

Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Engenharia e Gestão
do Conhecimento

**Um Modelo para Recuperação e Busca de
Informação Baseado em Ontologia e no
Círculo Hermenêutico**

Fabiano Duarte Beppler

Orientador: Roberto C. S. Pacheco, Dr.
Co-Orientador: Frederico T. Fonseca, Dr.

Florianópolis
2008

Fabiano Duarte Beppler

UM MODELO PARA RECUPERAÇÃO E BUSCA DE
INFORMAÇÃO BASEADO EM ONTOLOGIA E NO CÍRCULO
HERMENÊUTICO

Esta tese foi julgada e aprovada para a obtenção do título de Doutor em Engenharia do Conhecimento no Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina.

Florianópolis, 15 de dezembro de 2008.

Prof. Roberto Carlos dos Santos Pacheco, Dr.
Coordenador do programa

Roberto Carlos dos Santos Pacheco, Dr.
Orientador

Frederico Torres Fonseca, Ph.D.
Co-Orientador

Ronaldo dos Santos Mello, Dr.
Moderador

Gentil José de Lucena Filho, Dr.
Examinador Externo

Alexandre Leopoldo Gonçalves, Dr.
Membro

José Leomar Todesco, Dr.
Membro

À Leticia

AGRADECIMENTOS

Ao meu orientador Roberto C. S. Pacheco, que contribuiu para o desenvolvimento do trabalho e para minha formação como pessoa e pesquisador.

Ao meu co-orientador Fred Fonseca, que me ajudou a conceber e desenvolver esta pesquisa e abriu as portas para que fosse possível o doutorado sanduíche.

Aos professores José L. Todesco, Ronaldo S. Mello e Gentil J. Lucena pelas discussões e contribuições sobre o tema.

Aos grandes amigos do Instituto Stela, em especial ao Alexandre Gonçalves, Denilson Sell e Vinícius Kern, pelas nossas discussões em assuntos relacionados a esta pesquisa, e à Sandra R. Martins pelo auxílio na revisão deste trabalho.

Ao Andrey Soares e sua esposa Cris que foram grandes amigos durante minha estada em *State College* (EUA).

Ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (EGC), pela oportunidade de ser um dos desbravadores do curso.

À *Penn State University*, especialmente ao *College of Information Sciences and Technology*, pela oportunidade de passar um ano e quatro meses em contato com grandes pesquisadores.

Ao Instituto Stela, instituição à qual me dedico desde 1999 e que me proporcionou aprendizado e crescimento como ser humano e profissional.

À minha família, que mesmo acompanhando a minha luta de longe, foi fundamental para o meu triunfo.

E, em especial, à minha esposa Leticia, pela imensa compreensão, carinho, apoio e incentivo.

“And how will you enquire, Socrates, into that which you do not know? What will you put forth as the subject of enquiry? And if you find what you want, how will you ever know that this is the thing which you did not know?”

Meno, 400 a.C.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	CARACTERIZAÇÃO DO PROBLEMA	4
1.2	OBJETIVOS DO TRABALHO	5
1.2.1	<i>Objetivo geral</i>	5
1.2.2	<i>Objetivos específicos</i>	5
1.3	ESCOPO DO PROBLEMA NA ENGENHARIA DO CONHECIMENTO	6
1.4	JUSTIFICATIVA	7
1.5	INEDITISMO DO TRABALHO	9
1.6	DELIMITAÇÃO DE ESCOPO	10
1.7	METODOLOGIA DE PESQUISA	11
1.8	ORGANIZAÇÃO DA TESE	13
2	REVISÃO DA LITERATURA	15
2.1	INTRODUÇÃO	15
2.2	RECUPERAÇÃO DE INFORMAÇÃO	15
2.2.1	<i>Modelos de recuperação de informação</i>	17
2.2.2	<i>Interação em recuperação de informação</i>	21
2.3	INFORMATION SEEKING	23
2.3.1	<i>Modelos conceituais de information seeking</i>	24
2.3.2	<i>Information seeking e recuperação de informação</i>	31
2.4	ONTOLOGIA	32
2.4.1	<i>Tipos de ontologias</i>	33
2.4.2	<i>Inferência e raciocínio</i>	34
2.4.3	<i>Importância de ontologias</i>	36
2.5	CÍRCULO HERMENÊUTICO	38
2.6	CONSIDERAÇÕES FINAIS	41
3	MODELO PROPOSTO	42
3.1	INTRODUÇÃO	42
3.2	DESCRIÇÃO DO MODELO	43
3.2.1	<i>Atividades cognitivas</i>	44
3.2.2	<i>Camada de suporte</i>	46
3.2.3	<i>Camada intermediária</i>	47
3.3	A METÁFORA DO CÍRCULO HERMENÊUTICO APLICADA AO PROCESSO DE BUSCA POR INFORMAÇÃO	48
3.4	TIPOS DE INTERAÇÃO DISCUTIDAS NO ÂMBITO DO MODELO PROPOSTO	50
3.5	CONSIDERAÇÕES FINAIS	53
4	FRAMEWORK HERMENEUS	55

4.1	INTRODUÇÃO.....	55
4.2	ARQUITETURA DO HERMENEUS.....	56
4.2.1	<i>Módulo de indexação</i>	57
4.2.2	<i>Módulo de recuperação</i>	59
4.2.3	<i>Módulo de apresentação</i>	61
4.2.4	<i>Módulo de inferência</i>	63
4.3	ORGANIZAÇÃO DA ARQUITETURA.....	65
4.4	CONFIGURAÇÃO DO FRAMEWORK HERMENEUS	66
4.5	CONSIDERAÇÕES FINAIS.....	67
5	DEMONSTRAÇÃO DE VIABILIDADE E ANÁLISE COMPARATIVA	69
5.1	APRESENTAÇÃO DO PROTÓTIPO.....	69
5.1.1	<i>Exemplo de utilização do protótipo</i>	75
5.2	A UTILIZAÇÃO DE ONTOLOGIA COMO SUPORTE AO DESENVOLVIMENTO DE SISTEMAS DE BUSCA INTERATIVOS	83
5.2.1	<i>Construção de sistemas de busca baseados em ontologia</i>	83
5.2.2	<i>Extração de informação adicional</i>	85
5.2.3	<i>Interação gráfica com a ontologia</i>	86
5.2.4	<i>Índice semântico comparado ao índice invertido tradicional</i>	87
5.2.5	<i>Comparação da busca semântica com abordagens similares</i>	89
5.3	HERMENEUS COMO TRADUTOR DA INFORMAÇÃO REQUERIDA PELOS USUÁRIOS...96	
5.4	COMPARAÇÃO DO MODELO PROPOSTO COM OUTROS MODELOS DE RBI	98
5.4.1	<i>Modelo proposto e o modelo estratificado de Saracevic (1997)</i>	98
5.4.2	<i>Modelo proposto e o framework integrado de Ingwersen (1992) e revisto por Ingwersen e Järvelin (2005b)</i>	99
5.4.3	<i>Modelo proposto e o modelo episódico de Belkin (1993)</i>	101
5.4.4	<i>Análise comparativa entre o modelo proposto e os demais modelos</i> ...	103
5.5	CONSIDERAÇÕES FINAIS.....	108
6	CONCLUSÕES E TRABALHOS FUTUROS.....	109
6.1	CONCLUSÕES.....	109
6.2	TRABALHOS FUTUROS	110
7	REFERÊNCIAS BIBLIOGRÁFICAS.....	112

LISTA DE FIGURAS

Figura 1. Metodologia aplicada para o desenvolvimento do trabalho	13
Figura 2 - Elementos do modelo estratificado de interação em RI proposto por Saracevic (1997)	26
Figura 3 - Modelo interativo de RBI proposto por Ingwersen e Järvelin (2005a)	28
Figura 4 - Modelo de comportamento de informação (Wilson, 1999)	32
Figura 5 - Círculo hermenêutico	39
Figura 6 - Modelo proposto para RBI	43
Figura 7 – Exemplificação da metáfora do círculo hermenêutico aplicado ao processo de busca	50
Figura 8 - Especialista de um domínio configura o Hermeneus.....	57
Figura 9 - Índice invertido com informação semântica	58
Figura 10 - Exemplo: a partir de uma ontologia um índice para cada conceito é gerado	59
Figura 11 - Descrição em EBNF de uma consulta aplicada em uma busca	60
Figura 12 - Exemplo de ontologia de família com relacionamentos diretos e inferidos	63
Figura 13 - Exemplo de regras de nível de ontologia (esquerda) e regras de nível de instância	64
Figura 14 – Organização da arquitetura de módulos e componentes do framework.....	65
Figura 15 - Exemplo de citação disponível no site do JASIST	69
Figura 16. Ontologia: classes e suas relações.....	70
Figura 17 - Listagem contendo a ontologia em OWL	71
Figura 18. Exemplo de regra de nível de ontologia (parte superior) e de instâncias (parte inferior).....	72
Figura 19 - Tela inicial do protótipo	73
Figura 20 - Tela do protótipo que apresenta os resultados de uma busca	74
Figura 21 - Exemplo de busca pelo termo " <i>seeking</i> ".....	76
Figura 22 - Tela resultante da busca pelo termo " <i>seeking</i> "	77
Figura 23 - Tela resultante da busca por " <i>keyword: web search; seeking</i> "	78
Figura 24 - Tela que mostra informações adicionais sobre uma instância	80
Figura 25 - Tela resultante da busca por " <i>author:Amanda Spink</i> "	81
Figura 26 - Tela que apresenta a busca por " <i>author:Amanda Spink</i> " com <i>paper</i> como <i>conceito central</i>	82
Figura 27 - Exemplo de ontologia e instâncias para comparação do índice semântico com o índice invertido tradicional.....	88
Figura 28 - Exemplo de estrutura do índice invertido tradicional	88
Figura 29 - Exemplo da estrutura do índice semântico	89

LISTA DE TABELAS

Tabela 1 - Componentes da hermenêutica relacionados com os componentes do modelo	49
Tabela 2 - Exemplos de métodos de interação inerente a um processo de busca...	51
Tabela 3 - Exemplos de consultas válidas e não válidas	61
Tabela 4 - Configurações requeridas pelo framework Hermeneus	67
Tabela 5 - Classes da ontologia e suas respectivas propriedades	70
Tabela 6 - Exemplo de informação adicional	86

LISTA DE ABREVIATURAS

IA	Inteligência Artificial
JASIST	<i>Journal of the American Society for Information Science and Technology</i>
LSI	<i>Latent Semantic Indexing</i>
NER	<i>Named Entity Recognition</i>
OWL	<i>Web Ontology Language</i>
RBI	Recuperação e Busca de Informação
RDF	<i>Resource Description Framework</i>
RDQL	<i>RDF Data Query Language</i>
RI	<i>Recuperação de informação</i>
RII	Recuperação de Informação Interativa
RuleML	<i>Rule Markup Language</i>
SPARQL	<i>Simple Protocol and RDF Query Language</i>
SVD	<i>Singular Value Decomposition</i>
SWRL	<i>Semantic Web Rule Language</i>

RESUMO

BEPPLER, Fabiano Duarte. **Um modelo para recuperação e busca de informação baseado em ontologia e no círculo hermenêutico**, 2008. Tese de doutorado em Engenharia e Gestão do Conhecimento – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, UFSC, Florianópolis.

Os sistemas de busca, mesmo com avanços quanto à organização da informação e taxas de precisão e recuperação, ainda são baseados apenas na comparação de termos de busca, que ignora o possível aprendizado que existe nas iterações subseqüentes à primeira busca. Com base nesta constatação, propõe-se um modelo de recuperação e busca de informação com base nas teorias referentes ao círculo hermenêutico, *information seeking*, recuperação de informação e ontologias. Uma ontologia de domínio é usada para a construção de indexadores da base de conhecimento e para viabilizar a metáfora do círculo hermenêutico, ou seja, permitir que usuários naveguem nos conceitos da ontologia e nas instâncias desses conceitos em uma forma de *vai e vem*. A demonstração de viabilidade do modelo é realizada por meio do framework Hermeneus, que dá suporte à construção de sistemas de busca que alternam diferentes tipos de estratégias para o refinamento das buscas, e pela comparação com outros modelos de recuperação e busca de informação. Hermeneus atua como um intermediário entre o usuário e a informação requerida, conectando as questões que o usuário tem em mente às respostas que são encontradas nas iterações de busca com contextualização em ontologia. Dessa forma, o modelo proporciona que a dinamicidade requerida por sistemas de busca seja alcançada, pois em um processo de busca o significado está sempre sendo criado e recriado.

Palavras-chave: recuperação e busca de informação; ontologia; círculo hermenêutico.

ABSTRACT

BEPPLER, Fabiano Duarte. **Um modelo para recuperação e busca de informação baseado em ontologia e no círculo hermenêutico (*an information seeking and retrieval model based on ontology and hermeneutic circle*)**, 2008. Tese de doutorado em Engenharia e Gestão do Conhecimento – Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, UFSC, Florianópolis.

In spite of advances in information retrieval, search systems are still based only on the matching of query terms and ignore the learning of interactions after the first search. Taking into account such statement, we propose an information seeking and retrieval model based on the hermeneutic circle, information seeking, information retrieval, and ontology theories. A domain ontology is used to build the indexes of a knowledge base and to implement the metaphor of the hermeneutic circle, i.e., users are allowed to navigate in the ontology concepts and in the instances of these concepts in a back and forth way. In order to show the viability of our model, we conceive a framework called Hermeneus, which supports the construction of search systems that enable different kinds of strategies to refine searches, and compare our model to other information seeking and retrieval models. Hermeneus works as an intermediary that facilitates the user to move from the initial state of information need to the goal state of resolution. Our system intends to be the bridge between the user's question and the answer to be found while she/he interacts with the ontology concepts and their instances. Therefore, because the meaning is always being created and recreated in the searching process, our model allows building dynamic and interactive search systems.

Keywords: information seeking and retrieval; ontology; hermeneutic circle.

1 INTRODUÇÃO

A área de recuperação de informação (RI) tem progredido consideravelmente nos últimos anos (Witten, Moffat *et al.*, 1999; Lin e Demner-Fushman, 2006). Entretanto, quando se busca informação utilizando sistemas de RI, os usuários ainda têm dificuldade, o que resulta na necessidade de despende tempo considerável na tentativa de localizar a informação requerida (Lai e Soh, 2004; Allan, Carterette *et al.*, 2005). Alguns dos problemas comuns relacionados a essa dificuldade se referem à composição inadequada de consultas, à grande quantidade de informação resultante a partir de uma busca e à falta de interatividade que sistemas de busca oferecem. Algumas pesquisas na área de comportamento humano mostram que usuários quando estão procurando informação são preguiçosos para formular consultas, chegando a uma média de apenas 2,2 termos por consulta (Baeza-Yates e Ribeiro-Neto, 1999), além de raramente utilizar expressões booleanas (Spink, Wolfram *et al.*, 2001). Além disso, a maioria dos usuários analisa apenas os primeiros resultados, isto é, os usuários têm baixa tolerância a investigar mais detalhadamente a lista de documentos recuperados.

A área de pesquisa de *information seeking* se preocupa em como obter informação, seja ela a partir de seres humanos ou de sistemas computacionais. De acordo com Marchionini (1989), um processo de *information seeking* “inclui reconhecer e interpretar a informação requerida, estabelecer uma estratégia de busca, conduzir uma busca, analisar e avaliar os resultados e, caso necessário, interagir através de todo o processo novamente”. Dessa forma, para que sistemas de busca atendam aos requisitos definidos pela área de *information seeking* é necessário que tais sistemas sejam interativos de maneira a permitir buscas mais fáceis, rápidas e intuitivas (Järvelin e Ingwersen, 2004; Ingwersen e Järvelin, 2005b) – mais fáceis porque os usuários querem obter resultados com o mínimo de esforço possível; alto desempenho porque em sua maioria os usuários são impacientes, principalmente na Internet (Huberman e Lukose, 1997); e, também, esses sistemas precisam ser intuitivos porque os usuários não querem despende muito tempo tentando aprender como executar tarefas mais complexas para obter melhores resultados (Hearst, Elliott *et al.*, 2002).

Incerteza e ansiedade são características que fazem parte do processo de busca por informação (Kuhlthau, 1991; 1993). Usuários também enfrentam o problema de como descrever as suas necessidades. Essa inabilidade de expressar precisamente qual informação é re-

querida faz com que os usuários entrem em uma seqüência indeterminada de interações com um sistema de busca (Marchionini, 1989; Kuhlthau, 1993; Sonnenwald, Wildemuth *et al.*, 2001; Capra Iii e Pérez-Quiñones, 2005). Belkin e Croft (1992) descrevem essa “situação problemática” como um “estado anômalo de conhecimento que induz uma pessoa a engajar em um comportamento de *information seeking*”. De acordo com Ingwersen e Järvelin (2005a), o modelo cognitivo dos usuários pode ser dinâmico mas não auto-contido, isto é, independente. O conhecimento do usuário sofre alterações quando ele interage com a informação recuperada, o que reflete em alguma alteração no seu estado anômalo de conhecimento (Belkin, 1993).

Além disso, construir sistemas de busca não é uma tarefa fácil, visto que tal ação exige profundo conhecimento da área de RI. Os modelos de RI se preocupam basicamente com a organização da informação, ou com a visão lógica de documentos (Baeza-Yates e Ribeiro-Neto, 1999), de maneira a propiciar buscas mais rápidas e poderosas. Entretanto, esses modelos não oferecem mecanismos que podem melhorar a interatividade entre os usuários e um sistema de busca (Marchionini, 1989; Kuhlthau, 1993; Ingwersen e Järvelin, 2005b). Deve-se considerar, também, que para um sistema fornecer respostas mais precisas, o responsável por criar um sistema de busca deve ter conhecimento sobre as informações do domínio (Mcguinness, 1999; Hyvönen, Mäkelä *et al.*, 2005).

Considerando-se as questões discutidas previamente, propõe-se um modelo de recuperação e busca de informação (RBI). A área de *information seeking*, que serve de base para o modelo, sugere que o comportamento de seres humanos em um processo de busca, o qual não pode ser caracterizado como estático, mas sim dinâmico, pode ser facilitado por ferramentas interativas. Além disso, recorreu-se à Filosofia, especificamente ao círculo hermenêutico, no intuito de construir um modelo que esteja mais aproximado ao modo como seres humanos compreendem. O círculo hermenêutico é o círculo de interpretação que se estabelece entre aquele que interpreta e a realidade a ser interpretada – os seres humanos entendem um todo a partir de preconcepções sobre o significado das partes e suas inter-relações (Gadamer, 1989; Heidegger e Stambaugh, 1996). Um processo de busca está diretamente relacionado ao entendimento de um problema, e a cada interação com o sistema o usuário acessa novas informações que afetam o entendimento de suas necessidades e, conseqüentemente, aquilo que ele está buscando.

Para viabilizar que as características abordadas pelas áreas de *information seeking* e círculo hermenêutico fossem aplicadas de forma mais pragmática no modelo, foi utilizada a tecnologia de ontologias. Segundo Gruber (1993a), ontologia é “uma especificação formal e explícita de uma conceitualização compartilhada”. No modelo proposto, ontologia deve ser usada no âmbito da engenharia, para construção de mecanismos de RI mais inteligentes, e principalmente para enriquecer a interface com a construção de componentes mais interativos.

Baseado nas características discutidas pelo modelo propõe-se um framework para construção de um ambiente de busca interativo. Decidiu-se chamar o framework de *Hermeneus*, que em grego significa “interpretar” ou “traduzir” (Kingsley, 1993), seguindo Kuhlthau (1991; 1993), que considera que um sistema de busca é um intermediário que deve auxiliar o usuário na mudança de uma condição de incerteza para uma condição de solução. Hermeneus usa uma ontologia como guia para o desenvolvimento de um sistema de busca. Como um artefato da engenharia, ontologias podem ser usadas para validar componentes de sistemas de informação (Fonseca, 2007). Hermeneus é composto de quatro módulos responsáveis por indexar e recuperar instâncias de ontologia, inferir informação adicional e apresentar componentes visuais que facilitam a interação do usuário com o sistema de busca.

Como as instâncias de ontologia são indexadas e a ontologia é apresentada graficamente, os usuários podem formular consultas semânticas (e.g., “autor: pacheco”). O componente gráfico que apresenta a ontologia proporciona ao usuário a visualização do conhecimento do domínio, o que pode facilitar o entendimento de conceitos inerentes ao domínio e conseqüentemente auxiliar na composição de consultas mais precisas. Esse componente também pode ser utilizado para alterar a visualização da informação recuperada (i.e., instâncias de ontologia). Dessa forma, o usuário pode visualizar a informação recuperada em diferentes perspectivas, ou seja, as instâncias podem ser agrupadas de acordo com cada conceito disponível na ontologia. As instâncias recuperadas permitem a utilização de seu conteúdo pelos usuários, por meio de cliques de mouse, para refinar automaticamente a busca corrente. Com isso, os usuários ao identificarem uma informação relevante podem clicar sobre tal informação e automaticamente um novo conteúdo semântico é adicionado à consulta, o que produz um refinamento na busca.

Além disso, informação adicional pode ser visualizada para cada instância recuperada. O conteúdo dessa informação também pode ser utilizado pelos usuários para formular novas consultas ou refinar buscas. A extração de conhecimento adicional é possível porque a ontologia é descrita em uma linguagem formal que inclui axiomas para especificar relacionamentos entre conceitos (*e.g.*, João é pai de Maria e irmão de José; nesse caso pode ser inferido que José é tio de Maria e que Maria é sobrinha de José) (Decker, Erdmann *et al.*, 1999). Portanto, o *Hermeneus* pode também extrair informação adicional usando técnicas de inferência e regras previamente definidas e pode permitir que usuários visualizem e interajam com informação descrita apenas implicitamente na ontologia assim como nas suas instâncias. Com essas características disponíveis por meio de componentes interativos, visualiza-se o *Hermeneus* como uma ferramenta que permite aos usuários desenvolverem as suas perguntas enquanto em um processo de busca, além de encorajá-los a articular diferentes aspectos de suas necessidades.

Nas seções seguintes, descreve-se de forma mais detalhada a problemática e os objetivos que a pesquisa se propõe a investigar.

1.1 Caracterização do problema

Buscar informação é uma tarefa que seres humanos executam com certa frequência. Nesse processo de busca, os conceitos e as necessidades das pessoas mudam à medida que elas interagem com nova informação. Ou seja, um comportamento de busca é um processo, por natureza, dinâmico. Além disso, questões relacionadas à incerteza, dúvidas, ansiedade, confusão e relevância ajudam a tornar esse processo ainda mais complexo. A seguir são identificados os problemas discutidos anteriormente.

- ✚ Como criar um modelo de RBI que defina as características necessárias para concepção de um ambiente interativo que permita aos usuários desenvolverem as suas buscas de forma mais natural possível?
- ✚ Como a representação semântica de um domínio proporcionada por uma ontologia pode auxiliar a concepção de um modelo de RBI?
- ✚ Como definir um framework baseado no modelo proposto que possibilite a construção de sistemas de busca mais interativos?

- ✚ Como uma ontologia pode ajudar a engenharia de sistemas de busca e permitir que o conhecimento de um domínio possa ser utilizado diretamente em um processo de busca?

1.2 Objetivos do trabalho

1.2.1 Objetivo geral

O objetivo geral deste trabalho é apresentar um modelo de RBI, baseado em conceitos discutidos no âmbito do círculo hermenêutico, recuperação de informação, ontologia e *information seeking*, para subsidiar a construção de ambientes de busca interativos. Propõe-se, também, a concepção e o desenvolvimento de um framework que aplica as definições descritas pelo modelo.

1.2.2 Objetivos específicos

Visando-se atingir o objetivo principal, alguns objetivos específicos são requeridos, como os que se seguem.

- ✚ Definir um modelo de RBI baseado na teoria do círculo hermenêutico que utilize uma ontologia de domínio e viabilize a disponibilização de tais conceitos e as informações que os materializam por meio de componentes interativos;
- ✚ Conceber e desenvolver um framework baseado nas definições discutidas pelo modelo que possibilite a construção de sistemas de busca interativos;
- ✚ Utilizar uma ontologia no âmbito da engenharia de sistemas de busca que permita a construção automática dos mecanismos de indexação e recuperação requeridos por sistemas de RI;
- ✚ Criar componentes interativos baseados na ontologia que viabilizem a interação direta com o conhecimento do domínio;
- ✚ Permitir a composição de consultas que possam ser contextualizadas por meio da utilização dos conceitos definidos pela ontologia;
- ✚ Extrair e disponibilizar informação adicional a partir de inferências sobre a ontologia e sobre a sua base de conhecimento;

- ✚ Implementar um protótipo para demonstrar a viabilidade do framework e, por consequência, do modelo proposto.

1.3 Escopo do problema na Engenharia do Conhecimento

A nova engenharia do conhecimento, conforme descrevem Schreiber et al. (2000) e Studer et al. (2000), visa a construção de métodos e ferramentas que proporcionam a criação de técnicas avançadas que permitam que o conhecimento seja gerenciado e manipulado de maneira mais eficiente. De acordo com Schreiber et al. (2000), conhecimento é o conjunto completo de informações, dados, relações que levam as pessoas à tomada de decisão, à realização de tarefas e à criação de novas informações ou conhecimento. A escola cognitivista, proposta por Varela et al. (1991) e Krogh e Roos (1995), deixa claro que o “conhecimento é uma entidade universalmente representável que pode ser armazenada em computadores, bases de dados, arquivos e manuais”.

Uma ontologia, que é um dos componentes principais do modelo proposto, representa o conhecimento de um domínio, conforme descreve Gruber (1993a), e viabiliza a representação formal do conhecimento que, segundo Lai (2007) e Milton et al. (2006), é uma das características desejadas pela engenharia do conhecimento. A utilização da ontologia está em acordo, também, com o trabalho proposto por Shadbolt e Motta (1993), que afirmam que sistemas de conhecimento devem representar o conhecimento de um domínio independentemente da implementação. É importante enfatizar que ontologias vêm sendo utilizadas como componente principal em propostas para a construção de sistemas de conhecimento. Um exemplo é o trabalho descrito por Studer et al. (2000) que definem uma estrutura de três camadas para a nova engenharia do conhecimento utilizando a ontologia como um dos componentes centrais.

A organização e disponibilização de conhecimento são atividades que estão diretamente relacionadas com a gestão do conhecimento. Os trabalhadores da era do conhecimento, como denominam Schreiber et al. (2000), precisam cada vez mais ter condições de usufruir do conhecimento como subsídio para a tomada de decisões. Studer et al. (2000) destacam uma das tarefas relacionadas à engenharia e gestão do conhecimento como a possibilidade de disponibilizar ferramentas que permitam que o conhecimento possa ser explorado. Na presente proposta, além da engenharia de sistemas, que se beneficia com a utilização dos conceitos definidos pela nova engenharia do conhecimento, disponibilizam-se algumas características

que atingem diretamente os usuários de um sistema de conhecimento. O conhecimento é utilizado nesta proposta para o desenvolvimento de sistemas mais inteligentes (*i.e.*, no âmbito da engenharia) e ao mesmo tempo na concepção de componentes interativos mais elaborados que permitam aos usuários usufruir explicitamente desse conhecimento quando estão desenvolvendo um comportamento de busca.

Pode-se, também, relacionar as áreas de *information seeking* e círculo hermenêutico à engenharia do conhecimento, considerando que essas áreas oferecem suporte à construção de modelos que guiem o desenvolvimento de sistemas de conhecimento que estejam mais aptos à auxiliar os trabalhadores na era do conhecimento. O contexto da utilização dessas áreas está diretamente vinculado aos benefícios descritos por Schreiber et al. (2000) e Studer et al. (2000) discutidas anteriormente. Neste trabalho, as discussões inerentes a essas duas áreas de pesquisa forneceram subsídios para a definição de um modelo de RBI. O framework *Hermeneus*, que materializa o modelo proposto, tem como objetivo gerar sistemas de busca interativos, ou seja, sistemas de conhecimento que funcionem como intermediários entre trabalhadores e suas necessidades por informação e conhecimento.

A metáfora do círculo hermenêutico, descrita nesta pesquisa, pode também ser relacionada com a gestão do conhecimento. Uma organização, ao exercer a gestão do conhecimento, está em constante aprendizado, ou seja, a organização está inserida em um modelo cíclico de aprendizado (Schreiber, Akkermans *et al.*, 2000). Considera-se que o usuário, ao utilizar o Hermeneus, também desenvolve um comportamento semelhante caracterizado como um círculo de interpretação e entendimento. O usuário navega nos conceitos da ontologia e nas informações desses conceitos em uma forma de vai e vem. A cada interação o usuário adquire novas informações, reorganiza seu conhecimento anterior e, conseqüentemente, valida, ou seja, verifica e altera, caso necessário, suas estratégias e ações futuras. O Hermeneus, portanto, é uma ferramenta que proporciona que o conhecimento possa ser explorado de maneira mais eficiente e interativa.

1.4 Justificativa

Os modelos de RI têm evoluído consideravelmente nos últimos anos (Witten, Moffat *et al.*, 1999; Lin e Demner-Fushman, 2006), principalmente após o surgimento da Internet. Entretanto, usuários que tentam satisfazer as suas necessidades por informação ainda despendem um tempo considerável nesse processo (Lai e Soh, 2004; Allan, Carterette *et al.*, 2005). Esses

modelos, os chamados modelos tradicionais de RI (Baeza-Yates e Ribeiro-Neto, 1999; Witten, Moffat *et al.*, 1999; Greengrass, 2000) (*e.g.*, booleano, vetorial, probabilístico, difuso, etc.), preocupam-se com a organização da informação de modo a facilitar o acesso e prover informação mais acurada possível levando em consideração apenas os termos pertencentes a uma consulta. Oddy (1977) e Belkin *et al.* (1982) questionam a obrigatoriedade de se traduzir uma necessidade por meio de palavras de forma a permitir que um sistema de RI possa recuperar informação a partir de seus índices.

O problema é que esses modelos não incluem o usuário como parte integrante de modo a prover um ambiente onde seja possível haver um diálogo interativo entre usuários e um sistema. Bates (1989) faz a seguinte pergunta: por que um sistema de RI não oferece condições para que uma pessoa que esteja buscando informação possa expressar a sua necessidade de maneira mais natural ao invés de ter que representar sua necessidade por meio de uma consulta que o sistema possa interpretar?

Em um processo de recuperação que usa um modelo tradicional de RI, uma consulta é tratada como uma unidade unitária – representa a descrição de um problema em um determinado momento. Embora tal abordagem seja facilitadora do ponto de vista de sistemas de RI, buscas na vida real, em sua maioria, não funcionam dessa forma (Marchionini, 1997; Järvelin e Ingwersen, 2004; Ingwersen e Järvelin, 2005b). Na vida real, pessoas têm uma idéia geral sobre um determinado assunto e interagem com diferentes fontes de informação na tentativa de suprir uma necessidade. A cada interação novas idéias aparecem e conseqüentemente novas consultas são concebidas. De acordo com Bates (1989), a cada estágio de interação os usuários não estão apenas alterando os termos de uma consulta, eles estão em um processo de mudança contínua tanto das ações a serem tomadas, baseadas em um determinado resultado, quanto da necessidade em seu conjunto. Belkin (1993) deixa claro essa questão da dinamicidade inerente a um processo de busca citando que “a suposição de uma necessidade de informação estática é problemática do ponto de vista de interação, pois interação necessariamente implica em interpretação e interpretação necessariamente implica em mudança no interpretador”.

Para Rouse & Rouse (1984), “devido a mudanças nas necessidades dos usuários, os sistemas de RI devem ser suficientemente flexíveis de modo a permitir que usuários consigam adaptar o processo de busca por informação de acordo com suas atuais necessidades”.

Uma área de pesquisa que se preocupa com o processo de busca por informação na sua totalidade é *information seeking*. O processo de *information seeking*, conforme descreve Tissen (1991), é altamente exploratório, o que requer funcionalidades interativas bem elaboradas. Belkin (1993) e Kekäläinen e Järvelin (2002) sugerem que o problema de interfaces em sistemas de RI é construir métodos e formas que ofereçam suporte para diferentes tipos de interações que permitam desenvolver estratégias de *information seeking*.

É necessária a construção de modelos que visam tratar as questões previamente descritas visto que o processo de busca por informação é uma tarefa executada por seres humanos constantemente. De acordo com Saracevic (1997), não existe um modelo de interação que possa ser aceito universalmente como único. Cabe lembrar, ainda, que a quantidade de informação está aumentando exponencialmente (Greengrass, 2000; Kobayashi e Takeda, 2000), e modelos de RBI se tornam fundamentais no intuito de disponibilizar ambientes interativos que sejam facilitadores em um processo de busca.

Procura-se, neste trabalho, conceber um modelo de RBI mais pragmático embasado em conceitos discutidos pelas áreas de *information seeking* e círculo hermenêutico. As características debatidas por essas duas áreas são potencializadas por meio da utilização de ontologia, que é usada tanto no âmbito dos mecanismos requeridos por sistemas de RI quanto no âmbito da interação entre os usuários e os sistemas de busca. Também se propõe um framework que se fundamenta nas definições discutidas pelo modelo de forma a mostrar a aplicabilidade do modelo e munir atores que desejam construir sistemas de busca interativos.

1.5 Ineditismo do trabalho

A definição desta pesquisa foi baseada em trabalhos disponíveis na literatura, conforme referenciados no decorrer da tese. Entretanto, nota-se que ainda há necessidade de sistemas de busca que sejam mais adaptados ao modo como os seres humanos agem quando requerem informação. Ou seja, há falta de contextualização do conhecimento de um domínio juntamente com o processo de entendimento ou interpretação inerentes a um processo de busca que possui como facilitador um sistema de busca. A presente proposta visa contribuir nesse sentido por meio da definição de um modelo de RBI e a concepção de um framework baseado nesse modelo. A seguir, apresentam-se alguns aspectos inerentes ao trabalho que são considerados inéditos em relação à literatura pesquisada.

- ✚ A utilização das teorias referentes ao círculo hermenêutico e à ontologia juntamente com conceitos definidos pela área de *information seeking* para concepção de um modelo de RBI;
- ✚ A concepção e o desenvolvimento de um framework baseado no modelo que permite a geração de sistemas de busca interativos;
- ✚ A utilização de ontologia (*i.e.*, conhecimento do domínio) como um dos componentes principais do framework para auxiliar tanto a elaboração de mecanismos de RI mais inteligentes quanto a concepção de componentes interativos.

1.6 Delimitação de escopo

As áreas de recuperação de informação, *information seeking*, ontologia e círculo hermenêutico foram pesquisadas para conceber o modelo. Discutem-se algumas tarefas cognitivas que são descritas no decorrer da apresentação do modelo, relacionadas ao processo de compreensão ou entendimento de busca desenvolvido por seres humanos; contudo, a validação dessas tarefas não é foco deste estudo. Utilizam-se diferentes abordagens previamente publicadas que descrevem análises detalhadas sobre essas tarefas cognitivas no intuito de entender os conceitos que definem tais tarefas e de contextualizá-las na esfera do modelo proposto.

Valemo-nos de uma ontologia de domínio como um dos principais componentes do modelo. Conseqüentemente, o framework que materializa o modelo também utiliza uma ontologia. Entretanto, este trabalho não está focado em metodologias ou métodos para a criação de ontologias e em técnicas para construir bases de conhecimento através da população de instâncias de ontologia. Cabe ressaltar que, se a ontologia não for construída de maneira a representar a realidade inerente a um domínio, o framework irá utilizar os conceitos definidos pela ontologia e gerar um sistema de busca interativo que pode não estar adequado ao domínio – isso pode refletir em índices semânticos inconsistentes, respostas incoerentes e, conseqüentemente, em problemas na interação entre usuários e o sistema. Há uma série de trabalhos científicos que especificam metodologias e métodos especializados em fornecer subsídios para a construção de uma ontologia que possa refletir a realidade de um domínio (Shadbolt, Motta *et al.*, 1993; Uschold e Gruninger, 1996; López, Asunción *et al.*, 1999; Mcguinness, 1999; Noy e Mcguinness, 2001; Staab, Studer *et al.*, 2001; Corcho, Fernández-

López *et al.*, 2003; Pinto e Martins, 2004; Smith, 2004). Portanto, questões relacionadas ao desenvolvimento de uma ontologia estão fora do contexto do trabalho.

1.7 Metodologia de pesquisa

Visando à concepção e ao desenvolvimento da pesquisa, realizou-se uma pesquisa bibliográfica sobre o modo como seres humanos agem quando assumem um comportamento de busca. O estudo está focado em comportamento de busca aplicado ao uso de sistemas de busca. Também se recorreu à Filosofia, especificamente ao círculo hermenêutico, no intuito de obter um melhor entendimento sobre a forma com que seres humanos compreendem e interpretam uma nova informação. Assim, para que os objetivos sejam alcançados, o trabalho foi dividido nas etapas a seguir. Além disso, a Figura 1 apresenta, esquematicamente, a metodologia aplicada para o desenvolvimento do trabalho.

1. **Estudo sobre modelos de RI.** Nesta etapa, realiza-se um estudo sobre os modelos tradicionais de RI e de que forma tais modelos estão sendo aplicados para facilitar o processo de busca por informação desenvolvido por seres humanos. Também são analisadas as características que vêm sendo criticadas pela literatura da área sobre a forma com que esses modelos tradicionais de RI provêm suporte para que usuários consigam encontrar a informação requerida;
2. **Estudo sobre modelos de *information seeking*.** Nesta etapa, estudam-se os modelos de RBI analisando as suas principais características e fraquezas. Foi na literatura dessa área que se conseguiu obter um melhor entendimento sobre características inerentes ao processo de busca executado por seres humanos. Esta pesquisa auxiliou a verificar alguns recursos que não são tratados por modelos tradicionais de RI de modo a disponibilizar ambientes que estejam mais adequados à forma com que seres humanos agem no decorrer de um processo de busca por informação;
3. **Estudo sobre ontologia.** A área de ontologia é pesquisada com vistas a encontrar definições que facilitassem a concepção do modelo, pois o conhecimento de um domínio é um dos principais componentes da proposta. Também se buscou informações sobre a forma com que uma ontologia pode ser aplica-

da no âmbito da engenharia de sistemas, especificamente, na engenharia de sistemas de busca;

4. **Estudo sobre o círculo hermenêutico.** O estudo do círculo hermenêutico foi abordado por se tratar de uma área que possibilita um melhor entendimento de como os seres humanos compreendem ou interpretam novas informações. O foco, ao se estudar essa área, foi buscar sugestões que permitissem juntamente com as demais áreas descritas anteriormente, conceber o modelo proposto com características mais direcionadas ao modo como seres humanos interpretam e buscam novas informações;
5. **Definição do modelo de RBI.** A partir dos estudos descritos previamente, trabalhou-se na definição do modelo de RBI. Vários conceitos discutidos pelas áreas abordadas nos tópicos anteriores foram utilizados e serviram de base para a definição de cada componente do modelo. Visou-se conceber um modelo mais pragmático que considerasse o conhecimento de um domínio como integrante de destaque no modelo;
6. **Definição e desenvolvimento de um framework baseado no modelo.** Com a definição do modelo, decidiu-se conceber e desenvolver um framework que aplicasse as características apresentadas pelo modelo. Destaca-se a utilização da ontologia para representar o conhecimento de um domínio e possibilitar a sua utilização tanto para auxiliar usuários por meio de interação direta com esse conhecimento, como também para construir mecanismos mais inteligentes de sistemas de busca;
7. **Demonstração de viabilidade e análise comparativa.** Nesta etapa, desenvolveu-se um protótipo aplicando o framework e por conseqüência o modelo proposto. Uma ontologia referente ao domínio de artigos científicos foi definida e o framework, de modo a gerar um sistema de busca interativo, foi configurado. Também é feita uma análise de como uma ontologia pode ser usada para melhorar sistemas de busca, desde a extração de informação adicional até a interatividade gráfica. Uma comparação entre o índice semântico e o índice invertido também é apresentada. No contexto da hermenêutica, detalha-se algumas discussões de como o framework proposto atua como um tradutor da

informação requerida por usuários. E, por fim, é feita uma comparação do modelo proposto em relação a outros modelos de RBI discutidos no capítulo sobre revisão de literatura;

8. **Conclusões e trabalhos futuros.** Apresentam-se as conclusões da tese e descrevem-se algumas sugestões de possíveis trabalhos futuros.

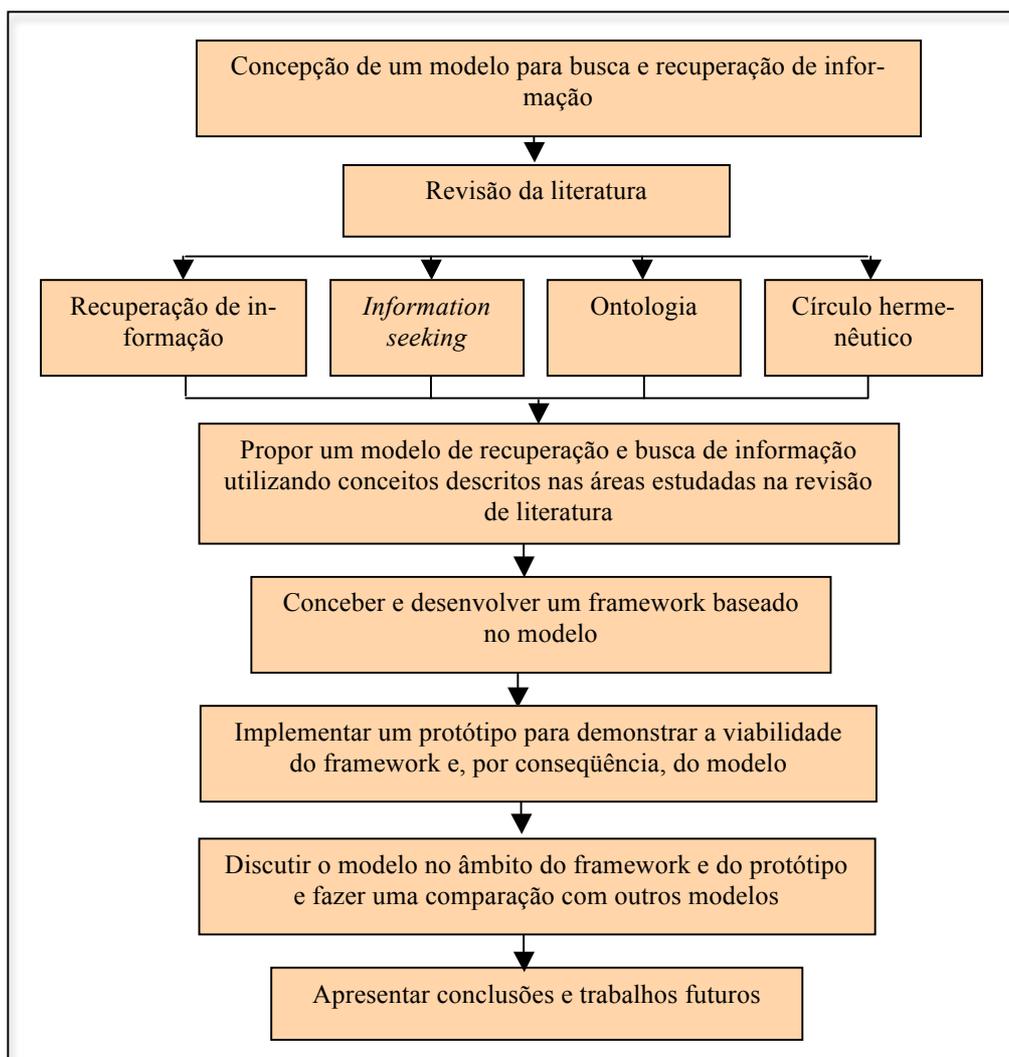


Figura 1. Metodologia aplicada para o desenvolvimento do trabalho

1.8 Organização da tese

Este trabalho é composto, além desta introdução, de mais cinco capítulos, como mostrado a seguir.

- ✚ Capítulo 2 - Revisão da literatura: neste capítulo descrevem-se as áreas de pesquisas (*i.e.*, RI, *information seeking*, ontologia e círculo hermenêutico) que serviram de embasamento para o trabalho;
- ✚ Capítulo 3 - Modelo proposto: neste capítulo apresenta-se o modelo de RBI;
- ✚ Capítulo 4 - Framework *Hermeneus*: neste capítulo apresenta-se o framework *Hermeneus*, que é a aplicação prática do modelo proposto;
- ✚ Capítulo 5 - Demonstração de viabilidade e análise comparativa: neste capítulo apresenta-se e discute-se um protótipo baseado no framework *Hermeneus* e, por consequência, no modelo. Também se discute o modelo proposto em comparação com outros modelos de RBI;
- ✚ Capítulo 6 - Conclusões e trabalhos futuros: neste capítulo destacam-se algumas conclusões sobre a pesquisa e descrevem-se sugestões para trabalhos futuros.

2 REVISÃO DA LITERATURA

2.1 Introdução

Vive-se uma era de transformação de como as pessoas interagem com informação e conhecimento. Os sistemas de informação propõem facilitar a gerência de informação para que as pessoas possam ser auxiliadas na solução de seus problemas. Contudo, com o aumento significativo da quantidade de informação disponível, principalmente após o advento da Internet, os sistemas de informação são requeridos para que forneçam alternativas que calçam os usuários a desenvolverem as suas necessidades de maneira mais eficiente e eficaz.

Um dos grandes problemas que ainda não possui uma solução definitiva se refere ao modo como encontrar somente a informação que se deseja. Há uma variedade de métodos e técnicas para recuperar informação, mas ainda não existe um sistema que consiga prover apenas as respostas que realmente possam ser utilizadas para solucionar uma ou mais necessidades. Além disso, como está descrito no trabalho de Teevan et al. (2004), haverá sempre o problema relacionado à forma como expressar uma necessidade por parte dos usuários, o que reflete na impossibilidade de se desenvolver um sistema que consiga apresentar respostas perfeitas. Entretanto, sistemas de busca devem prover ambientes em que os usuários consigam não somente buscar informações mas também desenvolver as suas questões ou dúvidas, *i.e.*, entender realmente o que eles próprios desejam.

As áreas de pesquisa descritas neste capítulo serviram de base para definição de um modelo de RBI. Além disso, um framework baseado no modelo foi concebido de forma a gerar sistemas de busca mais interativos. Um dos componentes do modelo é uma ontologia de domínio que oferece suporte para a construção de mecanismos mais inteligentes requeridos por ambientes de busca mais interativos e poderosos para RBI. A seguir, descreve-se as áreas de recuperação de informação, *information seeking*, ontologia e círculo hermenêutico, as quais estão envolvidas na realização desta pesquisa.

2.2 Recuperação de informação

Recuperação de informação (RI), de acordo com Salton (1968), “é a área de pesquisa que se preocupa com a estrutura, análise, organização, armazenamento, recuperação e busca de informação”. A representação e a organização dessa informação devem permitir que os usuários tenham fácil e rápido acesso à informação desejada. Kowalski (1997) cita que o principal

objetivo de um sistema de RI é minimizar a dificuldade do usuário em localizar a informação requisitada. Por essa dificuldade, pode-se entender o tempo gasto por um usuário e todos os passos necessários em um processo de busca até que a informação requerida possa ser acessada.

Entretanto, caracterizar uma necessidade por um usuário não é uma tarefa simples. Baeza-Yates e Ribeiro-Neto (1999) afirmam que “a recuperação efetiva de informação relevante é diretamente afetada pela tarefa do usuário e pela visão lógica dos documentos adotada por um sistema de recuperação”. A tarefa do usuário no âmbito da RI significa traduzir as necessidades dele em consultas que possam ser entendidas por sistemas de RI. A visão lógica dos documentos está relacionada ao modo como um documento é representado de forma a permitir que as buscas sejam possíveis. Normalmente, os documentos são representados por meio de índices formados pelas palavras, mais comumente conhecidas como termos no âmbito da RI, que compõem esses documentos (Rijsbergen, 1979).

A concepção de um sistema de RI se inicia, primeiramente, com a definição da fonte de informação, *i.e.* a relação de documentos a serem recuperados. Também é necessário determinar quais são as operações que podem ser executadas durante um processo de busca. Com essas definições, um desenvolvedor precisa estabelecer qual é a estrutura necessária, ou visão lógica dos documentos, que viabilize tais operações. Definida a visão lógica, o próximo passo é construir um índice que armazene os termos, contidos nos documentos. Segundo Baeza-Yates e Ribeiro-Neto (1999), “um índice é uma estrutura de dados crítica porque permite rápida busca sobre grandes volumes de dados”. A estrutura do índice também é definida pelo desenvolvedor, que analisa as operações a serem executadas sobre o índice, quais informações podem ser pesquisadas e como as informações são recuperadas. Com o índice criado, um processo de recuperação pode ser iniciado. Basicamente, esse processo começa com o usuário descrevendo a sua necessidade por meio de uma consulta, que é composta de termos. O sistema de RI, então, interpreta a consulta, traduzindo as possíveis operações em linguagem que possa ser aplicada sobre o índice. O resultado da aplicação da consulta sobre o índice retorna uma lista de documentos que são ordenados de acordo com algum critério, também definido pelo desenvolvedor, para então serem apresentados ao usuário.

A maior parte das pesquisas e do desenvolvimento de RI objetiva a melhoria da eficácia e eficiência da tarefa de recuperação (Rijsbergen, 1979). A eficiência é normalmente me-

dida em termos computacionais (*e.g.*, performance, tempo de CPU, etc.) enquanto que a eficácia, mesmo sendo subjetiva do ponto de vista do usuário (Kowalski, 1997), é comumente calculada por meio das medidas chamadas *precision* e *recall* (Rijsbergen, 1979; Baeza-Yates e Ribeiro-Neto, 1999; Witten, Moffat *et al.*, 1999). *Precision* é a relação do número de documentos relevantes recuperados do total do número de documentos recuperados. *Recall* é a relação do número de documentos relevantes recuperados sobre o número total de documentos relevantes. Essas medidas são usadas para avaliar o desempenho de um modelo de RI e para fazer a comparação de performance entre diferentes modelos. Nos tópicos a seguir descreve-se os principais modelos clássicos de RI, também conhecidos como tradicionais.

2.2.1 Modelos de recuperação de informação

Modelos de recuperação de informação formais têm sido a base da RI desde 1960. De acordo com Baeza-Yates e Ribeiro-Neto (1999), um modelo de RI é composto de: (a) um conjunto de visões lógicas, ou representações, de documentos em uma coleção; (b) um conjunto de visões lógicas, ou representações, da informação requerida por usuários; (c) um framework para modelar documentos, consultas e suas relações; e (d) uma função de ordenação que associa um número com uma consulta e um documento. Diferentes modelos foram e ainda estão sendo desenvolvidos, e descrevem aspectos inerentes à tarefa de recuperação, tais como estrutura e conteúdo do documento, consultas, necessidades de usuários e contexto no qual essa tarefa está inserida (Allan, Aslam *et al.*, 2002). Há uma variedade de modelos de RI (Korfhage, 1997) que oferecem diferentes estruturas e linguagens de busca e que disponibilizam especificidades que podem ser usadas em adequação à informação armazenada. Segundo Green-grass (2000), há duas categorias principais de modelos de RI: semânticos e estatísticos. Os modelos semânticos visam implementar análise semântica e sintática no intuito de “entender” um texto descrito em linguagem natural. Os modelos estatísticos atribuem medidas estatísticas que se referem à comparação entre uma consulta e um documento. A seguir são apresentados os modelos tradicionais mais comuns de RI.

2.2.1.1 Modelo booleano

Esse modelo é baseado na teoria dos conjuntos e na álgebra booleana (Korfhage, 1997; Baeza-Yates e Ribeiro-Neto, 1999). As consultas são descritas por meio de expressões booleanas usando conectores lógicos: AND, OR e NOT (*e.g.*, *a AND b OR c*). Os documentos são representados por meio de um conjunto de termos indexados. A recuperação de um documen-

to acontece se este responde verdadeiramente a uma expressão booleana, ou seja, a uma consulta. Como o resultado da recuperação de um documento é um valor binário, esse modelo não oferece um mecanismo de ordenação. A principal vantagem do modelo booleano se refere ao formalismo e à simplicidade. A principal desvantagem é a comparação exata entre consulta e documento, que pode levar à recuperação de poucos ou muitos documentos (Baeza-Yates e Ribeiro-Neto, 1999).

2.2.1.2 Modelo booleano estendido

O modelo booleano estendido diferencia-se do modelo booleano por usar diferentes operadores e por implementar uma função de ordenação (Lee, 1994; Baeza-Yates e Ribeiro-Neto, 1999; Greengrass, 2000). Esse modelo associa pesos aos termos de cada documento. O modelo booleano atribui valor 1 (um) aos termos de uma consulta, de acordo com a expressão lógica, que estão no documento e 0 (zero) caso contrário. Já o modelo estendido atribui valores que variam de 0 (zero) até 1 (um), os quais correspondem à estimativa de comparação de uma expressão lógica com um documento. Lee (1994) cita que modelos booleanos estendidos (*e.g.*, conjuntos difusos, *Waller-Kraft*, *Paice*, *P-Norm* e *Infinite-One*) têm sido propostos por oferecerem cálculos de similaridade mais robustos entre uma consulta e um documento.

2.2.1.3 Modelo vetorial

No modelo vetorial cada documento é representado por um vetor ou por uma lista de termos ordenados, por exemplo, pela frequência do termo no documento, em vez de apenas um conjunto de termos, como utiliza o modelo booleano (Salton, Wong *et al.*, 1975; Kowalski, 1997; Baeza-Yates e Ribeiro-Neto, 1999; Greengrass, 2000). A diferença entre o modelo booleano e o modelo vetorial está, além da representação dos termos, na determinação da medida de similaridade entre um documento e uma consulta. O propósito do modelo vetorial é avaliar o grau de similaridade de um documento em relação a uma consulta como uma correlação entre os vetores que representam o documento e a consulta (Baeza-Yates e Ribeiro-Neto, 1999; Witten, Moffat *et al.*, 1999). Essa correlação pode ser quantificada, por exemplo, pela medida do ângulo entre dois vetores, calculada por meio da fórmula do co-seno (Witten, Moffat *et al.*, 1999). Dessa forma, em vez de determinar se um documento é relevante ou não, o modelo vetorial ordena os documentos de acordo com o grau de similaridade com uma consulta.

2.2.1.4 Modelo probabilístico

O modelo probabilístico tenta resolver a questão de recuperar informação por meio de cálculos probabilísticos, usando para tal a teoria da probabilidade (Greengrass, 2000). Baeza-Yates e Ribeiro-Neto (1999) descrevem o modelo probabilístico considerando uma suposição: dados uma consulta e um documento em uma coleção, o modelo tenta calcular a probabilidade em que o usuário poderá achar o documento relevante. Para Rijsbergen (1979) e Salton (1989), o modelo probabilístico normalmente calcula a probabilidade condicional em que um determinado documento é visto sob uma ótica randômica em que esse documento é relevante a uma consulta. Tipicamente, consulta e documentos são representados por meio de um conjunto de termos, em que uma função probabilística calcula a probabilidade de ocorrência dos termos de uma consulta em documentos relevantes e não-relevantes. Essa função probabilística depende do modelo probabilístico a ser usado e de como os termos estão distribuídos entre os conjuntos de documentos relevantes e não-relevantes (Greengrass, 2000; Manning, Raghavan *et al.*, 2008).

2.2.1.5 Modelo difuso

O modelo difuso de RI é baseado na teoria dos conjuntos difusos criada por Zadeh (1965; 1983), a qual redefine o conceito clássico de pertinência de um elemento a um conjunto por meio da redefinição do conceito de função de pertinência do intervalo fechado e discreto $[0, 1] \in I$ para o intervalo contínuo $[0, 1] \in \mathbb{R}$. Ao generalizar a função característica dos conjuntos clássicos, a teoria dos conjuntos difusos fez com que todas as áreas que fundamentaram o seu desenvolvimento em teoria clássica de conjuntos pudessem ser generalizadas para uma teoria difusa. Esse é o caso da busca difusa, que generaliza a busca booleana. Para Baeza-Yates e Ribeiro-Neto (1999) esse modelo é uma extensão lógica do modelo booleano e possui uma função de ordenação cujo resultado da comparação entre um documento e uma consulta é aproximado. Entretanto, esse modelo tem sido criticado por gerar medidas de similaridades incorretas (Lee, 1994). Essa aproximação pode ser modelada por considerar que cada termo de uma consulta define um conjunto difuso e cada documento possui um grau de participação nesse conjunto (Baeza-Yates e Ribeiro-Neto, 1999). O procedimento para calcular a relevância de um documento em relação a uma consulta é similar ao adotado pelo modelo booleano – a diferença é que o cálculo é baseado em conjuntos difusos. A justificativa para a recuperação difusa de informação está no fato de que o sistema e freqüentemente o usuário não podem informar se um documento fornece toda a informação necessária ou não. Essa in-

certeza é modelada por meio de uma avaliação difusa do documento em relação a uma consulta (Korfhage, 1997).

2.2.1.6 Modelo de indexação semântica latente

A indexação semântica latente – LSI (do inglês *Latent Semantic Indexing*) é definida como uma técnica automática que analisa as co-ocorrências de termos em documentos textuais com vistas a descobrir relacionamentos latentes entre eles (Deerwester, Dumais *et al.*, 1990). Para identificar as relações semânticas, o LSI utiliza o modelo de Decomposição de Valores Singulares – SVD (do inglês *Singular Value Decomposition*) (Bartell, Cottrell *et al.*, 1992; Baeza-Yates e Ribeiro-Neto, 1999; Jurafsky e Martin, 2000). A estrutura de análise do LSI refere-se a uma matriz esparsa termo-documento. Essa matriz é analisada por SVD que a decompõe em três outras matrizes – (1) a primeira matriz possui colunas ortogonais e representa os termos; (2) a segunda, que também possui colunas ortogonais, representa os documentos; e (3) a terceira representa a matriz diagonal de valores singulares. É por meio do produto dessas três matrizes que os relacionamentos latentes são estabelecidos. No âmbito da RI, o SVD pode ser visto como uma técnica para derivar um conjunto de variáveis indexadas não-correlacionadas (Deerwester, Dumais *et al.*, 1990), em que cada termo e documento são representados por um vetor de pesos – o peso deve indicar a força da associação entre um termo e um documento. A definição de peso pode ocorrer por diferentes métodos, como, por exemplo, 0 e 1, indicando se um termo ocorre ou não no documento, ou um valor que indica a quantidade de ocorrências de um termo em um documento. Uma consulta é representada pela soma dos vetores dos termos que compõem a consulta. O conjunto de potenciais documentos é encontrado ao se calcular, por exemplo, o cosseno ou a distância do pseudo-documento (*i.e.*, os termos que formam a consulta) em relação ao conjunto total de documentos (Deerwester, Dumais *et al.*, 1990; Baeza-Yates e Ribeiro-Neto, 1999). Cabe ressaltar que LSI é um modelo que requer tempo elevado de processamento quando aplicado à grandes coleções de documentos (Ikehara, Murakami *et al.*, 2001) e não é um modelo semântico mesmo tendo a palavra semântica em seu nome – esse método é estatístico, pois tenta capturar dependências entre os termos que podem ter um significado semântico (Greengrass, 2000).

2.2.1.7 Modelo que utiliza processamento em linguagem natural

O modelo que usa processamento em linguagem natural pode ser caracterizado como modelo semântico, pois utiliza a estrutura e o significado de documentos textuais diretamente em vez

de utilizar apenas medidas estatísticas (Greengrass, 2000). Liddy (1998) classifica técnicas de processamento em linguagem natural nos seguintes níveis: fonológico, morfológico, léxico, sintático, semântico, discurso e pragmático. O nível fonológico se refere à interpretação de sons, fonemas; o nível morfológico se preocupa com a análise das diferentes formas de uma palavra (*e.g.*, prefixos, raiz e sufixos); o nível léxico está relacionado à análise da estrutura e ao significado puro da palavra; o nível sintático se preocupa em analisar a estrutura sintática de sentenças; o nível semântico tenta interpretar o significado de sentenças em vez de palavras individuais; o nível do discurso tenta interpretar a estrutura e o significado de unidades maiores (*e.g.*, parágrafo, documento completo, etc.); o nível pragmático é aquele no qual é aplicado conhecimento externo (*e.g.*, conhecimento de mundo, conhecimento de um determinado domínio, conhecimento sobre necessidades de usuários) tanto ao documento quanto à consulta. No processamento em linguagem natural, mesmo com a aplicação dessas etapas, a clareza da informação ainda pode estar implícita, sendo identificada por meio de deduções. Por exemplo, considerando-se a seguinte pergunta “*Você sabe a hora?*”, não significa que a resposta deve ser *sim* ou *não*. Mesmo com a aplicação dessas regras, é difícil determinar o que se quer buscar, pois uma consulta apresenta muito pouca informação, tornando a extração da informação relevante para a busca, muitas vezes, subjetiva (Korfhage, 1997). Além disso, técnicas de processamento em linguagem natural, conforme descreve Greengrass (2000), são raramente usadas em RI; tais técnicas são utilizadas mais comumente como suplemento a modelos estatísticos.

2.2.2 Interação em recuperação de informação

O processo de busca por informação, segundo Vakkari (2003), é cíclico, mas pode ser dividido nos seguintes componentes: o tipo de informação que se deseja buscar, a composição de consultas, táticas de busca, uso de ferramentas de busca e verificação da utilidade e relevância da informação encontrada. A busca por informação, segundo Toms (2002), que é normalmente realizada por meio de consultas e navegação, também é influenciada por situações e experiências relativas à interação com um sistema de informação. Ingwersen (1994) cita três características fundamentais sobre a importância da RI: (1) a incerteza está presente na interação em um sistema de RI associada com a interpretação tanto do usuário quanto do sistema; (2) pressuposições e intencionalidade entre as mensagens trocadas são vitais para a percepção e o entendimento de tais mensagens; e (3) recuperação de informação direta e real

executada pelos usuários. Um ambiente interativo permite que os usuários consigam tratar mais claramente questões de incerteza, utilizar pressuposições de maneira mais dinâmica e utilizar mais eficazmente um sistema de RI.

Segundo Saracevic (1997), “interatividade em RI é o diálogo entre os participantes – usuário e computador – por meio de uma interface, com o objetivo principal de influenciar o estado cognitivo do usuário para um uso efetivo da informação em conexão com uma aplicação”. Em um nível fundamental, recuperar informação é uma tarefa interativa (Savage-Knepshield e Belkin, 1999). Spink e Saracevic (1998) chegam a citar que “a interação se tornou a característica mais importante em RI”. Entretanto, ainda não há um entendimento completo sobre determinados aspectos de um processo interativo, apesar do número de estudos experimentais e teóricos já efetuados (Ingwersen, 1996; Saracevic, 1997; Spink e Saracevic, 1998; Ingwersen e Järvelin, 2005b).

Interação pode ser definida como uma seqüência de processos que ocorrem sobre vários níveis ou camadas e que estão de alguma forma conectados (Saracevic, 1997). Cada nível, segundo Spink e Saracevic (1998), envolve diferentes elementos e/ou processos específicos que, por sua vez, ainda estão divididos entre usuário e sistema. No lado do usuário, esses elementos são: fachada, cognitivo, situacional e emocional; já no lado do sistema, esses elementos são: fachada, engenharia, processo e conteúdo. “No nível de fachada, a interação é uma série de episódios no tempo no qual os usuários interagem através de uma interface com um sistema não somente para fazer buscas, mas também para engajar em um número de outros processos ou ‘coisas’ além de buscas, tais como, explorar os recursos de um sistema, navegar, visualizar resultados, efetuar feedback e assim por diante. Já os sistemas interagem com usuários com determinados processos no intuito de fornecer determinadas respostas a um episódio”. No nível cognitivo, os usuários interagem com a informação recuperada (*i.e.*, texto, imagens, dados e suas representações) como se fossem estruturas cognitivas, pois os usuários interpretam a informação obtida cognitivamente. No nível situacional os usuários se deparam com uma determinada situação ou problema o que gera uma necessidade por informação. No nível emocional, os usuários interagem com intenções e motivações associadas com sentimentos de satisfação, frustração, urgência, etc.

Durante um processo de busca, à medida que um usuário interage com o sistema, os aspectos relacionados à cognição, situação e emoção ou intenção são afetados e mudam cons-

tantemente, o que causa automaticamente uma redefinição do problema ou questão (Taylor, 1968; Belkin, 1980; Kelly e Belkin, 2002; Ingwersen e Järvelin, 2004). Kekäläinen e Järvelin (2002) citam que “a interação com necessidades dinâmicas pode ser vista como uma seqüência de ações onde em cada ação há algo que pode ser aprendido”. Dessa forma, à medida que a interação progride o nível de fachada também sofre alterações (*e.g.*, novos termos de busca são selecionados, outros são abandonados, táticas de busca são adaptadas, etc.), o que indica uma profunda ligação entre todos os níveis da interação. Um contexto de recuperação de informação interativa, de acordo com Ingwersen e Järvelin (2004), deve ser analisado em uma ampla perspectiva, num âmbito relacionado a comportamento de *information seeking*.

2.3 Information seeking

Information seeking é o processo de busca por informação (*e.g.*, em documentos textos, sistemas de informação ou até mesmo por meio de outras pessoas) que possa suprir uma necessidade sobre um determinado assunto (Marchionini, 1989, Marchionini, 1998 #188; Kuhlthau, 1991; Belkin, 1993). Belkin (1980) descreve o processo de *information seeking* como um estado anômalo de conhecimento (*i.e.*, uma lacuna de conhecimento). Esse processo ocorre especificamente quando uma pessoa se depara com a falta de conhecimento, por exemplo, para resolver um problema. Por necessidade, nesse caso, entende-se como uma condição motivacional interna que aciona pensamentos e ações com a intenção de atingir um objetivo (Taylor, 1968; Case, 2007).

Para Marchionini (1989), o processo de *information seeking* significa reconhecer e interpretar um problema, estabelecer um plano de busca, conduzir a busca, avaliar os resultados e, se necessário, repetir todo o processo novamente. A condição inicial de um processo de *information seeking*, de acordo com Wilson (2000) e Spink e Park (2006), é contextualizada assim como ocorre com uma falta de informação, e a construção desse processo evolui em estágios de acordo com certas variáveis, tais como, relevância e incerteza. Conforme aponta Kuhlthau (1991), o processo de *information seeking* deve incorporar três tipos de atividades: (1) física, que se refere às ações tomadas; (2) emocional, associada às experiências vividas; e (3) cognitiva, que está relacionada às funções mentais (*i.e.*, pensar, raciocinar e lembrar). Por meio de ações e interações entre essas atividades, uma pessoa que busca por informação consegue passar de uma condição de dúvida ou incerteza para uma condição de solução – o esta-

do de conhecimento de uma pessoa é dinâmico em vez de estático, sendo afetado durante todo um processo de *information seeking* (Kuhlthau, 1993).

Descrever ou especificar uma necessidade, principalmente logo após identificar a falta de conhecimento, está diretamente relacionado com experiências passadas e com o entendimento do problema. Além disso, a ansiedade, de acordo com Kuhlthau (1993), também faz parte de um processo de busca, o que faz com que sejam geradas incerteza e confusão. À medida que haja interação entre as atividades (*i.e.*, física, emocional e cognitiva), pertencentes ao processo de *information seeking*, o estado de conhecimento de uma pessoa é alterado e, portanto, as suas ações futuras são diretamente influenciadas. Dessa forma, conforme relatado por Belkin (1993), as alterações no estado de conhecimento de uma pessoa refletem automaticamente sobre o estado anômalo de conhecimento.

2.3.1 Modelos conceituais de *information seeking*

Tem havido considerável interesse nos últimos anos na concepção de modelos de *information seeking* (Spink, Wilson *et al.*, 2002; Järvelin e Wilson, 2003). Um modelo pode ser descrito como um framework para se pensar sobre um problema e pode desenvolver-se por meio de relacionamentos entre proposições teóricas (Wilson, 1999). O papel dos modelos, segundo Saracevic (1997), “é descrever os elementos essenciais e as relações de um objeto (sistema, processo, entidade, estrutura, idéia, etc.)”. De acordo com Järvelin e Wilson (2003) “modelos conceituais fornecem uma estratégia de trabalho, um esquema contendo conceitos gerais e suas inter-relações”. A maioria dos modelos que tentam descrever a atividade de *information seeking* estão mais diretamente relacionados a proposições teóricas e são descritos por meio de declarações normalmente na forma de diagramas, definições de causas e conseqüências e descrições do relacionamento entre as fases da atividade de *information seeking* (Järvelin e Wilson, 2003; Ingwersen e Järvelin, 2005b).

De acordo com Engelbart (1962), desenvolver modelos significa especificar: (a) objetos ou componentes essenciais de um sistema a ser estudado; (b) relacionamentos dos objetos que são identificados; (c) tipos de alterações nos objetos e seus relacionamentos que afetam o funcionamento do sistema e de que forma; e (d) definição de objetivos e métodos de pesquisa. Há vários tipos de modelos de *information seeking* (Järvelin e Wilson, 2003), como, por exemplo, os modelos propostos por Kuhlthau (1991), Wilson (1997), Leckie *et al.* (1996), Savolainen (2005) e Johnson (1997). Entretanto, para realização desta pesquisa apenas os

modelos que incluem especificamente a área de RI e são chamados modelos de recuperação e busca de informação (RBI) foram considerados. A seguir são apresentados os principais modelos de RBI disponíveis na literatura.

2.3.1.1 Modelo estratificado de interação em RI proposto por Saracevic (1997)

O modelo estratificado de interação em RI (Saracevic, 1996; 1997) inicia com duas suposições: (1) usuários interagem com sistemas de RI com o intuito de usar a informação; e (2) o uso da informação é conectado à cognição e então aplicado de modo situacional (Saracevic, 1997). São essas suposições que servem de foco e orientação para o modelo. Os principais elementos do modelo estratificado são os usuários e o computador, cada um com variáveis específicas e a interface propicia o diálogo entre eles. A interface, segundo Saracevic (1997), instancia uma variedade de interações, mas não é o foco das interações, apesar de que pode efetivamente suportar ou frustrar interações. O autor deixa claro que um processo de interação é uma seqüência de processos que ocorrem em diferentes níveis, mas que estão conectados. Cada nível envolve diferentes elementos e/ou processos específicos. Os processos no lado do usuário podem ser fisiológico (*e.g.*, visual, palpável, auditivo), psicológico e cognitivo; no lado do computador, os processos podem ser físico e simbólico. A Figura 2 mostra os elementos do modelo estratificado de interação em RI apresentado por Saracevic (1997).

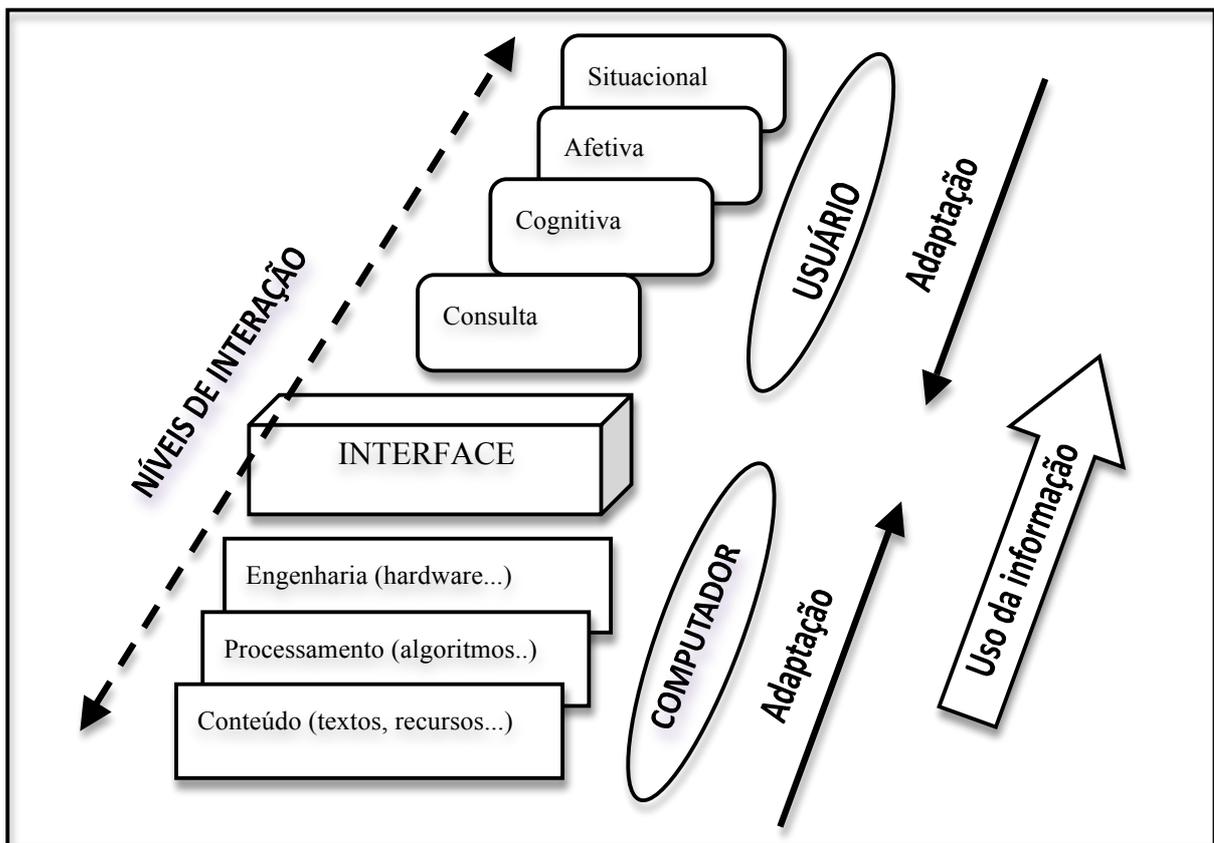


Figura 2 - Elementos do modelo estratificado de interação em RI proposto por Saracevic (1997)

A interface propicia a interação na qual: (a) os usuários realizam um diálogo executando comandos e recebendo respostas não somente para recuperação e comparação, como ocorre nos modelos tradicionais de RI, mas também para engajarem em outros processos além de recuperação e comparação, tais como, entender e descobrir atributos de um determinado componente do computador, navegar entre os recursos de informação, determinar o estado de um determinado processo, visualizar os resultados, obter e prover diferentes tipos de feedback, e assim por diante; (b) o computador interage com o usuário por meio de determinados processos predeterminados e fornece respostas nesse diálogo.

No lado do usuário, Saracevic (1997) sugere três processos: (1) no nível cognitivo os usuários interagem com textos e com suas representações considerando-as estruturas cognitivas – os usuários interpretam cognitivamente os textos obtidos e podem assimilá-los, também, cognitivamente; (2) no nível emotivo, os usuários interagem com as suas intenções e os conceitos relacionados à intencionalidade, tais como crenças, motivação, sentimentos, dese-

jos, urgência, etc.; (3) no nível situacional, os usuários interagem com uma determinada situação ou problema que foi o responsável por gerar uma necessidade.

No lado do computador, os processos descritos por Saracevic (1997) são: (1) o nível de engenharia se refere a determinadas características, como, por exemplo, capacidade, eficiência, poder de processamento; (2) o nível de processamento está vinculado a software, com interesse específico em algoritmos e modelos que manipulem textos, consultas, interface e outros processos críticos relacionados à interação entre os usuários e o computador e no próprio computador; (3) o nível de conteúdo concentra-se nos recursos de informação, os textos (textos, nesse contexto, referem-se aos diversos tipos de objetos de informação, como, por exemplo, imagens, sons, multimídia, etc.).

À medida que a interação ocorre, uma série de adaptações dinâmicas acontecem em ambos os elementos, usuários e computador. Essas interações ocorrem na interface, que é o ponto em que eles se encontram. Saracevic (1997) assume que o uso da informação se direciona para um nível situacional. Além disso, adaptações também podem significar alterações em uma variedade de níveis. É importante considerar também que diferentes tipos de feedback influenciam consideravelmente os tipos de alterações e as mudanças que podem ocorrer durante um processo de busca.

2.3.1.2 Framework integrado de RBI proposto por Ingwersen (1992) e revisto por Ingwersen e Järvelin (2005b)

Entre os primeiros modelos de interação definidos, o modelo apresentado por Ingwersen (1992), discutido por Järvelin e Ingwersen (2004) e mais detalhadamente revisto por Ingwersen e Järvelin (2005a) trata a interação em relação a componentes cognitivos e transformações em RI. O modelo, representado na Figura 3, trata a recuperação e busca de informação de maneira mais abrangente, em uma perspectiva global (Robins, 2000), pois considera desde a definição e construção de objetos computacionais até o ator cognitivo que possui uma necessidade por informação. O modelo, portanto, concentra-se em identificar os processos de cognição que podem ocorrer em todos os elementos, ou atores cognitivos, envolvidos que processam informação. Os números na Figura 3 representam basicamente os processos de interação (1 – 4), tais como interação social (1), ou referem-se à diferentes tipos de geração e transformação de cognição ou influência cognitiva (5 – 8).

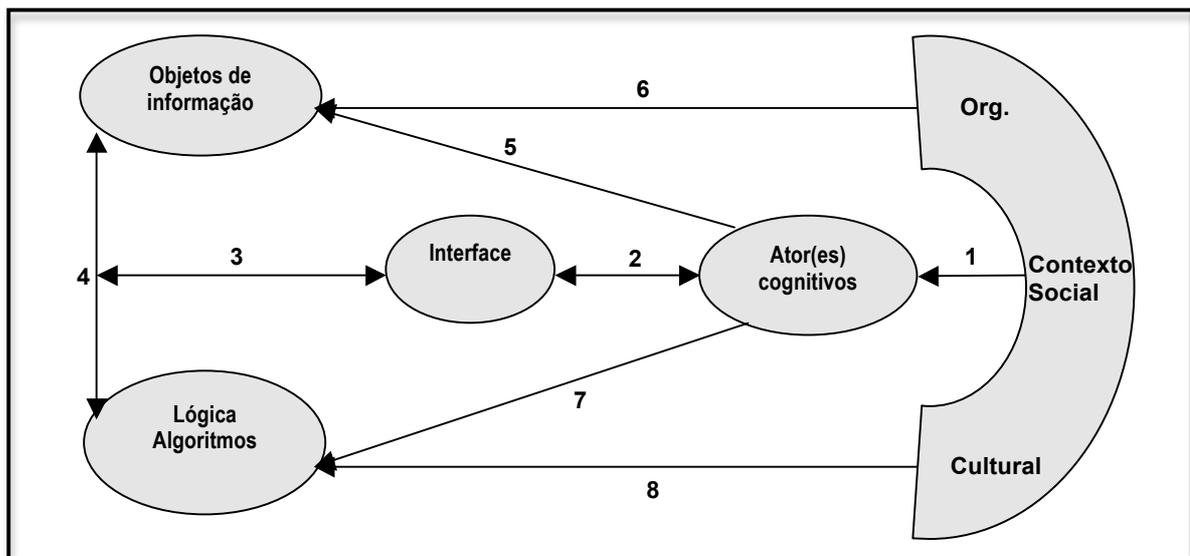


Figura 3 - Modelo interativo de RBI proposto por Ingwersen e Järvelin (2005a)

Há uma variedade de atores humanos de diferentes origens cognitivas no tempo e espaço que se comunicam durante a interatividade em RBI. De acordo com Ingwersen e Järvelin (2005b), os atores cognitivos são vistos em quatro formas de interação: com objetos de informação, com mecanismos computacionais, com interfaces e em um contexto organizacional, cultural e social. A seguir são descritos os atores humanos, ou atores cognitivos, como citados na Figura 3, abordados pelo modelo.

- a) atores que criam os tipos de objetos de informação;
- b) atores responsáveis por analisar índices dos objetos de informação;
- c) atores que desenvolvem as funcionalidades da interface e de recuperação;
- d) atores responsáveis por estruturar uma base de dados, implementar uma máquina de busca e algoritmos de indexação, definir mecanismos de filtragem e navegação, etc.;
- e) atores, chamados seletores, que decidem a disponibilidade pública ou comercial dos objetos de informação;
- f) atores que são na verdade os usuários que identificaram uma necessidade e buscam por informação; e
- g) atores que formam comunidades de indivíduos organizados em um contexto social, cultural ou organizacional.

O modelo enfatiza os processos de informação que são executados durante a RBI num contexto temporal: primeiro, os processos de interação social (interação 1 da figura) estão di-

retamente relacionados aos atores cognitivos por meio de um contexto organizacional ou sociocultural. A interação social pode instigar atividades de RBI, mas pode também ser o principal objetivo a ser alcançado. Segundo, a interação com a informação também ocorre entre os atores cognitivos e a manifestação cognitiva inserida nos mecanismos computacionais (*i.e.*, lógica, algoritmos) e nos objetos de informação por intermédio da interface (interações 2/3 da figura). Os últimos dois componentes interagem verticalmente (interação 4 da figura) e constituem a parte central de um sistema de informação. Terceiro, gerações e transformações cognitivas e emocionais de potenciais informações podem ocorrer como requerido pelo ator individual (interações 5/7 da figura) assim como a partir de contextos social, cultural ou organizacional em direção aos componentes relacionados aos mecanismos computacionais (*i.e.*, lógica, algoritmos) e objetos de informação (interações 6/8 da figura) ao longo do tempo.

Wilson (1999) enfatiza que o modelo apresentado por Ingwersen (1996) deixa claro outros elementos. Ingwersen, segundo Wilson (1999), demonstra que em cada componente do modelo as funções do usuário que tem uma necessidade por informação, o ator que gerencia os objetos de informação, o intermediário, a interface e o sistema de RI são o resultado de modelos cognitivos explícitos ou implícitos do domínio de interesse. Os usuários possuem modelos de suas necessidades, seus problemas ou seus objetivos, que são normalmente implícitos, mas que são capazes de serem explicitados. O sistema de RI, por exemplo, é uma explicação do modelo cognitivo do desenvolvedor do sistema que define o que o sistema deve fazer e como deve funcionar.

2.3.1.3 Modelo de interação episódico proposto por Belkin (1993)

O modelo de RBI apresentado por Belkin (1993) inicia com a hipótese de que um problema real em RI não é como representar textos, mas sim como representar o estado anômalo de conhecimento (Belkin, 1980) pertencente aos usuários – aspectos cognitivos e situacionais que levaram os usuários a assumirem um comportamento de *information seeking* (Saracevic, 1997). Belkin (1993), na verdade, propõe uma alternativa aos modelos tradicionais de RI utilizando conceitos da área de *information seeking* de forma a aprimorar a interação de usuários com a informação recuperada. Para tal, Belkin (1993) faz as seguintes suposições: (a) *information seeking* é essencialmente um processo interativo descrito por características relacionadas a interações de usuários com a informação recuperada; e (b) o objetivo dos sistemas de RI é dar suporte ao comportamento de *information seeking*. A partir dessas suposições, o

modelo, portanto, baseia-se em processos específicos relacionados ao comportamento de *information seeking* dos usuários e considera a interação do usuário com um sistema de RI como uma seqüência de diferenciadas interações em uma série de episódios de *information seeking*. Há diferentes métodos de interações para suportar a variedade de processos, tais como avaliação, interpretação, modificação, navegação e assim por diante.

Belkin (1993) fornece algumas sugestões que devem ser analisadas de modo a atender às suposições previamente descritas para o desenvolvimento de um sistema de RI. O primeiro ponto a ser considerado é como inserir o usuário como componente central do sistema de RI. Como sugestão para esse primeiro ponto, Belkin (1993) cita outros trabalhos que descrevem especificamente como inserir o usuário no âmbito de um sistema de informação. O trabalho apresentado por Belkin et al. (1983), por exemplo, cita que em um sistema de RI é necessário que o mecanismo que fornece a informação atue como um intermediário inteligente entre o usuário e a informação com a qual ele deve interagir. Os autores descrevem que esse mecanismo que fornece informação tem de entender as características do usuário e as suas necessidades, que poderão ser relevantes na identificação de informação que possa lhe interessar. Esse entendimento deve ser adquirido por meio de um diálogo (*i.e.*, interação) entre o usuário e o sistema.

O segundo ponto destacado por Belkin (1993) refere-se à identificação e ao suporte a estratégias de *information seeking*. Aqui, Belkin (1993) também cita outro trabalho (Belkin, Marchetti *et al.*, 1993) que caracteriza estratégias de *information seeking* como um conjunto de quatro dimensões ou facetas. Essas facetas, definidas como objetivo da interação (aprendizagem ou seleção), método de interação (inspeção ou busca), modo de recuperação (reconhecimento ou especificação) e tipo de recurso interagido (itens de informação ou metainformação), foram identificadas por meio da observação e da classificação de comportamentos de *information seeking* em uma variedade de cenários. Com essas facetas, os autores afirmam quais estratégias de *information seeking* podem ser mapeadas. Entretanto, os próprios autores afirmam que essas dimensões, ou facetas, podem não ser suficientes para representar todas as estratégias possíveis inerentes à realidade envolta em um processo de *information seeking*.

O último ponto descrito por Belkin (1993) destaca que a interação com a informação recuperada deve ser o processo central de RI. Belkin (1993) sugere que a definição dos papéis e das responsabilidades na interação constitui o ponto inicial. Além disso, promover e

oferecer acesso e manipulação direta à informação recuperada, prover suporte para navegação e criar estruturas que viabilizem o diálogo são exemplos de questões importantes que também devem ser consideradas.

De acordo com Saracevic (1997), a força do modelo proposto por Belkin (1993) está na definição direta de uma variedade de processos pertencentes à RI e não somente na comparação, como ocorre nos modelos clássicos de RI. Entretanto, o modelo apresenta dificuldades em identificar episódios individuais e seus efeitos sobre outros episódios.

2.3.2 *Information seeking* e recuperação de informação

Os modelos tradicionais de RI, de acordo com Allan et al. (2002) e Hert (1997), geralmente enfatizam tarefas de recuperação bem restritas (*e.g.*, recuperação e filtragem) e ignoram outras tarefas importantes (*e.g.*, navegação e sumarização). Saracevic (1997) descreve que os modelos tradicionais de RI consideram que a necessidade de um usuário é estática (Belkin, 1993) e, além disso, o processo de avaliação nesse tipo de modelo é consideravelmente mais fácil de ser realizado (Ingwersen e Järvelin, 2005b). Entretanto, a área de recuperação de informação interativa (RII) (Kelly e Belkin, 2002; White, Jose *et al.*, 2006) inclui a variável relativa à interação que ocorre entre sistemas de RI e usuários. Allan et al. (2002) afirmam que a RII abrange uma grande variedade de tarefas relativas à *information seeking*, mas os atuais modelos mapeiam apenas algumas dessas tarefas.

Information seeking é um processo ou atividade que visa obter informações tanto a partir de seres humanos quanto a partir de recursos tecnológicos (Belkin, Cool *et al.*, 1995), *i.e.*, *information seeking* é, portanto, uma área que estuda como uma pessoa busca informação não apenas em sistemas de informação, mas em um contexto mais completo. Há vários fatores estudados, por exemplo, a interação entre fatos pessoais tais como experiência e conhecimento e a necessidade em si, que requerem exploração da cognição humana (Marchionini, 1989). Esses diversos fatores que podem influenciar um processo de busca são pesquisados pela área de *information seeking*. Tissen (1991) cita que *information seeking* é uma atividade altamente exploratória.

A área de *information seeking* tem uma atuação mais abrangente que a área de RI e RII, pois considera o usuário como a variável principal na tarefa de busca por informação. Uma pessoa desenvolvendo um comportamento de *information seeking* não sabe se existe

uma resposta para uma necessidade e, segundo Case (2007) e Savolainen (2005), o próprio processo de busca pode ser suficiente para satisfazer uma necessidade. *Information seeking*, segundo Ingwersen e Järvelin (2005b) e Kuhlthau (2005), é orientado à pessoas enquanto que RII é orientada à tecnologia. A RI e principalmente a RII podem atuar como uma ferramenta de apoio a um processo de *information seeking*. Järvelin e Ingwersen (2004) citam que “a RII pode funcionar como um caso especial de *information seeking*”. A Figura 4, que apresenta o modelo de comportamento de informação definido por Wilson (1999), deixa claro a relação entre *information seeking* e RI e RII. Um comportamento de informação (do inglês *information behavior*), conforme define Ingwersen e Järvelin (2005b), “é o comportamento humano que se preocupa com a geração, comunicação, uso e outras atividades relacionadas à informação, tais como um comportamento de *information seeking* e RII”.

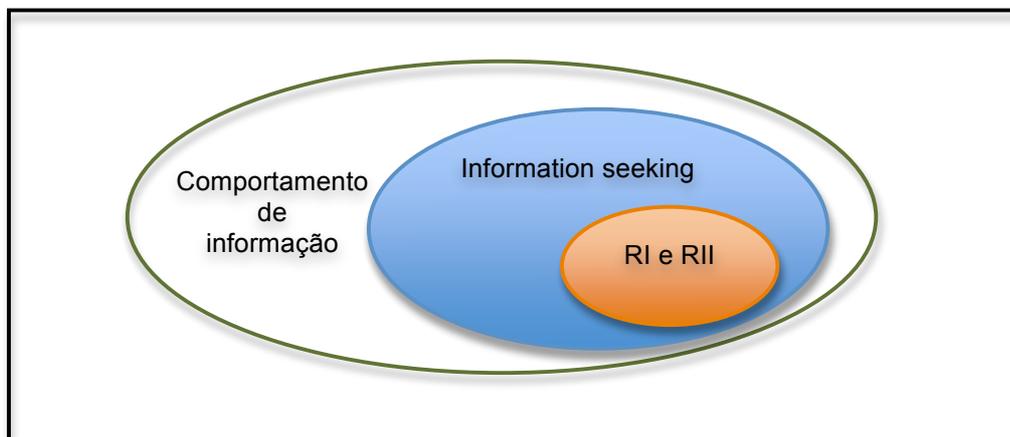


Figura 4 - Modelo de comportamento de informação (Wilson, 1999)

2.4 Ontologia

Ontologia, segundo a visão filosófica, é a ciência que estuda o que pode existir – dos tipos e estruturas de objetos, propriedades, eventos, processos e relações referente a uma única realidade de cada área (Smith e Welty, 2001; Smith, 2003; Fonseca, 2007). No âmbito da inteligência artificial (IA) e dos sistemas de informação, o termo ontologia refere-se à descrição de parte do conhecimento presente em um programa (Bruijn e Fensel, 2005). Gruber (1993a) define ontologia como “uma especificação formal e explícita de uma conceitualização compartilhada” que pode ser entendida como uma visão abstrata e simplificada do mundo (*i.e.*, domínio) que se deseja representar por meio da utilização de um conjunto de conceitos, suas definições e suas inter-relações (Ushold e Gruninger, 1996; Guarino, 1998). A ontologia é uma especificação porque representa uma conceitualização de forma concreta; também é explícita

porque representa todos os conceitos e restrições inerentes a um domínio. Uma ontologia deve ser expressa formalmente para que possa ser processada por programas de computador. Além disso, conforme descrito por Studer et al. (1998), uma ontologia deve representar o conhecimento de um domínio de maneira consensual, *i.e.*, o conhecimento compartilhado.

Basicamente, o papel da ontologia é facilitar a construção de um modelo de domínio por meio da representação de um vocabulário de termos e relações (Studer, Decker *et al.*, 2000). Guarino (1998) define ontologia como sendo uma teoria lógica que pretende representar ou explicar um determinado significado por meio de um vocabulário formal. Mais precisamente, conforme define Chandrasekaran et al. (1999), não é esse vocabulário que por si só define uma ontologia, mas sim as conceitualizações que os termos descritos nesse vocabulário pretendem capturar. Além disso, a formalização de uma ontologia é dependente da linguagem, ao passo que uma conceitualização independe de linguagem (Guarino, 1998).

O conhecimento em ontologias, de acordo com Gruber (1993b) e mais explicitamente descrito por Corcho e Gómez-Pérez (2000), pode ser especificado por meio de cinco tipos de componentes: conceitos, relações, funções, axiomas e instâncias. Um conceito é representado por uma classe e por suas respectivas propriedades (Noy e McGuinness, 2001; Staab, Studer *et al.*, 2001; Sure e Studer, 2002) – as propriedades de cada conceito descrevem os seus atributos e as suas características (*e.g.*, o conceito “autor” é definido pela classe “autor” que contém a propriedade “nome”). Entretanto, há alguns conceitos que compreendem subclasses e outros que são definidos por mais de uma classe – nesse caso, a relação entre classes, que estabelece o tipo de interação entre conceitos, também é fundamental para o entendimento de um conceito (Chandrasekaran, Josephson *et al.*, 1999). Uma função é um tipo especial de relação que considera um argumento especificamente. Axiomas são utilizados para criar restrições sobre as informações descritas nos conceitos e verificar a validade de conhecimento deduzido. Instância, também conhecida como indivíduo, são os termos usados para materializar um conceito.

2.4.1 Tipos de ontologias

Há diferentes tipos ou diferentes níveis de ontologia. A literatura apresenta vários tipos com descrições distintas, sendo alguns apenas com nomes diferentes mas que representam a mesma idéia. Studer et al. (2000), que expõem uma visão explicitamente direcionada para o de-

envolvimento de sistemas baseados em conhecimento, distinguem ontologia basicamente em três tipos:

- ✚ **Ontologias de domínio.** Representam o conhecimento de um domínio em particular (*e.g.*, medicina, eletrônica, engenharia mecânica, comércio eletrônico, biologia). É esse tipo de ontologia que interessa para utilização nesta proposta;
- ✚ **Ontologias genéricas ou de senso comum.** Armazenam o conhecimento geral sobre o mundo além de noções básicas e conceitos para coisas, como, por exemplo, tempo, espaço, estado, evento, ação, etc. Esse tipo de ontologia, também chamada de ontologia de alto nível (Guarino, 1998; Sure e Studer, 2002), é independente de um problema ou domínio em particular;
- ✚ **Ontologias representacionais.** Essas ontologias mantêm descrição de estruturas conceituais e metaestruturas que podem ser aplicadas em um âmbito geral; além disso, elas são baseadas em visões lógicas e filosóficas em vez de serem focadas especificamente em aplicações (Maedche e Staab, 2000).

2.4.2 Inferência e raciocínio

Inferência e raciocínio normalmente são conceitos que se confundem. Raciocínio pode ser visto como o processo de inferir um novo conhecimento enquanto que inferência é a derivação em si de novos dados, fatos ou conhecimentos, que podem ser tanto positivos quanto negativos, a partir de um conjunto de dados (Pearl, 1988). No âmbito da ontologia, a forma com que ela é representada, por meio de linguagem formal que inclui axiomas responsáveis por especificar relacionamentos entre conceitos (Decker, Erdmann *et al.*, 1999), permite que técnicas de inferência sejam utilizadas para extrair conhecimento definido apenas implicitamente (Corcho e Gómez-Pérez, 2000). Para Baader *et al.* (2005) raciocínio é importante por garantir a qualidade de uma ontologia e pode ser usado, por exemplo, para testar se conceitos não são contraditórios, derivar relações implícitas, verificar quais conceitos são especializações ou sinônimos de outros conceitos, testar a consistência na hierarquia de conceitos, inferir relacionamentos entre instâncias e assim por diante.

Uma máquina de inferência para ontologias oferece um tipo de serviço de raciocínio para aplicações que desejam utilizar uma ontologia apropriadamente (Staab, 2004), *i.e.*, utili-

zar processos de inferência para raciocinar sobre as definições explicitamente descritas e extrair conhecimento adicional representado não explicitamente. Além disso, o processo de inferência depende diretamente da expressividade da linguagem que formaliza uma ontologia (Staab, 2004) – quanto maior a expressividade oferecida pela linguagem mais complexa e poderosa uma máquina de inferência pode ser. Essa máquina emprega um conjunto de regras (e.g., *if (A and B) then infer C*) sobre um conjunto de fatos atômicos (Guha, Lassila *et al.*, 1998). A aplicação de regras para derivar novo conhecimento ocorre a partir de uma consulta ou quando novos fatos são obtidos.

Algumas vezes, a estrutura da ontologia por si só e as regras genéricas pré-configuradas em uma máquina de inferência não são suficientes para inferir um novo conhecimento. Contudo, há a possibilidade de se utilizarem regras personalizadas no intuito de, por exemplo, criar novos relacionamentos e ampliar o escopo de informação adicional a ser inferida (Golbreich, 2004). As regras são usadas normalmente para definir a parte de conhecimento que pode ser deduzido, como, por exemplo, expressar consultas mais complexas sobre uma ontologia e sobre a sua base de conhecimento (Golbreich, 2004). Dessa forma, a utilização de regras personalizadas pode ser caracterizada como uma ampliação da expressividade de uma linguagem, pois estende as possibilidades de inferência.

Um exemplo da importância em se utilizar inferência e raciocínio, como descreve Chandrasekaran *et al.* (1999), ocorre no âmbito de sistemas de resolução de problemas, pois esses sistemas normalmente requerem complexas seqüências de inferência para atingir objetivos específicos. Esse tipo de sistema, ainda segundo os autores, requer a possibilidade de escolha de diferentes caminhos num processo de raciocínio. Para tal, a ontologia, que representa o conhecimento de uma determinada realidade, ou domínio, juntamente com métodos empregados para resolução de problemas, possibilita que objetivos predeterminados ou não possam ser alcançados.

Um outro exemplo do uso de inferência é apresentado por Guha *et al.* (1998). O exemplo descrito pelos autores sugere a utilização de um sistema que permite dar títulos a fotografias. No caso de uma pessoa, por exemplo, buscar imagens que contêm decorações, um sistema de busca que utiliza comparação simples, *i.e.*, apenas sobre os títulos das fotografias, não conseguirá encontrar esse tipo de imagem, uma vez que dificilmente os títulos das imagens contêm descrição sobre decorações. Uma máquina de inferência configurada com um

conjunto de regras específicas sobre tipos e características pertencentes a determinados tipos de eventos pode inferir que em fotografias de aniversários há uma grande probabilidade de se encontrarem exemplos de decoração – portanto, há um aperfeiçoamento no processo de recuperação.

2.4.3 Importância de ontologias

Uma ontologia pode ser o coração de um sistema baseado em conhecimento, pois é por intermédio dela que o conhecimento de um domínio pode ser representado (Chandrasekaran, Josephson *et al.*, 1999). A utilização de ontologias, segundo Sowa (2002), possibilita o desenvolvimento de sistemas mais inteligentes, pois o conhecimento é representado formalmente de modo a suportar extração explícita e, mais importante, extração implícita de conhecimento. O surgimento da ontologia na área da computação, de acordo com Smith e Welty (2001), representa uma vitória do conteúdo sobre o processo. Essa vitória de alguma forma, mesmo que paradoxalmente, reforça o resultado do fato, que mesmo com a sofisticação dos softwares, os desenvolvedores e analistas de sistemas perceberam a importância de focar os dados sobre os quais os seus sistemas operam em vez de focar em funcionalidades e aspectos procedurais dos próprios sistemas.

Noy e McGuinness (2001) apresentam algumas razões que explicam o motivo de se desenvolver uma ontologia: (a) entendimento comum e compartilhado de uma estrutura de informação entre pessoas e programas; (b) possibilidade de reuso do conhecimento de um domínio; (c) explicitação de definições inerentes a um domínio; (d) separar o conhecimento de um domínio do conhecimento operacional; e (e) análise do conhecimento do domínio. O reconhecimento da importância de se usar ontologia ocorre em várias áreas, tais como engenharia e representação do conhecimento, modelagem de base de dados, integração de informação, recuperação e extração de informação, gestão do conhecimento e desenvolvimento de sistemas (Uschold e Gruninger, 1996; Guarino, 1998; Chandrasekaran, Josephson *et al.*, 1999; McGuinness, 1999).

Como uma ontologia deve representar um conhecimento compartilhado, além de ser definida de maneira consistente e não ambígua, Uschold e Gruninger (1996) descrevem que, portanto, a ontologia pode funcionar como um framework que unifica diferentes pontos de vista e serve de base para uma clara comunicação entre pessoas, interoperabilidade entre sistemas e, também, para engenharia de sistemas – especificamente sobre questões de reusabili-

dade, segurança e especificação. Nesse contexto, o uso de ontologias pode facilitar a identificação de requisitos de um sistema e a compreensão dos relacionamentos entre diferentes componentes do sistema.

Wand e Weber (2004) perceberam que as ontologias podem ser usadas no âmbito de sistemas de informação de diferentes maneiras: (a) uma ontologia fornece um conjunto de critérios para avaliar modelos usados no desenvolvimento de sistemas; (b) uma ontologia proporciona um conjunto de conceitos para modelar sistemas e raciocinar sobre suas características; e (c) uma ontologia pode ser usada para definir o significado da informação que estará disponível em um sistema de informação. Guarino (1998) também enfatiza o uso de ontologia no âmbito da engenharia de sistemas. Segundo ele, o uso de ontologia pode ocorrer não apenas como um artefato de um sistema, mas também como sendo parte do processo de desenvolvimento de um sistema de informação (*e.g.*, análise de requisitos, modelagem conceitual, design da interface). De acordo com Fonseca (2007), modelos e esquemas conceituais podem ser validados por meio de ontologias. Ontologias, portanto, podem ser concebidas de forma a representar o conhecimento relativo à construção de componentes de sistemas de informação (Smith e Welty, 2001). Segundo Guarino (1998), um importante benefício em utilizar uma ontologia na fase de desenvolvimento de um sistema é melhorar a transparência da aplicação devido à clara formalização das funcionalidades e induzir o desenvolvedor a pensar em um alto nível de reutilização, o que não acontece normalmente nos processos descritos pela engenharia de software. Isso permitiria a construção de uma biblioteca de ontologias especificamente concebida para a geração de componentes de sistemas de informação.

Chandrasekaran et al. (1999) afirmam que sistemas de RI, bibliotecas digitais, integração de fontes de informação heterogêneas e máquinas de busca na Internet são exemplos de áreas que estão usufruindo de ontologias para organizar o conhecimento de um domínio e proporcionar o desenvolvimento de sistemas com melhores funcionalidades. Em processamento de linguagem natural, por exemplo, os sistemas usam ontologias para tratar questões de ambigüidade e também para facilitar a identificação de categorias semânticas envolvidas no universo de discurso de um determinado domínio. A resolução de problemas (*problem solving*) baseado em conhecimento também é um exemplo de área que utiliza ontologia na busca de solução para uma variedade de problemas, tais como diagnóstico, planejamento e design. As ontologias podem, também, ser utilizadas até mesmo por usuários de forma a pos-

sibilitar um melhor entendimento do vocabulário usado por um sistema de informação (Guarino, 1998).

2.5 Círculo hermenêutico

Hermenêutica é uma teoria que discute a compreensão humana e a interpretação de textos escritos (Ricoeur, 1978; Gadamer, 1989; Heidegger e Stambaugh, 1996). A expressão “hermenêutica” é derivada do verbo grego *hermeneuen* que está relacionada com o substantivo *hermeneus*, o qual, por sua vez, refere-se ao nome do deus Hermes. Hermes é o mensageiro divino que traz a mensagem de destino; *hermeneuein* consegue interpretar a mensagem (Heidegger, 1982). A hermenêutica é uma tradição antiga que teve início nas interpretações grega e judaica de textos jurídicos e bíblicos. O papel da hermenêutica é entender o significado de textos quando este é confuso ou de alguma forma obscuro (Taylor, 1971; Winograd e Flores, 1987). Um texto pode ser caracterizado por ter conteúdo expresso e não expresso. Para que a compreensão do texto ocorra é necessário que haja interpretação tanto do conteúdo explícito quanto do conteúdo definido apenas implicitamente. A interpretação se inicia de uma forma ingênua em que o leitor identifica significados do texto em seu conjunto, partindo para um entendimento mais profundo por meio do reconhecimento do relacionamento das partes com o todo (Geanellos, 2000).

O princípio mais fundamental da hermenêutica é o círculo hermenêutico (Klein e Myers, 1999). A idéia do círculo hermenêutico, apresentado na Figura 5, é uma metáfora que descreve o círculo analítico que sugere que os seres humanos entendem um todo a partir de concepções sobre o significado das partes e suas inter-relações (Gadamer, 1989; Heidegger e Stambaugh, 1996). As concepções de cada pessoa influenciam o seu processo de entendimento e interpretação. Para facilitar o entendimento da interpretação do que significa o círculo hermenêutico, Gadamer (1989) usa o exemplo de tradução do significado de uma simples frase, como, por exemplo, “*eles estão jogando futebol*”. Para entender o significado das partes individuais da sentença (*i.e.*, se futebol usa uma bola redonda, ou uma bola em formato de ovo ou mesmo se não usa bola alguma), precisa-se entender o significado da sentença em sua totalidade. O processo de interpretação move de um entendimento precursor das partes para um todo e do entendimento global do todo contextualizado que afeta o entendimento de cada parte, *i.e.*, o significado de cada palavra. Além disso, a frase em seu conjunto é também parte de um contexto maior. Se o contexto deixa claro que ninguém gosta de es-

portes, a frase é utilizada apenas de forma metafórica e, dessa maneira, há uma nova interpretação sobre os conceitos que compõem a frase (e.g., o conceito “jogar” não necessariamente significa movimentos físicos desenvolvidos em um campo de grama). Gadamer (1989) descreve que essa mudança na compreensão do todo para a parte e de volta para o todo é constante e que pode ser representada como um círculo concêntrico.

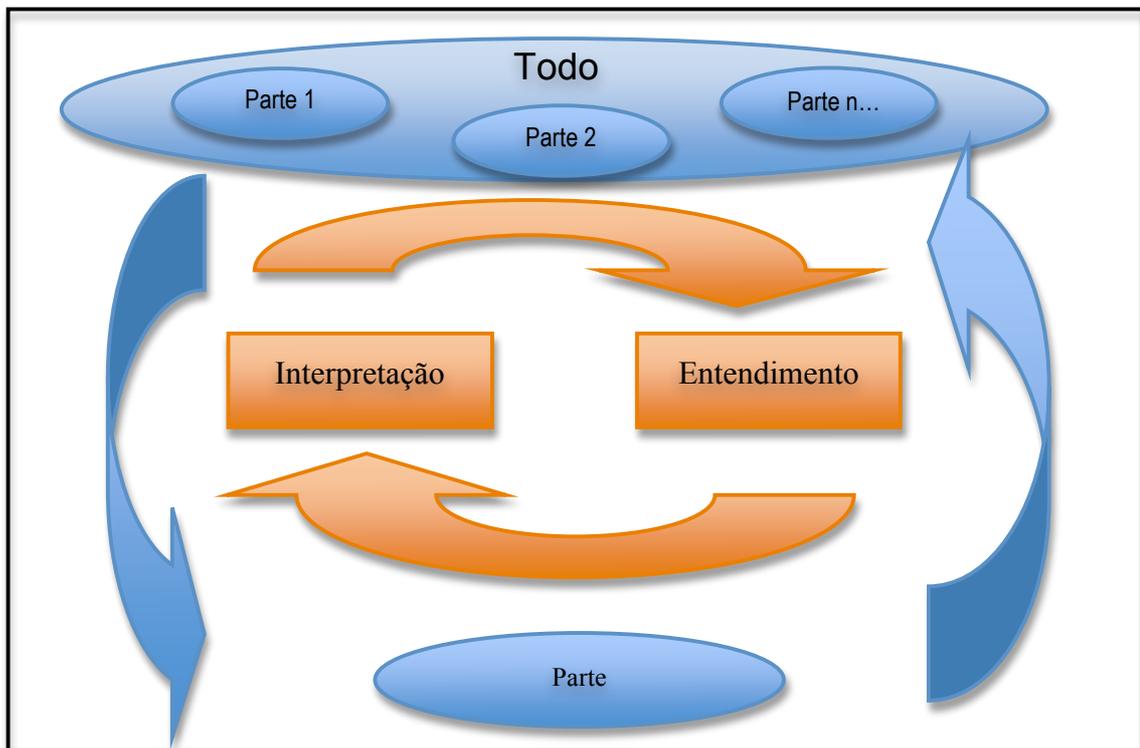


Figura 5 - Círculo hermenêutico

O processo de entendimento, segundo Heidegger (1982), está envolto em um diálogo contínuo entre as dimensões de análises e sínteses que visam compreender o todo e as dimensões que visam compreender as partes. Não se pode evitar o fato de que uma pessoa sempre traz concepções em um processo de entendimento tanto sobre o todo quanto sobre as partes. Entretanto, uma pessoa precisa refletir sobre suas pressuposições enquanto analisa um objeto de investigação. O processo de entender algo novo “é uma dialética contínua que muda de direção constantemente entre o mais local dos detalhes locais e a mais global da estrutura global de modo a induzir que ambos sejam entendidos simultaneamente” (Geertz, 1979) citado por (Bernstein, 1983).

Heidegger (1982) afirma que “cada interpretação que contribui para algum entendimento precisa obrigatoriamente já ter sido entendida sobre o que é para ser interpretado”. O

autor entende que essa proposição pode ser vista como um círculo vicioso. “Mas se a interpretação tem que operar dentro do que é entendido, como isso pode produzir resultados científicos sem entrar em um círculo, especialmente quando o entendimento pressuposto ainda opera em um conhecimento comum do ser humano e do mundo?” (Heidegger e Stambaugh, 1996) Contudo, Heidegger (1982) nega essa visão de círculo vicioso veementemente. Heidegger e Stambaugh (1996) afirmam que “a realização das condições fundamentais da possível interpretação não está na interpretação incorreta a priori das condições essenciais do que deve ser feito”. Além disso, o círculo do entendimento não é um círculo em que qualquer tipo de conhecimento opera, mas sim a expressão do conhecimento mais fundamental e intuitivo. Dessa forma, de acordo com Heidegger (1982), “o que é decisivo não é sair do círculo, mas entrar nele da forma correta”.

Para a descrição de Heidegger (1982) sobre o círculo hermenêutico, Gadamer (1989) adicionou dois elementos cruciais: (1) aplicação e (2) interatividade. Em vez de conceber o processo de entendimento no que se refere a análise e síntese por si só, Gadamer mostra a importância da dimensão da aplicação no processo. Entender acontece em um contexto no qual os conceitos envolvidos entram em uma espécie de prática. Isso requer um julgamento sobre as implicações por possíveis alternativas em situações locais. Além disso, Gadamer também enfatiza a interatividade referente ao *vai e vem* natural do círculo hermenêutico. Na interatividade, o participante não se depara com um subjetivismo no qual a interação é um tipo de predicado que atua contra o participante; pelo contrário, o participante é envolvido na interação como se ele mesmo fosse o predicado da interação. Os seus movimentos e decisões são limitados ao padrão *vai e vem* requerido pela interatividade.

Um dos fundamentos do trabalho de Gadamer em hermenêutica é a conexão entre a teoria e a prática (Winograd e Flores, 1987). O entendimento hermenêutico não é privado, não é uma atividade sem propósito. Para Gadamer, interpretação, entendimento e aplicação estão ligados como uma unidade. Eles não são eventos independentes no tempo, mas acontecem simultaneamente, com um influenciando e até mesmo possibilitando o outro. Gadamer (1989) afirma que “interpretação não é um fato ocasional suplementado pelo entendimento; na verdade, entendimento é sempre interpretação e, portanto, interpretação é uma forma explícita de entendimento”. A tarefa do interpretador é adaptar o texto à situação concreta na

qual o texto realmente se expressa. Entendimento é um processo dinâmico, e o significado está sempre sendo criado e recriado.

2.6 Considerações finais

Neste capítulo foram apresentadas as quatro áreas que forneceram subsídios para a concepção desta proposta. A área de RI se preocupa, principalmente, com a organização da informação e mecanismos que facilitem o processo de recuperar a informação requerida. Há vários modelos de RI sendo que cada um possui especificidades que podem ser utilizadas para determinadas aplicações.

Estudar o processo de busca é o centro da pesquisa de *information seeking*. Essa área auxilia a definição de sistemas de busca que sejam mais adaptados ao modo como os seres humanos assumem um comportamento de busca. Os modelos de RBI mostram a preocupação em mapear tarefas cognitivas juntamente com tarefas computacionais para definir ambientes em que usuários possam suprir as suas necessidades por informação.

Ontologia é usada para representar o conhecimento de um domínio e propiciar que sistemas computacionais possam ser desenvolvidos com mais inteligência. Uma ontologia possibilita que o conhecimento de um domínio pode ser utilizado por sistemas de busca, por meio da confecção de mecanismos mais inteligentes para organizar e recuperar informação, e por usuários, por intermédio de componentes interativos.

Pode-se considerar que um processo de busca envolve diversos tipos de variáveis, que são discutidas pelas áreas de *information seeking* e RI. Além disso, a Filosofia, especificamente o círculo hermenêutico, pode auxiliar no entendimento do processo de compreensão e interpretação inerentes ao ser humano. Um processo de busca envolve diretamente a compreensão e a interpretação de novas informações fruto de cada ação tomada. Tal compreensão e interpretação afetam o entendimento geral da necessidade de um usuário, o que reflete sobre novas ações a serem tomadas. O círculo hermenêutico, no âmbito desta pesquisa, também contribui para a concepção de um modelo de RBI. O próximo capítulo descreve o modelo proposto e o capítulo seguinte apresenta o framework *Hermeneus*, que aplica os conceitos descritos pelo modelo.

3 MODELO PROPOSTO

3.1 Introdução

A literatura da área de *information seeking* (Ingwersen, 1992; Belkin, 1993; Marchionini, 1997; Saracevic, 1997; Savage-Knepshield e Belkin, 1999; Wilson, 1999; Kelly e Belkin, 2002, Kuhlthau, 1993 #69; Spink, Wilson *et al.*, 2002; Järvelin e Ingwersen, 2004; Ingwersen e Järvelin, 2005b), conforme descrito anteriormente, questiona os modelos tradicionais de RI por não estarem adaptados ao processo de busca executado por seres humanos. Como o comportamento dos usuários não segue uma rotina predefinida em um processo de busca (Belkin, 1993; Kelly e Belkin, 2002; Teevan, Alvarado *et al.*, 2004; Ingwersen e Järvelin, 2005b), os usuários necessitam entrar em uma espécie de diálogo com o sistema de forma a entender realmente qual é a sua necessidade ou quais são as suas perguntas e, então, proceder de forma a tentar encontrar a informação que possa suprir essa falta de conhecimento.

Nesse contexto, propõe-se um modelo de RBI que está embasado nas áreas previamente pesquisadas. *Information seeking* auxiliou a entender o comportamento de seres humanos em um processo de busca que deve ser facilitado por meio de ferramentas – um processo de busca não pode ser caracterizado como estático mas sim dinâmico. Além disso, recorreu-se à Filosofia, especificamente ao círculo hermenêutico, no intuito de construir um modelo que esteja mais aproximado ao modo como os seres humanos entendem e/ou interpretam. Um processo de busca está diretamente relacionado ao entendimento de um problema, e a cada interação com um sistema de busca o usuário acessa novas informações que afetam o entendimento de suas necessidades e, conseqüentemente, daquilo que está buscando.

O modelo também utiliza ontologia, que visa dar suporte à construção de mecanismos que atuem tanto na esfera computacional, mais especificamente sobre os componentes de RI, quanto na esfera relativa à interação entre os usuários e o sistema. Cabe ressaltar que o objetivo desta pesquisa é promover a utilização de ontologias, juntamente com conceitos extraídos das áreas de RI, *information seeking* e círculo hermenêutico, para proporcionar a definição de um modelo de RBI. O modelo tem por objetivo subsidiar o desenvolvimento de ambientes de busca interativos e mais capazes de possibilitar que a interação em um processo de busca por informação ocorra da maneira mais natural possível.

3.2 Descrição do modelo

O modelo proposto é constituído pelos seguintes componentes, conforme descreve a Figura 6: usuário, informação requerida, comportamento de busca, componentes interativos (interface), ontologia, instâncias da ontologia, mecanismo de recuperação, índices semânticos e base de conhecimento. Além disso, o círculo hermenêutico também está representado, pois deve auxiliar no entendimento de ações executadas por usuários. Esses componentes visam descrever as tarefas cognitivas (*e.g.*, verificação, análise) e físicas (*e.g.*, formular ou reformular uma consulta) que estão diretamente relacionadas: (a) ao comportamento desenvolvido por usuários quando em um processo de busca por informação; (b) às ações executadas em sistemas de busca; e (c) ao círculo hermenêutico que pode ser utilizado para dar suporte à concepção de componentes interativos que facilitem a busca por informação.

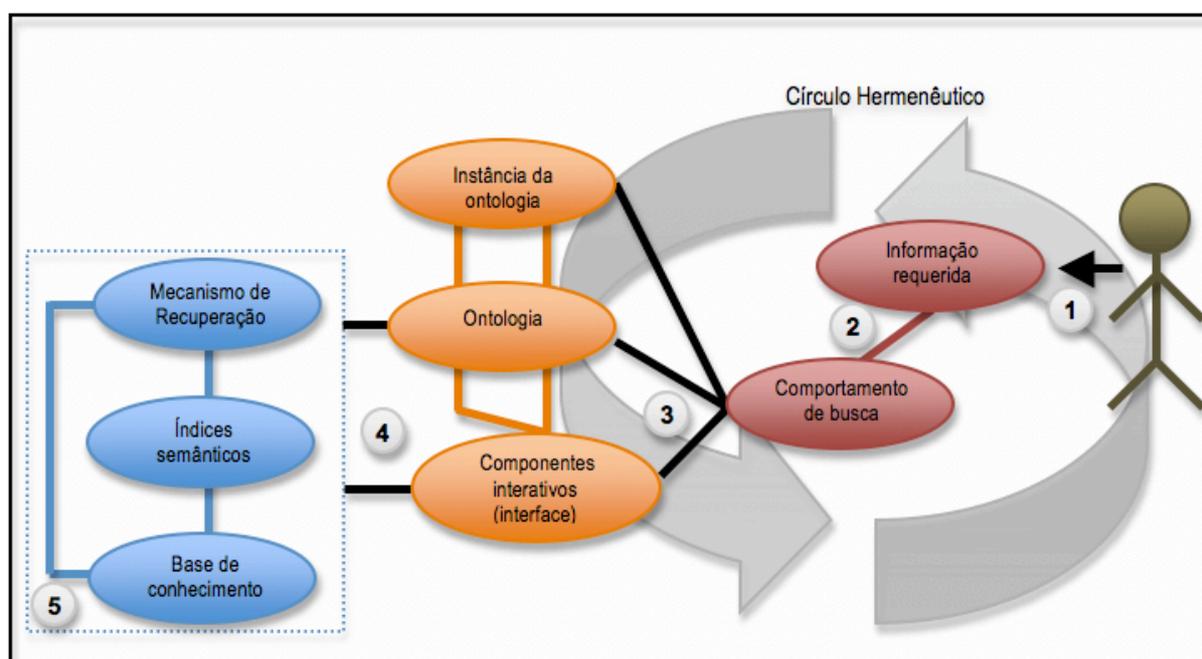


Figura 6 - Modelo proposto para RBI

Os componentes em vermelho, ilustrados na Figura 6, representam as atividades cognitivas (*i.e.*, informação requerida e comportamento de busca) que estão diretamente vinculadas à *information seeking* e ao círculo hermenêutico. Os componentes em laranja, que definem a camada intermediária, ou intermediário, como sugere a literatura de *information seeking* (Vakkari, 2003; Ingwersen e Järvelin, 2005b), são responsáveis por fazer a ponte entre o usuário e o sistema de busca. Os componentes em azul, que definem a camada de suporte, descrevem os módulos que dão suporte aos componentes da camada intermediária.

Um processo de busca inicia quando um usuário se depara com a falta de conhecimento (número 1 Figura 6) em um determinado assunto, entrando em um comportamento de busca (número 2). Nesse processo, o usuário interage com os componentes interativos (número 3), baseados na ontologia e nas instâncias, que compreendem a camada intermediária do modelo. Essa camada é subsidiada pelos componentes computacionais (número 4) descritos na última camada, que compreende o mecanismo de recuperação, índices semânticos e a base de conhecimento (número 5).

A seguir, as atividades cognitivas, inerentes à primeira camada, a camada de suporte e a camada intermediária, juntamente com seus respectivos componentes, são apresentadas em mais detalhes. Na seqüência é descrita a metáfora do círculo hermenêutico aplicada a um processo de busca e uma lista com tipos de interações é apresentada e discutida no âmbito do modelo proposto.

3.2.1 Atividades cognitivas

O usuário, a informação requerida e o comportamento de busca são componentes cognitivos que interferem na definição dos demais componentes. O modelo proposto tem como principal objetivo possibilitar que o usuário consiga desempenhar as atividades cognitivas inerentes a um processo de busca de uma maneira mais natural. Cada um desses três componentes é definido a seguir.

- ✚ **Usuário.** É o autor cognitivo que possui uma necessidade, ou identifica a falta de conhecimento, assumindo um comportamento de alguém que irá realizar uma busca. Esse usuário é responsável por interpretar os sinais gerados pela informação recuperada e pelas funcionalidades disponíveis na interface. Ele é o elemento que se comunica diretamente com a camada intermediária quando assume esse comportamento de busca;
- ✚ **Informação requerida.** Ocorre quando um usuário conscientemente identifica a falta de conhecimento sobre um determinado assunto (Ingwersen e Järvelin, 2005b). Case (2007) referenciando Grunig (1989) cita que a informação requerida é tipicamente caracterizada como um “estado motivacional interno” que produz pensamentos e ações. Belkin (1982) sugere que a percepção

da ausência de conhecimento induz um usuário a assumir um comportamento que o direciona à realização de uma busca;

- ✚ **Comportamento de busca (ou comportamento de *information seeking*).** É o comportamento humano baseado em pensamentos, sentimentos e ações (Kuhlthau, 1993), que se refere à geração, à comunicação, ao uso e a outras atividades relativas à busca por informação (Wilson, 1994). Case (2007) referenciando Marchionini (1997) cita que um comportamento de busca é “o processo no qual um ser humano intencionalmente engaja de modo a alterar seu estado de conhecimento e que está intimamente relacionado com aprendizagem e resolução de problemas”.

Cabe ressaltar que a informação requerida é diretamente influenciada por uma série de variáveis que implicitamente ajudam a moldar o contexto da informação desejada. Belkin (1993), Case (2007) e Ingwersen (2005b) descrevem que a informação requerida é diretamente influenciada pela situação em que ela ocorre e pelos contextos organizacional, social e cultural. Um usuário, de acordo com Saracevic (1997), também traz um determinado conhecimento, ou estado cognitivo, em relação a uma situação assim como uma intencionalidade que influencia, por exemplo, a especificação e a modificação de consultas, as táticas de busca e outros atributos ou interações inerentes a um processo de busca. Gadamer (1989) e Heidegger e Stambaugh (1996) descrevem, por meio do círculo hermenêutico, que uma pessoa possui seus preconceitos e que deve considerá-los durante um processo de interpretação ou entendimento de um novo assunto, ou nova informação. Dessa forma, o usuário, quando exercendo um comportamento de busca, age de acordo com essas influências no intuito de contextualizar a sua necessidade dentro das diversas variáveis inerentes ao processo de busca por informação.

Um comportamento de *information seeking*, aqui empregado especificamente no contexto de um sistema de busca, requer a utilização de diferentes procedimentos que visam auxiliar o usuário a entender uma necessidade e encontrar a informação requerida. Esses procedimentos são identificados especificamente como tarefas que afetam o usuário cognitivamente e permitem que ele consiga atravessar um processo de busca por informação da forma mais natural possível. Os tipos de interações descritos mais à frente na Tabela 2 (pg. 51) serviram

para a concepção do modelo proposto. Está claro que um ambiente de busca deve oferecer ferramentas interativas que possam atuar ativamente em um processo de *information seeking*.

3.2.2 Camada de suporte

A camada de suporte contém os componentes de *back-end* que devem oferecer suporte para a camada intermediária. O componente relativo ao mecanismo de recuperação é responsável por prover algoritmos ou recursos computacionais de forma a permitir que o conhecimento de um domínio, ou seja, a ontologia e a respectiva base de conhecimento, possa ser indexado e recuperado. Esse mecanismo precisa interpretar consultas que possam usufruir dos conceitos definidos na ontologia de forma a recuperar informação contextualizada.

O mecanismo de recuperação está diretamente conectado aos índices semânticos – índice, nesse caso, é a estrutura de dados responsável por organizar o conhecimento de domínio de modo a facilitar a recuperação. É importante enfatizar que a estrutura de dados do índice deve se basear nos conceitos descritos pela RI, pois um sistema de busca requer respostas rápidas e precisas. Quanto à precisão, especificamente, os índices semânticos devem armazenar o contexto de cada informação indexada e, portanto, possibilitar a recuperação de informação contextualizada.

A base de conhecimento deve armazenar as instâncias da ontologia. A estrutura de dados da base de conhecimento deve propiciar a recuperação total do conjunto de informações pertencentes a uma instância. Enquanto os índices semânticos devem ser concebidos com estruturas específicas que forneçam velocidade para a recuperação de informação, a base de conhecimento deve proporcionar a recuperação do conteúdo estruturado de cada instância.

A extração de informação adicional também é uma característica que os componentes da camada de suporte, especificamente o mecanismo de recuperação e a base de conhecimento, devem prover. Com a utilização de ontologia e suas instâncias, é possível utilizar técnicas de inferência, conforme descrito na revisão de literatura, para derivar conhecimento definido apenas implicitamente. Em uma ontologia sobre pesquisa, por exemplo, onde um usuário busca por pesquisadores de uma determinada área, o sistema de busca pode apresentar, além dos pesquisadores, as instituições e as linhas de pesquisa de tal área, mesmo que não haja uma relação direta entre os conceitos *instituição*, *linha de pesquisa* e *pesquisador*, definidos na ontologia.

3.2.3 Camada intermediária

Para que o usuário possa utilizar um sistema que facilite a busca por informação, é necessário que a interface contenha componentes interativos que possam agir como um intermediário de maneira a tornar exequível o processo de encontrar a informação requerida. Para Kuhlthau (1991), esse *intermediário* deve ser concebido de maneira a acomodar uma variedade de tarefas, em resposta a articulações relativas a problemas, desenvolvidas pelos usuários durante um processo de *information seeking*, tais como: buscas preliminares, exploratórias e compreensivas e sumários. Marchionini e Komlodi (1998) citam que “a interface serve como o canal de comunicação através do qual um processo de *information seeking* acontece”.

A formalização do conhecimento, a estrutura da ontologia e as suas instâncias possibilitam que a interface de um sistema de busca possa ser mais interativa através da construção de componentes mais bem elaborados. García e Sicília (2003) sugerem que uma ontologia pode ser usada em um sistema de busca de diferentes formas, como, por exemplo, na interface para auxiliar a composição de consultas, permitindo que usuários usem conceitos por ela definidos. Sugere-se, portanto, a utilização de ontologia como tecnologia que pode facilitar a construção de componentes interativos. Para Saracevic (1997), os usuários interpretam e julgam cognitivamente os resultados obtidos para então assimilá-los cognitivamente – os componentes interativos visam auxiliar esse processo de cognição. Com base nessa discussão, recomenda-se que os componentes interativos devem:

- ✚ disponibilizar a ontologia, preferencialmente em modo gráfico, para interação direta pelos usuários;
- ✚ recuperar informações e contextualizá-las no domínio de modo a possibilitar o refinamento de uma consulta com as informações recuperadas;
- ✚ permitir a composição de consultas que possam usufruir de conceitos definidos na ontologia (consultas semânticas) – formulação de consultas mais precisas;
- ✚ possibilitar a visualização das instâncias recuperadas agrupadas de acordo com as relações entre conceitos descritas na ontologia (e.g., em uma ontologia com dois conceitos relacionados, A e B, os usuários devem poder visuali-

zar as instâncias dos respectivos conceitos agrupadas tanto no formato A e B quanto no formato B e A);

- ✚ disponibilizar conhecimento adicional para cada informação recuperada e permitir que esse conhecimento possa ser utilizado de modo interativo; e
- ✚ ser dinâmicos, intuitivos e, principalmente, interativos.

Além disso, os componentes interativos também precisam estar interligados de modo a prover a dinamicidade requerida pela interação resultante de uma ação. Dessa forma, uma alteração em um componente automaticamente deve refletir sobre os demais componentes disponíveis na interface. Por exemplo, a composição de uma consulta deve, além de recuperar a informação específica, ser contextualizada na ontologia pois, dessa forma, o usuário se localiza no domínio e adquire uma nova informação que pode afetar a próxima ação a ser tomada.

A utilização de ontologia e instâncias também refletem sobre questões relativas ao círculo hermenêutico e à sua aplicação no âmbito do modelo proposto. Os componentes a serem disponibilizados na camada intermediária também devem atender as questões discutidas na próximo tópico, referente à utilização do círculo hermenêutico no contexto de um processo de busca por informação. E no tópico posterior há a apresentação de tipos de interação que também fornecem subsídios para a construção de um ambiente de busca mais interativo.

3.3 A metáfora do círculo hermenêutico aplicada ao processo de busca por informação

O atividade hermenêutica é o contexto no qual o usuário de um sistema de busca obtém novas idéias sobre a informação em que se está navegando. Para ser efetivo, os designers de um sistema de busca precisam estruturar um contexto em que permita aos usuários engajarem na espontaneidade de uma atividade interpretativa. O modelo proposto considera quatro componentes da hermenêutica em que cada um dá suporte aos componentes sugeridos na camada intermediária do modelo, como pode ser visto na Tabela 1.

Tabela 1 - Componentes da hermenêutica relacionados com os componentes do modelo

Componentes da hermenêutica	Componentes do modelo
Pré-estrutura	Ontologia computacional
Objeto	Instâncias
Interatividade	Interação dos usuários com as instâncias e a ontologia computacional
Aplicação	A questão do usuário

Primeiro, a *pré-estrutura* (Gadamer chama de pré-entendimento) que é a idéia inicial que o usuário possui sobre um assunto para iniciar a navegação sobre um conjunto de dados. No modelo, essa *pré-estrutura* corresponde à ontologia computacional. Segundo, é o *objeto*, que representa “as coisas” (do inglês *things*), que corresponde ao mundo objetivo, aos fatos, aos dados acumulados. No modelo, essas “coisas” são representadas pelas instâncias dos conceitos da ontologia. Finalmente, há a interatividade e a aplicação. A interatividade é representada pela interação do usuário com as instâncias e com a ontologia. Os usuários *vêm e vão* entre as instâncias e os conceitos sempre tendo em mente uma *aplicação*.

Por exemplo, um estudante que deseja fazer um pós-doutorado pode iniciar um processo de busca com um assunto em mente (*e.g.*, Web semântica). Com o resultado de uma busca, o estudante se depara com um pesquisador (*e.g.*, Tim Berners-Lee) que passa, então, a ser o seu objetivo. Dando prosseguimento à busca, o estudante encontra a instituição onde o pesquisador atua (*e.g.*, M.I.T.). O foco muda novamente, e agora o interesse não é mais sobre o pesquisador, mas sobre o local onde ele trabalha. Nesse processo o estudante altera constantemente o seu interesse como se estivesse em um ciclo – dos conceitos (*e.g.*, assunto, pesquisador e instituição) para as instâncias (*e.g.*, Web semântica, Tim Berners-Lee e M.I.T.) e de volta para os conceitos. A Figura 7 apresenta essa exemplificação graficamente.

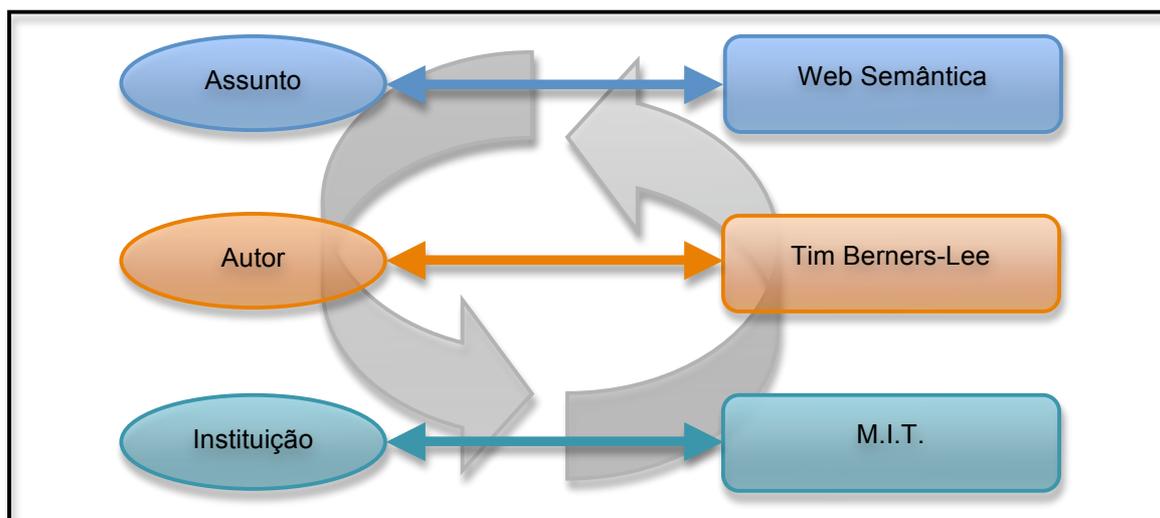


Figura 7 – Exemplificação da metáfora do círculo hermenêutico aplicado ao processo de busca

Entendimento e interpretação, conforme descreve Gadamer (1989), são aplicados a um texto e, aqui, aplicado a um assunto específico que é objetivo da questão inicial do usuário. A tarefa do interpretador é adaptar o texto à situação concreta na qual o texto realmente se expressa. O mesmo é válido aqui no entendimento de um assunto através do processo de *vai e vem* entre os conceitos e as instâncias. O modelo tem por objetivo oferecer ao usuário a mídia que permita esse tipo de entendimento, ou seja, o círculo hermenêutico é a metáfora da atividade do *vai e vem* entre as instâncias e os conceitos, que visa moldar as questões do usuário.

3.4 Tipos de interação discutidas no âmbito do modelo proposto

Como a interatividade é parte fundamental de um modelo de RBI, a Tabela 2 apresenta uma lista com tipos de interações e as contextualiza no âmbito do modelo proposto. É importante enfatizar que as interações desenvolvidas por um usuário, quando em um comportamento de *information seeking*, podem ser de cunho físico e cognitivo. As interações aqui descritas são referenciadas na literatura (Ide, 1971; Rocchio, 1971; Rouse e Rouse, 1984; Bates, 1989; Roberts, 1989; Salton e Buckley, 1990; Belkin e Croft, 1992; Ingwersen, 1992; Belkin, 1993; Belkin, Marchetti *et al.*, 1993; Kuhlthau, 1993; Hendry e Harper, 1996; Ingwersen, 1996; Leckie, Pettigrew *et al.*, 1996; Hearst e Karadi, 1997; Marchionini, 1997; Saracevic, 1997; Wilson, 1997; Spink e Saracevic, 1998; Savage-Knepshield e Belkin, 1999; Wilson, 1999; Browne e Rogich, 2001; Reiterer, Müller *et al.*, 2001; Sonnenwald, Wildemuth *et al.*, 2001; Spink, Wilson *et al.*, 2002; Toms, 2002; García e Sicilia, 2003; Järvelin e Wilson, 2003;

Vakkari, 2003; Järvelin e Ingwersen, 2004; Albertoni, Bertoni *et al.*, 2005; Ingwersen e Järvelin, 2005b; Kuhlthau, 2005; Case, 2007), principalmente em trabalhos específicos da área de *information seeking*.

Tabela 2 - Exemplos de métodos de interação inerente a um processo de busca

Interação	Descrição	Contextualização
Aprendizagem	É a aquisição de novo conhecimento por meio da identificação de recursos apropriados (e.g., informação recuperada). O aprendizado é uma tarefa cognitiva constante durante um processo de <i>information seeking</i> .	No modelo proposto, esse tipo de interação é melhor contextualizada com a utilização do círculo hermenêutico. O usuário possui uma idéia inicial, ou pré-concepções, e à medida que interage com um sistema de busca, o usuário desenvolve novas percepções. Como esse processo é iterativo, o usuário adquire novos conhecimentos.
Busca	Atividade desenvolvida com o objetivo de obter informação com vistas a suprir uma necessidade ou falta de conhecimento sobre um determinado assunto.	O modelo proposto tem como principal objetivo subsidiar a construção de um ambiente de busca que permita aos usuários encontrarem respostas para as suas questões ou necessidades agindo da forma mais natural possível.
Composição de consulta	Digitar termos que representam a informação requerida – tradução da informação requerida em termos. É a forma mais comum de interagir com um sistema de busca.	O modelo sugere componentes interativos que permitam aos usuários digitar termos que representem suas necessidades. Também se pressupõe a utilização de conceitos definidos na ontologia para a composição de consultas semânticas.
Escaneamento	Atividade em que uma pessoa olha cuidadosamente todas as partes de maneira a detectar alguma característica importante. Não está vinculado diretamente a um objetivo predeterminado, podendo ser apenas uma motivação para exploração, curiosidade, etc.	O usuário precisa se deparar com características disponíveis na interface de um sistema que lhe permita rapidamente identificar informações relevantes. Na camada intermediária do modelo há a definição de características que devem ser inseridas em componentes interativos, o que facilitam a atividade de escaneamento.
Intencionalidade	São razões emocionais/cognitivas que induzem uma pessoa a desenvolver atividades mentais e físicas, tais como, buscar informação.	Também está relacionada a um propósito ou objetivo. Tem uma relação forte com a aprendizagem. Essa questão de intencionalidade está discutida no âmbito da camada cognitiva do modelo. Um usuário descobre a falta de um conhecimento que o induz a entrar em um comportamento de busca.
Navegação	É uma atividade exploratória que varia entre uma busca com objetivo predeterminado a uma busca casual. É um comportamento de <i>information seeking</i> . Extremamente útil por viabilizar usuários a se moverem facilmente entre informações que podem estar conectadas. Não requer que usuários traduzam a informação (e.g., por meio de consultas) de modo a efetuar uma nova busca.	No modelo, os componentes da camada intermediária e a metáfora do círculo hermenêutico possibilitam, ou melhor, induzem a construção de um ambiente onde a navegação deve ser explorada em sua totalidade. Por exemplo, com a disponibilização das instâncias, representando a informação recuperada, o usuário pode clicar sobre qualquer parte dessa informação e automaticamente fazer um refinamento da busca. Além disso,

Interação	Descrição	Contextualização
Reformulação de consulta	Alterar a consulta inserindo novos termos ou excluindo os termos originais. É mais fácil reformular a consulta do que defini-la pela primeira vez.	o processo de <i>vai e vem</i> entre os conceitos e as instâncias, referente ao círculo hermenêutico, é uma forma de navegação mais inteligente, pois o usuário consegue contextualizar a informação no domínio além de facilitar o processo de entendimento, é claro. Cada informação recuperada pode ser utilizada para reformulação de uma consulta, como está definido pelo modelo. Um exemplo foi descrito na <i>navegação</i> sobre a recuperação de informação contextualizada. Outro exemplo é a disponibilização de conhecimento adicional por meio de um componente interativo. Com isso, o usuário pode reformular uma consulta com informação que foi inferida a partir da ontologia e suas instâncias – ou seja, composição de consultas mais precisas.
Refinamento	Procedimento que permite ao usuário utilizar o conteúdo, ou partes dele, para fazer uma redução de escopo automaticamente. O refinamento ocorre a partir de uma primeira busca. Esse procedimento pode ser facilitado, em um ambiente de busca, por ferramentas que permitam, por exemplo, que a informação recuperada possa ser utilizada de maneira dinâmica.	Exemplos descritos nos tipos de interação <i>navegação</i> e <i>reformulação de consulta</i> descrevem alguns tipos de refinamento que devem ser suportados pelos componentes interativos da camada intermediária descritas pelo modelo proposto.
Relevância	É uma avaliação cognitiva sobre a informação recuperada em relação a uma necessidade. Pode mudar dinamicamente à medida que um usuário procede em um comportamento de <i>information seeking</i> .	Esse tipo de interação também está vinculada ao ambiente em que o usuário atua. O modelo propõe a construção de um ambiente de busca que possibilite que tal tarefa ocorra de modo mais natural possível. Para tal, sugere-se o máximo possível de características interativas. Além disso, a interação também tem relação com o círculo hermenêutico, uma vez que o entendimento ou a interpretação de informação recuperada afeta diretamente as decisões dos usuários em definir o que é relevante.
Serendipidade	É a descoberta casual de informação. Uma pessoa buscando informações sobre um assunto A descobre informações sobre um assunto B.	Em um ambiente interativo de busca, conforme o modelo sugere, a possibilidade de navegação e escaneamento juntamente com as atividades de busca em si viabilizam a serendipidade. Um ambiente interativo torna a serendipidade mais fácil de ocorrer, pois os usuários conseguem visualizar e manipular mais informações e em diferentes perspectivas (<i>e.g.</i> , gráfico, agrupamento, etc.).
Tentativa e erro	Ações inerentes a um processo de busca e aprendizagem em que um usuário seleciona uma possível informação como resposta a um problema e, caso não esteja correta,	Um ambiente de busca deve proporcionar a seleção de informação de modo a permitir que um usuário consiga trabalhar a sua necessidade por meio da atividade de tentativa e erro. Reformulação de consultas, navega-

Interação	Descrição	Contextualização
	seleciona outra informação.	ção, refinamento e acesso à informação adicional são exemplos de tipos de interação que facilitam um processo de tentativa e erro. No âmbito do modelo, especificamente, a atividade de <i>vai e vem</i> entre as instâncias e os conceitos referentes a metáfora do círculo hermenêutico também facilitam esse processo.
Verificação	Essa é uma tarefa em que o usuário é responsável por avaliar a informação recuperada e tomar decisões em relação à necessidade de continuar o processo de busca ou qual caminho deve ser tomado.	Mesmo sendo uma tarefa estritamente cognitiva, ela deve ser considerada em um ambiente de busca, pois o usuário precisa definir diferentes caminhos em caso de continuar um processo de busca. Um ambiente de busca deve permitir essa mudança de direção da maneira mais natural possível, conforme descrito pelo modelo. A verificação também tem relação com o círculo hermenêutico no que diz respeito ao entendimento ou à interpretação sobre as informações recuperadas – o que direciona as novas ações a serem tomadas.

3.5 Considerações finais

O objetivo desse capítulo foi apresentar o modelo proposto de RBI. Recorreu-se à *information seeking*, ontologia, RI e círculo hermenêutico para conceber um modelo que esteja mais adaptado ao processo de busca realizado por seres humanos e que seja mais pragmático do ponto de vista de aplicação. Essas áreas de pesquisa forneceram subsídios para entender um comportamento de busca e a forma como seres humanos interpretam ou compreendem novas informações, o que facilitou a definição da proposta.

O modelo é composto de nove componentes que representam desde o usuário com o seu comportamento de busca, fruto da falta de conhecimento em algum assunto específico, até os mecanismos requeridos por um sistema de RI. Uma lista com tipos de interação também foi descrita. Cada tipo de interação apresenta uma ou mais características que devem ser consideradas no design de um ambiente de busca. A lista apresenta ainda a contextualização de cada tipo de interação no âmbito no modelo.

A ontologia, de acordo com esta proposta, possibilita a construção de um ambiente que aja como um intermediário no processo de busca, sendo a ponte entre a necessidade por informação e o acesso à solução para essa necessidade. Com acesso direto à ontologia, os usuários podem interagir com os conceitos e as suas instâncias. Essa interação entre os concei-

tos e as instâncias ocorre na forma de *vai e vem* – essa é a metáfora do círculo hermenêutico aplicada ao processo de busca por informação.

Como um dos aspectos levados em consideração na definição do modelo foi com relação à sua aplicabilidade, um framework foi concebido. O desenvolvimento do framework visa validar os conceitos definidos no modelo. Além disso, definições do modelo que porventura não estão claras, pelo menos em se tratando de sua aplicação prática, podem ser mais bem compreendidas por meio do framework. No próximo capítulo apresenta-se o framework *Hermeneus* e a sua arquitetura.

4 FRAMEWORK HERMENEUS

4.1 Introdução

A aplicação do modelo, descrito no tópico anterior, é proposta por meio de um framework que visa implementar os conceitos definidos pelo modelo de forma a permitir que um sistema de busca interativo possa ser um intermediário, conforme descrito por Kuhlthau (1991; 1993), no processo de busca por informação. Objetiva-se, portanto, proporcionar facilidades para que o usuário mude de uma condição de incerteza para uma condição de solução. Conforme previamente descrito na introdução deste trabalho, foi decidido chamar o framework de Hermeneus, que em grego significa “interpretar” ou “traduzir” (Kingsley, 1993).

O framework utiliza uma ontologia de domínio como artefato principal para a criação de módulos e componentes interativos. Ele é composto de quatro módulos, os quais são responsáveis por (1) indexar instâncias de uma ontologia, (2) recuperar instâncias de uma ontologia, (3) inferir conhecimento adicional e (4) apresentar componentes interativos. Esses módulos foram concebidos com base em definições discutidas pelo modelo. Segundo o modelo, as instâncias de uma ontologia devem ser indexadas de maneira a permitir a recuperação de informação contextualizada. O Hermeneus contém o módulo de indexação, o qual é responsável por extrair a informação contida na base de conhecimento, representada por meio de instâncias de ontologia, e por indexar tais informações visando à recuperação de instâncias e à aplicação de consultas contextualizadas. O módulo de recuperação do Hermeneus implementa os conceitos definidos pelo mecanismo de recuperação descrito pelo modelo. Esse módulo interpreta as consultas formuladas pelos usuários a partir da interface e aplica-as sobre os índices semânticos.

O modelo também descreve a necessidade de disponibilização de componentes interativos que, no Hermeneus, é implementado pelo módulo de apresentação. Há três componentes interativos: (1) o primeiro apresenta a ontologia e permite que o usuário possa interagir diretamente com esse conhecimento durante o processo de busca por informação; (2) o segundo é responsável por apresentar as instâncias recuperadas e possibilitar que o usuário utilize partes dessas instâncias para fazer refinamento na busca; e (3) o terceiro componente interativo está diretamente relacionado ao módulo de inferência. Esse componente visa extrair informação adicional de uma instância recuperada e apresentar a informação extraída de mo-

do a permitir que o usuário também consiga fazer refinamento ou até mesmo recuperar informação configurada a partir de consultas mais complexas previamente estabelecidas por um especialista de domínio.

O Hermeneus foi concebido por meio de módulos e de componentes. Essa modularidade permite que o framework seja flexível o suficiente para ser adaptado mais facilmente às especificidades requeridas por determinados domínios. Além disso, o Hermeneus pode ser configurado por especialistas de domínio. É importante enfatizar que se subentende como um especialista de domínio, no âmbito deste trabalho, um ator que tem conhecimento sobre o domínio da aplicação e que possui habilidades para construir uma ontologia e configurar o framework.

O framework proposto pode ser aplicado a qualquer domínio – um domínio é representado por meio de uma ontologia de domínio. Os módulos pertencentes à arquitetura do Hermeneus utilizam a ontologia para executarem automaticamente as suas funções predefinidas (*e.g.*, com a ontologia e a base de conhecimento os índices semânticos são gerados automaticamente) – por isso requerem apenas configuração por parte dos especialistas. A seguir, descreve-se a arquitetura do Hermeneus e, no capítulo sobre demonstração de viabilidade, apresenta-se um protótipo que auxilia a validação desta proposta tanto no âmbito do modelo quanto no âmbito do framework *per se*.

4.2 Arquitetura do Hermeneus

O principal objetivo em desenvolver um sistema de busca é facilitar o acesso do usuário à informação desejada (Rijsbergen, 1979; Korfhage, 1997; Baeza-Yates e Ribeiro-Neto, 1999; Witten, Moffat *et al.*, 1999). Entretanto, construir sistemas de busca exige, entre outras coisas, conhecimento sobre a área de RI. Além disso, um desenvolvedor necessita traduzir o conhecimento de um domínio em termos que podem conduzir à implementação de um sistema. O Hermeneus trata essa questão ao utilizar uma ontologia como guia para a construção de um sistema de busca interativo. Dessa forma, um especialista de domínio sem background na área de RI pode construir um sistema de busca interativo de forma a possibilitar que o conhecimento do domínio seja disponibilizado por meio de um ambiente interativo.

A Figura 8 apresenta uma visão geral da ideia do papel do especialista de domínio. Basicamente, ele é responsável por desenvolver uma ontologia, construir a base de conheci-

mento e configurar os módulos requeridos pelo Hermeneus. Os módulos precisam, essencialmente, ter acesso à ontologia e à base de conhecimento. O módulo de inferência requer configurações adicionais devido à necessidade de definição de regras específicas para se extrair informação adicional. Após a configuração dos módulos, os índices semânticos são gerados automaticamente. O sistema de busca resultante visa prover um ambiente interativo que permite aos usuários não apenas buscarem informação sobre um domínio, mas também desenvolverem as suas necessidades, isto é, entender qual é a falta de conhecimento que os obrigou a entrarem em um comportamento de busca por informação. Cada um dos quatro módulos são descritos nos subtópicos a seguir.

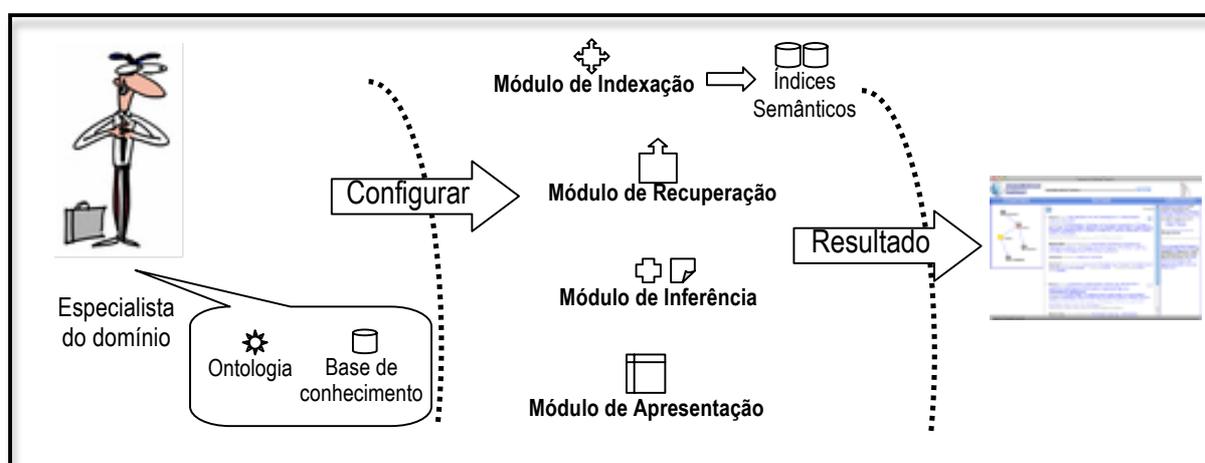


Figura 8 - Especialista de um domínio configura o Hermeneus

4.2.1 Módulo de indexação

O módulo de indexação utiliza uma ontologia de domínio e sua base de conhecimento (*i.e.*, as instâncias) para criar índices semânticos automaticamente. Basicamente, esse módulo acessa a ontologia com o objetivo de identificar os conceitos por ela definidos e as relações entre os conceitos – novamente, um conceito é definido por uma classe que contém propriedades. O módulo de indexação, então, lê as instâncias da ontologia que estão armazenadas na base de conhecimento com o intuito de indexar o seu conteúdo e, portanto, gerar os índices semânticos, que são baseados na estrutura de índice invertido.

Índice invertido, segundo Baeza-Yates e Ribeiro-Neto (1999), é a estrutura mais comum para indexar informação de maneira a permitir um melhor desempenho durante uma tarefa de busca. A estrutura do índice invertido foi adaptada para armazenar a informação semântica, alteração esta que adiciona uma nova dimensão à estrutura do índice, agora chama-

do de índice semântico, o qual permite a identificação de termos em contextos específicos. Ou seja, cada termo indexado, além de conter as informações requeridas pela estrutura do índice invertido clássico, também contém a descrição do conceito a que ele está vinculado. A Figura 9 mostra um exemplo de como a informação semântica é anexada à estrutura do índice invertido. No exemplo é possível visualizar uma instância de *paper* cujo identificador (*i.e.*, *Id*) contém o valor “*Inst_1*” e o título (*i.e.*, *title*) é representado pelos termos “*Ontology-Driven Information Retrieval*”. Cada termo do identificador e do título é armazenado em uma lista que possui apontamentos para outra estrutura, também em forma de lista, que contém as informações relativas à identificação de quais instâncias o termo está associado. Ainda na Figura 9 é possível identificar as letras (a), (b) e (c), que indicam as informações presentes na estrutura que vincula um termo a uma instância, ou seja, a identificação da instância à respectiva classe ou conceito e, por fim, a posição do termo na instância. O exemplo mostra que o termo “*Driven*” está associado à instância “*Inst_1*” e está relacionado ao conceito “*Paper*” na posição 2 (*i.e.*, “*Driven*” é o segundo termo). A descrição do conceito é a informação semântica que contextualiza o termo em uma instância.

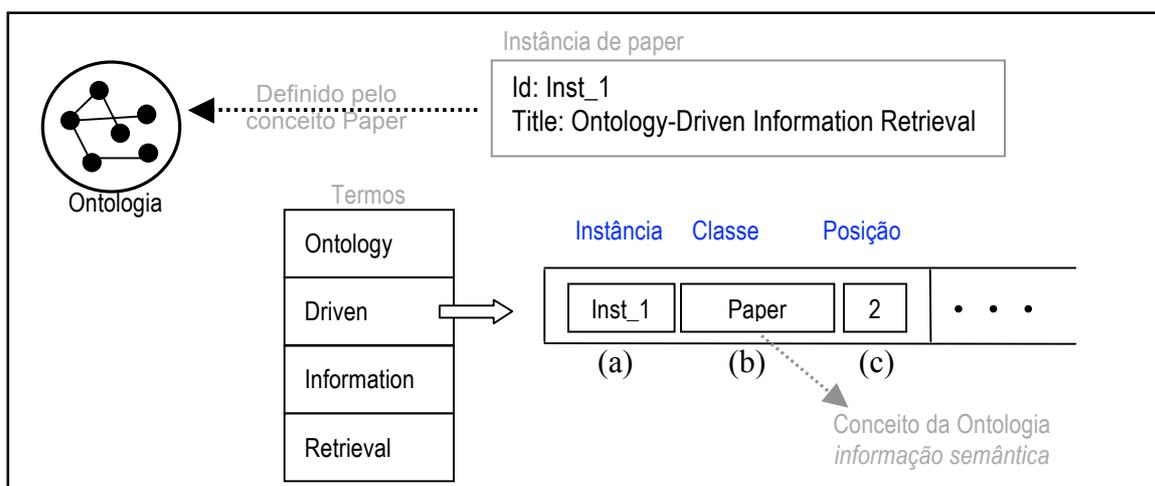


Figura 9 - Índice invertido com informação semântica

O Hermeneus cria um índice semântico para cada conceito definido pela ontologia – cada índice armazena instâncias específicas sobre o conceito e sobre as instâncias de conceitos que estão diretamente relacionados ao conceito. Por exemplo, em uma ontologia que possui apenas três conceitos, *autor*, *paper* e *instituição*, como pode ser visto na Figura 10, três índices semânticos são construídos. O índice semântico de *paper*, por exemplo, contém as informações de instâncias de *papers* e de instâncias de conceitos diretamente relacionados ao

conceito *paper*. Considerando-se que há uma relação direta entre os conceitos *paper* e *autor*, mas não entre *paper* e *instituição*, o índice de *paper* contém, também, as instâncias do conceito *autor*. Essa característica do Hermeneus permite que usuários consigam formular consultas contextualizadas e visualizar as instâncias recuperadas em diferentes perspectivas sem haver perda de performance, pois as instâncias estão armazenadas em índices estruturados de acordo com conceitos e suas relações declarados na ontologia (e.g., visualizar o resultado não somente agrupado pelas instituições e seus pesquisadores, mas também agrupado pelos pesquisadores e suas instituições).

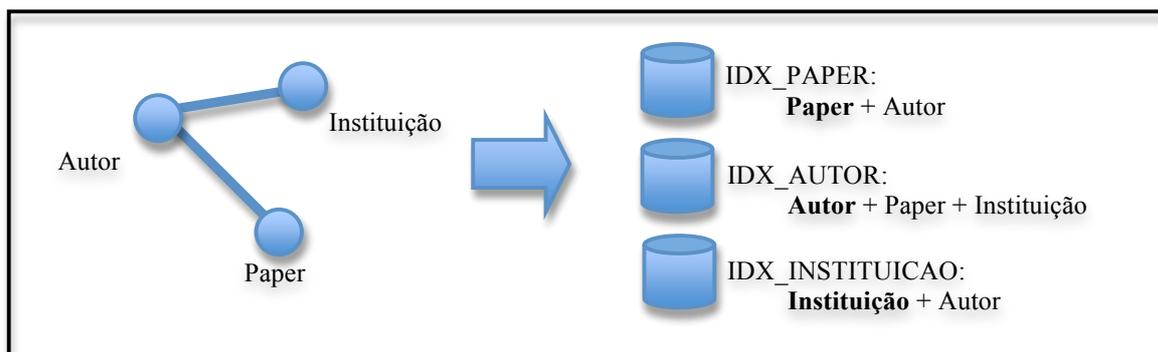


Figura 10 - Exemplo: a partir de uma ontologia um índice para cada conceito é gerado

4.2.2 Módulo de recuperação

O módulo de recuperação é responsável por recuperar as instâncias da ontologia que estão armazenadas nos índices semânticos em resposta a uma consulta. Esse módulo também é responsável por ordenar as instâncias recuperadas. Para tal, utiliza-se a medida de similaridade chamada *coordinate matching* descrita por Witten e Moffat (1999). Basicamente, essa medida de similaridade considera que quanto maior o número de termos de uma consulta em um documento maior a probabilidade do documento ser relevante – como no Hermeneus o índice é formado por instâncias, aquelas instâncias que possuem o maior número de termos da consulta são consideradas mais relevantes.

Como esse módulo também é responsável por interpretar consultas semânticas, restringiu-se as instâncias recuperadas apenas àquelas que respondem exatamente ao conteúdo descrito pela consulta semântica. Por exemplo, uma consulta semântica que vincula os termos “*engenharia do conhecimento*” ao conceito “*curso*” retorna, obrigatoriamente, apenas as instâncias que possuem os termos “*engenharia do conhecimento*” pertencentes as instâncias do conceito “*curso*”. Essa característica permite que não seja necessária nenhuma alteração na

medida de similaridade, pois as instâncias que serão ordenadas são aquelas que se adaptam tanto a uma consulta simples, que é composta por um ou mais termos sem contexto, quanto a uma consulta semântica. Uma consulta semântica é composta pela descrição de um conceito, que está definido na ontologia, seguido pelos respectivos termos (e.g., *autor:Pacheco*). A Figura 11 apresenta a descrição em EBNF¹ do possível formato de consulta a ser aplicado em uma busca.

```

search ::= query | semantic_query
query ::= term { term }
semantic_query ::= concept ":" term { term } ";" { concept ":" term { term } ";" }
concept ::= <classes described in the ontology>
term ::= letter { letter } | digit { digit }
letter ::= "a" | "b" | ... | "z"
digit ::= "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"

```

Figura 11 - Descrição em EBNF de uma consulta aplicada em uma busca

Além disso, uma consulta composta, contendo termos sem contexto e termos com descrição semântica, também pode ser interpretada. Nesse caso, uma instância para ser recuperada precisa conter os termos sem contexto e os termos contextualizados diretamente associados aos respectivos conceitos. Por exemplo, na consulta “*inteligência artificial autor:Fonseca*”, apenas as instâncias que contêm os termos “*inteligência*” e “*artificial*”, não importando em quais conceitos esses termos estão vinculados, e o termo “*Fonseca*” vinculado diretamente ao conceito “*autor*” são recuperadas. É importante enfatizar que também é possível utilizar os operadores lógicos que normalmente (Baeza-Yates e Ribeiro-Neto, 1999) estão disponíveis em sistemas de busca (i.e., OR, AND e NOT). Na presente proposta, em uma consulta que possui mais de um termo e não há operadores explicitados, o operador lógico padrão utilizado é o AND.

A Tabela 3 apresenta alguns exemplos de consultas válidas e não válidas. Como referência de ontologia é utilizado o exemplo apresentado na Figura 10, que descreve uma ontologia com apenas três conceitos: *autor*, *paper* e *instituição*. As consultas consideradas não válidas são aquelas que possuem uma nomenclatura que indica a utilização de um conceito, ou seja, uma consulta semântica, mas na verdade não é. Na quarta consulta da Tabela 3 há uma consulta semântica que não é válida por não conter um conceito, antes dos dois-pontos, definido na ontologia (i.e., UFSC não é um conceito definido pela ontologia da Figura 10). O

¹ EBNF – *Extended Backus-Naur Form* (Scowen, 1993): notação utilizada para descrever linguagens formais.

mesmo problema ocorre com as consultas das linhas 7 (*i.e.*, PSU não é conceito) e 8 (*i.e.*, 2004 não é conceito).

Tabela 3 - Exemplos de consultas válidas e não válidas

	Exemplos de consultas	Válida
1.	Engenharia OR Gestão AND Conhecimento	Sim
2.	Inteligência artificial AND Fonseca	Sim
3.	Autor:Pacheco 2008	Sim
4.	UFSC:instituição	Não
5.	Ontologia autor:Fonseca	Sim
6.	Autor:Pacheco Instituição:UFSC	Sim
7.	Paper:engenharia PSU:UFSC	Não
8.	IA NOT distribuída AND 2004:paper	Não

4.2.3 Módulo de apresentação

O módulo de apresentação utiliza uma ontologia de domínio para disponibilizar um ambiente interativo na interface. Esse módulo contém três componentes: navegador de ontologia, instâncias recuperadas e informação adicional. A ontologia é usada pelos três componentes, o que permite uma maior integração entre eles. São esses componentes que permitem aos usuários entrarem em um comportamento de busca atuando em um ambiente interativo. A seguir, são descritos cada um desses componentes e suas funcionalidades e, no capítulo 5, é apresentado um protótipo que contém algumas figuras que ajudam a entender melhor cada um dos componentes aqui descritos.

4.2.3.1 Componente de navegação de ontologia

O componente de navegação de ontologia apresenta a ontologia graficamente de forma a permitir que os usuários possam interagir com o conhecimento do domínio. O componente também auxilia a composição de consultas semânticas, pois os usuários podem visualizar os conceitos definidos pela ontologia e utilizá-los na composição de consultas – o que está de acordo com o modelo proposto, que sugere a possibilidade de utilização de consultas semânticas.

Outra funcionalidade também atribuída a esse componente refere-se à visualização da lista de resultados em diferentes perspectivas. Essa funcionalidade permite que usuários agrupem as instâncias resultantes de acordo com um conceito chamado *conceito central*. Essa

funcionalidade, ao se utilizar os exemplos de interatividade descritos na Tabela 2 (pg. 51), auxilia, por exemplo, na navegação, pois permite visualizar a informação recuperada agrupada por diferentes conceitos; também auxilia na execução de tarefas estritamente cognitivas, como, por exemplo, verificação, escaneamento e aprendizagem (mais detalhes sobre a importância dessa funcionalidade são discutidos no capítulo sobre demonstração de viabilidade). Além disso, por ser um componente visual, algumas características gráficas são implementadas visando auxiliar o processo cognitivo, no que diz respeito à interação direta com a ontologia, como, por exemplo, o escaneamento (exemplo de tipo de interação descrito na Tabela 2 – pg. 51): o *conceito central* é apresentado em forma de quadrado, e todos os conceitos que estão sendo utilizados em uma consulta semântica são pintados com uma cor diferente.

4.2.3.2 Componente de instâncias recuperadas

O componente de instâncias recuperadas tem como responsabilidade apresentar as instâncias recuperadas, *i.e.*, as instâncias que atendem a uma consulta. As informações das instâncias, que contêm conteúdo semântico, são disponibilizadas com o intuito de permitir que o conteúdo possa ser empregado em um processo de refinamento. Para tal, cada propriedade das instâncias pode ser utilizada para fazer a composição de uma consulta semântica. Esse procedimento é executado por meio de cliques de mouse, o que facilita a utilização do conteúdo recuperado de maneira mais dinâmica e evita que os usuários tenham de formular novas consultas manualmente quando se deparam com um conteúdo resultante que está diretamente relacionado à informação requerida. Além disso, a forma com que as instâncias recuperadas são disponibilizadas auxilia no processo de navegação, verificação e relevância e facilita o processo de tentativa e erro executado pelos usuários.

4.2.3.3 Componente de informação adicional

O componente de informação adicional permite que usuários visualizem, a partir da seleção de uma determinada instância, a informação que não está explicitamente definida tanto na ontologia quanto na base de conhecimento, e que interajam com essa informação. Para cada instância recuperada novas informações podem ser extraídas e disponibilizadas. Assim como o componente de instâncias recuperadas, esse componente permite que as informações adicionais possam ser utilizadas em um processo de refinamento. A extração de informação adicional auxilia, também, no processo de aprendizagem, uma vez que permite aos usuários interagir com informação não declarada explicitamente – visualização e interação com um novo

conhecimento. Cabe ressaltar que a informação adicional extraída está vinculada à configuração definida por especialistas de domínio. Esse componente está diretamente relacionado com o módulo de inferência que utiliza técnicas de inferência juntamente com regras pré-configuradas para a extração de conhecimento declarado apenas de maneira implícita.

4.2.4 Módulo de inferência

O módulo de inferência é responsável por extrair informação adicional de uma determinada instância recuperada por meio da aplicação de técnicas de inferência e regras pré-configuradas sobre a ontologia e a sua base de conhecimento. Como o especialista do domínio possui o conhecimento sobre o domínio, ele é responsável por configurar essas regras. De acordo com Golbreich (2004) e Hyvönen et al. (2005), as regras são úteis para representar a parte dedutiva do conhecimento.

O Hermeneus aceita dois tipos de regras, as quais são aplicadas no âmbito da ontologia e das instâncias. As regras de nível de ontologia são aplicadas diretamente sobre a estrutura da ontologia. Essa característica permite a criação de novos relacionamentos entre os conceitos e a especificação de novas restrições. Por exemplo, a Figura 12 apresenta uma ontologia sobre família que descreve alguns relacionamentos diretos (*i.e.*, pai, mãe, filho, filha e casal). Por meio de regras de nível de ontologia é possível definir relacionamentos indiretos (*i.e.*, genro/sogro e neto/avô).

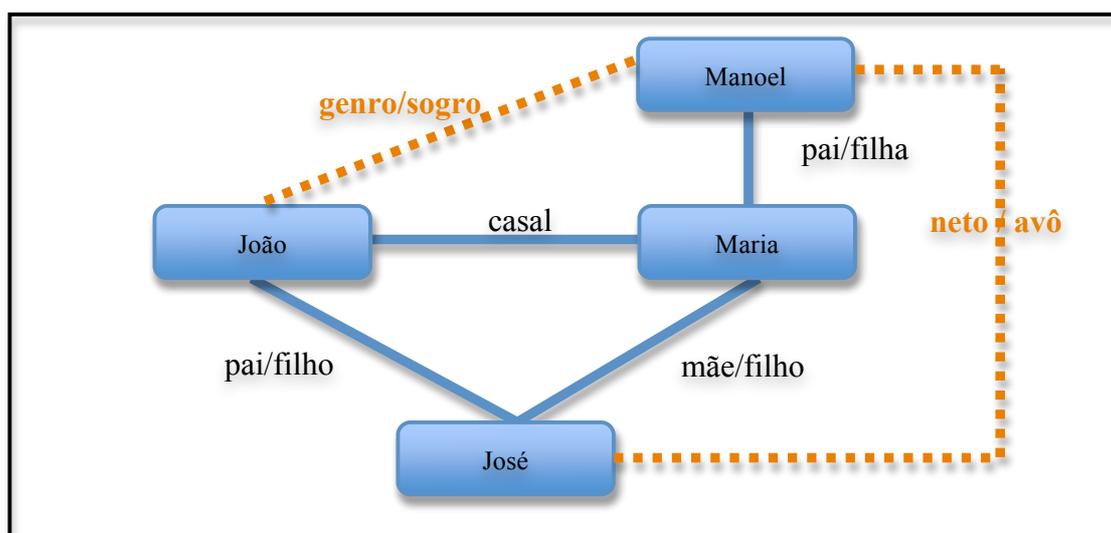


Figura 12 - Exemplo de ontologia de família com relacionamentos diretos e inferidos

As regras de nível de instância, também conhecidas como regras de nível de consulta (Golbreich, 2004), permitem a extração de informação adicional a partir das instâncias. Tais regras podem ser configuradas para utilizar as regras definidas no âmbito da ontologia. Por exemplo, considerando-se a ontologia de família descrita anteriormente na Figura 12, é possível configurar regras de nível de instância para extrair relacionamentos indiretos para uma determinada pessoa (e.g., extrair o gênero de “Manoel” e o avô de “José”).

A linguagem para representação de ontologia a ser usada no framework Hermeneus é a *Web Ontology Language* - OWL (Mcguinness e Harmelen, 2004; Cardoso, 2007), e a ferramenta Jena (Mcbride, 2002; Carroll, Dickinson *et al.*, 2004; Cardoso, 2007) é usada para manipular as instâncias e construir a base de conhecimento. Além disso, Jena também é usada para descrever e interpretar as regras usadas para inferir novo conhecimento. As regras de nível de ontologia devem ser descritas por meio da linguagem *Jena Rule Language* (Mcbride, 2002; Carroll, Dickinson *et al.*, 2004) enquanto as regras de nível de instâncias devem ser descritas por meio da linguagem SPARQL (Prud'hommeaux e Seaborne, 2007). A Figura 13 apresenta exemplos de regras de nível de ontologia, na parte esquerda, e regras de nível de instância aplicadas à ontologia da família, disponível na Figura 12.

<pre>[rule1: (?Manoel NS#pai ?Maria) (?Maria NS#mãe ?José) -> (?Manoel NS#avô ?José)] [rule2: (?João NS#casal ?Maria) (?Maria NS#filha ?Manoel) -> (?João NS#sogro ?Manoel)]</pre>	<pre>sparql1: SELECT ?VAR_PES WHERE { ?VAR_PES :filho ?Maria . ?Maria :filha ?Manoel . } sparql2: SELECT ?VAR_PES WHERE { ?VAR_PES :filha ?Manoel . ?Manoel :genro ?João . }</pre>
---	---

Figura 13 - Exemplo de regras de nível de ontologia (esquerda) e regras de nível de instância

A primeira regra de nível de ontologia (*i.e.*, rule1) cria um novo relacionamento, chamado “*avô*”, que conecta o “Manoel” diretamente ao “José” - isso é possível porque o “Manoel” é o pai da “Maria” que é mãe de “José”. A segunda regra de nível de ontologia (*i.e.*, rule2) cria outro relacionamento, chamado “*sogro*”, que vincula o “João” ao “Manoel” – esse relacionamento é viável porque o “João” é casado com a “Maria” que é filha do “Manoel”. Nos exemplos de regras de nível de instância, a primeira (*i.e.*, sparql1) extrai os filhos da “Maria” que tem como pai o “Manoel” – nesse caso a resposta é apenas “José”. Já a segunda

regra de instância (*i.e.*, sparql2) utiliza uma das regras de nível de ontologia para extrair as filhas do “Manoel” que tem como genro o “João” – nesse caso a resposta é “Maria”.

4.3 Organização da arquitetura

A estrutura dos módulos e componentes que compõem a arquitetura está descrita de modo geral na Figura 14. Os módulos de indexação (*i.e.*, *IndexModel*), recuperação (*i.e.*, *RetrieveModel*) e inferências (*i.e.*, *InferenceModel*) estão explicitamente definidos por suas respectivas classes, enquanto o módulo de apresentação está representado pelos componentes interativos de navegação da ontologia (*i.e.*, *OntologyNavigator*), instâncias recuperadas (*i.e.*, *RetrievedInstances*) e informação adicional (*i.e.*, *AdditionalInformation*). As classes *Controller* e *ReadOntology* auxiliam na organização e fluxo de comunicação da arquitetura. No lado esquerdo da Figura 14 encontra-se a ontologia e a base de conhecimento (*i.e.*, KB), fornecidas pelo especialista do domínio, e os índices semânticos.

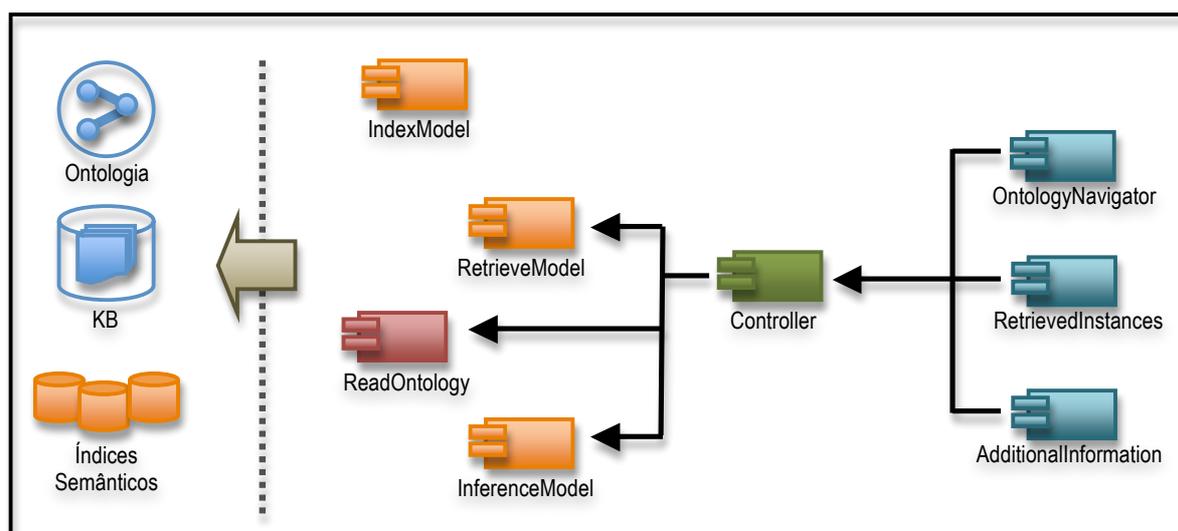


Figura 14 – Organização da arquitetura de módulos e componentes do framework

A linguagem de programação utilizada para implementação das classes *IndexModel*, *RetrieveModel*, *ReadOntology* e *InferenceModel* é Java². Para manipulação da ontologia e base de conhecimento a tecnologia utilizada foi o *Jena* (Mcbride, 2002; Carroll, Dickinson *et al.*, 2004; Cardoso, 2007). Já os componentes visuais foram implementados com tecnologia que lhes permitisse serem utilizados a partir de navegadores Web – o *OntologyNavigator* foi implementado também em Java, especificamente com a tecnologia *Applets*³, e os componen-

² <http://www.java.com>

³ <http://java.sun.com/applets/>

tes *RetrievedInstances* e *AdditionalInformation* foram implementados em HTML e *JavaScript*. A classe *Controller*, implementada com a tecnologia *Servlet*⁴, é responsável por interpretar as requisições dos componentes interativos e requisitar informações das classes responsáveis pela recuperação de instâncias (*i.e.*, *RetrieveModel*), leitura da ontologia (*i.e.*, *ReadOntology*) e inferência (*i.e.*, *InferenceModel*). O classe que representa o módulo de indexação (*i.e.*, *IndexModel*) não está conectada as demais classes por ter um comportamento autônomo e não ser utilizada diretamente por requisições efetuadas pelos componentes interativos. A classe que faz a leitura de ontologia (*i.e.*, *ReadOntology*) atende especificamente a requisição do componente de navegação de ontologia (*i.e.*, *OntologyNavigator*), que necessita conhecer toda a estrutura da ontologia.

4.4 Configuração do framework Hermeneus

O objetivo do framework é possibilitar a sua utilização em qualquer domínio do conhecimento. Para tal, conforme descrito em sua arquitetura, o Hermeneus foi concebido seguindo princípios definidos pelo modelo, para ser modular e configurável. A definição da ontologia e a construção da base de conhecimento são tarefas que estão fora do escopo do Hermeneus, as quais devem ser executadas por especialistas de domínio seguindo critérios amplamente discutidos pela literatura referente à engenharia de ontologia (Uschold e Gruninger, 1996; Noy e McGuinness, 2001; Smith e Welty, 2001; Corcho, Fernández-López *et al.*, 2003; Pinto e Martins, 2004). A seguir, são apresentadas as configurações necessárias para cada módulo da arquitetura do Hermeneus e, na Tabela 4, há um resumo das tarefas requeridas pelo framework.

- ✚ O módulo de indexação (*i.e.*, classe *IndexModel* da Figura 14) requer acesso à ontologia e à base de conhecimento. Uma vez configurados esses parâmetros, o módulo é concebido para gerar automaticamente os índices semânticos com as instâncias armazenadas na base de conhecimento. A definição de quais índices semânticos devem ser criados é extraída, automaticamente, da ontologia;
- ✚ O módulo de recuperação (*i.e.*, classe *RetrieveModel* da Figura 14) necessita de acesso à ontologia, à base de conhecimento e aos índices semânticos. O acesso à ontologia permite que esse módulo identifique os conceitos de modo a

⁴ <http://java.sun.com/products/servlet/>

interpretar consultas semânticas. A base de conhecimento é requerida para a apresentação das informações das instâncias recuperadas a partir do índice, o qual é usado para recuperar instâncias a partir de uma consulta;

- ✚ O módulo de inferência (*i.e.*, classe *InferenceModel* da Figura 14) necessita ter acesso à ontologia e à base de conhecimento. Esse módulo também requer a configuração das regras de nível de ontologia e de nível de instância. A inferência pode ocorrer em ambos – ontologia e base de conhecimento (*i.e.*, instâncias), estando vinculada aos tipos de regras utilizadas;
- ✚ O módulo de representação, que contém os componentes interativos (*i.e.*, componentes *OntologyNavigator*, *RetrievedInstances* e *AdditionalInformation* da Figura 14), precisa ter acesso à ontologia, no caso do componente de navegação de ontologia, acesso ao módulo de recuperação, no caso do componente de instâncias recuperadas, e acesso à base de conhecimento. O componente de informação adicional precisa ter acesso ao módulo de inferência e requer que as informações inferidas sejam configuradas para apresentação.

Tabela 4 - Configurações requeridas pelo framework Hermeneus

	Módulo de indexação	Módulo de recuperação	Módulo de inferência	Módulo de apresentação
Acesso à ontologia	x	x	x	x
Acesso à base de conhecimento	x	x	x	x
Acesso aos índices semânticos	x	x		
Acesso ao mecanismo de recuperação				x
Regras de nível de ontologia			x	
Regras de nível de instância			x	
Acesso ao mecanismo de inferência				x
Design da informação inferida			x	x

4.5 Considerações finais

O objetivo em definir um novo modelo de RBI foi conceber algo mais pragmático e que contivesse indicações que direcionassem a construção de um ambiente de busca interativo. O framework Hermeneus mostra a viabilidade de utilização do modelo na prática. O artefato gerado pelo Hermeneus é um sistema de busca interativo. Segundo Toms (2002), a forma

com que os usuários interagem com a informação em ambientes digitais é diretamente influenciada pela arquitetura oferecida pelo sistema de informação.

Este capítulo apresenta a arquitetura do framework e os itens pelos quais cada um de seus módulos e componentes são responsáveis. O módulo de indexação tem por função criar índices semânticos baseados na ontologia e em sua base de conhecimento. A recuperação e a ordenação das instâncias recuperadas são executadas pelo módulo de recuperação. Informações adicionais geradas a partir de técnicas de inferência e por meio da utilização de regras, tanto no nível de ontologia quanto no nível de instâncias, são descritas pelo módulo de inferência. No módulo de apresentação foi descrito os três componentes interativos que compõem a interface. O primeiro é o componente de navegação de ontologia, que permite que usuários interajam diretamente com o conhecimento do domínio e consigam gerar ações que afetam desde a composição de consultas semânticas até a visualização das instâncias recuperadas em diferentes perspectivas. O segundo é o componente de instâncias recuperadas, que é responsável por organizar as instâncias com vistas a composição de novas consultas dinamicamente. O terceiro componente está diretamente vinculado ao módulo de inferência, pois é o responsável por apresentar e permitir que informações adicionais sejam utilizadas para composição dinâmica de novas consultas.

No próximo capítulo, descreve-se mais detalhadamente de que forma o Hermeneus é usado para a construção de sistemas de busca interativos por meio da apresentação de um protótipo e conseqüente descrição de um exemplo. Mostra-se, também, a importância da ontologia no âmbito da engenharia de sistemas e de que forma o Hermeneus pode atuar como um tradutor da informação necessitada pelos usuários. Além disso, há uma comparação entre os modelos descritos na revisão da literatura com relação ao modelo proposto de RBI.

5 DEMONSTRAÇÃO DE VIABILIDADE E ANÁLISE COMPARATIVA

5.1 Apresentação do protótipo

Visando-se a validação do framework Hermeneus e conseqüentemente do modelo de RBI construiu-se um protótipo. Esse protótipo usa uma base com 622 citações de artigos do *Journal of the American Society for Information Science and Technology* (JASIST). Cada citação contém informações sobre um *paper*, incluindo título, autores, instituições, edição e local de publicação. A Figura 15 apresenta um exemplo de citação extraído do site do JASIST.

<p>AU: Frederico T. Fonseca, James E. Martin TI: Toward an alternative notion of information systems ontologies: Information engineering as a hermeneutic enterprise SO: Journal of the American Society for Information Science and Technology VL: 56 NO: 1 PG: 46-57 YR: 2005 CP: Copyright © 2004 Wiley Periodicals, Inc. ON: 1532-2890 PN: 1532-2882 AD: School of Information Sciences and Technology, Pennsylvania State University, University Park, PA 16802; Psychology Department, Pennsylvania State University, University Park, PA 16802 DOI: 10.1002/asi.20099 US: http://dx.doi.org/10.1002/asi.20099 AB: In this paper we discuss the construction of information systems ontologies. We summarize and discuss Barry Smith's review (2003a) of the field in the paper...</p>
--

Figura 15 - Exemplo de citação disponível no site do JASIST

Uma ontologia baseada nas informações disponíveis nas citações foi concebida contendo cinco conceitos: *paper*, *author*, *institution*, *journal* e *keyword*. Cabe ressaltar que cada conceito é representado por uma classe e suas respectivas propriedades – também conhecidas como *slots*. As propriedades de cada conceito descrevem as suas características e seus atributos (e.g., o conceito *autor* é definido pela classe *autor* e pela propriedade *nome*) (Noy e McGuinness, 2001). A Tabela 5 apresenta as classes da ontologia e as suas respectivas propriedades. Na Figura 16 é possível visualizar os conceitos da ontologia e o modo como estão relacionados uns aos outros e, na Figura 17, é apresentado o código fonte da ontologia em OWL. Definiu-se essa ontologia apenas com o propósito de mostrar a viabilidade da proposta. Qualquer discussão sobre a ontologia está além do escopo deste trabalho. Entretanto, cabe ressaltar que a ontologia foi concebida de acordo com princípios amplamente discutidos pela

literatura (López, Asunción *et al.*, 1999; Mcguinness, 1999; Noy e Mcguinness, 2001; Sure e Studer, 2002; Corcho, Fernández-López *et al.*, 2003; Pinto e Martins, 2004).

Tabela 5 - Classes da ontologia e suas respectivas propriedades

Classe	Propriedades
Paper	Identifier, Title e Abstract
Author	Name
Institution	Description
Journal	Volume, IssueNumber e Year
Keyword	KeywordName

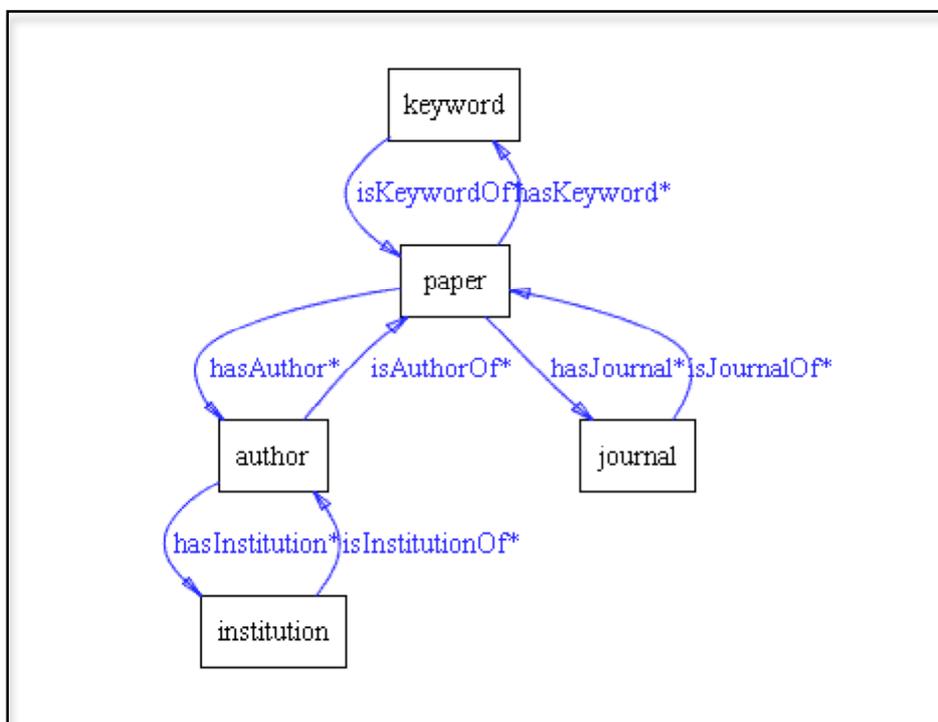


Figura 16. Ontologia: classes e suas relações

<pre> <?xml version="1.0"?> <rdf:RDF xmlns="http://www.owl-ontologies.com/interscience.owl#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base="http://www.owl-ontologies.com/interscience.owl"> <owl:Ontology rdf:about=""/> <owl:Class rdf:ID="keyword"/> <owl:Class rdf:ID="author"/> </pre>	<pre> <owl:ObjectProperty rdf:about="#hasJournal"> <owl:inverseOf rdf:resource="#isJournalOf"/> <rdfs:domain rdf:resource="#paper"/> <rdfs:range rdf:resource="#journal"/> </owl:ObjectProperty> <owl:DatatypeProperty rdf:ID="abstract"> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> <rdfs:domain rdf:resource="#paper"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="issueNumber"> <rdfs:range </pre>
--	---

<pre> <owl:Class rdf:ID="institution"/> <owl:Class rdf:ID="paper"/> <owl:Class rdf:ID="journal"/> <owl:ObjectProperty rdf:ID="isJournalOf"> <owl:inverseOf> <owl:ObjectProperty rdf:ID="hasJournal"/> </owl:inverseOf> <rdfs:domain rdf:resource="#journal"/> <rdfs:range rdf:resource="#paper"/> </owl:ObjectProperty> <owl:ObjectProperty rdf:ID="isInstitutionOf"> <owl:inverseOf> <owl:ObjectProperty rdf:ID="hasInstitution"/> </owl:inverseOf> <rdfs:domain rdf:resource="#institution"/> <rdfs:range rdf:resource="#author"/> </owl:ObjectProperty> <owl:ObjectProperty rdf:ID="hasKeyword"> <rdfs:domain rdf:resource="#paper"/> <rdfs:range rdf:resource="#keyword"/> <owl:inverseOf> <owl:ObjectProperty rdf:ID="isKeywordOf"/> </owl:inverseOf> </owl:ObjectProperty> <owl:ObjectProperty rdf:ID="hasAuthor"> <rdfs:range rdf:resource="#author"/> <rdfs:domain rdf:resource="#paper"/> <owl:inverseOf> <owl:ObjectProperty rdf:ID="isAuthorOf"/> </owl:inverseOf> </owl:ObjectProperty> <owl:ObjectProperty rdf:about="#isAuthorOf"> <rdfs:range rdf:resource="#paper"/> <owl:inverseOf rdf:resource="#hasAuthor"/> <rdfs:domain rdf:resource="#author"/> </owl:ObjectProperty> <owl:ObjectProperty rdf:about="#isKeywordOf"> <owl:inverseOf rdf:resource="#hasKeyword"/> <rdfs:range rdf:resource="#paper"/> <rdfs:domain rdf:resource="#keyword"/> </owl:ObjectProperty> <owl:ObjectProperty rdf:about="#hasInstitution"> <rdfs:range rdf:resource="#institution"/> <owl:inverseOf rdf:resource="#isInstitutionOf"/> <rdfs:domain rdf:resource="#author"/> </owl:ObjectProperty> </pre>	<pre> rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> <rdfs:domain rdf:resource="#journal"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="title"> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> <rdfs:domain rdf:resource="#paper"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="identifier"> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> <rdfs:domain rdf:resource="#paper"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="volume"> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> <rdfs:domain rdf:resource="#journal"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="year"> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> <rdfs:domain rdf:resource="#journal"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="keywordname"> <rdfs:domain rdf:resource="#keyword"/> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="description"> <rdfs:domain rdf:resource="#institution"/> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="journalname"> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> <rdfs:domain rdf:resource="#journal"/> </owl:DatatypeProperty> <owl:DatatypeProperty rdf:ID="name"> <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/> <rdfs:domain rdf:resource="#author"/> </owl:DatatypeProperty> </rdf:RDF> </pre>
---	---

Figura 17 - Listagem contendo a ontologia em OWL

A base de conhecimento foi construída com instâncias da ontologia. As propriedades das classes *paper*, *author*, *institution* e *journal* foram preenchidas com informações diretamente disponíveis nas citações. A propriedade da classe *keyword* foi preenchida com infor-

mações extraídas dos campos *title* e *abstract* por meio da técnica de extração de informação chamada *Named Entity Recognition*⁵ (NER) (Mikheev, Moens *et al.*, 1999).

O módulo de inferência foi configurado com regras tanto de nível de ontologia quanto de nível de instância e utiliza a máquina de inferência disponível no Jena (Mcbride, 2002; Carroll, Dickinson *et al.*, 2004; Cardoso, 2007). O protótipo foi configurado com uma regra de nível de ontologia que é responsável por criar um novo relacionamento entre os conceitos *author* e *keyword* – dessa forma pode-se facilmente recuperar as palavras-chave de um autor específico. A Figura 18 apresenta, na parte superior, uma regra de nível de ontologia expressa em Jena Rule Language e uma regra de nível de instância expressa em SPARQL, que extrai informações de instâncias utilizando a nova relação criada pela regra de nível de ontologia (*i.e.*, as *keywords* de um *author* específico).

<pre>[rule1: (?paper NS#hasAuthor ?author) (?paper NS#hasKeyword ?keyword) -> (?author NS#authorsKeyword ?keyword)]</pre>
<pre>SELECT ?keywordname WHERE ?keyword :keywordname ?keywordname . ?author :authorsKeyword ?keyword .?author :name 'Frederico Fonseca' .</pre>

Figura 18. Exemplo de regra de nível de ontologia (parte superior) e de instâncias (parte inferior)

O sistema de busca possui duas telas. A primeira tela, disponível na Figura 19, apresenta um campo-texto e o componente de navegação da ontologia. Nesse componente é possível definir o *conceito central*, que indica o tipo de informação a ser recuperada. Quando, por exemplo, o usuário seleciona o conceito *author* como *conceito central*, as instâncias sobre autores e seus respectivos *papers* e instituições (*i.e.*, *institution*) são apresentadas. A estrutura da ontologia indica quais tipos de instâncias devem ser apresentadas (na ontologia, Figura 16, há uma relação direta entre os conceitos *author*, *paper* e *institution*). Portanto, o *conceito central* indica o tipo principal de instância a ser recuperada. Além disso, as instâncias referentes a conceitos vinculados ao *conceito central* são também apresentadas.

Quando um usuário clica em um conceito no componente de navegação de ontologia a palavra que representa o conceito é automaticamente inserida no campo-texto seguida do símbolo de dois-pontos (*e.g.*, clicar no conceito *author*, a palavra “*author:*” é inserida no campo-texto). Dessa forma, os usuários podem digitar os termos a serem buscados especifi-

⁵ *Named Entity Recognition* é parte da área de extração de informação que visa a localização e classificação de elementos textuais em categorias predefinidas (*e.g.*, pessoas, locais, etc.).

camente para um conceito (*i.e.*, composição de consultas semânticas). Além disso, é possível combinar consultas simples e semânticas para se efetuar uma única busca. Em uma consulta simples, em que há termos não contextualizados, todas as instâncias do *conceito central* que contêm os termos pertencentes à consulta são recuperadas independentemente do contexto (*e.g.*, a busca por “*inteligência artificial*” recupera todas as instâncias que possuem esses dois termos em qualquer contexto). Já em uma consulta semântica, somente as instâncias do *conceito central* que possuem os termos contextualizados de acordo com a consulta são recuperadas (*e.g.*, a consulta semântica “*author:fonseca*” recupera apenas as instâncias que contêm o termo “*fonseca*” vinculado ao conceito *author*). É importante ressaltar que a busca é aplicada sobre todo o conjunto de instâncias armazenadas nos índices semânticos. O índice semântico a ser usado para uma busca é aquele vinculado ao *conceito central*.

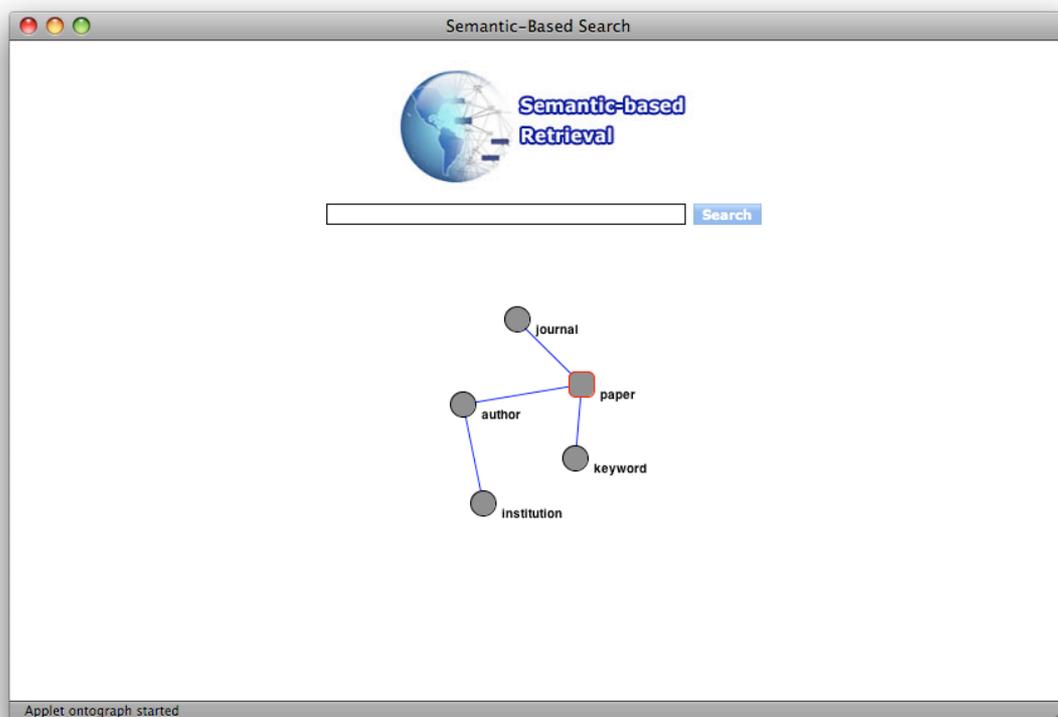


Figura 19 - Tela inicial do protótipo

Após a primeira busca, é apresentada a tela que mostra os resultados recuperados (Figura 20), a qual é basicamente dividida em quatro seções: (1) contém o campo-texto para composição de consultas; (2) apresenta o componente de navegação da ontologia; (3) mostra o componente de instâncias recuperadas; e (4) apresenta o componente de informação adicional. Na Figura 20 é apresentado o resultado de uma busca cuja consulta é “*ontology au-*

thor:fonseca” em que duas instâncias de *paper* são recuperadas. A consulta utilizada é composta de termos simples (*i.e.*, *ontology*) e termos semânticos (*i.e.*, *author:fonseca*). Também são apresentadas as instâncias dos conceitos (*i.e.*, *author*, *journal* e *keyword*) relacionados ao conceito *paper*. No componente de navegação da ontologia o *conceito central* é desenhado em um formato diferente (*i.e.*, quadrado) e o conceito que está sendo usado na consulta (*i.e.*, *author*) é pintado de amarelo. Ao se clicar na primeira instância recuperada, é apresentada informação adicional sobre essa instância na quarta seção. Essa informação mostra que o autor “*Frederico Fonseca*” tem como principais palavras-chave (*i.e.*, *keywords*) “*ontologies*”, “*information science*” e “*engineering*”; também é possível verificar que o autor coopera com “*Gilberto Câmara*” e “*James E. Martin*”. Além disso, são apresentadas informações sobre a instituição do autor, chamada “*The Pennsylvania State University*”, que possui 18 *papers* publicados por 36 autores e que possui como principais palavras-chave (*i.e.*, *keywords*): “*web search*”, “*web search engine*”, “*information retrieval*” e “*ontologies*”.

The screenshot shows a web application titled "Semantic-Based Search" with a search bar containing "ontology author:fonseca" and a "Search" button. The interface is divided into three main sections:

- Ontology Navigator:** A graph showing relationships between concepts: keyword, paper, author, journal, and institution. The 'paper' node is highlighted in red, and the 'author' node is highlighted in yellow.
- Result (paper):** Displays two search results. The first result is selected and shows:
 - 1. Paper:** (title) **The double role of ontologies in information science research**
 - (abstract)** In philosophy, Ontology is the basic description of things in the world. In information science, an ontology refers to an engineering artifact, constituted by a specific vocabulary used to describe a certain reality. Ontologies...
 - (identifier)** <http://doi.wiley.com/10.1002/asi.20565>
 - Keywords:** (keywordname) information systems; engineering; ontologies; science; ontologies of information; information science; ontology; ontology-driven information systems;
 - Authors:** (name) Frederico Fonseca;
 - Journal:** (journalname) Journal of the American Society for Information Science and Technology; (volume)9999; (issuenum)9999; (year)2007;
- Additional Information:**
 - Frederico Fonseca's most frequent keywords are: science, engineering, ontology, ontologies, information science; and s/he cooperates with:
 - Gilberto Câmara;
 - James E. Martin;
 - See Frederico Fonseca in Google Scholar.
 - Pennsylvania State University, University Park, PA 16802 has published 18 paper(s), written by 36 author(s), and its most frequent keywords are: web, web search engine, web search, information retrieval, engineering.

Figura 20 - Tela do protótipo que apresenta os resultados de uma busca.

Os usuários podem interagir com o conteúdo recuperado disponível no componente de instâncias recuperadas e, também, com as informações adicionais, disponível na quarta sessão. Na lista das instâncias recuperadas, quando um usuário seleciona uma propriedade de

uma instância uma nova busca é executada levando em consideração o conteúdo clicado – é o tipo de interação chamada de refinamento. No componente de informação adicional, essa funcionalidade pode se manifestar de maneira diferente, porque nesse componente o especialista do domínio pode definir comportamentos específicos para cada informação extraída. Por exemplo, no protótipo, ao se clicar no número que representa a quantidade de *papers* (*i.e.*, 18) é possível visualizar os dezoito *papers* no componente de instâncias recuperadas, publicados por autores vinculados à instituição “*The Pennsylvania State University*” (*i.e.*, a instituição que o autor “*Frederico Fonseca*” está vinculado).

5.1.1 Exemplo de utilização do protótipo

Neste tópico é apresentado um exemplo que visa demonstrar algumas características disponibilizadas pelo sistema de busca interativo resultante da aplicação do framework Hermeneus e, conseqüentemente, do modelo. Discute-se, também, algumas características à medida que o exemplo é apresentado. A Figura 21 mostra a tela inicial do sistema de busca interativo contendo apenas o termo “*seeking*” no campo-texto disponível para inserção da consulta.

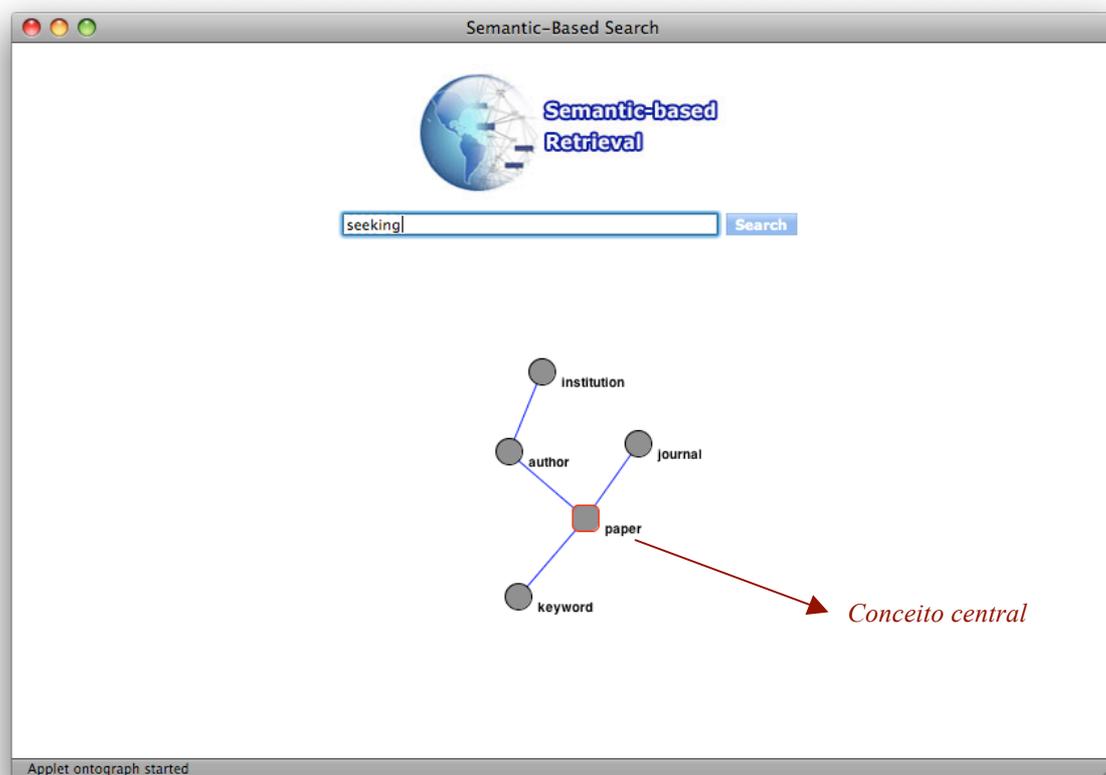


Figura 21 - Exemplo de busca pelo termo "seeking"

Optou-se apenas por definir uma consulta simples, mas nessa tela inicial também é possível alterar o *conceito central*, interagindo diretamente com a ontologia (o *conceito central* escolhido é *paper*, desenhado em forma de quadrado), e por compor consultas semânticas, utilizando os conceitos definidos na ontologia. O resultado da busca, como demonstra a Figura 22, apresenta as instâncias recuperadas juntamente com os demais componentes interativos. O componente de navegação da ontologia destaca apenas o *conceito central* com um desenho diferente. Essa característica permite facilitar, por exemplo, o escaneamento, pois visualmente o *conceito central* chama a atenção por ser representado em outro formato. Já no componente de instâncias recuperadas, as instâncias de *papers* são apresentadas e disponibilizadas para utilização em consultas a partir do seu conteúdo – tudo o que está em azul pode ser utilizado para efetuar refinamento da busca. É possível, também, identificar visualmente as instâncias de outros conceitos (*keywords*, *authors*, e *journal*) que estão diretamente vinculados com ao conceito *paper*. O componente de informação adicional não apresenta nenhuma informação por não haver uma instância selecionada.

The screenshot displays the 'Semantic-based Search' application window. The search term 'seeking' is entered in the search bar, and the results are displayed in a table-like format. The interface is divided into three main sections: 'Ontology Navigator', 'Result (paper)', and 'Additional Information'.

Ontology Navigator: A graph showing relationships between concepts: 'institution' is connected to 'author' and 'journal'; 'author' is connected to 'paper'; 'journal' is connected to 'paper'; and 'keyword' is connected to 'paper'. The 'paper' node is highlighted with a red square.

Result (paper): Shows 71 results. The first result is:

- Paper:** (title) **User information-seeking behavior in a medical Web portal environment: A preliminary study**
- (abstract) The emergence of information portal systems in the past few years has led to a greatly enhanced Web-based environment for users seeking information online. While considerable research has been conducted on user information-seeking...
- (identifier) <http://doi.wiley.com/10.1002/asi.20001>
- Keywords:** (keywordname) web-based environment; medical science; medical; medical web; web; science;
- Authors:** (name) Caroline Zambrowicz; Dongming Zhang; Nancy K. Roderer; Hong Zhou;
- Journal:** (journalname) Journal of the American Society for Information Science and Technology; (volume)55; (issuenumber)8; (year)2004;

The second result is:

- Paper:** (title) **Multitasking information seeking and searching processes**
- (abstract) Recent studies show that humans engage in multitasking behaviors as they seek and search information retrieval (IR) systems for information on more than one topic at the same time. For example, a Web search session by a single...
- (identifier) <http://doi.wiley.com/10.1002/asi.10124>
- Keywords:** (keywordname) web search engine; web systems; search information retrieval; web; web search;

The 'web search;' and 'web search;' keywords in the second result are highlighted with a red box. The 'Additional Information' section on the right contains a note: 'Click in one check box, available in the result set list, in order to retrieve additional information...'. At the bottom left, a status bar reads 'Applet ontograph started'.

Figura 22 - Tela resultante da busca pelo termo "seeking"

Dando continuidade ao exemplo, decidiu-se clicar em uma das instâncias de *keyword* relacionada com a segunda instância de *paper* recuperada, conforme mostra a Figura 22, com um destaque em vermelho na parte inferior da tela. Quando se clica no conteúdo de uma instância, esse conteúdo, que é contextualizado semanticamente, é automaticamente utilizado para efetuar um refinamento da busca. No exemplo, a consulta agora é formada pelos termos “*keyword: web search; seeking*”, e como resultado há apenas 11 instâncias de *papers*, enquanto havia 71 antes do refinamento. Visualmente, conforme mostra a Figura 23, além das instâncias recuperadas que mudaram, é possível também identificar uma alteração no componente de navegação da ontologia. O conceito usado na busca (*i.e.*, *keyword*) é pintado com outra cor (*i.e.*, amarela). Novamente, esse destaque permite que um usuário identifique facilmente quais conceitos estão sendo usados na busca e, ao mesmo tempo, também auxilie no processo de aprendizagem, uma vez que é possível relacionar o conceito usado na busca, que foi utilizado a partir de uma instância, com o conceito definido pela ontologia.

Essa funcionalidade que permite que um usuário utilize dinamicamente o conteúdo de uma instância recuperada para formular novas consultas é citada por Bates (1989) como uma necessidade em razão da mudança constante por causa da interação com a informação recuperada. No modelo clássico de RI, o usuário tem que ser direto e definir uma consulta que expressa a sua necessidade. Contudo, conforme afirma Bates (1989), um usuário, quando busca informação, tem um comportamento que se difere desse modelo tradicional de recuperação de informação em dois sentidos: (1) uma consulta está sempre sendo alterada, ela não é estática, ou seja, o usuário desenvolve a sua idéia à medida que interage com as informações recuperadas; e (2) não há um comportamento predefinido num processo de recuperação de informação, *i.e.*, não há um padrão de comportamento que possa condicionar um usuário – ele muda de direção constantemente. Cabe ressaltar que a primeira proposição citada por Bates (1989) tem uma relação direta com o círculo hermenêutico – discute-se essa questão mais à frente.

The screenshot displays a web application window titled "Semantic-Based Search". The interface is divided into three main sections: "Ontology Navigator", "Result (paper)", and "Additional Information".

- Ontology Navigator:** Shows a graph with nodes for "Institution", "author", "journal", "paper", and "keyword". "Institution", "author", and "journal" are grey circles, "paper" is a red square, and "keyword" is a yellow circle. Lines connect "Institution" to "author", "author" to "paper", "journal" to "paper", and "paper" to "keyword".
- Result (paper):** Displays search results for the query "keyword: web search; seeking". It shows 11 results. The first two are:
 - Paper:** (title) **Multitasking information seeking and searching processes**. (abstract) Recent studies show that humans engage in multitasking behaviors as they seek and search information retrieval (IR) systems for information on more than one topic at the same time. For example, a Web search session by a single... (identifier) <http://doi.wiley.com/10.1002/asi.10124>
Keywords: (keywordname) web search engine; web systems; search information retrieval; web; web search;
Authors: (name) Amanda Spink; Seda Ozmutlu; H. Cenk Ozmutlu;
Journal: (journalname) Journal of the American Society for Information Science and Technology; (volume)53; (issuenum)8; (year)2002;
 - Paper:** (title) **Web searching in Chinese: A study of a search engine in Hong Kong**. (abstract) The number of non-English resources has been increasing rapidly on the Web. Although many studies have been conducted on the query logs in search engines that are primarily English-based (e.g., Excite and AltaVista), only... (identifier) <http://doi.wiley.com/10.1002/asi.20592>
Keywords: (keywordname) web; search-query; search engine;
Authors: (name) Xiao Fang; Michael Chau; Christopher C. Yang;
- Additional Information:** Contains a message: "Click in one check box, available in the result set list, in order to retrieve additional information..."

At the bottom left, a status bar reads "Applet ontograph started".

Figura 23 - Tela resultante da busca por "keyword: web search; seeking"

Para utilizar o componente de informação adicional, basta selecionar uma instância recuperada. Escolheu-se a primeira instância visando verificar o tipo de informação adicional disponível para essa instância, como apresenta a Figura 24. Entre as informações recuperadas, nota-se que o primeiro autor do *paper*, *Amanda Spink*, tem as seguintes palavras-chave: *information retrieval*, *search engine*, *web search engine*. Como essa autora parece estar de acordo com o propósito de busca desse exemplo, decidiu-se clicar sobre o nome dela.

É importante atentar para o fato de que se começou a busca apenas digitando um termo sem contexto e uma idéia inicial sobre recuperar, encontrar, buscar (o termo usado foi “*seeking*”). Com as instâncias recuperadas, resultantes da primeira busca, uma das palavras-chave, pertencente à segunda instância recuperada, foi utilizada facilmente (apenas com o clique do mouse) para refinar a busca. O refinamento, nesse caso, foi contextualizado e já permitiu direcionar a busca, pois escolheu-se os termos “*web search*” vinculados ao conceito *keyword* – de um termo genérico (*i.e.*, *seeking*) para uma área mais específica, que é busca na Internet. Além disso, ao se escolher a primeira instância recuperada após o refinamento, foi identificado um autor que pode estar atuando na área em que se está interessado. Essa informação foi identificada a partir das informações adicionais extraídas pelo módulo de inferência. Pôde-se identificar a partir das informações adicionais outros termos que podem estar relacionados com a busca (*e.g.*, *information retrieval*, *search engine* e *web search engine*). Como informação adicional, tornou-se possível o acesso a informações referentes a instituições e a autores que cooperam diretamente com os autores relacionados à instância recuperada e selecionada para inferência.

As informações apresentadas para cada instância de *paper*, disponíveis no componente de instâncias recuperadas, não fornecem o conjunto de informações apresentadas no componente de informação adicional. Não é objetivo aqui discutir questões mais detalhadas sobre um processo de aprendizagem, mas pode-se citar que quando uma pessoa tem acesso a informações com maior valor agregado, essa pessoa possui mais condições de desenvolver o seu aprendizado (Doyle, 1984). Como consequência, ela pode influenciar as suas ações futuras, como, por exemplo, a identificação e a verificação de informações relevantes necessárias em um processo de busca ou em um comportamento de *information seeking*. Deve-se considerar, também, que essas informações adicionais possibilitam a sua utilização interativamente.

The screenshot shows a web application titled "Semantic-based Retrieval". At the top, there is a search bar containing the text "keyword:web search; seeking" and a "Search" button. Below the search bar, the interface is divided into three main sections:

- Ontology Navigator:** A network diagram showing relationships between concepts. The nodes are "institution", "author", "journal", "paper", and "keyword". "institution" is connected to "author" and "journal". "author" is connected to "journal" and "paper". "journal" is connected to "paper". "keyword" is connected to "paper". The "paper" node is highlighted in red.
- Result (paper):** A list of search results. The first result is "1. Paper: (title) **Multitasking information seeking and searching processes** (abstract) Recent studies show that humans engage in multitasking behaviors as they seek and search information retrieval (IR) systems for information on more than one topic at the same time. For example, a Web search session by a single... (identifier) http://doi.wiley.com/10.1002/asi.10124". Below the title, it lists "Keywords: (keywordname) web search engine; web systems; search information retrieval; web; web search;" and "Authors: (name) Amanda Spink; Seda Ozmutlu; H. Cenk Ozmutlu;". The second result is "2. Paper: (title) **Web searching in Chinese: A study of a search engine in Hong Kong** (abstract) The number of non-English resources has been increasing rapidly on the Web. Although many studies have been conducted on the query logs in search engines that are primarily English-based (e.g., Excite and AltaVista), only... (identifier) http://doi.wiley.com/10.1002/asi.20592". Below the title, it lists "Keywords: (keywordname) web; search-query; search engine;" and "Authors: (name) Xiao Fang; Michael Chau; Christopher C. Yang;".
- Additional Information:** A section providing more details about the authors. It includes "Amanda Spink's most frequent keywords are: information retrieval, web, search engine, evolutionary; and s/he cooperates with 20 distinct authors. See Amanda Spink in Google Scholar." and "H. Cenk Ozmutlu cooperates with: - Amanda Spink; - Seda Ozmutlu;". It also mentions "Seda Ozmutlu cooperates with: - H. Cenk Ozmutlu; - Amanda Spink;" and "Pennsylvania State University, University Park, PA 16802 has published 18 paper(s), written by 36 author(s), and its most frequent keywords are:".

Figura 24 - Tela que mostra informações adicionais sobre uma instância

Ao ser identificado que a autora *Amanda Spink* pode ter relação com o assunto em que se está interessado (*i.e.*, “*web search*”), pode-se apenas clicar sobre o nome da autora, disponível no componente de informações adicionais. A Figura 25 apresenta a tela resultante dessa ação, e nela é possível identificar que o componente de instâncias recuperadas mostra a instância da referida autora juntamente com os seus respectivos *papers*. O componente de navegação de ontologia atualiza visualmente as suas características para refletir as informações usadas na nova busca. Nesse caso, o *conceito central* foi automaticamente alterado para *author*, e o próprio conceito também foi marcado como sendo usado na busca (o conceito *author* está pintado em amarelo e desenhado em forma de quadrado – *conceito central*).

Durante a interação em um sistema de busca, como o diálogo avança por meio de episódios, aspectos cognitivos, emocionais e situacionais na interação tendem a mudar – um problema ou questão é redefinido, novos termos de busca são selecionados enquanto outros são abandonados, táticas de busca são adaptadas ou alteradas e assim por diante (Saracevic, 1997). O modelo proposto refletido no framework Hermeneus permite que os usuários trabalhem as suas questões e exerçam a capacidade de mudança de modo mais dinâmico ao passo

que desenvolvem um processo de busca. No exemplo aqui discutido, começou-se buscando um termo (*i.e.*, “*seeking*”) sem contexto e, por meio do sistema, encontrou-se um assunto (*i.e.*, “*web search*”), contextualizado por um conceito definido pela ontologia, em que se está interessado. Também foi encontrado um autor, contextualizado por um conceito definido pela ontologia, que pode estar diretamente relacionado com a necessidade. Essa é uma característica que está vinculada ao processo de entendimento ou interpretação, como discutido pelo círculo hermenêutico. Um usuário pode mudar constantemente as suas necessidades à medida que recebe e interpreta a nova informação. De um termo sem contexto, mudou-se para um assunto, descobriu-se um autor, sendo tudo isso executado de forma dinâmica. O sistema permite que essa mudança, que ocorre naturalmente em seres humanos, possa ser traduzida e facilitada por meio de componentes interativos que permitam aos usuários desenvolver as suas dúvidas ou questões enquanto experienciarem um processo de busca.

The screenshot displays the 'Semantic-based Retrieval' application window. At the top, there is a search bar containing 'author:Amanda Spink;' and a 'Search' button. Below the search bar, the interface is divided into three main sections:

- Ontology Navigator:** A graph showing relationships between concepts: 'author' (highlighted in yellow), 'institution', 'journal', and 'paper'. 'author' is connected to 'institution' and 'journal'. 'journal' is connected to 'paper'. 'paper' is connected to 'keyword'.
- Result (author):** Shows 1 result for the author 'Amanda Spink'. It lists three papers:
 - 1.1 Paper:** (title) **Defining a session on Web search engines**. Abstract: Detecting query reformulations within a session by a Web searcher is an important area of research for designing more helpful searching systems and targeting content to particular users. Methods explored by other researchers... (identifier) <http://doi.wiley.com/10.1002/asi.20564>. Institution: (description) Faculty of Information Technology, Queensland University of Technology, Gardens Point Campus, 2 George Street, GPO Box 2434, Brisbane QLD 4001, Australia.
 - 1.2 Paper:** (title) **A temporal comparison of AltaVista Web searching**. Abstract: Major Web search engines, such as AltaVista, are essential tools in the quest to locate online information. This article reports research that used transaction log analysis to examine the characteristics and changes in AltaVista... (identifier) <http://doi.wiley.com/10.1002/asi.20145>. Institution: (description) School of Information Sciences, University of Pittsburgh, 610 IS Building, 135 N. Bellefield Avenue, Pittsburgh, PA 15260.
 - 1.3 Paper:** (title) **Web searching on the Vivisimo search engine**. Abstract: The application of clustering to Web search engine...
- Additional Information:** A sidebar on the right providing further details:
 - Amanda Spink's most frequent keywords are: information retrieval, web, search engine, evolutionary; and she cooperates with 20 distinct authors. See Amanda Spink in Google Scholar.
 - H. Cenk Ozmutlu cooperates with: - Amanda Spink; - Seda Ozmutlu; See H. Cenk Ozmutlu in Google Scholar.
 - Seda Ozmutlu cooperates with: - H. Cenk Ozmutlu; - Amanda Spink; See Seda Ozmutlu in Google Scholar.
 - Pennsylvania State University, University Park, PA 16802 has published 18 paper(s), written by 36 author(s), and its most frequent keywords are:

At the bottom left, a status bar indicates 'Applet ontograph started'.

Figura 25 - Tela resultante da busca por "author:Amanda Spink"

Quando recuperou-se a autora com a sua respectiva lista de *papers*, o *conceito central* foi automaticamente alterado para *author*. Como são apresentadas além das instâncias do

conceito central, as instâncias sobre conceitos que estão diretamente relacionados com o *conceito central*, e de acordo com a ontologia, o conceito *author* está relacionado apenas com os conceitos *paper* e *institution*, para visualizar as informações dos *papers*, que é o objetivo, pode-se simplesmente trocar o *conceito central* no componente de navegação de ontologia.

Na Figura 26 é apresentada a tela em que se pode notar que o *conceito central* é *paper* e que no componente de instâncias recuperadas é apresentada uma lista contendo as 11 instâncias de *paper* em que *Amanda Spink* é autora. Esse tipo de funcionalidade permite que se visualize a informação recuperada em diferentes perspectivas. Nesse exemplo, da perspectiva de instâncias de autores passou-se para a perspectiva de instâncias de *paper*, e poder-se-ia visualizar, se assim fosse desejado, as perspectivas de instâncias de *institutions*, *keywords* ou *journals*. Esse agrupamento por conceitos estimula a interpretação sobre a informação recuperada, o que afeta diretamente outras características inerentes a um processo de busca, tais como aprendizado, verificação, identificação de relevância, escaneamento, serendipidade, e assim por diante.

The screenshot shows a web application titled "Semantic-based Retrieval". At the top, there is a search bar containing the query "author:Amanda Spink; paper:" and a "Search" button. Below the search bar is a navigation pane on the left labeled "Ontology Navigator" which displays a graph with nodes for "Institution", "author", "Journal", "paper", and "keyword". The "paper" node is highlighted in red. The main content area is titled "Result (paper)" and shows "11 results". Two results are visible:

- Paper: (title) Information seeking and mediated searching. Part 4. Cognitive styles in information seeking**
(abstract) This is the fourth in a series resulting from a joint research project directed by Professor Tom Wilson in the United Kingdom and Dr. Amanda Spink in the United States. The analysis reported here sought to test a number of...
(identifier) <http://doi.wiley.com/10.1002/asi.10084>
Keywords: (keywordname)
Authors: (name) Nigel Ford; David Ellis; Allen Foster; T.D. Wilson; Amanda Spink;
Journal: (journalname)Journal of the American Society for Information Science and Technology; (volume)53; (issuenumber)9; (year)2002;
- Paper: (title) A temporal comparison of AltaVista Web searching**
(abstract) Major Web search engines, such as AltaVista, are essential tools in the quest to locate online information. This article reports research that used transaction log analysis to examine the characteristics and changes in AltaVista...
(identifier) <http://doi.wiley.com/10.1002/asi.20145>
Keywords: (keywordname) web search; web;
Authors: (name) Bernard J. Jansen; Jan Pedersen; Amanda Spink;

On the right side, there is an "Additional Information" pane. It contains text about Amanda Spink's frequent keywords (information retrieval, web, search engine, evolutionary) and lists other authors she cooperates with: H. Cenk Ozmutlu and Seda Ozmutlu. It also mentions that Pennsylvania State University has published 18 papers by her.

Figura 26 - Tela que apresenta a busca por "author:Amanda Spink" com *paper* como *conceito central*

É importante ressaltar, também, que quando se está em um processo de busca por informação exercita-se a atividade descrita pela literatura como *tentativa-e-erro*. Essa é uma atividade comum e que é viabilizada pelo sistema, baseado no framework Hermeneus, por definir um ambiente interativo em que os usuários podem facilmente, conforme descrito anteriormente, utilizar os componentes para experimentar a busca por novos termos contextualizados. Eles podem, ainda, visualizar a informação recuperada por diferentes perspectivas e seguir caminhos sugeridos pela extração de informação adicional. Um usuário pode simplesmente digitar uma palavra que generalize a sua necessidade e, a partir daí, ter condições de apenas com cliques de mouse navegar nos componentes e nas instâncias recuperadas visando encontrar a informação requerida. Nesse exemplo mostrou-se essa capacidade – digitando apenas o termo “*seeking*”, encontrou-se os *papers* de uma determinada pesquisadora que possui assuntos relacionados diretamente com uma necessidade inicial.

5.2 A utilização de ontologia como suporte ao desenvolvimento de sistemas de busca interativos

5.2.1 Construção de sistemas de busca baseados em ontologia

Um dos objetivos da criação do framework Hermeneus visa possibilitar que especialistas de domínio com ou sem conhecimento da área de RI possam gerar sistemas de busca interativos. Isso permite que esses especialistas possam focar a sua atenção somente no conhecimento do domínio e não se preocupar com detalhes de implementação de sistemas. Dessa forma, problemas de comunicação entre possíveis atores (*e.g.*, especialistas de domínio, designers, desenvolvedores, etc.) envolvidos no desenvolvimento de um sistema podem ser evitados. Além disso, o tempo de desenvolvimento de um sistema de busca assim como aprimoramento da usabilidade e utilidade de tal sistema também é encurtado. O Hermeneus está de acordo com a proposição de Hendry e Harper (1996), que sugerem que “uma arquitetura pode reduzir a distância cognitiva entre conceitos em um problema de domínio e as abstrações necessárias em um programa de computador”.

O Hermeneus foi concebido para construir automaticamente os mecanismos requeridos por um sistema de busca interativo com base em uma ontologia de domínio e sua base de conhecimento. O módulo de indexação, por exemplo, gera índices semânticos de acordo com os conceitos definidos na ontologia. Maarek (1991) cita que “a maior vantagem na indexação automática, por exemplo, sobre a indexação manual, além do custo, obviamente, é a criação

de um esquema unificado que garante a compatibilidade entre índices”. A criação de índices semânticos possibilita a visualização do resultado (*e.g.*, mudar de um agrupamento de *papers* e autores para autores e seus *papers*) em diferentes perspectivas sem haver impacto na performance. Tais índices também viabilizam a manipulação de consultas semânticas – portanto, recuperação mais precisa. De acordo com Vallet et al. (2005), as ontologias podem superar as limitações de busca baseada em termos e proporcionar melhor *recall* e *precision*. Cabe ressaltar que o foco desta pesquisa está em propor um modelo mais pragmático de RBI em conjunto com o framework Hermeneus e não no desenvolvimento de algoritmos específicos de RI.

Desenvolveu-se o Hermeneus para ser modular, flexível, configurável e extensível. A sua arquitetura, genericamente apresentada na Figura 14 (pg. 65), é composta de quatro módulos e três componentes que facilitam a adaptação às especificidades exigidas por certos domínios. Por exemplo, um designer pode modificar a forma de visualização das instâncias recuperadas (*i.e.*, classe *RetrievedInstances* da Figura 14) alterando o componente de instâncias recuperadas ou, até mesmo, construir novos componentes interativos e se utilizar dos demais módulos da arquitetura. Regras de nível de ontologia e instâncias podem ser configuradas no módulo de inferência (*i.e.*, classe *InferedModel* da Figura 14). O protótipo, por exemplo, estendeu as relações entre conceitos por meio de regras específicas para extrair informação adicional da ontologia e da base de conhecimento.

A característica referente à modularidade indica que o Hermeneus pode ser mais facilmente adaptado e atualizado. Por exemplo, um novo algoritmo de ordenação ou medida de similaridade pode ser desenvolvido sendo necessário apenas alterar o módulo de recuperação (*i.e.*, classe *RetrieveModel* da Figura 14). As regras de ontologia na presente proposta precisam ser descritas em *Jena Rule Language*. Entretanto, pode-se alterar o módulo de inferência (*i.e.*, *InferedModel* da Figura 14) para descrever as regras na linguagem *RuleML*⁶ ou *SWRL*⁷ e utilizar outra máquina de inferência (*e.g.*, Pellet⁸) para interpretá-las. Esses tipos de modificações não requerem alterações em todos os módulos e componentes da arquitetura – portanto, a modularidade, a flexibilidade, a extensibilidade e a utilização de ontologia são características que tornam a arquitetura mais robusta.

⁶ <http://www.ruleml.org/>

⁷ <http://www.w3.org/Submission/SWRL/>

⁸ <http://pellet.owldl.com/>

5.2.2 Extração de informação adicional

Durante o processo de busca por informação os usuários normalmente focam em detalhes que podem ajudá-los a encontrar o que estão procurando. Uma lista resultante com apenas uma pequena parte do conteúdo de cada documento recuperado não é suficiente para que o significado do documento seja compreendido. Na presente proposta, a extração de informação adicional permite que usuários interajam com informação que não está definida nas instâncias recuperadas. Tal informação ajuda usuários a terem um entendimento melhor sobre o domínio e permite que eles aprendam mais sobre uma instância (Golbreich, 2004; Murdock, Pinheiro Da Silva *et al.*, 2005).

Inferência, de acordo com Decker et al. (1999), é um processo lógico de derivação de novos dados a partir de uma coleção de dados. Ao se utilizar ontologia, essa técnica combinada com regras possibilitam a derivação de novas informações. No protótipo, por exemplo, quando as instâncias recuperadas estão agrupadas pelo conceito *paper*, ao selecionar uma instância específica é apresentada informação inferida sobre os respectivos autores e suas instituições. Nesse caso, conforme mostra a Tabela 6, é apresentada a lista de palavras-chave mais freqüentes, considerando todos os *paper* armazenados na base de conhecimento, as instituições às quais os autores estão vinculados e os colaboradores dos autores. Além disso, é possível visualizar a lista de palavras-chave mais freqüentes para cada instituição e os *papers* que mencionam a instituição. A informação adicional pode também ser usada como uma atividade exploratória, pois é possível interagir com a informação inferida para refinar buscas dinamicamente. Esse processo interativo tem o mesmo princípio implementado no componente de instâncias recuperadas.

Tabela 6 - Exemplo de informação adicional

Instância recuperada (<i>paper</i>)	Informação adicional
<p>Title: The double role of ontologies in information science research</p> <p>Author: Frederico Fonseca</p> <p>Keywords: information systems; ontology; ontologies; engineering; science; information science; ontology-driven information science;</p>	<p>Frederico Fonseca is connected to Pennsylvania State University, University Park, PA 16802, his most frequent keywords are: ontologies, ontology, information science, engineering, science; and s/he cooperates with:</p> <ul style="list-style-type: none"> - Gilberto Câmara; - James E. Martin; <p>Pennsylvania State University, University Park, PA 16802 has published 18 paper(s), written by 36 author(s), and its most frequent keywords are: web, web search, web search engine, information retrieval, ontologies.</p>

5.2.3 Interação gráfica com a ontologia

O objetivo de uma ferramenta gráfica é facilitar que seres humanos consigam adquirir percepções mais aguçadas sobre dados. Visualizar uma ontologia significa visualizar seus conceitos e suas relações (Spence, 2000). A vantagem em apresentar uma ontologia graficamente está na sua expressividade, pois os conceitos e as suas relações (*i.e.*, o vocabulário de um domínio) tornam-se de fácil identificação (Fluit, Sabou *et al.*, 2002). No Hermeneus, a interface do componente de navegação de ontologia foi concebida de modo a estimular o usuário a explorar a informação disponibilizada. Esse componente pode auxiliar os usuários a entenderem melhor uma falta de conhecimento além de especificar a informação requerida mais precisamente. A possibilidade de interação direta com conceitos definidos pela ontologia, conforme descreve Tang (2007), estimula os usuários a articularem diferentes aspectos de suas necessidades.

Os conceitos usados em uma consulta semântica são disponibilizados com uma cor diferente na ontologia. Por exemplo, efetuando-se a busca por “*author: Pacheco*”, o conceito *author* é pintado com a cor amarela (os demais conceitos são cinza). Além disso, quando se posiciona o mouse sobre o conceito *author* uma caixa (*i.e.*, similar a um componente de dica) apresenta o conteúdo “*Pacheco*”. Esse tipo de característica gráfica ajuda os usuários a contextualizarem as suas necessidades ao mesmo tempo que o seu entendimento sobre o problema está sendo desenvolvido.

A ferramenta gráfica também habilita os usuários a utilizarem a ontologia para visualização dos resultados em diferentes perspectivas. Essa funcionalidade já foi descrita anteriormente, mas é importante enfatizar que tais mudanças de perspectiva sobre o resultado ocor-

rem por meio de cliques de mouse sobre os conceitos disponíveis graficamente. O tipo de informação que está sendo apresentado está relacionado com o *conceito central*. Tal conceito é desenhado com uma forma diferente (*i.e.*, quadrado, os demais são redondos) de modo a facilitar a identificação visual do tipo de informação apresentada.

Cabe ressaltar que o componente de navegação da ontologia, sugerido pelo framework, tem um tipo de apresentação que está mais adequada a uma ontologia que não possui muitos conceitos. A visualização gráfica de ontologias (Shneiderman, 1996; Fluit, Sabou *et al.*, 2002) é uma área de estudo dentro de visualização de informação (Card, Mackinlay *et al.*, 1999; Spence, 2000; Eidenberger e Breiteneder, 2003; Albertoni, Bertoni *et al.*, 2005). Há vários desafios para conceber um componente visual que permite a interação gráfica com ontologias que possuem centenas ou até milhares de conceitos. No âmbito desta proposta, em caso de haver a necessidade de uma ontologia com vários conceitos, sugere-se, por exemplo, a disponibilização dos conceitos em uma estrutura de facetas (Mcguinness, 1999; Albertoni, Bertoni *et al.*, 2005; Mäkelä, Hyvönen *et al.*, 2006; Tang, 2007), a utilização de apenas parte da ontologia (*i.e.*, apenas alguns conceitos) ou a definição de mecanismos dinâmicos de filtragem aplicados à ontologia. Contudo, esse tipo de desafio pode ser melhor contextualizado, no âmbito deste trabalho, como pesquisa futura, pois há uma série de questões que devem ser analisadas e contextualizadas.

5.2.4 Índice semântico comparado ao índice invertido tradicional

O framework Hermeneus contém um índice semântico que foi concebido para armazenar as instâncias da ontologia em uma estrutura que garanta a possibilidade de recuperar informação contextualizada de modo performático. A Figura 9 (página 58) apresenta graficamente a estrutura do índice. Para a construção do índice, como mencionado anteriormente, foi feita uma adaptação no modelo de índice invertido, sendo acrescentado o conceito, definido em uma ontologia, que está vinculado a cada termo pertencente a uma instância. O índice invertido tradicional não disponibiliza a possibilidade de inserir informação adicional aos termos indexados.

De modo a mostrar mais claramente a diferença entre ambas as abordagens, utiliza-se o exemplo a seguir, disponível na Figura 27, que considera uma ontologia relativa ao domínio de palestras com apenas três conceitos e duas instâncias. A primeira instância significa que o autor “*Roberto Pacheco*” proferiu a palestra “*A UFSC e o Conhecimento*” na “*PSU*”. A

segunda instância descreve a apresentação da palestra “A PSU e a filosofia”, na “UFSC”, pelo autor “Fred Fonseca”.

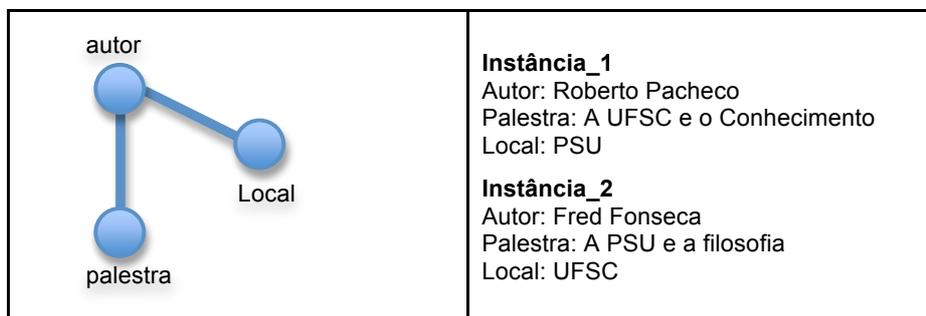


Figura 27 - Exemplo de ontologia e instâncias para comparação do índice semântico com o índice invertido tradicional

Um índice invertido tradicional com as informações das instâncias apresentadas na Figura 27 não armazena o contexto de cada termo no índice. A indexação nesse caso ocorre considerando apenas o texto e desconsidera a informação que contextualiza um termo no documento. Dessa forma, o conteúdo a ser indexado para a instância 1 é “*roberto pacheco a ufsc e o conhecimento psu*” e para a instância 2 é “*fred fonseca a psu e a filosofia ufsc*”. A seguir, na Figura 28, é possível visualizar como fica a estrutura do índice invertido tradicional. A primeira coluna contém a lista de termos, a segunda a quantidade de ocorrências do termo no conjunto de documentos e a terceira coluna apresenta a lista de documentos com a respectiva posição do termo no documento.

Termos	Freqüência	Lista de documentos
a	3	(Inst_1, 17) (Inst_2, 14) (Inst_2, 22)
conhecimento	1	(Inst_1, 28)
e	2	(Inst_1, 24) (Inst_2, 20)
filosofia	1	(Inst_2, 24)
fonseca	1	(Inst_2, 6)
fred	1	(Inst_2, 1)
o	1	(Inst_1, 26)
pacheco	1	(Inst_1, 9)
psu	2	(Inst_1, 41) (Inst_2, 16)
roberto	1	(Inst_1, 1)
ufsc	2	(Inst_1, 19) (Inst_2, 34)

Figura 28 - Exemplo de estrutura do índice invertido tradicional

Uma consulta em que o usuário deseja visualizar as palestras e autores que ocorreram na “UFSC”, ao se utilizar o índice invertido tradicional, receberia como resposta os dois documentos, pois o termo por si só, *i.e.*, sem contexto, é encontrado em ambas os documentos. Entretanto, ao se verificar as instâncias da ontologia nota-se claramente que há apenas uma

palestra (“*A PSU e a filosofia*”), feita por um autor (“*Fred Fonseca*”), que ocorreu na “*UFSC*”. Vale salientar, portanto, que ao se construir o índice invertido tradicional toda a referência semântica é perdida e não há de modo algum como saber com exatidão, por exemplo, quais palestras foram proferidas na “*UFSC*” e por quem.

Já o índice semântico permite que a busca por “*UFSC*” possa ser contextualizada de forma a retornar as palestras e autores relativos àquele local. A Figura 29 apresenta a estrutura do índice semântico para o exemplo da Figura 27. A primeira coluna contém a lista de termos, assim como no índice invertido tradicional, mas a segunda coluna, onde há a lista das instâncias relacionadas, há uma estrutura diferente. Nessa estrutura, há o identificador da instância, o conceito ou classe da ontologia a que a instância se refere e a posição do termo na instância. Dessa forma, uma consulta semântica (e.g., local:UFSC) recupera informação mais precisa sem haver comprometimento de desempenho no processo de recuperação.

Termos	Lista de instâncias
a	[inst_1 PALESTRA 1] [inst_2 PALESTRA 1] [inst_2 PALESTRA 9]
conhecimento	[inst_1 PALESTRA 12]
e	[inst_1 PALESTRA 8] [inst_2 PALESTRA 7]
filosofia	[inst_2 PALESTRA 11]
fonseca	[inst_2 autor 6]
fred	[inst_2 autor 1]
o	[inst_1 PALESTRA 10]
pacheco	[inst_1 AUTOR 9]
psu	[inst_1 LOCAL 1] [inst_2 PALESTRA 3]
roberto	[inst_1 AUTOR 1]
ufsc	[inst_1 PALESTRA 3] [inst_2 LOCAL 1]

Figura 29 - Exemplo da estrutura do índice semântico

5.2.5 Comparação da busca semântica com abordagens similares

Há várias abordagens que utilizam ontologia em sistemas de busca. Diferentemente da presente proposta, que utiliza a ontologia tanto para a construção dos mecanismos de RI quanto para o enriquecimento, e aqui mais especificamente com foco em *information seeking* e círculo hermenêutico, de componentes interativos disponibilizados na interface, os trabalhos pesquisados abordam a questão de ontologia e sistemas de busca em contextos bem específicos. A escolha das abordagens apresentadas aqui ocorreu em função de semelhanças com a presente proposta. Como citado anteriormente há várias abordagens que utilizam direta ou indiretamente conceitos relacionados à ontologia ou semântica no âmbito de sistemas de busca. Não é objetivo deste trabalho discutir mais profundamente e mais amplamente várias abordagens que utilizam busca semântica, até porque, a busca semântica é apenas uma parte

desta pesquisa. O objetivo é apresentar diferentes perspectivas e como cada uma dessas abordagens melhora sistemas de busca com a utilização de ontologia. A seguir são apresentados resumidamente alguns trabalhos que utilizam ontologia em sistemas de busca e na seqüência é apresentada uma análise comparativa entre essas abordagens e a presente proposta.

- ✚ MuseumFinland (Hyvönen, Saarela *et al.*, 2004; Hyvönen, Mäkelä *et al.*, 2005) é um portal para publicação de coleções culturais heterogêneas de museus que utiliza conceitos da Web Semântica. O sistema usa um conjunto de ontologias de forma a prover maior interoperabilidade e funcionalidades mais inteligentes de busca e navegação. Uma busca é baseada nos conceitos da ontologia e não pela comparação simples de termos que formam uma consulta. As bases de dados dos museus estão situadas em diferentes locais, o que requer que uma consulta tenha que ser executada de maneira distribuída. Utiliza a máquina de busca Ontogator (Hyvönen, Saarela *et al.*, 2003; Mäkelä, Hyvönen *et al.*, 2006) que é baseada no paradigma de multi-facetado (Hearst, Elliott *et al.*, 2002). Os conceitos usados para indexação são chamados de categorias e estão organizados em forma de uma taxonomia. Uma consulta pode ser feita de duas formas: (1) usando o paradigma de multi-facetado onde a consulta é formulada pela seleção de categorias de interesse de diferentes facetados; e (2) por meio da utilização de termos, em que cada termo é comparado com as categorias. Como resultado é apresentada a lista de categorias relacionadas à consulta e organizadas de forma hierárquica além dos itens que respondem à consulta em si;
- ✚ Textpresso (Müller, Kenny *et al.*, 2004) é um sistema de mineração de texto para literatura científica que disponibiliza uma máquina de busca que utiliza uma ontologia (em forma de taxonomia) baseada em categorias e subcategorias de termos, que são classes de conceitos biológicos, para recuperar documentos científicos. Há um mecanismo responsável por marcar essas categorias nas sentenças extraídas dos documentos para então indexá-las. Uma consulta é submetida ao sistema, que extrai sentenças para inspeção do usuário. Há uma busca simples e avançada, sendo que na avançada há as seguintes opções: buscar a combinação de categorias, subcategorias e palavras-chave; especifi-

car a frequência de ocorrências de itens em particular; e escolher onde a consulta deve ser aplicada (*e.g.*, título, resumo, corpo de texto). Quando uma consulta é aplicada sobre todo o texto de um documento, o mecanismo de busca tem a função de categorizar o texto enquanto encontra os termos da consulta. A combinação de categorias para descrever o significado de uma consulta, segundo os autores, representa a recuperação de melhores documentos do que a utilização de apenas palavras-chaves;

- ✚ SemSearch (Lei, Uren *et al.*, 2006) propõe uma busca semântica que produz respostas mais precisas por tirar vantagem da semântica explícita das informações que são descritas no contexto da Web Semântica. Esse sistema, de acordo com os autores, visa esconder a complexidade de consultas semânticas de usuários que não tenham necessariamente familiaridade com dados específicos de um domínio. A interface no estilo do Google permite a especificação explícita de consultas simples (*i.e.*, dois termos) e complexas (*i.e.*, três ou mais termos). Há um mecanismo que tenta entender *i.e.*, descobrir conceitos semânticos em uma consulta formatada pelo usuário e traduzir tal consulta em uma consulta formal. Basicamente, cada termo da consulta é aplicado a fontes de dados heterogêneas, no contexto da Web Semântica, na tentativa de encontrar conceitos, relações ou instâncias descritas nas ontologias de tais fontes de dados. Quando há casamento entre um ou mais termos da consulta com os termos que descrevem as ontologias, a consulta é alterada, tornando-se uma consulta descrita em linguagem formal, ou seja, uma consulta semântica;
- ✚ McGuinness (1999) apresenta um sistema de busca online que utiliza ontologia para recuperação de documentos da área médica. O sistema combina técnicas de representação de conhecimento com ferramentas de RI para recuperar documentos não estruturados e semi-estruturados. Uma taxonomia foi usada para auxiliar na navegação e na expansão de termos configurados em uma consulta. Usuários podem combinar consultas simples com a escolha de uma ou mais áreas, definidas na taxonomia. Essas áreas são apresentadas em uma estrutura de árvore onde os usuários podem navegar e escolher qualquer uma delas. Os termos usados na consulta são aplicados a todas as áreas escolhidas.

Contudo, os usuários podem apenas selecionar uma das áreas e automaticamente os documentos relacionados à área selecionada são recuperados. A página resultante apresenta apenas os documentos recuperados e não apresenta a taxonomia.

O Quadro 1 a seguir, disponibiliza uma lista de características e como cada uma das abordagens de busca semântica, apresentadas anteriormente, trata cada uma delas. O quadro compara as quatro abordagens, MuseumFinland, Textpresso, SemSearch e a pesquisa apresentada por McGuinness além da presente proposta.

Quadro 1 - Apresentação comparativa de características inerentes a propostas de sistemas que oferecem busca semântica

Característica	Presente proposta	MuseumFinland	Textpresso	SemSearch	McGuinness
Principal objetivo	Utilizar uma ontologia para construir componentes de RI que possam suprir as demandas geradas por componentes interativos a serem disponibilizados para usuários, que permitem a utilização da ontologia para composição de buscas semânticas, visualização de diferentes perspectivas das informações recuperadas e a extração de informação adicional para cada instância recuperada.	É um portal semântico que publica coleções heterogêneas de museus. Se preocupa com a interoperabilidade e a disponibilização de uma busca semântica, baseada em taxonomia, que permite aos usuários buscarem e navegarem itens das coleções.	É um sistema de mineração de texto que disponibiliza um mecanismo de busca que utiliza ontologia para recuperar documentos científicos da área biológica.	Tem o objetivo de criar um sistema de fácil utilização, que tenta extrair o significado semântico de termos digitados em uma consulta para recuperar documentos melhor contextualizados.	Disponibilizar uma busca semântica sobre dados não-estruturados e semi-estruturados da área da medicina.
Modelo de RI	Define e implementa um modelo que considera o contexto (<i>i.e.</i> , a semântica) de cada informação a ser indexada e conseqüentemente buscada. A ontologia é o componente que permite a construção dessa estrutura semântica que dá suporte à criação de uma estrutura de índice semântico e a um mecanismo de busca semântica.	Não específica. Entretanto, como se refere ao Ontogator, que é o mecanismo de busca semântica, há uma referência para buscas utilizando RDQL ⁹ , o que pode representar a não utilização de um modelo de RI.	Não específica.	Não define um específico. Utiliza o Lucene ¹⁰ para indexar e recuperar documentos. A ontologia e as instâncias são armazenadas no Sesame ¹¹ .	Não cita um modelo específico. Apenas afirma que faz busca conceitual.

⁹ Linguagem de consulta para RDF (RDF Query Language)

¹⁰ Biblioteca de código fonte aberto que implementa os conceitos de RI (<http://lucene.apache.org/>).

¹¹ Biblioteca de código fonte aberto que permite armazenar, inferir e buscar dados em RDF (<http://www.openrdf.org/>).

Característica	Presente proposta	MuseumFinland	Textpresso	SemSearch	McGuinness
Índice	Define um índice semântico. Faz uma adaptação no modelo de índice invertido e armazena o contexto de cada termo na estrutura do índice.	Faz citação para a utilização de um índice, mas não define como esse índice está organizado/ estruturado.	Não cita explicitamente. Como utiliza PERL em conjunto com expressões regulares para armazenar as instâncias, é dedutível que faz consulta direta por meio de expressões regulares.	Utiliza a funcionalidade disponível no Sesame.	Não descreve.
Ordenação	Utiliza a medida de similaridade <i>coordinate matching</i> que considera a quantidade de ocorrências de termos de consultas em documentos. Também faz um corte nas instâncias recuperadas levando em consideração as consultas semânticas (quando utilizadas).	Não menciona uma medida de similaridade ou critérios de ordenação. O resultado é disponibilizado em subcategorias.	O resultado é ordenado de acordo com o número de ocorrências dos termos de uma consulta que também ocorrem nos documentos.	Usa uma combinação entre a distância entre termos de uma consulta e seus termos semânticos e a quantidade de termos nos documentos resultantes.	Não especifica.
Raciocínio	Aplica regras de inferência sobre qualquer instância recuperada. Permite que tais regras possam ser configuradas.	Utilizado para projeção de categorias e recomendação de links semânticos.	Não utiliza.	Não menciona a utilização de raciocínio, apenas heurísticas para traduzir uma consulta simples em uma consulta formal.	Utiliza inferências simples por meio de regras genéricas (e.g., herança).
Componentes interativos	Apresenta um campo texto para inserção das consultas e utiliza a ontologia para construção dos componentes interativos: navegador da ontologia, instâncias recuperadas, informação adicional.	Apresenta um campo para inserção de consultas com termos simples e facetas de uma taxonomia que são disponibilizadas para busca direta.	Apresenta um campo para inserção de consultas e disponibiliza a possibilidade de escolher, no nível mais macro, um conceito específico definido na taxonomia.	Apresenta o campo texto para inserção da consulta e disponibiliza o conceito/ instância/ relação vinculado aos documentos recuperados que podem ser usados para nova busca.	Apresenta uma árvore de navegação textual com os conceitos definidos na taxonomia e um campo texto para inserção da consulta.
Resultado da informação	Como recupera instâncias, o resultado é contextualizado e disponível para utilização dinâmica e interativa	Como o domínio é de museu e há fotografias dos objetos, o resultado apresenta os itens resultantes	Apresenta a lista de instâncias encontradas de acordo com os termos da consulta. Nessa tela resultante não	Apresenta a lista de documentos recuperados e para cada documento também apresenta o conceito/ ins-	Apenas o conjunto de documentos. Exige que o usuário volte a tela inicial caso deseje efetuar nova

Característica	Presente proposta	MuseumFinland	Textpresso	SemSearch	McGuinness
	para, por exemplo, refinamento da busca.	com suas respectivas informações e fotografia. Também são apresentadas as categorias, na forma de facetas.	permite fazer mais nada a não ser escolher uma das instâncias.	tância/ relação.	busca.
Tipo de ontologia	Ontologia de domínio.	Ontologia de domínio (sete ao todo).	Ontologia (taxonomia) de domínio específica.	Não especifica.	Taxonomia de um domínio específico.
Linguagem da ontologia	OWL (<i>Web Ontology Language</i>)	OWL (<i>Web Ontology Language</i>) e RDF (<i>Resource Description Framework</i>)	A ontologia foi armazenada em um estrutura desenvolvida em PERL com expressões regulares.	RDF (<i>Resource Description Framework</i>)	Não informa.
Disponibilização da ontologia para o usuário	Apresenta a ontologia em um componente gráfico e interativo na página inicial e de resultado.	Disponibiliza a ontologia por meio de facetas na página inicial e de resultado.	Disponibiliza as categorias definidas na taxonomia para configuração de buscas.	Não disponibiliza.	Apresenta a taxonomia em um componente em forma de árvore apenas na página inicial.
Domínio	Qualquer domínio.	Aplicado ao domínio de museus (pode ser utilizada em outros domínios).	Aplicado ao domínio da biologia.	Qualquer domínio.	Aplicação específica para área médica.
Início da busca	Formulação de uma consulta com possibilidade de interação com os conceitos da ontologia e formulação de consulta semântica.	Formulação de consulta ou a escolha de uma categoria disponível nas facetas – relativas a uma taxonomia.	Formulação de uma consulta e a possível escolha de categorias definidas pela taxonomia.	Formulação de uma consulta.	Formulação de uma consulta ou escolha de uma das áreas disponíveis na taxonomia.
Consulta semântica formulada pelo usuário	Permite utilizar os conceitos da ontologia para composição de consultas semânticas (e.g., Instituição: Stela).	O campo de busca serve para busca simples, mas ao clicar nas categorias, disponíveis nas facetas, o usuário seleciona diretamente um conceito, disparando uma busca categorizada.	Não permite que um usuário formule uma consulta semântica. A consulta semântica acontece quando o usuário seleciona uma ou mais categorias. Nesse caso, o próprio sistema compõe a consulta semântica.	O sistema tenta automaticamente traduzir uma consulta simples em consulta formal (i.e., usando semântica). Permite que usuários formulem consultas semânticas (e.g., notícias: Barack Obama).	Não permite.

5.3 Hermeneus como tradutor da informação requerida pelos usuários

Uma das principais características do framework Hermeneus é auxiliar os usuários a entender e desenvolver as informações por eles requeridas. Embora possa ser estranho que os usuários não saibam exatamente o que estão buscando, a literatura trata essa incerteza como falta de conhecimento, o que induz a utilizarem intermediários, tais como um sistema de busca, para suprir uma necessidade. Os usuários, contudo, precisam desenvolver as suas próprias necessidades, dúvidas ou questões para entender o que desejam buscar. Nesse processo, o estado cognitivo dos usuários pode apresentar problemas na interpretação e, de acordo com Ingwersen e Järvelin (2005b), alcançar uma condição de incerteza. Entender a falta de conhecimento e buscar informação são tarefas cognitivas diretamente relacionadas uma com a outra durante o ciclo de aquisição de informação. O modelo proposto refletido no Hermeneus, conforme sugerem Hert (1997), Ingwersen e Järvelin (2005a), Kuhlthau (1993) e Marchionini (1989), permite que um usuário interaja com ferramentas que possam afetar a sua condição de incerteza de modo a estimular uma transformação no seu estado de conhecimento. Vakkari (2003) diz que “é evidente que o conhecimento anterior é alterado quando um indivíduo adquire nova informação”. Essa citação de Vakkari (2003) possui uma relação íntima com o círculo hermenêutico, pois foi utilizando os conceitos do círculo hermenêutico juntamente com *information seeking* que o modelo foi concebido.

Ao interagir com um sistema de busca, a forma mais comum de expressar uma informação desejada é por meio da composição de consultas. Entretanto, formular consultas precisas não é uma tarefa simples. Consultas precisas podem ser a diferença entre resultados adequados e não adequados em uma busca (Chen, Shankaranarayanan *et al.*, 1998). No âmbito desta proposta, tratou-se essa questão ao se permitir a composição de consultas semânticas, sendo facilitado pela interação direta com os conceitos da ontologia. O trabalho apresentado por Tang (2007) deixa clara a importância da visualização da ontologia como auxílio para a composição de consultas, principalmente quando os usuários não sabem ao certo o que desejam ou quando têm pouca familiaridade com um determinado tópico.

No protótipo, usuários podem interagir com a ontologia por meio de cliques de mouse; quando os usuários clicam sobre um conceito, automaticamente o nome desse conceito é

adicionado à consulta agindo como um assistente de composição de consultas. Nesta solução, foi seguido o princípio de que, por meio da interação com a ontologia, os usuários adquirem percepções sobre as suas necessidades e conseguem se expressar com mais precisão. Além disso, os usuários podem formular consultas menos ambíguas uma vez que podem expressar explicitamente o contexto para cada termo usado na consulta (García e Sicilia, 2003; Hyvönen, Mäkelä *et al.*, 2005).

Conforme descrito por Kuhlthau (1991) e Hert (1997), o estado de conhecimento do usuário é dinâmico em vez de estático, sendo alterado à medida que ele procede durante um processo de busca. Como as instâncias de ontologia são recuperadas e disponibilizadas em um ambiente contextualizado, em que cada informação contém um link semântico com os conceitos definidos pela ontologia, os usuários podem aprimorar as suas consultas adicionando automaticamente conteúdo semântico de instâncias relevantes. Por exemplo, em uma situação comum, em que um usuário digita apenas alguns termos (*e.g.*, “*semantic*”) para expressar a sua necessidade, a recuperação de instâncias pode ajudá-lo a encontrar descrições mais precisas e usá-las dinamicamente (*e.g.*, quando se busca pelo termo “*semantic*” no protótipo, os usuários podem visualizar e interagir com as seguintes instâncias do conceito *keyword*: *latent semantic indexing*, *semantic network*, *semantic information*, *semantic-based information retrieval*, *semantic-based web retrieval systems*, *semantic web* e *semantic search*). Com esse tipo de interação exploratória, os usuários são protegidos de detalhes requeridos pela formulação de consultas semânticas (*i.e.*, digitar explicitamente termos para compor consultas semânticas), além de construir expressões mais úteis (Capra Iii e Pérez-Quñones, 2005). Por exemplo, uma busca pelo termo “*web*” no protótipo produz como resultado 169 instâncias; mas, ao clicar no termo “*semantic web*”, apresentado em uma das instâncias recuperadas, relacionado com o conceito *keyword* (*i.e.*, a consulta agora é composta pelo termo “*web*” e pelo conteúdo semântico “*keyword: semantic web*”), retorna como resultado apenas 11 instâncias. Com esse tipo de interatividade os usuários podem facilmente experimentar o conteúdo recuperado para formular novas consultas como parte do processo de entendimento de suas necessidades.

A ontologia pode também ser usada para definir quais tipos de instâncias recuperadas devem ser apresentadas por meio da seleção de um *conceito central*. Essa funcionalidade permite que usuários tenham diferentes perspectivas de uma mesma consulta. Ontologias po-

dem ser usadas para produzir diferentes visões de conhecimento, particularmente para propósitos de exploração e navegação (Staab, Studer *et al.*, 2001). Como um processo de busca de informação é dinâmico e muda constantemente (Marchionini, 1989; Kuhlthau, 1993; Vakkari, 2003; Ingwersen e Järvelin, 2005a), visualizar a informação recuperada em diferentes perspectivas pode auxiliar os usuários a terem novas percepções sobre as suas necessidades. No protótipo, por exemplo, quando se busca por “*information science*”, há 58 instâncias agrupadas pelo conceito *paper* (*i.e.*, o *conceito central* é *paper*). Alterando-se a visualização dos resultados para o conceito *author* (*i.e.*, o *conceito central* agora é *author*), visualiza-se 73 instâncias de autores. Nesse caso, o usuário consegue visualizar a lista resultante de autores e as suas respectivas informações (*i.e.*, instâncias de *papers* e *institutions* para cada autor recuperado). Essa funcionalidade está de acordo com a teoria apresentada por Vakkari (2003), que cita que “modelagem e desenvolvimento de sistemas, como funcionalidades de interface que dão suporte a usuários em articular suas necessidades por usar termos de relevantes documentos agrupados de acordo com conceitos utilizados pelos usuários em suas consultas, podem aumentar o desempenho na recuperação”.

5.4 Comparação do modelo proposto com outros modelos de RBI

Neste tópico, demonstra-se uma comparação entre o modelo proposto e os três modelos de RBI apresentados no segundo capítulo: o modelo estratificado, proposto por Saracevic (1997), o framework integrado de RBI, proposto por Ingwersen (1992) e revisto por Ingwersen e Järvelin (2005b) e o modelo episódico, proposto por Belkin (1993). O objetivo aqui não é esmiuçar detalhes de cada modelo em relação à proposta. Até porque, infelizmente, conforme também sugere Kuhlthau (2005), muitas definições em torno da área de *information seeking* não vão além de implicações sugeridas para o desenvolvimento de uma aplicação que tenha impacto direto sobre um sistema. Um modelo conceitual, segundo Järvelin e Wilson (2003), não pode ser avaliado empiricamente, porque um modelo é a base para formulação de questões e hipóteses para a definição de testes empíricos; “um modelo só pode ser avaliado em termos de valores instrumentais e heurísticos”.

5.4.1 Modelo proposto e o modelo estratificado de Saracevic (1997)

Assim como o modelo estratificado supõe, também considerou-se que os usuários interagem com um sistema de busca com o intuito de usar a informação e que o uso dessa informação está diretamente conectado com a cognição e com um modo situacional. Entretanto, Sarace-

vic (1997) não deixa claro esse processo inerente à cognição humana em que um usuário interage com uma informação e a contextualiza de um modo situacional. Com relação aos usuários, também acredita-se que em um processo de busca os níveis cognitivo, emotivo e situacional, descritos por Saracevic (1997), são características que geram influências sobre as ações a serem tomadas no processo. Recorreu-se, mais especificamente, ao círculo hermenêutico para viabilizar descrições que possam acrescentar explicações relacionadas àquelas definidas por Saracevic (1997).

A descrição do círculo hermenêutico oferece explicações de como um ser humano interpreta e compreende um novo conceito. Esse processo de compreensão leva em consideração predefinições e contextualiza as novas definições em um procedimento de *vai e vem* realizado entre as dimensões participantes da análise. O resultado dessa interação visa produzir um efeito sobre as partes individuais de um contexto geral e conseqüentemente sobre o próprio contexto em seu conjunto.

Saracevic (1997) deixa claro que a interface é o ponto de intersecção entre o usuário e o computador, mas não descreve claramente o que uma interface deve possuir. Entretanto, o autor cita apenas de modo geral que a interface instancia uma variedade de interações, mas não sugere de que forma prática tais interações devem ser realizadas. No modelo proposto há uma ontologia que é usada tanto no âmbito do “computador”, seguindo a linha de pensamento de Saracevic (1997), quanto no âmbito do usuário. Nesse contexto, é a camada intermediária que descreve a utilização de uma ontologia e suas instâncias para a concepção de componentes interativos que atendam à determinados requisitos. O modelo também utiliza a metáfora do círculo hermenêutico, relativa ao *vai e vem* da interação que ocorre a partir dos conceitos para as instâncias e de volta para os conceitos, para a concepção de ambientes de busca que estejam mais aproximados ao modo como o ser humano compreende. O framework Hermeneus é um exemplo de demonstração pragmática desse ambiente, com componentes que permitem uma variedade de tipos de interação (detalhes sobre as interações foram descritos na Tabela 2, pg. 51).

5.4.2 Modelo proposto e o framework integrado de Ingwersen (1992) e revisto por Ingwersen e Järvelin (2005b)

O modelo de RBI proposto por Ingwersen (1992) e revisto por Ingwersen e Järvelin (2005b) descreve vários atores cognitivos que processam informação – sete no total. Essa é a primeira

característica que se difere da presente proposta, pois apenas o usuário, que é aquele que possui uma necessidade por informação, é considerado. Decidiu-se apenas criar um modelo mais direto e objetivo, ou seja, um modelo que atenda o usuário que busca informação. Entretanto, cabe enfatizar que as características definidas pelo modelo requerem certos componentes para a construção de um ambiente que proporcione um comportamento de *information seeking*. Isso significa que os atores envolvidos na definição desse tipo de ambiente precisam estar sintonizados com tais características. A utilização de índices semânticos, por exemplo, induz o ator responsável por criar índices a construir uma estrutura que permita a utilização de consultas semânticas.

Ingwersen e Järvelin (2005b) definem que os usuários sofrem influências de cunho cultural, social e organizacional. O modelo estratificado apresentado por Saracevic (1997) ainda mostra que os usuários sofrem influências em um contexto situacional. Heidegger (1996) e Gadamer (1989), quando descrevem o círculo hermenêutico, mencionam que quando as pessoas estão em um processo de interpretação elas não podem esquecer os seus preconceitos. O círculo hermenêutico foi utilizado para tentar encontrar explicações sobre como um processo de compreensão ou interpretação ocorre, com vistas a aprimorar o entendimento em torno de um usuário que desenvolve um processo de busca – que está intimamente ligado ao processo de compreensão ou interpretação executado por seres humanos. Ao se pesquisar o círculo hermenêutico, entendeu-se que a informação recuperada deve ser organizada de forma a induzir os usuários a interpretá-la como ela “deseja ser interpretada” – a utilização de ontologia auxilia nesse processo uma vez que representa formalmente o vocabulário que descreve o conhecimento de um domínio.

Descreveu-se, da mesma forma como o fazem Ingwersen e Järvelin (2005b), que um usuário, quando interage com a interface de um sistema de busca por intermédio de componentes interativos, sofre influência em seu estado cognitivo. Essa influência é refletida diretamente em ações do usuário sobre esses componentes. Além disso, tal interação também afeta o estado do sistema, uma vez que os objetos de informação estão diretamente alinhados com a requisição ou interação desse usuário. Uma ação do usuário afeta o estado do sistema de busca que, por sua vez, também afeta o estado cognitivo do usuário. No modelo proposto, descreve-se esse processo como um processo em que o usuário busca um melhor entendimento sobre seu problema ou necessidade para mais claramente definir quais são as suas

questões. Novamente, o círculo hermenêutico serve para ajudar a explicar e entender esse processo. Nesse caso, é importante enfatizar a importância da interação entre ontologia e as suas instâncias no *vai e vem* descrito pelo círculo hermenêutico.

Os objetos de informação (*e.g.*, dicionários, conteúdo, estruturas, etc.), segundo o modelo de Ingwersen e Järvelin (2005b), podem influenciar um ator cognitivamente. Na presente proposta tais objetos são representados por instâncias de uma ontologia que, por meio da interface, permite que o conhecimento de um domínio seja manipulado diretamente por usuários. É possível relacionar os objetos de informação definidos por Ingwersen e Järvelin (2005b) como sendo a base de conhecimento que está definida na camada de suporte. A ontologia e as suas instâncias estão explicitamente disponíveis para os usuários o que pode influenciá-los cognitivamente de maneira mais direta.

5.4.3 Modelo proposto e o modelo episódico de Belkin (1993)

As interações descritas por Belkin (1993) em seu modelo de RBI, como, por exemplo, avaliação, interpretação, modificação e navegação, são consideradas tarefas cognitivas exercidas por usuários. Belkin (1993) referencia também outro trabalho, desenvolvido por Belkin et al. (1993), para descrever algumas dimensões ou facetas que caracterizam as estratégias de *information seeking*. Conforme descrito na Tabela 2 (pg. 51), tais características foram referenciadas como tarefas inerentes aos tipos de interações. No trabalho apresentado por Belkin et al. (1993) os próprios autores descrevem que as dimensões por eles citadas não são suficientes para representar toda a realidade presente em estratégias de *information seeking*. Acrescentou-se, ao definir o presente modelo, a utilização da teoria do círculo hermenêutico para auxiliar no entendimento de um processo de busca.

O modelo de Belkin (1993) baseia-se em duas suposições: (1) um processo de *information seeking* é essencialmente interativo; e (2) sistemas de busca têm por objetivo dar suporte a um comportamento de *information seeking*. O modelo proposto também leva em consideração essas suposições deixando claro que um sistema de busca precisa, obrigatoriamente, disponibilizar componentes interativos que possibilitam aos usuários exercerem um comportamento de *information seeking*.

Belkin (1993) descreve, ainda, que um sistema de busca precisa entender as características de um usuário visando oferecer informação que possa ser interessante para a sua bus-

ca. Entretanto, nesta proposta acredita-se que um usuário precisa ter liberdade para interagir com toda a informação disponível em um domínio e percorrer um caminho que se adapte à forma de entendimento sobre o seu problema. Um sistema de busca que utiliza informação de acordo com alguns passos executados pelo usuário, por exemplo, pode direcionar uma busca para caminhos que não necessariamente levam à informação desejada, principalmente durante a fase em que um usuário está desenvolvendo as suas perguntas. Esse tipo de “restrição”, mesmo que em sistemas de busca interativos, pode influenciar negativamente um usuário durante um processo de busca.

Diferentemente do modelo proposto por Belkin (1993), definiu-se claramente que o conhecimento de um domínio deve ser mapeado explicitamente. Uma informação contextualizada é mais fácil de ser digerida cognitivamente do que uma informação sem contexto (*e.g.*, o termo “Java” pode representar diferentes significados, mas ao se acrescentar um contexto, como, por exemplo, “Localidade” ou “Linguagem de Programação”, a compreensão do significado do termo é facilitada consideravelmente). Também é importante considerar que a ontologia permite que os usuários consigam identificar mais facilmente quais são os conceitos definidos pelo domínio.

Por exemplo, um usuário que deseja comprar um automóvel mas que não tem muito conhecimento sobre modelos e características que devem ser analisadas durante um processo de compra. Em um sistema tradicional de RI, o usuário teria de transcrever as suas dúvidas por meio de palavras-chave, encontrar algum documento que pudesse fornecer algumas dicas relacionadas a modelos e características para, então, refazer a sua consulta com outras palavras-chave que melhor expressassem a sua “nova” necessidade. Um sistema de busca interativo que estivesse de acordo com a proposta de Belkin (1993) pode facilitar essa busca oferecendo recursos mais avançados e que permitam que o usuário navegue na informação recuperada. Entretanto, ainda assim o usuário pode ter dificuldades em interpretar a informação recuperada se essa informação não estiver contextualizada. Nesse caso, se, além de um ambiente interativo, o usuário também conseguisse visualizar os conceitos definidos pelo domínio de automóveis várias dúvidas poderiam ser resolvidas apenas com a interação com esse conhecimento (*e.g.*, o conceito “modelo” tem como instâncias “Golf”, “307”, “Omega”, “Tigra”, “Brasília”, “Apollo”, etc.).

5.4.4 Análise comparativa entre o modelo proposto e os demais modelos

Neste tópico é apresentado uma comparação, disponível no Quadro 2, que descreve algumas características do modelo proposto e dos demais modelos de RBI descritos anteriormente. O que fica claro com essa comparação é a abstração que cada modelo aplica quando define características que devem ser analisadas ao se desenvolver um sistema de busca. O modelo proposto visa ser mais pragmático, direcionando à concepção de um sistema de busca que ofereça componentes interativos tendo como base o conhecimento de um domínio, e que possua uma camada de suporte robusta e organizada.

Quadro 2 - Comparação entre os modelos de RBI

Característica	Modelo proposto	Modelo estratificado	Framework integrado	Modelo episódico
Interatividade	A interação acontece em componentes interativos que estão baseados em uma ontologia e suas instâncias. O modelo contextualiza os seguintes tipos de interatividade: aprendizagem, busca, composição de consulta, escaneamento, intencionalidade, navegação, reformulação de consulta, refinamento, relevância, serendipidade, tentativa e erro e verificação. Também utiliza a metáfora do círculo hermenêutico para mostrar que usuários podem navegar nos conceitos da ontologia e nas instâncias desses conceitos em uma forma de <i>vai e vem</i> .	É uma seqüência de processos em vários níveis que estão conectados. O modelo também cita que a interatividade é o diálogo entre o usuário e o computador que ocorre por meio da interface e que tem por objetivo principal afetar o estado cognitivo do usuário. O modelo também considera a interatividade como o principal componente em sistemas de busca.	A interação ocorre por meio de componentes cognitivos e transformações em um sistema de busca.	A interação do usuário é o processo central em um sistema de busca. A interação ocorre em uma série de episódios de <i>information seeking</i> . Há uma série de interações tais como avaliação, interpretação, modificação, navegação, etc.
O usuário	É o autor cognitivo que possui uma necessidade ou identifica uma falta de conhecimento. É o elemento que se comunica diretamente com a camada intermediária. Ele interpreta os sinais gerados por componentes disponíveis na interface e reage a eles, tendo sempre em mente uma necessidade (que pode ser dinâmica – é redefinida à medida que novas informações são compreendidas).	É um dos principais elementos do modelo. O modelo sugere três níveis de “coisas” que usuários realizam: (a) cognitivo: os usuários interpretam cognitivamente os textos recuperados; (b) emotivo: usuários interagem com suas intenções (e.g., crenças, motivação, desejos, etc.); e (c) situacional: usuários interagem em função de uma determinada situação ou problema.	É o autor cognitivo que necessita de informação e que possui modelos de suas necessidades, seus problemas e seus objetivos que são normalmente implícitos, mas que são capazes de serem explicitados. Busca por informação que deseja possuir baseando-se em problemas, interesses ou tarefas.	Deve ser o componente central de um sistema de busca e possui o chamado estado anômalo de conhecimento que o induz a engajar em um comportamento de <i>information seeking</i> . Usuários não são apenas recipientes passivos de mensagens, mas ativos buscadores e ativos construtores de significado.
Principais componentes	Usuário, informação requerida, comportamento de busca, com-	Usuário (níveis cognitivo, afetivo e situacional), interface e compu-	Contexto organizacional, cultural e social, ator cognitivo, interface,	É o mais abstrato dos modelos. Cita apenas o usuário, interativi-

Característica	Modelo proposto	Modelo estratificado	Framework integrado	Modelo episódico
	ponentes interativos (interface), ontologia, instâncias, mecanismo de recuperação, índices semânticos e base de conhecimento.	tador (níveis de engenharia, processamento e conteúdo).	objetos de informação e lógica /algoritmos.	dade e os modelos tradicionais de RI.
A interface	Mecanismo localizado na camada intermediária. É composta por componentes interativos baseados na ontologia e suas instâncias. Permite que os usuários executem uma variedade de tarefas e articulem ações relativas a suas questões e necessidades. Enfatiza a necessidade de componentes baseados na ontologia e suas instâncias que permitam que a metáfora do círculo hermenêutico possa acontecer.	É responsável por propiciar a interação na qual os usuários realizam um diálogo executando comandos e recebendo respostas, navegando entre os recursos de informação, visualizando resultados e provendo feedbacks.	Há um ator cognitivo específico para o design da interface. Esse ator sofre influência organizacional, cultural e social e precisa adquirir conhecimento de modo a conceber componentes de impacto para a interface.	Não define um papel explícito para a interface. Entretanto, é possível identificar que as interações, divididas em facetas, devem ocorrer na interface. Além disso, o modelo deixa claro que um sistema de busca deve oferecer acesso e manipulação direta à informação recuperada, prover suporte para a navegação e criar estruturas que viabilizem o diálogo.
Aspectos computacionais	São explicitamente definidos nas camadas intermediária e de suporte. Na camada intermediária descreve a necessidade de componentes interativos que devem seguir certas características além de permitir que a metáfora do círculo hermenêutico possa acontecer. Na camada de suporte descreve a necessidade de geração de índices semânticos, mecanismo de recuperação que usufrua dos índices semânticos e de uma base de conhecimento.	São descritos em três níveis: (a) engenharia: refere-se à características computacionais, como, por exemplo, capacidade e eficiência; (b) processamento: vinculado a software, mas com interesse específico em algoritmos que manipulam texto, consultas, e processos que ocorrem na interface; e (c) conteúdo: refere-se aos recursos de informação (e.g., textos, imagens, sons, etc.).	Dos sete atores descritos pelo modelo, quatro são responsáveis por aspectos computacionais do framework. Esses autores são responsáveis por: (1) criar os tipos de objetos de informação; (2) analisar os índices; (3) estruturar uma base de dados, implementar os algoritmos de indexação e busca, etc.; e (4) desenvolver as funcionalidades da interface e recuperação. Para cada autor há um detalhamento de suas obrigações.	Não há definição de mecanismos formais, apenas a citação dos mecanismos de modelos tradicionais de RI que devem oferecer suporte à interação e um mecanismo que possa entender as características do usuário de modo a auxiliá-lo em um processo de <i>information seeking</i> . Entretanto, os autores sugerem que aspectos computacionais devem considerar: como incorporar o usuário como componente central, como identificar e suportar comportamentos de <i>information seeking</i> e como tornar a interação como o processo central de um sistema de busca.

Característica	Modelo proposto	Modelo estratificado	Framework integrado	Modelo episódico
Busca	O usuário se depara com a falta de conhecimento em um determinado assunto. Essa necessidade exige que o usuário entre em um comportamento de busca. O usuário define os termos que possam representar a necessidade e desenvolve uma série de tipos de interações sobre os componentes interativos, que vão desde a reformulação de uma consulta, navegação e refinamento, até a verificação, análise de relevância e aprendizagem.	Os usuários definem comandos e recebem respostas por meio da interface. Sobre o resultado é possível navegar nos recursos de informação e prover feedbacks.	Não define explicitamente como um usuário faz uma busca. Cita como exemplo uma análise de relevância e feedbacks sobre o resultado de um sistema. Enfatiza que o sistema precisa ser interativo.	No decorrer dos episódios de <i>information seeking</i> , os usuários mudam de um tipo de interação para outro, e à medida que vão resolvendo o problema eles engajam em diferentes tipos de interações considerando objetivos, conhecimento e intenções.
Informação requerida	Ocorre quando um usuário conscientemente identifica uma falta de conhecimento. Essa ausência de conhecimento induz um usuário a entrar em um comportamento de busca.	Uma situação é a razão que influencia uma necessidade. Todavia, o problema pode não estar bem definido o que pode induzir a redefinição durante a interação.	Interação social, organizacional e social são contextos que instigam usuários a necessitar de informação.	É definida como o desejo de gerenciar um problema, resolver uma situação problemática, responder a falta de algo. Aspectos cognitivos e situacionais induzem a uma necessidade que é representada por meio do estado anômalo de conhecimento.
Comportamento de busca	Comportamento humano baseado em pensamentos, sentimentos e ações no qual um usuário engaja de modo a alterar seu estado cognitivo. O comportamento de busca está intimamente ligado à compreensão e interpretação de novas informações. Utiliza, também, a teoria do círculo hermenêutico que explica o processo de entendimento inerente a seres humanos.	O usuário se depara com uma situação que o induz a um processo de diálogo através de episódios que ocorre com um sistema de busca. À medida que o usuário interage novos termos são escolhidos e outros abandonados, táticas são adaptadas e alteradas e assim por diante.	O usuário inicialmente tem um objetivo ou talvez uma idéia e até uma teoria. Com a teoria ele gera uma hipótese. Cada hipótese é testada no contexto da teoria visando validar a teoria ou redefini-la. Usuário é influenciado pelo seu contexto no domínio que pode incluir o conhecimento anterior, prevalecendo crenças e tradições sobre o domínio e a partir de seus colegas.	Os usuários entram em um comportamento de busca para conhecer sobre o mundo, para aprender, entender, se informar, fazer o trabalho e resolver seus problemas. Um comportamento de busca é essencialmente um processo interativo. Os autores vão mais longe e afirmam que o comportamento de busca pode ser caracterizado de acordo com quatro facetas: objetivo da interação (aprender, selecionar), método da interação (escanear, bus-

Característica	Modelo proposto	Modelo estratificado	Framework integrado	Modelo episódico
				car), modo de recuperação (reconhecer, especificar) e tipo de recurso interagido (informação, meta-informação).
Componentes interativos	Os componentes interativos devem: (a) disponibilizar a ontologia para interação direta pelos usuários preferencialmente em modo gráfico; (b) recuperar informações e contextualizá-las no domínio de modo a possibilitar o refinamento de uma consulta com as informações recuperadas; (c) permitir a composição de consultas que possam usufruir de conceitos definidos na ontologia (consultas semânticas); (d) disponibilizar conhecimento adicional para cada informação recuperada e permitir que esse conhecimento possa ser utilizado de modo interativo; e (e) ser dinâmicos, intuitivos e, principalmente, interativos.	Não especifica componentes interativos. Cita apenas que a interface deve propiciar o diálogo entre o usuário e o computador.	Não define componentes. Sugere que um sistema de busca deve prover mecanismos para que o usuário se torne mais informativo.	Não deixa claro como os componentes interativos devem ser desenvolvidos. Apenas cita a importância da interatividade e da necessidade de interação direta com o conteúdo.
Conhecimento do domínio	É um dos componentes centrais do modelo, representado pela ontologia de domínio e responsável por guiar a construção dos componentes pertencentes a camada de suporte e intermediária.	Não menciona.	Não menciona.	Não menciona.
Modelo de RI	Não define um modelo específico. Entretanto, caracteriza a necessidade de indexação e recuperação de informação contextualizada.	Não sugere um modelo de RI específico.	Não cita modelos de RI.	Não sugere um modelo de RI específico.

5.5 Considerações finais

Neste capítulo foi demonstrada a viabilidade do modelo proposto de RBI por meio de um protótipo baseado no framework Hermeneus. O protótipo disponibiliza um ambiente interativo em que os usuários conseguem navegar no conhecimento do domínio, formular consultas semânticas, refinar buscas dinamicamente, visualizar diferentes perspectivas da informação recuperada e extrair informação adicional. Também foi mostrado que os usuários conseguem navegar entre os conceitos e as instâncias em uma forma de *vai e vem*, de acordo com a metáfora do círculo hermenêutico. A descrição de um exemplo deixa claro que o Hermeneus possibilita a construção de um ambiente que atua como um intermediário entre usuários e a informação desejada.

O pragmatismo, cabe ressaltar, é uma das características abordada pelo modelo proposto – o framework Hermeneus valida essa questão – ao se comparar com os outros três modelos de RBI. A ontologia viabilizou, principalmente, a implementação de conceitos tratados em *information seeking* e círculo hermenêutico além de aperfeiçoar os mecanismos requeridos por RI. Enfim, o modelo proposto juntamente com o framework Hermeneus explicitam concretamente características que sistemas de busca interativos devem possuir, enquanto os demais modelos definem características apenas conceitualmente.

No próximo capítulo, apresenta-se a conclusão, que discute de uma maneira geral o trabalho proposto, e os trabalhos futuros.

6 CONCLUSÕES E TRABALHOS FUTUROS

6.1 Conclusões

Descreve-se na presente pesquisa que questões relacionadas à incerteza e à ansiedade são parte de um processo de busca. Se um usuário não sabe exatamente o que precisa, como um modelo de RBI pode auxiliar em um processo de *information seeking*? Recorreu-se à Filosofia, especificamente ao círculo hermenêutico, visando buscar ajuda para a definição da proposta. No círculo hermenêutico os usuários desenvolvem as suas idéias enquanto navegam na informação e nos conceitos que a representam.

A solução que se apresenta aqui é um modelo de RBI. Além disso, concebeu-se, por meio da aplicação dos conceitos definidos pelo modelo, o framework Hermeneus. A definição do modelo e conseqüentemente do framework segue o conceito estabelecido por Kuhlthau (1991; 1993), que afirma que um sistema de busca deve ser um intermediário entre a necessidade de um usuário por informação e o acesso à solução para essa necessidade. A teoria de *information seeking* auxiliou a entender os requisitos de um usuário para a construção de um ambiente em que ele pode intervir e interagir mais ativamente. Buscar informação é uma atividade exploratória na qual os usuários aplicam conhecimento e intuição combinados com estratégias e ferramentas para encontrar a informação desejada (Capra Iii e Pérez-Quiñones, 2005). O modelo proposto refletido no Hermeneus permite que os usuários entrem em uma seqüência de interações que os auxiliam a entender o que eles realmente precisam buscar.

O modelo utiliza ontologia como tecnologia que fornece suporte para criação de um ambiente de busca mais interativo. A ontologia é utilizada no âmbito dos mecanismos requeridos por um sistema de RI. Na presente proposta, um sistema de busca deve viabilizar a construção desse ambiente interativo. Ao utilizar o Hermeneus um usuário pode navegar nos conceitos da ontologia e nas instâncias desses conceitos em forma de *vai e vem*. Interagindo assim, os usuários podem mudar a visualização do resultado por meio dos conceitos descritos na ontologia. Eles podem, também, interagir com as instâncias recuperadas de modo a usar o conteúdo para refinar ou iniciar novas buscas. Além disso, esses usuários também podem interagir com informação adicional extraída a partir de cada instância recuperada – para tal utiliza-se técnicas de inferência combinadas com regras previamente definidas.

O modelo e o framework consideram os requisitos descritos por Järvelin e Ingwersen (2004) e Ingwersen e Järvelin (2005b) que afirmam que uma atividade de busca deve ser fácil, rápida e intuitiva. No que se refere à facilidade, permite-se que usuários acessem o conhecimento do domínio por meio de consultas e interação direta com uma ontologia. Dessa forma, um usuário pode formular consultas usando termos que definem a sua condição inicial de conhecimento e ter intuição quando navega nos conceitos da ontologia. Com relação à velocidade, os usuários podem utilizar o conteúdo das instâncias recuperadas para refinar buscas e extrair informação adicional para cada instância recuperada. As vantagens de tais operações, principalmente no que se refere à utilização dinâmica do conteúdo recuperado, estão em acordo com Ingwersen e Järvelin (2005b). Os autores citam que “relevância é um fenômeno cognitivo, situacional e dinâmico”. Finalmente, a proposta tem como objetivo conceber um ambiente intuitivo de modo a permitir que os usuários utilizem os conceitos descritos na ontologia para a composição de consultas simples e semânticas, que utilizem o conteúdo de instâncias recuperadas para refinar buscas e que utilizem a ontologia para visualizar a informação resultante em diferentes perspectivas. Essas funcionalidades, de acordo com Card et al. (1999) e Albertoni et al. (2005), facilitam a cognição de usuários o que, portanto, auxilia o alcance de uma condição de solução mais rapidamente.

6.2 Trabalhos futuros

Como trabalhos futuros, entende-se que é possível apresentar a informação recuperada de formas diferentes, com vistas a tirar vantagem do conteúdo semântico descrito nas instâncias recuperadas. Mostrar os resultados recuperados graficamente pode enriquecer ainda mais a interatividade entre os usuários e um sistema de busca. Também visualiza-se a construção de um ambiente gráfico em que especialistas de domínio possam formular regras sem precisar entender uma linguagem específica. Esse ambiente também poderia ser utilizado pelos próprios usuários para que eles mesmos configurassem as suas regras.

Entende-se, também, que o caminho (*i.e.*, ações) percorrido pelos usuários pode ser armazenado de modo a auxiliar outros usuários quando estão buscando assuntos similares. Nesse caso, o sistema poderia “entender” o comportamento desses novos usuários e oferecer-lhes uma melhor assistência durante todo o processo de busca por meio de dicas sobre as suas buscas. Visualiza-se, ainda, a criação de um mecanismo que possa ser utilizado por usuários para definir as suas próprias ontologias dinamicamente e, dessa forma, configurar um sistema

de busca de acordo com a noção da realidade deles próprios sobre um domínio específico. Os usuários, nesse caso, poderiam criar um sistema de busca personalizado e compartilhá-lo com outros usuários.

Uma tecnologia que está ganhando cada vez mais corpo é a TV digital. Esse ambiente trás novos tipos de interatividade que precisam ser utilizados para adaptar aplicações de modo a torná-las aptas a funcionar e tirar vantagem dessa nova tecnologia. No presente trabalho descreveu-se 12 tipos de interatividade, apresentados na Tabela 2 (pg. 51), e contextualizados no âmbito do modelo proposto. Montez e Becker (2005), referenciando Lemos (1997), apresentam 8 tipos específicos de interatividade relacionada à TV digital que vão desde a simples troca de canais até o nível mais elevado, que permite até mesmo que o telespectador produza seus programas e os envie à emissora. Sugere-se, portanto, que o modelo proposto possa ser adaptado a essa nova tecnologia de modo a propiciar que os usuários possam usufruir de ferramentas mais inteligentes para fazer buscas nesse tipo de mídia.

No campo da filosofia, nota-se que o Hermeneus pode ser mais profundamente explorado, por exemplo, no processo de comunicação que ocorre entre o usuário e o sistema. Flores (1998), ao discutir um trabalho anterior (Winograd e Flores, 1987), descreve a importância do computador como um mecanismo de comunicação e não computação. A comunicação, contudo, não é apenas a transferência e processamento de informação, mas um mecanismo que influencia e altera o mundo, conforme define a teoria dos atos de fala (do inglês *Theory of Speech Acts*) (Austin, 1975; Searle, 1979). O ambiente semântico que o Hermeneus proporciona entrega mais do que a necessidade por informação. De que forma a comunicação entre o usuário e o sistema influencia a formação de novas identidades, induz a ação e caracteriza-se como um sistema de coordenação de compromissos, como declara Flores (1998)?

7 REFERÊNCIAS BIBLIOGRÁFICAS

- Albertoni, R., A. Bertoni, *et al.* Information Search: The Challenge of Integrating Information Visualization and Semantic Web. Sixteenth Int. Workshop on Database and Expert Systems Applications, p.529-533. 2005.
- Allan, J., J. Aslam, *et al.* Challenges in Information Retrieval and Language Modeling: Report of a Workshop Held at the Center for Intelligent Information Retrieval. ACM SIGIR Forum, v.37, n.1, p.31-47. 2002.
- Allan, J., B. Carterette, *et al.* When will Information Retrieval be "Good Enough?". Proceedings of the 28th annual ACM SIGIR, p.433-440. 2005.
- Austin, J. L. How to Do Things with Words: Harvard University Press. 1975
- Baader, F., I. Horrocks, *et al.* Description Logics as Ontology Languages for the Semantic Web. In: (Ed.). Mechanizing Mathematical Reasoning: Springer Berlin/Heidelberg, v.2605/2005, 2005. Description Logics as Ontology Languages for the Semantic Web
- Baeza-Yates, R. A. e B. A. Ribeiro-Neto. Modern Information Retrieval: ACM Press, Addison-Wesley. 1999. 513 p.
- Bartell, B. T., G. W. Cottrell, *et al.* Latent Semantic Indexing is an Optimal Special Case of Multidimensional Scaling. ACM SIGIR Conference, 1992. 161-167 p.
- Bates, M. J. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Review, v.13, n.5, p.407-431. 1989.
- Belkin, N. J. Anomalous States of Knowledge as a Basis for Information Retrieval. The Canadian Journal of Information Science, v.5, p.133-143. 1980.
- _____. Interaction with Texts: Information Retrieval as Information-Seeking Behavior. Information Retrieval '93: Von derModellierung zu Anwendung, 1993. 55-66 p.
- Belkin, N. J., C. Cool, *et al.* Cases, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems. Expert Systems with Applications, v.9, n.3, p.379-395. 1995.
- Belkin, N. J. e W. B. Croft. Information Filtering and Information Retrieval: Two Sides of the Same Coin? Communications of the ACM, v.35, n.12, p.29-38. 1992.
- Belkin, N. J., P. G. Marchetti, *et al.* BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval. Information Processing and Management, v.29, n.3, p.325-344. 1993.
- Belkin, N. J., R. N. Oddy, *et al.* ASK for Information Retrieval: Part I: Background and Theory. Journal of Documentation, v.38, n.2, p.61-71. 1982.

Belkin, N. J., T. Seeger, *et al.* Distributed Expert Problem Treatment as a Model for Information Systems Analysis and Design. Journal of Information Science, v.5, n.5, p.152-167. 1983.

Bernstein, R. J. Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis. Philadelphia, PA, USA: University of Pennsylvania Press. 1983

Browne, G. J. e M. B. Rogich. An Empirical Investigation of User Requirements Elicitation: Comparing the Effectiveness of Prompting Techniques. Journal of Management Information Systems, v.17, n.4, p.223-249. 2001.

Bruijn, J. D. e D. Fensel. Ontology Definitions. In: (Ed.). Encyclopedia of Library and Information Science: Taylor & Francis, 2005. Ontology Definitions

Capra Iii, R. G. e M. A. Pérez-Quñones. Using Web Search Engines to Find and Refind Information. IEEE Computer Society, v.38, n.10, p.36-42. 2005.

Card, S. K., J. D. Mackinlay, *et al.* Readings in Information Visualization: Using Vision to Think: Morgan Kaufmann. 1999

Cardoso, J. The Semantic Web Vision: Where are We? IEEE Intelligent Systems, v.22, n.5, p.84-88. 2007.

Carroll, J. J., I. Dickinson, *et al.* Jena: Implementing the Semantic Web Recommendations. 13th Int. WWW Conference. New York: ACM Press, 2004. 74-83 p.

Case, D. O. Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior: Elsevier. 2007

Chandrasekaran, B., J. R. Josephson, *et al.* What are Ontologies, and Why Do We Need Them? IEEE Intelligent Systems, v.14, n.1, p.20-26. 1999.

Chen, H., G. Shankaranarayanan, *et al.* A Machine Learning Approach to Inductive Query by Examples: An Experiment Using Relevance Feedback, ID3, Genetic Algorithms, and Simulated Annealing. Journal of the American Society for Information Science, v.49, n.8, p.693-705. 1998.

Corcho, O., M. Fernández-López, *et al.* Methodologies, Tools and Languages for Building Ontologies: Where is their Meeting Point? Data & Knowledge Engineering, v.46, n.1, p.41-64. 2003.

Corcho, O. e A. Gómez-Pérez. A Roadmap to Ontology Specification Languages. In: (Ed.). Knowledge Engineering and Knowledge Management. Methods, Models, and Tools: Springer Berlin / Heidelberg, 2000. A Roadmap to Ontology Specification Languages

- Decker, S., M. Erdmann, *et al.* Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. DS-8: Kluwer Academic Publisher, 1999. 351-369 p.
- Deerwester, S., S. T. Dumais, *et al.* Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science, v.41, n.6, p.391-407. 1990.
- Doyle, C. S. Information Literacy in an Information Society: A Concept for the Information Age. Syracuse, NY: Eric. 1984
- Eidenberger, H. e C. Breiteneder. VizIR: A Framework for Visual Information Retrieval. Journal of Visual Languages and Computing, v.14, p.443-469. 2003.
- Engelbart, D. C. Augmenting Human Intellect: A Conceptual Framework. Stanford Research Institute. 1962. (AFOSR-3233)
- Flores, F. Reflections since Understanding Computers and Cognition. Emerald Group Publishing Limited, v.11, n.4, p.351-372. 1998.
- Fluit, C., M. Sabou, *et al.* Ontology-Based Information Visualization. In: (Ed.). Visualising the Semantic Web: Springer Verlag, 2002. Ontology-Based Information Visualization
- Fonseca, F. The Double Role of Ontologies in Information Science. Journal of the American Society for Information Science and Technology, v.58, n.6, p.786-793. 2007.
- Gadamer, H.-G. Truth and Method (Original publication 1960, as Wahrheit und Methode): Crossroad. 1989
- García, E. e M.-Á. Sicilia. User Interface Tactics in Ontology-Based Information Seeking. PsychNology Journal, v.1, n.3, p.242-255. 2003.
- Geanellos, R. Exploring Ricoeur's Hermeneutic Theory of Interpretation as a Method of Analysing Research Texts. Nursing Inquiry, v.7, n.2, p.112-119. 2000.
- Geertz, C. From the Native's Point of View: On the Nature of Anthropological Understanding. In: P. Rabinow e W. M. Sullivan (Ed.). Interpretive Social Science: A Reader. Berkeley: University of California Press, 1979. From the Native's Point of View: On the Nature of Anthropological Understanding, p.225-241
- Golbreich, C. Combining Rule and Ontology Reasoners for the Semantic Web. RuleML. Hiroshima, Japan: Springer-Verlag, 2004. 155-169 p.
- Greengrass, E. Information Retrieval: A Survey. 2000. (TR-R52-008-001)
- Gruber, T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Formal Ontology in Conceptual Analysis and Knowledge Representation. 1993a.

_____. A Translation Approach to Portable Ontology Specification. Knowledge Acquisition. 1993b.

Grunig, J. Publics, Audience and Market Segments: Segmentation Principles for Campaigns. In: (Ed.). Information Campaigns: Balancing Social Values and Social Change. Beverly Hills, CA: Sage, 1989. Publics, Audience and Market Segments: Segmentation Principles for Campaigns, p.199-288

Guarino, N. Formal Ontology and Information Systems. Formal Ontology in Information Systems, Netherlands: IOS Press, p.3-15. 1998.

Guha, R. V., O. Lassila, *et al.* Enabling Inferencing. W3C Query Languages Workshop (QL'98). Cambridge 1998.

Hearst, M., A. Elliott, *et al.* Finding the Flow in Web Site Search. Communications of the ACM, The consumer side of search, v.45, n.9. 2002.

Hearst, M. e C. Karadi. Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results Using a Large Category Hierarchy. ACM SIGIR Forum, v.31, n.SI, p.246-255. 1997.

Heidegger, M. On the Way to Language. San Francisco: Harper & Row. 1982. 200 p.

Heidegger, M. e J. Stambaugh. Being and Time: A Translation of Sein und Zeit. Albany, NY: State University of New York Press. 1996. xix, 487 p. p. (SUNY series in contemporary continental philosophy)

Hendry, D. G. e D. J. Harper. An Architecture for Implementing Extensible Information-Seeking Environments. 19th ACM SIGIR. Switzerland: ACM Press, 1996. 94-100 p.

Hert, C. A. Understanding Information Retrieval Interactions: Theoretical and Practical Implications. London, England: Ablex Publishing Corporation. 1997

Huberman, B. A. e R. M. Lukose. Social Dilemmas and Internet Congestion. Science. American Association for the Advancement of Science, v.277, n.5325, p.535-537. 1997.

Hyvönen, E., E. Mäkelä, *et al.* MuseumFinland: Finnish Museums on the Semantic Web. Journal of Web Semantics, v.3, n.2, p.25. 2005.

Hyvönen, E., S. Saarela, *et al.* Ontogator: Combining View- and Ontology-Based Search with Semantic Browsing. XML Finland 2003, Open Standards, XML, and the Public Sector, 2003. p.

_____. Application of Ontology Techniques to View-Based Semantic Search and Browsing. The Semantic Web: Research and Applications (First European Semantic Web Symposium), 2004. p.

Ide, E. New Experiments in Relevance Feedback. In: (Ed.). In the Smart system: experiments in automatic document processing: Prentice Hall, 1971. New Experiments in Relevance Feedback, p.337-354

Ikehara, S., J. I. Murakami, *et al.* Vector Space Model Based on Semantic Attributes of Words. Pacific association for computational linguistics (PACLING). Kitakyushu, Japan, 2001. p.

Ingwersen, P. Information Retrieval Interaction. London: Taylor Graham. 1992

_____. Polyrepresentation of Information Needs and Semantic Entities: Elements of a Cognitive Theory for Information Retrieval Interaction. ACM SIGIR Conference, p.101-110. 1994.

_____. Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory. Journal of Documentation, v.52, n.1, p.3-50. 1996.

Ingwersen, P. e K. Järvelin. Information Retrieval in Context. ACM SIGIR Workshop, p.6-8. 2004.

_____. The Sense of Information: Understanding the Cognitive Conditional Information Concept in Relation to Information Acquisition. Lecture Notes in Computer Science, Springer, v.3507, p.7-19. 2005a.

_____. The Turn: Integration of Information Seeking and Retrieval in Context. Netherlands: Springer. 2005b (The Information Retrieval Series)

Järvelin, K. e P. Ingwersen. Information Seeking Research Needs Extension Towards Tasks and Technology. Information Research: An International Electronic Journal, v.10, n.1. 2004.

Järvelin, K. e T. D. Wilson. On Conceptual Models for Information Seeking and Retrieval Research. Information Research, v.9, n.1. 2003.

Johnson, J. D. Cancer-Related Information Seeking. Cresskill: Hampton Press. 1997

Jurafsky, D. e J. H. Martin. Speech and Language Processing: Prentice Hall. 2000

Kekäläinen, J. e K. Järvelin. Evaluating Information Retrieval Systems Under the Challenges of Interaction and Multidimensional Dynamic Relevance. 4th Int. conf. on conceptions of library and information science: emerging frameworks and methods. Seattle WA, 2002. p.

Kelly, D. e N. J. Belkin. A User Modeling System for Personalized Interaction and Tailored Retrieval in Interactive IR. Annual Conference of the American Society for Information Science and Technology. Philadelphia, PA, 2002. 316-325 p.

Kingsley, P. Poimandres: The Etymology of the Name and the Origins of the Hermetica. Journal of the Warburg and Courtauld Institutes, v.56, p.1-24. 1993.

Klein, H. K. e M. D. Myers. A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Quarterly, v.23, n.1, p.67-94. 1999.

Kobayashi, M. e K. Takeda. Information Retrieval on the Web. ACM Computing Surveys (CSUR), v.32, n.2, p.144-173. 2000.

Korfhage, R. R. Information Storage and Retrieval. New York: Wiley Computer Publishing. 1997

Kowalski, G. Information Retrieval Systems: Theory and Implementation: Kluwer Academic Publishers. 1997 (Multimedia Information Retrieval: Content-Based Information Retrieval from Large Text and Audio Databases)

Krogh, G. V. e J. Roos. Organizational Epistemology. New York, NY: St. Martin's Press. 1995

Kuhlthau, C. C. Inside the Search Process: Information Seeking from the User's Perspective. Journal of the American Society for Information Science, v.42, n.5, p.361-371. 1991.

_____. Seeking Meaning: A Process Approach to Library and Information Services. USA: Ablex Publishing Corporation. 1993

_____. Towards Collaboration between Information Seeking and Information Retrieval. Information Research, v.10, n.2. 2005.

Lai, J. e B. Soh. Similarity Score for Information Filtering Thresholds. ISCIT 2004. IEEE International Symposium on Communications and Information Technology, 2004. 216-221 p.

Lai, L. F. A Knowledge Engineering Approach to Knowledge Management. Information Sciences, v.177, p.4072-4094. 2007.

Leckie, G. J., K. E. Pettigrew, *et al.* Modeling the Information Seeking of Professionals: A General Model Derived from Research on Engineers, Health Care Professionals, and Lawyers. Library Quarterly, v.66, p.161-193. 1996.

Lee, J. H. Properties of Extended Boolean Models in Information Retrieval. ACM SIGIR Forum, 1994. 182-190 p.

Lei, Y., V. Uren, *et al.* SemSearch: A Search Engine for the Semantic Web. 3rd European Semantic Web Conference (ESWC). Montenegro, 2006. p.

Lemos, A. L. M. Anjos Interativos e Retribalização do Mundo: Sobre Interatividade e Interfaces Digitais. UFBA. 1997. (s.n.)

Liddy, E. D. Enhanced Text Retrieval Using Natural Language Processing. Bulletin of the American Society for Information Science, v.24, n.4. 1998.

- Lin, J. e D. Demner-Fushman. The Role of Knowledge in Conceptual Retrieval: A Study in the Domain of Clinical Medicine. Proceedings of the 29th ACM SIGIR, p.99-106. 2006.
- López, M. F., G.-P. Asunción, *et al.* Building a Chemical Ontology Using Methontology and the Ontology Design Environment. Intelligent Systems and their Applications IEEE, v.14, n.1, p.37-46. 1999.
- Maarek, Y. S., D. M. Berry, *et al.* An Information Retrieval Approach for Automatically Constructing Software Libraries. IEEE transactions on Software Engineering, v.17, n.8, p.800-813. 1991.
- Maedche, A. e S. Staab. Discovering Conceptual Relations from Text. ECAI-2000 13th European Conference on Artificial Intelligence: IOS Press, 2000. 321-325 p.
- Mäkelä, E., E. Hyvönen, *et al.* Ontogator - A Semantic View-Based Search Engine Service for Web Applications. 5th International Semantic Web Conference (ISWC 2006), 2006. p.
- Manning, C. D., P. Raghavan, *et al.* Introduction to Information Retrieval: Cambridge University Press. 2008
- Marchionini, G. Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia. Journal of the American Society for Information Science, v.40, n.1, p.54-66. 1989.
- _____. Information Seeking in Electronic Environments: Cambridge University Press. 1997
- Marchionini, G. e A. Komlodi. Design of Interfaces for Information Seeking. Annual Review of Information Science and Technology, v.33, p.89-130. 1998.
- Mcbride, B. Jena: A Semantic Web Toolkit. Internet Computing, IEEE, v.6, n.6, p.55-59. 2002.
- Mcguinness, D. L. Ontology-Enhanced Search for Primary Care Medical Literature. Int. Medical Informatics Association. Phoenix, Arizona, 1999. p.
- Mcguinness, D. L. e F. V. Harmelen. OWL Web Ontology Language Overview. W3C World Wide Web Consortium (<http://www.w3.org/TR/owl-features/>). 2004.
- Mikheev, A., M. Moens, *et al.* Named Entity Recognition without Gazetteers. 9th Int. Conf. of the European Chapter of the Association for Computational Linguistics, 1999. p.
- Milton, N., D. Clarke, *et al.* Knowledge Engineering and Psychology: Towards a Closer Relationship. Human-Computer Studies, v.64, p.1214-1229. 2006.
- Montez, C. e V. Becker. TV Digital Interativa. Florianópolis, SC: Editora da UFSC. 2005

- Müller, H.-M., E. E. Kenny, *et al.* Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature. PLoS Biol, v.2, n.11. 2004.
- Murdock, J. W., P. Pinheiro Da Silva, *et al.* Encoding Extraction as Inferences. AAAI Spring Symposium on Metacognition on Computation: AAAI Press, Stanford University, 2005. 92-97 p.
- Noy, N. F. e D. L. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology. Knowledge Systems Laboratory. 2001.
- Oddy, R. N. Information Retrieval through Man-Machine-Dialogue. Journal of Documentation, v.33, n.1, p.1-14. 1977.
- Pearl, J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference: Morgan Kaufmann. 1988
- Pinto, H. S. e J. P. Martins. Ontologies: How Can They Be Built? Knowledge and Information Systems, v.6, p.441-464. 2004.
- Prud'hommeaux, E. e A. Seaborne. SPARQL Query Language for RDF. W3C - World Wide Web Consortium (<http://www.w3.org/TR/rdf-sparql-query/>). 2007.
- Reiterer, H., G. Müller, *et al.* A Visual Information Seeking System for Web Search. Mensch & Computer conference. Bad Honnef, Germany, 2001. p.
- Ricoeur, P. The Task of Hermeneutics. In: (Ed.). Heidegger and Modern Philosophy: Critical Essays. London: Yale University Press, 1978. The Task of Hermeneutics
- Rijsbergen, C. J. V. Information Retrieval. Glasgow, Scotland, UK: University of Glasgow. 1979
- Roberts, R. M. Serendipity: Accidental Discoveries in Science: Willey. 1989
- Robins, D. Interactive Information Retrieval: Context and Basic Notions. Informing Science, v.3, n.2. 2000.
- Rocchio, J. J. Relevance Feedback in Information Retrieval. In: (Ed.). The Smart system: experiments in automatic document processing: Prentice Hall, 1971. Relevance Feedback in Information Retrieval, p.313-323
- Rouse, W. B. e S. H. Rouse. Human Information Seeking and Design of Information Systems. Information Processing & Management, v.20, n.1-2, p.129-138. 1984.
- Salton, G. Automatic Information Organization and Retrieval. New York: McGraw-Hill. 1968

_____. Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Boston: Addison-Wesley Longman Publishing Co. 1989

Salton, G. e C. Buckley. Improving Retrieval Performance by Relevance Feedback. Journal of the American Society for Information Science, v.41, n.4, p.288-297. 1990.

Salton, G., A. Wong, *et al.* A Vector Space Model for Automatic Indexing. Communications of the ACM, v.18, n.11, p.613-620. 1975.

Saracevic, T. Modeling Interaction in Information Retrieval (IR): A Review and Proposal. American Society of Information Science, v.33, p.3-9. 1996.

_____. The Stratified Model of Information Retrieval Interactions: Extension and Applications. American Society for Information Science, 1997. 313-327 p.

Savage-Knepshield, P. A. e N. J. Belkin. Interaction in Information Retrieval: Trends Over Time. Journal of the American Society for Information Science, v.50, n.12, p.1067-1082. 1999.

Savolainen, R. Everyday Life Information Seeking. In: (Ed.). Theories of Information Behaviour. Medford: Information Today Inc., 2005. Everyday Life Information Seeking

Schreiber, G., H. Akkermans, *et al.* Knowledge Engineering and Management: The CommonKADS Methodology: Massachusetts Institute of Technology. 2000

Scowen, R. S. Extended BNF - A Generic Base Standard. Software Engineering Standards Symposium. Brighton, UK: IEEE Computer Society Press, 1993. p.

Searle, J. R. Expression and Meaning: Studies in the Theory of Speech Acts: Cambridge University Press. 1979

Shadbolt, N., E. Motta, *et al.* Constructing Knowledge-Based Systems. Software IEEE, v.10, n.6, p.34-38. 1993.

Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. IEEE Symposium. Proceedings on Visual Languages, p.336-343. 1996.

Smith, B. Ontology. In: L. Floridi (Ed.). Blackwell Guide to the Philosophy of Computing and Information. Oxford: Blackwell, 2003. Ontology, p.155-166

_____. Beyond Concepts: Ontology as Reality Representation. International Conference on Formal Ontology and Information Systems (FOIS). 2004.

Smith, B. e C. Welty. Ontology: Towards a New Synthesis. Formal Ontology in Information Systems, ACM press. 2001.

Sonnenwald, D. H., B. M. Wildemuth, *et al.* A Research Method to Investigate Information Seeking Using the Concept of Information Horizons: An Example from a Study of Lower Socio-Economic Students' Information Seeking Behaviour. The New Review of Information Behaviour Research, v.2, n.November, p.65-86. 2001.

Sowa, J. F. Architectures for Intelligent Systems. Systems Journal, Artificial Intelligence, v.41, n.2, p.331. 2002.

Spence, R. Information Visualization: Addison Wesley. 2000. 206 p.

Spink, A. e M. Park. Multitasking and Co-ordinating Framework for Human Information Behavior. In: (Ed.). New Directions in Human Information Behavior: Springer, 2006. Multitasking and Co-ordinating Framework for Human Information Behavior, p.137-154

Spink, A. e T. Saracevic. Interaction in Information Retrieval: Selection and Effectiveness of Search Terms. Journal of the American Society for Information Science, v.48, n.8, p.741-761. 1998.

Spink, A., T. D. Wilson, *et al.* Information-Seeking and Mediated Searching. Part 1. Theoretical Framework and Research Design. Journal of the American Society for Information Science, v.53, n.9, p.695-703. 2002.

Spink, A., D. Wolfram, *et al.* Searching the Web: The Public and their Queries. Journal of the American Society for Information Science, v.53, n.2, p.226-234. 2001.

Staab, S. Knowledge Representation with Ontologies: The Present and Future. IEEE Intelligent Systems, v.10, n.1, p.72-81. 2004.

Staab, S., R. Studer, *et al.* Knowledge Processes and Ontologies. Intelligent Systems, IEEE, v.16, n.1, p.26-34. 2001.

Studer, R., V. R. Benjamins, *et al.* Knowledge Engineering: Principles and Methods. Data Knowledge Engineering, v.25, n.1-2, p.161-197. 1998.

Studer, R., S. Decker, *et al.* Situation and Perspective of Knowledge Engineering. Knowledge Engineering and Agent Technology. IOS Press. 2000.

Sure, Y. e R. Studer. On-To-Knowledge Methodology. Institute AIFB, University of Karlsruhe. 2002. (EU-IST Project IST-1999-10132 - OTK/2002/D18/v1.0)

Tang, M.-C. Browsing and Searching in a Faceted Information Space: A Naturalistic Study of PubMed Users' Interaction with a Display Tool. Journal of the American Society for Information Science, v.58, n.13, p.1998-2006. 2007.

Taylor, C. Interpretation and the Sciences of Man. Review of Metaphysics, v.23, p.3-51. 1971.

Taylor, R. S. Question-Negotiation and Information Seeking in Libraries. College & Research Libraries, v.29, p.178-194. 1968.

Teevan, J., C. Alvarado, *et al.* The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. Proceedings of the SIGCHI conference on Human factors in computing systems, p.415-422. 2004.

Tissen, A. A Case-Based Architecture for a Dialogue Manager for Information-Seeking Processes. Acm SIGIR Conference, p.152-161. 1991.

Toms, E. G. Information Interaction: Providing a Framework for Information Architecture. Journal of the American Society for Information Science, v.53, n.10, p.855-862. 2002.

Uschold, M. e M. Gruninger. Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, v.11, n.2. 1996.

Vakkari, P. Task-Based Information Search. Annual Review of Information Science and Technology, v.37, p.413-464. 2003.

Vallet, D., M. Fernández, *et al.* An Ontology-Based Information Retrieval Model. In: (Ed.). The Semantic Web: Research and Applications: Springer, 2005. An Ontology-Based Information Retrieval Model, p.455-470. (Lecture Notes in Computer Science)

Varela, F., E. Thompson, *et al.* Embodied Mind: Cognitive Science and Human Experience. Cambridge, MA: MIT Press. 1991

Wand, Y. e R. Weber. Reflection: Ontology in Information Systems. Journal of Database Management, v.15, n.2. 2004.

White, R. W., J. M. Jose, *et al.* An Implicit Feedback Approach for Interactive Information Retrieval. Information Processing and Management, v.43, p.166-190. 2006.

Wilson, T. D. Information Needs and Uses: Fifty Years of Progress? Journal of Documentation Review, p.15-51. 1994.

_____. Information Behaviour: An Interdisciplinary Perspective. Information Processing and Management, v.33, n.4, p.551-572. 1997.

_____. Models in Information Behavior Research. The Journal of Documentation, v.55, n.3, p.249-270. 1999.

_____. Human Information Behavior. Information Science, v.3, n.2, p.49-56. 2000.

Winograd, T. e F. Flores. Understanding Computers and Cognition: A New Foundation for Design: Addison Wesley. 1987

Witten, I. H., A. Moffat, *et al.* Managing Gigabytes: Compressing and Indexing Documents and Images. San Francisco, CA: Morgan Kaufmann. 1999

Zadeh, L. A. Information and Control. Fuzzy Sets, v.8, p.338-353. 1965.

_____. A Fuzzy-Set-Theoretic Approach to the Compositionality of Meaning: Propositions, Dispositions and Canonical Forms. Journal of Semantics, v.2, n.1, p.253-272. 1983.