percorrer links e dos botões mencionados, outras ferramentas também são geralmente fornecidas pelos navegadores atuais, como o Mozilla Firefox e o Internet Explorer, e podem ser assim denominadas: menus dos botões para retroceder e avançar; mecanismo para gerenciamento de uma lista de favoritos (*bookmarks*), mecanismo de visualização do histórico de navegação (*history*); mecanismo para digitar, selecionar e completar automaticamente o campo de endereço (URL); e mecanismo de pesquisa (*search toolbar*). Em geral, o navegador também fornece recursos para orientar o usuário durante a navegação, por exemplo, a apresentação do título da página carregada e do endereço (URL) de destino de um link.

O usuário também pode utilizar outras ferramentas para auxiliar a sua navegação na Web, por exemplo, aquelas oferecidas por meio de um programa externo (*plug-in*) ou de uma extensão (*add-on*) para o navegador. Adicionalmente, uma página carregada no navegador também pode conter elementos projetados para auxiliar o usuário durante a navegação (por exemplo, barras de navegação e mapas de site), além de interfaces para ferramentas desenvolvidas como aplicações no lado cliente (por exemplo, um programa em linguagem de script) ou como aplicações cliente-servidor (por exemplo, uma ferramenta de busca).

Existem muitas ferramentas propostas na literatura para auxiliar a navegação na Web e enumerar todas é uma tarefa difícil. Uma das dificuldades da área consiste na distribuição dos resultados das pesquisas sobre um número grande de publicações diferentes. No entanto, algumas das técnicas subjacentes às ferramentas podem ser categorizadas segundo a sua função no processo de navegação. Através das categorias e exemplos é possível ter uma noção aproximada da abrangência e interdisciplinaridade da área. A classificação apresentada na seqüência foi elaborada a partir de revisão da literatura, e também foi influenciada pelos trabalhos de Bates [1990], Hearst [1999] e de Teevan et al. [2004]. As categorias são as seguintes:

- *Busca ou recuperação.* As técnicas desta categoria têm a função de auxiliar o usuário a realizar efetivamente a navegação, isto é, executar passos para obter (representações de) objetos de informação. Esta categoria pode ser dividida nas seguintes subcategorias:
 - *Início.* As técnicas desta categoria são aquelas para auxiliar o usuário a entender o conteúdo de uma coleção (ou fonte) e decidir como começar o processo de navegação. São exemplos de técnicas para o início da navegação: categorias hierárquicas (Yahoo! Directory¹⁰ e Open Directory Project¹¹); visão geral textual ou gráfica [Mukherjea 1999; Nielsen 1999; Danielson 2002]; e roteiro guiado (*guided tour*) [Furuta et al. 1997; Jühne et al. 1998; Sandvad et al. 2001].
 - *Especificação e seleção.* As técnicas desta categoria são aquelas para auxiliar o usuário a especificar em algum grau de abstração os (tipos de) objetos desejados e obter um conjunto correspondente àquela especificação. São exemplos de técnicas de especificação e seleção: especificação de consulta por meio de operadores booleanos com palavras-chave e/ou metadados [Baeza-Yates e Ribeiro-Neto 1999]; previsão de consulta (*query preview*) [Plaisant et al. 1999; Tanin et al. 2007]; navegação por facetas de metadados [Sinha e Karger 2005; Hearst 2006].
- *Análise.* As técnicas desta categoria têm a função de auxiliar o usuário durante a atividade de análise dos objetos de informação encontrados como resultado de algum passo de navegação realizado (por exemplo, por meio de alguma técnica da categoria de busca e recuperação). Esta categoria pode ser dividida nas seguintes subcategorias:
 - *Visualização de contexto.* As técnicas desta categoria visam auxiliar o usuário a entender o(s) objeto(s) de informação por meio da apresentação de informações adicionais relacionadas. São exemplos de técnicas para visualização de contexto: apresentação dos metadados que descrevem os obje-

¹⁰<http://dir.yahoo.com/>

¹¹<http://www.dmoz.org/>

tos (*details-on-demand*) [Shneiderman 1996]; realce dos termos da consulta (*highlighting of query terms*) [Marchionini 1995; Hearst 1999; Nielsen 2005].

- *Organização*. As técnicas desta categoria visam auxiliar o usuário a visualizar os diferentes relacionamentos entre os objetos de informação. São exemplos de técnicas para organização: classificação (*sorting*) [Nielsen e Loranger 2006]; agrupamento (*clustering*) [Hearst 2006].
- *Controle do processo*. As técnicas desta categoria têm a função de auxiliar o usuário a manter o controle do processo de navegação à medida que são realizados passos de navegação. São exemplos de técnicas: *breadcrumb navigation* [Ahmed e Blustein 2005; Nielsen 2007]; armazenamento e reuso do estado anterior entre diversas sessões [Cousins et al. 1997; Jhaveri e Rähkä 2005]; manutenção de um espaço de trabalho para armazenar uma lista de itens (*working list*) [Jhaveri e Rähkä 2005]; anotação de objetos de informação [Vasudevan e Palmer 1999].

2.2.3 A Web Semântica

A Web Semântica é um esforço colaborativo liderado pelo W3C, com a participação de pesquisadores e da indústria, para melhorar as potencialidades da Web através da criação de padrões e ferramentas que permitam atribuir significado ao conteúdo das páginas para que usuários e programas possam trabalhar de maneira cooperativa [Berners-Lee et al. 2001; W3C 2001].

Nesse contexto, um conjunto de padrões tem sido especificado para a identificação de recursos na Web, assim como para a representação sintática, estrutural, semântica e lógica de informações referentes a esses recursos. Esse conjunto de padrões forma as diversas camadas em que se divide a arquitetura da Web Semântica, conforme ilustrada na Figura 2.3. Adicionalmente, esses padrões também estão distribuídos em uma linha do tempo conforme representada na Figura 2.4.

A camada denominada de *Unicode / URI* fornece a interoperabilidade em relação à codificação de caracteres e ao endereçamento e nomeação de recursos da Web Semântica. O Unicode [Dürst e Freytag 2007] é um padrão de codificação para fornecer uma representação numérica universal e sem ambigüidade para cada caractere de maneira independente da plataforma de software e do idioma. O URI [Berners-Lee et al. 1998] é um padrão para identificar um recurso físico ou abstrato de maneira única e global. Um identificador URL é um caso específico de URI, formado pela concatenação de seqüências de caracteres para identificar o protocolo de acesso ao recurso, o endereço da máquina na qual o recurso pode ser encontrado, e o próprio recurso em questão.

A camada denominada de *XML / Namespaces / XML Schema* na Figura 2.3 fornece a interoperabilidade em relação à sintaxe de descrição de recursos da Web Semântica. A XML [W3C 2006a] é uma linguagem para a representação sintática de recursos de

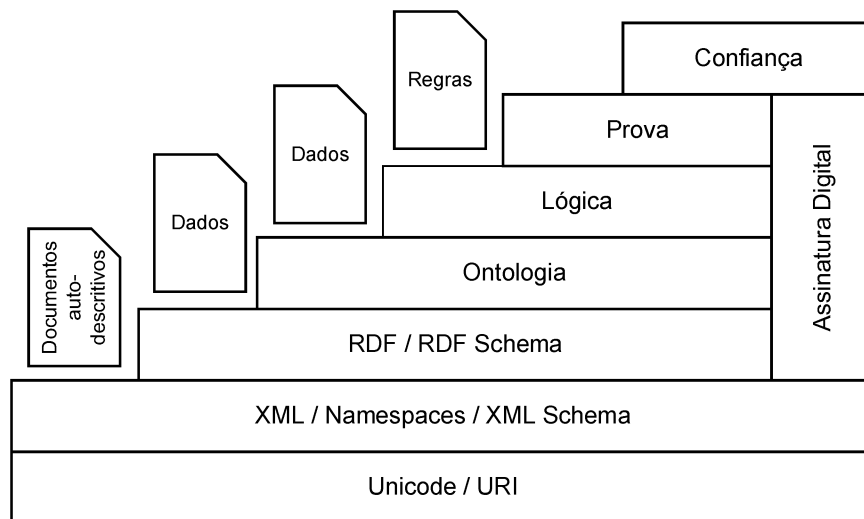


Figura 2.3: As camadas da Web Semântica (adaptado de Berners-Lee [2000])

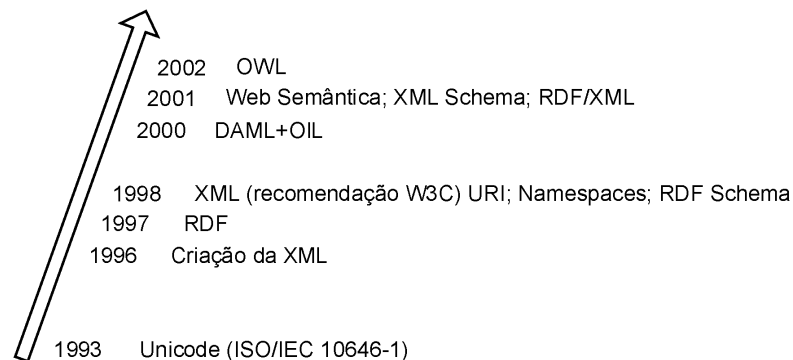


Figura 2.4: Linha do tempo para a Web Semântica

maneira independente de plataforma. Os documentos que têm a sua estrutura e conteúdo representados na linguagem XML são denominados de documentos XML. Os documentos XML utilizados no intercâmbio entre aplicações podem ser submetidos a um processo de validação, no qual é realizada a verificação se a sua estrutura e conteúdo seguem as regras de formação da gramática definida para a sua classe de documentos. A XML Schema [W3C 2004f] é uma linguagem de definição para descrever uma gramática (ou esquema) para uma classe de documentos XML. A linguagem XML Schema fornece elementos para descrever a estrutura [W3C 2004g] e restringir o conteúdo [W3C 2004h] de documentos XML. Os documentos auto-descritivos da Figura 2.3 são documentos XML; e eles são válidos se estão em conformidade com a sua respectiva gramática especificada na linguagem XML Schema.

Os espaços de nomes (*namespaces*) [W3C 2006c] fornecem um método para qualificar os nomes de elementos e atributos, utilizados nos documentos XML, através da associação destes nomes com espaços de nomes identificados por *referências de URI* (URIref)¹². Os espaços de nomes são úteis para distinguir entre dois elementos defi-

¹²Uma referência de URI (ou URIref) é uma URI com um identificador de fragmento opcional no final. Por exemplo, a referência de URI “http://www.example.org/index.html#section2” consiste na URI

nidos com um mesmo nome, mas que pertencem a esquemas diferentes. Além disso, um documento pode associar elementos previamente definidos à sua estrutura, desde que utilize referências aos esquemas que definem esses elementos.

A camada denominada de *RDF / RDF Schema* na Figura 2.3 fornece um *framework* para representar informação (metadados) sobre recursos. As principais especificações do *Resource Description Framework* (RDF) [W3C 2004a] abrangem um modelo de dados (para expressar declarações sobre os recursos), uma sintaxe baseada na linguagem *Extensible Markup Language* (XML) (para o intercâmbio das declarações) e uma linguagem de definição de esquemas para vocabulários.

O RDF fornece um modelo de dados fundamentado na idéia de expressar declarações simples sobre recursos; cada declaração consiste em uma tripla (sujeito, predicado e objeto). Por exemplo, na declaração “a data de criação da página <http://www.example.org/index.html> é 16/08/1999”, o sujeito é “<http://www.example.org/index.html>”, o predicado é “data de criação” e o objeto é “16/08/1999”. Um conjunto de triplas (ou declarações) é chamado de grafo RDF [W3C 2004c], que pode ser ilustrado como um diagrama de nós e arcos orientados, no qual cada tripla é representada como uma ligação nó-arco-nó. O RDF fornece uma sintaxe baseada na linguagem XML, denominada de RDF/XML [W3C 2004d], para realizar o intercâmbio desses grafos.

Além do modelo e da sintaxe, o RDF também fornece uma linguagem, denominada RDF Schema [W3C 2004b], para a definição de esquemas para os vocabulários (termos) utilizados nas declarações. A RDF Schema estende a especificação básica do RDF para permitir a definição de vocabulários. Em outras palavras, a RDF Schema é uma linguagem mínima para a representação de ontologias simples. Basicamente, essa linguagem fornece o suporte necessário para descrever classes e propriedades, e também para indicar quais propriedades são utilizadas para a descrição de uma classe.

A camada denominada de *Ontologia* na Figura 2.3 fornece suporte para a evolução de vocabulários e para processar e integrar a informação existente sem problemas de indefinição ou conflito de terminologia. A linguagem RDF Schema permite a construção de ontologias com expressividade e inferência limitadas, pois fornece um conjunto básico de elementos para a modelagem, e poucos desses elementos podem ser utilizados para inferência. A *Web Ontology Language* (OWL) estende o vocabulário da RDF Schema para incluir elementos com maior poder com relação à expressividade e inferência. Além disso, a linguagem OWL fornece três módulos (ou dialetos), OWL Lite, OWL DL e OWL Full, para permitir o uso da linguagem por aplicações com diferentes requisitos de expressividade e inferência. O desenvolvedor pode escolher o módulo OWL adequado, de acordo com os requisitos da sua aplicação.

A camada denominada de *Lógica* na Figura 2.3 fornece suporte para a descrição

“<http://www.example.org/index.html>” e, separado pelo caractere “#”, o identificador de fragmento “section2”.

de regras para expressar relações sobre os conceitos de uma ontologia, as quais não podem ser expressas com a linguagem de ontologia utilizada. As linguagens *Rule Markup Language* (RuleML)¹³ e *Semantic Web Rule Language* (SWRL)¹⁴ são exemplos de linguagens propostas para a descrição de regras para a Web Semântica. Nesse sentido, o W3C iniciou o trabalho sobre o *Rule Interchange Format* (RIF)¹⁵ para fornecer suporte para o intercâmbio das diversas tecnologias baseadas em regras.

As camadas denominadas de *Prova* e de *Confiança* na Figura 2.3 fornecem o suporte para a execução das regras, além de avaliar a correção e a confiabilidade dessa execução. Essas camadas ainda estão em desenvolvimento e dependem da maturidade das camadas inferiores.

2.3 Modelos de *information seeking* e *information searching*

Nesta seção são descritos alguns dos modelos mais conhecidos na literatura que tentam representar o comportamento das pessoas em atividades de *information seeking* e *information searching*¹⁶. Os modelos foram selecionados de acordo com a sua influência na literatura, isto é, os modelos considerados são aqueles mais frequentemente citados e/ou referenciados em artigos de revisão ou levantamento (*survey*). Além disso, os modelos específicos para o contexto da Web encontrados na literatura também são descritos.

Antes de descrever os modelos, alguns dos principais conceitos envolvidos são apresentados (Subseção 2.3.1). Em seguida, os modelos são descritos, organizados em modelos de *information seeking* (Subseção 2.3.2) e modelos de *information searching* (Subseção 2.3.3).

2.3.1 Conceitos

Existem muitas definições para o conceito de informação [Case 2002; Capurro e Hjørland 2003; Ingwersen e Järvelin 2005]. Entretanto, o conceito de informação não respeita os limites das áreas do conhecimento e não encontra consenso sobre qual deveria ser sua definição, que varia de uma área para outra e em relação a diferentes contextos [Case 2002; Capurro e Hjørland 2003]. Portanto, o objetivo desta seção não é realizar uma revisão ampla da literatura, mas somente apresentar uma definição para auxiliar na compreensão deste e de outros conceitos diretamente relacionados com o contexto deste trabalho. Uma revisão mais completa das diversas abordagens existentes para o conceito de informação pode ser encontrada no trabalho de Bates [2005].

De acordo com Ingwersen e Järvelin [2005], o conceito de informação tem de sa-

¹³<http://www.ruleml.org/>

¹⁴<http://www.w3.org/Submission/SWRL/>

¹⁵<http://www.w3.org/2005/rules/>

¹⁶Os termos *information seeking* e *information searching* são mantidos no idioma original (inglês) para facilitar a identificação com a literatura científica em geral, já que em português tanto a palavra *seek* quanto a palavra *search* são traduzidas como busca.

tisfazer duas condições simultaneamente:

1. Informação sendo algo que é o resultado de uma transformação das estruturas do conhecimento de um gerador.
2. Informação sendo algo que, quando percebido, afeta e transforma o estado do conhecimento do receptor.

Essa visão cognitiva permite distinguir entre a informação como objeto ou coisa (por exemplo, documentos) e a informação como dependente da interpretação de um agente cognitivo. Se somente a primeira condição é satisfeita, a informação pode ser considerada informação potencial, isto é, dados ou entidades similares armazenadas em fontes de informação, que são de valor potencial para os receptores (pessoas ou máquinas). Se somente a segunda condição é satisfeita, a informação pode ser considerada dados sensoriais que são percebidos, ou sinais de outra natureza.

De maneira semelhante ao conceito de informação, também existem muitas concepções para o conceito de necessidade de informação (*information need*) [Case 2002]. De acordo com Ingwersen e Järvelin [2005], necessidade de informação significa uma lacuna identificada de maneira consciente por uma pessoa no conhecimento disponível a ela. A necessidade de informação pode motivar uma pessoa a realizar atividades de *information seeking* para adquirir informação nova.

O conceito de *information seeking* se refere ao contexto mais amplo de uma necessidade de informação, enquanto que *information searching* se refere à interação com um sistema para uma atividade mais específica. No entanto, essa distinção entre *information seeking* e *information searching* pode ser obscura [Marchionini 1995; Wilson 1999; Spink et al. 2002]. Ingwersen e Järvelin [2005], por exemplo, não fazem distinção entre esses conceitos.

Information seeking considera as questões do ambiente, por exemplo, a profissão da pessoa ou a estrutura da organização na qual acontecem as atividades, assim como as razões subjacentes para a tarefa. Além disso, também pode incluir a aquisição de informação a partir de fontes não eletrônicas, como aquelas provenientes dos colegas de trabalho e de anotações em papel.

Information searching é utilizado para expressar atividades relacionadas com a interação entre uma pessoa, com uma necessidade de informação, e um sistema de informação. Neste caso, o interesse está na interação do usuário com o sistema.

2.3.2 Modelos de *information seeking*

Nesta subseção, os principais modelos de *information seeking* encontrados na literatura são descritos. Esses modelos tentam representar uma visão global do comportamento das pessoas quando estão procurando por informação, e consideram o uso de fontes de informação formais e informais.

Modelo Wilson

O modelo Wilson [1997, 1999] é uma revisão do seu próprio modelo proposto anteriormente (1981) sobre o comportamento humano em relação a fontes de informação e aos inter-relacionamentos entre conceitos utilizados no campo da ciência da informação. O seu modelo mais atual de *information seeking*, ilustrado na Figura 2.5, é centrado em como uma pessoa dentro de um contexto trata a sua necessidade de informação.

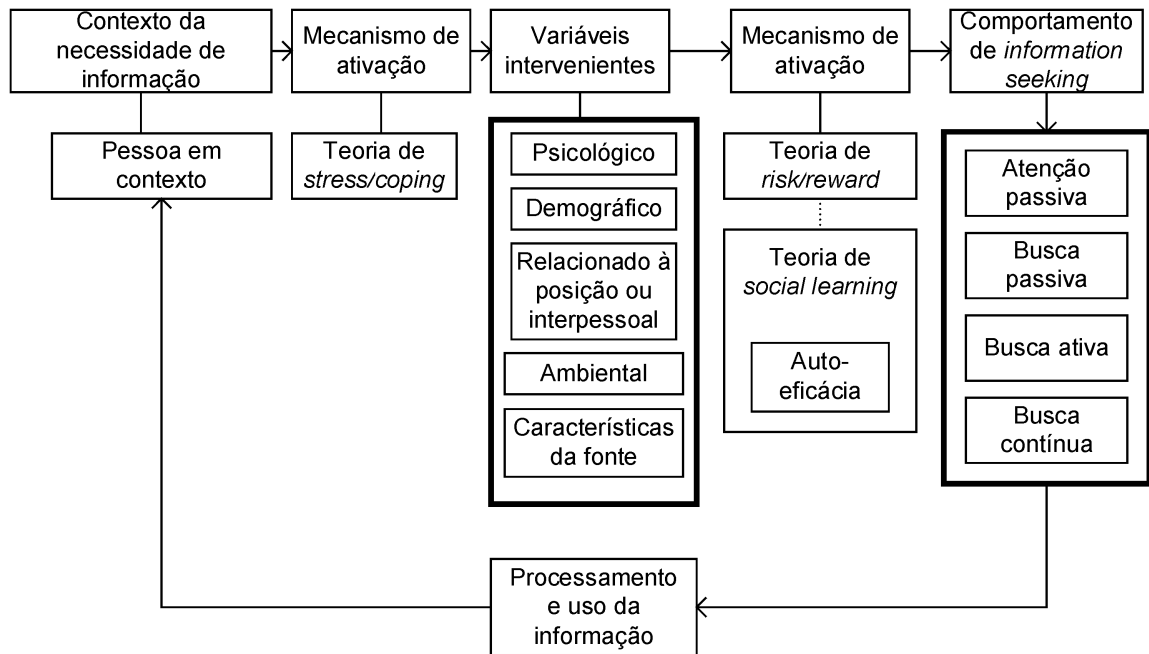


Figura 2.5: Modelo Wilson [1997] de *information seeking*

Wilson introduz algumas variáveis que podem interferir no processo, além de considerar a pessoa em contexto como o foco da necessidade de informação. Essas variáveis (*intervening variables*) foram identificadas em uma revisão extensiva da pesquisa de outros campos diferentes do campo da ciência da informação, por exemplo, da psicologia. Wilson sugere que as características pessoais, como o nível educacional, influenciam a escolha de fontes de informação. Além disso, no modelo é considerado como as teorias de “*stress/coping*”, “*risk/reward*” e “*social learning*”, podem explicar a continuação ou não do processo de *information seeking*.

O modelo Wilson é um modelo muito geral para ser utilizado como referência no projeto e desenvolvimento de sistemas para facilitar as atividades de *information searching*, pois muitas informações oriundas de disciplinas diferentes estão contidas no modelo. No entanto, o modelo pode ser utilizado para identificar abordagens e áreas problemáticas no contexto de *information seeking*.

Modelo Ellis

Ellis [1989] identificou oito características principais (ou *features*) do comporta-

mento de *information seeking*, as quais descrevem os padrões básicos de comportamento de cientistas sociais, físicos e químicos, engenheiros e de cientistas no ambiente de uma empresa industrial. Esse conjunto de características é baseado em uma sequência de estudos empíricos. As características principais identificadas são denominadas e definidas da seguinte maneira [Ellis 1989; Ellis e Haugan 1997]:

- Iniciar (*starting*): realizar atividades para iniciar a pesquisa por informação.
- Encadear (*chaining*): seguir citações e notas do material conhecido ou do material obtido por meio de índices.
- Explorar (*browsing*): pesquisar de maneira direcionada e estruturada as fontes primárias e secundárias da área de interesse.
- Diferenciar (*differentiating*): utilizar as diferenças conhecidas entre as fontes de informação como uma maneira de filtrar a natureza, quantidade e a qualidade da informação obtida.
- Monitorar (*monitoring*): verificar periodicamente os desenvolvimentos do campo (ou área) através de determinadas fontes de informação formais e informais.
- Extrair (*extracting*): identificar sistematicamente o material relevante em uma fonte de informação.
- Verificar (*verifying*): verificar a precisão da informação.
- Finalizar (*ending*): realizar atividades para finalizar a pesquisa por informação.

Ellis não estabelece relacionamentos entre essas características e admite que a inter-relação detalhada das características, em qualquer padrão individual de *information seeking*, depende das circunstâncias das atividades (de *information seeking*) da pessoa em relação a um determinado momento. Entretanto, Wilson [1999] propõe relacionamentos entre essas categorias de atividades e sugere uma representação gráfica para o modelo, mostrado na Figura 2.6.

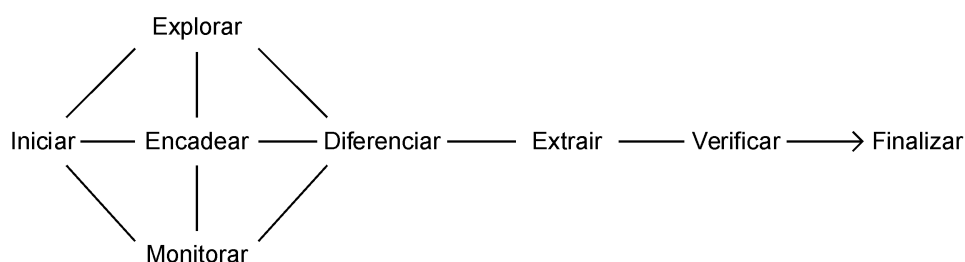


Figura 2.6: Um modelo de processo baseado no conjunto de características de Ellis [Wilson 1999]

Ellis [1989] discute o quanto essas categorias estão disponíveis em sistemas de recuperação de informação existentes (na época) e propõe o uso dessas categorias

como um *framework* para organizar recomendações para o projeto e desenvolvimento desses sistemas. Entretanto, as categorias não fornecem diretamente quaisquer recomendações ou especificações que poderiam servir de referência para o desenvolvimento de sistemas. Ao contrário, elas representam tipos de atividades que os usuários poderiam querer realizar por meio de sistemas.

Modelo Kuhlthau

Kuhlthau [1991] mostrou, por meio de uma série de estudos empíricos, que as tarefas de aprendizagem e a resolução de problemas por estudantes e usuários de bibliotecas consistem em vários estágios. No modelo obtido a partir desses estudos é considerada a informação com relação às experiências emocionais, cognitivas e físicas, nos diferentes estágios do processo. As pessoas procuram e usam informação de maneira diferente, dependendo do estágio do processo. Os seis estágios do modelo Kuhlthau do processo de *information seeking* são os seguintes:

1. Início (*initiation*): tornar-se ciente da necessidade de informação.
2. Seleção (*selection*): identificar e selecionar o tópico geral a ser pesquisado.
3. Exploração (*exploration*): procurar informação sobre o tópico geral.
4. Formulação (*formulation*): formar um foco a partir da informação encontrada.
5. Coleta (*collection*): coletar informação relacionada ao tópico enfocado.
6. Apresentação (*presentation*): completar a pesquisa, apresentar e/ou usar os resultados.

Kuhlthau associa ao longo dos estágios do processo os sentimentos, pensamentos e ações relacionadas, além das tarefas adequadas para cada estágio, como mostrado na Tabela 2.1. Por exemplo, o estágio de Iniciação (*initiation*) do modelo é caracterizado pelos sentimentos de incerteza, por pensamentos vagos e gerais sobre a área do problema e está associado com a pesquisa por informação de fundamentação (*background*); a tarefa adequada para este estágio é simplesmente reconhecer a necessidade de informação. Através do modelo é possível observar que os sentimentos, pensamentos e ações se modificam à medida que os estágios avançam, por exemplo, passando da incerteza inicial para a satisfação ou decepção no final do processo indicado pelo modelo.

Kuhlthau também afirma que as questões básicas de pesquisa sobre as razões do comportamento de pessoas que procuram por informação, as quais foram tratadas em seu estudo, produzem resultados que também têm aplicação prática para o projeto e desenvolvimento de sistemas. Entretanto, ele também admite que estudos sobre como os resultados da pesquisa apresentada poderiam ser aplicados no projeto e desenvolvimento de sistemas ainda são necessários.

Tabela 2.1: Modelo Kuhlthau [1991] de *information seeking*

Estágios do processo de <i>information seeking</i>	Sentimentos comuns a cada estágio	Pensamentos comuns a cada estágio	Ações comuns a cada estágio	Tarefa apropriada de acordo com o Modelo de Kuhlthau
1. Início	Incerteza	Geral/ Vago	Buscar informação básica	Reconhecer
2. Seleção	Otimismo			Identificar
3. Exploração	Confusão/ Frustração/ Dúvida		Buscar informação relevante	Investigar
4. Formulação	Clareza	Específico/ Claro		Formular
5. Coleta	Senso de direção/ Confiança	Interesse	Buscar informação relevante ou focalizada	Coletar
6. Apresentação	Alívio/ Satisfação ou Frustração	Claro e muito focalizado		Completar

2.3.3 Modelos de *information searching*

Nesta subseção, os principais modelos de *information searching* encontrados na literatura são descritos. Esses modelos tentam descrever as atividades relacionadas com a interação entre uma pessoa com uma necessidade de informação e um sistema de informação.

Modelo Bates

Bates [1989] propõe um modelo de *information searching*, denominado de “*berry-picking*” em analogia ao processo de colher pequenas frutas na floresta, as quais devem ser colhidas uma por vez. Segundo Bates, o modelo está mais próximo ao comportamento real das pessoas que procuram por informação do que o modelo tradicional de recuperação de informação.

No princípio subjacente ao modelo, ilustrado na Figura 2.7, é pressuposto que cada parte nova de informação que as pessoas encontram fornece a elas novas idéias e a direção a seguir e, conseqüentemente, uma nova concepção de sua necessidade de informação. Em cada estágio da atividade, o usuário pode identificar itens de informação úteis que o conduzem a seguir adiante. Em vez de versões alternativas de uma consulta (por exemplo, a modificação dos termos utilizados), ocorrem modificações nas estruturas cognitivas (associadas com a necessidade de informação) devido à confrontação com o contexto da informação. Bates denomina esse comportamento de “busca desenvolvida gradualmente” (*evolving search*), no sentido de que a necessidade de informação não é satisfeita por um único item no final da atividade, mas por uma série de seleções de itens individuais ao longo do caminho percorrido.

Bates sugere diversas técnicas manuais que poderiam auxiliar melhor os usuários, nas suas estratégias de *information seeking*, do que aquelas fornecidas pelos sistemas da época. A discussão de cada uma dessas técnicas é acompanhada por sugestões

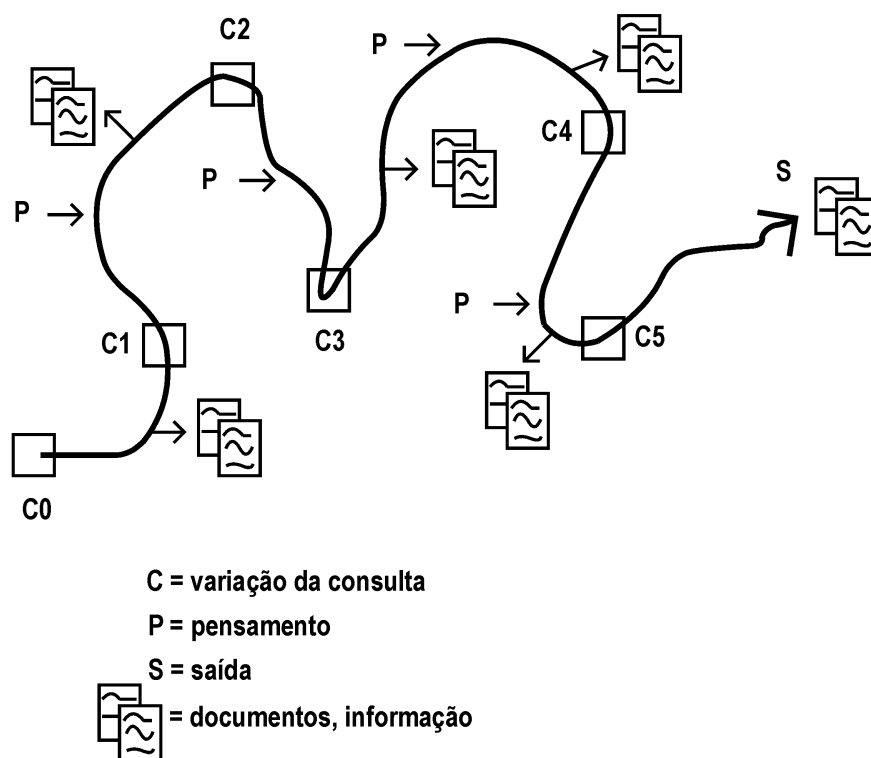


Figura 2.7: Modelo “berrypicking” proposto por Bates [1989]

de como incorporar a técnica no projeto de interface. No entanto, essas técnicas são específicas para o domínio de bibliotecas digitais: seguir uma nota de rodapé (*footnote chasing*); buscar o conjunto de referências que citam uma determinada referência (*citation searching*); percorrer as edições de um periódico (*journal run*); buscar um assunto em diversas fontes (*subject searching*); examinar sistematicamente uma área (*area scanning*); e buscar a produção de um autor (*autor searching*).

Modelo Ingwersen

Ingwersen [1996] propôs um modelo cognitivo, baseado em estudos analíticos e empíricos orientados para o usuário, para representar a interação entre o usuário e sistemas de recuperação de informação. No modelo mostrado na Figura 2.8, a interação é um conjunto de processos cognitivos, que podem ocorrer em todos os elementos envolvidos no processo. Por exemplo, o usuário não interage somente com o sistema, mas também com os objetos de informação, que são estruturas cognitivas considerados como um espaço de informação.

O modelo fornece uma visão geral da localização lógica de um sistema de recuperação de informação. Além disso, o modelo também representa as diversas transformações cognitivas que acontecem desde o momento em que o usuário identifica uma necessidade de informação até a situação na qual ele pode pesquisar um conjunto de objetos de informação e identificar aqueles considerados úteis. A interface funciona como o elemento para intermediar a interação entre usuário e sistema.

O modelo não leva à aplicação direta para o projeto e desenvolvimento de sistemas.

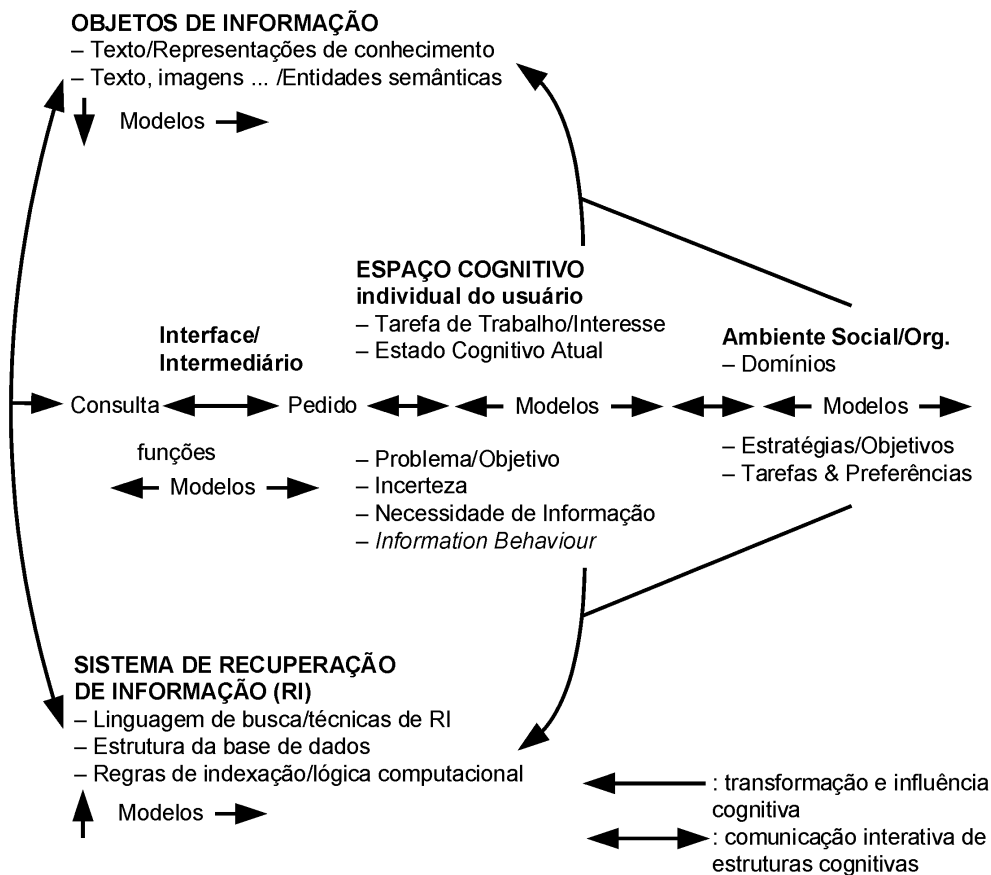


Figura 2.8: Modelo cognitivo proposto por Ingwersen [1996]

No entanto, o modelo mostra uma representação de como os elementos relacionados com *information searching* e recuperação de informação estão integrados. A visão de um processo, entretanto, não está explícita no modelo, e também não são apresentadas informações de como ocorre o processo.

Modelo Saracevic

Saracevic [1996, 1997] propõe um modelo estratificado, mostrado na Figura 2.9, que descreve a interação entre usuário e sistema (computador) como um diálogo entre os participantes, realizado por meio de uma interface no nível de superfície. Além disso, na representação do modelo cada um dos participantes (usuário e sistema) tem níveis/estratos diferentes. Na parte correspondente ao usuário estão envolvidos os seguintes níveis: cognitivo, afetivo e situacional. Na parte do sistema (computador) existem os níveis: engenharia, processamento e conteúdo. As mudanças que ocorrem nos níveis de uma das partes (usuário e sistema) influenciam as ações que são tomadas no nível de superfície. Por exemplo, a escolha dos termos para uma nova consulta é resultado dos documentos recuperados (ou da falta de documentos) a partir da consulta previamente realizada.

Saracevic [1997] discute a extensão do modelo para incluir elementos mais específicos que representam outros papéis na interação: noção de relevância, modelagem

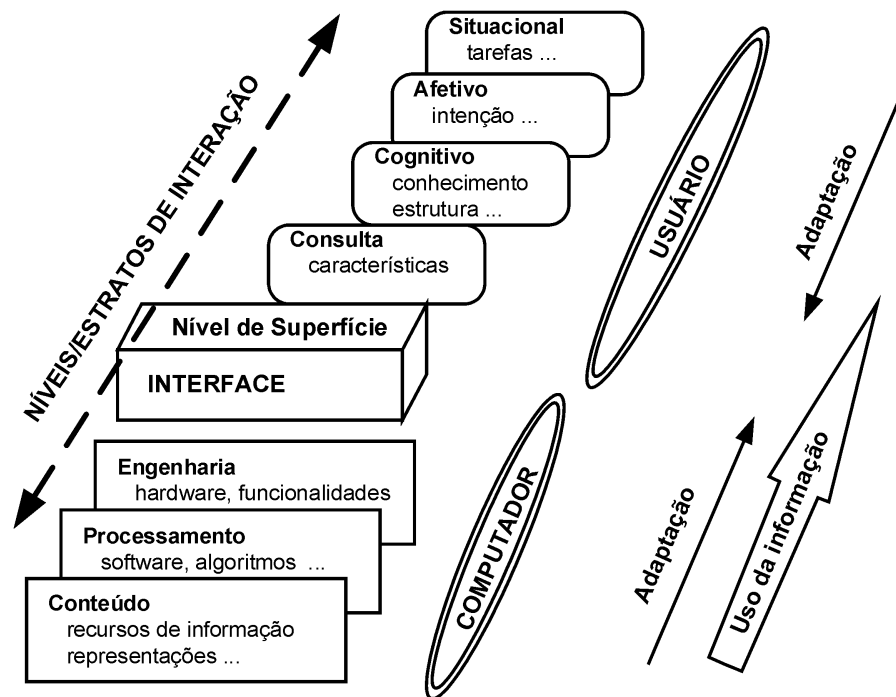


Figura 2.9: Modelo estratificado proposto por Saracevic [1996]

do usuário, seleção de termos e tipos de realimentação (*feedback*). Saracevic afirma que a vantagem do modelo é separar os elementos complexos envolvidos na interação em partes mais fáceis de manipular, descrever, representar e de compreender. Além disso, ele também afirma que a melhor compreensão da interação é um pré-requisito para o projeto e desenvolvimento de sistemas. No entanto, não se encontram quaisquer indicações da aplicação específica do modelo no desenvolvimento de sistemas, mas somente sugestões gerais em alto nível de abstração.

Modelo Spink

Spink [1997] propõe um modelo de *information searching*, baseado em estudos empíricos, como um processo cíclico de laços interativos de realimentação (*interactive feedback loops*) no qual um usuário interage com um sistema de recuperação de informação. O modelo, mostrado na Figura 2.10, é formado pelos seguintes elementos: ciclo, laço interativo de realimentação, tática ou movimento de busca e julgamento do usuário.

Segundo Spink, cada estratégia de *information searching* pode consistir em um ou mais ciclos, por exemplo, consultas ao sistema e as respostas correspondentes. Um ciclo, por sua vez, pode consistir em uma ou mais ocorrências de laços interativos de realimentação. Da mesma maneira, cada laço interativo de realimentação pode ser composto por uma ou mais táticas (ou movimentos) de busca, destinadas a melhorar ou avançar o processo. Em cada tática de busca, composta por uma consulta do usuário e da correspondente resposta do sistema, podem ocorrer vários julgamentos do usuário com relação à resposta apresentada pelo sistema (por exemplo, conside-

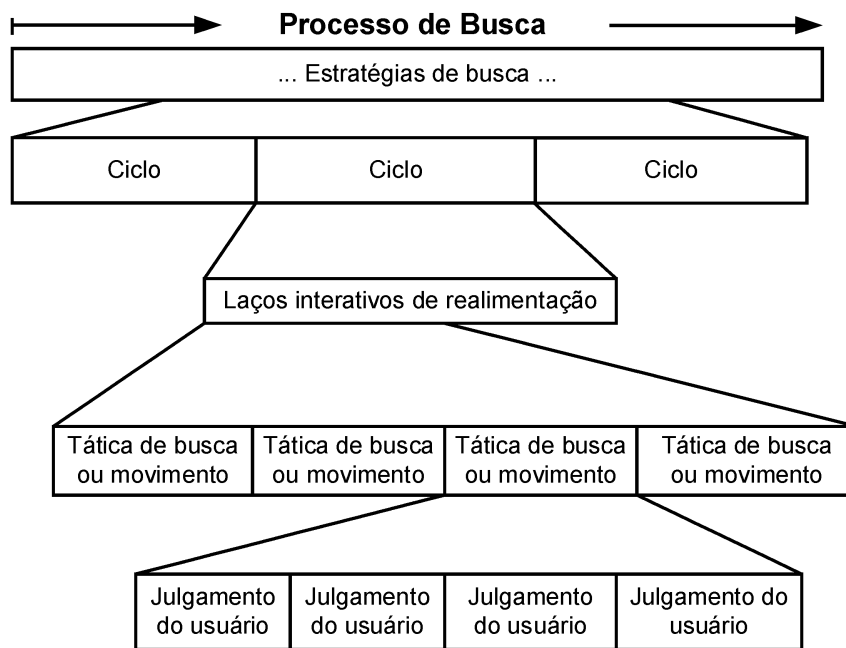


Figura 2.10: Modelo Spink de *information searching*

rando a relevância ou a quantidade de documentos encontrados).

O foco do modelo é auxiliar na análise das diferentes categorias de laços interativos de realimentação em sistemas de recuperação de informação. As categorias identificadas envolvem questões importantes que devem ser consideradas no projeto e desenvolvimento de sistemas, pois a realimentação é parte da dinâmica da interação entre o usuário e sistema. No entanto, nenhuma outra orientação com relação a desenvolvimento de sistemas é fornecida além das relacionadas a essas categorias.

Modelo Navarro-Prieto, Scaife, e Rogers

Navarro-Prieto et al. [1999] identificaram estratégias cognitivas relacionadas a *information searching* na Web. Os autores realizaram um estudo de comparação entre usuários experientes e novatos e concluíram que o comportamento dos usuários com experiência é de planejar primeiro a maneira de realizar a tarefa com base no seu conhecimento sobre a Web, enquanto que os usuários com pouca experiência dificilmente planejam e são conduzidos por representações externas (isto é, o que eles vêem no navegador).

A partir desse estudo, três padrões diferentes de *information searching* na Web foram identificados:

- Estratégia *top-down*: esse padrão ocorre quando o usuário procura em uma área geral (por exemplo, um *website*) e depois direciona o foco da busca a partir dos links disponíveis.
- Estratégia *bottom-up*: nesse padrão, o usuário procura utilizando palavras-chave específicas (por exemplo, por meio de *search engines*).

- Estratégia *mixed*: o usuário utiliza as duas estratégias anteriores em paralelo (por exemplo, alternando várias janelas).

Navarro-Prieto et al. construíram dois modelos a partir da análise dos resultados conforme representado na Figura 2.11: (a) modelo para usuários experientes, e (b) modelo para usuários novatos. No modelo para usuários experientes, a busca inicia com um planejamento que considera como a informação (com relação ao objetivo) está organizada na Web e o próprio objetivo em questão (neste caso, a tarefa é exploratória ou de encontrar fatos); esses passos não são necessariamente seqüenciais. No modelo para usuários novatos, a busca inicia sem qualquer tipo de planejamento.

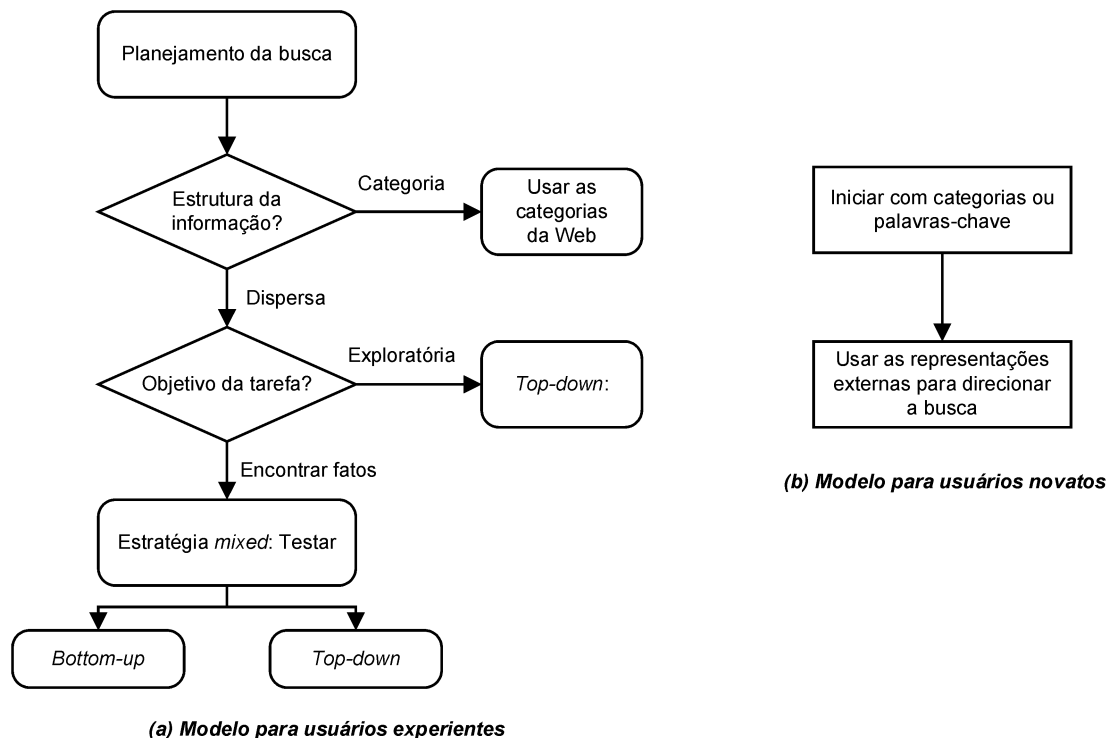


Figura 2.11: Modelo Navarro-Prieto, Scaife, e Rogers [1999] de *information searching* na Web para (a) usuários experientes e (b) usuários novatos

O objetivo dos autores foi desenvolver um modelo com base em estudos empíricos, para auxiliar a explicar como as pessoas procuram por informação na Web e para desenvolver diretrizes para auxiliar as atividades de *information searching* na Web. Nesse sentido, o modelo não tem aplicação direta para o projeto e desenvolvimento de sistemas.

Modelo Choo, Detlor, e Turnbull

Choo et al. [2000] desenvolveram um modelo para descrever o comportamento de *information searching*¹⁷ na Web observado em um estudo sobre como profissionais procuram por informação no seu trabalho diário. O modelo, mostrado na Tabela 2.2,

¹⁷Os autores fazem referência ao modelo como de *information seeking*. No entanto, de acordo com a definição apresentada na Subseção 2.3.1, o modelo de Choo et al. [2000] deve ser considerado de

é uma matriz com exemplos genéricos dos episódios de comportamento observados no estudo. As linhas correspondem aos modos de *information searching* identificados: visualização não dirigida (*undirected viewing*), visualização condicionada (*conditioned viewing*), busca informal (*informal search*), busca formal (*formal search*). Os autores do modelo caracterizam cada um desses modos de acordo com a generalidade/especificidade da necessidade de informação, das estratégias gerais e do uso pretendido da informação. As colunas correspondem às categorias de comportamento de *information seeking* identificadas por Ellis [1989]: iniciar (*starting*), encadear (*chaining*), explorar (*browsing*), diferenciar (*differentiating*), monitorar (*monitoring*) e extrair (*extracting*).

Tabela 2.2: Modelo Choo, Detlor, e Turnbull [2000] de *information searching* na Web

	Iniciar	Encadear	Explorar	Diferenciar	Monitorar	Extrair
Visualização não dirigida	Identificar, selecionar, e iniciar páginas e sites	Seguir links nas páginas iniciais				
Visualização condicionada			Explorar páginas de entrada, cabeçalhos, mapas de site	Marcar páginas, imprimir, copiar; Ir direto a site conhecido	Visitar novamente sites marcados para procurar informação nova	
Busca informal				Marcar páginas, imprimir, copiar; Ir direto a site conhecido	Visitar novamente sites marcados para procurar informação nova	Usar search engine (local) para extrair informação
Busca formal					Visitar novamente sites marcados para procurar informação nova	Usar search engine para extrair informação

As táticas de *information seeking* que caracterizam cada modo foram identificadas por meio das seqüências recorrentes de ações do navegador iniciadas pelo usuário. Portanto, visualização não dirigida é caracterizada por ações de iniciar e encadear; visualização condicionada é caracterizada por ações de diferenciar, explorar e monitorar; busca informal é caracterizada por diferenciar e extrair (local); e busca formal consiste em monitorar e extrair de maneira sistemática e detalhada.

Os autores desenvolveram um modelo do comportamento de *information searching* na Web com base em modos de navegação e busca que podem ser diferenciados de acordo com a necessidade de informação e a atividade de *information seeking*. Além disso, os autores estavam interessados em métodos operacionais para medir as atividades na Web e métodos complementares para a coleta de dados qualitativos e quantitativos sobre o comportamento dos participantes no seu próprio local de trabalho. O modelo desenvolvido pode ser considerado um mapeamento do modelo de Ellis [1989] para o contexto da Web.

information searching, pois o foco do estudo é a interação do usuário com o navegador para a pesquisa de informação na Web.

Modelo Hölscher e Strube

Hölscher e Strube [2000] conduziram um estudo sobre o comportamento de usuários especialistas e novatos nas atividades de *information searching* na Web para compreender melhor as características desses usuários. Os autores do estudo realizaram dois experimentos. A partir do primeiro experimento, realizado com usuários especialistas, um modelo foi derivado e depois utilizado no segundo experimento como ferramenta para comparar o comportamento dos usuários. Os participantes do segundo experimento foram divididos em quatro grupos de usuários, entre aqueles considerados especialistas ou novatos e aqueles com/sem conhecimento do domínio.

O modelo, mostrado na Figura 2.12, é dividido em dois níveis: (a) nível global, que representa o comportamento que abrange tanto atividades de *browsing* quanto de *searching*; e (b) nível local, que representa a interação direta com *search engines*. Os valores representam as probabilidades de transição de uma ação (retângulo) para outra.

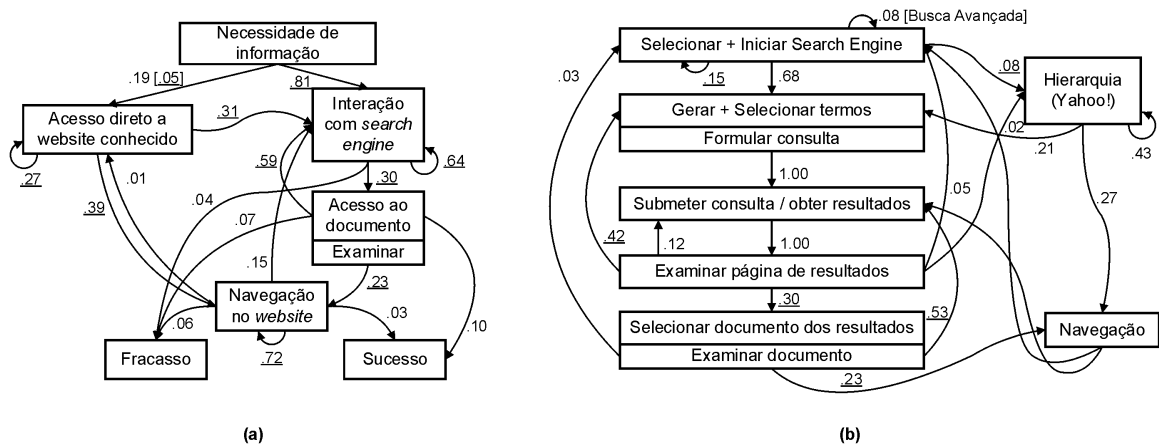


Figura 2.12: Modelo Hölscher e Strube de *information searching* na Web

Os autores identificaram que os participantes especialistas e com conhecimento do domínio foram superiores nas suas atividades. A falta de uma ou outra especialidade (Web ou domínio) conduzia a um comportamento compensatório, por exemplo, os participantes novatos e com conhecimento do domínio confiavam muito na terminologia e evitavam a especificação detalhada de uma consulta. Os participantes com nível de conhecimento baixo (Web e domínio) foram menos flexíveis nas suas estratégias e retornaram aos estágios anteriores da sua busca mais freqüentemente ao contrário de tentar novas abordagens.

Esse estudo foi realizado com o objetivo de melhorar as atividades de *information searching* na Web com base em uma melhor compreensão das características dos usuários, isto é, do seu comportamento quando estão procurando por informação na Web. Portanto, Hölscher e Strube [2000] não propõem o modelo para o projeto e desenvolvimento de sistemas e a sua aplicação nesse contexto não é direta.

2.4 Considerações finais

Neste capítulo, foi apresentada a revisão bibliográfica, dividida em duas partes. Na primeira parte foram abordados os primeiros sistemas hipertexto, a Web, e a Web Semântica, com o objetivo de definir a infra-estrutura para a qual o trabalho foi direcionado e apresentar uma classificação das técnicas de auxílio à navegação.

Na segunda parte, foram descritos os modelos que tentam representar o comportamento das pessoas quando estão procurando por informação.

Os modelos de *information seeking*, Wilson [1997, 1999], Ellis [1989] e Kuhlthau [1991], podem fornecer informação com relação ao suporte necessário para auxiliar atividades de *information seeking*. No entanto, é difícil a aplicação direta desses modelos no projeto e desenvolvimento de sistemas, pois ou os modelos são excessivamente generalistas, como o modelo Wilson, ou não fornecem orientações sobre como utilizá-lo, como os modelos Ellis e Kuhlthau. Ellis e Kuhlthau reconhecem a possibilidade de aplicar o modelo nesse contexto, mas a informação disponível consiste apenas na divisão do processo em estágios e as definições para esses estágios.

Os modelos de *information searching* são focalizados na interação do usuário com o sistema. Ingwersen [1996] apresenta um modelo que não leva à aplicação direta para o projeto e desenvolvimento de sistemas, pois não existe um processo explícito ou informações de como ocorre o processo. O modelo de Saracevic [1996, 1997] é muito parecido com o de Ingwersen, pois considera a dinâmica da interação entre usuário e sistema e especifica os fatores subjacentes a essa dinâmica. O modelo de Saracevic também não possui um processo definido explicitamente.

Bates [1989] e Spink [1997] desenvolveram modelos que poderiam ser utilizados no contexto pretendido, uma vez que fornecem um processo e também estratégias subjacentes bem definidas. No entanto, as orientações com relação ao projeto e desenvolvimento de sistemas estão restritas a bibliotecas digitais, no caso de Bates, e a laços interativos de realimentação, no caso de Spink. Essas informações são importantes e devem ser consideradas, mas não são suficientes para o projeto e desenvolvimento de sistemas.

Navarro-Prieto et al. [1999], Choo et al. [2000] e Hölscher e Strube [2000]) desenvolveram modelos de *information searching* específicos para a Web. Os estudos que serviram de base para os modelos têm o seu foco direcionado para a compreensão do comportamento dos participantes com o objetivo de melhorar o desempenho nas suas atividades de *information searching* na Web. Portanto, não havia interesse explícito no projeto e desenvolvimento de sistemas, apesar de que essa compreensão é importante e deve ser considerada quando esse for o interesse.

Portanto, a partir da revisão desses modelos, a existência de uma lacuna no conhecimento científico foi verificada, para a qual foi direcionada a contribuição deste trabalho.

Metodologia

3.1 Considerações iniciais

A finalidade deste capítulo é descrever a metodologia empregada no desenvolvimento dos trabalhos e pesquisas realizados. Inicialmente, o problema prático que motivou este trabalho é apresentado na Seção 3.2, assim como o problema de pesquisa identificado. Em seguida, as hipóteses e os objetivos são apresentados nas Seções 3.3 e 3.4 respectivamente. Na Seção 3.5, a pesquisa realizada é caracterizada de acordo com classificações encontradas na literatura. Na Seção 3.6 são descritos os aspectos referentes ao método de pesquisa utilizado. Finalmente, as etapas do trabalho são apresentadas na Seção 3.7.

3.2 Motivação e justificativa

A motivação para este trabalho surgiu durante um estudo de caso [Fumagalli et al. 2003] de análise da documentação de projetos de interface, para a abstração de uma estrutura comum da documentação. O estudo envolveu a análise da documentação de 20 projetos, produzida como parte das tarefas da disciplina de Interação Usuário-Computador. A coleção de documentos analisada corresponde àquela referente aos projetos desenvolvidos nas disciplinas ministradas para a graduação e pós-graduação do Instituto de Ciências Matemáticas e de Computação / Universidade de São Paulo (ICMC/USP) no período de 2001 a 2002.

Os alunos da disciplina desenvolveram a documentação usando a CoTeia [Arruda Jr et al. 2002; Arruda Jr e Pimentel 2001], uma ferramenta para aprendizagem colaborativa apoiada por computador, *Computer Supported Collaborative Learning* (CSCL), utilizada no ICMC/USP para complementar as aulas presenciais com atividades colaborativas. A CoTeia é uma ferramenta colaborativa assíncrona, baseada em Wiki [Leuf e Cunningham 2001] e análoga à CoWeb [Guzdial 1999]. Macedo et al. [2001] descrevem experiências de utilização da CoWeb original de Guzdial no ICMC/USP como uma ferramenta para CSCL.

A documentação dos projetos, portanto, estava disponível para acesso como pági-

nas wiki em um repositório¹ CoTeia. O trabalho de análise da documentação exigia a navegação (*browsing and exploration*) através das páginas wiki referentes aos projetos. Durante esse trabalho foi observado que a “meta informação” disponível no conteúdo da página inicial da CoTeia (disciplina e administrador) e nas demais páginas wiki (título da página, grupo de alunos, nome do projeto, fase do projeto) exercia um papel fundamental no processo de navegação, principalmente no sentido de orientação. Nesse contexto levantou-se a hipótese de desenvolver um suporte com base em metadados para auxiliar/facilitar a navegação nas páginas wiki dos repositórios CoTeia.

A ferramenta CoTeia tem sido utilizada no contexto educacional desde 2001 e contém uma quantia significativa de material educacional. Embora a CoTeia forneça o recurso de busca por palavra-chave², observam-se muitos casos nos quais os usuários engajam em uma navegação exaustiva para encontrar o que eles estão procurando. A partir dos levantamentos realizados posteriormente, na etapa de avaliação, foi confirmado que a maneira usual de procurar por informação na CoTeia é por meio do acesso (e navegação) a partir das entradas (links) correspondentes às disciplinas na página inicial. Portanto, ficou evidente a necessidade de melhor suporte para a navegação nas páginas wiki da CoTeia.

Inicialmente, duas questões importantes surgiram ao considerar a hipótese de desenvolver uma navegação com base em metadados como solução para o problema de procurar por informação na CoTeia. A primeira questão, com relação aos metadados:

Quais elementos de metadados utilizar na descrição das páginas wiki e como criar/produzir os valores para esses metadados?

A segunda questão, com relação ao suporte para a navegação:

Quais ferramentas e estratégias poderiam ser utilizadas para uma navegação com base em metadados?

A partir da revisão bibliográfica realizada sobre os principais padrões de metadados encontrados na literatura para descrever objetos de informação (por exemplo, *Dublin Core* (DC) [DCMI 2003] e *Learning Object Metadata* (LOM) [IEEE LTSC 2002]), foi possível decidir por um conjunto de elementos de metadados (*application profile*) suficiente para descrever páginas wiki utilizadas no contexto educacional. Adicionalmente, a questão da criação dos valores para os metadados foi superada por meio de estratégias e técnicas para a geração automática e semi-automática de metadados, como apresentado nos trabalhos desenvolvidos [Pansanato e Fortes 2005a,b]. Outra

¹No contexto da CoWeb e CoTeia, o termo repositório é utilizado para fazer referência ao conjunto de páginas wiki sob uma mesma entrada (link) na página inicial da ferramenta. Neste trabalho, o termo repositório também é frequentemente utilizado para fazer referência ao conjunto de todas as páginas wiki disponíveis para acesso na ferramenta CoTeia.

²Esse recurso é implementado como uma consulta SQL no conteúdo das páginas wiki armazenadas na base de dados da CoTeia.

decisão importante nesse contexto foi a de utilizar uma infra-estrutura de metadados com base na Web Semântica [W3C 2001]. A idéia foi trabalhar com uma web semântica delimitada e não com toda a Web Semântica (no contexto global) para restringir o escopo do trabalho e evitar inicialmente o problema/requisito de escalabilidade. Portanto, esta atividade aplicada do trabalho desenvolvido ficou restrita a repositórios baseados na Web, isto é, páginas Web dinâmicas que derivam parte ou todo o seu conteúdo a partir de arquivos ou bases de dados, semelhantemente aos repositórios CoTeia. Dessa maneira, segundo a visão de Marshall e Shipman [2003], as diversas contribuições parciais podem se consolidar futuramente na Web Semântica.

A segunda questão era com relação ao suporte (ferramentas e estratégias) que poderia ser utilizado para uma navegação com base em metadados. Na Subseção 2.2.2 foram categorizadas as principais técnicas subjacentes às ferramentas encontradas na literatura que podem ser utilizadas como suporte para a navegação. A diversidade de opções encontrada motivou a investigação sobre como as pessoas procuram por informação, para escolher melhor as ferramentas que poderiam fornecer auxílio efetivo durante o processo de navegação. Em especial, a estratégia identificada no estudo observacional conduzido por Teevan et al. [2004], denominado pelos autores de *orienteering*, tem motivado bastante este trabalho.

Segundo Teevan et al., *orienteering* é uma estratégia (ou classe de estratégias) na qual as pessoas utilizam uma série de passos (ações) para encontrar a informação desejada, sem especificar completamente o objetivo com antecedência. Os autores do estudo também observaram que, para efetuar esses passos, as pessoas geralmente utilizam técnicas e ferramentas diferentes. Além disso, essa estratégia também envolve o uso da informação anterior e de contexto local para guiar a escolha do próximo passo. Outros trabalhos [Aula et al. 2005; Hölscher e Strube 2000; Navarro-Prieto et al. 1999] também têm descrito estratégias semelhantes, mas não foi atribuída qualquer denominação especial para o comportamento identificado. No entanto, apesar dessa estratégia ser utilizada pelas pessoas, pouco ou nenhum suporte a estratégias é fornecido pelas ferramentas geralmente disponíveis para a navegação na Web.

Um sistema interativo poderia fornecer diversas ferramentas diferentes para permitir ao usuário escolher a ferramenta adequada ou a melhor combinação de ferramentas de acordo com a sua estratégia, provavelmente ajustada ao seu nível de habilidade e conhecimento, à sua preferência, e ao tipo de informação que ele está procurando no momento. Seguindo os conceitos de Booth et al. [2005] sobre a formulação do problema prático, uma questão que poderia ser assim caracterizada é a seguinte:

Como desenvolver um sistema de navegação para o usuário criar estratégias usando ferramentas com base em metadados (infra-estrutura da Web Semântica)?

Na Seção 2.3 foi apresentada a revisão bibliográfica sobre modelos no domínio de

information seeking e de *information searching*. Segundo Wilson [1999], esses modelos podem sugerir relacionamentos entre teorias que poderiam ser explorados ou testados. No entanto, é difícil a aplicação direta desses modelos no projeto e desenvolvimento de ferramentas e sistemas. Além disso, os modelos para *information searching* discutidos também na Seção 2.3 não fornecem diretrizes (ou orientações) para o projeto e desenvolvimento de sistemas interativos para navegação (por exemplo, baseada em metadados). Portanto, existe a necessidade de investigar modelos de navegação adequados nesse contexto para servir de referência para desenvolvedores de sistemas. Recentemente, White et al. [2006b] apresentou um modelo de processo, mas ele está em um nível de abstração muito alto e deixa em aberto a maioria das questões descritas no modelo NAVE.

A existência dessa lacuna impede o desenvolvimento de sistemas de navegação que poderiam ser extremamente úteis para a melhoria das atividades das pessoas que envolvem a pesquisa por informação na Web Semântica. Portanto, essa necessidade poderia ser considerada um problema de pesquisa na visão de Booth et al. [2005]. O problema de pesquisa identificado nesta tese pôde ser formulado da seguinte maneira:

Proposição de um modelo de navegação exploratória para auxiliar a criação de sistemas para pesquisa de informação no contexto da Web Semântica.

3.3 Hipóteses

Uma vez formulado o problema de pesquisa, a seguinte *hipótese básica* é proposta como solução do problema:

É viável a criação de um modelo de navegação exploratória para auxiliar o projeto e desenvolvimento de sistemas para pesquisa de informação no contexto da Web Semântica.

A hipótese básica é complementada pela seguinte *hipótese secundária*:

Um sistema segundo esse modelo é “melhor” do que sistemas tradicionais. Usando um sistema (para navegação exploratória em documentos) segundo esse modelo, os usuários têm desempenho superior do que usando sistemas tradicionais.

Segundo Silva e Menezes [2005], essa hipótese pode ser considerada como secundária porque envolve um aspecto não especificado na hipótese básica. Nesse caso, a questão da qualidade dos produtos resultantes da aplicação do modelo. A inclusão desse aspecto diretamente na hipótese básica comprometeria a sua simplicidade e clareza. Sistemas tradicionais são considerados aqueles com interface e comportamento bem conhecidos, por exemplo, uma ferramenta de busca por palavra-chave (*keyword search*), que é geralmente fornecida como suporte para a pesquisa no conteúdo de um *website*.

3.4 Objetivos

O objetivo geral deste trabalho é definido da seguinte maneira:

Criar um modelo de navegação exploratória para servir de referência para o desenvolvimento de sistemas para pesquisa de informação no contexto da Web Semântica.

Esse é o objetivo que, ao ser alcançado, propicia uma variedade de formas de navegação que podem ser implementadas considerando a infra-estrutura da Web Semântica, em especial, utilizando o potencial dos metadados.

A abordagem adotada para atingir esse objetivo envolveu a proposta de um modelo inicial, o emprego deste modelo proposto para o desenvolvimento de um sistema específico para um determinado domínio e a avaliação deste sistema com os usuários. A vantagem de adotar essa abordagem é ter meios para aprimorar no futuro o modelo, o sistema e a avaliação, utilizando novas investigações que venham demonstrar todo o potencial do modelo de navegação exploratória definido.

A partir da definição do objetivo geral e da abordagem a ser seguida, os objetivos específicos podem ser assim definidos:

1. *Propor um modelo para servir de referência para o desenvolvimento de sistemas para a navegação na Web Semântica.*
2. *Desenvolver um sistema de navegação com base no modelo proposto para pesquisa de informação no contexto da Web Semântica.*
3. *Avaliar o sistema desenvolvido.*

3.5 Caracterização da pesquisa

Nesta seção são descritas, de maneira sucinta, as principais características da pesquisa realizada. Segundo Cervo et al. [2007], cada processo de pesquisa admite níveis diferentes de aprofundamento e enfoques específicos conforme o objeto de estudo, os objetivos definidos e a qualificação do pesquisador. Portanto, o objetivo não é descrever todos os aspectos que a pesquisa possa abordar ou todas as classificações existentes.

A natureza da pesquisa é aplicada, pois objetiva gerar conhecimentos para a aplicação prática dirigidos à solução de um problema concreto [Silva e Menezes 2005; Cervo et al. 2007]. A pesquisa pura (ou básica) objetiva gerar conhecimentos para o avanço da ciência, sem aplicação prática prevista.

A pesquisa pode ser classificada, segundo a forma de abordagem do problema, entre quantitativa e qualitativa [Silva e Menezes 2005]. Basicamente, na abordagem quantitativa as variáveis subjacentes ao problema podem ser definidas e medidas

para classificação e análise. Na qualitativa, por outro lado, a ênfase é a interpretação dos fenômenos e atribuição de significados, considerando o vínculo existente entre o mundo objetivo e a subjetividade dos indivíduos. Essas abordagens não são excludentes e podem coexistir conjuntamente em uma mesma pesquisa [Pádua 2006]. A combinação dessas abordagens é oportuna devido à natureza do problema de pesquisa identificado. Assim, a pesquisa realizada utiliza tanto a abordagem quantitativa quanto a qualitativa, principalmente na etapa de avaliação, para apreender mais sobre a dinâmica do processo de navegação (predominantemente qualitativa) e para comparar dois sistemas (predominantemente quantitativa).

A pesquisa também pode ser classificada, a partir do ponto de vista de seus objetivos, em exploratória, descritiva e explicativa [Silva e Menezes 2005]. A pesquisa realizada é exploratória, principalmente devido à revisão bibliográfica necessária para a familiarização e explicitação do problema identificado. Adicionalmente, a pesquisa realizada também é descritiva, pois visa descrever as características de um modelo de navegação que pode ser utilizado para auxiliar o desenvolvimento de sistemas.

3.6 Método de pesquisa

A pesquisa realizada incluiu a criação de um modelo, o seu uso para o desenvolvimento de um sistema de navegação exploratória para um domínio específico, e a avaliação desse sistema com os usuários quanto à sua aplicação nesse domínio. Essa organização reflete a adoção do método de pesquisa denominado de método hipotético-dedutivo, proposto em 1935 por Karl Popper [Silva e Menezes 2005; Marconi e Lakatos 2004].

A pesquisa científica, segundo a abordagem desse método, inicia com a identificação de um problema ou lacuna no conhecimento científico e com a sua descrição clara e precisa. Para tentar explicar as dificuldades expressas no problema identificado são formuladas conjecturas ou hipóteses. As hipóteses formuladas são utilizadas para deduzir conseqüências que deverão ser testadas ou falseadas. Falsear significa tentar tornar falsas as conseqüências deduzidas das hipóteses. Quando não se consegue demonstrar qualquer caso concreto capaz de falsear a hipótese, tem-se a sua corroboração, que não excede o nível do provisório. A hipótese se apresenta válida porque superou todos os testes, mas não definitivamente confirmada, pois a qualquer momento poderá surgir um fato que a invalide.

3.7 Etapas do trabalho

A pesquisa realizada pode ser dividida em quatro etapas assim denominadas: (1) preparação da infra-estrutura, (2) desenvolvimento de um sistema, (3) avaliação do sistema com usuários e (4) descrição do modelo. O conhecimento gerado é a soma de todas as informações (e dos resultados) obtidas em cada uma dessas etapas. As etapas são apresentadas a seguir.

3.7.1 Preparação da infra-estrutura

Uma etapa de preparação foi necessária para tratar principalmente duas questões levantadas durante o Exame de Qualificação. A primeira questão está relacionada ao problema de se produzir valores adequados para os metadados. Geralmente, os autores de objetos de informação não preenchem suficientemente os campos reservados aos metadados (por exemplo, como nos problemas identificados por Ryan e Walmsley [2003] com relação a objetos de aprendizagem). A direção adotada foi a de investigar métodos para a geração automática de metadados [Pansanato e Fortes 2005a,b]. Resumidamente, o trabalho envolveu a análise do conteúdo das páginas do repositório da CoTeia, a definição de um *application profile* para páginas wiki educacionais, a extensão do suporte a metadados da CoTeia e o desenvolvimento de um sistema, denominado MeGen [Pansanato e Fortes 2006], para a geração automática de valores para os elementos de metadados do *application profile*. Nesse contexto, as discussões ocorridas no início dos projetos *Tecnologia da Informação para o Desenvolvimento da Internet Avançada – Aprendizado Eletrônico* (TIDIA-Ae)³ e *Software Engineering Available for Everyone* (SAFE)⁴ [Fortes et al. 2004] foram importantes para o avanço do trabalho e a investigação de alternativas para pesquisa, em especial com relação a metadados [Pansanato e Fortes 2004].

A segunda questão está relacionada à forma de avaliação do sistema que foi desenvolvido com base no modelo proposto. Essa questão foi tratada como um problema de avaliar sistemas interativos para recuperação de informação [Belkin et al. 2004; Borlund 2003]. Uma revisão bibliográfica foi conduzida para identificar as principais técnicas de avaliação, variáveis, e medidas que poderiam ser utilizadas [White et al. 2006b; Dix et al. 2004; Borlund 2003; Preece et al. 2002; Wohlin et al. 2000; Baeza-Yates e Ribeiro-Neto 1999; Dumas e Redish 1999]. Considerando essa revisão, a decisão foi utilizar na avaliação tanto a abordagem qualitativa quanto a quantitativa, além de diversos métodos de coleta de dados para maximizar a aquisição de informações sobre o sistema. Adicionalmente, para auxiliar a avaliação, foi decidido pelo desenvolvimento de um sistema para registro e análise da interação do usuário, com funcionalidades semelhantes ao sistema WebTracker⁵.

3.7.2 Desenvolvimento de um sistema

A proposta inicial do modelo consistiu em apenas um esboço da dinâmica do processo de navegação pretendido. Além disso, foram definidos alguns requisitos considerados necessários para um sistema de navegação baseado nesse modelo inicial. Isso foi suficiente para servir de base para o desenvolvimento de um protótipo de software. Pansanato e Fortes [2005c] apresentaram as idéias iniciais do trabalho realizado à época. O modelo de navegação foi elaborado à medida que o desenvolvimento do protótipo ocorria e os conceitos eram testados na prática, o que caracteriza uma

³<http://tidia-ae.iv.fapesp.br/>

⁴<http://safe.icmc.usp.br>

⁵<http://www.ischool.utexas.edu/~donturn/research/webtracker/index.html>

abordagem de investigação do específico para o genérico.

O desenvolvimento do sistema de navegação exploratória, denominado de ENS [Pansanato e Fortes 2007a], assim como as soluções apresentadas, consiste na verificação inicial da viabilidade do modelo proposto, pois a partir dessa experiência de desenvolvimento é realizada uma primeira avaliação das dificuldades e reais possibilidades de criação de um modelo de navegação exploratória.

Os resultados dessa etapa estão relacionados com a hipótese básica (Seção 3.3), de que é possível criar um modelo de navegação exploratória para servir de referência para o desenvolvimento de sistemas para pesquisa de informação na Web Semântica. Assim, a meta desta etapa foi verificar a capacidade de empregar o modelo na construção de soluções. Caso o modelo não permitisse a construção de uma solução, a hipótese não seria verificada, mesmo que fosse encontrado um modelo qualquer anteriormente.

3.7.3 Avaliação do sistema com usuários

O objetivo desta etapa foi avaliar o sistema e analisar os resultados encontrados. A avaliação realizada foi centrada em usuários, e envolveu tanto a abordagem qualitativa quanto a quantitativa. Uma maneira considerada para avaliar o sistema foi conduzir um estudo de usabilidade (Seção 6.2) [Pansanato e Fortes 2007a], no qual predominou a abordagem qualitativa. A outra maneira considerada foi realizar um estudo experimental (Seção 6.3) [Pansanato e Fortes 2007b], no qual predominou a abordagem quantitativa. A partir da experiência com a avaliação do sistema com usuários, o modelo de navegação foi revisado.

Os resultados do estudo experimental conduzido nesta etapa estão relacionados com a hipótese secundária (Seção 3.3), de que os usuários têm desempenho superior utilizando o sistema desenvolvido do que utilizando sistemas tradicionais. Além disso, foi possível realizar uma nova avaliação da hipótese básica (Seção 3.3), neste caso com relação aos usuários participantes dos estudos conduzidos.

3.7.4 Descrição do modelo

Para descrever o modelo proposto foi elaborada uma documentação, a qual consiste em uma representação gráfica do modelo e nas descrições dos seus elementos (estágios, decisões e entidades). Além disso, as orientações com relação ao tipo de suporte para a navegação (ferramentas), que um sistema desenvolvido segundo o modelo deve oferecer ao usuário, foram organizadas de acordo com cada estágio do modelo. Essas orientações estão relacionadas com a classificação (Seção 2.2.2) das técnicas para auxiliar a navegação na Web.

A documentação do modelo também inclui um conjunto de recomendações para o projeto e desenvolvimento de sistemas de navegação com base no modelo proposto. Essas recomendações são derivadas não somente do modelo, mas também da experiência (lições aprendidas) obtida com o desenvolvimento e avaliação do sistema.

3.8 Considerações finais

Neste capítulo, a metodologia empregada no desenvolvimento dos trabalhos e pesquisas realizados foi apresentada. De maneira resumida, a abordagem adotada para atingir esse objetivo envolveu a proposta de um modelo inicial, o emprego deste modelo proposto para o desenvolvimento de um sistema específico para um determinado domínio e a avaliação deste sistema com os usuários. A vantagem de adotar essa abordagem é ter meios para aprimorar no futuro o modelo, o sistema e a avaliação, utilizando novas investigações que venham demonstrar todo o potencial do modelo.

Um Modelo de Navegação Exploratória para a Web Semântica

4.1 Considerações iniciais

Neste capítulo é proposto um modelo de navegação exploratória, denominado de NAVE, para auxiliar o projeto e desenvolvimento de sistemas de apoio a atividades de *information searching* na Web Semântica. O principal objetivo deste modelo é servir como referência para o projeto e desenvolvimento de sistemas de navegação para auxiliar atividades de busca exploratória na Web Semântica. O termo “navegação” é utilizado (ao invés de “busca”) para enfatizar a liberdade do usuário de decidir qual estratégia utilizar para satisfazer a sua necessidade de informação.

O modelo se mostra adequado para a Web Semântica, pois o sistema desenvolvido como parte da validação (prova de conceito) foi construído sobre uma infra-estrutura com base em tecnologias da Web Semântica. Entretanto, devido à sua independência com relação a tecnologias específicas, o modelo pode ser utilizado no contexto de outros ambientes, por exemplo, bibliotecas digitais, e da própria Web. O modelo pode evoluir, no futuro, para um *framework* de diretrizes, recomendações, e práticas, direcionadas para o projeto e desenvolvimento de sistemas de navegação exploratória para a Web Semântica.

O capítulo está organizado da seguinte maneira: na Seção 4.2 são apresentados os conceitos subjacentes ao modelo; na Seção 4.3 são apresentadas as características relacionadas ao desenvolvimento do modelo; na Seção 4.4, o modelo é descrito; na Seção 4.5, a aplicação do modelo e as recomendações são abordadas; na Seção 4.6 são apresentadas as considerações finais com relação ao modelo.

4.2 Conceitos subjacentes ao modelo

Os modelos de *information searching*, descritos anteriormente na Seção 2.3, auxiliam na compreensão do processo no qual o usuário ingressa, motivado por uma

necessidade de informação. No entanto, esses modelos não fornecem orientações práticas (ou que possam ser aplicadas na prática) para o projeto e desenvolvimento de sistemas. O modelo proposto fornece orientações para uma abordagem de navegação fundamentada nos níveis de atividade de *information seeking* descritos por Bates (Subseção 4.2.1) e na estratégia de *orienteering* observada em estudos de comportamento de *information seeking* (Subseção 4.2.2), de maneira a fornecer suporte para um tipo de atividade conhecida por “busca exploratória” (Subseção 4.2.3).

4.2.1 Níveis de atividade de *information seeking*

Os níveis de atividade de *information seeking* descritos por Bates [1990] são explorados neste trabalho para estabelecer uma terminologia com relação ao termo estratégia e os seus componentes básicos. Bates, com base em estudos empíricos do comportamento de usuários de bibliotecas, distingue quatro níveis (ou tipos) de atividade:

1. Movimento (*move*): consiste em uma ação ou pensamento identificável que é parte de *information seeking*. Por exemplo, localizar algum trecho de texto em uma página wiki. Segundo Bates, o termo “movimento” é neutro, e pode ser aplicado a qualquer tipo de atividade. O movimento é a unidade básica de análise de comportamento de *information seeking*.
2. Tática (*tactic*): consiste em um movimento ou vários movimentos realizados com a finalidade de melhorar ou acelerar uma atividade. Por exemplo, ampliar ou restringir uma consulta para recuperar um número maior ou menor de páginas wiki.
3. Estratagema (*stratagem*): consiste em um conjunto de movimentos e/ou táticas e envolve geralmente um determinado domínio identificado para a atividade, e também uma maneira de tirar proveito da forma como esse domínio está organizado/estruturado. Por exemplo, procurar primeiro nas páginas wiki de uma determinada disciplina e depois naquelas relacionadas a um autor específico.
4. Estratégia (*strategy*): consiste em um plano para satisfazer uma necessidade de informação e pode incluir combinações de todos os tipos de atividades mencionados anteriormente (estratagema, tática e movimento). Por exemplo, realizar uma busca por palavra-chave para recuperar páginas wiki relacionadas a termos específicos que descrevem um assunto, selecionar para leitura aquelas consideradas relevantes, anotar as referências no texto para os artigos que podem fornecer mais detalhes e recuperar esses artigos a partir de outra fonte se necessário.

De acordo com essa ordem, uma tática pode envolver um ou mais movimentos, um estratagema pode incluir táticas e movimentos, e uma estratégia pode incluir estratagemas, táticas e movimentos. Entretanto, Bates ressalta que esses quatro

níveis de atividade não são somente tamanhos diferentes da mesma natureza. Não é necessário combinar alguns movimentos para obter uma tática, algumas táticas para obter um estratagema e alguns estratagemas para obter uma estratégia. Uma tática pode consistir em um único movimento ou em vários movimentos. Um estratagema pode ser composto de táticas ou apenas de movimentos; cada nível de atividade é conceitualmente diferente.

4.2.2 *Orienteering*

Nesta seção, as principais características da estratégia que motivou o desenvolvimento do modelo são apresentadas. A estratégia de *orienteering* pode ser representada por meio da seguinte metáfora:

Orienteering¹ é um esporte no qual uma pessoa ou equipe deve completar um percurso através de uma área aberta, passando por determinados pontos, e utilizando um mapa e uma bússola para a navegação. A principal estratégia utilizada nesse esporte consiste em considerar as informações da posição atual para escolher o caminho para alcançar o próximo ponto ou o objetivo final.

O termo *Orienteering* também é atribuído a um comportamento observado em estudos de *information seeking* apresentados na literatura [Teevan et al. 2004; O'Day e Jeffries 1993], por causa da semelhança das estratégias utilizadas. *Orienteering* é uma estratégia (ou classe de estratégias) na qual as pessoas utilizam uma série de passos (ou ações) para satisfazer a sua necessidade de informação, sem especificar completamente o objetivo com antecedência. Os autores também observaram que, para efetuar esses passos, as pessoas utilizaram técnicas e ferramentas diferentes. Além disso, essa estratégia envolve o uso da informação anterior e de contexto local para guiar a escolha do próximo passo. Outros trabalhos [Aula et al. 2005; Hölscher e Strube 2000; Navarro-Prieto et al. 1999] também têm descrito estratégias semelhantes, mas não foi atribuída qualquer denominação especial para o comportamento identificado.

As principais características do comportamento observado que representam a abordagem da estratégia de *orienteering* podem ser assim resumidas [Teevan et al. 2004; O'Day e Jeffries 1993]:

- As pessoas utilizam uma série de passos interconectados para procurar a informação desejada, sem especificar com antecedência toda a necessidade de informação. Geralmente, esses passos são tomados de maneira a limitar o espaço de informação na direção da informação desejada.

¹No Brasil é chamado de corrida de orientação, Confederação Brasileira de Orientação, <http://www.cbo.org.br/>

- Existe variação no “tamanho” (complexidade) dos passos tomados ao longo do caminho e nos métodos escolhidos para efetuar esses passos. O significado de passos pequenos é relacionado ao uso de ferramentas simples. Diversas técnicas de consulta e navegação são utilizadas para efetuar os passos.
- Cada passo da navegação é seguido por leitura (*scanning*), assimilação e análise das informações obtidas. Essas atividades podem ser consideradas como passos de transição, isto é, atividades para auxiliar a decisão sobre os próximos passos que serão tomados. Dessa maneira, a informação contextual é utilizada para guiar a escolha do próximo passo.
- A necessidade de informação se modifica e evolui à medida que os passos são efetuados, provavelmente devido à melhor compreensão do problema que motivou a necessidade de informação e/ou do espaço de informação.

4.2.3 Busca exploratória

Definir o que é uma busca exploratória (*exploratory search*) consiste em um desafio [White et al. 2006a]. Na realidade, quase todas as buscas são exploratórias de alguma maneira. Geralmente, as atividades de busca exploratória são motivadas por um problema bastante complexo, no qual a terminologia utilizada não é conhecida ou é mal compreendida, e ainda quando falta conhecimento da estrutura do espaço de informação. A busca exploratória representa uma mudança da abordagem analítica da recuperação de informação tradicional (*query-document matching*) na direção da orientação em todos os estágios do processo de *information seeking*.

Marchionini [2006] classifica as atividades de busca exploratória em dois tipos principais: aprendizagem e investigação. As atividades do tipo aprendizagem envolvem o processamento cognitivo e a interpretação de conhecimento novo; as do tipo investigação requerem avaliação crítica antes da integração nas bases de conhecimento. Além disso, Marchionini também relaciona diversas atividades de busca exploratória com relação a esses dois tipos principais. Essas atividades são as seguintes:

- Aprendizagem:
 - Aquisição de conhecimento;
 - Compreensão/Interpretação;
 - Comparação;
 - Agregação/Integração;
 - Socialização.
- Investigação:
 - Crescimento;
 - Análise;

- Exclusão/Negação;
- Síntese;
- Avaliação;
- Descoberta;
- Planejamento/Previsão;
- Transformação.

Os sistemas desenvolvidos para busca exploratória (por exemplo, RB++ [Marchionini e Brunk 2003], mSpace [m. c. schraefel et al. 2006], entre outros [White et al. 2007]) têm por objetivo tornar a busca mais efetiva, fornecendo uma variedade grande de funcionalidades na interface e alterando dinamicamente como os resultados são apresentados. Outros sistemas também incluem o uso de interfaces que empregam categorização ou agrupamento (*clustering*) [Hearst 2006].

Os pesquisadores de diversas comunidades, por exemplo, de recuperação de informação, visualização de informação, biblioteca digital, interação homem-computador, sociologia, antropologia, etnografia, ciência da informação, têm trabalhado em tecnologias para fornecer suporte para as atividades de busca exploratória. No entanto, esforços são necessários com o objetivo de desenvolver/elaborar um conjunto de diretrizes para o projeto e desenvolvimento de sistemas para suporte de atividades de busca exploratória [White et al. 2007].

4.3 Desenvolvimento

O modelo de navegação exploratória (NAVE) foi desenvolvido considerando os níveis de atividades descritos por Bates (Subseção 4.2.1), a estratégia de *orientteering* (Subseção 4.2.2), assim como a experiência e as lições aprendidas com o desenvolvimento de um sistema, e também os resultados dos estudos conduzidos para avaliar o sistema desenvolvido e investigar a abordagem proposta no modelo.

Os modelos de *information searching* (descritos na Subseção 2.3.3) foram utilizados como ponto de partida para o desenvolvimento do modelo. O foco em modelos de *information searching* (ao invés de *information seeking*) é devido à interação do usuário com o sistema. A proposta inicial do modelo consistiu em apenas um esboço da dinâmica do processo de navegação pretendido para servir de referência para o desenvolvimento de um sistema de navegação exploratória. Os estágios do modelo que tratam com a definição do problema, a formulação de consulta e a avaliação de resultados são aqueles que foram considerados no esboço inicial do modelo. Além disso, foram definidos requisitos considerados necessários para um sistema que pudesse ser desenvolvido com base nesse modelo inicial. O modelo de navegação exploratória foi elaborado à medida que o desenvolvimento do sistema ocorria e os conceitos eram testados na prática, o que caracteriza uma abordagem de investigação do específico para o genérico.

O estudo de usabilidade com usuários potenciais do sistema de navegação exploratória foi considerado para verificar o modelo, fazer correções e expandir o modelo por meio do detalhamento dos estágios, decisões e entidades. Os dados coletados foram analisados, principalmente os registros do “*think-aloud*” e das entrevistas, assim como os arquivos de *log* da interação. As diferentes abordagens (episódios interessantes) adotadas pelos participantes foram analisadas e comparadas com o modelo. Algumas das interações específicas que são freqüentemente realizadas no sistema que estava em desenvolvimento puderam ser generalizadas.

4.4 Descrição

A documentação do modelo de navegação exploratória consiste em quatro partes:

1. Um fluxograma que apresenta os principais estágios e decisões que são partes do processo de navegação proposto. O foco desse diagrama é fornecer uma representação visual da seqüência do processo.
2. A descrição detalhada dos elementos (estágios, decisões e entidades) do modelo, e a indicação da categoria de técnica de navegação adequada para cada parte do modelo.
3. Uma matriz derivada diretamente das descrições para facilitar a associação das partes do modelo com as categorias de técnicas de navegação indicadas para cada uma dessas partes.
4. Um conjunto de recomendações para o projeto e desenvolvimento de sistemas de navegação exploratória.

A seqüência dos estágios do modelo NAVE é representada no fluxograma da Figura 4.1. Os estágios são representados por retângulos (com seqüência numérica), as decisões são representadas por losangos (com seqüência alfabética) e o fluxo seqüencial pode ser acompanhado através das setas. Os retângulos arredondados (sem seqüência indicativa) indicam o início e o fim dos estágios.

4.4.1 Estágios

Os estágios são definidos sob o ponto de vista do usuário, isto é, de acordo com as suas atividades. Existe uma pergunta associada ao nome do estágio para a qual um sistema deve fornecer suporte. Em cada estágio são citadas as categorias de técnicas de navegação (Subseção 2.2.2) que são indicadas para fornecer suporte às atividades do usuário naquele estágio. Essas informações são utilizadas para a construção de uma matriz, mostrada na Tabela 4.1, na qual as categorias de ferramentas estão representadas nas colunas e os estágios do modelo nas linhas.

Os oito estágios do modelo NAVE são assim definidos:

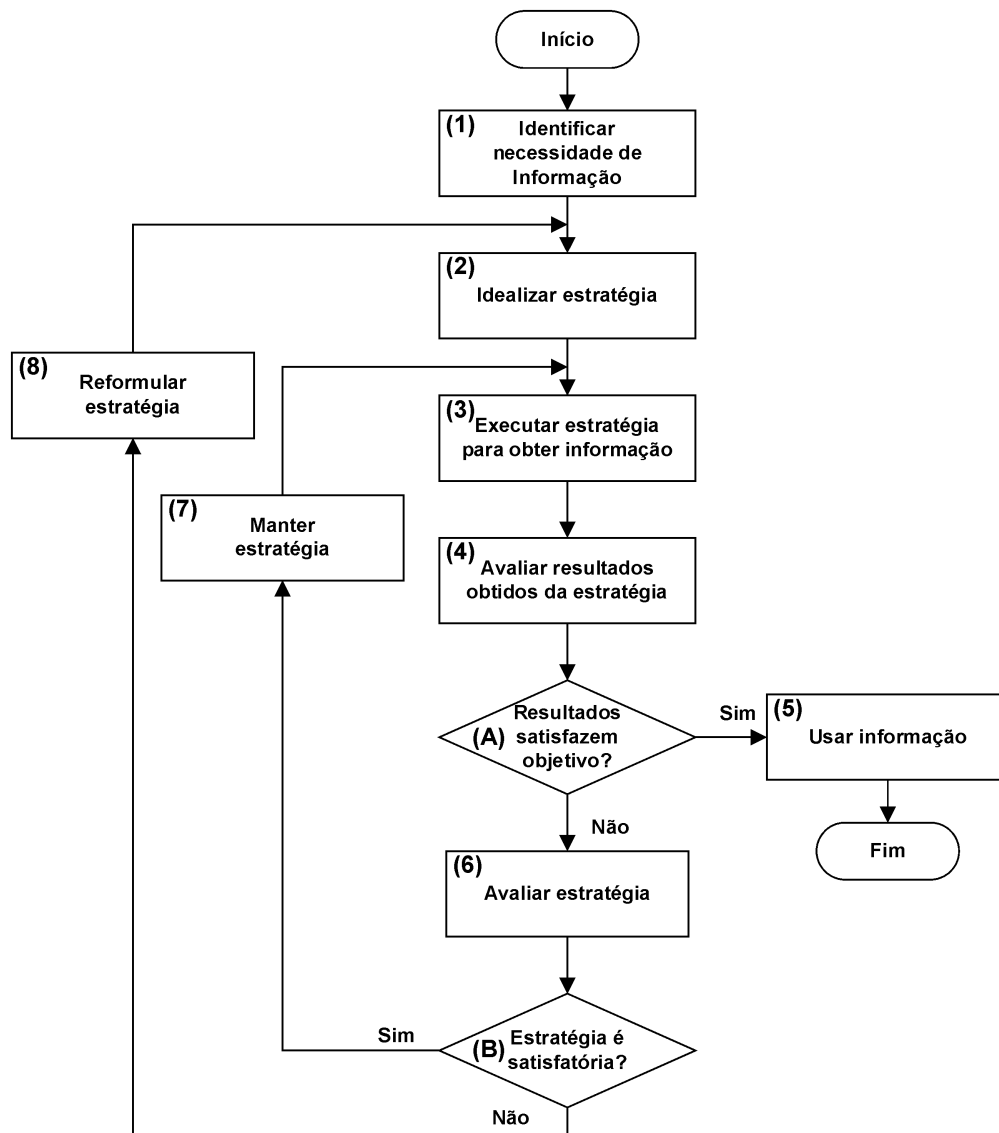


Figura 4.1: Modelo de Navegação Exploratória (NAVE)

1. *Identificar necessidade de informação (Quais são os meus objetivos?).* Motivado pela necessidade de informação, o usuário pode definir um ou mais objetivos (que podem ser alterados durante os estágios) e iniciar o processo do modelo NAVE para tentar satisfazer a sua necessidade de informação. O modelo não representa a formação e o desenvolvimento da necessidade de informação, por serem inerentes ao processo cognitivo do ser humano, mas considera implicitamente que as necessidades de informação são dinâmicas.

– Categorias: Não se aplica.

2. *Idealizar estratégia (Como eu poderia atingir os meus objetivos?).* Neste estágio, o usuário planeja o modo de abordar o espaço de informação e o domínio. A estrutura do espaço de informação pode ser considerada (por exemplo, os tipos de objetos disponíveis e como estão organizados), assim como as ferramentas disponíveis. O usuário pode dispor da opção de não idealizar qualquer estratégia

Tabela 4.1: Matriz dos estágios e decisões em relação às categorias de técnicas de navegação

Estágios	Categorias de Técnicas de Navegação				
	Busca ou recuperação Início	Especificação e seleção	Análise Visualização de contexto	Organização	Controle do processo
1: Identificar necessidade de informação					
2: Idealizar estratégia	X				X
3: Executar estratégia para obter informação	X	X	X	X	X
4: Avaliar resultados obtidos da estratégia			X	X	X
5: Usar informação					X
6: Avaliar estratégia			X	X	X
7: Manter estratégia					X
8: Reformular estratégia					X
Decisões					
A: Resultados satisfazem objetivo?			X	X	
B: Estratégia é satisfatória?			X	X	X

e se engajar em uma navegação serendipista².

– Categorias: Busca ou recuperação (início), Controle do processo.

3. *Executar estratégia para obter informação (Eu estou avançando na direção dos meus objetivos?)*. Neste estágio, o usuário efetua um ou mais passos de acordo com a estratégia idealizada. Inicialmente, o usuário pode contar com as ferramentas disponíveis para início do processo. Os usuários experientes podem executar vários passos de uma vez antes de passar para o próximo estágio.

– Categorias: Todas.

4. *Avaliar resultados obtidos da estratégia (Eu preciso de mais informação?)*. Neste estágio, o usuário avalia os resultados obtidos dos passos efetuados. O foco da avaliação são os resultados. Dessa maneira, a magnitude (quantidade), relevância (qualidade) e a natureza (mídia) dos resultados são avaliadas. O usuário pode selecionar itens individuais para que sejam examinados em detalhes. Os objetivos podem ser satisfeitos por meio do uso da informação obtida nos resultados. O usuário tenta compreender os resultados e visualizar os diferentes relacionamentos entre eles. A estratégia conduzida até este ponto pode também ser analisada, mas somente para tentar compreender como os resultados foram obtidos.

– Categorias: Análise (visualização e organização), Controle do processo.

²Serendipismo é a descoberta por acidente e sagacidade.

5. *Usar informação (Como eu poderia usar a informação obtida para satisfazer a necessidade de informação?).* Neste estágio, o usuário utiliza as facilidades fornecidas pelas ferramentas de controle do processo para encaminhar os resultados obtidos para o uso pretendido. A navegação passa a ser um aspecto secundário para o usuário até o aparecimento de uma nova necessidade de informação.

– Categorias: Controle do processo.

6. *Avaliar estratégia (A maneira como eu estou tentando atingir os meus objetivos está apresentando bons resultados?).* Neste estágio, o usuário avalia a estratégia. O foco são os resultados e a estratégia. Os resultados são confrontados com a estratégia utilizada para identificar problemas, por exemplo, aqueles relacionados com a sequência dos passos e os parâmetros utilizados (por exemplo, palavras-chave, facetas e instâncias).

– Categorias: Análise (visualização e organização), Controle do processo.

7. *Manter estratégia (Quais são os próximos passos da estratégia?).* Neste estágio, o usuário mantém a estratégia e retorna (estágio 3) para efetuar mais um passo da sua estratégia.

– Categorias: Controle do processo.

8. *Reformular estratégia (De que outra maneira eu poderia atingir os meus objetivos?).* Neste estágio, o usuário reformula a estratégia e retorna (estágio 2). O usuário tem a opção de idealizar uma estratégia inteiramente nova.

– Categorias: Controle do processo.

4.4.2 Decisões

As decisões também são definidas sob o ponto de vista do usuário. As categorias de técnicas de navegação (Subseção 2.2.2) também são citadas e essas informações constam na matriz mostrada na Tabela 4.1. As duas decisões do modelo NAVE são assim definidas:

A. *Resultados satisfazem objetivo?* A partir da avaliação realizada (estágio 4), o usuário deve decidir se os objetivos podem ser satisfeitos por meio do uso da informação obtida nos resultados. O usuário tem a opção de usar a informação (estágio 5) ou avaliar a estratégia (estágio 6).

– Categorias: Análise (visualização e organização).

B. *Estratégia é satisfatória?* A partir da avaliação realizada (estágio 6), o usuário deve decidir se a informação necessária pode ser obtida por meio da estratégia utilizada, isto é, se o processo avança na direção dos resultados adequados para

os objetivos. O usuário tem a opção de manter a estratégia (estágio 7) ou reformular a estratégia (estágio 8) parcialmente ou totalmente (por exemplo, reiniciar o processo).

– Categorias: Análise (visualização e organização), Controle do processo.

Apesar de não constar explicitamente no processo, o usuário pode abandonar a navegação a qualquer momento. O usuário é soberano sobre essa decisão.

4.4.3 Entidades

As entidades correspondem aos estados internos do modelo que podem ser alterados no transcorrer dos estágios.

- a. *Passo*. Um passo corresponde a uma ação ou atividade de *information searching* (no nível de movimento ou tática).
- b. *Estratagema*. Um estratagema corresponde a uma seqüência de passos que pode ser parte de uma estratégia ou até mesmo a própria estratégia.
- c. *Estratégia*. Uma estratégia corresponde a várias seqüências de passos (ou estratagemas) que são efetuados para satisfazer a necessidade de informação.
- d. *Resultados*. Um conjunto de objetos de informação fornecido pelo sistema de acordo com a estratégia realizada.
- e. *Coleção de objetos relevantes*. Um conjunto de objetos de informação armazenados pelo usuário para uso subsequente no processo ou para satisfazer a necessidade de informação.

4.5 Recomendações

No projeto e desenvolvimento de sistemas, o modelo NAVE pode servir de referência para a especificação das funcionalidades de um sistema e para sua interface. As entidades podem ser representadas através de componentes na interface, a qual pode fornecer suporte para o processo de navegação da maneira apresentada no modelo.

As recomendações formam um conjunto de orientações para o projeto e desenvolvimento de sistemas baseados no modelo NAVE. O objetivo é manter mínimo esse conjunto para evitar a sobreposição com outras recomendações, por exemplo, aquelas relacionadas com uma ferramenta específica para suporte à navegação. As recomendações apresentadas a seguir estão diretamente relacionadas com a dinâmica do processo subjacente ao modelo:

- R1. Os elementos do modelo (estágios, decisões e entidades) são referência para o desenvolvedor e não precisam ser explicitados ao usuário.

- R2. Ao usuário deve ser fornecido suporte a cada um dos estágios e decisões do processo sem impor qualquer seqüência de interação.
- R3. As categorias de técnicas de navegação devem ser cobertas pelo suporte fornecido ao usuário.
- R4. As ferramentas devem ser integradas ao sistema de maneira que os resultados obtidos em uma delas possam ser utilizados como entrada para outra.
- R5. Ao usuário deve ser permitido visualizar e manipular a seqüência de ações realizadas que é parte da sua estratégia.
- R6. Ao usuário deve ser oferecida liberdade de escolha das ferramentas que fazem parte de sua estratégia.

4.6 Considerações finais

Neste capítulo, um modelo de navegação exploratória foi proposto, o modelo NAVE, para auxiliar o projeto e desenvolvimento de sistemas de apoio a atividades de *information searching* na Web Semântica. Os conceitos subjacentes ao modelo e as características relacionadas com o seu desenvolvimento e aplicação foram apresentados, como também a descrição dos elementos que compõem o modelo.

O modelo foi utilizado como base para o desenvolvimento de um sistema, descrito no próximo capítulo, para verificar a viabilidade do modelo NAVE proposto.

Sistema de Navegação Exploratória

5.1 Considerações iniciais

Neste capítulo é apresentado o sistema de navegação exploratória desenvolvido no contexto deste trabalho, denominado de *Exploratory Navigation System*(ENS). O sistema ENS foi desenvolvido com o propósito de auxiliar a validação do modelo NAVE, proposto no Capítulo 4. Além disso, o ENS também tem auxiliado na verificação da aplicabilidade de extensões para as ferramentas utilizadas no sistema.

O capítulo está organizado da seguinte maneira: na Seção 5.2 são apresentadas as características da arquitetura e implementação do ENS e as decisões de projeto com relação à navegação e independência de domínio; na Seção 5.3 são descritos os sistemas de apoio; na Seção 5.4 é apresentada uma visão geral da interface gráfica do ENS; na Seção 5.5, cada uma das ferramentas do ENS é descrita, assim como são comentadas as técnicas subjacentes às ferramentas e as extensões desenvolvidas; na Seção 5.6 é apresentado um cenário de uso do ENS para procurar por informação; na Seção 5.7 são apresentadas as considerações finais.

5.2 Arquitetura e implementação

O sistema ENS, por ser uma aplicação web, foi concebido de acordo com a arquitetura cliente-servidor (Figura 5.1). Além disso, o padrão de projeto (*design pattern*) *Model-View-Controller* (MVC) [Buschmann et al. 1996] foi adotado na implementação, representado na Figura 5.1 pelos componentes Modelo, Visão e Controlador. O objetivo com o uso desse padrão é facilitar a separação entre a interface com o usuário e a lógica da aplicação (*separation of concerns*), assim como maximizar a manutenibilidade e a flexibilidade do código.

No lado da máquina cliente, as responsabilidades específicas de cada um dos componentes (Modelo, Visão e Controlador) são descritas conforme a seguir:

- *Modelo*: representar os estados do sistema para um usuário individual.

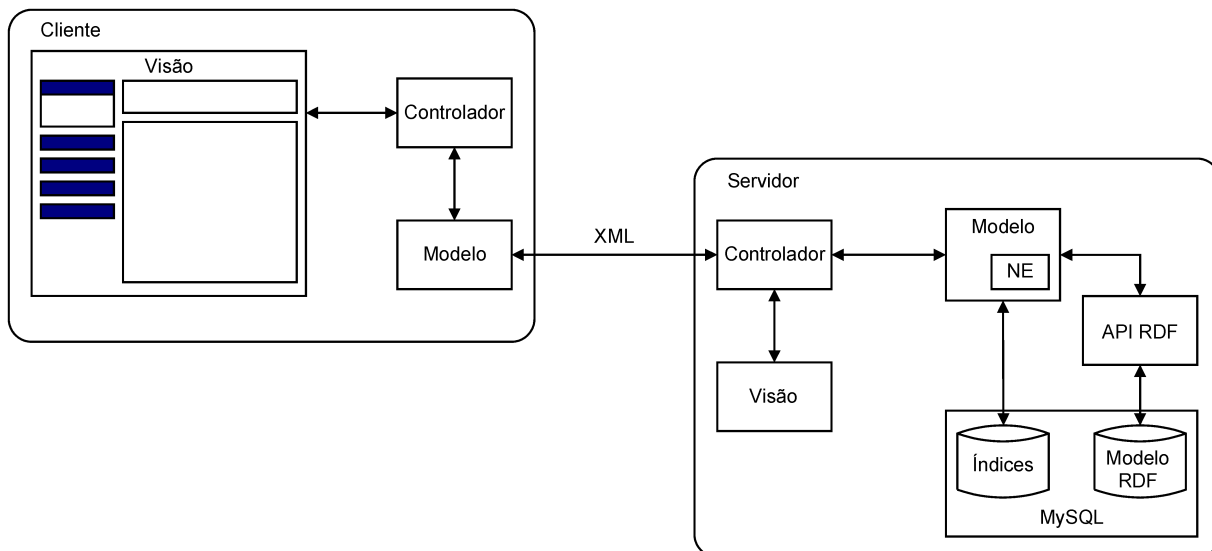


Figura 5.1: Arquitetura do ENS

- **Visão:** (a) fornecer uma interface visível para o usuário disparar eventos que serão tratados pelo Controlador; e (b) alterar a interface devido a mudanças ocorridas no Modelo, que são comunicadas pelo Controlador.
- **Controlador:** funcionar como intermediário entre Modelo e Visão. Este componente consiste no código responsável por encaminhar cada evento do usuário ao módulo adequado do Modelo para processamento. O resultado é encaminhado pelo Controlador ao componente Visão para a modificação da informação apresentada na interface.

No lado da máquina servidor, as responsabilidades específicas de cada um dos componentes (Modelo, Visão e Controlador) são descritas conforme a seguir:

- **Modelo:** (a) fornecer acesso aos dados do domínio armazenados na base de dados (representados na Figura 5.1 pelo Índice e Modelo RDF); e (b) processar as solicitações do usuário. O Índice é uma representação para os documentos de uma coleção. O Modelo RDF é uma representação para grafos RDF e os elementos deste modelo são declarações RDF ou triplas (sujeito, predicado e objeto). O acesso ao Modelo RDF é realizado por meio de uma *Application Programming Interface* (API) para a manipulação de grafos RDF (representada na Figura 5.1 pela API RDF). A *Navigation Engine* (NE), incluída no Modelo e descrita na Subseção 5.2.1, é responsável pelo processamento das ações do usuário.
- **Visão:** gerar o formato da resposta.
- **Controlador:** funcionar como intermediário entre Modelo e Visão. Este componente consiste no código responsável por encaminhar cada solicitação ao módulo adequado do Modelo para processamento. O resultado é encaminhado pelo Controlador ao componente Visão para a geração da resposta.

Na implementação do lado cliente foram utilizadas as linguagens JavaScript e *Extensible HyperText Markup Language* (XHTML). No lado servidor foi utilizada a linguagem PHP¹, o gerenciador de base de dados MySQL² e a biblioteca *RDF API for PHP* (RAP)³ [Oldakowski et al. 2005]. A biblioteca RAP é um conjunto de ferramentas (*toolkit*) de código aberto (*open source*) para a linguagem PHP que fornece uma *Application Programming Interface* (API) para análise (*parse*), manipulação, armazenamento e consulta a grafos RDF.

O uso da técnica *Asynchronous Javascript And XML* (AJAX) [Garret 2005], para permitir melhor interação do usuário com a interface, adicionou complexidade no código do lado cliente devido ao tratamento dessa interação e ao processamento da comunicação com o lado servidor. Para uma aplicação funcionar com AJAX é necessário manter um modelo de dados do domínio no cliente (para responder rapidamente à interação do usuário) e no servidor (para acesso aos dados armazenados). Visualmente, o uso dessa técnica permite a construção de uma interface no Cliente baseada em *widgets*⁴ em vez da interface tradicional, baseada em documentos. No caso do ENS, as interfaces das ferramentas (descritas na Seção 5.5) foram desenvolvidas para funcionar como *widgets*.

A comunicação entre os subsistemas Modelo, no Cliente, e o Controlador, no Servidor, tem papel importante na arquitetura. As mensagens são trocadas de maneira assíncrona usando o objeto `XMLHttpRequest` da linguagem JavaScript, e são estruturadas usando a linguagem XML. Nesse caso, pode-se dizer que a abordagem é orientada a dados: o cliente (subsistema Modelo) envia os dados da solicitação no formato XML, que são processados (*parsed*) no servidor; e o servidor (subsistema Controlador) retorna os dados de resposta também no formato XML, que são processados (*parsed*) no cliente e utilizados para atualizar os estados do sistema e/ou a interface do usuário.

Na Figura 5.2 é mostrado um exemplo de solicitação. Uma solicitação é composta do passo a ser executado, da seqüência de passos executada até o momento (isto é, a estratégia atual), e da lista das referências (*links*) para as páginas encontradas que são resultadas da estratégia atual. Nesse exemplo, o usuário escolheu utilizar a ferramenta de busca por palavra-chave (linhas 4-7) depois de ter utilizado as ferramentas de navegação por facetas (linhas 12-15) e de agrupamento (linhas 16-18). Para simplificar o exemplo, apenas duas referências são mostradas na lista (linhas 21-27).

Na Figura 5.3 é mostrado um exemplo de resposta. Uma resposta é composta principalmente da nova lista das referências (*links*) para as páginas (linhas 2-20). Cada item contém todos os dados necessários para a construção da parte da in-

¹<http://www.php.net/>

²<http://www.mysql.com/>

³[\(http://sourceforge.net/projects/rdfapi-php/](http://sourceforge.net/projects/rdfapi-php/)

⁴O termo *widget* é usado para fazer referência a um componente gráfico de uma *Graphical User Interface* (GUI) ou ao seu programa de controle, ou ainda para fazer referência à combinação de ambos.

```

01<request>
02  <step>
03    <strategy>
04      <action tool="keywordSearch">
05        <option name="words" value="estruturas de dados"></option>
06        <option name="operator" value="and"></option>
07      </action>
08    </strategy>
09  </step>
10  <history>
11    <strategy>
12      <action tool="facetBrowse">
13        <option name="facet" value="source"></option>
14        <option name="instance" value="coteia.icmc.usp.br"></option>
15      </action>
16      <action tool="group">
17        <option name="facet" value="course"></option>
18      </action>
19    </strategy>
20  </history>
21  <uriList>
22    <item uri="http://coteia.icmc.usp.br/coteia/mostra.php?ident=161.2" rank="">
23    </item>
24    <item uri="http://coteia.icmc.usp.br/coteia/mostra.php?ident=161.3" rank="">
25    </item>
26    ...
27  </uriList>
28</request>

```

Figura 5.2: Exemplo de mensagem de solicitação (*request*) enviada pelo cliente

terface correspondente à área de resultados. Além disso, a resposta também pode conter outras informações necessárias para a atualização da interface das ferramentas. Nesse exemplo, a lista de facetas da ferramenta deve ser atualizada com as novas informações com relação a *Source*, pois no primeiro passo da estratégia o repositório “2004-2006” foi selecionado e a quantidade de itens reduzida para 113 (linha 23) depois de executada a busca por palavra-chave.

```

01<response>
02  <uriList>
03    <item uri="http://coteia.icmc.usp.br/coteia/mostra.php?ident=24.9" rank="1">
04      <title>Bibliografia</title>
05      <summary>SZWARCFITER, J. L.; MARKENZON, L. Estruturas de Dados e
06      seus Algoritmos,
07      J. L.; MARKENZON, L. Estruturas de Dados e seus Algoritmos, Livros
08      Técnicos e
09      </summary>
10    </item>
11    <item uri="http://coteia.icmc.usp.br/coteia/mostra.php?ident=24.2" rank="2">
12      <title>Programa do Curso</title>
13      <summary> registros, cadeias de caracteres, estruturas dinâmicas
14      (ponteiros), e
15      Aula 02 - Variáveis e Tipos de Dados básicos em C . Aula 03 - Mais
16      sobre
17      </summary>
18    </item>
19    ...
20  </uriList>
21  <facetList>
22    <facet name="source" label="Repository">
23      <instance name="coteia.icmc.usp.br" label="2004-2006" occurrences="113">
24      </instance>
25    </facet>
26  </facetList>
27</response>

```

Figura 5.3: Exemplo de mensagem de resposta (*response*) enviada pelo servidor

A maior parte do processamento das funções de apresentação é mantida no navegador e desenvolvida na linguagem JavaScript. Por essa razão, o subsistema Visão no Servidor tem apenas a função de gerar o formato da resposta. Geralmente, uma aplicação web comum tem pelo menos duas camadas principais: a camada de negócio, que modela o domínio (e realiza o acesso à base de dados); e a camada de apresentação, que apresenta ao usuário os dados da camada de negócio. A função do navegador é de somente mostrar a aparência da página. O uso de AJAX permitiu que a maior parte das funções de apresentação atribuída ao servidor fosse movida para o cliente. No entanto, a adição de uma camada AJAX acarretou no aumento da carga de processamento devido ao *parse* realizado nas mensagens (por exemplo, aquelas mostradas nas Figuras 5.2 e 5.3), tanto no cliente quanto no servidor.

Nas subseções a seguir são descritas as duas principais decisões de projeto que possibilitaram flexibilidade e extensibilidade do sistema ENS, visando se adequar ao modelo NAVE para atender os estágios definidos.

5.2.1 *Navigation Engine*

O mecanismo do sistema ENS que representa a sua funcionalidade principal é a *Navigation Engine* (NE). Nesse mecanismo estão integradas as diferentes ferramentas de navegação que compõem o ENS. Essas ferramentas podem ser escolhidas e são descritas em detalhes na Seção 5.5.

A NE é parte do subsistema Modelo no lado servidor da arquitetura apresentada na Figura 5.1. A função principal da NE é processar o novo passo (de uma estratégia) contido em uma solicitação (por exemplo, aquela da Figura 5.2), além de computar o novo conjunto de resultados (com base no conjunto atual contido na solicitação).

A NE também é capaz de processar uma estratégia completa, isto é, uma sequência composta de vários passos. Na versão atual, essa funcionalidade é utilizada para implementar o recurso de desfazer um passo da estratégia. Ao invés de manter os estados anteriores e permitir desfazer somente a última ação executada, a interface do ENS permite desfazer qualquer passo da estratégia, inclusive um passo intermediário. Quando uma ação é desfeita, o que realmente ocorre é uma solicitação para processar uma estratégia completa, mas sem o passo referente àquela ação. Essa funcionalidade é importante para, por exemplo, permitir ao usuário colecionar estratégias bem sucedidas e submeter/repetir uma estratégia periodicamente em uma mesma fonte de informação ou para outras.

A NE também pode determinar qual o modo de visualização dos resultados:

- Simples: determinado pela última ferramenta.
- Complexo: combinação de todas as ferramentas.
- Otimizado: personalizado de acordo com as preferências do usuário.

A informação sobre o modo de visualização dos resultados é utilizada pelo compo-

nente Visão na geração do formato da resposta. Atualmente, somente o modo simples está disponível na versão implementada.

5.2.2 Independência de domínio

Uma importante decisão de projeto para o sistema ENS foi que possuisse características genéricas e independência do domínio de aplicação. O ENS foi configurado para permitir a pesquisa de informação nos repositórios de páginas wiki da CoTeia (domínio educacional). Esta subseção tem por objetivo descrever as características e configurações que permitem algum grau de independência de domínio.

As configurações básicas são especificadas em um arquivo de inicialização (`config.ini`), por exemplo, a configuração do nome das bases de dados criadas no MySQL para o Índice, assim como o nome do usuário e a senha de acesso. As bases de dados são criadas usando um arquivo (`databases.sql`) com comandos da linguagem *Structured Query Language* (SQL). O Modelo RDF é criado por meio de um programa PHP que utiliza o suporte fornecido pela RAP. Outras configurações úteis de nível básico incluídas no arquivo de inicialização são: a indicação se a função de registrar o *log* da interação está ativada e o arquivo no formato *Comma Separated Value* (CSV)⁵ utilizado para armazenar os registros da interação.

As partes da interface que são dependentes do domínio da aplicação também podem ser configuradas no arquivo de inicialização. Por exemplo: os metadados utilizados nas ferramentas para navegação por facetas, classificação e agrupamento; e as entidades e padrões utilizados nas ferramentas para marcar resultados e realce de termos. Essas configurações podem ser automatizadas facilmente a partir das informações armazenadas no Modelo RDF. A exceção são os padrões que não podem ser armazenados no modelo porque são definidos através de expressões regulares, por exemplo, os padrões para e-mails, datas e horários. Essas configurações poderiam ser incluídas no arquivo de inicialização, mas ficariam separadas das outras configurações associadas às ferramentas. Atualmente, essas configurações são especificadas diretamente no código (*hard coded*) do subsistema Modelo no lado servidor.

O ENS permite a navegação em mais de um repositório da CoTeia. Neste caso, o repositório desejado pode ser “selecionado” na ferramenta para navegação por facetas por meio do elemento de metadados *Source* (do padrão Dublin Core [DCMI 2003]), que é definido como “o recurso a partir do qual o recurso descrito é derivado”. Nesse caso, cada repositório no todo é considerado como um recurso, assim como as páginas wiki nele contidas também são individualmente recursos.

Para garantir a independência de domínio com relação aos repositórios (da CoTeia), o ENS não é acoplado aos sistemas responsáveis pela “alimentação” das bases de dados (Índice e Modelo RDF), denominados de sistemas de apoio. Os sistemas de

⁵*Comma Separated Value* (CSV) é um formato de arquivo utilizado freqüentemente para o intercâmbio de dados com sistemas legados (*legacy systems*). A linguagem PHP tem funções específicas para tratar o formato, e o CSV também é um padrão de arquivo reconhecido pelo Microsoft Excel.

apoio e suas características genéricas são descritos brevemente na próxima seção.

5.3 Sistemas de apoio

Os sistemas de apoio são responsáveis pelo processamento necessário para inserir os dados referentes às páginas wiki nas bases de dados (Índice e Modelo RDF) utilizadas pelo sistema ENS. A partir desse processamento, o ENS trata as páginas wiki como recursos, segundo o conceito de recurso utilizado pelos padrões da arquitetura da Web Semântica (Figura 2.3).

Os sistemas de apoio, mostrados na Figura 5.4, foram implementados na linguagem PHP e são descritos nas subseções seguintes: o sistema *Crawler* é descrito na Subseção 5.3.1, o *Indexer* na Subseção 5.3.1 e o *Metadata Generator* na Subseção 5.3.1.

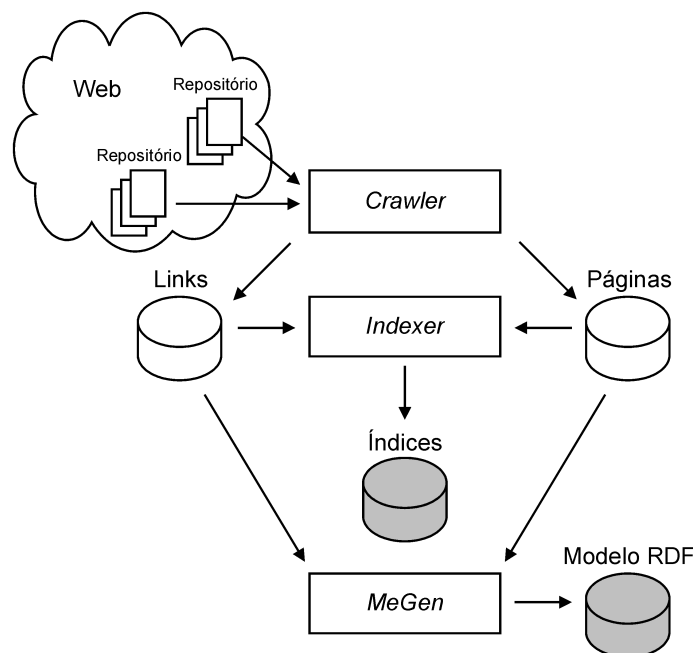


Figura 5.4: Arquitetura dos sistemas de apoio

5.3.1 Crawler

Um *crawler* (também chamado de *spider* ou *robot*) é um programa para “navegar” a Web, isto é, para percorrer recursivamente os links contidos em um conjunto de páginas iniciais [Baeza-Yates e Ribeiro-Neto 1999; Pokorný 2004]. Em geral, as páginas recuperadas são armazenadas em um repositório de páginas.

O sistema *Crawler* da arquitetura mostrada na Figura 5.4 é um “*site crawler*”, isto é, consiste em um *crawler* que não percorre os links externos contidos nas páginas. Dessa maneira, o processo fica restrito ao escopo dos *websites* especificados pelo conjunto de páginas iniciais. Esse comportamento do programa é adequado para processar os repositórios de páginas wiki da CoTeia. Os links são armazenados em

uma base de dados de links e cópias das páginas são armazenadas em um diretório para o uso posterior pelos sistemas *Indexer* e *MeGen*.

O programa permite dois modos de processamento, que são importantes para a realização de testes:

- modo normal, no qual todos os registros da base de dados de links são excluídos antes do processamento; e
- modo incremental, no qual os registros não são excluídos para permitir o processamento parcial de um conjunto de sites.

As configurações básicas são especificadas em um arquivo de inicialização (`config.ini`), por exemplo, a configuração do nome da base de dados criada no MySQL para o armazenamento dos links, assim como o nome do usuário e a senha de acesso. Outras configurações que podem ser realizadas incluem: lista de páginas iniciais; lista de links que devem ser ignorados pelo programa (por exemplo, para JavaScript, e-mail, e arquivos); diretório de armazenamento das páginas; e modo de processamento.

5.3.2 *Indexer*

No contexto da arquitetura mostrada na Figura 5.4, o *Indexer* processa as páginas coletadas pelo *Crawler*. Basicamente, o programa *Indexer* analisa o conteúdo de cada documento, remove as palavras muito comuns (*stopwords*⁶) e indexa as demais palavras sem diferenciar letras maiúsculas e minúsculas (*case insensitive*). O *Indexer* cria um índice invertido (*inverted index*) que contém uma lista de referências para cada palavra encontrada nas páginas.

As configurações básicas são especificadas em um arquivo de inicialização (`config.ini`). Da mesma maneira que no *Crawler*, esse arquivo contém a configuração do nome da base de dados criada no MySQL para o armazenamento dos links, assim como o nome do usuário e a senha de acesso. Além disso, também é possível especificar o tamanho mínimo que uma palavra deve ter para ser indexada e uma lista de arquivos com as *stopwords*. A versão atual do *Indexer* utiliza dois arquivos de *stopwords*, um arquivo com as palavras do idioma Português e outro com as do Inglês.

5.3.3 *Metadata Generator*

No contexto da arquitetura mostrada na Figura 5.4, o gerador de metadados, denominado de *Metadata Generator* (MeGen) [Pansanato e Fortes 2006], é um sistema para a geração de metadados que implementa diversos métodos de acordo com as seguintes categorias de métodos para a geração automática de metadados [Pansanato e Fortes 2005a,b]:

⁶<http://www.ranks.nl/stopwords/>

- *Geração de Metadados Baseada no Recurso*: os métodos dessa categoria implementam a geração com base na análise do conteúdo do próprio recurso. No caso da CoTeia, os recursos são as páginas wiki. Nesse sentido, podem ser utilizadas técnicas típicas, por exemplo: classificação de idioma (*language classification*) e extração de palavras-chave (*keyword extraction*).
- *Geração de Metadados Baseada no Contexto*: os métodos dessa categoria implementam a geração com base na informação do contexto no qual o recurso é usado. A informação de contexto pode ser obtida a partir da estrutura dos recursos (por exemplo, estrutura de composição), do gerenciamento dos usuários relacionados ao recurso e a partir do sistema de arquivos ou outro software responsável pelo gerenciamento do armazenamento do recurso.
- *Geração de Metadados Baseada em Template*: os métodos dessa categoria implementam a geração com base na informação obtida a partir de *templates* ou *profiles* associadas ao recurso, e são considerados semi-automáticos porque requerem a intervenção do usuário em algum momento. As *templates* podem ser mantidas no nível de sistema, de grupo de recursos e no nível de usuário. A idéia é manter um conjunto de metadados válido para todos os recursos criados (sistema), um conjunto válido para um grupo de recursos e um conjunto válido para todos os recursos criados por um mesmo usuário.
- *Geração de Metadados Baseada em Regra*: os métodos dessa categoria implementam a geração com base em regras que permitem explorar relacionamentos entre os metadados do recurso. Por exemplo, determinados metadados podem ser derivados de outros metadados. Isso é especialmente indicado quando é necessário mapear um elemento de metadados local para algum elemento de um esquema padronizado ou quando existe uma regra heurística conhecida que pode ser aplicada.

O MeGen utiliza os links e as páginas armazenadas pelo *Crawler* no processo de geração automática dos metadados. No arquivo de configuração (*config.ini*) do MeGen são especificadas as diretivas necessárias para especializar os métodos implementados para um determinado domínio. Por exemplo, podem ser especificados os arquivos com tabelas de palavras-chave (códigos, nomes, etc.), as informações de contexto e as regras de mapeamento.

A hierarquia de classes simplificada do MeGen é mostrada na Figura 5.5. O diagrama de classe fornece uma visão geral do gerador de metadados apresentando a classe *MetadataGenerator* como uma superclasse das classes que correspondem às categorias de métodos descritos anteriormente nesta subseção. A partir desse diagrama pode-se notar que alguns elementos de metadados podem ser gerados por dois ou mais métodos (operações), por exemplo, o tamanho (*Extent*) e o título (*Title*) de uma página wiki. Nesse caso o método definido para manipular conflitos (especificado pelo

método `setConflictHandler()` deve ser capaz de resolver quais valores devem ser considerados.

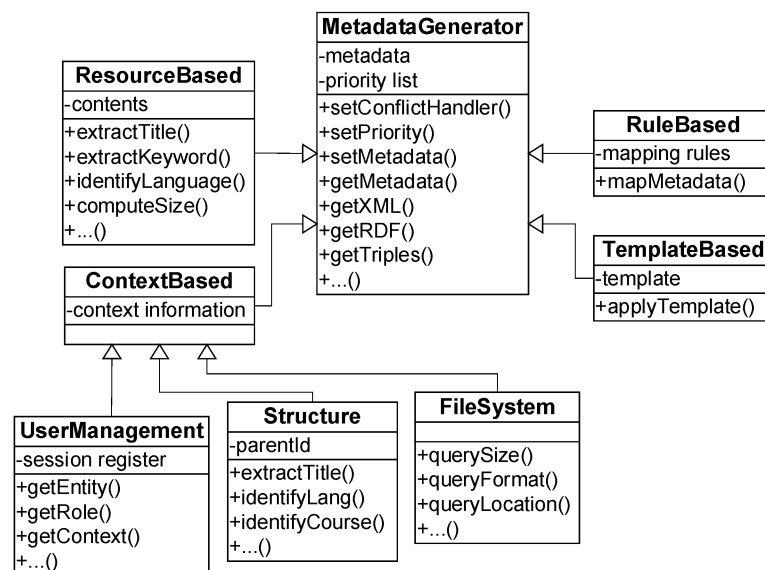


Figura 5.5: A hierarquia de classes simplificada do MeGen

O registro de metadados (*metadata record*) gerado pelo MeGen para cada página wiki da CoTeia é um *application profile* [Duval e Hodgins 2003] baseado no padrão LOM [IEEE LTSC 2002] e no padrão DC [DCMI 2003], ou seja, o esquema de metadados usado na CoTeia contém elementos desses dois padrões. Além disso, a esse *application profile* também foram adicionados cinco elementos de metadados para fornecer suporte a um domínio de aplicação específico; o de páginas wiki educacionais no ICMC/USP. A idéia é fornecer metadados adicionais com vocabulário conhecido pela comunidade do ICMC/USP para melhor apoiar os usuários da CoTeia na descoberta de páginas wiki (recursos educacionais). Na Tabela 5.1 são apresentados os elementos de metadados do *application profile* da CoTeia. Nessa tabela também é mostrada a sobreposição da semântica dos elementos dos padrões DC e LOM e dos novos elementos (Local). Cada registro de metadados é codificado em RDF segundo diretrizes específicas [DCMI 2002; Nilsson et al. 2003] e armazenado no Modelo RDF (Figura 5.4) no formato de triplas [W3C 2004e].

O MeGen foi utilizado para gerar os elementos de metadados do *application profile* da CoTeia de 4.234 páginas wiki, divididas em dois repositórios. Na Tabela 5.2 são mostrados dados referentes a esses repositórios: o nome pelo qual o repositório é conhecido, o período principal que as páginas foram criadas, a URL de acesso e a quantidade de páginas wiki disponíveis. Aproximadamente 84,9% de todos os valores foram gerados pela versão do MeGen utilizada. Pansanato e Fortes [2006] descrevem os detalhes da aplicação do MeGen aos repositórios da CoTeia e apresentam algumas análises com relação aos resultados obtidos.

Tabela 5.1: Um *application profile* para páginas wiki educacionais

LOM	Dublin Core	Local
General.Title	Simple DC: title	-
General.Language	Simple DC: language	-
General.Description	Simple DC: description	-
General.Keyword	Simple DC: subject	-
LifeCycle.Version	-	-
LifeCycle.Contribute.Entity (Role = author)	Simple DC: creator	-
LifeCycle.Contribute.Entity (Role = publisher)	Simple DC: publisher	-
LifeCycle.Contribute.Date	Simple DC: date	-
Technical.Format	Simple DC: format	-
Technical.Size	Qualified DC: extent	-
Technical.Location	-	-
Educational.LearningResourceType	Simple DC: type	Learning Material Type
Educational.Context	-	Brazilian Context
Educational.TypicalAgeRange	Qualified DC: audience	-
-	Simple DC: source	-
-	-	Knowledge Area*
-	-	Course
-	-	Learning Activity Type

*Anteriormente denominado de Topic, este elemento de metadados foi renomeado.

Tabela 5.2: Repositórios utilizados para avaliar o MeGen

Nome	Período	URL	Quantidade de páginas
CoTeia	2004-2006	http://coteia.icmc.usp.br/coteia/	1.731
CoWeb	2001-2003	http://safedevl.icmc.usp.br/coweb/	2.503
Total:			4.234

O MeGen também é responsável pela alimentação do Modelo RDF com conhecimento sobre nomes de pessoas e locais, com relação à comunidade do ICMC/USP. Esses metadados são coletados/extraídos do website do ICMC/USP. Algumas das ferramentas descritas na Seção 5.5 utilizam esse conhecimento na sua funcionalidade.

5.4 Visão geral da Interface gráfica

O projeto da interface do sistema ENS iniciou com a análise de usuários e tarefas (que constitui a análise de requisitos) e foi conduzido considerando um processo iterativo no qual cada etapa apresentou evoluções a partir da etapa anterior [Dix et al. 2004]. Cada etapa envolveu a especificação de funcionalidades segundo o modelo NAVE proposto no Capítulo 4, a prototipação da interface (que possibilitava realizar a interação de acordo com o modelo especificado) e a avaliação da interface. A partir dessa avaliação seguia uma nova etapa de especificação, prototipação e avaliação. A avaliação era realizada por meio de testes de usabilidade informais [Preece

et al. 2002], com ênfase em contribuições rápidas ao invés de testes formais, devido à característica de evolução dinâmica para identificação das interações possíveis na interface. Além disso, eram também comuns as sessões de *brainstorming* com especialistas em usabilidade (alunos da pós-graduação do ICMC/USP envolvidos com desenvolvimento de aplicações web), as quais foram muito importantes para atingir a aparência e o modo de atuar (*look and feel*) da interface atual.

A evolução do *layout* da interface é ilustrada na Figura 5.6 através da apresentação das interfaces obtidas durante as etapas do processo de projeto do ENS. As interfaces mostradas nessa figura correspondem à primeira interface proposta (**a**), a interfaces intermediárias (**b**, **c**), e à interface atual (**d**).



Figura 5.6: Evolução da interface do ENS

A interface do ENS é formada por quatro partes (Figura 5.7): o conjunto de ferramentas à esquerda (**A**), o histórico de ações na parte superior-direita (**B**), e as áreas de resultados (**C**) e de conteúdo (**D**) na parte inferior-direita. Na área de resultados é apresentada uma lista com os itens encontrados como resposta às ações tomadas pelo usuário; o título de cada item da lista é um link para o conteúdo. Na área de conteúdo é apresentado o conteúdo de um item selecionado pelo usuário a partir da área de resultados.

As entidades do modelo NAVE (Subseção 4.4.3) estão representadas na interface do ENS. O conjunto de ferramentas (**A**) é utilizado para efetuar passos, que formam o

estratagema (seqüência de passos) mostrado no histórico de ações **(B)**. O estratagema é parte da estratégia do usuário para procurar por informação. A área de resultados **(C)** mostra os resultados da seqüência de passos efetuados pelo usuário. A área de conteúdo **(D)** apresenta o conteúdo de um resultado selecionado pelo usuário. Os resultados da área de resultados **(C)** também podem ser selecionados pelo usuário para fazer parte de uma coleção de objetos relevantes por meio do uso de uma ferramenta específica (Subseção 5.5.3).

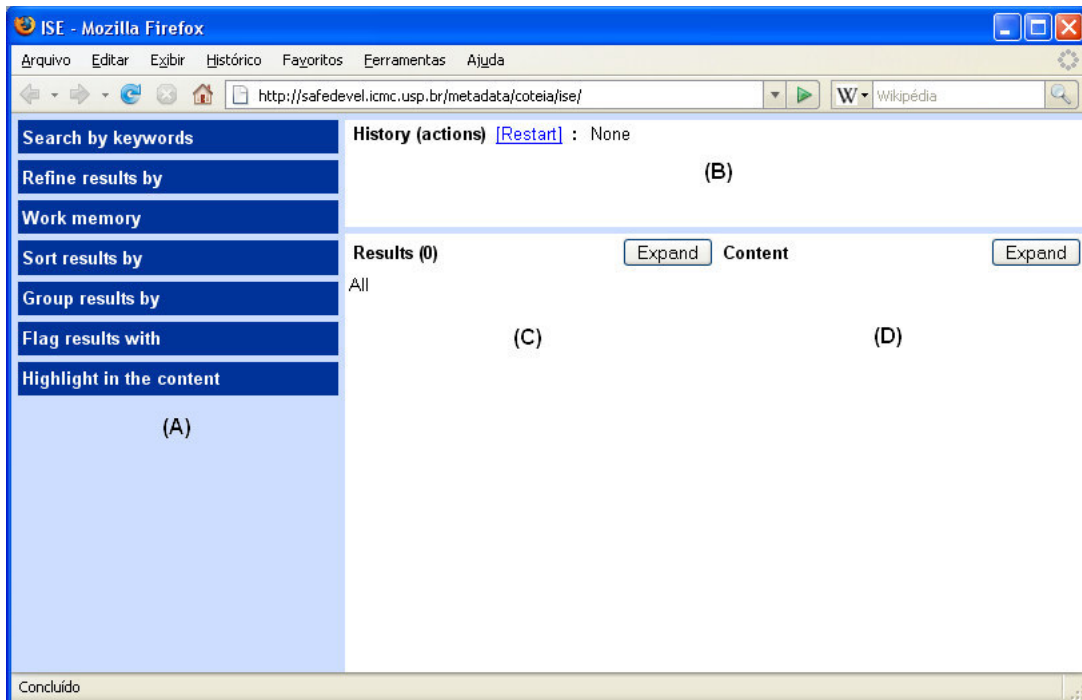


Figura 5.7: Interface gráfica do ENS

Os botões “*Expand*” disponíveis nas áreas de resultados **(C)** e de conteúdo **(D)** servem para estender uma das áreas para ocupar o espaço da outra. Depois de pressionado, esse botão é alterado para um botão “*Split*” para permitir que a interface possa ser restaurada ao seu estado original, ou seja, com as duas áreas visíveis.

A barra de título de cada uma das ferramentas disponíveis à esquerda funciona da mesma maneira que *toggle buttons*. Um *toggle button* funciona como um botão comum exceto pelo fato de que ele conserva o seu estado pressionado. Ao pressionar uma barra de título surge a interface da ferramenta correspondente, e essa interface desaparece quando a barra de título é pressionada novamente. A interface de cada uma das ferramentas é apresentada a seguir (Seção 5.5), assim como detalhes de funcionamento e implementação.

5.5 Ferramentas

A maturidade da interface e a familiaridade do usuário com a funcionalidade da ferramenta estão entre as barreiras que devem ser superadas para que novas ferra-

mentas para navegação sejam adotadas amplamente por desenvolvedores. As lições aprendidas durante este trabalho foram de que o desenvolvimento de idéias para ferramentas novas como extensões daquelas usadas freqüentemente pode auxiliar na superação dessas barreiras. Nesse sentido, o ENS é uma oportunidade para testar a aplicação de algumas idéias simples, mas que se mostram úteis nas atividades do usuário que envolve pesquisa de informação (Capítulo 6).

Na Tabela 5.3, as ferramentas que fazem parte da interface da versão atual do ENS são categorizadas de acordo com a proposta para classificação de técnicas de navegação apresentada na Subseção 2.2.2. Adicionalmente, por meio da comparação da Tabela 5.3 com a matriz mostrada na Tabela 4.1 (Subseção 4.4.1) é possível verificar que o sistema ENS fornece ferramentas para suporte a cada um dos estágios e decisões do processo de navegação proposto pelo modelo NAVE.

Uma ferramenta pode pertencer a mais de uma categoria devido às suas características próprias (técnica) ou de acordo com a implementação. Na Tabela 5.3, por exemplo, a ferramenta “Navegação por facetas” é classificada em mais de uma categoria porque a visualização das instâncias organizadas por facetas pode auxiliar o usuário na decisão de como iniciar a exploração, além de permitir a seleção de uma instância para filtrar a lista de resultados.

Tabela 5.3: Classificação das ferramentas da interface do ENS

Ferramentas	Categorias de Técnicas de Navegação			
	Busca ou recuperação <i>Início</i>	Especificação e seleção	Análise <i>Visualização de contexto</i>	Organização Controle do processo
Busca por palavra chave		X		
Navegação por facetas	X	X		
Memória de trabalho				X
Classificação				X
Agrupamento				X
Marcar resultados			X	
Realce de termos			X	
Histórico				X

As ferramentas são descritas nas próximas subseções na seguinte sequência: Busca por palavra-chave (Subseção 5.5.1), Navegação por facetas (Subseção 5.5.2), Memória de trabalho (Subseção 5.5.3), Classificação (Subseção 5.5.4), Agrupamento (Subseção 5.5.5), Marcar resultados (Subseção 5.5.6), Realce de termos (Subseção 5.5.7) e Histórico (reiniciar e desfazer) (Subseção 5.5.8).

5.5.1 Busca por palavra-chave

A ferramenta Busca por palavra-chave é uma máquina de busca (*search engine*) de arquitetura *crawler-indexer* centralizada [Baeza-Yates e Ribeiro-Neto 1999; Pokorný 2004], isto é, em um mesmo sistema local, um programa (*crawler*) envia solicitações aos servidores Web para recuperar os documentos e outro programa (*indexer*) realiza a indexação destes para a consulta posterior a partir de uma interface. Dessa maneira, a ferramenta possui duas partes distintas: uma que consiste nos módulos *Crawler* e *Indexer* (descritos anteriormente na Seção 5.3) e outra parte, que consiste na interface do usuário e um programa para processar a consulta (geralmente denominado de *query engine*).

A interface apresentada na Figura 5.8 consiste em uma caixa de texto na qual a consulta pode ser especificada por meio de uma sequência de palavras-chave. Além disso, é possível selecionar o comportamento do programa de consulta (*query engine*) por meio de botões, entre a opção de considerar os documentos que contêm todas as palavras especificadas (AND) e a de considerar aqueles com qualquer uma das palavras (OR). As máquinas de busca tradicionais geralmente oferecem essa opção somente nas interfaces de pesquisa avançada. O programa de consulta fornece uma lista dos documentos nos quais as palavras-chave especificadas foram encontradas, que é apresentada na área de resultados. Para simplificar a implementação, na versão atual não foi incluído suporte para operadores booleanos, pesquisa por frase (*phrase search*) e para caracteres-coringa (*wildcards characters*).

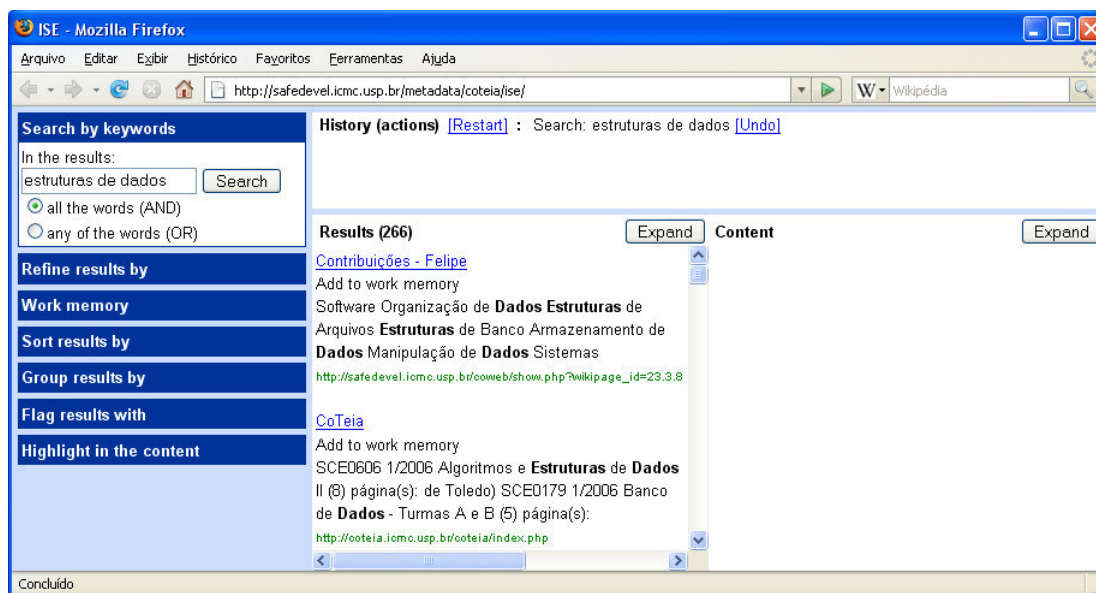


Figura 5.8: Interface da ferramenta Busca por palavra-chave no ENS

Apesar do mecanismo adotado ser bastante simples, a ferramenta funciona bem para o propósito de obter uma lista das páginas wiki relacionadas com as palavras-chave especificadas pelo usuário. As páginas dessa lista são classificadas de acordo

com a frequência das palavras-chave especificadas pelo usuário, isto é, as páginas que contêm o maior número de palavras-chave são consideradas mais relevantes. As máquinas de busca tradicionais podem utilizar uma classificação de relevância que muitos usuários podem ficar confusos: algumas páginas podem ter classificação alta mesmo quando elas não possuem todas as palavras-chave. Isso ocorre quando as máquinas de busca consideram a popularidade de uma página para melhorar a sua classificação.

5.5.2 Navegação por facetas

A navegação por facetas (*faceted browsing*) é uma maneira de utilizar metadados facetados (*faceted metadata*) para permitir aos usuários encontrar informação. Os metadados podem ter várias facetas, isto é, atributos em diversos conjuntos ortogonais de categorias. Por exemplo, no domínio de páginas wiki educacionais, algumas facetas possíveis poderiam ser autor (professor ou aluno) e disciplina. O uso de categorias hierárquicas para a navegação foi popularizado por *websites* semelhantes ao Yahoo! Directory⁷ e ao Open Directory Project⁸. O Epicurious⁹ é um exemplo de *website* comercial que utiliza a navegação por facetas.

Uma ferramenta de navegação por facetas permite aos usuários filtrar um conjunto de itens (por exemplo, páginas wiki ou outro tipo de objeto de informação) por meio da seleção progressiva somente dos valores válidos (instâncias) das facetas. A lista de valores válidos é filtrada para apresentar somente aqueles que possuem itens disponíveis. Portanto, é impossível obter uma lista de itens vazia como resultado da seleção de uma instância.

A interface da ferramenta Navegação por facetas apresentada na Figura 5.9 consiste na lista de facetas disponíveis e links para expandir (*expand*) ou contrair (*collapse*) as instâncias de cada uma das facetas. Ao lado de cada instância é apresentada a quantidade de itens categorizados naquela instância.

A visualização das facetas e instâncias pode auxiliar os usuários na decisão de como iniciar e explorar uma coleção de itens. Qualquer faceta pode ser escolhida para o início da navegação. No entanto, o uso eficiente depende da familiaridade do usuário com o vocabulário utilizado para as instâncias das facetas.

5.5.3 Memória de trabalho

A técnica utilizada na ferramenta Memória de trabalho consiste em fornecer uma memória especial, que pode armazenar uma lista de itens (por exemplo, páginas wiki), e também fornecer algumas operações para manipular essa lista. Basicamente, as operações implementadas correspondem às funções de uma calculadora típica: M+ (*sum*), M- (*difference*), MR (*recall*), e MC (*clear*).

⁷<http://dir.yahoo.com/>

⁸<http://www.dmoz.org/>

⁹<http://www.epicurious.com/recipes/find/browse/>

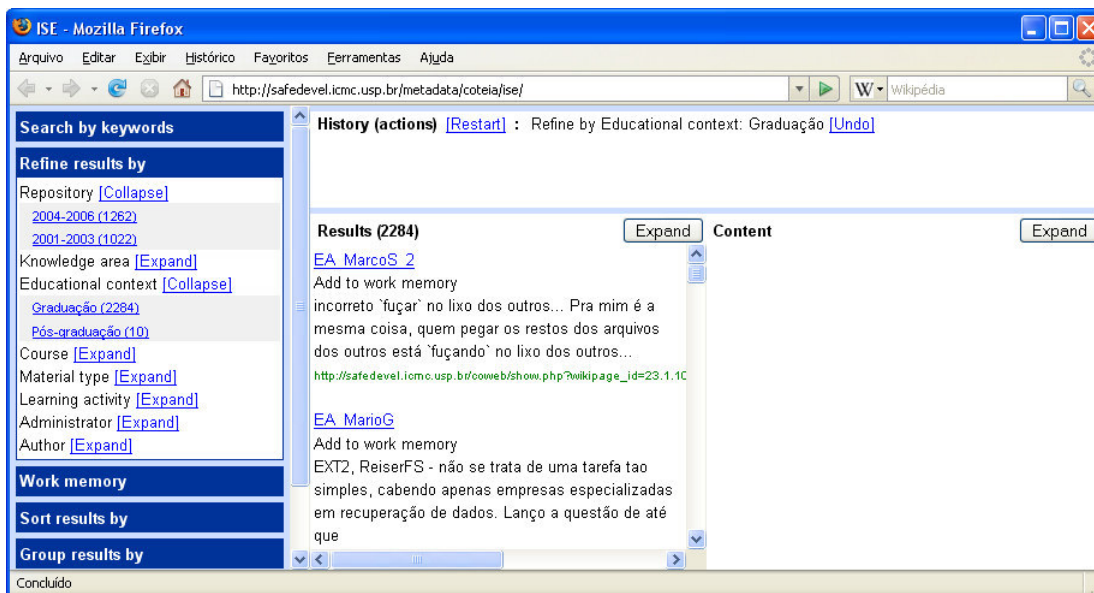


Figura 5.9: Interface da ferramenta Navegação por facetas no ENS

A interface apresentada na Figura 5.10 consiste na quantidade de itens da lista armazenados na memória e dos botões correspondentes às operações. A operação M+ computa a união entre a lista de itens apresentada na área de resultados e a lista de itens armazenada na memória; os itens duplicados são removidos. A operação M- computa a diferença entre a lista de itens na memória e a lista de itens da área de resultados (isto é, na memória ficarão armazenados todos os itens da sua lista que não estão presentes na lista da área de resultados). A operação MC limpa (ou elimina) todos os itens armazenados na lista da memória. A operação MR recupera os itens da lista da memória para a lista da área de resultados (isto é, os itens da memória ficam disponíveis na área de resultados).

A lista de itens armazenada na memória é estática e não reflete as ações subsequentes realizadas que alteram a lista de itens da área de resultados. Essa característica da ferramenta é útil nos casos em que o objetivo consiste em uma coleção de itens que devem ser obtidos a partir de estratégias diferentes e não somente no final de uma única estratégia. Por exemplo, a tarefa de coletar os links das páginas wiki que contêm material didático pode envolver o armazenamento de resultados intermediários para formar uma lista final daqueles considerados relevantes.

A ferramenta Memória de trabalho pode ser implementada totalmente no lado cliente, desde que sejam mantidos todos os dados necessários para recompor a parte da interface que corresponde à área de resultados. Na versão atual do ENS, a ferramenta mantém o acesso ao servidor devido ao registro da interação do usuário, e principalmente devido à decisão de projeto de implementar a *Navigation Engine* (Subseção 5.2.1).

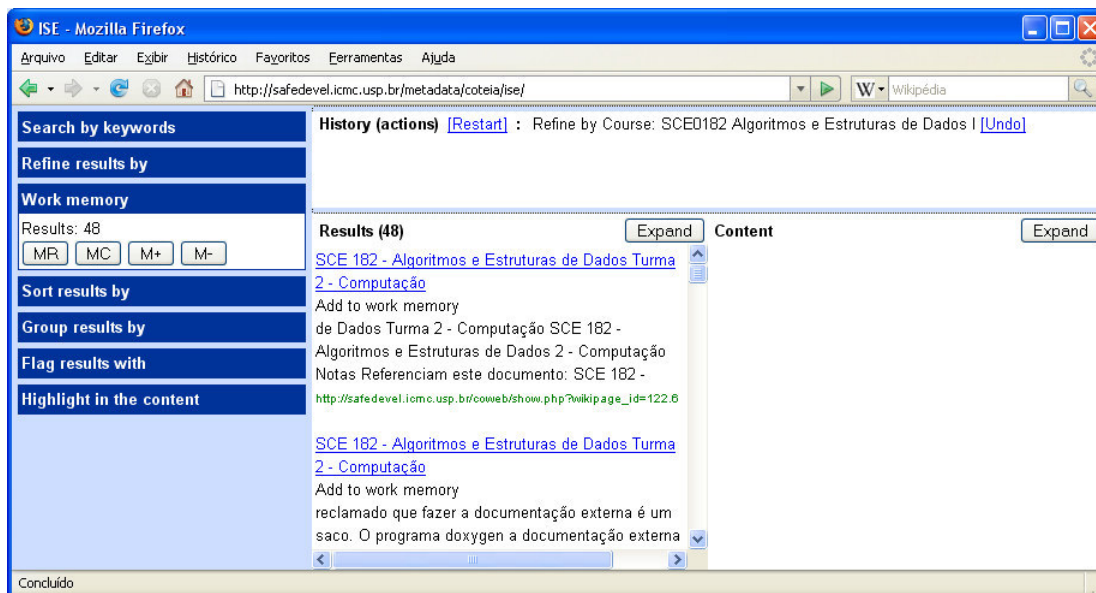


Figura 5.10: Interface da ferramenta Memória de trabalho no ENS

5.5.4 Classificação

Classificação é o processo de organizar uma lista de itens de acordo com um critério específico. As *search engines* classificam a lista de resultados de acordo com a relevância determinada pelos seus algoritmos proprietários de classificação (*ranking*). Existem outras opções que poderiam ser utilizadas para organizar os resultados, por exemplo, data, título, nome do autor (se conhecido) e domínio. Entretanto, enquanto o padrão de todas as *search engines* é classificar os resultados por relevância, poucas fornecem mais do que uma ou duas opções para classificação.

A interface da ferramenta Classificação apresentada na Figura 5.11 permite organizar a lista de resultados por qualquer um dos elementos de metadados disponíveis no Modelo RDF, e também pela relevância dos resultados da mesma maneira que a ferramenta Busca por palavra-chave descrita na Subseção 5.5.1 (isto é, pela frequência das palavras-chave especificadas pelo usuário). Além disso, a interface também permite escolher a opção de ordenação, entre ascendente (padrão) e descendente.

5.5.5 Agrupamento

Agrupamento (*grouping* ou *clustering*) é o processo de organizar uma lista de itens dentro de classes ou categorias. Algumas ferramentas de busca tentam agrupar automaticamente os resultados dentro de conceitos, como na *metasearch engine* Clusty¹⁰. Geralmente, os resultados dessas ferramentas de busca são organizados em um *layout* horizontal (árvores hierárquicas), com as categorias mais gerais na parte superior da árvore e aquelas mais específicas na parte inferior.

A vantagem da técnica de agrupamento é que o usuário pode examinar primeiro

¹⁰<http://clusty.com/>

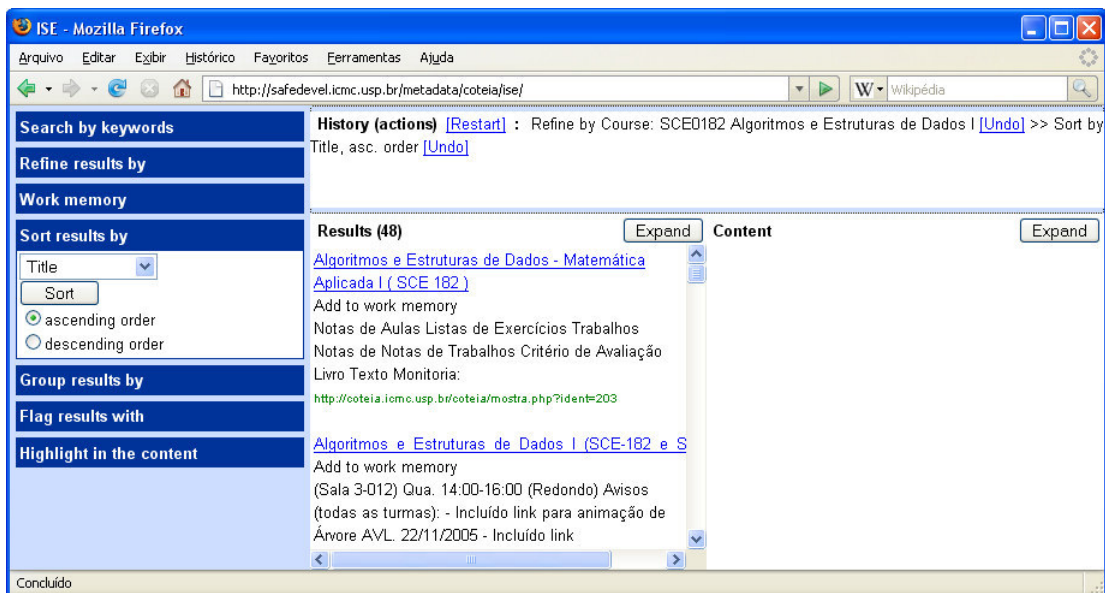


Figura 5.11: Interface da ferramenta Classificação no ENS

as categorias antes de verificar os resultados dentro de uma determinada categoria. Isso contrasta com o *layout* vertical comum da lista de resultados das *search engines* tradicionais, no qual o usuário pode precisar examinar cada um dos resultados para determinar se ele está relacionado à categoria ou tópico de interesse. A desvantagem dessa técnica é que as categorias geradas automaticamente não são sempre bem organizadas.

A interface da ferramenta Agrupamento apresentada na Figura 5.12 permite organizar os resultados de acordo com as mesmas categorias que são utilizadas pela ferramenta Navegação por facetas descrita na Subseção 5.5.2 (isto é, são utilizados os elementos de metadados disponíveis no Modelo RDF para organizar os resultados). A razão para essa adaptação (ou variação) da técnica é que os usuários podem compreender melhor a organização quando as categorias ou tópicos são conhecidos. A interface também permite agrupar por somente uma determinada categoria selecionada pelo usuário.

5.5.6 Marcar resultados

A técnica de *flagging*, desenvolvida no contexto deste trabalho, é uma técnica de *preview* que consiste em marcar (ou sinalizar) automaticamente um ou mais itens que possuem uma determinada característica dentro um conjunto de itens. O objetivo dessa técnica é facilitar a localização e seleção dos itens de interesse entre os demais itens do conjunto.

A ferramenta Marcar resultados implementa a técnica de *flagging* para marcar os resultados que contêm um determinado conteúdo. A interface apresentada na Figura 5.13 permite ao usuário selecionar o tipo de conteúdo de seu interesse. Na versão atual, os seguintes tipos de conteúdo estão disponíveis na ferramenta: nomes

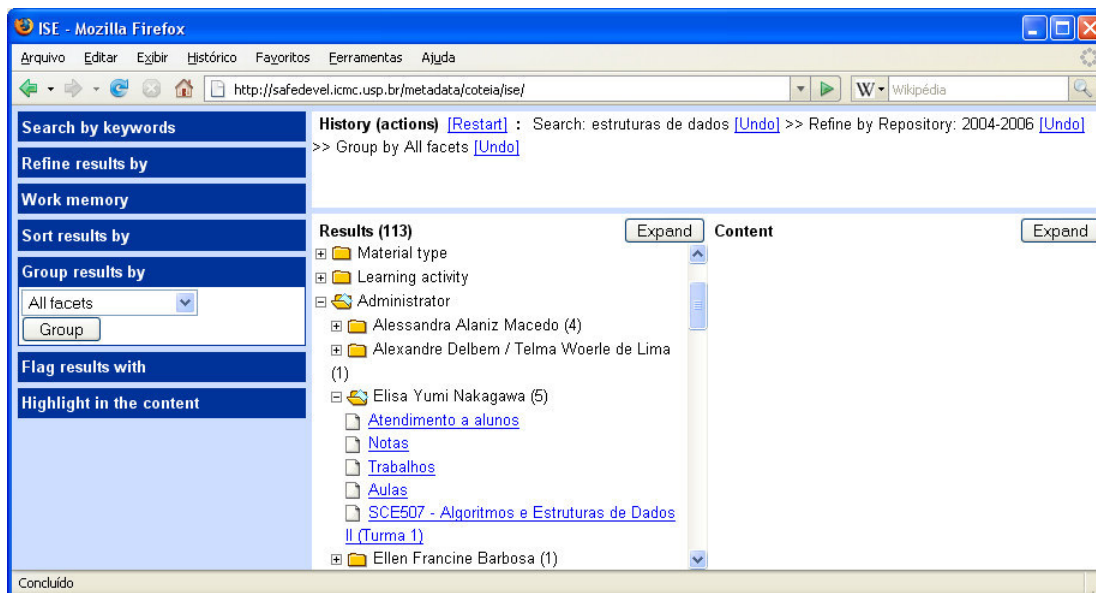


Figura 5.12: Interface da ferramenta Agrupamento no ENS

de pessoas, locais (por exemplo, salas), e-mails, datas e horários. A ferramenta insere um marcador (neste caso, um losango vermelho) no início do texto do título do resultado (área de resultados na Figura 5.13). Essa ferramenta é útil para que os usuários possam identificar à primeira vista se existem itens com um determinado conteúdo de interesse na lista de resultados fornecida como resposta às suas ações (ou estratégias).

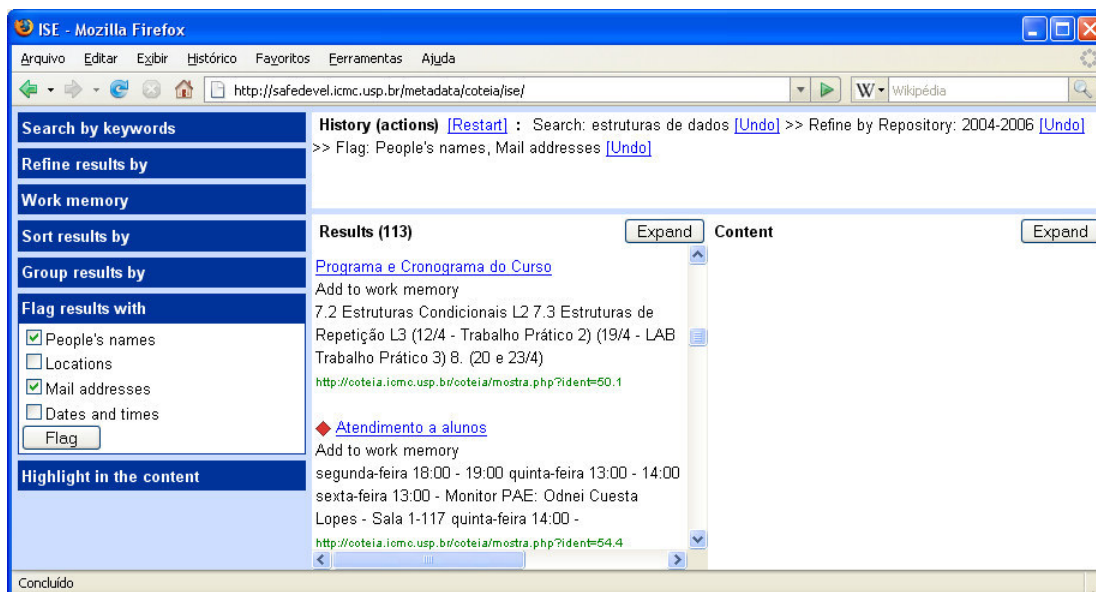


Figura 5.13: Interface da ferramenta Marcar resultados no ENS

5.5.7 Realce de termos

A técnica *keywords in context* é uma técnica de contexto [Baeza-Yates e Ribeiro-Neto 1999] que consiste em apresentar as ocorrências dos termos especificados na consulta dentro do conteúdo (ou de uma parte do conteúdo) dos documentos recuperados. Os termos geralmente são realçados através da alteração da cor do fundo (*background*) para outra cor diferente da que aparece no documento. Essa técnica é útil para auxiliar a exploração local porque muitos usuários não lêem as páginas cuidadosamente quando eles percorrem (*scan*) o texto para encontrar o que estão procurando. Essa técnica é utilizada nas *search engines*, como no sistema de *caching* do Google.

As técnicas “*entities in context*” e “*patterns in context*” são extensões da técnica descrita anteriormente. No caso de “*entities in context*”, as entidades contidas no Modelo RDF podem ser realçadas dentro do conteúdo, por exemplo: nomes de pessoas, locais e e-mails. No caso de “*patterns in context*”, as partes de texto que combinam com os padrões previamente especificados por meio de expressões regulares, podem ser realçadas dentro do conteúdo, por exemplo: e-mails, datas e horários.

A interface da ferramenta Realce de termos apresentada na Figura 5.14 permite ao usuário selecionar as opções que ele quer visualizar realçadas no conteúdo da página wiki apresentada na área de conteúdo. Além disso, o usuário pode digitar palavras em uma caixa de texto para que também sejam realçadas no conteúdo. Essa última funcionalidade é semelhante ao recurso “Localizar” disponível na maioria dos navegadores, e foi incluída na ferramenta a pedido dos participantes do estudo de usabilidade conduzido na avaliação (Capítulo 6).

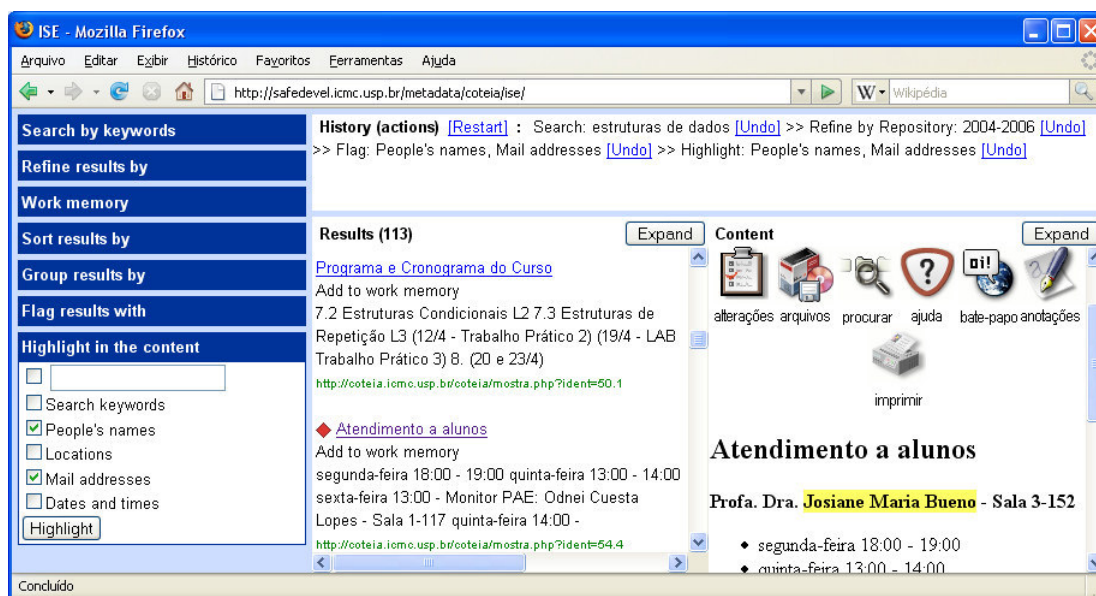


Figura 5.14: Interface da ferramenta Realce de termos no ENS

5.5.8 Histórico (reiniciar e desfazer)

A interface da ferramenta Histórico apresentada na Figura 5.15 permite ao usuário visualizar a seqüência de ações tomadas que fazem parte da sua estratégia atual de maneira semelhante à técnica conhecida por “*breadcrumb navigation*”. Essa técnica é geralmente utilizada para proporcionar ao usuário um meio de localização dentro da estrutura de um *website*, e consiste em mostrar o caminho percorrido pelo usuário (neste caso, a seqüência de links selecionados) até a página apresentada no navegador. Entretanto, ao contrário dos links do caminho percorrido pelo usuário, as ações mostradas no histórico não podem ser navegadas.

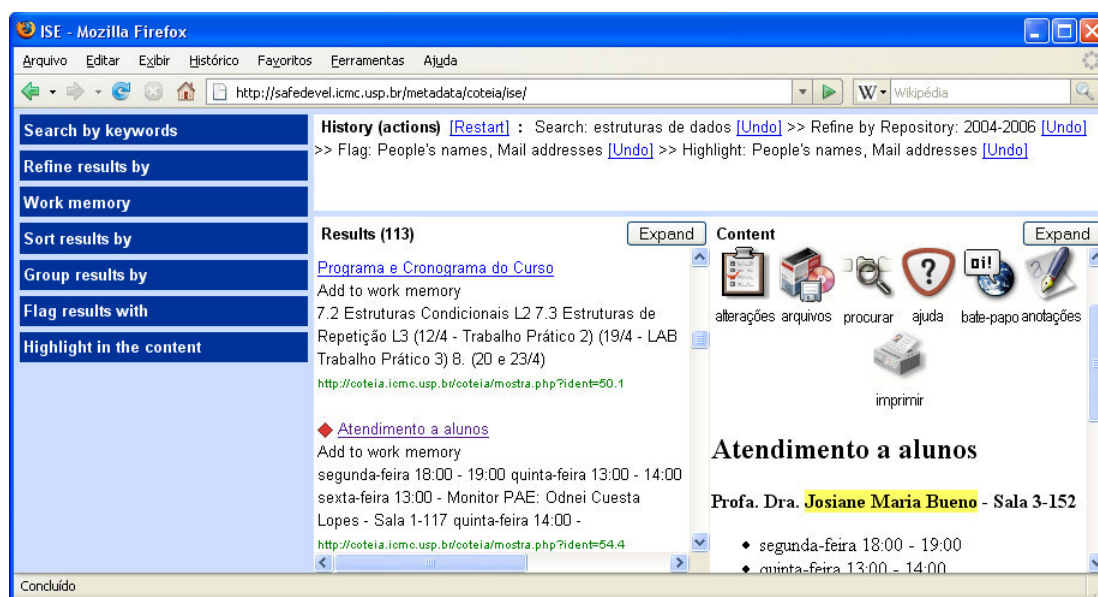


Figura 5.15: Interface da Ferramenta Histórico no ENS

A ferramenta também fornece suporte para reverter ações realizadas pelo usuário por meio de opções na interface para reiniciar a navegação (*restart*) e desfazer uma ação (*undo*). A opção de reiniciar permite ao usuário cancelar todas as ações e iniciar novamente a navegação. A opção de desfazer permite ao usuário cancelar qualquer uma das ações executadas ao invés de permitir somente a última ação executada (comportamento comum de um comando desfazer de edição). Ao desfazer uma ação, conseqüentemente os resultados refletirão a nova seqüência de ações (resultante) como se fosse aquela a seqüência executada pelo usuário.

O suporte fornecido pela ferramenta Histórico auxilia a manter o processo de navegação sob o controle do usuário. A visualização da estratégia utilizada até o momento pode auxiliar o usuário na compreensão de como os resultados foram obtidos. Além disso, se os resultados não são satisfatórios, o usuário pode utilizar a ferramenta para avaliar e reformular a sua estratégia, por exemplo, para desfazer uma ou mais ações ou até mesmo recomeçar “a partir do zero” (*restart*).

5.6 Cenário de uso

Nesta seção é apresentado um exemplo de uso do ENS para procurar por informação nos repositórios CoTeia do ICMC/USP. O objetivo principal é ilustrar a abordagem proposta no modelo NAVE dentro do contexto do ENS (prova de conceito) em atividades de pesquisa por informação. Além disso, este exemplo mostra como estratégias¹¹ diferentes podem ser utilizados para formar uma estratégia, isto é, mostra as diferentes possibilidades de escolha de ferramentas para realizar a mesma atividade.

O exemplo é apresentado através de uma descrição narrativa informal, denominada de cenário segundo Carroll [2000], e de uma descrição em passos das ações do usuário segundo a notação proposta para cenários por Leite et al. [1997a,b]. A primeira descrição está centrada nas atividades de alto nível de abstração realizadas pelo usuário e a segunda está centrada na interação entre usuário e sistema.

A notação para cenários de Leite et al. foi ligeiramente modificada para também mostrar números associados aos episódios com a finalidade de indicar a sequência dos episódios e fazer referência no texto. Além disso, a linguagem utilizada para descrever um episódio é menos formal e abstrata do que a proposta pela notação; as ações do episódio são instanciadas ao invés de representar um caso mais geral que o cenário deveria abranger.

Segundo os levantamentos realizados durante a avaliação do ENS (Capítulo 6), a abordagem geralmente utilizada pelos usuários da CoTeia para procurar informação é a navegação a partir da entrada (link) correspondente à disciplina na página inicial do repositório. Nesse contexto, um cenário em alto nível para uma necessidade de informação e a estratégia utilizada pode ser assim descrito usando a concepção de Carroll [2000]:

*Quero encontrar exercícios sobre filas para uma aula de estruturas de dados.
Vou então procurar as páginas das disciplinas relacionadas a esse tópico e
verificar aquelas com exercícios.*

A estratégia descrita no cenário é válida para a abordagem geralmente utilizada para procurar informação da CoTeia: navegar a partir dos links correspondentes às disciplinas de estruturas de dados na página inicial e visualizar todas as páginas que possam conter exercícios. Um dos problemas dessa abordagem ocorre quando o usuário utiliza apenas um dos repositórios, por exemplo, aquele com as páginas das disciplinas mais recentes (Tabela 5.2). O usuário pode desconsiderar um material valioso contido nas páginas wiki de outro repositório.

Na Figura 5.16 é mostrado um cenário de uso do ENS segundo a estratégia apresentada no cenário descrito informalmente. Nesse cenário de uso, o usuário utiliza as ferramentas Busca por palavra-chave (episódios 1 e 3) e Agrupamento (episódios

¹¹Segundo a definição de estratégia proposta por Bates [1990], descrita na Subseção 4.2.1

2 e 4). Nessa abordagem, os dois repositórios são explorados durante a navegação como se fosse um único repositório. No caso do usuário querer navegar em apenas um dos repositórios, o repositório desejado pode ser selecionado por meio do elemento de metadados *Source* (do *application profile*, Tabela 5.1) na ferramenta Navegação por facetas.

Título	Um professor procurando exercícios sobre filas para uma aula de estruturas de dados.
Objetivo	Mostrar um exemplo de uso do ENS para procurar por informação nos repositórios de páginas wiki da CoTeia.
Contexto	Local de trabalho do usuário
Ator	Professor.
Recurso	Navegador.
Episódios	1. <u>Buscar</u> por palavras-chave: estruturas dados. 2. <u>Agrupar</u> por Course. 3. <u>Buscar</u> por palavras-chave: filas. 4. <u>Agrupar</u> por Materialtype. 5. <u>Seguir links</u> selecionados.

Figura 5.16: Cenário de uso do ENS com busca por palavra-chave e agrupamento

Na Figura 5.17 também é mostrado um cenário de uso do ENS no contexto da estratégia apresentada para satisfazer a necessidade de informação. No entanto, outras ferramentas foram utilizadas pelo usuário: Navegação por facetas (episódios 1, 4 e 8) e Memória de trabalho (episódios 2, 5 e 8), além do Histórico (episódios 3 e 6). A estratégia em alto nível é a mesma, mas a escolha que o usuário faz das ferramentas disponíveis no ENS para realizar a navegação é diferente.

5.7 Considerações finais

Neste capítulo, o sistema de navegação exploratória, denominado de ENS, foi apresentado. As características da arquitetura e implementação do sistema foram apresentadas, assim como as decisões de projeto com relação à navegação e independência de domínio e os sistemas de apoio (*Crawler*, *Indexer* e *MeGen*). As ferramentas de navegação integradas ao sistema ENS foram descritas junto com as técnicas subjacentes e as extensões que foram desenvolvidas. No final, foi apresentado um cenário de uso do ENS.

Título	Um professor procurando exercícios sobre filas para uma aula de estruturas de dados.
Objetivo	Mostrar um exemplo de uso do ENS para procurar por informação nos repositórios de páginas wiki da CoTeia
Contexto	Local de trabalho do usuário
Ator	Professor.
Recurso	Navegador.
Episódios	1. <u>Refinar</u> por Course: SCE0182 Algoritmos e Estruturas de Dados I. 2. <u>Adicionar</u> os resultados à <u>memória</u> de trabalho. 3. <u>Desfazer</u> Refinar por Course: SCE0182 Algoritmos e Estruturas de Dados I. 4. <u>Refinar</u> por Course: SCE0603 Algoritmos e Estruturas de Dados I. 5. <u>Adicionar</u> os resultados à <u>memória</u> de trabalho. 6. <u>Desfazer</u> Refinar por Course: SCE0603 Algoritmos e Estruturas de Dados I. 7. <u>Recuperar da memória</u> . 8. <u>Refinar</u> por Material type: Exercícios. 9. <u>Seguir links</u> selecionados.

Figura 5.17: Cenário de uso do ENS com navegação por facetas e memória de trabalho

Avaliação

6.1 Considerações iniciais

Neste capítulo são apresentados o estudo de usabilidade e o estudo experimental conduzidos para avaliar o sistema ENS (Capítulo 5), desenvolvido com base no modelo NAVE (Capítulo 4). Adicionalmente, conhecimentos sobre o sistema e o modelo subjacente também foram obtidos por meio do estudo do registro da interação dos usuários com o sistema capturado durante o estudo de usabilidade.

O capítulo está organizado da seguinte maneira: na Seção 6.2 é apresentado o estudo de usabilidade conduzido com professores; na Seção 6.3 é apresentado o estudo experimental realizado com alunos; na Seção 6.4 são apresentadas as considerações finais deste capítulo.

6.2 Estudo de usabilidade

Os testes de usabilidade, ou testes com usuários, são uma maneira aplicada de experimentação utilizada por desenvolvedores para testar se o produto que desenvolveram é usável pela população pretendida de usuários na realização de suas tarefas [Dumas e Redish 1999; Preece et al. 2002]. A realização de um estudo de usabilidade foi uma das maneiras consideradas para avaliar o sistema desenvolvido.

A avaliação do sistema por meio de um estudo de usabilidade foi realizada de acordo com as recomendações do padrão ISO/IEC 14598-5 [ISO/IEC 1998], que divide o processo de avaliação em quatro estágios: especificação, planejamento, avaliação piloto e avaliação. A apresentação do estudo realizado está assim organizada: na Seção 6.2.1 são apresentados os objetivos deste estudo de usabilidade; na Seção 6.2.2 é apresentada a metodologia empregada abrangendo os participantes, as tarefas projetadas e o procedimento conduzido; e na Seção 6.2.3 são apresentados os resultados e observações deste estudo.

6.2.1 Objetivos

São três os objetivos deste estudo de usabilidade:

1. Validar se os usuários conseguem encontrar informação relevante nos repositórios CoTeia usando o sistema.
2. Confirmar se os usuários usam estratégias diferentes na abordagem proposta (modelo NAVE) no sistema.
3. Analisar passo a passo as estratégias que os usuários seguem durante a navegação pelas páginas wiki usando o sistema, e as suas preferências com relação às ferramentas disponíveis.

6.2.2 Metodologia

Diversos métodos foram utilizados para a coleta de dados sobre a experiência do usuário com o sistema com o objetivo de obter mais informações a partir de perspectivas diferentes. Os seguintes métodos foram combinados: (a) questionário, (b) *think-aloud*¹, (c) entrevista e (d) registro da interação do usuário.

Participantes

O sistema foi avaliado com 15 professores do ICMC/USP. Embora quanto mais participantes houver nesse tipo de estudo, mais representativos da população de usuários são os resultados, especialistas de usabilidade recomendam que de 6 a 12 participantes são o suficiente [Dumas e Redish 1999].

Um levantamento (*survey*) realizado no início da avaliação mostra que esses participantes são usuários experientes de ferramentas de busca (*search engines*) e pesquisam por informação diariamente. Os participantes também são usuários regulares da CoTeia (com 2-6 anos de experiência) e visitam as páginas wiki semanalmente para leitura e/ou edição. Portanto, os participantes podem ser considerados usuários potenciais do sistema.

Tarefas

Os participantes completaram um conjunto de tarefas que envolveram a pesquisa por material educacional para reuso. Foram projetadas cinco tarefas (Tabela 6.1) em quatro tipos de cenários diferentes baseados na definição de Shneiderman [1997]: pesquisa específica (*specific fact-finding*), pesquisa estendida (*extended fact-finding*), navegação aberta (*open-ended browsing*) e exploração de disponibilidade (*exploration of availability*).

As tarefas foram apresentadas aos participantes de acordo com o conceito de situação simulada da tarefa de trabalho (*simulated work task situation*) [Borlund

¹Técnica de pensar “em voz alta”. Essa técnica exige que as pessoas digam em voz alta tudo o que estiverem pensando e tentando fazer; dessa maneira, seus processos de raciocínio são exteriorizados.

Tabela 6.1: Cinco tarefas em quatro tipos de cenários

Tarefas		Cenários
T1	Procurar por um determinado exercício.	pesquisa específica
T2	Procurar pelo e-mail de um monitor de disciplina.	
T3	Procurar pelos critérios de avaliação recentemente divulgados pelos outros professores.	pesquisa estendida
T4	Procurar por uma nova abordagem usada para ensinar um determinado tópico.	navegação aberta
T5	Procurar pelo material educacional disponível para uma determinada disciplina.	exploração de disponibilidade

2003]. Uma situação simulada da tarefa de trabalho é uma descrição aberta do contexto/cenário de uma determinada situação. Essa descrição funciona como o gatilho da necessidade de informação do participante e a base para o julgamento da relevância (*relevance judgment*). O objetivo é garantir o maior realismo possível por meio do envolvimento de usuários potenciais que, baseados na situação simulada, desenvolvem interpretações individuais e subjetivas para a necessidade de informação. Individualmente, os participantes usam o sistema e avaliam a relevância dos resultados obtidos com relação às suas percepções da necessidade de informação e da situação simulada subjacente.

Procedimento

O procedimento completo teve duração, em média, de aproximadamente 1 hora e 33 minutos e foi realizado no próprio local de trabalho do participante. O material preparado para os participantes (Apêndice A) contém um questionário pré-avaliação, uma visão geral do processo de avaliação, a descrição das tarefas e questionário pós-tarefa, e entrevista pós-avaliação. O procedimento realizado compreendeu quatro sessões:

1. *Sessão de levantamento.* Um questionário de levantamento, denominado de questionário pré-avaliação, foi projetado para coletar informação demográfica básica (por exemplo, gênero, idade, tempo de serviço) e detalhes da experiência do participante com relação a ferramentas de busca (*search engines*), navegadores (*browsers*), catálogos *on-line* de biblioteca e experiência com relação à CoTeia.
2. *Sessão de treinamento.* Uma sessão de treinamento foi ministrada para o participante familiarizar-se com o sistema e as ferramentas. O conteúdo fornecido nessa sessão incluiu: (a) uma visão geral do sistema; (b) uma introdução às ferramentas; e (c) uma oportunidade para usar o sistema.
3. *Sessão de tarefa.* Após a sessão de treinamento, os participantes completaram as tarefas enquanto pensavam “em voz alta” sobre suas estratégias. Os participantes foram instruídos a articular o que eles estavam pensando e sentindo

enquanto trabalhavam com o sistema. As elocuições foram registradas usando um gravador de áudio. As interações dos participantes com o sistema, associadas com cada tarefa, foram registradas em um arquivo de *log*. Todos os participantes completaram as mesmas tarefas e nenhum limite de tempo foi atribuído para qualquer uma das tarefas. Após cada tarefa, os participantes responderam questões sobre a relevância dos resultados obtidos e a utilidade do sistema com relação àquela tarefa.

4. *Sessão de entrevista.* Após executar as tarefas, os participantes foram entrevistados sobre suas percepções acerca do sistema que eles experimentaram. A entrevista semi-estruturada foi projetada para fornecer *feedback* sobre o sistema (no sentido geral), as ferramentas, os metadados, o processo de navegação e sobre a questão da “desorientação do usuário”. Além disso, os participantes foram incentivados a indicar quaisquer preferências, limitações e dificuldades que eles tiveram enquanto usavam o sistema.

6.2.3 Resultados e observações

Nesta subseção são apresentados os principais resultados e observações do estudo de usabilidade realizado.

Relevância e utilidade

A avaliação de situações simuladas é difícil porque não existem respostas corretas e o objetivo não é necessariamente minimizar o tempo para completar tarefas [Borlund 2003]. Após cada tarefa, os participantes responderam um pequeno questionário sobre a relevância dos resultados encontrados e sobre a utilidade do sistema para a tarefa. Os julgamentos foram feitos em uma escala Likert de 7 pontos (1 = nenhuma, 7 = extrema). Na Tabela 6.2 são apresentados os resultados estatísticos dos julgamentos com relação a relevância e utilidade. Os participantes atribuíram uma avaliação média de 5,92 para relevância (desvio padrão de 0,90) e de 6,07 para utilidade (desvio padrão de 0,81). Esses resultados fornecem indicadores para a afirmação do primeiro objetivo: os usuários conseguem encontrar informação relevante em repositórios CoTeia usando o sistema.

Estratégias

A partir da análise do arquivo de *log* da interação do usuário foi encontrado que os participantes utilizaram ferramentas diferentes para completar as tarefas. O número médio de ferramentas diferentes utilizadas por tarefa foi 3,24, e o desvio padrão (DP = 1,48), assim como os valores de mínimo (MIN = 1) e máximo (MAX = 8), também indicam essa variabilidade no uso das ferramentas. Adicionalmente, também foram observadas diferenças no número de vezes que as ferramentas foram usadas, na escolha de ferramentas e diferenças nas seqüências escolhidas para cada tarefa.

Na Figura 6.1 é mostrado um exemplo de análise da interação de um usuário

aleatório. Cada marca (x) significa que a ferramenta da linha correspondente foi utilizada na tarefa indicada pela coluna; a seqüência de interação é representada da esquerda para a direita. Por exemplo, para completar a Tarefa 3, o usuário utilizou a seguinte seqüência de ferramentas: Busca por palavra-chave, Desfazer (da ferramenta Histórico); Navegação por facetas; e novamente a Busca por palavra-chave.

A ferramenta Realce de termos (linha 7) e a Reiniciar (da ferramenta Histórico, linha 9) foram utilizadas apenas uma vez, e a Memória de trabalho (linha 3) não foi utilizada. Analisando as interações dos demais usuários, é possível verificar que também existem diferenças entre os participantes considerando uma mesma tarefa. Os resultados da análise fornecem indicadores para o segundo objetivo: os usuários usam estratégias diferentes na abordagem de navegação proposta.

Tabela 6.2: Resultados estatísticos dos julgamentos com relação a relevância e utilidade

	Relevância			Utilidade		
	Média	Mediana	DP	Média	Mediana	DP
T1	6,07	6,00	0,59	6,20	6,00	0,77
T2	6,07	6,00	0,96	6,13	6,00	1,13
T3	5,87	6,00	0,99	6,20	6,00	0,56
T4	5,27	6,00	0,88	5,53	6,00	0,74
T5	6,33	6,00	0,72	6,27	6,00	0,59
Total	5,92	6,00	0,90	6,07	6,00	0,81

Julgamentos foram feitos em uma escala Likert de 7 pontos (1 = nenhuma; 7 = extrema)

Na Figura 6.2 é mostrado o quanto as ferramentas foram utilizadas de fato. As ferramentas Navegação por facetas e a Busca por palavra-chave foram as mais utilizadas. Isso ocorreu porque essas ferramentas são as que fornecem maneiras de obter (buscar ou refinar) uma lista de resultados e os participantes escolheram iniciar suas tarefas com Navegação por facetas (62,5%) mais freqüentemente do que com Busca por palavra-chave (37,5%). Inesperadamente, a Memória de trabalho teve porcentagem alta de uso; entretanto, apenas três participantes usaram intensivamente essa ferramenta, enquanto outros nove nem sequer a consideraram para completar as suas tarefas.

Uso dos metadados

Os elementos de metadados mais utilizados na navegação por facetas foram: *Course* (31,55%), *Learning Material Type* (16,67%), e *Learning Activity Type* (11,90%). Esses resultados motivam a necessidade de vocabulários mais específicos para o contexto educacional, por exemplo, do que aqueles propostos na categoria Educacional do esquema de metadados LOM [IEEE LTSC 2002].

											→ tempo	
Ferramentas	Tarefa 1			Tarefa 2				Tarefa 3		Tarefa 4		Tarefa 5
1 Busca por palavra-chave	x	x	x	x	x	x	x	x	x	x		
2 Navegação por facetas		x	x	x				x		x	x	x
3 Memória de trabalho												
4 Classificação												
5 Agrupamento					x		x				x	
6 Marcar resultados												x
7 Realce de termos							x			x		
8 Desfazer (Histórico)			x	x		x		x			x	x
9 Reiniciar (Histórico)		x										

Figura 6.1: Exemplo de análise da interação de um usuário

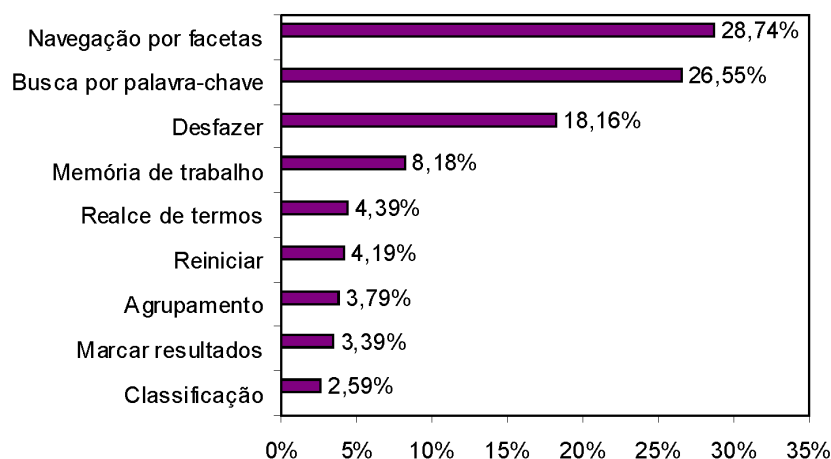


Figura 6.2: Porcentagem de vezes que as ferramentas foram utilizadas

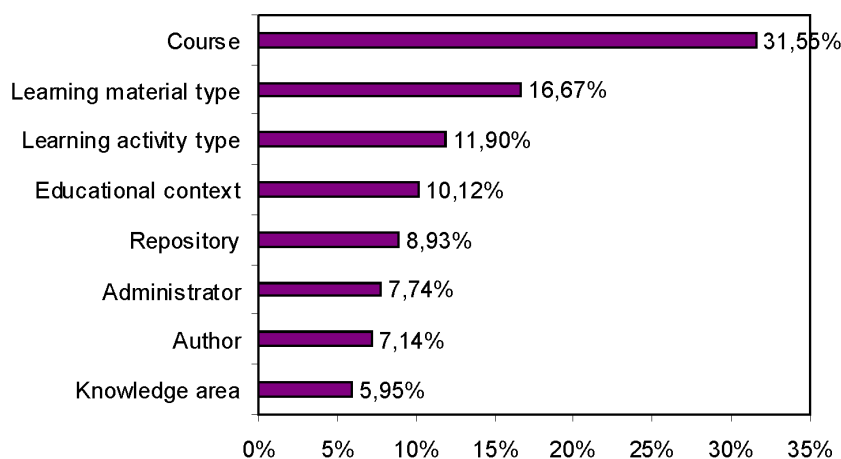


Figura 6.3: Porcentagem de vezes que os metadados foram utilizados

Análise qualitativa das entrevistas

Os participantes comentaram favoravelmente sobre experiência de utilizar o sistema para procurar por informação nos repositórios da CoTeia. As dificuldades citadas pelos participantes foram: lembrar da funcionalidade de cada uma das ferramentas; e compreender o funcionamento da ferramenta Histórico (principalmente a opção de desfazer).

Em geral, os participantes consideraram que o sistema foi útil para a realização das tarefas e que os resultados encontrados foram relevantes. Entre as ferramentas mais citadas como as preferidas dos participantes estão a Memória de trabalho e a Destaque de termos. A Navegação por facetas e a Busca por palavra-chave foram consideradas as ferramentas mais úteis.

Os metadados disponíveis foram considerados representativos das páginas wiki. Um dos participantes solicitou que o termo ano/semestre com relação ao elemento de

metadados *Course* fosse incluído na interface da ferramenta Navegação por facetas.

Os participantes foram questionados especificamente para comentar sobre o controle do processo de navegação. Os participantes (todos) declararam que podiam controlar a navegação através do uso da ferramenta Histórico. Além disso, quando questionados, os participantes não afirmaram terem ficado perdido enquanto estavam usando o sistema e que compreendiam como os resultados eram obtidos. Ao final da entrevista, a maioria dos participantes afirmou que gostaria de usar o sistema no futuro.

Episódios relevantes

Os resultados qualitativos da análise dos registros das elocuções do *think-aloud* são resultado de observações, e assim como em muitos estudos qualitativos, somente as partes consideradas interessantes são analisadas. Portanto, os episódios que são particularmente ilustrativos das características e vantagens da abordagem subjacente ao ENS são apresentados.

Os exemplos da abordagem envolvem a navegação realizada em passos (passo a passo) utilizando diferentes técnicas/ferramentas em cada passo. A seguir, é transcrito um trecho da mensagem de áudio gravada de um participante. (os termos que poderiam identificar o participante foram substituídos por letras maiúsculas na transcrição do episódio).

Nesse semestre eu vou ministrar uma nova disciplina: AB. Como é a primeira vez que essa disciplina é ministrada não vai ter nada especificamente sobre ela na CoTeia, mas tem material sobre A e tem material sobre B. Eu vou procurar por curso, selecionar as disciplinas similares e verificar uma por uma. Melhor, eu vou escolher o repositório mais recente antes. [...] Vieram muitos resultados. Agora eu preciso refinar mais, pois eu quero somente sobre A e B. Por tipo de material ou contexto [...] mas também não vai adiantar. Pelo administrador e selecionar X e Y, e daí eu chego onde quero. Agora teria de agrupar isso por administrador. Posso utilizar a memória de trabalho [...] então eu adiciono. Eu quero selecionar as disciplinas com A e B, ministradas pelos professores (administradores) X e Y. [...] Agora eu vou percorrer os links para selecionar o material e ver se o conteúdo é o que quero. Poderia também refinar novamente por palavras para tentar diminuir a quantidade de resultados.

6.3 Estudo experimental

A motivação para a realização de um estudo experimental foi verificar se os usuários atuam melhor na abordagem proposta no ENS (que segue o modelo NAVE descrito no Capítulo 4) ou quando os usuários utilizam somente uma ferramenta, neste caso, foi escolhida a ferramenta de busca por palavra-chave (que segue o modelo clássico de recuperação de informação [Baeza-Yates e Ribeiro-Neto 1999]). A busca por palavra-chave foi escolhida porque é muito comum encontrar em *websites* uma ferramenta semelhante incorporada na página principal. Inicialmente, também era pretendido

realizar a comparação com os usuários utilizando somente o navegador. No entanto, um teste piloto revelou a impossibilidade de aplicar o experimento em um terceiro grupo devido ao tempo e ao número de pessoas disponíveis. A solução encontrada foi incluir uma pergunta no questionário de fechamento para obter alguma informação de maneira subjetiva.

Na Figura 6.4 é mostrada a interface do sistema ENS, denominado de Sistema A na sequência desta seção. Uma parte da interface (a área das ferramentas à esquerda) foi traduzida para o Português para facilitar o uso pelos participantes. A outra parte não foi traduzida porque demandaria uma quantidade significativa de re-trabalho de codificação somente para esta etapa de experimentação.

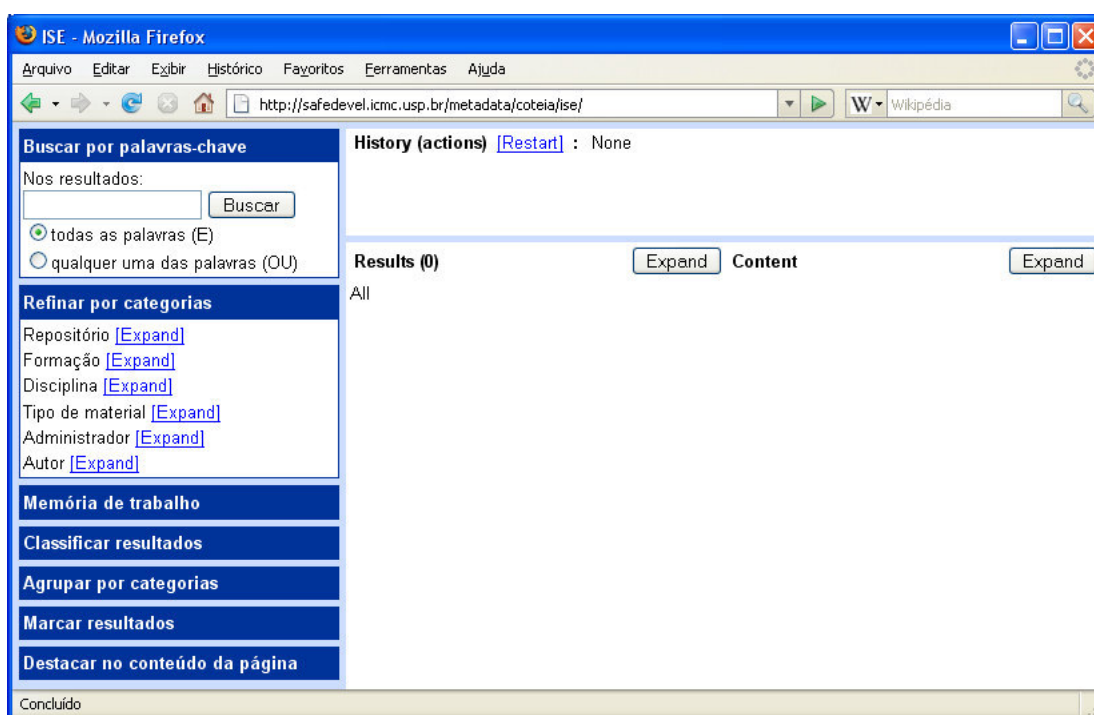


Figura 6.4: Interface do ENS (Sistema A)

Na Figura 6.5 é mostrada a interface da busca por palavra-chave, denominada de Sistema B na sequência desta seção. A aparência e o modo de atuar (*look and feel*) da interface foram mantidas para que não tivessem influência no desempenho dos participantes. A própria ferramenta de busca por palavra-chave do ENS foi utilizada no estudo para tentar igualar os tempos de resposta (ou pelo menos minimizar a diferença).

As próximas subseções seguem a estrutura proposta por Wohlin et al. [2000] para a apresentação de experimentos: definição do experimento (Subseção 6.3.1), planejamento do experimento (Subseção 6.3.2), operação do experimento (Subseção 6.3.3) e análise e interpretação dos resultados (Subseção 6.3.4). Os principais documentos elaborados neste estudo experimental estão no Apêndice B.

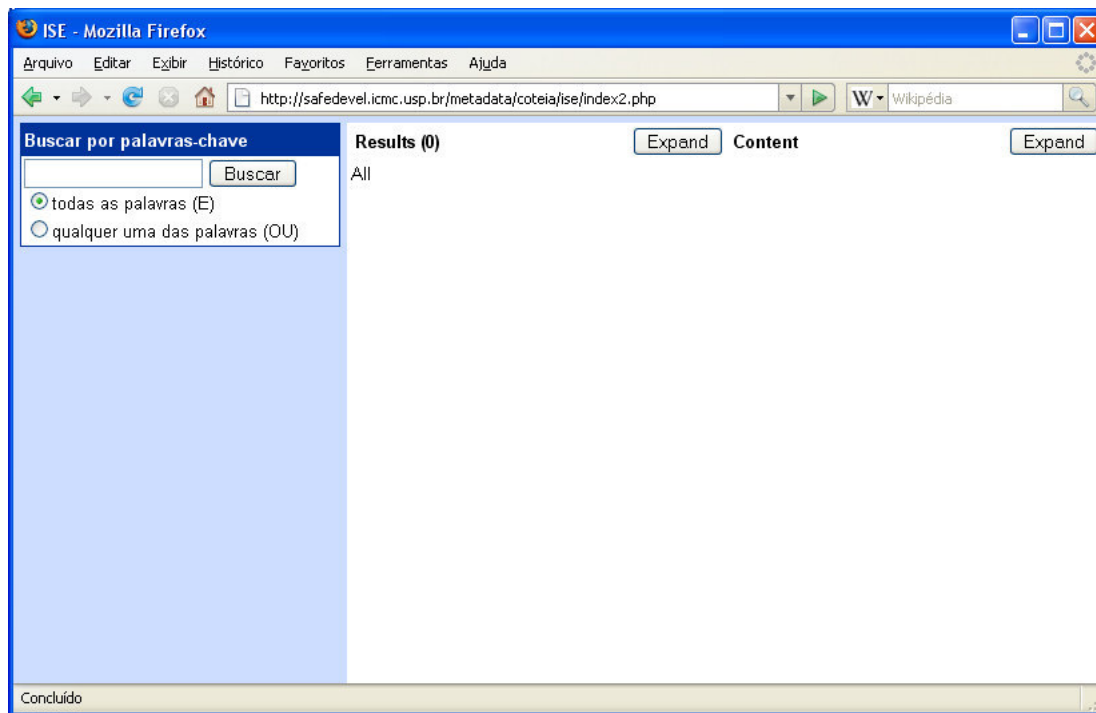


Figura 6.5: Interface da busca por palavra-chave (Sistema B)

6.3.1 Definição do experimento

Nesta subseção é apresentada a definição do experimento. O objeto de estudo é definido para delimitar o escopo do experimento. O propósito e o foco de qualidade formam a base para a definição das hipóteses. A perspectiva define a visão e o contexto apresenta de maneira sucinta os participantes e a caracterização do experimento.

Objeto de estudo

Os objetos de estudo são o Sistema A (ENS) e o Sistema B (busca por palavra-chave).

Propósito

O propósito é comparar o desempenho dos usuários na realização de tarefas para encontrar informação nos repositórios CoTeia usando o Sistema A com o desempenho daqueles usando o Sistema B.

Foco de qualidade

O principal efeito estudado no experimento é o desempenho na realização de tarefas para encontrar informação nos repositórios CoTeia.

Perspectiva

A perspectiva é em relação ao ponto de vista do pesquisador, isto é, o pesquisador gostaria de saber se existe alguma diferença significativa no desempenho dos usuários

no uso de cada um dos objetos de estudo.

Contexto

O experimento foi realizado com alunos de graduação do ICMC/USP, no contexto de uso dos dois sistemas para encontrar informação nos repositórios CoTeia. O estudo é caracterizado como um *blocked subject-object study* [Wohlin et al. 2000], no qual vários participantes foram randomicamente alocados a cada um dos objetos de estudo.

6.3.2 Planejamento do experimento

Nesta subseção é apresentado o planejamento do experimento. O ambiente do experimento é apresentado, as hipóteses e as variáveis são definidas, e a técnica de seleção dos participantes é descrita. O projeto do experimento é definido baseado na definição das hipóteses e variáveis. A instrumentação necessária para a realização do experimento também é descrita. Finalmente, as ameaças com relação à validade dos resultados do experimento são avaliadas.

Seleção do contexto

O experimento foi conduzido em condições de laboratório e de forma controlada (com a presença de um supervisor). No entanto, não houve supervisão no sentido de auxiliar os participantes na realização das tarefas. O ambiente escolhido foi o laboratório da graduação (conhecido dos participantes) e a sua escolha teve por objetivo minimizar a influência do ambiente na realização das tarefas. O experimento é dirigido para um problema prático e específico, pois o foco são sistemas que podem ser utilizados para encontrar informação nos repositórios CoTeia.

Definição da hipótese

A hipótese pode ser declarada informalmente da seguinte maneira:

A ferramenta CoTeia tem sido utilizada no ICMC/USP desde 2001 e contém uma quantidade significativa de material didático. Os usuários têm dificuldade de encontrar a informação que estão procurando usando apenas os recursos do navegador. Acredita-se que os usuários têm desempenho melhor usando a abordagem do ENS (Sistema A) do que a abordagem tradicional da busca por palavra-chave (Sistema B).

A partir da declaração informal pode-se definir a hipótese formalmente da seguinte maneira:

- Hipótese nula, H_0 : o desempenho dos usuários na realização de tarefas utilizando o Sistema A é inferior que utilizando o Sistema B.

- Hipótese alternativa, H1: o desempenho dos usuários na realização de tarefas utilizando o Sistema A é superior que utilizando o Sistema B.

Seleção das variáveis

Uma vez que a hipótese foi declarada em termos de desempenho, as variáveis foram identificadas. As variáveis dependentes de desempenho consideradas para o experimento são o tempo gasto para completar as tarefas e a relevância dos resultados encontrados. A relevância é determinada pelas medidas de revocação (*recall*) e precisão (*precision*). A medida de revocação (*r*) é a fração dos resultados relevantes que foi encontrada, enquanto que a medida de precisão (*P*) é a fração dos resultados encontrados que é relevante [Baeza-Yates e Ribeiro-Neto 1999]. O conjunto de resultados relevantes considerado para cada uma das tarefas foi determinado da seguinte maneira: um conjunto inicial foi obtido a partir de navegação (*browsing*); os resultados apresentados pelos participantes foram analisados e, se um resultado considerado relevante por um participante não constava no conjunto inicial, esse resultado era acrescentado ao conjunto inicial.

Essas variáveis (tempo e relevância) podem competir uma com a outra em cada uma das tarefas. Por exemplo, um usuário pode gastar muito tempo para completar as tarefas, mas os resultados encontrados podem apresentar relevância alta; ou um resultado com relevância baixa pode ser consequência de pouco tempo gasto para completar as tarefas. Nesse sentido, para rejeitar a hipótese nula, o critério adotado considera que a relevância dos resultados encontrados é mais importante do que o tempo gasto para completar as tarefas.

Para a relevância também é considerada uma terceira medida, que é obtida a partir de um questionário aplicado após cada uma das tarefas. O objetivo desse questionário é obter um julgamento subjetivo do participante com relação à relevância dos resultados encontrados na realização da tarefa. Os resultados dessa medição subjetiva foram comparados posteriormente com a relevância determinada pelas medidas de revocação e precisão (Subseção 6.3.4).

Seleção dos participantes

A técnica de seleção dos participantes utilizada foi a amostragem por conveniência (*convenience sampling*). Nessa técnica, as pessoas mais próximas e mais convenientes são selecionadas para participar do experimento. No entanto, a partir da amostra tomada foram alocados, aleatoriamente, participantes diferentes para cada um dos sistemas (condições). Nessa abordagem, a vantagem é que não há efeitos de ordenamento causados pela influência da experiência dos participantes em um conjunto de tarefas no desempenho de outras, uma vez que cada participante atua apenas em uma condição.

Projeto do experimento

A amostra dos participantes foi dividida randomicamente em dois grupos e cada grupo atribuído a um dos sistemas (condições experimentais, Tabela 6.3). O projeto do experimento consiste em um fator com dois tratamentos (*one factor with two treatments*), projetado para comparar a média de dois tratamentos. Neste caso, o *unpaired t-test* foi utilizado. Segundo Wohlin et al. [2000], o *t-test* é um dos testes paramétricos mais frequentemente utilizados para comparar médias nesse tipo (*design*) de experimento.

Tabela 6.3: Projeto do experimento

	Sistema A	Sistema B
Grupos	Grupo I	Grupo II

Instrumentação

O acesso aos sistemas (objetos do estudo) foi realizado a partir dos computadores disponíveis no ambiente escolhido (laboratório da graduação). Um treinamento foi elaborado com o objetivo de apresentar aos participantes uma visão geral da abordagem e do sistema correspondente a ser utilizado para realizar as tarefas. Um roteiro para o experimento também foi elaborado de modo que quaisquer diferenças entre os desempenhos dos participantes não poderiam ser atribuídas à utilização de procedimentos diferentes.

Uma carta de consentimento foi utilizada para explicar os objetivos do estudo e confirmar o compromisso de que os dados pessoais e de desempenho dos participantes não seriam divulgados e que seriam utilizados apenas para os propósitos especificados. Além disso, essa carta de consentimento também serve para estabelecer um acordo entre o avaliador e os participantes com o objetivo de auxiliar a confirmar o compromisso profissional de participar seriamente do experimento [Preece et al. 2002].

A caracterização dos participantes foi definida por meio de um questionário de levantamento (*survey*). Esse questionário consiste em várias perguntas para coletar informação demográfica básica (por exemplo, gênero, idade, tempo de serviço) e detalhes da experiência do participante com relação a ferramentas de busca (*search engines*), navegadores (*browsers*), catálogos *on-line* de biblioteca e experiência com relação à CoTeia.

As descrições das tarefas foram baseadas na classificação de Shneiderman [1997]: pesquisa específica (*specific fact-finding*), pesquisa estendida (*extended fact-finding*), navegação aberta (*open-ended browsing*) e exploração de disponibilidade (*exploration of availability*). O objetivo foi cobrir o maior número possível de cenários usando o menor número possível de tarefas. As tarefas envolveram a pesquisa por informação

no contexto educacional (Tabela 6.4).

Para obter uma avaliação subjetiva com relação a cada uma das tarefas, principalmente sobre a satisfação do usuário com os resultados encontrados e sobre a facilidade de uso do sistema para a tarefa realizada, um pequeno questionário foi elaborado para ser respondido logo após a conclusão de cada tarefa.

Para obter uma avaliação final acerca do sistema que os participantes experimentaram, um questionário de fechamento foi elaborado. Esse questionário foi definido com base na entrevista semi-estruturada realizada com os participantes do estudo de usabilidade apresentado na Seção 6.2.

Tabela 6.4: Tarefas para quatro tipos de cenários

Tarefas		Cenários
T1	Encontrar páginas com informação sobre os tipos de dados das linguagens Java e Python.	pesquisa específica
T2	Encontrar páginas com nomes e e-mails dos monitores da disciplina “Algoritmos e Estruturas de Dados I” dos semestres anteriores.	pesquisa estendida
T3	Procurar por informações sobre como projetar aplicações para pessoas portadoras de necessidades especiais.	navegação aberta
T4	Procurar pelo material dos seminários apresentados pelos alunos das disciplinas relacionadas a Engenharia de Software da pós graduação.	exploração de disponibilidade

Avaliação da validade

As ameaças relacionadas à validade dos resultados de um experimento podem ser avaliadas segundo os seguintes tipos de validade [Wohlin et al. 2000]: conclusão, interna, construção e externa. A validade de conclusão (*conclusion validity*) está relacionada com as questões que afetam a habilidade de extrair a conclusão correta sobre as relações entre os dados. A validade interna (*internal validity*) está relacionada com as influências que podem afetar as variáveis independentes e que não foram consideradas no projeto do experimento. A validade de construção (*construct validity*) está relacionada com a generalização dos resultados para o conceito ou teoria definida no experimento. A validade externa (*external validity*) também está relacionada com a generalização, mas no sentido de verificar se as conclusões do estudo podem ser generalizadas para fora do contexto considerado no experimento. As maneiras utilizadas para minimizar os efeitos dos prováveis problemas são apresentadas a seguir:

- *Validade de conclusão.* Os problemas devidos às diferenças individuais (variabilidade dos participantes) não são esperados porque a amostra é bastante homogênea, considerando o tempo de curso dos participantes e os seus níveis de

experiência. A confiabilidade das medidas é obtida através da elaboração disciplinada dos instrumentos de coleta dos dados e dos avisos para preenchimento correto incluídos no roteiro a ser seguido na execução do experimento. Qualquer problema identificado com relação ao preenchimento desses instrumentos implicou na posterior remoção dos dados do participante para a análise e interpretação dos resultados (Subseção 6.3.3). A confiabilidade da implementação dos tratamentos é obtida através da padronização dos procedimentos realizados na execução do experimento. A execução do experimento em condições de laboratório e de forma controlada (com a presença de um supervisor) evita que quaisquer elementos fora do ambiente experimental pudessem causar distúrbios nos resultados (por exemplo, devido a interrupções).

- *Validade interna.* A simplicidade do experimento e a maneira como foi projetado e conduzido evitam muitos dos problemas que afetam esse tipo de validade, por exemplo, aqueles relacionados à execução do experimento e seleção dos participantes. A realização de um teste piloto também teve papel importante para ajustar a instrumentação e determinar o tempo aproximado de duração de cada etapa do experimento. A dinâmica obtida com as diversas sessões evitou que houvesse reações negativas dos participantes ao longo da execução do experimento. De fato, não foram observados casos aparentes de que os participantes estivessem afetados negativamente. Ao contrário, os participantes estavam motivados a participar seriamente do experimento.
- *Validade de construção.* A definição do experimento foi especificada com antecedência, segundo a estrutura proposta por Wohlin et al. [2000], para evitar os problemas decorrentes de declarações pré-operacionais inadequadas e/ou obscuras. As variáveis consideradas para o desempenho, tempo para completar tarefas e relevância dos resultados encontrados, foram definidas claramente. A relevância, em especial, além de ser determinada a partir das medidas de revocação e precisão, também é verificada por um julgamento subjetivo no questionário aplicado após cada uma das tarefas para evitar desvios (*bias*) das estimativas estatísticas devido ao uso de um único método de medição. Os resultados estatísticos desses tipos diferentes de medidas podem ser confrontados.
- *Validade externa.* A maior dificuldade em generalizar os resultados para outros alunos poderia estar relacionada com a homogeneidade da amostra considerando o tempo de curso dos participantes e os seus níveis de experiência. No entanto, é perfeitamente possível e interessante replicar o experimento para comparar o desempenho de alunos com níveis de experiência diferentes, por exemplo, a execução do experimento com calouros. Além disso, a ferramenta CoTeia, sem considerar as suas funcionalidades associadas a Wikis, consiste em um conjunto de páginas que derivam seu conteúdo dinamicamente a partir de arquivos ou base de dados, de maneira semelhante a muitos *websites*. Portanto, as páginas wiki armazenadas pela ferramenta podem ser consideradas

um conjunto de recursos disponíveis a partir da Web. Ainda com relação à validade externa, o pequeno número de tarefas poderia ser considerado um aspecto negativo para a generalização. No entanto, um possível efeito negativo na validade é minimizado por meio da diversificação das tarefas segundo os tipos de cenários descritos por Shneiderman [1997].

6.3.3 Operação do experimento

Nesta subseção são descritas as etapas da operação do experimento: preparação, execução e validação dos dados.

Preparação

A preparação envolveu a elaboração com antecedência do treinamento, da carta de consentimento, dos formulários dos questionários, do texto das tarefas e do roteiro a ser seguido na execução do experimento. O material impresso que foi fornecido aos participantes para a realização do experimento foi um livreto contendo: carta de consentimento; questionário de levantamento; descrições das tarefas e questionários pós-tarefa; e questionário de fechamento (Apêndice B).

Execução

O experimento completo foi realizado em três dias devido aos horários nos quais os participantes estavam disponíveis. Em cada dia ocorreram duas etapas, uma para cada um dos sistemas. Cada etapa do experimento foi realizada em cinco sessões, resumidas a seguir:

1. Os participantes foram informados dos objetivos do estudo. A informação transmitida incluiu uma visão geral do experimento, a quantidade de tempo aproximada do procedimento, os tipos de dados que seriam coletados e como eles seriam analisados. Além disso, os participantes também foram motivados acerca da importância da participação profissional no estudo.
2. O treinamento foi aplicado aos participantes. Além da apresentação do sistema que seria utilizado, também foi fornecida uma oportunidade para usar o sistema de maneira que os participantes pudessem familiarizar-se com o objeto de estudo.
3. Os participantes assinaram a carta de consentimento e preencheram o questionário de levantamento.
4. Os participantes completaram as tarefas, anotando o tempo de início e fim, e também salvando os resultados encontrados. Após cada tarefa, o questionário correspondente aquela tarefa foi respondido.
5. Os participantes responderam o questionário de fechamento.

Validação dos dados

Os dados foram coletados de 48 participantes. No entanto, os dados de 14 participantes foram removidos do conjunto porque não foram considerados válidos para o estudo. Os 14 participantes foram removidos devido ao:

- Não preenchimento correto dos questionários. Seis (6) dos participantes esqueceram-se de responder questões, principalmente do questionário pós-tarefa.
- Não anotação do tempo de início e fim da tarefa. Após concluir a tarefa, três (3) dos participantes não preencheram os dados com relação ao tempo de duração da tarefa.
- Não armazenamento dos resultados das tarefas. Dois (2) dos participantes não armazenaram corretamente os resultados encontrados.
- Perfil discrepante da média dos participantes. O tempo de graduação de dois (2) dos participantes era muito acima da média dos demais participantes e o nível de experiência com navegadores, ferramentas de busca e CoTeia de um (1) dos participantes era muito abaixo da média.

Para a análise estatística e interpretação dos resultados foram considerados, portanto, os dados de 34 participantes.

6.3.4 Análise e interpretação dos resultados

Nesta subseção é apresentada a análise e interpretação dos resultados. Inicialmente, a estatística descritiva é apresentada para melhor compreensão da natureza dos dados e para identificar anormalidades no conjunto de dados (denominadas de *outliers*). Na redução do conjunto de dados, as anormalidades encontradas são descritas e discutidas. No teste de hipóteses são realizados os testes estatísticos.

Estatística descritiva

Neste primeiro passo para analisar os dados, a estatística descritiva é utilizada para resumir e descrever as principais características dos dados coletados. Primeiro são apresentados os resultados estatísticos dos grupos com relação ao tempo para completar as tarefas, e depois aqueles com relação ao desempenho na realização das tarefas.

Na Tabela 6.5 são apresentados os resultados estatísticos dos grupos para a média e desvio padrão com relação ao tempo para completar as tarefas. Na Figura 6.6 é apresentado o gráfico *boxplot* com relação ao total geral desses dados. A partir da Tabela 6.5 pode-se notar que os participantes do Grupo I tiveram desempenho melhor (tempo menor) do que os do Grupo II com relação ao tempo total para completar todas

as tarefas. Entretanto, o desvio padrão do Grupo I é bastante alto. Além disso, a partir dos gráficos da Figura 6.6 pode-se notar que os dados do Grupo I contêm o menor valor e o maior valor dos dois grupos, como também apresentam variabilidade bastante alta. Essas informações são indicações de que uma nova investigação dos dados é necessária (ou pelo menos recomendada). Neste caso, os dados correspondentes a cada uma das tarefas são analisados na sequência.

Tabela 6.5: Resultados estatísticos dos grupos com relação ao tempo para completar as tarefas; os valores estão em minutos e segundos

Tarefas	Grupo I		Grupo II	
	Média	DP	Média	DP
T1	7:45	3:46	9:03	2:32
T2	7:56	2:55	7:42	2:10
T3	7:45	3:04	6:31	2:20
T4	6:14	3:22	8:07	3:35
Total	29:42	9:46	31:24	6:41

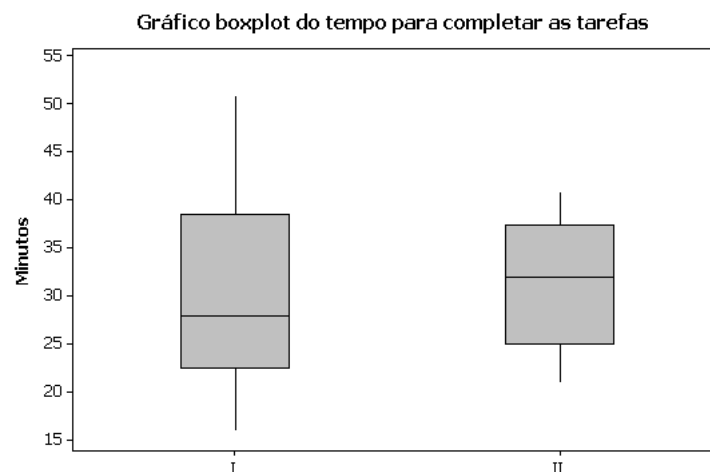


Figura 6.6: Gráficos *boxplot* dos dados dos grupos com relação ao tempo total para completar todas as tarefas

A partir da Tabela 6.5 pode-se notar que os participantes do Grupo I tiveram desempenho melhor do que os do Grupo II para as tarefas 1 e 4, enquanto que o inverso ocorre com relação às tarefas 2 e 3. Na Figura 6.7 são apresentados os gráficos *boxplot* do tempo para completar cada uma das tarefas. A partir do gráfico relativo à tarefa 1 pode-se notar que existe uma anormalidade (*outlier*) nos dados do Grupo II; entretanto, além de não ser o caso de um valor extremo (mais de dez vezes menor do que o valor mais alto), esse *outlier* parece não ter influenciado estatisticamente o resultado (isto é, mesmo sendo um valor muito baixo, a média do Grupo II para esta tarefa é maior do que a do Grupo I). A partir do gráfico relativo à tarefa 2 pode-se notar que o padrão das caixas é bem parecido (observe a mediana). Os gráficos relativos às tarefas 3 e 4 não revelam nada relevante.

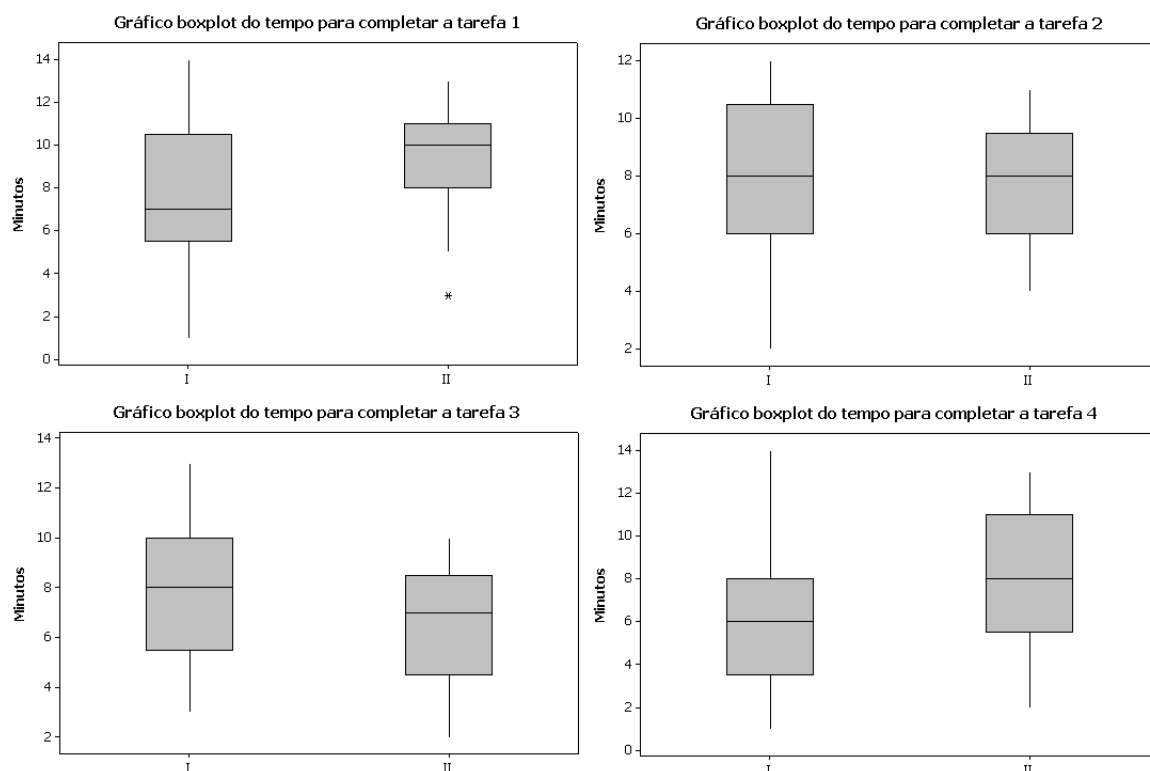


Figura 6.7: Gráficos *boxplot* dos dados dos grupos com relação ao tempo para completar as tarefas

Na Tabela 6.6 são apresentados os resultados estatísticos dos grupos para a média e desvio padrão com relação às medidas consideradas para a relevância dos resultados das tarefas: revocação (r) e precisão (P). Na Figura 6.8 são apresentados os gráficos *boxplot* com relação ao total geral desses dados. A partir da Tabela 6.6 pode-se notar que os participantes do Grupo I também tiveram desempenho melhor do que os do Grupo II com relação à relevância dos resultados encontrados. Além disso, a partir dos gráficos da Figura 6.8 pode-se notar que existe um padrão de que os participantes do Grupo I têm desempenho melhor para ambas as medidas. Adicionalmente, os gráficos *boxplot* dos dados dos grupos com relação a cada uma das tarefas também foram analisados e não revelaram nada que fosse relevante (por exemplo, *outliers*); portanto, esses gráficos não são apresentados.

Tabela 6.6: Resultados estatísticos dos grupos com relação às medidas de revocação (r) e precisão (P) para as tarefas

Tarefas	Grupo I				Grupo II			
	r		P		r		P	
	Média	DP	Média	DP	Média	DP	Média	DP
T1	0,27	0,12	0,97	0,10	0,31	0,20	0,93	0,16
T2	0,32	0,17	0,56	0,21	0,08	0,07	0,34	0,37
T3	0,30	0,18	0,66	0,33	0,23	0,13	0,70	0,37
T4	0,37	0,26	0,65	0,23	0,18	0,12	0,40	0,31
Total	0,31	0,12	0,66	0,14	0,22	0,10	0,58	0,17

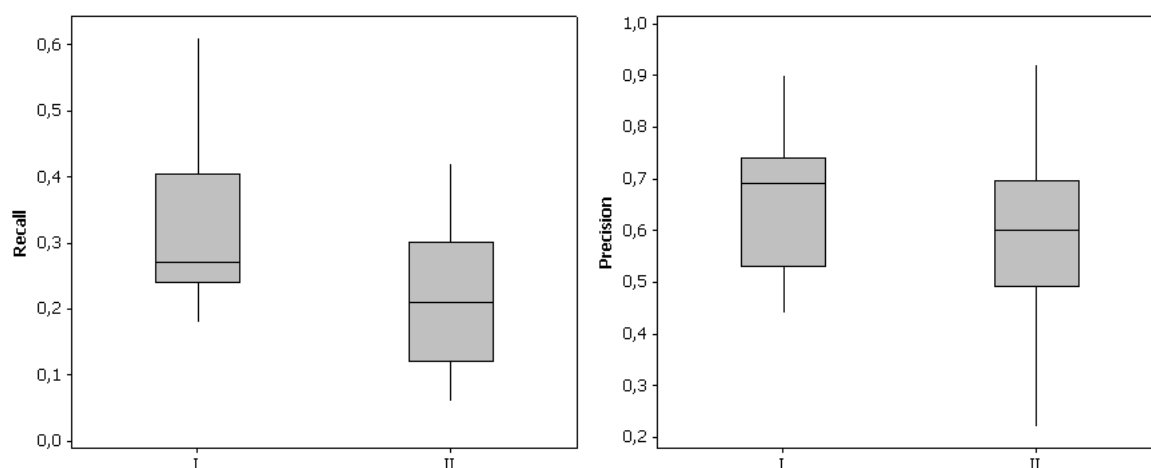


Figura 6.8: Gráficos *boxplot* dos dados dos grupos com relação às medidas de revocação (r) e precisão (P) do total de todas as tarefas

Na Tabela 6.7 são apresentados os resultados estatísticos dos grupos com relação ao julgamento subjetivo da relevância dos resultados das tarefas, realizado pelos participantes após cada uma das tarefas (Apêndice B). A partir dessa tabela pode-se notar que a relevância atribuída pelos participantes do Grupo I aos resultados obtidos nas tarefas realizadas é maior do que a do Grupo II.

Para confrontar os resultados estatísticos com relação à relevância obtida pelos dois métodos de medição (Tabela 6.6 e 6.7) é adequado utilizar uma única medida que combine revocação e precisão. Nesse caso, uma medida que pode ser utilizada consiste na média harmônica (F) dessas duas medidas [Baeza-Yates e Ribeiro-Neto 1999]. A função F assume valores dentro do intervalo $[0, 1]$. Quando nenhum documento relevante é recuperado, a função assume valor zero; e quando todos os documentos recuperados são relevantes, a função assume valor um. Além disso, a média harmônica (F) assume um valor alto quando revocação e precisão são valores altos. Os resultados da média harmônica (F) são apresentados na Tabela 6.8. A partir da comparação entre a Tabela 6.7 e a Tabela 6.8 pode-se notar que os resultados possuem a mesma tendência com relação ao total, mas são conflitantes com relação às tarefas 1 e 3.

Tabela 6.7: Resultado estatístico dos grupos com relação ao julgamento subjetivo de relevância dos resultados das tarefas (questionário pós-tarefa)

Tarefas	Grupo I		Grupo II	
	Média	DP	Média	DP
T1	5,65	0,86	4,41	1,50
T2	5,59	0,94	4,35	1,54
T3	4,00	1,37	4,06	1,85
T4	5,12	1,05	4,59	1,66
Total	5,07	1,25	4,34	1,63

Finalmente, na Tabela 6.9 é apresentado um resumo das questões e dos resulta-

Tabela 6.8: Resultados da média harmônica (F) entre as medidas de revocação e precisão

Tarefas	Grupo I		Grupo II	
	Média	DP	Média	DP
T1	0,41	0,14	0,44	0,23
T2	0,37	0,14	0,13	0,11
T3	0,35	0,13	0,32	0,15
T4	0,41	0,19	0,23	0,15
Total	0,40	0,07	0,31	0,12

dos estatísticos do questionário de fechamento aplicado na última sessão do estudo experimental. O questionário consistiu em nove questões e os julgamentos foram feitos em escalas Likert de 7 pontos.

Tabela 6.9: Resumo das questões e dos resultados estatísticos do questionário de fechamento

Questões	Grupo I			Grupo II		
	Média	Mediana	DP	Média	Mediana	DP
Q1. Qual a frequência que você realiza atividades relacionadas a essas tarefas? ^a	4,24	4,00	0,97	4,41	4,00	1,37
Q2. Qual a relevância dos resultados obtidos? ^b	5,06	5,00	0,75	4,47	5,00	1,01
Q3. Qual a facilidade de compreender o sistema? ^b	5,41	5,00	0,94	5,59	6,00	0,87
Q4. Qual a facilidade de usar (operar) o sistema? ^b	5,24	5,00	0,97	5,47	6,00	0,87
Q5. Qual a utilidade do sistema? ^b	5,29	5,00	1,16	4,47	5,00	1,18
Q6. Quantas vezes você ficou perdido durante a navegação? ^a	2,82	3,00	0,95	2,88	3,00	1,05
Q7. Você considera que o sistema forneceu auxílio para completar as tarefas em menos tempo do que usando um navegador comum? ^b	5,29	5,00	1,40	4,12	4,00	1,27
Q8. Você considera o reuso futuro do sistema? ^b	5,47	6,00	1,33	4,41	5,00	1,58
Q9. Qual a sua satisfação pessoal em usar o sistema? ^b	5,00	5,00	0,71	4,29	5,00	1,45

^a Julgamentos foram feitos em uma escala Likert de 7 pontos (1 = nunca; 7 = muito frequentemente)

^b Julgamentos foram feitos em uma escala Likert de 7 pontos (1 = nenhuma; 7 = extrema)

Embora os resultados da Tabela 6.9 não possam ser interpretados conclusivamente, as médias obtidas nas questões Q5, Q7 e Q8 indicam uma boa impressão de receptividade dos participantes ao uso do Sistema A.

Redução do conjunto de dados

A redução do conjunto de dados consiste em identificar as anormalidades (*outliers*) através da análise dos dados coletados pelos formulários e da estatística descritiva, ao invés de somente na execução do experimento (realizada durante a validação dos dados). O procedimento realizado quando uma anormalidade é identificada pode ser o seguinte [Wohlin et al. 2000]:

- Se a anormalidade ocorre devido a um evento estranho ou raro que nunca pode acontecer novamente, o dado poderia ser excluído.
- Se a anormalidade ocorre devido a um evento raro que pode acontecer novamente, não é aconselhável excluir o dado da análise porque existe informação relevante contida nele.
- Se a anormalidade ocorre devido a uma variável que não foi considerada antes, ela pode passar a fazer parte do modelo, ou dois modelos separados podem ser derivados.

Para este experimento foram analisados os dados dos formulários referentes a todas as anormalidades (*outliers*) identificadas por meio de gráficos *boxplot*. Essa análise não revelou qualquer ocorrência devido a um evento estranho ou raro que nunca poderia acontecer novamente. Portanto, todos os dados referentes às anormalidades identificadas foram mantidos no conjunto de dados.

Teste de hipóteses

Formalmente, a hipótese foi definida da seguinte maneira:

- Hipótese nula, H_0 : O desempenho dos usuários na realização de tarefas utilizando o Sistema A é inferior que utilizando o Sistema B.
- Hipótese alternativa, H_1 : O desempenho dos usuários na realização de tarefas utilizando o Sistema A é superior que utilizando o Sistema B.

As variáveis dependentes de desempenho consideradas para o experimento são o tempo gasto para completar as tarefas e a relevância dos resultados encontrados, esta última determinada pelas medidas de revocação e precisão. Além disso, também são consideradas as medidas de relevância subjetiva (do questionário pós-tarefa) e da média harmônica (F) entre revocação e precisão. Os testes foram conduzidos para comparar as médias das amostras para cada uma dessas medidas (tempo, revocação e precisão). Nesse caso, o *unpaired t-test* foi utilizado.

De acordo com os testes realizados, não existem diferenças significativas com relação ao tempo para completar as tarefas ($t(32) = -0,59$, $p = 0,557$) e com relação à relevância dos resultados para a medida de precisão ($t(32) = 1,35$, $p = 0,187$). Para a medida de revocação, a hipótese nula (H_0) foi rejeitada ($t(32) = 2,41$, $p = 0,022$), considerando o nível de significância de 95%. A hipótese nula (H_0) também foi rejeitada para as medidas de relevância subjetiva ($t(32) = 2,37$, $p = 0,012$) e média harmônica (F) ($t(25) = 2,67$, $p = 0,006$). Em última análise, os testes fornecem indicações estatísticas de que os participantes tiveram melhor desempenho usando o Sistema A (ENS) do que usando o Sistema B (busca por palavra-chave).

6.3.5 Resumo e conclusões

A finalidade deste estudo experimental é comparar o desempenho dos usuários na realização de tarefas para encontrar informação nos repositórios CoTeia utilizando o ENS com o desempenho daqueles utilizando a busca por palavra-chave. Nesse contexto, a seguinte hipótese foi investigada:

- O desempenho dos usuários na realização de tarefas utilizando o sistema ENS é superior que utilizando a ferramenta tradicional de busca por palavra-chave.

Os resultados do estudo fornecem indicações estatísticas para supor que os participantes do experimento, alunos de graduação do ICMC/USP, atuam melhor na abordagem do ENS. Os resultados mostraram que os participantes tiveram desempenho maior na realização das tarefas propostas em três das cinco medidas consideradas no estudo. Além disso, a hipótese de igualdade nas outras duas medidas não foi rejeitada.

Portanto, é importante a realização de novos experimentos para explorar os tipos de tarefa nos quais a abordagem do ENS é eficiente e os tipos nos quais ela é ineficiente.

6.4 Considerações finais

Neste capítulo, foram apresentados os estudos de usabilidade e experimental conduzidos para avaliar o sistema ENS, desenvolvido com base no modelo NAVE. Os resultados das análises desses estudos apóiam a proposta relativa às questões formuladas na metodologia (Capítulo 3).

No próximo capítulo, as conclusões desta tese são apresentadas, principalmente as contribuições originadas da investigação realizada e as possibilidades no que se refere a extensões do modelo e outros trabalhos futuros relacionados com os sistemas desenvolvidos.

Conclusão

7.1 Considerações iniciais

A Web Semântica é tratada como a próxima geração da Web, na qual a informação recebe significado para oferecer novas possibilidades de comunicação e interação. À medida que ocorre a proliferação de informação disponibilizada segundo os padrões e tecnologias da Web Semântica, novas abordagens para a navegação nesse contexto são necessárias.

Nesta tese foi apresentado um modelo de navegação exploratória, denominado de *Navigation and Exploration Model* (NAVE), para auxiliar o desenvolvimento de sistemas de suporte a atividades de *information searching* na Web Semântica. O principal objetivo deste modelo é servir como referência para o projeto e desenvolvimento de sistemas de navegação exploratória para a Web Semântica.

O modelo NAVE propõe orientações para o desenvolvimento de uma abordagem de navegação para fornecer suporte a um tipo de atividade conhecida por busca exploratória. A abordagem de navegação do modelo NAVE é fundamentada em níveis de atividades de *information seeking* (estratégia, estratagema, tática e movimento) e na estratégia de *ienteering*, observada em estudos de comportamento de *information seeking*.

O modelo NAVE é composto de diversos elementos: estágios, decisões e entidades. Os estágios e decisões são organizados em uma sequência representada por meio de um fluxograma. As entidades correspondem a estados internos do modelo que podem ser alterados no transcorrer dos estágios. Os estágios e decisões do modelo estão relacionados às categorias de técnicas de navegação, que são indicadas para fornecer suporte a cada um desses elementos. O modelo NAVE também relaciona um conjunto de recomendações para o projeto e desenvolvimento de sistemas de navegação exploratória segundo a abordagem proposta no modelo.

Um sistema de navegação exploratória, denominado de *Exploratory Navigation System* (ENS), foi desenvolvido com a finalidade de auxiliar a validação do modelo NAVE. Os estudos conduzidos para avaliar o sistema ENS forneceram informações funda-

mentais para compreender melhor a abordagem de navegação subjacente ao sistema e, conseqüentemente, auxiliar no desenvolvimento do modelo NAVE.

Neste capítulo são apresentadas as conclusões gerais deste trabalho, enfatizando suas principais contribuições, na Seção 7.2, e as sugestões para trabalhos futuros, na Seção 7.3. As considerações finais são apresentadas na Seção 7.4.

7.2 Contribuições

As contribuições da investigação conduzida neste trabalho podem ser destacadas nos seguintes itens:

- O *Navigation and Exploration Model* (NAVE), apresentado no Capítulo 4, para servir de referência para o projeto e desenvolvimento de sistemas de navegação exploratória para a infra-estrutura da Web Semântica. A contribuição consiste em um modelo composto de estágios, decisões, entidades e recomendações relacionadas com atividades de pesquisa por informação, com o qual os desenvolvedores de sistemas podem considerar diversas oportunidades para suporte automatizado à navegação em geral e para a Web Semântica em particular, de maneira a utilizar o potencial dos metadados.
- O *Exploratory Navigation System* (ENS), apresentado no Capítulo 5 desenvolvido com base no modelo NAVE, para auxiliar a pesquisa por informação na Web Semântica. O sistema ENS foi configurado para permitir a navegação nos repositórios de páginas wiki da CoTeia (domínio educacional). No ENS, foi desenvolvida uma *navigation engine* que integra diversas ferramentas de suporte à navegação para permitir ao usuário escolher a ferramenta adequada ou a melhor combinação de ferramentas de acordo com a sua estratégia, provavelmente ajustada ao seu nível de habilidade e conhecimento, à sua preferência e ao tipo de atividade que ele está realizando no momento.
- A avaliação do sistema ENS, apresentada no Capítulo 6, considerada um estudo de caso de avaliação de sistemas interativos para a pesquisa de informação. A avaliação do ENS combinou abordagens tanto qualitativas quanto quantitativas, além de diversos métodos para coleta de dados para tornar mais eficaz a aquisição de informações sobre esse sistema. As etapas da avaliação realizadas e a instrumentação produzida para suporte a essas etapas são exemplos que podem ser utilizados no contexto da avaliação de sistemas de recuperação de informação e/ou busca exploratória.
- O *Metadata Generator* (MeGen), apresentado na Subseção 5.3.3 que implementa diversos métodos automáticos de geração de metadados para páginas wiki educacionais. O sistema MeGen foi utilizado para gerar os elementos de metadados do *application profile* definido para a CoTeia para as páginas wiki dos repositórios disponíveis. A descrição de métodos para a geração automática de metadados e as classes do MeGen auxiliam a construção de sistemas geradores de

metadados, pois fornecem uma estrutura básica que pode ser utilizada para a implementação gradual e sistemática.

As contribuições não estão limitadas aos quatro itens apresentados anteriormente. Existiram outros resultados durante o desenvolvimento do trabalho que merecem registro, por exemplo: as contribuições nas inúmeras discussões iniciais dos projetos *Tecnologia da Informação para o Desenvolvimento da Internet Avançada – Aprendizado Eletrônico* (TIDIA-Ae)¹ e *Software Engineering Available for Everyone* (SAFE)² [Fortes et al. 2004]; a extensão da CoTeia com uma interface de metadados [Pansanato e Fortes 2005a,b]; a definição, com base nos padrões de metadados *Dublin Core* e LOM, de um *application profile* para páginas wiki educacionais, utilizada na CoTeia [Pansanato e Fortes 2006]; o esboço inicial de ontologia para tipo de material de aprendizagem (*Learning Material Type*), tipo de atividade de aprendizagem (*Learning Activity Type*) e área do conhecimento (*Knowledge Area*) [Pansanato e Fortes 2005a]; o estudo sobre acessibilidade, visando incorporar essa característica na navegação, e que teve como resultado dois capítulos de livro publicados [Fortes et al. 2005a,b]; os sistemas *Crawler* e *Indexer* desenvolvidos; o sistema para auxílio ao registro e análise da interação do usuário, utilizado na etapa de avaliação do ENS; as ferramentas Memória de trabalho e Marcar resultados, além de outras extensões (por exemplo, “*entities in context*” e “*patterns in context*”) para as ferramentas de navegação do ENS, as quais também são oportunidades para continuidade de trabalhos futuros; e as participações em cursos ministrados para a graduação, eventos e em seminários do grupo.

7.3 Trabalhos futuros

Alguns dos trabalhos que podem dar continuidade às contribuições apresentadas nesta tese são resumidos a seguir:

1. Estender o modelo NAVE para acomodar aprimoramentos que podem aparecer à medida que for utilizado no projeto e desenvolvimento de outros sistemas e também na análise de sistemas de suporte a atividades de busca exploratória. A classificação das técnicas de navegação também pode ser estendida para incorporar outras categorias à sua hierarquia principal, além de exemplos de aplicação.
2. Estender o modelo NAVE para incorporar um processo de avaliação para identificar falhas de projeto em sistemas desenvolvidos com base no modelo e para avaliar a conformidade de sistemas de busca exploratória com o modelo. Esse processo deve fornecer orientações para: identificar componentes na interface que correspondem às entidades do modelo; identificar as funcionalidades que fornecem suporte aos estágios e decisões; construir uma matriz para relacionar

¹<http://tidia-ae.iv.fapesp.br/>

²<http://safe.icmc.usp.br>

as funcionalidades identificadas com as categorias de técnicas de suporte; comparar as informações coletadas com as recomendações. O modelo NAVE não é completo, e esse não é o seu objetivo, mas uma avaliação com base no modelo pode auxiliar na identificação da ausência de características importantes que não foram consideradas no projeto de um sistema. Nessa direção, também são vislumbrados esforços para a especificação formal do modelo NAVE.

3. Especializar o modelo NAVE para o domínio de *e-learning*. As atividades de aprendizagem são classificadas como um tipo de busca exploratória [Marchionini 2006]. Este trabalho futuro consiste em incorporar características ao modelo NAVE específicas para o projeto e desenvolvimento de sistemas de suporte a atividades de aprendizagem. Essa proposta também consiste na investigação do uso da infra-estrutura da Web Semântica para atividades de aprendizagem. Nessa direção, uma possibilidade seria adotar a abordagem do modelo para um ambiente de *e-learning*, por exemplo, o TIDIA-Ae.
4. Evoluir a arquitetura do sistema ENS para um padrão de arquitetura baseada em componentes. A investigação proposta envolve a elaboração de padrões para a interação do usuário com ferramentas de navegação. À medida que as interfaces das ferramentas se tornam padrão de fato (ou maduras), elas podem evoluir para componentes de interface, com uma API para a comunicação padrão com uma aplicação (por exemplo, um *Web Service*). Por exemplo, a interface esperada para a busca por palavra-chave tem três componentes [Nielsen 2005]: uma caixa para digitar palavras, um botão “Buscar” para executar a busca e uma lista dos resultados. A interação do usuário com outras ferramentas não estão tão bem definidas. No entanto, algumas interfaces são fortes candidatas à padronização, por exemplo, a navegação por facetas. A padronização dessas interfaces (e os mecanismos associados) é difícil, mas é provável que seja satisfatória proporcionando reuso. Uma consequência dessa transformação em componentes é viabilizar, no futuro, que o sistema ENS seja disponibilizado como software livre.
5. Agregar novas ferramentas ao sistema ENS. Basicamente, o objetivo da investigação é estender o sistema ENS com ferramentas para recomendação [Adomavicius e Tuzhilin 2005], personalização [Teevan et al. 2005] e busca semântica [Guha et al. 2003]. Por um lado, a capacidade de executar estratégias completas, facilitada pela *navigation engine*, pode ficar mais poderosa. Por outro lado, com o aumento da quantidade de ferramentas (diferentes) disponíveis, o impacto no desempenho do usuário merece ser investigado.
6. Realizar mais experimentos com o sistema ENS. Essa investigação estaria relacionada com a validade externa, isto é, a capacidade de generalizar os resultados. A validade externa será maior e mais forte à medida que mais experimentos são realizados, por exemplo, com usuários e domínios diferentes. Nessa linha de investigação, um experimento interessante consiste em configurar o sistema ENS para permitir a navegação na documentação de linguagens de programação

(por exemplo, na documentação da PHP) e comparar o desempenho de usuários novatos e experientes utilizando uma ferramenta de busca *desktop* (por exemplo, Google Desktop³ ou MSN Desktop Search⁴) com o desempenho utilizando o sistema ENS.

7. Desenvolver ferramentas de suporte à avaliação de sistemas interativos de navegação. Durante a realização dos estudos (de usabilidade e experimental) para avaliar o sistema ENS, surgiram diversas questões relacionadas com a análise dos dados coletados. A dificuldade encontrada para analisar e comparar os dados coletados por métodos diferentes e inter-relacionados requer o desenvolvimento de ferramentas mais efetivas. Por exemplo, selecionar um trecho de áudio gravado em uma sessão de *think-aloud* e associar esse trecho com o registro da interação correspondente armazenado em um arquivo de *log*, para a apresentação simultânea de ambos quando necessário. O sistema para registro e análise da interação pode ser estendido para auxiliar a estatística descritiva, por exemplo, gerando gráficos automaticamente a partir dos arquivos de *log*. Além disso, *templates* padrão podem ser elaboradas para questionários *on-line*.
8. Estender o sistema MeGen. O reuso dos métodos implementados no MeGen é variável e é necessário um estudo da capacidade de adaptação e/ou extensão para outros domínios. Por um lado, os métodos genéricos podem ser reusados, por exemplo, classificação de idioma, extração de palavras-chave, exploração do contexto do sistema de arquivos, mecanismo de mapeamento. Por outro lado, os métodos dependentes do domínio (páginas wiki educacionais) são mais difíceis de serem reusados. Nesse contexto, o uso da especificação *Gleaning Resource Descriptions from Dialects of Languages* (GRDDL) [W3C 2006b] deve ser considerada, pois define um mecanismo para obter dados RDF a partir de documentos XML e de páginas XHTML.

7.4 Considerações finais

A pesquisa por informação é uma atividade fundamental na vida das pessoas. As tecnologias da Web Semântica fornecem a infra-estrutura necessária para o desenvolvimento de novas abordagens para auxiliar as pessoas a encontrar a informação que elas estão procurando.

A investigação de modelos de navegação para a Web Semântica pode fornecer informações para melhorar as interfaces e os mecanismos da próxima geração de sistemas de navegação. O modelo NAVE, proposto nesta tese, é um passo nessa direção.

A pesquisa e os trabalhos realizados, descritos nesta tese, serviram de apoio para dar prosseguimento a novas pesquisas, assim como viabilizaram um conjunto base para os conhecimentos do estado da arte relacionados com a navegação exploratória

³<http://desktop.google.com/>

⁴<http://desktop.msn.com/>

para a infra-estrutura da Web Semântica. Dessa maneira, os resultados, descritos como contribuições, representam avanços para o estado da arte em navegação exploratória.

Referências Bibliográficas

- Adomavicius, G. e Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749. DOI=<<http://doi.ieeecomputersociety.org/10.1109/TKDE.2005.99>>.
- Ahmed, I. e Blustein, J. (2005). Navigation in information space: How does spatial ability play a part? In *Proceedings of the IADIS International Conference on Web Based Communities 2005*, p. 119–125, Algarve, Portugal. IADIS. Disponível em: <http://www.iadis.net/dl/Search_list_open.asp?code=1286>. Acesso em: 4 jul. 2007.
- Arruda Jr, C. R. E., Izeke, C. A., e Pimentel, M. G. C. (2002). CoTeia: Uma Ferramenta Colaborativa de Edição Baseada na Web. In *Proceedings of the 8th Brazilian Symposium on Multimedia and Hypermedia Systems*, p. 371–374, Fortaleza, CE, Brazil. SBC.
- Arruda Jr, C. R. E. e Pimentel, M. G. C. (2001). Projeto e Implementação de um Sistema Colaborativo de Edição. *Revista Eletrônica de Iniciação Científica (REIC), Sociedade Brasileira de Computação (SBC)*, 1(2). Disponível em: <<http://www.sbc.org.br/index.php?language=1&subject=101>>. Acesso em: 25 abr. 2007.
- Aula, A., Jhaveri, N., e Käkä, M. (2005). Information search and re-access strategies of experienced web users. In *Proceedings of the 14th International Conference on World Wide Web*, p. 583–592, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/1060745.1060831>>.
- Baeza-Yates, R. e Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. New York: ACM Press e Addison-Wesley, first ed. 513 p.
- Balasubramanian, V. (1993). State of the Art Review on Hypermedia Issues And Applications. Website, E-Papyrus, Inc. Disponível em: <http://www.e-papyrus.com/hypertext_review/index.html>. Acesso em: 9 jan. 2007.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the on-line search interface. *Online Review*, 13(5):407–424. Disponível em: <<http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>>. Acesso em: 23 mar. 2005.

- Bates, M. J. (1990). Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5):575–591.
- Bates, M. J. (2005). Information and knowledge: an evolutionary framework for information science. *Information Research*, 10(4). Disponível em: <<http://InformationR.net/ir/10-4/paper239.html>>. Acesso em: 22 jan. 2007.
- Belkin, N., Dumais, S., Scholtz, J., e Wilkinson, R. (2004). Evaluating Interactive Information Retrieval Systems: Opportunities and Challenges. In *Proceedings of the Extended Abstracts on Human Factors in Computing Systems*, p. 1594–1595, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/985921.986162>>.
- Berners-Lee, T. (1989). Information Management: A Proposal. CERN, Geneva. Disponível em: <www.w3.org/People/Berners-Lee/WorldWideWeb.html>. Acesso em: 11 jul. 2006.
- Berners-Lee, T. (2000). Semantic Web – XML2000: Architecture. Website. Disponível em: <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>>. Acesso em: 15 mai. 2005.
- Berners-Lee, T. e Cailliau, R. (1990). WorldWideWeb: Proposal for a HyperText Project. CERN, Geneva. Disponível em: <www.w3.org/Proposal.html>. Acesso em: 11 jul. 2006.
- Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F., e Secret, A. (1994). The World-Wide Web. *Communications of the ACM*, 37(8):76–82. DOI=<<http://doi.acm.org/10.1145/179606.179671>>.
- Berners-Lee, T., Fielding, R., e Masinter, L. (1998). Uniform Resource Identifiers (URI): Generic Syntax. Request for Comments (RFC) 2396, The Internet Society. Disponível em: <<http://www.ietf.org/rfc/rfc2396.txt>>. Acesso em: 15 fev. 2004.
- Berners-Lee, T., Hendler, J., e Lassila, O. (2001). The Semantic Web. *Scientific American*. Disponível em: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>. Acesso em: 30 mar. 2005.
- Booth, W. C., Colomb, G. G., e Williams, J. M. (2005). *A arte da pesquisa*. São Paulo: Martins Fontes, segunda ed. 351 p.
- Borlund, P. (2003). The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3). Disponível em: <<http://informationr.net/ir/8-3/paper152.html>>. Acesso em: 21 fev. 2006.
- Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P., e Stal, M. (1996). *Pattern-Oriented Software Architecture: A System of Patterns*. Chichester, England: John Wiley & Sons, first ed. 476 p.
- Bush, V. (1945). As We May Think. *The Atlantic Monthly*. Disponível em: <<http://www.theatlantic.com/doc/194507/bush>>. Acesso em: 17 jan. 2007.

- Capurro, R. e Hjørland, B. (2003). The concept of information. *Annual Review of Information Science & Technology (ARIST)*, 37(1):343–411.
- Carroll, J. M. (2000). Introduction to the special issue on “Scenario-Based Systems Development”. *Interacting with Computers*, 13(1):41–42.
- Case, D. O. (2002). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Academic Press, San Diego, CA, USA. 352 p.
- Cervo, A. L., Bervian, P. A., e Da Silva, R. (2007). *Metodologia científica*. São Paulo: Pearson Prentice Hall, sexta ed. 176 p.
- Choo, C. W., Detlor, B., e Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday*, 5(2). Disponível em: <http://firstmonday.org/issues/issue5_2/choo/index.html>. Acesso em: 26 jan. 2004.
- Conklin, J. (1987). Hypertext: An Introduction and Survey. *IEEE Computer*, 10(9):17–41. DOI=<<http://dx.doi.org/10.1109/MC.1987.1663693>>.
- Cousins, S. B., Paepcke, A., Winograd, T., Bier, E. A., e Pier, K. (1997). The digital library integrated task environment (DLITE). In *Proceedings of the 2nd ACM International Conference on Digital Libraries*, p. 142–151, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/263690.263807>>.
- Danielson, D. R. (2002). Web navigation and the behavioral effects of constantly visible site maps. *Interacting with Computers*, 14(5):601–618. DOI=<[http://dx.doi.org/10.1016/S0953-5438\(02\)00024-3](http://dx.doi.org/10.1016/S0953-5438(02)00024-3)>.
- DCMI (2002). Expressing Qualified Dublin Core in RDF / XML. DCMI Proposed Recommendation, Dublin Core Metadata Initiative. Disponível em: <<http://dublincore.org/documents/dcq-rdf-xml/>>. Acesso em: 14 jun. 2004.
- DCMI (2003). Dublin Core Metadata Element Set, Version 1.1: Reference Description. DCMI Recommendation, Dublin Core Metadata Initiative. Disponível em: <<http://dublincore.org/documents/dces/>>. Acesso em: 24 mai. 2004.
- Dix, A., Finlay, J., Abowd, G., e Beale, R. (2004). *Human-Computer Interaction*. Prentice-Hall, Harlow, England, third ed. 834 p.
- Dürst, M. e Freytag, A. (2007). Unicode in XML and other Markup Languages. Unicode Technical Report #20 / W3C Working Group Note, World Wide Web Consortium / Unicode Consortium. Disponível em: <<http://www.w3.org/TR/unicode-xml/>>. Acesso em: 30 mar. 2005.
- Dumas, J. S. e Redish, J. C. (1999). *A Practical Guide to Usability Testing*. Exeter, UK: Intellect, rev. ed. 416 p.

- Duval, E. e Hodgins, W. (2003). A LOM Research Agenda. In *Alternate Paper Tracks Proceedings of the 12th International Conference on World Wide Web*, Budapest, Hungary. Disponível em: <<http://www2003.org/cdrom/papers/alternate/P659/p659-duval.html.html>>. Acesso em: 29 mar. 2005.
- Ecma (1999). ECMAScript Language Specification. Standard ECMA-262, 3rd edition, Ecma International. Disponível em: <<http://www.ecma-international.org/publications/standards/Ecma-262.htm>>. Acesso em: 16 jul. 2007.
- Ellis, D. (1989). A behavioural model for information retrieval system design. *Journal of Information Science*, 15(4-5):237–248.
- Ellis, D. e Haugan, M. (1997). Modeling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment. *The Journal of Documentation*, 53(4):384–403.
- Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., e Berners-Lee, T. (1999). Hypertext Transfer Protocol – HTTP/1.1. Request for Comments (RFC) 2616, The Internet Society. Disponível em: <<http://www.ietf.org/rfc/rfc2616.txt>>. Acesso em: 20 fev. 2004.
- Fortes, R. M., Lara, S. M. A., Freire, A. P., e Pansanato, L. T. E. (2005a). *Web e Multimídia: Desafios e Soluções*, Capítulo 7: Acessibilidade no Projeto de Aplicações Web, p. 197–225. PUC Minas, Poços de Caldas, MG.
- Fortes, R. M., Pansanato, L. T. E., Lara, S. M. A., e Freire, A. P. (2005b). *Minicursos do XVI Simpósio Brasileiro de Informática na Educação (SBIE 2005)*, Capítulo 2: Universalização do Acesso a Conteúdo Educacional em Aplicações Web, p. 21–42. Editar Editora Associada, Juiz de Fora, MG.
- Fortes, R. P. M., Silva, M. A. G., Freire, A. P., e Junqueira, D. C. (2004). SAFE - Software Engineering Available for Everyone. In *Anais do V Workshop sobre Software Livre (WSL 2004), parte do V Fórum Internacional de Software Livre (FISL 2004)*, p. 203–206, Porto Alegre-SP, Brasil.
- Fumagalli, L. C., Pansanato, L. T. E., e Fortes, R. P. M. (2003). Documentation Process in Interactive Systems: A Study Case to Abstract its Structure. In *Proceedings of the 2nd International Information and Telecommunication Technologies Symposium (I2TS 2003)*, Florianópolis, SC, Brazil. Fundação Barddal de Educação e Cultura.
- Furuta, R., F. M. Shipman, I., Marshall, C. C., Brenner, D., e Hsieh, H. (1997). Hypertext paths and the World-Wide Web: experiences with Walden's Paths. In *Proceedings of the 8th ACM Conference on Hypertext*, p. 167–176, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/267437.267455>>.

- Gloor, P. A. (1997). *Elements of Hypermedia Design: Techniques for Navigation and Visualization in Cyberspace*. Birkhäuser Boston, Birkhäuser Boston, first ed. 416 p. Disponível em: <<http://www.ickn.org/elements/>>. Acesso em: 30 jul. 2005.
- Guha, R., McCool, R., e Miller, E. (2003). Semantic search. In *Proceedings of the 12th International Conference on World Wide Web*, p. 700–709, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/775152.775250>>.
- Guzdial, M. (1999). Supporting learners as users. *The Journal of Computer Documentation*, 23(2):3–13.
- Hearst, M. A. (1999). Baeza-Yates, R., Ribeiro-Neto, B., *Modern Information Retrieval*, Capítulo User Interfaces and Visualization, p. 257–323. New York: ACM Press e Addison-Wesley, first ed. 513 p.
- Hearst, M. A. (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM*, 49(4):59–61. DOI=<<http://doi.acm.org/10.1145/1121949.1121983>>.
- Hölscher, C. e Strube, G. (2000). Web Search Behavior of Internet Experts and Newbies. In *Proceedings of the 9th International World Wide Web Conference on Computer Networks*, p. 337–346, Amsterdam, The Netherlands. Disponível em: <<http://www9.org/w9cdrom/81/81.html>>. Acesso em: 26 jan. 2004.
- IEEE LTSC (2002). IEEE 1484.12.1:2002 Draft Standard for Learning Object Metadata. Draft Standard, IEEE Learning Technology Standards Committee. Disponível em: <http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf>. Acesso em: 25 jun. 2004.
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52(1):3–50.
- Ingwersen, P. e Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer-Verlag New York, Inc., Secaucus, NJ, USA. 448 p.
- ISO (1986). ISO Standard 8879:1986 Information processing – Text and office systems – Standard Generalized Markup Language (SGML). ISO International Standard, International Organization for Standardization.
- ISO/IEC (1998). Information technology – Software product evaluation – Part 5: Process for evaluators. ISO/IEC International Standard 14598-5:1998, International Organization for Standardization/International Electrotechnical Commission.
- Jhaveri, N. e Räihä, K. (2005). The advantages of a cross-session web workspace. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, p. 1949–1952, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/1056808.1057064>>.

- Jühne, J., Jensen, A. T., e Grønbæk, K. (1998). Ariadne: a Java-based guided tour system for the World Wide Web. In *Proceedings of the 7th International Conference on World Wide Web*, p. 131–139, Amsterdam, The Netherlands, The Netherlands. Elsevier Science Publishers B. V. DOI=<[http://dx.doi.org/10.1016/S0169-7552\(98\)00124-X](http://dx.doi.org/10.1016/S0169-7552(98)00124-X)>.
- Johansson, R. (2006). Developing With Web Standards: Recommendations and best practices. Website, 456 Berea Street. Disponível em: <http://www.456bereastreet.com/lab/developing_with_web_standards/>. Acesso em: 22 out. 2006.
- Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5):361–371. Disponível em: <<http://www.viktoria.se/~dixi/BISON/resources/kuhlthau1991.pdf>>. Acesso em: 20 jun. 2005.
- Leite, J. C. S. P., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G. N., Hadad, G. D. S., e Oliveros, A. (1997a). Enhancing a Requirements Baseline with Scenarios. In *Proceedings of the 3rd IEEE International Symposium on Requirements Engineering (RE'97)*, p. 44–53, Washington, DC, USA. IEEE Computer Society. Disponível em: <www-di.inf.puc-rio.br/julio/Slct-pub/baseline.pdf>. Acesso em: 26 fev. 2007.
- Leite, J. C. S. P., Rossi, G., Balaguer, F., Maiorana, V., Kaplan, G. N., Hadad, G. D. S., e Oliveros, A. (1997b). Enhancing a Requirements Baseline with Scenarios. *Requirements Engineering Journal*, 2(4):184–198.
- Leuf, B. e Cunningham, W. (2001). *The Wiki Way: Quick Collaboration on the Web*. Boston: Addison-Wesley. 440 p.
- Lowe, D. e Hall, W. (1999). *Hypermedia & the Web: An engineering approach*. John Wiley & Sons, Chichester, England. 626 p.
- m. c. schraefel, Wilson, M., Russell, A., e Smith, D. A. (2006). mSpace: improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, 49(4):47–49. DOI=<<http://doi.acm.org/10.1145/1121949.1121980>>.
- Macedo, A. A., Bulcão Neto, R. F., e Pimentel, M. G. C. (2001). Autoria Colaborativa na Web: Experiências e Reflexões sobre a CoWeb. *Revista Brasileira de Informática na Educação (RBIE), Sociedade Brasileira de Computação (SBC)*, (9):9–25.
- Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge, UK: Cambridge University Press. 236 p.
- Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46. DOI=<<http://doi.acm.org/10.1145/1121949.1121979>>.

- Marchionini, G. e Brunk, B. (2003). Towards a General Relation Browser: A GUI for Information Architects. *Journal of Digital Information*, 4(1). Disponível em: <<http://jodi.tamu.edu/Articles/v04/i01/Marchionini/>>. Acesso em: 20 jan. 2007.
- Marconi, M. A. e Lakatos, E. M. (2004). *Metodologia Científica*. Atlas, São Paulo, SP, quarta ed. 305 p.
- Marshall, C. C. e Shipman, F. M. (2003). Which semantic web? In *Proceedings of the 14th ACM Conference on Hypertext and Hypermedia*, p. 57–66, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/900051.900063>>.
- Müller-Prove, M. (2002). Vision and Reality of Hypertext and Graphical User Interfaces. Technical Report FBI-HH-B-237/02, University of Hamburg, Department of Informatics. 122 p. Disponível em: <<http://www.mprove.de/diplom/>>. Acesso em: 25 jan. 2005.
- Mukherjea, S. (1999). Information Visualization for Hypermedia Systems. *ACM Computing Surveys*, 31(4). DOI=<<http://doi.acm.org/10.1145/345966.345984>>.
- Navarro-Prieto, R., Scaife, M., e Rogers, Y. (1999). Cognitive Strategies in Web Searching. In *Proceedings of the 5th Conference on Human Factors & the Web*, Gaithersburg, Maryland, USA. Disponível em: <<http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html>>. Acesso em: 26 jan. 2004.
- Nielsen, J. (1990). *Hypertext and Hypermedia*. Academic Press, Boston, MA. 272 p.
- Nielsen, J. (1995). *Multimedia and Hypertext: The Internet and Beyond*. Academic Press Professional, Boston, MA. 480 p.
- Nielsen, J. (1999). User Interface Directions for the Web. *Communications of the ACM*, 42(1):65–72. DOI=<<http://doi.acm.org/10.1145/291469.291470>>.
- Nielsen, J. (2005). Reviving Advanced Hypertext. Website, Jakob Nielsen's Alertbox. Disponível em: <<http://www.useit.com/alertbox/20050103.html>>. Acesso em: 11 jul. 2007.
- Nielsen, J. (2007). Breadcrumb Navigation Increasingly Useful. Website, Jakob Nielsen's Alertbox. Disponível em: <<http://www.useit.com/alertbox/breadcrumbs.html>>. Acesso em: 11 jul. 2007.
- Nielsen, J. e Loranger, H. (2006). *Prioritizing Web Usability*. New Riders, Berkeley, CA, USA. 432 p.
- Nilsson, M., Palmér, M., e Brase, J. (2003). The LOM RDF binding – principles and implementation. In *Proceedings of the 3rd Annual ARIADNE Conference*, p. 1–7, Katholieke Universiteit Leuven, Belgium. Disponível em: <<http://cid.nada.kth.se/pdf/CID-243.pdf>>. Acesso em: 16 jun. 2004.

- O'Day, V. L. e Jeffries, R. (1993). Orienteering in an information landscape: how information seekers get from here to there. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '93)*, p. 438–445, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/169059.169365>>.
- Oldakowski, R., Bizer, C., e Westphal, D. (2005). RAP: RDF API for PHP. In *Proceedings of the 2nd European Semantic Web Conference (ESWC), Workshop on Scripting for the Semantic Web*, p. 1–10, Heraklion, Greece. Disponível em: <<http://www.semanticscripting.org/SFSW2005/papers/Oldakowski-RAP.pdf>>. Acesso em: 17 jan. 2006.
- Pansanato, L. T. E. e Fortes, R. P. M. (2004). Uma Análise de Metadados para o Acesso Unificado às Informações dos Repositórios de Ferramentas de Software Livre. In *Anais do II Congresso de Tecnologias para Gestão de Dados e Metadados do Cone Sul*, Ponta Grossa, PR. Universidade Tuiuti do Paraná. Disponível em: <<http://conged.deinfo.uepg.br/iiconged/Artigos/artigo2.pdf>>. Acesso em: 5 set. 2007.
- Pansanato, L. T. E. e Fortes, R. P. M. (2005a). Strategies for automatic LOM metadata generating in a web-based CSCL tool. In *Proceedings of the 11th Brazilian Symposium on Multimedia and the Web (WebMedia'05)*, p. 1–8, Pocos de Caldas, MG, Brazil. ACM Press. DOI=<<http://doi.acm.org/10.1145/1114223.1114231>>.
- Pansanato, L. T. E. e Fortes, R. P. M. (2005b). Strategies for filling out LOM metadata fields in a Web-based CSCL tool. In *Proceedings of the Third Latin American Web Congress (LA-WEB 2005)*, p. 187–190, Buenos Aires, Argentina. IEEE Computer Society. DOI=<<http://dx.doi.org/10.1109/LAWEB.2005.35>>.
- Pansanato, L. T. E. e Fortes, R. P. M. (2005c). Uma pesquisa em documentos Web baseada em estratégia de orientação. In *Anais do XI Simpósio Brasileiro de Multimídia e Web (WebMedia'05)*, p. 238–240, Poços de Caldas, MG, Brazil. SBC.
- Pansanato, L. T. E. e Fortes, R. P. M. (2006). MeGen: Um gerador automático de metadados para a CoTeia. Relatório técnico 287, Instituto de Ciências Matemáticas e de Computação / Universidade de São Paulo (ICMC/USP), São Carlos, SP. Disponível em: <http://www.icmc.usp.br/biblio/BIBLIOTECA/rel_tec/RT_287.pdf>. Acesso em: 12 set. 2007.
- Pansanato, L. T. E. e Fortes, R. P. M. (2007a). System Description: An Orienteering Strategy to Browse Semantically-Enhanced Educational Wiki Pages. In Franconi, E., Kifer, M., e May, W., editors, *4th European Semantic Web Conference (ESWC 2007), Lecture Notes in Computer Science (LNCS)*, volume 4519, p. 809–818, Innsbruck, Austria. Berlin Heidelberg: Springer-Verlag. DOI=<http://dx.doi.org/10.1007/978-3-540-72667-8_59>.
- Pansanato, L. T. E. e Fortes, R. P. M. (2007b). Using an Orienteering Strategy to Browse Semantically-Enhanced Educational Wiki Pages. In *Proceedings of the IADIS*

- International Conference e-Learning*, volume 1, p. 11–18, Lisbon, Portugal. IADIS Press.
- Pádua, E. M. M. (2006). *Metodologia da pesquisa: Abordagem teórico-prática*. Campinas, SP: Papirus, 12^a ed. 124 p.
- Plaisant, C., Shneiderman, B., Doan, K., e Bruns, T. (1999). Interface and data architecture for query preview in networked information systems. *ACM Transactions on Information Systems*, 17(3):320–341. DOI=<<http://doi.acm.org/10.1145/314516.314522>>.
- Pokorný, J. (2004). Web Searching and Information Retrieval. *Computing in Science and Engineering*, 6(4):43–48. DOI=<<http://doi.ieeecomputersociety.org/10.1109/MCSE.2004.24>>.
- Preece, J., Rogers, Y., e Sharp, H. (2002). *Interaction Design: Beyond Human-Computer Interaction*. New York: John Wiley & Sons. 544 p.
- Reich, S., Carr, L., Roure, D. D., e Hall, W. (1999). Where have you been from here? trails in hypertext systems. *ACM Computing Surveys*, 31(4):11. DOI=<<http://doi.acm.org/10.1145/345966.345994>>.
- Ryan, B. e Walmsley, S. (2003). Implementing metadata collection: a projects problems and solutions. *IEEE Learning Technology*, 5(1). Disponível em: <http://lthf.ieee.org/learn_tech/>. Acesso em: 25 jun. 2004.
- Sandvad, E. S., Grønbæk, K., Sloth, L., e Knudsen, J. L. (2001). A metro map metaphor for guided tours on the Web: the Webwise guided tour system. In *Proceedings of the 10th International Conference on World Wide Web*, p. 326–333, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/371920.372079>>.
- Saracevic, T. (1996). Modeling Interaction in Information Retrieval (IR): A Review and Proposal. In *Proceedings of the 59th American Society for Information Science (ASIS) Annual Meeting*, volume 33, p. 3–9.
- Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the 60th American Society for Information Science (ASIS) Annual Meeting*, volume 34, p. 313–327.
- Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, p. 336–343, Los Alamitos, CA, USA. IEEE Computer Society. DOI=<<http://doi.ieeecomputersociety.org/10.1109/VL.1996.545307>>.
- Shneiderman, B. (1997). Designing information-abundant web sites: issues and recommendations. *International Journal of Human-Computer Studies*, 47(1):5–29. Disponível em: <<http://ijhcs.open.ac.uk/shneiderman/shneiderman.pdf>>. Acesso em: 5 mar. 2005.

- Silva, E. L. e Menezes, E. M. (2005). *Metodologia da pesquisa e elaboração de dissertação*. Florianópolis: UFSC, quarta ed. 138 p.
- Sinha, V. e Karger, D. R. (2005). Magnet: supporting navigation in semistructured data environments. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, p. 97–106, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/1066157.1066169>>.
- Spink, A. (1997). Study of Interactive Feedback During Mediated Information Retrieval. *Journal of the American Society for Information Science (JASIS)*, 48(5):382–394. DOI=<[http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199705\)48:5<382::AID-ASI2>3.0.CO;2-R](http://dx.doi.org/10.1002/(SICI)1097-4571(199705)48:5<382::AID-ASI2>3.0.CO;2-R)>.
- Spink, A., Wilson, T., Ford, N., Foster, A., e Ellis, D. (2002). Information-Seeking and Mediated Searching. *Journal of the American Society for Information Science and Technology*, 53(9):695–735.
- Tanin, E., Shneiderman, B., e Xie, H. (2007). Browsing large online data tables using generalized query previews. *Information Systems*, 32(3):402–423. DOI=<<http://dx.doi.org/10.1016/j.is.2005.12.006>>.
- Teevan, J., Alvarado, C., Ackerman, M. S., e Karger, D. R. (2004). The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2004)*, p. 415–422, Vienna, Austria. Disponível em: <<http://haystack.lcs.mit.edu/papers/chi2004-perfectse.pdf>>. Acesso em: 4 mai. 2004.
- Teevan, J., Dumais, S. T., e Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*, p. 449–456, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/1076034.1076111>>.
- Vasudevan, V. e Palmer, M. (1999). On web annotations: promises and pitfalls of current webinfrastructure. In *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*, volume 2, p. 2012–2020, Los Alamitos, CA, USA. IEEE Computer Society. DOI=<<http://doi.ieeecomputersociety.org/10.1109/HICSS.1999.772663>>.
- W3C (1999). HTML 4.01 Specification. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/html401>>. Acesso em: 28 de jun. 2004.
- W3C (2000a). Document Object Model (DOM) Level 1 Specification (Second Edition). W3C Working Draft, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/REC-DOM-Level-1>>. Acesso em: 5 de ago. 2007.

- W3C (2000b). Document Object Model (DOM) Level 2 Core Specification. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/DOM-Level-2-Core>>. Acesso em: 5 de ago. 2007.
- W3C (2001). Semantic Web Activity. Website, World Wide Web Consortium (W3C). Disponível em: <<http://www.w3.org/2001/sw/>>. Acesso em: 30 mar. 2004.
- W3C (2002). XHTML 1.0 The Extensible HyperText Markup Language (Second Edition): A Reformulation of HTML 4 in XML 1.0. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/xhtml1>>. Acesso em: 5 ago. 2007.
- W3C (2004a). RDF Primer. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/rdf-primer/>>. Acesso em: 30 mar. 2005.
- W3C (2004b). RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>. Acesso em: 30 mar. 2005.
- W3C (2004c). RDF/XML Syntax Specification (Revised). W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/rdf-syntax-grammar/>>. Acesso em: 30 mar. 2005.
- W3C (2004d). Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/rdf-concepts/>>. Acesso em: 30 mar. 2005.
- W3C (2004e). Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/rdf-concepts/>>. Acesso em: 7 jul. 2004.
- W3C (2004f). XML Schema Part 0: Primer Second Edition. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/xmlschema-0/>>. Acesso em: 30 mar. 2005.
- W3C (2004g). XML Schema Part 1: Structures Second Edition. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/xmlschema-1/>>. Acesso em: 30 mar. 2005.
- W3C (2004h). XML Schema Part 2: Datatypes Second Edition. W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/xmlschema-2/>>. Acesso em: 30 mar. 2005.
- W3C (2006a). Extensible Markup Language (XML) 1.0 (Fourth Edition). W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/xml>>. Acesso em: 5 ago. 2007.
- W3C (2006b). GRDDL Primer. W3C Working Draft, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/grddl-primer/>>. Acesso em: 5 ago. 2007.

- W3C (2006c). Namespaces in XML 1.0 (Second Edition). W3C Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/xml-names>>. Acesso em: 5 ago. 2007.
- W3C (2007). Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification. W3C Candidate Recommendation, World Wide Web Consortium. Disponível em: <<http://www.w3.org/TR/CSS21>>. Acesso em: 5 ago. 2007.
- White, R. W., Drucker, S. M., Marchionini, G., Hearst, M., e m. c. schraefel (2007). Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, p. 2877–2880, New York, NY, USA. ACM Press. DOI=<<http://doi.acm.org/10.1145/1240866.1241100>>.
- White, R. W., Kules, B., Drucker, S. M., e m. c. schraefel (2006a). Supporting exploratory search: Introduction. *Communications of the ACM*, 49(4):36–39. DOI=<<http://doi.acm.org/10.1145/1121949.1121978>>.
- White, R. W., Muresan, G., e Marchionini, G. (2006b). Report on ACM SIGIR 2006 Workshop on Evaluating Exploratory Search Systems. *SIGIR Forum*, 40(2):52–60. DOI=<<http://doi.acm.org/10.1145/1189702.1189711>>.
- Wilson, M. L. e m. c. schraefel (2007). Bridging the Gap: Using IR Models for Evaluating Exploratory Search Interfaces. In *Proceedings of SIGCHI 2007 Workshop on Exploratory Search and HCI*, p. 19–22, New York, NY, USA. ACM Press. Disponível em: <<http://research.microsoft.com/~ryenw/proceedings/ESI2007.pdf>>. Acesso em: 10 ago. 2007.
- Wilson, T. D. (1997). Information behaviour: an interdisciplinary perspective. *Information Processing & Management*, 33(4):551–572. Disponível em: <<http://informationr.net/tdw/publ/infbehav/>>. Acesso em: 21 fev. 2006.
- Wilson, T. D. (1999). Models in information behavior research. *Journal of Documentation*, 55(3):249–270. Disponível em: <<http://informationr.net/tdw/publ/papers/1999JDoc.html>>. Acesso em: 21 fev. 2006.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., e Wesslén, A. (2000). *Experimentation in Software Engineering: An Introduction*. Kluwer Academic Publishers, Boston/Dordrecht/London, 1st ed. 228 p.
- Zeldman, J. (2006). *Designing with Web Standards*. New Riders, Berkeley, CA, USA, second ed. 410 p.

Material do Estudo de Usabilidade

Avaliação do Novo Sistema de Navegação para a CoTeia

Este material contém:

- Questionário pré-avaliação;
- Visão geral do processo de avaliação;
- Recomendações aos participantes;
- Descrição das tarefas;
- Questionário pós-tarefa;
- Entrevista pós-avaliação.

(Página propositalmente deixada em branco)

Questionário pré-avaliação

Informação Geral

Q1. Qual o seu gênero (sexo):

a. () Feminino b. () Masculino

Q2. Qual a sua idade:

_____ anos

Q3. Qual o seu tempo de profissão como docente:

_____ anos

Experiência com Ferramentas

1. Ferramentas de busca (*search engines*):

Q1.1 Qual a sua experiência de uso:

Principiante									Especialista
1	2	3	4	5	6	7	8	9	10

Q1.2 Qual você prefere:

a. () Google b. () Yahoo! Search c. () MSN Search d. () Outra: _____

Q1.3 Qual a sua frequência de uso:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

2. Ferramenta de Diretório (*directory*) ou Sites de Listas de Categorias:

Q2.1 Qual a sua experiência de uso:

Principiante									Especialista
1	2	3	4	5	6	7	8	9	10

Q2.2 Qual você prefere:

a. () Yahoo! Directory b. () Open Directory () c. Outra: _____

Q2.3 Qual a sua frequência de uso:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

3. Catálogos on-line de biblioteca e/ou outros sistemas:

Q3.1 Qual a sua experiência de uso (por exemplo, com DEDALUS-SIBi/USP, ACM Digital Library, IEEE Explore):

Principiante									Especialista
1	2	3	4	5	6	7	8	9	10

Q3.2 Qual a sua frequência de uso:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

Experiência com a CoTeia

Q1. Há quanto tempo você conhece a CoTeia:

_____ anos

Q2. Qual a sua experiência de uso:

Principiante									Especialista
1	2	3	4	5	6	7	8	9	10

Q3. Qual repositório de páginas você acessa:

- a. () Páginas de disciplinas de 2004-2006 <http://coteia.icmc.usp.br/coteia/>
b. () Páginas de disciplinas de 2001-2003 <http://safedevol.icmc.usp.br/coweb/>

Q4. Qual a sua frequência de acesso às páginas para consulta (visita):

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

Q5. Qual a sua frequência de acesso às páginas para edição:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

Q6. Qual a sua frequência de uso especificamente para reuso de material das disciplinas:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

Q7. Qual a abordagem que você geralmente usa para procurar pelo material (ou outro tipo de informação) disponível nas páginas da CoTeia:

- a. () Acesso a partir da entrada correspondente à disciplina na página inicial.
b. () Busca por meio de uma search engine, especifique: _____ .
c. () Outra: _____

(Página propositalmente deixada em branco)

Processo de avaliação

A avaliação consiste de testes com usuários. Cada participante é convidado a executar tarefas usando o sistema em estudo e a responder às perguntas de um pequeno questionário após cada tarefa. Ao final da execução das tarefas, o participante também é entrevistado.

Toda a interação do participante com a interface do sistema é registrada em um arquivo (log). O áudio da entrevista também é gravado. O nome de cada um dos participantes é mantido em sigilo e todas as ligações entre o participante e os seus dados serão destruídas após a análise da sua participação. Portanto, todos os resultados do estudo serão apresentados de maneira anônima.

As tarefas são situações simuladas, isto é, consistem de uma descrição aberta de um cenário que tem por objetivo provocar uma necessidade de informação (*information need*). A partir da leitura dessa descrição o participante define o que ele gostaria de procurar de acordo com o seu próprio contexto. A maioria das tarefas estão relacionadas ao reuso do material disponível na CoTeia. O questionário pós-avaliação é igual para todas as tarefas propostas.

Você tem alguma pergunta?

Recomendações

Durante a realização das tarefas:

- Tenha certeza de que a tarefa foi claramente compreendida.
- Fale em voz alta (*think aloud*) o que você está procurando e como pretende realizar a tarefa.

Você tem alguma pergunta?

Obrigado por participar da avaliação do novo sistema de navegação para a CoTeia
As 5 tarefas que você irá realizar durante a avaliação são descritas a seguir

Por favor, espere até que você seja informado para iniciar a avaliação.

1ª Tarefa

Você foi designado para ministrar uma disciplina nova, isto é, uma disciplina de uma área diferente das áreas que está acostumado a trabalhar. Inicialmente, você quer procurar (“dar uma olhada”) pelo material que existe disponível na CoTeia relacionado ao programa (ou conteúdo programático) dessa nova disciplina.

Por favor, avise quando você concluir essa tarefa.

Espere até que você seja informado para continuar.

Questionário pós-tarefa 1

Q1. Qual a frequência que você realiza (ou espera realizar) atividades relacionadas a essa tarefa?

Não realiza (ou espera realizar)	Uma ou duas vezes por ano	Uma ou duas vezes por semestre	Uma ou duas vezes por mês	Uma ou duas vezes por semana	Uma ou duas vezes por dia	Mais do que duas vezes por dia
1	2	3	4	5	6	7

Q2. Qual a relevância dos resultados obtidos? Expresse o quanto você está satisfeito com os resultados obtidos.

Nenhuma relevância	Pouca relevância	Alguma relevância	Razoável relevância	Considerável relevância	Muita relevância	Extrema relevância
1	2	3	4	5	6	7

Q3. Qual a dificuldade de compreensão da tarefa? Expresse o grau de dificuldade em compreender a tarefa e definir o seu objetivo.

Nenhuma dificuldade	Pouca dificuldade	Alguma dificuldade	Razoável dificuldade	Considerável dificuldade	Muita dificuldade	Extrema dificuldade
1	2	3	4	5	6	7

Q4. Qual a utilidade do sistema para essa tarefa? Expresse o grau de facilitação do sistema para essa tarefa?

Nenhuma utilidade	Pouca utilidade	Alguma utilidade	Razoável utilidade	Considerável utilidade	Muita utilidade	Extrema utilidade
1	2	3	4	5	6	7

Por favor, avise quando você concluir esse questionário.

Espere até que você seja informado para continuar com a próxima tarefa.

2ª Tarefa

Você precisa elaborar uma questão para uma prova dissertativa sobre um determinado assunto apresentado nas aulas. Especificamente, você quer procurar por um enunciado de um determinado exercício que foi proposto anteriormente e está disponível na CoTeia.

Por favor, avise quando você concluir essa tarefa.

Espere até que você seja informado para continuar com a próxima tarefa.

Questionário pós-tarefa 2

Q1. Qual a frequência que você realiza (ou espera realizar) atividades relacionadas a essa tarefa?

Não realiza (ou espera realizar)	Uma ou duas vezes por ano	Uma ou duas vezes por semestre	Uma ou duas vezes por mês	Uma ou duas vezes por semana	Uma ou duas vezes por dia	Mais do que duas vezes por dia
1	2	3	4	5	6	7

Q2. Qual a relevância dos resultados obtidos? Expresse o quanto você está satisfeito com os resultados obtidos.

Nenhuma relevância	Pouca relevância	Alguma relevância	Razoável relevância	Considerável relevância	Muita relevância	Extrema relevância
1	2	3	4	5	6	7

Q3. Qual a dificuldade de compreensão da tarefa? Expresse o grau de dificuldade em compreender a tarefa e definir o seu objetivo.

Nenhuma dificuldade	Pouca dificuldade	Alguma dificuldade	Razoável dificuldade	Considerável dificuldade	Muita dificuldade	Extrema dificuldade
1	2	3	4	5	6	7

Q4. Qual a utilidade do sistema para essa tarefa? Expresse o grau de facilitação do sistema para essa tarefa?

Nenhuma utilidade	Pouca utilidade	Alguma utilidade	Razoável utilidade	Considerável utilidade	Muita utilidade	Extrema utilidade
1	2	3	4	5	6	7

Por favor, avise quando você concluir esse questionário.

Espere até que você seja informado para continuar com a próxima tarefa.

3ª Tarefa

Você deseja melhorar (ou simplesmente modificar) o critério de avaliação atualmente adotado para uma determinada disciplina. Inicialmente, você quer procurar pelos critérios de avaliação divulgados recentemente na CoTeia pelos professores de outras disciplinas.

Por favor, avise quando você concluir essa tarefa.

Espere até que você seja informado para continuar com a próxima tarefa.

Questionário pós-tarefa 3

Q1. Qual a frequência que você realiza (ou espera realizar) atividades relacionadas a essa tarefa?

Não realiza (ou espera realizar)	Uma ou duas vezes por ano	Uma ou duas vezes por semestre	Uma ou duas vezes por mês	Uma ou duas vezes por semana	Uma ou duas vezes por dia	Mais do que duas vezes por dia
1	2	3	4	5	6	7

Q2. Qual a relevância dos resultados obtidos? Expresse o quanto você está satisfeito com os resultados obtidos.

Nenhuma relevância	Pouca relevância	Alguma relevância	Razoável relevância	Considerável relevância	Muita relevância	Extrema relevância
1	2	3	4	5	6	7

Q3. Qual a dificuldade de compreensão da tarefa? Expresse o grau de dificuldade em compreender a tarefa e definir o seu objetivo.

Nenhuma dificuldade	Pouca dificuldade	Alguma dificuldade	Razoável dificuldade	Considerável dificuldade	Muita dificuldade	Extrema dificuldade
1	2	3	4	5	6	7

Q4. Qual a utilidade do sistema para essa tarefa? Expresse o grau de facilitação do sistema para essa tarefa?

Nenhuma utilidade	Pouca utilidade	Alguma utilidade	Razoável utilidade	Considerável utilidade	Muita utilidade	Extrema utilidade
1	2	3	4	5	6	7

Por favor, avise quando você concluir esse questionário.

Espere até que você seja informado para continuar com a próxima tarefa.

4ª Tarefa

Você gostaria de alterar o método de ensino atualmente adotado para um determinado tópico de uma disciplina. Inicialmente, você quer verificar se existe proposto na CoTeia algum material didático alternativo ou alguma abordagem inovadora ou diferente da geralmente utilizada para ministrar esse tipo de tópico.

Por favor, avise quando você concluir essa tarefa.

Espere até que você seja informado para continuar com a próxima tarefa.

Questionário pós-tarefa 4

Q1. Qual a frequência que você realiza (ou espera realizar) atividades relacionadas a essa tarefa?

Não realiza (ou espera realizar)	Uma ou duas vezes por ano	Uma ou duas vezes por semestre	Uma ou duas vezes por mês	Uma ou duas vezes por semana	Uma ou duas vezes por dia	Mais do que duas vezes por dia
1	2	3	4	5	6	7

Q2. Qual a relevância dos resultados obtidos? Expresse o quanto você está satisfeito com os resultados obtidos.

Nenhuma relevância	Pouca relevância	Alguma relevância	Razoável relevância	Considerável relevância	Muita relevância	Extrema relevância
1	2	3	4	5	6	7

Q3. Qual a dificuldade de compreensão da tarefa? Expresse o grau de dificuldade em compreender a tarefa e definir o seu objetivo.

Nenhuma dificuldade	Pouca dificuldade	Alguma dificuldade	Razoável dificuldade	Considerável dificuldade	Muita dificuldade	Extrema dificuldade
1	2	3	4	5	6	7

Q4. Qual a utilidade do sistema para essa tarefa? Expresse o grau de facilitação do sistema para essa tarefa?

Nenhuma utilidade	Pouca utilidade	Alguma utilidade	Razoável utilidade	Considerável utilidade	Muita utilidade	Extrema utilidade
1	2	3	4	5	6	7

Por favor, avise quando você concluir esse questionário.

Espere até que você seja informado para continuar com a próxima tarefa.

5ª Tarefa

Você precisa se comunicar com um aluno que é/foi monitor de uma determinada disciplina. Especificamente, você quer procurar pelo e-mail desse aluno que está disponível na CoTeia.

Por favor, avise quando você concluir essa tarefa.

Espere até que você seja informado para continuar com a próxima tarefa.

Questionário pós-tarefa 5

Q1. Qual a frequência que você realiza (ou espera realizar) atividades relacionadas a essa tarefa?

Não realiza (ou espera realizar)	Uma ou duas vezes por ano	Uma ou duas vezes por semestre	Uma ou duas vezes por mês	Uma ou duas vezes por semana	Uma ou duas vezes por dia	Mais do que duas vezes por dia
1	2	3	4	5	6	7

Q2. Qual a relevância dos resultados obtidos? Expresse o quanto você está satisfeito com os resultados obtidos.

Nenhuma relevância	Pouca relevância	Alguma relevância	Razoável relevância	Considerável relevância	Muita relevância	Extrema relevância
1	2	3	4	5	6	7

Q3. Qual a dificuldade de compreensão da tarefa? Expresse o grau de dificuldade em compreender a tarefa e definir o seu objetivo.

Nenhuma dificuldade	Pouca dificuldade	Alguma dificuldade	Razoável dificuldade	Considerável dificuldade	Muita dificuldade	Extrema dificuldade
1	2	3	4	5	6	7

Q.4 Qual a utilidade do sistema para essa tarefa? Expresse o grau de facilitação do sistema para essa tarefa?

Nenhuma utilidade	Pouca utilidade	Alguma utilidade	Razoável utilidade	Considerável utilidade	Muita utilidade	Extrema utilidade
1	2	3	4	5	6	7

Por favor, avise quando você concluir esse questionário.

Espere até que você seja informado para continuar com a próxima tarefa.

Entrevista pós-avaliação

Q1. Como você poderia descrever a experiência de usar o sistema para navegar as páginas da CoTeia? Quais as dificuldades?

Q2. O sistema foi útil na execução das tarefas?

☐ Sim ☐ Não ☐ Não soube responder

Q3. Os resultados obtidos foram relevantes?

☐ Sim ☐ Não ☐ Não soube responder

Q4. Qual foi a ferramenta preferida? Por quê?

☐ Busca por palavra-chave ☐ Navegação por facetas
☐ Memória de trabalho ☐ Classificação
☐ Agrupamento ☐ Marcar resultados
☐ Destaque de termos ☐ Desfazer e reiniciar

Q5. Quais ferramentas foram as mais úteis? Por quê?

☐ Busca por palavra-chave ☐ Navegação por facetas
☐ Memória de trabalho ☐ Classificação
☐ Agrupamento ☐ Marcar resultados
☐ Destaque de termos ☐ Desfazer e reiniciar

Q6. Os metadados (e facetas) são representativos do conteúdo das páginas da CoTeia?

☐ Sim ☐ Não ☐ Não soube responder

Q7. Existe algum elemento de metadados não considerado que poderia ser útil? Qual?

Q8. Existe alguma ferramenta que você considera com potencial para auxiliar o usuário, mas que deveria ser melhorada ou modificada?

Q9. Você sentiu que tinha o controle sobre o processo de navegação? (por meio do histórico de ações e da capacidade de desfazer uma ação).

Q10. Você manteve o sentido de orientação?

☐ Sim ☐ Não ☐ Não soube responder

Q11. Você compreendeu como os resultados foram obtidos? (mesmo quando não existiam páginas que correspondessem à sequência de ações executada).

☐ Sim ☐ Não ☐ Não soube responder

Q12. Sobre a avaliação, as escalas utilizadas no questionário pré-avaliação e no questionário pós-tarefa são adequadas?

☐ Sim ☐ Não ☐ Não soube responder

Q13. Sobre a avaliação, a descrição da tarefa na forma de situação simulada é difícil de compreender? Por quê?

☐ Sim ☐ Não ☐ Não soube responder

Material do Estudo Experimental

Avaliação de Sistemas de Navegação para a CoTeia

Este material contém:

- Carta de consentimento;
- Questionário de levantamento;
- Descrição das tarefas;
- Questionário pós-tarefa;
- Questionário de fechamento.

(Página propositalmente deixada em branco)

Carta de consentimento

Afirmo que sou maior de 18 anos e desejo participar do experimento conduzido no Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, São Carlos, pela Profa Dra Renata Pontin de Mattos Fortes e seu aluno de doutorado Luciano Tadeu Esteves Pansanato. O propósito do experimento é avaliar sistemas de navegação para a CoTeia.

Todas as informações coletadas neste experimento são confidenciais, e meu nome não será identificado em momento algum. Portanto, todos os resultados do experimento serão apresentados de maneira anônima.

Estou ciente de que posso fazer perguntas ou desistir da colaboração em qualquer momento, sem qualquer tipo de penalidade.

Assinatura do participante

Data

(Página propositalmente deixada em branco)

Questionário de levantamento

Informação Geral

Q1. Qual o seu gênero (sexo):

a. () Feminino b. () Masculino

Q2. Qual a sua idade:

_____ anos

Q3. Qual a sua ocupação (se não for estudante em tempo integral):

Q4. Qual curso de graduação você está matriculado:

Q5. Há quanto tempo você está cursando a graduação:

_____ anos e _____ meses

Experiência com Ferramentas

1. Navegadores Web (*browsers*):

Q1.1 Qual você prefere:

a. () Internet Explorer b. () Mozilla Firefox c. () Opera d. () Outro: _____

Q1.2 Qual a sua experiência de uso:

Principiante									Especialista
1	2	3	4	5	6	7	8	9	10

Q1.3 Qual a sua frequência de uso:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

2. Ferramentas de busca (*search engines*):

Q2.1 Qual você prefere:

a. () Google b. () Yahoo! Search c. () MSN Search d. () Outra: _____

Q2.2 Qual a sua experiência de uso:

Principiante									Especialista
1	2	3	4	5	6	7	8	9	10

Q2.3 Qual a sua frequência de uso:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

3. Catálogos on-line de biblioteca e/ou outros sistemas:

Q3.1 Qual a sua experiência de uso:

Principiante									Especialista
1	2	3	4	5	6	7	8	9	10

Q3.2 Qual a sua frequência de uso (por exemplo, com DEDALUS-SIBi/USP, ACM Digital Library, IEEE Explore):

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

Experiência com a CoTeia

Q1. Há quanto tempo você conhece a CoTeia:

_____ anos

Q2. Qual a sua experiência de uso:

Principiante									Especialista
1	2	3	4	5	6	7	8	9	10

Q3. Qual repositório de páginas você acessa:

- a. () Páginas de disciplinas de 2004-2006 <http://coteia.icmc.usp.br/coteia/>
b. () Páginas de disciplinas de 2001-2003 <http://safedevol.icmc.usp.br/coweb/>

Q4. Qual a sua frequência de acesso às páginas para consulta (visita):

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

Q5. Qual a sua frequência de acesso às páginas para edição:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

Q6. Qual a sua frequência de uso especificamente para usar material das disciplinas:

Nunca	Anualmente	Semestralmente	Mensalmente	Semanalmente	Diariamente
1	2	3	4	5	6

Q7. Qual a abordagem que você geralmente usa para procurar pelo material (ou outro tipo de informação) disponível nas páginas da CoTeia:

- a. () Acesso a partir da entrada correspondente à disciplina na página inicial.
b. () Busca por meio de uma search engine, especifique: _____ .
c. () Outra: _____

(Página propositalmente deixada em branco)

Obrigado por participar da avaliação do Sistema de Navegação para a CoTeia
As 4 tarefas que você irá realizar durante a avaliação são descritas a seguir

Por favor, espere até que você seja informado para iniciar a avaliação.

(Página propositalmente deixada em branco)

1ª Tarefa

Início: Término:

Descrição:

Considere que você foi selecionado para ser monitor da disciplina “Algoritmos e Estruturas de Dados I”. Inicialmente, você quer entrar em contato com os ex-monitores da disciplina para conversar sobre detalhes da monitoria.

A tarefa consiste em encontrar páginas com os nomes e e-mails dos monitores dos semestres anteriores.

Questionário pós-tarefa:

Q1.1 Completar essa tarefa foi:	Fácil						Difícil
	1	2	3	4	5	6	7
Q1.2 O sistema para essa tarefa foi:	Fácil de usar						Difícil de usar
	1	2	3	4	5	6	7
Q1.3 A relevância dos resultados encontrados é:	Nenhuma						Extrema
	1	2	3	4	5	6	7

Por favor, avise quando você concluir essa tarefa.

Espere até que você seja informado para continuar.

(Página propositalmente deixada em branco)

2ª Tarefa

Início: Término:

Descrição:

Considere que você precisa apresentar um seminário sobre um tópico avançado de Engenharia de Software de sua escolha. Inicialmente, você gostaria de “dar uma olhada” no material que existe disponível dos seminários apresentados na pós-graduação.

A tarefa consiste em encontrar páginas das disciplinas relacionadas a Engenharia de Software da pós-graduação que contenham links para o material dos seminários apresentados pelos alunos.

Questionário pós-tarefa:

Q2.1 Completar essa tarefa foi:	Fácil						Difícil
	1	2	3	4	5	6	7
Q2.2 O sistema para essa tarefa foi:	Fácil de usar						Difícil de usar
	1	2	3	4	5	6	7
Q1.3 A relevância dos resultados encontrados é:	Nenhuma						Extrema
	1	2	3	4	5	6	7

Por favor, avise quando você concluir essa tarefa.

Espere até que você seja informado para continuar com a próxima tarefa.

(Página propositalmente deixada em branco)

3ª Tarefa

Início: Término:

Descrição:

Considere que você gostaria de saber a diferença entre os tipos de dados das linguagens Java e Python.

A tarefa consiste em encontrar páginas com essa informação ou com links para algum material que contenha essa informação.

Questionário pós-tarefa:

Q3.1 Completar essa tarefa foi:	Fácil						Difícil
	1	2	3	4	5	6	7
Q3.2 O sistema para essa tarefa foi:	Fácil de usar						Difícil de usar
	1	2	3	4	5	6	7
Q1.3 A relevância dos resultados encontrados é:	Nenhuma						Extrema
	1	2	3	4	5	6	7

Por favor, avise quando você concluir essa tarefa.

Espera-se até que você seja informado para continuar com a próxima tarefa.

(Página propositalmente deixada em branco)

4ª Tarefa

Início: Término:

Descrição:

Considere que você gostaria de obter informações sobre como projetar aplicações para pessoas portadoras de necessidades especiais.

A tarefa consiste em encontrar páginas com essa informação.

Questionário pós-tarefa:

Q4.1 Completar essa tarefa foi:	Fácil						Difícil
	1	2	3	4	5	6	7
Q4.2 O sistema para essa tarefa foi:	Fácil de usar						Difícil de usar
	1	2	3	4	5	6	7
Q1.3 A relevância dos resultados encontrados é:	Nenhuma						Extrema
	1	2	3	4	5	6	7

Por favor, avise quando você concluir essa tarefa.

Espere até que você seja informado para continuar com a próxima tarefa.

(Página propositalmente deixada em branco)

Questionário de fechamento

Q1. Qual a freqüência que você realiza (ou espera realizar) atividades relacionadas a essas tarefas?

Nunca	Raramente	Algumas vezes	Razoavelmente	Consideravelmente	Freqüentemente	Muito freqüentemente
1	2	3	4	5	6	7

Q2. Qual a relevância dos resultados obtidos? Expresse o quanto você está satisfeito com os resultados obtidos.

Nenhuma relevância	Pouca relevância	Alguma relevância	Razoável relevância	Considerável relevância	Muita relevância	Extrema relevância
1	2	3	4	5	6	7

Q3. Qual a facilidade de compreender o sistema? Expresse o grau de facilidade em compreender as funcionalidades do sistema.

Nenhuma facilidade	Pouca facilidade	Alguma facilidade	Razoável facilidade	Considerável facilidade	Muita facilidade	Extrema facilidade
1	2	3	4	5	6	7

Q4. Qual a facilidade de usar (operar) o sistema? Expresse o grau de facilidade de usar as funcionalidades do sistema para realizar as tarefas.

Nenhuma facilidade	Pouca facilidade	Alguma facilidade	Razoável facilidade	Considerável facilidade	Muita facilidade	Extrema facilidade
1	2	3	4	5	6	7

Q5. Qual a utilidade do sistema? Expresse o grau de facilitação do sistema para realizar as tarefas.

Nenhuma utilidade	Pouca utilidade	Alguma utilidade	Razoável utilidade	Considerável utilidade	Muita utilidade	Extrema utilidade
1	2	3	4	5	6	7

Q6. Quantas vezes você ficou perdido durante a navegação? Indique a freqüência que você ficou perdido navegando pelas páginas usando o sistema.

Nunca	Raramente	Algumas vezes	Razoavelmente	Consideravelmente	Freqüentemente	Muito freqüentemente
1	2	3	4	5	6	7

Q7. Você considera que o sistema forneceu auxílio para completar as tarefas em menos tempo do que usando um navegador (browser) comum?

Nenhum auxílio 1	Pouco auxílio 2	Algum auxílio 3	Razoável auxílio 4	Considerável auxílio 5	Muito auxílio 6	Extremo auxílio 7
------------------------	-----------------------	-----------------------	--------------------------	------------------------------	-----------------------	-------------------------

Q8. Você considera o reuso futuro do sistema? Indique o quanto você pretende reusar o sistema para navegar pelas páginas.

Nenhuma pretensão 1	Pouca pretensão 2	Alguma pretensão 3	Razoável pretensão 4	Considerável pretensão 5	Muita pretensão 6	Extrema pretensão 7
---------------------------	-------------------------	--------------------------	----------------------------	--------------------------------	-------------------------	---------------------------

Q9. Qual a sua satisfação pessoal em usar o sistema? Expresse o quanto você ficou satisfeito em utilizar o sistema para realizar as tarefas.

Nenhuma satisfação 1	Pouca satisfação 2	Alguma satisfação 3	Razoável satisfação 4	Considerável satisfação 5	Muita satisfação 6	Extrema satisfação 7
----------------------------	--------------------------	---------------------------	-----------------------------	---------------------------------	--------------------------	----------------------------