

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO INTERDISCIPLINAR DE NOVAS TECNOLOGIAS NA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA NA EDUCAÇÃO

Luís Carlos Costa Fonseca

**UM ASSISTENTE PESSOAL DE
APRENDIZAGEM CONTINUADA NA WEB**

Porto Alegre
2009

Luís Carlos Costa Fonseca

**UM ASSISTENTE PESSOAL DE
APRENDIZAGEM CONTINUADA NA WEB**

Tese apresentada ao Programa de Pós-Graduação em Informática na Educação do Centro Interdisciplinar de Novas Tecnologias na Educação da Universidade Federal do Rio Grande do Sul, como requisito para obtenção do título de Doutor em Informática na Educação.

Orientadora:
Profa. Dra. Rosa Maria Vicari

Coorientador:
Prof. Dr. Crediné Silva de Menezes

Porto Alegre
2009

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitor: Prof. José Carlos Alexandre Netto
Vice-Reitor: Prof. Rui Vicente Oppermann
Pró-Reitor de Pós-Graduação: Prof. Aldo Bolten Lucion
Diretor do CINTED: Profa. Rosa Maria Vicari
Coordenador do PPGIE: Prof. José Valdeni de Lima

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

F676a Fonseca, Luís Carlos Costa

Um assistente pessoal de aprendizagem continuada na web / Luís Carlos Costa Fonseca; orientador: Rosa Maria Vicari; coorientador: Crediné Silva de Menezes. – Porto Alegre, 2009.

126 f.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Centro Interdisciplinar de Novas Tecnologias na Educação. Programa de Pós-Graduação em Informática na Educação, 2009, Porto Alegre, BR-RS.

1. Ambiente de aprendizagem. 2. Educação continuada. 3. Internet. 4. Recuperação da informação. 5. Pedagogia de projetos. 6. Tecnologias da informação e comunicação. I. Vicari, Rosa Maria. II. Menezes, Crédine Silva de. III. Título.

CDU – **371.694:681.3:371.334**

Luís Carlos Costa Fonseca

**UM ASSISTENTE PESSOAL DE
APRENDIZAGEM CONTINUADA NA WEB**

Tese apresentada ao Programa de Pós-Graduação em Informática na Educação do Centro Interdisciplinar de Novas Tecnologias na Educação da Universidade Federal do Rio Grande do Sul, como requisito para obtenção do título de Doutor em Informática na Educação.

Aprovada em 15 dez. 2009.

Prof. Dra. Rosa Maria Vicari – Orientadora

Prof. Dr. Crediné Silva de Menezes – UFRGS

Prof. Dr. Leandro Krug Wives – UFRGS

Prof. Dra. Magda Bercht – UFRGS

Prof. Dra. Cecília Dias Flores – UFSC

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus por sempre me acompanhar e ajudar em todos os momentos.

Agradeço a meus pais e familiares, que sempre me apoiaram e deram forças para continuar nesta batalha do doutorado.

À professora Rosa Maria Vicari pelas sábias orientações e principalmente por sua paciência e incentivo.

Ao professor Crediné Silva Menezes pelo longo tempo dedicado à minha co-orientação.

Aos amigos Sofiane Labidi, Othon Bastos Filho, Nadie, Cleber Pereira, Cenidalva, Christiane Abas, Cristiane Tavares, Fabio, Maria do Carmo, Antonio, Maximira, Jaqueline Picceti e Luciane que sempre estiveram presentes nos momentos mais difíceis e que de uma forma ou de outra contribuíram para a concretização deste trabalho.

RESUMO

O conhecimento de uma pessoa pode se tornar obsoleto ou incompleto rapidamente, visto que, na chamada sociedade da informação e do conhecimento, a todo instante surgem novas teorias e tecnologias que são rapidamente publicadas e disseminadas pelos novos meios de comunicação e informação, notadamente a Internet. Visando contribuir para a atenuação desse problema, o presente trabalho propõe o desenvolvimento de um ambiente baseado na Web que dê apoio à aprendizagem continuada de um indivíduo dentro de algum campo de interesse, facilitando, assim, a atualização de seus conhecimentos. Do ponto de vista tecnológico, o trabalho se apóia na recuperação de informações em contexto e na tecnologia de agentes de software; do ponto de vista pedagógico, inspira-se na pedagogia de projetos de aprendizagem.

Palavras-chave: **Ambiente de aprendizagem. Educação continuada. Internet. Recuperação de informação. Pedagogia de Projeto. Tecnologia de informação e comunicação.**

ABSTRACT

The knowledge of a person can become rapidly obsolete or incomplete, since the called society of information and knowledge, all the time there are new theories and technologies that are quickly published and disseminated by new media and information, notably the Internet. Aiming to reduce this problem, this work proposes the development of a Web-based environment which provides support for lifelong learning of an individual within a certain field of interest, thus facilitating the upgrade of their knowledge. From a technological standpoint, the work relies on Information Retrieval in Context and the Technology of Software Agents, the pedagogical point of view, inspired by the Pedagogy of Learning Projects.

Keywords: **Learning environment. Lifelong learning. Internet. Information retrieval. Learning projects pedagogy. Information technology and communication.**

SUMÁRIO

1	Introdução	10
1.1	Motivação	13
1.2	Objetivos.....	16
1.3	Questão de Pesquisa	17
1.4	Conteúdo deste texto.....	17
2	Problema	19
3	Aspectos pedagógicos	21
3.1	Educação continuada	21
3.2	Information Literacy	24
3.3	Autonomia	26
3.3.1	Heutagogia	29
3.4	Construtivismo	31
3.5	Projetos de Aprendizagem	33
3.6	Ambientes de aprendizagem	39
4	Aspectos tecnológicos	41
4.1	Agentes de Software	42
4.1.1	Conceito de agente.....	42
4.1.2	Características de agentes	43
4.1.3	Algoritmos Genéticos.....	45
4.2	Recuperação de Informações.....	47
4.2.1	A Visão Lógica de um Documento.....	52
4.2.2	Técnicas de Recuperação de Informações	54
4.2.2.1	Modelo Booleano.....	56
4.2.2.2	Modelo Vetorial.....	59
4.2.2.3	LSI	63
4.2.2.4	Probabilístico	66
4.2.2.5	CNG.....	67
4.2.3	Recuperação de Informação baseada em Contexto.....	71
4.3	Ontologias	76
5	A Proposta de um Ambiente Mediador.....	81

5.1	Metodologia	83
5.1.1	Tipologia da Pesquisa	83
5.1.2	Técnicas de Coleta de Dados.....	85
5.1.3	Universo e Amostra	86
5.1.4	Da Amostra Intencional	86
5.1.5	Dados e Fontes de Informação Utilizadas	87
5.1.6	Algumas considerações sobre projetos de aprendizagem	87
5.2	O Protótipo ARPA.....	88
5.3	Selecionando documentos	96
5.4	Adaptação de um modelo de RI considerando contexto	101
5.5	Agentes do sistema	105
5.6	Testes.....	111
6	Considerações Finais.....	114
	Referências.....	118

Lista de Figuras

Figura 1: Linha do tempo da quantidade de documentos disponíveis (FIELDEN, 2002)	11
Figura 2: Crescimento anual da quantidade de domínios registrados (ISC, 2009) ...	12
Figura 3: Problema de mediação entre o usuário e a Internet.....	16
Figura 4: Interação do usuário com o sistema de recuperação através de diferentes tarefas (BAEZA-YATES, 1999: 4)	48
Figura 5: Visão Lógica de um Documento: do Texto Completo ao Conjunto de Termos Indexados (BAEZA-YATES, 1999: 6).....	53
Figura 6: Componentes conjuntivos da consulta [$q = ka \wedge (kb \vee \sim kc)$].....	57
Figura 7: Documentos Indexados com os pesos de cada documento.	59
Figura 8: O cosseno do ângulo Θ é adotado como $sim_{dj, q}$	60
Figura 9: CNG Exemplo indicando conexões entre os documentos da Tabela 5 e os termos da Tabela 6.	69
Figura 10: Framework cognitivo geral da busca e recuperação de informações.....	73
Figura 11: Modelo Aninhado das Dimensões do Contexto em RI (INGWERSEN, 2005b: 281)	75
Figura 12: Ontologia dos projetos de aprendizagem.....	80
Figura 13: Modelo de um Gestor Pessoal de Informações (JENNINGS & WOOLDRIDGE, 1996)	82
Figura 14: Casos de Uso do ARPA	90
Figura 15: Lista de Itens de Conhecimento.....	92
Figura 16: Cadastro de Certezas e Dúvidas.....	93
Figura 17: Ciclo de Interação do Sistema	93
Figura 18: Listagem de Rótulos.....	95
Figura 19: Tarefas do usuário no ambiente ARPA	96
Figura 20: Evolução dos Itens de Conhecimento do Aprendiz	97
Figura 21: Processo decisório de seleção de documentos	98
Figura 22: Modelo de seleção de documentos.....	100
Figura 23: DED da base de dados do sistema.	108
Figura 24: Interfaces de conexão com motores de busca.....	110
Figura 25: Arquitetura do Sistema de Agentes.....	111

1 Introdução

Por muitos anos, acreditou-se que a simples conclusão de um curso de nível superior era formação necessária e suficiente para o desempenho competente das atividades profissionais. Com a evolução das recentes tecnologias computacionais, observou-se que as pessoas não poderiam limitar-se aos conhecimentos obtidos em cadeiras escolares, faculdades e cursos extracurriculares. À medida que o fluxo de acesso a informações se tornou extremamente dinâmico, determinou-se um novo ritmo de trabalho e impuseram-se novas exigências às pessoas para manterem um desempenho profissional, um convívio social e cultural mínimos. Em outras palavras, nos dias atuais o conhecimento de uma pessoa pode se tornar obsoleto ou incompleto rapidamente, visto que, na “sociedade da informação e do conhecimento”, a todo instante surgem novas teorias e tecnologias que são rapidamente publicadas e disseminadas pelos novos meios de comunicação e informação, notadamente a Internet.

Para iniciar as discussões sobre este trabalho, imagine o seguinte cenário: um indivíduo que quer se manter atualizado sobre os assuntos de seu interesse tem acesso a uma grande biblioteca. Imagine algo na proporção da Biblioteca de Alexandria¹, ou maior, e que tenha um conteúdo dinâmico, ou seja, a cada dia chegam mais e mais documentos.

Agora imagine a dificuldade de descobrir alguma informação relevante dentro desse emaranhado de livros. Obviamente seria necessária alguma forma de organização/catalogação bibliográfica. Uma solução para o problema seria a criação de um índice remissivo para se procurar por palavras-chave o conteúdo de cada um dos livros, mas mesmo assim seria uma forma bastante trabalhosa de se manter atualizado; pessoas que fizessem uso desse método demorariam muito tempo até encontrar algo relevante.

1 A Biblioteca de Alexandria foi uma das maiores bibliotecas do mundo e se localizava na cidade de Alexandria. Considera-se que tenha sido fundada no início do século 3 a.C., durante o reinado de Ptolomeu II do Egito, após seu pai ter construído o Templo das Musas (Museum). É atribuída a Demétrio de Falero sua organização inicial. Uma nova biblioteca foi inaugurada em 2003 próxima ao sítio da antiga que no ano de 646 foi queimada por completo pelos árabes.

De outro modo, a Internet (rede mundial de computadores), mais precisamente a *Web* (W3C, 2007), pode ser vista como um exemplo desse cenário, só que em proporções muito maiores. Todo esse material disponível na Web, aliado ao seu fácil acesso, se constitui valiosa fonte de informação. Observa-se na Figura 1 a evolução do volume de documentos disponíveis no decorrer da história. A variedade desses documentos que se pode encontrar *on-line* é das mais diferentes naturezas e fontes: artigos científicos, notícias de jornais, leis, documentos históricos, entrevistas, livros, entre tantos outros.

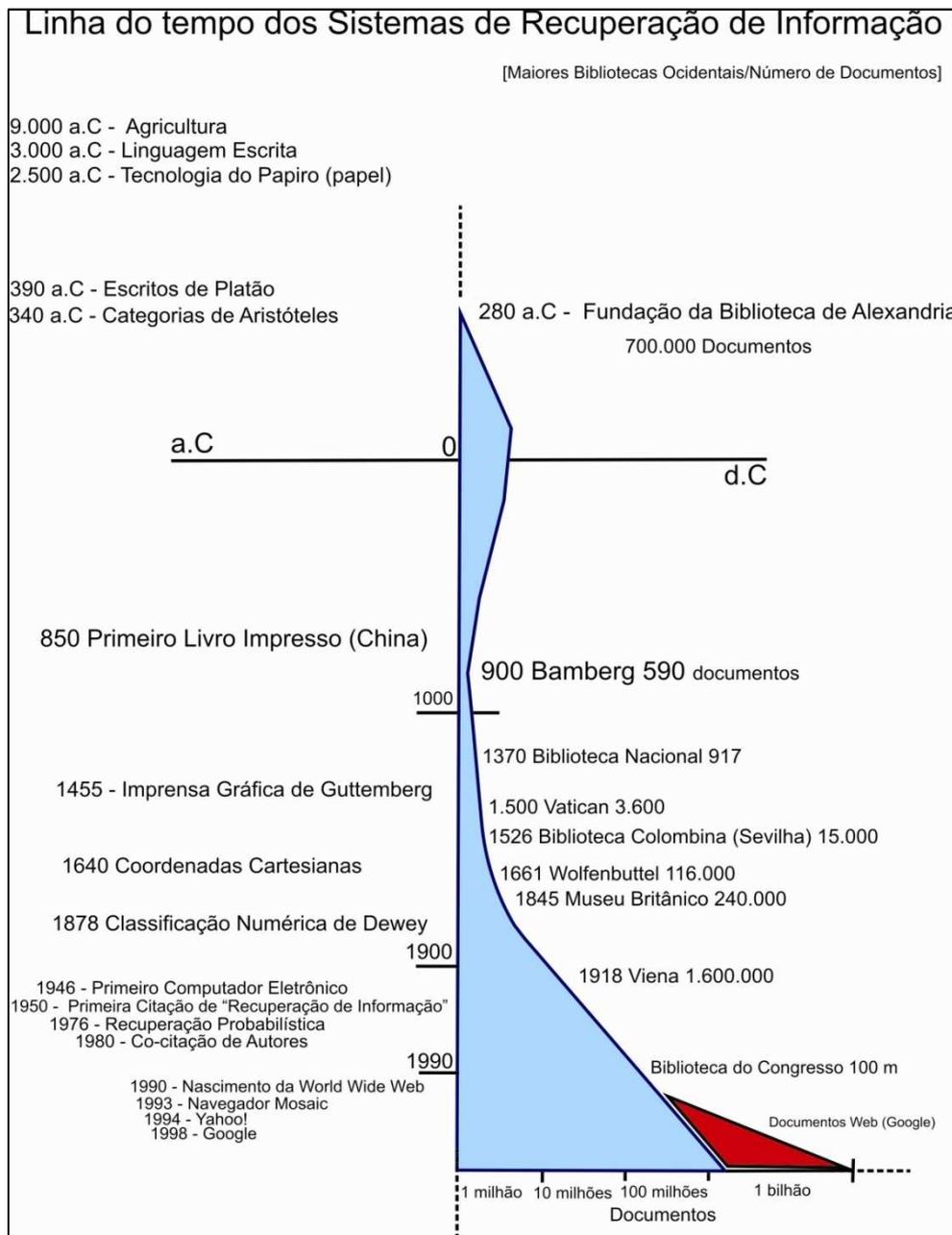


Figura 1: Linha do tempo da quantidade de documentos disponíveis (FIELDEN, 2002)

As informações na Internet estão distribuídas em *sites* hospedados em servidores localizados nas mais diversas partes do planeta, tornando-se praticamente impossível conhecer o que existe e localizar uma informação de interesse sem o auxílio de um mecanismo computacional.

Outro dado estatístico interessante de ser verificado é o crescimento do volume de domínios registrados na Web anualmente (cf. Figura 2). Um domínio é um nome que representa de forma mais amigável um endereço da Internet. Eles podem ser vistos como provedores de informação, ou seja, locais em que as pessoas podem publicar documentos.

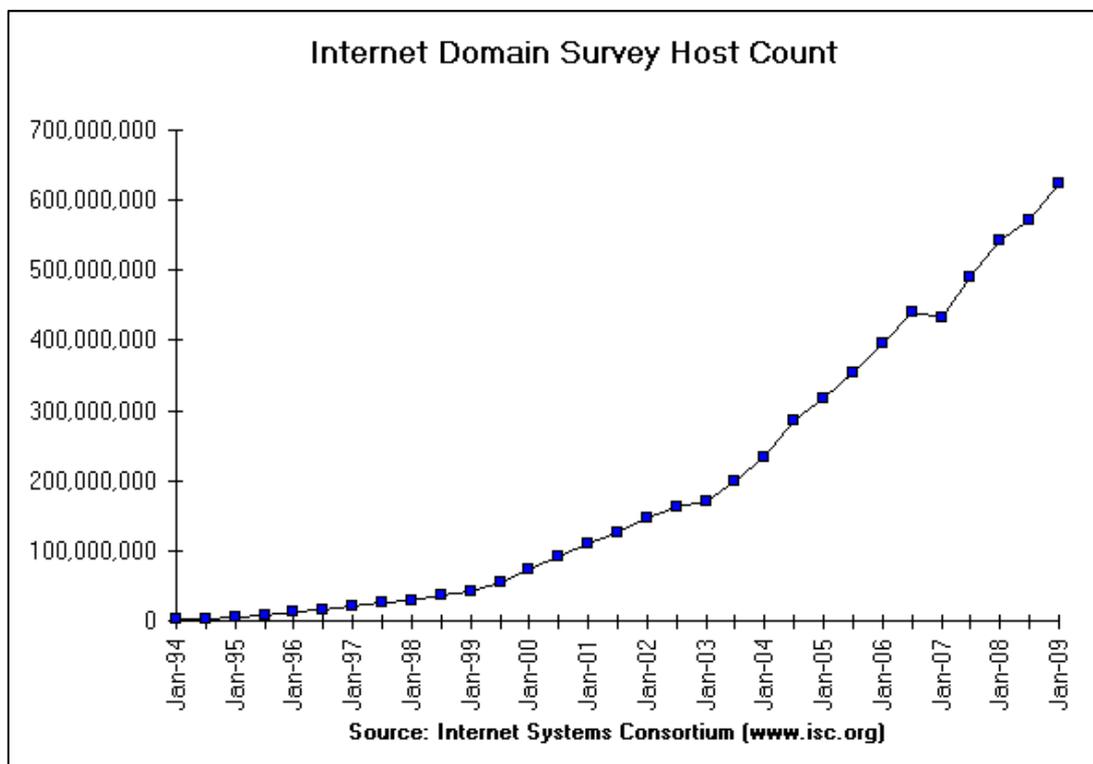


Figura 2: Crescimento anual da quantidade de domínios registrados (ISC, 2009)

Um dos problemas associados ao uso da Internet, e que persiste até os dias atuais, é chamado na literatura de “problema de acesso à informação”. Consiste na dificuldade de estimar a quantidade de documentos na Web. Isso se deve à natureza dinâmica e pode ser mais bem visualizado quando se tenta imaginar a quantidade de novos documentos que surgem e somem da Web diariamente. Tomando-se por base a quantidade de domínios existente hoje e supondo que em

cada um é publicado um novo documento de uma página por dia, então teremos diariamente um conjunto de documentos de cerca de quinhentos milhões de páginas. Para ilustrar melhor o problema: isso equivale a uma produção diária de 2,5 milhões de livros de 200 páginas. Uma biblioteca de uma universidade de grande porte tem quantos documentos?

O problema de acesso à informação foi diminuído a partir da aplicação dos resultados de pesquisas em Recuperação de Informações (RI) nos repositórios da Web. Essa evolução foi gradativa; inicialmente existiam os diretórios, que ofereciam uma árvore de indexação na qual era/é possível navegar em busca de documentos específicos. Depois surgiram os *sites* de busca, que oferecem uma linguagem com a qual o usuário pode descrever em palavras-chave o que pretende recuperar.

Entretanto, o problema de acesso à informação ainda não foi solucionado completamente. Mesmo com os mais novos mecanismos de indexação e recuperação, os motores de busca ainda não conseguiram englobar todo o conteúdo da Web. Além disso, outro fator que dificulta é o fato de que as máquinas de busca tratam as requisições do usuário de modo isolado, sendo que os resultados para determinada consulta são idênticos, independentemente do usuário ou do contexto no qual este se encontra (LAWRENCE, 2000).

1.1 Motivação

Apesar de existirem tantas fontes de informação, sabe-se que o indivíduo, na sua busca diária por conhecimento, se direciona a subconjuntos de documentos que lhe parecem ter maior relação com sua área de atuação profissional. Em outras palavras, nem tudo o que circula pela Web é interessante a todos os usuários.

Há algumas décadas, antes do surgimento das novas tecnologias de informação e telecomunicação e da Internet, o conjunto de documentos suficientes e necessários para alguém se manter atualizado em seu campo profissional era pequeno. Apenas algumas publicações de referência, sendo que a maioria dos

trabalhos relevantes e as novas tendências tecnológicas pertinentes eram facilmente encontradas. Entretanto, o conhecimento dos trabalhos desenvolvidos por outras comunidades de interesse era muitas vezes tolhido por conta da impossibilidade proveniente da distância geográfica. Essa realidade foi abruptamente revertida com o surgimento da Internet. Atualmente, os profissionais de todo o mundo têm acesso a produtos, serviços, tecnologias, técnicas, métodos, desenvolvidos em todas as partes do mundo; em contrapartida, a concorrência se tornou mais agressiva, obrigando as pessoas a manter seus conhecimentos atualizados.

Retomando a modelagem do cenário inicialmente proposto, imagine agora que um indivíduo poderia se manter atualizado continuamente apenas fazendo uso da Internet. Entretanto, levando-se em conta o volume de informações contidas na rede, seria desejável que fossem devolvidas ao usuário as possibilidades de ler apenas o que lhe interessa (é necessário e suficiente) e de ter como encontrar isso.

Um grande problema associado a esse desejo está relacionado à natureza das ferramentas de recuperação de informações, que evidentemente não foram planejadas nem consideraram o desejo do usuário gostaria de fazer uso da Internet para dar continuidade a sua formação.

Muitas outras dificuldades poderiam ser elencadas; são consideradas mais relevantes, entretanto, as seguintes:

- mesmo depois de muitas interações com os motores de busca, muitos dos sistemas disponíveis não guardam quaisquer informações sobre as necessidades dos usuários;
- muitas vezes, os usuários não têm tempo ou mesmo não dominam a linguagem utilizada para recuperar documentos em motores de busca² na Web;
- mesmo sabendo buscar na Web, muitas vezes o usuário tem dificuldade de escolher as palavras-chave mais adequadas.

Para que houvesse a recuperação satisfatória das informações, as pessoas deveriam ter acesso a ferramentas computacionais que dessem um suporte adequado a esse processo. Essas ferramentas poderiam tomar como base as

² Termo técnico dado aos sistemas dos sites de busca de informação na Internet (Google, Altavista, etc.).

anotações e os vestígios deixados pelo usuário durante o uso. Este trabalho, portanto, parte exatamente dessa curiosidade de definir *como organizar informações e extrair esses vestígios* a partir das “pegadas” deixadas pelos usuários nas suas interações com um sistema de RI. A seguir, apresentamos os objetivos deste trabalho.

1.2 Objetivos

Podemos observar que um usuário da Internet tem à sua disposição um universo de informações, representado hoje por uma quantidade quase incalculável de documentos. Para atenuar, desejamos construir mediadores que possibilitem, de forma ágil, o acesso à informação desejada.

Com esse trabalho, pretende-se desenvolver um ambiente mediador que sirva de apoio à aprendizagem continuada, baseado na recuperação de documentos disponíveis na Internet. Em linhas gerais, o problema abordado neste trabalho pode ser ilustrado na Figura 3.

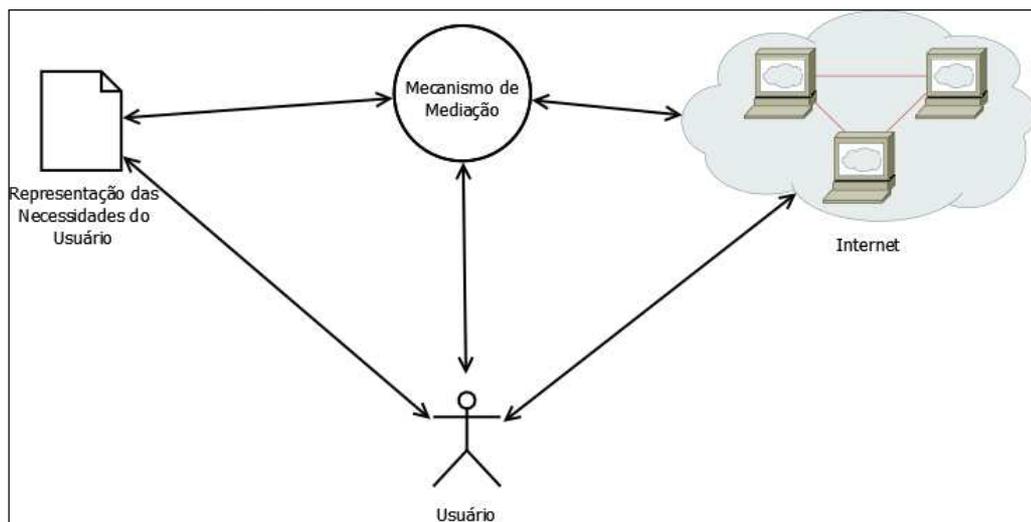


Figura 3: Problema de mediação entre o usuário e a Internet

Para tanto, estabelece-se os seguintes objetivos específicos:

1. desenvolver mecanismos de meta-busca para Recuperar Informações da Internet baseadas no contexto do usuário;
2. investigar o uso da tecnologia de agentes de software para da suporte ao mecanismo de meta-busca e extração de termos para as buscas tomando por base os Projetos de Aprendizagem como representação das necessidades do usuário;

3. estender e adaptar o modelo de registro dos Projetos de Aprendizagem de forma a que eles possuam informações mais sistematizadas das necessidades dos aprendizes, e que estas possam ser processadas pelos agentes do sistema, e que se possa estudar como esses agentes podem extrair dados sobre o contexto, a partir das ações do usuário.

1.3 Questão de Pesquisa

A questão principal deste trabalho é:

- que características de um ambiente mediador de acesso à informação podem contribuir para facilitar ao indivíduo manter-se atualizado com o uso dos documentos disponíveis na internet?

Além dela, podemos elencar as seguintes questões secundárias:

- como desenvolver um ambiente mediador tendo por base a tecnologia de recuperação de informação a partir de projetos de aprendizagem?
- quais critérios devem ser levados em consideração para a apresentação de textos ao usuário e como sugerir os documentos nos pontos corretos do projeto de aprendizagem?

1.4 Conteúdo deste texto

Este trabalho se organiza com a seguinte estrutura: no primeiro capítulo é feita uma introdução apresentando as idéias iniciais, objetivos, motivação e

questionamentos que motivaram o desenvolvimento desta pesquisa. No Capítulo 2, são definidos os problemas associados ao desenvolvimento do trabalho. A seguir, o Capítulo 3 traz a metodologia utilizada até o presente momento e futuramente até o término da pesquisa. Os Capítulos 4 e 5 tratam respectivamente das fundamentações teórica pedagógica e tecnológica. O Capítulo 6 apresenta uma proposta de ambiente mediador como resposta às problemáticas apresentadas no Capítulo 2. E, finalmente, no Capítulo 7 são tecidos os comentários finais e apresentadas as conclusões.

2 Problema

Até o presente instante foi definido que seria desejável desenvolver um ambiente que fosse mais amigável para a atualização continuada de indivíduos que desejam se manter atualizados em alguma área de interesse. Entretanto, ainda não foram detalhados, neste trabalho, os requisitos que um sistema dessa natureza deveria atender, ou seja, os problemas associados ao desenvolvimento e à implementação de uma ferramenta com essas características.

Após a realização de um levantamento sobre trabalhos de objetivos similares, foram verificados alguns trabalhos com características complementares, e que serviram de suporte para lapidar esta proposta.

Um primeiro trabalho que chama bastante atenção é o xFIND (GÜTL, 1999). Os autores o caracterizam como um sistema de descoberta de conhecimento adaptado para um Sistema de Treinamento Baseado na Web chamado GENTLE (DIETINGER & MAURER, 1998). A ideia é que esses dois sistemas, trabalhando em conjunto, funcionem como um ambiente de aprendizagem sob demanda (aprendizagem continuada).

O sistema na forma descrita no trabalho tem características interessantes e desejáveis ao desenvolvimento desta proposta, principalmente em relação à sua arquitetura. Entretanto, o mesmo não preenche a lacuna relacionada à independência do aprendiz no que se relaciona a ele ter autonomia sobre os rumos de seu aprendizado. O próprio trabalho faz menção a essa questão quando relata sua experiência enquanto trabalhando com Sistemas de Treinamento Baseados na Web, onde se afirma: “[...] aprender via Web é muito mais do que ler lições navegando por cursos multimídia preparados e responder exercícios [...]”.

De forma similar, Ramquist (RAMQUIST, 2007) cita outros elementos como comunicação, colaboração e acesso a bibliotecas dinâmicas e estáticas; além disso, ele menciona como ingredientes importantes o uso de motores de busca e a exploração de sites na Internet. Várias formas de aprendizagem também são citadas por (CAPPSS, 1997): por observação, por pesquisa e investigação, por ação,

individualmente, face a face e em grupos, por experimentos, por avaliação e reflexão.

Outro trabalho interessante é o pacote de ferramentas CmapTools (CMAPTOOLS, 2007), que facilitam a construção de mapas conceituais, utilizados para a organização e representação visual das relações existentes entre os conceitos, contribuindo, assim, para uma melhor compreensão destes. O *kit* de ferramentas ainda conta com um mecanismo de busca que faz consultas à Internet a partir dos conceitos colocados nos mapas. Entretanto, essas consultas são simples e pontuais, baseadas apenas no mapa atual, não considerando variáveis de contexto mais complexas.

Considerando as ferramentas pesquisadas, é interessante ressaltar que apesar de constituírem iniciativas interessantes, nenhuma se presta completamente à tarefa de dar suporte continuado aos estudos de um indivíduo, portanto nos propomos nesta tese a investigar tecnologias que possam contribuir na tarefa de construir um ambiente mediador para estudos continuados e autônomos.

3 Aspectos pedagógicos

Para Candy (2002), aprender é uma das características humanas universais. Certamente, a curiosidade e a capacidade de aprender estão entre as características que definem o que significa *ser humano*. Em muitas centenas de gerações, os povos aprenderam com sua própria experiência e, em menor escala, com a observação dos outros. Entretanto, com o passar do tempo, a língua tornou-se gradualmente mais complexa e sofisticada, e surgiu a possibilidade de codificar o que tinha sido aprendido por gerações anteriores e em localidades distantes, com a finalidade de passar informações sobre fenômenos percebidos.

Na maioria das culturas antigas que confiaram na transmissão oral do conhecimento, os povos desenvolveram maiores capacidades para escutar e para recordar. Entretanto, comparado com a maioria de povos que vivem hoje em países ocidentais avançados, a quantidade e a complexidade da informação a ser tratada por um homem adulto era limitada e com mudanças relativamente lentas. Não havia a necessidade de uma formação tão dinâmica e contínua como agora. A partir desse contexto, a seguir apresentaremos alguns conceitos associados a essa necessidade dos indivíduos das sociedades contemporâneas.

3.1 Educação continuada

O conceito de educação continuada (*LifeLong Learning – LLL*), como uma estratégia educacional, surgiu há cerca de três décadas através dos esforços de organizações como OECD (Organização para Cooperação Econômica e Desenvolvimento), UNESCO (Organização das Nações Unidas para a Educação, a Ciência e a Cultura) e o Conselho da Europa (COE). Sendo que esse último entende o LLL como sendo “[...] toda atividade de aprendizagem empreendida durante a vida, com o objetivo de melhorar os conhecimentos, as habilidades e as competências,

dentro de uma perspectiva pessoal, cívica, social e/ou relacionada ao emprego”.

Mas por que a educação continuada é importante? Segundo o relatório da OECD (OECD, 2004), existe um grande número de aspectos socio-econômicos que tem apontado para a abordagem de educação continuada. O ritmo crescente da globalização e das mudanças tecnológicas, as mudanças na natureza do trabalho e no mercado de trabalho, e o envelhecimento das populações estão entre as forças que enfatizam a necessidade da continuidade de atualização das habilidades de trabalho durante toda a vida.

Sendo assim, a educação continuada também pode ser vista como uma “segunda chance” de atualização de habilidades básicas e de oferecer oportunidades de aprendizagem em níveis mais elevados.

Uma classificação bem útil é proposta em Candy (2002), em que são definidas categorias de aprendizagem:

- *aprendizagem formal*: ocorre em ambientes de educação formal, como escolas, faculdades e universidades. Distingue-se pela presença de um currículo formal, um professor (tutor ou outro facilitador) e, usualmente, por seus dispositivos, como grades curriculares, palestras ou aulas, listas de leitura e provas. Existem regras distintas que guiam o aprendiz nessas circunstâncias, especialmente as que indicam como ele ou ela devam se comportar e o que se constitui uma aprendizagem valorosa;
- *aprendizagem não formal*: Aprendizagem que é organizada e planejada por alguém, mas não leva a nenhuma qualificação ou reconhecimento. Isso inclui a maioria dos cursos de curta duração empreendidos para o enriquecimento pessoal ou avanço profissional/vocacional. Grande parte dessa aprendizagem é planejada e empreendida pelo próprio aprendiz de forma autônoma. Desde o advento da Internet, e mais recentemente da WWW (*Word Wide Web*), uma grande parte dessa forma de aprendizagem vem ocorrendo de forma mediada por tecnologia. Vem se tornando necessário, portanto, que as pessoas possam se conectar e terem fluência tecnológica para que possam acessar e

usar a tecnologia de forma natural;

- *aprendizagem incidental*: tão importante quanto as aprendizagens formal e não formal. Esta representa apenas uma fração de minuto da aprendizagem que ocorre durante a vida de uma pessoa. De fato, é uma evidência que sugere que, tanto em termos de duração e de impacto, a aprendizagem informal ultrapassa com folga os outros dois tipos. Isso ocorre quase que continuamente e coincide com a vida. Não requer uma preparação particular, embora aqueles que são mais atenciosos e/ou tenham um círculo mais largo de experiências provavelmente tenham uma aprendizagem incidental mais extensiva e mais diversa. A aprendizagem informal não ocorre exclusivamente em ambientes informais; ela ocorre também em ambientes formais, como escolas, faculdades, universidades e centros de aprendizado. Não é, entretanto, planejada nem controlada explicitamente como parte do currículo ou do programa do estudo.

Com relação ao público-alvo, Michael Kendall (2003) considera que a educação continuada pode ser subdividida conforme as suas perspectivas: pessoal, cívica, social e/ou laboral.

Em uma perspectiva relacionada a emprego, os estudantes da educação continuada podem ser criadores e aplicadores de conhecimento. Em ambas as modalidades os estudantes aprendem no trabalho em “grupos de aprendizagem”. Para o primeiro grupo, a aprendizagem é natural; para o segundo, isso talvez tenha de ser organizado quando a aprendizagem não fizer parte do trabalho de forma integral.

Para as perspectivas sociais e cívicas, os estudantes são pessoas que aprendem em “comunidades de aprendizagem”. Essas comunidades podem ser comunidades locais, de interesses ou de passatempo (*hobbies*), incluindo comunidades regionais.

Já em uma perspectiva pessoal, os estudantes podem ser “aprendizes solitários”. Mesmo assim, a “aprendizagem em times” e a “aprendizagem em comunidades”, ou seja, as outras perspectivas, não deixam de ser importantes, pois

ainda fazem parte da sua vida social e das atividades relacionadas ao trabalho. Aprendizes solitários podem, de forma notável, apoderar-se das TICs (Tecnologias de Informação e Comunicação) como forma de se engajarem como membros de times de aprendizagem autossustentáveis ou de comunidades de aprendizagem.

Em todos os casos, os aprendizes necessitam de certas habilidades e competências para obter progresso de forma satisfatória em sua aprendizagem. Essas habilidades formam parte do conceito de *Information Literacy* (CANDY, 2002), que será tratado a seguir.

3.2 *Information Literacy*

Praticamente em paralelo ao conceito de Educação Continuada surgiu o conceito de *Information Literacy* (por volta de 1974). Para Candy (2002), são dois conceitos que se misturam de modo insolúvel.

Alguns autores trazem uma nova luz para a compreensão dos conceitos de alfabetização, escolarização e letramento. Este último pode ter surgido para preencher uma lacuna, decorrente de atualizações no conceito de alfabetização. Soares (2003a, 2003b, 2003c), assim como Tfouni (1995), têm discutido letramento em perspectivas que muito contribuíram para esclarecer não apenas o conceito em si, mas as relações estabelecidas entre alfabetização, escolarização e letramento.

O conceito de letramento é relativamente novo em português e, segundo Magda Soares (2003a: 15), só foi incluído no Dicionário Houaiss da Língua Portuguesa em 2001. O alfabetizado, “[...] nomeia aquele que apenas aprendeu a ler e a escrever, não aquele que adquiriu o estado ou a condição de quem se apropriou da leitura e da escrita, incorporando as práticas sociais que as demandam [...]” (SOARES, 2003: 19).

O termo *letramento* corresponderia, em inglês, a *literacy* – *literacia*, em português europeu –, mas vai além da aprendizagem básica da leitura e da escrita, ou seja, do domínio da técnica, compreendendo a aquisição “[...] de habilidades,

conhecimentos e atitudes necessários ao uso efetivo e competente da leitura e da escrita nas práticas sociais que envolvem a língua escrita [...]” (SOARES, 2003a: 89).

No que se refere a *literacy*, segundo Lyman (1979), pode ser definida como “[...] a habilidade de compreender matérias, ler criticamente, usar materiais complexos e aprender por si mesmo [...]”. Entretanto, o termo não apresenta um significado preciso. Novas ênfases têm emergido, acomodando novos significados.

Na última década, uma ampla variedade de *literacies* tem sido proposta (DOYLE, 1994), (RAFFERTY, 1999), incluindo a cultural aspectos compartimentalizados de *literacy*, termos exclusivos, como cultura, tecnológica, acadêmica, marginal, etc. A *information literacy*, ao contrário, é um termo inclusivo, englobando todas as demais. Praticamente inexplorada no Brasil, a expressão ainda não possui tradução para a língua portuguesa. Porém, algumas expressões possíveis seriam: alfabetização informacional, letramento, literacia, fluência informacional, competência em informação. A partir de agora, neste trabalho, consideraremos fluência informacional como o significado de *Information Literacy*.

O relatório final da Associação Americana de Biblioteconomia (AAB) de 1989 da Comissão Presidencial sobre Fluência Informacional define que, para ter fluência informacional, “uma pessoa deve reconhecer quando a informação é necessária, possuir a habilidade de encontrá-la de forma eficaz, além de avaliar e usar essas informações” (ALA, 1989). Essa definição de fluência informacional não é limitada a nenhuma forma de informação; mais adiante, o relatório adiciona, “[...] quer a informação selecionada venha de um computador, de um livro, de uma agência de governo, de um filme, ou de qualquer outra fonte possível”.

Assim como a Educação Continuada, o conceito de fluência informacional tem gerado um grande volume literário, com muitas definições ligeiramente diferentes. Entretanto, da mesma forma que Webber e Johnson (2000) definem em seu trabalho intitulado “*Conceptions of Information Literacy: New Perspectives and Implications*”, muitas dessas definições contam os seguintes elementos:

- procura eficaz de informação;
- escolha de fontes de informações;
- avaliação e escolha da informação;

- conforto em usar uma grande variedade de meios para obter mais vantagens;
- manipulação consciente crítica e confiável da informação; e
- eficácia na transmissão de informação para os outros.

Muitas dessas definições, incluindo a da AAB, tendem a retratar a fluência informacional como uma realização genérica, uma maneira de aplicar certas habilidades como recuperação de informações ou do pensamento crítico, desprezando o assunto (disciplina) ou o meio em que está contido ou em que é feito o saber. Considera simplesmente a implicação de um aspecto genérico à fluência informacional; uma potencialidade de distinguir os recursos úteis dos inúteis, fontes de informação confiáveis das não confiáveis e sensibilidade a necessidades de informações tolas. Em alguns casos, isso é provavelmente verdadeiro, porém existem elementos fortemente dependentes do contexto, onde a fluência informacional é influenciada pelo domínio, e pela forma da informação, se é digital ou não.

O sistema proposto neste trabalho tem alguns critérios para os usuários poderem fazer uso do dele. Um deles é que eles devem possuir fluência informacional, conceito este que depende de certa autonomia por parte do utilizador do sistema. A seguir, o conceito de autonomia será mais bem detalhado.

3.3 Autonomia

Pretende-se apresentar, aqui, uma ferramenta que dê suporte a educação continuada de forma autônoma. Mas o que se entende por autonomia? Como este trabalho se apoia em idéias construtivistas, para responder a essa questão será feita, inicialmente, uma verificação sobre como Piaget entendia o juízo moral e, por fim, a autonomia.

Em 1932, portanto no início de sua carreira, Jean Piaget publicou *Le jugement moral chez l'enfant*. O eixo principal do livro trata das idéias do autor sobre o juízo

moral, deixando implícita sua concepção sobre a relação entre afetividade e cognição.

Piaget afirmou que:

“Toda moral consiste num sistema de regras e a essência de toda moralidade deve ser procurada no respeito que o indivíduo adquire por estas regras.” (PIAGET, 1932: 2).

Piaget segue sua pesquisa escolhendo um campo muito peculiar da atividade humana e que facilita consideravelmente o entendimento de suas idéias: o jogo de regras.

Como as pesquisas de Piaget foram geralmente baseadas em observações do comportamento de crianças, ele escolheu para esse estudo dois jogos infantis: o de bola de gude, para meninos, e a amarelinha, para meninas. Para cada sujeito, pesquisou a prática e a consciência das regras, solicitando inicialmente que a criança lhe ensinasse a jogar e posteriormente jogasse com ele. Em seguida, procedeu a questionamentos sobre: a origem das regras, quem eram seus autores e se poderiam ser modificadas, etc.

Para Piaget, a evolução da prática e da consciência das regras pode ser dividida em três etapas desconexas de padrões etários dos indivíduos, mas que podem ser observados mais comumente em determinadas idades.

A primeira delas é a anomia. Entre zero e cinco anos, a criança não manifesta noção de regras coletivas ou consciência moral. O que predomina nesse momento é a dependência da regularidade motriz, fundamentalmente os gestos de repetição imitativa. Interessa-se, por exemplo, por bolas de gude, mas para satisfazer seus interesses motores ou suas fantasias simbólicas, e não tanto para participar de uma atividade coletiva.

A segunda etapa é a da heteronomia. Variando geralmente dos cinco aos oito anos de idade, é caracterizada pelo realismo moral, em que a regra tem uma validade absoluta por vir de Deus, dos pais, dos professores, de amigos e de outros adultos, merecendo respeito absoluto. Apesar de um interesse em participar de atividades coletivas e regradas, percebe-se aí uma ausência de poder de crítica às normas, considerando-se os níveis de consciência delas.

Entretanto, apesar desse respeito quase religioso às regras, a criança,

quando joga, mostra-se bastante “liberal”, no que tange à aplicação das regras: frequentemente introduz, sem qualquer consulta prévia ao seu adversário, alguma variante que lhe possibilite ter melhor desempenho, e não acha estranho afirmar no final de uma partida que “todo mundo ganhou”.

Observa-se, então, que a criança heterônoma não assimilou o sentido da existência de regras. Por não as conceber como necessárias, não as aplica à risca; e justamente por não as conceber, atribui-lhes uma origem totalmente estranha à atividade e aos membros do grupo, quase que divina.

A terceira e última etapa é a autonomia. Verificada normalmente após os treze anos de idade, quando há uma maior conscientização e um maior detalhamento da regra. É o momento de rever, debater, discutir e analisar as regras morais, modificando-as segundo os limites do consenso na vida em grupo, entre iguais. Piaget escreveu que:

“[...] esta noção de autonomia é empregada aqui sem uma conotação filosófica. Ela designa somente a possibilidade do sujeito de elaborar, ao menos em parte, suas próprias normas [...]” (PIAGET, 1932: 534).

A autonomia se contrapõe à heteronomia, colocando-se como uma concepção adulta do jogo. Primeiramente, as crianças jogam seguindo as regras com esmero. Em segundo lugar, o respeito pelas regras é compreendido como decorrente de acordos mútuos entre os participantes do jogo, em que, cada um concebe a si próprio como possível “legislador”, ou seja, criador de novas regras que serão submetidas à apreciação e à aceitação dos outros.

Para Correia (2003), o processo que compreende o percurso do egocentrismo ao altruísmo, da heteronomia à cooperação intragrupal, iniciada por volta de 7-8 anos de vida, conduz a criança a secundarizar o papel da coação social em benefício da compreensão dos conceitos de justiça, reciprocidade e cooperação. É nesse estágio que ela começa a entender que as coisas nem sempre são como devem ser. Aí a pergunta: se, de antemão, já está decidido como as coisas devem ser, onde há possibilidade da autonomia? Aí o entendimento de que *autonomia é a capacidade de o sujeito elaborar, ao menos em parte, suas próprias normas*.

Um indivíduo autônomo age livremente de acordo com um plano próprio, de forma análoga a um governo independente, que administra seu território e estabelece suas políticas. Em um contexto educacional, é possível que alguém

queira estudar de forma independente, escolhendo seus horários de trabalho e temas de estudo.

Essa ultima concepção é muito importante para este trabalho, pois, para alguém que se preste a fazer uso do sistema proposto nesta tese, as capacidades de elaboração e respeito às regras criadas são cruciais. Nesse contexto, a seguir serão tratadas questões práticas sobre o papel da autonomia na Educação a Distancia, suas vantagens e soluções.

3.3.1 Heutagogia

Em EaD, um conceito largamente utilizado atualmente é o da heutagogia, que compreende a aprendizagem autodirecionada, em que o aluno é o gestor e programador de seu processo de aprendizagem – lembrando muito as idéias de Paulo Freire (FREIRE, 1996) em sua preocupação de trazer a realidade do aprendiz para seu contexto de aprendizagem.

A proposta central da heutagogia na EaD é que as novas tecnologias de comunicação e informação facilitem ao aluno essa autonomia na condução de sua aprendizagem, contando, quando possível, com o suporte de professores, colegas, ou especialistas (tutores) no assunto tratado.

Carmem Maia (MAIA, 2007: 85) afirma que o ensino a distancia exige, portanto um aprendiz autônomo e independente, ou seja, mais responsável pelo processo de aprendizagem e disposto à autoaprendizagem. Mais ainda, com a alteração da cultura do ensino para a cultura da aprendizagem, o estudo passou a ser auto-administrado e automonitorado por um aprendiz autônomo.

Nesse novo ambiente, a autora ainda cita novos termos para falar de aprendizagem, tais como autorresponsável, autoplanejada, auto-organizada, independente e autorregulada, além de não linear e não sequencial, em que os aprendizes trilham seus próprios caminhos e alcançam seus próprios objetivos.

Cabe aqui, então, destacar uma característica crucial para esta tese no que

se refere a esse novo modelo de aprendiz, que é a mudança em sua postura de estudo. Agora, os indivíduos terão como atividades principais buscar, encontrar, selecionar e aplicar, e não mais receber e memorizar.

Para Otto Peters (2004: 170), essa nova forma de aprendizagem coloca os alunos no primeiro plano, e não os professores. Nela, acredita-se que eles são capazes de planejar, organizar, controlar e avaliar eles mesmos o seu trabalho.

Afirma esse autor (2004: 80), que também trata desse tema como um modelo do aluno autônomo, que esse não é realmente um modelo de educação a distancia, porque não tem a vantagem de ensinar um número muito grande de alunos. Um professor conseguirá orientar e apoiar apenas de vinte a trinta alunos.

Entretanto, existem outras formas e demandas de aprendizagem, de acordo com Litwin (2001). O autodidata é o estudante que seleciona os conteúdos e não conta com uma proposta pedagógica e didática para o estudo. Na modalidade a distância, por exemplo, apesar de se permitir uma organização autônoma dos estudantes em relação à escolha de espaço e tempo para o estudo, isso não significa que não haja uma proposta didática, uma seleção de conteúdo, orientações de prosseguimento dos estudos e propostas de atividades.

Trabalhando como autodidata, o próprio aprendiz seria o guia de sua aprendizagem. Nesse caso ele mesmo poderia determinar sua proposta didática, fazer a seleção de conteúdos e escolhendo, assim, os rumos de sua formação e caracterizando uma heutagogia “mais independente”. Certamente os caminhos tomados nessa aprendizagem poderiam ser fortemente influenciados por fatores ou interesses externos, como assuntos relacionados ao trabalho, assuntos associados a comunidades de interesses e/ou, assuntos de interesse próprio do aprendiz, remontando às perspectivas da Educação Continuada vistas anteriormente.

Uma pergunta que pode ser formulada com a leitura desse texto é: como organizar esse estudo seguindo uma metodologia de trabalho? A resposta talvez já tenha sido encontrada em alguns projetos no Reino Unido, em que é comum a elaboração de um “contrato de aprendizagem” que envolve não apenas professor e aluno, mas muitas vezes a própria instituição de ensino e, em alguns casos, como em projetos de *work-based learning*, a empresa em que o aluno trabalha.

Este trabalho, entretanto, presta-se a discutir a criação de uma ferramenta

que dê suporte à educação continuada de forma autônoma. Como visto na seção anterior, um indivíduo autônomo age livremente, de acordo com um plano próprio, não cabendo aqui algo como um contrato, sendo mais um acerto entre as partes envolvidas no projeto pedagógico. Sendo assim, outra característica importante para este trabalho é a possibilidade de desenvolver um mecanismo que facilite a organização e o planejamento do aprendiz na forma de um plano pessoal de aprendizagem. Essa idéia ficará mais esclarecida na seção que trata especificamente dos projetos de aprendizagem. Antes disso serão apresentados alguns conceitos básicos para este trabalho, como o de construtivismo.

3.4 Construtivismo

Como não poderia deixar de ser, em um trabalho que envolve a implementação de um sistema com finalidades pedagógicas, torna-se sumamente importante expor as teorias que estão por trás dos modelos pedagógicos utilizados. É o que se propõe a seguir. E, na sequência, uma definição mais completa do que é um projeto de aprendizagem.

Muito se fala sobre a obra de Piaget, entretanto, depois de variadas análises de seu trabalho, surgiram algumas ideias equivocadas sobre ele e seu legado. Por exemplo, que ele teria sido um pedagogo ou um psicólogo do desenvolvimento e que teria dedicado sua vida ao estudo da criança. Na verdade, essa leitura do famoso pensador diz muito pouco da sua atividade de da sua vida.

Segundo Ramozzi-Chiarotino (1988), Piaget, com o intuito de formular uma abordagem da construção do conhecimento em bases biológicas, habituou-se a observar a Natureza, a descobrir as suas regularidades e a tentar formular leis que regem a vida de animais e vegetais. Após anos de pesquisas, e cultivando o plano de escrever uma epistemologia do conhecimento, resolveu observar o comportamento humano, da mesma forma que fez com plantas e animais. Isso o conduziu ao estudo das ações das crianças, surgindo então sua teoria a respeito do

nascimento da inteligência.

A principal questão de investigação de Piaget era responder “Como é possível alcançar o conhecimento?”. Esta remete a outras questões, em especial a que interroga: “Conhecimento de quê?”. É, entretanto, simples a resposta a esse último questionamento: “conhecimento do mundo em que vivemos, do meio que nos circunda”, aponta-nos Piaget. Entretanto, para ele, o entendimento da palavra “meio” é abrangente, abarcando tudo: natureza, objetos construídos pelo homem, idéias, valores, relações humanas, em suma, História e Cultura.

E o que significa “conhecimento”? Para Piaget, esse conceito é claramente associado a organizar, estruturar e explicar, porém, a partir do vivido (do experimentado). Sendo assim, ele entendia que conhecer não é somente explicar ou viver, mas, sim, vivenciar experiências – agir sobre o objeto do conhecimento, para que esse objeto seja imerso em um sistema de relações.

Segundo Piaget (1929-1979), o fenômeno da aprendizagem depende de um processo construtivo que ocorre por meio de construções e reconstruções dos sistemas de significação e dos sistemas lógicos de cada indivíduo. Para que isso ocorra, é fundamental que esse indivíduo possa interagir com os objetos (no sentido piagetiano: natureza, mundo físico, cultura, artes, ciências, linguagens...), com outros sujeitos (sociedade, instituições...) e, agora, com tecnologias. Entendemos que a aprendizagem ocorre de forma mais eficiente, quando o sujeito se envolve com a resolução de problemas que lhe dizem respeito, quando busca entendimento das coisas que lhe importem que povoem seu universo.

O professor Fernando Beker (1993) define o construtivismo da seguinte forma:

[...] a idéia de que nada, a rigor, está pronto, acabado, e de que, especificamente, o conhecimento não é dado, em nenhuma instância, como algo terminado. Ele se constitui pela interação do indivíduo com o meio físico e social, com o simbolismo humano, com o mundo das relações sociais; e se constitui por força de sua ação e não por qualquer dotação prévia, na bagagem hereditária ou no meio, de tal modo que podemos afirmar que antes da ação não há psiquismo nem consciência e, muito menos, pensamento (BEKER, 1993: 2).

Mais adiante, no mesmo trabalho, ele afirma:

[...] O conhecimento é uma construção. O sujeito age, espontaneamente – isto é, independentemente do ensino mas não independentemente dos estímulos sociais –, com os esquemas ou estruturas que já tem, sobre o

meio físico ou social. Retira (abstração) deste meio o que é do seu interesse. Em seguida, reconstrói (reflexão) o que já tem, por força dos elementos novos que acaba de abstrair. Temos, então, a síntese dinâmica da ação e da abstração, do fazer e do compreender, da teoria e da prática. É dessas sínteses que emerge o elemento novo, sínteses que o apriorismo e o empirismo são incapazes de processar porque só valorizam um dos pólos da relação. Na visão construtivista, sujeito e meio têm toda a importância que se pode imaginar, mas essa importância é radicalmente relativa (BEKER, 1993: 4).

Um dos discípulos de Piaget, Papert (1994), sugere que uma boa estratégia para a ocorrência da aprendizagem é a construção de artefatos. O construcionismo compartilha a visão da aprendizagem do construtivismo sobre “a construção de estruturas cognitivas” pela internalização das ações... Ele adiciona, porém, a idéia de que isso ocorre especialmente em um contexto no qual o aprendiz esteja engajado continuamente na construção de uma entidade pública, quer seja um castelo de areia na praia ou uma teoria do universo.

A seguir apresentamos a pedagogia que embasa este trabalho e que é fundamentada na construção de artefatos (os “projetos de aprendizagem”), conforme a sugestão de Papert.

3.5 *Projetos de Aprendizagem*

Desde meados dos anos 80 até hoje, desenvolvem-se projetos envolvendo Informática na Educação no Laboratório de Estudos Cognitivos (LEC) da Universidade Federal do Rio Grande do Sul (UFRGS), em parceria com o Governo do Brasil pelo Ministério da Educação (MEC), conforme se pode ler no documento “Aprendizes do Futuro – As Inovações Começaram” (FAGUNDES, 1998), que apresenta o relato das experiências com “Informática Educativa” desenvolvidas por aquele grupo de pesquisa.

Um dos pontos centrais e também mais interessantes das pesquisas desenvolvidas no LEC está relacionado ao uso da atividade construtiva de elaborar e desenvolver projetos como uma metodologia de aprendizagem. Para eles:

A atividade de fazer projetos é simbólica, intencional e natural do ser humano. Por meio dela, o homem busca a solução de problemas e desenvolve um processo de construção de conhecimento, que tem gerado

tanto as artes quanto as ciências naturais e sociais. (FAGUNDES, 1998: 15)

A partir desse conceito, foi desenvolvida uma metodologia baseada em Aprendizagem por Projetos que se diferencia do Ensino por Projeto, no que tange a vários aspectos, tais como (FAGUNDES, 1998: 17):

Tabela 1: Comparativo entre Ensino por Projetos e Aprendizagem por Projetos (FAGUNDES, 1998: 17).

	<i>Ensino por Projetos</i>	<i>Aprendizagem por Projetos</i>
Autoria: quem escolhe o tema?	Professores, coordenação pedagógica	Alunos e professores individualmente e, ao mesmo tempo, em cooperação
Contextos	Arbitrado por critérios externos e formais	Realidade da vida do aluno
A quem satisfaz	Arbítrio da sequência de conteúdos do currículo	Curiosidade, desejo, vontade do aprendiz
Decisões	Hierárquicas	Heterárquicas
Definições de regras, direções e atividades	Impostas pelo sistema, cumpre determinações sem optar	Elaboradas pelo grupo, consenso de alunos e professores
Paradigma	Transmissão do conhecimento	Construção do conhecimento
Papel do Professor	Agente	Estimulador / orientador
Papel do aluno	Receptivo	Agente

Considerando as características acima citadas sobre Aprendizagem por Projetos, conclui-se que um documento gerado na forma de *Projeto de Aprendizagem* poderia servir como um registro sistematizado das atividades, podendo este ser utilizado como indicador dos interesses do aprendiz em um dado instante. Surge daí o interesse pelo uso da metodologia de Aprendizagem por Projetos como fundamentação pedagógica para este trabalho.

Em paralelo a isso, o grupo de pesquisa em Inteligência Artificial da UFRGS (grupo no qual este trabalho está inserido) já tem trabalhos relacionados ao uso de sistemas de RI para ambientes educacionais (NAKAYAMA, 2004, 2005). Faz-se,

entretanto, uso de Redes Bayesianas³ (SEIXAS, 2005) como modelo de representação do conhecimento.

O desenvolvimento de Projetos de Aprendizagem é uma pedagogia fundamentada nos princípios acima apresentados. Nela, o aprendiz parte de suas interações com o mundo, da falta de explicações para os fenômenos físicos, sociais e culturais de seu mundo, ou seja, parte de uma *questão de investigação*.

Em um momento seguinte, o indivíduo se dedica a fazer um inventário das idéias que ele associa à sua questão. Essas idéias podem ser agrupadas em dois blocos: *certezas e dúvidas*. Elas farão parte do suporte metodológico para a “viagem” que será feita. As dúvidas serão temporárias, e durante o processo iremos buscar elementos para esclarecê-las. As certezas serão provisórias, dado que, posteriormente, elas poderão ser validadas ou refutadas.

Para fazer essa aventura, os aprendizes fazem planos, organizam as certezas e dúvidas de uma forma lógica, escolhem uma ordem para a investigação. A cada etapa do processo, colhem dados, analisam, fazem conexões, sintetizam, permitindo aflorarem novas dúvidas e certezas.

Como exemplo desse processo, pode-se considerar o seguinte caso hipotético: um indivíduo deseja aprofundar seus conhecimentos sobre colônias de formigas. Ele pode escolher como questão de investigação: “Como funciona a organização de um formigueiro?”, assim como elaborar algumas certezas e dúvidas iniciais, como as que seguem:

Tabela 2: Exemplo de Certezas para o Projeto de Aprendizagem Formigueiro

Certezas:
c1: existe uma subdivisão de tarefas no formigueiro
c2: as formigas trabalham em função de proteger a rainha
c3: as formigas responsáveis por buscar comida e manter o formigueiro são as operárias
c4: as formigas podem se comunicar umas com as outras
c5: existe uma hierarquia no formigueiro
c6: o formigueiro tem compartimentos com funções bem definidas

3 Para Seixas (2005), uma rede Bayesiana é formada por um grafo acíclico, onde os nós representam as variáveis que se inter-relacionam e representam a estrutura de raciocínio de um especialista ou aluno aprendiz num domínio de aplicação. Os nós são ligados por arcos, direcionados no sentido do nó de saída para o nó de entrada; para cada nó, devem ser estimados os valores de probabilidade condicional.

c7: um dos compartimentos do formigueiro é o depósito de lixo
c8: o depósito de lixo é o compartimento mais fundo do formigueiro

Tabela 3: Exemplo de Dúvidas para o Projeto de Aprendizagem Formigueiro

Dúvidas:
d1: qual a quantidade de formigas em um formigueiro?
d2: como se escolhe a rainha de um formigueiro?
d3: existe apenas uma rainha no formigueiro?
d4: como as formigas se comunicam umas com as outras?
d5: a rainha pode comandar as outras formigas, tomar decisões?
d6: já foi comprovado algum nível de inteligência em um formigueiro?
d7: que atividades inteligentes já foram identificadas em formigueiros?
d8: podem ocorrer desentendimentos (brigas) no formigueiro?
d9: como é a reação das formigas em caso de invasão de predadores?
d10: as formigas podem se adaptar, aprender e se comportar de forma diferente, conforme problemas que elas encontrem?
d11: existe um compartimento para depositar alimento, ou as formigas o largam em qualquer lugar?

Segundo Fagundes (1998: 17), existem diferentes caminhos que podem levar à construção do projeto, a partir das necessidades do aluno. Inventando e decidindo é que os estudantes/autores vão ativar e sustentar sua motivação. Para tanto, precisa-se respeitar e orientar a sua autonomia para:

- Decidir critérios de julgamento sobre relevância em relação a determinado contexto;
- Busca/localizar/selecionar/recolher informações;
- Definir/escolher/inventar procedimentos para testar a relevância das informações escolhidas em relação aos problemas e às questões formuladas;
- Organizar e comunicar o conhecimento construído.

Como sugestão de procedimentos subseqüentes, pode-se considerar que o estudante gostaria de classificar essas certezas e dúvidas em grupos contextuais como o exemplo que se segue:

1. G1 (tarefas): C1, C3.
2. G2 (hierarquia): C2, C5, D2, D3, D5.

3. G3 (comunicação): C4, D4, D5.
4. G4 (inteligência): D6, D7, D10.
5. G5 (compartimentos): C6, C7, C8, D11.
6. G6 (conflitos): D8, D9.
7. G7 (população): D1.
8. G8 (sem classificação): Nenhum.

A coleta de dados, o ponto de partida para a validação das certezas e para o esclarecimento das dúvidas ocorre de várias maneiras, usando-se diversas fontes. São as mais comuns: a observação, as entrevistas, as enquetes, as leituras e as simulações. Neste trabalho, focamos nossa atenção na busca por material de leitura na Internet.

Um recurso que pode ser utilizado, antes de se partir para a coleta de informações, é a definição de um planejamento individual de estudos. Esse plano será também sempre inacabado, flexível, ajustável às novas demandas e oportunidades, e contínuo, pois sua essência é justamente a aprendizagem constante e continuada. Sem um plano, torna-se difícil avaliar nossos progressos, planejar nosso tempo, organizar nossas ações, etc.

Fazer um plano requer certa sistematização e uma reflexão que vai além de apenas identificar necessidades e desejos. É preciso também estipular metas, ações e recursos. Resumidamente, um plano de trabalho deveria registrar as respostas às seguintes perguntas:

- O quê? – tema a ser pesquisado ou desenvolvido. Pode ser algo já estudado dentro do contexto do projeto de aprendizagem ou algo novo;
- Quando? – período de tempo a ser dedicado ao trabalho;
- Como? – procedimentos a serem adotados no estudo;
- Quais as metas? – definição do que se quer alcançar, determinando-se também a forma de avaliação;
- Com quem? – definição de possíveis colaboradores para o estudo, até mesmo de um grupo de estudos;

- Como apresentar? – formas de socializar o que foi estudado.

A função dessa etapa é organizar a sequência de passos da pesquisa, que pode ser registrada na forma de um texto descritivo ou de outras formas, por exemplo, tabelas, esquemas, etc.

A partir dessa primeira rodada podem surgir muitas novas dúvidas e certezas formando assim um ciclo contínuo de aprendizagem. Para Schlemmer (2001), a metodologia de Projetos de Aprendizagem Baseados em Problemas favorece a cooperação, que se dá por trocas recíprocas e respeito mútuo. Busca-se aprender conteúdos, aprofundar conceitos, por meio de procedimentos que ajudam o sujeito a desenvolver a própria capacidade de continuar aprendendo, num processo construtivo e simultâneo de questionar-se. Mais adiante, no mesmo trabalho a autora ainda afirma que:

Assim, a metodologia de Projetos de Aprendizagem Baseados em Problemas poderá ser uma porta que permitirá: aproximar-se da identidade dos estudantes; evidenciar e desenvolver talentos; favorecer o desenvolvimento de concepções humanistas; desenvolver a cooperação, comunicação, autonomia, criatividade, senso crítico; repensar a estrutura de cursos, a organização do currículo por disciplinas e a maneira de estabelecê-lo no tempo e no espaço educacional; resgatar o que ocorre fora do espaço educacional formal, as transformações derivadas da imensa produção de informação; questionar a idéia de uma versão única da realidade; saber onde buscar a informação, como buscar e o que fazer com ela. (SCHLEMMER, 2001: 16)

A pedagogia de Projetos de Aprendizagem aplicada a Ambientes Virtuais de Aprendizagem é um recurso que já vem sendo utilizado desde 2002 por pesquisadores da Universidade Federal do Rio Grande do Sul (UFRGS). O projeto AMADIS (BASSO, 2005) é um exemplo de ferramenta dessa natureza e já é realidade em escolas da rede pública. A principal crítica à ferramenta é a ausência de um mecanismo que recupere documentos da Web diretamente, a partir do conteúdo cadastrado nos Projetos de Aprendizagem, e em seguida ofereça esses documentos ao usuário.

3.6 *Ambientes de aprendizagem*

Com uma abundância de novos espaços eletrônicos, de ferramentas de interação e com a explosão da educação à distância, há a tendência de que espaços eletrônicos sejam cada vez mais utilizados para facilitar a aprendizagem, tanto como suporte para distribuição, aquisição e recuperação de materiais didáticos quanto como complemento aos espaços presenciais de aprendizagem.

Segundo Otto Peters (2004: 133), o conceito de *ambiente de aprendizagem* foi criado com base na proposição de mudança do paradigma educacional de instrução, com uma finalidade voltada para metas e de bases empíricas, para a aprendizagem construtivista. Os estudantes não são mais vistos como *objetos*, mas sim como *sujeitos* do processo de aprendizagem. Sua aprendizagem não consiste mais em receber e processar o conhecimento oferecido, mas em debater ativamente com um objeto de aprendizagem que eles mesmos selecionaram em um contexto que é definido a partir da interação simultânea com outros estudantes e no qual eles mesmos desenvolvem ou alteram estruturas cognitivas individuais.

Entende-se por ambiente de aprendizagem qualquer sistema que forneça suporte a qualquer tipo de atividade realizada pelo aluno, isto é, um conjunto de ferramentas usadas em diferentes situações do processo de aprendizagem.

[...] um ambiente de aprendizagem poderá ser muito rico, porém, se o aluno não desenvolve atividades para o aproveitamento de seu potencial, nada acontecerá [...]. (GALVIS, 1992: 52).

Qualquer ambiente deve permitir diferentes estratégias de aprendizagem, não só para se adequar ao maior número possível de pessoas, que terão certamente estratégias diferentes, mas também porque as estratégias utilizadas individualmente variam de acordo com fatores como interesse, familiaridade com o conteúdo, estrutura dos conteúdos, motivação e criatividade. Além disso, deve proporcionar uma aprendizagem colaborativa, interação e autonomia – esta última é tema central deste trabalho.

Ademais, se esse ambiente de aprendizagem é informatizado e disposto em rede, uma imensa extensão do campo educacional acontece na imaginação dos alunos (PETERS, 2004: 134). Essa nova configuração provê uma nova dimensão de

oportunidades e chances, especialmente ao se considerar a população atingida – crianças e jovens da escola pública, no geral das classes média e baixa –, que agora poderá dispor de mais acesso à informação.

4 Aspectos tecnológicos

O principal trabalho desta pesquisa encontra-se na área tecnológica, pois se trata de uma proposta computacional para auxiliar em ambientes pedagógicos.

Segundo Solange Oliveira Rezende (2003, 7), os pontos-chave nos Sistemas Inteligentes são:

- habilidade para usar conhecimento para desempenhar tarefas ou resolver problemas;
- a capacidade para aproveitar associações e inferências para trabalhar com problemas complexos que se assemelham a problemas reais.

A mesma autora diz que, entre as habilidades inteligentes, está a habilidade para armazenar e recuperar eficientemente grande quantidade de informação, para resolver problemas ou tomar decisões, e para conectar nossos pensamentos e nossas idéias de maneira não linear, ou seja, de modo associativo.

O centro desse tipo de comportamento refere-se não somente à nossa habilidade organizacional complexa, mas também à nossa habilidade de adaptar ou modificar nosso comportamento (com base na racionalidade, no bom senso) e empregar várias habilidades a uma dada situação.

Para um sistema computacional ser “inteligente”, ele precisa possuir um subconjunto dessas habilidades e saber como elas modelam tarefas específicas. (REZENDE, 2003: 7)

A partir dessas definições, serão introduzidos a seguir os conceitos relacionados à tecnologia de agentes de software. Outros conceitos tecnológicos necessários ao empreendimento deste trabalho também serão apresentados posteriormente, tais como o conceito de recuperação de informações em contexto e ontologias.

4.1 Agentes de Software

Todos os itens descritos na revisão de literatura, até o momento, pretendem embasar o desenvolvimento de um conjunto de agentes assistentes de aprendizes autônomos, baseados na recuperação de informações em contexto. Incrementando assim a motivação e qualificando o processo de aprendizagem autônoma dos mesmos.

O termo agente apresenta conceituações variadas, de acordo com grupos e pesquisadores diferentes. Assim, inicialmente, percebe-se a importância de se definir esse termo, em função da abordagem do projeto, destacando as características mais comuns que os identificam e apresentando definições de tipos de agentes utilizados no escopo do trabalho.

4.1.1 Conceito de agente

No mundo físico, um agente é alguém que age em função de interesses alheios, desempenhando papéis específicos (FERREIRA, 1994). Atualmente, sistemas baseados no paradigma de agentes têm recebido atenção especial, despertando o interesse por estudos mais aprofundados e para a possibilidade da sua utilização na construção de variadas aplicações, em diversas áreas da computação, embora não exista, ainda, uma definição precisa do que seja um agente nesse contexto (HEDBERG, 1995).

Uma das definições mais aceita considera que um agente trata-se, normalmente, de um *software* atuando num papel corriqueiramente desempenhado por um usuário humano, podendo, ainda, agir em função de outro agente computacional (JOHANSEN, 1995), com o propósito de ajudar na realização de determinadas tarefas. Formalmente, podemos definir os agentes de software como

uma peça de software que executa uma determinada tarefa, empregando informações extraídas de seu ambiente para agir de forma adequada, no sentido de completar sua tarefa de modo bem sucedido. O software deve ser capaz de adaptar-se a eventuais modificações ocorridas em seu ambiente de maneira que o resultado pretendido seja independentemente alcançado. (LECKY-THOMPSON, 1997)

Podemos considerar ainda que:

Chama-se de agente uma entidade abstrata que é capaz de agir sobre ela mesma e sobre seu próprio ambiente, que dispõe de uma representação parcial deste ambiente, e que, em um universo multiagente, pode comunicar-se com outros agentes, e cujo comportamento é consequência de suas observações, de seu conhecimento e das interações com outros agentes. (FERBER, 1991)

Assim, percebe-se que, para a realização de suas tarefas, surge a necessidade de especificar o comportamento desses agentes, bem como promover a comunicação entre eles, com os seus usuários e, também, com o ambiente em geral. Tais agentes podem realizar trabalhos rotineiros para os seus usuários, bem como auxiliá-los em tarefas complexas.

4.1.2 Características de agentes

O termo agente é empregado, em sua forma mais geral, para denotar uma entidade baseada em hardware ou, mais freqüentemente, em *software*, caracterizada por algumas das seguintes propriedades (WOOLDRIDGE, 1995; ETZIONI, 1995; PORTO, 1997):

- autonomia: os agentes funcionam sem a intervenção direta de operadores de qualquer tipo e possuem algum tipo de controle sobre as suas ações e sobre seu estado interno;
- habilidade social: os agentes interagem com outros agentes e/ou com humanos, por meio de alguma linguagem de comunicação, para obter informações ou ajuda na realização de suas tarefas;
- continuidade temporal: os agentes são processos em execução contínua, que tanto podem estar ativos quanto adormecidos;

- iniciativa: diferente dos programas padrões que são diretamente disparados pelos usuários, um agente pode perceber modificações em seu ambiente e decidir quando atuar, de acordo com seus objetivos;
- reatividade: os agentes respondem aos estímulos do ambiente;
- racionalidade: é a hipótese de que os agentes irão agir de forma a atingir seus objetivos e não contra eles, pelo menos dentro do alcance de suas crenças;
- adaptabilidade: um agente deve ser capaz de se adaptar aos hábitos, métodos de trabalho e preferências de seus usuários e, também, às modificações em seu ambiente;
- colaboração: um agente não deve aceitar (e executar) instruções impensadamente. Deve levar em conta que seres humanos (ou outros agentes) cometem erros, como dar uma ordem com objetivos conflitantes, omitir informações importantes ou fornecer informações ambíguas. Assim deve também ser capaz de recusar ordens que, por exemplo, produzam uma sobrecarga inaceitável na rede ou que ocasionem danos a outros usuários;
- e outras propriedades como: orientação a objetivos, benevolência, mobilidade.

Embora nenhum agente tenha todas essas características, vários protótipos apresentam uma boa parte delas (ETZIONI, 1995). Não há, até o momento, um consenso sobre a importância relativa dessas diferentes propriedades, mas a maioria dos pesquisadores concorda que essas características possibilitam a diferenciação entre agentes e programas quaisquer.

Para a construção de um Assistente Pessoal de Aprendizagem, ou de um agente de software que se preste a isso, espera-se que o agente ou o sistema de agentes apresentem mais fortemente as seguintes características: autonomia, adaptabilidade e colaboração. Os estudos da tecnologia de agentes já se mostraram promissores em áreas como software de mensagens, ferramentas de desenvolvimento, softwares de interface com o usuário, controle de processos, gestão do fluxo de trabalho, gerenciamento de redes e gestão e recuperação de

informações (JENNINGS & WOOLDRIGDE, 1998). Esta última aplicação se alinha às pretensões deste trabalho e a ela, portanto, daremos atenção especial a seguir.

4.1.3 Algoritmos Genéticos

Em programação genética (PG), está envolvida uma população de programas de computadores. Isso é, geração por geração, PG transforma estocasticamente a população de programas em uma nova, com esperança de melhorias. PG, tal como a natureza, é um processo aleatório, e não pode garantir resultados (POLI, LANGDON E MCPHEE, 2008). Apesar de seu caráter essencialmente aleatório, podem ser utilizados com métodos determinísticos, e assim como na natureza, os algoritmos genéticos (AG) têm sido muito bem sucedidos na evolução de formas novas e inesperadas de se resolver problemas.

Os passos básicos de um algoritmo genético estão apresentados a seguir. Um AG tenta descobrir como um programa funciona, executando-o, e, em seguida, comparando o seu comportamento a algum ideal (linha 3). Alguém pode estar interessado, por exemplo, na forma como um programa prevê uma série temporal ou controla um processo industrial. Esta comparação é quantificada por um valor numérico chamado aptidão. Os programas mais bem sucedidos são escolhidos (linha 4) e produzem novos programas para a nova geração (linha 5). As operações genéticas primárias que são usadas para gerar novos programas a partir dos preexistentes são:

- recombinação (*crossover*): Criação de um programa filho, combinando aleatoriamente duas partes de programas pais selecionados;
- mutação: A criação de um novo programa filho aleatoriamente alterando randomicamente uma parte de um programa pai selecionado.

Os passos básicos para a implementação de um AG seguem abaixo:

1. Criação aleatória de uma população inicial de programas a partir das primitivas iniciais;
2. Repita
 - a. Executar cada programa e determinar a sua aptidão;
 - b. Selecionar um ou dois programas a partir da população com probabilidade baseada em sua aptidão para participar das operações genéticas;
 - c. Criar novos programas individuais pela aplicação das operações genéticas com as probabilidades especificadas;
3. Até encontrar uma solução aceitável ou alguma outra condição de parada (isto é, um número máximo de gerações é alcançado);
4. Retorna assim o melhor indivíduo.

A aptidão dos indivíduos avaliados por um algoritmo genético é calculada por uma função objetivo. Essa função é o objetivo da otimização de um AG. Ela pode referir-se a um problema de otimização, um conjunto de testes para identificar os indivíduos mais aptos, ou mesmo uma "caixa preta" onde sabemos apenas o formato das entradas dos dados e que retorna um valor que se quer aperfeiçoar. A grande vantagem dos algoritmos genéticos está no fato de não se precisar saber como funciona esta função objetivo, apenas tê-la disponível para ser aplicada aos indivíduos e comparar os resultados.

A seleção é o processo de eliminação dos indivíduos menos aptos ao meio. É nesse momento que é utilizada a aptidão, pois é ela que definirá quais indivíduos sobreviverão nas próximas gerações e quais serão extintos. Os indivíduos que compõem a população são classificados conforme seu grau de adaptação, os últimos indivíduos são excluídos. A quantidade excluída é pré-definida.

O cromossomo é o componente principal de um algoritmo genético e representa uma solução (indivíduo) candidata para o problema em questão. Um cromossomo é composto de genes, os quais descrevem a solução candidata. Os genes de um cromossomo podem ser compostos por uma string de bits (e.g. 10111011110), por valores inteiros (1, 2, 3, 4), ou por valores reais (e.g. 33.9, 11.0,

ABC etc).

Como citado acima, nos passos básicos de um AG, a geração inicial de uma população é dada através de primitivas iniciais. De forma geral, o cálculo da aptidão inicial é dado pela submissão de cada indivíduo à função objetivo que se deseja otimizar (BIANCONI, 2007). O indivíduo que obtiver o maior valor resultante da função será classificado como mais apto à busca, seguido do segundo maior valor, e assim por diante.

Este trabalho lança mão de AG para fazer a representação do conhecimento dos agentes utilizados para recuperação de informações. A seguir é apresentada uma introdução aos sistemas de recuperação de informações, teoria que embasa o trabalho apresentado nesta tese. Posteriormente apresentaremos como foi modelado o AG utilizado neste trabalho.

4.2 Recuperação de Informações

A área de Recuperação de Informações (RI) remete a representação, armazenamento, organização e acesso a itens de informação. A representação e a organização dos itens deve prover, para o usuário, um acesso fácil àquelas informações pelas quais ele se interessa.

Retomando o exemplo do formigueiro, poderíamos imaginar que alguém envolvido naquele projeto de aprendizagem (cf. seção 4.6) poderia se interessar por encontrar páginas com a seguinte descrição:

Encontre todas as páginas contendo informações sobre formigueiros e que: (1) detalhem a sua organização social e (2) descreva as funções de cada classe de formiga.

Obviamente, essa descrição completa da necessidade do usuário não pode ser utilizada diretamente para requisitar informações usando as interfaces atuais dos motores de busca na Web. É necessário que o usuário primeiramente traduza essa necessidade de informação em uma “consulta” (*query*) que possa ser processada

por um motor de busca.

Normalmente essa tradução se resume a um conjunto de palavras-chave (ou termos indexados) que representam a descrição da necessidade de informação do usuário. Sendo assim, dada uma consulta, o objetivo de um sistema de RI é recuperar informações que sejam relevantes para o usuário.

Segundo Baeza-Yates (1999: 3), a recuperação efetiva de informações relevantes é diretamente afetada pelas tarefas com as quais o usuário está envolvido (*user tasks*) e pela visão lógica dos documentos adotada pelo sistema de RI. A Figura 4 ilustra a interação do usuário nas diferentes tarefas identificadas.

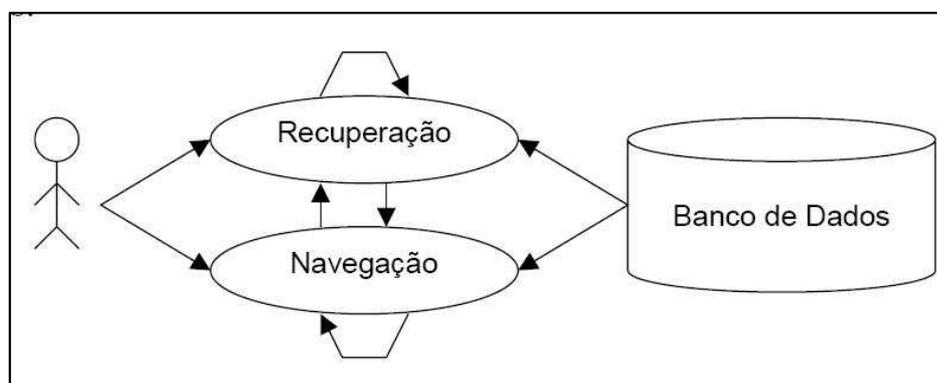


Figura 4: Interação do usuário com o sistema de recuperação através de diferentes tarefas (BAEZA-YATES, 1999: 4)

Para Peter Ingwersen (2005a: 19), existem alguns fatores importantes associados ao processo de RI, tais como: os participantes ou componentes do processo, as tarefas, o conteúdo, a necessidade de informação e sua relevância, e o processo.

Como participantes do processo de Busca e Recuperação de Informações (B&RI), o autor considera seis componentes:

- Ator Cognitivo: pessoa responsável pela interpretação ou provisão de informação em potencial ou de sinais representados como objetos de informação, TI, funcionalidades de interface e comunicação. As categorias principais de atores em RI são: usuários, autores, indexadores, projetistas de sistemas de algoritmos, seletores (tais como publicadores, editores e

empregadores);

- **Objetos de Informação:** entidades físicas (digitais) que podem ser de mídias variadas e que pertençam ao espaço do Sistema de RI, provendo informações em potencial. Os Objetos de Informação são usados alternadamente com os termos dos documentos;
- **Espaço de Informação:** representado pelos objetos de informação e consistindo de informação potencial, é geralmente estruturado de acordo com o conjunto de Tecnologias da Informação (TI) do Sistema de Informação;
- **Sistema de RI:** um Sistema de Informação constituído por processos interativos entre o Espaço de Informação, o Conjunto de TI, as funcionalidades de interface e o seu ambiente. Ele é capaz de buscar e encontrar informações de valor potencial para o seu usuário;
- **Interface:** é um mecanismo posicionado como mediador entre dois componentes eletrônicos ou humanos de um sistema de informações;
- **Contexto:** em sistemas de RI, os atores e os outros componentes funcionam (proveem informações) como contexto uns para os outros no processo de interação. Existem contextos sociais, organizacionais, culturais, assim como contextos sistêmicos, que se envolvem todo tempo.

Para Ingwersen (2005a: 19), as tarefas com as quais o usuário está envolvido são os fatores que definem o conceito de contexto (discutido nas seções que se seguem). Para o autor, essas tarefas podem ser de três tipos:

- **tarefas do dia a dia ou interesse:** todas as tarefas e interesses que não são tarefas de trabalho ou tarefas de consulta. Essas atividades ou tarefas devem ser de natureza social e cultural, incluindo lazer e entretenimento;
- **tarefas de trabalho:** uma tarefa relacionada a trabalho ou não, ou o interesse a ser cumprido pelo ator cognitivo. Eventualmente,

algumas tarefas podem não ser resolvidas imediatamente pelo(s) ator(es), então elas podem levá-lo(s) a um estado de incerteza ou a uma situação que demande uma tarefa de pesquisa.

- tarefas de pesquisa: tarefa executada por um ator com a intenção de obter informações associadas ao cumprimento de alguma tarefa de trabalho. Tarefas de busca podem ser tanto tarefas de procura (navegação) como de recuperação, dependendo do envolvimento com o sistema de RI.

O conteúdo manipulado por esses sistemas é classificado de duas formas:

- Informação: esse conceito, na perspectiva da Ciência da Informação, deve atender a dois requisitos: ser o resultado de uma transformação nas estruturas cognitivas de um ator gerador, e por outro lado, ser algo que, quando percebido, afete e transforme o estado de conhecimento de um receptor;
- Conhecimento: o entendimento total do indivíduo sobre si e sobre o universo que o rodeia em um dado ponto no tempo, incorporando pensamento e cognição, assim como propriedades intuitivas e emocionais e memória (sub)consciente (conhecimento tácito). RI pode operar com conhecimento declarativo e procedural em uma dimensão e com conhecimento relacionado ao domínio em outra.

O mesmo autor ainda define conceitos para o que ele chama de necessidades e relevância:

- necessidade de informação: significa uma lacuna identificada conscientemente pelo ator em seu conhecimento. Essa necessidade pode levar a uma procura por informações e formulações de requisições de informações;
- requisição: a formulação de uma necessidade de informação ou de estados de intencionalidades subjacentes, percebidos, e fornecidos em um dado ponto a um sistema de RI;
- consulta: uma transformação de uma formulação de requisição feita pelo próprio ator a um mediador ou interface, com a finalidade

de interrogar um sistema de RI em concordância com a indexação e com algoritmos de recuperação do sistema;

- relevância: é a avaliação da pertinência, utilidade, etc., das fontes de informação, feita pelo ator cognitivo ou pelos dispositivos algorítmicos, com referência a uma dada informação em um dado ponto no tempo. Isso pode se modificar dinamicamente durante o tempo pelo mesmo ator. A relevância pode ser avaliada segundo muitas variáveis, tornando-se, assim, de natureza multidimensional e subjetiva.

Para esclarecer as formas de manipulação de informação, Ingwersen (Idem, 2005a: 21) ainda define os processos da informação como sendo:

- Comportamento Informativo: comportamento humano relacionado com a geração, a comunicação, o uso e outras atividades ligadas a informações, tais como Comportamento de Procura e RI Interativa;
- Interação Informativa: refere-se à troca entre dois ou mais atores cognitivos em um contexto de RI;
- Recuperação de Informação: o processo que envolve representação, armazenagem, busca, encontro, filtragem e apresentação de informações que podem ser potencialmente relevantes a uma requisição desejada por um usuário em dado contexto. A Recuperação de Informações é comumente dividida em RI algorítmica e RI Interativa;
- RI Interativa: é definido como o processo de comunicação interativa que ocorre durante a recuperação de informações envolvendo todos os principais componentes de um sistema de RI, ou seja, o usuário, o contexto sócio-organizacional, o conjunto de Tecnologias da Informação, a interface e o conjunto de Informações.

Após a definição desses conceitos centrais sobre RI, serão apresentados a seguir maiores detalhes sobre os documentos manipulados em um sistema de RI e sua visão lógica.

4.2.1 A Visão Lógica de um Documento

Por questões históricas e de limitações computacionais, os documentos de uma coleção são frequentemente representados por um conjunto de termos indexadores ou por palavras-chave. Essas palavras-chave podem ser extraídas diretamente de um documento de texto ou podem ser especificadas por um indivíduo. Desconsiderando a origem dessas palavras-chave, quer sejam extraídas automaticamente, quer sejam geradas por um especialista, elas provêm uma visão lógica do documento.

Os computadores modernos têm possibilitado a representação de documentos por seu conjunto completo de palavras. Nesse caso, é dito (BAEZA-YATES, 1999: 5) que o sistema de recuperação utiliza uma visão lógica baseada no texto completo (uma representação). Entretanto, mesmo os computadores mais modernos, em coleções muito grandes, precisam reduzir o conjunto de palavras-chave representativas. Isso pode ser acompanhado da remoção de *stopwords* (como artigos e conetivos), o uso de *stemming* (que substituem palavras flexionadas por seus respectivos radicais), identificação de substantivos (que eliminam adjetivos, advérbios e verbos) e a compressão das estruturas lógicas internas (índices e índices invertidos, por exemplo). Essas operações, chamadas de *operações textuais*, reduzem a complexidade da representação dos documentos, além de possibilitarem a conversão de um texto completo para um conjunto de termos indexadores.

Para Baeza-Yates (1999: 5), o documento completo é claramente a mais completa visão lógica de um documento, mas seu uso implica custos computacionais altos. Um conjunto resumido de categorias (geradas por um especialista humano) provê a visão lógica mais concisa de um documento, mas seu uso pode levar a uma recuperação de baixa qualidade. Em alguns casos, visões lógicas intermediárias (de um documento) podem ser utilizadas por um sistema de RI como ilustra a Figura 5. Portanto, adotando algumas das representações

intermediárias, o sistema de RI pode reconhecer estruturas internas normalmente encontradas em um documento (capítulos, seções, subseções, etc.).

Conforme ilustrado na Figura 5, o problema de representar logicamente um documento é contínuo. Essa representação pode variar desde o texto completo até uma especificação de alto nível feita por um especialista humano. A escolha de uma representação mais ou menos elaborada depende da necessidade do sistema de RI a ser implementado.

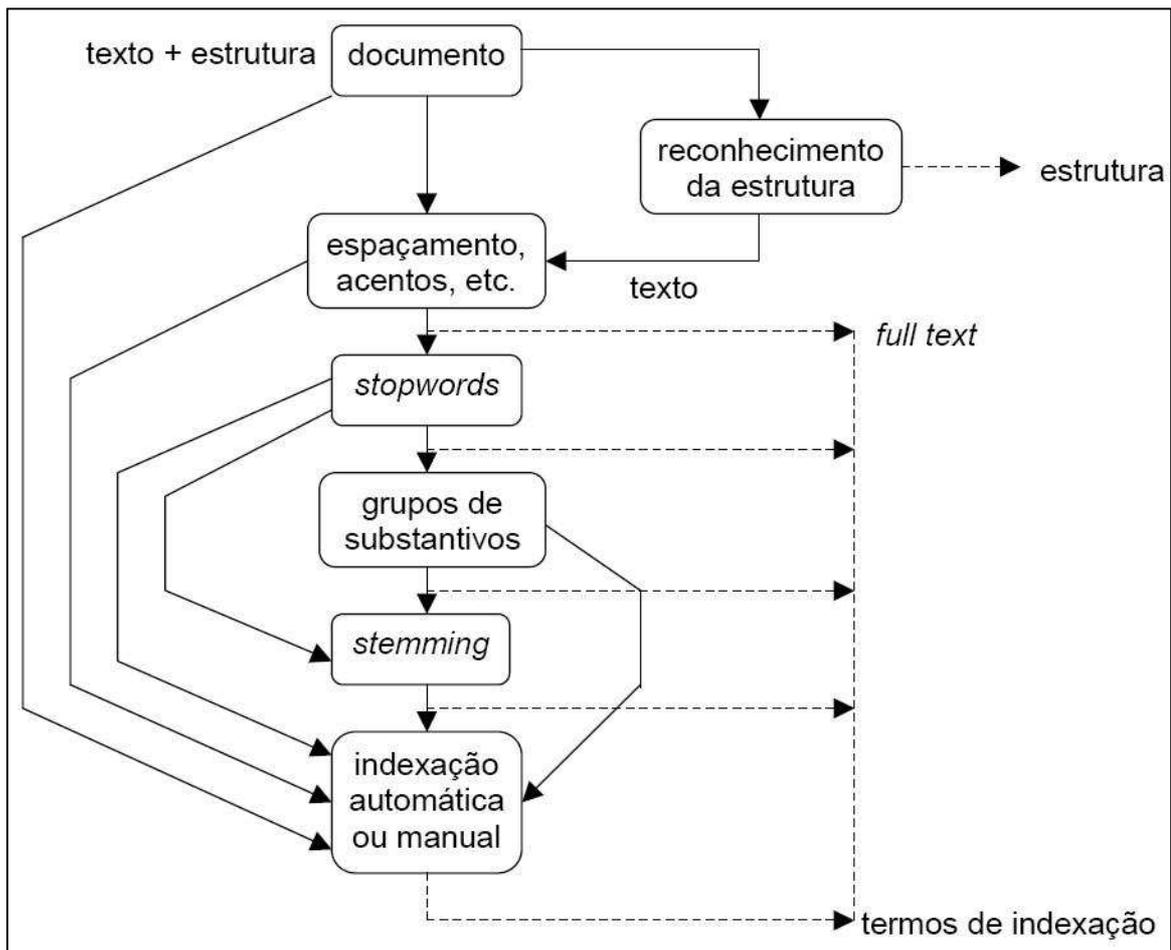


Figura 5: Visão Lógica de um Documento: do Texto Completo ao Conjunto de Termos Indexados (BAEZA-YATES, 1999: 6)

A seguir, serão tratadas algumas questões técnicas sobre como os documentos de uma coleção podem ser recuperados. Serão apresentados os principais modelos de RI, bem como seus princípios de funcionamento.

4.2.2 Técnicas de Recuperação de Informações

Nesta seção serão tratados alguns aspectos técnicos sobre as principais técnicas de recuperação de informações. Para isso, será necessária a definição de alguns conceitos básicos.

Conforme citado na seção anterior, em que foi discutida a visão lógica de um documento dentro de um Sistema de RI, será retomada a discussão sobre a importância isolada de cada termo dentro de um documento, ou seja, o seu peso. Para um conjunto de termos indexados em um documento nem todos são igualmente úteis. Decidir sobre a importância de um termo para resumir o conteúdo de um documento não é tarefa trivial. Desconsiderando essas dificuldades, existem propriedades de um termo indexador que podem ser facilmente mensuradas e que são úteis para avaliar o potencial de um termo como tal. Por exemplo, considerando uma coleção com cem mil documentos. Uma palavra que apareça em cada um dos cem mil documentos não tem utilidade alguma como termo indexador, pois ela não diz nada sobre quais documentos o usuário pode se interessar. Por outro lado, uma palavra que apareça em apenas cinco documentos se torna bastante útil, pois ela reduz consideravelmente o espaço de documentos que podem ser de interesse do usuário.

Segundo Baeza-Yates (1999: 24), termos indexadores distintos têm relevâncias variadas quando utilizados para descrever o conteúdo de documentos. Esse efeito é capturado pela associação de pesos numéricos para cada termo indexador de um documento.

Formalmente, considerando que k_i seja um termo indexador, d_j seja um documento e $w_{i,j} \geq 0$ seja um peso associado com o par (k_i, d_j) . Esse peso quantifica a importância do termo indexador na descrição semântica de um documento.

Definição: considere que t seja o número de termos indexados no sistema e k_i seja um termo genérico. $K = \{k_1, \dots, k_t\}$ é o conjunto de todos os termos indexados. Um peso $w_{i,j} > 0$ é associado a cada termo k_i de um documento d_j , para um termo que não apareça no texto de um documento temos $w_{i,j} = 0$. Com o documento d_j

associado a um vetor de termos indexados \vec{d}_j , representado por $\vec{d}_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j})$. Posteriormente, consideremos g_i como sendo a função que calcula o peso associado com um termo indexado k_i em qualquer vetor t-dimensional (i.e. $g_i(\vec{d}_j) = w_{i,j}$).

Para Baeza-Yates (1999: 25), os pesos dos termos indexados são normalmente considerados mutuamente independentes. Isso significa que, conhecendo o peso $w_{i,j}$ associado com o par (k_i, d_j) , isso não nos diz nada sobre o peso $w_{i+1,j}$ associado ao par (k_{i+1}, d_j) . Isso é obviamente uma simplificação, porque as ocorrências de termos indexados em um documento são correlacionadas. Considere, por exemplo, que os termos “computador” e “rede” são utilizados para indexar um dado documento que trata da área de “redes de computadores”. Frequentemente, neste documento, o aparecimento de uma dessas duas palavras atrai o aparecimento da outra. Então, tais palavras são correlacionadas e seus pesos podem refletir essa correlação. Baeza-Yates (1999:25) ainda afirma que “a independência mútua parece ser uma simplificação forte, mas ela também diminui a tarefa de computar os pesos dos termos indexados, possibilitando um procedimento de ranking mais rápido”.

As definições anteriores darão suporte para a melhor compreensão e o entendimento dos modelos de recuperação de informação detalhados a seguir. São eles: Booleano, Vetorial, Latent Semantic Indexing, Probabilísticos e Context Network Graphs.

4.2.2.1 Modelo Booleano

O Modelo Booleano é um modelo de recuperação simples baseado na Teoria dos Conjuntos e na Álgebra Booleana. Sendo o conceito de Conjunto bastante intuitivo, o modelo Booleano provê um *framework* fácil de manipular por um usuário comum de um sistema de RI. Em sistemas dessa natureza, as consultas são definidas como expressões Booleanas com semântica precisa. Por conta de sua simplicidade e formalismo, o modelo Booleano recebeu atenção especial no passado, sendo adotado em muitos sistemas de bibliotecas.

Esse modelo é muito mais utilizado para recuperação de dados do que para recuperação de informação. É bem utilizado apenas pelo usuário que entende bem a álgebra booleana.

Exemplo: numa consulta com 3 termos: t_1 , t_2 e t_3 , as possibilidades de ocorrência desses termos em documentos pertence a uma das seguintes opções:

Tabela 4: Combinações possíveis de mintermos para uma consulta com 3 termos.

$m_1 = t_1, t_2, t_3$	$m_5 = t_1', t_2', t_3$
$m_2 = t_1', t_2, t_3$	$m_6 = t_1, t_2', t_3'$
$m_3 = t_1, t_2', t_3$	$m_7 = t_1', t_2, t_3'$
$m_4 = t_1, t_2, t_3'$	$m_8 = t_1', t_2', t_3'$

Onde $m_1 \dots m_8$ são mintermos – conjuntos que descrevem todas as possibilidades para o conjunto resposta da consulta. Com 4 termos, isso fica bem mais complicado, e assim por diante. O número de mintermos cresce exponencialmente (número de mintermos = 2^n , onde n é o número de termos da consulta).

O modelo Booleano considera que os termos indexados estão presentes ou não em um documento. Uma consulta q é composta de termos indexados ligados por três conectivos “não”, “e” e “ou”. Deste modo, uma consulta é essencialmente uma expressão booleana convencional, que pode ser representada por uma disjunção de vetores conjuntivos (i.e., forma normal disjuntiva – DNF). Por exemplo,

a consulta $[q = k_a \wedge (k_b \vee \sim k_c)]$ pode ser escrita na forma normal disjuntiva como $[\vec{q}_{dnf} = (1,1,1) \vee (1,1,0) \vee (1,0,0)]$, onde cada componente é um vetor de pesos binários associados com a tupla (k_a, k_b, k_c) . Esses vetores de pesos binários são chamados de componentes conjuntivos de \vec{q}_{dnf} .

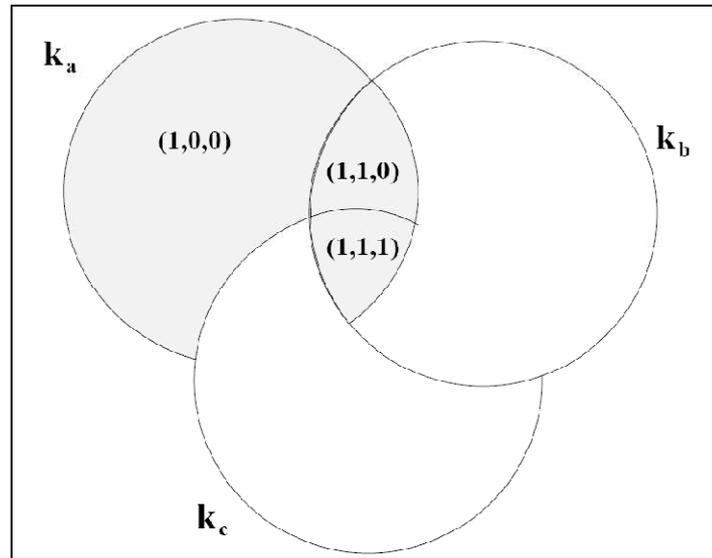


Figura 6: Componentes conjuntivos da consulta $[q = k_a \wedge (k_b \vee \sim k_c)]$.

Definição: para o modelo Booleano, o peso das variáveis do vetor de termos indexados é binários, i.e., $w_{i,j} \in \{0,1\}$. Uma consulta q é uma expressão booleana convencional. Considere que \vec{q}_{dnf} seja a forma normal disjuntiva da consulta q . Posteriormente, considere que \vec{q}_{cc} seja algum dos componentes conjuntivos de \vec{q}_{dnf} . A similaridade de um documento d_j para a consulta q é definida como:

$$sim(d_j, q) = \begin{cases} 1, & \text{se } \exists \vec{q}_{cc} \mid (\vec{q}_{cc} \in \vec{q}_{dnf}) \wedge (\forall k_i, g_i(\vec{d}_j) = g_i(\vec{q}_{cc})) \\ 0, & \text{Caso Contrário} \end{cases}$$

Se $sim(d_j, q) = 1$, então o modelo Booleano considera que o documento d_j é relevante para a consulta q (podendo não ser). Caso contrário, o resultado é que o documento não é relevante.

As principais vantagens do modelo booleano são:

- expressividade completa se o usuário souber exatamente o que quer;
- é facilmente programável e exato;

Como desvantagens, têm:

- pessoas lidam com conhecimento parcial;
- saída pode ser nula ou haver sobrecarga (problema que dá um falso senso de segurança);
- a saída não é ordenada.

Apesar de suas limitações, o modelo é altamente utilizado em sistemas comerciais. Existem algumas formas de tentar melhorar os resultados gerados nesse modelo:

- atribuindo pesos aos termos: a carga semântica dos termos é completamente diferente quando se tem, por exemplo, uma consulta com dois termos ou duas consultas distintas com um termo cada;
- usando conjuntos *fuzzy*: a pertinência ou não de um elemento a um conjunto varia entre 0 e 1, não é exata;
- categorização em RI: dividir a consulta em classes e conceitos, tentar encontrar os documentos baseados nos conceitos;
- *passage retrieval*: o conjunto de termos a ser procurado deve aparecer o mais próximo possível, por exemplo, da mesma página (uma possível passagem). É uma técnica mais eficiente que a de *document retrieval*, porém muito mais difícil de ser implementada. A proximidade é importante.
- ordenando a saída: uma vez que haja alguma forma de determinar quais termos são mais importantes para determinada consulta, é possível ordenar o resultado.

Como contrapartida às desvantagens do modelo booleano, existem outros modelos de RI que já proveem soluções para as lacunas de ranqueamento e atribuição de pesos.

4.2.2.2 Modelo Vetorial

O Modelo Vetorial reconhece que o uso de pesos binários é muito limitado e propõe um *framework* pelo qual é possível mensurar uma similaridade parcial (BAEZA-YATES, 1999: 27). Isso é alcançado pela associação de pesos não binários aos termos indexados das consultas e dos documentos. Esses pesos são utilizados, então, para medir o grau de similaridade entre cada documento armazenado no sistema e a consulta do usuário. Classificando os documentos recuperados em ordem decrescente desse grau de similaridade, o Modelo Vetorial leva em conta documentos que se igualam apenas parcialmente com a consulta. O principal efeito resultante é que o conjunto de documentos ranqueados é bem mais preciso que o apresentado pelo Modelo Booleano.

No modelo vetorial, cada documento é representado como um vetor de termos. Cada termo possui um valor associado, que indica o grau de importância (peso - *weight*) deste no documento. Em outras palavras, cada documento possui um vetor associado, constituído por pares de elementos na forma {(palavra_1, peso_1), (palavra_2, peso_2), ..., (palavra_n, peso_n)}.

Esses documentos podem ser organizados, por exemplo, num arquivo invertido:

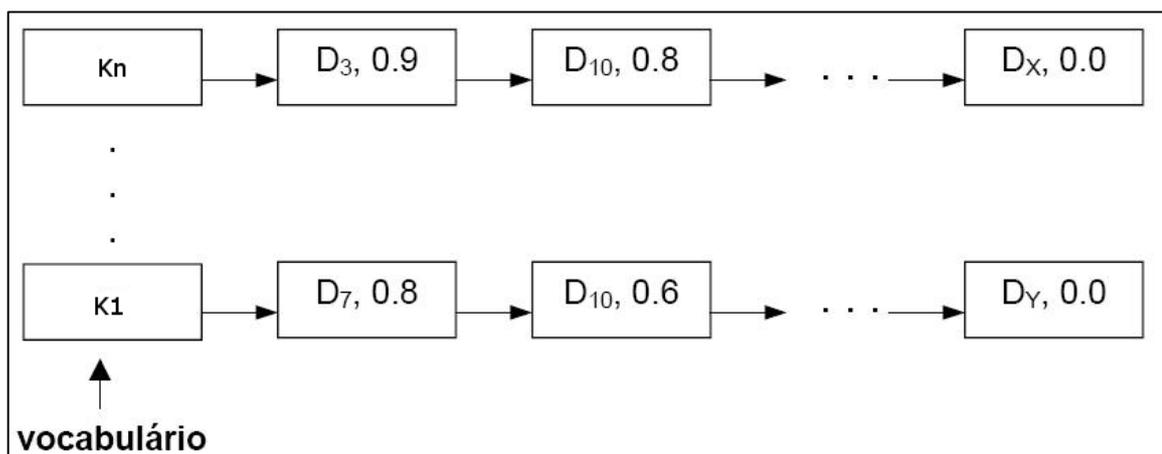


Figura 7: Documentos Indexados com os pesos de cada documento.

Definição: para o modelo vetorial, o peso $w_{i,j}$ associado com o par (k_i, d_j) é positivo, e não binário. A seguir, também são atribuídos pesos aos termos da consulta. Considere que $w_{i,q}$ seja o associado ao par $[k_i, q]$, onde $w_{i,q} \geq 0$. Então, o

vetor \vec{q}_j é definido como $\vec{q}_j = (w_{1,q}, w_{2,q}, \dots, w_{t,q})$, onde t é o número total de termos indexados no sistema. Assim como antes, o vetor de um documento \vec{d}_j é representado por $\vec{d}_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j})$.

Então, um documento d_j e uma consulta do usuário q são representados como vetores t -dimensionais, como demonstrado na figura 8. Portanto, o modelo vetorial propõe avaliar o grau de similaridade de um documento d_j com a consulta q como sendo a correlação entre os vetores \vec{d}_j e \vec{q} . Essa correlação pode ser quantificada, por exemplo, pelo cosseno do ângulo entre esses dois vetores.

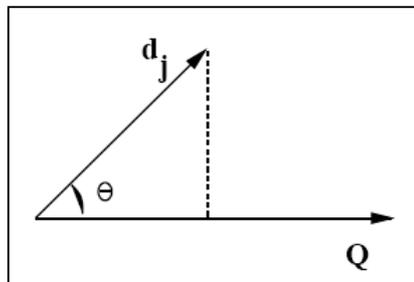


Figura 8: O cosseno do ângulo Θ é adotado como $sim(d_j, q)$.

Então a $sim(d_j, q)$ é calculada pela fórmula:

$$sim(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|}$$

$$sim(d_j, q) = \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}}$$

Partindo do princípio que $w_{i,j} \geq 0$ e $w_{i,q} \geq 0$, $sim(d_j, q)$ varia de 0 até 1, então, em vez de tentar dizer qual documento é relevante ou não, o modelo vetorial classifica os documentos de acordo com o seu grau de similaridade, com a consulta. Um documento pode, assim, ser recuperado mesmo que ele só atenda parcialmente a uma determinada consulta. Entretanto, para computar esse ranque é necessário primeiro estabelecer o modo de obter os pesos dos termos indexados.

Existem muitas formas de computar ranques; o trabalho de Salton e McGill (1983) cita muitos exemplos. Como o objetivo deste trabalho não é discutir em

detalhes tais técnicas, nos deteremos apenas a explicar a ideia principal por trás de uma das mais conhecidas. Os princípios a seguir dão suporte a técnicas de criação de *clusters*.

Dada uma coleção C de objetos e uma vaga descrição de um conjunto A , o objetivo de um algoritmo simples de *clustering* deve ser separar a coleção C em dois conjuntos: um primeiro, composto de objetos relacionados ao conjunto A , e um segundo composto de objetos que não são relacionados a A . Uma vaga decisão quer dizer aqui que não se tem uma informação completa para se decidir precisamente quais objetos são e quais não são do conjunto A . Por exemplo, alguém poderia estar buscando um conjunto A de carros com preços comparáveis ao de um Fusca. Sabendo que o significado comparável não é exato, não existe uma descrição precisa (única) do conjunto A .

Alguns algoritmos de *clustering* (SALTON E MCGRILL, 1983) podem tentar separar os objetos em várias classes, de acordo com suas propriedades. Por exemplo, pacientes de um médico especializado em câncer podem ser classificados em cinco classes: terminal, avançado, em metástase, diagnosticado ou curado. Novamente, as possíveis classes continuam imprecisas e o problema é alguém decidir a qual dessas classes o paciente pertence. Portanto, discutiremos apenas a versão mais simples do problema de *clustering* (i.e. a que considera a existência de apenas duas classes), por conta de que tudo que é requerido é uma decisão sobre quais documentos são relevantes e quais não são (no que diz respeito a uma dada consulta do usuário).

Salton e McGill (1983), em um de seus trabalhos, apresentam a visão da área de RI como um problema de *clustering*. O autor considera que os documentos são uma coleção C de objetos e que a consulta do usuário seja uma (vaga) especificação de um conjunto A de objetos. Sendo assim, o problema de RI pode ser reduzido a determinar quais documentos são do conjunto A e quais não são (i.e., o problema de RI pode ser visto como um problema de *clustering*).

Segundo Baeza-Yates (1999), em um problema de *clustering*, duas principais questões precisam ser resolvidas. Primeiro, alguém precisa definir quais as características que melhor descrevem os objetos do conjunto A . Segundo, alguém precisa determinar quais as características que melhor distinguem do conjunto A dos demais da coleção C . Do primeiro conjunto de características proveem a

quantificação para a similaridade *intracluster*, enquanto o segundo provê a quantificação para a dissimilaridade *intercluster*. Os algoritmos de *clustering* que obtêm mais sucesso são aqueles que tentam balancear esses dois efeitos.

No modelo vetorial, a similaridade *intraclustering* é quantificada pela frequência crua dos termos k_i no documento d_j . Essa frequência de termos é usualmente referida como o *fator tf* (*tf-factor*) e provê uma medida de quão bem esses termos descrevem o conteúdo do documento (i.e., caracterização intradocumento). Ademais, a dissimilaridade *intercluster* é quantificada pela frequência inversa de um termo k_i entre os documentos da coleção. Esse fator é geralmente referido como a frequência inversa do documento ou o *fator idf* (*idf-factor*). A motivação para o uso do fator *idf* é que termos que aparecem em muitos documentos não são muito úteis para distinguir um documento relevante de um não relevante. Assim como um bom algoritmo de *clustering*, os esquemas mais efetivos de balanceamento de termos (atribuição de pesos) para RI tentam balancear esses dois efeitos.

Definição: considere que N seja o número total de documentos no sistema e n_i o número de documentos nos quais o termo indexado k_i aparece. Considere que $freq_{i,j}$ seja a frequência crua do termo k_i no documento d_j (i.e., o número de vezes que o termo k_i é mencionado no texto do documento d_j). Então, a frequência normalizada $f_{i,j}$ do termo k_i no documento d_j é dada por:

$$f_{i,j} = \frac{freq_{i,j}}{\max_l freq_{j,l}}$$

Onde o máximo é computado sobre todos os termos mencionados no texto do documento d_j . Se o termo k_i não aparecer no documento d_j então $f_{i,j} = 0$. Ademais, considere idf_i , a frequência inversa de documento para k_i , sendo dada por:

$$idf_i = \log \frac{N}{n_i}$$

Segundo Baeza-Yates (1999), os melhores esquemas de atribuição de pesos a termos utilizam pesos atribuídos por:

$$w_{i,j} = f_{i,j} \times idf_i$$

Ou por variações dessa fórmula. Tais estratégias de atribuição de pesos são conhecidas como esquemas *tf-idf*.

As principais vantagens do modelo vetorial são:

- seu esquema de atribuição de pesos, que melhora a qualidade da recuperação;
- sua estratégia de correspondência parcial, que possibilita a recuperação de documentos que se *aproximam* da consulta;
- sua fórmula de ranque, baseada em cosseno, que classifica os documentos de acordo com seu grau de similaridade.

Apesar de sua simplicidade, o modelo vetorial é uma estratégia de ranque bastante eficaz. Por sua popularidade, tem sido comparado com vários modelos alternativos de RI. Em geral, o modelo vetorial, no entanto, tem-se mostrado superior ou tão bom quanto os modelos alternativos conhecidos.

4.2.2.3 LSI

Conforme citado anteriormente, resumir o conteúdo dos documentos e consultas por um conjunto de vetores de termos pode levar à diminuição do desempenho da busca por conta de dois efeitos. Primeiramente, muitos documentos sem relação contextual podem ser incluídos no conjunto da resposta. Segundo, documentos relevantes que não são indexados por nenhuma das palavras-chave da consulta não são recuperados. Para Baeza-Yates (1999: 44), ambos os efeitos são gerados pela vagância dos processos de recuperação de informações baseados em conjuntos de palavras-chave.

Para o mesmo autor, as idéias contidas em um texto são mais relacionadas aos conceitos descritos nele do que com os termos indexados utilizados em sua descrição. Então, o processo de encontrar documentos associados a uma dada consulta pode ser baseado em buscar conceitos em vez de termos indexados. Essa idéia possibilita a recuperação de documentos mesmo quando estes não são indexados por termos de uma consulta. Por exemplo, um documento poderia ser recuperado porque ele compartilha conceitos com outro documento que é relevante

para uma dada consulta.

A idéia principal do modelo LSI (*latent semantic indexing*) (ROSARIO, 2000) é mapear cada vetor de documento e consulta em um espaço dimensional reduzido, que é associado aos conceitos. Isso se dá pelo mapeamento dos vetores de termos indexados nesse espaço dimensional reduzido. Segundo Baeza-Yates (1999, 44), a alegação é que a recuperação em um espaço reduzido pode ser melhor que recuperar espaço dos termos indexados. O autor provê a seguinte definição matemática para o modelo LSI:

Definição: assim como antes, considere t como sendo o número de termos indexados na coleção e que N seja o número total de documentos. Defina $\vec{M} = M_{ij}$ como uma matriz de associação entre termos e documentos com t linhas e N colunas. Para cada elemento M_{ij} dessa matriz será associado um peso $w_{i,j}$ ao par $[k_i, d_j]$. Esse peso $w_{i,j}$ pode ser gerado usando-se a técnica de atribuição de pesos *tf-idf*, comumente utilizada no modelo de espaço vetorial.

Na matriz citada acima, os elementos (i, j) quantificam o relacionamento entre os documentos d_i e d_j .

O modelo LSI propõe a decomposição da matriz de associação \vec{M} em três componentes, utilizando a decomposição de valor singular, como segue:

$$\vec{M} = \vec{K} \vec{S} \vec{D}^t$$

A matriz \vec{K} é a matriz dos autovetores derivados da matriz de correlações termo a termo dada por $\vec{M} \vec{M}^t$. A matriz \vec{D}^t é a matriz de autovetores derivada da transposta da matriz dada por $\vec{M}^t \vec{M}$. A matriz \vec{S} é uma matriz diagonal $r \times r$ de valores singulares, em que $r = \min(t, N)$ é o ranque de \vec{M} .

Considerando agora que apenas os s maiores valores singulares de \vec{S} são mantidos, juntamente com as respectivas colunas em \vec{K} e \vec{D}^t (i.e., os valores singulares restantes de \vec{S} são excluídos). A matriz \vec{M}_s é a matriz de ranque s que mais se aproxima da matriz original \vec{M} , no sentido da média dos mínimos quadrados. Essa matriz é dada por:

$$\vec{M}_s = \vec{K}_s \vec{S}_s \vec{D}_s^t$$

Onde s , $s < r$, é a dimensionalidade de um espaço conceitual reduzido. A

seleção de um valor para s tenta balancear dois efeitos opostos. Primeiro, o s deveria ser grande o suficiente para possibilitar o ajuste de toda a estrutura aos dados reais. Segundo, o s deveria ser pequeno o suficiente para possibilitar a filtragem de todo o conteúdo não relevante.

O relacionamento entre qualquer par de documentos no espaço reduzido de dimensionalidade s pode ser obtido pela matriz $\vec{M}_s^t \vec{M}_s$, dada por:

$$\begin{aligned}\vec{M}_s^t \vec{M}_s &= (\vec{K}_s \vec{S}_s \vec{D}_s^t)^t \vec{K}_s \vec{S}_s \vec{D}_s^t \\ &= \vec{D}_s \vec{S}_s \vec{K}_s^t \vec{K}_s \vec{S}_s \vec{D}_s^t \\ &= \vec{D}_s \vec{S}_s \vec{S}_s^t \vec{D}_s^t \\ &= (\vec{D}_s \vec{S}_s) (\vec{S}_s \vec{D}_s^t)^t\end{aligned}$$

Para classificar um documento em relação a uma consulta do usuário, simplesmente se modela a consulta como um *pseudodocumento* na matriz \vec{M} termo-documento original. Assume-se que a consulta é modelada como o documento de número 0. Então, a primeira linha da matriz $\vec{M}_s^t \vec{M}_s$ provê o ranque de todos os documentos em relação a essa consulta.

Uma das principais críticas ao uso da decomposição em valores singulares é o seu alto custo computacional quando se utilizam grandes quantidades de documento. Segundo Baeza-Yates (1999: 45), o modelo LSI introduziu uma conceituação interessante para o problema de Recuperação de Informações baseado na teoria de decomposição de valores singulares, tendo seu valor como um novo *framework* teórico. Se ele funciona de forma superior com coleções gerais ainda está por ser verificado.

4.2.2.4 Probabilístico

O modelo probabilístico aborda a Recuperação de Informações como um problema estatístico. A idéia fundamental é que, dada uma consulta do usuário, existe um conjunto de documentos o qual contem exatamente os documentos relevantes e nenhum outro. Vamos chamar esse conjunto de documentos de conjunto resposta ideal. Portanto, dada a descrição desse conjunto ideal, não teríamos problemas em recuperar esses documentos.

Então, pode-se pensar no processo de busca como um processo de especificação das propriedades de um conjunto resposta ideal (o que é análogo a interpretação do problema de *clustering*). O problema é descobrir exatamente o que essas propriedades são. Tudo que se sabe é que elas são termos indexados dos quais a semântica pode ser utilizada para caracterizar essas propriedades.

O modelo probabilístico é baseado no seguinte pressuposto: dada uma consulta q e um documento d_j em uma coleção, o modelo probabilístico tenta estimar a probabilidade do usuário considerar o documento d_j interessante (i.e. relevante). O modelo assume que essa probabilidade de relevância depende apenas da representação da consulta e do documento. Além disso, o modelo assume que existe um subconjunto de documentos que o usuário prefere definir como a resposta para a consulta q . Esse conjunto de resposta ideal é chamado de R e deve maximizar a probabilidade global de relevância para o usuário. Documentos pertencentes ao conjunto R são ditos relevantes para a consulta. Documentos fora dele são ditos não relevantes.

Segundo Baeza-Yates (1999), essa pressuposição é bastante problemática porque ela não define explicitamente como computar as probabilidades de relevância. Na verdade, nem mesmo o espaço amostral utilizado para a definição de tais probabilidades é dado.

Dada uma consulta q , o modelo probabilístico associa a cada documento d_j , como uma medida de sua similaridade a consulta, a razão $P(d_j \text{ relevante para } q)/P(d_j \text{ não relevante para } q)$ que computa as probabilidades do documento d_j ser relevante para a consulta q . Tomando as probabilidades de relevância como o ranque

minimiza-se a probabilidade de um julgamento errôneo.

Definição: para o modelo probabilístico, o peso dos termos indexados são todos binários, isto é, $w_{i,j} \in \{0, 1\}$, $w_{i,q} \in \{0, 1\}$. A consulta q é um subconjunto de termos indexados. Consideremos R como sendo o conjunto de documentos conhecidos (ou inicialmente imaginados) como relevantes. Considere \bar{R} sendo o complemento de R (i.e., o conjunto de documentos não relevantes). Considere $P(R|\vec{d}_j)$ com sendo a probabilidade do documento d_j ser relevante para a consulta q e $P(\bar{R}|\vec{d}_j)$ a probabilidade de d_j não ser relevante para q . então, a similaridade $sim(d_j, q)$ do documento d_j para a consulta q é definida pela razão:

$$sim(d_j, q) = \frac{P(R|\vec{d}_j)}{P(\bar{R}|\vec{d}_j)}$$

A maior vantagem do modelo probabilístico, teoricamente, é que os documentos são ranqueados em ordem decrescente de suas probabilidades de serem relevantes. As desvantagens incluem: (1) a necessidade de adivinhar a separação inicial de documentos em conjuntos relevantes e não-relevantes; (2) o fato de o método não considerar a frequência com a qual um termo indexado ocorre dentro de um documento (i.e., todos os pesos são binários); e (3) a adoção da suposição de independência entre os termos indexados. Entretanto, como sugere Baeza-Yates (1999), não está claro se a suposição de independência entre termos indexados é uma má idéia em situações práticas.

A seguir será apresentado um modelo pouco citado na literatura, mas igualmente interessante. As Redes de Grafos Conceituais aparecem como uma alternativa com custo computacional baixo e fácil de ser implementada.

4.2.2.5 CNG

As *Context Network Graphs* (CNG) constituem uma técnica que serve tanto para classificar documentos quanto para expandir consultas. Conforme relata Maciej

Ceglowski e colegas (2003), CNG é um algoritmo de busca que pode recuperar coleções de documentos potencialmente grandes. O autor sugere que a técnica obtém resultados comparáveis às LSI e com um custo computacional menor.

As CNGs são baseadas na criação de um grafo bipartido de nós de termos e documentos, conectando um nó termo a um nó do documento. Neste modelo, cada termo está ligado a todos os documentos em que o mesmo aparece, e cada documento tem um link para cada um de seus termos. Os valores de frequências ponderadas de outros modelos correspondem aos pesos colocados nas extremidades do grafo.

Como exemplo, é apresentada uma coleção de documentos em miniatura na Tabela 5, e é associada com a lista de termos da Tabela 6:

Tabela 5: Exemplo de documentos associadas a nós.

Nó	Conteúdo do Documento
1	Gelo glacial, muitas vezes aparece em azul.
2	As geleiras são compostas de neve caída.
3	Neve granulada é um estado intermediário entre a neve e o gelo glacial.
4	As plataformas de gelo ocorrem quando as calotas polares se desprendem sobre o mar.
5	Geleiras e das calotas polares criam icebergs no mar.
6	A neve granulada tem a metade da densidade da água do mar.
7	Icebergs são blocos de gelo glacial flutuando sobre a água do mar.

Abaixo na Tabela 6 a lista de termos relevantes encontrados na Tabela 5 associados às suas ocorrências na coleção de documentos.

Tabela 6: Lista rotulada de termos extraída da Tabela 5

Nó	Termo	Ocorrências
a	Gelo glacial	3
b	Gelo	5
c	Geleira	2
d	Neve	2
e	Neve Granulada	2
f	Calota polar	2
g	Mar	3
h	Água	2
i	Iceberg	2
j	Calota	2

Baseado nos dados acima, a Figura 9 apresenta a ligação entre os documentos da coleção, os termos e suas ocorrências.

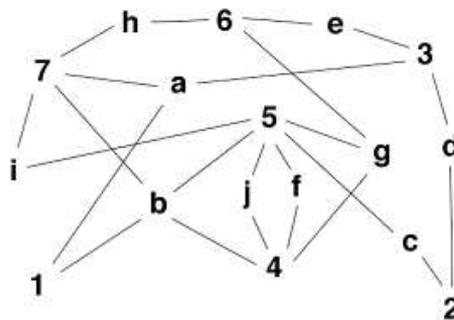


Figura 9: CNG Exemplo indicando conexões entre os documentos da Tabela 5 e os termos da Tabela 6.

Conforme observado na demonstração, não existe codificação de nenhuma informação sobre gramática ou relacionamentos hierárquicos entre termos. Sua estrutura é determinada unicamente pela co-ocorrência de termos na coleção.

Cada relacionamento do gráfico tem uma força que lhe é atribuída, e cuja magnitude depende da nossa escolha do esquema de atribuição de pesos local e global utilizado na geração da Matriz de Termos e Documentos. A única restrição sobre os regimes de ponderação é a de que todos os pesos das bordas estejam no intervalo (0,1).

Segundo Maciej Ceglowski e colegas (2003), podemos consultar a coleção representada por esse grafo através da energização de um nó de consulta, permitindo que a energia se propague para os outros nós ao longo das bordas do grafo com base em um conjunto de regras simples. A energia total depositada em qualquer nó do grafo dependerá tanto do número de caminhos entre ele e o nó de consulta, quanto da força relativa das ligações ao longo desses caminhos. Isso corresponde à intuição de que os documentos que compartilham muitos termos raros são susceptíveis a serem semanticamente relacionados. Também apresenta o mesmo tipo de *recall* aprimorado observado no modelo LSI, uma vez que uma consulta em por determinada palavra-chave pode chegar a um documento que não contém a palavra em si, mas esteja que intimamente ligado a outros documentos que a contem.

Para o exemplo acima, uma pesquisa por 'iceberg' iria começar ativando o nó correspondente ao termo da consulta, neste caso, nó i . Para tanto, é atribuído a este nó uma energia padrão de consulta E , que é distribuída entre nós vizinhos de acordo com o seguinte algoritmo:

```

1 procedure energize( energy E, node  $n_k$  ) {
2   energy( $n_k$ ) := energy( $n_k$ ) + E
3    $E' := E / \text{degree of } n_k$ 
4   if (  $E' > T$  ) {
5     for each node  $n_j$  in  $N_k$  {
6        $E'' := E' * e_{jk}$ 
7       energize(  $E''$ ,  $n_j$  )
8     }
9   }
10 }
```

Na próxima seção será apresentada uma visão da área de Recuperação de Informações mais centrada no usuário, mais preocupada com as tarefas com as quais eles estão envolvidos. A seguir trataremos sobre RI em Contexto, bem como, suas variáveis e dimensões.

4.2.3 Recuperação de Informação baseada em Contexto

Nos últimos 20 anos, a área de Recuperação de Informação (RI) cresceu além de seus objetivos iniciais, que eram indexar textos e buscar documentos relevantes dentro de uma coleção (YATES-RIBEIRO, 1999). Atualmente, as pesquisas em RI incluem também: modelagem, classificação de documentos e categorização, arquiteturas de sistemas, interfaces com o usuário, visualização de dados, filtragem, linguagens, etc.

Tradicionalmente, a medida de avaliação da qualidade desses mecanismos é baseada apenas em medidas orientadas ao sistema (*recall e precision*⁴) e frequentemente irrelevantes para o usuário. Esses parâmetros não mensuram, por exemplo, o sucesso do usuário no processo de solução de seus problemas (INGWERSEN, 2005a).

Peter Ingwersen e colegas, em sua mais recente obra (idem, 2005a), o livro *The Turn*, fazem uma crítica aos modelos de avaliação dos mecanismos de busca, sugerindo uma extensão dos mesmos no sentido de englobar também as tarefas de trabalho do usuário (*work tasks*), além de concepções de mais alto nível. Entretanto, isso demanda que características – representando pessoas e suas interpretações/percepções, tarefas de trabalho, interações, situações e contexto – sejam de alguma forma incorporadas no modelo. A simples incorporação de algumas dessas características nos algoritmos de recuperação por si só podem se mostrar grandes desafios.

Segundo Ingwersen (2005b), as pesquisas em RI estão tendendo para ambientes multimídias, com independência linguística, e de modo, mas sem considerar o contexto. Entretanto, a recuperação dessas informações vai depender de fatores como tempo, lugar, histórico de interação, tarefa atual do usuário, além de outros fatores que não podem ser medidos explicitamente, mas apenas implicitamente nos ambientes de RI. A essas informações é dado o nome de contexto. A informação contextual pode ser usada efetivamente para auxiliar na recuperação de informações através da redução da complexidade do processo de

⁴ *Recall*: fração dentre os documentos relevantes que foi recuperada; *Precision*: fração dos documentos recuperados que é relevante.

recuperação.

Recuperação de Informações Interativa (RII) é definida como sendo o processo interativo de comunicação que ocorre durante a recuperação de informação, envolvendo os principais participantes do processo de RI, isto é, o usuário, o mediador e o sistema de RI (INGWERSEN, 1992).

Então, Contexto implica RI de forma Interativa, considerando as dimensionalidades dos contextos em associação com os motores e sistemas de RI. Essa dimensionalidade varia conforme as características de conteúdo tradicionais que existem dentro e entre os objetos de informação. Como exemplos de dimensões podem ser considerados: as palavras próximas a parágrafos e hiperlinks, os movimentos do mouse durante a seção de interativa, etc.

Ao investigar RI em Contexto, alguém poderia focar em pelo menos uma dimensão contextual, explorando sua natureza e questionando as suas implicações da seguinte forma:

1. Quais os elementos de contexto que são potencialmente importantes para a recuperação de informações? (Ver Figura 10)
 - a. Trabalho ou tarefas diárias ou características do interesse;
 - b. característica do usuário;
 - c. características da interação;
 - d. características do sistema;
 - e. características dos documentos;
 - f. características físicas e ambientais;
 - g. características temporais.
2. Quais dessas características ou elementos são, ou podem ser úteis na melhoria da recuperação de informações? Como fazer isso?
 - a. Isso depende da situação;
 - b. também depende do objetivo do usuário;
 - c. depende da criação de métodos de investigação, de testes de mesa e de medidas e técnicas de avaliação, assim como;

- d. depende ainda da exploração da operacionalização de aspectos de contexto.

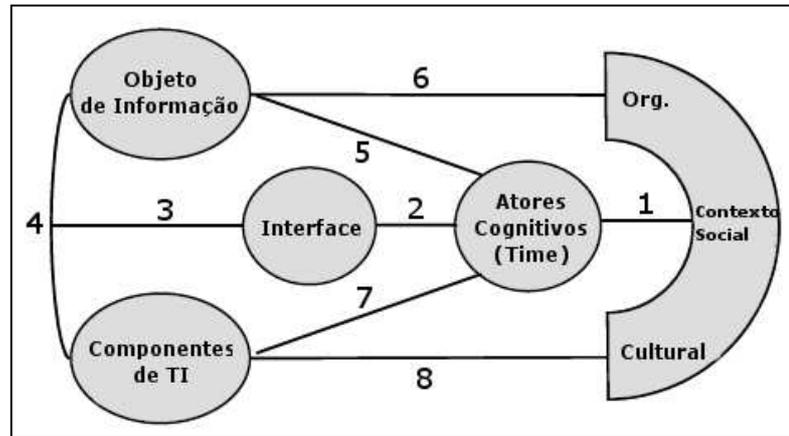


Figura 10: Framework cognitivo geral da busca e recuperação de informações.

Os números no *framework* basicamente mostram o processo de interação (1-4), como a interação social (1), ou se referem aos diferentes tipos de produção e transformação de cognição ou influência cognitiva (5-8).

O *framework* enfatiza os processos de informação que são executados durante a B&RI em contexto no decorrer do tempo. Em primeiro lugar, o processo de interação social (1) ocorre entre atores e seu contexto sócio-cultural e organizacional no decorrer do tempo. A interação social pode instigar atividades de B&RI, e pode também fazer parte do seu cumprimento. Em segundo, a interação com a informação ocorre entre Ator Cognitivo, Componentes de Tecnologia da Informação (TI) e Objeto de Informação via Interfaces (2/3). Os últimos dois componentes interagem verticalmente (4) e constituem o núcleo do sistema de informação. Por último, as transformações cognitivas e emocionais e a geração de informações em potencial podem ocorrer na medida exigida por cada ator individualmente (5/7), assim como, a partir dos contextos sociais, culturais ou organizacionais, através dos Componentes de TI e dos Objetos de Informação (6/8) no decorrer do tempo.

Isso implica uma influência sobre o Comportamento Informacional dos atores – e, por conseguinte, sobre as estruturas cognitivo-emocionais que os representa. O impacto implica que o ator pode colaborar em times – coletivamente – e adaptar-se ao seu entorno.

Peter Ingwersen (2005b) propõe nove grandes classes de variáveis que interagem no processo de B&RI, as quais podem ser chamadas também de dimensões. Essas dimensões estão ilustradas na Figura 10:

1. Dimensão da Tarefa de Trabalho Natural: tarefa de trabalho⁵, organização (social) do trabalho ou atividades cotidianas, colaboração e o ambiente cultural e sistêmico;
2. Tarefa de Pesquisa natural, i.e., tarefa de busca e recuperação de informação, no entendimento das organizações;
3. Dimensão do Ator: o conhecimento declarativo do ator e competências procedimentais;
4. Dimensão da Tarefa de Trabalho Percebida: a percepção do ator relativa à tarefa de trabalho;
5. Percepção da Tarefa de Pesquisa: a percepção do ator sobre a tarefa de pesquisa, incluindo tipos de necessidades de informações relativas à tarefa e ao desempenho do processo, às emoções;
6. Dimensão do Documento: documento, gêneros e coleções em várias linguagens e meios de comunicação, que podem conter informações relevantes para a tarefa na percepção do ator;
7. Dimensão dos Componentes de TI: as representações de documentos/informações e as necessidades de informação, ferramentas e suporte para formulação da consulta e métodos de correspondência.
8. Dimensão da Interface com os Componentes de TI: ferramentas de visualização e apresentação;
9. Dimensão do Acesso e Interação: as estratégias de acesso à informação e a interação entre o ator e a interface (tanto no contexto social como no sistêmico).

Cada uma dessas dimensões é complexa e contém um número muito grande

⁵ A noção de “Tarefa de trabalho” implica também em tarefas não relacionadas ao trabalho e da vida cotidiana e/ou interesses.

de variáveis, que por sua vez têm uma quantidade considerável de valores possíveis. Um detalhamento maior sobre essas variáveis pode ser encontrado em (INGWERSEN, 2005b).

Para cada componente do Framework Cognitivo de Busca e Recuperação de Informações, existem objetos representativos que são ligados a alguma mídia. Esses objetos são, por exemplo, entidades de *software* nos Componentes de TI e nas interfaces, ou os documentos (objetos de informação). Dentro de cada objeto existe uma série de elementos contextuais: as estruturas intra-objetos, a dimensão de contexto (1) (Figura 11). Por exemplo, em um objeto de informação, as suas figuras são contextuais para o texto ao seu redor ou outras estruturas anexas a ela, e vice-versa. Os parágrafos servem como contexto para suas próprias sentenças e palavras, e assim por diante.

Os próprios objetos são contextuais entre si – formando relacionamentos ou redes interobjetos – dimensão contextual (2) (Figura 11). Propriedades dos documentos, como referências ou links a outros objetos de informação, assim como as citações, são consideradas provedoras de contexto para o conteúdo de outros objetos.

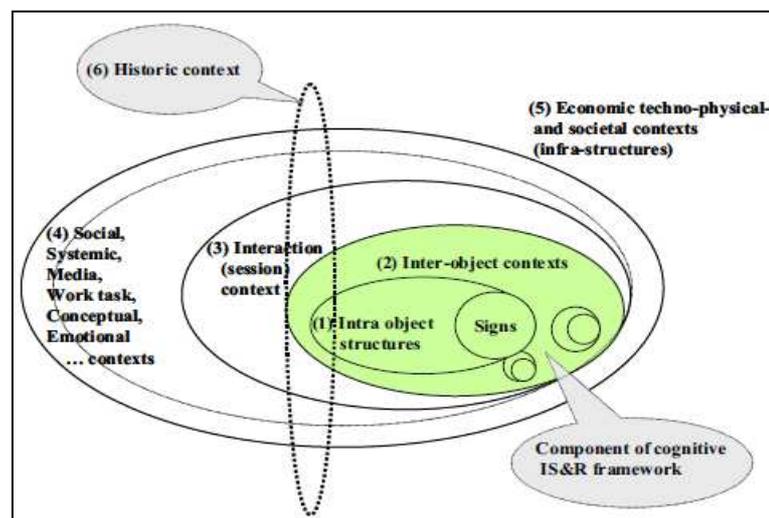


Figura 11: Modelo Aninhado das Dimensões do Contexto em RI (INGWERSEN, 2005b: 281)

O tipo de componente do *Framework* Cognitivo de B&RI colocado como o cerne do Modelo Aninhado, as dimensões (1-2), determinam o tipo de contexto da (seção de) interação (3) e dos componentes restantes do modelo, que serve como uma dimensão de contexto maior (4). Considerando o usuário como o cerne, a

interação consiste tanto em interação social quanto em atividades de RI interativas. Por sua vez, estas são encaixadas no contexto dos outros componentes do modelo (dimensão contextual 4), tais como as sócio-organizacionais e sistêmicas (Componentes de TI, interface e documentos): o primeiro, contendo tarefas reais e relacionadas ao trabalho e à vida cotidiana; e o segundo, sendo dependente da mídia e do domínio. Entretanto, no caso da interface ser o componente central do modelo (Figura 11), a interação (dimensão contextual 3) significa a seção de recuperação. Isso inclui as interações internas do sistema tais como os processos de expansão de consultas: os componentes de TI e objetos de informação do framework, assim como os atores da busca com suas características conceitual-emocionais, adjacentes da dimensão contextual 4 (o círculo interno). O contexto atual, o mais remoto sócio-organizacional, constitui a dimensão 4 (o círculo externo). Todos os elementos aninhados são influenciados por infra-estruturas sociais dominantes (5). Em toda essa estratificação opera uma dimensão adicional, ou seja, a do contexto das experiências históricas (6) de todos os participantes, formando as suas expectativas. Portanto, todos os processos e atividades de B&RI estão sob influência desse tipo de contexto temporal.

Como já foi descrito nesta seção, o trabalho de Inwersen (2005b) define um Framework Conceitual para o problema da B&RI em contexto, e o mesmo autor, ainda afirma que um sistema hipotético pode considerar apenas algumas variáveis (ou classes) dessas como sendo contexto. Sendo assim, esses conceitos poderiam ser especificados na forma de uma conceituação compartilhada entre todos os componentes do sistema. Na seção que se segue, trataremos de definir e aproximar o uso de ontologias nos sistemas de RI.

4.3 Ontologias

Segundo a etimologia da palavra ontologia, ela tem origem grega e é composta de dois outros termos: *ontos*, que deriva de dois substantivos gregos, *tà onta* (“os bens e as coisas realmente possuídas por alguém” e “as coisas realmente

existentes”); e “*logos*”, estudo. Marilena Chaui (2006) estuda a definição etimológica da palavra ontologia para a filosofia e conclui que significa:

estudo ou conhecimento do Ser, dos entes ou das coisas tais como são em si mesmas, real e verdadeiramente, correspondendo ao que Aristóteles chamara de filosofia primeira, isto é, o estudo do Ser enquanto Ser (CHAUI, 2006: 183).

Apesar de ser um conceito que se confunde com o que é utilizado na filosofia, o termo ontologia tem um sentido “especial” na área da organização da informação, diferente daquele tradicional adotado pela filosofia.

Desde o início dos anos noventa, as ontologias vêm se tornando um tópico de pesquisa popular e bastante estudado por várias comunidades, incluindo Engenharia do Conhecimento, Web Semântica, processamento de linguagem natural e representação do conhecimento. Recentemente, a noção de ontologia vem se tornando comum em campos como integração inteligente de informações, sistemas cooperativos de informações, recuperação de informações, comércio eletrônico e gestão do conhecimento.

A razão da popularização das ontologias está relacionada a o que elas prometem: um entendimento comum e compartilhado de algum domínio que pode ser comunicado entre pessoas e sistemas de aplicação. Pelo fato de as ontologias almejarem um domínio do conhecimento consensual, o desenvolvimento delas é muitas vezes um processo cooperativo envolvendo pessoas diferentes, e possivelmente em locais diferentes.

Filósofos e engenheiros de software têm pontos de vista diferentes sobre ontologias. Em Guarino (1998), é apresentada a diferença entre ontologia no sentido filosófico e na maneira como o termo é usado pela comunidade de Inteligência Artificial. Para a engenharia de software, as ontologias descrevem uma determinada realidade com um vocabulário específico, usando um conjunto de premissas de acordo com o sentido intencional das palavras do vocabulário. No sentido filosófico, ontologia é um sistema específico de categorias que reflete uma visão específica do mundo. Gruber (1992) define uma ontologia como uma especificação explícita de uma conceituação. Guarino (1998) revê e amplia essa definição: uma ontologia é uma teoria lógica que corresponde ao significado intencional de um vocabulário formal, ou seja, um comprometimento ontológico com uma conceituação específica do mundo. Os modelos intencionais de uma linguagem lógica, usando este

vocabulário, são controlados por seu comprometimento ontológico. Esse comprometimento e a conceituação subentendida são refletidos na ontologia pela aproximação desses modelos intencionais.

Dessa forma, pode-se dizer que, para o engenheiro de software existem diversas ontologias, enquanto para o filósofo existe apenas uma, a Ontologia, com letra maiúscula. Para resolver essa questão, Smith (1998) sugere uma distinção terminológica entre uma ontologia baseada na realidade (ontologia-R) e uma ontologia epistemológica (ontologia-E). Ontologia-R é uma teoria que explica como o universo é organizado e corresponde ao mundo dos filósofos. Uma ontologia-E serve aos propósitos dos engenheiros de software e pode ser definida como uma teoria que explica como um indivíduo, um grupo, uma linguagem ou uma ciência entendem determinado domínio.

As ontologias têm diversas aplicações para a computação, tais como: gestão do conhecimento, comércio eletrônico, processamento de linguagens naturais, recuperação de informações na Web, de cunho educacional, entre outras. Para mais detalhes sobre aplicações de ontologias consultar Almeida & Bax (2003). Neste trabalho fazemos uso das ontologias para a área de recuperação de informações, que vem sendo utilizadas como mecanismo de desambiguação em consultas. Logo abaixo, a Tabela 7 apresenta projetos relacionados ao uso de ontologias em recuperação de informações:

Tabela 7 Projetos relacionados à Recuperação de Informação na Web (IDEM, 2003)

Projeto	Descrição
<i>OntoSeek</i>	Recupera informações de catálogos de produtos on-line utilizando um sistema de agentes inteligentes, um mecanismo de casamento de padrões baseado em ontologias para tratar o conteúdo e um formalismo para a representação (BORGIO, 1997).
<i>WebKB-2</i>	Permite que usuários da Web recuperem e adicionem conhecimento em uma base compartilhada. Permite a publicação de informações automaticamente recuperáveis e compatíveis com as de outros usuários (MARTIN & EKLUND, 2001).
<i>C-Web – Community Web</i>	Formaliza o conhecimento comum utilizado por comunidades da Web. A limitação é conseguir um ponto de acesso único para as várias fontes de informação das comunidades (ALEXAKI, 2002).
<i>Seal (Semantic Portal)</i>	Possibilita o desenvolvimento de portais semânticos a

	partir de abordagem baseada em ontologias. Explora o aspecto semântico por meio do fornecimento e do acesso à informação em um portal (MAEDEHE, 2001)
--	---

A Tabela 8 apresenta alguns projetos relacionados à utilização de ontologias na educação:

Tabela 8 Projetos relacionados à Educação (IDEM, 2003)

Projeto	Descrição
<i>RichODL</i>	Ambiente de aprendizado na Web desenvolvido para treinar estudantes na modelagem e simulação de ambientes dinâmicos. Ontologias são usadas para descrever o domínio físico dos sistemas modelados, além de suas correlações (ZDRAHAL, 2000).
<i>Smartrainer</i>	Sistema de treinamento automático no domínio do fornecimento de energia. Treina funcionários de empresas de energia para a recuperação de acidentes em subestações elétricas (JIN, 1997).
<i>SchoolOnto (Scholarly Ontologies Project)</i>	Biblioteca digital baseada em ontologias que possibilita interpretar domínios. Auxilia na modelagem de pesquisas dinâmicas que carecem de ferramentas para tratar inconsistências e permite discutir a contribuição do documento para a literatura da área por meio de uma rede semântica (SHUM, MOTTA & DOMINGUE, 2000).

Por enquanto, este trabalho faz um uso mais restrito de ontologias. Elas são utilizadas apenas para definir um vocabulário comum de comunicação entre os agentes do sistema. Abaixo é apresentada a ontologia utilizada pelos agentes do sistema na Figura 12.

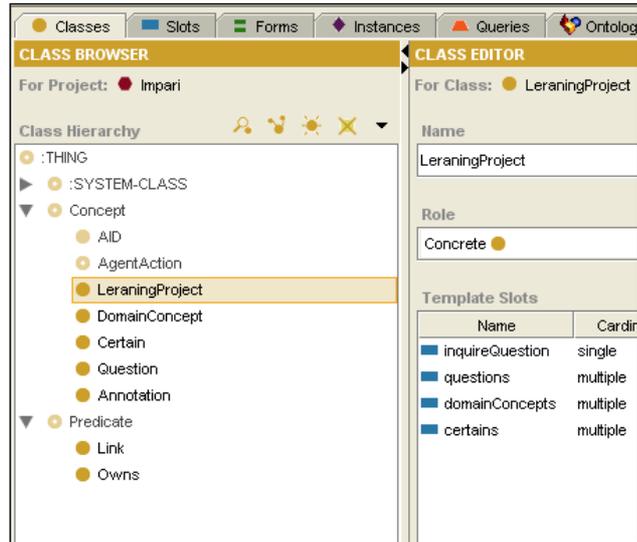


Figura 12: Ontologia dos projetos de aprendizagem.

Nas considerações finais serão apresentadas algumas idéias para estender o uso de Ontologias na forma de trabalhos futuros, buscando assim melhorar os resultados desta proposta. Nos próximo capítulo será apresentado o protótipo criado para este trabalho.

5 A Proposta de um Ambiente Mediador

Ainda que as ferramentas de busca e diretórios atualmente existentes não suportem nem tenham sido desenvolvidos para serem utilizados na Educação Continuada Autônoma (ECA), nota-se uma demanda por algo que possa ajudar as pessoas a fazer uso da Internet no sentido da continuidade do seu processo de aprendizagem. Portanto, acredita-se ser possível desenvolver um modelo de ambiente de aprendizagem mais cooperativo para a educação continuada autônoma, considerando-se os interesses explicitados (certezas, dúvidas e anotações) nas escolhas (opiniões, conclusões, observações) e nas ações (seleção, avaliação, arquivamento de documentos).

Um ambiente dessa natureza deveria ainda acompanhar o desenvolvimento do estudante de forma continuada, analisando as “pegadas” deixadas pelo usuário durante as sessões de uso do ambiente. Portanto, também faz parte deste estudo verificar se os projetos de aprendizagem podem fornecer as informações necessárias para que os agentes artificiais autônomos possam recuperar documentos em uma base.

Mas o que se espera de tais agentes? No ano de 1996, Nicholas Jennings e Michael Wooldridge (1996) apresentaram um trabalho em que discutem sobre um Gestor Pessoal de Informações. Naquele trabalho, os autores consideraram que, para obter algo dessa natureza seria necessária a construção de um assistente treinado com as seguintes tarefas:

- constantemente procurar e crivar fontes de informações disponíveis;
- conhecer o que nos interessa e o que não nos interessa, procurar temas interessantes, ou seguir links na Web de novos sites, para verificar se eles contêm algo de interesse;
- procurar metodicamente um artigo fracamente especificado em sites da Web sem se ficar ou entediado ou distraído por links para sites não relacionados;

- monitorar e-mails, classificá-los em uma ordem que seja interessante e que não acumule mensagens desnecessárias.

Certamente, no mundo real, esse tipo de assistente não existe. Mesmo que se peça ao melhor amigo, um dia ele irá dizer que tem outras coisas a fazer. É a partir daí, portanto, que surgiu a ideia de um Gestor Pessoal de Informações (GPI) e, mais especificamente, neste trabalho, de um Assistente Pessoal de Aprendizagem (Figura 13).

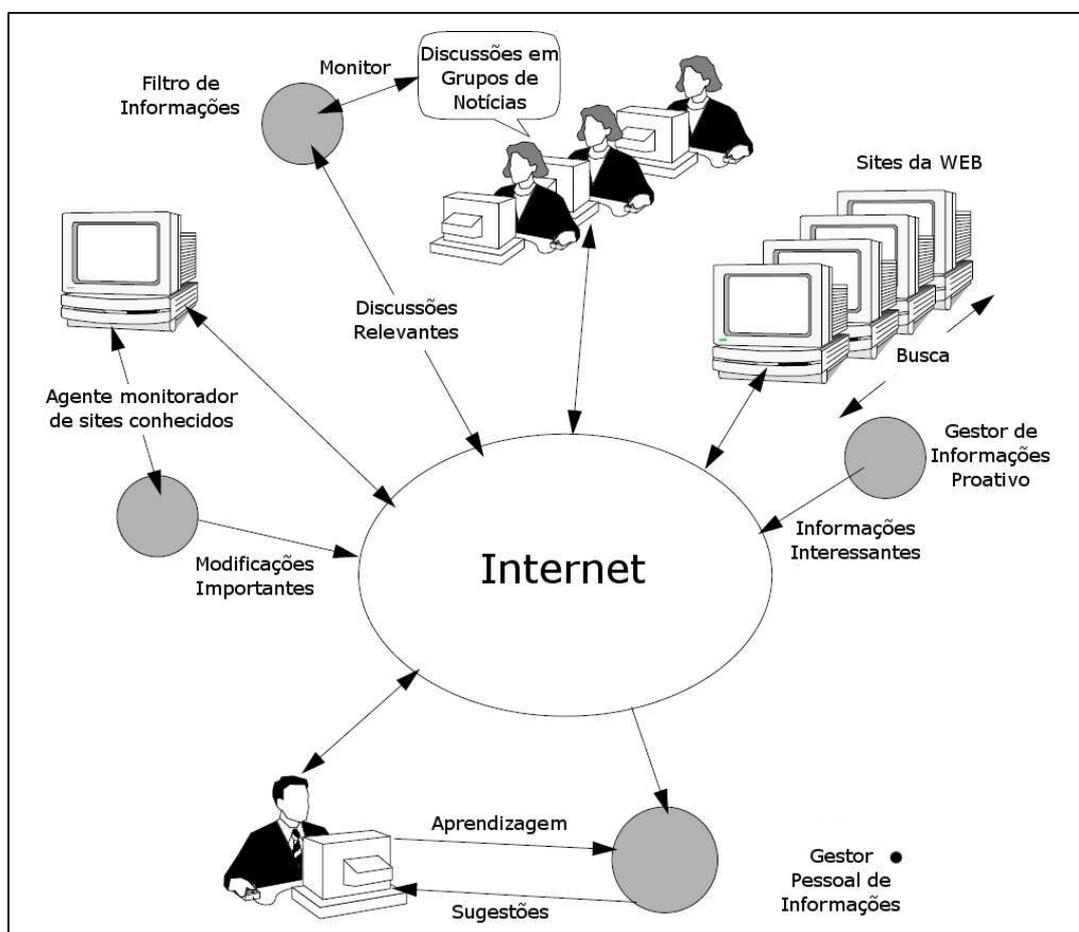


Figura 13: Modelo de um Gestor Pessoal de Informações (JENNINGS & WOOLDRIDGE, 1996)

Um Assistente Pessoal de Aprendizagem seria uma entidade ou alguém que pudesse fornecer documentos relacionados aos contextos de estudo de um indivíduo. Ele se diferenciaria de um GPI, pois levaria em consideração uma forma de trabalho estruturada, baseada na organização ou na forma de estudo de uma pedagogia específica.

Apropriando-se dessa ideia, alguém que pretende se manter atualizado em um determinado assunto poderia lançar mão de um Assistente Pessoal de Aprendizagem para auxiliar no processo de aquisição de documentos. Portanto, esse assistente poderia se responsabilizar por analisar também os projetos de aprendizagem com a finalidade de classificar os documentos encontrados conforme os interesses de estudos de um usuário.

A seguir, será apresentada a estratégia metodológica utilizada para dar cabo deste trabalho, em seguida será apresentado o protótipo criado para fins de testes bem como e a forma de manipular os Projetos de Aprendizagem dentro de um ambiente de testes.

5.1 Metodologia

Quanto aos aspectos metodológicos, será caracterizada a seguir a tipologia da pesquisa, as técnicas a serem utilizadas para a coleta de dados, o universo e amostra a serem utilizados.

5.1.1 Tipologia da Pesquisa

A pesquisa, conforme a taxonomia proposta por Vergara (2009, p.41) pode ser classificada em dois critérios básicos: quanto aos fins e quanto aos meios.

Seguindo o critério quanto aos fins, a pesquisa foi metodológica, pois utilizou instrumentos de captação e manipulação da realidade investigada para a realização dos testes do sistema e aplicada pela necessidade de resolver problemas concretos.

Conforme o critério quanto aos meios, Vergara (2009. p.42), o tipo de

investigação foi pesquisa experimental em laboratório por poder ser caracterizada como simulação em meio computacional e por permitir a análise do fenômeno sob condições determinadas.

A fundamentação teórica do trabalho foi realizada por meio de pesquisa bibliográfica baseada em escritos de fontes publicadas, sendo estas: livros, artigos científicos técnicos e profissionais, dissertações, revistas eletrônicas, acervos eletrônicos, internet, sites educacionais, sites do governo, isto é, material disponível para acesso ao público.

Vergara (2009, p.43) afirma que a pesquisa bibliográfica:

É o estudo sistematizado desenvolvido com base em material publicado em livros, revistas, jornais, redes eletrônicas, isto é, material acessível ao público em geral. Fornece instrumental analítico para qualquer outro tipo de pesquisa, mas também pode esgotar-se em si mesma.

Continuando ainda aos conceitos teóricos da autora Vergara (2009, p.43), a parte relativa aos testes comparativos entre a ferramenta, a busca nos motores disponíveis pela internet e o levantamento manual, ocorreu na forma de pesquisa de campo e utilizar-se-á de instrumentos de coleta de dados previamente definidos.

Pesquisa de campo é investigação empírica realizada no local onde ocorre ou ocorreu um fenômeno ou que dispõe de elementos para explicá-lo. Pode incluir entrevistas, aplicação de questionários, testes e observação participante ou não.

Para Marconi e Lakatos (2009, p.188) a pesquisa de campo, é aquela utilizada com o objetivo de conseguir informações e/ou conhecimentos acerca de um problema, para o qual se procura uma resposta, ou de uma hipótese, que se queira comprovar, também se pode, descobrir novos fenômenos ou as relações entre eles.

A escolha dessa metodologia tem como sustentação as particularidades do trabalho proposto.

5.1.2 Técnicas de Coleta de Dados

Por se tratar de uma pesquisa de campo, fica evidente a necessidade de se utilizar técnicas ou instrumentos que se mostrassem compatíveis com esse trabalho. Sendo estes classificados por Vergara (2009, p.51) como: a observação, o questionário, o formulário e a entrevista.

Vale ressaltar que, para o desenvolvimento de coleta de dados dessa pesquisa utilizar-se-á a aplicação de exercício investigativo, fazendo uso da Pedagogia de Projetos de Aprendizagem para os dois grupos de usuários. Esse formato se assemelha a aplicação de um formulário. O formulário neste caso foi mais efetivo devido a possibilidade de interação entre os avaliados e o aplicador do teste.

O principal instrumento de coleta aplicado aos usuários constou de uma proposta de elaboração de um Projeto de Aprendizagem com temática direcionada a área de formação dos alunos.

Como já citado anteriormente, o teste foi aplicado em dois grupos de usuário, que foram compostos por alunos de duas turmas distintas de graduação do Curso de Enfermagem. Cada turma foi dividida em equipes para confecção de PA's.

O primeiro grupo elaborou seus projetos de forma manual, sem a ferramenta desenvolvida neste trabalho, utilizando apenas editores de texto, planilha eletrônica, e consultas à internet. O segundo grupo elaborou seus projetos no ambiente ARPA, para isto foi realizada uma breve explicação do funcionamento da ferramenta. Ambos seguiram um documento norteador com um PA exemplo.

5.1.3 Universo e Amostra

Sobre Universo ou População, Vergara (2009, p. 46), prerroga que deve-se entender por população não o número de habitantes de um local, como é largamente conhecido o termo, mas um conjunto de elementos (empresas, produtos, pessoas) que possuem as características que serão objeto de estudo.

Quanto à questão da delimitação do universo, Marconi e Lakatos (2009, p.225) enfatizam que consiste em explicar que pessoas ou coisas ou fenômenos serão pesquisados, enumerando suas características comuns, como, por exemplo, sexo, faixa etária, organização a que pertencem comunidade onde vivem, etc.

5.1.4 Da Amostra Intencional

Para Marconi e Lakatos (2009, p. 225), amostra pode ser uma porção ou parcela, convenientemente selecionada do universo ou população como é um subconjunto do universo. Representa-se em expressão a amostra sendo que: n o numero de elementos da amostra, esta pode ser representada pela letra latina minúscula x , tal que $x_n = x_1; x_2; x_3; \dots x_n$ onde $x_n < X_N$ e $n \leq N$.

A amostra intencional desta coleta de dados levará em consideração como população parte dos usuários que desejam realizar a tarefa de educação continuada de forma autônoma divididos em 2 grupos com as mesmas atividades, porém utilizando ferramentas distintas, permitindo assim, o estudo avaliativo e comparativo de resultados alcançados.

Foram consideradas como amostra duas turmas do Curso de Graduação em Enfermagem da Universidade Federal do Maranhão, Campus de Imperatriz. Trata-se de alunos que já haviam concluído mais de 70% do curso e com conhecimento necessário para atuação na área, além de adequada fluência tecnológica.

5.1.5 Dados e Fontes de Informação Utilizadas

Para efeito deste trabalho, dado diz respeito a tudo que foi coletado e ainda não sofreu qualquer espécie de tratamento ou análise. Informação significa conhecimento fundamentado, resultante da análise e combinação de vários dados.

Os dados podem ser classificados, dependendo das fontes que servirão para sua coleta, em primários e secundários. Dados primários são aqueles que não foram ainda coletados, portanto estão a cargo do pesquisador. Dados secundários são aqueles que já foram coletados por outra pessoa ou estão disponíveis de alguma forma.

Na pesquisa de campo, os dados foram exclusivamente primários devido à necessidade de se elaborar e aplicar um instrumental para obter informações sobre os resultados alcançados.

Para a realização do teste, foi necessária apresentação da Pedagogia de Projetos de Aprendizagem, bem como um modelo de PA para servir de instrumento norteador durante o desenvolvimento das atividades.

5.1.6 Algumas Considerações Sobre o Trabalho

O desafio metodológico para dar conta dessa tese teve início com um convite para auxiliar no desenvolvimento do projeto AMADIS (BASSO, 2005), e vem ocorrendo de forma gradativa seguindo algumas etapas.

Inicialmente, alguns trabalhos relacionados ao desenvolvimento do projeto AMADIS foram estudados. Nesse período, foi possível verificar o embasamento pedagógico utilizado pelo projeto, englobando, assim, estudos sobre a Epistemologia Genética de Jean Piaget e sobre a Pedagogia de Projetos de Aprendizagem que inspira esta tese.

Em seguida, foi desenvolvido pelo nosso grupo de estudos um protótipo para o AMADIS com a finalidade de coletar dados na forma de Projetos de Aprendizagem. Essa etapa serviu também para a apropriação da forma de trabalhar via Internet com os projetos.

Foi percebido que isso implicaria uma visão humana com vistas à autonomia, numa sociedade em que o rápido desenvolvimento dos meios de comunicação e informação demanda uma formação continuada para sua inclusão e manutenção nesta sociedade.

A partir daí se começou a pensar em um protótipo que registrasse mais informações sobre o usuário e que pudesse oferecer ou recomendar documentos. O levantamento desses requisitos do sistema começou a ser feito a partir de uma experiência inicial, que foi a de desenvolver um Projeto de Aprendizagem na forma tradicional, ou seja, fazendo as buscas de forma não automatizada, apenas utilizando consultas *ad hoc* na Internet ou em outros repositórios de conteúdo.

Considerando o novo conjunto de requisitos, foi desenvolvido outro protótipo de interface *Web* para a catalogação de Projetos de Aprendizagem, que foi chamado de ARPA (Ambiente de Apoio a Recuperação de Informações para Projetos de Aprendizagem) (FONSECA, 2005). Este segundo protótipo já foi completamente desenvolvido em linguagem de programação Java (2009) e com suporte para a comunicação com a plataforma de agentes JADE (2009). O protótipo continua em aperfeiçoamento e será descrito em detalhes nas seções que se seguem.

5.2 O Protótipo ARPA

Os Projetos de Aprendizagem, em sua concepção original foram idealizados para serem utilizados em escolas sob a orientação de professores especialmente treinados para trabalhar a nova pedagogia. Isso fica evidenciado em Fagundes (1998: 19):

Em nossas experiências-piloto no Projeto EducaDi/CNPq, alunos não

precisavam estudar os mesmos conteúdos ao mesmo tempo. Os projetos eram diversificados, mas interdisciplinares. Havia temas que atravessavam transversalmente as atividades de todos. Cada aluno explorava melhor os conteúdos no seu tempo, segundo seu ritmo; e podia ser atendido em suas necessidades, que apareceram com maior clareza. Mas ao mesmo tempo, se conectava com outros alunos e professores, com quem tinha interesses e necessidades afins, em outros espaços/tempos diferentes – de modo síncrono, ou assíncrono. Essa troca entre parceiros proporciona uma constante atividade operatória de construção e reflexão.

Entretanto, para que um indivíduo trabalhe de forma autônoma foram necessárias algumas adaptações na metodologia de trabalho, com o intuito de organizar o conteúdo que passa a ter um volume bem maior. Basicamente, foram introduzidas duas etapas no procedimento de criação de um PA: Agrupar as certezas e dúvidas em grupos temáticos e fazer um planejamento das atividades de pesquisa. Essas duas etapas já foram exemplificadas na seção que trata especificamente dos Projetos de Aprendizagem.

A construção de uma Interface do Usuário deve levar em conta alguns fatores, principalmente quando se pensa em automatizar algo que envolve organização e catalogação de informação, como os Projetos de Aprendizagem. Portanto, este trabalho se preocupou em projetar uma interface exclusivamente para o sistema proposto, em vez de aproveitar outros Ambientes Virtuais de Aprendizagem que já implementaram a pedagogia de Projetos de Aprendizagem (Ex.: o AMADIS já citado na seção que trata dos PA's).

Do ponto de vista do projeto de interfaces do usuário, as pessoas têm habilidades e preferências muito diferentes. As diferenças importantes para Interfaces de Acesso à Informação incluem relação espacial e capacidade de memória, habilidades de raciocínio, amplitude verbal e (potencialmente) diferenças de personalidade. Idade e diferenças culturais podem contribuir para a aceitação ou rejeição de uma interface. Uma interface inovadora pode ser útil e agradável para alguns usuários e estranhas e complexas para outros. Desse modo, a concepção do software deve permitir certa flexibilidade no estilo de interação, e não se deve esperar que os novos recursos sejam igualmente úteis para todos os usuários.

Nesse sentido, foi desenvolvido um protótipo de Ambiente Virtual de Aprendizagem baseado nestas concepções. O ARPA (Ambiente de Apoio à Recuperação de Informações para Projetos de Aprendizagem) foi concebido especialmente para eventuais testes que se fizerem necessários no sistema de

Recuperação de Informações proposto para este trabalho.

Na Figura 14, são apresentados alguns dos casos de uso já identificados para uma versão funcional do ambiente ARPA.

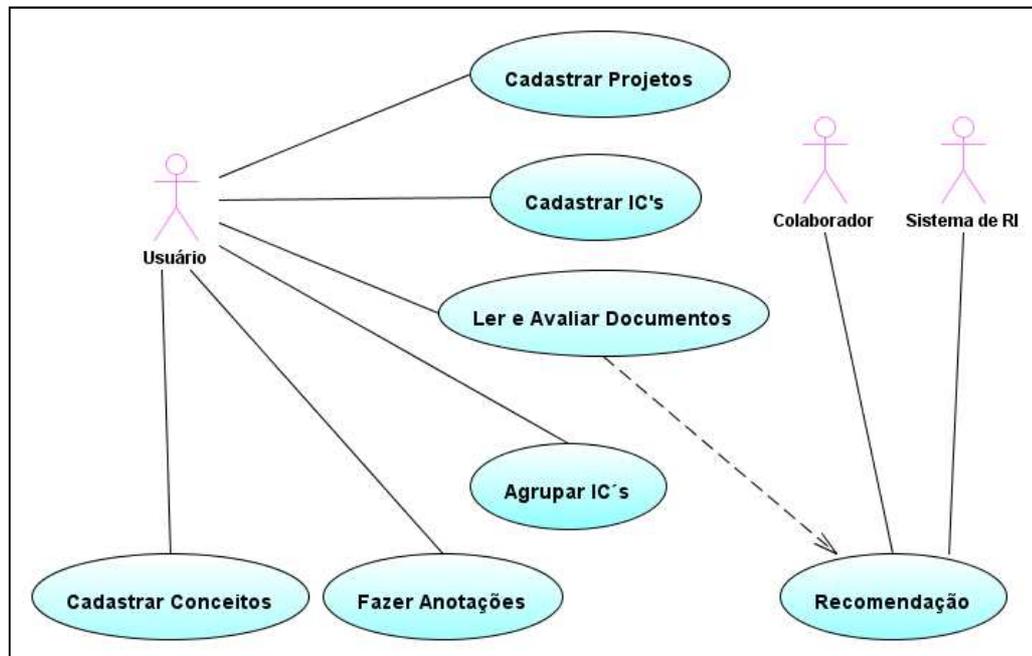


Figura 14: Casos de Uso do ARPA

Foram definidos três possíveis atores/papéis dentro do ARPA e eles são:

- o de usuário: no contexto deste trabalho, é o indivíduo que utiliza o sistema e que pretende manter-se atualizado fazendo uso da Pedagogia de Projetos de Aprendizagem;
- o de colaborador: professor, tutor, ou outro aprendiz que queira recomendar documentos para um grupo de usuários com interesses comuns;
- Sistema de Meta-busca: conjunto de agentes responsáveis por submeter a consultas os Motores de Busca da Internet a partir dos projetos de aprendizagem e recomendar documentos relevantes, provendo assim uma recomendação de documentos automatizada.

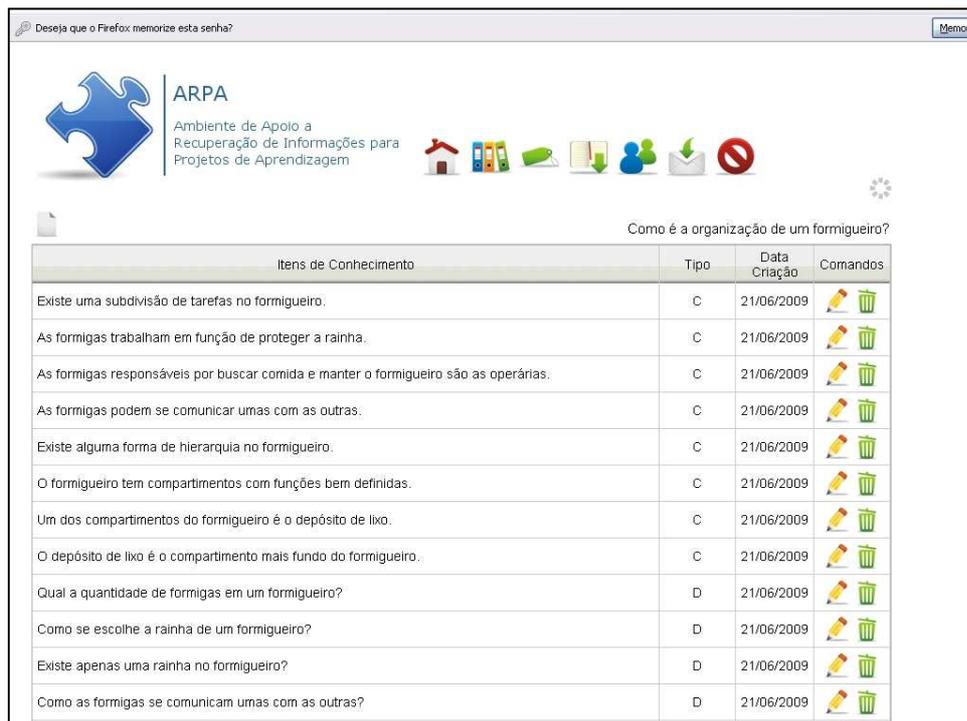
Cada ator do ARPA possui casos de uso associados conforme demonstrado na Figura 14. De forma geral, eles podem ser descritos da seguinte forma:

- cadastrar Projetos de Aprendizagem: criação e manutenção dos Projetos de Aprendizagem com os quais o aprendiz está envolvido;
- cadastrar IC's: uso que define a criação das certezas e dúvidas para um Projeto de Aprendizagem;
- agrupar IC's: o usuário do sistema poderá classificar as suas certezas e dúvidas segundo grupos temáticos, criando, assim, correlações de contexto entre os itens de conhecimento do projeto de aprendizagem;
- ler e avaliar documentos: nesse momento, o usuário poderá escolher quais documentos armazenar, fazer anotações sobre os documentos, além de correlacioná-los com certezas ou dúvidas específicas;
- fazer anotações: o usuário do sistema poderá fazer anotações sobre os documentos relacionados aos ICs;
- cadastrar conceito: o usuário também poderá criar definições para termos ou conceitos relevantes dentro de cada projeto de aprendizagem, criando assim uma espécie de glossário de termos;
- recomendação: essa função do sistema pode ser exercida por dois tipos de atores. Em um primeiro caso um Colaborador poderia sugerir documentos que ele julgue pertinentes aos interesses de algum usuário aprendiz. No segundo caso, essa recomendação poderia ser feita pelos agentes do sistema. Isso se daria por meio de um processo de identificação de contexto e depois recuperação de informações. Esse processo será mais bem explicado nos capítulos que se seguem.

A seguir são apresentadas algumas interfaces relacionadas a esses casos de uso. Para a concepção dessas interfaces, foi utilizada a linguagem de programação Java (2009) em combinação com a tecnologia JSF (*Java Server Faces*) (JSF, 2009). Outro recurso utilizado foi a biblioteca de componentes *IceFaces*, que implementa a tecnologia AJAX (*Asynchronous Javascript And XML*) (OLSON, 2007), que serve para tornar páginas Web mais interativas com o usuário, utilizando-se de

solicitações assíncronas de informações.

A interface apresentada na Figura 15 ilustra a listagem de Itens de Conhecimento para o projeto de aprendizagem do formigueiro, conforme explanado na seção que trata dos Projetos de Aprendizagem. Procurou-se fazer uma interface o mais limpa possível, permitindo acesso à manipulação completa dos Itens de Conhecimento de um Projeto de Aprendizagem.



Itens de Conhecimento	Tipo	Data Criação	Comandos
Existe uma subdivisão de tarefas no formigueiro.	C	21/06/2009	 
As formigas trabalham em função de proteger a rainha.	C	21/06/2009	 
As formigas responsáveis por buscar comida e manter o formigueiro são as operárias.	C	21/06/2009	 
As formigas podem se comunicar umas com as outras.	C	21/06/2009	 
Existe alguma forma de hierarquia no formigueiro.	C	21/06/2009	 
O formigueiro tem compartimentos com funções bem definidas.	C	21/06/2009	 
Um dos compartimentos do formigueiro é o depósito de lixo.	C	21/06/2009	 
O depósito de lixo é o compartimento mais fundo do formigueiro.	C	21/06/2009	 
Qual a quantidade de formigas em um formigueiro?	D	21/06/2009	 
Como se escolhe a rainha de um formigueiro?	D	21/06/2009	 
Existe apenas uma rainha no formigueiro?	D	21/06/2009	 
Como as formigas se comunicam umas com as outras?	D	21/06/2009	 

Figura 15: Lista de Itens de Conhecimento.

Para dar uma ideia de como funciona a inserção de informações no ARPA, apresentamos na Figura 16 a interface para cadastro das certezas ou dúvidas dos PA's.

Um conceito importante introduzido por esta tela é o de Item de Conhecimento (IC), que é uma generalização dos termos certeza e dúvida; o usuário, portanto, passa a cadastrar itens de seu conhecimento, que podem ser certezas ou dúvidas. Posteriormente, esse conceito pode ser estendido, englobando outros itens, como conceitos, anotações, etc.

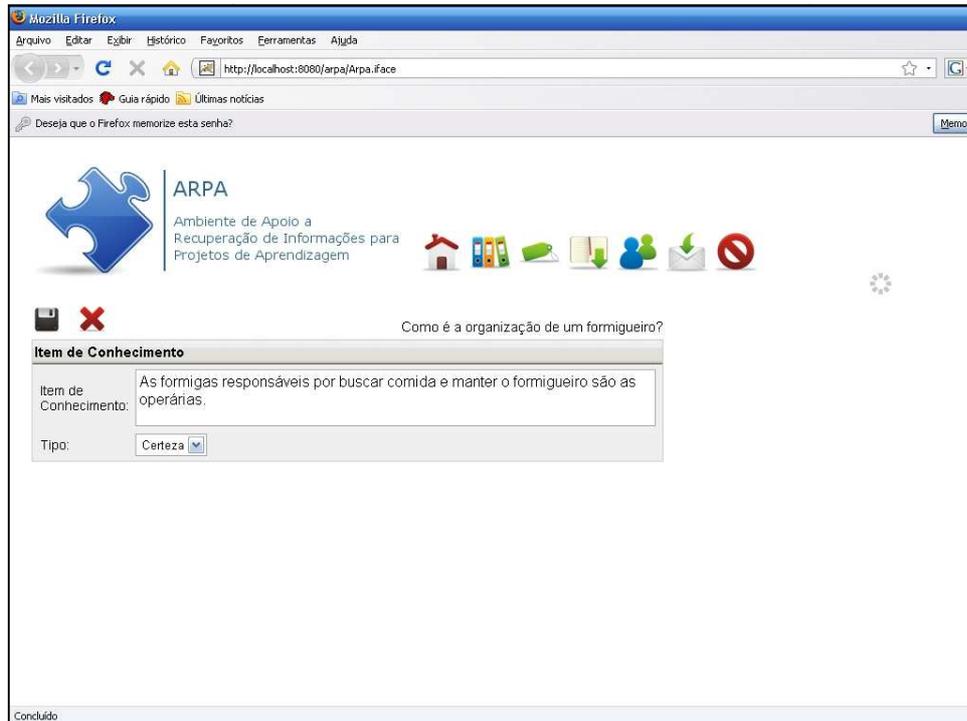


Figura 16: Cadastro de Certezas e Dúvidas

A Figura 17 ilustra o ciclo de trabalho básico do ARPA. A ilustração apresenta um conceito novo que é o de “rótulos”, na verdade o usuário pode fazer o agrupamento temático dos itens de conhecimento através de rótulos atribuídos a cada um deles.

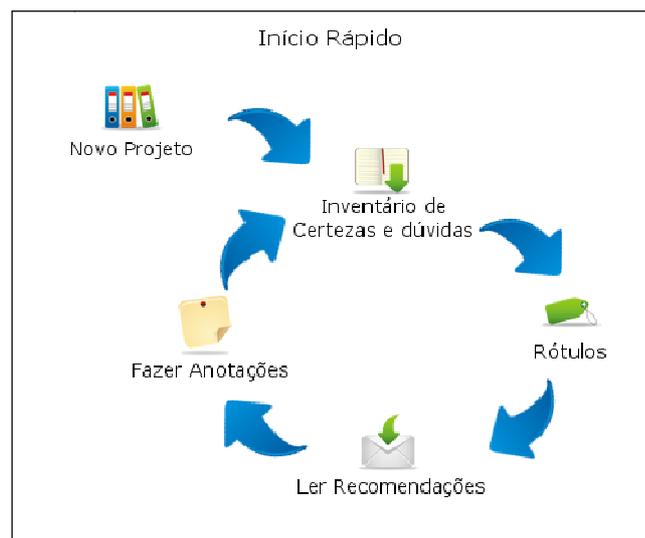


Figura 17: Ciclo de Interação do Sistema

Portanto, a atividade de um usuário em um ambiente que dê suporte a sua

atualização continuada, baseada na criação de projetos de aprendizagem, seleção de documentos, e retroalimentação do projeto de aprendizagem; forma uma atividade cíclica e não necessariamente síncronas, ou seja, o aprendiz poderia iniciar uma interação com o ambiente, modificar seu projeto de aprendizagem, avaliar textos recomendados, fazer novas anotações, etc; fechar essa seção e posteriormente retomar seu trabalho.

O trabalho dos Projetos de Aprendizagem pode ocorrer continuamente, o que sugere que os indivíduos que fizeram uso do ambiente puderam catalogar assuntos de interesse mais geral, ou seja, questões de investigação e itens de conhecimento mais abrangentes e correlacionados ao seu cotidiano, de forma que os seus projetos de aprendizagem reflitam suas preocupações diárias. Dessa forma, os agentes do sistema auxiliam no seu processo de formação autônoma continuada.

Como é demonstrado na seção que apresenta os projetos de aprendizagem, é possível também fazer um agrupamento das certezas e dúvidas em grupos temáticos, para o ARPA, denomina-se rótulos dos itens de conhecimento.

Essa estratégia é interessante do ponto de vista do usuário, pois serve para confrontar certezas e duvida semelhantes, trabalhando assim, o usuário terá a chance de visualizar itens faltosos ou redundantes, permitindo a criação de correlações de contexto entre os itens de conhecimento do projeto de aprendizagem. Do ponto de vista do protótipo, esta abordagem é útil na composição dos índices de busca para o Agente de Recuperação de Informações (ARI), conforme explica a seção que trata dos agentes do sistema.

A Figura 18, demonstra a interface de criação dos rótulos dos IC's no ambiente ARPA.

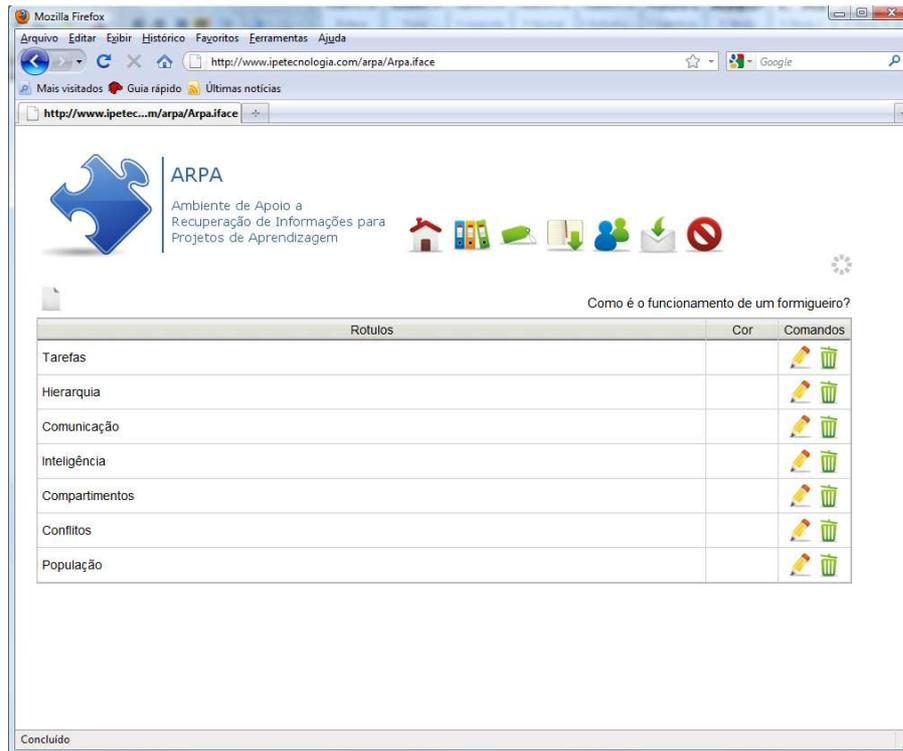


Figura 18: Listagem de Rótulos

As recomendações podem ser introduzidas no sistema por agentes humanos (outros aprendizes ou tutores do processo de aprendizagem) ou por agentes artificiais que as fariam a partir de textos encontrados na Web pelo processo de recuperação, esse último mecanismo de recomendação, será mais bem detalhado nas seções que se seguem.

Existem duas observações relevantes a serem citadas antes de darmos prosseguimento à descrição desse trabalho. São elas:

- Observação 1: o modelo apresentado não representa o que o aprendiz sabe ou não sabe, mas pode dar indícios de sua necessidade de informação. O que se tenta isolar é o conjunto de interesses específicos do indivíduo, manifestados em suas certezas, dúvidas e anotações.
- Observação 2: o que se pretende construir não é um STI (Sistema Tutor Inteligente), mas sim um Sistema de Apoio ao Trabalho Inteligente. Isso ocorre na medida em que o sistema proposto neste trabalho não tem qualquer informação sobre o domínio de conhecimento. Ele apenas facilita a vida do usuário, ajudando-o a

realizar tarefas repetitivas ou a memorizar informações importantes e identificar padrões recorrentes em documentos e em atividades realizadas pelo usuário.

São basicamente duas tarefas para a o usuário na tela do ARPA, catalogar e organizar IC's e selecionar os documentos que forem recomendados. A Figura 19 ilustra essas operações.

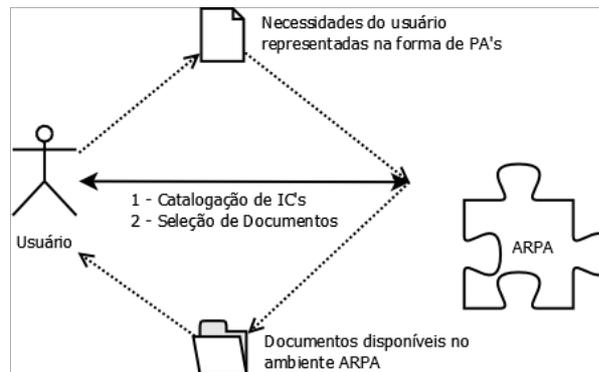


Figura 19: Tarefas do usuário no ambiente ARPA

Conforme já dito, o ARPA é um protótipo desenvolvido especialmente para fins de testes deste trabalho. Na próxima seção, serão apresentados comentários sobre o processo de seleção de documentos.

5.3 Selecionando documentos

Um dos grandes desafios deste trabalho é definir uma forma de aproveitar os dados de um Projeto de Aprendizagem, tornando essas informações consumíveis por agentes artificiais.

Para que esse problema seja solucionado, é proposto um modelo organizacional para os projetos de aprendizagem. Para tanto foi feito um estudo sobre a interação dos usuários com os Projetos de Aprendizagem através de suas Tarefas de Usuário (Recuperação e Navegação). Além disso, foi feita uma pesquisa

bibliográfica sobre modelos e estudos que tratem da seleção de documentos por parte dos usuários.

Uma certeza, no contexto do trabalho, pode ser entendida como a denotação de uma crença do aprendiz; uma certeza, portanto, pode assumir três estados para o sistema: não verificada, refutada ou validada. Da mesma forma que uma certeza, as dúvidas do aprendiz podem ter seus estados modificados, e eles variam entre: não verificada e solucionada.

É importante ressaltar que, do ponto de vista do aprendiz, a natureza do resultado de uma operação de validação/esclarecimento não é apenas uma resposta sim ou não. Na verdade, é uma síntese que associa pré-condições e conclusões. Portanto, o estudo de uma certeza ou dúvida pode gerar outras certezas ou dúvidas, e isso se dá através da interação do aprendiz com os documentos recuperados pelo usuário conforme a sua tarefa, quer seja de busca ou de navegação.

Conhecimento prévio
Inventário de certezas e dúvidas
Pesquisa bibliográfica
Seleção de documentos
Interação com os documentos (Leitura)
Novas certezas e dúvidas

Figura 20: Evolução dos Itens de Conhecimento do Aprendiz

A evolução desse inventário de certezas ou dúvidas pode ser associada a anotações. Essas anotações podem ser classificadas. Entretanto, essas classificações ainda estão em estudo, mas a princípio poderiam registrar coisas tais como: Esclarecimento, Observações Contextuais, Indícios ou Hipóteses, Conclusões, etc. Cada tipo de anotação pode ocorrer mais freqüentemente conforme ações específicas do aprendiz dentro do ambiente. Uma possibilidade que também é considerada no trabalho é deixar essas classificações livres, permitindo aos usuários dizerem o que suas anotações são.

Segundo (WANG E SOERTEL, 1998) o processo de seleção de documentos a partir de um banco de dados bibliográfico passa por três pontos de decisão até ser considerado por um usuário. No primeiro estágio, o usuário avalia uma informação resumida do documento, decidindo por obter ou não uma cópia do documento. Em seguida, a tomada de decisão é relacionada à leitura do documento, e posteriormente ele decide por citar ou não o documento. A seleção de documentos é um processo decisório no qual um usuário avalia um documento recuperado é ou não de seu interesse. A Figura 21 representa esse processo.

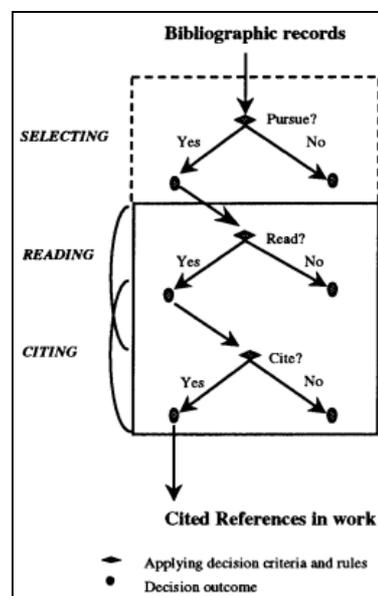


Figura 21: Processo decisório de seleção de documentos

Especificamente, o primeiro estágio do processo de seleção de documentos, proposto por Peiling Wang e Dagobert Soergel (1998), as informações avaliadas pelo usuário são os Elementos Informativos do Documento (EID), mais precisamente o título, autor, resumo, introdução, local e data de publicação, palavras-chaves, tipo do documento, etc. Essas informações são utilizadas pelo usuário no processo de escolha dos documentos seguindo alguns critérios como: atualidade, orientação, qualidade, inovação, disponibilidade, fonte, relação, etc. Esses critérios são combinados com alguns valores. Segundo o mesmo autor, existe uma analogia clara entre a compra de mercadorias e a leitura de documentos, ambas são direcionadas por necessidades; ambas envolvem tomadas de decisão baseadas em julgamentos de valor; e por último, ambas consideram o custo (dinheiro ou tempo) de aquisição.

O autor ainda cita cinco valores de consumo que podem influenciar as decisões dos consumidores:

- valor funcional: a utilidade de um documento potencialmente contribuir com uma tarefa específica atual;
- valor social: utilidade obtida a partir de uma associação com um ou mais grupos sociais específicos, ou com indivíduos como um orientador acadêmico, figuras ilustres na área, etc. Quando o usuário é associado com o autor de alguma forma, ele pode ser motivado a saber o que o documento diz;
- valor emocional: utilidade de um documento decorrente de sua capacidade de despertar sentimentos ou estados afetivos. A leitura de um documento pode causar sentimentos positivos ou negativos;
- valor epistêmico: utilidade de um documento para satisfazer um desejo de conhecimento ou de informação que é desconhecido;
- valor condicional: utilidade obtida como resultado de uma situação específica, ou conjunto de circunstâncias, enfrentadas pelo usuário.

Por exemplo, a compra de um carro japonês por razões de economia de combustível é uma consideração do valor funcional do veículo, enquanto a escolha de um Mercedes Benz pode ser baseada no valor social desse carro.

A combinação desses valores com os critérios de seleção de documentos, citados anteriormente, corroboram para a tomada de decisão do usuário em relação a aceitar, rejeitar ou ter incerteza sobre um documento. A Figura 22 apresenta o modelo de seleção de documentos proposto por Peiling Wang e Dagobert Soergel (1998).

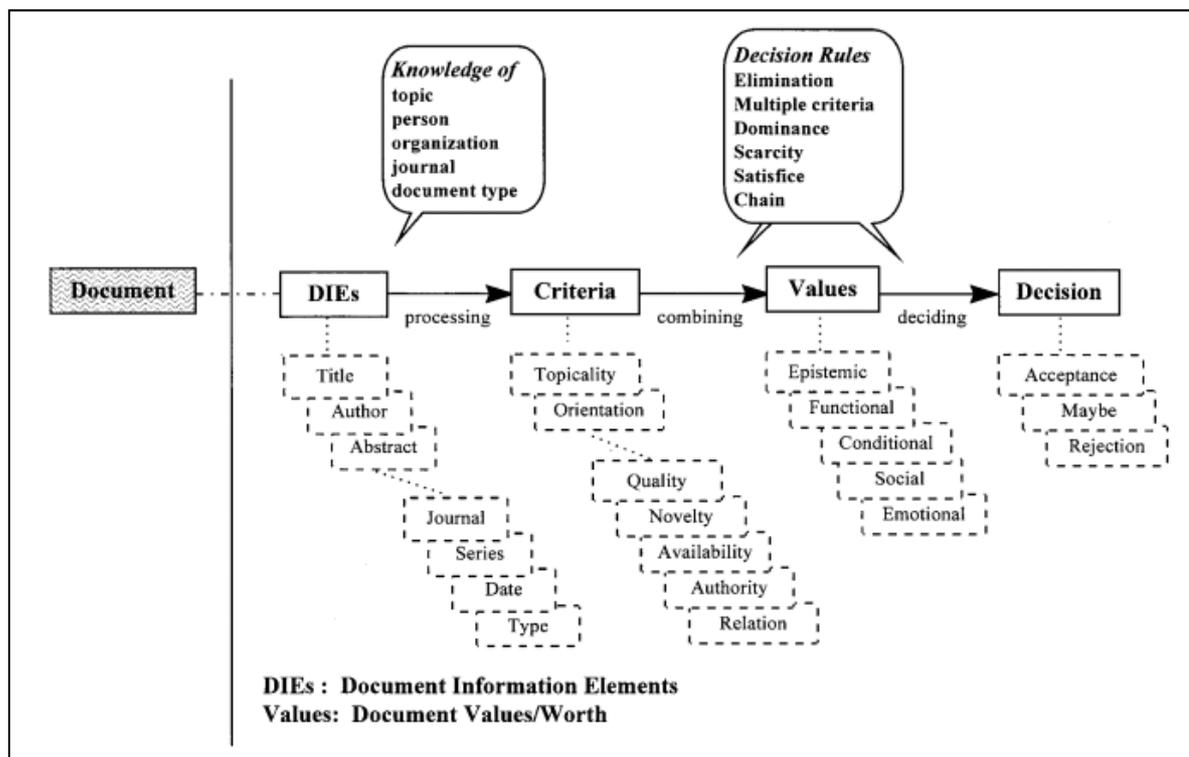


Figura 22: Modelo de seleção de documentos

No trabalho de Peiling Wang e Dagobert Soergel (1998), o objetivo da seleção de documentos é a sua citação para a produção de um texto, neste trabalho, o uso desses documentos tem uma variedade maior de finalidades. Após um documento ser selecionado, o usuário poderá utilizá-lo para criar novas certezas e dúvidas, fazer anotações sobre outros elementos de um Projeto de Aprendizagem, ou mesmo produzir um texto, citando assim o documento selecionado.

A seguir, serão tratadas algumas questões mais técnicas de como serão processados os itens de conhecimento dos projetos de aprendizagem cadastrados no ambiente.

5.4 Adaptação de um modelo de RI considerando contexto

Como já foi citado anteriormente, o desafio desse trabalho é organizar os itens de conhecimento de um projeto de aprendizagem de forma que eles possam ser utilizados pelos agentes do sistema, produzindo respostas adequadas às necessidades de informação do usuário.

Para tanto, foi desenvolvido um modelo de funcionamento para o sistema baseado em Algoritmos Genéticos e RI. Esse modelo parte do princípio que existem documentos relevantes na Internet que podem ajudar o usuário do sistema a satisfazer suas necessidades de informações. Esses documentos podem facilmente serem recuperados através de motores de busca da Internet, entretanto é preciso criar uma estratégia de seleção de palavras-chaves para criar consultas e recuperá-los. Portanto, o problema aqui é descobrir quais conjuntos de consultas podem dar acesso a essa coleção de documentos?

Sendo assim, vamos considerar os grupos temáticos de Itens de Conhecimento (Certezas e Dúvidas) como sendo os melhores candidatos a nos fornecer essas consultas. Note que cada grupo temático, funciona como se fosse um documento de uma coleção. Então o que teríamos, seria algo similar a uma amostra de documentos que detalha os interesses do usuário.

Então, temos uma coleção C composta de documentos d_j que representam as necessidades do usuário e que podem ser utilizadas como fonte de consultas para obtermos documentos a partir da Internet. Agora falta definir como essas consultas serão geradas.

O primeiro passo então é transformar esses documentos em vetores de termos indexados \vec{d}_j . Para isso, este trabalho propõe uma sequência de tratamentos sobre cada documento da coleção, para que essas informações possam ser aproveitadas. São eles:

- Remoção de *Stop Words*;
- *Stemming*;
- Criação dos vetores de termos com peso para cada documento;

- Seleção de uma quantidade de n termos k_i em cada vetor para formar uma consulta q_j que represente o documento d_j .

Em uma primeira rodada, a escolha dos termos é feita seguindo o critério do maior peso, ou seja, os primeiros n termos a representarem o documento serão aqueles que obtiverem os maiores pesos a partir da função *tf-idf* (detalhada no capítulo sobre Recuperação de informações) de atribuição de pesos.

Definição: Dada uma coleção C de documentos d_j , será gerada uma consulta q_j representativa de cada vetor de termos indexados do documento \vec{d}_j , considerando os n primeiros termos k_i classificados por seus respectivos maiores pesos $w_{i,j}$.

Agora temos uma coleção de consultas que podem ser submetidas aos nossos motores de busca, entretanto não sabemos como avaliá-las e nem como melhorá-las a partir das escolhas dos usuários sobre os documentos recuperados.

Nesse instante, é que se propõe o uso dos Algoritmos Genéticos como estratégia de seleção natural dos termos das consultas, a idéia é que, geração a geração, as consultas consigam recuperar documentos mais próximos do que o usuário necessita.

Especificamente, as fases do algoritmo utilizado são: inicialização da população, busca de documentos, avaliação da população, seleção, recombinação e mutação; que são explicadas a seguir:

1. Geração dos índices e das consultas possíveis considerando os grupos de Itens de Conhecimento do Projeto de Aprendizagem disponíveis para criação de uma geração no AG;
2. Submissão das consultas com os n termos indexados mais relevantes;
3. Classificação dos documentos recuperados considerando apenas os m primeiros em cada consulta;
4. Escolha de documentos relevantes pelo usuário;
5. Seleção natural dos genes (consultas) que mais se adaptaram, isto é, aquelas que devolveram documentos que o usuário aproveitou;
6. Operação de mutação genética, criando assim, uma nova geração

de genes;

7. Observação de novas anotações para inserção de novos genes;
8. Retorna ao passo 2.

A partir daí o sistema entra em um processo cíclico de operações genéticas sobre os termos de busca, criando um processo de Recuperação de Informações Interativo baseado nas anotações de Certezas e Dúvidas de um estudante utilizador do sistema. A idéia é que isso funcione como uma ferramenta de suporte ao estudo continuado. Então, quanto mais o sistema for alimentado, mais ele recomenda documentos, e mais ele aprende e se aproxima de consultas que retornam documentos relacionados ao que o usuário se interessa. Na seção seguinte, a que trata sobre os agentes do sistema, esse processo será melhor detalhado.

A escolha de qual grupo temático deve receber a recomendação é feita através de um classificador, foi escolhido para este trabalho o classificador baseado em LSI. Então, será mantida também uma matriz de valores singulares para todos os Grupos Temáticos de certezas e dúvidas, e quando os documentos forem trazidos da Internet, eles serão submetidos a esse classificador, definindo assim um critério de recomendação de documentos.

Cada um desses tratamentos faz uso de uma técnica específica, por isso, cada uma delas foi estudada, e para cada uma foi feita uma pesquisa por bibliotecas de código fonte que pudessem ser aproveitadas. Todas elas em linguagem de programação Java (2009). No decorrer desse levantamento foram encontradas várias bibliotecas dentre elas são destacadas algumas mais relevantes conforme a tabela abaixo:

Tabela 9: Bibliotecas de software estudadas

<i>Nome</i>	<i>Descrição</i>
Classifier4J	É uma biblioteca Java projetada para fazer classificação textual. Ela provê a implementação de um Classificador Bayesiano e de um baseado em Espaço Vetorial. A biblioteca oferece também um conjunto de facilidades como: Resumo textual e

	remoção de <i>stopwords</i> . (CLASSIFIER4J, 2009)
JTMT	O <i>Java Text Mining Toolkit</i> é um conjunto de classes Java que implementa várias funcionalidades para mineração de texto, entre elas está um classificador baseado no algoritmo LSI. Ela também provê facilidades como criação de <i>tokens</i> : abreviações, números, frases, palavra, endereços de internet, <i>stopwords</i> , etc. (PAL, 2009)
Lucene	O Apache Lucene, ou simplesmente Lucene, é uma API de indexação de documentos e um motor de busca, escrito na linguagem de programação Java. É um software de código aberto da Apache Software Foundation. (LUCENE, 2009)
WordNet	Base de dados léxico de palavras, essa base está disponível em vários idiomas, entre eles o português e pode ser utilizada como desambiguador de termos, ou mesmo para a recomendação de termos. (WORDNET, 2009)

Os termos submetidos aos motores de busca são gerados a partir do uso das palavras com maior carga semântica. Segundo Baeza-Yates (1999), os substantivos são os melhores candidatos a esse papel. Também podem ser utilizados os substantivos extraídos das anotações associadas a um IC, eles podem servir para confirmar o uso de palavras já encontradas no IC ou para expandir a consulta adicionando palavras que melhorem a descrição das necessidades do usuário, ou seja, acrescentando contexto à busca.

Para a obtenção dos documentos da *WEB* será necessário um gerenciamento de consultas feitas a motores de busca da Internet, o protótipo comunica-se com os sites Google (2009) e Yahoo (2009), aproveitando de cada, uma quantidade limitada de documentos (ajustável pelos usuários, mas aproximadamente de 10 respostas).

A combinação desse conjunto de estratégias apresentadas se mostrou uma

atividade complexa para um sistema simples de computação. Portanto, apresentaremos, na próxima seção, uma arquitetura baseada em agentes para a distribuição dessas tarefas.

5.5 Agentes do sistema

Para dar conta de todas as características citadas anteriormente para o sistema, optou-se por distribuir algumas dessas atribuições entre agentes de software. Portanto, nesta seção será apresentada a arquitetura de agentes utilizada e as interações feitas entre estes componentes.

Basicamente, o sistema proposto neste trabalho é composto por dois módulos, um primeiro que é o protótipo de Ambiente Virtual de Aprendizagem que já foi explicado na seção anterior, e um segundo que é composto pelos Agentes do Sistema. Atualmente, são propostos dois agentes um para RI e outro para recomendação de documentos.

O primeiro agente do sistema é o Agente de Recomendação de Documentos (ARD), ele classifica os documentos recuperados pelo ARI e posteriormente recomenda os documentos que forem considerados mais relevantes para cada elemento dos Projetos de Aprendizagem dos usuários do sistema.

O ARD também é composto de dois componentes principais, que são: Classificador LSI e o Recomendador de Documentos. O classificador LSI, como o próprio nome já faz referência, faz uso do modelo baseado em Semântica Latente, de forma prática consideremos os agrupamentos de Certezas e Dúvidas para o experimento que se segue:

Tabela 10: Agrupamento temático de Itens de Conhecimento

Grupo Temático	Itens de Conhecimento
Tarefas	c1: existe uma subdivisão de tarefas no formigueiro; c3: as formigas responsáveis por buscar comida e manter o

	formigueiro são as operárias;
Hierarquia	c2: as formigas trabalham em função de proteger a rainha; c5: existe uma hierarquia no formigueiro; d2: como se escolhe a rainha de um formigueiro? d3: existe apenas uma rainha no formigueiro? d5: a rainha pode comandar as outras formigas, tomar decisões?
Comunicação	c4: as formigas podem se comunicar umas com as outras d4: como as formigas se comunicam umas com as outras? d5: a rainha pode comandar as outras formigas, tomar decisões?
Inteligência	d6: já foi comprovado algum nível de inteligência em um formigueiro? d7: que atividades inteligentes já foram identificadas em formigueiros? d10: as formigas podem se adaptar, aprender e se comportar de forma diferente, conforme problemas que elas encontrem?
Compartimentos	c6: o formigueiro tem compartimentos com funções bem definidas c7: um dos compartimentos do formigueiro é o depósito de lixo c8: o depósito de lixo é o compartimento mais fundo do formigueiro d11: existe um compartimento para depositar alimento, ou as formigas o largam em qualquer lugar?
Conflitos	d8: podem ocorrer desentendimentos (brigas) no formigueiro? d9: como é a reação das formigas em caso de invasão de predadores?
População	d1: qual a quantidade de formigas em um formigueiro?

Conforme já citado anteriormente, a primeira coisa a ser feita para criar uma

visão lógica desses grupos temáticos é a remoção das chamadas *stopwords* que são palavras não indexadas pelos motores de busca, por exemplo: de, com, a, ao, etc. Em seguida é feita a atribuição de pesos para os termos dos documentos, e é partir daí que obtemos a primeira rodada de consultas, escolhendo os n primeiros termos desse cálculo.

Considerando cada um desses grupos temáticos será obtida a seguinte primeira rodada de consultas:

Tabela 11: Primeira rodada de consultas extraídas do sistema

Grupo Temático	Palavra-chave
Tarefas	formigueiro manter formigas existe operarias
Hierarquia	rainha formigueiro existe formigas forma
Comunicação	formigas comunicam tomar podem comunicar
Inteligência	foram inteligentes encontrem diferente podem
Compartimentos	deposito formigueiro compartimento lixo compartimentos
Conflitos	brigas podem predadores formigueiro formigas
População	quantidade formigueiro formigas

Pode-se observar que ainda existem algumas palavras que poderiam ser classificadas como *stopwords*, isso se deve ao banco de dados restrito de palavras utilizado. Não é tão fácil encontrar bancos de *stopwords* em português na Internet, mesmo os disponíveis são muito pequenos e falhos. O que foi utilizado neste protótipo já foi bastante incrementado com palavras novas, mas ainda está longe de ser satisfatório.

O segundo agente proposto neste trabalho é o de Recuperação de Informações (ARI), que é responsável por criar, gerenciar e submeter consultas aos motores de busca da Internet. Este agente engloba dois componentes: o Gerenciador de Consultas Genético (GCG) e o Motor de Meta-Busca (MMB).

O GCG é o componente do sistema que faz uso dos algoritmos genéticos para selecionar quais gerações de consultas tem maior aptidão para atender as necessidades do usuário.

A estratégia desse agente é modelar as consultas como genes, e no decorrer das gerações selecionar quais são os genes mais aptos, ou seja, as consultas que continuam retornando documentos que o usuário utiliza. O GCG cadastra as informações sobre os genes e o histórico da evolução das gerações na própria base de dados de Projetos de Aprendizagem, como demonstra o Diagrama de Estruturas de Dados (DED) da Figura 23:

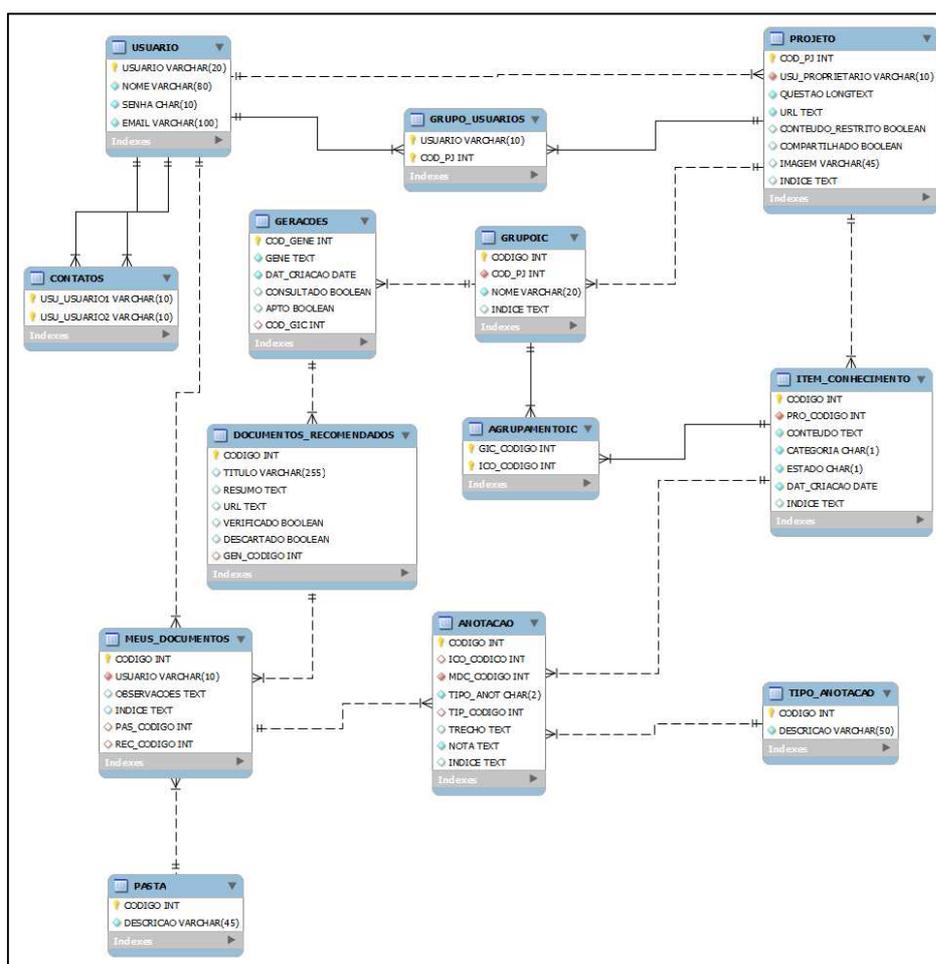


Figura 23: DED da base de dados do sistema.

Conforme apresentado no DED, existe uma tabela de “GERACOES” de consultas utilizada pelo GCG, levando em consideração os dados gerados pelo Classificador LSI para a primeira rodada de consultas a Tabela 12 apresenta a seguinte projeção de dados:

Tabela 12: Primeira geração de genes

GENE	DAT_CRIACAO	CONSULTADO	APTO
formigueiro manter formigas existe operarias	11/11/2009	Não	Sim
rainha formigueiro existe formigas forma	11/11/2009	Não	Sim
formigas comunicam tomar podem comunicar	11/11/2009	Não	Sim
foram inteligentes encontrem diferente podem	11/11/2009	Não	Sim
deposito formigueiro compartimento lixo compartimentos	11/11/2009	Não	Sim
brigas podem predadores formigueiro formigas	11/11/2009	Não	Sim
quantidade formigueiro formigas	11/11/2009	Não	Sim

Após essa primeira rodada, os documentos classificados são cadastrados na tabela “DOCUMENTOS_RECOMENDADOS” para aguardar a análise do usuário. Conforme o aproveitamento de documentos do usuário, o GCG inicia o seu trabalho; que é basicamente modificar as gerações de genes, conforme a utilização dos documentos.

O passo seguinte é descartar genes que não tiveram documentos aproveitados e criar novas gerações de documentos, as novas gerações são criadas através da operação de mutação, onde substituídas as palavras aleatoriamente dos genes, a taxa de substituição é pré-definida e varia conforme ajustes no sistema. O exemplo abaixo ilustra como essa operação é feita:

Tabela 13: Exemplo de mutação

Mutação				
rainha	formigueiro	existe	formigas	forma
Novas palavras →			manter	tarefas

A partir daí, é criada uma nova geração de consultas que são submetidas novamente aos Motores de Busca da Internet.

O MMB é o componente que serve unicamente para submeter as consultas geradas pelo GCG aos motores de busca da Internet, basicamente foram desenvolvidas para este componente interfaces para dois sites o Google (2009) e Yahoo (2009). A Figura 24 apresenta o diagrama de classes para as interfaces criadas.

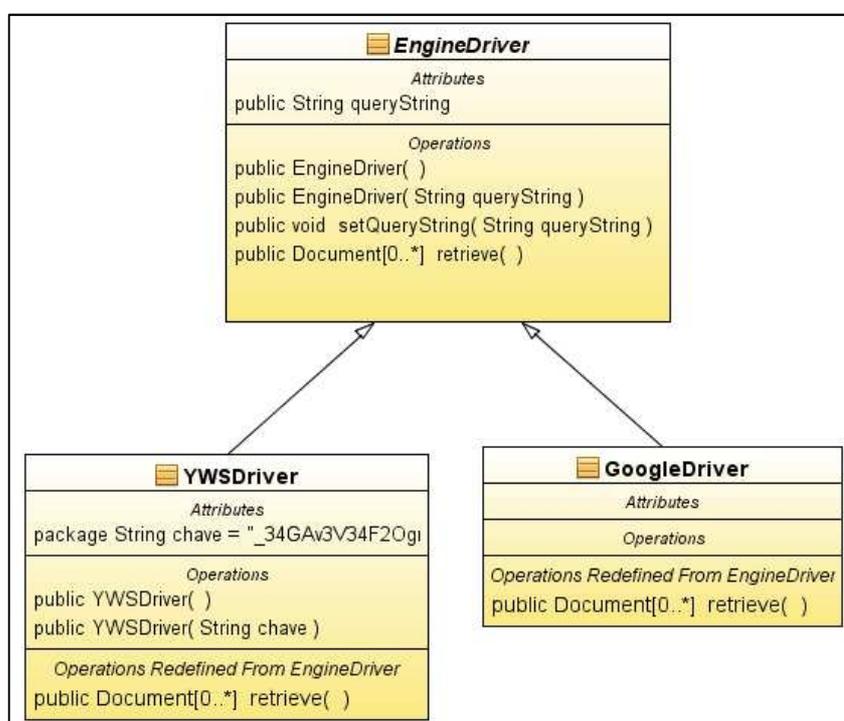


Figura 24: Interfaces de conexão com motores de busca.

Inicialmente, o sistema foi programado para ter apenas uma instância de cada agente para todos os projetos de aprendizagem cadastrados na base, sendo assim, a comunidade de agentes analisa ciclicamente os projetos. Encontrando modificações, elas são tratadas, não encontrando verifica-se o próximo projeto.

A arquitetura dos agentes descrita neste trabalho é ilustrada na Figura 25. A figura demonstra as interações de cada agente, e como estes extraem e inserem informações no ambiente virtual de aprendizagem ARPA. Toda a comunicação entre AVA e agentes é feita pelo compartilhamento de bases de dados comuns.

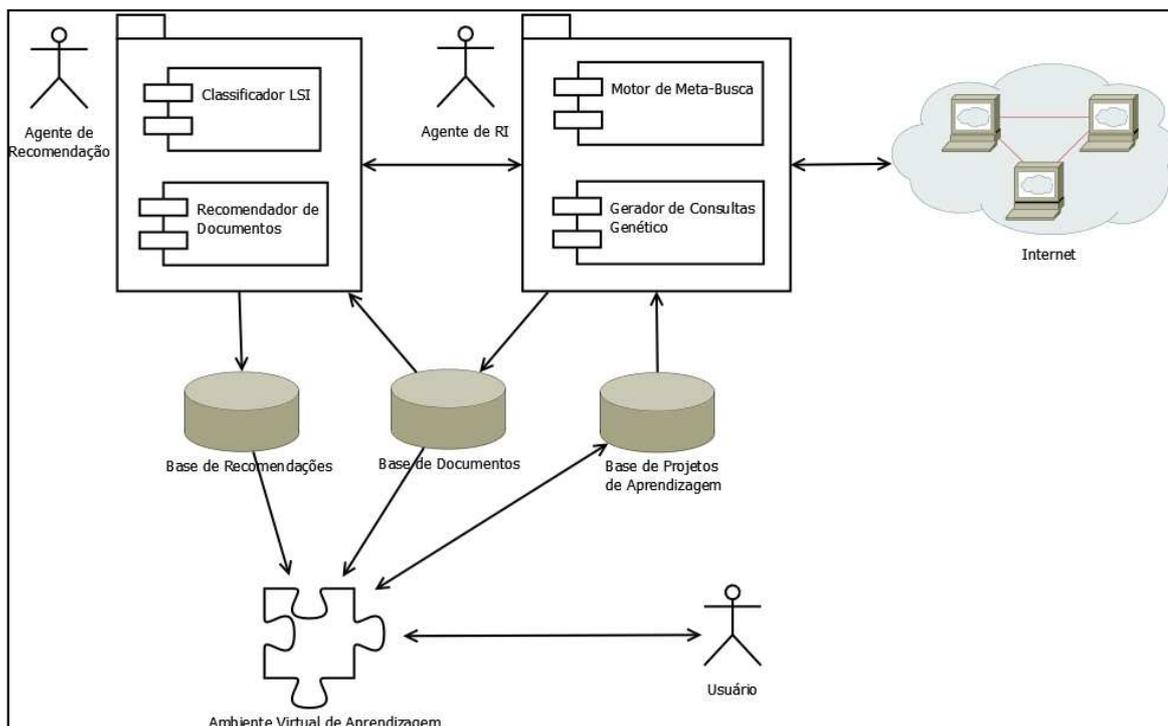


Figura 25: Arquitetura do Sistema de Agentes

O Ambiente de Agentes é a plataforma necessária para a execução dos agentes e é onde foi desenvolvido o Ambiente Mediador deste trabalho, atualmente os protótipos criados para os agentes são feitos na plataforma de agentes JADE (2009).

A seguir serão apresentados alguns testes relacionados aos protótipos desenvolvidos. Em seguida estão enumeradas as considerações finais deste trabalho, bem como uma lista de possíveis trabalhos futuros.

5.6 Testes

Recapitulando a metodologia, o teste foi aplicado em dois grupos de usuário, que foram compostos por alunos de duas turmas distintas de graduação do Curso de Enfermagem. Cada turma foi dividida em equipes para confecção de PA's.

O primeiro grupo elaborou seus projetos de forma manual, sem a ferramenta

desenvolvida neste trabalho, utilizando apenas editores de texto, planilha eletrônica, e consultas à internet. Denominaremos esse primeiro grupo de “Grupo de Controle”. O segundo grupo elaborou seus projetos no ambiente ARPA, para isto foi realizada uma breve explicação do funcionamento da ferramenta, esse segundo grupo chamaremos de “Grupo Experimental”. Ambos seguiram um documento norteador com um PA exemplo. A seguir temos um quadro que demonstra a caracterização das duas turmas:

Tabela 14: Caracterização dos grupos avaliados

Característica	Grupo de Controle	Grupo Experimental
Quantidade de alunos	23	21
Numero de equipes	6	5
Tempo de treinamento na pedagogia de PA	2 h/a	2 h/a
Tempo estipulado para o trabalho	1 semana	1 semana

Os dois testes foram executados em tempos diferentes, sendo que o grupo de controle foi testado no período letivo 2009/2 – semestre normal – e o grupo experimental no período 2009/2 – período de férias. Ambos os grupos foram avisados do experimento e todos os alunos concordaram com o mesmo, sendo que nos dois casos os trabalhos consistiam de avaliação valendo nota. Todas as equipes entregaram os trabalhos nos prazos estabelecidos. A estrutura que o trabalho devia atender era a seguinte:

1. Questão de investigação: Questão central do trabalho;
2. Itens de Conhecimento (Certezas e Dúvidas): Inventário inicial do que os alunos conhecem;
3. Agrupamentos Temáticos: Agrupamentos temáticos das certezas e dúvidas;
4. Documentos Recuperados: Esta etapa devia correlacionar o link para o documento, o grupo temático e anotações relevantes;

5. Sínteses: Para esta etapa os grupos eram livres para escrever utilizando mais de uma fonte por síntese.

Uma primeira impressão foi quanto às variações na qualidade dos trabalhos para os dois grupos. Denotando claramente a dedicação maior de algumas equipes nas duas turmas.

Para este primeiro teste do uso do ambiente ARPA foi feita apenas uma comparação estatística da produção dos dois grupos considerando a média aritmética simples das quantidades de elementos produzidos por seção em cada equipe dos dois grupos. O quadro abaixo demonstra os resultados dessa análise:

Tabela 15: Quadro estatístico da produção de elementos nos PA's avaliados

Item	Grupo de Controle	Grupo Experimental
Média do Numero de Itens de Conhecimento	15	22,5
Media do Numero de Agrupamentos	5,75	8,25
Media de Documentos Recuperados	6,5	9,66
Media de Sínteses	5	7

Para a finalidade do teste, não foi avaliada a qualidade do conteúdo dos trabalhos feitos pelos grupos. Entretanto, é interessante observar o crescimento do volume da produção no grupo experimental em relação ao grupo de controle. É sabido que essa informação não é determinante, mas pode dar indícios que de alguma forma o trabalho no ambiente ARPA facilita a organização e catalogação dos Itens de Conhecimento e dos demais elementos considerados em um PA.

Os testes no ambiente ARPA estão no início e ainda precisarão ser aprimorados. Esse trabalho terá continuidade com melhorias no protótipo construído e a publicação de novos artigos apresentando mais resultados, conforme é citado nas considerações finais deste trabalho.

6 Considerações Finais

No decorrer deste trabalho de doutoramento foram considerados vários tópicos relacionados à pedagogia e a tecnologia, portanto, neste ultimo capítulo será apresentado um fechamento dessas idéias.

Primeiramente foi citada a educação continuada como uma preocupação de vários órgãos mundiais e como uma tendência observada nos dias atuais. O ritmo crescente da globalização e das mudanças tecnológicas, as mudanças na natureza do trabalho e no mercado de trabalho, e o envelhecimento das populações estão entre as forças que enfatizam a necessidade da continuidade de atualização das habilidades de trabalho durante toda a vida.

Em seguida foi observado que para uma sociedade que demanda atualização constante, os indivíduos agora necessitam de novas habilidades, conhecimentos e atitudes necessários ao uso efetivo e competente da leitura e da escrita nas práticas sociais que envolvem a língua escrita. Para tanto, conclui-se que os indivíduos necessitam ter letramento para a manutenção dessa atualização.

Mais que isso, os indivíduos necessitam ser autônomos detendo assim a capacidade de elaborar, ao menos em parte, suas próprias normas. Assim, um indivíduo autônomo age livremente de acordo com um plano próprio, de forma análoga a um governo independente, que administra seu território e estabelece suas políticas. Em um contexto educacional, é possível que alguém queira estudar de forma independente, escolhendo seus horários de trabalho e temas de estudo.

Essa ultima concepção é muito importante para este trabalho, pois, para alguém que se preste a fazer uso do sistema proposto nesta tese, as capacidades de elaboração e respeito às regras criadas são cruciais.

Outro conceito importante foi o da heutagogia, que compreende a aprendizagem autodirecionada, em que o aluno é o gestor e programador de seu processo de aprendizagem. Nessa proposta, o aluno de EaD lança mão das novas tecnologias de comunicação e informação para facilitar a autonomia na condução de sua aprendizagem, contando, quando possível, com o suporte de professores,

colegas, ou especialistas (tutores) no assunto tratado. Nesse novo ambiente, foram citados também novos termos para falar de aprendizagem, tais como autorresponsável, autoplanejada, auto-organizada, independente e autorregulada, além de não linear e não sequencial, em que os aprendizes trilham seus próprios caminhos e alcançam seus próprios objetivos.

Uma das idéias chave deste trabalho citada no decorrer do texto foi a construção de artefatos como estratégia da ocorrência da aprendizagem. O construcionismo compartilha a visão da aprendizagem do construtivismo sobre “a construção de estruturas cognitivas” pela internalização das ações. Ele adiciona, porém, a idéia de que isso ocorre especialmente em um contexto no qual o aprendiz esteja engajado continuamente na construção de uma entidade pública, quer seja um castelo de areia na praia ou uma teoria do universo.

Portanto, observou-se a importância da Pedagogia de Projetos de Aprendizagem como uma ferramenta interessante de construção de artefatos. Lembrando que a atividade de fazer projetos é simbólica, intencional e natural do ser humano. Por meio dela, o homem busca a solução de problemas e desenvolve um processo de construção de conhecimento, que tem gerado tanto as artes quanto as ciências naturais e sociais.

Considerando as características acima citadas sobre Aprendizagem por Projetos, concluiu-se, então, que um documento gerado na forma de Projeto de Aprendizagem poderia servir como um registro sistematizado das atividades, podendo este ser utilizado como indicador dos interesses do aprendiz em um dado instante. Surge daí o interesse pelo uso da metodologia de Aprendizagem por Projetos como fundamentação pedagógica para este trabalho.

Retomando os objetivos deste trabalho, que é o de desenvolver um ambiente mediador que sirva de apoio à aprendizagem continuada, baseado na recuperação de documentos disponíveis na Internet. Partiu-se então do pressuposto de que os Projetos de Aprendizagem são bons candidatos a representar ou dar indícios – dentro de um ambiente de aprendizagem – das necessidades do usuário.

Nesse sentido, foi desenvolvido um protótipo de Ambiente Virtual de Aprendizagem, ao qual foi atribuído o nome de ARPA – Ambiente de Apoio a Recuperação de Informações e Projetos de Aprendizagem.

Então, a pesquisa se encaminhou para a área de recuperação de informações, que remete a representação, armazenamento, organização e acesso a itens de informação. Mais precisamente a Recuperação de Informações Interativa (RII) que é definida como sendo o processo interativo de comunicação que ocorre durante a recuperação de informação, envolvendo os principais participantes do processo de RI, isto é, o usuário, o mediador e o sistema de RI.

Mais adiante, observou-se que seria importante considerar o contexto do usuário durante o processo de Recuperação de Informações de forma Interativa, considerando as dimensionalidades dos contextos em associação com os motores e sistemas de RI. Essa dimensionalidade varia conforme as características de conteúdo tradicionais que existem dentro e entre os objetos de informação. Como exemplos de dimensões podem ser considerados: as palavras próximas a parágrafos e hiperlinks, os movimentos do mouse durante a seção de interativa, etc.

Para facilitar a vida do usuário do sistema, verificou-se que seria interessante utilizar a idéia de um Assistente Pessoal de Aprendizagem, que seria uma entidade ou alguém que pudesse fornecer documentos relacionados aos contextos de estudo de um indivíduo. Ele se diferenciaria de um Gestor Pessoal de Informações, pois levaria em consideração uma forma de trabalho estruturada, baseada na organização ou na forma de estudo de uma pedagogia específica.

Apropriando-se dessa ideia, alguém que pretende se manter atualizado em um determinado assunto poderia lançar mão de um Assistente Pessoal de Aprendizagem para auxiliar no processo de aquisição de documentos. Portanto, esse assistente poderia se responsabilizar por analisar também os projetos de aprendizagem com a finalidade de classificar os documentos encontrados conforme os interesses de estudos de um usuário.

Essa última idéia viabilizou a utilização da Tecnologia de Agentes de Software para modelar esse Assistente Pessoal de Aprendizagem. Para tanto, foi construída uma arquitetura de agentes, que em testes de uso se mostrou viável e promissora.

Os agentes utilizados no protótipo conseguiram retornar documentos relacionados aos Projetos de Aprendizagem dos usuários, demonstrando que conseguem extrair palavras-chaves contextualizadas aos projetos de aprendizagem com os quais o usuário está envolvido.

Como o mecanismo aqui proposto é de meta-busca, essas palavras-chaves são submetidas a motores de busca da Internet. Os documentos devolvidos são posteriormente classificados e recomendados aos usuários que tem o poder de decidir se determinado documento é relevante ou não.

O processo de organização dos projetos de aprendizagem, recomendação de documentos e catalogação dos mesmos, é uma atividade cíclica que funciona como ferramenta de suporte aos estudos continuados. Portanto este trabalho cumpriu seu objetivo que era o de criar um protótipo de ambiente mediador de estudos continuados de forma autônoma, esse mecanismo faz uso apenas de documentos extraídos da Internet.

Pretende-se dar continuidade a esta pesquisa na forma de trabalhos futuros, então seguem abaixo algumas sugestões de aprimoramentos possíveis para o sistema:

- aprimoramento da interface do ARPA;
- estudo de viabilidade do uso de outras operações genéticas para a criação de novas gerações de consultas no algoritmo genético;
- testes comparativos de desempenho utilizando outros algoritmos de classificação, tais como: Probabilístico, CNG e Espaço Vetorial;
- avaliação de outras variáveis de contexto do usuário que possam ser consideradas para a melhoria da qualidade da busca;
- escrita de mais artigos sobre o trabalho para a publicação dos resultados da pesquisa;
- realização de testes em escala maior com grupos de usuários do sistema e grupo de controle;

Referências

- ALA. **American Library Association's Presidential Committee on Information Literacy**. Final Report. Chicago, American Library Association. 1989.
- ALEXAKI, S. et al. **Managing RDF metadata for community webs**. In: International Workshop on The World Wide Web And Conceptual Modeling. 2. 2000. [S. l. : s. n.], 2000. p. 140-151. Disponível em: <<http://139.91.183.30:9090/RDF/publications/>>. Acesso em: 11 dez. 2007.
- ALMEIDA, Mauricio B.; BAX, Marcello P. **Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção**. Ci. Inf., Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003.
- BAEZA-YATES, Ricardo, RIBEIRO-NETO, Berthier. **Modern Information Retrieval**. New York: ACM Press, 1999.
- BASSO, M. V. A.; FAGUNDES, L. C.; NEVADO, R. A.; BETENCOURT, J.; MENEZES, C. S. **AMADIS Um Ambiente Virtual para apoio ao Desenvolvimento de Projetos de Aprendizagem**. In: XVI Simpósio Brasileiro de Informática na Educação, 2005, Juiz de Fora - MG. Proceedings XVI SBIE, 2005.
- BECKER, Fernando. **O que é construtivismo?** Idéias. n. 20, São Paulo: 1993.
- BRATMAN, M. E.; ISRAEL, D. J.; POLLACK, M. E. **Plans and Resource-Bounded Practical Reasoning**, Computational Intelligence, [S.l.], v.4, p.349–355, 1988.
- BIANCONI, Leonardo. **Recuperação de Informação da Web usando Algoritmos Genéticos**. Trabalho de Conclusão de Curso. (Graduação em Engenharia de Computação) - Pontifícia Universidade Católica de Campinas. Orientador: Juan Manuel Adán Coello. 2007.
- BORGO, S. et al. **Using a large linguistic ontology for Internet-based retrieval of object-oriented components**. In: Proceedings of The International Conference on Software Engineering and Knowledge Engineering. 9. 1997. Madrid. Madrid : [s. n.], 1997.
- CANDY, Philip C., **Information Literacy and Lifelong Learning**, July 2002, White Paper prepared for UNESCO, the U.S. National Commission on Libraries and

Information Science, and the National Forum on Information Literacy, for use at the Information Literacy Meeting of Experts, Prague, The Czech Republic.

CAPPSS. (Centre for Applied Social Studies). **Means of Learning. National University of Ireland (1997)**, Maynooth, Ireland. Disponível em: <http://www.may.ie/cappss/courses/hdcw/hdcw_2.htm. Acessado em 03/07/2007>.

CAZELLA, Sílvio César; REATEGUI, Eliseo. Mini-course: **Recommender Systems**. In: ENIA, 2005, São Leopoldo, RGS, Brazil. 2005.

CEGLOWSKI M.; COBURN A.; CUADRADO, J. **Semantic Search of Unstructured Data using Contextual Network Graphs**. 2003. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.4283>>.

CHARLESWORTH, M. **La bioética en una sociedad liberal**. Cambridge: Cambridge, 1996.

CHAUI, Marilena. **Convite à Filosofia**. São Paulo: Ática. 13ª Ed. 2006.

CLASSIFIER4J. **Classifier4J**. Disponível em: <<http://classifier4j.sourceforge.net/>>. Acessado em 10/11/2009.

CMAPTOOLS. **CmapTools – Home Page**. Disponível em: <<http://cmap.ihmc.us/>>. Acessado em 03/07/2007.

COOL, Colleen; SPINK, Amanda. Issues of context in information retrieval (IR): an introduction to the special issue. **Information Processing and Management: an International Journal**. Volume 38 , 2002.

CORREIA W. **Piaget: que diabos de autonomia é essa?** In: Currículo sem fronteiras, v.3, n.2, pp. 126-145, Jul/Dez. 2003.

DIETINGER, T., MAURER, H.: **GENTLE - General Network Training an Learning Environment**; Proc. EDMEDIA, Freiburg, Germany, 274–280. 1998.

DOYLE, C. S. **Information literacy in an Information Society: a concept for the information age**. New York : Syracuse University, 1994.

ETZIONI, Oren; WELD, Daniel S. **Intelligent Agents on the Internet: Fact, Fiction, ant Forecast**. IEEE Expert, New York, p. 44-49, Aug. 1995.

FAGUNDES, L.C. et. al.; **Aprendizes do Futuro: as inovações já começaram**. MEC. Coleção Informática para mudança na educação, Brasília, 1998.

FERBER, Jacques; GASSER, Les. Intelligence Artificielle Distribuée. In: **Conference on Expert Systems and Their Applications**, 11., 1991, Avignon. Tutorial Notes... [S.l.:s.n.], 1991

FERREIRA, Aurélio Buarque de Holanda. **Dicionário Aurélio Eletrônico**. Versão 1.4. Rio de Janeiro: Nova Fronteira, 1994. Software baseado no Novo Dicionário da Língua Portuguesa.

FENSEL, D. **Ontologies: silver bullet for knowledge management and electronic commerce**. Berlin: Springer-Verlag, 2001.

FIELDEN, N. **History of Information Retrieval Systems & Increase of Information Over Time**. Biennial California Academic & Research Librarians (CARL) Conference. Asilomar, Monterey California, 2002.

FLORES, Cecília. **Negociação Pedagógica Aplicada a um Ambiente Multiagente de Aprendizagem Colaborativa**. Tese de doutorado. Porto Alegre: PPGC / UFRGS, 2005.

FONSECA, L. C. C.; VICARI, Rosa Maria; MENEZES, Crediné Siva de; SOARES, Jonatas. **An Intelligent and Contextual Information Retrieval Environment for Lifelong Learning**. In: Frontiers in Education (FIE) 2008, New York. Frontiers in Education Proceedings, 2008.

FONSECA, L. C. C.; MENESES, C. S.; VICARI, R. M. **Towards an Intelligent and Contextual Information Retrieval Environment for Lifelong Learning**. Revista Scientia: Estudos Interdisciplinares em Computação. São Leopoldo - RS. v. 17, n. 1. 2007.

FONSECA, L. C. C.; MENEZES, Crediné Siva de; VICARI, Rosa Maria; FAGUNDES, Lea da Cruz. **Esclarecendo Dúvidas e Validando Certezas - Uma ferramenta de apoio à construção de conhecimento no contexto de Projetos de Aprendizagem**. Rernote Revista Novas Tecnologias na Educação, Porto Alegre - RS, v. 3, 2005.

FREIRE, Paulo. **Pedagogia da Autonomia. Saberes necessários a Prática Educativa**. São Paulo: Paz e Terra. 1996.

GIRAFFA, L. M. **Uma arquitetura de tutor utilizando estados mentais**, Tese (Doutorado em Ciência da Computação) — UFRGS, Porto Alegre, 1999.

GOOGLE. **Google**. Disponível em: <<http://www.google.com>>. Acesso em: 10/11/2009.

GRUBER, T. **A Translation Approach to Portable Ontology Specifications**. Knowledge Systems Laboratory. Stanford University, Stanford, CA, Technical Report KSL, p.92-71.

GUARINO, N. **Formal Ontology and Information Systems**. In: Guarino, N. (Ed.) Formal Ontology in Information Systems, IOS Press, Amsterdam, Netherlands, p. 3-15, 1998.

GÜTL Ch., MAURER H., PIVEC M. **Learning on demand using xFIND: an improved way for ongoing and lifelong learning as a smart module for the GENTLE learning environment**. Proceedings of the ICCE `99, 7th International conference on computers in education, Chiba, Japan, (Frontiers in artificial intelligence and applications, Vol. 55). Amsterdam: IOS Press, Vol. 2, p.592-599. 1999.

HEDBERG, Sara Reese. **Intelligent Agents: The First Harvest of Softbots Looks Promising**. IEEE Expert, New York, p. 6-9, Aug. 1995.

INGWERSEN, P. **Information Retrieval Interaction**. Reino Unido: Taylor Graham Publishing. 1992.

INGWERSEN, P., JÄRVELIN, K. **The Turn: Integration of Information Seeking and Retrieval in Context**. Netherlands: Springer. 2005a.

INGWERSEN, Peter; JARVELIN, Kalervo. **Selected variables for IR Interaction in context: Introduction to IRiX SIGIR 2005 Workshop**. IRiX ACM SGIR Workshop. pp. 6-9. 2005b.

ISC. **The ISC Domain Survey**. Disponível em: <<https://www.isc.org/solutions/survey>>. Acesso em: 10/11/2009

JADE. **Java Agent DEvelopment Framework**. Disponível em: <<http://jade.tilab.com/>>. Acesso em: 10/11/2009.

JADEX. **JADEX BDI Agent System**, Disponível em: <<http://vsis-www.informatik.uni-hamburg.de/projects/jadex/>>. Acesso em 10/11/2009.

JAVA. **Java Technology**. Disponível em: <<http://java.sun.com/>>. Acesso em: 10/11/2009.

- JENNINGS, N. R.; WOOLDRIGDE, M. **Applications of intelligent agents**. 1998.
- JIN, L. et al. **Role explication of simulation in intelligent training systems by training task ontology**. In: Workshop on Architectures for Intelligent Simulation-Based Learning Environments, 1997, Kobe, Japan. [S. l. : s. n.], 1997.
- JOHANSEN, Dag et al. **An Introduction to the TACOMA Distributed System – Version 1.0**. 1995. Disponível por WWW em: <<http://www.cs.uit.no/DOS/Tacoma/>> (12 de dezembro de 2004).
- JSF. **Java Server Faces Technology**. Disponível em: <<http://java.sun.com/javaee/javaserverfaces/index.jsp>>. Acessado em: 10/11/2009.
- KAMII, C. **A criança e o número**. Campinas: Papirus, 1985:103,108.
- KANT, I. **Fundamentos da metafísica dos costumes**. Trad. P. Quintela. Lisboa: Edições 70. 1986.
- KENDALL, M. & B. SAMWAYS, T. van Weert (Ed.), J. Wibe (2003). **IFIP TC3 Position Paper on Lifelong Learning**. In: (Weert, Tom J. van & Mike Kendall, 2003).
- LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de Metodologia Científica**. 6ª. ed. 7 reimpr. São Paulo: Atlas, 2009.
- LAWRENCE, Stevie. Context in Web Search. IEEE Data Engineering Bulletin, Volume 23, Number 3, pp. 25–32, 2000.
- LECKY-THOMPSON, J. W. **Software Agents Mail List**. 1997. Disponível em: http://www.ee.mcgill.ca/~agent_faq.html (30 de julho de 1997).
- LITWIN, E. Das tradições à virtualidade. In: Litwin, Edith (Org.), **Educação a distância**. Porto Alegre: Artmed, 2001.
- LUCENE. **Welcome to Lucene**. Disponível em: <<http://lucene.apache.org/>>. Acessado em: 11/10/2009.
- LYMAN, L. Literacy education as library community service. **Library Trends**, v. 28, n. 2, p. 193-217, 1979.
- MAEDCHE, A.; VOLZ, R. **The text-to-onto ontology extraction and maintenance environment to appear**. In: Proceedings of the ICDM Workshop on Integrating Data Mining and Knowledge Management, 2001, San Jose, California. [S. l. : s.n.], 2001.

MAIA, Camem; MATTAR, João. **ABC da EaD**. São Paulo: Pearson Prentice Hall, 1. Ed., 2007.

MARTIN, P. H.; EKLUND, P. **Large-scale cooperatively-built heterogeneous KBs**. In: ICCS'01 International Conference on Conceptual Structures, 9. 2001. Disponível em: <<http://meganesia.int.gu.edu.au/~phmartin/WebKB/doc/papers/iccs01/>>. Acesso em: 6 out. 2008.

MILL, J. S. **On Liberty**. Boston: Collier, 1909.

NAKAYAMA, Lauro; ALMEIDA, Vinícius Nóbile; VICCARI, Rosa Maria. **A Contextual Information Retrieval Service for Educational Environments**. In: International Conference on Service Systems and Service Management (ICSSSM'04), 2004, Pequim. Beijing: International Academic Publishers, 2004. v. II, p. 636-641.

NAKAYAMA, Lauro; VICCARI, Rosa Maria; COELHO, Helder. An Information Retrieving Service for Distance Learning. **Transactions On Internet Research**, v. 1, p. 49-56, 2005.

OECD. Organization for Economic Co-operation and Development. **Lifelong Learning. Policy Brief**. 2004.

OLSON, Steven Douglas. **AJAX com Java**. Editora Alta Books. Ed. 1. 224p. São Paulo. 2007.

PAL, Sujit. **Jtmt - Java Text Mining Toolkit**. Disponível em: <<http://jtmt.sourceforge.net/>>. Acessado em 10/11/2009.

PIAGET, J. **Conjunto de sua Obra entre 1929 e 1979**. Université de Genève. Presses Universitaires de France.

PIAGET, J. (1994). **O juízo moral na criança** (2ª ed., E. Lenardon, Trad.). São Paulo: Summus. (Trabalho original publicado em 1932)

PAPET, S. A. **Máquina das Crianças: repensando a escola na era da informática**. 1ª ed., Porto Alegre. Artes Médicas. 1994.

PETERS, Otto, **A Educação a Distância em Transição**. São Leopoldo – Rio Grande do Sul, Ed. Unisinos, 2004.

POLI, R.; LANGDON, W. B.; Mcphee, N. F. **A field guide to genetic programming**. Publicado via <http://lulu.com>. Disponível em: <<http://www.gp-field-guide.org.uk>>. ISBN: 978-1-4092-0073-4. Ed. Softcover. San Francisco - USA. 2008.

PORTO, Paulo Ricardo Prestes; PALAZZO, Luiz Antonio Moro; CASTILHOS, José Mauro Volkmer de. **Agentes de Informação Inteligentes**. In: **Oficina de Inteligência Artificial**, Pelotas, 1., 1997. Anais. Pelotas: Ed. EDUCAT, 1997. p. 81-107.

QUARESMA, P. **Inferência de Atitudes em Diálogos**, Tese (Doutorado Engenharia de Informática) – Universidade Nova de Lisboa, 1997.

RAFFERTY, C. D. Literacy in the information age. **Educational Leadership**, v. 5, n. 2, p. 22-25, 1999.

RAMOZZI-CHIAROTTINO, Zélia. **Psicologia e epistemologia genética de Jean Piaget**. Zélia Ramozzi-Chiarottino. - São Paulo: EPU, 1988.

RAMQUIST, R. **Definition of Learning Activity Pages - LAPS**, James Madison University, Harrisonburg, VA, USA, disponível em: <<http://www.eml.jmu.edu/LAPS/LAPS.defined.htm>>. Acessado em 03/07/2007.

REZENDE, S. O. (Org.) . **Sistemas Inteligentes: Fundamentos e Aplicações**. 1. ed. Barueri, SP: Editora Manole Ltda, 2003. v. 1. 560 p.

ROSÁRIO, Bárbara. **Latent Semantic Indexing: An overview**. Infosys 240, Final Paper. 2000.

SALTON, G. MCGILL, M. J. **Introduction to Modern Information Retrieval**. McGraw-Hill Book Co. New York. 1983.

SEARLE, J. R. (2002), **Intencionalidade**, 2. ed.. São Paulo: Martins Fontes, 390p., 2002.

SEIXAS, Louise Margueite Jeanty. **Estratégias Pedagógicas para um Ambiente Multi-agente Probabilístico Inteligente de Aprendizagem – AMPLIA**. 2005. 150p. Tese (Doutorado em Informática na Educação) - Programa de Pós-graduação em Informática na Educação, UFRGS, Porto Alegre.

SCHLEMMER, Eliane. **Projetos de Aprendizagem Baseados em Problemas: uma metodologia interacionista/construtivista para formação de comunidades em Ambientes Virtuais de Aprendizagem**. Colabor@ - Revista digital da CVA – RICESU. Colabora, Curitiba, v.1, n.2 - p. 10-21 , novembro 2001.

SHUM, S. B. E, MOTTA; DOMINGUE, J. **ScholOnto: An ontology-based digital**

library server for research documents and discourse. *International Journal on Digital Libraries*, v. 3, n. 3, p. 237-248, Sept. 2000.

SILVEIRA, Ricardo A. **Modelagem Orientada a Agentes Aplicada a Ambientes Inteligentes Distribuídos de Ensino: JADE**. Tese de doutorado. Porto Alegre: CPGCC da UFRGS, 2001.

SMITH, B. **An Introduction to Ontology**. In: PEUQUET, D.; PEUQUET, B.; BROGAARD, B. (Eds.). *The Ontology of Fields*, NCGIA, Bar Harbor, ME. p. 10-14.

SOARES, Magda. **Alfabetização e Letramento**. São Paulo. Contexto. 2003a.

SOARES, Magda. Letramento e Escolarização. In RIBEIRO. V. M. et. al. (org) **Letramento no Brasil: reflexões a partir do INAF 2001**. São Paulo: Global. P. 89 – 113. 2003b.

SOARES, Magda. **Letramento: um tema de três gêneros**. Belo Horizonte. Autêntica. 2003c.

TFOUNI, Leda. **Letramento e Alfabetização**. São Paulo: Cortez. 1995.

VERGARA, Sylvia Constant. **Projetos e relatórios de pesquisa em administração**. 10. ed. São Paulo: Atlas. 2009.

WANG, Peiling. SOERTEL, Dagobert. **A Cognitive Model of Document Use During a Research Project. Study I. Document Selection**. *JASIS* 49(2): 115-133. 1998.

WEBBER, S. and JOHNSTON, B. **Conceptions of information literacy: new perspectives and implications**. *Journal of information science*. 26 (6). 381-397. 2000.

WOOLDRIDGE, M.; JENNINGS, N. **Agents Theory, Architectures and Languages: A Survey**. In: WOOLDRIDGE, M.; JENNINGS, N. (Eds.). 1995.

WOOLDRIDGE, M.; JENNINGS, N. R. **Software Agents**. Published in *IEE Review*, Janeiro 1996, pp 17-20.

WORDNET. **WordNet – About**. Disponível em: <<http://wordnet.princeton.edu/>>. Acessado em: 11/10/2009.

W3C. **World Wide Web Consortium**. Disponível em <<http://www.w3.org/>>. Acessado em 02/06/2007.

YAHOO. **Yahoo! Brasil**. Disponível em: <<http://www.yahoo.com.br>>. Acesso em: 10/11/2009.

ZDRAHAL, Z. et al. **Sharing engineering design knowledge in a distributed environment**. Behaviour and Information Technology, v. 19, n. 3, p. 189-200, 2000.