

## Conhecimento e Internet: busca, exploração e a familiaridade

**Fernando Marques Figueira Filho**

Este exemplar corresponde à redação final da Tese devidamente corrigida e defendida por Fernando Marques Figueira Filho e aprovada pela Banca Examinadora.

Campinas, 21 de Fevereiro de 2011.



Prof. Paulo Lício de Geus (Orientador)

Tese apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

FICHA CATALOGRÁFICA ELABORADA POR  
ANA REGINA MACHADO – CRB8/5467  
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E  
COMPUTAÇÃO CIENTÍFICA – UNICAMP

F469c Figueira Filho, Fernando Marques, 1981-  
Conhecimento e Internet : busca, exploração e a  
familiaridade / Fernando Marques Figueira Filho. –  
Campinas, SP : [s.n.], 2011.

Orientador: Paulo Lício de Geus.  
Tese (doutorado) - Universidade Estadual de Campinas,  
Instituto de Computação.

1. Recuperação da informação. 2. Gerenciamento da  
informação pessoal. 3. World Wide Web (Sistema de  
recuperação da informação). 4. Internet – Aspecto sociais.  
5. Teoria fundamentada em dados. I. Geus, Paulo Lício de,  
1956- II. Universidade Estadual de Campinas. Instituto de  
Computação. III. Título.

Informações para Biblioteca Digital

**Título em inglês:** Knowledge and the Internet : search, exploration and  
the familiarity.

**Palavras-chave em inglês:**

Information retrieval

Personal information management

World Wide Web

Internet - Social aspects

Grounded theory

**Área de concentração:** Ciência da Computação

**Titulação:** Doutor em Ciência da Computação

**Banca examinadora:**

Paulo Lício de Geus [Orientador]

Cleidson Ronald Botelho de Souza

Sérgio Roberto Pereira da Silva

Maria Cecília Calani Baranauskas

Célio Cardoso Guimarães

**Data da defesa:** 21-02-2011

**Programa de Pós-Graduação:** Ciência da Computação

## TERMO DE APROVAÇÃO

Tese Defendida e Aprovada em 21 de fevereiro de 2011, pela Banca examinadora composta pelos Professores Doutores:



---

**Prof. Dr. Cleidson Ronald Botelho de Souza**  
IBM-Research / Brasil



---

**Prof. Dr. Sérgio Roberto Pereira da Silva**  
Departamento de Informática / UEM



---

**Prof. Dr. Celso Cardoso Guimarães**  
IC / UNICAMP



---

**Profa. Dra. Maria Cecília Galani Baranauskas**  
IC / UNICAMP



---

**Prof. Dr. Paulo Lício de Geus**  
IC / UNICAMP



# Conhecimento e Internet: busca, exploração e a familiaridade

**Fernando Marques Figueira Filho**

21 de Fevereiro de 2011

## **Banca Examinadora:**

- Prof. Paulo Lício de Geus (Orientador)
- Prof. Cleidson de Souza – IBM Research (membro titular externo)
- Prof. Sérgio Roberto da Silva – Universidade Estadual de Maringá (membro titular externo)
- Profa. Cecília Baranauskas (membro titular interno)
- Prof. Célio Cardoso Guimarães (membro titular interno)



# Resumo

A Internet parece estar realizando o sonho de muitos visionários ao angariar parcela crescente do conhecimento humano e tornar seu acesso mais fácil. Entretanto, o modelo de interação para busca e exploração desse conhecimento é ainda similar àquele utilizado, por exemplo, por bibliotecários para consultar acervos de livros e publicações. Usuários entram com uma consulta constituída por palavras-chave e recebem como resposta uma lista de resultados que contém os termos utilizados na consulta. A exploração do conhecimento disponível através desse modelo apresenta dois problemas. Primeiro, os sistemas de busca são implementados sob o pressuposto que o conjunto de palavras-chave capaz de filtrar documentos relevantes é familiar ao usuário. No entanto, a heterogeneidade dos usuários da Internet frente ao imenso volume de informação disponível implica em diferentes níveis de familiaridade dos usuários com o conhecimento disponível. Segundo, a interpretação dos resultados retornados pelo sistema, de acordo com esse modelo, é delegada ao indivíduo, de modo que o usuário torna-se responsável por analisar e sintetizar o conteúdo filtrado para cumprir com os seus objetos de busca. Isto torna-se um problema em atividades de busca de longo prazo, que visam explorar o conhecimento para fins de aprendizado, revisão bibliográfica, dentre outras.

A presente tese baseia-se em arcabouços teóricos das ciências sociais e em estudos empíricos para caracterizar e propor soluções para esses problemas. Os resultados mostram que a exploração do conhecimento na Internet depende de um conjunto de artefatos externos aos sistemas de busca atuais. Também mostram que a interação dos usuários com esses artefatos e com agentes humanos é parte integrante das atividades cognitivas desempenhadas para exploração do conteúdo disponível na Internet. Diferentes níveis de familiaridade demandam alternativas ao modelo de interação consulta-resposta e sugerem a integração de tecnologias da web social no desenvolvimento de sistemas de informação que visam suportar tarefas de busca exploratórias.



# Abstract

The Internet seems to be accomplishing the dream of visionaries by collecting part of the growing human knowledge and making its access easier. However, the interaction model for search and exploration of this knowledge is still similar to the one used, for example, by librarians to query over collections of books and publications. Users formulate queries that are constituted by keywords and receive a list of results in return, which contains the terms used in the entered query. The exploration of the available knowledge through this model presents two main problems. First, search systems are implemented under the assumption that the set of keywords to filter relevant documents is familiar to the user. But the heterogeneity of Internet users when faced to the huge amount of available information implies in different levels of familiarity of these same users regarding the available knowledge. Second, the interpretation of results returned by the system, according to this model, is delegated to the individual, and the user becomes responsible for analyzing and synthesizing the filtered content in order to meet his/her search goals. This becomes a problem in the case of long term search activities, that aims at exploring knowledge for purposes of learning, bibliography review and so on.

This thesis is based on theoretical frameworks of the social sciences and on empirical studies to characterize and come up with solutions to those problems. The findings show that knowledge exploration on the Internet depends on a set of artifacts that are external to the search systems currently available. They also show that the user interaction with these artifacts and also with human agents is a fundamental part of the cognitive activities performed for knowledge exploration on the Internet. Different levels of familiarity demand alternatives to the query-response model and suggest the integration with social web technologies to develop information systems that support exploratory search tasks.



# Agradecimentos

Foi uma longa jornada até aqui e, por esta razão, eu não poderia deixar de agradecer. Minha escola foi o Instituto de Computação da UNICAMP, que me ensinou a pensar de forma sistemática e onde tive o primeiro contato com pessoas que possuíam a paixão pelo conhecimento e pelos computadores. Esta paixão está, agora, dentro de mim. Mas algo estava faltando. O ser humano, seja no papel de desenvolvedor ou no de usuário, preocupava-me deveras. Como tornar a máquina menos ríspida e mais útil aos nossos propósitos? Foi nas aulas da Profa. Cecília Baranauskas que aprendi os primeiros métodos. Isto motivou uma mudança definitiva: a minha saída da área de segurança de computadores e a entrada em um mundo completamente novo à mim, o da interação humano-computador.

Fui transferido do mestrado para o doutorado e iniciei a busca por um novo tema. Não foi uma tarefa fácil, pois ainda tinha dentro de mim aquela paixão pelos computadores, pelos sistemas formais e seus algoritmos. Como entender um sistema tão complexo, i.e. a mente humana e a sociedade, bem como todas as implicações da interação do homem com a tecnologia. Este entendimento necessitava de uma mudança na minha forma de pensar, e o Prof. João Porto de Albuquerque auxiliou-me através da sugestão de literatura, da co-autoria em artigos e, sobretudo, através de muitas conversas. Mas o processo de mudança ainda não havia terminado.

Foi na minha qualificação de doutorado que conheci pessoalmente o Prof. Cleidson de Souza. Ele, na época, nem imaginara o impacto que seus atos iriam causar. Ele recomendou que eu fosse estudar na Universidade da Califórnia, em Irvine, e através da sua indicação, eu fui aceito para trabalhar por 7 meses em período colaborativo com o Prof. Gary M. Olson. Mais do que uma relação orientador-aluno, o Prof. Olson proporcionou o meu contato com metodologias novas de pesquisa, novas maneiras de encarar problemas e, acima de tudo, um contato com uma nova maneira de fazer ciência. Agradeço a Steve Abrams, a Roberto Silva Filho e à Profa. Judith Olson, que me receberam com muito carinho e foram fundamentais à minha adaptação como aluno visitante. Agradeço também a outros alunos da mesma universidade: Mark Villegas, meu companheiro de apartamento, que me introduziu à cultura filipino-americana e, sobretudo, a uma nova

maneira de enxergar os processos culturais. A Ruy Cervantes, que gentilmente cedeu seu quarto durante o período e à Sharon Quinsaas, pelas diversas conversas confortadoras. Agradeço especialmente aos estudantes brasileiros Nilmax Moura e Tiago Proença, pela nossa amizade. Não poderia deixar de agradecer também a todos os estudantes que participaram nos dois estudos que lá foram realizados, a Matthew Bietz pelos seus conselhos e também ao Prof. David Redmiles, uma figura humana maravilhosa.

Um agradecimento especial deve ser dado ao suporte que tive nesse período de meu orientador principal, Prof. Paulo Lício de Geus, e também de três alunos de graduação que na época me auxiliaram com a implementação dos sistemas propostos e avaliados nessa tese. São eles: Thiago Cabral, Eric Velten de Melo e André Resende. Eles contribuíram com idéias, suor e, principalmente, ao acreditar nesta tese.

Mas minha jornada ainda não tinha terminado. Em fevereiro de 2010 participei do colóquio de doutorado da conferência CSCW e recebi conselhos de diversos professores, especialmente dos professores Saul Greenberg, Mark Ackerman e Geraldine Fitzpatrick. Agradeço a todos os participantes daquele colóquio, incluindo os estudantes com os quais pretendo manter uma longa e duradoura amizade.

O período após meu retorno ao Brasil não foi simples. Precisava analisar os dados coletados nos dois estudos realizados na Califórnia e não tinha idéia de como fazê-lo. Foi então que solicitei para que o Prof. Olson entrasse em contato com a Profa. Wendy Mackay, me indicando para um período colaborativo de 2 meses na Universidade de Paris Sud 11, Laboratório de Pesquisa em Informática. As primeiras reuniões com a Profa. Mackay não foram fáceis. Hoje agradeço a ela por ter me explicado a função da elaboração sistemática de perguntas de pesquisa. Agradeço também pelos seus ensinamentos de como realizar um bom estudo de campo. Neste período, eu também recebi a ajuda de outros pesquisadores: Ilaria Liccardi, que contribuiu mais do que ela possa imaginar, Julie Wagner, pelos seus conselhos, e aos professores Theophanis Tsandilas, Olivier Chapuis e Michel Beaudouin-Lafon, pelo suporte concedido. Agradeço também a todos os que gentilmente aceitaram participar do estudo sobre documentos pessoais e familiares apresentado nesta tese.

Meu segundo retorno ao Brasil foi decisivo. Eu precisava sintetizar as lições aprendidas através da colaboração com diferentes mentores. Os resultados estavam, de certa forma, espalhados de acordo com diferentes metodologias e procedimentos. Por isso, eu e meu orientador decidimos que seria necessário um estudo final. Os últimos seis meses foram marcados por muito trabalho e reflexão. Neste período, agradeço novamente a Thiago Cabral pela implementação das extensões de navegador que visavam monitorar pesquisadores em suas atividades de busca por artigos científicos. Agradeço também aos alunos Eric Velten de Melo e André Resende pela ajuda com algumas das figuras incluídas nesta tese. E, é claro, aos pesquisadores do Instituto de Computação da UNICAMP que

aceitaram participar no último estudo.

Por fim, gostaria de agradecer a Deus e ao meu anjo-guardião, por concederem todas as condições para que eu escrevesse esta tese. Agradeço também a cinco pessoas muito importantes na minha vida. Primeiro, ao meu orientador, Prof. Paulo Lício de Geus, pelo suporte e por acreditar em mim desde a minha iniciação científica, em 2003, até hoje. Segundo, à minha esposa, por todo apoio e compreensão nos momentos difíceis. Eu te amo e sempre amarei. Terceiro, à minha avó, que mesmo com a idade avançada foi capaz de cuidar de mim. Quarto, à minha mãe, o meu eterno amor. Por toda as suas palavras, gestos e sentimentos que estas meras palavras nunca serão capazes de exprimir. Você estará comigo sempre. E, finalmente, ao meu falecido avô, o pai que eu nunca tive. Esta tese é dedicada a você, onde quer que esteja.



# Sumário

Resumo	vii
Abstract	ix
Agradecimentos	xi
<b>1 Introdução</b>	<b>1</b>
<b>2 O processo de busca no trabalho de pesquisadores acadêmicos</b>	<b>5</b>
2.1 Introdução . . . . .	5
2.2 Estudo . . . . .	6
2.3 Resultados . . . . .	7
2.3.1 Organização em arquivos vs. a prontidão da informação na Web . .	7
2.3.2 Planos e situações reais de busca . . . . .	9
2.4 Discussão . . . . .	10
2.5 Conclusão . . . . .	12
<b>3 Monitorando a busca de pesquisadores acadêmicos</b>	<b>13</b>
3.1 Introdução . . . . .	13
3.2 Resultados . . . . .	15
3.2.1 Buscas EF- . . . . .	16
3.2.2 Buscas EF+ . . . . .	18
3.3 Discussão . . . . .	19
3.4 Conclusão . . . . .	21
<b>4 O processo de busca de novatos em Linux</b>	<b>23</b>
4.1 Introdução . . . . .	23
4.2 Estudo . . . . .	24
4.2.1 Participantes . . . . .	24
4.2.2 Procedimento . . . . .	24

4.3	Resultados . . . . .	26
4.3.1	Atividades de procura de informação . . . . .	26
4.3.2	Falta de experiência e a incerteza . . . . .	26
4.3.3	Informação e suas fontes preferenciais . . . . .	28
4.4	Discussão . . . . .	28
4.5	Conclusão . . . . .	29
<b>5</b>	<b>O papel da familiaridade na exploração do conhecimento</b>	<b>31</b>
5.1	Introdução . . . . .	31
5.2	Busca e seus diversos níveis de atividade . . . . .	32
5.2.1	O modelo consulta-resposta e a recuperação de informação . . . . .	32
5.2.2	A procura de informação . . . . .	33
5.2.3	A busca no contexto de trabalho . . . . .	34
5.2.4	Busca exploratória . . . . .	34
5.3	Uma visão sociotécnica sobre a busca na Web . . . . .	35
5.3.1	Comunidades de prática e sistemas de informação . . . . .	35
5.3.2	Espaços informacionais comuns . . . . .	38
5.3.3	Categorias como objetos de fronteira . . . . .	41
5.4	O modelo consulta-resposta e suas limitações inerentes . . . . .	43
5.4.1	Máquinas de busca e a cegueira de significado . . . . .	44
5.4.2	A metáfora da sala chinesa . . . . .	46
5.5	Teoria fundamentada em dados . . . . .	48
5.5.1	Trabalhando com informação . . . . .	49
5.5.2	Procurando informação . . . . .	51
5.5.3	Recuperando informação . . . . .	54
5.6	Propostas de soluções para o <i>design</i> de sistemas de busca . . . . .	54
5.6.1	Anotação colaborativa . . . . .	55
5.6.2	Hipertexto e o conceito de caminhos . . . . .	56
5.6.3	Redes sociais em suporte à busca . . . . .	57
<b>6</b>	<b>Conclusão</b>	<b>61</b>
6.1	Contribuições em artigos . . . . .	63
6.2	Trabalhos futuros . . . . .	65
	<b>Bibliografia</b>	<b>67</b>
<b>A</b>	<b>Materiais</b>	<b>73</b>
A.1	Estudo: Cap. 2 . . . . .	73
A.2	Estudo: Cap. 3 . . . . .	75

A.2.1	Material para o <i>survey</i> . . . . .	75
A.2.2	Material para entrevista semi-estruturada . . . . .	76
<b>B</b>	<b>Resultados do <i>survey</i> (Cap. 3)</b>	<b>77</b>



# Capítulo 1

## Introdução

A Internet apresenta-se como a maior fonte de conhecimento da atualidade. Vannevar Bush, no artigo seminal “*As We May Think*” [Bush, 1945], antecipou brilhantemente a invenção do hipertexto, do computador pessoal, da Web e das bibliotecas *online*.

*“Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and, to coin one at random, “memex” will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory”.* ([Bush, 1945], p. 6)

A criatividade de Bush propôs um sistema de utilidade pública, assim como é a Internet nos dias atuais. Entretanto, a utilidade da Internet está intimamente relacionada com a maneira como o conhecimento disponível é produzido e consumido. Sistemas de busca por palavras-chave, e.g. Google<sup>1</sup> realizam a tarefa de indexar documentos da Web e oferecem a funcionalidade de consulta a esses documentos. Por um lado, a eficiência desses sistemas sob o ponto de vista do usuário ainda depende, pelo menos em parte, das palavras-chave e termos utilizados na consulta. Por outro lado, ainda cabe ao usuário a análise, interpretação e síntese dos resultados retornados como resposta a uma consulta.

Estes não são problemas *per se*, mas características de um modelo de interação. Sistemas de busca seguem o modelo consulta-resposta, i.e. palavras-chave como *consulta* e uma lista de documentos como *resposta*. Seus algoritmos de indexação e ranqueamento fazem, sem dúvida, um trabalho grandioso, mas para aquilo que foram projetados: a filtragem de informações de acordo com os termos da consulta formulada pelo usuário. Vannevar Bush também previu esse modelo de interação:

---

<sup>1</sup><http://www.google.com>

*“There is, of course, provision for consultation of the record by the usual scheme of indexing. If the user wishes to consult a certain book, he taps its code on the keyboard, and the title page of the book promptly appears before him, projected onto one of his viewing positions. Frequently-used codes are mnemonic, so that he seldom consults his code book; but when he does, a single tap of a key projects it for his use”.* ([Bush, 1945], p. 6)

Os problemas deste modelo de interação emergem quando a interação do usuário com o sistema de busca não visa apenas a recuperação de informação, na forma de documentos individuais, mas uma atividade cognitiva de mais alto nível: a exploração de conhecimento. O objetivo desta tese é o de estudar em profundidade essa atividade cognitiva, que tem sido chamada na literatura científica de busca exploratória [White and Roth, 2009]. Uma busca exploratória é um tipo de atividade de procura por informação (IS, ou *“information-seeking”*), que pode durar diversos dias, meses e até anos, e.g. exploração de um novo tema de pesquisa científica [Renear and Palmer, 2009], busca de informação técnica por novatos, atividades de aprendizado. Tais atividades têm por característica a realização de múltiplas sessões de busca, visto que múltiplos documentos precisam ser analisados para cumprir com o objetivo de uma tarefa de trabalho mais geral. A atividade de IS torna-se exploratória quando o usuário não tem *familiaridade* com o assunto, enquanto que nas atividades de IS típicas, o usuário possui uma grande familiaridade com o domínio de conhecimento relacionado às suas buscas.

Esta tese tem como principal enfoque o estudo das atividades de exploratórias de busca, propondo, portanto, as seguintes perguntas de pesquisa:

1. Como usuários buscam, organizam e re-encontram documentos em seus ambientes de trabalho?
  - (a) Que estratégias são utilizadas e quais as considerações influenciam a escolha destas estratégias?
  - (b) Quais são os benefícios/limitações percebidos de cada estratégia a partir da perspectiva do usuário?
  - (c) Como essas estratégias podem variar de acordo com a motivação da busca e a familiaridade do usuário no assunto?
2. Quais os problemas enfrentados durante o processo de busca nos casos em que o usuário possui pouca familiaridade com o assunto?
  - (a) Quais as estratégias tomadas pelo usuário quando a busca torna-se difícil?
  - (b) Quais são as etapas mais problemáticas durante o processo?

3. Quais conclusões teóricas podem ser utilizadas de modo a propor soluções para os problemas antes identificados?
  - (a) Como podemos descrever os processos de busca de acordo com a variação de familiaridade do usuário para com o assunto buscado?
  - (b) Quais as implicações da teoria proposta no que tange ao *design* de sistemas de busca?

O Cap. 2 procura responder à parte da pergunta (1) e investiga as estratégias utilizadas por pesquisadores acadêmicos com a finalidade de re-encontrar documentos, especialmente quando estes usuários trabalham com vários documentos simultaneamente. Nestas situações, as estratégias dos usuários podem ter que se adaptar para atender a novos requisitos, e.g. mudanças de cronograma, surgimento de novos objetivos de busca. Os acadêmicos podem se beneficiar do uso generalizado da indexação digital e o surgimento em muitas disciplinas científicas de ontologias interoperáveis, que permitem que eles naveguem em paralelo através de artigos que estão disponíveis em diferentes bibliotecas digitais na Web [Renear and Palmer, 2009]. Ao mesmo tempo, estes mesmos pesquisadores possuem a necessidade de organizar documentos de acordo, por exemplo, com projetos em andamento. Os resultados do Cap. 2 revelam as estratégias e considerações tomadas por esses usuários ao preencher a lacuna entre espaços informacionais de uso geral, e.g. bibliotecas online, e de uso pessoal, e.g. pastas em seus computadores pessoais.

O Cap. 3 avança nas investigações apresentadas no Cap. 2 através do monitoramento das atividades de busca de pesquisadores acadêmicos, complementando a abordagem à pergunta (1). O objetivo deste monitoramento era o de coletar informações mais específicas sobre suas pesquisas, como a data em que a busca ocorreu e as palavras-chave que os usuários utilizaram em suas consultas. Deste modo, esperava-se que eles seriam capazes de lembrar mais detalhes de cada evento de busca quando confrontados com os registros de suas buscas. Os resultados identificaram dois tipos gerais de busca exploratória:  $EF+$ , buscas baseadas em estratégias mais elaboradas e em um menor grau de familiaridade com o assunto buscado, e  $EF-$ , buscas baseadas em estratégias menos elaboradas e em um menor grau de familiaridade.

O Cap. 4 procura responder à pergunta (2) e apresenta os resultados de um estudo realizado em laboratório. A mudança de método permitiu que observássemos buscas  $EF-$  mais diretamente, o que nos auxiliou a compreender as principais dificuldades enfrentadas pelos usuários durante o processo e *in situ*. Para este efeito, escolhemos a população de usuários novatos no sistema operacional Linux. Participantes foram convidados a utilizar o sistema de busca Google<sup>2</sup> para encontrar informações que os auxiliassem a resolver

---

<sup>2</sup><http://www.google.com>

problemas no ambiente operacional Linux. Observamos que a falta de familiaridade desses usuários com o sistema tem um impacto significativo na forma como eles interagem com o sistema de busca e em como eles analisam uma grande coleção de documentos da Web para comparar, agregar e sintetizar informações. A falta de familiaridade com o vocabulário peculiar ao assunto buscado pode levar à imprecisão, resultados ambíguos, delegando aos usuários o trabalho de analisar extensivamente os resultados fornecidos pelo sistema de busca em resposta a uma consulta.

Finalmente, o Cap. 5 procura responder à pergunta (3) e introduz uma teoria que é baseada nos resultados empíricos dos três estudos apresentados ao longo desta tese. A teoria visa identificar as principais condições, consequências e propriedades que estão relacionadas às variações de familiaridade dos usuários com o domínio de conhecimento relacionado às suas buscas. As atividades desempenhadas por esses usuários ocorrem num contexto mais amplo e por muitas vezes o sucesso ou a falha na realização de uma tarefa de busca depende de fatores externos ao sistema computacional utilizado. Comunidades *online*, grupos de trabalho e outras organizações têm influência direta no processo de obtenção, organização e aquisição de conhecimento através da Web e seus mecanismos de busca. Os resultados desta tese mostram que artefatos diversos, como notas em papel, artigos impressos, mapas e diversos outros documentos agem em conjunto com os sistemas de busca, com o objetivo de completar uma tarefa de busca complexa. Além disso, a comunicação com outros agentes humanos é fundamental para tomada de decisões sobre qual estratégia seguir durante o processo de busca. A presente tese termina com suas conclusões gerais e direções para trabalhos futuros no Cap. 6.

# Capítulo 2

## O processo de busca no trabalho de pesquisadores acadêmicos

### 2.1 Introdução

Usuários manipulam um grande número de documentos em seus ambientes de trabalho e muitas vezes têm dificuldade para encontrar tais documentos novamente [Boardman and Sasse, 2004]. Estudos têm argumentado que os usuários preferem navegar utilizando estruturas de arquivos como pastas em vez de utilizar sistemas de busca por palavras-chave para encontrar novamente os seus documentos [Barreau and Nardi, 1995, Jones et al., 2005, Bergman et al., 2008], mesmo que um motor de busca “ideal” esteja disponível [Teevan et al., 2004]. Por exemplo, uma descoberta predominante destes estudos é que as pessoas preferem não perder tempo recordando palavras-chave relevantes para recuperar um documento, dado que elas já conhecem a sua localização exata em seus desktops [Bergman et al., 2008]. Outros podem sentir-se desconfortáveis em abrir mão de suas estruturas de arquivos e muitas vezes preferem manter certos documentos pessoais juntos em pastas nas suas estruturas de arquivos [Jones et al., 2005].

No entanto, a web tem fornecido aos usuários uma grande variedade de fontes de informação e mecanismos de busca poderosos. A disponibilidade imediata das informações fornecidas por estes mecanismos através de uma simples consulta por palavras-chave pode potencialmente aliviar os usuários da necessidade de manutenção e organização de informações recuperadas a partir da web, uma vez que a mesma informação pode ser facilmente encontrada novamente usando os motores de busca disponíveis na web [Jones et al., 2001]. Além disso, o re-encontro de documentos pode não estar limitado às páginas web: outros tipos de documentos podem ser necessários novamente durante a rotina no ambiente de trabalho, tais como imagens e arquivos pdf, que são explicitamente baixados para manipulação em um ambiente de computação pessoal.

Neste capítulo, estamos interessados nas estratégias utilizadas por usuários com a finalidade de re-encontrar documentos, especialmente quando se trabalha com vários arquivos simultaneamente. Nestas situações, as estratégias dos usuários podem ter que se adaptar para atender a novos requisitos, e.g. mudanças de cronograma, surgimento de novos objetivos de busca. Para esta finalidade, focamos na população de pesquisadores acadêmicos. Os acadêmicos podem se beneficiar do uso generalizado da indexação digital e o surgimento em muitas disciplinas científicas de ontologias interoperáveis, que permitem que eles naveguem em paralelo através de artigos que estão disponíveis em diferentes bibliotecas digitais na Web [Renear and Palmer, 2009].

Trabalhos anteriores têm tentado responder a algumas das perguntas de pesquisa enunciadas na introdução desta tese. Boardman e Sasse [Boardman and Sasse, 2004] investigaram as estratégias para armazenamento de arquivos no ambiente do *desktop*. Bergman et al. [Bergman et al., 2008] levantaram a hipótese de que a disponibilidade de um mecanismo de busca por palavras-chave mais poderoso aumentaria o uso da função de busca. Nenhum destes estudos, no entanto, consideram o papel dos mecanismos de busca disponíveis na Web para o re-encontro de informação. Outros estudos foram além do ambiente de trabalho, investigando as práticas dos usuários no gerenciamento de informações obtidas através da Web [Jones et al., 2001, Aula et al., 2005, Capra et al., 2010]. No entanto, embora estes estudos fizeram observações importantes, eles focaram nas estratégias de usuários para re-encontrar as páginas web. Em particular, esperamos que os usuários se comportem diferentemente em relação a outros tipos de documentos adquiridos a partir da web: enquanto as páginas web são normalmente acessadas novamente, outros documentos podem precisar ser armazenados e, eventualmente, organizados para a manipulação direta no ambiente computacional de cada usuário. Isto traz outras possibilidades para o re-encontro de informação que não estão limitadas somente aos motores de busca da Web, mas podem incluir também mecanismos de busca baseados no *desktop*, e.g. Spotlight para Mac, e estratégias de navegação, e.g. pastas.

## 2.2 Estudo

A pesquisa foi conduzida como um estudo qualitativo. O método utilizado foi o de Teoria Fundamentada em Dados (*Grounded Theory* ou GT) [Strauss and Corbin, 1998]. Foram realizadas 11 entrevistas semi-estruturadas (ver material utilizado na Seção A.1), com pesquisadores acadêmicos trabalhando em tempo integral (10 homens e 1 mulher) em um laboratório de pesquisa na França. Dois eram pós-doutorandos, seis estudantes (quatro de doutorado, 2 de mestrado), dois professores permanentes e um era um programador em tempo integral, que também realizava atividades de pesquisa. Nove utilizavam Macintosh, um usava PC e um usava Linux.

Cada entrevista durou cerca de 25 minutos e foi realizada no local de trabalho dos participantes ou em uma sala separada, se eles compartilhavam um escritório. Perguntamos a cada participante sobre suas experiências recentes em re-encontrar documentos. Nosso objetivo foi o de caracterizar os possíveis problemas que poderiam surgir a partir da interação entre estratégias baseadas na Web e baseadas no *desktop* para re-encontrar documentos. Por exemplo, nós pedimos para os participantes lembrarem momentos em que eles tiveram problemas para encontrar um documento novamente. Tomamos nota e gravamos cada entrevista com uma câmera de vídeo. As entrevistas foram transcritas e analisadas utilizando codificação aberta (*open coding*), de modo a delinear conceitos que representavam blocos de dados em estado bruto [Strauss and Corbin, 1998]. Não houve categorias pré-definidas ou teorias estabelecidas antes da análise.

Como resultado, identificou-se um fenômeno central de interesse: quando se trata de artigos científicos, pesquisadores acadêmicos preferiram usar motores de busca baseados na web e bibliotecas digitais online para re-encontrar a informação, embora soubessem ter armazenado a informação ou documento no passado em seus computadores pessoais. A partir da identificação desse fenômeno, foi iniciado o processo de codificação axial (*axial coding*) [Strauss and Corbin, 1998] para identificar as categorias de codificação que poderiam explicar o nosso fenômeno central. Os resultados são apresentados na próxima seção.

## 2.3 Resultados

### 2.3.1 Organização em arquivos vs. a prontidão da informação na Web

Os participantes de nosso estudo muitas vezes encontravam-se às voltas com a seguinte decisão: (a) armazenar e organizar documentos adquiridos a partir da Web para posterior recuperação utilizando estruturas de arquivos, tais como pastas ou (b) pesquisar documentos diretamente na Web usando motores de busca baseados em palavras-chave.

Alguns participantes desenvolveram suas próprias estratégias para lidar com essas situações no seu trabalho diário. Por exemplo, alguns participantes foram mais consistentes que outros e seguiram estratégias pré-definidas de acordo com suas necessidades:

*“I don’t store stuff on my computer, only stuff that I’ve made myself, because I’m too messy... also papers I rather look them up on Google Scholar or in the digital library then actually have them on my computer”.* [P1]

*“I don’t store them [papers] on my computer, because papers have this strange naming scheme, they use a letter and a number... I could rename them, but I’m*

*lazy to do that and don't know how to rename them because maybe [sometimes] I don't know the title of the paper, I just know the content, sometimes I just know the title, not the content... I just want to type anything I want... this is what search such as Google does for me". [P11].*

Ao contrário de hierarquias de pastas, as bibliotecas digitais disponíveis na Internet oferecem aos usuários mais do que apenas um meio de para re-encontrar documentos. Os participantes relataram o uso de vários recursos que não estão prontamente disponíveis em seus *desktops*, como a recuperação formatada para citações de artigos científicos e a navegação através de *links* para referências relacionadas:

*"I use the ACM Directory because... it gives me another information which is the bibTex, and often I have to put the bibTex too in my paper". [P3]*

*"I'm used to research them [scientific articles] on the ACM [Library], rather than finding them in my personal files. Because I'm used to brainstorm over the subject using keywords, you know, when you are looking for [scientific] papers and then you can get the related ones". [P5]*

No entanto, outros participantes demonstraram enfrentar constantemente o *trade-off* entre o que deve ser armazenado em seus computadores e que não deve. Muitos deles relatam o desejo de manter os documentos juntos em pastas, por exemplo, se o participante está atualmente trabalhando sobre estes documentos [Barreau and Nardi, 1995] ou se tais documentos estão relacionadas a um determinado projeto [Jones et al., 2005]. No entanto, para obter informações de uso efêmero (*ephemeral information*) [Barreau and Nardi, 1995], ou seja, documentos que são necessários por apenas um curto período de tempo, os participantes revelaram que ainda preferem usar a pesquisa por palavras-chave na Web para re-encontrar documentos:

*"I save them all together, for example, if they are related to a particular project. In fact I would say that many times I try to look on the web first, so I don't save them [papers] on the computer, unless it is... a group of things... Because I won't use bookmarks". [P2]*

A organização de arquivos pode ser complicada de manter, especialmente para um grande número de documentos. Nestes casos, as necessidades da organização são voláteis, e os participantes muitas vezes desistem de manter seus documentos organizados em algum ponto nas suas rotinas de trabalho. Alguns participantes reconheceram os motores de busca baseados na Web como ferramentas úteis para re-encontrar documentos que são perdidos ou desaparecidos em seus computadores:

*“When I’m starting a project or when I need related work, I install the pdf files in folders, I use colors, to know the ones that I have already read... the problem is that this [organization] never goes really far, at some point I just forget to keep this updated... very often I go back to the ACM website to get a paper that I knew it was [stored] on my computer, but I can’t find”. [P7]*

Outros participantes revelaram que raramente usam a pesquisa na Web para re-encontrar documentos, ainda que esta apareça como uma opção confiável para eventos inesperados:

*“Usually if I know it is stored on my computer, I rarely use the web, but sometimes, specifically for papers, I forget that I already download it [a research article], and just re-download it... it is just faster”. [P4]*

### 2.3.2 Planos e situações reais de busca

Uma das surpresas do nosso estudo é o fato de que muitas vezes os participantes criam estereótipos de si mesmos usando termos como “confuso”, “preguiçoso” e assim por diante. No entanto, esses estereótipos não condiziam com seus comportamentos em se tratando de situações do mundo real. Por exemplo, na seguinte transcrição o participante demonstra ser orgulhoso de sua própria maneira de organizar as coisas, em comparação com os outros. No entanto, logo após o participante recorda uma condição sob a qual ele se comportaria de modo não usual:

*“I know there is a lot of people that use search to find a document, but I use organization, you see, I have bookmarks... I know that there is a lot of things that are not useful anymore, so I keep it and then I rebuild a more clean bookmarks... I like bookmarks”. [P3]*

*“[Q: But does it happen for you to search for a document using keywords?] Yes... for example, I need to found back paper. I know I downloaded this paper, so I can start to look at my hierarchy... but sometimes it is difficult, I do not find it here... so I prefer going to the ACM directory”. [P3]*

Planos pré-concebidos também podem atuar contra a eficiência no re-encontro de documentos. Os participantes que demonstraram aderir com frequência a suas estratégias, também relataram gastar muito tempo agindo de acordo com seus hábitos, assim, mais tarde, percebendo que o plano executado não foi eficiente em determinada situação:

*“I always take a huge amount of time renaming [scientific articles] with the title... even that is not completely clear, I mean, sometimes I have to use the spacebar just to see the first the images to see if it’s really what I’m looking for. [Q: Why don’t you go to the Web to find the paper again?] Usually is stubbornness, I know it will probably be faster for me to find it on my laptop... but once I’m searching, at some point I will say, okay, the website will be faster”. [P7]*

Outro resultado interessante é que algumas vezes participantes não planejam suas ações durante suas atividades de pesquisa. O comportamento do usuário é tão familiar e integrado em sua rotinas de trabalho que acaba se tornando automatizado e transparente a ele:

*“I don’t really remember what I do when I do this kind of thing because it is fairly well integrated in my way of looking for documents. So I don’t pay attention to it”. [P9]*

Além disso, as pessoas vão e voltam entre as estratégias nas situações do mundo real. Portanto, o comportamento de busca não é frequentemente ligado à dicotomia “navegar por pastas” ou “pesquisar usando palavras-chave”. Isto também pode ser definido como um tipo de comportamento de orientação (*orienting behavior*) [Teevan et al., 2004], através do qual as ações são motivadas por eventos emergentes ao contexto *in situ*:

*“I wouldn’t want to separate things. I mean, it is more like sort of a spectrum. On one hand, you know precisely where to go and where to start with to find and [on the other hand] you don’t know anything at all. So here [the former] you just simply use the folder structure and over here [the latter] you are only able to use the search mechanism. But most of the things are in between, where you can use the folder structure to automatically limit the search mechanism, so I think I use them very fluently”. [P9]*

## 2.4 Discussão

Nossos resultados indicam alguns pontos que devem ser cuidadosamente considerados na elaboração de conclusões sobre as práticas dos usuários para gerir informação no ambiente de trabalho. Primeiro, o uso disseminado de aplicações baseadas na web tem modificado o modo de trabalhar com documentos digitais. Como tal, ao concentrar a investigação exclusivamente sobre as ferramentas disponíveis no *desktop* dos computadores pessoais, tais como estruturas de arquivos, os trabalhos científicos têm ignorado uma série

de práticas que são realizadas fora da própria estrutura de arquivos. Em segundo lugar, a população do estudo foi escolhida com um propósito claro: pesquisadores acadêmicos têm a necessidade de gerenciar vários documentos simultaneamente para realizar o seu trabalho. Embora a escolha desta população traga ameaças à validade dos nossos resultados, ela nos ajudou a destacar um fenômeno de importância crescente: a fronteira do ambiente do *desktop* atual não pode ser considerada da mesma forma que há duas décadas. Novas ferramentas estão surgindo no ambiente web. No caso dos acadêmicos, verificou-se uma significativa transparência entre o uso de ferramentas do *desktop* e da web. Este comportamento é tão comum para eles que muitas vezes torna-se naturalizado: os usuários não percebem a mudança de um ambiente para o outro. Em terceiro lugar, alguns dos trabalhos anteriores na área concentraram a investigação sobre o comportamento dos usuários para re-encontrar páginas web. Nossos resultados indicam que os usuários comportam-se de maneira diferente quando comparamos o uso de documentos digitais como artigos científicos e páginas da web: no caso do primeiro, o documento muitas vezes precisa ser mantido em estruturas de arquivos para a manipulação durante um determinado período de tempo, ao contrário do último caso, em que o documento é tipicamente re-acessado e muitas vezes dispensa qualquer tipo de organização nos computadores pessoais dos usuários.

No entanto, problemas de integração entre ferramentas do *desktop* e da web podem surgir quando os usuários precisam re-encontrar documentos que foram previamente manipulados no ambiente de computação pessoal. Por exemplo, bibliotecas digitais podem ser consideradas como ferramentas de gerenciamento de informação gerais (*general information management* ou GIM), i.e. sistemas em que profissionais de informação como bibliotecários e editores, gerem a informação disponível para um público diverso, e devem acrescentar atributos que são objetivos e gerais para atender às necessidades de muitos usuários [Bergman et al., 2003]. Isto representa uma diferença significativa entre os ambientes de computação pessoal e sistemas de GIM. Enquanto no primeiro os usuários exercem um controle total sobre os documentos, no segundo esse controle é delegado a outras pessoas. Assim, apesar de muitos participantes de nosso estudo relatarem a necessidade de manter as coisas em suas estruturas de arquivos, por exemplo, para um dado projeto em andamento, nenhum participante conseguiu reportar uma maneira eficiente de fazer o mesmo usando as ferramentas da Web. Além disso, nossos participantes consideraram os motores de busca e bibliotecas digitais baseados na Web como um meio de re-encontrar a informação de forma eficiente, mas o estado-da-arte dessas ferramentas não fornece aos usuários maneiras de organizar a informação encontrada de forma similar àquela utilizada em seus ambientes computacionais pessoais. Esta é uma indicação de que futuras ferramentas para gestão da informação deverão exigir o poder dos atuais mecanismos de busca baseados na Web para re-encontrar a informação, em conjunto com meios eficazes para

categorizar a informação recuperada em coleções que atendem às necessidades pessoais dos usuários.

Para pesquisadores acadêmicos, ferramentas como o *Zotero*<sup>1</sup> e *Papers* para Macintosh<sup>2</sup> têm fornecido um ambiente integrado para recuperar, organizar e re-encontrar artigos científicos. No entanto, nenhum dos nossos participantes relataram usá-los. Pode ser o caso de que essas ferramentas ainda não podem substituir as ferramentas disponíveis no *desktop* e àquelas baseadas na web simultaneamente. Ou talvez as estratégias utilizadas pelos acadêmicos com ferramentas convencionais, e.g. pastas e bibliotecas digitais, são tão bem integradas nas suas rotinas de trabalho, que eles não sentem a necessidade de mudar seus métodos atuais de busca.

## 2.5 Conclusão

Nós apresentamos os resultados de um estudo que visa preencher a lacuna na literatura (i) investigando como pesquisadores acadêmicos gerenciam múltiplos documentos adquiridos a partir da Web para posterior recuperação e (ii) compreendendo o papel da pesquisa por palavra-chave, considerando os motores de busca da Web como ferramentas em potencial para recuperação de documentos. Descobrimos que as ferramentas de busca por palavras chave são úteis não somente quando a localização do documento no ambiente computacional pessoal é desconhecida, mas também quando os custos de recuperação são tão baixos que os usuários não se incomodam em armazenar documentos que podem ser facilmente recuperados usando os motores de busca e bibliotecas digitais disponíveis na Web.

A pesquisa por palavras-chave na Web é uma ferramenta que pode ser usada para diversas aplicações e atividades, o que aumenta seu poder e usabilidade. No entanto, descobrimos que os usuários ainda sentem a necessidade de organizar certos tipos de documentos, especialmente se eles estão relacionados a um projeto. Neste caso, os mecanismos de busca e bibliotecas digitais da Web falham ao fornecer aos usuários maneiras eficientes de organizar aquelas informações.

---

<sup>1</sup><http://www.zotero.org/>

<sup>2</sup><http://www.mekentosj.com/papers/>

# Capítulo 3

## Monitorando a busca de pesquisadores acadêmicos

### 3.1 Introdução

O objetivo do estudo<sup>1</sup> apresentado neste capítulo foi o de avançar e complementar a investigação cujos resultados foram apresentados no capítulo anterior (Cap. 2). O método utilizado foi o mesmo, i.e. Teoria Fundamentada em Dados ou Grounded Theory (GT) [Strauss and Corbin, 1998]. O método GT recomenda que os períodos de coleta e análise de dados sejam entrelaçados. Como tal, os resultados do capítulo anterior serviram como suporte para o planejamento de uma segunda fase de coleta e análise de dados, cujos resultados são apresentados neste capítulo. Foi realizada uma amostragem teórica para investigar as condições que influenciam o fenômeno central antes identificado, i.e. pesquisadores acadêmicos preferem acessar serviços Web para recuperar documentos, mesmo que já tenham esses documentos armazenados em seus computadores pessoais. Strauss e Corbin [Strauss and Corbin, 1998] propõem que a técnica de amostragem teórica deve localizar os diferentes *contextos* em que o fenômeno ocorre, identificando as *condições* e as *consequências* que geram tais contextos e que, portanto, provocam variações no fenômeno estudado.

Uma das dificuldades encontradas no estudo anterior foi a de que os participantes mostraram certa dificuldade ao recordar detalhes de suas atividades de pesquisa, apesar de as terem realizado recentemente. Decidimos então monitorar as atividades de busca de pesquisadores acadêmicos usando um *logger* desenvolvido como uma extensão para os navegadores para os navegadores Mozilla Firefox<sup>2</sup> e Google Chrome<sup>3</sup>. Este *logger* foi

---

<sup>1</sup>Estudo número 1103/2010 aprovado pelo Comitê de Ética em Pesquisa (CEP) da UNICAMP.

<sup>2</sup><http://www.mozilla.com>

<sup>3</sup><http://www.google.com/chrome>

instalado em seus computadores pessoais. O objetivo deste monitoramento era o de coletar informações mais específicas sobre suas pesquisas, como a data em que a busca ocorreu e as palavras-chave que os usuários utilizaram em suas consultas. Deste modo, esperava-se que eles seriam capazes de lembrar mais detalhes de cada evento de busca durante entrevistas semi-estruturadas. No entanto, a fim de lidar com questões de privacidade, nós restringimos os dados coletados para serviços da Web de interesse: os motores de busca da Web e bibliotecas digitais *online* tipicamente usadas para recuperar artigos científicos.

Para este efeito, foi realizada uma pesquisa preliminar na forma de um *survey* (ver material utilizado na Seção A.2.1 e resultados no apêndice B) com pesquisadores acadêmicos de um instituto de pesquisa no Brasil para determinar os serviços Web usados com mais frequência para recuperar artigos científicos. O questionário apresentado aos participantes continha um conjunto de bibliotecas digitais nas áreas relacionadas à pesquisa em computação, juntamente com os motores de pesquisa na Web de uso generalizado. Eles poderiam selecionar a frequência de utilização de cada serviço, utilizando uma escala Likert. O questionário também forneceu os respondentes um campo de texto, no qual eles podem fornecer informações complementares, e.g. serviços não disponíveis entre as opções do questionário. 76 acadêmicos responderam ao *survey*, o que nos deu uma ideia geral das ferramentas de busca tipicamente usadas dentre os pesquisadores acadêmicos que trabalham no instituto escolhido.

Após a realização do *survey*, foi realizado um estudo qualitativo que consistiu em uma segunda fase de coleta e análise de dados. 20 pesquisadores acadêmicos (19 homens e 1 mulher) foram selecionados de acordo com a suas disponibilidades de tempo e interesse para participar no estudo. Foram monitoradas as suas atividades de navegação na Web por um período de seis semanas usando o *logger* desenvolvido para o estudo. O *logger* coletou (a) as consultas enviadas a um motor de busca na web, e.g. Google Scholar, ou a uma biblioteca digital, e.g. ACM Portal, e (b) os cliques em resultados retornados como resposta à consulta. Por questões de privacidade, os dados só eram coletados quando o clique levava a uma biblioteca digital. As bibliotecas monitoradas foram selecionadas baseado nos resultados do *survey* que serão apresentados na próxima seção. Apenas as bibliotecas com maior utilização foram escolhidas, i.e. ACM Portal<sup>4</sup>, IEEE Xplore<sup>5</sup> e SpringerLink<sup>6</sup>. Foram monitorados também os sistemas de busca Google, o sistema de indexação CiteSeer<sup>7</sup> e o Google Scholar<sup>8</sup>. Os participantes tinham a oportunidade de

---

<sup>4</sup><http://portal.acm.org>

<sup>5</sup><http://ieeexplore.ieee.org>

<sup>6</sup><http://www.springerlink.com>

<sup>7</sup><http://citeseer.ist.psu.edu>

<sup>8</sup>Os sistemas Google e CiteSeer não estavam disponíveis explicitamente como opção no questionário, mas muitos participantes indicaram utilizá-los para busca e recuperação de artigos científicos com bastante frequência, o que motivou a inclusão destes sistemas no monitoramento.

revisar os dados coletados antes de enviá-los para um servidor central através de um ambiente seguro. O *logger* lembrava cada participante para que eles enviassem os dados coletados a cada intervalo de dois dias.

Entre os 20 participantes, 11 (10 homens e 1 mulher) enviaram dados dentro do período de coleta de dados. 583 consultas e 134 cliques foram coletados. Com base nos dados registrados, foram realizadas entrevistas semi-estruturadas (ver material utilizado na Seção A.2.2). Cada entrevista durou de 10 a 15 minutos e foi marcada não mais que duas semanas depois do primeiro evento registrado de cada participante, o que significa que as entrevistas foram realizadas durante todo o período de coleta de dados. Nesta fase, os eventos registrados foram utilizados para lembrar os participantes sobre suas atividades recentes de busca acadêmica. Nós perguntamos a eles sobre as suas motivações por trás de cada consulta, sua familiaridade com o tema da pesquisado, os artefatos criados e utilizados durante o processo de busca e como a interação com outros pesquisadores auxiliaram nas estratégias de busca tomadas.

Após a transcrição das entrevistas, foi realizada uma segunda fase de codificação aberta e axial. O objetivo nesta fase de análise foi o de (1) saturar as principais categorias que se relacionam com o nosso fenômeno central (ver capítulo anterior) e (2) iniciar a codificação seletiva para apresentar um modelo teórico do processo em estudo [Strauss and Corbin, 1998].

## 3.2 Resultados

Como resultado deste estudo, foram identificados dois padrões de comportamento no que tange à busca exploratória:

**Estratégia menos elaborada e/ou menor familiaridade (EF−)** O usuário ainda não adquiriu conhecimento o suficiente para produzir consultas precisas, que tipicamente retornam resultados bastante relevantes. É também o caso de usuários que não estipularam *a priori* estratégias mais elaboradas de busca e terminam por usar alguma estratégia de forrageamento da informação [Pirolli and Card, 1995], ou mesmo algum comportamento de orientação pelo espaço informacional [Teevan et al., 2004].

**Estratégia mais elaborada e/ou maior familiaridade (EF+)** O usuário já conhece a terminologia relacionada ao assunto buscado e encontra-se num estágio mais avançado com relação às melhores estratégias para obter documentos relevantes. Estes usuários também utilizam artefatos auxiliares diversos como suporte às suas atividades de busca, assim como dicas de colegas de trabalho e/ou orientadores que compartilham de alguma familiaridade com o assunto.

### 3.2.1 Buscas EF–

A Tabela 3.1 apresenta um exemplo de busca EF–. Da esquerda para direita, são mostrados (1) os dias, (2) o sistema utilizado para realizar a busca, (3) a consulta na forma de palavras-chave, (4) o número da página retornada pelo sistema de busca com os resultados e (5) o número de resultados clicados nesta página. Os dias servem apenas para indicar que o usuário realizou a busca em dias distintos.

Tabela 3.1: Exemplo de busca EF–.

Dia	Sistema	Consulta	Pág.	Cli.
1	Google Scholar	virtual environments with kvm	1	2
2	Google Scholar	virtual enviroments for research	1	0
2	Google Scholar	virtual environments for research	1	0
2	Google Scholar	virtual environments for research kvm	1	3
2	Google Scholar	virtual environments for research	1	0
2	Google Scholar	virtual environments for research	2	0
2	Google Scholar	virtual environments for research	3	0
2	Google	KVM.whitepaper filetype:pdf	1	1
2	Google Scholar	virtual environments for research kvm	1	1
2	Google Scholar	virtual environments for research kvm	2	0
2	Google Scholar	virtual environments for research kvm	3	0
2	Google Scholar	virtual environments for research kvm	4	0
2	Google Scholar	virtual environments for research kvm	5	0
2	Google Scholar	virtual environments for research kvm	6	0
3	Google Scholar	virtualization benefits	1	0
3	Google Scholar	“virtualization benefits”	1	1
3	Google Scholar	“research and educational” testbeds	1	0
3	Google Scholar	“research and educational” testbeds virtual machines	1	2
4	Google Scholar	“virtualization benefits”	1	0
4	Google Scholar	scenario-based virtualization	1	0

*“Porque a gente mandou uma proposta de mini-curso [...] para montar ambientes virtuais que auxiliam à pesquisa. [...] O que eu estava procurando eram referências para fundamentar a proposta. [...] Que benefícios você tem usando virtualização para ensino ou pesquisa [...] Eu queria mais uma parte introdutória”. [P02]*

O participante também revelou sua familiaridade com o assunto relacionado à busca:

*“Razoavelmente eu conheço [...] outra vez que eu fui procurar esse tipo de coisa, era mais a parte prática, mais a parte de implementação [...] agora foi um pouco mais pro lado acadêmico, porque para o curso ser aprovado, a gente precisa fundamentá-lo”. [P02]*

Na Tabela 3.1, pode-se observar algumas características das buscas do tipo EF-. Primeiro, a maioria das atividades de recuperação de informação gera nenhum ou poucos cliques em resultados, e.g. artigos científicos que são baixados para o computador do usuário. Segundo, termos utilizados para consulta são inicialmente bastante gerais e tornam-se mais específicos à medida que o usuário aprende com o conteúdo recuperado ao longo de várias sessões de busca. O participante, no exemplo, realizou diversas atividades de recuperação espalhadas ao longo de quatro dias. Apenas no terceiro ele introduziu o termo “*testbeds*”, que foi aprendido depois da análise de um dos documentos recuperados. Note que a adição do termo possibilitou a recuperação de dois novos artigos (antepenúltima consulta).

*“O termo “testbeds” já foi oriundo do artigo que eu li [...] eu aprendi o termo no artigo”. [P02]*

A Tabela 3.2 mostra outro exemplo de busca do tipo EF-. A motivação da busca foi explicitada pelo participante, bem como a sua familiaridade com o assunto:

*“Esse é o tema [...] de doutorado. Não está totalmente definido, eu estou fazendo o levantamento bibliográfico para ver o que já tem, principalmente em cima de frameworks e modelos. [...] Eu não domino o assunto, mas não é totalmente desconhecido”. [P05]*

Pode-se observar que mesmo adicionando e removendo palavras-chave de suas consultas, o participante não clicou em nenhum resultado. Entretanto, nas duas últimas consultas o participante parece ter mudado de estratégia, colocando nomes de artigos como consulta. Durante a entrevista, o participante revelou que tais buscas foram realizadas após sugestão de artigo provida pelo orientador. Este artefato, i.e. um artigo científico relacionado ao tema, motivou novas buscas por parte do participante:

*“Eu usei porque a minha orientadora me passou um artigo porque eu não tinha achado em nenhuma dessas buscas e a partir das referências desse artigo eu cheguei a fazer algumas buscas por palavras-chave ou até por nome”. [P05]*

Tabela 3.2: Outro exemplo de busca EF- (algumas consultas foram parcialmente omitidas).

Dia	Sistema	Consulta	Pág.	Cli.
1	ACM Portal	framework emotion tic education	1	0
1	ACM Portal	framework emotion ict education	1	0
1	ACM Portal	framework emotion education design with children	1	0
1	ACM Portal	framework emotion education design with children	2	0
1	ACM Portal	framework emotion education design with children	3	0
1	ACM Portal	framework emotion education design with children	4	0
1	ACM Portal	framework affect design with children learning	1	0
1	Google	Developing [...] for capturing self reports of affect	1	1
1	Google	“User [...] behavior model in an educational game”	1	1

### 3.2.2 Buscas EF+

A Tabela 3.3 mostra um exemplo de busca exploratória com maior familiaridade (EF+). O participante foi perguntado sobre a sua motivação para realizar a busca:

Tabela 3.3: Exemplo de busca EF+ (algumas consultas foram parcialmente omitidas).

Dia	Sistema	Consulta	Pág.	Cli.
1	Google Scholar	“Lighting analysis [...] in realist paintings”	1	1
1	Google Scholar	“Estimating the directions [...] human face”	1	1
1	Google Scholar	“Color Constancy through [...] Space”	1	1
1	Google Scholar	“Statistical multiple light source detection”	1	1
1	Google Scholar	“A model for image splicing”	1	1
1	Google Scholar	forgery + illumination + light + doctored	1	2
1	Google Scholar	forgery + illumination + light + doctored	2	1

*“Estou buscando informações relacionadas à detecção e análise de iluminação em imagens de forma geral”. [P07]*

O participante usou uma estratégia mais elaborada para explorar o tema:

“Eu fui olhando nos sites de revistas conceituadas e bons congressos na área de visão computacional e computação gráfica, daí os artigos que pareciam ter algo haver com o meu interesse eu buscava”. [P07]

As duas últimas consultas do participante, ao contrário das demais, não utilizam frases de consulta (“*phrase queries*”), i.e. consultas utilizando aspas duplas. Perguntado sobre a sua familiaridade com os termos utilizados, o participante respondeu:

“[Q: Vejo que aqui você não estava procurando por um artigo específico, ou estava?] Não, estava buscando por artigos que contivessem esses termos. [Existia algum termo que você aprendeu recentemente, ou todos já eram familiares a você?] Todos já eram familiares. [Q: E como você aprendeu esses termos?] Doutorado, uma disciplina”. [P07]

O exemplo da Tabela 3.3 mostra que o usuário fez uso de outros artefatos informacionais, i.e. artigos científicos listados em *proceedings* de conferências, para organizar suas buscas. De fato, o participante revelou usar sistematicamente planilhas para cobrir a literatura publicada nos principais veículos científicos relacionados à sua motivação de busca. A Fig. 3.1 mostra uma planilha que foi utilizada para este propósito.

AÇÃO	TÍTULO	LOCAL	AUTORES	ANO	RESUMO
LER AGORA	Illumination and Spatially Va	CVPR	Kenji Hara and Ko Nishino	2009	
LER AGORA	(De) Focusing on Global Ligt	CVPR	Mohit Gupta† , Yuandong Tian†, Srinivasa G. Narasi	2009	
LER AGORA	Learning to Recognize Shad	CVPR	Jiejie Zhu, Kegan G.G. Samuel, Syed Z. Masood, M	2010	
LER AGORA	Face Re-Lighting from a Sing	CVPR	Yang Wang1 , Zicheng Liu2 , Gang Hua3 , Zhen We	2007	
LER AGORA	Robust Shadow and Illumina	CVPR	Alexandros Panagopoulos1 , Dimitris Samaras1 and	2009	
LER AGORA	Separating Reflections in Hu	ICCV	Huiqiong Wang - Stephen Lin - Xiaopei Liu - Sing Bir	2005	
LER AGORA	Using Eye Reflections for Fa	ICCV	Ko Nishino - Peter N. Belhumeur - Shree K. Nayar	2005	

Figura 3.1: Trecho da planilha utilizada pelo participante P07 para organizar a exploração do seu tema de doutorado.

Desse modo, embora estivesse realizando uma busca exploratória, as consultas na forma de frases permitiram que ele fosse diretamente ao documento desejado. Mesmo quando realizou consultas por termos, sua eficiência para recuperar conteúdo relevante foi maior, dado seu conhecimento prévio no tema e a uma maior familiaridade com as principais fontes de informação.

### 3.3 Discussão

Descobriu-se que toda busca de caráter exploratório possui uma *motivação* que, por sua vez, está atrelada a uma ou mais necessidades informacionais dos usuários. Durante as

entrevistas, participantes foram perguntados sobre a motivação que os levaram a realizar suas buscas, usando como referência os dados coletados pelo *logger*. Um exemplo de motivação típica encontrada entre os pesquisadores é a realização de revisões bibliográficas para formulação de propostas de tese ou para escrita de artigos científicos.

Assim como foi apresentado no Cap. 2, participantes tomaram diferentes *estratégias* para realizar suas buscas. Neste quesito, a *familiaridade* com o assunto buscado mostrou-se bastante influente sobre tais estratégias. Mais do que isso, podemos afirmar que a familiaridade é uma *condição contextual* central com relação ao fenômeno estudado. É uma condição pois causa variações nas estratégias adotadas pelos usuários, e é contextual, pois ao mesmo tempo influencia e é influenciada pelo processo de busca. Usuários seguem uma determinada estratégia de acordo com sua familiaridade inicial, i.e. antes do início da busca, e modificam suas estratégias ao longo do processo, uma vez que aprendem mais sobre o assunto buscado. Este ciclo de aprendizado é fundamental em buscas exploratórias, pois cria novas condições de familiaridade com relação ao tópico buscado.

Por outro lado, artefatos auxiliares e agentes humanos mostraram funcionar como condições significativas na elaboração de estratégias exploratórias de busca. Artefatos, a exemplo de planilhas e notas, auxiliam o usuário a tomar estratégias mais elaboradas e a guiarem-se com mais precisão pelo espaço informacional. Agentes humanos, a exemplo de orientadores e colegas de trabalho, auxiliam usuários com menor familiaridade no assunto buscado através de dicas na forma de estratégias eficientes, e.g. anais de conferências a serem investigados, e também, na sugestão de referências, e.g. outros colegas ou mesmo outros artefatos como artigos científicos que servem como ponto de partida para buscas mais direcionadas.

Em resumo, os resultados mostram que, embora os mecanismos de busca e bibliotecas digitais da Web sejam preferidos para recuperação de documentos científicos, o processo de busca é influenciado por uma série de condições que, por sua vez, produzem consequências na estratégia seguida pelos usuários. Podemos identificar as seguintes relações entre condições e consequências para o processo de busca exploratória de pesquisadores acadêmicos:

1. A falta de familiaridade com o assunto buscado pode tornar o processo de busca exploratória penoso e ineficaz.
2. Processos de busca de caráter exploratório podem exigir diversas sessões de busca, de modo que todo o planejamento estratégico, bem como o auxílio de outros agentes humanos são fundamentais para a eficiência e a eficácia no processo.
3. Estratégias exploratórias costumam empreender processos de aprendizado por parte de quem busca, sendo que quanto maior a aquisição de conhecimento sobre o assunto durante o processo, mais eficazes torna-se-ão as consultas dos usuários.

4. A utilização de artefatos diversos, como planilhas, atas de reunião e outros documentos auxiliam o usuário a planejar suas buscas de forma sistemática. Os resultados indicam que o uso desses artefatos colabora no processo de busca, especialmente quando o usuário não possui uma maior familiaridade no assunto.

## 3.4 Conclusão

Baseado nos resultados apresentados, conclui-se que a Web e seus mecanismos de busca são mais eficazes à medida que cresce a familiaridade do usuário com o domínio de conhecimento relacionado às suas buscas. Por um lado, a busca por palavras-chave é uma eficiente ferramenta, pois permite a recuperação quase que imediata de documentos. Por outro, os sistemas de busca da atualidade não proveem ferramentas que permitam o planejamento de tarefas de busca com caráter exploratório, deixando ao usuário o papel de organizar, sintetizar e escolher as melhores estratégias de busca que o auxiliem a cumprir com seus objetivos.



# Capítulo 4

## O processo de busca de novatos em Linux

### 4.1 Introdução

Os capítulos anteriores revelam os sistemas de busca da Web como ferramentas essenciais para procura de informação (Cap. 2). Entretanto, conforme foi mostrado através dos resultados do estudo empírico apresentado no Cap. 3, o modelo de interação consulta-resposta utilizado atualmente para procura de informação pode apresentar dificuldades aos usuários, especialmente quando eles possuem pouca familiaridade com o assunto buscado. Enquanto que no estudo realizado no Cap. 3, usuários foram monitorados através de um *logger* instalado em seus computadores, este capítulo apresenta um estudo realizado em laboratório. A mudança de método permitiu que observássemos o processo de busca *EF*- (Seção 3.2.1) mais diretamente, o que nos auxiliou a compreender as principais dificuldades enfrentadas pelos usuários durante o processo e *in situ*.

Para este efeito, escolhemos a população de usuários novatos no sistema operacional Linux. Participantes foram convidados a utilizarem o sistema de busca Google<sup>1</sup> para encontrar informações que os auxiliassem a resolver problemas no ambiente operacional Linux. Observamos que a falta de familiaridade desses usuários com o sistema tem um impacto significativo na forma como eles interagem com o sistema de busca e em como eles analisam uma grande coleção de documentos da Web para comparar, agregar e sintetizar informações. A falta de familiaridade com o vocabulário peculiar ao assunto buscado pode levar à imprecisão, resultados ambíguos, delegando aos usuários o trabalho de analisar extensivamente os resultados fornecidos pelo sistema de busca em resposta a uma consulta.

Em síntese, a análise quantitativa dos dados coletados nos experimentos mostra que

---

<sup>1</sup><http://www.google.com>

usuários gastam um tempo significativamente maior fazendo a triagem dos resultados retornados pelo sistema de busca e avaliando o conteúdo destes resultados, do que propriamente formulando suas consultas. Uma análise qualitativa dos dados revela que a falta de familiaridade com o assunto pode levar a dificuldades para determinar a relevância de cada resultado retornado pelo sistema de busca. Assim, incapazes de escolher qual resultado da lista pode ser mais útil para completar a tarefa do experimento, usuários tendem a analisar o conteúdo de cada página retornada separadamente, tornando o processo de busca penoso e frustrante sob a perspectiva do usuário. Além disso, usuários costumam preferir fontes de informação com as quais eles possam estabelecer um certo grau de confiança para com o conteúdo oferecido. Por fim, os resultados mostram que usuários com pouca familiaridade com o assunto buscado preferem analisar fontes que contenham toda a informação necessária para completar as tarefas experimentais. Este capítulo descreve o estudo realizado na Seção 4.2, seguindo com a apresentação dos resultados na Seção 4.3. O capítulo termina com uma discussão (Seção 4.4) e uma conclusão (Seção 4.5).

## 4.2 Estudo

### 4.2.1 Participantes

Seis estudantes de pós-graduação do Departamento de Informática da Universidade da Califórnia, Irvine, participaram em experimentos no laboratório de usabilidade da mesma universidade. A experiência de cada voluntário foi aferida qualitativamente em entrevista pré-teste, na qual cada um foi perguntado se havia ou não utilizado o sistema no passado, se continuava utilizando-o e para quais tarefas. Participantes foram selecionados de acordo com o grau de experiência que tinham na utilização do sistema operacional Linux. Cada voluntário era elegível a participar no estudo se possuísse pouca experiência com o sistema. Nenhum dos participantes selecionados utilizava Linux como seu sistema operacional principal. A maioria dos participantes havia utilizado o sistema apenas durante a graduação para realizar tarefas de disciplinas curriculares.

### 4.2.2 Procedimento

A Fig. 4.1 mostra o ambiente utilizado no experimento. Sessões individuais foram realizadas e a atividade de cada participante foi gravada usando um *software* para captura de tela e gravação do áudio ambiente. A gravação de áudio e o protocolo *think-aloud* (TAP) contribuíram para a avaliação de etapas importantes no que tange ao raciocínio por trás das ações dos participantes. Cada participante também tinha a sua disposição no seu computador uma versão da distribuição Ubuntu Linux em execução.

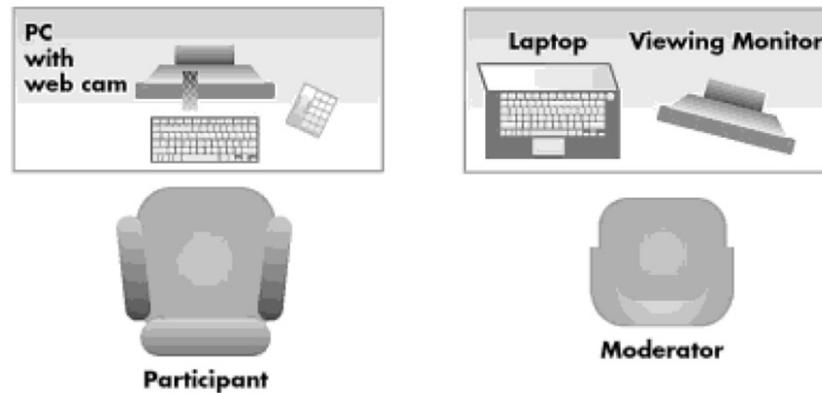


Figura 4.1: Ambiente utilizado para o experimento. Participante e moderador situam-se lado a lado em uma sala, cada um utilizando um computador com o aplicativo *Morae* instalado e em funcionamento. O moderador pode observar as ações do participante através do modo observador do aplicativo e por observação direta, bem como tomar notas em papel e através do aplicativo. Ao participante é mostrada a instrução da tarefa, já pré-programada no aplicativo.

Durante cada sessão, os participantes foram convidados a completar duas tarefas que representavam problemas técnicos comuns enfrentados por usuários dos sistemas Linux. Os enunciados de cada tarefa são mostrados abaixo:

**Tarefa 1** *Suponha que você tenha um MacBook Pro (processador Intel Core 2 Duo, 2 GB RAM) e queira instalar uma distribuição Linux. Suponha também que você não terá acesso à Internet durante a instalação, então você precisa salvar as instruções e levá-las com você onde o seu Macintosh está. Você precisa também escolher qual distribuição Linux será instalada. Indique ao moderador quando você encontrar um conjunto de informações que julga adequado para completar a tarefa. Você deve usar o sistema de busca Google através do navegador disponível na máquina deste teste para obter informações. Não é necessário realizar a instalação, no entanto. Você tem um máximo de 10 minutos encontrar todas as informações que você julga necessárias para instalar uma distribuição Linux em seu Macintosh.*

**Tarefa 2** *Suponha que você deseja encontrar o número de ocorrências de uma determinada palavra dentro do conteúdo de arquivos de texto localizados em uma pasta específica na hierarquia de arquivos da máquina utilizada por você neste teste. Como existem milhares de arquivos, você precisa de alguma ferramenta para realizar a tarefa. Utilize o sistema de busca Google para obter instruções sobre como usar ferramentas de linha de comando para executar a tarefa. Você deve utilizar o terminal aberto na sua máquina para completar a tarefa. Não é necessário instalar qualquer*

*ferramenta para realizar essa tarefa e todos os softwares que você precisa já estão disponíveis. Você tem um máximo de 10 minutos para executar esta tarefa.*

Cada participante teve a sua disposição um navegador e foi convidado a utilizar o sistema de busca Google para procurar por informações. Durante cada sessão, um moderador esteve presente na mesma sala, na qualidade de observador, fazendo perguntas ocasionais ao participante, quando uma explicação verbal ajudaria a esclarecer pontos de interesse. Cada sessão incluiu também um tempo pré-teste dedicado a explicações sobre o ambiente para o participante, apresentação de detalhes sobre o estudo e coleta de assinaturas em termos de consentimento<sup>2</sup>. Havia também, no final de cada sessão individual, um tempo dedicado a revisão das dificuldades encontradas.

## 4.3 Resultados

### 4.3.1 Atividades de procura de informação

Foram identificadas três atividades principais durante a observação dos participantes ao executarem tarefas de busca de informação técnica, particularmente no domínio de conhecimento dos sistemas Linux, no qual os sujeitos possuíam pouca familiaridade. São elas: (1) *formulação de consultas*, que consistia na atividade de elaborar e teclar palavras-chave que servem como consulta ao sistema de busca; (2) *triagem* (ou “*screening*”), que consistia na atividade de analisar a lista de resultados retornada pelo sistema e realizar julgamentos de relevância sobre essa lista e (3) *análise de conteúdo*, que representa a atividade de examinar um ou mais documentos escolhidos pelo usuário baseado na informação contida em cada um deles. A Fig. 4.2 mostra a porcentagem do tempo gasto em cada atividade, por participante. No caso da segunda tarefa apresentada anteriormente, foi contabilizado também o tempo que cada participante passou verificando (“*verification*”) as informações encontradas na Web, e.g. rodando comandos usando o terminal do sistema Linux para completar a tarefa.

### 4.3.2 Falta de experiência e a incerteza

Em nosso estudo, observamos que os usuários novatos geralmente têm dificuldade em determinar a qualidade de uma fonte de informação, e.g. utilidade e abrangência, apenas pela triagem sobre a lista de resultados fornecida pelo sistema de busca. Esta página mostra tipicamente uma página de resultados na forma de uma lista ordenada,

---

<sup>2</sup>Protocolo número 2009-7195 aprovado pelo Comitê de Revisão Institucional (IRB) da Universidade da Califórnia, Irvine.

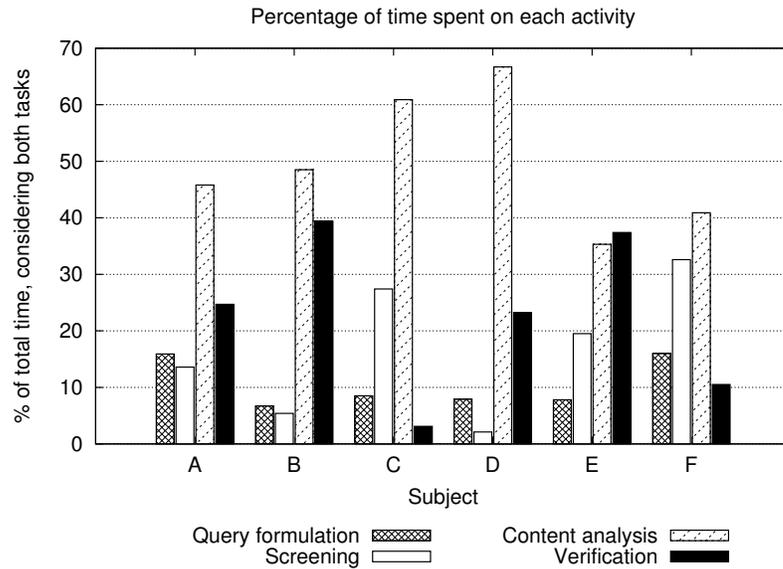


Figura 4.2: Percentagem do tempo total gasto em cada atividade, considerando as duas tarefas do estudo.

cada resultado representado por um título e um breve sumário que destaca os termos correspondentes à consulta realizada pelo usuário. Porque os participantes novatos não conseguiam determinar com antecedência se um resultado era útil para a tarefa em mãos, eles muitas vezes tinham que clicar em cada resultado, carregar a página referida, e analisar o conteúdo da página, o que nossos experimentos revelaram ser a etapa mais demorada do processo de busca (Fig. ??). Por exemplo, durante a triagem sobre a página de resultados contendo uma série de *links* para discussões em fóruns, um participante disse:

*“There are a lot of links. Now, whether the information in those links is meaningful at all is just a matter of reading one by one.”* [P03]

Portanto, um resultado significante é que a falta de familiaridade com o assunto leva a maiores níveis de incerteza com relação a quais *links* devem ser visitados e analisados a partir da página de resultados retornada pelo sistema de busca. Em muitos casos em que os títulos e sumários não ajudam, uma estratégia comum foi a de analisar cada resultado separadamente, melhorando assim a compreensão sobre o domínio da tarefa. Depois de aprender algo útil, como uma nova palavra-chave ou o nome de um comando, os participantes muitas vezes reformulavam a consulta, o que possibilitava que um conjunto de resultados mais refinados pudessem ser obtidos.

### 4.3.3 Informação e suas fontes preferenciais

Byström [Bystrøm, 2002] aponta que existe uma relação de preferência relativamente forte entre os tipos de informações e de tipos de fonte de informação. Do modo similar, percebemos este efeito durante a execução da primeira tarefa pelos participantes. Antes de encontrar instruções sobre como instalar uma distribuição Linux, o participante era convidado a encontrar qual distribuição poderia satisfazer suas necessidades. Por exemplo, após várias tentativas infrutíferas para encontrar uma fonte de informação confiável, um participante encontrou um artigo e mencionou “*é isso que eu estou procurando*”, logo que a página terminara de carregar. Antes de analisar o conteúdo em detalhes, o participante foi perguntado da razão de tal comentário. Ele elogiou como a informação era apresentada e estruturada, com um sumário de conteúdo contendo *links* para as diferentes seções da página. Os usuários novatos em Linux mostraram uma maior preferência por documentos bem estruturados e produzidos por especialistas na forma de artigos, em detrimento a documentos na forma de discussões em fóruns. Como um exemplo, um participante criticou a forma como as informações são apresentadas em fóruns na web:

*“I actually randomly look at the tech forums. Because it is like the same format, the titles... ah! That is one of the things that I would say pro, in general, commercial software... in the open-source [community] more people comment to help you out with the problem, but it is so time-consuming to look for what you need”.* [P05]

## 4.4 Discussão

Os resultados deste capítulo mostram que ao enfrentar dificuldades no processo de busca, usuários tendem a gastar mais tempo fazendo a triagem dos resultados retornados pelo sistema de busca e analisando o conteúdo destes resultados. Neste contexto, durante as buscas *EF-*, usuários novatos em Linux mostraram fazer uso de estratégias de orientação [Teevan et al., 2004], navegando passo a passo pelo espaço informacional e ao mesmo tempo angariando um maior conhecimento no assunto buscado. Este comportamento de orientação também fora identificado entre os pesquisadores acadêmicos, conforme os resultados apresentados na Seção 2.3.2.

Entretanto, os estudos anteriores haviam revelado apenas uma parte dos problemas típicos enfrentados por usuários com pouca familiaridade. Através de uma observação direta do processo de busca desses usuários, o estudo apresentado neste capítulo possibilitou identificar que a etapa de formulação de consultas é apenas uma das dificuldades enfrentadas pelos usuários. De acordo com a análise quantitativa dos dados coletados em laboratório, as etapas de triagem e análise de resultados mostraram ser penosas aos

usuários. Através da análise qualitativa dos dados, descobrimos que a forma de apresentação do conteúdo buscado e a própria estrutura hipertextual da Web, são fontes de desorientação e sobrecarga cognitiva para os usuários com pouca familiaridade, o que já fora reconhecido em trabalho anterior [Conklin, 1987].

Em nosso estudo, ao comparar as informações encontradas em serviços de pergunta-resposta (*question-answering* ou QA) com as encontradas em fóruns de discussões, participantes revelaram uma maior preferência pelo primeiro tipo de fonte de informação. Acreditamos que as razões para este comportamento são duas. Primeiro, em serviços de QA, o conteúdo melhor classificado pela comunidade é apresentado antecipadamente, isentando o usuário de analisar inúmeras páginas com informações. Segundo, usuários novatos depositaram uma maior credibilidade na informação que possui o aval de especialistas.

Baseado na análise qualitativa, foi possível encontrar, portanto, duas relações entre condição e consequência. Primeiro, conteúdo bem estruturado e informações detalhadas contam para uma maior *completude* da fonte de informação. Como consequência, usuários novatos mostraram uma maior preferência por fontes que continham toda a informação necessária para completar suas tarefas. Segundo, níveis mais elevados de *expertise*, quando associados à fonte de informação e ao seu conteúdo, contam para uma maior *credibilidade* por parte do usuário com relação à informação encontrada. Em consequência, diante de múltiplas fontes potencialmente relevantes, usuários novatos preferiram fontes em cuja informação transmitia maior credibilidade.

## 4.5 Conclusão

Este capítulo apresenta os resultados de um estudo realizado em laboratório com usuários novatos em Linux. Embora os resultados tenham sido obtidos a partir de uma pequena amostragem, a análise qualitativa dos dados coletados aponta na direção de que a falta de familiaridade tem efeitos potenciais sobre o comportamento dos usuários ao utilizarem sistemas de busca baseados no modelo consulta-resposta. Grande parte do trabalho de triagem é delegado ao usuário, que não possui familiaridade o suficiente para desempenhar essa etapa com eficiência e eficácia. Os resultados sugerem que os sistemas de busca devem oferecer maneiras alternativas para os usuários realizarem essa triagem. Uma possível alternativa seria implementar mecanismos que permitissem a triagem colaborativa de conteúdo, tendo em mente fatores como a credibilidade e a completude das fontes de informação selecionadas.



# Capítulo 5

## O papel da familiaridade na exploração do conhecimento

### 5.1 Introdução

*“...in an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it”*. ([Simon, 1971], p. 40-41)

O problema caracterizado por Simon pode ser facilmente encontrado nos sistemas de busca atualmente disponíveis na Web. A busca do Google<sup>1</sup>, por exemplo, retorna ao usuário um conjunto de documentos da Web em resposta a uma consulta por palavras-chave. Embora os algoritmos de ranqueamento tenham evoluído de forma significativa nos últimos anos, a análise dos resultados retornados é, ainda, uma tarefa delegada ao usuário que realiza a consulta. Mais do que isso, os sistemas de busca atuais não provêem mecanismos de organização, de modo que usuários possam gerenciar a informação já analisada de acordo, por exemplo, com uma dada tarefa ou projeto.

Desta forma, por vezes não basta conhecer um conjunto de palavras-chave apropriado para obter a informação desejada: em tarefas de busca complexas, o sistema de busca delega ao usuário o trabalho de triagem e a análise de múltiplos documentos, o que pode demandar múltiplas sessões de busca, i.e. tarefas que duram dias ou semanas para serem completadas, e.g. revisões bibliográficas, aprendizado e síntese de conhecimento não familiar ao usuário, dentre outros tipos de atividades.

---

<sup>1</sup><http://www.google.com>

Este capítulo baseia-se nos resultados empíricos dos capítulos anteriores para o desenvolvimento de uma teoria sobre a influência da familiaridade na exploração e no gerenciamento de conhecimento. Primeiramente, um arcabouço teórico é introduzido para ampliar a compreensão da Web como um sistema sociotécnico de produção, consumo e gerenciamento de informação. Depois, este capítulo baseia-se nos resultados empíricos dos capítulos anteriores para desenvolver a teoria e identificar as principais condições, consequências e propriedades que estão relacionadas às variações de familiaridade dos usuários com o domínio de conhecimento explorado.

## 5.2 Busca e seus diversos níveis de atividade

Diversos modelos no campo da ciência da informação procuraram modelar o comportamento humano em atividades de busca. A Fig. 5.1 mostra o modelo apresentado por Järvelin e Ingwersen [Järvelin and Ingwersen, 2004], que contempla a atividade de busca por informação em diferentes níveis ou contextos (do mais amplo ao mais específico): o sócio-organizacional e cultural, o da tarefa de trabalho, o da procura (*seeking*) e o da recuperação (*retrieval*) de informação. No modelo apresentado, a busca pode ser modelada como uma hierarquia de objetivos e tarefas, de modo que cada tarefa provê objetivos que servem como contexto para tarefas subsidiárias.

### 5.2.1 O modelo consulta-resposta e a recuperação de informação

A recuperação de informação (IR) encontra-se no nível de menor granularidade da hierarquia e, dessa forma, prevê tarefas que podem ser realizadas a partir de uma simples consulta a um banco de dados de informações ou, no caso da Web, a um sistema de busca, e.g. Google. Consultas, nesse caso, são geralmente compostas por palavras-chaves, de forma que uma tarefa típica de IR envolve encontrar documentos cujos termos correspondem às palavras-chave apresentadas pela consulta. O elemento comum à todas tarefas de IR é o ciclo que consiste na (i) consulta por palavras-chave e (ii) recebimento de uma resposta do sistema de busca na forma de uma lista de resultados.

A imensa maioria dos sistemas de busca, e.g. Google, provêem os usuários com mecanismos que suportam o ciclo de IR. Tal ciclo é exercido através de um modelo de interação ficou conhecido como paradigma “consulta-resposta” [White and Roth, 2009] (Fig. 5.2). O usuário pode ter a necessidade de exercitar este ciclo por diversas vezes, de modo a cumprir tarefas de mais alto nível, como a tarefa de procura de informação.

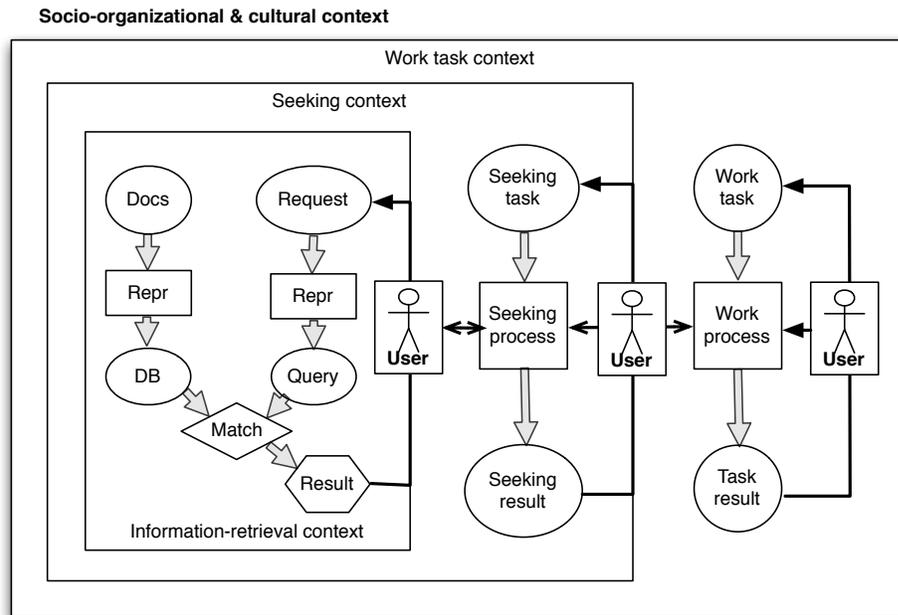


Figura 5.1: Modelo para as atividades de busca segundo Järvelin e Ingwersen [Järvelin and Ingwersen, 2004].

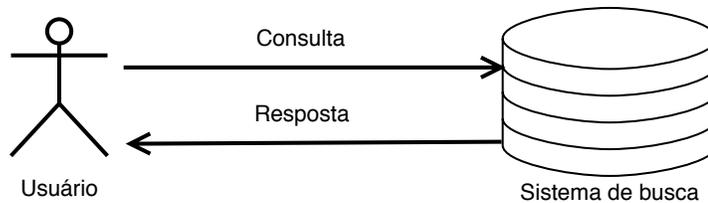


Figura 5.2: Ilustração do modelo de interação consulta-resposta.

### 5.2.2 A procura de informação

O objetivo de uma tarefa de procura de informação (IS), por sua vez, é o de satisfazer uma necessidade informacional ou problema [Marchionini, 1997]. Tarefas de IS englobam uma ou mais tarefas de IR (*seeking context* na Fig. 5.1). No nível de IS, usuários fazem decisões estratégicas sobre como, onde e, ainda, se devem ou não encontrar a informação relacionada às suas necessidades [Wilson et al., 2010]. Eles podem adotar também uma estratégia de navegação através da qual, por exemplo, eles começam num ponto conhecido do espaço informacional (possivelmente os resultados de uma consulta de IR) e navegam por uma estrutura de *links* para localizar outros documentos relacionados. Os documentos encontrados durante a navegação ou através de consultas de IR são exa-

minados, analisados e sintetizados de modo a satisfazer uma necessidade informacional particular. A estrutura do hipertexto provida pela Web oferece suporte às estratégias de navegação: enquanto navegam, cada documento oferece informações que podem suplantam novas idéias, sugerir novas direções e por vezes modificar a natureza da necessidade informacional [Bates, 1993]. Em outras palavras, as decisões realizadas pelo usuário que busca são guiadas pela avaliação, passo-a-passo, do que é mais provável que vá satisfazer suas necessidades no processo de forrageamento pela informação [Pirulli and Card, 1995].

### 5.2.3 A busca no contexto de trabalho

O processo de IS é motivado e iniciado dentro de um contexto ainda mais amplo, i.e. o de trabalho [Byström and Hansen, 2005] (*work task context* na Fig. 5.1). Tarefas de trabalho são situadas no contexto de uma organização [Wilson et al., 2010] (Fig. 5.1) ou, ainda, de uma comunidade de prática [Wenger, 1999]. As decisões do usuário, nesse nível, são influenciadas pela cultura organizacional, normas sociais, bem como pelos recursos informacionais disponíveis (e.g. bibliotecas *online*) e por outros agentes humanos, e.g. colegas de trabalho ou outros membros numa comunidade de prática.

### 5.2.4 Busca exploratória

A atividade de busca exploratória pode ser entendida como um tipo de atividade de procura de informação (IS). Entretanto, atividades de IS *típicas* fazem parte da rotina de trabalho dos usuários que as exercem, enquanto atividades de IS de caráter *exploratório* não fazem. Com a finalidade de melhor compreender essa distinção, Belkin [Belkin, 1982] define o conceito de estados anômalos de conhecimento (ou ASK):

*“The ASK hypothesis is that an information need arises from a recognized anomaly in the user’s state of knowledge concerning some topic or situation and that, in general, the user is unable to specify precisely what is needed to resolve that anomaly”*. ([Belkin, 1982], p. 2)

Em ambos os casos de atividade de procura de informação, ASKs são um fator motivacional: o usuário tem uma necessidade informacional que é percebida por ele, usando as palavras de Belkin, através do reconhecimento de uma anomalia no estado de conhecimento do próprio usuário com relação a algum tópico ou situação. Schütz e Luckmann [Schütz and Luckmann, 1973] definem tais situações como *rotineiras* ou *problemáticas*. Em uma situação rotineira, os elementos necessários para resolver um ASK são conhecidos. A estratégia de busca é baseada nas habilidades e no conhecimento prévio do usuário, de modo que cada passo em direção ao cumprimento das suas necessidades informacionais é conhecido *a priori*. Situações problemáticas, por outro lado, emergem quando o

conhecimento prévio do usuário o auxilia apenas parcialmente nas tomadas de decisão associadas à estratégia que pode ser seguida na sua finalidade de satisfazer suas necessidades informacionais. Nesses casos, atividades de procura de informação demandam estratégias de caráter exploratório:

*“Exploratory strategies are often motivated by a poor understanding of terminology and information space structure, and a desire to learn”.*  
([White and Roth, 2009])

Portanto, a busca exploratória pode ser definida por situações em que pessoas: (i) não são familiares com o domínio do conhecimento associado às suas necessidades informacionais; (ii) precisam primeiramente aprender sobre um tópico para depois formular suas estratégias de busca ou, ainda, (iii) não estão certas sobre seus objetivos de busca [White and Roth, 2009].

## 5.3 Uma visão sociotécnica sobre a busca na Web

### 5.3.1 Comunidades de prática e sistemas de informação

O conceito de *comunidades de prática* (CoP) tem suas origens nas teorias sobre o aprendizado e a cognição desenvolvidas por cientistas sociais no final da década de 80. Lave [Lave, 1988] faz uma análise sobre uma forma particular de atividade cognitiva — a resolução de problemas aritméticos. Sua análise entende a cognição humana como um processo amplo, que atravessa os limites individuais do corpo e da mente, trazendo para o plano frontal da análise empírica os aspectos sociais, ecológicos e históricos que interagem com o indivíduo *in situ* — na situação em que a atividade cognitiva ocorre. Deste modo, Lave apresenta uma crítica aos modelos cognitivos tradicionais:

*“I have argued that cognitive experimental research relies on a priori normative models as source and inspiration for the development of experimental tasks and the interpretation of activity in experiments. (...) Cognitive processes, viewed in this light, become obvious candidates for reexamination as culturally constituted phenomena”* ([Lave, 1988], p. 92).

Esta observação estimulou uma mudança de caráter epistemológico no estudo da cognição humana: teóricos situacionais sugerem que a aquisição de conhecimento é uma prática *social*, e não *individual*. Desta forma, *atividades* não podem ser entendidas separadamente do *contexto* em que estão *situadas* [Suchman, 1987]. Em outras palavras, o aprendizado individual não pode ser compreendido sem levar em conta as práticas sociais, bem como alguns fatores que antes — e de acordo com a tradição cognitivista —

eram considerados externos ou mutuamente exclusivos ao indivíduo, como a cultura e a linguagem.

Baseado nessa perspectiva, não se pode entender o conceito de CoP, sem antes tomar o conceito de *participação legítima periférica*:

*“Legitimate peripheral participation provides a way to speak about relations between newcomers and old-timers, and about activities, identities, artifacts, and communities of knowledge and practice. It concerns the process by which newcomers become part of a community of practice”.*

([Lave and Wenger, 1991], p. 29)

CoPs, por definição, não são necessariamente submetidas a um regime rígido organizacional, nos quais seus membros possuem posições ou papéis definidos *a priori*. Ao contrário, elas emergem da participação de indivíduos que compartilham de uma prática e aprendem à medida que interagem entre si [Wenger, 1999]. Novatos interagem periféricamente, adquirindo conhecimento através dos mais experientes e eventualmente vindo a participar no núcleo da comunidade de prática (Fig. 5.3).

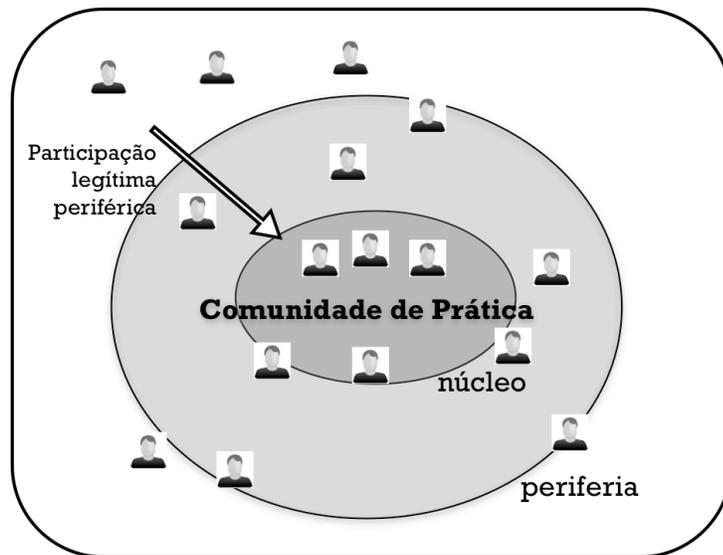


Figura 5.3: Ilustração do conceito de CoP. Novatos interagem periféricamente com indivíduos mais experientes situados no núcleo da comunidade.

A legitimidade da participação dá-se pelo interesse comum — em diferentes níveis — dos indivíduos com relação a um domínio do conhecimento particular. Um bom exemplo de comunidade de prática pode ser encontrado nos usuários dos fóruns de discussão da Web (Fig. 5.4). Usuários mais experientes auxiliam novatos através do sistema, respondendo a perguntas ou dúvidas referentes a um tema compartilhado pelos participantes da comunidade.

## Adding ttf fonts in Ubuntu

new thread post reply OOoForum.org Forum Index -> OpenOffice.org Impress

View previous topic :: View next topic

Author	Message
<b>flipz</b> Newbie 000000 Joined: 01 Oct 2008 Posts: 2	Posted: Wed Oct 01, 2008 11:42 am Post subject: Adding ttf fonts in Ubuntu <span style="float:right">quote</span> I recently installed Ubuntu 8.04 and am using OpenOffice 2.4 which comes with it. I have 2 ttf fonts that I need to install for use with a presentation (impress). I have tried 2 different methods. First was what I found on the forum <a href="#">here</a> The "typing fonts:///" in to nautilus doesn't work, I just get an error basically saying that it can't find that directory. I also tried the one that has a bunch of steps. That completed with no errors however the fonts did not show up in OpenOffice. I then checked OpenOffice's help file and it says you can add fonts using the /spadmin program in their directory. That let me add the fonts and they show up in the OpenOffice font list now, however they aren't the right fonts. They both show up identical and are just some generic font. They aren't the right ones. Please help! Thanks. Back to top <span style="float:right">profile PM</span>
<b>Robert Tucker</b> Moderator 000000 Joined: 16 Aug 2004 Posts: 2750 Location: Manchester UK	Posted: Wed Oct 01, 2008 11:48 am Post subject: <span style="float:right">quote</span> Copying the .ttfs to (a folder in) /usr/share/fonts/truetype usually works on Linux systems. Have you checked they are there. LibreOffice 3.3.2 on Ubuntu 10.10 / OOo 3.3.0 (Fedora version) on Fedora 14 (VirtualBox) Back to top <span style="float:right">profile PM</span>
<b>flipz</b> Newbie 000000 Joined: 01 Oct 2008 Posts: 2	Posted: Wed Oct 01, 2008 11:51 am Post subject: <span style="float:right">quote</span> Yup, they are definitely there. And I did <pre>Code: sudo chown root.root *.ttf sudo mkfontdir cd .. fc-cache</pre> On that directory which it says needs to be done to make the fonts available to other applications and that didn't work. I also tried putting them in /usr/share/fonts/truetype/openoffice/ and still no luck. They are showing up in the font list in Open Office but aren't functioning. They aren't the right font. I can also say that when I copy them to my desktop the icon for them shows up fine, and I can open them in font editor fine. When I copy them to any font directory, they show as a different icon from all other ttfs. It's a lowercase a, and has an orange and white X in the top right corner? Back to top <span style="float:right">profile PM</span>

Figura 5.4: Exemplo de uma discussão em um fórum da Web.

Entretanto, enquanto os usuários e suas atividades constituem CoPs, a tecnologia utilizada para interação entre esses usuários e para mediação da informação por eles produzida vai além do conceito de CoP. Ao incluir a tecnologia, devemos primeiramente definir esse tipo de sistema sociotécnico como um sistema de informação. Podemos afirmar que as atividades de produção de informação e a interação entre agentes humanos numa CoP, quando mediadas pela tecnologia, constituem, por definição, um *sistema de informação* (SI). A definição a seguir ilustra bem a natureza sociotécnica desses sistemas:

*“An information system is fundamentally concerned with communication in support of human activity using artifacts to represent, store, manipulate and transmit data. The essence of an information system therefore lies not purely in the technology or in the activity: it lies in the way in which technology is used in support of purposeful action. (...) Information systems constitute*

*communication systems designed to support human activity with the aid of technology*". ([Beynon-Davies, 2009], p. 10, col. 2)

Alguns aspectos desses sistemas devem ser enfatizados. Primeiro, SIs embutem formas de representação para os artefatos gerados pela atividade humana, em particular, para informação produzida pelos agentes humanos participantes do sistema. Segundo, atividades em um SI têm um propósito compartilhado. Em uma CoP, o propósito é o aprendizado e a troca de informações entre seus membros. Terceiro, SIs oferecem mecanismos para manipulação dos artefatos produzidos no sistema, e.g. buscadores, filtros e classificadores de informação. Quarto, e mais importante, formas de representação e mecanismos para manipulação são utilizados para transmissão de artefatos entre os agentes humanos participantes de um SI. Por isso, SIs são também considerados *sistemas de comunicação*.

Porém, a comunicação no interior dos SIs abrange necessariamente diferentes contextos, o que confere *heterogeneidade* para informação transmitida. Bowker e Star [Bowker and Star, 1999] caracterizam a transmissão de informação sob essa perspectiva:

*"Even seemingly simple, replication and transmission of information from one place to another involves encoding and decoding as time and place shift. Thus the context of information shifts in spite of its continuities; and this shift in context imparts heterogeneity to the information itself."*. ([Bowker and Star, 1999], p. 290)

Segundo Bowker e Star, a representação padronizada da informação apresenta-se como uma forma de regularizar, em parte, o movimento da informação de um contexto ao outro, isto é, de prover meios de acesso à informação que permeiam as diferenças espaço-temporais impostas por contextos distintos:

*"If both people and information objects inhabit multiple contexts and a central goal of information systems is to transmit information across contexts, then [an standardized] representation is a kind of pathway that includes (...) people, things-objects, previous representations, and information about its own structure"*. ([Bowker and Star, 1999], p. 293)

### 5.3.2 Espaços informacionais comuns

Os conceitos de CoP e SIs são importantes para definição do conceito de espaços informacionais comuns (EICs). Segundo Schmidt e Bannon [Schmidt and Bannon, 1992], o compartilhamento e a comunicação de informação não são, entretanto, condições suficientes para construção de EICs. Os autores utilizam o exemplo de uma base de dados compartilhada entre diversos usuários, i.e. um sistema de informação, para argumentar

que a importância de um EIC dá-se no compartilhamento de um entendimento comum com relação a informação transmitida de um agente humano a outro, através do sistema de informação:

*“A common database is not a common information space. Objects in a database are perceived and manipulated at different semantic levels. They can be manipulated qua objects, but are more usually perceived and manipulated as carriers of representations. Their importance lies in the interpretation human actors place on the meaning of the representational object. The distinction between the material carrier of information - the object - and its meaning is crucial. The material representation of information in the common space (e.g., a letter, memo, drawing, file) exists as an objective phenomenon and can be manipulated as an artifact. The semantics of the information carried by the artifact, however, is, put crudely, ‘in the mind’ of the beholder, and the acquisition of information conveyed by the artifacts requires an interpretive activity on the part of the recipient. Thus, a common information space encompasses the artifacts that are accessible to a cooperative ensemble as well as the meaning attributed to these artifacts by the actors. [...] Now, what happens if the information object accessed by one actor is produced by another and vice versa, that is, if the set of information objects are produced and accessed by multiple actors? At the level of the objects themselves, shareability may not be a problem, but in terms of their interpretation, the actors must attempt to jointly construct a common information space which goes beyond their individual personal information spaces”. ([Schmidt and Bannon, 1992], p. 20)*

A exemplo dos fóruns de discussão, EICs demandam o *trabalho de interpretação* dos artefatos neles produzidos. O trabalho de interpretação pode ser realizado face-a-face, mas frequentemente requer o uso de outros meios de comunicação, dado que os agentes humanos que se comunicam no sistema de informação podem estar distribuídos no tempo e no espaço. Deste modo, mecanismos de interação suportados por computadores são aplicados para permitir da comunicação desses agentes [Bossen, 2002]. É novamente o caso dos fóruns de discussão na Web: o mecanismo de interação que permite a publicação de perguntas e respostas funciona como um meio para cooperação assíncrona entre usuários na produção compartilhada de informação e tem por objetivo permitir a comunicação entre os múltiplos indivíduos envolvidos.

Entretanto, a base de dados compartilhada contendo as postagens desses usuários não constitui, por si só, um EIC. Além disso, ainda que tais usuários do fórum de discussão sejam membros de uma CoP, esta condição não é suficiente para que a CoP *per se* cons-

titua um EIC. A informação armazenada e compartilhada em um fórum de discussão, por exemplo, é percebida e manipulada em diferentes níveis semânticos e, por isso, a sua importância reside na interpretação realizada pelos seus usuários. De acordo com Schmidt e Bannon [Schmidt and Bannon, 1992], a representação material da informação contida em um espaço comum existe como um fenômeno objetivo e pode ser manipulada como um artefato. A semântica carregada pelo artefato, entretanto, está na mente dos que o manipulam, e a aquisição da informação trazida por ele requer uma atividade interpretativa por parte de quem a recebe. A Fig. 5.5 mostra a relação entre os três conceitos apresentados, a saber, SI, CoP e EIC.

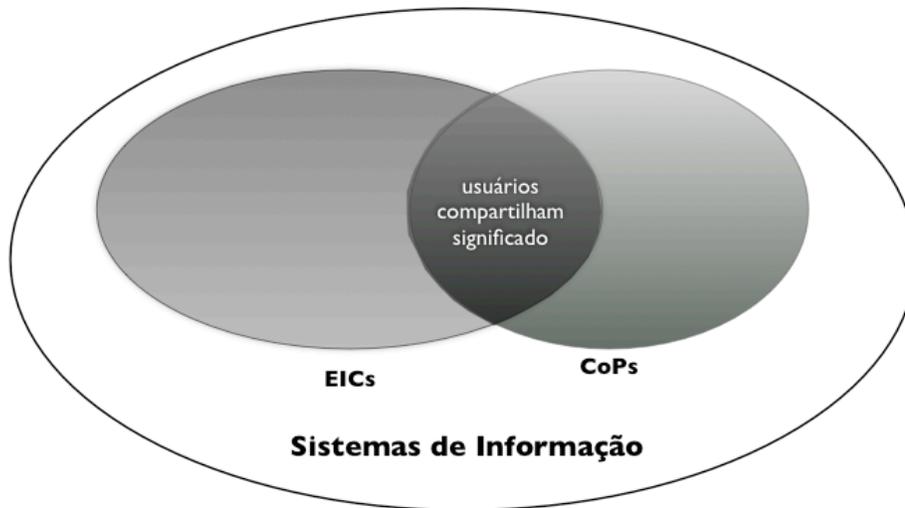


Figura 5.5: Relação entre os conceitos de SI, CoP e EIC. A comunicação entre membros de uma comunidade de prática realizada através de um sistema de informação contribui para construção de um EIC quando os usuários participantes compartilham um significado comum com relação à informação transmitida entre eles.

Dessa maneira, a natureza distribuída do trabalho cooperativo gera dificuldades para construção de EICs. Isto porque a informação que é produzida cooperativamente por um ator é acessada, na maioria das vezes, por outro ator. Ainda segundo Schmidt e Bannon [Schmidt and Bannon, 1992], o compartilhamento entre os atores, no nível da informação em si mesma, não necessariamente constitui um problema. A interpretação que os atores fazem da informação compartilhada, entretanto, é o grande desafio para construção dos EICs, o que ultrapassa os limites dos espaços informacionais de cada ator.

### 5.3.3 Categorias como objetos de fronteira

O trabalho cooperativo, com o objetivo da produção e consumo de informação, dá-se em diferentes contextos e atravessa as bordas de diferentes comunidades [Figueira Filho et al., 2010a]. Esta relação entre produção e consumo entre diferentes CoPs situados em um mesmo SI é ilustrada na Fig. 5.6.



Figura 5.6: Produção e consumo de informação entre diferentes comunidades de prática situadas em um mesmo SI.

Categorias, quando associadas à informação transmitida, funcionam como objetos de fronteira (“*boundary objects*”):

*“[Boundary object] is an analytic concept [...] which both inhabit several intersecting social worlds and satisfy the informational requirements of each of them. Boundary objects [...] are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. [...] They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation. The creation and management of boundary objects is a key process in developing and maintaining coherence across intersecting social worlds”.*  
 ([Star and Griesemer, 1989], p. 393)

O conceito de *social worlds* concebido por Strauss [Strauss, 1978] equivale, segundo Bowker e Star [Bowker and Star, 1999] ao conceito de CoP [Lave and Wenger, 1991]. Ca-

tegorias, desse modo, surgem do trabalho cooperativo em espaços informacionais. Contudo, Bowker e Star alertam para a complexidade desse gerenciamento:

*“The processes by which CoPs manage divergent and conflicting classification systems are complex, the more so as people are all members in fact of many CoPs, with varying levels of commitment and consequence”.* ([Bowker and Star, 1999], p. 293)

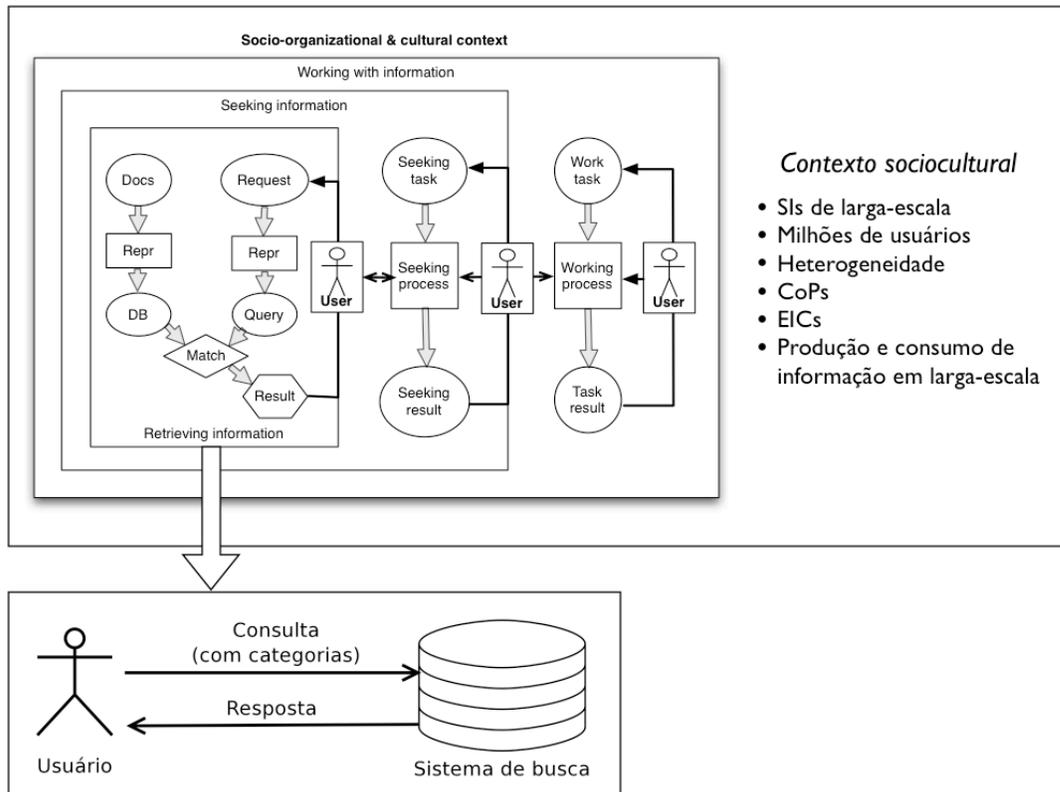


Figura 5.7: Níveis de atividade de busca e modelo de recuperação de informação consultaresposta apresentados na Seção 5.2.1 imersos no contexto sociocultural.

Neste ponto, retornamos ao modelo em níveis de atividade de busca (Seção 5.2) de Järvelin e Ingwersen [Järvelin and Ingwersen, 2004] e também ao modelo consultaresposta apresentado na Seção 5.2.1 (Fig. 5.7). Quando imerso no contexto sociocultural apresentado até aqui, podemos compreender de forma mais ampla o papel dos sistemas de busca que implementam tal modelo em SIs de larga-escala como a Web. A partir da participação em diferentes CoPs simultaneamente, usuários podem possuir diferentes níveis de familiaridade com o espaço informacional e com as categorias atreladas à informação compartilhada. Deste modo, o modelo consulta-resposta passa a ser encarado

em um contexto sociocultural, servindo a grupos heterogêneos de usuários que, por sua vez, produzem consultas na forma de palavras-chaves. Estas palavras-chaves incluem categorias, que nesta seção são apresentadas como objetos de fronteira. Assim, a *relação de familiaridade* que tais usuários possuem com o assunto buscado e o entendimento das categorias relacionadas não podem ser considerados separadamente de um nível mais amplo, sociocultural, conforme mostra a Fig. 5.7. A figura representa, ainda, a interação de acordo com o modelo consulta-resposta, implementado pela maioria dos sistemas de busca atuais, e.g. Google. O suporte destes sistemas existe predominantemente no nível de recuperação de informação (IR) (Seção 5.2.1) e está ausente nos demais níveis.

A ausência de suporte dos sistemas de busca nos níveis superiores ao da recuperação de informação apresenta uma série de problemas. Estes problemas serão analisados no decorrer deste capítulo e finalmente contrastados com os resultados empíricos obtidos nos capítulos anteriores. A seção seguinte faz uma análise filosófica dos sistemas de busca e do modelo consulta-resposta por eles implementado, pontuando o problema do significado das categorias que são atreladas à informação e as limitações inerentes desse modelo.

## 5.4 O modelo consulta-resposta e suas limitações inerentes

*“If you didn’t experience the meaning of the words, then how could you laugh at puns?”* ([Wittgenstein, 1996], p. 711)

Sistemas de busca que implementam o modelo consulta-resposta são *máquinas manipuladoras de símbolos*, i.e. sistemas formais que tomam como entrada símbolos, os combinam em expressões (ou estruturas simbólicas) e os manipulam através de processos para produzir novas expressões. Em meados dos anos 70, Newell e Simon propuseram o que ficou conhecido como a *hipótese dos sistemas simbólicos físicos*:

*“A physical symbol system has the necessary and sufficient means for general intelligent action”.* ([Newell and Simon, 1976], p. 4, col. 2)

Por *“general intelligent action”*, os autores querem indicar algo no mesmo escopo de inteligência observado nas ações humanas, i.e. *“any real situation behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity”*. Newell e Simon deixam claro, entretanto, a natureza da hipótese colocada: *“This is an empirical hypothesis (...) Intelligent behavior is not so easy to produce that any system will exhibit it willynilly. Indeed, there are people whose analyses lead them to conclude either on philosophical or on scientific grounds that*

*the hypothesis is false. Scientifically, one can attack or defend it only by bringing forth empirical evidence about the natural world*". ([Newell and Simon, 1976], p. 4, col. 2).

Dreyfus em [Dreyfus, 1992] sumariza os pressupostos históricos da hipótese de Newell e Simon, que se tornou base para o que mais tarde ficaria conhecido como “*Good Old Fashioned Artificial Intelligence*” (GOFAI):

*“GOFAI is based on the Cartesian idea that all understanding consists in forming and using appropriate symbolic representations. For Descartes, these representations were complex descriptions built up out of primitive ideas or elements. Kant added the important idea that all concepts are rules for relating such elements and Frege showed that the rules could be formalized so that they could be manipulated without intuition or interpretation”*. ([Dreyfus, 1992])

A hipótese proposta por Newell e Simon era, no mínimo, tentadora. Se fosse possível construir o sistema formal correto, seres humanos poderiam então descrever o estado geral das coisas (“*states of affairs*”) no mundo de forma que antes seria impossível ou impraticável de observar diretamente. Tais estados gerais das coisas poderiam ser algo no futuro, antes imprevisível, mas que o advento dos sistemas simbólicos formais e das máquinas manipuladoras desses símbolos tornaria factível de prevê-los [Hutchins, 1995].

Neste capítulo, poder-se-ia apresentar diversos outros contra-argumentos aos pressupostos da hipótese de Newell e Simon. Em vez disso, utiliza-se uma analogia que serve a um dos principais objetivos desta revisão da literatura teórica: mostrar a principal limitação dos sistemas de busca que seguem o paradigma *consulta-resposta*, cujas consultas são realizadas interativamente por usuários através da escolha de *palavras-chave*. Essa analogia é ilustrada através da idéia de cegueira de significado.

### 5.4.1 Máquinas de busca e a cegueira de significado

A idéia de cegueira de significado (do inglês, *meaning-blindness*) proposta por Ludwig Wittgenstein em *Philosophical Investigations* ([Wittgenstein, 1960]) está em consonância com as afirmações de críticos como Hubert Dreyfus com relação às limitações dos computadores.

Wittgenstein supõe um ser-humano incapaz de sentir e compreender os diversos significados das coisas através das suas relações com essas mesmas coisas em diferentes situações. Ele inicia com a definição do conceito de cegueira de “aspecto” através de uma série de questionamentos, e introduz o conceito de cegueira de significado:

*“Could there be human beings lacking in the capacity to see something as something—and what would that be like? What sort of consequences would it*

*have?—Would this defect be comparable to colour-blindness or to not having absolute pitch?—We will call it ‘aspect-blindness’* ” ([Wittgenstein, 1960], p. 213). *“Drawings of cubes, triangles and duck-rabbits can be seen in different ways, and someone who is not “aspect-blind” can experience the switch. Similarly, words can have different meanings (...) and someone who can “experience the meaning of a word” also experiences the switch between those meanings, whereas a “meaning-blind” person cannot. The switch escapes him; he cannot make it happen”.* ([Wittgenstein, 1980], p. 166–268)

Segundo Wittgenstein, o problema está também na ‘representação’ do pensamento humano através da forma que temos de nos expressar, seja através da linguagem corporal ou verbal:

*“Let us assume there was a man who always guessed right what I was saying to myself in my thoughts. (It does not matter how he manages it.) But what is the criterion for his guessing right? Well, I am a truthful person and I confess that he has guessed right—But might I not be mistaken, can my memory not deceive me? And might it not always do so when—without lying—I express what I have thought within myself”?* ([Wittgenstein, 1960], p. 226)

O argumento de Wittgenstein é que se torna impossível a expressão dos nossos pensamentos, da forma exata como pensamos, através de qualquer linguagem. A principal característica da interação entre seres-humanos, entretanto, é que somos capazes de interpretar formas de expressão pois vivenciamos o significado dessas expressões. Uma pessoa cega de significado pode conhecer a linguagem utilizada em uma conversação, mas não possui a *sensibilidade* necessária para experimentar o significado da linguagem. As reações e gestos compartilhados pelos seres-humanos durante a comunicação são inseparáveis da linguagem verbal, e carregam consigo diferentes significados, tornando a expressão humana significativa aos diferentes interlocutores.

A analogia proposta nesta seção é a de que computadores, como máquinas manipuladoras de símbolos, sofrem do problema da cegueira de significado. Primeiro porque, embora eles possam representar a linguagem através de símbolos, eles não podem experimentar os diferentes significados da linguagem. Segundo, porque computadores são incapazes de vivenciar o significado da expressão humana, falta a eles a sensibilidade necessária para compreender as diferentes nuances dessa expressão, seja ela verbal ou gestual.

Tomemos como exemplo o “sorriso” representado na Fig. 5.8. Se uma pessoa “sorrisse”, mas tivesse apenas cinco posições para sua face, mudando de uma posição à outra diretamente, então nós poderíamos não responder a essa pessoa da mesma forma que fazemos a um sorriso “verdadeiro”. Em outras palavras, tanto um ser humano quanto

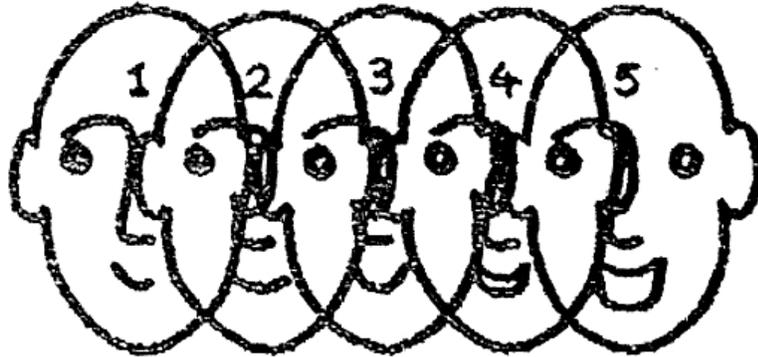


Figura 5.8: Ilustração de um “sorriso” representado simbolicamente (extraído de [Heaton and Groves, 2005])

um computador podem reconhecer que a expressão ilustrada é um sorriso, mas apenas os seres humanos podem sentir o seu significado.

#### 5.4.2 A metáfora da sala chinesa

O modelo de interação consulta-resposta introduzido anteriormente consiste em uma pessoa manipulando símbolos de um lado, i.e. consultas, e um sistema manipulando esses mesmos símbolos do outro lado de modo a gerar respostas. Hutchins em [Hutchins, 1995] apresenta uma crítica a esse modelo de interação:

*“Notice that when symbols are in the environment of the human, and the human is manipulating the symbols, the cognitive properties of the human are not the same as the properties of the system that is made up of the human in interaction with these symbols. The properties of the human in interaction with the symbols produce some kind of computation. But that does not mean that computation is happening inside people’s head”.* ([Hutchins, 1995], p. 361)

A crítica de Hutchins está de acordo com o argumento de Wittgenstein apresentado anteriormente. Em citação ao clássico exemplo de John Searle referente à “sala chinesa”, ilustrado na Fig. 5.9, Hutchins explora o problema:

*“Imagine a room inside of which sits the philosopher Searle. Chinese people come up to the room and push strings of Chinese characters through a slot in the door. Searle slips back other strings of characters, which the Chinese take*



Figura 5.9: Ilustração da “sala chinesa”.

*to be clever responses to their questions. Now, Searle does not understand Chinese. He doesn't know the meaning of any Chinese character. To him, the characters of written Chinese are just a bunch of elaborate squiggles. However, Searle has with him in the room baskets of Chinese characters, and he has a rulebook which says that if he gets certain sequences of characters he should create certain other sequences of characters and slide them out the slot. Searle intends his thought experiment as a demonstration that syntax is not sufficient to produce semantics”. ([Hutchins, 1995], p. 361)*

A sala parece comportar-se como se entendesse Chinês. Entretanto, nem o filósofo, nem nada dentro da sala pode ser considerado como se entendesse o *significado* da linguagem trocada com os chineses. Em outras palavras, a “sala chinesa” é cega de significado. Sob o ponto de vista da cognição distribuída de Hutchins, quando a manipulação de símbolos é automatizada em um sistema formal, e.g. sistemas de busca baseados no modelo consulta-resposta, nem os processos cognitivos, nem as atividades da pessoa que manipula símbolos para interagir com o sistema, são modelados. Os símbolos *per se* são desmaterializados e colocados dentro do sistema através de formas ou estruturas de representação internas.

Em resumo, parte dos problemas enfrentados pelo usuário que interage através do modelo consulta-resposta para realizar suas atividades de busca são colocados “fora da sala”, ou seja, são externos ao sistema. A Fig. 5.10 mostra, por exemplo, o posicionamento externo dos sistemas de busca com relação ao contexto sociocultural dos usuários. Embora sistemas de busca sejam parte de sistemas de informação de larga-escala, a interação no modelo consulta-resposta não captura aquele contexto, pois sistemas formais são cegos de significado. Por mais sofisticados sejam os processos rodando no sistema de busca, eles

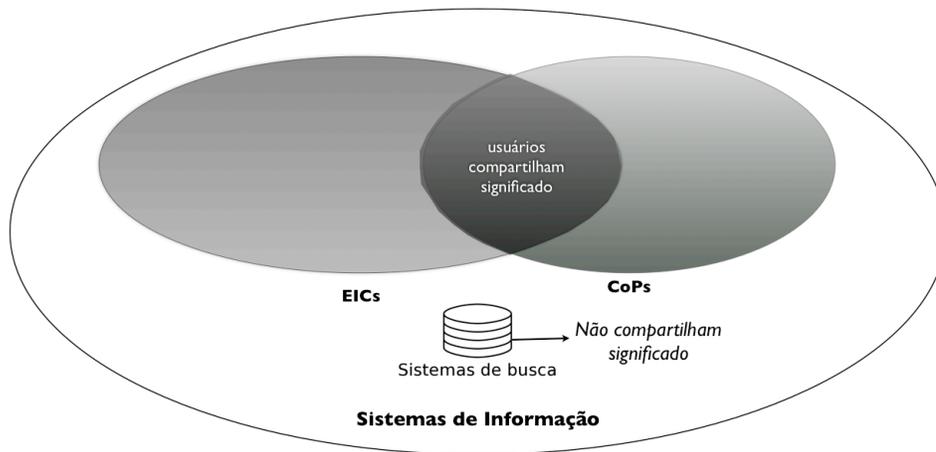


Figura 5.10: O posicionamento dos sistemas de busca com relação ao contexto sociocultural dos usuários.

sempre serão incapazes de compreender o significado das consultas dos agentes humanos que as produzem, como os próprios agentes humanos as compreendem.

## 5.5 Teoria fundamentada em dados

Esta seção apresenta uma teoria sobre a influência da familiaridade nos processos de busca. Esta teoria é fundamentada nos resultados empíricos apresentados em capítulos anteriores. O processo de desenvolvimento dessa teoria segue as tendências construtivistas de Glaser [Glaser, 1978] e Charmaz [Charmaz, 2006]. Segundo os autores, um aspecto importante no processo de construção de uma teoria fundamentada em dados é a aquisição de *sensibilidade teórica*:

*“When you theorize [using Grounded Theory], you reach down to fundamentals, up to abstractions, and probe into experience. The content of theorizing cuts the core of studied life and poses new questions about it. [...] Taking a closer look at processual analysis may aid your efforts to construct theory. [In] studying a process... you define and conceptualize relationships between experiences and events”.* ([Charmaz, 2006], p. 135–136)

Não por acaso, a teoria aqui apresentada modifica a terminologia originalmente apresentada na Seção 5.2 (ver Fig. 5.1), colocando no gerúndio os termos (1) tarefa de trabalho (*“work task”*), (2) procura de informação (*“seeking”*) e (3) recuperação de informação (*“information-retrieval”*). O objetivo desta mudança é o de salientar os subprocessos en-

volvidos em cada nível da atividade de busca. A mudança não é meramente cosmética e tem raiz nas orientações para a construção de uma teoria fundamentada a dados:

*“Adopting gerunds fosters theoretical sensitivity because these words nudge us out of static topics and into enacted processes. [...] Thus, I suggest renewed emphasis on actions and processes [...] as a strategy in constructing theory and moving beyond categorizing types of individuals”.* ([Charmaz, 2006], p. 136)

Dessa forma, a teoria aqui apresentada estende o modelo de Järvelin e Ingwersen [Järvelin and Ingwersen, 2004], focando nos seguintes subprocessos de busca:

1. Trabalhando com informação (*Working with information*)
2. Procurando informação (*Seeking information*)
3. Recuperando informação (*Retrieving information*)

### 5.5.1 Trabalhando com informação

O Cap. 2 apresentou os resultados de um estudo sobre os processos de busca no trabalho de pesquisadores acadêmicos. Um fenômeno central foi identificado, a saber, o de que pesquisadores acadêmicos preferiam encontrar novamente artigos científicos na Web usando bibliotecas digitais, mesmo sabendo que tais documentos haviam sido armazenados previamente em seus computadores pessoais. Pesquisadores acadêmicos revelaram, no entanto, ter a necessidade de organizar tais documentos, por exemplo mantendo-os em pastas que refletiam a relação entre documentos e uma dada tarefa de trabalho. Tal necessidade gerava uma lacuna entre as funcionalidades oferecidas pelos sistemas de gerenciamento de informações pessoais, e.g. estruturas de arquivos, e os de gerenciamento de informações gerais, e.g. bibliotecas digitais *online*. A parte superior da Fig. 5.11 mostra as características principais do processo de trabalho dos pesquisadores acadêmicos que participaram do estudo.

A parte inferior da Fig. 5.11 detalha duas propriedades principais desse processo, i.e. encontrando e reencontrando documentos. Pesquisadores acadêmicos estão constantemente em busca de documentos e procuram mantê-los juntos em estruturas de arquivos. Entretanto, ao perderem um documento, acadêmicos enfrentam a escolha de encontrá-los novamente em seus ambientes computacionais pessoais ou buscá-los novamente diretamente na Web. Um resultado do estudo apresentado no Cap. 2 é que pesquisadores acadêmicos acabam por reformular suas estratégias de gerenciamento de artigos científicos e muitas vezes passam a procurar por esses documentos diretamente na Web, deixando

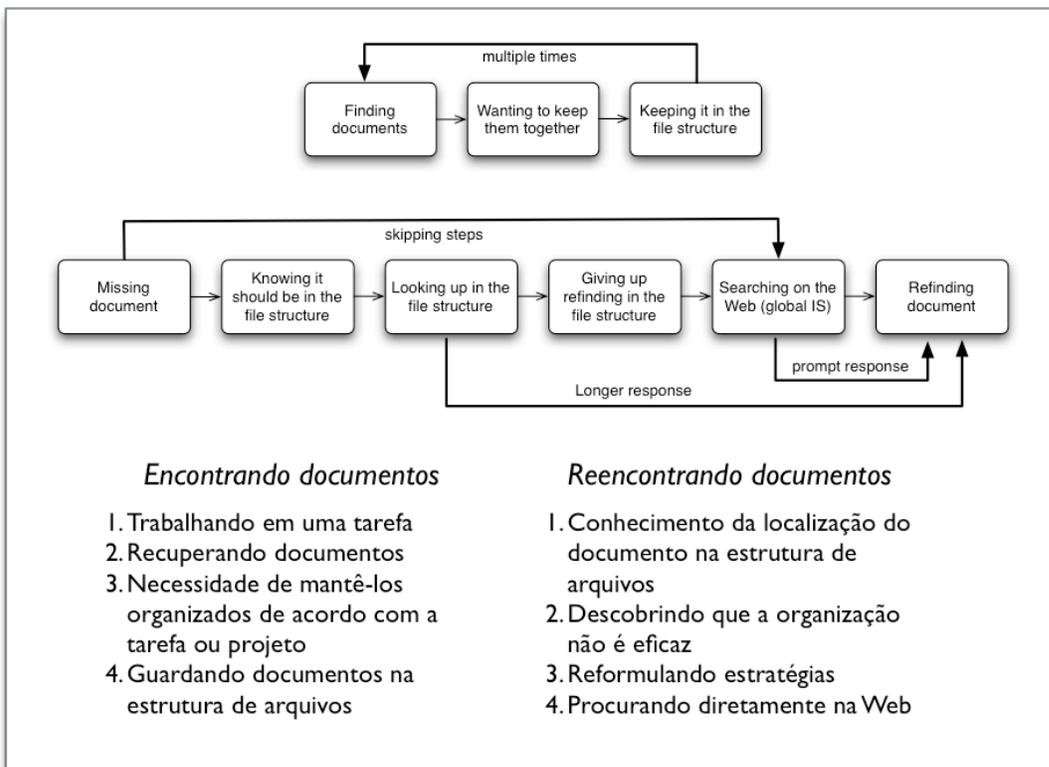
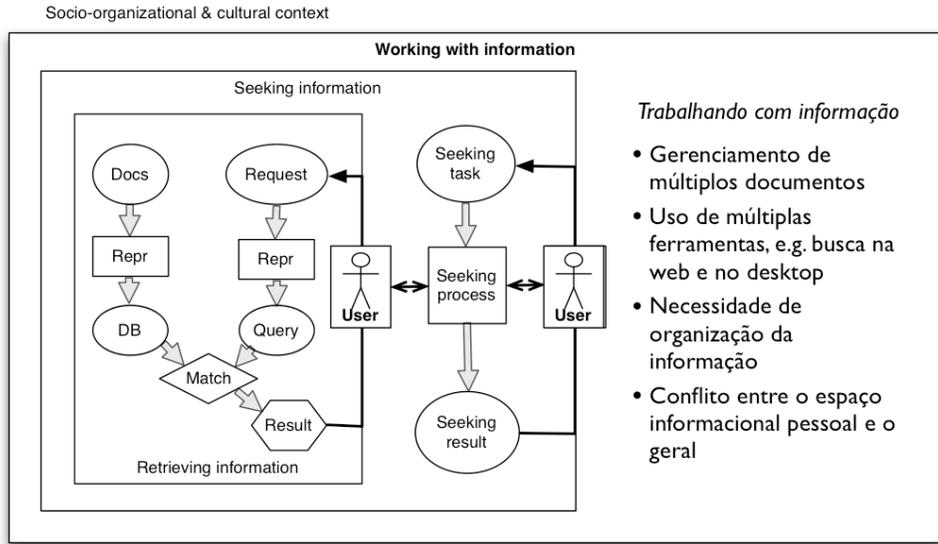


Figura 5.11: Características do trabalho de pesquisadores acadêmicos com informação (acima) e seus subprocessos (abaixo).

de organizá-los em suas estruturas de arquivos. Portanto, as seguintes condições foram encontradas no trabalho de acadêmicos com múltiplos documentos:

**Lacuna entre os espaços de informação pessoal e geral** O usuário tem controle sobre seu espaço de informação pessoal, porém este controle demanda tempo de manutenção, de modo a manter tal espaço organizado para o reencontro eficiente de documentos. Por outro lado, espaços de informação gerais como bibliotecas digitais são organizados para o público e não oferecem, portanto, ferramentas eficientes para personalização, e.g. organização de documentos em pastas referentes a um projeto. Desta forma, usuários adotaram estratégias para lidar com a lacuna existente entre ambos os espaços. Tais estratégias foram influenciadas pelo grau de familiaridade dos usuários para com o assunto relacionado aos documentos.

**Grau de familiaridade** Nos casos em que pesquisadores acadêmicos mostraram grande familiaridade com as palavras-chave relacionadas ao assunto buscado, e.g. pedaço do nome ou autor do artigo científico, eles puderam ter o benefício da prontidão com a qual um dado documento poderia ser encontrado e recuperado diretamente usando mecanismos de busca da Web. Essa condição mostrou influenciar as estratégias tomadas para manter documentos em seus espaços informacionais pessoais, de modo que muitos pesquisadores preferiram manter organizados em pastas apenas os documentos de sua própria autoria, recuperando os demais documentos através de buscas por palavras-chave na Web.

## 5.5.2 Procurando informação

O Cap. 3 apresenta os resultados de um estudo que monitorou o comportamento de busca de pesquisadores acadêmicos por períodos regulares de tempo. Participantes submeteram os dados de utilização de bibliotecas digitais e sistemas de busca para um servidor de coleta. Estes dados serviram como base para condução de entrevistas com esses participantes. Tal estudo nos auxiliou a compreender o subprocesso de procura de informação de pesquisadores acadêmicos, cujas características principais são mostradas na parte superior da Fig. 6.1.

Os resultados apresentados no Cap. 3 mostram que a relação de familiaridade dos usuários com o assunto buscado influencia significativamente as estratégias adotadas durante o subprocesso de procura de informação. Desta forma, foi possível caracterizar dois extremos no espectro da relação entre a familiaridade e o processo de procura de informação, o que chamamos de buscas  $EF+$  e  $EF-$ . A variação de familiaridade ao longo desse espectro é influenciada e, ao mesmo tempo, influencia dois tipos de comportamento:

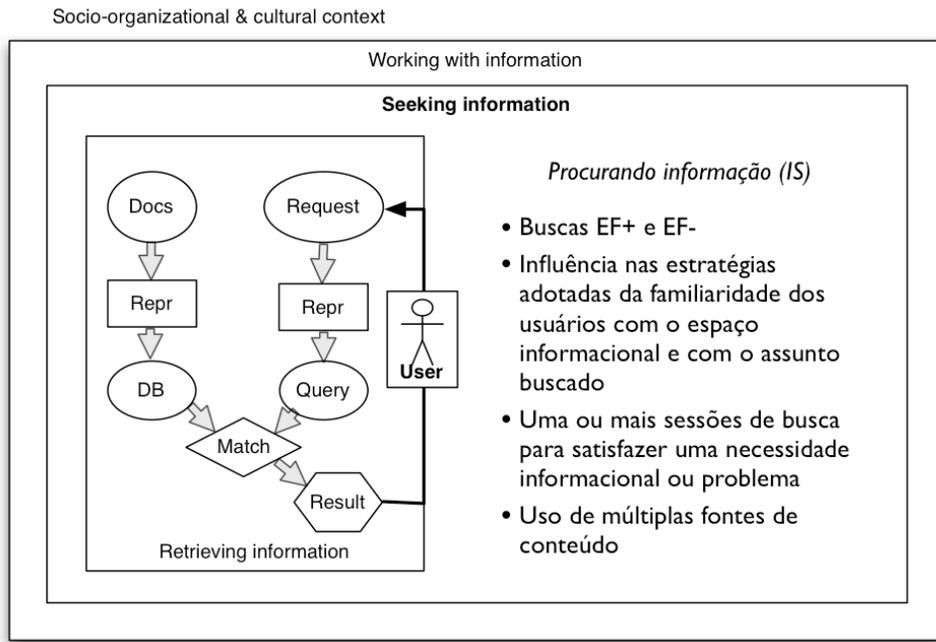


Figura 5.12: Características do processo de procura de informação.

**Uso de artefatos auxiliares** Artigos científicos, anotações, emails, dentre outros, são peças importantes no processo de aumento da familiaridade do usuário com relação ao assunto buscado. Tais artefatos provêm os usuários com informações adicionais que são utilizadas em estratégias mais eficazes de busca. Baseado nesses artefatos, usuários podem tomar decisões mais acertadas com relação a quais palavras-chave podem ser utilizadas, quais fontes de informação podem levá-los a melhores resultados, quais documentos já foram analisados etc.

**Cooperação com outros agentes humanos** De forma complementar, a cooperação com outros agentes humanos, seja utilizando formas de comunicação síncrona, e.g. face a face, ou assíncrona, e.g. email, produz insumos que levaram a mudanças nas estratégias adotadas por usuários. No caso de acadêmicos, por exemplo, sugestões como anais de conferências ou mesmo de novos artigos científicos auxiliou usuários a adquirirem familiaridade com o assunto buscado de forma mais rápida, tornando o processo de procura de informação mais eficiente quando comparado ao de usuários que não receberam tal auxílio.

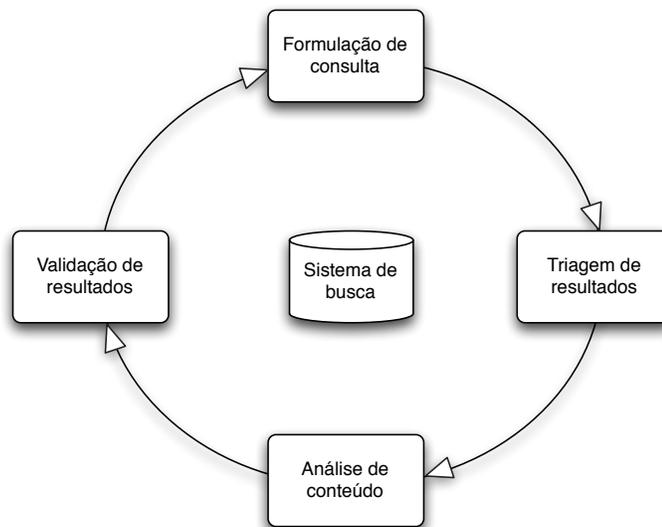
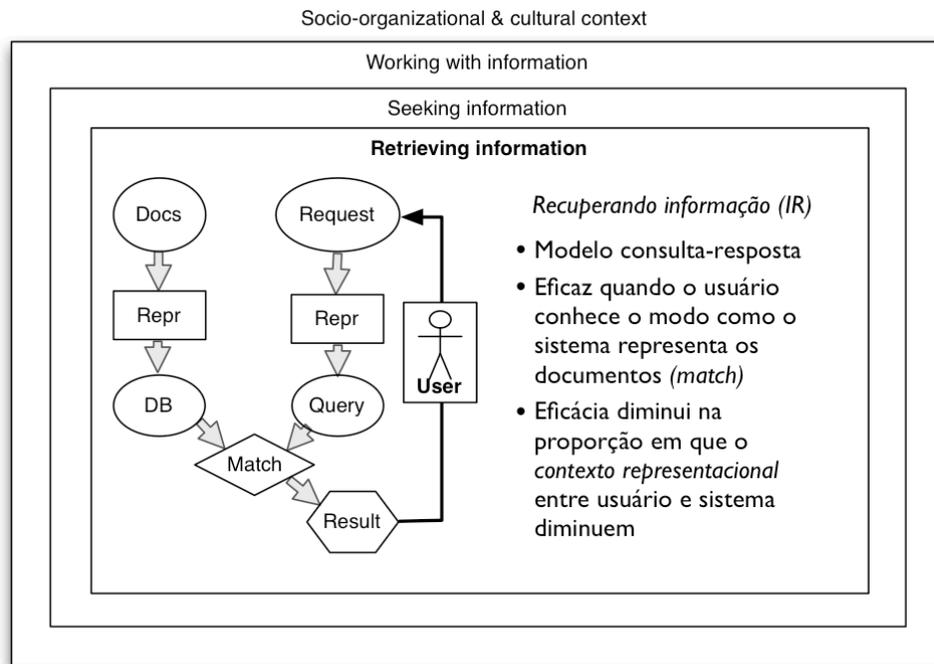


Figura 5.13: Processo de recuperação de informação (acima) e suas etapas (abaixo).

### 5.5.3 Recuperando informação

O resultados apresentados no Cap. 4 aprofundaram a investigação no nível de recuperação de informação, através da observação de novatos em Linux durante buscas utilizando o sistema de busca Google, que implementa o modelo consulta-resposta introduzido na Seção 5.2.1. A Fig. 5.13 mostra as características do processo de recuperação de informação (acima) e suas etapas (abaixo). A baixa familiaridade de usuários novatos em Linux produziu as seguintes consequências:

**Incerteza** A baixa familiaridade com o assunto buscado leva a incerteza com relação a quais resultados devem ser selecionados para serem posteriormente analisados, aumentando o tempo gasto na triagem e dificultando a extração de informação relevante a partir de páginas da Web. A incerteza também mostrou-se evidente através da dificuldade mostrada pelos usuários ao estabelecer uma relação de confiança para com as páginas retornadas pelo sistema de busca. Uma estratégia comum nesses casos foi a de analisar diversos resultados separadamente em um processo de aprendizado contínuo durante cada sessão de busca. Depois de aprender mais sobre o domínio da tarefa de busca, participantes tipicamente reformulavam suas consultas de modo a refinar os resultados previamente obtidos [Figueira Filho et al., 2010c].

**Necessidade de fontes bem estruturadas** A baixa familiaridade também levou a uma maior necessidade por fontes de informação com conteúdo bem estruturado e organizado. No caso particular dos novatos em Linux, essa necessidade mostrou-se evidente na busca de informação técnica: usuários mostraram maior preferência por sites de pergunta-resposta (*Q&A*) em que a melhor resposta para uma dada pergunta era mostrada em destaque na página de resultado, o que poupava esforço na etapa de análise de conteúdo [Figueira Filho et al., 2010c].

## 5.6 Propostas de soluções para o *design* de sistemas de busca

A teoria apresentada na seção anterior permite compreender que embora os sistemas de busca atuais sejam eficientes ao suportar atividades de recuperação de informação (Seção 5.5.3), eles falham ao suportar atividades de mais alto nível, como a de busca exploratória (Seção 5.2.4). Os resultados do Cap. 3 indicam que o uso de artefatos auxiliares e a colaboração com agentes humanos durante a procura de informação são importantes para o fortalecimento da relação de familiaridade entre usuários e o assunto buscado. Os resultados do Cap. 2 mostram que o trabalho com informação requer algum nível de organização, principalmente quando a informação procurada serve a algum objetivo de

longo prazo ou, ainda, quando a tarefa de busca possui um caráter exploratório e não pode ser resolvida, por exemplo, com uma única sessão de busca por palavras-chave. O suporte a essas atividades deve ser integrado ao modelo de interação prevalente nas ferramentas de busca atuais, oferecendo aos usuários ferramentas para (1) organização do espaço informacional de acordo com tarefas específicas e (2) colaboração com outros usuários. As seguintes tecnologias podem ser combinadas de modo a oferecer tais ferramentas:

### 5.6.1 Anotação colaborativa

A funcionalidade de anotação vem sendo usada em diversas aplicações Web, dentre elas as de *social bookmarking*<sup>2</sup>, e.g. delicious<sup>3</sup>, de *autoria colaborativa*, e.g. Wikipedia<sup>4</sup>) e de *filtragem colaborativa*, i.e. sistemas de recomendação que utilizam o conhecimento da comunidade para ranquear conteúdo Web [Resnick and Varian, 1997, Terveen et al., 1997].

No contexto dos sistemas de busca, estamos particularmente interessados em um tipo de anotação que vem crescendo em popularidade na Web. Trata-se dos mecanismos que permitem realçar pedaços de texto (“*highlighting*”) e criar notas que se fixam (“*sticky notes*”) em pontos especificados pelo usuário na página Web. Esse tipo de funcionalidade é implementada, por exemplo, em serviços como o Diigo<sup>5</sup> (Fig. 5.14).

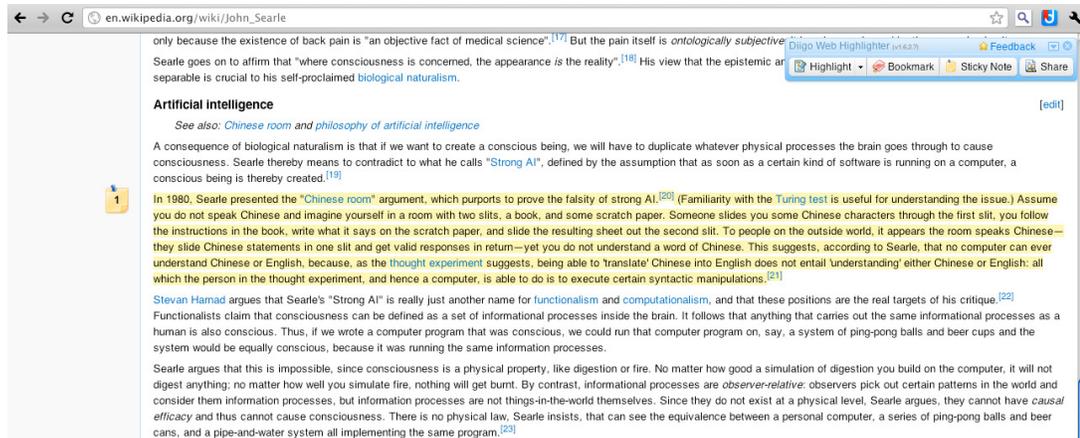


Figura 5.14: Ilustração da funcionalidade apresentada pelo serviço Diigo. Uma extensão de navegador permite que o usuário realce partes do texto e adicione notas (retângulo amarelo ao lado do parágrafo) que ficam permanentemente afixadas em uma determinada posição da página.

<sup>2</sup>Ver artigo [Figueira Filho et al., 2010a], Seção 2.2, para uma caracterização do termo.

<sup>3</sup><http://delicious.com>

<sup>4</sup><http://www.wikipedia.org>

<sup>5</sup><http://www.diigo.com>

O interessante sobre essa tecnologia é seu potencial colaborativo para filtragem, sumarização e agregação de conteúdo no contexto de tarefas exploratórias de busca. Nelson et al. [Nelson et al., 2009] avaliou os efeitos no aprendizado de usuários ao utilizarem ferramentas de anotação em diferentes condições. O experimento comparou o aprendizado de dois grupos. O primeiro tinha acesso a ferramentas convencionais para tomar notas, e.g. papel, caneta e editores de texto como o *MS Word*. O segundo grupo teve acesso às ferramentas digitais de anotação e um terceiro grupo tinha acesso às anotações de um amigo com maior experiência no assunto. Os resultados mostram que os ganhos de aprendizado do terceiro grupo foi significativamente maior que os ganhos dos demais grupos. O experimento mostrou, ainda, que os participantes do terceiro grupo visitaram menos páginas, passaram mais tempo lendo as que visitaram e escreveram mais no ensaio para avaliar o conhecimento adquirido durante a tarefa.

### 5.6.2 Hipertexto e o conceito de caminhos

Hammond e Allinson [Hammond and Allinson, 1988] mostram, no contexto de sistemas de suporte ao aprendizado, que os objetivos do usuário, sua familiaridade com o assunto e a estrutura dos documentos utilizada pelo sistema são fatores relevantes para um aprendizado mais eficaz:

*“The optimal level of [system] control is likely to depend on the nature of the learners, on their familiarity with the materials, on their learning goals and not least on the nature of the system they are using. Handing complete control to the beginner may be as ineffective as forcing the expert through a drill-and-practice tutorial (...) there seems to be a consensus that more knowledgeable learners can capitalise on self-directed learning whereas novices should be provided with more structure. There is evidence, too, of strong individual differences in the optimal organization of information. Finally, the nature of the learner’s task may also determine the most appropriate level of control. A student preparing an essay may have different requirements from one seeking a particular point of information”.* ([Hammond and Allinson, 1988], p. 1, col. 2)

De forma semelhante, os resultados do Cap. 4 mostram que usuários novatos têm preferência por documentos bem estruturados, como tutoriais, em vez de páginas de um fórum de discussão, por exemplo. Em tempo, o problema de sites de pergunta e resposta (Q&A) é que parte do conteúdo é preparado por usuários experientes em resposta às perguntas específicas, o que dificulta a exploração de um conhecimento mais geral que é muitas vezes necessário por parte de usuários novatos.

Em parte, esse problema está relacionado à estrutura do hipertexto. Uma estrutura hipertextual típica provê um conjunto pré-definido de *links* entre documentos, mas isso não implica que esse mesmo conjunto seja igualmente eficaz para todos os contextos de uso que incluem tais documentos, e.g. uma mesma estrutura hipertextual pode ser eficiente para um usuário experiente e não ser para um usuário novato, e vice-versa. A Fig. 5.15 mostra um exemplo de estrutura hipertextual, como a utilizada na Web.

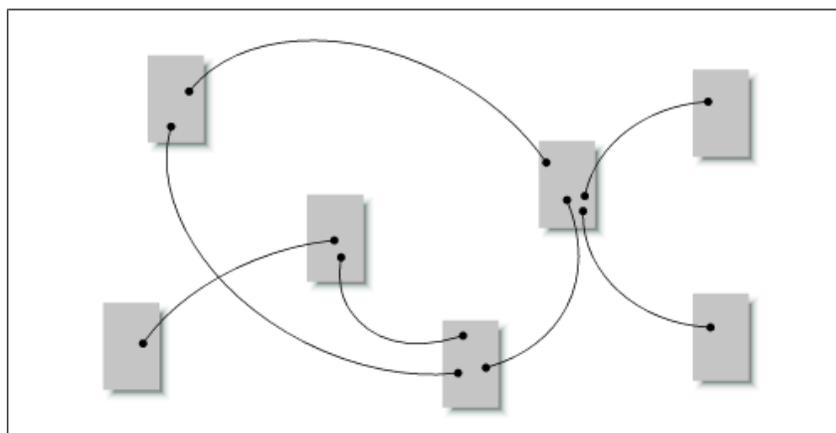


Figura 5.15: Documentos interligados por uma estrutura hipertextual.

Em busca de uma solução para o problema, Zellweger [Zellweger, 1989] implementa o conceito de *caminhos* através de um mecanismo chamado de “*Scripted Documents*”. *Caminhos* consistem em um conjunto de *links* que transvertem de forma ordenada um conjunto de documentos, como em uma apresentação concebida para transmitir uma ideia integrada sobre um tema particular. Caminhos são independentes da estrutura de *links* do hipertexto, de modo que a existência de um *link* entre dois documentos não necessariamente implica na formação de um caminho entre os mesmos documentos. A Fig. 5.16 ilustra esse conceito.

Dessa forma, uma possível solução é a criação de ferramentas de anotação que possibilitem a criação colaborativa de caminhos que são independentes da estrutura hipertextual existente. Estes caminhos podem ser customizados de acordo com a tarefa, objetivos de busca, e principalmente, com a familiaridade dos usuários com o assunto buscado.

### 5.6.3 Redes sociais em suporte à busca

Outra solução pouco explorada utiliza uma abordagem colaborativa para distribuir o problema da filtragem de informação e análise de conteúdo entre pares no escopo de uma rede social. Um espaço para interação social similar aos disponíveis em serviços

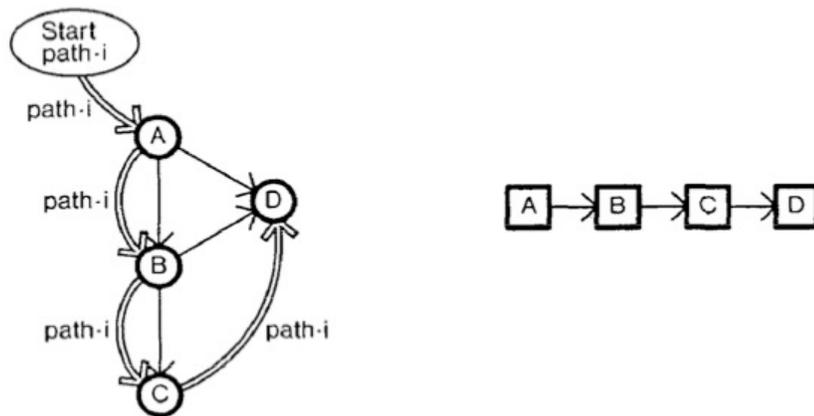


Figura 5.16: Ilustração do conceito de caminhos, retirada de [Zellweger, 1989]. À esquerda, os nós representam documentos no hipertexto. *Links* entre documentos são representados pelas setas simples. Caminhos são representados pelas setas duplas. À direita, a figura ilustra o caminho produzido.

como Facebook<sup>6</sup> serviria para que usuários pudessem compartilhar tarefas de busca. O contexto de busca, dessa forma, é construído pela interação entre usuários e não mais pelas palavras-chave utilizadas na consulta a um sistema de busca que implementa o modelo consulta-resposta. A atividade cognitiva torna-se *distribuída*, e não mais *individualizada*. Em outras palavras, o sistema de interação serve como uma veículo para comunicação entre pares, sendo que o significado das tarefas de busca é interpretado por outros usuários e não mais por um sistema.

A solução aqui proposta é diferente, entretanto, daquele encontrada nos fóruns de discussão da Web e em sites de pergunta-resposta (*Q&A*) em geral. Primeiro, fóruns não tiram proveito dos benefícios das redes sociais, nas quais usuários compartilham de um *contexto social* que, por sua vez, permitem a criação de uma cultura organizacional própria [Boyd, 2007] através do compartilhamento de conhecimento tácito, interesses e, até mesmo, afeições. Segundo, a credibilidade na fonte de informação mostrou-se fundamental no estudo realizado com usuários novatos em Linux (Cap. 4). Nelson et al. [Nelson et al., 2009] mostram que usuários confiam mais nas informações providas por “amigos”. Ao contrário dos fóruns, a confiança entre os pares que interagem em uma rede social é estabelecida pelo contexto social que esses pares compartilham. Terceiro, fóruns não resolvem o problema da análise de conteúdo. Discussões são muitas vezes extensas demais para serem interpretadas corretamente por usuários novatos. Nesse aspecto, a anotação colaborativa apresentada anteriormente pode auxiliar na sumarização e interpretação compartilhada da informação contida em páginas da Web. Essas anotações são

<sup>6</sup><http://www.facebook.com>

integradas na forma de caminhos de modo a diminuir a sobrecarga cognitiva dos usuários inexperientes.



# Capítulo 6

## Conclusão

Com o advento da tecnologia da informação e a criação de ambientes virtuais para interação social, novas possibilidades para a gestão do conhecimento estão surgindo. No entanto, em sistemas de informação de larga escala como a Web, a troca de conhecimento é uma tarefa desafiadora por vários motivos. Primeiro, espaços de informação na Web tem para satisfazer necessidades de informação distintas, que são originadas a partir de grupos heterogêneos de usuários. Segundo, a articulação do conhecimento entre os usuários com níveis semelhantes de familiaridade com relação a um determinado assunto ocorre de uma maneira diferente da articulação de conhecimento entre usuários com níveis diferentes de familiaridade, e.g. a partilha de conhecimentos entre especialistas acontece de uma maneira diferente do que com os novatos. Esta tese procurou compreender, através da investigação das práticas de busca e de gerenciamento de informação, o cenário complexo composto pelos sistemas de busca, categorias, pessoas e outras fontes de informação que são tipicamente utilizadas em tarefas de busca exploratórias. Baseado em resultados empíricos, a teoria apresentada no Cap. 5 caracteriza dois tipos de busca, i.e.  $EF+$  e  $EF-$ , de acordo com as estratégias adotadas e com o grau de familiaridade adquirido pelos usuários durante seus processos de busca. As características gerais de ambos os tipos de busca são sumarizadas na Fig. 6.1.

Buscas  $EF+$  são acompanhadas por uma maior familiaridade com o espaço informacional. Por exemplo, para re-encontrar um dado artigo científico, pesquisadores acadêmicos tem a sua disposição o espaço informacional pessoal e o geral (Fig. 6.2). Ao utilizar um espaço informacional geral, e.g. biblioteca digital na Web, um alto grau de familiaridade com as palavras-chave resulta no re-encontro quase que imediato do documento — um efeito da prontidão da Web como chamamos no Cap. 2. Como vantagem, usuários também têm acesso a outros tipos de informação não organizados por eles, e.g. referências a outros artigos científicos. Como desvantagem, usuários não conseguem organizar a informação disponível num espaço informacional de uso geral de acordo com as suas necessidades

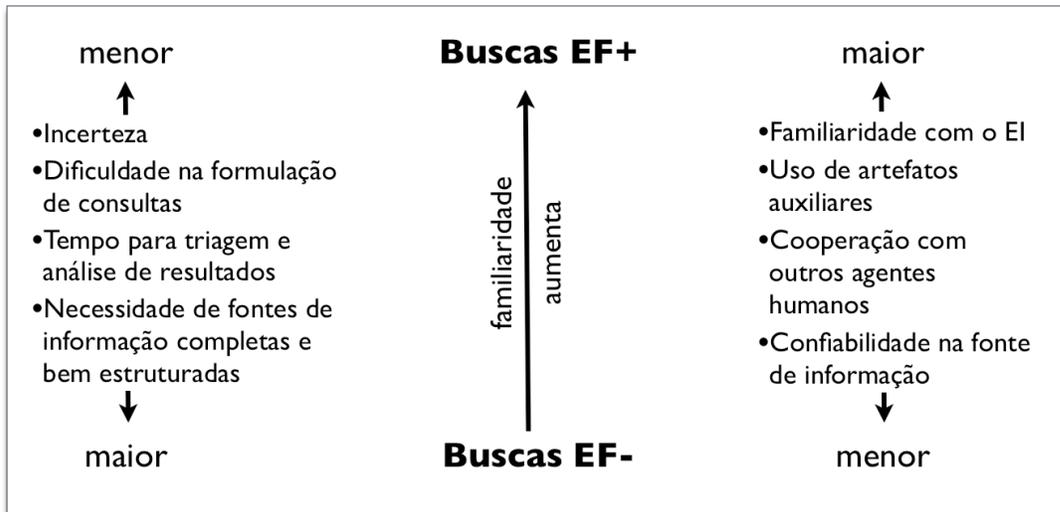


Figura 6.1: Sumário das propriedades do processo de busca de acordo com a variação de familiaridade.

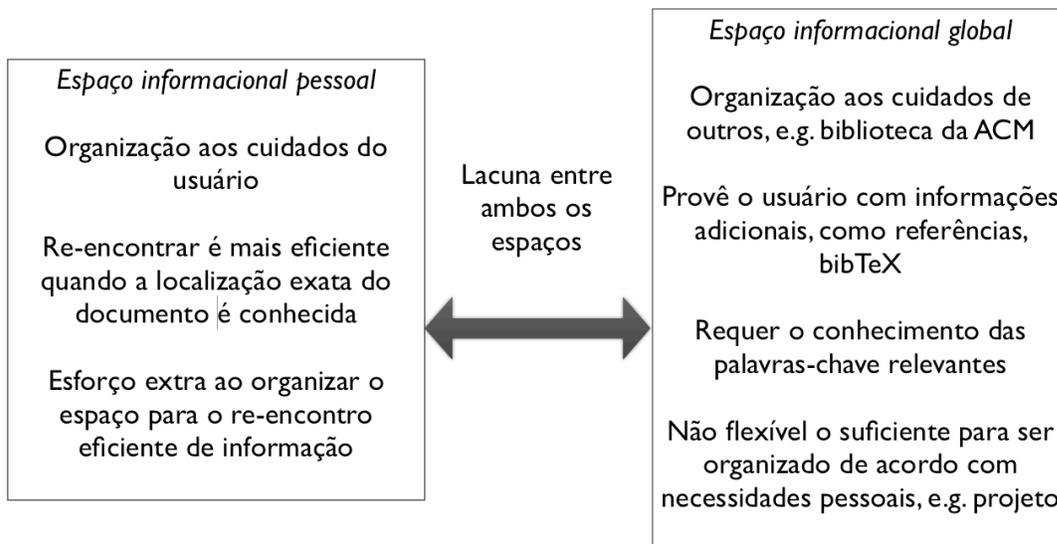


Figura 6.2: Características dos espaços informacionais *pessoal* e *geral*, bem como a lacuna existente entre ambos.

personais. Por outro lado, o re-encontro de documentos baseado em seus espaços informacionais pessoais também é quase que imediato, i.e. o usuário sabe exatamente onde guardou um dado documento. Como vantagem, usuários podem relacionar documentos de acordo com um projeto, criar notas pessoais, referências entre arquivos etc. Como desvantagem, usuários precisam gerenciar tal organização de modo eficiente para o re-encontro e muitas vezes são forçados a lembrar a localização de um dado documento com o qual não têm trabalhado recentemente. Ao lidar com múltiplos documentos, usuários mostraram preferir os mecanismos de busca da Web, mas mostraram também insatisfação ao não conseguirem organizar o espaço informacional geral de acordo com as suas necessidades pessoais. Como resultado, usuários estão constantemente enfrentando a decisão entre migrar a informação buscada de um espaço ao outro, o que gera uma sobrecarga nas tarefas de trabalho desempenhadas por esses usuários.

Buscas  $EF^-$ , por outro lado, são acompanhadas por uma menor familiaridade com o espaço informacional. Usuários podem apresentar uma maior incerteza ao avaliar resultados de busca e também dificuldades na triagem de conteúdo. Os resultados do Cap. 3 mostram que diversas fontes de informação externas aos mecanismos de busca podem auxiliar os usuários na transição entre buscas  $EF^-$  para  $EF^+$ . Trata-se de planilhas, *emails*, mas também de outros atores humanos que, através da interação em suas rotinas de trabalho, produzem os insumos para estratégias de busca mais elaboradas que servem de suporte às atividades exploratórias, também contribuindo para o aumento do nível de familiaridade de usuários menos experientes com o domínio de conhecimento relacionado às suas motivações e objetivos de busca.

A presente tese mostra, através de diversos estudos e diferentes metodologias de investigação, as características da busca exploratória no contexto da Web atual. Tais atividades ocorrem a partir de objetivos e motivações de mais alto nível e demandam alternativas aos sistemas de busca atuais, no que tange às interfaces empregadas, aos mecanismos de comunicação entre usuários com diferentes níveis de familiaridade e ao gerenciamento de documentos. Os resultados mostram que a exploração do conhecimento na Internet depende de um conjunto de artefatos que são parte integrante das atividades cognitivas desempenhadas pelos usuários durante a realização de suas buscas. A tese apresenta alternativas ao modelo de interação utilizado na maioria dos sistemas de busca atuais e sugere a integração de sistemas de anotação colaborativa e redes sociais, no desenvolvimento de sistemas de informação que visam suportar tarefas de busca exploratórias.

## 6.1 Contribuições em artigos

Os seguintes trabalhos foram publicados durante o desenvolvimento desta tese. A lista abaixo apresenta-se por tipo de publicação, i.e. trabalhos completos, trabalho publicado

em colóquio de doutorado e trabalhos publicados em *workshops*. As publicações de cada tipo estão listadas de acordo com ordem cronológica crescente:

### Trabalhos Completos publicados em anais de congressos

1. [Figueira Filho et al., 2008a] Baseado no referencial teórico de comunidades de prática, o artigo apresenta uma análise dos sistemas de classificação na Web, i.e. ontologias e classificação social (“*social tagging*”). Os resultados derivam dois princípios para o *design* para sistemas de classificação na Web: (a) colaboração, que permite com que usuários de diferentes comunidades de prática participem ativamente na categorização de documento e (b) multidimensionalidade, que permite a co-existência e o gerenciamento de múltiplas perspectivas de classificação, o que é fundamental para comunicação, consumo e produção de conteúdo em sistemas complexos, heterogêneos e de larga-escala como a Web.
2. [Figueira Filho et al., 2010a] O artigo amplia os resultados do artigo acima, abordando as categorias utilizadas na produção e consumo de informação na Web como objetos de fronteira. O artigo argumenta, ainda, que a comunicação entre membros de uma comunidade de prática ocorre de forma distinta no interior destas comunidades e através suas bordas. Estes resultados teóricos são discutidos à luz do conceito de contexto como (a) representação e (b) interação [Dourish, 2004].
3. [Figueira Filho et al., 2010c] O artigo identifica problemas com a busca exploratória e propõe um protótipo de interface que implementa um novo modelo de interação baseado na interação social e em tecnologias de anotação da web social.

### Trabalho publicado em Colóquio de Doutorado

4. [Figueira Filho et al., 2010b] O trabalho foi apresentado no colóquio de doutorado da conferência CSCW e discute, sob uma perspectiva sociotécnica, os problemas do paradigma consulta-resposta sob a perspectiva do conceito de comunidades de prática *online*, como também as implicações das decisões de *design* desse paradigma para implementação de sistemas de busca.

### Trabalhos publicados em *Workshops*

5. [Figueira Filho et al., 2008b] O trabalho representa uma investigação sobre o papel dos sistemas de recomendação da web social com o objetivo de reduzir

a sobrecarga cognitiva dos seus usuários. O artigo recomenda que informações estruturais, como os elos entre usuários numa rede social, devem ser unificadas com informações semânticas, o que poderia conceder mais significado aqueles elos e produzir melhores recomendações.

6. [Figueira Filho et al., 2009a] Baseado no trabalho anterior, este trabalho apresenta uma proposta de sistema de recomendação que utiliza algoritmos de ranqueamento baseados nos elos de uma rede social — em vez de elos entre documentos da Web — para filtrar conteúdo produzido por usuários com maior expertise em sistemas como fóruns de discussão técnica na Web. A implementação e avaliação do referido algoritmo de recomendação não foi levada a cabo, pois estudos preliminares mostraram que as dificuldades dos usuários novatos vão além da filtragem eficaz de conteúdo relevante, i.e. mesmo sendo relevante à tarefa de busca, os usuários não sabem como interpretar tal conteúdo de forma satisfatória.
7. [Figueira Filho et al., 2009b] Este trabalho apresenta uma proposta de interface para o refinamento progressivo de consultas.

## 6.2 Trabalhos futuros

A presente tese aponta algumas direções para pesquisas futuras. Primeiro, a teoria proposta precisa ser validada com usuários da mesma população, e.g. pesquisadores acadêmicos e novatos em Linux, mas também perante usuários com outras atividades, e.g. engenheiros de *software*. Isto permitirá reconhecer em que pontos a teoria apresentada pode ser generalizável para outras populações e em que pontos ela precisa ser refinada, expandida e até, delimitada, de acordo com o contexto em análise.

Segundo, as práticas de gerenciamento de informação têm mudado com o advento de dispositivos móveis, como os *iPads*<sup>1</sup>. Sob esta perspectiva, as diferenças entre o espaço informacional pessoal e global tendem a sofrer mudanças, uma vez que qualquer documento poderá ser armazenado nas “nuvens”, termo que ganhou notoriedade com o aumento dos sistemas de informação nos quais os dados estão guardados de forma distribuída, fora do computador pessoal do usuário. Por exemplo, muitos estudos em gerenciamento de informação pessoal apontam a força da metáfora da localização ao descrever a preferência dos usuários em re-encontrar documentos usando pastas. Entretanto, esta mesma localização está cada vez menos associada a um dispositivo em particular, uma vez que uma mesma pasta poderá estar acessível através de qualquer dispositivo. Desta forma, espera-se as práticas de gerenciamento utilizadas até hoje pelos usuários sofram mudanças

---

<sup>1</sup><http://www.apple.com/ipad/>

à medida em que os espaços pessoais e globais sejam gerenciáveis a partir de qualquer dispositivo.

Terceiro, os resultados desta tese mostram o potencial da busca colaborativa, como um mecanismo para elaboração conjunta de processos de busca, re-uso e compartilhamento de conhecimento. Uma solução de busca poderia ser implementada na forma de um protótipo e o *design rationale* dessa ferramenta deve estar baseado na teoria proposta nesta tese. Em outras palavras, os resultados teóricos aqui apresentados devem guiar os requisitos para construção de espaços informacionais colaborativos que possam permitir a busca colaborativa de informação.

# Referências Bibliográficas

- [Aula et al., 2005] Aula, A., Jhaveri, N., and Käki, M. (2005). Information search and re-access strategies of experienced web users. In *Proceedings of the 14th international conference on World Wide Web, WWW '05*, pages 583–592, New York, NY, USA. ACM.
- [Barreau and Nardi, 1995] Barreau, D. and Nardi, B. A. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bull.*, 27(3):39–43.
- [Bates, 1993] Bates, M. (1993). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424.
- [Belkin, 1982] Belkin, N. J. (1982). Ask for information retrieval: Part i. background and theory. *Journal of Documentation*, 38(2):61–71.
- [Bergman et al., 2003] Bergman, O., Beyth-Marom, R., and Nachmias, R. (2003). The user-subjective approach to personal information management systems. *J. Am. Soc. Inf. Sci. Technol.*, 54:872–878.
- [Bergman et al., 2008] Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., and Whittaker, S. (2008). Improved search engines and navigation preference in personal information management. *ACM Trans. Inf. Syst.*, 26(4):1–24.
- [Beynon-Davies, 2009] Beynon-Davies, P. (2009). The language of informatics: The nature of information systems. *International Journal of Information Management*, 29(2):92 – 103.
- [Boardman and Sasse, 2004] Boardman, R. and Sasse, M. A. (2004). Stuff goes into the computer and doesn't come out: a cross-tool study of personal information management. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 583–590, Vienna, Austria. ACM.
- [Bossen, 2002] Bossen, C. (2002). The parameters of common information spaces:: the heterogeneity of cooperative work at a hospital ward. In *Proceedings of the 2002 ACM*

*conference on Computer supported cooperative work, CSCW '02*, pages 176–185, New York, NY, USA. ACM.

- [Bowker and Star, 1999] Bowker, G. and Star, S. (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA.
- [Boyd, 2007] Boyd, D. (2007). The significance of social software. In Burg, T. N. and Schmidt, J., editors, *BlogTalks reloaded: Social software research & cases*, pages 15–30. Norderstedt: Books on Demand.
- [Bush, 1945] Bush, V. (1945). As we may think. *Atlantic Magazine*, pages 1–10.
- [Bystrøm, 2002] Bystrøm, K. (2002). Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science and Technology*, 53:581–591.
- [Byström and Hansen, 2005] Byström, K. and Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology*, 56(10):1050–1061.
- [Capra et al., 2010] Capra, R., Marchionini, G., Velasco-Martin, J., and Muller, K. (2010). Tools-at-hand and learning in multi-session, collaborative search. In *Proceedings of the 28th international conference on Human factors in computing systems, CHI '10*, pages 951–960, New York, NY, USA. ACM.
- [Charmaz, 2006] Charmaz, K. (2006). *Constructing Grounded Theory*. Sage Publications.
- [Conklin, 1987] Conklin, J. (1987). Hypertext: An introduction and survey. *Computer*, 20(9):17–41.
- [Dourish, 2004] Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8:19–30.
- [Dreyfus, 1992] Dreyfus, H. L. (1992). *What computers still can't do: a critique of artificial reason*. The MIT Press.
- [Figueira Filho et al., 2008a] Figueira Filho, F. M., de Albuquerque, J. P., and de Geus, P. L. (2008a). Analisando sistemas de classificação na web sob a perspectiva da interação social em comunidades de prática. In *Proceedings of the VIII Brazilian Symposium of Human Factors on Computer Systems (IHC 08)*, pages 40–49. Editora da SBC.

- [Figueira Filho et al., 2008b] Figueira Filho, F. M., de Albuquerque, J. P., and de Geus, P. L. (2008b). Sistemas de recomendação e interação na web social. In *Proceedings of the 1st Workshop on Human-Computer Interaction Aspects in the Social Web, in conjunction with the VIII Brazilian Symposium of Human Factors on Computer Systems (IHC 08)*, pages 24–27.
- [Figueira Filho et al., 2009a] Figueira Filho, F. M., de Albuquerque, J. P., and de Geus, P. L. (2009a). Um sistema de recomendação para fóruns de discussão na web baseado na estimativa da expertise e na classificação colaborativa de conteúdo. In *Proceedings of the 2nd Workshop on Dissertations and Thesis on Information Systems, in conjunction with the V Brazilian Symposium on Information Systems (SBSI)*, pages 306–313.
- [Figueira Filho et al., 2010a] Figueira Filho, F. M., de Albuquerque, J. P., and de Geus, P. L. (2010a). Broadening the perspective on classification systems in the web: Analyzing web classification as a situated activity within communities of practice. In *Proceedings of the IADIS International Conference Collaborative Technologies (CT 2010)*, pages 117–124. IADIS Press.
- [Figueira Filho et al., 2009b] Figueira Filho, F. M., de Albuquerque, J. P., Resende, A., de Geus, P. L., and Olson, G. M. (2009b). A visualization interface for interactive search refinement. In *Proceedings of the 3rd Annual Workshop on Human-Computer Interaction and Information Retrieval (HCIR '09)*, pages 46–49.
- [Figueira Filho et al., 2010b] Figueira Filho, F. M., Olson, G. M., de Albuquerque, J. P., and de Geus, P. L. (2010b). Empirical studies of the searching behavior of novice users when looking for technical information. In *Proceedings of the Doctoral Colloquium in conjunction with the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW)*, pages 507–508.
- [Figueira Filho et al., 2010c] Figueira Filho, F. M., Olson, G. M., and de Geus, P. L. a. (2010c). Kolline: a task-oriented system for collaborative information seeking. In *Proceedings of the 28th ACM International Conference on Design of Communication (ACM SIGDOC 2010)*, pages 89–94.
- [Glaser, 1978] Glaser, B. G. (1978). *Theoretical Sensitivity: Advances in the methodology of Grounded Theory*. Sociology Press.
- [Hammond and Allinson, 1988] Hammond, N. and Allinson, L. (1988). Travel around a learning support environment: rambling, orienteering or touring? In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '88*, pages 269–273, New York, NY, USA. ACM.

- [Heaton and Groves, 2005] Heaton, J. and Groves, J. (2005). *Introducing wittgenstein*. Icon Books.
- [Hutchins, 1995] Hutchins, E. (1995). *Cognition in the Wild*. MIT Press, Cambridge, MA.
- [Järvelin and Ingwersen, 2004] Järvelin, K. and Ingwersen, P. (2004). Information seeking research needs extension towards tasks and technology. *Information Research*, 10(1):paper 212.
- [Jones et al., 2001] Jones, W., Bruce, H., and Dumais, S. (2001). Keeping found things found on the web. In *Proceedings of the tenth international conference on Information and knowledge management, CIKM '01*, pages 119–126, New York, NY, USA. ACM.
- [Jones et al., 2005] Jones, W., Phuwanartnurak, A. J., Gill, R., and Bruce, H. (2005). Don't take my folders away!: organizing personal information to get things done. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*, pages 1505–1508, Portland, OR, USA. ACM.
- [Lave, 1988] Lave, J. (1988). *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge University Press.
- [Lave and Wenger, 1991] Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- [Marchionini, 1997] Marchionini, G. (1997). *Information seeking in electronic environments*. Cambridge University Press.
- [Nelson et al., 2009] Nelson, L., Held, C., Pirolli, P., Hong, L., Schiano, D., and Chi, E. H. (2009). With a little help from my friends: examining the impact of social annotations in sensemaking tasks. In *Proceedings of the 27th international conference on Human factors in computing systems, CHI '09*, pages 1795–1798, New York, NY, USA. ACM.
- [Newell and Simon, 1976] Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. ACM*, 19:113–126.
- [Pirolli and Card, 1995] Pirolli, P. and Card, S. (1995). Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '95*, pages 51–58, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.
- [Renear and Palmer, 2009] Renear, A. H. and Palmer, C. L. (2009). Strategic reading, ontologies, and the future of scientific publishing. *Science*, 325:828–832.

- [Resnick and Varian, 1997] Resnick, P. and Varian, H. R. (1997). Recommender systems. *Commun. ACM*, 40(3):56–58.
- [Schmidt and Bannon, 1992] Schmidt, K. and Bannon, L. (1992). Taking csw seriously. *Computer Supported Cooperative Work (CSCW)*, 1:7–40.
- [Schütz and Luckmann, 1973] Schütz, A. and Luckmann, T. (1973). *The structure of the life world*. Northwestern University Press.
- [Simon, 1971] Simon, H. A. (1971). *Computers, Communication, and the Public Interest*, chapter Designing Organizations for an Information Rich World, pages 40–41. The Johns Hopkins Press.
- [Star and Griesemer, 1989] Star, S. and Griesemer, J. (1989). Institutional Ecology, Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3):387.
- [Strauss, 1978] Strauss, A. (1978). A social world perspective. *Studies in symbolic interaction*, 1(1):119–128.
- [Strauss and Corbin, 1998] Strauss, A. L. and Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications.
- [Suchman, 1987] Suchman, L. A. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press, New York, NY, USA.
- [Teevan et al., 2004] Teevan, J., Alvarado, C., Ackerman, M. S., and Karger, D. R. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422, Vienna, Austria. ACM.
- [Terveen et al., 1997] Terveen, L., Hill, W., Amento, B., McDonald, D., and Creter, J. (1997). Phoaks: a system for sharing recommendations. *Commun. ACM*, 40(3):59–62.
- [Wenger, 1999] Wenger, E. (1999). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.
- [White and Roth, 2009] White, R. and Roth, R. (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–98.
- [Wilson et al., 2010] Wilson, M., Kules, B., scharaefel, m. c., and Shneiderman, B. (2010). From keyword search to exploration: Designing future search interfaces for the web. *Foundations and Trends in Web Science*, 2(1):1–97.

- [Wittgenstein, 1960] Wittgenstein, L. (1960). *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigations'*. Harper Torchbooks.
- [Wittgenstein, 1980] Wittgenstein, L. (1980). *Remarks on the philosophy of psychology*, volume 1. Oxford: Blackwell.
- [Wittgenstein, 1996] Wittgenstein, L. (1996). *Last writings on the philosophy of psychology*, volume 1. The University of Chicago Press.
- [Zellweger, 1989] Zellweger, P. T. (1989). Scripted documents: a hypermedia path mechanism. *HYPertext '89: Proceedings of the second annual ACM conference on Hypertext*, pages 1–14.

# Apêndice A

## Materiais

### A.1 Estudo: Cap. 2

Questionário usado como guia para as entrevistas semi-estruturadas do estudo:

1. Tell me about the last time that you easily found a document you knew to be stored on your desktop.
  - (a) Can you reproduce the steps you performed to find it?
  - (b) Are these part of your typical steps to find documents?
2. Tell me about the last time that you used a very atypical strategy to find again a document you knew to be stored on your desktop.
  - (a) Can you reproduce the steps you made?
  - (b) Are these part of your typical steps to find documents?
3. Tell me about a time that you had problems to find again a document.
  - (a) What document was it?
  - (b) Did you manage to find it?
  - (c) Can you reproduce the steps you made to find it?
  - (d) Are these part of your typical steps to find documents?
4. Have you ever used the Web to find a document without relying on any desktop search, even though you knew that the document needed was already stored on your machine?
  - (a) What document was it?

- (b) Can you reproduce the steps you performed to find it?
  - (c) Are these part of your typical steps to find documents?
  - (d) Can you give me another example of a similar situation?
5. Tell me about a time that you needed multiple strategies find again a document.
- (a) What document was it?
  - (b) Can you reproduce the steps you made to find it?
  - (c) Is it common for you to use multiple strategies to search again for documents?
  - (d) Can you give me another example?
6. Are you satisfied with your strategies to search for documents you already have stored on your machine?
- (a) Give me an example of a strategy that you are satisfied with.
  - (b) Why does this particular strategy work for you?
  - (c) Now give me an example in which you are unsatisfied.
  - (d) Why does this particular strategy do not work for you?
7. Are you satisfied with the tools you use to search for documents?
- (a) Give me an example of a tool that you are satisfied with.
  - (b) Why does this particular tool work for you?
  - (c) Now give me an example of a tool that you are unsatisfied.
  - (d) Why does this particular tool do not work for you?

## A.2 Estudo: Cap. 3

### A.2.1 Material para o *survey*

Quais *websites* você utiliza para busca de artigos científicos? Por favor, marque apenas as opções que você utiliza, indicando a sua frequência de uso. Não é necessário marcar todas as opções para submeter. Sua participação é anônima e por isso solicitamos que você responda apenas uma vez.

Caso você utilize um outro *website* ou mecanismo de busca não listado abaixo, por favor adicione o nome do mesmo no campo de texto situado no final desse *survey*. Você também pode adicionar livremente comentários que julgar relevantes para a pesquisa.

Obrigado pela sua participação!

Sistema	1 (pouco)	2	3	4	5 (muito)
Google Scholar	<input type="radio"/>				
ACM Portal	<input type="radio"/>				
IEEE Xplore	<input type="radio"/>				
SpringerLink	<input type="radio"/>				
ScienceDirect	<input type="radio"/>				
Scopus	<input type="radio"/>				
ISI Web of Knowledge	<input type="radio"/>				
IngentaConnect	<input type="radio"/>				
JSTOR	<input type="radio"/>				

Outros

### A.2.2 Material para entrevista semi-estruturada

Questionário usado como guia para as entrevistas semi-estruturadas do estudo:

1. Por que você realizou essa busca?
2. Você encontrou o que estava procurando?
3. Como o artigo buscado influenciou ou deveria ter influenciado no seu trabalho?
4. O quão familiar a você é o assunto relacionado ao artigo?
5. Por quanto tempo você vem investigando o assunto?
  - (a) Você poderia estimar quantas vezes você já realizou buscas no assunto?
  - (b) Se você pudesse contar uma história sobre sua relação com esse tópico, como seria sua história?
6. A estratégia de busca que você utilizou é típica pra você?
7. Você se baseou em algum outro documento para realizar essa busca? Notas em papel, notas digitais, outros artigos científicos, referências na Web?
  - (a) Se sim, pedir para ver o(s) documento(s).
8. Você já tinha esse artigo armazenado no seu computador?
  - (a) Se sim, por que você preferiu utilizar a Web para buscá-lo novamente?
  - (b) Como ele estava armazenado?
  - (c) O modo como você o armazenou teve alguma influência na sua decisão em utilizar a Web?
9. O artigo buscado serve para [motivação dada pelo participante]. Você está colaborando com alguém nesse projeto?
  - (a) Se sim, como vocês estão compartilhando informações? Que ferramentas são utilizadas?
  - (b) Como essas ferramentas lhe ajudam para coordenar suas atividades com outros colaboradores?
  - (c) O que poderia ser modificado no processo e nas ferramentas de colaboração que facilitaria o seu trabalho?

## Apêndice B

### Resultados do *survey* (Cap. 3)

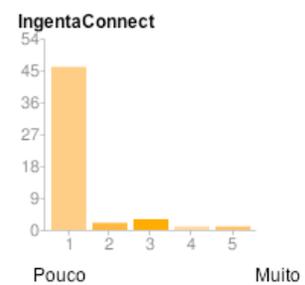
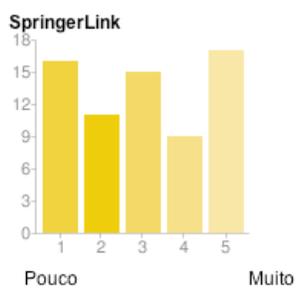
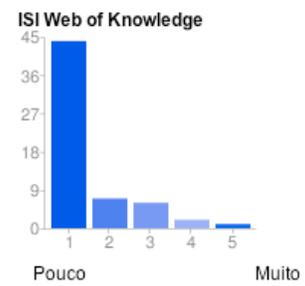
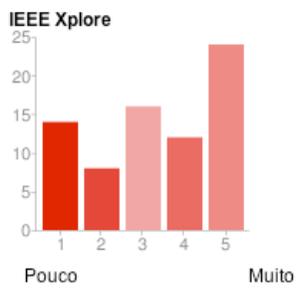
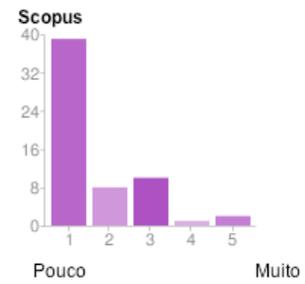
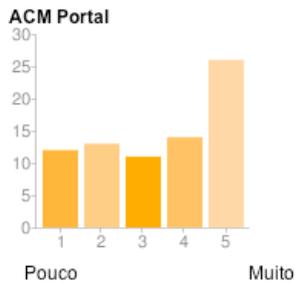
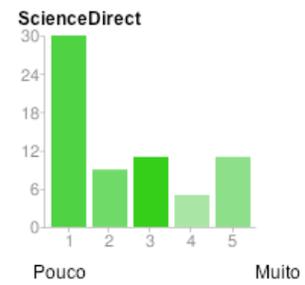
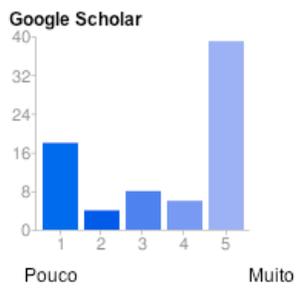


Figura B.1: Utilização de cada sistema de busca ou biblioteca digital, de acordo com a Escala Likert.