



UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO
ENGENHARIA DE COMPUTAÇÃO E SISTEMAS

FRANCISCO ADELTON ALVES RIBEIRO

**RECOMENDAÇÃO DE DOCUMENTOS PARA OS USUÁRIOS
DO AVA MOODLE A PARTIR DAS HASHTAGS POSTADAS
NOS FÓRUNS**

São Luís
2013

Ribeiro, Francisco Adelton Alves.

Recomendação de documentos para os usuários do AVA Moodle a partir das hashtags postadas nos fóruns / Francisco Adelton Alves Ribeiro. – São Luís, 2013.

108 f

Dissertação (Mestrado) – Curso de Engenharia de Computação e Sistemas, Universidade Estadual do Maranhão, 2013.

Orientador: Prof. Dr. Luís Carlos Costa Fonseca

FRANCISCO ADELTON ALVES RIBEIRO

**RECOMENDAÇÃO DE DOCUMENTOS PARA OS USUÁRIOS
DO AVA MOODLE A PARTIR DAS HASHTAGS POSTADAS
NOS FÓRUNS**

Dissertação de Mestrado em Engenharia da Computação e Sistemas, na área de concentração Informática na Educação, apresentada à Coordenadoria do Programa de Pós-Graduação em Engenharia da Computação e Sistemas da Universidade Estadual do Maranhão como requisito parcial à obtenção do título de mestre.

Orientador: Prof. Dr. Luís Carlos Costa Fonseca

São Luís

2013

FOLHA DE APROVAÇÃO

FRANCISCO ADELTON ALVES RIBEIRO

RECOMENDAÇÃO DE DOCUMENTOS PARA OS USUÁRIOS
DO AVA MOODLE A PARTIR DAS HASHTAGS POSTADAS
NOS FÓRUNS

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia da Computação e Sistemas da Universidade Estadual do Maranhão, como requisito parcial para obtenção do título de Mestre em Engenharia da Computação e Sistemas pela Banca Examinadora composta pelos membros:

BANCA EXAMINADORA

Prof. Dr. Luís Carlos Costa Fonseca
Universidade Estadual do Maranhão – PECS/UEMA

Prof. Dr. Fábio de Jesus Lima Gomes
Instituto Federal de Educação, Ciência e Tecnologia do Piauí – IFPI

Prof^a. Dr. Eveline de Jesus Viana Sá
Instituto Federal de Educação, Ciência e Tecnologia do Maranhão – IFMA
PECS/UEMA

Prof. Dr. Henrique Mariano Costa do Amaral
Universidade Estadual do Maranhão – PECS/UEMA

Aprovada em: 30 de outubro de 2013.

Dedico este trabalho aos meus pais:
Francisco Paulo Ribeiro (*in memoriam*) e
Maria Odinéa Alves Ribeiro.
E a Deus.

AGRADECIMENTOS

Agradeço primeiramente a Deus, a quem devo tudo o que sou.

Aos meus pais, Francisco Paulo Ribeiro e Maria Odinéa Alves Ribeiro, que me ensinaram com amor valores sem os quais não chegaria até aqui.

Ao meu orientador, Professor Dr. Luís Carlos da Costa Fonseca, que sempre demonstrou disponibilidade, paciência, amizade, pessoa que levo como inspiração e exemplo para meu futuro.

Aos professores do PECS/UEMA, por seus ensinamentos.

À minha esposa Adriana de Oliveira Ribeiro, por sempre me incentivar e apoiar em todas as minhas decisões, pela paciência e compreensão nos momentos que estive ausente.

Aos meus filhos Vanessa de Oliveira Ribeiro e Murilo de Oliveira Ribeiro, pelo incentivo, amor, inspiração e compreensão nos momentos de ausência.

À amiga Lidinalva Almada Coutinho com quem convivi durante estes dois anos em momentos decisivos e de estudo.

Aos amigos Elmo de Jesus Nery Júnior, José Wilker Pereira Luz, Marcelo Eugênio Castro Gonçalves, Miguel de Sousa Freitas, Otílio Paulo da Silva Neto e Wilker Magalhães de Jesus Noleto, por estarem comigo nos momentos de trabalho e diversão.

À banca examinadora, pelas valiosas críticas e sugestões.

E a todos aqueles que de alguma forma contribuíram para que eu pudesse iniciar, desenvolver e/ou concluir este trabalho.

RESUMO

A pesquisa tem como objetivo o desenvolvimento de um módulo adicional ao Ambiente Virtual de Aprendizagem (AVA) Moodle, a partir do conjunto de hashtags atribuídas nas postagens dos fóruns, com a adequação da medida estatística TF-IDF (Term Frequency – Inverse Document Frequency) utilizada para avaliar o quanto uma palavra é importante para um documento em relação a uma coleção (*corpus*), permitindo a execução de um algoritmo para classificação e ranqueamento de conteúdos a partir dos termos mais utilizados nos fóruns. As recomendações foram geradas utilizando a técnica de filtragem baseada em conteúdo, além da medida de similaridade, para determinar o quanto as *hashtags* estão relacionadas com o conteúdo do repositório local de arquivos do AVA Moodle. O trabalho apresenta diferentes técnicas usadas em sistema de recomendação, além de descrever e explicar como o novo módulo foi criado para o referido ambiente. Foram realizados experimentos com alunos de um curso técnico em informática que permitiu a verificação do sistema de recomendação e o nível de satisfação dos usuários.

Palavras-chave: Hashtag. Sistema de Recomendação. Objetos de Aprendizagem. Ambiente Virtual de Aprendizagem. Re却itório de Objetos de Aprendizagem.

ABSTRACT

The research has as main purpose the development of an additional module to the Virtual Environment of Learning (VEL) Moodle, from the set of hashtags attributed at the forum posts, with the proper statistical measure TF-IDF (Term Frequency-Inverse Document Frequency) used to evaluate how important a word is to a document in relation to a collection (corpus) allowing the execution of an algorithm for classification and ranking of content from the most commonly used terms in the forums. The recommendations were generated using a filtering technique based on content of the local repository file Moodle VEL. This paper presents different techniques used in a recommended system, and besides, it describes and explains how the new module was created for that environment. Experiments were performed with students of a computer technician course who allowed the verification of the recommendation system and the level of the users satisfaction.

Keywords: Hashtag. Recommendation System. Learning Objects. Virtual Environment of Learning. Repository Objects of Learning.

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Applications Program Interfaces</i>
AVA	Ambiente Virtual de Aprendizagem
CHIC	Classificação Hierárquica, Implicativa e Coesitiva
CI	Ciência da Informação
EaD	Educação a Distância
E-TEC	Escola Técnica Aberta do Brasil
GNU	General Public License
IBOPE	Instituto Brasileiro de Opinião Pública e Estatística
IDF	<i>Inverse Document Frequency</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
IESB	Instituto de Educação Superior de Brasília
LDB	Lei de Diretrizes e Bases da Educação
LMS	<i>Learning Management System</i>
LOM	<i>Learning Object Metadata</i>
LorSys	<i>Learning Object Recommender System</i>
LTSC	<i>Learning Technology Standards Committee</i>
MEC	Ministério da Educação e Cultura
MOODLE	<i>Modular Object-Oriented Dynamic Learning Environment</i>
NIED	Núcleo de Informática Aplicada à Educação
OA	Objetos de Aprendizagem
PHP	<i>Personal Home Page</i>
RI	Recuperação de Informação
SGBD	Sistema Gerenciador de Banco de Dados
SR	Sistema de Recomendação
SRD	Sistema de Recomendação de Documentos
TF	<i>Term Frequency</i>
TIC	Tecnologias da Informação e Comunicação
UAB	Universidade Aberta do Brasil
UCB	Universidade Católica de Brasília

LISTA DE FIGURAS

Figura 1.1 - Etapas da Pesquisa	17
Figura 2.1 - Fórmula do Coeficiente de Pearson	27
Figura 2.2 - Cálculo de predição	28
Figura 2.3 - Representação do processo de recuperação da informação	31
Figura 2.4 - Representação vetorial de um arquivo com dois vetores de indexação	34
Figura 2.5 - Espaço vetorial contendo dois documentos	36
Figura 2.6 - Função cosine vector similarity	38
Figura 2.7 - Autenticação do TelEduc	52
Figura 2.8 - Página inicial do Blackboard	54
Figura 2.9 - Ambiente Virtual de Aprendizagem do Curso de Engenharia da Computação	56
Figura 2.10 - Dez países por registro	56
Figura 4.1 - Diretório e arquivos do MOODLE	67
Figura 4.2 - Diretório de dados moodledata	68
Figura 4.3 - Repositório de plugins do Moodle	70
Figura 4.4 - Habilitando Repositório Dropbox	70
Figura 4.5 - Habilitar/Desabilitar Repositório Local	71
Figura 4.6 - Instância para pasta do servidor web	71
Figura 4.7 - Janela do File Picker do MOODLE	72
Figura 4.8 - Plugin Sistema de Arquivos	72
Figura 4.9 - Visão Geral do Sistema de Recomendação	73
Figura 4.10 - Grupo de tabelas fórum	74
Figura 4.11- Diagrama de Classe do grupo Fórum	75
Figura 4.12- Diagrama de Banco de Dados do Fórum	77
Figura 5.1 - Componentes do SRD	79
Figura 5.2 - Esquema TF/IDF	81
Figura 5.3 - Código TF/IDF	82
Figura 5.4 - Diagrama dos componentes do sistema	83
Figura 5.5 - Diagrama de Entidade e Relacionamentos - Novas tabelas	84
Figura 5.6 - Diagrama de Atividades	85
Figura 5.7 - Diagrama de sequência das recomendações	86
Figura 5.8 - Mensagem contendo hashtag do fórum MOODLE	87
Figura 5.9 - Bloco com recomendações	88
Figura 5.10 – Visualização e avaliação do documento recomendado	89
Figura 5.11 - Página inicial do curso ABC da Informática	90
Figura 5.12 - Gráfico alunos inscritos e desistentes	91

LISTA DE QUADROS

Quadro 1.1 - Organização dos capítulos frente às etapas de pesquisa	19
Quadro 2.1 - interesse do usuário	24
Quadro 2.2 - Avaliação de usuários em relação a livros de diversas áreas	26
Quadro 2.3 - Similaridade do usuário A	28
Quadro 2.4 - Sistemas que utilizam folksonomia	45
Quadro 2.5 – Estatística MOODLE	55
Quadro 3.1 - Analise dos trabalhos correlacionados	64
Quadro 4.1 - Dicionário de Dados mdl_forum-discussions	75
Quadro 4.2 - Dicionário de Dados mdl_forum-post	76
Quadro 5.1 – Exemplo de ranqueamento das hashtags postadas	80
Quadro 5.2 – Conceitos das avaliações	92
Quadro 5.3 - Publicação	92

LISTA DE FÓRULAS

Fórmula 2.1 – Cálculo de relevância	21
Fórmula 2.2 – Cálculo da média de pontuação	27
Fórmula 2.3 – Similaridade	38
Fórmula 5.1 - Adequação da Similaridade	82

SUMÁRIO

1 INTRODUÇÃO	14
1.1 CONTEXTUALIZAÇÃO E PROBLEMÁTICA	14
1.2 OBJETIVO	15
1.3 METODOLOGIA	16
1.4 ROTEIRO DA PESQUISA	16
1.4.1 <i>Etapa A – Teórica</i>	17
1.4.2 <i>Etapa B – Estrutura do AVA MOODLE</i>	18
1.4.3 <i>Etapa C – Sistema Proposto</i>	18
1.4.4 <i>Etapa D – Análise</i>	19
1.5 ORGANIZAÇÃO DO TEXTO.....	19
2 REFERENCIAL TEÓRICO	20
2.1 SISTEMA DE RECOMENDAÇÃO.....	20
2.1.1 <i>Recomendação Baseada em Conteúdo</i>	22
2.1.2 <i>Recomendação Baseada em Filtragem Colaborativa</i>	25
2.1.3 <i>Métodos de Recomendação Híbridos</i>	28
2.2 RECUPERAÇÃO DA INFORMAÇÃO	31
2.2.1 <i>Modelos para Recuperação da Informação</i>	33
2.3 OBJETOS DE APRENDIZAGEM	39
2.4 REPOSITÓRIO DIGITAL	42
2.5 ETIQUETAGEM E FOLKSONOMIA	43
2.6 EDUCAÇÃO A DISTÂNCIA.....	47
2.7 AMBIENTES VIRTUAIS DE APRENDIZAGEM	49
2.7.1 <i>TelEduc</i>	51
2.7.2 <i>Blackboard</i>	53
2.7.3 <i>MOODLE</i>	54
2.8 CONSIDERAÇÕES DO CAPÍTULO	57
3 TRABALHOS CORRELATOS.....	58
3.1 SISTEMA DE RECOMENDAÇÃO DE OBJETOS DE APRENDIZAGEM	58
3.2 MODELO DE SISTEMA DE RECOMENDAÇÃO DE MATERIAIS DIDÁTICOS PARA AMBIENTES VIRTUAIS DE APRENDIZAGEM	59
3.3 MODELO DE SISTEMA DE ALERTAS E RECOMENDAÇÃO PARA MEDIAR APRENDIZAGEM EM TURMAS HETEROGÊNEAS NOS CURSOS DE EAD	60
3.4 RECOMENDAÇÃO DE MENSAGENS NO FÓRUM DO MOODLE.....	62
3.5 SUMÁRIO DOS TRABALHOS	63

3.6	CONSIDERAÇÕES DOS TRABALHOS	63
4 ARQUITETURA DO MOODLE	66	
4.1	REQUISITOS DO SISTEMA	68
4.2	O NÚCLEO DO MOODLE.....	69
4.3	REPOSITÓRIOS DE ARQUIVOS DO MOODLE	69
4.4	BASE DE DADOS DO MOODLE	74
4.5	FÓRUNS DO MOODLE.....	77
4.6	CONSIDERAÇÕES DO CAPÍTULO	78
5 ARQUITETURA DO SISTEMA DE RECOMENDAÇÃO	79	
5.1	VISÃO GERAL	79
5.2	SISTEMA PROPOSTO.....	80
5.3	DESENVOLVIMENTO DO SISTEMA.....	86
5.3.1	<i>Aquisição da informação</i>	87
5.3.2	<i>Recomendação</i>	87
5.3.3	<i>Visualização das Recomendações</i>	88
5.3.4	<i>Visualização e avaliação do documento recomendado</i>	89
5.4	EXPERIMENTAÇÃO DO SISTEMA DE RECOMENDAÇÃO	90
5.5	RESULTADOS	91
5.6	CONSIDERAÇÕES DO CAPÍTULO	93
6 CONSIDERAÇÕES FINAIS	94	
REFERÊNCIAS	96	

1 INTRODUÇÃO

Neste capítulo estarão presentes aspectos relacionados à contextualização/problemática, objetivo da pesquisa, roteiro da pesquisa e organização do texto.

1.1 Contextualização e Problemática

A Internet é uma grande fonte de informação que vem crescendo de forma exponencial nos últimos anos. Segundo pesquisa do IBOPE (2013), o total de pessoas com acesso à internet no Brasil, no primeiro trimestre de 2013, chegou a 102,3 milhões, representa 9% a mais sobre os 94,2 milhões divulgados no terceiro trimestre de 2012, fato que caracteriza o aumento de usuários e de conteúdos. Apesar do enorme volume de dados disponíveis, muitas vezes as pessoas não conseguem localizar uma determinada informação, ou levam muito tempo para fazê-lo.

Para ter acesso às informações disponíveis na internet, os usuários contam com ferramentas auxiliares, que são os sistemas de busca como *Google*, que utiliza o algoritmo de Page Rank, publicado em Page et al (1998), realizando buscas de forma não automatizada, através das palavras-chave ou trechos de textos informados pelos usuários, as quais são denominadas de consultas com acesso casual único *ad hoc*. Apesar de eficazes, tais sistemas retornam muitos conteúdos que, em sua maioria, não atendem as necessidades dos usuários. Para o domínio educacional, isto não é necessariamente a melhor opção, pois os alunos precisam de conteúdos específicos, por exemplo, em uma disciplina de matemática do ensino médio onde logaritmo seja o tema central de busca, talvez seja interessante criar alguma associação com funções exponenciais, com o propósito de mostrar conceitos fundamentais relacionados com o tema pesquisado.

O processo que envolve a busca por informações que possam auxiliar os usuários na internet, tanto em termos gerais, como quando utilizados para fins educacionais, tornou-se um desafio. Atualmente, buscar e sugerir informações relacionadas à sugestão de conteúdos para alunos tornou-se uma tarefa pouco trivial.

Nesse sentido, o trabalho de Cazella, Drumm, Barbosa (2010) aborda este problema, mencionando que o usuário não é capaz de selecionar as informações

mais adequadas à sua necessidade, seja por desconhecimento do tema, ou pela sobrecarga de informação provocada pelas ferramentas de busca utilizadas (exemplo, *google*). Tais necessidades vêm aos poucos sendo incorporadas através de plataformas educacionais que objetivam delimitar seus resultados de consultas através de grupos com características específicas.

Sobretudo, no caso da Internet, podemos citar os mecanismos de busca, que procuram facilitar a localização de informações, mas sabidamente acabam por gerar mais sobrecarga de informações, geralmente com informação irrelevante para o usuário. Nessa perspectiva, os usuários do *Modular Object – Oriented Dynamic Learning Environment* (MOODLE), necessitam de uma ferramenta para coletar e sugerir conteúdos relevantes relacionados ao tema em estudo. Autores como Burke (2007), O'Donovan, Smyth (2005) definem SR como sendo ferramentas de *software* personalizadas que fornecem recomendações e sugestões de itens que possam ser de utilidade para o usuário. Tais recomendações e sugestões devem ser relacionadas com músicas, livros, filmes, artigos entre outros.

Nesse sentido, esta pesquisa justifica-se diante da escassez de mecanismos que tratam a recomendação de Objetos de Aprendizagem (OA) no Ambiente Virtual de Aprendizagem (AVA) MOODLE, automatizando a indicação de materiais relevantes para o processo de ensino-aprendizagem dos alunos de forma mais satisfatória, minimizando a necessidade de eles recorrerem a *sites* de busca à procura de materiais complementares ao estudo.

Considerando esta lacuna, recomendamos adicionar ao AVA MOODLE um bloco contendo Sistema de Recomendação de Documentos (SRD), o qual será discutido ao longo deste trabalho.

1.2 Objetivo

Objetivando propor o desenvolvimento de um modelo de SRD, em forma de módulo, que será adicionado ao AVA MOODLE. As recomendações serão realizadas a partir do conjunto de *hashtags* atribuídas pelos usuários nas postagens dos fóruns, empregando a medida estatística TF-IDF (Term Frequency – Inverse Document Frequency) utilizada para avaliar o quanto uma palavra é importante para um documento em relação a uma coleção (corpus), permitindo a execução de um algoritmo para classificação e ranqueamento de conteúdos a partir dos termos utilizados nos fóruns.

1.3 Metodologia

Atendendo aos procedimentos metodológicos foram criadas, nesse estudo, as seguintes etapas:

1. Pesquisa bibliográfica sobre sistema de recomendação, objetos de aprendizagem, etiquetagem, folksonomia e ambiente virtual de aprendizagem;
2. Utilização da técnica de atribuição de pesos a documentos utilizadas na área da Recuperação da Informação (RI);
3. Adequação do esquema a frequência de ocorrência do termo no documento (Term Frequency – TF) e do inverso da frequência do termo entre documentos da coleção (Inverse Document Frequency – IDF) à estrutura dos cursos do AVA Moodle, fazendo uso de hashtags;
4. Criar e integrar o sistema de recomendação ao AVA Moodle;
5. Realização de teste com 30 alunos, o que possibilitou o agrupamento das hashtags postadas nos fóruns, fornecendo ao sistema informações necessárias para recomendar OA.

O conjunto destes requisitos é uma proposta de desenvolvimento de um modelo de SRD aos usuários do AVA MOODLE, baseado no conteúdo dos fóruns hashtags, de forma a criar um banco de informações relevantes para o sistema.

1.4 Roteiro da pesquisa

A presente pesquisa está dividida em 4 etapas, a Etapa A – Teórica, a Etapa B – Estrutura do AVA MOODLE, a Etapa C – Sistema Proposto, e a Etapa D – Análise. Cada etapa é composta por determinados conjuntos de atividades e produtos.

Para interpretar a Figura 1, as Etapas são delimitadas por círculos preenchidos com uma letra e acompanhados por uma legenda, desse modo, o círculo com a letra A e legenda teórica representa a Etapa A – Teórica. As atividades são organizadas em quadrados com cantos arredondados, paralelos à cada etapa, de forma a representar esta relação de pertinência; são compostas por caixas e com o título da atividade. Os produtos são representados por retângulos com os cantos arredondados em paralelo às atividades relacionadas. São compostos por itens, que podem ser estruturados.

1.4.1 Etapa A – Teórica

QP 1- Como os atuais ambientes virtuais de aprendizagem ligam com os aspectos relacionados a recomendação de documentos?

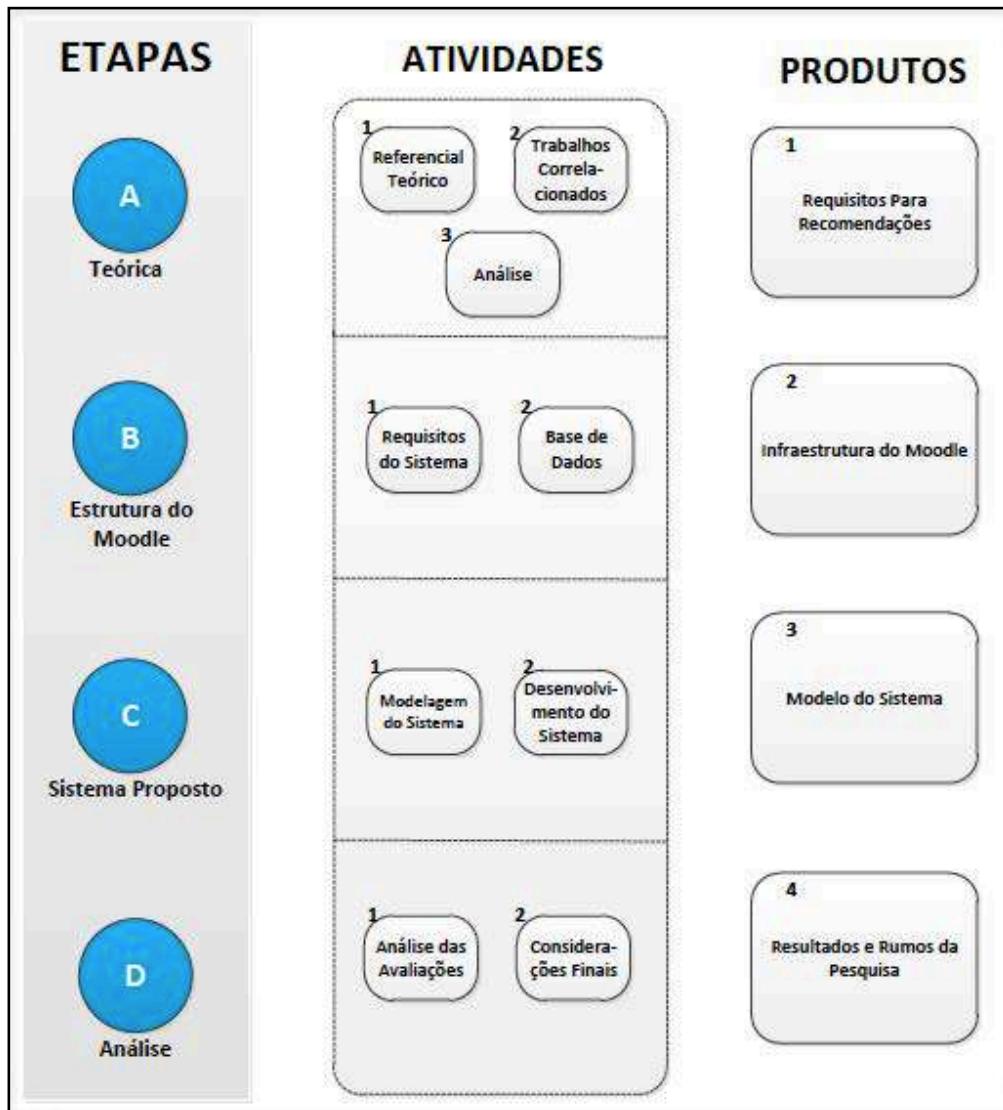


Figura 1.1: Etapas da Pesquisa.

De maneira a trabalhar nesta questão de pesquisa, a atividade **1 - Referencial Teórico**, apresenta conceitos referentes a sistemas de recomendação, técnicas de recomendação, recuperação da informação, objetos de aprendizagem, repositório digital, etiquetagem, folksonomia, educação a distância e ambientes virtuais de aprendizagem.

A atividade **2 – Trabalhos Correlacionados**, apresenta trabalhos importantes e, de maneira interdisciplinar, faz uso dos conceitos mencionados na atividade 1 e

que tem foco direcionado à Sistema de recomendação, Objetos de aprendizagem e ambientes virtuais de aprendizagem.

O conjunto de atividades proposto na etapa **A** resulta no produto **1 - Requisitos para Recomendação**, que, apoiado pela atividade **2**, enumera um conjunto de aspectos a serem considerados quando se trabalha com Sistema de Recomendação.

1.4.2 Etapa B – Estrutura do AVA MOODLE

QP 2 - De que forma pode ser modelado um sistema de recomendação de documentos para ser utilizado dentro do AVA MOODLE?

De modo a trabalhar com esta questão de pesquisa, faz-se necessário construir um modelo tecnológico baseado no produto **1** da etapa **A**.

Sendo assim, a atividade **1 – Requisitos do Sistema**, descreve os requisitos mínimos para instalação do AVA MOODLE. Tal descrição é importante para o funcionamento do sistema de recomendação proposto, que será detalhado na etapa **C**.

A atividade **2 – Estrutura das tabelas**, apresenta o grupo de tabelas pertencentes aos fóruns, o diagrama de classes e o dicionário de dados, bem como o relacionamento entre as principais tabelas que auxiliam os desenvolvedores na customização da plataforma.

1.4.3 Etapa C – Sistema Proposto

Descreve o sistema sistema proposto e a adequação da técnica TF/IDF à estrutura do AVA MOODLE.

A atividade **1 – Modelagem do Sistema**: apresenta a visão geral do sistema e seus principais componentes.

A atividade **2 – Desenvolvimento do Sistema**: define como as informações serão tratadas e comparadas com os documentos do repositório, apresentando o fluxo de atividades, diagrama de sequência e a descrição do sistema.

1.4.4 Etapa D – Análise

QP 3 - De que forma pode ser analisada as recomendações realizadas pelo sistema?

Apresenta a análise das avaliações realizadas com os alunos e os resultados obtidos.

A atividade **1** – Apresenta os resultados oriundos das avaliações realizadas pelos usuários, os quais são identificados, bem como o nível de satisfação dos usuários com o sistema.

A atividade **2** – Apresenta as considerações finais e sugestões de trabalhos futuros.

1.5 Organização do texto

Os capítulos são organizados conforme o quadro 1. A coluna **Etapa** apresenta a etapa da pesquisa, e a coluna **Capítulo** ao seu correspondente no texto.

Quadro 1.1: Organização dos capítulos frente às etapas de pesquisa.

Etapa	Capítulo
A	2 e 3
B	4
C	5
D	6

Elaborado pelo autor (2013).

2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados os conceitos que compõem a pesquisa realizada por esta dissertação. Dessa forma, os conceitos mencionados terão como base os sistemas de recomendação, técnicas de recomendação, recuperação da informação, objetos de aprendizagem, repositório digital, etiquetagem, folksonomia, educação à distância e ambientes virtuais de aprendizagem. Os conceitos que neste capítulo são apresentados visam descrever aspectos técnicos e teóricos que formam o núcleo da pesquisa realizada.

Conforme o capítulo 1, serão apresentadas as atividades e produtos referentes à **Etapa A Teórica, Atividade 1 – Referencial Teórico**.

2.1 Sistema de Recomendação

Sistemas de recomendação são agentes de informação personalizados que fornecem recomendações e sugestões de itens que possam ser de utilidade para o usuário (BURKE, 2007).

Segundo Cazella et al (2012), os Sistemas de Recomendação vêm auxiliar o usuário no processo de seleção de conteúdo. Em um sistema típico de filtragem de informação, as pessoas fornecem recomendações como entradas e o sistema agrupa e direciona para os indivíduos, que são considerados potenciais interessados.

A origem dos sistemas de recomendação pode ser traçada a partir de trabalhos em ciências cognitivas, teorias de aproximação, recuperação de informações, teorias sobre previsão de tempo, bem como, algumas ligações com ciências de gerenciamento e marketing (ADOMAVICIUS e TUZHILIN, 2005).

Dentre as áreas abordadas por Gediminas Adomavicius e Alexander Tuzhilin (2005), para este trabalho, menciona-se também como vertente crucial a área de Filtragem de Informações (sub-área de recuperação de informações) como de grande contribuição para a origem de tais sistemas.

Os sistemas de recomendação podem ser entendidos como qualquer sistema que procura recomendações individualizadas como saída, ou que tenha o efeito de guiar o usuário de forma personalizada os objetos interessantes e úteis, diante de uma grande variedade de opções (BURKE, 2002).

Combinando ideias como perfis de usuários, filtragem de informação e aprendizagem de máquina (O'DONOVAN; SMYTH, 2005), os sistemas de recomendação emergiram como uma área de pesquisa independente a partir da década de 1990. Isto ocorreu no momento em que, tendo por base avaliações anteriormente feitas, foram realizadas tentativas para estimar qual a avaliação seria feita por um usuário em relação a um item que não houvesse sido ainda avaliado (ADOMAVICIUS et al, 2005).

Entre os trabalhos pioneiros na área, está o de Tapestry (GOLDBERG et al, 1992), que visava a sugestão de memorando de acordo com características de usuários específicos; o de Shardanand, Maes (1995), que usa um processo de Filtragem Colaborativa na recomendação de álbuns e artistas de música; e de Resnick et al (1994), que recomenda artigos de notícias provenientes da Usenet.

Para (ADOMAVICUS & TUZHILIN, 2005) o problema da relevância nos SR pode ser formulado da seguinte forma: considerando que "C" seja o conjunto de todos os usuários e "S" o conjunto de todos os itens possíveis de serem recomendados – como músicas, livros, filmes, etc. A quantidade de "S" (possíveis itens a serem recomendados) pode ser muito grande, referindo-se a milhares de centenas ou milhões de itens em alguns modelos de aplicações, tais como em sistemas de recomendação de músicas e textos.

De forma semelhante, a quantidade de usuários pode também ser muito extensa, dezenas ou centenas de milhões em alguns casos. Consideremos "u" a utilidade ou relevância da informação em uma função matemática que mensura a relevância do item "s" para o usuário c, ou seja, $u: C \times S \rightarrow R$, onde R é o conjunto total de recomendações ordenadas. Então, para cada usuário $c \in C$ o sistema deve escolher o item $s' \in S$ que maximize a relevância da recomendação ao usuário. Em termos formais temos:

$$\forall c \in C, s'_c = \arg \max_{s \in S} u(c, s) \quad (1)$$

Fórmula 2.1: Cálculo da relevância.

Fonte: (TORRES, 2004).

A função de utilidade é normalmente representada pelo ordenamento crescente obedecendo ao ranking que indica a probabilidade de um usuário específico gostar ou se interessar por um item em particular. Pode, por exemplo, ser

representada por uma nota que evidencie quanto um usuário gostou de um item. Exemplificando, o filme “A Rede Social” foi avaliado por um usuário como **nota 8** em um escala ranqueada de 1 a 10.

Normalmente, os sistemas de recomendação encontram-se classificados em categorias, conforme a abordagem utilizada para realizar as recomendações, isto é, o tipo de recomendação adotado pelo sistema.

Os tipos de recomendações mais referenciados na literatura são, conforme Balabanovic; Shoham (1997); Adomavicius; Tuzhilin (2005):

- Recomendação baseada em conteúdo: o usuário recebe recomendação de itens similares aos que preferiu no passado;
- Recomendação baseada na filtragem colaborativa: o usuário recebe recomendação de itens que pessoas com gostos e preferências similares ao seu preferiram no passado;
- Recomendação baseada na filtragem híbrida: o sistema combina filtragem baseada em conteúdo e colaborativa.

Nas próximas sessões serão abordadas as três classificações de sistemas de recomendação citadas, de forma a detalhar o funcionamento de cada uma delas.

2.1.1 Recomendação Baseada em Conteúdo

Os sistemas baseados em conteúdo têm sua origem nos sistemas de recuperação de informação (BALABANOVIC; SHOHAM, 1997; ANAND; MOBASHER, 2005). Em um sistema de recomendação baseado em conteúdo, cada item a ser recomendado é definido por um conjunto de características associadas a ele mesmo (BURKE, 2002). Um livro, por exemplo, pode ser representado pelo título, autor, gênero, resenha etc., enquanto um texto pode ser representado pelas palavras que o compõem. É com base em suas características que os itens podem ser comparados e a semelhança entre eles estabelecida (OLIVEIRA, 2007).

Em geral, um sistema de recomendação baseado em conteúdo analisa documentos avaliados individualmente por um usuário e utiliza o conteúdo destes documentos e a avaliação recebida para inferir um perfil que pode ser usado para recomendar itens relevantes. Sistemas baseados em filtragem de conteúdo foram

desenvolvidos principalmente para recomendações fundamentadas em itens textuais, pois o conteúdo nesses sistemas é usualmente descrito com palavras-chave.

Trata-se de uma técnica que, definida por Shafer et al (1999), é geradora de recomendações apoiadas na correlação entre os itens. Para que isso aconteça, essa técnica deve possuir meios de construir o perfil do usuário conforme as suas interações, ou seja, de acordo com as características dos produtos que ele adquire com o passar do tempo. Adomavicius e Tuzhilin (2005) definiram que um modo de se construir esse perfil é armazenar um vetor de pesos, onde cada peso representa uma característica do produto.

Sistemas de recomendação baseados no conteúdo - Content-Based Methods tentam premeditar o grau de utilidade de um determinado produto para um determinado usuário, com base em seu passado. A utilidade $u(c, s)$ de um produto s para um usuário c tem seu cálculo baseado na utilidade $u(c, s_i)$, tal que si são os itens pertencentes ao conjunto S , adquiridos em algum momento do passado (Adomavicius e Tuzhilin, 2005). Ou seja, para que uma recomendação seja feita, são comparadas as características dos itens já adquiridos com o item alvo de uma possível recomendação.

A existência de uma base de dados contendo informações relativas às características dos produtos a serem recomendados é de fundamental importância para fazer recomendações dos itens aos usuários. Além disso, o perfil do usuário também é necessário, já que essa técnica de recomendação baseia-se na realização do cruzamento das características do produto com o perfil do usuário, o qual é construído a partir da análise do histórico de suas aquisições anteriores (FELFERNIG et al, 2008).

Para exemplificar essa forma de construção do perfil do usuário, suponha que uma loja virtual possua departamentos de eletrodomésticos, informática, eletrônicos e livros, e que o usuário tenha adquirido ao longo do tempo quatro produtos da área de informática, dois eletrônicos e um livro, conforme quadro a seguir:

Quadro 2.1: interesse do usuário.

Área	Histórico de compras	Total
Informática	Notebook	4
	Impressora	
	Ipad	
	Computador	
Eletrônicos	TV	2
	Home Theater	
Livros	Ciências Exatas	1
Eletrodomésticos	<<nenhum>>	0

Elaborado pelo autor (2013).

Conforme os dados apresentados, um sistema de recomendação baseado no conteúdo pode usar o número total de produtos de cada departamento para representar o peso de cada característica do produto para o usuário em questão, ou seja, pesos_{usuário}(informática, eletrônicos, livros e eletrodomésticos) corresponderia a pesos_A(4,2,1,0), o que permite ao sistema de recomendação identificar que a área de maior preferência do usuário é a área de informática, seguida da área de eletrônicos e, logo após, a área de livros. Dessa forma, cabe ao sistema de recomendação, com base no conteúdo, analisar quais produtos que deverão ser recomendados, tomando-se como base essas informações.

Essa abordagem, de geração de recomendações, possui uma limitação, de não poder gerar “surpresas” para o usuário, ou seja, não recomendar um produto que em um determinado momento esteja sendo bastante procurado por outros usuários. Para o usuário do exemplo anterior, que em seu histórico de aquisições apresenta o total 0 (zero) para o departamento de eletrodomésticos, a recomendação de um produto dessa área provavelmente não seria gerada, já que o grau de utilidade („algum_produto_do_departamento_de_eletrodomésticos”) é igual a zero.

Esse tipo de recomendação limita e reduz as chances do sistema identificar novas características do usuário, por não oferecer novidades, mesmo que seja um item muito utilizado entre os usuários, ou um item que outros usuários com as mesmas características tenham adquirido ultimamente.

A seguir abordaremos a recomendação baseada na filtragem colaborativa, descrevendo suas principais características.

2.1.2 Recomendação Baseada em Filtragem Colaborativa

Os sistemas de recomendação baseados em filtragem colaborativa procuram predizer a utilidade de um item para um usuário particular com base nas avaliações realizadas por outros usuários (ADOMAVICIUS; TUZHILIN, 2005). Consequentemente, não exige a compreensão ou reconhecimento das características dos itens para realizar as recomendações, e sim a definição de similaridade entre os gostos e as preferências dos usuários. Nos sistemas colaborativos, a essência está na troca de experiências entre as pessoas que possuem interesses comuns (REATEGUI; CAZELLA, 2005).

A técnica de Filtragem Colaborativa funciona da seguinte forma:

1. Os usuários avaliam os produtos em questão, e as avaliações são armazenadas no sistema;
2. Com base nas avaliações já armazenadas, o sistema compara-as com as dos outros usuários; os usuários com maior similaridades entre si são agrupados por perfis semelhantes;
3. O sistema retorna uma predição dos produtos e/ou conteúdos que possam ser mais relevantes ao usuário ativo (SARWAR, 2000).

Tais algoritmos não consideram o conteúdo dos itens, mas sim, a similaridade entre as pessoas para realizar o processo de recomendação, visando recriar o “boca a boca” – *word of mouth* (SHARDANAND; MAES, 1995) realizado entre amigos quando sugerem “coisas” entre si. Torres (2004) assinala que o nome Filtragem Colaborativa teve origem no sistema *tapestry* proposto por Goldberg et al (1992).

A ideia básica deste tipo de algoritmo é fazer a recomendação considerando as avaliações deste por pessoas que tenham um perfil semelhante ao do usuário que irá receber uma sugestão. Parte-se do princípio de que os usuários que concordaram sobre determinados itens no passado irão concordar no futuro (RESNICK et al, 1994; SHARDANAND, MAES, 1995; HILL et al, 1995; ou CACHEDA et al, 2011).

O problema pode ser formulado com uma matriz de usuários x itens, onde cada célula da matriz armazena o grau de utilidade $u(c, s)$ do item em relação a um usuário c . Esse poderá ser obtido pela avaliação do usuário em relação ao item recomendado. O trabalho está em premeditar essa utilidade para as células vazias, ou seja, premeditar a utilidade de um item para um usuário (MANOUSELIS e COSTOPOULOU, 2008),

permitindo ao sistema de recomendação avaliar o grau de utilidade e, assim, decidir se o item será recomendado ou não.

A formação da vizinhança é o passo mais importante para os sistemas baseados em filtragem colaborativa. Este foca o problema de como identificar os outros usuários mais próximos. Após calcular a proximidade entre todos os usuários, a tarefa é formar a vizinhança utilizando-se algum esquema que selecione e agrupe os usuários mais próximos daquele. Por fim, o próximo passo é derivar as recomendações obtidas pela geração da vizinhança e obter os produtos mais adequados para aquele usuário.

Para tornar mais fácil a compreensão sobre essa técnica, foi construída a tabela abaixo, a qual possui dados fictícios de uma loja virtual de venda de livros também fictícia. Cada célula representa o grau de utilidade de um livro adquirido por um determinado usuário, com números entre 0 e 2. Neste exemplo foi definida a seguinte forma de avaliação:

- Valor 2 para um livro já adquirido.
- Valor 1 para um livro cujos detalhes foram acessados, indicando certo grau de interesse.
- Valor 0 para um livro não adquirido e não acessado.

Para exemplificar, considera-se a quadro abaixo:

Quadro 2.2: Avaliação de usuários em relação a livros de diversas áreas.

	Livros						
Usuário	Português	Direito Constitucional	Geografia	Programação Java	Gramática	Informática para Concursos	
A	2	2	0	0	1	1	
B	0	1	1	2	0	2	
C	0	2	2	1	0	2	
D	1	2	0	0	1	1	
E	1	1	0	1	2	2	

Elaborado pelo autor (2013).

De acordo com o quadro acima, o usuário B não acessou o livro de Português, por outro lado, adquiriu o livro de Programação em Java. Conforme o exemplo, é possível notar que o usuário A tende a concordar com o usuário D, mas tende a discordar com o usuário E. Dessa forma, a técnica tenderia a recomendar ao usuário A os livros que o usuário D tenha se interessado.

Para implementar esta técnica, é preciso fazer uso de um cálculo matemático, sendo o mais comumente utilizado o da correlação de Pearson ou Coeficiente de Pearson (HERLOCKER, 2000; RESNICK, 1994; TORRES, 2004), entendendo que as avaliações precisam ser numéricas para serem armazenadas. Essas avaliações são, então, expressas por um número inteiro, e quanto maior o valor do número, mais o usuário gostou do produto. Essas avaliações permitem que a filtragem colaborativa possa ser utilizada para recomendar os mais variados tipos de produtos (TORRES, 2004). O cálculo é feito de acordo com a fórmula mostrada na Figura 2.1. A representação é feita por um número real dentro do intervalo de [-1, +1], sendo que -1 indica a ausência de correlação, e +1 indica uma forte correlação.

$$w_{a,u} = \frac{\sum_{i=1}^m [(r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)]}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2 \sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}}$$

Figura 2.1: Fórmula do Coeficiente de Pearson
Fonte: (TORRES, 2004).

De acordo com a Figura 2.1, a equação, $w_{a,u}$ é a correlação do usuário ativo a com um determinado usuário u , $r_{a,i}$ é a avaliação que o usuário ativo fez para o item i , e \bar{r}_a é a média de todas as avaliações do usuário ativo a .

Conforme Quadro 2.2, as pontuações que os usuários têm em comum são representadas conforme os seguintes vetores: usuário A = [2, 2, 1, 1]; usuário D = [1, 2, 1, 1]. Fazendo análise dos perfis, têm-se as médias das pontuações, sendo para o usuário A e o usuário D as médias 1,5 e 1,25, respectivamente. Desenvolvendo a fórmula, tem-se:

$$w_{usuário\ A,usuário\ D} = \frac{(2 - 1,5)(1 - 1,25) + (2 - 1,5)(2 - 1,25) + (1 - 1,5)(1 - 1,25) + (1 - 1,5)(1 - 1,25)}{\sqrt{(2 - 1,5)^2 + (2 - 1,5)^2 + (1 - 1,5)^2 + (1 - 1,5)^2} \sqrt{(1 - 1,25)^2 + (2 - 1,25)^2 + (1 - 1,25)^2 + (1 - 1,25)^2}} = 0,58 \quad (2)$$

Fórmula 2.2: Cálculo da média de pontuação.

Calculando a similaridade do usuário A com todos os outros usuários, obtém-se os valores apresentados no quadro 2.3 abaixo:

Quadro 2.3: Similaridade do usuário A.

Usuários	Pearson
B	-0,50
C	-0,50
D	0,58
E	-1,00

Elaborado pelo autor (2013).

Portanto, o usuário A é similar ao usuário D. Por outro lado, o usuário D não é similar ao usuário E.

A partir de então, calcula-se a predição, que consiste em definir quais produtos serão recomendados. Segundo Cazella (2006), o cálculo de predição de um produto para um determinado usuário pode ser feito com a equação abaixo:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^n |w_{a,u}|}$$

Figura 2.2: Cálculo de predição.
Fonte: (TORRES, 2004).

No momento de determinar quais produtos recomendar a um usuário, os valores de predição devem ser calculados para todos os livros, e os que possuem maior valor de predição devem ser recomendados.

A seguir, abordaremos a recomendação baseada no método de recomendação híbridos, descrevendo suas principais características.

2.1.3 Métodos de Recomendação Híbridos

Os sistemas híbridos combinam mais de uma técnica de recomendação para construir sistemas de recomendação mais eficientes (TORRES, 2004; BURKE,

2002; ADOMAVICIUS; TUZHILIN, 2005). Tais sistemas procuram reduzir ou até eliminar deficiências difíceis de superar dentro dos limites de uma abordagem de recomendação simples (ANAND; MOBASHER, 2005; TORRES, 2004).

Donaldson (2007) definiu que Sistema de Recomendação Híbrido é aquele que combina a recomendação com base no passado dos usuários e a recomendação com base na colaboração dos usuários, ou seja, com base na correlação item-item e na correlação pessoa-pessoa.

Os primeiros sistemas de recomendação híbridos combinaram as filtragens colaborativa e baseada em conteúdo: era possível alcançar as vantagens das técnicas baseadas em conteúdo, incluindo as previsões prévias para cobrir todos os itens e usuários, enquanto se ganhavam as vantagens das previsões precisas da filtragem colaborativa, como o aumento do número de usuários e das avaliações (CLAYPOOL et al., 1999). Ambas as técnicas são as mais comuns e utilizadas em combinações híbridas, pois ajudam a evitar certas limitações dos sistemas baseado em conteúdo e colaborativos (ANAND; MOBASHER, 2005; ADOMAVICIUS; TUZHILIN, 2005; BORGES; LORENA, 2010).

Burke (2002) e Adomavicius e Tuzhilin (2005) definiram alguns modos de como se obter um sistema de recomendação híbrido. São eles:

- Implementando a recomendação com base no passado dos usuários e com base na colaboração dos usuários separadamente, combinando as pontuações das recomendações para gerar uma única recomendação.
- Realizando a escolha sobre o modo como a recomendação será feita, se baseada no conteúdo ou baseada na filtragem colaborativa. A escolha pode ser feita de forma aleatória, ou por critérios estipulados dentro do sistema.
- Gerando várias recomendações ao mesmo tempo, sendo que todas as técnicas de recomendação são utilizadas, o que permite ao usuário obter várias recomendações geradas por diversas abordagens.
- Por meio da geração de recomendação em cascata, em que uma técnica é usada para selecionar alguns itens, que são avaliados por outra técnica de recomendação.
- Construindo um modelo unificado que incorpore ambas as abordagens de recomendação.

Para criar um Sistema de Recomendação híbrido, os perfis de usuários baseados em análise de conteúdo são mantidos e comparados para determinar usuários similares para recomendação colaborativa. Um usuário recebe uma recomendação quando um item tem uma nota alta em relação ao seu perfil ou quando ele recebeu uma nota alta de um usuário com perfil similar ao seu (BALABANOVIC, SHOHAM, 1997).

Na visão de Adomavicius e Tuzhilin (2005), existem diferentes maneiras de combinar métodos de filtragem colaborativa e filtragem baseada em conteúdo:

- Implementar os métodos colaborativos e baseados em conteúdos separadamente e combinar suas previsões.
- Incorporar algumas características das abordagens baseadas em conteúdo em uma abordagem colaborativa.
- Incorporar algumas características das abordagens colaborativas em uma abordagem baseada em conteúdo.
- Construir um modelo de unificação que incorpore as características das duas abordagens.

Algumas das principais características herdadas pela Filtragem Híbrida de cada uma das abordagens anteriores são, conforme Cazella & Reategui, 2005:

- Descoberta de novos relacionamentos entre usuários.
- Recomendação de itens diretamente relacionados ao histórico.
- Bons resultados para usuários incomuns.
- Precisão, independente do número de usuários.

A descoberta de novos conhecimentos entre usuários e a recomendação de itens diretamente relacionados ao histórico do usuário são herdadas da Filtragem Colaborativa, pois trabalha com a ideia de “perfis similares”. Já os bons resultados para usuários incomuns e a precisão são herdadas da Filtragem Baseada em Conteúdo, que não leva em consideração qualquer relacionamento entre perfis de usuários e não recomenda itens não relacionados ao perfil do usuário, ou seja, não realiza recomendações “surpresa”.

Essa abordagem favorece a possibilidade de um sistema recomendar produtos com base nas informações de cada usuário e com base nas informações de um conjunto de usuários, ampliando as chances de obter acertos em suas recomendações.

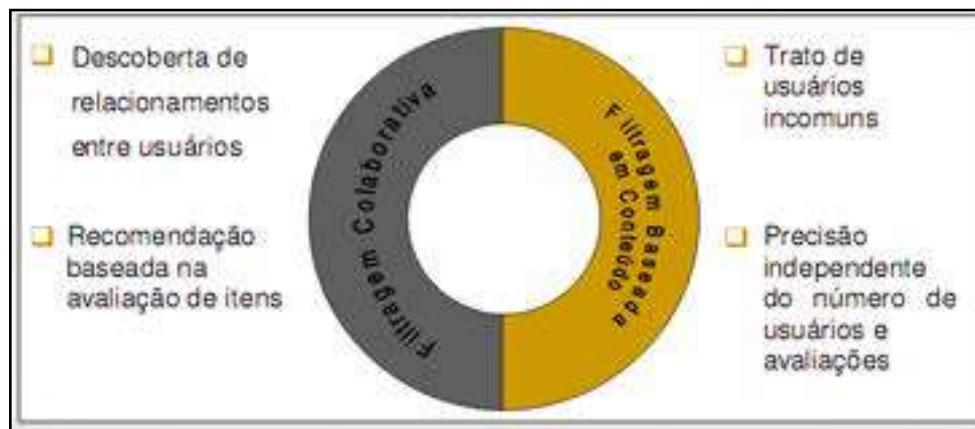


Figura 2.3: Características herdadas pela Filtragem Híbrida.

Fonte: (adaptado de CAZELLA; REATEGUI, 2005).

Segundo Huang et al (2002), sistemas híbridos podem obter diferentes graus de ganho em exatidão de predição, por utilizarem múltiplas fontes de informação (informações do item, do usuário e transacionais), variando de modestos benefícios a melhorias significativas. Porém, esta adição de informação nem sempre conduz a melhores resultados. A análise da variação na qualidade da recomendação, em função da multidimensionalidade da informação, requer um estudo muito aprofundado.

Os métodos citados são fortemente fundamentados em técnicas de recuperação de informação, que abordaremos a seguir.

2.2 Recuperação da informação

Recuperação de informação é uma subárea da ciência da computação que estuda o armazenamento e recuperação automática de documentos, que são objetos de dados, geralmente textos. No contexto da Ciência da Informação, o termo “recuperação da informação” significa, para uns, a operação pela qual se seleciona documentos, a partir do acervo, em função da demanda do usuário. Para outros, “recuperação da informação” consiste no fornecimento, a partir de uma demanda definida pelo usuário, dos elementos de informação documentária correspondente.

A recuperação da informação pode ser entendida como um processo de recuperação de referências de documentos em resposta a alguma solicitação, em que os sistemas de recuperação da informação são sistemas de operações

interligadas para identificar, entre um grande conjunto de informações, aquelas que são realmente úteis, isto é, que estão de acordo com a demanda expressa pelo usuário (ARAÚJO JÚNIOR, 2007).

Os sistemas de recuperação de informação devem apresentar o conteúdo dos documentos do corpus¹ e apresentá-los ao usuário de uma maneira que lhe permita uma rápida seleção dos itens que satisfazem total ou parcial a sua necessidade de informação, formalizada através da expressão de busca.

Buckland (1997, p. 806) define documento como “qualquer signo físico ou símbolo, preservado ou registrado, com a intenção de representar, reconstruir ou demonstrar um fenômeno físico ou abstrato”. Esta definição generaliza o conceito de documento a qualquer tipo de suporte, seja ele material ou digital.

No ambiente digital, os acervos de objetos digitais se multiplicam tanto no que se refere à sua topologia quanto à sua complexidade. Nesse cenário, textos, imagens, sons, vídeos, páginas Web e diversos outros objetos digitais requerem diferentes tipos de tratamento e representação para uma recuperação da informação eficaz (BURKER, 1999).

A recuperação de informação está diretamente ligada ao modelo e técnica de recuperação usada pelo Sistema de Recuperação da Informação (SRI), que, segundo Manning et al (2009), são programas de computador responsáveis pelas atividades de indexação, busca e classificação de grandes coleções de documentos em formato digital.

Para Araújo Júnior (2007), os SRI dizem respeito a um sistema de operações interligadas para identificar, dentre um grande conjunto de informações (uma base de dados, por exemplo), aquelas que são de fato úteis, ou seja, que estão de acordo com a demanda expressa pelo usuário. Nesse contexto, os SRI têm como principal objetivo recuperar a informação contida na coleção de documento disponibilizando apenas o que é útil ou relevante para o usuário. Em geral, a necessidade de informação é expressa em uma linguagem de consulta, composta por palavras-chave e operadores lógicos.

Buscadores, ferramentas de busca ou mecanismos de busca são ferramentas que auxiliam os usuários no processo de busca de informação na Web. Eles coletam informações sobre os conteúdos dos sites e os armazenam em bancos de dados

¹ Corpus significa conjunto de documentos.

que estão disponíveis para consulta. Realizando uma busca, o usuário poderá descobrir a localização exata das informações que deseja. Conforme dados de Alexa (2012), o mecanismo de busca do *Google* é o sítio com maior audiência na Web, seguido pelo sítio da plataforma de rede social *Facebook*. Este fato nos faz refletir sobre a importância dos buscadores no cotidiano das pessoas.

Os mecanismos de busca são fundamentais para a recuperação das informações na Web. Entretanto, uma busca à informação relevante é uma tarefa difícil, que depende da habilidade do usuário no uso das ferramentas de busca e a capacidade do buscador em, a partir de um termo ou conceito, compreender as necessidades do usuário e recuperar as informações adequadas.

A eficiência de um sistema de recuperação da informação está diretamente ligada ao modelo que ele utiliza. Um modelo, por sua vez, influencia diretamente no modo de operação do sistema.

2.2.1 Modelos para Recuperação da Informação

Os modelos para Recuperação da Informação (RI) especificam como os documentos e a necessidade de informação do usuário devem ser representados e comparados, a fim de produzir um resultado de busca (CROFT, 2000). Baeza-Yates e Ribeiro-Neto (1999) definem formalmente um modelo para RI, o qual é caracterizado na definição seguinte.

Definição Um modelo de Recuperação de Informação é uma 4-upla $\{D, Q, F, R(q_i, d_j)\}$ onde:

- D é um conjunto formado pelas representações dos documentos de uma coleção;
- Q é um conjunto formado pelas representações das necessidades de informação dos usuários;
- F é um arcabouço para a modelagem da representação dos documentos, das consultas e das suas relações;
- R(q_i, d_j) é uma função de ranking, a qual associa um número real com uma consulta $q_i \in Q$ em uma representação do documento $d_j \in D$.

De acordo com Baeza-Yates e Ribeiro-Neto (1999), os três modelos clássicos da RI são o modelo booleano, o modelo de espaço vetorial e o modelo probabilístico. Esses modelos são diferenciáveis pelos elementos da 4-upla. Por exemplo, o modelo booleano utiliza a teoria de conjuntos para representar seus documentos (FRAKES e BAEZA-YATES, 1992), o modelo de espaço vetorial utiliza ferramentas da álgebra linear (BERRY et al, 1999), e o modelo probabilístico utiliza a estimativa da probabilidade de um documento ser julgado como relevante (FUHR, 1992).

Esta pesquisa trabalha com o modelo de espaço vetorial proposto por Salton e Buckley (1998), por possuir técnicas capazes de inferir se um documento é similar a outro ou não realizando cálculos a partir de um vetor criado para cada documento. Para Salton e Buckley (1998), o modelo vetorial é a forma mais comum de representar documentos em mineração de textos. Neste modelo, um documento é representado por um vetor no qual cada elemento representa o peso, ou a relevância, do respectivo termo indexado para o documento. Cada vetor descreve a posição do documento em um espaço multidimensional, onde cada termo de indexação representa uma dimensão ou eixo. Cada elemento do vetor (peso) é normalizado de forma a assumir valores entre zero e um. Os pesos mais próximos de um (1) indicam termos com maior importância para a descrição do documento. A figura 2.4 mostra a representação gráfica de um documento **DOC₁** com termos de indexação e com pesos 0.4 e 0.7, respectivamente.

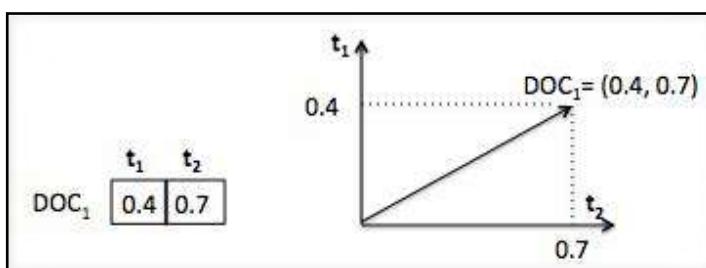


Figura 2.4: Representação vetorial de um arquivo com dois vetores de indexação.
Fonte: Elaborado pelo autor (2013).

O peso de um termo em um documento pode ser calculado de diversas formas. Os pesos são usados para computar a similaridade entre cada documento armazenado e uma consulta feita pelo usuário. Esses métodos de cálculo de peso

geralmente se baseiam no número de ocorrências do termo no documento (frequência).

Uma das formas de se calcular o peso, dada por Salton e Buckley (1998), é o modelo estatístico conhecido por TF/IDF (*Term Frequency–Inverse Document Frequency*). Seu principal objetivo é identificar termos que, apesar de muito frequentes em um documento, também aparecem com frequência em textos de outras áreas. Sendo assim, são considerados de pouco valor discriminatório, pois não ajudam a classificar um texto em áreas.

O modelo TF/IDF corresponde a uma medida estatística utilizada para avaliar o quanto uma palavra é importante para um documento em relação a uma coleção (*corpus*). Essa importância aumenta proporcionalmente com o número de vezes em que a palavra apareça no documento, e diminui de acordo com a frequência da palavra na coleção.

A função *TF* corresponde ao número de ocorrências de um termo em um documento. Uma palavra frequente no texto recebe assim uma maior relevância do que uma palavra menos frequente. A função *IDF* tenta modelar o poder discriminatório da palavra no âmbito do conjunto. Quanto menos documentos conterem a palavra, maior relevância lhe é atribuída. Unindo-se os dois conceitos em uma formulação logarítmica é possível produzir uma função para cálculo de pesos de termos de uma coleção de documentos e selecionar automaticamente índices adequados para eles.

Definição:

$TF = freq(k, D)$ (frequência do termo k no documento/consulta D);

$IDF = \log \frac{N}{n}$ (*inverse document frequency*), onde N é o número de termos na coleção e n número de vezes que o termo ocorre na coleção.

O produto de *TF* e *IDF* ($TF \times IDF$) seria portanto uma medida razoável:

$$W_t = freq(k, D) \cdot \log \frac{N}{n}$$

Para exemplificar a equação acima, suponha que um documento D contém os termos: *apresentação* 18 vezes, e *informação* 8. Na base de 2048 documentos, ocorrem os termos: *apresentação* em 128 deles; *história* em 16 e *informação* em 1024, fazendo o cálculo ($TF \times IDF$) temos:

$$TF \times IDF = (\text{apresentação}) = 18 * \log \frac{2048}{128} = 18 * 1.2 = 21.6;$$

$$TF \times IDF = (\text{história}) = 0 * \log \frac{2048}{16} = 0 * 2.1 = 0;$$

$$TF \times IDF = (\text{informação}) = 8 * \log \frac{2048}{1024} = 8 * 0.3 = 2.4.$$

Logo temos o vetor de termos: <21.6, 0, 2.4>, dessa forma o termo *informação* é ranqueado na primeira posição e *apresentação* na segunda. Se alterarmos a quantidade de vezes que o termo *informação* aparece no documento D de 8 para 90 vezes, o resultado do seu cálculo ($TF \times IDF$) mudaria para 27.0, provocando inversão na posição.

Ferneda (2003, p. 39) mostra uma figura com “dois documentos **DOC₁** e **DOC₂** representados no mesmo espaço vetorial. Os números positivos representam os pesos de seus respectivos termos. Termos que não estão presentes em um determinado documento possuem peso igual a zero” (Figura 2.5).

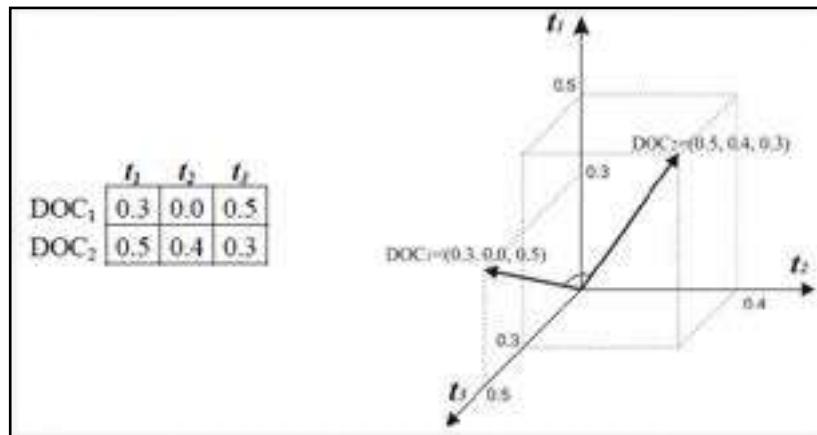


Figura 2.5: Espaço vetorial contendo dois documentos
Fonte: (FERNEDA, 2003).

Para calcular a similaridade entre documentos, os seguintes passos são seguidos:

- Definição da Coleção de Textos: consiste em definir a coleção que será utilizada no processo (RIZZI et al, 2000). Nesta pesquisa, essa coleção é composta das hashtags postadas nos fóruns.
- Identificação de Termos: o objetivo é identificar as palavras dos documentos, excluindo do texto símbolos, figuras, caracteres especiais,

etc., com intuito de facilitar a análise do conteúdo (RIZZI et al, 2000). Como nesta pesquisa vamos trabalhar com informações textuais, serão removidos os símbolos de pontuação e também os acentos das palavras. Tais remoções foram feitas porque alguns usuários não usam acentuação nas suas postagens e, para não atrapalhar o processo, fizemos as referidas remoções. Além disso, todo o texto foi convertido em letras minúsculas, para garantir que as comparações efetuadas vão retornar resultados realmente corretos, independente de terem sido escritos com letras maiúsculas ou minúsculas.

- Remoção de *stopwords*: consiste em ignorar as palavras que costumam aparecer em todos os documentos. Essas palavras são denominadas *stopwords*, uma expressão que pode ser traduzida por “palavras-negativas”, “palavras-ferramentas” ou “palavras-vazias”. Essas palavras são preposições, artigos, conjunções e demais palavras utilizadas para auxiliar na construção sintática das orações (WIVES, 2001). Nesta pesquisa, usou-se a lista *stopwords* para português (SNOWBALL, 2011).
- Seleção de Termos ou Características: busca identificar através de métodos específicos (como Frequência Absoluta, Frequência Relativa, etc.) os termos que mais caracterizam e representam o conteúdo do texto em questão (RIZZI et al, 2000).
- Método de Similaridade: aplicação de algum método para o cálculo de similaridade entre documentos.

As distâncias entre um documento e outro indicam seu grau de similaridade, ou seja, documentos que possuem os mesmos termos acabam sendo colocados em uma mesma região do espaço e, em teoria, tratam de assuntos similares.

A consulta do usuário também é representada por um vetor. Dessa forma, os vetores dos documentos podem ser comparados com o vetor da consulta e o grau de similaridade entre cada um deles pode ser identificado.

Os documentos mais similares (mais próximos no espaço) à consulta são considerados relevantes para o usuário e retornados como resposta para ela.

Uma das formas de calcular a proximidade entre os vetores é testar o ângulo entre estes vetores. No modelo original, é utilizada uma função batizada de *cosine vector similarity*, que calcula o produto dos vetores de documentos através da fórmula:

$$\text{similaridade } (Q, D) = \frac{\sum_{k=1}^n w_{qk} \cdot w_{dk}}{\sqrt{\sum_{k=1}^n (w_{qk})^2} \cdot \sqrt{\sum_{k=1}^n (w_{dk})^2}}$$

Figura 2.6: Função cosine vector similarity.
Fonte: (FERNEDA, 2003).

Onde,

- Q representa o vetor de termos da consulta;
- D é o vetor de termos do documento;
- W_{qk} são os pesos dos termos da consulta; e
- W_{dk} são os pesos dos termos do documento.

O grau de similaridade entre o documento **DOC₁** e o documento **DOC₂**, representados na Figura 2.5, é calculado como:

$$sim(DOC_1, DOC_2) = x = \frac{(0.3 \times 0.5) + (0.0 \times 0.4) + (0.5 \times 0.3)}{\sqrt{0.3^2 + 0.0^2 + 0.5^2} \times \sqrt{0.5^2 + 0.4^2 + 0.3^2}} = \frac{0.15 + 0.0 + 0.15}{\sqrt{0.34} \times \sqrt{0.5}} = 0.73 \quad (3)$$

Fórmula 2.3: Similaridade.

Fonte: Elaborado pelo autor (2013).

Depois dos graus de similaridade terem sido calculados, é possível montar uma lista ordenada (*ranking*) de todos os documentos e seus respectivos graus de relevância à consulta, da maior para a menor relevância.

Neste estudo, faremos adequação do tradicional esquema de *ranking* denominado de TF/IDF (BAEZA-YATES e RIBEIRO-NETO, 1999; MANNING et al., 2009), para quantificar um termo como fator discriminatório para todos os documentos da coleção, fazendo uso de um conjunto de hashtags postadas nos fóruns do AVA MOODLE, ordenando e recomendando OA a seus utilizadores. A seguir, será apresentada uma perspectiva para isto no domínio educacional, visando a possibilidade de recuperar informações que registrem relações entre usuários, OA e ambientes de ensino.

2.3 Objetos de aprendizagem

A proliferação de cursos usando educação a distância demanda uma produção intensa de recursos educacionais digitais para apoiar as atividades de ensino aprendizagem. A reusabilidade de recursos existentes, dentro de uma filosofia de conteúdos educacionais abertos, é importante e necessária para reduzir custos e aumentar a eficiência das equipes envolvidas neste processo. Todavia, um problema com o qual estas equipes se deparam é a dificuldade para encontrar os recursos existentes na própria instituição ou em repositórios externos. Tais recursos, estruturados como objetos de aprendizagem, podem estar dispersos em grande número de servidores e locais (TAROUCO et al, 2003).

Na tentativa de alcançar a eficiência e eficácia do processo de aprendizagem, a tecnologia de objetos de aprendizagem tem sido utilizada como uma estratégia de ensino, possibilitando ao professor criar componentes de materiais pedagógicos e organizá-los de modo a possibilitar sua reutilização, economizando tempo e custos na produção de conteúdos on-line.

Behar et al (2009, p. 65) entendem que os objetos de aprendizagem são “qualquer material digital, como, por exemplo, textos, animações, vídeos, imagens, aplicações, páginas web, de forma isolada ou em combinação, com fins educacionais”.

Segundo Longmire (apud BETTIO; MARTINS, 2002), objetos de aprendizagem são considerados como:

Uma entidade, digital ou não digital, que pode ser usada e reutilizada ou referenciada durante um processo de suporte tecnológico ao ensino e aprendizagem. Exemplos de tecnologia de suporte ao processo de ensino e aprendizagem incluem aprendizagem interativa, sistemas instrucionais assistido por computadores inteligentes, sistemas de educação à distância, e ambientes de aprendizagem colaborativa. Exemplos de objetos de aprendizagem incluem conteúdos de aplicações multimídia, conteúdos instrucionais, objetivos de aprendizagem, ferramentas de software e software instrucional, pessoas, organizações ou eventos referenciados durante o processo de suporte da tecnologia ao ensino e aprendizagem (LONGMIRE, 2001 apud BETTIO; MARTINS, 2004).

Silva (2005) assinala que, apesar de se estar usando, atualmente, um grande número de conceitos sobre Objetos de Aprendizagem, o conceito proposto pelo

Learning Technology Standards Committee (LTSC) do Institute of Electrical and Electronics Engineers (IEEE), que os descreve como os menores componentes instrucionais, define como sendo Objeto de Aprendizagem:

Qualquer entidade, digital ou não digital, que pode ser usada, reutilizada ou referenciada durante aprendizagem suportada por tecnologia. Exemplos de aprendizagem suportada por tecnologia incluem: sistemas de treinamento baseado em computador; ambientes de aprendizagem interativa; sistema inteligente de instrução auxiliado por computador; sistemas de aprendizagem a distância; e ambientes de aprendizagem colaborativa. Exemplos de objetos de aprendizagem incluem: conteúdo multimídia, conteúdo instrucional, objetivos de aprendizagem, ferramentas de software e software instrucional, e pessoas, organizações ou eventos referenciados durante aprendizagem suportada por tecnologia (LOM, 2000 apud WILEY, 2000).

Os objetos de aprendizagem também podem ser compreendidos como peças disponíveis em ambientes virtuais de ensino. Seu funcionamento em conjunto é equivalente ao jogo/brinquedo LEGO, pois também pode ser reconectado e inventar-se uma infinidade de configurações (VIEIRA, 2007).

Deve-se considerar que os objetos digitais são projetados especificamente para fins pedagógicos, apresentando-se em formatos diferentes, tais como textos, animações, apresentações, *home pages* e *applets*. No entendimento de Mendes, Souza e Caregnato (2004), para que um OA possa ser inserido em um ambiente de aprendizado, é necessário que tenha as seguintes características:

- Reusabilidade: deve ser projetado para ser reutilizado em contextos diversos.
- Adaptabilidade: deve ser adaptável a qualquer ambiente de ensino.
- Granularidade: o conteúdo de um OA deve ser fragmentado em pequenos “pedaços” para compor unidades maiores.
- Acessibilidade: deve estar disponível via web para ser usado em diversos locais.
- Durabilidade: independente da mudança de tecnologia, deve ter a possibilidade de continuar a ser usado.
- Interoperabilidade: pode ser usado em qualquer plataforma, independente de *hardware*, sistemas operacionais e navegadores.

Além dessas características, outras são apresentadas a seguir, conforme Longmire (2001 apud MOURA, 2005):

- Facilidade para atualização: a fragmentação do conteúdo do OA em pequenas porções facilita a localização e a atualização de conteúdo.
- Customização: os OAs podem ser personalizados para diferentes cursos ou usuários.
- Aumento do valor de um conhecimento: à medida que os OAs são reutilizados, eles se tornam melhores, aumentando o valor do conhecimento.
- Indexação e busca: possibilidade da criação de um banco de OAs para pesquisa de elementos que possam vir a constituir outro conteúdo.

Com base em Singh, citado por Bettio e Martins (2002), um objeto de aprendizagem, para ser bem estruturado, é dividido em três partes bem definidas:

- Objetivos: esta parte do objeto tem como intenção demonstrar ao aprendiz o que pode ser aprendido a partir do estudo desse objeto, além do pré-requisito para um bom aproveitamento do conteúdo.
- Conteúdo instrucional: parte que apresenta todo o material didático necessário para que, no término, o aluno possa atingir os objetivos definidos.
- Prática e *feedback*: uma das características importantes do paradigma objeto de aprendizagem é que, a cada final de utilização, julga-se necessário que o aprendiz verifique se o seu desempenho atingiu as expectativas.

Os objetos com as características acima citadas são normalmente armazenados em grandes bases de dados disponíveis na internet chamados de repositórios. Na tentativa de superar o problema de encontrar recursos ou repositórios externos, convém investigar e analisar alternativas metodológicas e tecnológicas para criar, gerenciar e reaproveitar os objetos de aprendizagem em diversos contextos e dispositivos.

2.4 Repositório digital

Na área da educação, são criados e disponibilizados muitos materiais, porém, acessá-los torna-se uma tarefa cansativa e muitas vezes fadada ao fracasso. De modo geral, a internet oferece uma gama de informações, mas não satisfaz as expectativas do usuário na mesma medida. Por essa razão, são criados materiais parecidos com propósitos similares.

Diante da crescente interação social via redes sociais, iniciada no final do século passado, a qual tem atingido a escola e seus agentes, não se pode ignorar a influência e a necessidade de uso dessas TIC como um suporte efetivo ao processo de ensino-aprendizagem, sobretudo em ambientes virtuais de aprendizagem, uma vez que cresce o uso dessas redes como meio de circulação das informações e nos diálogos entre pessoas, modificando-lhes o comportamento e os hábitos nas suas relações sociais. Sendo assim, por que não fazer uso dos recursos tecnológicos no processo de formação dos sujeitos, através da educação formal, inclusive mantendo a preocupação com a constante atualização tecnológica?

Para tanto, são inúmeras as iniciativas de várias instituições de ensino, no sentido de elaborar, armazenar e disponibilizar materiais pedagógicos produzidos por meio de repositórios digitais.

Para Nascimento (2009), o objetivo crucial dos repositórios digitais que armazenam ferramentas educacionais é disponibilizar os conteúdos e permitir a sua reutilização da forma que for mais conveniente. Contudo, para desenvolver um repositório digital, é fundamental a criação de um padrão de metadados como IEEE LOM e Dublin Core, que se baseiam na taxionomia para garantir a consistência entre os metadados dos diferentes autores de objetos de aprendizagem.

Ao apropriar-se desses recursos, esperamos que o professor seja capaz de utilizá-los e reutilizá-los de forma crítica, dialogando com teorias que tratam do tema em questão, para refletir e transformar suas experiências e, dessa forma, apresentar estratégias que possibilitem a apropriação e a utilização desses recursos, impulsionando um fazer pedagógico capaz de envolver os educandos, incentivando-os a serem autores e produtores do seu próprio conhecimento.

A disseminação de repositórios educacionais de objetos de aprendizagem no contexto educacional propicia o surgimento de um volume expressivo de recursos

pedagógicos, que contribuirá para avanços significativos nas práticas educacionais. Isso possibilita ao professor organizar, avaliar quais recursos pedagógicos (objetos de aprendizagem) serão utilizados e qual a melhor estratégia. Ao expor todo esse material disponível em um repositório digital, o professor estará colaborando e cooperando com seus pares. Dessa forma, é imprescindível um planejamento cuidadoso, com ênfase na confiança e melhoria na qualidade e utilidade dos recursos.

Apesar dos repositórios educacionais digitais serem considerados tecnologias inovadoras, são muitas as iniciativas de acadêmicos para a construção de tais ferramentas. Porém, existe a dificuldade quanto à licença e restrição de autores, no que diz respeito à publicação e a reutilização de recursos educacionais. Ressalta-se ainda o fato de que a implementação de um repositório está ligada à criação de padrão de metadados que vai garantir a busca e a recuperação de recursos no repositório.

2.5 Etiquetagem e folksonomia

O processo de etiquetagem no mundo real é amplamente utilizado como uma maneira de rotular, identificar, localizar produtos, serviços e conteúdos, proporcionando desenvolvimento econômico e social entre os mais diversos países e suas respectivas culturas. Diante desse amplo uso, o modelo de etiquetagem foi levado para o mundo virtual, onde a internet através da Web 2.0 apropriou-se para localizar, classificar e organizar itens de interesses do usuário. Seguindo essa tendência, o presente estudo faz uso de etiquetas para localização de informações relacionadas à temática de interesse dos usuários, propiciando a aprendizagem colaborativa, cooperativa e a construção do conhecimento de modo rápido, com o propósito de motivar o pesquisador em rede.

Etiquetagem (em inglês, *tagging*) é uma forma de indexação em que as próprias pessoas, no caso, os usuários da informação, classificam os documentos. A marcação de conteúdo com termos descritivos, também chamados palavras-chave ou etiquetas (em inglês *tagging*), é uma forma comum de organizar conteúdo para futura navegação, filtragem ou busca. Embora essa forma de organizar conteúdo eletrônico não seja algo novo, uma forma colaborativa desse processo que foi

chamada de “etiquetagem”, está ganhando popularidade na web (GOLDER e HUBERMAN, 2006, p. 198, tradução nossa). Sendo assim, repositórios de documentos ou bibliotecas digitais muitas vezes permitem que documentos e suas coleções sejam organizados pelas palavras-chaves.

Nos AVAs, os usuários anotam estas palavras-chaves nas distintas ferramentas de comunicação, como os fóruns, chat, diário de bordo, etc. Essas anotações podem ser utilizadas pelo SR de modo a disponibilizar OA aos usuários, ou simplesmente para classificar conteúdos. Esse processo de etiquetagem no mundo virtual é denominado de folksonomia, e é largamente utilizada em redes sociais, haja vista o *twitter*, *facebook*, *Google+*, entre outras. Normalmente, são termos acompanhados de algum “caractere coringa”, por exemplo, o “#”, ou “jogo da velha”.

Nesse sentido, etiquetar é uma das ações relacionadas à classificação mais incorporada ao dia a dia das pessoas. Está nos rótulos dos produtos, nos alimentos e em uma variedade de objetos de uso pessoal ou coletivo, e contém informações importantes sobre determinado objeto. Na “sociedade 2.0”, etiquetar configura ação de disseminação da informação, utilizando-se da folksonomia para a atribuição de termos para representação do conhecimento.

Folksonomia é a tradução do termo *folksonomy*, um neologismo criado em 2004 por Thomas Vander Wal, a partir da junção de *folk* (povo, pessoas) com *taxonomy*. Para Wal (2006), Folksonomia é o resultado da atribuição livre e pessoal de etiquetas (*tagging*) a informações ou objetos, visando à sua recuperação. A atribuição de etiquetas é feita num ambiente social.

As Folksonomias podem ser encontradas em diversos tipos de sites web que permitem o armazenamento de recursos por parte de seus usuários, tais como o armazenamento de imagens, de vídeos, documentos, músicas e links (URLs) e são também chamados de “tagging systems”. O quadro 2.4 apresenta a relação de alguns sistemas que utilizam *tags* para identificar os recursos armazenados por seus usuários.

Quadro 2.4 - Sistemas que utilizam folksonomia.

SITE	TIPO DE RECURSO	ENDEREÇO
Amazon	Loja virtual	http://www.amazon.com/
Bibsonomy	Informações bibliográficas	http://www.bibsonomy.org
CiteULike	Links acadêmicos	http://www.citeulike.org
Connotea	Informações bibliográficas	http://connotea.org/
Delicious	Links favoritos	http://delicious.com
Diigo	Links	http://www.diigo.com/
Flickr	Fotos	http://www.flickr.com
Scribd	Compartilhamento de arquivos em diversos formatos (ppt, pdf, doc, txt, etc.)	http://www.scribd.com/
SlideShare	Arquivos em formato ppt	http://www.slideshare.net
Twitter	Microblogging	https://twitter.com/
Wikipédia	Enciclopédia livre	http://pt.wikipedia.org/wiki/P%C3%A1gina_principal
YouTube	Vídeos	http://www.youtube.com

Fonte: Elaborado pelo autor (2013).

Portanto, Folksonomia é o resultado da etiquetagem dos recursos da Web num ambiente social (compartilhado e aberto a outros) pelos próprios usuários da informação visando a sua recuperação. Destacam-se, portanto, três fatores essenciais: 1) é resultado de uma indexação livre do próprio usuário do recurso; 2) objetiva a recuperação *a posteriori* da informação, e 3) é desenvolvida num ambiente aberto que possibilita o compartilhamento e, até, em alguns casos, a sua construção conjunta.

A iniciar-se pela definição de Wal (2006), o criador do termo, que entende a folksonomia como o resultado da etiquetagem de recursos digitais da Web, ou seja, um produto que existe em função de uma ação, a de etiquetar. Já em Lund (2005), a folksonomia se refere a um vocabulário, ou lista de termos, que surge da sobreposição de etiquetas definidas por vários usuários ao marcar seus *links* favoritos, ou marcadores para posterior recuperação. Para este autor, portanto, o produto seria uma lista de termos, ou vocabulário. Na visão de Mathes (2004), a folksonomia é um conjunto de termos que um grupo de usuários utilizou para etiquetar os conteúdos de recursos digitais da Web. Trant (2006 apud SMITH, 2006) afirma que é o resultado de um sistema de classificação socialmente construído, ou,

coleção de conceitos expressos em um sistema de classificação desenvolvido de forma cooperativa (TRANT, 2006). Ou ainda, um conjunto informal e orgânico de terminologia relacionada (TRANT, 2006).

Segundo Zhou et al (2008) e Pereira e da Silva (2008), a utilização da técnica de folksonomia pode esbarrar em alguns problemas. Os autores afirmam que pelo fato de as *tags* serem geradas pelos próprios usuários (e não baseadas no conteúdo do documento, como geralmente acontece com os termos de índice), a diferença entre o conhecimento dos usuários deveria ser considerada no momento da categorização. Alguns usuários podem se diferenciar em seu conhecimento sobre determinado assunto e, consequentemente, isso afeta a qualidade de suas categorizações sobre os documentos (ROSCHE, 1978).

Schenkel et al (2008) apresentam uma revisão sobre trabalhos na área de *tagging/folksonomia* e discutem algumas dimensões sobre a técnica de folksonomia, que podem ser levadas em consideração ao se projetar um sistema baseado nessa técnica. Os autores enfatizam que a indexação manual, conquanto bastante popular, não representa de fato uma inovação. No entanto, eles apontam como vantagem da utilização dessa técnica a facilidade e a simplicidade de sua utilização, fazendo com que “um esquema tradicional de organização de informação pareça uma técnica da idade da pedra, efetivo, porém muito desconfortável”.

De forma geral, os autores citados têm interpretado a folksonomia não apenas como uma lista de termos gerados e compartilhados livremente, mas algo como uma abordagem, uma metodologia, um sistema de classificação ou até um novo paradigma de classificação.

É possível concordar com Quintarelli (2005), quando o autor assinala que as folksonomias podem fazer contribuições aos estudos da organização, em Ciência da Informação, devido ao fato de surgirem como um meio de autoexpressão de um grupo e, num contexto mais geral, sugerir possibilidades de agregação, de análise e de funcionamento de comunidades que podem acelerar o compartilhamento e a especificação de linguagens de referência.

A folksonomia surgiu como uma alternativa inovadora para estabelecer uma organização às inúmeras e diversas informações publicadas na web de forma vertiginosa e, ainda que limitada sob diversos aspectos, tornou-se prática corrente em diversos *sites*, exibindo uma performance satisfatória para a classificação de conteúdos disponíveis em ambientes virtuais colaborativos. Porém, muitos são os

aspectos que envolvem a prática da folksonomia e que devem ser analisados atentamente.

Pesquisar novas formas de classificação de conteúdos nos ambientes virtuais de aprendizagem, onde os usuários podem organizar informações de uma forma simples e intuitiva, permitindo que sejam encontradas rapidamente e independente do suporte em que estão armazenadas, torna-se tarefa instigante e necessária para melhor compreensão das transformações advindas dos conteúdos postados nos AVAs e suas implicações no campo da Ciência da Informação (CI).

Aplicado ao contexto do potencial educativo da web 2.0, permite aos alunos etiquetar conteúdos. Este é um mecanismo de análise que possibilita a identificação de temas locais ou globais, e o filtro de informação conforme esses temas, e de acordo com o interesse do professor e do aluno.

A utilização de etiquetas para a classificação de conteúdos tem-se alargado naquilo que se denomina de “nuvem de hashtags” (ou *tag cloud*), construída a partir de “informações sobre a frequência com que as hashtags são usadas em particular” (ANDERSON, 2007).

Nesse sentido, usaremos nessa pesquisa as hashtags criadas, pelos usuários, nos fóruns do AVA MOODLE, para classificar, ranquear e sugerir conteúdos relacionados ao tema em estudo.

2.6 Educação a distância

São diversos os conceitos de Educação a Distância (EaD), sendo a maioria de caráter descritivo com base na educação. Para Moran et al (2006), “é o processo de ensino-aprendizagem, mediado por tecnologias, onde professores e alunos estão separados espacial e/ou temporalmente”.

Segundo Maia; Mattar (2007, p. 6) “[...] a EaD é uma modalidade de educação em que professores e alunos estão separados, planejada por instituições e que utiliza diversas tecnologias de comunicação”.

A educação a distância está se expandindo cada vez mais, por ser uma modalidade de ensino que busca atender as novas demandas educacionais, decorrentes das mudanças na nova ordem econômica mundial, que vêm acontecendo em ritmo acelerado, sendo visíveis no crescente avanço das

tecnologias de comunicação e informação, tendo como consequências mudanças no campo educacional (BELLONI, 2008).

Filatro (2007) chama EaD o ensino que:

Supõe a separação espacial e temporal entre professor e aluno. A maior parte da comunicação entre professor e aluno é indireta, mediada por recursos tecnológicos, mas não depende exclusivamente da comunicação on-line (ver, por exemplo, tradição do ensino por correspondência baseado apenas em mídia impressa) (FILATRO, 2007, p. 48).

Por consequência desse fator, o tradicional ensino por correspondência deu origem à modalidade de ensino a distância. Carneiro (2009) diz que “a implantação dos sistemas de correios foi que propiciou o surgimento dos cursos por correspondência, permitindo a efetiva interação dos alunos/estudantes e professores entre si” (CARNEIRO, 2009, p. 36).

Conforme a Lei de Diretrizes e Bases da Educação – LDB, Decreto nº 5.622, de 19 de dezembro de 2005, a Educação a Distância:

Caracteriza-se como modalidade educacional na qual a mediação didático-pedagógica nos processos de ensino e aprendizagem ocorre com a utilização de meios e tecnologias de informação e comunicação, com estudantes e professores desenvolvendo atividades educativas em lugares ou tempos diversos (BRASIL, 2008).

Para Silva (2005, p. 34), o sistema de EaD encontra-se em um período de mudanças, principalmente na forma de organizar os recursos de acordo com a tecnologia, “o que prevalecia historicamente nas instituições de EaD, torna-se insustentável num ambiente com múltiplas tecnologias”. Esta alteração passa pelo conceito de tecnologia da TIC e o uso de OA como forma de criar, organizar e reutilizar materiais pedagógicos.

No Brasil, esse sistema está em fase de transição, isto é, começando a ser adotado na educação. O governo Federal tem incentivado o EaD, fazendo uso de programas como: Universidade Aberta do Brasil (UAB), Escola Técnica Aberta do Brasil (E-Tec) e ainda o Plano Nacional de Formação Continuada de Professores da Educação Básica (Plataforma Freire).

Em 2005, o Ministério da Educação criou e regulamentou o artigo 80 da LDB, determinando os procedimentos que devem ser adotados pelas instituições para obter o credenciamento do MEC para a oferta de cursos a distância, podendo as

universidades, faculdades e os centros tecnológicos oferecer até 20% da carga horária total de qualquer um de seus cursos presenciais na modalidade a distância, desde que o referido curso seja reconhecido pelo MEC.

Como forma de ampliar o acesso e diversificar a oferta do ensino superior no Brasil, o MEC criou em 2005 a Universidade Aberta do Brasil (UAB), através da metodologia EaD, provendo acesso à formação especializada para camadas da população que são excluídas do processo educacional.

Em 2007, foi criado, por meio da modalidade de educação a distância, o E-TEC – Escola Técnica Aberta do Brasil, provendo o acesso ao ensino técnico público e incentivando os jovens a concluírem o Ensino Médio. Dessa forma, objetivando levar cursos técnicos para regiões distantes das instituições de ensino e para a periferia das grandes cidades brasileiras.

A eficácia dos cursos oferecidos pelas instituições de ensino que atuam em EaD depende de uma proposta pedagógica bem elaborada e de uma estrutura administrativa e tecnológica que mantenha o suporte necessário para o atendimento ao aluno.

A utilização das ferramentas da web 2.0 como recurso no processo de ensino e aprendizagem vem se consolidando através dos AVAs, que são plataformas de criação e gerenciamento de cursos a distância. Moita e Silva (2006, p. 5) afirmam que “da mudança no processo educacional, emergem práticas de aprendizagem em ambientes virtuais”. A adoção destes AVAs depende da proposta pedagógica e da metodologia aplicada pelas Instituições de Ensino. Na seção a seguir, serão apresentados alguns AVAs e suas principais ferramentas.

2.7 Ambientes virtuais de aprendizagem

Para Vavassori e Raabe (2003), os ambientes virtuais de aprendizagem (AVAs) “(...) é uma plataforma de ensino a distância que contém muitas ferramentas e recursos, permitindo o gerenciamento de cursos on-line e potencializa a utilização de atividades através da Internet”.

Os AVAs são sistemas computacionais disponíveis na internet, destinados ao suporte de atividades mediadas pelas TIC, que permitem integrar múltiplas mídias, ordenar informações, desenvolver interações entre pessoas e objetos de

conhecimento. Tornar o AVA um ambiente mais humanizado deve ser uma meta constante em todo projeto de EaD, seja através do uso das mídias sociais, de outras ferramentas da Internet ou mesmo de qualquer estratégia que esteja ao alcance dos tutores e dos professores no ambiente virtual (COELHO, 2012).

Os AVAs reúnem uma grande quantidade de ferramentas como: fóruns, chat, e-mail, grupo de discussão, estando estruturado a partir de uma plataforma de interação, permitindo gerenciar, desenvolver e acompanhar conteúdos a distância, possibilitando aos administradores, professores e criadores de cursos o uso de ferramentas pedagógicas de geração e disponibilização de conteúdos aos alunos, de forma a potencializar a criação de cursos nessa modalidade.

Atualmente os AVAs oferecem ferramentas de comunicação síncronas e assíncronas. A comunicação síncrona pode ser entendida como forma de comunicação *on-line*, ou seja, comunicação feita de modo instantâneo; já a comunicação assíncrona pode ser entendida como a forma de comunicação *off-line*, ou seja, forma de comunicação desconectada do tempo e do espaço. Podemos citar como exemplo de comunicação assíncrona os fóruns, chats, glossários, questionários, blogs e wikis.

Os AVAs podem ser empregados como suporte para sistemas de EaD realizados exclusivamente *on-line*, para apoio às atividades presenciais de sala de aula, permitindo expandir as interações da aula para além do espaço-tempo, do encontro face a face, ou para suporte a atividades de formação semipresencial nas quais o ambiente virtual poderá ser utilizado tanto nas ações presenciais como nas atividades a distância.

Existem vários AVAs que podem ser utilizados pelas Instituições de Ensino em seus cursos na modalidade a distância, alguns são softwares livres, o que possibilita sua customização e acréscimo de novas funcionalidades a outros proprietários.

Nesse contexto, serão abordados alguns ambientes de aprendizagem, atualmente utilizados por várias instituições de ensino, identificando as suas principais funcionalidades. Os ambientes a serem apresentados são: *TelEduc*, *Blackboard* e *MOODLE*.

2.7.1 TelEduc

De acordo com o site oficial, o TelEduc é um ambiente para a criação, participação e administração de cursos na Web. Ele foi concebido tendo como alvo o processo de formação de professores para informática educativa, baseado na metodologia de formação contextualizada, desenvolvida por pesquisadores do Núcleo de Informática Aplicada à Educação – Nied, da Unicamp. O TelEduc foi desenvolvido de forma participativa, ou seja, todas as suas ferramentas foram idealizadas, projetadas e depuradas segundo necessidades relatadas por seus usuários. Com isso, ele apresenta características que o diferenciam dos demais ambientes para educação a distância disponíveis no mercado, como a facilidade de uso por pessoas não especialistas em computação, a flexibilidade quanto a como usá-lo, e um conjunto enxuto de funcionalidades.

O TelEduc é um *software* livre; você pode redistribuí-lo e/ou modificá-lo sob os termos da *General Public License* – GNU, versão 2, como publicada pela *Free Software Foundation*, e foi concebido tendo como elemento central a ferramenta que disponibiliza Atividades. Isso possibilita a ação onde o aprendizado de conceitos em qualquer domínio do conhecimento é feito a partir da resolução de problemas, com o subsídio de diferentes materiais didáticos como textos, softwares, referências na internet, dentre outros, que podem ser colocados para o aluno usando ferramentas como: material de apoio, leituras, perguntas frequentes, etc.

Esse ambiente possui as seguintes ferramentas:

- Agenda: página onde traz a programação do curso.
- Avaliações: local onde há a lista de avaliações a serem realizadas pelos alunos.
- Material de apoio: local onde apresenta informações úteis sobre o curso.
- Leituras: local onde os artigos relacionados ao curso são armazenados. Nesse local podem ser armazenadas sugestões de revistas, jornais, endereços da web, etc.
- Perguntas frequentes: local onde possui resposta para as perguntas sobre o curso que são frequentemente questionadas.
- Enquete: ferramenta para criação de enquetes.

- Parada obrigatória: local no qual há materiais que visam desencadear reflexões e discussões entre os alunos do curso.
- Mural: local onde os alunos podem deixar mensagens relacionadas ao curso.
- Fóruns de discussão: local que permite que tópicos sejam discutidos pelos alunos.
- Bate-papo: ferramenta que permite a conversa em tempo real entre os usuários.
- Correio: ferramenta de correio eletrônico interno ao ambiente.
- Grupos: local que permite a criação de grupos de pessoas para facilitar a distribuição de tarefas.
- Perfil: local onde cada participante pode falar de si mesmo.
- Portfólio: local onde os participantes podem armazenar textos e arquivos utilizados no curso. Trata-se de um local particular ao usuário, que pode ser compartilhado com os formadores ou participantes do curso.
- Acessos: local onde é permitido verificar a frequência dos usuários do curso.
- Busca: local que permite a realização de busca por todas as ferramentas do ambiente.



Figura 2.7: Autenticação do TelEduc.
Fonte: <http://www.teleduc.org.br/>. Acesso em: 03 de julho de 2013.

2.7.2 Blackboard

A plataforma Blackboard é um *software* proprietário, desenvolvido pela Blackboard Inc, um provedor de *softwares* e serviços para educação *on-line*. O Blackboard é um Learning Management System – LMS, com funcionalidades de instrução e comunicação, bastante utilizado por instituições de ensino privadas no Brasil, a exemplo da Universidade Católica de Brasília e Instituto de Educação Superior de Brasília – IESB.

O Blackboard é uma ferramenta avançada que enriquece o processo de ensino-aprendizagem com o uso das tecnologias de informação e comunicação. No Blackboard, professores e alunos participam de um ambiente interativo que permite novas possibilidades de planejamento, execução e avaliação das atividades acadêmicas.

Por meio dessa plataforma de aprendizagem, é possível que professores e alunos interajam e compartilhem várias mídias, como textos, imagens, filmes entre outras. Nesse ambiente são realizadas atividades de comunicação síncrona e assíncrona, criando condições para a formação de comunidades de aprendizagem colaborativa.

O ambiente permite diversas formas de avaliação, autoavaliações, atividades individuais e em grupo. Com o Blackboard, os alunos a distância têm flexibilidade de tempo e de local para estudar. O acompanhamento da turma, pelo docente, é facilitado pela plataforma, que gera relatórios e estatísticas sobre o desempenho acadêmico dos alunos.

Com a utilização do Blackboard, o conhecimento produzido para os cursos é sistematizado e potencializado, resultando num padrão elevado de ensino-aprendizagem, sendo utilizado nos cursos a distância e também como ferramenta de apoio para os alunos do ensino presencial.



Figura 2.8: Página inicial do Blackboard.

Fonte: <http://www.blackboard.com/>, acessado em 03 de julho de 2013.

2.7.3 MOODLE

O Ambiente Virtual de Aprendizagem – AVA MOODLE ou Virtual Learning Environment – VLE, também conhecido por Sistema de Gerenciamento de Cursos (Course Management System ou CMS) ou Sistema de Gerenciamento de Aprendizagem (Learning Management System ou LMS), tornou-se popular entre os educadores de todo o mundo, sendo bastante utilizado pelas instituições de ensino em seus cursos de educação a distância, por oferecer diversos recursos, onde cursos podem ser cadastrados, gerenciados e utilizados por seus usuários.

Foi desenvolvido pelo australiano Martin Dougiamas, objetivando ofertar cursos não presenciais auxiliados pela rede mundial de computadores. Em virtude da facilidade de manuseio e controle das ações desenvolvidas pelos seus usuários, o MOODLE vem sendo utilizado por inúmeros países, sobretudo, os subdesenvolvidos. Tal fato ganha mais importância em função de ser um ambiente gratuito e seu código está disponível (*opensource*). Ademais, é baseado na Web e foi desenvolvido fundamentado em princípios pedagógicos, possibilitando a participação de uma grande comunidade de usuários que contribuem para a melhoria do ambiente, além de ser acessível a diversos idiomas.

O AVA MOODLE tem diferentes funcionalidades, entre elas podemos destacar: acesso móvel, alimentador RSS remoto, busca Google, repositório de objeto de aprendizagem móvel, servidor de rede, usuários móveis e *on-line*, livro, criação de página Web, pacote IMS CP, tags, tags do blog, entre outros. Sendo

assim, o AVA MOODLE disponibiliza funcionalidade de forma interativa, dinamizando o processo de aprendizagem. A cada nova versão, novas atividades são incorporadas, como: base de dados, chat, escolha, *flash-trainer*, fórum, glossário, lição, *móBILE-tags*, objeto de aprendizagem móvel, pesquisa de avaliação, questionário, Scorm/ACC, tarefas e Wiki.

A utilização do MOODLE como Ambiente Virtual de Aprendizagem vem crescendo rapidamente. Segundo dados extraídos do site oficial do ambiente, até julho de 2013, 237 países e 84,505 sites usavam a plataforma.

Quadro 2.5 – Estatística MOODLE.

Indicador	Quantitativo
Sites registrados	84,505
Número de países	237
Número de cursos	7,649,547
Número de usuários	71,829,870
Número de professores	1,294,379
Postagens no fórum	127,201,733
Recursos utilizados	68,144,978
Questões do quiz	188,516,909

Fonte: <http://MOODLE.org/stats/>, acessado em 23 de julho de 2013.

Esses dados confirmam a preferência pelo uso do MOODLE, *software* de distribuição gratuita, constantemente atualizado pelos colaboradores da comunidade, aumentando a quantidade de módulos disponíveis para a implementação de diversas funcionalidades, assim como melhorando módulos existentes.

O ambiente MOODLE é atualmente utilizado pela Universidade Estadual do Maranhão em seus cursos de Educação a Distância. Os cursos são cadastrados no ambiente virtual, com as suas respectivas disciplinas, sob o controle do professor e dos tutores. Os alunos acessam as disciplinas e realizam as atividades planejadas pelo professor da disciplina, sob acompanhamento dos tutores, que auxiliam os alunos nos fóruns de discussão, chats e trocas de mensagens.

Figura 2.9: Ambiente Virtual de Aprendizagem do Curso de Engenharia da Computação.
Fonte: <http://www.engcomp.uema.br/MOODLE/login/index.php>. Acesso: 09 de julho de 2013.

Desde o início da pesquisa até agora, o Brasil passou do quarto para o terceiro lugar na lista dos países com maior número de sites registrados, conforme podemos observar na figura a baixo, extraída e adaptada do site oficial do MOODLE (<http://MOODLE.org/stats/>).



Figura 2.10: Dez países por registro.
Fonte: <http://MOODLE.org/stats/>, acessado em 23 de julho de 2013.

Para Ribeiro (2010), o número de adeptos da plataforma MOODLE no Brasil cresceu em função do Sistema de Universidade Aberta do Brasil (UAB), do Projeto da Escola Técnica Aberta do Brasil (E-Tec) e, ainda, do Plano Nacional de Formação Continuada de Professores da Educação Básica (Plataforma Freire).

O AVA MOODLE reúne uma grande quantidade de objetos de aprendizagem, permitindo gerenciar, desenvolver e acompanhar conteúdos a distância, possibilitando aos administradores, professores, tutores e criadores de cursos o uso de ferramentas pedagógicas de geração e disponibilização de conteúdos aos alunos, de forma a potencializar a criação de cursos nessa modalidade. Nesse sentido, promove a interação entre professor-aluno, aluno-aluno, aluno-tutor, aluno-conteúdo.

Apesar de o MOODLE auxiliar no processo de ensino/aprendizagem, apresenta algumas dificuldades, porquanto é um *software* que limita professores e alunos ao armazenamento, por vezes estático, de informações de ensino, tornando o processo de aprendizagem passivo, sem que exista uma interação *on-line* ativa entre docentes e discentes (CORRADI, 2009). Objetivando melhorias e ampliação dos recursos do MOODLE, a fim de auxiliar as estratégias de ensino, com novas formas de comunicação, exercício da interatividade na aprendizagem, trabalho cooperativo e na construção do conhecimento, esta pesquisa propõe a criação de um módulo recomendador de OA para o AVA MOODLE, sugerindo aos alunos conteúdos relevantes baseados nas hashtags postadas nos fóruns da plataforma.

2.8 Considerações do capítulo

Neste capítulo foram apresentados os conceitos teóricos que formam a base do trabalho realizado. Foi apresentada a revisão teórica sobre sistema de recomendação, sistema de recuperação de informação, repositórios digitais, objetos de aprendizagem, etiquetagem, folksonomia e ambientes virtuais de aprendizagem.

3 TRABALHOS CORRELATOS

No capítulo 2 foram apresentados os conceitos que fundamentam o núcleo teórico desta pesquisa. Agora, serão apresentados alguns trabalhos que fazem uso de sistema de recomendação inseridos em ambientes virtuais de aprendizagem e, nesse contexto, terão como ponto de interseção SR para ambientes virtuais de aprendizagem.

Conforme o capítulo 1, serão apresentadas as atividades e produtos referentes à Etapa **A Teórica e Atividade 2 – Trabalhos correlatos**.

Esta seção apresenta trabalhos correlacionados ao modelo do sistema proposto. Ao final desta seção, encontra-se uma comparação entre os trabalhos explanados.

3.1 Sistema de Recomendação de Objetos de Aprendizagem

O sistema foi desenvolvido para recomendar objetos de aprendizagem no para o ambiente virtual de aprendizagem Sophia, projetado pela UNIVALI na forma de um bloco para o MOODLE. O bloco LorSys (Learning Object Recommender System) tem aplicação individual para cada disciplina em que está inserido e gera recomendações baseando-se nas informações dos usuários e objetos de aprendizagem da disciplina (FERREIRA, 2009).

O sistema adquire as informações através de duas funções: obter e armazenar as informações dos metadados no padrão Learning Object Metadata – LOM, permitindo especificar o esquema conceitual de dados para a estrutura do metadado de um objeto de aprendizagem no repositório de objetos de aprendizagem do ambiente Sophia; bem como armazenar as informações sobre a utilização destes objetos de aprendizagem.

As informações contidas nos metadados dos objetos de aprendizagem não são extraídas e armazenadas automaticamente pelo MOODLE. Para realizar essa atividade, foi desenvolvida uma rotina no LorSys, que extrai os metadados dos objetos de aprendizagem e os armazena na base de dados do MOODLE (FERREIRA, 2009).

O módulo de recomendação tem a função de acessar os dados no banco de dados, utiliza a técnica de recomendação híbrida para recomendar objetos de aprendizagem. O ambiente Sophia registra a navegação do usuário, permitindo ao módulo identificar similaridade entre usuários e os objetos de aprendizagem.

Ferreira (2009), realizou-se um experimento com alunos de um curso de Ciências da Computação o qual permitiu verificar a adequação das funcionalidades desenvolvidas e a pertinência das recomendações. Daí, foi criada uma disciplina denominada “Revisão Computação” onde foram inseridos 28 objetos de aprendizagem, para realização dos testes, A seguir, 4 alunos foram convidados para interagir com os objetos de aprendizagem disponibilizados na disciplina pelo período de 3 dias. A avaliação foi realizada de duas formas: quantitativa e qualitativa.

Através de testes realizado com a rotina desenvolvida para o LorSys, Ferreira (2009), verificou o pleno funcionamento do bloco, pois não foi identificado nenhum erro pelos usuários. Esta conclusão torna possível a utilização do sistema em larga escala a fim de avaliar o seu caráter pedagógico. A técnica de recomendação utilizadas pelo LorSys mostram-se eficientes, apresentando recomendações personalizadas de caráter diversificado aos usuários.

3.2 Modelo de sistema de recomendação de materiais didáticos para ambientes virtuais de aprendizagem

Ferro (2010), propõe um modelo de sistema de recomendação de materiais didáticos para ambientes virtuais de aprendizagem tendo com base o perfil do aluno. Trabalha com a seguinte hipótese: a aplicação de um Sistema de Recomendação em um Ambiente Virtual de Aprendizagem pode sugerir recomendações de materiais didáticos com características que satisfaçam aos interesses de estudos dos alunos.

O modelo foi implementado na forma de um módulo do ambiente MOODLE, configurado e publicado na internet no domínio www.ambientemoodle.com.br, com a finalidade de observar e analisar as recomendações que foram sugeridas aos alunos.

Neste trabalho, adotou-se a filtragem híbrida como técnica de recomendação, fazendo uso de três abordagens diferentes, na tentativa de propor sugestões de materiais didáticos com conteúdos de interesse dos alunos.

A arquitetura do Sistema de Recomendação proposta compõe-se de cinco componentes: Perfil do Usuário, Perfil do Material Didático, Gerador de Recomendações, Avaliação/Feedback e Monitoramento de Aquisições. Cada componente possuirá papel específico, de forma que a atuação em conjunto desses componentes proporcionará aos usuários uma visualização de recomendações de materiais didáticos (FERRO, 2010).

A ferramenta foi desenvolvida na linguagem de programação PHP e o Sistema Gerenciador de Banco de Dados (SGBD) MySQL, ambos utilizados também pelo ambiente MOODLE. O modelo foi implementado sob a versão 1.9.9 do ambiente MOODLE, na forma de um bloco, podendo ser reutilizado e aperfeiçoado por qualquer instituição de ensino.

Para experimentos, foi criado um curso de extensão, disponibilizado para alunos de perfis heterogêneos, ou seja, estudantes de várias áreas de conhecimento. Esse curso, denominado Formatação de Trabalhos Científicos, teve como público alguns alunos concluentes de cursos de graduação e pós-graduação, oferecido na modalidade a distância, com carga horária de 20 horas.

Todas as recomendações criadas foram armazenadas no banco de dados, independentemente de serem avaliadas ou não pelos alunos. Assim, foi possível efetuar análises sobre esses dados, e, particularmente, verificar como cada abordagem de recomendação se comportou durante o experimento.

3.3 Modelo de sistema de alertas e recomendação para mediar aprendizagem em turmas heterogêneas nos cursos de EaD

A pesquisa realizada por Morais (2011), apresenta uma análise sobre a questão da interação dos alunos em turmas heterogêneas em ambientes virtuais de aprendizagem, focalizando as derivações em ações de mediação pedagógica capazes de mostrar as interações nesses espaços. O autor buscou responder ao seguinte problema: Como as mediações pedagógicas, através das técnicas de sistema de recomendação e de sinalizador de alertas, poderiam auxiliar nas

estratégias de ensino e no processo de aprendizagem em turmas heterogêneas em um ambiente de EaD? Assim, o objetivo proposto foi verificar a possibilidade de se desenvolver estratégias para o professor trabalhar com a Zona de Desenvolvimento Proximal dos alunos através do uso de técnicas de recomendação e da aplicação do software Classificação Hierárquica, Implicativa e Coesitiva (CHIC).

A ferramenta leva em consideração o perfil dos alunos, a formação acadêmica, os diferentes níveis de conhecimento, estratégias de ensino e técnicas de recomendações. Para tanto, o material didático foi preparado prevendo atividades que foram disponibilizadas através de Ambiente Virtual de Aprendizagem. O desafio da pesquisa foi observar e analisar o comportamento e perfil dos alunos através das atividades que derivaram ações de mediação com alertas e recomendações, e assim medir seus níveis de participação e conhecimento (MORAIS, 2011).

Os sujeitos da pesquisa foram os alunos, professor e tutor do curso de especialização em Informática na Educação. O ambiente virtual de aprendizagem foi o local onde ocorreu todo o processo de ensino e aprendizagem. Nele, foi apresentado a disciplina a todos os alunos no primeiro encontro presencial, bem como a aplicação de um questionário que ajudou a colher subsídios para a validação de algumas hipóteses. Esse instrumento de coleta foi aproveitado do questionário do sistema de matrícula na forma de um questionário estruturado, com objetivo de obter informações do perfil do aluno. Esses dados foram exportados em XML junto com os relatórios de acesso ao ambiente, fórum, bate-papo, e-mail, atividades e material de apoio, maneira de identificar a presença ativa ou não.

Para avaliar e validar a construção do modelo, o autor aproveitou a disciplina para simular o ambiente de alertas e recomendação, simulando duas etapas: na primeira, os alunos tiveram a oportunidade de familiarizar-se com o material da disciplina, possibilitando-lhes contribuir na construção do material, disponibilizado em WIKI e no fórum discussão sobre o texto, para que ele possa recomendar e contribuir com novos textos. A cada participação do aluno, foram disparadas mensagens de alertas e recomendações enviadas por e-mail, tanto para os alunos como para o professor. Na segunda etapa: “Como trabalhar com alunos de formações diferentes, e como nivelar o seu conhecimento e aproximar sua zona desenvolvimento proximal dentro da disciplina em vigor?” Para isso, o autor fez uso das atividades elaboradas junto ao professor, em que foram construídas três atividades/estágios e aplicadas em momentos distintos.

Os experimentos foram realizados no ambiente TELEDUC e MOODLE. A técnica de recomendação utilizada foi a filtragem colaborativa; o cálculo da similaridade (BADRUL et al, 2001) entre os alunos permite encontrar um conjunto de alunos mais próximos com sua conduta dentro do ambiente. Dessa forma, os históricos dos acessos da primeira turma serviram como base de recomendação para a turma seguinte. Outro ponto relevante foi para os itens não acessados pela análise dos vizinhos mais próximos. Essa abordagem proporcionou entender a conduta e a aprendizagem dos alunos e turmas heterogêneas, em três momentos distintos.

O autor conclui que é muito complicado aplicar uma metodologia que respeite o princípio de um sistema de recomendação sem automatização do processo, mas que, ao mesmo tempo, pudessem surgir efeitos nas ações e reações dos alunos, ainda que não se aderisse a ele completamente, afinal, era preciso discorrer sobre as técnicas de comportamentos similares. O maior problema era resolver a integração entre os dados informativos e históricos de cada ferramenta, pois apresentar um estudo sobre este enfoque exigia um mínimo de contextualização histórica do todo. Entretanto, a diferença seria percebida nos resultados da análise estatística aplicada a cada item.

3.4 Recomendação de mensagens no fórum do MOODLE

A pesquisa realizada por Zamberlan (2011), propõe o desenvolvimento de uma ferramenta de recomendação de mensagens para o fórum do AVA MOODLE, usando técnicas de sistema de recomendação.

Seu principal objetivo é, considerando fóruns de discussão do AVA MOODLE, utilizar técnicas de análise de similaridade entre textos e recomendação para recomendar mensagens que contenham respostas à mensagem que está sendo enviada por um usuário. Dessa forma, a ferramenta responde a dúvida do usuário antes mesmo de ela ser postada em um fórum, diminuindo a quantidade de mensagens e, por conseguinte, o trabalho do tutor, que passa a atender mais alunos de forma qualificada.

Os principais componentes da ferramenta são: os usuários, o fórum e a ferramenta de recomendação.

Quando o usuário envia uma mensagem para o fórum, essa mensagem não é automaticamente inserida nele, ela é primeiramente capturada e analisada. Se houver mensagens semelhantes no fórum, o usuário receberá essas recomendações. Por isso, o usuário pensa que está interagindo diretamente com o fórum, mas a sua interação na verdade é com a ferramenta desenvolvida. Quem interage com o fórum é a ferramenta (ZAMBERLAN, 2011).

A técnica de recomendação utilizada foi a filtragem baseada em conteúdos, bem como a medida de similaridade *Cosine* para determinar o quanto as mensagens estão relacionadas. O trabalho apresenta diferentes técnicas usadas em sistema de recomendação, além de descrever o AVA MOODLE e explicar como o novo módulo pode ser criado para tal ambiente.

Com o intuito de validar a ferramenta proposta, o autor criou um curso de Teste no AVA MOODLE, adicionando o módulo teste. Após a inclusão e configuração do módulo, ele criou ainda algumas categorias para iniciar os testes, as mensagens contidas em cada categoria foram extraídas de outros fóruns da internet e manualmente inseridas no curso de Teste, como tópicos do fórum. Ao total foram inseridos 18 tópicos, a fim de simular 18 perguntas, as quais algumas foram respondidas. Em seguida foi criada uma mensagem que simulou o encaminhamento de uma nova pergunta ao fórum e, após o envio da mensagem, a ferramenta buscou mensagens pertencentes à categoria e respondeu a dúvida do usuário. Depois de visualizar a mensagem, o usuário poderá marcar a opção “Sim” ou “Não” para dizer se a recomendação foi útil ou não.

3.5 Sumário dos trabalhos

O sumário dos trabalhos correlacionados aqui apresentados pode ser observado e comparado com esta pesquisa conforme dados do quadro 3.1. Onde a coluna **Título** apresenta o título do trabalho e a referência analisada, a coluna **Algoritmo** apresenta em seu conteúdo o tipo de algoritmo de recomendação utilizado ou “-“ quando nenhum algoritmo específico foi utilizado, coluna **Medida Estatística**, apresenta a técnica usada para atribuir pesos a documentos utilizado ou “-“ quando nenhuma técnica específica foi utilizada. As demais colunas apresentam “ \odot ” quando o trabalho analisado da ênfase a tal área.

Quadro 3.1: Analise dos trabalhos correlacionados.

Título	Algoritmo	Media estatística	Similaridade	Contexto	Usuário
Sistema de Recomendação de Objetos de Aprendizagem (FERREIRA, 2009).	H	-		◎	◎
Modelo de Sistema de Recomendação de Materiais Didáticos para Ambientes Virtuais de Aprendizagem (FERRO, 2010).	H	-		◎	◎
Modelo de Sistema de Alertas e Recomendação Para Mediar Aprendizagem em Turmas Heterogêneas nos Cursos de EAD (MORAIS, 2011).	FC	-		◎	◎
Recomendação de Mensagens no Fórum do Moodle (ZAMBERLAN, 2011).	BC	-	◎	◎	◎
Recomendação de Documentos para os Usuários do AVA Moodle a partir das Hashtags Postas nos Fóruns (RIBEIRO, 2013).	BC	TF/IDF	◎	◎	◎

Fonte: Elaborado pelo autor (2013).

Através do quadro 3.1 é possível observar que os algoritmos mais utilizados pelos atuais sistemas de recomendação são os baseados em filtragem colaborativa e em conteúdos. O modelo de sistema de recomendação desenvolvido nesta pesquisa faz uso da medida estatística TF/IDF para atribuir pesos a documentos e da calculo de similaridade para classificar os conteúdos a partir dos termos mais postadas nos fóruns da plataforma.

Desta forma, SRD utiliza o algoritmo baseado em conteúdo, para fazer o agrupamento dos termos “hashtags” mais postados, criando vetores de termos e pesos, o qual possibilita a realização do cálculo de similaridade entre as hashtags melhor ranqueadas e os documentos armazenados no repositório do AVA MOODLE.

3.6 Considerações do capítulo

Neste capítulo foram apresentados estudos relacionados ao uso de sistema de recomendação desenvolvidos para ambientes virtuais de aprendizagem. Os trabalhos citados ressaltam as atuais pesquisas e os principais algoritmos utilizados pelos sistemas de recomendação.

No próximo capítulo são apresentados a arquitetura do AVA MOODLE, os requisitos do sistema, a estrutura do diretório moodledata, o núcleo do MOODLE, a base de dados e as principais tabelas utilizadas pelo sistema para fazer as recomendações.

4 ARQUITETURA DO MOODLE

No capítulo 2 foram apresentados os conceitos que fundamentam o núcleo teórico dessa pesquisa. Até o presente instante foi definido que seria desejável desenvolver um modelo contendo SRD baseado no conjunto de hashtags mais postadas nos fóruns do AVA MOODLE. Entretanto, ainda não foram detalhados os requisitos que um sistema dessa natureza deveria atender, ou seja, os problemas associados ao desenvolvimento e à implementação de uma ferramenta com essas características.

Esse capítulo descreve a arquitetura do MOODLE versão 2.2.8+. Tal descrição é importante para o entendimento do módulo de recomendação proposto, que será detalhado no capítulo 5.

O MOODLE é dos ambientes mais populares e utilizados no mundo para gerenciar cursos a distância. No que se refere à integração com outros sistemas, o MOODLE recentemente passou por uma reformulação e a partir da versão 2.0, apresenta a possibilidade de integrar repositórios externos através da implementação de *plugins*. Para isso, é oferecida uma as *Applications Program Interfaces (APIs)*² chamadas *Repository*, a qual possui a classe “*Repository*” com diversos métodos implementados de modo a facilitar a integração. Cada plugin de repositório fica armazenado no diretório padrão /Repository/plugin_name (RODRIGUES, 2012).

O diretório do MOODLE, contém vários arquivos e subdiretórios, que fica disponível em http://o_seu_servidor/MOODLE, ou diretamente em http://o_seu_servidor. Abaixo são listados e descritos os principais arquivos e diretórios:

- config.php - contém as configurações básicas. Este arquivo não vem no MOODLE, mas é criado durante o processo de instalação.
- install.php - o script que deverá executar para criar config.php;
- version.php - define a versão atual do código do MOODLE;
- index.php - a página principal do site;

² Uma API é um conjunto de programas responsável por prover uma estrutura com as finalidades de comunicação e reutilização de software.

- admin/ - pasta contendo código para administração de todo o servidor;
 - auth/ - pasta contendo módulos para autenticação de utilizadores;
 - blocks/ - pasta contendo módulos para os pequenos blocos laterais em várias páginas;
 - calendar/ - pasta contendo todo o código para mostrar e gerenciar calendários;
 - course/ - pasta contendo código para mostrar e gerenciar disciplinas;
 - doc/ - pasta contendo documentação e ajuda do MOODLE;
 - files/ - pasta contendo código para mostrar e gerenciar arquivos enviados;
 - lang/ - pasta contendo com as mensagens de texto em diferentes línguas, um diretório por cada língua;
 - lib/ - pasta contendo librarias do código básico do MOODLE;
 - login/ - pasta contendo código para criação e acesso às contas de utilizadores;
 - mod/ - pasta contendo todos os módulos de disciplina no MOODLE;
 - pix/ - pasta contendo imagens genéricas do sítio;
 - theme/ - pasta contendo com os pacotes temáticos (theme/skins) para mudar a aparência do site;
 - user/ - pasta contendo com o código para mostrar e gerencia a lista de utilizadores da plataforma.

A figura abaixo representa toda a sua estrutura de arquivos e diretórios.

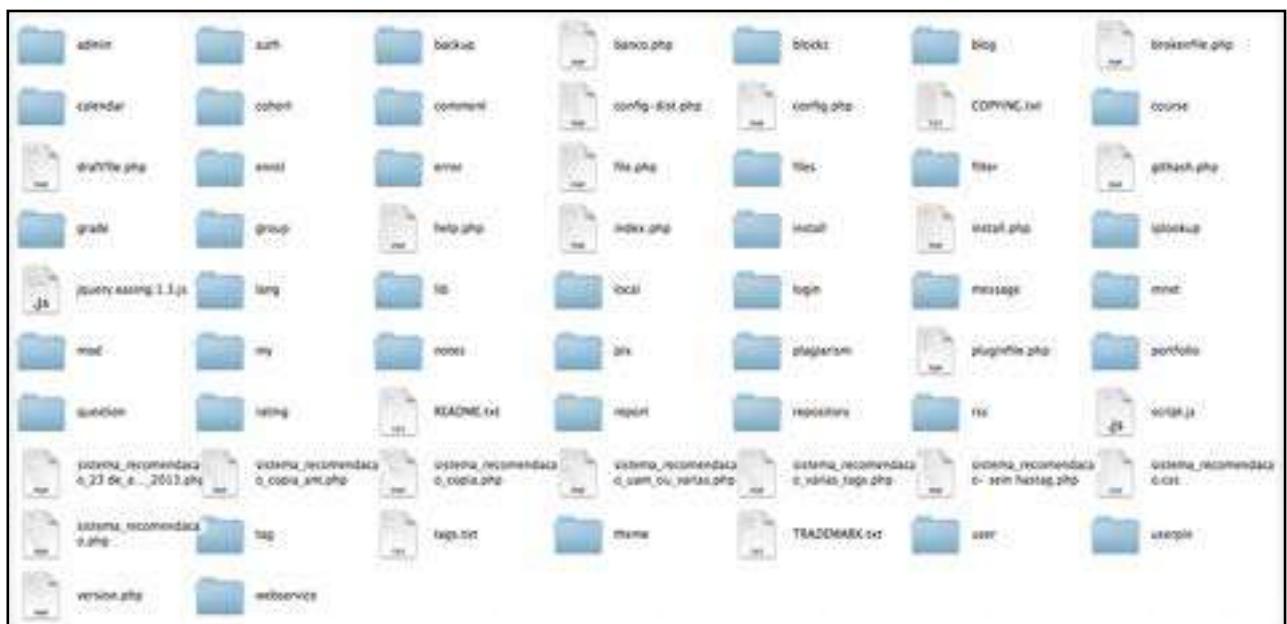


Figura 4.1: Diretório e arquivos do MOODLE.
Fonte: Elaborado pelo autor (2013).

No ponto de vista técnico, o MOODLE é um ambiente aberto, flexível, que pode ser ajustado para atender qualquer demanda pedagógica, pois, possibilita a criação de novos módulos/plugins. Para isso, basta seguir as diretrizes de desenvolvimento, e seu trabalho pode ser facilitado graças a diversidades de APIs disponíveis (MOODLE, 2011).

4.1 Requisitos do Sistema

O MOODLE é um sistema desenvolvido na linguagem PHP executado em um servidor Web, para sua instalação é necessário ter um servidor web que suporte PHP; uma base de dados MySQL ou Postgres ou MSSQL 20 ou Oracle; e um diretório de dados (moodledata) para salvar arquivos enviados pelos usuários da plataforma, conforme figura 4.2.

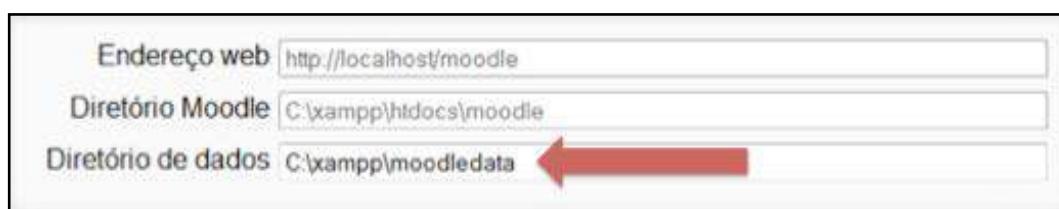


Figura 4.2 – Diretório de dados moodledata.
Fonte: Elaborado pelo autor (2013).

Analizando a figura 4.2 observa-se que o caminho do diretório de dados é C:\xampp\moodledata, porém o mesmo não pode ser acessado por uma aplicação pois, não esta no caminho C:\xampp\htdocs, dessa forma o conteúdo desta pasta não possui permissões de leitura e escrita para o usuário do servidor web.

O MOODLE foi desenvolvido para ser simples de instalar em qualquer servidor que atenda seus requisitos. Para tanto é necessário fazer o download da versão desejada no site oficial do MOODLE (<http://download.moodle.org/>), extrair os arquivos para dentro da pasta MOODLE no diretório do servidor e acessá-la para iniciar a instalação.

4.2 O Núcleo do MOODLE

O núcleo do MOODLE fornece toda a infraestrutura necessária para criar um AVA. Novos *plugins* podem ser implementados baseados nas seguintes funcionalidades:

- Cursos e Atividades;
- Usuários;
- Grupos;
- Matriculas e Controle de Acesso;
- Navegação e Configuração;
- Bibliotecas de Forms;
- Biblioteca javascript;
- Instalação e Atualização;
- Logs e Estatísticas.

Por possuir código aberto muitos programadores desenvolvem novos módulos e os encaminham para aprovação de Martin Dougiamas e sua equipe. Isto funciona como uma empresa de desenvolvimento e controle de qualidade.

4.3 Reppositórios de Arquivos do MOODLE

A Instalação do MOODLE, a partir da versão 2.0, os repositórios ganham um lugar de destaque entre os módulos originais do MOODLE, onde ficam disponíveis *plugins* para os repositórios previamente integrados como Flickr, Google Docs, Dropbox, Sistema de arquivos, Alfresco, Box.net, Flickr público, Merlot.org, Web álbum do Picasa, Amazon S3, Downloader URL, repositório WebDAV, Wikimedia e Videos Youtube, conforme mostra a figura 4.3. Além disso, oferecem opções de gerenciamento de OAs através de um repositório local, funcionalidade anteriormente implementada pelo módulo adicional MrCute. O repositório local do MOODLE permite fazer upload dos recursos para uma área privada ou para uma área comum do servidor que pode ser acessada por todos os cursos e usuários professores ou administradores. Além de armazenar é possível acessar, recuperar, usar e reusar os OAs disponíveis no repositório local.

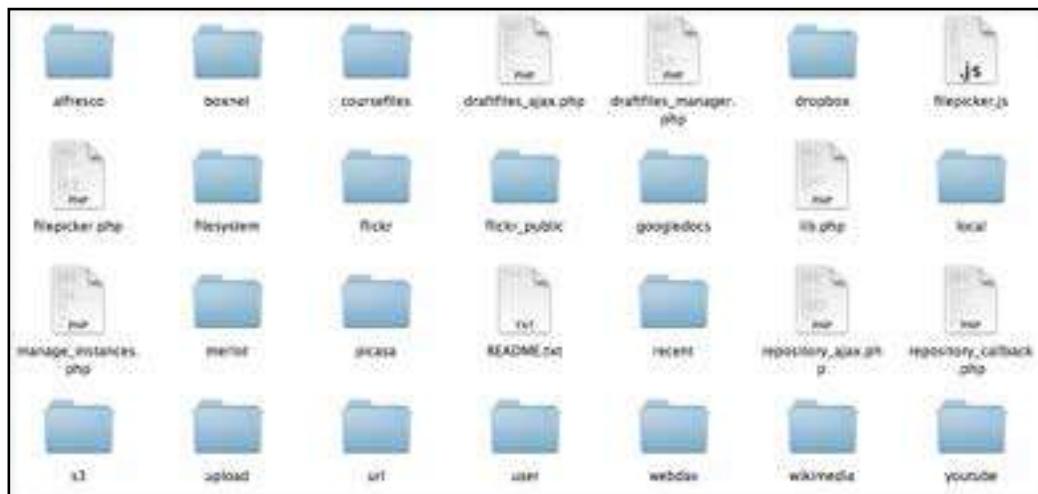


Figura 4.3: Repositório de plugins do moodle.

Fonte: Elaborado pelo autor (2013).

Como administrador do MOODLE é possível gerenciar os repositórios com as opções: Desativar, Habilitado mas Escondido, Habilitado e Visível. Ainda nesse mesmo perfil, é possível realizar configurações referente a cada plugin, como, por exemplo, inserir uma chave de licença para acesso ao repositório Dropbox.



Figura 4.4: Habilitando Repositório Dropbox.

Fonte: Elaborado pelo autor (2013).

Com relação ao repositório local é possível criar instância para o mesmo. Na prática isso é feito através da criação de pastas dentro do servidor web onde os arquivos ficam armazenados, o sistema de recomendação proposto faz o uso de um repositório local do MOODLE ilustrado na figura 4.5, para isso foi habilitado o plugin sistema de arquivos e criado uma instância para a mesma chamada sistema de arquivos mostrada na figura 4.6.

The screenshot shows the 'Administración do site' (Site Administration) section of the Moodle control panel. On the left, there is a sidebar with various menu items like 'Aulas' (Lessons), 'Inscrições' (Enrolments), 'Opções avançadas' (Advanced options), 'Usuários' (Users), 'Cursos' (Courses), 'Notas' (Notes), 'Local' (Local), 'Home', 'Plugins', 'Módulos de atividades' (Activity modules), 'Blocos' (Blocks), 'Saídas de mensagens' (Message outputs), 'Autenticação' (Authentication), 'Inscrições' (Enrolments), 'Editores de texto' (Text editors), 'Licenças' (Licences), 'Efeitos' (Effects), 'Repositórios' (Repositories), and 'Verificar repositórios' (Check repositories). The 'Plugins' and 'Repositórios' sections are highlighted with red boxes.

The main area displays several repository configurations:

- Dropbox:** Status: Habilitado mas escondido (Enabled but hidden). Buttons: Configurações (Configurations), Desativar (Disable).
- Sistema de arquivos:** Status: Habilitado e visível (Enabled and visible). Buttons: Configurações (Configurations), Desativar (Disable). Below it, a note says: 1 instância(s) comum que abrangem o site, 0 instância(s) comum que abrangem o curso, 0 instâncias privativas do usuário.
- Wikimedia:** Status: Habilitado e visível (Enabled and visible). Buttons: Configurações (Configurations), Desativar (Disable).
- Downloader URL:** Status: Habilitado e visível (Enabled and visible). Buttons: Configurações (Configurations), Desativar (Disable).
- Repositório Alfresco:** Status: Desativar (Disable). Buttons: Configurações (Configurations), Desativar (Disable).
- Box.net:** Status: Desativar (Disable). Buttons: Configurações (Configurations), Desativar (Disable).
- Arquivos de curso legados:** Status: Desativar (Disable). Buttons: Configurações (Configurations), Desativar (Disable).

Figura 4.5: Habilitar/Desabilitar Repositório Local.

Fonte: Elaborado pelo autor (2013).

A figura 4.5 mostra as configurações necessárias para habilitar o plugin para o repositório chamado de sistema de arquivos.

The screenshot shows the 'Instâncias de repositórios do site' (Site repository instances) page. It has a table with columns: Nome (Name), Plugins de repositório (Repository plugins), Configurações (Configurations), and Excluir (Delete). There is one entry: 'Sistema de Arquivos'. Below the table is a button labeled 'Cria uma instância de repositório' (Create a repository instance).

Figura 4.6: Instância para pasta do servidor web.

Fonte: Elaborado pelo autor (2013).

A inserção de novos recursos é uma opção disponível apenas para o repositório local do MOODLE e suas instâncias. Os plugins são trazidos para a interface do MOODLE através de uma única aplicação chamada Seletor de Arquivos (File Picker) onde é permitido anexar arquivos, inserir imagem, vídeo, adicionar arquivos; o File Picker (Figura 4.7) é apresentado com todos os repositórios locais ativados por padrão na instalados no MOODLE, que são:

- Arquivos do servidor – possibilita o acesso aos arquivos em outro lugar no MOODLE de acordo com suas permissões;

- Arquivos recentes – mostra os últimos 50 arquivos enviados, de acordo com o contexto;
 - Enviar um arquivo - usado para navegar e fazer o upload de um arquivo do seu computador;
 - Arquivos privados – é a área de armazenamento de arquivos privada do usuário.

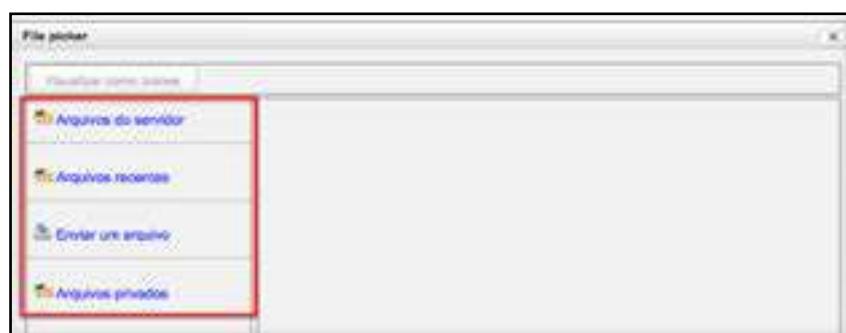


Figura 4.7: Janela do File Picker do MOODLE.
Fonte: Elaborado pelo autor (2013).

Conforme mencionado e ilustrado nas figuras 4.6 e 4.7 o sistema de recomendação proposto fará uso do repositório local do MOODLE chamado sistema de arquivos, mostrado na figura 4.8.

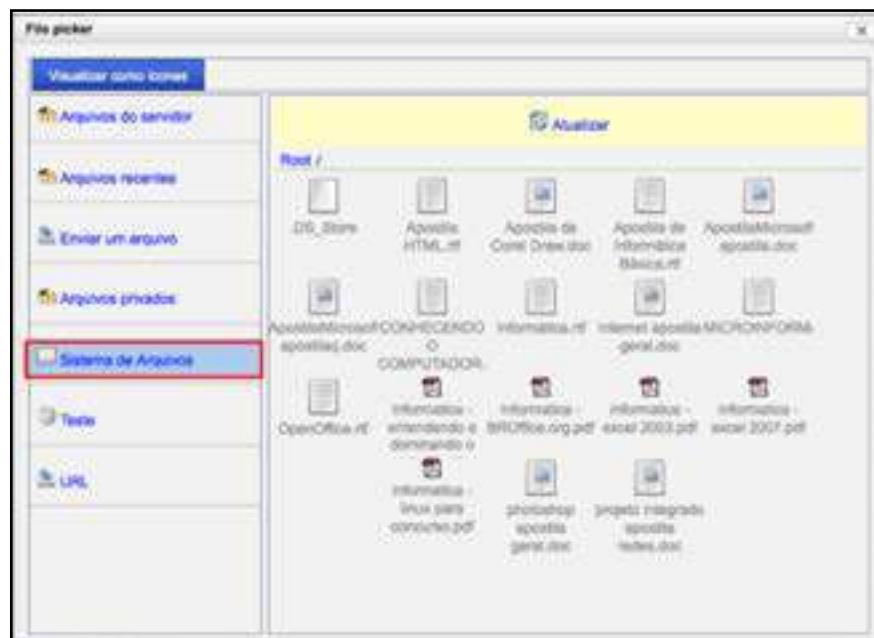


Figura 4.8: Plugin Sistema de Arquivos.
Fonte: Elaborado pelo autor (2013).

A figura 4.9 apresenta uma visão geral do modelo proposto do sistema de recomendação de objetos de documentos para os usuários do AVA MOODLE, o qual será detalhado nas próximas seções.

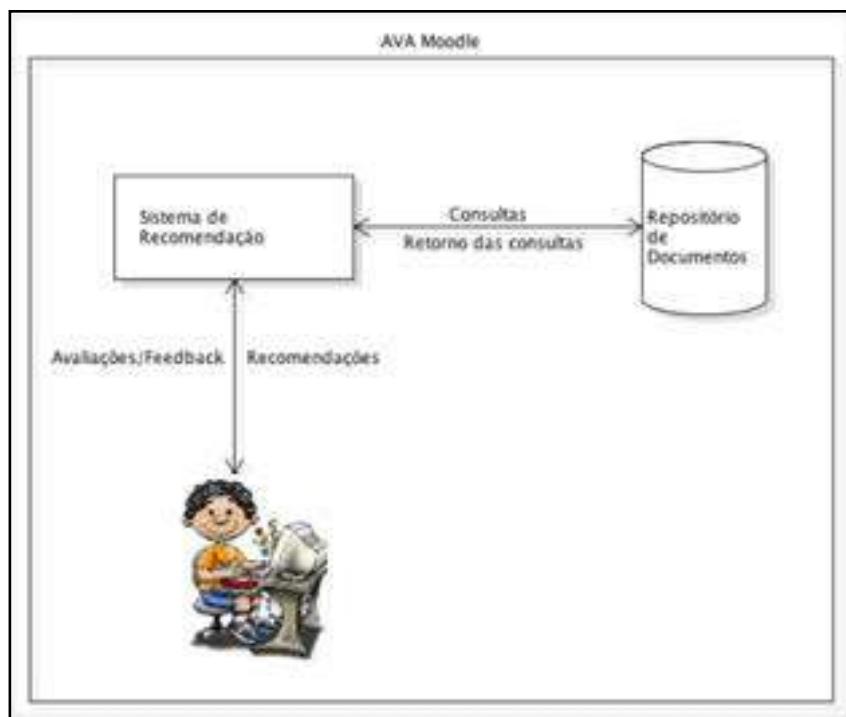


Figura 4.9: Visão Geral do Sistema de Recomendação.

Fonte: Elaborado pelo autor (2013).

Esse modelo, permite a integração do SRD com repositório local do MOODLE mostrado na figura 4.8.

Nesse contexto, os repositórios são sistemas que permitem disponibilizar conteúdos através do cadastro de características desses conteúdos (metadados) e/ou seus arquivos, as quais facilitam a sua posterior localização pelos usuários. Dentre as vantagens no uso de repositórios está a facilidade de localização de conteúdos (YALCINALP; EMIROGLU, 2012). Na seção seguinte são apresentadas a base de dados do MOODLE o conjunto de tabelas utilizadas para desenvolvimento do SRD e seus relacionamentos.

4.4 Base de Dados do MOODLE

O MOODLE conta com um banco de dados para guardar todo conteúdo do ambiente, desde informações dos usuários e logs de atividades até configurações internas de sua instalação. Na instalação da versão 2.2.8+ são criadas 285 tabelas.

O modelo do SRD em desenvolvimento faz uso de um conjunto de hashtags postadas pelos usuários nos fóruns do AVA MOODLE, para melhor compreensão desse sistema a figura 4.10 e 4.11, mostram o grupo de tabelas Fórum e o Diagrama de Classe do grupo Fórum, já os quadros 4.1 e 4.2 mostram o Dicionário de Dados das principais tabelas dos fóruns. As figuras citadas são apresentadas por fornecer ao SRD informações necessárias para sua implementação. A extração dos modelos foi realizada utilizando o programa MySQL Workbench 6.0.

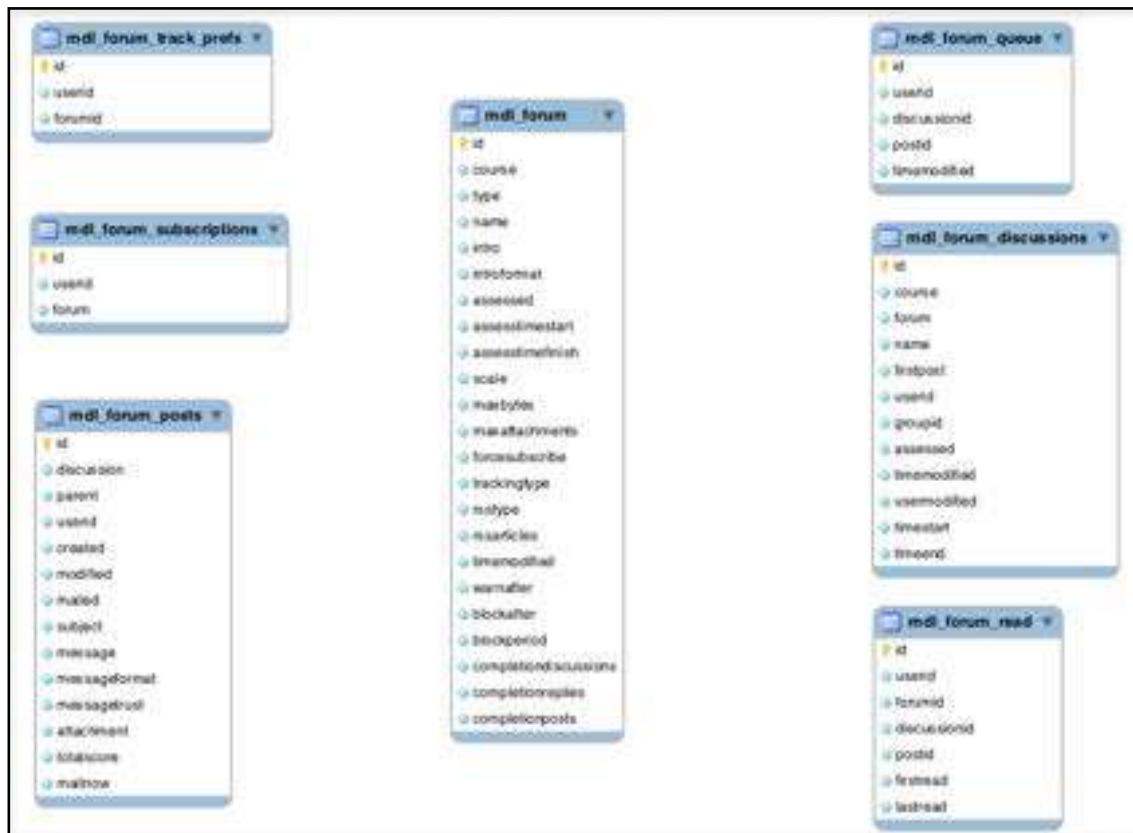


Figura 4.10: Grupo de tabelas fórum.

Fonte: Elaborado pelo autor (2013).

A completa implementação do MOODLE engloba o desenvolvimento de diversas classes correlacionadas. O diagrama de classe do grupo Fórum apresentado pela figura a seguir.

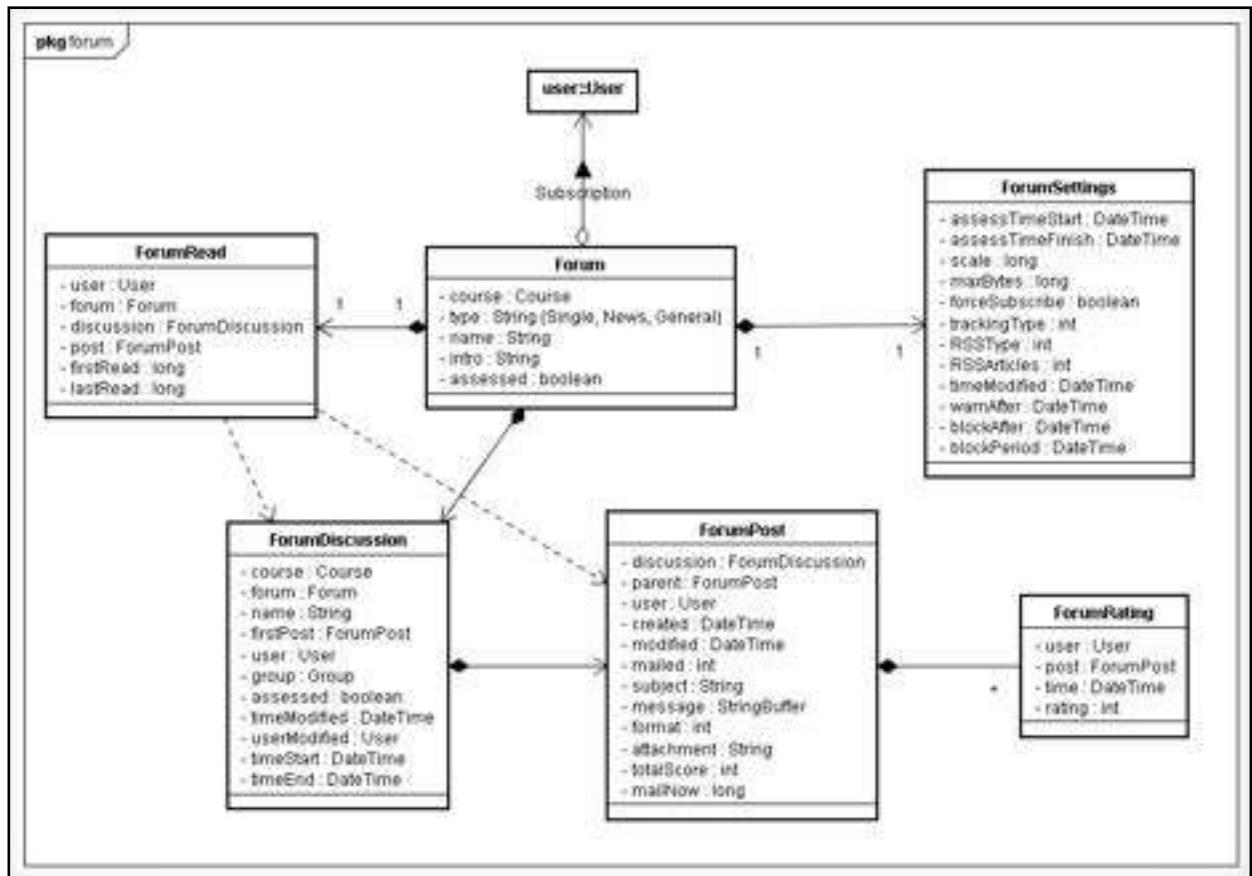


Figura 4.11: Diagrama de Classe do grupo Fórum.

Fonte: Elaborado pelo autor (2013).

Os quadros 4.1 e 4.2 mostram o Dicionário de Dados relacionados com a tabela fórum_discussions e fórum_post, as quais são utilizadas pelo sistema para armazenar as postagens de todos os usuários, fornecendo as informações necessárias para o sistema realizar as recomendações.

Quadro 4.1: Dicionário de Dados mdl_forum-discussions.

mdl_forum_discussions						
ColumnName	DataType	PrimaryKey	NotNull	Flags	Default Value	Comment
id	BIGINT(10)	PK	NN	UNSIGNED		
course	BIGINT(10)		NN	UNSIGNED		
forum	BIGINT(10)		NN	UNSIGNED		

utolInc

name	VARCHAR(255)	NN	
firstpost	BIGINT(10)	NN	UNSIGNED
userid	BIGINT(10)	NN	UNSIGNED
groupid	BIGINT(10)	NN	-1
assessed	TINYINT(1)	NN	1
timemodified	BIGINT(10)	NN	UNSIGNED
usermodified	BIGINT(10)	NN	UNSIGNED
timestamp	BIGINT(10)	NN	UNSIGNED
timeend	BIGINT(10)	NN	UNSIGNED
IndexName	IndexType	Columns	
PRIMARY	PRIMARY	id	
mdl_forudisc_use_ix	Index	userid	
mdl_forudisc_for_ix	Index	forum	

Fonte: Elaborado pelo autor (2013).

Quadro 4.2: Dicionário de Dados mdl_forum-post.

mdl_forum_posts							
ColumnName	DataType	PrimaryKey	NotNull	Flags	Default Value	Comment	AutoInc
id	BIGINT(10)	PK	NN	UNSIGNED			
discussion	BIGINT(10)		NN	UNSIGNED			
parent	BIGINT(10)		NN	UNSIGNED			
userid	BIGINT(10)		NN	UNSIGNED			
created	BIGINT(10)		NN	UNSIGNED			
modified	BIGINT(10)		NN	UNSIGNED			
mailed	TINYINT(2)		NN	UNSIGNED			
subject	VARCHAR(255)		NN				
message	TEXT		NN				
format	TINYINT(2)		NN				
attachment	VARCHAR(100)		NN				
totalscore	SMALLINT(4)		NN				
mailnow	BIGINT(10)		NN	UNSIGNED			
IndexName	IndexType	Columns					
PRIMARY	PRIMARY	id					
mdl_forupost_use_ix	Index	userid					
mdl_forupost_cre_ix	Index	created					
mdl_forupost_mai_ix	Index	mailed					
mdl_forupost_dis_ix	Index	discussion					
mdl_forupost_par_ix	Index	parent					

Fonte: Elaborado pelo autor (2013).

O SRD está sendo desenvolvido baseado nas informações contidas no grupo de tabelas relacionadas aos fóruns, Diagrama de Classe e Dicionário de Dados das principais tabelas dos fóruns.

A arquitetura do sistema de recomendação bem com o relacionamento com as tabelas serão descritas no capítulo 5.

4.5 Fóruns do MOODLE

O fórum é uma ferramenta de comunicação assíncrona, que possibilita a interação e discussão entre os participantes do curso sobre determinado assunto. As mensagens são estruturadas de forma hierárquica, apresentando os assuntos em destaque. Apesar dessa hierarquia, o fórum traz o potencial do meio digital, por permitir dinâmicas hipertextuais e agregação de várias mídias. Além disso, os usuários são notificados via e-mail de cada nova mensagem enviada ao fórum. A estrutura do seu banco de dados é mostrada na figura 4.12:

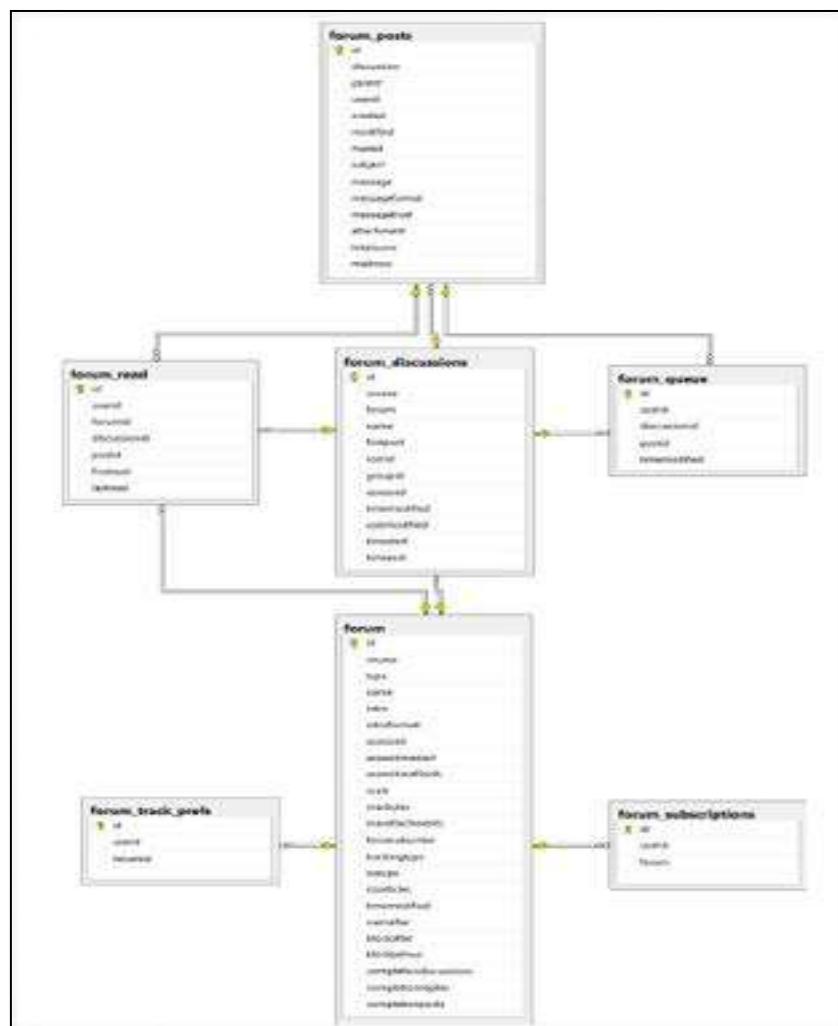


Figura 4.12: Diagrama de Banco de Dados do Fórum.
Fonte: Elaborado pelo autor (2013).

As principais tabelas desse módulo são: fórum_post, fórum e fórum_discussions. Por ser modular cada tabela do MOODLE contém as principais informações de tópico iniciado, como título, curso, usuário que o postou, data de envio, etc. O SRD fará uso principalmente da tabela fórum_post, pois a mesma contém a mensagem propriamente dita que será utilizada pelo sistema de recomendação.

Além dos relacionamentos mostrados na figura 4.12, algumas tabelas também se relacionam com as tabelas “course” e “user”, através dos campos “userid” e “course”. Como o foco dessa pesquisa são as mensagens contidas nos fóruns, mostramos apenas o diagrama de banco de dados do fórum.

4.6 Considerações do capítulo

Neste capítulo foram apresentados estudos relacionados a arquitetura do MOODLE versão 2.2.8+, bem como os requisitos do sistema, repositórios de arquivos, base de dados, diagrama do classe, dicionário de dados e as principais tabelas do AVA.

O próximo capítulo descreve o módulo de Recomendação desenvolvido, tendo como proposta principal localizar o conjunto de hashtags mais postadas nos fóruns do AVA MOODLE, classificando-as, ranqueando-as e sugerindo conteúdos relacionados ao tema em estudo do aluno. Tal solução utilizará a técnica de atribuição de pesos a documentos TF/IDF.

5 ARQUITETURA DO SISTEMA DE RECOMENDAÇÃO

Este capítulo apresenta as definições quanto ao modelo do sistema desenvolvido neste trabalho. São apresentados a visão geral do sistema e os aspectos relacionados ao desenvolvimento do sistema de recomendação, como o diagrama de entidade e relacionamentos, diagrama de atividades, diagrama de sequência, o desenvolvimento do modelo e finalizando o capítulo apresentamos a análise do questionário avaliativo.

5.1 Visão Geral

Os três principais componentes do SRD são: os usuários, os fóruns e o sistema de recomendação.

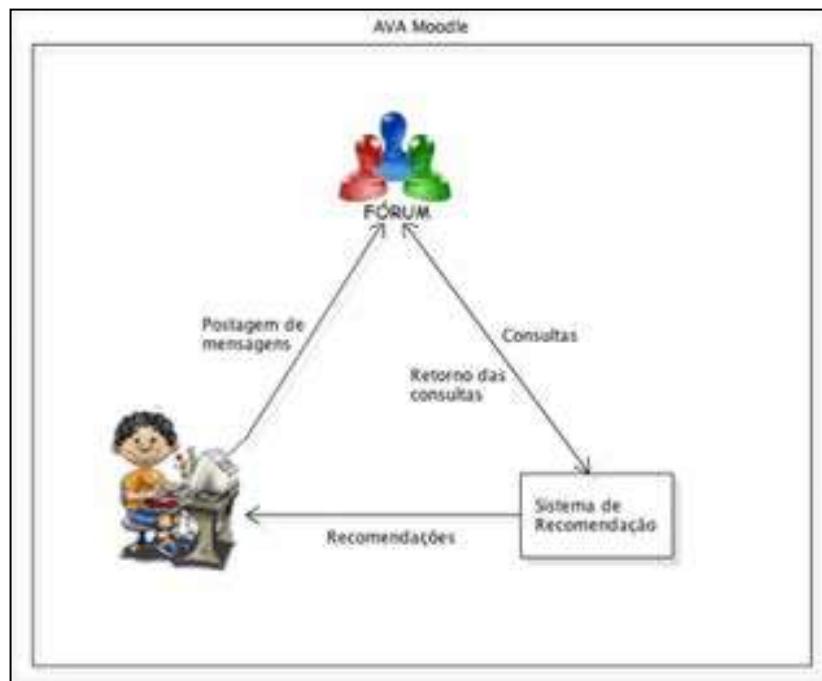


Figura 5.1: Componentes do SRD.
Fonte: Elaborado pelo autor (2013).

Para receber as recomendações, os usuários do AVA MOODLE deverão atribuir hashtags às suas postagens nos fóruns, com a finalidade de criar um

agrupamento textual em torno da marcação, por exemplo, o usuário A faz a seguinte postagem: “*a palavra informática é a junção de outras duas palavras, uma delas é informação e a outra é automática, ou seja, obter informação de maneira automática #informatica.*”

O sistema fará um agrupamento textual em torno das hashtags postadas no fóruns, criando vetores de termos e pesos, o qual possibilita a realização do cálculo de similaridade e o ranqueamento das hashtags. Para exemplificar, visualize o quadro a seguir:

Quadro 5.1: exemplo de ranqueamento das hashtags postadas.

Hashtags Usuário	Vetor	Similaridade	Ranking
A	<0.3, 0.0, 0.5>	Sim(A,B) = 0.73	2º lugar
B	<0.5, 0.4, 0.3>	Sim(A,C) = 1.21	3º lugar
C	<0.0, 0.0, 0.8>	Sim(A,B) = 0.38	1º lugar

Fonte: Elaborado pelo autor (2013).

Concluído do ranqueamento, a hashtags, melhor posicionada, é comparada com o conteúdo de todos os documentos armazenados no repositório do AVA MOODLE, logo após o usuário recebe a recomendação.

5.2 Sistema proposto

O ponto de partida deste trabalho remete ao capítulo 2 seção 2.1 Sistema de Recomendação. Já está claro que o sistema fará uso do conjunto de *hashtags* postadas nos fóruns do AVA MOODLE para recomendar OA aos usuários da plataforma. Entretanto, ficou para esta seção a explicação mais clara de como isto se dará.

Primeiramente, define-se um pressuposto baseado na finalidade básica das *hashtags*, que é classificar:

Pressuposto: A atribuição de *hashtags* às postagens de um usuário cria um agrupamento contextual em torno da marcação, ou seja, as postagens fornecem contexto às *hashtags*.

Essa observação ganha relevância quando se visualiza cada *hashtag* como um agregado de postagens. Formando assim, um documento do usuário composto de recortes desconexos em torno de um mesmo assunto.

O problema desta abordagem é definir como as informações serão tratadas e comparadas com os documentos contidos em repositório. Para tanto, será criado para cada hashtag h_j um vetor de termos e pesos, contendo todos os termos das postagens de um determinado usuário que esteja marcado com h_j . Para esta formulação, utilizar-se-á uma técnica de atribuição de pesos a documentos conhecida na área de RI que é o esquema a frequência de ocorrência do termo no documento (TF) e o inverso da frequência do termo entre documentos da coleção (IDF). Desta forma, para cada termo (palavra) das postagens associadas a uma hashtag h_j , teremos o seguinte esquema de cálculo dos pesos W_{ij} :

$$w_{i,j} = \begin{cases} (1 + \log f_{i,j}) \times \log \frac{N}{n_i} & \text{se } f_{i,j} > 0 \\ 0 & \text{se } f_{i,j} \leq 0 \end{cases}$$

Figura 5.2: Fórmula TF/IDF.
Fonte: (FERNEDA, 2003).

As variáveis da fórmula são detalhadas abaixo:

- $W_{i,j}$ é o peso de cada termo k_i associado à uma hashtag h_j ;
- $f_{i,j}$ é a frequência de ocorrência da um termo k_i nas postagens de h_j ;
- N é o número total de hashtags utilizadas em uma determinada sala de aula; e
- n_i é o número de hashtags (vistas como agrupamento de posts de um usuário) que fazem referência ao termo k_i nos fóruns de uma determinada sala de aula, disciplina ou curso (esta terminologia varia conforme a estrutura de cursos adotada).

Convém salientar que este estudo faz adequação da medida estatística denominada TF/IDF (BAEZA-YATES e RIBEIRO-NETO, 1999; MANNING et al, 2009), à estrutura dos cursos do MOODLE, conforme mostrado na figura 5.3. No esquema original, a variável N é o número de documentos de uma coleção e n_i é a quantidade de documentos nos quais o termo k_i ocorre.

```
$sql4 = "SELECT id, filename FROM mdl_files";
$sql_result4 = mysql_query($sql4,$conn);
$qtd_arq_geral = mysql_num_rows($sql_result4);

foreach ($hashtag_total as $chave => $valor){
    if($hashtag_total_arquivo[$chave] > '0'){
        $div = $qtd_arq_geral / $hashtag_total_arquivo[$chave];
        $log = log($div);
        $hashtag_termo_freq[$chave] = $hashtag_total[$chave] * $log;
    }
    else{
        $hashtag_termo_freq[$chave] = '';
    }
}

asort($hashtag_termo_freq);

foreach ($hashtag_termo_freq as $chave => $valor){
    $cont++;
    if ($cont <= '1'){

        $hashtag_busca = $hashtag_nome[$chave];

        $sql5 = "SELECT contextid, filename FROM mdl_files WHERE filename like '%$hashtag_busca%'";
        $sql_result5 = mysql_query($sql5,$conn);
        $qtd_arq = mysql_num_rows($sql_result5);

        echo "<span class=titulo>Você possui $qtd_arq recomendações:<br><br></span>";
    }
}
```

Figura 5.3: Código TF/IDF.
Fonte: Elaborado pelo autor (2013).

Desta forma, é possível comparar a estrutura de dados das postagens (hashtags) com uma coleção de documentos em repositório. A princípio, as hashtags de melhor ranqueamento, serão aferidas com os documentos existentes no repositório local do MOODLE, de modo a calcular suas similaridades. Para realizar o cálculo fazemos uso do Modelo Vetorial, conforme a fórmula abaixo:

Fórmula 5.1: Adequação da Similaridade.

$$sim(H_j, D) = \frac{\sum_{t=1}^t W_{i,j} \times W_{i,d}}{\sqrt{\sum_{t=1}^t W_{i,j}^2} \times \sqrt{\sum_{j=1}^t W_{i,d}^2}} \quad (1)$$

Fonte: Elaborado pelo autor (2013).

Nessa fórmula, H_j é o vetor de termos das postagens associadas à hashtag H_j , D é o vetor de termos do documento que está sendo comparado, $W_{i,j}$ são os pesos dos termos das postagens associadas à hashtag H_j e $W_{i,d}$ são os pesos dos termos do documento.

Para a fórmula acima se observa uma nova variável D , que representa os itens (documentos) que potencialmente podem ser recomendados a um usuário. Estes itens são ranqueados de acordo com o seu grau de similaridade com alguma hashtag H_j . Portanto, torna-se possível recomendar os mais variados documentos de um repositório, sejam planilhas eletrônicas, arquivos de textos, apresentações, arquivos PDF, etc, bastando apenas que eles contenham alguma informação textual que seja indexável.

A figura 5.4, é um diagrama de componentes o qual ilustra a arquitetura do sistema de recomendação, bem como sua forma de interação com o Moodle.

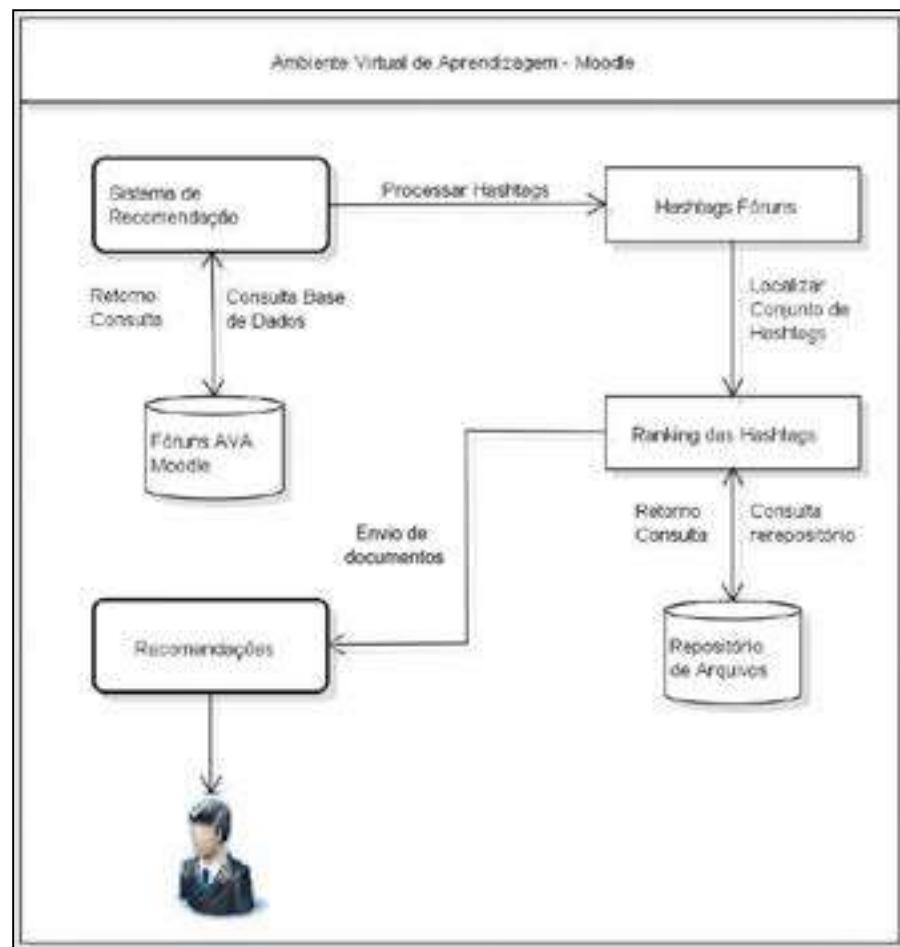


Figura 5.4: Diagrama dos componentes do sistema.
Fonte: Elaborado pelo autor (2013).

O sistema proposto é baseado na técnica de recomendação de conteúdos, no qual o algoritmo recebe consultas textuais, para extrair todas as hashtags presentes nos fóruns do AVA MOODLE de forma idêntica ao processo de indexação, retornando, então, a lista de documentos do repositório que possuem nomes relacionados com as hashtags. Para cada documento retornado, é calculado um escore entre o documento e a consulta através do somatório de pesos dos termos gerais via esquema vetorial TF-IDF. Finalizando o processo, os documentos recuperados são ordenados através dos pesos vetoriais calculados e normalizados em relação ao maior peso encontrado no repositório.

Como já era de se esperar, a arquitetura atual do MOODLE não dá suporte à esta implementação, portanto será necessária a modificação da base de dados do mesmo para viabilizar a construção desta ferramenta. A princípio serão incluídas duas novas tabelas no Banco de Dados, conforme o Diagrama de Entidade e Relacionamentos abaixo:

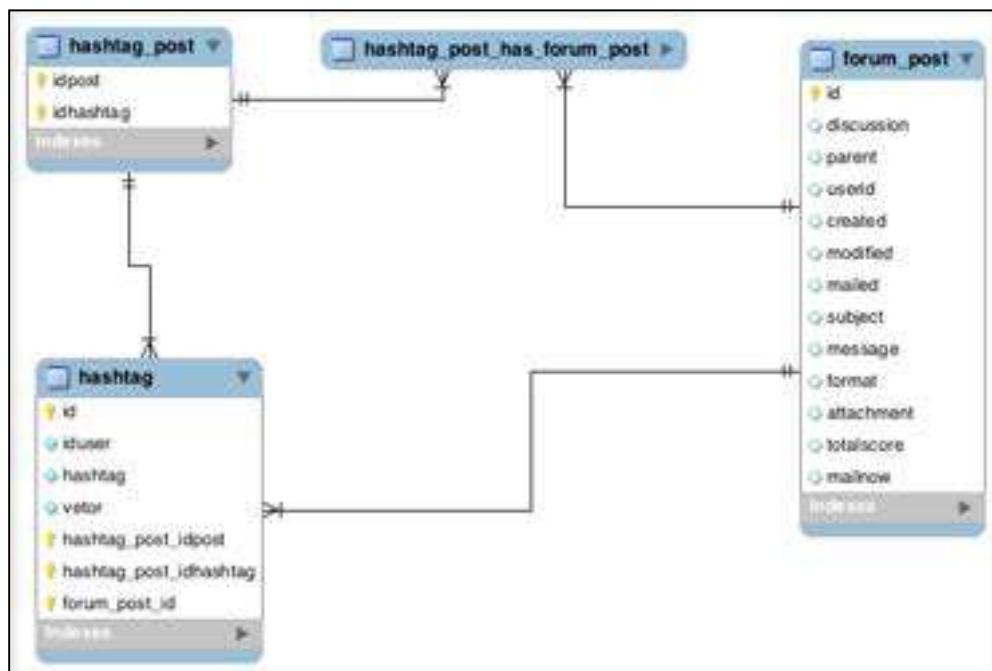


Figura 5.5: Diagrama de Entidade e Relacionamentos - Novas tabelas.
Fonte: Elaborado pelo autor (2013).

O momento da postagem é o marco inicial de todo o processo de recomendação, é a partir dele que o SRD verifica se foram utilizadas hashtags e

posteriormente efetua os cálculos necessários para realizar o ranqueamento dos elementos do repositório indexado e posteriormente recomendar.

Para entender o funcionamento do sistema de recomendação, a figura 5.6 mostra o diagrama de atividade, que apresenta o fluxo das atividades para o SRD realizar as recomendações que utiliza a técnica de filtragem baseada em conteúdo.

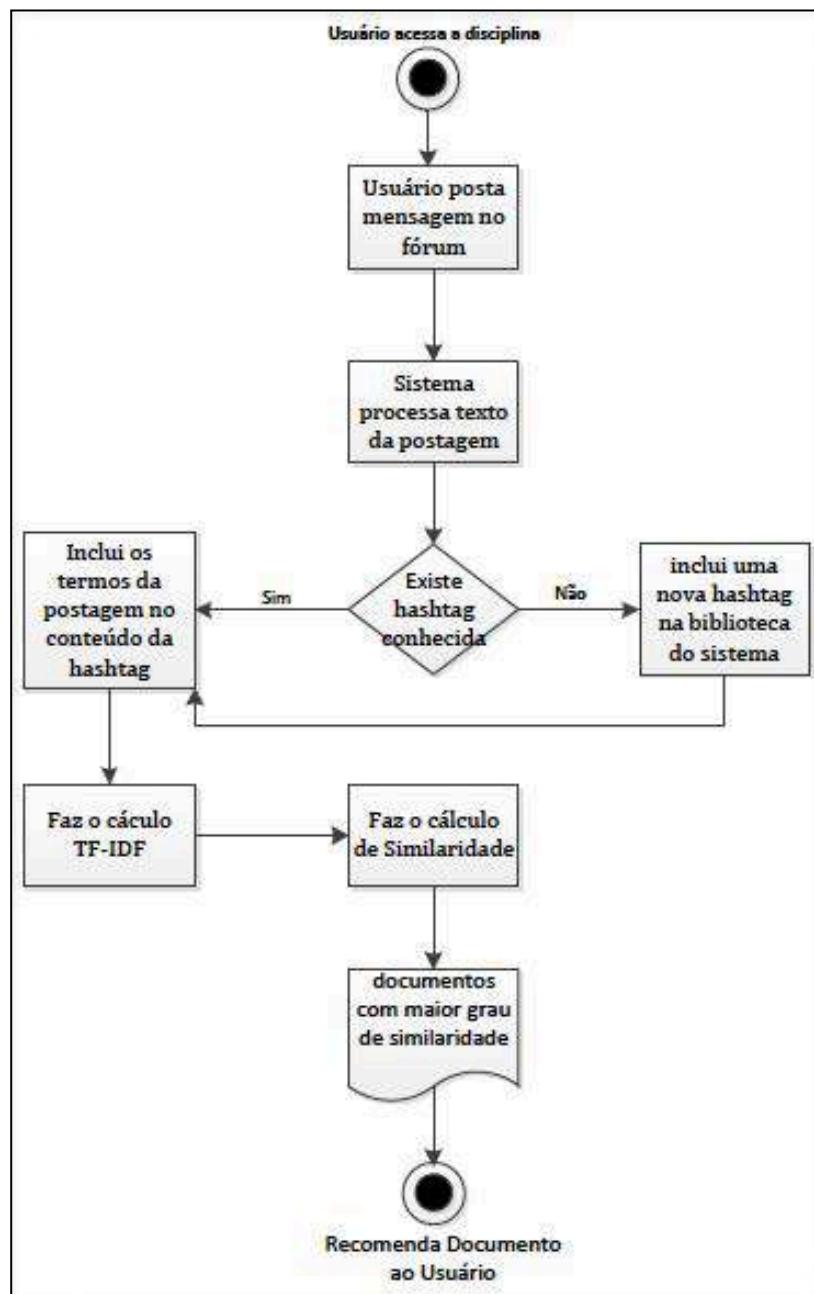


Figura 5.6: Diagrama de Atividades.
Fonte: Elaborado pelo autor (2013).

A figura 5.7 mostra o diagrama de sequência das recomendações, demonstrando que, para o usuário receber sugestões documentos é necessário a existência de hashtags nos fóruns do AVA MOODLE.

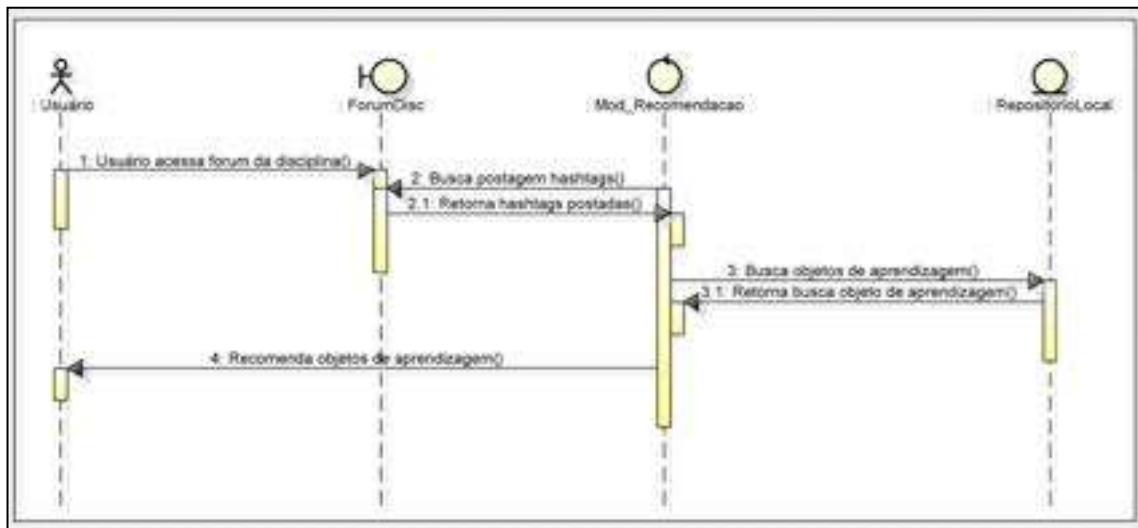


Figura 5.7: Diagrama de sequência das recomendações.
Fonte: Elaborado pelo autor (2013).

No diagrama de sequência apresentado é possível verificar que ao usuário acessar os fóruns da disciplina, o sistema irá buscar as hashtags postadas nos fóruns. Feito isso o sistema irá buscar no repositório local os documentos relacionados com as hashtags postadas para então recomendá-los ao usuário.

5.3 Desenvolvimento do sistema

O objetivo desta pesquisa é o desenvolvimento do modelo de um sistema de recomendação de documentos para os usuários do AVA MOODLE em forma de bloco. O bloco tem aplicação individual para cada disciplina que está inserido e gera recomendações baseando-se no conjunto de hashtags postadas pelos usuários nos fóruns da plataforma.

A descrição do SRD apresentada a seguir está organizada da seguinte forma:
(a) aquisição da informação; (b) recomendação; (c) visualização da recomendação.

5.3.1 Aquisição da informação

Este módulo tem a função de obter e armazenar as hashtags postadas nos fóruns. Para realizar esta atividade foi desenvolvida uma rotina no SRD que extrai as hashtags postadas nos fóruns e as armazena nas tabelas hashtag e hashtag_post da base de dados do MOODLE.

As informações que o SRD armazena sobre as hashtags postadas pelos usuários, na disciplina em que o bloco está instalado, relaciona-se com todos os fóruns. Sempre que o usuário acessa a disciplina e faz nova postagem o sistema atualiza as tabelas hashtag, hashtag_post da base de dados.

5.3.2 Recomendação

Essa funcionalidade é sem dúvida a mais importante do sistema. Para efetuar as recomendações, foi usada a abordagem de Filtragem Baseada em Conteúdo, por ser a mais adequada para uso em fóruns de comunicação, pois o algoritmo realiza as recomendações baseado nas hashtags postadas pelos usuários na plataforma.

Quando um usuário envia uma mensagem ao fórum, a(s) hashtag(s) contida(s) na mensagem (figura 5.8) é inserida em um vetor (conjunto de elementos formados por pares de chaves e valores) para realização do cálculo de atribuição de pesos a documentos – TF/IDF, conforme explicado no capítulo 2 seção 2.2.1. Em seguida a hashtag melhor ranqueada é comparada com o conteúdo do repositório local de arquivos do MOODLE e SRD faz a recomendação dos documentos.

A implementação da técnica de filtragem baseada em conteúdo utiliza os dados da tabela form_post. A relação entre as postagens e os documentos recomendados é feita através do conteúdo dos fóruns, conforme mostra figura 5.8.

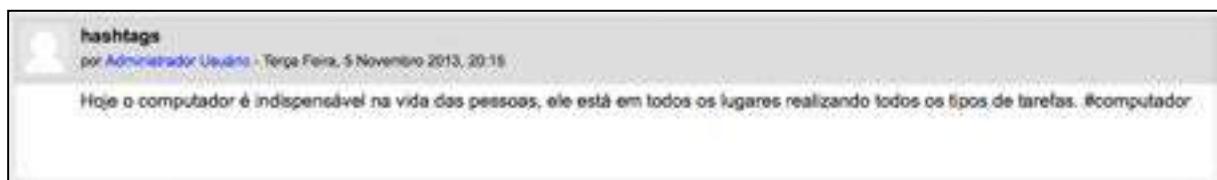


Figura 5.8: Mensagem contendo hashtag do fórum MOODLE.
Fonte: Elaborado pelo autor (2013).

A agregação das postagens formam um documento para o usuário composto de recortes desconexos em torno de um mesmo assunto, a partir do mesmo é feito o cálculo de atribuição de pesos a documentos tornando possível comparar a estrutura de dados das postagens (hashtags) com uma coleção de documentos em repositório.

5.3.3 Visualização das Recomendações

As recomendações geradas automaticamente pelo SRD ficando visíveis em forma de bloco na tela inicial do curso em que o módulo está instalado. Assim, o usuário poderá visualizá-las sem a necessidade de dar muitos cliques. As recomendações são exibidas em forma de texto, quando clicada é aberta uma janela informando ao usuário o procedimento necessário para fazer o download.

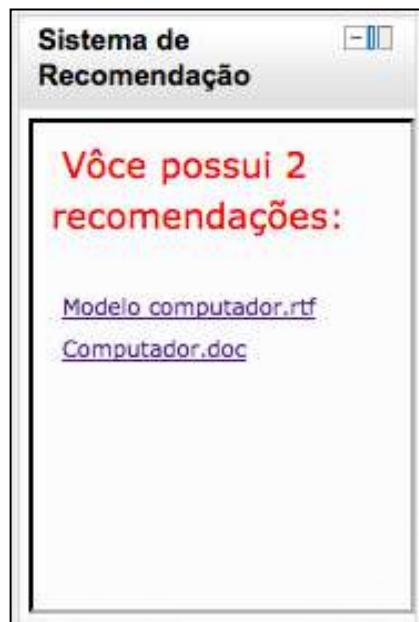


Figura 5.9: Bloco com recomendações.
Fonte: Elaborado pelo autor (2013).

A figura 5.9 mostra o bloco Sistema de Recomendação desenvolvido para os usuários do AVA MOODLE.

5.3.4 Visualização e Avaliação do Documento Recomendado

Ao clicar na recomendação desejada, a janela de visualização será aberta, a qual informa ao que para fazer o download é necessário a visualizar todo documento e ao final avaliar a recomendação como: Boa, Regular ou Ruim.



Figura 5.10: Visualização e Avaliação do documento recomendado.

Fonte: Elaborado pelo autor (2013).

Ao avaliar o documento o aluno estará dando um feedback ao sistema sobre a recomendação recebida. Em seguida poderá fazer o download do documento.

5.4 Experimentação do Sistema de Recomendação

Para realizar os primeiros testes com o modelo de sistema de recomendação, foi criado um curso no ambiente MOODLE intitulado “ABC da Informática” para 30 alunos do primeiro ano do curso Técnico de Informática do Instituto Federal de Educação, Ciência e Tecnologia do Maranhão IFMA – Campos Codó, ofertado na modalidade a distância, com três encontros presenciais e carga horaria de 20 horas, objetivando apresentar conteúdos relacionados aos tipos de computadores, seu periféricos, tipos de sistemas operacionais, editores de textos, planilhas eletrônicas e navegar na internet. Toda criação e formatação foi feita diretamente na plataforma, usando as diversas ferramentas e recursos disponíveis no ambiente, conforme mostra a figura 5.11.

Figura 5.11: Página inicial do curso ABC da Informática.

Fonte: Elaborado pelo autor (2013).

Antes do inicio do curso os alunos foram instruídos a fazer uso de hashtags nos fóruns, pois as mesmas fornecem ao sistema as informações necessárias para gerar as recomendações.

O curso foi organizado em tópicos distribuídos em temas e ao final de cada tópico foi disponibilizado um fórum de dúvidas, onde os alunos podem iniciar uma discussão em torno de um tema e inserir hashtags relacionadas ao mesmo.

5.5 Resultados

Durante os encontros presenciais e através de e-mails, alguns alunos relataram, que nunca haviam participado de um curso na modalidade a distância e que nunca tinham utilizado um ambiente virtual. Dos 30 alunos inscritos apenas 18 frequentaram regularmente o curso, resultando numa evasão de aproximadamente 40%, conforme mostrado no gráfico da figura 5.12.



Figura 5.12: Gráfico alunos inscritos e desistentes.

Fonte: Elaborado pelo autor (2013).

A maioria dos alunos desistente, alegaram não ter computador em casa, falta de tempo e dificuldade em interagir com o AVA MOODLE.

Ao finalizar o curso verificamos que o sistema gerou 41 recomendações, o que representa uma média pouco superior a duas recomendações por aluno. O feedback das avaliações feitas pelos alunos revelou que apenas 19 recomendações foram avaliadas, indicando que nem todas foram visualizadas até o final, ou que algumas não despertaram o interesse dos alunos.

De acordo com o quadro 5.2, das 19 avaliações realizadas pelos alunos, foi analisado que mais de 63,1% foram consideradas boas, número satisfatório devido a exiguidade temporal da disponibilização do curso. Esta limitação não permitiu avaliar de forma considerável a evolução das recomendações em decorrência da limitação do tempo de uso. Porém, a fim de avaliar o funcionamento do sistema, relegando o aspecto pedagógico, as avaliações foram eficientes.

Quadro 5.2: conceitos das avaliações.

Avaliação	Total de Avaliações	Porcentagem
Boa	12	63,1%
Regular	5	26,6%
Ruim	2	10,3%

Fonte: Elaborado pelo autor (2013).

As recomendações ocorreram a contento, pois os resultados revelaram o bom funcionamento do sistema, bem como sua possível utilização em larga escala. As técnicas utilizadas mostraram-se eficientes, apresentando recomendações padronizadas em caráter diversificado, segundo interesse do usuário.

A partir do trabalho desenvolvido nesta dissertação foi originada a publicação, descrita no Quadro 5.3:

Quadro 5.3 : Publicação.

Título	Autores	Conferência
Recomendando Objetos de Aprendizagem a partir das hashtags postadas no MOODLE	Francisco Adelton Alves Ribeiro, Luis Carlos Costa Fonseca, Miguel de Sousa Freitas	Congresso Brasileiro de Informática na Educação (CBIE), São Paulo 2013

Fonte: Elaborado pelo autor (2013).

A análise dos dados obtidos na avaliação mostrou um bom nível de satisfação dos usuários com as recomendações recebidas e demonstrou a possibilidade do mesmo ser utilizado por uma instituição de ensino, visando adequar suas necessidades para averiguar sua possível utilização em larga escala.

5.6 Considerações do capítulo

Neste capítulo foi descrita a arquitetura do sistema de recomendação, bem como o diagrama de entidade e relacionamentos, o diagrama de sequência, a descrição do desenvolvimento do sistema, experimentos e resultados obtidos. O próximo capítulo apresenta as considerações finais.

6 CONSIDERAÇÕES FINAIS

A EaD vem sendo a cada dia mais utilizada pelas instituições de ensino, tanto para apoio aos cursos presenciais como para cursos a distância. Nesse cenário, uma ferramenta muito importante para gestão de cursos é o AVA MOODLE, por possuir uma grande variedade de recursos para comunicação, entre eles destacamos os fóruns, que são utilizados por professores e alunos para esclarecer dúvidas e sugerir discussões em torno de um tema. É grande o número de mensagens postadas nos fóruns da plataforma e o MOODLE não possui mecanismo de gerência para essas informações.

Como forma de classificar as informações contidas nos fóruns de comunicação do AVA MOODLE, foi desenvolvido o modelo de um sistema para recomendar documentos, baseado no conjunto de hashtags postadas pelos usuários nos fóruns. O sistema foi desenvolvido em forma de bloco e adequado ao tradicional esquema de atribuição de pesos a documentos, conhecido na área da recuperação da informação o TF/IDF, permitindo a execução de um algoritmo para classificação e ranqueamento de conteúdos a partir dos termos mais utilizados, determinando assim, o quanto as hashtags estão relacionadas com o conteúdo do repositório de arquivos.

De acordo com os resultado obtidos, o modelo do sistema de recomendação de documentos para os usuário do AVA MOODLE apresentou-se satisfatoriamente nos testes iniciais, envolvendo os alunos do curso Técnico de Informática do Instituto Federal de Educação, Ciência e Tecnologia do Maranhão IFMA – Campos Codó. A arquitetura, em geral, permitiu ao sistema buscar informações relevantes para os usuários e recomendar documentos relacionados ao tema discutido.

Um dos pontos mais desafiadores desta pesquisa, foi criar o algoritmo gerador das recomendações, adequando-o ao esquema de atribuição de pesos TF/IDF e ao cálculo de similaridade. O sistema faz uso da técnica de Filtragem Baseada em Conteúdo, porém antes de executar essa tarefa, foi necessário analisar as diferentes abordagens de recomendações e criar a arquitetura geral do sistema. Outro desafio foi estudar o núcleo do AVA MOODLE, para conhecer suas 285 tabelas e entender seus relacionamentos, permitindo criar novas tabelas e seus relacionamentos, sem alterar a performance da plataforma.

Durante o desenvolvimento e o período de teste do SRD foi possível vislumbrar a ampliação de suas funções através de trabalhos futuros. Dentre estas aspirações, está o estudo de outros algoritmos para o cálculo de similaridade entre os documentos dos repositórios e as postagens, levando-se em consideração o custo computacional. Faz parte do escopo desta pesquisa também, estender a aplicação do sistema de recomendação ora proposto a outros repositórios de documentos (ou Objetos de Aprendizagem). Outra funcionalidade que pode ser implementada é a vinculação das hashtags ao usuário, tornando possível ao avaliador de um curso verificar especificamente, para cada aluno o que foi dito nos fóruns, configurando-se uma ferramenta interessante de avaliação. Entretanto, é importante ressaltar que a cultura de uso das hashtags deve estar devidamente difundida aos usuários da plataforma MOODLE.

REFERÊNCIAS

ALEXA., The top 500 sites on the web, 2012. <http://www.alexa.com/topsites>.

ADOMAVICIUS, Gediminas; TUZHILIN, Alexander. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering. **Piscataway**, v. 17, n. 6, p. 734-749, jun. 2005.

ANAND, S. S. e MOBASHER, B. (2005). **Intelligent techniques for web personalization**. Disponível em: <<http://maya.cs.depaul.edu/mobasher/papers/am-itwp-springer05.pdf>>. Acesso em: 25 de março de 2013.

ANDERSON, P. (2007). **What is Web 2.0**: ideas, technologies and implications for education. JISC Tecnology and Standards Watch. Disponível em: <<http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf>>. Acesso em: 13 jun. 2012.

ARAÚJO JÚNIOR, R. H. **Precisão no processo de busca e recuperação da informação**. Brasília: Thesaurus, 2007.

BAEZA-YATES, R. A.; Ribeiro-Neto, B. **Modern information retrieval**. Addison-Wesley Longman Publishing Co., Inc., 1999.

Badrul, S., George, K., Joseph, K., & John, R. (2001). **Item-Based Collaborative Filtering Recommendation Algorithms**. Paper presented at the Proceedings of the 10th International Conference on World Wide Web.

BALABANOVIC, M.; SHOHAM, Y. (1997). Fab: content-based, collaborative recommendation. **Communications of the ACM**, 40(3):66-72.

BEHAR, P. A. et al. Objetos de aprendizagem para educação a distância. In: _____. (Col.). **Modelos pedagógicos em educação a distância**. Porto Alegre: Artmed, v. 1, p. 66-92, 2009.

BELLONI, Maria Luiza. Educação a distância.5. ed. Campinas: Autores Associados, 2008.

BERRY, M. W.; DRMAC, Z.; JESSUP, E. R. Matrices, vector spaces and information retrieval. **SIAM Review**, v. 41, n. 2, p. 335, 1999.

BETTIO, Raphael Winckler de; MARTINS, Alejandro. **Objetos de aprendizagem – um novo modelo direcionado ao ensino a distância**. Disponível em: <<http://www.universia.com.br/materia/materia.jsp?id=5938>>. Acesso em: 15 mar 2012.

BRASIL. Referenciais de Qualidade para Cursos a Distância. Brasília. Ministério da Educação Secretaria de Educação a Distância (2003). Disponível em: <<http://portal.mec.gov.br/seed/index.php?option=content&task=view&id=62&Itemid=191>>. Acesso em: 10 jun. 2013.

BUCKLAND, M. K. What is a “document”? **Journal of the American Society of Information Science**, v. 48, n 9, p. 804-809, 1997.

BURKE, R. Hybrid web recommender systems. In: BRUSILOVSKY, P., KOBSA, A., NEJDL, W. (eds.). **The adaptive web: methods and strategies of web personalization.**, lecture notes in computer science, v. 4321, p. 377-408, Springer, Berlin-Heidelberg, 2007.

BURKE, R. Hybrid web recommender systems. Surver and experiments. **User modeling and user adapted interaction**, v.12, n.6, p. 331-370, nov. 2002. Disponível em: < <http://josquin.cti.depaul.edu/~rburke/pubs/burke-umuai02.pdf>>. Acesso em: 20/mai. 2013.

CACHEDA, F. et al. Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems. **ACM Trans. Web**, [S.I.], v.5, n.1, p.1–33, Feb. 2011.

CARNEIRO, M. L. F. **Instrumentalização para o ensino a distância**. Porto Alegre: UFRGS, 2009.

CAZELLA, S. C. **Aplicando a relevância da opinião de usuários em sistemas de recomendação para pesquisadores**. 2006. 180f. Tese (Programa de Pós-Graduação em Computação). Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

CAZELLA, S. C.; BHEAR, P.; SCHNEIDER, D.; SILVA, K. K.; FREITAS, R. **Desenvolvendo um sistema de recomendação de objetos de aprendizagem baseado em competências para a educação: relato de experiências**. In: Anais do XXIII Simpósio Brasileiro de Informática na Educação, SBIE. Rio de Janeiro, RJ, 2012.

_____ ; DRUMM, J. V; BARBOSA, J. L. V. Um serviço para recomendação de artigos científicos baseado em filtragem de conteúdo aplicado a dispositivos móveis.

Novas tecnologias na educação. Florianópolis: CINTED-UFRGS, v. 8, nº 3, 2010. Disponível em: <http://www.lume.ufrgs.br/bitstream/handle/10183/51601/Resumo_200801528.pdf?sequence=1>. Acesso em: agosto/2013.

CLAYPOOL, M. et al. Combining content-based and collaborative filters in an online newspaper. Miscellaneous, [S.I.], 1999.

COELHO, G. W.; Uso dos recursos de mídias sociais na Educação a Distância: impactos na percepção da presença social. In: **Anais do XXIII Simpósio Brasileiro de Informática na Educação**, SBIE, Rio de Janeiro, RJ, 2012.

CORRADI, B. S. Avaliação de novas tecnologias da informação e da comunicação no processo de ensino aprendizagem em uma disciplina experimental da UFV. 2009. Disponível em: <http://www.tede.ufv.br/tedesimplificado/tde_arquivos/39/TDE-2010-02-08T102543Z-2141/Publico/texto%20completo.pdf>. Acesso em: 2012.

CROFT, W.B.: Combining approaches to information retrieval. In: W.B. Croft (ed.). Advances in information retrieval. **Kluwer Academic Publishers**, 1-36, 2000.

DONALDSON, J. A hybrid social-acoustic recommendation system for popular music. **ACM Conference On Recommender Systems**. p. 187-190, 2007. Disponível em: <<http://portal.acm.org/citation.cfm?id=1297271>>. Acesso: 07 jun. 2012.

FELFERNIG, A.; BURKE, R. D. Constraint-based recommender systems: technologies and research issues. **ACM International Conference Proceeding Series**, v. 342, 2008. Disponível em: <<http://portal.acm.org/citation.cfm?id=1409544>>. Acesso em: 10 jan. 2011.

FERNEDA, Eduino. **Recuperação da informação:** análise sobre a contribuição da ciência da computação para a ciência da informação. 2003. Tese (Doutorado) Programa de Pós-Graduação da Escola de Comunicação e Arte da Universidade de São Paulo. São Paulo: USP, 2003.

FERREIRA, V. H. **Sistema de recomendação de objetos de aprendizagem.** Itajaí, 2009. TCC (Graduação) – Centro de Ciências Tecnologias da terra e do Mar, Curso de Ciência da Computação , UNIVALI, Itajaí.

FERRO, M. R. C. **Modelo de Sistema de Recomendação de materiais Didáticos para Ambientes Virtuais de Aprendizagem.** Maceió, 2010. Dissertação (Mestrado) – Instituto de Computação, Programa Multidisciplinar de Pós-Graduação em Modelagem Computacional de Conhecimento de Computação , UFAL, Maceió.

FILATRO, A. **Design instrucional contextualizado:** educação e tecnologia. São Paulo: SENAC, 2007.

FRAKES, W. B., and BAEZA-YATES, R. Information retrieval: data structures and algorithms. englewood cliffs, NJ: Prentice Hall. See more at: <<http://www.maa.org/publications/periodicals/loci/joma/the-linear-algebra-behind-search-engines-references-and-resources#sthash.Y9aKj1PN.dpuf>, 1992.

FUHR, N. Probabilistic models in information retrieval. **The Computer Journal**, v. 35, p. 243-255, 1992.

GOLDBERG, David et al. Using collaborative filtering to weare an information tapestry comm. **Communications of the ACM.** New York, v. 35, n. 12, p. 61-70, dec. 1992.

GOLDER, S.; HUBERMAN, B.A. Usage patterns of collaborative tagging Systems. **Journal of Information Science**, v. 32, n. 2, p. 198-208, 2006.

HERLOCKER, J. L. **Understanding and improving automated collaborative filtering systems**. 2000. 144p. Tese. University of Minnesota, Minnesota, EUA.

HILL, W. et al. Recommending and evaluating choices in a virtual community of use. In: **Proceedings...** [S.I.: s.n.], 1995, p.194-201.

HUANG, Z. et al. A Graph-based Recommender System for Digital Library. In: JCDL, 2002. **Proceedings...** [S.I.: s.n.], 2002.

IBOPE – Instituto Brasileiro de Opinião Pública e Estatística. Disponível em: <http://www.appbrasil.org.br/ibope/free/acesso_internet.pdf>. Acesso em: 24/09/2013.

LONGMIRE, W. A. A primer on learning objects. In: _____. **American Society for Training & Development**. Virginia, USA: [s.n.], 2001.

LUND, Ben et al. Social Bookmarking Tools (II): a case study: Connotea. D-Lib Magazine, v. 11, n. 4, apr. 2005. Disponível em: <<http://www.dlib.org/dlib/apr05/lund/04lund.html>>. Acesso em: 14 nov. 2012.

MAIA, C.; MATTAR, João. **ABC da EaD**: a educação a distância hoje. São Paulo: Pearson Prentice Hall, 2007.

MANNING, C.; RAGHAVAN, p.; SCHUTZE, H. **Introduction to information retrieval.** Cambridge University Press, 2009.

MANOUSELIS, N.; COSTOPOULOU, C. Experimental analysis of multiattribute utility collaborative filtering on a synthetic data set. **Personalization techniques and recommender systems**, 2008, cap. 5, p. 111-133.

MATHES, Adam. **Folksonomies** - cooperative classification and communication through shared metadata. Computer mediated communication – LIS590CMC, Urbana: University of Illinois, 2004. Disponível em: <<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>>. Acesso em: 25 mar. 2013.

MENDES, R. M.; SOUZA, V. I.; CAREGNATO, S. E. A propriedade intelectual na elaboração de objetos de aprendizagem. In: BAHIA, E. da Universidade Federal da (Ed.). CINFORM - Encontro Nacional de Ciência da Informação. Salvador: **Anais**, 2004. v. 5. Disponível em: <[http://www.ciform.ufba.br/v_anais/artigos/rozimaramendes.html](http://www.cinform.ufba.br/v_anais/artigos/rozimaramendes.html)>. Acesso em: 25 mar. 2013.

MOITA, F. e SILVA A. **Os games no contexto de currículo e aprendizagens colaborativas on-line.** Comunicação apresentada no III congresso luso brasileiro sobre questões curriculares. Publicado nos anais em CD-Rom. Braga, PT, de 09 a 11 de fevereiro de 2006.

MORAIS, C. T. Q. **Modelo de Sistema de Alertas e Recomendação Para Mediar Aprendizagem em Turmas Heterogêneas nos Cursos de EAD.** Porto Alegre, 2011. Tese (Doutorado) – Programa de Pós-Graduação em Informática na Educação, Centro de Estudos Interdisciplinares em Novas Tecnologias da Educação, UFRGS, Porto Alegre.

MORAN, J. M. Ensino e aprendizagem inovadores com tecnologias audiovisuais e telemáticas. In: ____; MASETTO, M.T.; BEHRENS, M.A. **Novas tecnologias e mediação pedagógica**. Campinas: Papirus. 10. ed. p. 11-65, 2006.

MOURA, S. L. Uma arquitetura para integração de repositórios de objetos de aprendizagem baseados em mediadores e serviços WEB. 2005. Dissertação (Dissertação de Mestrado) Departamento de Informática, PUC-Rio, 2005.

NASCIMENTO, Anna Christina Aun de Azevedo. Aprendizagem por meio de repositórios digitais e virtuais. In: LITTO, Frederic Michael; FORMIGA, Manuel Marcos Maciel (Org.). **Educação a distância: o estado da arte**. São Paulo: Person Educacion do Brasil, 2009. Cap. 49, p. 352-357.

O'DONOVAN, J.; SMYTH, B. Trust in recommender systems. In: **Proceedings of the international conference on intelligent user interfaces (IUI'05)**. 10, São Diego, Estados Unidos. Nova Iorque: ACM Press, 2005.

OLIVEIRA, L.G. **Proposal of the methodological structure to implement recommender systems**. (2007). Disponível em:
[<http://www.intelog.net/artigosnoticias/arquivos/artigo_leonardo_oliveira.pdf>](http://www.intelog.net/artigosnoticias/arquivos/artigo_leonardo_oliveira.pdf).
Acesso em: 02 de março de 2012.

PAGE, L.; BRIN, S.; MOTWANI, R. & WINOGRAD, T. **The pagerank citation: bringing order to the web**. Computer Science Departament, Standford University, 1998.

PEREIRA, R.; DA SILVA, S. Folksonomias: uma análise crítica focada na interação e na natureza da técnica. In: de Computação, S. B., ed. Proceedings of the VIII

brazilian symposium on human factors in computing systems. **Association for Computing Machinery**, 2008, p. 126-135.

QUINTARELLI, Emanuele. Folksonomies: power to the people. In: INCONTRO ISKO ITALIA -UNIMIB, Milão, 2005. **Papers...** Milan: Universitá di Milano, 2005. Disponível em: <<http://www.iskoi.org/doc/folksonomies.htm>>. Acesso em: 05 jun. 2012.

REATEGUI, E. B.; CAZELLA, S. C. Sistemas de recomendação. XXV CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, XXV. 2005. São Leopoldo. **Anais...**, São Leopoldo: Unisinos, 2005, p. 306-348.

RESNICK, P. et al. Group Lens: an open architecture for collaborative filtering of netnews. In: **Proceedings...** [S.l.: s.n.], 1994, p.175-186.

RIBEIRO, F. A. A. **Moodle e as novas formas do aprender**. In: V Congresso Norte e Nordeste de Pesquisa e Inovação, 19, 2010, Maceió. **Anais...** 2010 p. 71-80, nov., 2010.

RIZZI C. et al. Fazendo uso de categorização de textos em atividades empresariais. In: INTERNATIONAL SYMPOSIUM ON KNOWLEDGE MANAGEMENT/DOCUMENT MANAGEMENT (ISKM/DM 2000), III, 2000, Curitiba, **Proceedings...** Curitiba: PUC-PR, 2000, p. 125-147. Disponível em: <<http://www.leandro.wives.nom.br/pt-br/publicacoes/iskmdm2000-2.pdf>>. Acesso em: 28 de jul. 2013.

RODRIGUES, A. P. **Integração de Ambiente Virtual de Aprendizagem com Repositório Digital**. Tese (doutorado em Informática na Educação) Programa de Pós-graduação em Informática na Educação, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2012.

ROSCH, E. Principles of categorization. In: ROSCH, E.; LLOYD, B., eds. **Cognition and categorization**. Hillsdale, New Jersey: Erlbaum, p. 27-48, 1978.

SALTON, Gerard; BUCKLEY, Christopher. Term-weighting approaches in automatic text retrieval. In: **Information processing & management**, vol. 24, pages 513-523, 1988.

SARWAR, B.; KARYPIS, G.; KONSTAN, J.; RIEDL, J. Analysis of recommendation algorithms for e-commerce. **Electronic commerce**. 2000, p. 158-167. Disponível em: <<http://portal.acm.org/citation.cfm?id=352887>>. Acesso em: 10 jan. 2012.

SCHAFER, J. B.; KONSTAN, J.; RIEDL, J. Recommender systems in e-commerce. **Electronic commerce: proceedings of the 1st ACM conference on Electronic commerce**, 1999, p.158-166. Disponível em: <<http://portal.acm.org/citation.cfm?id=337035>>. Acesso em: 11 jan. 2012.

SCHENKEL, R.; CRECELIUS, T.; KACIMI, M.; MICHEL, S.; NEUMANN, T.; PARREIRA, J. X.; WEIKUM, G. Efficient top-k querying over social-tagging networks. In: **Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval**, New York, NY, USA: ACM, 2008, p. 523-530.

SHARDANAND, U.; MAES, P. Social information filtering: algorithms for automating “word of mouth”. In: **Proceedings...** [S.l.: s.n.], 1995, p. 210-217.

SILVA, R. P. Avaliação da perspectiva cognitivista como ferramenta de ensino/aprendizagem da geometria descritiva a partir do ambiente hipermédia hypercal GD. Florianópolis: UFSC, 2005, 214p. Tese (Doutorado) – Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina, Florianópolis, 2005.

SILVA, T. L. K. **Produção flexível de materiais educacionais personalizados: o caso da geometria descritiva.** Florianópolis: UFSC, 2005, 185p. Tese (Doutorado) – Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina, Florianópolis, 2005.

SMITH, Martha Kellogg. Viewer tagging in art museums: comparisons to concepts and vocabularies of art museum visitors. In: SIG/CR CLASSIFICATION RESEARCH WORKSHOP, 17., Austin (USA), 2006. **Papers...** American Society for Information Science and Technology, 2006. Disponível em: <<http://www.slais.ubc.ca/users/sigcr/sigcr-06smith.pdf>>. Acesso em: 06 nov. 2012.

SNOWBALL. 2011. **Lista de stop-words para português.** Disponível em: <<http://snowball.tartarus.org/algorithms/portuguese/stop.txt>>. Acesso: em 10 de jul. 2013.

TAROUCO, Liane; FABRE, Marie-Christine Julie Mascarenhas; TAMUSIUNAS, Fabrício Raupp. Reusabilidade de objetos educacionais. **Revista Novas Tecnologias na Educação.** Porto Alegre, p. 1-11, 2003.

TORRES, R. Combining collaborative and content-based filtering to recommend research papers. Miscellaneous, [S.I.], 2004.

TRANT, Jennifer. Exploring the potential for social tagging and folksonomy in art museums: proof of concept. **New Review of Hypermedia and Multimedia**, v.12, n.1, p.63-81, jun. 2012.

_____. Social classification and folksonomy in art museums: early data from the steve.museum tagger prototype. In: SIG/CR CLASSIFICATION RESEARCH WORKSHOP, 17. , Austin (USA), 2006. **Papers...** American Society for Information Science and Technology, 2006. Disponível em: <<http://www.slais.ubc.ca/users/sigcr/sigcr-06trant.pdf>>. Acesso em: 06 jan. 2013b.

VAVASSORI, Fabiane B. e RAABE, André L. A. Organização de atividades de aprendizagem utilizando ambientes virtuais: um estudo de caso. In: SILVA, Marco (org). **Educação online:** teorias, práticas, legislação, formação corporativa . São Paulo: Loyola, 2003, p. 311-325.

VIEIRA, Alessandro. **Objetos de aprendizagem:** uma abordagem lúdica para ensino de matemática. 2007. Disponível em: <www.gamecultura.com.br/index.php?option=com_content&task=view&id=404>. Acesso em: 01 jul. 2011.

WAL, Thomas Vander. **Folksonomy definition and wikipedia.** Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>. Acesso em: 22 nov. 2012.

ZAMBERLAN, D. M. **Recomendação de mensagens no fóruns do MOODLE.** Porto Alegre, 2011. TCC (Trabalho de Graduação) - Instituto de Informática, UFRGS, Porto Alegre.

WIVES, L. K. **Utilizando conceitos como descritores de textos para o processo de identificação de conglomerados (clustering) de documentos.** 2004. 136f. Tese (Doutorado em Ciência da Computação) – Instituto de Informática, UFRGS, Porto Alegre. Disponível em: <<http://www.leandro.wives.nom.br/pt-br/publicacoes/Tese.pdf>>. Acesso em: 20 de jul. 2013.

YALCINALP S., EMIROGLUB.G. (2012) “**Through Efficient Use of LORs: Prospective Teachers’ Views on Operational Aspects of Learning Object Repositories**” British Journal of Educational Technology, Cilt 43, Sayı 3, 474 – 488 (SSCI Index).

ZHOU, D.; BRIAN, J.; ZHENG, S.; ZHA, H.; GILES, C. L. Exploring social annotations for information retrieval. In: **proceedings of the www/internet conference:** social networks and web 2.0 Track, 2008, p. 715-724.