

Universidade de Brasília - UnB
Faculdade UnB Gama - FGA
Engenharia de Software

Sistema de Buscas em Ambiente Virtual de Aprendizagem baseada em Perfis

Autor: Pedro Henrique Potiguara Carvalho
**Orientador: Professor Doutor Sérgio Antônio Andrade de
Freitas**

Brasília, DF
2014



Pedro Henrique Potiguara Carvalho

Sistema de Buscas em Ambiente Virtual de Aprendizagem baseada em Perfis

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Orientador: Professor Doutor Sérgio Antônio Andrade de Freitas

Brasília, DF

2014

Pedro Henrique Potiguara Carvalho
Sistema de Buscas em Ambiente Virtual de Aprendizagem baseada em Perfis/
Pedro Henrique Potiguara Carvalho. – Brasília, DF, 2014-
86 p. : il. (algumas color.) ; 30 cm.

Orientador: Professor Doutor Sérgio Antônio Andrade de Freitas

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB
Faculdade UnB Gama - FGA , 2014.

1. Palavra-chave01. 2. Palavra-chave02. I. Professor Doutor Sérgio Antônio
Andrade de Freitas. II. Universidade de Brasília. III. Faculdade UnB Gama. IV.
Sistema de Buscas em Ambiente Virtual de Aprendizagem baseada em Perfis

CDU 02:141:005.6

Pedro Henrique Potiguara Carvalho

Sistema de Buscas em Ambiente Virtual de Aprendizagem baseada em Perfis

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Trabalho aprovado. Brasília, DF, 17 de novembro de 2014:

**Professor Doutor Sérgio Antônio
Andrade de Freitas**
Orientador

**Professor Doutor Edson Alves da
Costa Júnior**
Convidado 1

**Professor Doutora Carla Silva Rocha
Aguiar**
Convidado 2

Brasília, DF
2014

Eu dedico esse trabalho à minha família, meus amigos e meus professores, que durante essa jornada puderam me ensinar e me conduzir em sabedoria.

*“O temor do Senhor é o principio da sabedoria;
todos os que cumprem os seus preceitos
revelam bom senso.
Ele será louvado para sempre! ”
(Bíblia Sagrada, Salmos 111, 10)*

Resumo

Este trabalho tem como objetivo a modelagem e a implementação da parte da busca de um ambiente virtual de aprendizagem, levando em consideração tanto o perfil de grupo do usuário que está realizando a busca, quanto o seu perfil individual. Este tipo de modelagem da busca permite que sejam apresentados ao usuário, como resultado da busca, documentos mais próximos da sua necessidade e portanto, mais relevantes para ele. Ao decorrer do trabalho é descrita a proposta do trabalho, bem como o processo de desenvolvimento adotado para o desenvolvimento do trabalho. A metodologia utilizada para a realização do trabalho gira em torno de estudos em implementações já realizadas para a definição de perfil de grupo e perfil individual, além da composição (integração) dessas implementações e da construção da busca associada a essa composição. Como resultado deste trabalho foi desenvolvido um ambiente capaz de apresentar páginas html com conteúdos relevantes ao usuário, a partir de uma busca realizada por ele e baseada em seu perfil.

Palavras-chaves: Perfis de busca. Perfil de Grupo. Perfil Individual. Ambiente virtual de aprendizagem.

Abstract

This paper aims at modeling and implementing the part of the search for a virtual learning environment, considering the group profile of the user performing the search and his individual profile. This type of modeling allows the search to be presented to the user as a result of the search, closest documents on his need and therefore more relevant to him. In the course of this work is described in the proposed work as well as the development process adopted for the engagement. The methodology used to carry out the work revolves around studies on implementations already undertaken to define individual and group profile profile in addition to the composition (integration) of these implementations and the construction of the search associated with this composition. As a result of this work we developed an environment capable of displaying html pages with relevant content to the user from a search conducted by him, and based on your profile.

Key-words: Search profile. Group Profile. Personal Profile. Virtual Learning Environment.

Lista de ilustrações

Figura 1 – Execução dos elementos de um motor de busca	30
Figura 2 – Relação de número de cursos à distância (virtuais) por ano no Brasil (ABRAEAD, 2008)	35
Figura 3 – Representação visual dos objetos de aprendizagem e áreas de conhecimento	40
Figura 4 – Representação visual dos objetos de aprendizagem relacionados com as áreas de conhecimento	41
Figura 5 – Representação visual da intersecção de objetos de aprendizagem entre duas áreas de conhecimento	41
Figura 6 – Caso de uso do sistema FGAVirtual	43
Figura 7 – DER do Sistema <i>SmartQuizz</i>	47
Figura 8 – DER representando as tabelas do Sistema FGAVirtual juntamente com as tabelas do <i>SmartQuizz</i>	50
Figura 9 – Trecho de código do procedimento que atualiza o perfil individual do usuário na área de conhecimento de matemática	53
Figura 10 – Diagrama de blocos que representa a sequência dos passos realizados durante o processo de uma busca no FGAVirtual	54
Figura 11 – Página de login do sistema FGAVirtual	56
Figura 12 – Mensagem lançada quando há um erro nos dados de login inseridos pelo usuário	56
Figura 13 – Página de busca do sistema FGAVirtual	57
Figura 14 – Página de apresentação dos resultados do sistema FGAVirtual	57

Lista de tabelas

Tabela 1 – Tabela de relação entre os pesos das áreas de conhecimento associados aos perfis de grupo	32
Tabela 2 – Tabela representando a estória de usuário relacionada à funcionalidade de Realizar Busca	44
Tabela 3 – Tabela representando a estória de usuário relacionada à funcionalidade de Realizar Login	44
Tabela 4 – Tabela representando as tarefas criadas para a implementação da funcionalidade de Realizar login	45
Tabela 5 – Tabela representando as tarefas criadas para a implementação da funcionalidade de Realizar Busca	46
Tabela 6 – Tabela de ferramentas utilizadas para a elaboração e execução do sistema	47
Tabela 7 – Tabela associando as áreas de conhecimento do perfil de Grupo com os conceitos de teste inseridos no sistema	52
Tabela 8 – Tabela que representa o resultado final da ordenação utilizando o termo “quimica” em uma busca com o perfil de grupo de Automotiva e perfil individual de 25% e 50%	60
Tabela 9 – Tabela que representa o resultado final da ordenação utilizando o termo “quimica” em uma busca com o perfil de grupo de Automotiva e perfil individual de 75%	60
Tabela 10 – Tabela que representa o resultado final da ordenação utilizando o termo “computação” em uma busca com o perfil de grupo de Software e perfil individual de 25%	61
Tabela 11 – Tabela que representa o resultado final da ordenação utilizando o termo “computação” em uma busca com o perfil de grupo de Software e perfil individual de 50%	61
Tabela 12 – Tabela que representa o resultado final da ordenação utilizando o termo “computação” em uma busca com o perfil de grupo de Software e perfil individual de 75%	62
Tabela 13 – Tabela que representa os documentos mais relevantes para cada uma das áreas de conhecimento envolvidas na busca “matemática computação”	63
Tabela 14 – Tabela que representa os documentos mais relevantes para cada uma das áreas de conhecimento envolvidas na busca “matemática elétrica” .	63

Tabela 15 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática computação” em uma busca com o perfil de grupo de Software e perfis individuais de 0% para ambas as áreas de conhecimento de matemática e computação	64
Tabela 16 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática computação” em uma busca com o perfil de grupo de Software e perfis individuais de 0% e 100% para as áreas de conhecimento de matemática e computação, respectivamente	64
Tabela 17 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática computação” em uma busca com o perfil de grupo de Software e perfis individuais de 100% e 0% para as áreas de conhecimento de matemática e computação, respectivamente	65
Tabela 18 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática computação” em uma busca com o perfil de grupo de Software e perfis individuais de 100% e 100% para as áreas de conhecimento de matemática e computação, respectivamente	65
Tabela 19 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática elétrica” em uma busca com o perfil de grupo de Energia e perfis individuais de 0% para ambas as áreas de conhecimento de matemática e elétrica e 0% e 100% para as áreas de conhecimento de matemática e elétrica, respectivamente	66
Tabela 20 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática elétrica” em uma busca com o perfil de grupo de Energia e perfis individuais de 100% e 0% para as áreas de conhecimento de matemática e elétrica, respectivamente	67
Tabela 21 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática elétrica” em uma busca com o perfil de grupo de Energia e perfis individuais de 100% e 100% para as áreas de conhecimento de matemática e elétrica, respectivamente	68
Tabela 22 – Tabela que representa o resultado final da ordenação utilizando o termo “química” em uma busca com o perfil de grupo de Eletrônica e perfil individual de 25%	75
Tabela 23 – Tabela que representa o resultado final da ordenação utilizando o termo “química” em uma busca com o perfil de grupo de Eletrônica e perfil individual de 50%	76
Tabela 24 – Tabela que representa o resultado final da ordenação utilizando o termo “química” em uma busca com o perfil de grupo de Eletrônica e perfil individual de 75%	76

Tabela 25 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Energia e perfil individual de 25%	77
Tabela 26 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Energia e perfil individual de 50%	77
Tabela 27 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Energia e perfil individual de 75%	78
Tabela 28 – Tabela que representa os documentos mais relevantes para cada uma das áreas de conhecimento envolvidas na busca "química física"	79
Tabela 29 – Tabela que representa o resultado final da ordenação utilizando os termos "química física" em uma busca com o perfil de grupo de Energia e perfis individuais de 0% e 0% respectivamente	80
Tabela 30 – Tabela que representa o resultado final da ordenação utilizando os termos "química física" em uma busca com o perfil de grupo de Energia e perfis individuais de 0% e 100% respectivamente	81
Tabela 31 – Tabela que representa o resultado final da ordenação utilizando os termos "química física" em uma busca com o perfil de grupo de Energia e perfis individuais de 100% e 0% respectivamente	82
Tabela 32 – Tabela que representa o resultado final da ordenação utilizando os termos "química física" em uma busca com o perfil de grupo de Energia e perfis individuais de 100% e 100% respectivamente	83

Lista de abreviaturas e siglas

AVA	Ambiente Virtual de Aprendizagem
JSF	Java Server Faces
JSP	Java Server Pages
FGA	Faculdade do Gama
MVC	Model View Control

Sumário

1	INTRODUÇÃO	23
1.1	Objetivos	23
1.2	Motivação	24
1.3	Metodologia	24
1.4	Estrutura da Monografia	24
2	ESTADO DA ARTE	27
2.1	Algoritmos de Ordenação	27
2.1.1	Algoritmos de Ordenação Baseada em termos	27
2.1.2	Algoritmos de Ordenação Estática	28
2.2	Motor de Busca	28
2.2.1	Funcionamento de um Motor de Busca	29
2.3	Frameworks de Desenvolvimento Web	30
2.4	Perfis de Busca em Grupo	31
2.5	Perfis de Busca Individual	33
2.6	Ambientes Virtuais de Aprendizagem	34
3	A PROPOSTA	37
3.1	Introdução	37
3.1.1	Perfil de Grupo	38
3.1.2	Perfil Individual	39
3.1.3	União do perfil de Grupo e Individual	39
4	DESENVOLVIMENTO	43
4.1	Planejamento	43
4.2	Implementação do Sistema	46
4.3	Funcionamento do sistema	53
4.4	Testes do Sistema	55
4.5	Avaliação do Sistema	58
4.5.1	Avaliação Utilizando um Termo	59
4.5.2	Avaliação Utilizando dois Termos	62
5	CONCLUSÃO	69
6	TRABALHOS FUTUROS	71

APÊNDICES	73
APÊNDICE A – CONSULTAS UTILIZANDO UM TERMO	75
APÊNDICE B – CONSULTAS UTILIZANDO DOIS TERMOS . . .	79
Referências	85

1 Introdução

Com os avanços tecnológicos atuais, é possível encontrar informações facilmente na internet utilizando grandes motores de buscas, como o Google (GOOGLE, 2004). Mas encontrar informações que sejam de fato relevantes não é uma tarefa fácil.

A grande quantidade de artigos, trabalhos e vídeos presentes no mundo virtual tornou possível encontrar qualquer documento que contenha alguma informação na internet. Todavia, a intenção do usuário ao buscar um documento não é encontrar qualquer documento, mas aquele que contenham informações úteis para a sua necessidade.

Apesar da dificuldade de se filtrar informações na internet, a utilização de ambientes virtuais tem se mostrado como uma forma eficaz para o aprendizado, principalmente pela quantidade e diversidades de conteúdos presentes nela. Construindo-se uma forma de busca de conteúdos em mídia textual digital que garanta a recuperação de informações coerentes ao perfil do usuário, é possível tornar a utilização de ambientes virtuais de aprendizagem mais eficiente.

Através desse trabalho, foi realizada a modelagem e a implementação da parte da busca de um Ambiente Virtual de Aprendizagem (AVA), que seja capaz de apresentar documentos para um usuário, baseando essas buscas em critérios de busca que integrem o perfil de grupo com o perfil individual do usuário.

A metodologia utilizada para a realização deste trabalho envolveu estudos relacionados ao: funcionamento de um motor de busca; trabalhos já existentes relacionados à implementação de perfis; e frameworks de desenvolvimento web. Foi possível através desses estudos definir qual a modelagem para a parte da busca do AVA que foi implementada e as demais características desse sistema.

1.1 Objetivos

Os objetivos deste trabalho podem ser resumidos em:

- Apresentar a modelagem utilizada para desenvolver a parte da busca de um AVA;
- Realizar a integração entre os trabalhos já desenvolvidos para a definição de perfil de grupo e perfil individual de usuário;
- Construir um ambiente de busca capaz de apresentar documentos, levando em consideração o perfil do usuário que realizar a busca.

1.2 Motivação

São motivações para a realização deste trabalho:

- Compreender o funcionamento de motores de buscas;
- Buscar formas de se auxiliar na aprendizagem dos estudantes da FGA;
- Auxiliar professores da FGA a acompanhar os alunos individualmente, identificando dificuldades e acompanhando o seu rendimento durante as disciplinas.

1.3 Metodologia

A metodologia para o desenvolvimento deste trabalho está dividida em 4 fases distintas. A primeira fase diz respeito à fase de estudos, onde foi possível compreender os trabalhos já realizados com relação à implementação de perfis. A segunda fase diz respeito à fase de modelagem, onde foram propostas formas distintas de se realizar a implementação da parte da busca do AVA, levando em consideração as implementações de perfis previamente estudados. A terceira fase diz respeito à integração dos trabalhos estudados, onde foram adaptados os trabalhos de implementação de perfil já realizados, para a sua utilização no AVA. A quarta fase foi a fase de implementação, onde ocorreu a construção de uma interface web capaz de realizar buscas para o usuário, comunicando-se com os diferentes componentes de apoio à busca adaptados.

1.4 Estrutura da Monografia

Esse trabalho possui 4 partes distintas: O Estado da Arte, a Proposta o Desenvolvimento e a Conclusão.

No Estado da Arte (Capítulo 2) são descritas, baseadas em referencial teórico, todas as tecnologias, ferramentas e estudos que foram utilizados para o desenvolvimento deste trabalho. São detalhados alguns tipos de algoritmos de ordenação, o funcionamento de um motor de busca, o que são frameworks de desenvolvimento web, a diferença entre perfis de grupo e perfil individual e ao final é descrito um pouco sobre ambientes virtuais de aprendizagem.

A segunda parte do trabalho trata da proposta (Capítulo 3). Nessa parte, é descrito sistematicamente a solução encontrada para o desenvolvimento do trabalho, a fim de se alcançar o objetivo final.

No desenvolvimento (Capítulo 4) é descrito, em termos de engenharia de Software, todos os passos de construção do sistema, da fase de planejamento até a fase de testes, e

é apresentado o funcionamento do sistema, com a apresentação de imagens representando as telas do sistema e suas funcionalidades.

Na Conclusão (Capítulo 5) é comentado sobre os resultados obtidos neste trabalho, com uma análise geral de cada um desses resultados, ressaltando as vantagens da modelagem utilizada.

Por fim são apresentados os Trabalhos Futuros (Capítulo 6), com uma descrição do que é planejado ser realizado adiante levando em consideração melhorias da aplicação.

2 Estado da Arte

Neste capítulo serão descritos, baseado em referencial teórico, os conceitos, tecnologias e ferramentas necessárias para o desenvolvimento deste trabalho.

2.1 Algoritmos de Ordenação

Ordenar documentos tem sido um tópico de consideração por pesquisadores há muito tempo, e continua sendo foco de pesquisa até os dias atuais (SHARMA; ADHAO; MISHRA, 2013).

Os algoritmos de ordenação são elementos dentro de uma busca responsáveis por organizar os documentos em uma ordem de relevância ao usuário. Isso permite uma maior eficiência na busca do ponto de vista do usuário, pois é possível encontrar aquilo que é buscado gastando-se menos tempo. Apesar de organizar os elementos de uma busca em ordem de relevância, nem sempre é possível garantir que os algoritmos de ordenação trarão para o usuário os documentos específicos para a sua necessidade, ou seu perfil.

Existem diversos tipos de algoritmos de ordenação, e cada um possui suas características para atingir um objetivo dentro de um contexto específico. Eles podem ser divididos em 3 categorias distintas: algoritmos de ordenação estática, algoritmos de ordenação baseada em termos e algoritmo de ordenação dinâmica (MARTINS, 2013). Neste trabalho, serão explorados apenas os algoritmos de ordenação baseados em termos e os algoritmos de ordenação estática.

2.1.1 Algoritmos de Ordenação Baseada em termos

Os algoritmos de ordenação baseada em termos são algoritmos que atribuem uma pontuação a cada documento dentro de um conjunto de documentos, através de cálculos envolvendo distribuição de probabilidade (MARTINS, 2013).

O Terrier (TERRIER, 2011) é um sistema que trabalha com ordenação utilizando algoritmos de ordenação baseada em termos. Suas buscas são baseadas utilizando o paradigma de Divergence from Randomness (DFR). Este paradigma segue o seguinte princípio: quanto maior a divergência da frequência do termo dentro de um documento específico a partir da sua frequência na coleção, maior a informação carregada pela palavra no documento. O peso do termo é inversamente proporcional à probabilidade de encontrar o termo dentro do documento (MARTINS, 2013).

Uma forma de otimização desses tipos de algoritmos é a inclusão de pesos refe-

rentes à perfis específicos de usuários nos cálculos de ordenação. É uma forma eficiente de obter resultados melhores utilizando esses algoritmos, embora nem todos os algoritmos permitam essa configuração, como é o caso do algoritmo BM25, e alguns apresentam aderência parcial e esse tipo de configuração, como é o caso do DFI0. Entre os algoritmos DFRee, DLH e PL2, o algoritmo que apresenta melhores resultados aderindo à uma configuração de busca voltada para perfis é o algoritmo DFRee ([MARTINS, 2013](#)).

2.1.2 Algoritmos de Ordenação Estática

Os primeiros algoritmos de ordenação de páginas web foram baseados em informação prévia sobre os sites. PageRank, HITS, SALSA, RankNet e Frank são exemplos desses algoritmos. Todos eles são algoritmos que utilizam características estáticas de páginas web e, portanto, são denominados como algoritmos estáticos([SHARMA; ADHAO; MISHRA, 2013](#)).

De forma geral, esses algoritmos ordenam os documentos a partir de quantidade de referências que esses documentos recebem, ou seja, aquele documento que é mais popular, ou mais referenciado, acaba se colocando em uma posição de destaque entre os demais. Isso permite que, probabilisticamente, o documento que é de necessidade do usuário seja apresentado entre os primeiros dentro da ordenação. Mas o usuário muitas vezes não tem conhecimento da estrutura da internet, tornando as consultas abstratas e nada específicas, uma vez que são determinadas pela estrutura das páginas web ([SHARMA; ADHAO; MISHRA, 2013](#)). Além do mais, não é possível afirmar que a busca ordenará os documentos especificamente para o perfil do usuário, ou seja, é possível que o usuário nunca encontre aquele documento que procura utilizando esse tipo de ordenação.

2.2 Motor de Busca

Bibliotecas foram as primeiras instituições a adotar sistemas de buscas para a recuperação de informações. Tudo começou com o aumento progressivo de livros e informações que eram antes armazenadas as bibliotecas, tornando cada vez mais difícil a busca de uma determinada informação com eficiência. Usualmente, esses tipos de sistemas especializados em recuperar informações foram desenvolvidos por instituições acadêmicas e posteriormente por fornecedores comerciais. Basicamente, este avanço pode ser determinado por 3 gerações: a primeira geração se refere a quando os sistemas consistiam basicamente em automações das tecnologias já existentes (como cartões catálogo) e basicamente permitiam pesquisas baseadas no nome do autor e o título do livro; na segunda geração foram adicionadas funcionalidades que permitiam a busca por detalhes subjetivos do livro, como seu assunto ou palavras-chave, e foram implementadas também facilidades na busca; por fim, na terceira geração, que atualmente ainda está sendo desenvolvida, é

possível perceber a preocupação com interfaces gráficas, formulários eletrônicos, arquiteturas abertas, etc. (YATES; NETO et al., 1999)(SALTON; MCGILL, 1986).

Motores de busca são ferramentas voltadas para a recuperação de informações para um usuário, a partir de uma base de dados. Utilizando critérios específicos, a sua função é encontrar dados, locais ou na rede, de forma rápida e organizada. Existem algumas diferenças entre motores de busca web e padrão. A principal diferença entre eles é que nos motores de busca web não há a necessidade de se acessar ao próprio texto, e sim ao index, que é uma estrutura de dados que permite agilização de buscas. Sem os índices, seria necessário o armazenamento de todas as páginas web localmente no próprio computador do usuário, ou fazer com que a pesquisa fosse realizada em todas as páginas web durante o tempo que foi realizado a consulta, sem que houvesse um index pronto, ocasionando em uma lentidão intensa(MARTINS, 2013) (YATES; NETO et al., 1999).

Existem diversos motores de buscas. O Terrier (TERRIER, 2011) é um motor de busca em java flexível, implementando as funções de indexação e de recuperação, além de fornecer uma ótima plataforma para desenvolvimento, avaliação e experimentação com relação à recuperação de textos(MARTINS, 2013). Ele possui três formas de execução: a execução interativa, desktop e web (TERRIER, 2011). Todas elas executam utilizando algoritmos de ordenação baseadas em termos, e portanto, a única diferença entre elas é a própria interface, permitindo ao desenvolvedor utilizar aquela interface que melhor se adequar ao seu projeto e detalhes da forma de indexação dos documentos e acesso às classes principais.

2.2.1 Funcionamento de um Motor de Busca

Um motor de busca contém 3 componentes básicos: *web crawler*, *index* e ordenação.

A partir da Figura 1 é possível perceber que os elementos do motor de busca possuem uma ordem de execução. O primeiro elemento a ser acionado após a busca do usuário é o *web crawler*. Em seguida os índices são acessados e por fim ocorre a ordenação para a apresentação dos dados ao usuário.

O *web crawler* (também conhecido como *web spider*), é o componente que auxilia na otimização da performance de uma busca. Sua função é rastrear a internet identificando os *links* de páginas webs. Muitas vezes esses *links* são armazenados localmente para serem utilizados de forma instantânea, sem a necessidade de se acessar a internet novamente. Logo após esse rastreamento, o *index* é o componente responsável por otimizar as buscas. Em termos gerais, ele acessa cada documento rastreado, identificando, recolhendo e analisando alguns dados necessários para a recuperação de informação mais rápida. Por fim, a ordenação é o componente responsável por organizar sequencialmente os documentos em ordem de relevância ao usuário.

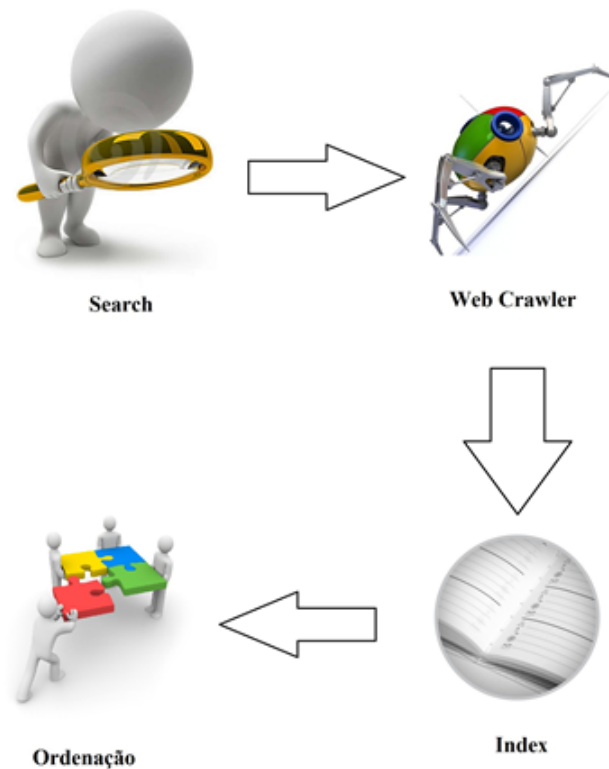


Figura 1 – Execução dos elementos de um motor de busca

2.3 Frameworks de Desenvolvimento Web

Com o avanço dos sistemas de informação na internet surge à necessidade de se desenvolver aplicações web cada vez mais complexas, as quais envolvem fatores como conversões, validações, persistência no banco de dados, mecanismos de segurança, acessibilidade, suporte a Cascading Style Sheets (CSS), suporte a criação e uso de templates, entre diversas outras funcionalidades. Para auxiliar no desenvolvimento de alguns desses fatores existem diversos *frameworks* de desenvolvimento Web (CARMISINI; VAHLICK, 2012).

Segundo Dirk (RIEHLE, 2000), trata-se um modelo de classe, somado com um conjunto de classes integradas e um conjunto de classe de extensão de ponto, sendo que o núcleo seu núcleo é o modelo de classes.

Framework define como os objetos colaboram entre si para representar o domínio que está sendo modelado. Ele capta os aspectos de um domínio que são considerados invariantes dentre um conjunto de colaborações entre objetos que representam situações de domínio concreto. Como qualquer outro modelo, ele sempre se concentra em aspectos específicos do domínio e ignora aqueles considerados irrelevantes para a tarefa de modelagem (RIEHLE, 2000).

Eles podem ser classificado em dois tipo: caixa branca e caixa preta. Aquele que

permite a personalização através de extensões é considerado um caixa-branca. Nesse tipo de framework deve ter um conjunto não vazio de pontos de extensão de classes; caso contrário, é caixa-preta. Em outras palavras, *frameworks* caixa branca são baseados em sistemas de herança presentes no contexto de orientação a objetos, pela implementação de *templates methods*, *hook methods* ou *abstract methods*. Já os caixa preta são instanciados por classes especializadas na implementação de interfaces.

Segundo Dirk, o desenvolvimento de software em geral, e desenvolvimento de aplicações baseadas em *framework* orientado a objeto em particular, é um ramo não-trivial que vem com muitos problemas. Estes incluem questões técnicas, questões sociais e de comunicação, questões de gerenciamento de projetos e questões relacionadas à organização de desenvolvimento (RIEHLE, 2000). Algumas pesquisas indicam que a adoção de técnicas de objeto apresentam grandes problemas dentro de uma organização, até mesmo com relação à produtividade, pois a curva de aprendizado para os desenvolvedores é íngreme, tornando longo o processo até que haja trabalho produtivo.

Um *framework* de desenvolvimento web para o domínio de uma aplicação particular pode também ser mapeado como um framework orientado a objeto (SCHWABE; ESMERALDO, 2001). Inicialmente eram utilizados somente os servlets, e as JSPs (JavaServer Pages) para desenvolvimento web, mas logo alguns padrões começaram a ser utilizados, como o MVC e o *Front Controller*, uma vez que essas formas de trabalhar eram improdutivas e desorganizadas. Atualmente, podem ser encontrados diversos frameworks na internet, como o Struts (STRUTS, 2000), o VRaptor (VRAPTOR, 2003) e o JSF (JSF, 2004).

2.4 Perfis de Busca em Grupo

A inserção de perfis em uma busca permite um relacionamento entre pesos e a ordenação dos resultados de uma busca. Traçando-se um perfil de busca em grupo em uma determinada busca, é possível estabelecer um conjunto de áreas de conhecimentos comuns para um grupo específico de usuários, tornando a busca mais objetiva para esse grupo.

De acordo com Chidlovskii (CHIDLOVSKII; GLANCE; GRASSO, 2001), uma forma de se definir um perfil de grupo de usuários é categorizar documentos de um conjunto de documentos entre os diferentes grupos e, em seguida, classificar os vários documentos de acordo com os perfis individuais dos usuários pertencentes ao grupo. A classificação de documentos dentro do grupo para o usuário varia de acordo com seus níveis de especialização. Vários métodos para aproximar o nível de um usuário de especialização dentro de um grupo pode ser utilizado (por exemplo por meio de estatística, etc.). O sistema denominado System of invention (CHIDLOVSKII; GLANCE; GRASSO,

2001) fornece a capacidade de classificar os resultados retornados em vários motores de busca e a capacidade de levar em conta o contexto de usuário (ou perfil de grupo) através do uso de perfis de usuários, comunidade ou usuário especialista.

Outra forma de se definir perfil de grupo de usuário foi descrito por Martins (MARTINS, 2013), onde foram classificados perfis de grupo baseados nos cursos de engenharia da FGA, determinando diferentes pesos para cada área de conhecimento dentro de um determinado curso de Engenharia, como pode ser visualizado na Tabela 1.

Percebe-se que a primeira coluna da Tabela 1 apresenta os diferentes perfis, ou seja, os diferentes cursos de graduação da FGA (automotiva, eletrônica, energia e de software), e as colunas seguintes apresentam o peso de cada área de conhecimento relacionada ao perfil. Além disso, observa-se também a área de conhecimento de Engenharias aparecendo em 3 dos perfis como a segunda com maior peso, e para o perfil de engenharia automotiva como a terceira com maior peso. Essa área se refere ao conhecimento comum de engenharia entre os 4 cursos, e por isso aparece em todos os perfis com um peso maior.

Perfis	10.000	5.000	1.000	500	100	1
Engenharia Automotiva	Engenharia Mecânica	Matemática	Engenharias	Física	Engenharia Elétrica	Demais Áreas de Conhecimento
Engenharia Eletrônica	Engenharia Elétrica	Engenharias	Matemática	Física	Engenharia Mecânica	Demais Áreas de Conhecimento
Engenharia de Energia	Matemática	Engenharias Mecânica	Engenharias	Física	Química	Demais Áreas de Conhecimento
Engenharia de Software	Computação	Engenharias	Matemática	Engenharia Elétrica	Engenharia Mecânica	Demais Áreas de Conhecimento

Tabela 1 – Tabela de relação entre os pesos das áreas de conhecimento associados aos perfis de grupo

A determinação dos pesos para cada perfil associado à uma área de conhecimento foi realizada a partir de uma análise das grades curriculares de cada engenharia da Faculdade (automotiva, eletrônica, energia e de software), e através desses currículos foi possível analisar qual a porcentagem de matérias que cada um dos cursos possui em determinada área de conhecimento (MARTINS, 2013). Utilizando a ferramenta Terrier (TERRIER, 2011) em sua versão Desktop, esses pesos foram inseridos nos algoritmos padrões da ferramenta (BM25, DFI0, DFRee, DLH, PL2), a fim de se determinar qual dentre esses

algoritmos se adequaria melhor à implementação de perfis em grupo. Por fim, o algoritmo que melhor se adaptou para a implementação de perfis foi o DFRee (MARTINS, 2013).

Este tipo de modelagem permitiu distinguir os pesos de determinadas áreas de conhecimento em cada perfil de aluno dentro do contexto da universidade, ou seja, para um aluno de engenharia automotiva, é possível determinar que a área de conhecimento de engenharia mecânica tem um maior peso do que a área de conhecimento de matemática, que por sua vez possui maior peso do que conceitos gerais de engenharia, e assim por diante.

Dentro do contexto de buscas, é possível perceber que, determinando-se os pesos para cada área de conhecimento relacionada a cada curso e conhecendo-se o perfil do usuário, ou seja, a qual curso (ou perfil de grupo) o usuário está incluído, uma busca pode trazer resultados mais precisos para o usuário como um coletivo, assumindo que a sua necessidade é maior em determinada área de conhecimento do que em outra e, portanto, um documento deve possuir maior prioridade do que outro considerando a ordenação de resultados.

O trabalho desenvolvido nesta monografia utilizada a abordagem de perfil de grupo desenvolvida no trabalho desenvolvido por Martins (MARTINS, 2013).

2.5 Perfis de Busca Individual

Perfis de busca individual tem como objetivo traçar um perfil específico para um determinado usuário, levando em consideração diversos aspectos individuais e características pessoais, a fim de se realizar buscas levando em consideração a necessidade individual do usuário. Dentro de um contexto de aprendizado, por exemplo, através de perfis de busca individual, é possível determinar quais objetos de aprendizagem um aluno necessita maior reforço ou quais o aluno possui maior domínio.

De acordo com o Cavallucci (CAVELLUCCI, 2005), o perfil individual de usuário pode ser definido através da definição de oito características básicas: extroversão ou introversão, sensação ou intuição, pensamento ou sentimento, e julgamento ou percepção. A extroversão indica que a pessoa está ligada ao mundo exterior e às coisas presentes nele. A introversão, pelo contrário, indica que a pessoa está mais voltada ao seu mundo interior, idéias e impressões. A sensação indica que o indivíduo está mais voltado ao presente e às informações obtidas por meio dos seus sentidos. Em contrapartida, a intuição indica que a pessoa está mais voltada ao futuro, aos padrões e possibilidades. O pensamento indica que a pessoa prefere se apoiar na impessoalidade e suas decisões são baseadas na lógica e na análise de causa e efeito. Por outro lado, o sentimento indica que a pessoa baseia suas decisões em valores e avaliações subjetivas. Por fim, o Julgamento indica que o indivíduo prefere abordagens planejadas e organizadas com relação à vida, já na Percepção

ele aprecia abordagens flexíveis e espontâneas, preferindo propostas e opções abertas. A combinação dessas características em cada um de seus polos define o perfil individual do usuário (CAVELLUCCI, 2005).

A experiência adquirida em trabalhos anteriores realizados no âmbito do i-GPDAM (Grupo de Pesquisa e Desenvolvimento de Ambientes Virtuais), com relação à criação de objetos de aprendizagem interativa para um ambiente de aprendizagem virtual para engenheiros, foi utilizada na criação de uma nova ferramenta para a educação: o *SmartQuiz*. Basicamente se refere a uma série de *quizzes* desenvolvidos por estudantes de engenharia da Universidade de Brasília, que atribuem uma pontuação a um indivíduo dependendo da resposta fornecida à uma pergunta. A utilização desse tipo de técnica apresentou resultados bastante úteis: além do desafio de se responder diversas perguntas categorizadas em níveis, os questionários são considerados como uma forma interativa de aprendizagem (FREITAS et al., 2012).

A metodologia de criação do *SmartQuiz* permite traçar o perfil individual de usuário a partir de três partes distintas. A primeira delas é com relação à criação de um mapa conceitual para a correção das respostas do *quiz*, sinalizando as respostas corretas e os conceitos que estão mal compreendidos por trás das respostas incorretas. Em segundo lugar, são desenvolvidas metodologias para a análise e o diagnóstico através das respostas capturadas. Por fim, a última parte trata da criação do perfil individual do usuário. Os resultados preliminares da utilização do *SmartQuiz* suportam a ideia de que essa ferramenta fornece um bom diagnóstico de desempenho individual ou coletivo de estudantes, além de contribuir para estabelecer estratégias de aprendizagem (FREITAS et al., 2012).

O trabalho desenvolvido nesta monografia utiliza o trabalho do *SmartQuiz* (FREITAS et al., 2012) para traçar o perfil individual do usuário que utiliza o sistema.

2.6 Ambientes Virtuais de Aprendizagem

Os avanços tecnológicos estão cada vez mais modificando a forma de se ensinar e aprender. Atualmente, a necessidade de se aprender continuamente deixou de ser de responsabilidade apenas dos mais jovens, pois todos devem se manter continuamente atualizados para permanecer dentro do contexto competitivo e social da sociedade (PEREIRA; SCHMITT; DIAS, 2007).

Segundo Clark (CLARK, 2001), estudos de meta-análise sobre pesquisa de mídia têm mostrado que os alunos ganham benefícios significativos de aprendizagem quando se aprende a partir da mídia áudio-visual ou do computador, ao contrário de ensinamentos convencionais; no entanto, os mesmos estudos sugerem que a motivo desses benefícios não é o meio de instrução, mas sim as estratégias construídas nos materiais de aprendizagem.

Cada vez mais, as organizações estão adotando aprendizagem on-line como o principal método para treinar os funcionários (SIMMONS, 2002). Apesar de ser uma forma muitas vezes cara de se fornecer aprendizagem, deve-se perceber de que o uso de aprendizagem on-line oferece grandes vantagens. Para os alunos, a internet pode ser utilizada para serem obtidos materiais relevantes de aprendizagem atualizados de forma rápida e precisa, além de fornecer um meio direto de comunicação com especialistas na área em que eles estão estudando (ANDERSON, 2008).

De acordo com Fahy (FAHY, 2004), a utilização de diferentes didáticas de aprendizagem, como vídeos, áudios, gráficos e textos, podem apresentar diversas vantagens, entre elas: promover habilidades e formação de conceitos; permite uma melhor compreensão do conteúdo; e facilita a aprendizagem, uma vez que animações e áudio permitem uma compreensão mais consistente do que textos em telas.

Ambientes virtuais de aprendizagem tratam de formas que envolvem opções de mídia como forma de aprendizagem à distância. Pode-se dizer que esta forma de ensino iniciou-se no século XVIII, a partir do oferecimento de cursos por correspondência. Com a popularização da internet, foi possível o desenvolvimento desses ambientes virtuais de forma em que a comunicação entre os participantes não exigisse a presença de ambos, permitindo o ensino a qualquer modalidade, em qualquer momento e em qualquer lugar. (PEREIRA; SCHMITT; DIAS, 2007).

	Ano Declarado																
	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008*
Nº de Cursos	1	1	1	3	6	9	6	10	14	9	10	22	45	78	194	320	8

Figura 2 – Relação de número de cursos à distância (virtuais) por ano no Brasil (ABRAEAD, 2008)

De acordo com o Anuário Brasileiro Estatístico de Educação Aberta e a Distância (ABRAEAD, 2008), houve um aumento significativo de cursos superiores fornecidos à distância desde 1992, como pode ser visualizado na Figura 2. Essa figura indica a quantidade de cursos à distância oferecidos entre os anos de 1992 até 2008. É possível perceber que, em 1992 havia apenas 1 curso à distância oferecido em todo o Brasil. Com o passar o tempo, o número de cursos cresceu e em 2007 já apresentava 320 cursos oferecidos à distância. Isto significa que há uma demanda cada vez mais crescente de estudantes e universidades que adotam esse tipo de aprendizagem, e que a tendência é que se desenvolva uma estrutura cada vez mais consistente que apoie a esses ambientes virtuais de aprendizagem.

As tecnologias e definições aqui apresentadas permitem a construção de um ambiente virtual de buscas, capaz de modelar o perfil de um usuário presente no sistema, e

utilizar este perfil para apresentar resultados que sejam mais próximos da sua necessidade.

3 A proposta

Neste capítulo é descrita a proposta do trabalho, detalhando os passos a serem realizados e considerando o referencial teórico apresentado e estudos já realizados.

3.1 Introdução

A utilização de perfis dentro do contexto de uma busca tem como objetivo tornar seus resultados cada vez mais próximos das características já existentes do usuário e, no caso deste trabalho, das características que seriam desejáveis para o usuário no futuro. No contexto acadêmico, a proposta desse trabalho é modelar e implementar a parte de um AVA em interface web que seja responsável pela busca de conteúdos. Os elementos necessários para a busca devem ser implementados baseados na utilização de perfis individual e perfil de grupo que permitam a modelagem de um usuário existente no ambiente. Além disso, o sistema deve contar com um módulo responsável por realizar o download de documentos diretamente do Google, de acordo com as palavras-chave inseridas pelo usuário no sistema, a fim de construir a base de documentos a serem apresentados ao final da busca dentro de um contexto real de aplicação.

A complexidade deste trabalho se dá principalmente pela definição de como relacionar o perfil individual do usuário com o seu perfil de grupo com a finalidade de apresentar documentos específicos para este usuário. O perfil de grupo representa o perfil esperado que o aluno tenha ao final do curso, ou seja, se a área de computação é a mais importante dentro do curso de engenharia de software, é esperado que um aluno de engenharia de software ao final do curso possua um conhecimento maior na área de computação. Em outras palavras, o perfil de grupo indica um parâmetro de conhecimentos que um aluno em determinada engenharia necessita ter. Em contrapartida, o perfil individual indica o conhecimento atual do aluno nas diversas áreas de conhecimento. Com este perfil, é possível identificar quais áreas de conhecimento o aluno possui maior domínio, e quais ele possui menor conhecimento. Comparando-se portanto o perfil individual do usuário com o perfil de grupo deste mesmo usuário, é necessário que a busca do AVA retorne certos documentos para o usuário com o intuito de transformar o perfil individual do usuário o mais próximo possível do seu perfil de grupo, uma vez que ele aprenda o conteúdo dos documentos apresentados pela busca. Por exemplo, considerando que um aluno de engenharia de software possua bastante conhecimento em computação, mas em contrapartida seu conhecimento em conceitos básicos de engenharia é relativamente fraco. Se for observado o perfil de grupo definido para engenharia de software, a área de conhecimento de computação é a área mais importante, seguido de engenharias (MARTINS, 2013). Por-

tanto, se a busca fosse associada apenas ao perfil de grupo do usuário, os documentos recuperados seriam prioritariamente para a área de conhecimento de computação, seguidos da área de conhecimento de engenharias. Mas se for observado o perfil individual do usuário e for comparado ao seu perfil de grupo, seria mais coerente que fossem recuperados documentos relacionados à engenharias e não à computação, visto que o usuário já possui conhecimento na área de computação. A questão, portanto, é modelar como essa busca deveria ser realizada para atender à essa necessidade e direcionar o resultado da busca para a área de conhecimento que seja realmente mais coerente ao usuário. Essa busca deve então ponderar os dados coletados do perfil individual do usuário e do perfil de grupo do usuário e incluir na busca de forma a aumentar os pesos dos documentos satisfazendo a necessidade do usuário.

É importante ressaltar que, apesar do trabalho desenvolvido para a definição de perfil de grupo (MARTINS, 2013) ser um trabalho já relacionado à busca de documentos, o *SmartQuiz* (FREITAS et al., 2012) apenas modela os potes de conhecimento do usuário, e portanto não possui nenhuma implementação de buscas associada. Em outras palavras, o trabalho que define o perfil de grupo (MARTINS, 2013) além de definir o que são os perfis de grupo também já possui um sistema de busca que utiliza os perfis de grupo para apresentar documentos para o usuário, já o *SmartQuiz* (FREITAS et al., 2012) é um trabalho que apenas modela o perfil individual do usuário. É necessário que o sistema associe o perfil individual do usuário à buscas, ou seja, a partir dos potes de aprendizagem definidos pelo *SmartQuiz*, o sistema inclua esses potes na busca de tal forma que sejam apresentados documentos específicos para o perfil individual do usuário. O sistema portanto deve possuir uma lógica de pesos, assim como aqueles que definem o perfil de grupo, e esses pesos serão utilizados na busca para garantir que, em uma porcentagem aceitável, uma busca para um usuário específico irá recuperar documentos que satisfaçam a sua necessidade individual. Por fim com esta lógica de pesos implementada, é possível realizar uma relação entre os pesos dos perfis individual e de grupo para garantir a ordenação correta dos documentos específicos para o usuário.

Nos tópicos a seguir são discutidas as definições dos perfis de grupo e individual utilizados nos algoritmos de buscas neste trabalho e qual foi a lógica utilizada neste trabalho para unir os perfis individual e de grupo de usuário para a realização da busca.

3.1.1 Perfil de Grupo

O perfil de grupo é proveniente do trabalho desenvolvido (MARTINS, 2013) que aplicou determinados pesos para cada área de conhecimento relacionada a uma engenharia definida, dentre as 4 engenharias da FGA (Eletrônica, Software, Automotiva, Eletrônica). Os pesos foram classificados a partir de uma análise curricular das disciplinas de cada engenharia, ou seja, aquelas áreas de conhecimento que possuíam maior número de disci-

plinas para um determinado curso receberam um peso maior, em contrapartida, as áreas com menor número de disciplinas receberam um peso menor. Por exemplo, no domínio da engenharia de software, a área de computação é mais importante do que a área de química, uma vez que, ao analisar a matriz curricular de engenharia de software haviam muitas disciplinas voltadas à área de computação do que disciplinas voltadas para a área de química, e assim por diante. Esses pesos aplicados aos algoritmos de busca permitiram, em uma margem satisfatória, garantir que fossem apresentados em uma busca primeiramente os documentos cuja área de conhecimento fosse mais importante ao perfil do usuário que realiza a busca.

3.1.2 Perfil Individual

O perfil individual de usuário é definido a partir do sistema *SmartQuiz*, ou seja, o Sistema de buscas inteligentes (FREITAS et al., 2012). O *SmartQuiz* é um sistema onde são apresentadas ao usuário uma série de perguntas, e a cada resposta é realizada uma avaliação de conhecimento, identificando conceitos aprendidos ou necessários para o usuário por trás de sua resposta. Este sistema é composto de potes de aprendizagem, que representam o quanto o usuário já aprendeu e o que o usuário ainda necessita aprender sobre um determinado objeto de aprendizagem. Dessa forma, é possível atribuir uma pontuação para cada pote de aprendizagem através da avaliação de cada pergunta que o usuário responda e por fim é possível determinar quais são os conceitos melhores compreendidos pelo usuário e quais é necessário uma melhor aprendizagem.

3.1.3 União do perfil de Grupo e Individual

Esta sessão apresenta o relacionamento entre os perfis de grupo e individual no processo da busca, garantindo uma ordenação dos documentos finais de uma forma mais relevante ao usuário.

Neste trabalho, o conhecimento é o que define os perfis de grupo e o perfil individual. O perfil de grupo indica quais áreas de conhecimento são mais importantes para o usuário, ou seja, em quais áreas de conhecimento o usuário deve possuir maior conhecimento, considerando qual o seu curso. Portanto, é produto do perfil de grupo quais áreas de conhecimento são mais relevantes para o usuário específico. No mesmo sentido, o perfil individual nos apresenta qual é o conhecimento do usuário em cada objeto de aprendizagem, considerando o seu desempenho no sistema *SmartQuiz*, e o valor de cada pote de aprendizagem em um dado instante.

A complexidade do trabalho se dá inicialmente ao integrar sistemas diferentes que trabalham com formas diferentes de classificar conhecimentos. Isto é claro ao observar como os diferentes sistemas armazenam a informação de conhecimento referente ao usuá-

rio. Enquanto o perfil individual representado pelo sistema *SmartQuiz* é definido através de potes de aprendizagem associados a objetos de aprendizagem, o perfil de grupo é definido por áreas de conhecimento, ou seja, o perfil de grupo é apresentado através de uma representação mais geral e ampla de conhecimentos. Apesar da diferença, é possível unir os diferentes sistemas realizando um mapeamento de objetos de aprendizagem e relacionando-os a cada área de conhecimento. Por exemplo, considere a Figura 3.

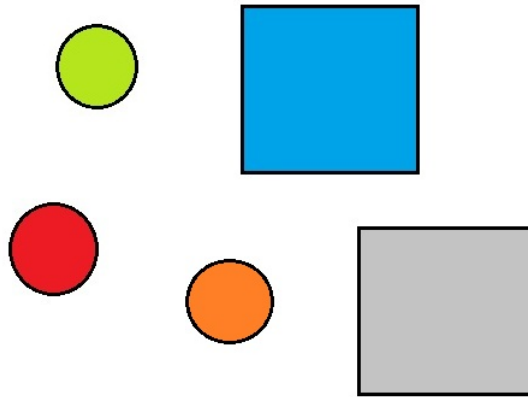


Figura 3 – Representação visual dos objetos de aprendizagem e áreas de conhecimento

A Figura 3 contém uma representação de objetos de aprendizagem e áreas de conhecimento. Considere que os quadrados representam as áreas de conhecimento e os círculos os objetos de aprendizagem. Apesar de objetos de aprendizagem e áreas de conhecimento serem unidades diferentes, neste trabalho estas entidades foram relacionadas assim como no trabalho desenvolvido (LEFFA, 2006), levando em consideração que um objeto de aprendizagem pertence a uma ou mais áreas de conhecimento. Por exemplo, a área de conhecimento de matemática é formada pelo objeto de aprendizagem de álgebra, geometria, e diversos outros objetos. Dessa forma são identificados quais os objetos de aprendizagem que definem cada área de conhecimento. Isso permite uma associação entre os objetos de aprendizagem presentes no sistema *SmartQuiz* e as distintas áreas de conhecimento representadas pelo perfil de Grupo, como pode ser visualizado na Figura 4.

Percebe-se na Figura 4 um mapeamento dos objetos de aprendizagem referentes a cada área de conhecimento. Os objetos de aprendizagem representados pelos círculos das cores verde e vermelho definem a área de conhecimento representada pelo quadrado de cor azul, enquanto o círculo laranja representa o objeto de aprendizagem que define a área de conhecimento representada pelo quadrado de cor cinza. Dessa forma, pode-se estimar uma representação numérica de quanto o usuário já conhece de uma área de conhecimento específica através de uma razão da quantidade de conhecimento que um usuário possui nos objetos de aprendizagem que definem esta área de conhecimento. O Sistema *SmartQuiz* possui, para cada objeto de aprendizagem, um pote de aprendizagem. Este pote representa

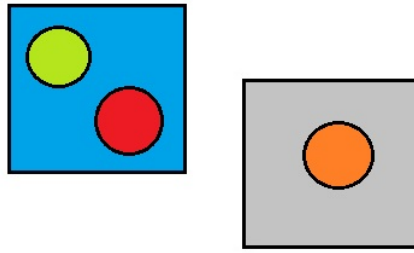


Figura 4 – Representação visual dos objetos de aprendizagem relacionados com as áreas de conhecimento

quanto o usuário já aprendeu sobre o objeto de aprendizagem em questão. Realizando a soma de quanto o usuário aprendeu sobre cada objeto de aprendizagem associado à uma área de conhecimento específica, e dividindo o resultado pela quantidade de potes que definem esta área de conhecimento, o resultado disto será a representação numérica da média de quanto o usuário já aprendeu acerca desta área de conhecimento.

Muitas vezes um objeto de aprendizagem não pertence apenas a uma área de conhecimento, e sim uma ou mais. Por exemplo, ao analisar o objeto de aprendizagem de derivada, identifica-se que este objeto pode pertencer às áreas de conhecimento de matemática e ao mesmo tempo à área de física. A Figura 5 representa esta situação.

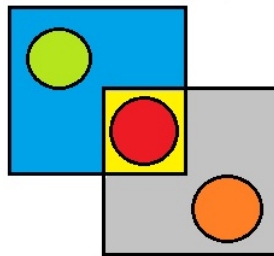


Figura 5 – Representação visual da intersecção de objetos de aprendizagem entre duas áreas de conhecimento

A área demarcada de amarelo da figura acima representa a intersecção de conhecimento entre áreas de conhecimento representadas pelos quadrados de cor azul e cinza. Apesar dessas áreas de conhecimento possuírem seus próprios objetos de aprendizagem, determinados pelas cores verde e laranja, há um objeto de aprendizagem na área amarela representado pela cor vermelha que pertence às duas áreas de conhecimento. Para este caso, é necessário um cuidado maior ao se realizar a média de conhecimento do usuário em uma área de conhecimento. Este cuidado é necessário pois, já que este objeto é compartilhado pelas duas áreas de conhecimento, este objeto possivelmente será mais importante para uma das áreas de conhecimento do que para a outra, e neste caso deve-se adotar uma forma de pesos para distribuir a importância deste objeto de aprendizagem nas áreas de conhecimento que o compartilham, como será explicado mais à frente neste trabalho.

4 Desenvolvimento

Neste capítulo é descrito todo o processo de desenvolvimento do trabalho, detalhando os passos que foram realizados desde o início do projeto e o funcionamento da busca no sistema.

4.1 Planejamento

Para o sistema FGAVirtual, foi elaborado um diagrama de caso de uso, como pode ser visualizado na Figura 6.

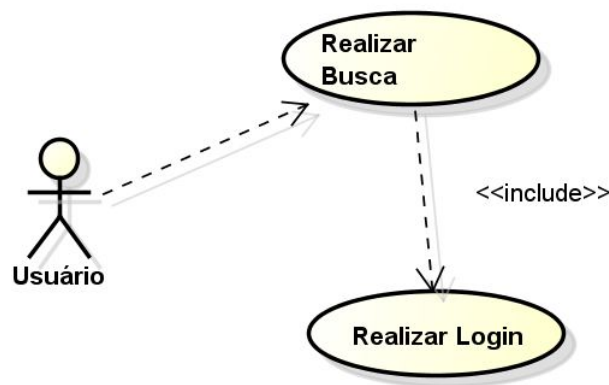


Figura 6 – Caso de uso do sistema FGAVirtual

O diagrama de caso de uso possui como objetivo demonstrar visualmente as funcionalidades do sistema e a interação do usuário com o sistema. Cada personagem do caso de uso representa uma pessoa ou um outro sistema que participa das funcionalidades do sistema e cada círculo representa uma funcionalidade do sistema.

No caso de uso do sistema FGAVirtual demonstrado na Figura 6, existe apenas um personagem que é o usuário. Este usuário possui duas funcionalidades correlacionadas no sistema, são elas: Realizar Busca e Realizar Login. Essas funcionalidades estão correlacionadas pois, para que seja possível realizar uma busca no sistema, é necessário saber quem é o usuário que está realizando a busca, e através disso, identificar qual o seu perfil de grupo e o seu perfil individual, ou seja, quais são os pesos a serem aplicados durante a busca. Esta relação é demonstrada através do “include” presente no diagrama, o que indica que na funcionalidade de Realizar Busca está inclusa a funcionalidade de Realizar Login.

Apesar de ser um diagrama simples, a complexidade do sistema está na aplicação dos pesos representados pelo perfil individual e de grupo do usuário no momento da busca. A funcionalidade do sistema limita-se à realização de buscas e apresentação dos resultados como resultado da busca.

O sistema FGAVirtual foi desenvolvido utilizando a metodologia ágil. Para isto, foram criadas algumas histórias de usuário. As histórias de usuário indicam quais são as funcionalidades do sistema do ponto de vista do próprio usuário. Elas podem ser escritas por tabelas contendo na primeira linha qual é o personagem envolvido na funcionalidade, na segunda linha qual é a meta daquela funcionalidade e por fim na última linha qual é o motivo da funcionalidade. Esta formalização auxilia no momento de criação das tarefas durante o processo de desenvolvimento, pois é ressaltado qual o objetivo de cada funcionalidade, além de auxiliar no momento de verificar se a funcionalidade foi desenvolvida corretamente e se ela está atendendo às necessidades do envolvido.

Como o sistema é composto apenas das funcionalidades de Realizar Busca e Realizar login, foram criadas duas histórias de usuário, cada uma relacionada à uma das funcionalidades, que pode ser visualizada nas Tabelas 2 e 3.

Como	Usuário
Quero	Buscar documentos específicos para a minha necessidade
Para que eu possa	aprender um conteúdo através desses documento.

Tabela 2 – Tabela representando a história de usuário relacionada à funcionalidade de Realizar Busca

Como	Usuário
Quero	Realizar login no sistema
Para que eu possa	Realizar a busca de documentos.

Tabela 3 – Tabela representando a história de usuário relacionada à funcionalidade de Realizar Login

Com as histórias de usuário desenvolvidas, foi utilizada a técnica de *Planning Poker*. Esta técnica busca atribuir um esforço para cada uma das histórias de usuário, a fim de se dividir o trabalho ao longo do processo de desenvolvimento do sistema. Para este trabalho, foi utilizada a sequência 0, 1, 2, 3, 5, 8, 13, 20, 40, 100 para atribuir os pontos à cada uma das histórias, resultando na pontuação 20 para a estória de usuário de “Realizar Busca” e 8 para a estória de usuário de “Realizar login”. A diferença se deu por causa

da complexidade associada à utilização dos perfis durante a busca, o que exigiria mais esforço do que a funcionalidade de autenticar o usuário no sistema.

Após a realização das estimativas de trabalho utilizando o *Planning Poker*, foram divididas as tarefas a serem realizadas durante cada uma das funcionalidades. Para a funcionalidade de realizar login, foram criadas as tarefas que podem ser visualizadas na Tabela 4.

Tarefa	Dificuldade
Criar tabela do usuário no banco de dados	1
Criar tela de login	2
Realizar comunicação entre o banco de dados e a interface	4
Criar tabela no banco de dados que representa o perfil de grupo do usuário	1
Criar tabela no banco de dados que representa o perfil de individual do usuário	1
Apresentar visualmente erro de login na interface ao usuário	2

Tabela 4 – Tabela representando as tarefas criadas para a implementação da funcionalidade de Realizar login

Como pode ser observado na Tabela 4, foram criadas 6 tarefas para o desenvolvimento da funcionalidade de Realizar login, e cada uma das tarefas possuem uma dificuldade associada. A dificuldade foi definida utilizando valores entre 1 e 5, onde 1 representa uma tarefa de baixa dificuldade e 5 de muita dificuldade. Os valores foram atribuídos levando em consideração a necessidade de estudo e tempo de desenvolvimento para a implementação correta da tarefa.

Na Tabela 5 são representadas as tarefas referentes à implementação da funcionalidade de Realizar busca.

Percebe-se que a funcionalidade de Realizar busca possui mais tarefas do que a funcionalidade de Realizar login, e em geral suas tarefas possuem maior dificuldade. Isto justifica o fato da funcionalidade de Realizar busca possuir uma maior pontuação no *planning poker*, uma vez que suas atividades possuem uma maior complexidade de execução.

Além da definição de tarefas para as histórias de usuário, o desenvolvimento deste trabalho contou com reuniões semanais com o orientador do trabalho de conclusão de

Tarefa	Dificuldade
Criar tela de busca	2
Criar módulo para download dos documentos em formato HTML do Google	5
Realizar indexação automática dos documentos HTML utilizando o terrier	4
Criar procedure para atualização da tabela do perfil individual do usuário	5
Identificar e armazenar dados do perfil individual e de grupo do usuário durante a busca	2
Estruturar comunicação entre a interface do FGAVirtual e a interface do terrier Web	3
Inserir perfil individual na busca utilizando a metodologia desenvolvida neste projeto	3
Criar tela de apresentação dos resultados	2

Tabela 5 – Tabela representando as tarefas criadas para a implementação da funcionalidade de Realizar Busca

curso. Isso permitiu um desenvolvimento direcionado e alinhado com os objetivos finais do sistema.

4.2 Implementação do Sistema

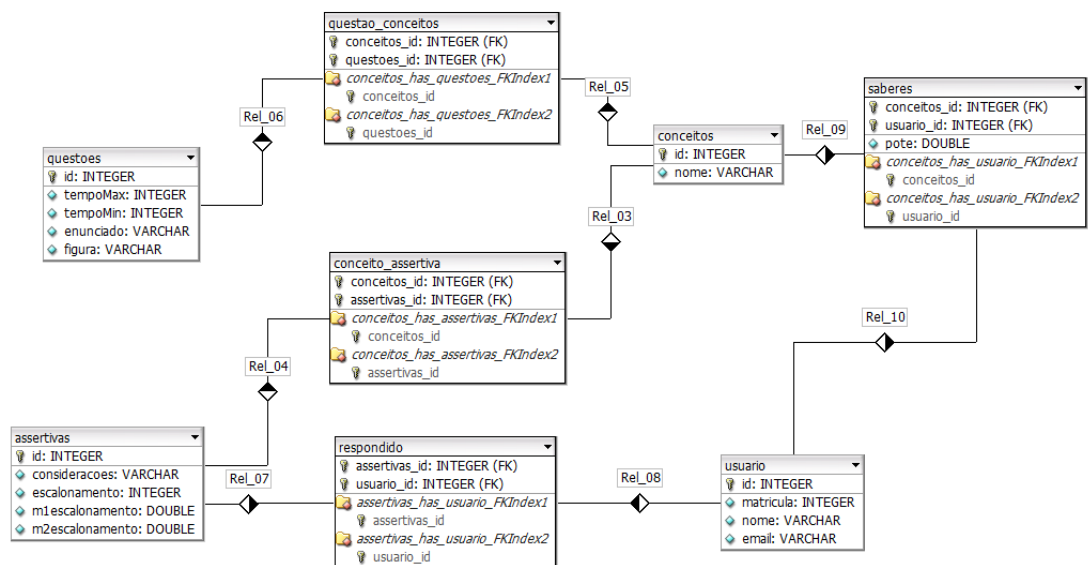
A Tabela 6 apresenta todas as ferramentas utilizadas para a implementação do sistema e para a execução do mesmo, incluindo sua versão.

O estudo deste trabalho considerou como os potes de aprendizagem armazenam as informações do perfil individual do usuário no banco de dados. O DER (diagrama de entidade-relacionamento) do *SmartQuiz* que pode ser visualizado na Figura 7 apresenta como essas informações são armazenadas.

O DER representado na Figura 7 apresenta todas as tabelas do sistema *SmartQuiz*. É visível além das tabelas, os seus atributos e como cada tabela se relaciona uma com as outras.

Ferramenta	Descrição
Java JDK 1.7	Ferramenta utilizada para a execução do aplicativo do Terrier
Terrier 3.5	Aplicativo utilizado para reordenar os documentos do Google utilizando o perfil de grupo e individual do usuário
DBDesigner 4	Ferramenta utilizada para modelagem do Banco de Dados
pgAdmin version 1.18.1	Ferramenta de controle do banco de dados PostgreSQL
Astah Community 6.7.0	Ferramenta utilizada para elaboração da modelagem UML do sistema
NetBeans IDE 8.0	Ferramenta para implementação JAVA da interface web e comunicação com os outros sistemas

Tabela 6 – Tabela de ferramentas utilizadas para a elaboração e execução do sistema

Figura 7 – DER do Sistema *SmartQuiz*

Os potes de aprendizagem são representados por um atributo chamado “pote” na tabela saberes, que é resultado do relacionamento entre a tabela de usuário, que representa o usuário, e a tabela de conceito, que representa cada um dos conceitos de aprendizagem.

Uma característica importante do atributo “pote” é que ele representa a porcentagem de conhecimento que o usuário necessita conhecer de certo conceito, ou objeto de aprendizagem. Ou seja, quando o sistema é iniciado para um novo usuário, todos os potes de aprendizagem são iniciados com 100%, representando que o usuário ainda necessita aprender 100% do conteúdo daquele pote. A cada pergunta respondida pelo usuário, é aplicado um fator de correção nesta porcentagem, e ela é reduzida, caso o usuário acerte a resposta, e aumentada, caso o usuário erre a pergunta. Quando esta porcentagem atinge o limite inferior, ou seja, 0 %, isto significa que o usuário já aprendeu 100% do conhecimento deste pote. Dessa forma, calcula-se quanto um usuário já aprendeu em determinada área de conhecimento (FREITAS et al., 2012).

Suponha que a área de conhecimento de matemática seja determinada pelos objetos de aprendizagem de álgebra e geometria, e que os potes referentes a esses objetos possuem valores de 100% e 50% no banco de dados. Isto significa que o usuário ainda não conhece nada sobre álgebra, mas sobre geometria ele já aprendeu metade do conteúdo. Para calcular o conhecimento do usuário na área de conhecimento de matemática é necessário primeiramente subtrair de 100 a porcentagem de conhecimento em cada pote de aprendizagem, para descobrir o quanto o usuário já aprendeu de cada pote. Em outras palavras, em álgebra o usuário conhece $100 - 100 = 0$ % do conteúdo, e em geometria ele conhece $100 - 50 = 50$ % do conteúdo. Por fim, basta fazer a média dos resultados, ou seja, neste caso basta somar os conhecimentos do usuário em álgebra e geometria ($0 + 50 = 50$) e dividir o resultado pela quantidade de objetos de aprendizagem que definem a área de conhecimento, que neste caso são 2, resultando no valor 25. Isto significa que o usuário já aprendeu 25% do conteúdo da área de conhecimento de matemática, nesta situação hipotética.

Suponha que o usuário não possua conhecimento algum em certa área de conhecimento; neste caso, uma busca em um termo referente a esta mesma área de conhecimento deveria manter os pesos do perfil de grupo do usuário, pois eles representam quais áreas de conhecimento são mais importantes para um usuário de uma determinada engenharia sem considerar o seu perfil individual. Agora suponha que o usuário já conheça 50% desta área de conhecimento, neste caso espera-se que o peso representado pelo perfil de grupo desta área de conhecimento seja reduzido para este usuário, pois ele já conhece metade desta área de conhecimento. Em outras palavras, aplicando-se um peso que represente o perfil individual do usuário sobre o peso do perfil de grupo deste mesmo usuário, podem-se obter resultados de uma busca mais coerentes para aquilo que o usuário já conhece ou que ele necessita conhecer. Por exemplo, suponha que o peso do perfil de grupo de um usuário para a área de conhecimento de matemática seja de 1000. Além disso, suponha que o seu conhecimento na área de conhecimento de matemática seja de 50%, representado por seu perfil individual. A ideia é diminuir o peso de grupo do usuário para esta área de conhecimento uma vez que o usuário já conhece metade desta área de conhecimento.

Uma forma de se reduzir o peso de grupo desta área de conhecimento sabendo o seu perfil individual seria subtrair do seu peso de grupo uma razão relacionada à porcentagem de conhecimento que o usuário já aprendeu, ou seja, neste caso, subtrair de 1000 o valor 500, ou:

$$P_F = 1000 - 1000 \cdot 0.5$$

Em termos gerais, a equação para redução dos pesos é:

$$P_F = P_G - P_G \cdot P_I$$

Onde P_F é o peso final, P_G é o perfil de grupo e P_I é o perfil individual do usuário, ou seja, a porcentagem de conhecimento que o usuário já aprendeu sobre esta área de conhecimento.

A forma de calcular o conhecimento individual do usuário através de porcentagens apresenta uma vantajosa forma de unir os perfis de grupo e individuais para a busca, pois calculando-se a porcentagem média de conhecimento do usuário em cada área de conhecimento, é obtido diretamente o valor que é necessário ser reduzido sobre o perfil de grupo do usuário, a fim de se obter uma ordenação dos documentos mais específica para o usuário ao final da busca.

Uma vez que os valores dos pesos de aprendizagem do usuário estão armazenadas no banco de dados, este foi utilizado para a comunicação entre o sistema FGAVirtual e do *SmartQuizz*. Portanto, para que esta comunicação funcionasse, foi exigida a criação de 3 novas tabelas no banco de dados do *SmartQuizz*, como pode ser visualizado no DER da Figura 8.

As três novas tabelas inseridas no sistema são: login, perfil, e área, localizadas à direita do DER. A tabela de login armazena as informações de usuário e senha para acesso do sistema FGA virtual. Ela se relaciona com a tabela perfil, onde é definida qual o curso de engenharia do usuário em questão, dentre os 4 cursos de engenharia da FGA. Por fim, a tabela área armazena a média de conhecimento do usuário em porcentagem referente cada área de conhecimento, auxiliando diretamente na aplicação do perfil individual na fórmula que calcula o peso final. Essas áreas de conhecimento estão alinhadas com as áreas de conhecimento do perfil de grupo, e é essa tabela que permite unir os dois tipos diferentes de perfis utilizados nesse trabalho.

Existem outras formas de se unir o perfil individual e de grupo na busca de documentos pelo usuário. Uma delas seria realizar uma redução dos pesos de cada objeto de aprendizagem associados a uma determinada área de conhecimento a partir do perfil de grupo do usuário nesta área de conhecimento. Isto significa que o direcionamento dos pesos seria o inverso, ou seja, seria aplicado um peso partindo do perfil de grupo para o

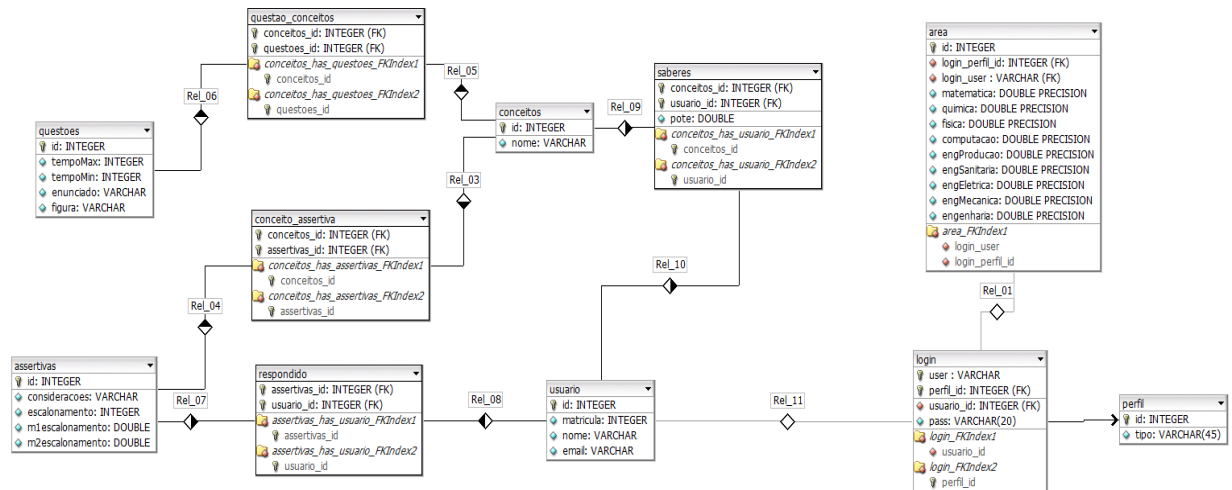


Figura 8 – DER representando as tabelas do Sistema FGAVirtual juntamente com as tabelas do *SmartQuiz*.

perfil individual do usuário, diferentemente daquele que foi adotado neste trabalho, onde o peso parte do perfil individual e é aplicado no perfil de grupo do usuário. Em vista da complexidade da aplicação do peso de grupo para cada objeto de aprendizagem de uma determinada área de conhecimento, foi decidido aplicar o perfil individual sobre o perfil de grupo, por apresentar uma solução mais simples e com vantagens semelhantes.

Na sessão anterior deste trabalho foi discutido sobre o caso em que um objeto de aprendizagem é compartilhado por duas ou mais áreas de conhecimento. Neste caso, existem portanto objetos de aprendizagem pertencentes a não apenas uma área de conhecimento, e sim duas ou mais áreas de conhecimento. Portanto, assumindo que para o cálculo do perfil individual do usuário é necessário considerar este objeto de aprendizagem, deve ser utilizada a utilização de pesos para dividir este objeto de aprendizagem nas áreas de conhecimento que o compartilham. Isto ocorre pois este objeto provavelmente possui maior relevância para uma das áreas de conhecimento que o compartilha, e por isto deve ser tratado diferentemente. Dessa forma, existem algumas soluções para este caso que serão discutidas a seguir.

A solução adotada, e talvez a mais simples, seria a divisão de pesos desse pote para as áreas de conhecimento de forma igualitária, ou seja, se fosse um objeto de aprendizagem pertencente a duas áreas de conhecimento, essa porcentagem seria dividida em peso 1 para cada uma das áreas de conhecimento, e o cálculo do perfil individual de uma dessas áreas de conhecimento seria normal, como se o objeto de aprendizagem pertencesse a apenas uma das áreas de conhecimento. Por exemplo, suponha que a área de conhecimento de matemática é formada pelos objetos de aprendizagem “limite” e “derivada”, e o objeto de aprendizagem de derivada pertence também à área de conhecimento de física. Além disso, considere também que o usuário possui 100% de conhecimento no objeto de aprendizagem de limite e 25% de conhecimento no objeto de aprendizagem de derivada. Neste caso, o perfil individual do usuário na área de conhecimento de matemática seria:

$$\frac{100 + 25}{2} = 62.5\%$$

No cálculo é considerado os 25% de conhecimento no objeto de aprendizagem de “derivada” como se este objeto de aprendizagem fosse pertencente apenas a esta área de conhecimento, e os 100% de conhecimento no objeto de aprendizagem de limite. Por fim, o conhecimento deste usuário na área de matemática é calculado como uma média simples, portanto, 62.5%.

Existem outras formas mais complexas de resolução deste problema. Uma delas seria considerar uma proporção dos pesos mais real deste objeto de aprendizagem em cada área de conhecimento, levando em consideração o fato do objeto de aprendizagem ser mais relevante a uma das áreas de conhecimento que o compartilham. Estes pesos podem ser abstraídos dos pré-requisitos deste objeto de aprendizagem, considerando o seguinte raciocínio: se para aprender sobre “derivada” é necessário de conhecer 3 objetos de conhecimento pertencentes à área de conhecimento de matemática e 2 de física, a área de conhecimento de matemática possui portanto uma razão de peso de 3/5 sobre o objeto de aprendizagem de “derivada”, enquanto a área de conhecimento de física possui 2/5. Estes dados de pré-requisitos são obtidos a partir da ontologia criada no *SmartQuiz*, pois para que o sistema funcione existe um relacionamento de dependência entre os objetos de conhecimento, de forma que seja possível identificar quais são os objetos de conhecimento envolvidos em cada uma das perguntas e respostas do *Quiz*. Embora exista essa e outras soluções, elas são mais complexas e requerem que a ontologia dos objetos de aprendizagem do *SmartQuiz* esteja definida para que seja possível sua implementação. Por este motivo, foi aplicada neste trabalho a solução atribuindo pesos iguais aos objetos de aprendizagem compartilhados entre uma ou mais área de conhecimento e entre os objetos dentro de uma mesma área de conhecimento.

Uma característica importante desta modelagem adotada é que o peso determinado pelo perfil de grupo do usuário representa aquilo que é esperado que o usuário saiba, ou

seja, quanto maior for o peso de uma área de conhecimento, mais o usuário necessita aprender sobre essa área. Quando é aplicado o fator do perfil individual e o peso de grupo é reduzido, a realidade por trás deste cálculo é que, quanto mais o usuário conhece de certa área de conhecimento, menos ele necessita aprender. Isto não significa que a área de conhecimento se torna menos importante para o usuário, mas no contexto de aprendizagem, isto significa que o usuário já aprendeu o bastante sobre o assunto, e portanto outras áreas de conhecimento devem possuir maior relevância para ele.

Para que a tabela denominada área (a qual armazena o perfil do usuário) seja atualizada, foi criado um procedimento de atualização na linguagem PL/SQL no banco de dados utilizado. Esse procedimento permite que sejam obtidos os valores de cada um dos potes e a partir disso seja calculado o valor médio de aprendizagem do usuário em cada uma das áreas de conhecimento, utilizando a lógica descrita neste capítulo.

Para fins de testes, foram inseridos 17 dados na tabela de conceitos, e foi realizado um mapeamento para identificar a qual área de conhecimento está associado cada um dos conceitos, como pode ser visualizado na Tabela 7.

Área de conhecimento	Conceitos Inseridos
Matemática	Derivada e Geometria
Química	Elemento e Reação
Física	Empuxo e Derivada
Computação	Algoritmo e Software
Engenharia de Produção	Produção e Máquina
Engenharia Sanitária	Bacia e Canais
Engenharia Elétrica	Eletricidade e Pilha
Engenharia Mecânica	Engrenagem e Motor
Engenharia	Processo e Protótipo

Tabela 7 – Tabela associando as áreas de conhecimento do perfil de Grupo com os conceitos de teste inseridos no sistema

Percebe-se na Tabela 7 que foram inseridos dois conceitos para cada uma das áreas de conhecimento, com exceção do conceito derivada, que pertence às áreas de matemática e de física. Dessa forma, o procedimento criado para atualizar a tabela de áreas acessa cada um dos registros da tabela “saberes”, onde é mantido o valor do pote de cada objeto de aprendizagem, e realiza uma média desses potes para descobrir a quantidade que o usuário já aprendeu sobre cada área de conhecimento. A Figura 9 apresenta um trecho do procedimento criado, onde é atualizado o valor do perfil individual do usuário na área de conhecimento de matemática.

O primeiro select do código tem a função de buscar o valor do pote do usuário relativo ao conceito com o número identificador “1”, que se refere ao conceito de derivada do sistema, e atribuir para uma variável temporária chamada aprendido. Portanto, a

```
-- atualiza área de matemática
soma = 0;
SELECT pote into aprendido FROM saberes where usuario_id = 1 and conceito_id = 1;

soma = soma + 100 - aprendido;
soma = soma + 100;

SELECT pote into aprendido FROM saberes where usuario_id = 1 and conceito_id = 2;

soma = soma + 100 - aprendido;
soma = soma/(2*100);

update area
set matematica = soma
where user_login = 'admin';
```

Figura 9 – Trecho de código do procedimento que atualiza o perfil individual do usuário na área de conhecimento de matemática

variável `aprendido` representa quanto o usuário ainda necessita aprender com relação ao pote de aprendizagem. Já que o conceito de derivada é compartilhado entre duas áreas de conhecimento (matemática e física) e foi utilizada a divisão de pesos igualmente para as áreas de conhecimento, o cálculo ocorre normalmente considerando o valor que há no banco de dados. Portanto é necessário primeiramente subtrair 100 do valor do pote, para descobrir quanto o usuário já aprendeu, e atribuir o valor calculado à variável `soma`. Por exemplo, se o valor do pote no banco de dados (representado pela variável `aprendido`) é de 25%, então o usuário já aprendeu 75% deste pote e este valor é somado à variável denominada `soma`.

O segundo select da imagem busca o valor do pote de geometria e subtrai de 100 este valor, para descobrir quanto o usuário já aprendeu. Estes valores são somados à variável denominada `soma`, e por fim o valor desta variável é dividida por 2×100 , já que existem dois conceitos que definem esta área de conhecimento e a lógica do sistema exige um valor decimal da porcentagem. O resultado final é atualizado na tabela área, onde o atributo é chamado de matemática.

4.3 Funcionamento do sistema

Nesta sessão é descrito cada um dos passos realizados pelo sistema a partir do momento em que o usuário deseja realizar a busca.

O diagrama de blocos representado pela Figura 10 apresenta cada um dos objetos e sistemas utilizados pelo sistema desde o início da busca até a apresentação dos resultados ao usuário. As setas do diagrama representam a ordem de chamada dos objetos, e os retângulos representam cada um dos objetos e sistemas.

Inicialmente, o usuário decide realizar a busca inserindo as palavras-chave por ele escolhidas. Essas palavras-chave indicam qual é o tema ou o assunto que o usuário deseja

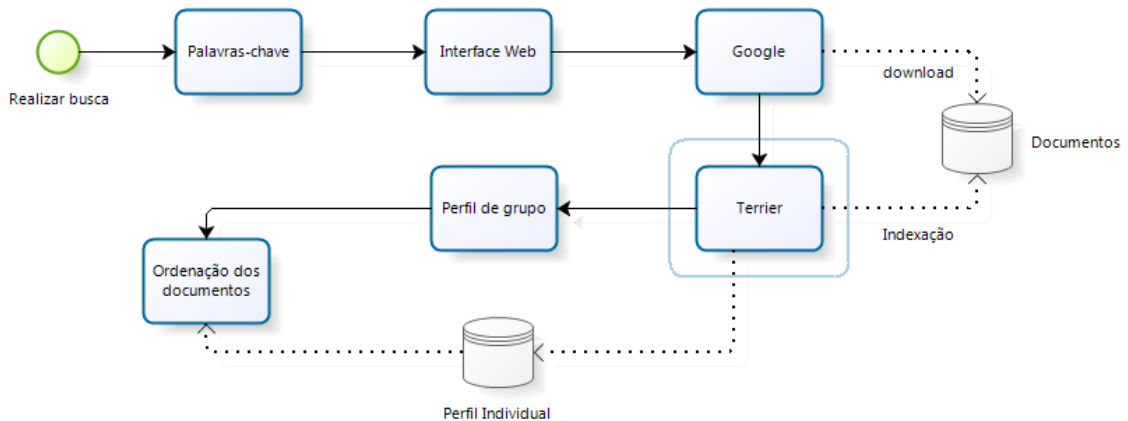


Figura 10 – Diagrama de blocos que representa a sequência dos passos realizados durante o processo de uma busca no FGAVirtual

aprender. Portanto, essas palavras-chave são inseridas na busca e é iniciado o processo de indexação. O sistema não possui inicialmente nenhum documento indexado para a busca, ou seja, não existe uma base de arquivos para serem recuperados no processo da busca. Para construir essa base, a interface web se comunica com o Google através de uma classe responsável por fazer downloads dos documentos. Nessa etapa, é realizada em tempo real uma busca no google utilizando as mesmas palavras-chave inseridas pelo usuário e são realizados os downloads localmente de todos os arquivos com extensão “html” presentes nas duas primeiras páginas do resultado da busca do Google. Cada um dos arquivos são gravados com o nome da url do site, sem a utilização dos símbolos de “/” e “.”.

O segundo passo é responsável pela indexação dos documentos para que eles possam ser utilizados na busca do sistema FGAVirtual. Para isso, são executados os arquivos de indexação do Terrier, passando como parâmetro o diretório onde os arquivos resultado da busca do Google estão. Infelizmente o sistema operacional não permite que o aplicativo do Terrier seja executado e ao mesmo tempo haja uma reindexação dos arquivos para uma nova busca, portanto esse processo é realizado com o aplicativo do Terrier sem

funcionamento. Com os arquivos indexados, é realizada a ordenação dos documentos para o usuário. Neste momento, o servidor do Terrier é iniciado, são identificados os perfis de grupo do usuário e perfis individual e é iniciada uma busca no terrier utilizando as mesmas palavras-chave inseridas pelo usuário. Os perfis são adquiridos através de uma comunicação entre o sistema e o banco de dados, onde são mantidas as informações do perfil individual do usuário. Sabendo-se portanto qual o peso de grupo a ser utilizado pelo sistema durante a busca, e o perfil individual do usuário na área de conhecimento que o usuário está realizando a busca, o Terrier atribui uma nota para cada documento utilizando o peso calculado, e apresenta os resultados em sua interface web. Por fim, o sistema FGAVirtual adquire o resultado da ordenação dos documentos gerada pela interface do Terrier, e apresenta para o usuário na interface web.

Uma característica interessante do sistema é a união de diferentes algoritmos de buscas durante a recuperação dos documentos. Inicialmente é realizada uma busca utilizando o Google, que trabalha com um algoritmo de ordenação estática, ou seja, ele calcula quantas referências ou “links” cada um dos documentos da busca possuem, e apresenta como resultado da busca primeiramente aqueles documentos que possui maior popularidade na internet. Por fim, é aplicado outro algoritmo de ordenação sobre os documentos resultado da busca do Google, conhecido como ordenação baseada em termos, através do Terrier. Esses algoritmos trabalham com a frequência de ocorrência do termo dentro dos documentos, e a partir da divergência desta frequência, eles atribuem uma pontuação para cada documento, gerando a ordenação final.

Para auxiliar no desenvolvimento das interfaces para a aplicação web, foi utilizado o framework Prime Faces, tornando a implementação das páginas mais rápida, e mantendo o esforço do trabalho no desenvolvimento da lógica das buscas focadas nos perfis. O Prime faces é um componente JSF livre, que possui suporte para ajax e uma ampla variedade de temas visuais (PRIMEFACES, 2009). Portanto, o sistema como um todo funciona com a aplicação web executando como front-end da aplicação, onde são coletadas as informações necessárias para as buscas, e os demais sistemas (Google, Terrier e *SmartQuizz*) como apoio à busca.

4.4 Testes do Sistema

O objetivo dos testes do sistema foi verificar a execução do sistema, com a apresentação de resultados, sem considerar se os resultados estavam realmente associados ao perfil do usuário.

Basicamente, o sistema FGAVirtual é composto por 3 páginas web: A página de login, a página de busca e a página de resultados. Na página de resultados, o usuário possui a opção de realizar uma nova busca, inserindo as palavras-chave novamente e pressionando

o botão de buscar.

A página de login pode ser visualizada na Figura 11.



Figura 11 – Página de login do sistema FGAVirtual

Nesta página, o usuário insere os seus dados registrados de usuário e senha, e dessa forma é realizada uma busca. Esse processo é necessário porque com isso é identificado qual é o usuário que está solicitando a busca, e assim recuperar do banco de dados qual é o seu perfil de grupo, ou seja, qual é o seu curso e portanto quais áreas de conhecimento são mais importantes para ele, além de saber qual é o seu conhecimento em cada objeto de aprendizagem distinto através do perfil individual traçado pelo sistema *SmartQuizz*.

O primeiro teste buscou identificar a mensagem de erro, caso o usuário não insira corretamente os dados de login. Neste caso, o sistema apresenta uma mensagem que pode ser visualizada na Figura 12.



Figura 12 – Mensagem lançada quando há um erro nos dados de login inseridos pelo usuário

Dessa forma, o usuário pode inserir novamente os seus dados até que os dados estejam corretos. Quando os dados inseridos estão corretos, a próxima mensagem apresentada pelo sistema é página de busca, como pode ser visualizado na Figura 13.

Nesta página, percebe-se que o sistema já possui algumas informações do usuário registrada, como o seu login, apresentado após a frase de “Bem-vindo”, e a sua engenharia, logo abaixo desta mesma mensagem. Nesta página, há um campo para o usuário inserir as



Figura 13 – Página de busca do sistema FGAVirtual

suas palavras-chave da busca e um botão que permite que ele realize a busca, e o sistema inicie o processo de download dos documentos do Google, a indexação através do Sistema Terrier, e a busca utilizando os perfis individual e de grupo do usuário. Nesta tela ainda há um botão de sair, o qual permite que o usuário volte para a tela de login, caso deseje entrar com outro usuário no sistema.

Assim que as palavras-chave são inseridas e o botão de buscar é pressionado, todas as etapas descritas na sessão de funcionamento do sistema são executadas, e por fim é apresentada a última tela do sistema ao usuário, como pode ser visualizado na Figura 14.



Figura 14 – Página de apresentação dos resultados do sistema FGAVirtual

Na página de resultados são apresentados ao usuário o produto final da busca, com os respectivos documentos em extensão html obtidos do google. Neste resultado, é apresentado o diretório onde os documentos estão armazenados juntamente com o seu nome, que representa o endereço na internet do documento html, com a ausência de caracteres como a “/” e o “.”. Os documentos estão enumerados, e a numeração indica, em ordem crescente, qual o documento com maior importância para o usuário.

Nem todos os documentos resultado da busca do google são apresentados como resultado da busca pois, durante a busca através do Terrier, muitos documentos não alcançam a pontuação mínima para serem apresentados. De qualquer forma, caso o usuário

deseje visualizar os documentos que não são apresentados na busca, é possível encontrá-los na pasta de download onde os arquivos estão armazenados.

4.5 Avaliação do Sistema

Esta sessão apresenta os resultados da avaliação do sistema, com análise qualitativa dos dados, isto é, nesta etapa foi buscado avaliar o resultado das buscas considerando o perfil de grupo de usuário e o seu perfil individual, analisando se os resultados estão coerentes com o seu perfil.

Para a avaliação do sistema foram utilizados 30 documentos diferentes. São eles: Álgebra, Biorrefinaria, Circuito Digital, Computação, Computação Quântica, Desenvolvimento Ágil de Software, Eletromagnetismo, Engenharia, Engenharia Automóvel, Engenharia de Energia, Engenharia de Produção, Engenharia de Software, Engenharia do Ambiente, Engenharia Elétrica, Engenharia Eletrônica, Engenharia Mecânica, Engenharia Sanitária, Física, Física Moderna, Gestão Ambiental, Indústria, Máquina Térmica, Matemática, Matemática Discreta, Motor de Combustão Interna, Petróleo, Política Energética do Brasil, Projeto, Química e Transmissão de Energia Elétrica.

Para a avaliação dos resultados, não foi utilizado o mecanismo de download de documentos do Google desenvolvido neste trabalho pois com um corpus restrito de documentos a análise dos resultados se torna mais objetiva.

Além disso, a avaliação do sistema foi dividida em duas partes. A primeira parte buscou avaliar o resultado de uma busca utilizando apenas um termo, e a segunda parte utilizando dois termos em áreas de conhecimento diferentes. Essas duas partes contaram com uma análise direta dos resultados pelo autor e outra pessoa, considerando a semântica própria de cada documento. Como algoritmo de busca foi utilizado o DFRee em ambas as partes, por ter sido considerado o algoritmo que melhor se adaptou à implementação de buscas baseada em perfis ([MARTINS, 2013](#)).

A análise também contou com uma variação direta do perfil individual do usuário em números fixos, a fim de se verificar se ocorreriam mudanças na ordenação dos documentos e se as mudanças estavam coerentes com o perfil individual adotado.

Para as buscas utilizando apenas um termo, ou uma palavra-chave, a metodologia adotada foi alterar o perfil individual do usuário em 3 valores: 25%, 50% e 75%. Isto significa que, quando o perfil individual de 25% foi adotado, isto indica que o usuário já sabe apenas 25% da área de conhecimento do termo que está sendo utilizado na busca. O mesmo ocorre para os valores de 50% e 75%.

Para as buscas utilizando dois termos, houve uma maior complexidade na avaliação. Neste caso, ao invés do perfil variar entre 25%, 50% e 75% eles variaram entre os

extremos opostos, ou seja: 0% e 100%. Neste caso, existiam duas áreas de conhecimento diferentes envolvidas na busca e, portanto, a variação do perfil individual se deu em 4 cenários diferentes: o primeiro deles atribuiu para os perfis individual nas duas áreas de conhecimento o valor de 0%, ou seja, o usuário não conhecia nada das duas áreas de conhecimento; o segundo cenário atribuiu o valor de 100% para uma das áreas de conhecimento e 0% para a outra área de conhecimento; o terceiro cenário atribuiu os pesos inversos ao cenário anterior, ou seja, aquela área de conhecimento que possuía perfil individual de 0% passou a ter 100% e aquela que possuía 100% passou a ter 0% e por fim o último cenário atribuiu a porcentagem de 100% do perfil individual para as duas áreas de conhecimento.

Na avaliação utilizando dois termos é importante ressaltar que quando o perfil individual nas duas áreas de conhecimento recebe o valor de 0%, a busca é realizada utilizando o perfil de grupo puro desenvolvido no trabalho (MARTINS, 2013). Isto ocorre pois quando o perfil individual recebe valor de 0%, o peso não é reduzido, e portanto é mantido o peso do perfil de grupo definido. Quando o perfil individual nas duas áreas de conhecimento recebe valor de 100%, a busca é realizada utilizando o algoritmo puro, sem atribuir o peso do perfil do usuário, uma vez que há uma redução de 100% do valor do perfil de grupo e o peso é reduzido para o próprio peso original.

4.5.1 Avaliação Utilizando um Termo

A avaliação utilizando apenas um termo na busca utilizou o termo química. Para esta consulta, foi considerado como documentos importantes o documento de química e biorrefinaria, que estão destacados de verde nas tabelas. As Tabelas 8 e 9 representam o resultado final da ordenação do sistema utilizando o perfil de 25%, 50% e 75% na área de conhecimento de química e o perfil de grupo de Engenharia Automotiva, onde a área de conhecimento possui peso de 1, ou seja, é uma das áreas menos relevantes para o curso.

É importante ressaltar que não houve mudança nos resultados utilizando 25% e 50% de perfil individual para a avaliação utilizando este cenário, por isso os resultados estão representados na Tabela 8.

Percebe-se que nos três testes realizados (variando o perfil individual entre 25%, 50% e 75%) não houve alteração na ordenação nem dos documentos importantes para a busca, nem para os demais documentos, com exceção do teste utilizando perfil individual de 75% onde o documento referente à Engenharia de Energia (destacado pela cor cinza) trocou de posição com o documento referente à Engenharia, subindo assim 1 posição. Isto indica que, para este caso onde a área de conhecimento não é expressiva para este perfil de grupo (com o peso de apenas 1), independente da quantidade que o usuário já aprendeu acerca desta área de conhecimento, não há modificação relevante nos resultados da busca. Este resultado era esperado, pois para áreas de conhecimento onde o perfil de grupo é baixo, ou seja, para áreas de conhecimento que não são tão relevantes para o usuário,

Ordem	Documento
1	Química
2	Biorrefinaria
3	Física
4	Engenharia
5	Engenharia de Energia
6	Engenharia de Produção
7	Motor de Combustão Interna
8	Máquina Térmica
9	Engenharia Elétrica
10	Petróleo
11	Matemática discreta
12	Matemática

Tabela 8 – Tabela que representa o resultado final da ordenação utilizando o termo “química” em uma busca com o perfil de grupo de Automotiva e perfil individual de 25% e 50%

Ordem	Documento
1	Química
2	Biorrefinaria
3	Física
4	Engenharia de Energia
5	Engenharia
6	Engenharia de Produção
7	Motor de Combustão Interna
8	Máquina Térmica
9	Engenharia Elétrica
10	Petróleo
11	Matemática discreta
12	Matemática

Tabela 9 – Tabela que representa o resultado final da ordenação utilizando o termo “química” em uma busca com o perfil de grupo de Automotiva e perfil individual de 75%

a variação do peso traçada pelo perfil individual é muito baixo, e portanto, não afeta a ordenação dos documentos.

Apesar de não haver alterações nos resultados, percebe-se que os documentos mais importantes para a busca, ou seja, os documentos que representam a área de conhecimento relacionada a química (que é a palavra chave da busca), que são “Química” e “Biorrefinaria”, estão nas duas primeiras posições da busca, independente do perfil individual utilizado.

As Tabelas 10, 11 e 12 representam os resultados utilizando o termo de “computação” em um perfil de grupo de usuário de Engenharia de software. Neste caso, foi utilizado

uma área de conhecimento bastante relevante para o perfil de grupo adotado, onde o peso de grupo para esta área é de 10000.

Para esta busca, foi considerado com o documentos importantes os documentos de Computação e Computação Quântica, que estão destacados de verde nas tabelas.

Ordem	Documento
1	Computação
2	Computação Quântica
3	Engenharia Elétrica
4	Matemática discreta
5	Engenharia Eletrônica
6	Circuito Digital
7	Física
8	Engenharia de Software
9	Engenharia
10	Matemática
11	Engenharia de Produção

Tabela 10 – Tabela que representa o resultado final da ordenação utilizando o termo “computação” em uma busca com o perfil de grupo de Software e perfil individual de 25%

Ordem	Documento
1	Computação
2	Computação Quântica
3	Engenharia Elétrica
4	Matemática discreta
5	Engenharia Eletrônica
6	Circuito Digital
7	Engenharia de Software
8	Engenharia
9	Matemática
10	Física
11	Engenharia de Produção

Tabela 11 – Tabela que representa o resultado final da ordenação utilizando o termo “computação” em uma busca com o perfil de grupo de Software e perfil individual de 50%

Percebe-se que neste caso, utilizando um termo de uma área de conhecimento de maior relevância ao usuário há uma maior variação da ordenação dos documentos em cada um dos casos de aplicação do perfil individual.

Apesar de não haver alteração com relação ao posicionamento do documento referente à computação (computação e computação quântica), nota-se algo interessante. Através da definição de perfil de grupo a área de computação é a mais relevante para

Ordem	Documento
1	Computação
2	Computação Quântica
3	Engenharia Elétrica
4	Matemática discreta
5	Física
6	Engenharia Eletrônica
7	Engenharia de Produção
8	Matemática
9	Engenharia
10	Engenharia de Software
11	Circuito Digital

Tabela 12 – Tabela que representa o resultado final da ordenação utilizando o termo “computação” em uma busca com o perfil de grupo de Software e perfil individual de 75%

um usuário de Engenharia de software, acompanhada em segundo lugar da área de matemática. Neste cenário, é perceptível um aumento graduativo de posição do documento referente à matemática (destacado com a cor cinza) com a variação do perfil individual do usuário. Ao utilizar o perfil individual de 25%, o documento referente à matemática é apresentado na décima posição. Ao aumentar o perfil individual para 50% este mesmo documento sobe para a nona posição e por fim, com o perfil de 75%, o documento sobe mais uma posição, se apresentando na oitava.

Isto significa que, ao se utilizar um termo de uma área de conhecimento que é relevante para o usuário, o aumento graduativo de conhecimento individual do usuário nesta área de conhecimento permite que outras áreas de conhecimento obtenham maior importância na busca, se apresentando portanto cada vez mais no topo dos resultados.

4.5.2 Avaliação Utilizando dois Termos

Para avaliação dos testes utilizando dois termos foram realizadas buscas utilizando os termos: “matemática computação” e “matemática elétrica”.

Para cada um dos testes, foram definidos os documentos que são relevantes para a busca para serem analisados. Por exemplo: para a busca “matemática computação”, os documentos relevantes são: “matemática discreta”, “computação quântica”, “computação” e “matemática” e para a busca “matemática elétrica” os documentos relevantes são: “engenharia elétrica”, “matemática”, “álgebra”, “circuito digital” e “eletromagnetismo”.

Além disso, foi considerado qual documento é mais relacionado à cada uma das áreas de conhecimento envolvida na busca através de uma contagem de termos. Por exemplo: para o documento de “matemática discreta” foram encontradas 71 ocorrências da

palavra “matemática” e 16 da palavra “computação”. Por isto, na busca envolvendo as palavras “matemática computação” o documento de “matemática discreta” está mais associado a área de matemática do que com a área de computação.

As Tabelas 13 e 14 representam quais documentos são mais relevantes pra cada áreas de conhecimento envolvidos nas buscas.

Área de conhecimento	Documentos mais relevantes
Matemática	Matemática Discreta e Matemática
Computação	Computação Quântica e Computação

Tabela 13 – Tabela que representa os documentos mais relevantes para cada uma das áreas de conhecimento envolvidas na busca “matemática computação”

Área de conhecimento	Documentos mais relevantes
Matemática	Matemática, Matemática Discreta e Álgebra
Elétrica	Engenharia Elétrica, Circuito Digital e Eletromagnetismo

Tabela 14 – Tabela que representa os documentos mais relevantes para cada uma das áreas de conhecimento envolvidas na busca “matemática elétrica”

As Tabelas 15, 16, 17 e 18 representam os resultados da busca “matemática computação” utilizando o perfil de grupo de engenharia de software. Para este grupo, as áreas de conhecimento de matemática e computação são ambas relevantes, com o peso de 1000 e 10000 na busca, respectivamente.

Os documentos destacados com a cor verde representam aqueles importantes para a área de conhecimento de matemática, e os destacados com a cor amarela, à área de computação.

Ao observar as tabelas, identifica-se que os documentos importantes para essa busca (matemática discreta, matemática, computação quântica e computação) sempre aparecem nas 5 primeiras posições. Isto significa que, independente do perfil individual do usuário, não há alteração significativa dos resultados dentro dos documentos importantes para a busca, quando as duas áreas de conhecimento utilizadas são relevantes para o usuário. Apesar disso, percebe-se uma modificação de posição do documento de Engenharia elétrica, que representa a área de conhecimento de Elétrica, quando o perfil individual é modificado para 100% nas duas áreas de conhecimento envolvidas na busca (matemática e computação) na Tabela 18. A área de engenharia elétrica é a quarta mais importante para Engenharia de Software, de acordo com o trabalho desenvolvido (MARTINS, 2013), logo atrás de Engenharia. Apesar de não haverem mudanças significativas no documento de “Engenharia”, o documento de “Engenharia Elétrica” (destacado na cor cinza), que estava na sexta posição, passa portanto para a quarta quando o usuário já possui conhecimento pleno das duas áreas de conhecimento mais importantes (Tabela 18). Novamente

Ordem	Documento
1	Computação
2	Matemática
3	Matemática discreta
4	Álgebra
5	Computação Quântica
6	Engenharia Elétrica
7	Engenharia Eletrônica
8	Engenharia de Produção
9	Engenharia Mecânica
10	Circuito Digital
11	Engenharia
12	Física
13	Engenharia de Energia
14	Engenharia de Software
15	Química
16	Transmissão de Energia Elétrica

Tabela 15 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática computação” em uma busca com o perfil de grupo de Software e perfis individuais de 0% para ambas as áreas de conhecimento de matemática e computação

Ordem	Documento
1	Computação
2	Matemática
3	Matemática discreta
4	Álgebra
5	Computação Quântica
6	Engenharia Elétrica
7	Química
8	Física
9	Transmissão de Energia Elétrica
10	Engenharia de Software
11	Engenharia de Energia
12	Engenharia
13	Circuito Digital
14	Engenharia Mecânica
15	Engenharia de Produção
16	Engenharia Eletrônica

Tabela 16 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática computação” em uma busca com o perfil de grupo de Software e perfis individuais de 0% e 100% para as áreas de conhecimento de matemática e computação, respectivamente

Ordem	Documento
1	Computação
2	Matemática
3	Matemática discreta
4	Álgebra
5	Computação Quântica
6	Engenharia Elétrica
7	Engenharia Eletrônica
8	Engenharia de Produção
9	Engenharia Mecânica
10	Circuito Digital
11	Engenharia
12	Física
13	Engenharia de Energia
14	Engenharia de Software
15	Química
16	Transmissão de Energia Elétrica

Tabela 17 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática computação” em uma busca com o perfil de grupo de Software e perfis individuais de 100% e 0% para as áreas de conhecimento de matemática e computação, respectivamente

Ordem	Documento
1	Computação
2	Matemática discreta
3	Matemática
4	Engenharia Elétrica
5	Computação Quântica
6	Álgebra
7	Engenharia Eletrônica
8	Engenharia de Produção
9	Circuito Digital
10	Engenharia Mecânica
11	Engenharia
12	Engenharia de Software
13	Engenharia de Energia
14	Transmissão de Energia Elétrica
15	Física
16	Química

Tabela 18 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática computação” em uma busca com o perfil de grupo de Software e perfis individuais de 100% e 100% para as áreas de conhecimento de matemática e computação, respectivamente

é possível perceber, portanto, um destaque para uma área de conhecimento importante para o usuário, através da variação do perfil individual.

Por fim, as Tabelas 19, 20 e 21 apresentam o resultado das consultas para a busca “matemática elétrica” para o perfil de grupo de Engenharia de energia. Para este perfil a área de conhecimento de matemática é a mais importante, e a área de conhecimento de elétrica é a de menor importância, com pesos de 10000 e 1 respectivamente.

O resultado da busca para o perfil individual de 0% nas duas áreas de conhecimento e 0% para a área de matemática e 100% para a área de elétrica foram os mesmos, e estão representados na Tabela 19.

Os documentos destacados com a cor verde representam aqueles importantes para a área de conhecimento de matemática, e os destacados com a cor amarela, à área de elétrica.

Ordem	Documento
1	Matemática
2	Matemática discreta
3	Álgebra
4	Engenharia Elétrica
5	Engenharia Eletrônica
6	Eletromagnetismo
7	Transmissão de Energia Elétrica
8	Engenharia de Produção
9	Computação
10	Engenharia Mecânica
11	Máquina Térmica
12	Física
13	Política energética do Brasil
14	Engenharia
15	Engenharia de Energia
16	Petróleo
17	Indústria
18	Engenharia de Software
19	Computação Quântica
20	Química
21	Circuito Digital

Tabela 19 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática elétrica” em uma busca com o perfil de grupo de Energia e perfis individuais de 0% para ambas as áreas de conhecimento de matemática e elétrica e 0% e 100% para as áreas de conhecimento de matemática e elétrica, respectivamente

Este teste envolvendo a área de conhecimento mais importante para o perfil de grupo do usuário e a menos importante apresentou resultados interessantes. Observando

Ordem	Documento
1	Engenharia Elétrica
2	Matemática discreta
3	Engenharia Eletrônica
4	Matemática
5	Engenharia de Produção
6	Álgebra
7	Transmissão de Energia Elétrica
8	Eletromagnetismo
9	Física
10	Engenharia
11	Engenharia de Energia
12	Máquina Térmica
13	Política energética do Brasil
14	Circuito Digital
15	Computação
16	Engenharia Mecânica
17	Engenharia de Software
18	Computação Quântica
19	Indústria
20	Petróleo
21	Química

Tabela 20 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática elétrica” em uma busca com o perfil de grupo de Energia e perfis individuais de 100% e 0% para as áreas de conhecimento de matemática e elétrica, respectivamente

os documentos importantes para a busca (destacados de verde e amarelo), todos eles, com exceção de Eletromagnetismo (que em uma das buscas aparece na oitava posição) e de Circuito Digital, sempre aparecem nas 6 primeiras posições da busca, alterando as posições entre si ao se variar o perfil individual.

Percebe-se na Tabela 19, que os resultados da ordenação dos documentos com o perfil individual em elétrica equivalente a 0% e 100% foram os mesmos. Isto era esperado, pois a área de conhecimento em elétrica no perfil de grupo de Energia possui pouca relevância, então a quantidade de conhecimento que o usuário aprende sobre esta área não afeta a busca.

Ao comparar as tabelas onde há a inversão do perfil individual, ou seja, as tabelas onde o usuário conhece 100% de uma das áreas de conhecimento de 0% da outra (Tabela 19), e o contrário (Tabela 20), percebe-se uma inversão de posição entre os documentos relevantes a cada uma das áreas de conhecimento. Por exemplo, na Tabela 19 o usuário conhece 0% da área de conhecimento de matemática e 100% de elétrica, e portanto, são apresentados primeiramente os documentos de Matemática, Matemática Discreta e

Ordem	Documento
1	Engenharia Elétrica
2	Matemática discreta
3	Matemática
4	Engenharia Eletrônica
5	Álgebra
6	Eletromagnetismo
7	Engenharia de Produção
8	Transmissão de Energia Elétrica
9	Máquina Térmica
10	Física
11	Engenharia de Energia
12	Política energética do Brasil
13	Engenharia
14	Computação
15	Engenharia Mecânica
16	Circuito Digital
17	Engenharia de Software
18	Computação Quântica
19	Petróleo
20	Indústria
21	Química

Tabela 21 – Tabela que representa o resultado final da ordenação utilizando os termos “matemática elétrica” em uma busca com o perfil de grupo de Energia e perfis individuais de 100% e 100% para as áreas de conhecimento de matemática e elétrica, respectivamente

Álgebra, e em seguida, os documentos de Engenharia elétrica e Eletromagnetismo. Por outro lado, na Tabela 20 onde o usuário conhece 100% de matemática e 0% de elétrica o documento de Engenharia Elétrica aparece na primeira colocação, seguido de Matemática Discreta.

Este comportamento é interessante pois, visto que a áreas de conhecimento de matemática é a mais importante para o perfil de Grupo de energia, quando a busca atribui perfil individual de 0% para esta área de conhecimento, todos os três documentos relevantes para matemática são apresentado nas três primeiras posições da busca. Já ao atribuir 100% de perfil individual para a área de conhecimento de matemática e 0% para a área de conhecimento de elétrica, apenas o documento de Engenharia Elétrica aparece na primeira posição, e os demais documentos importantes para a área de conhecimento de Elétrica (Circuito digital e Eletromagnetismo) permanecem em posições mais baixas. Isto ocorre pois apesar do usuário não conhecer nada de Engenharia elétrica, esta área de conhecimento possui peso mínimo para o perfil de Grupo de Energia, indicando que não é uma área tão relevante para este perfil.

5 Conclusão

A construção de um sistema capaz de realizar buscas considerando não apenas o perfil de grupo de usuário, mas também o seu perfil individual apresentou-se como uma solução com aspectos positivos. O perfil de grupo de usuários demonstra qual conhecimento é esperado para que o usuário tenha, enquanto que o perfil individual demonstra o estado de conhecimento do usuário em um dado instante de tempo. A união desses perfis representou a possibilidade de se dar prioridade a outras áreas de conhecimento dependendo de quanto o usuário conhece, e assim tornar a busca mais dinâmica e específica para o perfil do usuário que a realiza. Uma vez que uma área de conhecimento é muito relevante para um usuário mas seu conhecimento individual é grande nessa área, outras áreas devem possuir maior relevância na busca.

A avaliação de resultados comprovou uma mudança significativa na ordenação dos documentos, variando-se o perfil individual do usuário, principalmente com a utilização de dois termos representando áreas de conhecimento opostamente relevantes ao usuário, isto é, uma área de bastante relevância e outra com menor importância. Os testes neste sentido apresentaram uma reordenação dos resultados priorizando aquela área em que o usuário não conhecia o bastante, e portanto, atingindo o objetivo final deste trabalho.

Apesar de alguns testes não terem causado bastante diferença na ordenação dos documentos que era mais relevantes à busca, a modificação do posicionamento dos demais documentos refletiu, de forma geral, positivamente com as alterações realizadas no perfil individual, de forma que fossem apresentados documentos semanticamente de acordo com as áreas de conhecimento que o usuário possuía maior necessidade de conhecimento, com base em seu perfil individual e de grupo.

O Google possuiu bastante relevância para no contexto das buscas do sistema, pois além de fornecer os documentos para a construção do corpus, seus documentos apresentam conteúdos semanticamente relevantes, trazendo o Sistema FGAVirtual de um ambiente de testes para um ambiente real de aprendizagem.

6 Trabalhos Futuros

Para trabalhos futuros, serão realizadas avaliações comportamentais do sistema, para identificar quais são as leis de formação da busca com um corpus mais rico de documentos e definição mais complexa de relacionamento entre os documentos e as diversas áreas de conhecimento envolvidas no trabalho. Este trabalho exige uma análise sistemática de especialistas em cada uma das áreas de conhecimento, para que a relação seja o próxima do real e a análise possa ser mais exata.

Outra iniciativa de trabalhos futuros será a inclusão de outros tipos de arquivos ao sistema além de páginas HTML, como arquivos de vídeo e slides. Isso auxiliará o aprendizado do aluno que realiza a pesquisa, pois é mais fácil aprender com mídia audiovisual do que apenas lendo conteúdos em páginas da internet.

Por fim, como trabalhos futuros visa-se a implantação do sistema FGAVirtual como extensão do trabalho desenvolvido SmartQuizz. O FGAVirtual será um auxiliador nos estudos dos usuários que utilizarem o SmartQuizz, uma vez que a cada resposta incorreta que o usuário fornecer, o sistema fornecerá conteúdos e arquivos que possibilitem o estudante a aprender os conteúdos necessários, tornando o estudo mais dinâmico para o usuário.

Apêndices

APÊNDICE A – Consultas utilizando um termo

Nesta sessão são apresentados os resultados de todas as consultas realizadas no sistema utilizando um termo, além daquelas já apresentadas neste trabalho.

As Tabelas 22, 23 e 24 representam o resultado da busca "química" utilizando como perfil de Grupo o curso de Eletrônica.

Ordem	Documento
1	Química
2	Biorrefinaria
3	Física
4	Engenharia
5	Engenharia de Energia
6	Engenharia de Produção
7	Motor de Combustão Interna
8	Máquina Térmica
9	Engenharia Elétrica
10	Petróleo
11	Matemática Discreta
12	Matemática

Tabela 22 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Eletrônica e perfil individual de 25%

As Tabelas 25, 26 e 27 representam o resultado da busca "química" utilizando como perfil de Grupo o curso de Energia.

Ordem	Documento
1	Química
2	Biorrefinaria
3	Física
4	Engenharia
5	Engenharia de Energia
6	Engenharia de Produção
7	Motor de Combustão Interna
8	Máquina Térmica
9	Engenharia Elétrica
10	Petróleo
11	Matemática Discreta
12	Matemática

Tabela 23 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Eletrônica e perfil individual de 50%

Ordem	Documento
1	Química
2	Biorrefinaria
3	Física
4	Engenharia de Energia
5	Engenharia
6	Engenharia de Produção
7	Motor de Combustão Interna
8	Máquina Térmica
9	Engenharia Elétrica
10	Petróleo
11	Matemática Discreta
12	Matemática

Tabela 24 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Eletrônica e perfil individual de 75%

Ordem	Documento
1	Física
2	Engenharia
3	Engenharia de Energia
4	Engenharia de Produção
5	Motor de Combustão Interna
6	Biorrefinaria
7	Máquina Térmica
8	Engenharia Elétrica
9	Petróleo
10	Matemática Discreta
11	Matemática
12	Química

Tabela 25 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Energia e perfil individual de 25%

Ordem	Documento
1	Biorrefinaria
2	Física
3	Engenharia
4	Engenharia de Energia
5	Engenharia de Produção
6	Motor de Combustão Interna
7	Máquina Térmica
8	Engenharia Elétrica
9	Petróleo
10	Matemática Discreta
11	Matemática
12	Química

Tabela 26 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Energia e perfil individual de 50%

Ordem	Documento
1	Biorrefinaria
2	Física
3	Engenharia
4	Engenharia de Energia
5	Química
6	Engenharia de Produção
7	Motor de Combustão Interna
8	Máquina Térmica
9	Engenharia Elétrica
10	Petróleo
11	Matemática Discreta
12	Matemática

Tabela 27 – Tabela que representa o resultado final da ordenação utilizando o termo "química" em uma busca com o perfil de grupo de Energia e perfil individual de 75%

APÊNDICE B – Consultas utilizando dois termos

Nesta sessão são apresentados os resultados de todas as consultas realizadas no sistema utilizando dois termos, além daquelas já apresentadas neste trabalho.

O próximo teste aplicou a busca "química física" para usuário com perfil de grupo em Energia. A Tabela 28 representa quais são os documentos importantes para cada uma das áreas de conhecimento envolvidas:

Área de conhecimento	Documentos mais relevantes
Química	Química e Biorrefinaria
Física	Física e Motor de Combustão Interna

Tabela 28 – Tabela que representa os documentos mais relevantes para cada uma das áreas de conhecimento envolvidas na busca "química física"

Para este perfil de grupo (Engenharia de Energia), as áreas de conhecimento são razoavelmente relevantes, com pesos de 100 para a área de química e 500 para a área de física. As Tabelas 29, 30, 31 e 32 apresentam o resultado final da busca.

Ordem	Documento
1	Física Moderna
2	Química
3	Petróleo
4	Matemática
5	Engenharia Elétrica
6	Motor de Combustão interna
7	Engenharia de Energia
8	Matemática discreta
9	Transmissão de Energia Elétrica
10	Engenharia
11	Circuito Digital
12	Engenharia de Produção
13	Máquina Térmica
14	Física
15	Engenharia Eletrônica
16	Computação Quântica
17	Engenharia Mecânica
18	Eletromagnetismo
19	Biorrefinaria

Tabela 29 – Tabela que representa o resultado final da ordenação utilizando os termos "química física" em uma busca com o perfil de grupo de Energia e perfis individuais de 0% e 0% respectivamente

Ordem	Documento
1	Engenharia
2	Engenharia de Energia
3	Engenharia de Produção
4	Motor de Combustão interna
5	Máquina Térmica
6	Engenharia Elétrica
7	Matemática
8	Petróleo
9	Engenharia Eletrônica
10	Circuito Digital
11	Transmissão de Energia Elétrica
12	Computação Quântica
13	Engenharia Mecânica
14	Matemática Discreta
15	Eletromagnetismo
16	Biorrefinaria
17	Física
18	Química
19	Física Moderna

Tabela 30 – Tabela que representa o resultado final da ordenação utilizando os termos "química física" em uma busca com o perfil de grupo de Energia e perfis individuais de 0% e 100% respectivamente

Ordem	Documento
1	Física Moderna
2	Petróleo
3	Matemática
4	Engenharia Elétrica
5	Motor de Combustão interna
6	Engenharia de Energia
7	Engenharia
8	Engenharia de Produção
9	Matemática Discreta
10	Máquina Térmica
11	Transmissão de Energia Elétrica
12	Circuito Digital
13	Química
14	Engenharia Eletrônica
15	Computação Quântica
16	Engenharia Mecânica
17	Eletromagnetismo
18	Física
19	Biorrefinaria

Tabela 31 – Tabela que representa o resultado final da ordenação utilizando os termos "química física" em uma busca com o perfil de grupo de Energia e perfis individuais de 100% e 0% respectivamente

Ordem	Documento
1	Química
2	Física
3	Física Moderna
4	Biorrefinaria
5	Eletromagnetismo
6	Engenharia Mecânica
7	Computação Quântica
8	Engenharia de Energia
9	Engenharia Eletrônica
10	Máquina Térmica
11	Engenharia de Produção
12	Motor de Combustão interna
13	Engenharia
14	Engenharia Elétrica
15	Circuito Digital
16	Transmissão de Energia Elétrica
17	Matemática
18	Matemática Discreta
19	Petróleo

Tabela 32 – Tabela que representa o resultado final da ordenação utilizando os termos "química física" em uma busca com o perfil de grupo de Energia e perfis individuais de 100% e 100% respectivamente

Referências

- ABRAEAD. *Anuario Brasileiro Estatístico de Educação Aberta e a Distância*. 2008. Citado 2 vezes nas páginas 13 e 35.
- ANDERSON, T. *The theory and practice of online learning*. [S.l.]: Athabasca University Press, 2008. Citado na página 35.
- CARMISINI, A.; VAHLDICK, A. Comparativo entre frameworks de javaserver faces: apache tobago, primefaces e richfaces. *Revista eletrônica do Alto Vale do Itajai*, 2012. Citado na página 30.
- CAVELLUCCI, L. C. B. Estilos de aprendizagem: em busca das diferenças individuais. *Curso de Especialização em Instrucional Design*, 2005. Citado 2 vezes nas páginas 33 e 34.
- CHIDLOVSKII, B.; GLANCE, N.; GRASSO, A. *System and method for collaborative ranking of search results employing user and group profiles derived from document collection content analysis*. Google Patents, 2001. US Patent 6,327,590. Disponível em: <<http://www.google.com/patents/US6327590>>. Citado 2 vezes nas páginas 31 e 32.
- CLARK, R. E. *Learning from media: Arguments, analysis, and evidence*. [S.l.]: IAP, 2001. Citado na página 34.
- FAHY, P. J. Media characteristics and online learning technology. *Theory and practice of online learning*, p. 137, 2004. Citado na página 35.
- FREITAS, S. A. A. et al. Smart quizzes in the engineering education. *International Mechanical Engineering IMECE*, 2012. Houston, Texas, USA. Citado 4 vezes nas páginas 34, 38, 39 e 48.
- GOOGLE. *Google Inc.* 2004. <<https://www.google.com.br/>>. [Online; accessed 3-Junho-2014]. Citado na página 23.
- JSF. *Oracle JSF*. 2004. <<http://www.oracle.com/technetwork/java/javasee/javaserverfaces-139869.html>>. [Online; accessed 30-Maio-2014]. Citado na página 31.
- LEFFA, V. J. Nem tudo o que balança cai: objetos de aprendizagem no ensino de línguas. *Polifonia*, v. 12, n. 12 (2), 2006. Citado na página 40.
- MARTINS, G. O. Avaliação de algoritmos para ordenação de documentos digitais recuperados em busca. 2013. Citado 10 vezes nas páginas 27, 28, 29, 32, 33, 37, 38, 58, 59 e 63.
- PEREIRA, A. T. C.; SCHMITT, V.; DIAS, M. R. A. C. *Ambientes Virtuais de Aprendizagem: em diferentes contextos*. 2007. Rio de Janeiro: Ciência Moderna Ltda. Citado 2 vezes nas páginas 34 e 35.
- PRIMEFACES. *JSF PrimeFaces*. 2009. <<http://www.primefaces.org/>>. [Online; accessed 30-Maio-2014]. Citado na página 55.

- RIEHLE, D. *Framework Design: A Role Modeling Approach*. 2000. Ph.D. Thesis, No. 13509. Zurich, Switzerland, ETH Zurich. Citado 2 vezes nas páginas 30 e 31.
- SALTON, G.; MCGILL, M. J. *Introduction to modern information retrieval*. McGraw-Hill, Inc., 1986. Citado na página 29.
- SCHWABE, D.; ESMERALDO, L. *Engineering Web Applications for Reuse*. 2001. Catholic University of Rio de Janeiro PUC. Citado na página 31.
- SHARMA, A.; ADHAO, N.; MISHRA, A. A survey: Static and dynamic ranking. *International Journal of Computer Applications*, Foundation of Computer Science, 244 5th Avenue, # 1526, New York, NY 10001, USA India, v. 70, n. 14, 2013. Citado 2 vezes nas páginas 27 e 28.
- SIMMONS, D. E. *The forum report: E-learning adoption rates and barriers*. 2002. Citado na página 35.
- STRUTS. *Apache Struts*. 2000. <<http://struts.apache.org/>>. [Online; accessed 30-Maio-2014]. Citado na página 31.
- TERRIER. *Terrier Website*. 2011. <<http://terrier.org/>>. [Online; accessed 26-Maio-2014]. Citado 3 vezes nas páginas 27, 29 e 32.
- VRAPTOR. 2003. <<http://www.vraptor.org/pt/>>. [Online; accessed 30-Maio-2014]. Citado na página 31.
- YATES, R. B.; NETO, B. R. et al. *Modern information retrieval*. [S.l.]: ACM press New York, 1999. Citado na página 29.