

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

Breno Fabrício Terra Azevedo

**MINERAFÓRUM: UM RECURSO DE APOIO
PARA ANÁLISE QUALITATIVA EM FÓRUMS DE DISCUSSÃO**

Porto Alegre
2011

Breno Fabrício Terra Azevedo

**MINERAFÓRUM: UM RECURSO DE APOIO
PARA ANÁLISE QUALITATIVA EM FÓRUMS DE DISCUSSÃO**

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

Orientadora: Dra. Patricia Alejandra Behar

Coorientador: Dr. Eliseo Berni Reategui

Linha de Pesquisa: Ambientes Informatizados e Ensino a Distância

Porto Alegre
2011

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. José Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Aldo Bolten Lucion

Diretora do CINTED: Profa. Dra. Liane Margarida Rockenbach Tarouco

Coordenadora do PPGIE: Profa. Dra. Maria Cristina Villanova Biazus

CIP - Catalogação na Publicação

Azevedo, Breno Fabrício Terra

MineraFórum: um recurso de apoio para análise qualitativa em fóruns de discussão / Breno Fabrício Terra Azevedo. -- 2011.

204 f.

Orientadora: Patricia Alejandra Behar.

Coorientador: Eliseo Berni Reategui.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Centro de Estudos Interdisciplinares em Novas Tecnologias na Educação, Programa de Pós-Graduação em Informática na Educação, Porto Alegre, BR-RS, 2011.

1. Mineração de textos. 2. Fórum de discussão. 3. Análise qualitativa. 4. Relevância temática. I. Behar, Patricia Alejandra, orient. II. Reategui, Eliseo Berni, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Breno Fabrício Terra Azevedo

**MINERAFÓRUM: UM RECURSO DE APOIO
PARA ANÁLISE QUALITATIVA EM FÓRUNS DE DISCUSSÃO**

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

Aprovada em 13/12/2011.

Profa. Dra. Patricia Alejandra Behar - Orientadora

Prof. Dr. Eliseo Berni Reategui - Coorientador

Profa. Dra. Liane Margarida Rockenbach Tarouco

Profa. Dra. Mára Lúcia Fernandes Carneiro

Prof. Dr. Sílvio César Cazella

AGRADECIMENTOS

A minha esposa, Flávia Maria, pelo amor, carinho, amizade, dedicação e pelo enorme apoio para realizar este trabalho.

A minha orientadora, Professora Patricia Behar, pela amizade, confiança, incentivo, energia e pelo exemplo de profissional.

Ao meu coorientador, Professor Eliseo Reategui, pelo apoio, amizade e pelas excelentes contribuições.

Aos meus pais, Cláudio e Regina, pelo amor, dedicação, carinho e exemplo de vida.

A minha irmã Betina pela amizade, amor e alegria.

Aos Professores Liane Tarouco, Sílvio Cazella e Mara Carneiro pelas valiosas contribuições a esta pesquisa e por aceitarem participar da banca desta Tese.

A equipe do NUTED, pela parceria no desenvolvimento deste trabalho.

Aos amigos da turma do Doutorado, pelo auxílio e por propiciaram diversos momentos de lazer e alegria.

Ao IFF pelo apoio e oportunidade em permitir meu aperfeiçoamento profissional.

A UFRGS, instituição pública, gratuita e de qualidade.

RESUMO

Esta tese aborda o desenvolvimento, uso e experimentação do MineraFórum. Trata-se de um recurso para auxiliar o professor na análise qualitativa das contribuições textuais registradas por alunos em fóruns de discussão. A abordagem desta pesquisa envolveu técnicas de mineração de textos utilizando grafos. As interações proporcionadas pelas trocas de mensagens em um fórum de discussão representam uma importante fonte de investigação para o professor. A partir da análise das postagens, o docente pode identificar quais alunos redigiram contribuições textuais que contemplam conceitos relativos ao tema da discussão, e quais discentes não o fizeram. Desta forma, é possível ter subsídios para motivar a discussão dos conceitos importantes que fazem parte do tema em debate. Para atingir o objetivo do presente estudo, foi necessário realizar uma revisão da literatura onde foram abordados temas como: a Educação a Distância (EAD); Ambientes Virtuais de Aprendizagem; os principais conceitos da área de Mineração de Textos e, por último, trabalhos correlacionados a esta tese. A estratégia metodológica utilizada no processo de desenvolvimento do MineraFórum envolveu uma série de etapas: 1) a escolha de uma técnica de mineração de textos adequada às necessidades da pesquisa; 2) verificação da existência de algum software de mineração de textos que auxiliasse o professor a analisar qualitativamente as contribuições em um fórum de discussão; 3) realização de estudos preliminares para avaliar a técnica de mineração escolhida; 4) definição dos indicadores de relevância das mensagens; elaboração de fórmulas para calcular a relevância das postagens; 5) construção do sistema; 6) integração do MineraFórum a três Ambientes Virtuais de Aprendizagem e, por último, 7) a realização de experimentos com a ferramenta.

Palavras-chave: mineração de textos; fórum de discussão; análise qualitativa; relevância temática; ambientes virtuais de aprendizagem.

ABSTRACT

This thesis presents the development, use and experimentation of the MineraFórum software. It is a resource that can help teachers in doing qualitative analyses of text contributions in discussion forums. This research included the use of text mining techniques with graphs. Message exchange in discussion forums are an important source of investigation for teachers. By analyzing students' posts, teachers can identify which learners wrote contributions that have concepts related to the debate theme, and which students did not succeed to do so. This strategy may also give teachers the necessary elements to motivate discussion of concepts relevant to the topic being debated. To accomplish the objectives of this study, a review of the literature was carried on topics such as: Distance Learning; Virtual Learning Environments; main concepts in Text Mining; and studies related to this thesis. The methodological strategy used in the development of MineraFórum followed these steps: 1) choosing a text mining technique suitable to the needs of the research; 2) checking whether there was software available to help teachers to do qualitative analysis of contributions in discussion forums; 3) doing preliminary studies to evaluate the selected mining technique; 4) defining indicators of relevance in the messages; elaborating formulas to calculate relevance in posts; 5) building the system; 6) integrating MineraFórum to three Virtual Learning Environments, and 7) carrying experiments with the tool.

Keywords: text mining; discussion forum; qualitative analysis; thematic relevance; virtual learning environments.

LISTA DE FIGURAS

FIGURA 1 - PLANO DE TRABALHO	23
FIGURA 2 - MOODLE	42
FIGURA 3 - INTERFACE PRINCIPAL DO ROODA	43
FIGURA 4 - INTERFACE PRINCIPAL DO ETC.....	46
FIGURA 5 - FÓRUM DE DISCUSSÃO DO AVA ROODA.....	48
FIGURA 6 - FÓRUM DE DISCUSSÃO DO ETC	49
FIGURA 7 - FÓRUM DE DISCUSSÃO DO MOODLE.....	49
FIGURA 8 - ARQUITETURA DE UM SISTEMA PARA MINERAÇÃO DE TEXTOS.....	56
FIGURA 9 - RESULTADOS DE UMA BUSCA NO SISTEMA FACTA	62
FIGURA 10 - INTERFACE DO EUREKHA.....	67
FIGURA 11 - INTERFACE DO BOW2NNBRs.....	67
FIGURA 12 - INTERFACE DO TAGCROWD	68
FIGURA 13 - INTERFACE DO SOBEK.....	69
FIGURA 14 - EXEMPLO DO AGENTE INTELIGENTE DENTRO DO FÓRUM DE DISCUSSÃO	80
FIGURA 15 - GRAFO GERADO A PARTIR DE UMA MENSAGEM DO FÓRUM DO EXPERIMENTO 1	88
FIGURA 16 - DIAGRAMA DE CASOS DE USO DA VERSÃO 1.0 DO MINERAFÓRUM	112
FIGURA 17 - DIAGRAMA DE CLASSES DA VERSÃO 1.0 DO MINERAFÓRUM.....	114
FIGURA 18 - ETAPAS DO MINERAFÓRUM.....	115
FIGURA 19 - TELA PRINCIPAL DA VERSÃO 1.0 DO MINERAFÓRUM.....	116
FIGURA 20 - TELA DA VERSÃO 1.0 PARA INDICAR A FREQUÊNCIA MÍNIMA NO TEXTO DE REFERÊNCIA	117
FIGURA 21 - TELA PRINCIPAL DA VERSÃO 1.0 COM A ABA “BASE DE CONCEITOS” SELECIONADA	118
FIGURA 22 - TELA PRINCIPAL DA VERSÃO 1.0 COM O MENU “ARQUIVO” SELECIONADO	120
FIGURA 23 - TELA PRINCIPAL DA VERSÃO 1.0 COM A ABA “MINERAR FÓRUM” SELECIONADA	121
FIGURA 24 - TELA DA VERSÃO 1.0 PARA INFORMAR O VALOR MÍNIMO DO CRT	122
FIGURA 25 - QUANTIDADE DE CONTRIBUIÇÕES RELEVANTES DO PRIMEIRO EXPERIMENTO	126
FIGURA 26 - QUANTIDADE DE CONTRIBUIÇÕES RELEVANTES DO PRIMEIRO GRUPO.....	126
FIGURA 27 - QUANTIDADE DE CONTRIBUIÇÕES RELEVANTES DO SEGUNDO GRUPO.....	127
FIGURA 28 - DIAGRAMA DE CASOS DE USO DA VERSÃO 2.0 DO MINERAFÓRUM	130
FIGURA 29 - DIAGRAMA DE CLASSES DA VERSÃO 2.0 DO MINERAFÓRUM.....	131
FIGURA 30 - TELA PRINCIPAL DA VERSÃO 2.0 COM O MENU “ARQUIVO” SELECIONADO	132
FIGURA 31 - TELA PRINCIPAL DA VERSÃO 2.0 COM A ABA “MINERAR FÓRUM” SELECIONADA	133
FIGURA 32 - TELA DA VERSÃO 2.0 PARA INFORMAR O VALOR MÍNIMO DA RELEVÂNCIA TEMÁTICA	134
FIGURA 33 - RELEVÂNCIA TEMÁTICA DAS MENSAGENS POSTADAS PELOS ALUNOS NO PRIMEIRO FÓRUM ANALISADO	137
FIGURA 34 - RELEVÂNCIA TEMÁTICA DAS MENSAGENS POSTADAS PELOS ALUNOS NO SEGUNDO FÓRUM ANALISADO	137

FIGURA 35 - RELEVÂNCIA TEMÁTICA DAS MENSAGENS POSTADAS PELOS ALUNOS NO TERCEIRO FÓRUM ANALISADO	138
FIGURA 36 - RELEVÂNCIA TEMÁTICA DAS MENSAGENS POSTADAS PELOS ALUNOS NO QUARTO FÓRUM ANALISADO	138
FIGURA 37 - RELEVÂNCIA TEMÁTICA DAS MENSAGENS POSTADAS PELOS ALUNOS NO QUARTO FÓRUM ANALISADO	139
FIGURA 38 - DIAGRAMA DE CASOS DE USO DA VERSÃO 3.0 DO MINERAFÓRUM	142
FIGURA 39 - DIAGRAMA DE CLASSES DA VERSÃO 3.0 DO MINERAFÓRUM.....	144
FIGURA 40 - TELA PRINCIPAL DA VERSÃO 3.0 COM O MENU “ARQUIVO” SELECIONADO	145
FIGURA 41 - TELA PRINCIPAL DA VERSÃO 3.0 COM A ABA “MINERAR FÓRUM” SELECIONADA	146
FIGURA 42 - VARIAÇÃO DO GRAU DE SIMILARIDADE NOS EXPERIMENTOS	157
FIGURA 43 - INTERFACE DO ROODA COM <i>LINK</i> PARA EXECUTAR O MINERAFÓRUM.....	162
FIGURA 44 - INTERFACE DO ETC COM <i>LINK</i> PARA EXECUTAR O MINERAFÓRUM	163
FIGURA 45 - INTERFACE DO MOODLE COM <i>LINK</i> PARA EXECUTAR O MINERAFÓRUM	164

LISTA DE QUADROS

QUADRO 1 - INTER-RELAÇÃO ENTRE ENTRADAS E SAÍDAS EM UM SISTEMA DE EAD.....	27
QUADRO 2 - COMPETÊNCIAS MÍNIMAS NECESSÁRIAS AOS PROFESSORES, ALUNOS E TUTORES...	35
QUADRO 3 - RELAÇÃO ENTRE A UNIDADE MÍNIMA DO TEXTO E ALGUMAS REPRESENTAÇÕES INTERMEDIÁRIAS	56
QUADRO 4 - REFERÊNCIAS BIBLIOGRÁFICAS QUE CITARAM OS INDICADORES DESCRITOS NO QUESTIONÁRIO	93
QUADRO 5 - REQUISITOS FUNCIONAIS DA VERSÃO 1.0 DO MINERAFÓRUM.....	111
QUADRO 6 - DESCRIÇÃO DOS CASOS DE USO DA VERSÃO 1.0 DO MINERAFÓRUM.....	113
QUADRO 7 - DESCRIÇÃO DAS CLASSES DA VERSÃO 1.0 DO MINERAFÓRUM.....	114
QUADRO 8 - REQUISITOS FUNCIONAIS DA VERSÃO 2.0 DO MINERAFÓRUM.....	129
QUADRO 9 - DESCRIÇÃO DOS CASOS DE USO DA VERSÃO 2.0 DO MINERAFÓRUM.....	131
QUADRO 10 - REQUISITOS FUNCIONAIS DA VERSÃO 3.0 DO MINERAFÓRUM.....	142
QUADRO 11 - DESCRIÇÃO DOS CASOS DE USO DA VERSÃO 3.0 DO MINERAFÓRUM.....	143
QUADRO 12 - NOMES E TEMAS DOS FÓRUNS UTILIZADOS NOS EXPERIMENTOS DA VERSÃO 3.0	150
QUADRO 13 - NÍVEL DE ENSINO, MODALIDADE E CURSO DOS ALUNOS QUE PARTICIPARAM DOS FÓRUNS	151

LISTA DE TABELAS

TABELA 1 - QUANTIDADE DE CONTRIBUIÇÕES RELEVANTES DO PRIMEIRO EXPERIMENTO	89
TABELA 2 - QUANTIDADE DE CONTRIBUIÇÕES RELEVANTES DO PRIMEIRO GRUPO.....	90
TABELA 3 - QUANTIDADE DE CONTRIBUIÇÕES RELEVANTES DO SEGUNDO GRUPO.....	90
TABELA 4 - RESPOSTAS INFORMADAS PELOS PROFESSORES	94
TABELA 5 - ANÁLISE DO MINERAFÓRUM E DO PROFESSOR 1 NOS FÓRUNS 1 A 5.....	152
TABELA 6 - ANÁLISE DO MINERAFÓRUM E DO PROFESSOR 1 NOS FÓRUNS 6 A 10.....	152
TABELA 7 - ANÁLISE DO MINERAFÓRUM E DO PROFESSOR 1 NOS FÓRUNS 11 A 15.....	153
TABELA 8 - ANÁLISE DO MINERAFÓRUM E DO PROFESSOR 1 NOS FÓRUNS 16 A 20.....	153
TABELA 9 - ANÁLISE DO MINERAFÓRUM E DO PROFESSOR 2 NOS FÓRUNS 1 A 5.....	154
TABELA 10 - ANÁLISE DO MINERAFÓRUM E DO PROFESSOR 2 NOS FÓRUNS 6 A 10.....	154
TABELA 11 - ANÁLISE DO MINERAFÓRUM E DO PROFESSOR 2 NOS FÓRUNS 11 A 15.....	155
TABELA 12 - ANÁLISE DO MINERAFÓRUM E DO PROFESSOR 2 NOS FÓRUNS 16 A 20.....	155
TABELA 13 - GRAU DE SIMILARIDADE ENTRE A MÉDIA DOS PROFESSORES E A MÉDIA DO MINERAFÓRUM.....	156
TABELA 14 - COMPARAÇÃO ENTRE SISTEMAS/SOLUÇÕES E O MINERAFÓRUM	160

LISTA DE SIGLAS

AVA	Ambiente Virtual de Aprendizagem
CDU	Caso de Uso
CRT	Coeficiente de Relevância Temática
EAD	Educação a Distância
MEC	Ministério da Educação
MT	Mineração de Textos
RF	Relevância da Mensagem
RT	Relevância Temática da Mensagem
SVM	Support Vector Machine
TIC	Tecnologias da Informação e Comunicação
UML	Unified Modeling Language

SUMÁRIO

1. INTRODUÇÃO	17
2. CONTEXTUALIZAÇÃO DA PESQUISA	19
2.1 PLANO DE TRABALHO	22
3. REVISÃO DA LITERATURA	24
3.1 EDUCAÇÃO A DISTÂNCIA	24
3.1.1 PAPÉIS DOS ATORES ENVOLVIDOS NA EDUCAÇÃO A DISTÂNCIA.....	30
3.1.2 INTERAÇÃO ENTRE OS ATORES ENVOLVIDOS NA EDUCAÇÃO A DISTÂNCIA	36
3.2 AMBIENTES VIRTUAIS DE APRENDIZAGEM	37
3.2.1 FERRAMENTAS DISPONÍVEIS NOS AMBIENTES VIRTUAIS DE APRENDIZAGEM	41
3.2.2 FÓRUNS DE DISCUSSÃO	47
3.3 USO PEDAGÓGICO DE FÓRUNS DE DISCUSSÃO.....	50
3.4 MINERAÇÃO DE TEXTOS.....	54
3.4.1 PRINCÍPIOS BÁSICOS	55
3.4.2 UTILIZAÇÃO DE INFORMAÇÕES ADICIONAIS PARA AUXILIAR A MINERAÇÃO DE TEXTOS	60
3.4.3 ÁREAS DE APLICAÇÃO	61
3.4.4 APLICAÇÕES NA EDUCAÇÃO	63
3.5 SOFTWARES PARA MINERAÇÃO DE TEXTOS	65
3.6 TRABALHOS CORRELACIONADOS	71
3.6.1 PESQUISAS SOBRE ANÁLISE DE FÓRUNS DE DISCUSSÃO.....	71
3.6.2 PESQUISAS SOBRE ANÁLISE DE FÓRUNS DE DISCUSSÃO REALIZADAS COM TÉCNICAS DE MINERAÇÃO DE TEXTOS	78
4. METODOLOGIA.....	83
4.1 ESCOLHA DE UMA TÉCNICA DE MINERAÇÃO DE TEXTOS.....	84
4.2 SELEÇÃO DO SOFTWARE PARA MINERAÇÃO DE TEXTOS.....	85
4.3 ESTUDOS PRELIMINARES	86

4.4 DEFINIÇÃO DOS INDICADORES DE RELEVÂNCIA A SEREM DIAGNOSTICADOS PELO MINERAFÓRUM	91
4.5 CONSTRUÇÃO DO MINERAFÓRUM.....	95
4.5.1 CÁLCULO DA RELEVÂNCIA DAS MENSAGENS	98
5. MINERAFÓRUM	105
5.1 PROCESSO DE DESENVOLVIMENTO	105
5.2 VERSÃO 1.0	107
5.2.1 ANÁLISE DO SISTEMA.....	108
5.2.2 INTERFACE DO SISTEMA.....	116
5.2.3 ALGORITMO PARA CALCULAR A RELEVÂNCIA TEMÁTICA DE UMA MENSAGEM.....	122
5.2.4 EXPERIMENTOS COM A VERSÃO 1.0	125
5.2.5 ANÁLISE E DISCUSSÃO DOS RESULTADOS DA VERSÃO 1.0	127
5.3 VERSÃO 2.0	128
5.3.1 MODIFICAÇÕES NO SISTEMA.....	128
5.3.2 EXPERIMENTOS COM A VERSÃO 2.0	136
5.3.3 ANÁLISE E DISCUSSÃO DOS RESULTADOS DA VERSÃO 2.0.....	139
5.4 VERSÃO 3.0	140
5.4.1 DESCRIÇÃO DAS ALTERAÇÕES	140
5.4.2 EXPERIMENTOS COM A VERSÃO 3.0	149
5.4.3 ANÁLISE E DISCUSSÃO DOS RESULTADOS DA VERSÃO 3.0.....	157
5.5 COMPARAÇÃO ENTRE O MINERAFÓRUM E OUTROS SISTEMAS/SOLUÇÕES	159
5.6 INTEGRAÇÃO DO MINERAFÓRUM AOS AMBIENTES VIRTUAIS DE APRENDIZAGEM.....	161
5.6.1 ROODA.....	161
5.6.2 ETC	162
5.6.3 MOODLE.....	163
6. CONSIDERAÇÕES FINAIS.....	165
6.1 CAMINHOS PERCORRIDOS.....	165
6.2 CONTRIBUIÇÕES DA TESE	167

6.3 DESAFIOS E LIMITAÇÕES.....	167
6.4 TRABALHOS FUTUROS	169
6.5 A TÍTULO DE CONCLUSÃO	169
REFERÊNCIAS BIBLIOGRÁFICAS	171
APÊNDICE A - PRODUÇÃO CIENTÍFICA.....	184
APÊNDICE B – QUESTIONÁRIO APLICADO COM PROFESSORES PARA DIAGNÓSTICO DE INDICADORES DE RELEVÂNCIA.....	186
APÊNDICE C - TERMO DE CONSENTIMENTO PARA A ANÁLISE DAS MENSAGENS DOS FÓRUNS	188
APÊNDICE D – TUTORIAL DA VERSÃO 3.0 DO MINERAFÓRUM.....	189
APÊNDICE E – INSTALAÇÃO DA VERSÃO 3.0 DO MINERAFÓRUM.....	200

1. INTRODUÇÃO

Na sociedade atual, devido aos avanços tecnológicos das telecomunicações e da informática, a comunicação entre as pessoas tornou-se mais fácil e rápida. A *World Wide Web* facilitou a disseminação das informações. Com estas mudanças, algumas áreas do conhecimento humano estão repensando seus paradigmas, e também se expandindo. A educação a distância (EAD) é uma delas e tem usufruído da evolução das tecnologias da informação e comunicação para melhorar os seus recursos de apoio. A utilização da *Web*, multimídia, videoconferência, realidade virtual, softwares educacionais, Ambientes Virtuais de Aprendizagem (AVAs), são exemplos de ferramentas que podem auxiliar a EAD.

Os Ambientes Virtuais de Aprendizagem são utilizados através da *Web* (SCHLEMMER, 2002). Eles oferecem ferramentas para gerenciar os materiais didáticos, para analisar o progresso dos discentes, e para facilitar a comunicação entre as pessoas (ONRUBIA et al., 2010). Estes ambientes possuem recursos que favorecem o compartilhamento de informações, a interação entre o docente e os discentes, e entre eles mesmos.

Com estes ambientes os professores e alunos possuem a sua disposição diversas ferramentas que podem auxiliar na organização dos materiais produzidos e facilitar a interação entre os sujeitos participantes (BEHAR, 2007a; FRANCISCATO et. al, 2008). Entre estas, podem ser citadas as seguintes: espaço para disponibilização de material didático, espaço pessoal do aluno, bate-papo, fórum de discussão, envio de mensagens, entre outras.

Observando os recursos citados anteriormente que fazem parte dos AVAs, a presente pesquisa tem como foco o fórum de discussão. Com a troca de mensagens escritas registradas neste tipo de recurso, docentes e discentes podem se comunicar uns com os outros, obtendo um espaço importante para interação. Para o professor, a análise das mensagens postadas representa um *feedback* valioso dos discentes. A partir dos textos produzidos, o docente pode ter acesso a diferentes relatos e perceber alguns aspectos, como: quais alunos estão com dúvidas, quais estão contribuindo e discutindo melhor os conceitos abordados pelo tema da discussão, entre outros.

O envolvimento/participação em fóruns de discussão é uma parte importante das atividades dos alunos que estudam através da modalidade de educação a distância pois, em

muitos casos, os fóruns são considerados no processo de avaliação dos discentes. A análise deste tipo de ferramenta permite que o professor possa diagnosticar informações sobre os alunos. No entanto, se o docente possuir uma grande quantidade de alunos, o tempo necessário para que ele consiga analisar as discussões será grande. Um recurso que permita fazer, de forma automatizada e rápida, uma análise qualitativa das mensagens postadas em um fórum de discussão, pode auxiliar o professor, no sentido de otimizar o seu tempo. Isto permitirá que o docente identifique os alunos que estão (e aqueles que não estão) discutindo os conceitos do tema em debate. Desta forma, o docente poderá dedicar um tempo maior para descobrir o motivo pelo qual alguns discentes não discutiram os conceitos referentes ao tema. Caso o professor identifique alunos com dificuldade para compreender os conceitos importantes, ele poderá oferecer ajuda. Se for necessário, o docente também pode motivar os discentes a participar mais do debate.

Para auxiliar na análise da relevância das contribuições produzidas pelos alunos em um fórum de discussão, este estudo propôs a construção, utilização e experimentação de um recurso denominado MineraFórum.

O MineraFórum utiliza técnicas de mineração de textos para analisar as mensagens de um fórum de discussão. O objetivo da ferramenta é auxiliar o docente a verificar de forma mais otimizada os conteúdos das contribuições textuais redigidas pelos alunos. A partir dos resultados fornecidos pelo MineraFórum, o professor pode direcionar imediatamente seu auxílio para os alunos que não discutiram os conceitos relativos ao tema em debate.

A fim de atingir o objetivo traçado, este estudo foi organizado em seis capítulos. O capítulo dois descreve a contextualização da pesquisa e o plano de trabalho da tese. O capítulo três apresenta uma revisão da literatura sobre os seguintes temas: educação a distância, Ambientes Virtuais de Aprendizagem, papéis dos atores na educação a distância, uso pedagógico de fóruns de discussão, mineração de textos, alguns softwares de mineração de textos que foram pesquisados, e trabalhos correlacionados. O capítulo quatro aborda a metodologia utilizada nesta pesquisa. O capítulo cinco apresenta os resultados da presente tese, neste caso, o desenvolvimento do MineraFórum, as diferentes versões do software, a integração com os Ambientes Virtuais de Aprendizagem. Também são descritos os experimentos realizados, a análise e discussão dos resultados. O capítulo seis relata as considerações finais desta pesquisa.

2. CONTEXTUALIZAÇÃO DA PESQUISA

De acordo com Coll e Monereo (2010a), a Sociedade da Informação surgiu graças à interligação dos computadores e à internet. Segundo os autores, ela pode ser definida como um novo estágio de desenvolvimento das sociedades humanas. Sob o aspecto das tecnologias da informação e comunicação, a Sociedade da Informação é caracterizada pela capacidade das pessoas em obter e compartilhar informações rapidamente, de qualquer lugar e na forma preferida.

Observando as transformações no mundo atual, é necessário que a educação promova o desenvolvimento de competências necessárias ao indivíduo deste novo milênio. De acordo com Coll e Monereo (2010a), a Sociedade da Informação apresenta características que precisam ser consideradas no contexto educacional. Para Lévy (1999), a educação deve considerar as mudanças contemporâneas da relação com o saber.

Na sociedade atual, é muito grande a facilidade e a velocidade com que a informação torna-se disponível às pessoas. No contexto educacional, a busca de dados por parte dos alunos depende cada vez menos do professor, pois o acesso às tecnologias permite encontrá-los rapidamente. Desta forma, é importante que o professor auxilie o aluno a selecionar os elementos que são relevantes, ajudando na interpretação e contextualização dos mesmos.

A evolução das tecnologias da informação e comunicação (TIC) proporcionou um enorme aumento na velocidade com a qual as pessoas realizam determinadas atividades, e também nas formas de interação. Para a educação, o avanço das TIC pode representar um conjunto de ferramentas auxiliares ao trabalho do professor e dos alunos. O docente é quem deve definir, planejar, direcionar o uso pedagógico das tecnologias, mantendo-as inter-relacionadas com o processo educativo.

O progresso das TIC também propiciou mudanças na educação a distância (EAD), permitindo que a mesma pudesse ser mais difundida, facilitando a comunicação entre os sujeitos participantes, possibilitando a construção de recursos com multimídia, entre outros. A utilização da *Web* na EAD também oferece benefícios, pois permite que os alunos possam acessar materiais didáticos e utilizar recursos de comunicação a qualquer momento que desejarem, e em qualquer lugar que disponibilize acesso a *Web*.

Este estudo defende a ideia de que a EAD, mediada através das TIC e realizada através da *Web*, pode ser uma das soluções para auxiliar a educação a desenvolver nos alunos algumas características necessárias para este novo milênio. Entre elas podem ser destacadas: incentivo a autonomia; interação/comunicação entre as pessoas; desenvolvimento da aprendizagem colaborativa/cooperativa, entre outras.

A EAD pode utilizar recursos tecnológicos através da *Web* como, por exemplo, os Ambientes Virtuais de Aprendizagem (AVAs). Estes possuem ferramentas que facilitam o compartilhamento de informações e a comunicação entre as pessoas. Estes ambientes também podem oferecer ao professor, opções para gerenciar o material didático e acompanhar o progresso dos alunos.

Estes recursos possibilitam a comunicação síncrona e assíncrona entre os participantes como, por exemplo, o bate-papo e o fórum de discussão. No bate-papo, todos participam ao mesmo tempo. O fórum funciona de forma assíncrona, ou seja, em tempos diferentes e, por este motivo, necessita de um tempo maior para que as discussões possam ocorrer. Este pode se tornar um espaço bastante interessante para a comunicação entre os discentes e entre eles e o docente, de acordo com a metodologia/estratégia utilizada pelo professor. As mensagens postadas no fórum representam um valioso recurso, para que este possa analisar o *feedback* dos alunos em relação ao tópico proposto para o debate. Ao investigar o conteúdo das mensagens postadas, o docente pode observar como os discentes estão trabalhando os conceitos relativos ao tema da discussão. Atualmente, nos AVAs esta análise é realizada de forma manual pelo professor.

Em um fórum, dependendo do tema proposto, do perfil dos alunos, do interesse despertado pelo tema, podem ser disparadas centenas de contribuições. Alguns textos redigidos possuem relevância com relação ao tema tratado e, outros não. Observa-se que as contribuições significativas abordam conceitos importantes relativos ao assunto em discussão, onde estes conceitos são citados e relacionados entre si.

Esta pesquisa se justifica devido à importância que existe em avaliar qualitativamente as contribuições registradas pelos alunos em um fórum de discussão. A análise quantitativa é mais simples de ser executada, como por exemplo, a taxa de participação dos estudantes e a quantidade de mensagens das discussões (MAZZOLINI, 2007). Para o docente a análise quantitativa é importante, mas neste estudo, entende-se que a qualitativa é fundamental para que o mesmo possa realizar um acompanhamento detalhado do processo de construção do

conhecimento dos discentes. Além disso, ela permite que o professor possa verificar a real contribuição de cada aluno e observar a própria participação e discussão.

Foi construído, como parte integrante deste trabalho, um recurso que tem potencial para auxiliar o professor a realizar rapidamente uma análise qualitativa das mensagens produzidas pelos alunos em um fórum de discussão. Este foi denominado MineraFórum e utiliza técnicas de mineração de textos.

A mineração de textos é uma área da Ciência da Computação cujo objetivo é identificar informações novas, ou desconhecidas, através da extração automática das mesmas a partir de documentos escritos (GUPTA e LEHAL, 2009). Esta área também é conhecida como análise inteligente de textos, mineração de dados de textos, ou descoberta de conhecimento em textos, refere-se ao processo de extrair informações interessantes e conhecimento a partir de textos. Para analisar os textos são utilizadas técnicas de recuperação de informação, mineração de dados, aprendizado de máquina, linguística computacional e estatística. Nos procedimentos envolvidos há semelhança com a mineração de dados, cujo objetivo é descobrir padrões interessantes a partir de grandes bases de dados. No entanto, as ferramentas de mineração de dados são projetadas para trabalhar com dados estruturados provenientes de bancos de dados, e as ferramentas de mineração de textos trabalham com informações não estruturadas ou semi-estruturadas, como: documentos de textos, mensagens de correio eletrônico, páginas “html”, entre outros (GUPTA e LEHAL, 2009).

Logo, entende-se que, com a aplicação da mineração de textos nas contribuições textuais produzidas pelos alunos, o professor pode identificar quais conceitos relevantes ao tema da discussão estão sendo trabalhados pelos discentes.

Desta forma, levanta-se a seguinte questão de pesquisa:

Como um recurso de mineração de textos, incorporado a um ambiente virtual de aprendizagem, pode auxiliar na análise qualitativa das contribuições em um fórum de discussão ?

O objetivo geral proposto para esta pesquisa é identificado abaixo:

Analisar de que forma o MineraFórum poderia contribuir na análise qualitativa das mensagens postadas em um fórum de discussão.

A partir deste objetivo geral, são derivados os seguintes objetivos específicos:

- a) Definir indicadores que possam ser utilizados para identificar contribuições relevantes em um fórum de discussão.
- b) Construir o MineraFórum para auxiliar no acompanhamento qualitativo e quantitativo das contribuições em um fórum de discussão.
- c) Integrar o MineraFórum aos Ambientes Virtuais de Aprendizagem Rooda, ETC e Moodle.

2.1 PLANO DE TRABALHO

O plano de trabalho desta tese, apresentado na figura 1, aborda o estudo do referencial teórico envolvendo temas como a EAD e a Mineração de Texto, entre outros. O MineraFórum foi projetado, implementado e, por fim, analisado através de vários experimentos, a fim de responder ao problema de pesquisa.

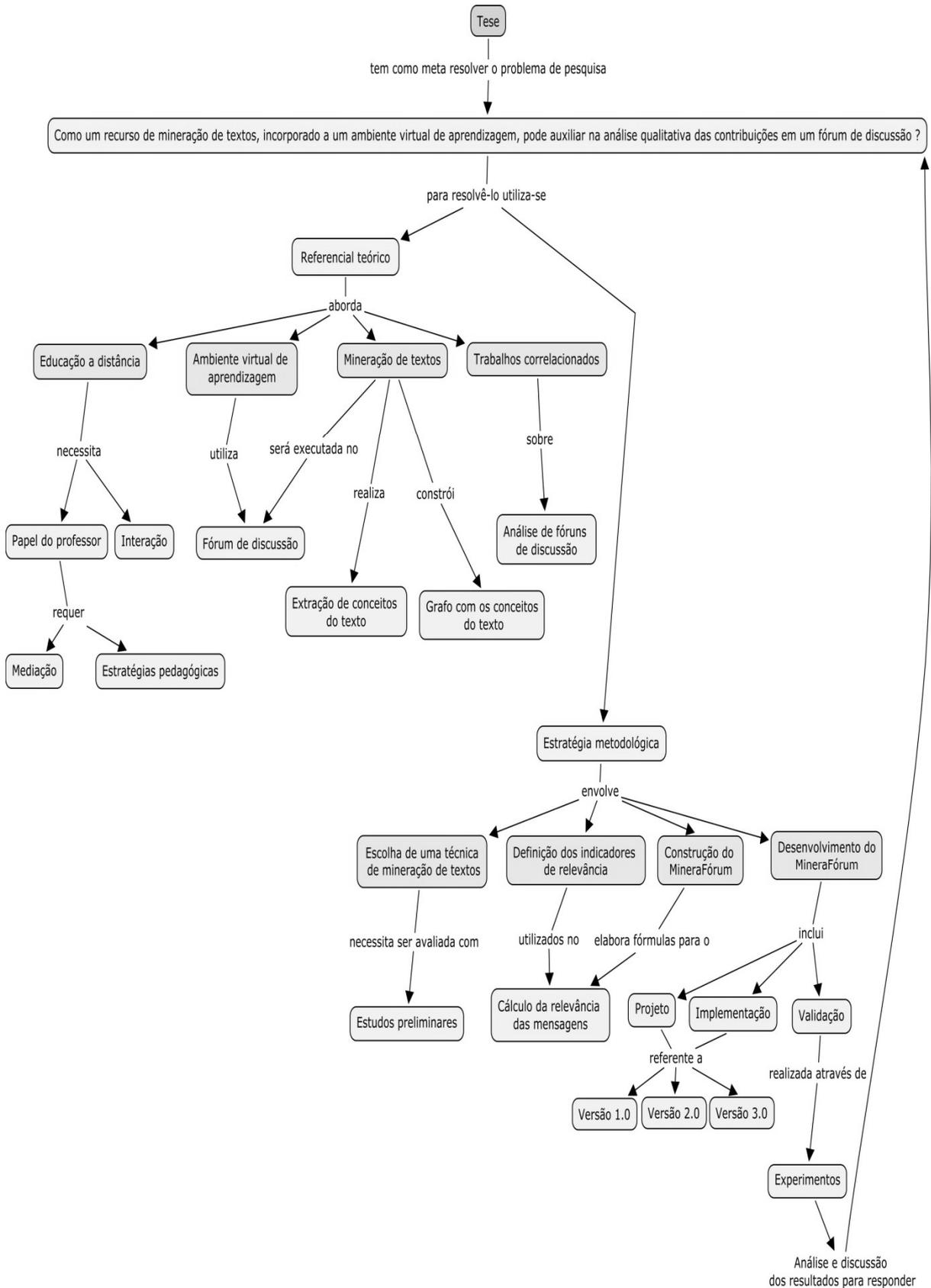


Figura 1 - Plano de trabalho
Fonte: o autor

3. REVISÃO DA LITERATURA

Este capítulo apresenta a base teórica utilizada nesta tese. Em relação a educação a distância foram pesquisadas algumas obras de Michael G. Moore, César Coll, Fredric Litto e Marcos Formiga, Maria Luiza Belloni, Rena M. Palloff e Keith Pratt, apresentadas na seção 3.1. A seção 3.1.1 descreve os papéis dos atores envolvidos no processo de educação a distância, a partir das obras de Maria Luiza Belloni, Michael G. Moore e Rena M. Palloff e Keith Pratt. O estudo sobre Ambientes Virtuais de Aprendizagem, na seção 3.2, pautou-se nas obras dos pesquisadores Eliane Schlemmer, Patricia Alejandra Behar, Pierre Dillenbourg, Javier Onrubia, Caroline Ramsey. A seção 3.2.1 apresenta algumas ferramentas disponíveis nos AVAs. As características dos fóruns de discussão são relatadas na seção 3.2.2. A seção 3.3 apresenta o uso pedagógico de fóruns de discussão. O aporte teórico sobre mineração de textos, apresentado na seção 3.4, foi estudado a partir das obras de Ronen Feldman e James Sanger, Michael W. Berry e colaboradores, Min Song e Yi-fang Brook Wu e colaboradores. A seção 3.5 apresenta alguns softwares utilizados na área de mineração de textos. A seção 3.6 finaliza o capítulo apresentando os trabalhos correlacionados a esta pesquisa.

3.1 EDUCAÇÃO A DISTÂNCIA

A evolução das TIC proporcionou a mudança na execução de diversas atividades realizadas pelo ser humano. No cenário educacional, as TIC causaram modificações nos papéis do professor e do aluno, nas formas de interação, na organização do tempo e do espaço, na forma de apresentação dos conteúdos, e o acesso aos recursos (Coll, 2010a). Destaca-se que o potencial das TIC pode ser utilizado com maior ou menor profundidade no cenário educacional. Para Coll (2010b), são as finalidades desejadas com a adoção das TIC e o uso efetivo das mesmas por docentes e discentes, que determinam seu maior ou menor impacto nas práticas educacionais e na transformação do processo de ensino-aprendizagem.

Os avanços dos recursos tecnológicos da informática e das telecomunicações também causaram mudanças na educação a distância (EAD), permitindo que a mesma pudesse ser mais difundida. A aparição da *Web* e sua posterior evolução, com a apresentação de recursos

multimídias, ferramentas de comunicação, entre outros, proporcionou um valioso caminho para a utilização da educação a distância.

Destaca-se que a realização da EAD com a utilização da *Web* permite uma quebra nas barreiras de tempo e espaço. Os alunos podem acessar materiais didáticos e utilizar recursos de comunicação a qualquer momento que desejarem, e em qualquer lugar que disponibilize acesso a *Web*.

De acordo com Litto (2009), a segunda geração da *Web* (*Web 2.0*) oferece possibilidades para a colaboração e o compartilhamento de informações entre docentes e discentes. A *Web 2.0* utiliza ferramentas úteis para a aprendizagem, como agendas *online*, ambientes para colaboração, recursos para gerenciamento de projetos, sistemas multimídia.

Segundo Formiga (2009), a educação a distância está estreitamente relacionada às TIC devido ao seu dinamismo. O uso crescente das tecnologias favorece a inovação na EAD. A sociedade da informação e do conhecimento reflete-se na EAD pela apropriação dos conceitos e inovações. Para Litto (2009) há um horizonte de enormes expectativas quando se discute novas formas de ensinar e aprender.

Moore e Kearsley (2008) definem a educação a distância como uma área de natureza multidimensional.

Educação a distância é o aprendizado planejado que ocorre normalmente em um lugar diferente do local do ensino, exigindo técnicas especiais de criação do curso e de instrução, comunicação por meio de várias tecnologias e disposições organizacionais e administrativas especiais (MOORE e KEARSLEY, 2008, p. 2).

A EAD não é uma modalidade educacional nova. Moore e Kearsley (2008) citam as cinco gerações pelas quais a EAD passou, na seguinte ordem: correspondência, transmissão por rádio e televisão, universidades abertas, teleconferência, internet/*web*.

De acordo com Moore e Kearsley (2008) cada uma destas gerações apresentou características marcantes. A primeira utilizou a correspondência e possibilitou o fundamento para a educação individualizada a distância. A segunda usou a transmissão por rádio e televisão, agregando recursos de comunicação oral e visual à apresentação das informações aos alunos. A terceira, marcada pelas universidades abertas, surgiu de experiências norte-americanas, unindo recursos de áudio/vídeo e correspondência com orientação face a face. A quarta foi baseada na teleconferência, com recursos de áudio/vídeo e computador para oferecer interação em tempo real de alunos com alunos e professores. A quinta geração, da

internet/web, aborda a teoria construtivista, aprendizado em colaboração, e realiza a integração de texto, áudio e vídeo para efetuar a comunicação.

Os aspectos importantes que caracterizam a EAD dizem respeito à separação tanto espacial, quanto temporal, entre professor e aluno. Atualmente, ao se falar em EAD, entende-se que os principais mecanismos a serem utilizados para interação entre professor e aluno envolvem tecnologias de informação e comunicação. Uma das principais características da EAD é proporcionar a diferentes pessoas, de lugares distintos, o acesso a educação a qualquer momento.

O Decreto 5.622, de 19 de dezembro de 2005¹, em seu artigo 1º define a educação a distância.

Para os fins deste Decreto, caracteriza-se a educação a distância 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, 2005).

Nas instituições de ensino, a implantação e utilização da EAD devem observar a interligação e a qualidade de alguns processos. Segundo Moore e Kearsley (2008), os principais componentes de um sistema de EAD são: uma fonte de conhecimento; um subsistema para estruturar esse conhecimento em materiais e atividades para os alunos; um subsistema para transmissão do material; professores que interajam com os alunos; alunos em seus ambientes distintos; um subsistema para controlar e avaliar os resultados, possibilitando intervenções; uma organização com uma política e estrutura administrativa. De acordo com os Referenciais de Qualidade para Educação Superior a Distância do MEC (BRASIL, 2007), a utilização da EAD deve considerar os aspectos pedagógicos, os recursos humanos, e a infraestrutura.

Moore e Kearsley (2008) examinam as inter-relações entre aspectos considerados como variáveis de entrada e os resultados em um sistema de EAD (quadro 1). Para os autores, as entradas afetam de alguma maneira os resultados.

¹ Disponível em http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2005/Decreto/D5622.htm

Quadro 1 - Inter-relação entre entradas e saídas em um sistema de EAD

Entradas em um sistema de EAD	Resultados em um sistema de EAD
<ul style="list-style-type: none"> • Características individuais do aluno. • Experiência do professor de EAD. • Entendimento dos alunos por parte da equipe administrativa. • Qualidade na elaboração e produção do curso. • Custo de criação e produção do curso. • Tecnologia escolhida. • Acessibilidade dos serviços de apoio. • Frequência e qualidade dos dados de avaliação. • Investimento financeiro. 	<ul style="list-style-type: none"> • Índices de satisfação do aluno. • Resultados do aluno. • Índice de finalização do curso. • Número total de matrículas. • Avaliações da qualidade. • Resultados da certificação. • Mensalidades e outras receitas. • Reputação e rotatividade dos colaboradores.

Fonte: Moore e Kearsley (2008, p. 21)

Para Moore e Kearsley (2008), na EAD a estrutura dos cursos é importante, mas também é essencial que haja colaboração entre docentes e discentes. A EAD deve considerar as características individuais e a independência dos alunos. A autonomia do estudante é fundamental, pois significa que as pessoas possuem capacidades diferentes para tomar decisões sobre seu próprio aprendizado. Esta autonomia envolve a capacidade para desenvolver um plano de ação, para pesquisar recursos para seu estudo, e a capacidade para decidir sozinho quando o avanço foi satisfatório.

Segundo Belloni (2008), o princípio norteador das ações de EAD deve ser um processo de ensino-aprendizagem centrado no aluno. Para a autora, é importante conhecer as características sócio-culturais, os conhecimentos, as experiências, as necessidades e as expectativas dos alunos. Estes aspectos devem ser integrados nas metodologias, estratégias, e materiais de ensino, criando possibilidades de auto-aprendizagem.

Para Palloff e Pratt (2002), o processo de aprendizagem a distância por computador traz desafios, originados da relação do usuário com o computador, e da relação entre os alunos uns com os outros, e entre eles e o professor. Algumas questões que devem ser tratadas são: a análise da participação do aluno, o envolvimento do aluno com o tema de estudo, a disposição dos sujeitos, o estado emocional do aluno, a personalidade desenvolvida por meio da interação. De forma semelhante a estes desafios, Moore e Kearsley (2008) citam que na EAD, devido a distância entre docentes e discentes, é necessário haver mecanismos de *feedback* dos alunos, para que o professor possa perceber as ações/reações dos mesmos. Os autores também citam que a tecnologia utilizada como mediadora da EAD também deve ser dominada.

Os professores e alunos possuem comportamentos distintos na sala de aula tradicional e na sala de aula no ciberespaço² (PALLOFF e PRATT, 2002; TELES, 2009; MOORE e KEARSLEY, 2008). Na aprendizagem a distância por computador, existem meios para que os alunos explorem os conteúdos de forma colaborativa ou para que busquem suas necessidades. Neste processo, é muito importante destacar as interações que ocorrem entre os discentes e entre eles e o docente, e a colaboração resultante destas interações. Para os autores, é por meio destas interações que o conhecimento é produzido.

A aprendizagem no ambiente de educação a distância não pode ser passiva. Se os alunos não entram em sua sala de aula online – se não enviam uma colaboração para a discussão -, o professor não terá como saber que eles estiveram presentes. Assim, os estudantes não são apenas responsáveis pela sua conexão, mas também devem contribuir com o processo de aprendizagem por meio do envio de mensagens com seus pensamentos e suas ideias (PALLOFF e PRATT, 2002, p. 28).

De acordo com Palloff e Pratt (2002), na aprendizagem a distância por computador, é interessante que o professor solicite a execução de tarefas colaborativas, atue como facilitador em discussões ativas, motive o desenvolvimento do pensamento crítico, e promova a capacidade de pesquisa.

Segundo Harasim (2006), a aprendizagem colaborativa pode ser o fator mais relevante para a aprendizagem através da *Web*, pois este princípio motiva o desenvolvimento de habilidades sócio-afetivas e cognitivas. O ambiente *online* assíncrono da *Web* permite e requer aprendizagem colaborativa, pois a colaboração prevê a motivação e a união social de uma comunidade que envolve alunos e incentiva-os a participar e contribuir para alcançar objetivos comuns.

Para Gunawardena e McIsaac (2004), os avanços tecnológicos conduzem a um aumento na popularidade da EAD, mas é o bom planejamento da sua condução que permite ao aluno interagir com a tecnologia na construção do conhecimento. É a interação efetiva de professor, aluno, e tecnologia que concebe à educação a distância sua importância dentro da comunidade educativa.

É importante destacar que são as concepções pedagógicas que definirão a abordagem e o uso dos recursos proporcionados pela EAD. A construção do material didático, os papéis do

² O ciberespaço é o meio de comunicação que surge da interconexão mundial dos computadores (LÉVY, 1999, p. 17).

professor, do tutor e do aluno são delineados pelas compreensões estabelecidas no processo de ensino-aprendizagem.

Para Lévy (1999), os sistemas de educação devem considerar as modificações contemporâneas da relação com o saber. Estas mudanças referem-se a alguns aspectos como: a velocidade com que surgem e renovam os saberes; a nova forma de conceber o trabalho, significando cada vez mais, aprender, transmitir saberes, e produzir conhecimentos; e as características do ciberespaço, suportando tecnologias intelectuais que ampliam e modificam diversas funções cognitivas humanas.

Segundo Lévy (1999), o ciberespaço, com suas comunidades virtuais, seus recursos de textos, sons, imagens, suas simulações interativas, será o mediador da inteligência coletiva da humanidade. Com este novo apoio à informação e comunicação, surgem diferentes gêneros de conhecimento, novas formas de orientar o saber, novos papéis na produção e tratamento dos conhecimentos. Os sistemas de educação precisam considerar estes aspectos.

Analisando as mudanças necessárias à educação na era do ciberespaço, Ramal (2000) cita que é necessário desenvolver nos alunos processos mais abrangentes.

Educar será, portanto, desenvolver processos abrangentes, segundo critérios como consistência, previsibilidade, motivação, envolvimento, performance, capacidade de articular conhecimentos, de comunicar-se e estabelecer relações. Isso ajudará a preparar o cidadão da era do ciberespaço: como a matéria-prima da produção será a informação, e os conteúdos da formação inicial se tornarão rapidamente obsoletos, ele deverá ser um profissional capaz de aprender sempre; um ser consciente e crítico, que dialogue com as diferentes culturas e os diversos saberes, que saiba trabalhar de forma cooperativa e que seja flexível, empreendedor e criativo para administrar sua carreira e sua vida pessoal, social e política (RAMAL, 2000, p. 2-3).

Com relação à educação, Schlemmer (2002) enfatiza que a transformação a ser realizada é passar da cultura do ensino, com foco na concepção empirista, para a cultura da aprendizagem, com foco na concepção interacionista.

Na cultura da aprendizagem, o foco do processo educacional está na interação, na construção do conhecimento, no desenvolvimento de competências e habilidades, na aprendizagem. Há respeito ao ritmo de desenvolvimento do sujeito, pois se acredita que a aprendizagem é um processo coletivo, significado individualmente, estando relacionada às construções/significações anteriores do sujeito (SCHLEMMER, 2005, p. 136).

Para Kellner (2006), os indivíduos devem ser ajudados a desenvolver as competências para compreender, criticar, e transformar as suas condições sociais e culturais, adquirindo a

capacidade de serem sujeitos criativos e transformadores. Isso requer o desenvolvimento de pensamento crítico e de reflexão. Sujeitos ativos e engajados são construídos a partir da interação social com outros, e também com métodos e técnicas. As habilidades sociais e as capacidades individuais de comunicação, criatividade e ação devem ser parte dos múltiplos objetivos necessários a uma reconstrução da educação.

A partir das mudanças contemporâneas indicadas por Lévy (1999) e Ramal (2000), e as competências necessárias aos indivíduos citadas por Kellner (2006), verifica-se a necessidade de repensar os sistemas de educação. Considerando a evolução e papel das TIC indicadas por Coll e Monereo (2010a), Coll et al. (2010b); os aspectos relacionados a EAD citados por Moore e Kearsley (2008), Belloni (2008); as mudanças nos papéis do aluno e do professor apresentados por Palloff e Pratt (2002, 2004), Moore e Kearsley (2008), Teles (2009); observa-se que a EAD, mediada através das TIC (tecnologias da informação e comunicação), e realizada através da *Web*, pode ser uma das soluções para auxiliar a educação nas transformações necessárias a este novo milênio. Entre as características que podem ser promovidas a partir desta união, destacam-se:

- a) Incentivo a autonomia do aluno.
- b) Atenção às características individuais de cada sujeito.
- c) Favorecimento da interação entre as pessoas.
- d) Redução das barreiras de tempo e espaço.
- e) Desenvolvimento da aprendizagem colaborativa.

3.1.1 PAPÉIS DOS ATORES ENVOLVIDOS NA EDUCAÇÃO A DISTÂNCIA

Esta seção discute os papéis desempenhados pelos atores envolvidos na EAD, relacionados ao contexto deste trabalho, os professores, alunos e tutores. A função destes possui algumas diferenças em relação à educação em sala de aula tradicional. Estas distinções surgem devido à separação tanto espacial, quanto temporal, entre as pessoas envolvidas. A utilização e domínio das tecnologias empregadas na EAD também é um aspecto a ser considerado.

De acordo com Belloni (2008), surgem papéis para o professor na educação a distância, como:

- a) Oferecer auxílio, quando o aluno sente necessidade;
- b) Elaborar materiais didáticos;
- c) Atualizar-se constantemente;
- d) Trabalhar com um diálogo dinâmico;
- e) Promover a construção coletiva do conhecimento;
- f) Realizar trabalhos em equipes interdisciplinares e complexas;
- g) Refletir sobre sua prática pedagógica;
- h) Orientar e participar da pesquisa de seus alunos.

Na EAD, o papel desempenhado pelo professor no processo de ensino-aprendizagem possui algumas características marcantes. Moore e Kearsley (2008) citam alguns fatores relacionados aos professores, que diferenciam a educação a distância da educação em sala de aula tradicional. Entre estes fatores, o autor observa a importância do professor utilizar mecanismos para capturar o feedback dos alunos, e o desafio para alguns professores na utilização da tecnologia. Para Palloff e Pratt (2004), se um aluno que utiliza recursos digitais para interagir com o professor perceber a falta de atividades, de participação, isto pode fazer com o que o mesmo fique desmotivado.

Segundo Teles (2009), há mudança na postura do professor e do aluno na sala de aula virtual, apontando para indicadores do papel do docente no processo de ensino-aprendizagem que ocorre de maneira *online*. Nas pesquisas realizadas por Teles (2009), ele analisou as quatro funções do professor *online* propostas por Berge (1995, 1996), que são: pedagógica, social, gerencial, e suporte técnico. A função pedagógica envolve todas as estratégias utilizadas para apoiar o processo de aprendizagem do aluno ou grupo de alunos. A função gerencial envolve todas as atividades executadas para que o curso seja realizado de forma eficiente, a nível administrativo. A função de suporte técnico envolve a escolha dos softwares e a ajuda para utilização dos mesmos. Com relação à função social, Teles (2009) expõe que a primeira tarefa do professor *online* é criar um ambiente de comunicação fácil e confortável. Neste espaço, os alunos não devem sentir isolamento e as respostas aos mesmos precisam ser rápidas, sendo importante reconhecer e valorizar os comentários dos alunos. No contexto onde alunos e professores estão distantes fisicamente, também faz parte da função social, que o professor gerencie a comunicação entre os alunos, para tentar evitar a impessoalidade e a hiperpessoalidade.

Na educação a distância, o papel desempenhado pelos alunos possui características importantes que devem ser observadas. De acordo com Palloff e Pratt (2004), o aluno virtual sabe trabalhar em conjunto com outros para atingir seus objetivos de aprendizagem. Na realização de trabalhos em grupo, o aluno compreende que a construção do conhecimento ocorre de forma colaborativa, permitindo que ele aumente sua capacidade crítica. A capacidade de refletir é também uma característica importante do aluno virtual.

Os alunos de educação a distância compreendem as mudanças relativas a tempo e espaço que acontecem nesta modalidade de ensino. Segundo Palloff e Pratt (2004), o aluno virtual acredita que a aprendizagem pode ocorrer em qualquer momento e lugar.

No processo de ensino-aprendizagem, o professor deve estar atento aos alunos que possam estar passando por alguma dificuldade. Na educação a distância, o docente precisa observar, caso ocorram, mudanças no nível de participação dos discentes, pois isto pode indicar que eles estão com algum tipo de dificuldade (PALLOFF e PRATT, 2004). Na educação a distância, a interação entre o professor e os alunos, e entre eles mesmos, é fundamental.

Outro papel importante num processo de educação a distância é desempenhado pelo tutor. De acordo com os Referenciais de Qualidade para a Educação Superior a Distância propostos pelo MEC (BRASIL, 2007), o tutor deve participar ativamente da prática pedagógica. As atividades desenvolvidas pelo tutor devem auxiliar o processo de ensino-aprendizagem. Nesta seção, aborda-se o papel do tutor a distância, isto é, aquele que está geograficamente distante dos alunos.

[...] principal atribuição deste profissional é o esclarecimento de dúvidas através de fóruns de discussão pela Internet, pelo telefone, participação em videoconferências, entre outros, de acordo com o projeto pedagógico. O tutor a distância tem também a responsabilidade de promover espaços de construção coletiva de conhecimento, selecionar material de apoio e sustentação teórica aos conteúdos e, freqüentemente, faz parte de suas atribuições participar dos processos avaliativos de ensino-aprendizagem, junto com os docentes (BRASIL, 2007).

Para Niskier (1999), o papel do tutor envolve:

- Comentar as atividades realizadas pelos alunos;
- Corrigir as avaliações dos discentes;
- Auxiliar os alunos a compreender os materiais do curso através de discussões e

explicações;

- Ajudar os discentes no planejamento das atividades;
- Organizar círculos de estudo;
- Efetuar comunicação com os alunos;
- Supervisionar trabalhos práticos e projetos;
- Prover informações sobre o progresso dos estudantes;
- Fornecer *feedback* aos coordenadores sobre os materiais dos cursos e as dificuldades dos discentes;
- Atuar como intermediário entre a instituição e os alunos.

Segundo Vergara (2007), a principal responsabilidade do tutor é motivar e orientar os alunos. Como não existe o contato presencial na EAD, é importante ter agilidade para retornar respostas aos alunos. O tutor deve despertar nos alunos a vontade consciente de compartilhar reflexões e compreensões. O tutor deve estar atento a participação dos alunos, monitorando-os, incentivando-os a prosseguir.

Duran e Vidal (2007) destacam a tutoria entre iguais, uma modalidade de aprendizagem baseada na criação de duplas de alunos. Estes assumem o papel de tutor e tutorado, de acordo com o nível de conhecimento sobre o assunto. Os dois possuem um objetivo comum, conhecido e compartilhado. Os autores observam que entre os recursos pedagógicos do século XXI não podem faltar métodos de aprendizagem cooperativa. Estes procedimentos baseiam-se na heterogeneidade dos discentes. As diferenças entre os estudantes transformam-se em um componente facilitador da aprendizagem. A tutoria entre iguais é um exemplo destes métodos.

De acordo com Mauri e Onrubia (2010) as competências³ dos professores virtuais podem ser descritas como:

1. Competências relacionadas ao projeto tecnológico:
 - a. Analisar e incentivar a integração das TIC na educação.

³ Competência é um saber agir responsável e reconhecido, que implica mobilizar, integrar, transferir conhecimentos, recursos e habilidades, que agreguem valor econômico à organização e valor social ao indivíduo (Fleury, 2001).

- b. Valorizar o ensino do uso das TIC.
 - c. Conhecer as implicações e consequências das TIC na vida cotidiana.
 - d. Conhecer as ferramentas tecnológicas disponíveis.
 - e. Analisar as características tecnológicas de propostas instrucionais, de materiais didáticos e de conteúdos educacionais.
2. Competências referentes ao projeto pedagógico:
- a. Elaborar propostas educacionais virtuais que promovam a construção significativa e com sentido do conhecimento do aluno.
 - b. Criar propostas instrucionais que respondam às exigências de flexibilidade da aprendizagem.
 - c. Projetar atividades, avaliações e tarefas de aprendizagem eficaz.
 - d. Elaborar condições que facilitem a presença social dos envolvidos no processo educacional.
 - e. Construir oportunidades de orientação e acompanhamento do discente para que ele se aproprie do conteúdo e reflita sobre o processo de aprendizagem, aumentando o controle e a autogestão deste.
 - f. Criar oportunidades de comunicação entre professor e aluno.
3. Competências relativas ao uso técnico-pedagógico das TIC:
- a. Construir conjuntamente com o aluno uma representação inicial da situação virtual.
 - b. Consultar, gerenciar, armazenar e apresentar informação para responder às necessidades dos alunos de aprendizagem significativa e com sentido.
 - c. Potencializar a exploração ativa e a seleção da informação por parte do aluno, ajudando-o a compreender o que é essencial.
 - d. Mediar a leitura de diferentes mídias para informar-se e aprender.
 - e. Ajudar o discente a gerenciar e apresentar informação.

- f. Organizar e gerenciar o processo de ensino e aprendizagem, de modo a promover a participação, que as contribuições sejam necessárias e que se facilite a construção da interatividade.
- g. Gerenciar o tempo e o ritmo de trabalho com os alunos, combinando exigência e flexibilidade para facilitar o processo de aprendizagem.
- h. Promover o incremento da atuação conjunta nas atividades de ensino e aprendizagem ao longo do processo.

O quadro 2 apresenta as competências mínimas necessárias aos professores, alunos, e tutores, considerando-se o contexto da EAD, citados por Konrath et al. (2009).

Quadro 2 - Competências mínimas necessárias aos professores, alunos e tutores

Ator	Competências na EAD
Professor	<ul style="list-style-type: none"> a) Conhecer e dominar o conteúdo. b) Conhecer e dominar a tecnologia utilizada. c) Realizar a organização pedagógica e gerenciar as ações do curso. d) Administrar discussões e trabalhos em grupo. e) Trabalhar as regras e expectativas do curso. f) Avaliar os alunos. g) Desenvolver atos pedagógicos. h) Desenvolver técnicas para facilitar a aprendizagem. i) Acompanhar o trabalho do tutor. j) Oferecer suporte em tarefas cognitivas. k) Promover a auto-reflexão nos alunos. l) Acompanhar o processo de aprendizagem dos discentes. m) Guiar os alunos. n) Instruir os discentes. o) Efetuar comunicação com os alunos (opinar, informar, estabelecer e manter contato, convidar alunos para participar, corrigir). p) Avaliar os efeitos sociais da comunicação (impessoal, interpessoal, hiperpessoal).
Aluno	<ul style="list-style-type: none"> a) Ser responsável pela sua aprendizagem. b) Organizar seu tempo para estudo, realização das atividades e interação com o grupo. c) Possuir postura ativa, criativa e comprometida. d) Participar. e) Questionar. f) Posicionar-se. g) Refletir. h) Utilizar as ferramentas do curso.
Tutor	<ul style="list-style-type: none"> a) Conhecer e dominar o conteúdo. b) Conhecer e dominar a tecnologia utilizada. c) Acompanhar o trabalho pedagógico, orientando, e incentivando os alunos. d) Efetuar diálogo constante com os alunos e o professor. e) Avaliar os efeitos sociais da comunicação (impessoal, interpessoal, hiperpessoal).

Fonte: Konrath et al. (2009)

3.1.2 INTERAÇÃO ENTRE OS ATORES ENVOLVIDOS NA EDUCAÇÃO A DISTÂNCIA

Na EAD, é muito importante que o professor construa atividades que favoreçam momentos de interação entre os alunos. No processo de ensino-aprendizagem, este processo entre aluno, professor, tutor, é de suma importância.

Para Perraudeau (2009), o professor pode ter uma postura interativa de mediação ou tutoria. As duas são modos de interação que implicam em incentivar o interesse e as trocas de ideias entre os alunos. Na tutoria, há a presença colaborativa do tutor durante a realização da tarefa, fornecendo informações, respondendo às perguntas, enquanto a atividade é realizada pelo aluno. Na mediação, o professor observa os alunos trabalhando, podendo incentivar, mas não intervindo na tarefa.

[...] a assistência para com o aluno é uma questão pedagógica renovada incessantemente, que varia conforme um conjunto de parâmetros: a época, a demanda institucional, o contexto, o estilo do professor, o objetivo, etc (PERRAUDEAU, 2009, p. 202).

Segundo Moore e Kearsley (2008), o conceito de interação surgiu com John Dewey e foi desenvolvido por Boyd (1980), “a interação implica a inter-relação do ambiente e das pessoas com os padrões de comportamento em uma situação”.

A interação a distância é o hiato de compreensão e comunicação entre professores e alunos causados pela distância geográfica que precisa ser suplantada por meio de procedimentos diferenciadores na elaboração da instrução e na facilitação da interação (MOORE e KEARSLEY, 2008, p. 240).

Na educação a distância, Moore e Kearsley (2008) citam três tipos de interação que podem ocorrer: aluno-conteúdo, aluno-professor e aluno-aluno.

A interação aluno-conteúdo envolve o discente e o material a ser estudado. É essencial que o professor possa facilitar este tipo de interação, pois a partir da mesma, o aluno realiza a construção do conhecimento.

A interação aluno-professor é considerada fundamental por grande parte dos alunos e professores. Ela envolve alguns pontos relativos às ações do docente, como a ajuda ao discente na sua interação com o conteúdo. O professor auxilia na resolução de dúvidas, oferece esclarecimentos e explicações. O docente também deve motivar o estudante e avaliar o progresso do mesmo. É importante destacar que a reação de cada pessoa ao conteúdo é

distinta e, portanto, o direcionamento do professor a cada discente também será diferente. Dependendo da necessidade de cada um, o docente pode: oferecer explicações sobre algo que não foi compreendido, realizar elaborações, construir analogias, sugerir leituras complementares, entre outras ações.

A interação aluno-aluno é considerada como motivadora pelos discentes. Esta interação pode ocorrer internamente nos grupos de estudantes, entre os grupos, e individualmente. As discussões entre eles são importantes, pois os ajudam a refletir sobre o conteúdo estudado.

De acordo com Mattar (2009), a interação do aluno com o professor, de forma síncrona ou assíncrona, fornece motivação e *feedback* aos discentes, favorecendo o aprendizado. Com tecnologias modernas, a interação do estudante com o conteúdo pode ocorrer de diversas formas, através de diferentes mídias. A interação entre os alunos pode acontecer de forma síncrona ou assíncrona, favorecendo um processo de aprendizagem colaborativo e cooperativo. Ela gera motivação e diminui a sensação de isolamento da educação a distância, desenvolvendo também o senso crítico e habilidade de trabalhar em equipe.

A interação entre alunos, professores, tutores, pode ser promovida mediante a utilização de Ambientes Virtuais de Aprendizagem.

3.2 AMBIENTES VIRTUAIS DE APRENDIZAGEM

Esta seção apresenta as características dos Ambientes Virtuais de Aprendizagem (AVAs). No cenário da EAD, dentre os recursos tecnológicos relacionados à *Web*, estes ambientes merecem grande destaque.

Segundo Schlemmer (2002), a maior parte dos AVAs é baseada na arquitetura cliente-servidor. O cliente é um navegador *Web* que acessa as páginas no servidor. O servidor é um computador com software específico para esta finalidade, com capacidade para atender às solicitações de vários clientes.

Os AVAs representam um recurso tecnológico para a educação. Estes ambientes possuem recursos que favorecem o compartilhamento de informações e a comunicação entre

os alunos e entre eles e o professor. Os recursos disponíveis nos AVAs facilitam o gerenciamento, o controle e a interação entre os atores envolvidos.

De acordo com Onrubia et al. (2010), estes ambientes, também chamados de Sistemas de Gestão da Aprendizagem (*Learning Management Systems - LMS*) passaram por duas gerações. A primeira baseava-se na distribuição de conteúdos e na avaliação dos resultados dos alunos. A segunda geração pautou-se em oferecer recursos para a comunicação entre os participantes e para acompanhar o processo de ensino-aprendizagem. De acordo com o autor, estes reúnem os recursos necessários para gerenciar os materiais de aprendizagem, para acompanhar e avaliar o progresso dos discentes, e para disponibilizar a comunicação entre os participantes.

Moisey e Hughes (2008) destacam que um ambiente de aprendizagem que ofereça ferramentas de apoio aos alunos é fundamental para garantir o sucesso, a promoção da persistência, e evitar o abandono. O ambiente de aprendizagem ideal visa desenvolver a independência do aluno e facilitar o processo de aprendizagem, fornecendo recursos que são flexíveis, acessíveis e disponíveis quando necessários.

Para Dillenbourg et al. (2002), os AVAs possuem as seguintes características:

- a) Constituem um espaço planejado de informações.
- b) Representam um espaço social onde ocorrem interações educacionais.
- c) O espaço virtual é explicitamente representado de diversas formas, desde textos até imersões em três dimensões.
- d) Os alunos não são somente ativos, mas também atores, pois eles participam da construção do espaço virtual.
- e) Eles não são restritos a EAD, pois também podem enriquecer atividades na sala de aula.
- f) Podem integrar diversas tecnologias e múltiplas abordagens pedagógicas.

Segundo Behar (2007b), um AVA é um espaço que favorece a interação dos sujeitos, possibilitada através de uma plataforma⁴.

⁴ A plataforma é uma infra-estrutura tecnológica composta pelas funcionalidades e interface gráfica que compõe o AVA (BEHAR, 2006a).

Considera-se um AVA como ambiente coletivo que favorece a interação dos sujeitos participantes, sendo este um todo constituído pela plataforma e por todas as relações estabelecidas pelos sujeitos usuários a partir do uso das ferramentas de interação, tendo como foco principal a aprendizagem. O AVA é um espaço na Internet formado pelos sujeitos e suas interações e formas de comunicação que se estabelecem através de uma plataforma (BEHAR, 2007b, p. 6).

Para Schlemmer (2002), os AVAs são softwares construídos com o objetivo de gerenciar a aprendizagem via *Web*. Eles são sistemas que sintetizam a Comunicação Mediada por Computador (CMC), e possibilitam a entrega do material dos cursos realizados.

De acordo com Bassani (2006), os AVAs oferecem diversos recursos. Alguns deles permitem o acompanhamento da frequência, assiduidade, e produção de cada aluno (como os resultados das atividades realizadas, trabalhos publicados). Outros recursos permitem verificar as mensagens trocadas entre os alunos, professores e tutores.

Barreto et al. (2006) explicam que os AVAs conduzem a uma mudança de dimensão do ensinar e do aprender, permitindo que o espaço e o tempo de aprendizagem sejam ampliados. Nesse redimensionamento, o aluno e o professor assumem novos papéis, onde o aluno precisa de mais autonomia para aprender, e o professor torna-se um moderador e facilitador do processo de aprendizagem. Para Araújo Jr. e Marquesi (2009), com este papel, o professor encontra novos desafios sobre como trabalhar o processo de ensino-aprendizagem de forma eficiente nesses ambientes e como produzir conteúdos e atividades adequados. As atividades realizadas nos AVAs podem motivar o desenvolvimento da autonomia, sistematizar o conhecimento, e possibilitar a exploração de espaços/recursos virtuais.

Segundo Ramsey (2003), os AVAs podem contribuir para melhorar e desenvolver as interações entre professores e alunos. Estes podem facilitar relações conversacionais e participativas entre docentes e discentes, e relações mais engajadas entre os discentes. Para a autora, a aprendizagem é um processo construtivo. A partir das relações entre os alunos e professores, há a elaboração de formas para construir o conhecimento. O uso de AVAs pode enriquecer as relações de aprendizagem, pois eles representam um lugar de encontro.

De acordo com Schlemmer (2005), o ponto mais importante a ser observado na utilização de um AVA é o critério didático-pedagógico com o qual ele foi construído. Segundo a autora, o desenvolvimento de um produto para a educação é permeado por uma concepção epistemológica.

Schlemmer (2005) propõe uma estratégia para a avaliação de AVAs sob uma perspectiva técnica, didático-pedagógica, comunicacional-social e administrativa, de acordo com os critérios apresentados em Schlemmer e Fagundes (2001) e Schlemmer (2002).

A perspectiva técnica envolve o suporte tecnológico, com disponibilidade de ferramentas para: gerenciamento e gestão do ambiente, gestão de comunidades, autoria (professor e/ou estudante), área individual, área da comunidade, serviços de apoio, interação síncrona, interação assíncrona, avaliação.

A perspectiva didático-pedagógica envolve a concepção epistemológica, o foco do ambiente para a aprendizagem, a compreensão de ensino. Ela está relacionada com a forma de organizar o currículo, a metodologia didática, a forma de avaliação. A capacitação dos professores, a concepção sobre a aquisição de conhecimento, a possibilidade de trabalhar a aprendizagem autônoma, e os recursos para promover a reflexão dos estudantes, também devem ser considerados.

A perspectiva comunicacional-social destaca aspectos relacionados com: o espaço para as informações, as fontes de negociação, as ferramentas discursivas, as possibilidades de interação, a concepção do papel do aluno e do professor.

A perspectiva administrativa envolve: os atores que utilizam o AVA e suas funções, adaptabilidade, possibilidades de colaboração, recursos para monitoramento por parte do professor, recursos para auto-organização dos alunos, a concepção do sistema educacional, e a forma de estruturar os conteúdos a serem estudados.

Schlemmer e Fagundes (2001) enfatizam que uma estratégia de avaliação de AVAs deve considerar três aspectos importantes:

- O AVA deve oferecer possibilidades para a melhoria da qualidade da aprendizagem.
- O ambiente deve possuir recursos que suportem a interação entre os atores do processo de ensino-aprendizagem.
- O AVA deve diminuir a sobrecarga administrativa dos docentes, possibilitando que eles dediquem mais tempo com as necessidades educacionais individuais dos discentes.

Neste trabalho, entende-se o conceito de AVA como um ambiente para auxiliar o processo de ensino-aprendizagem. Ele disponibiliza um conjunto de ferramentas que facilitam

a interação/comunicação entre docentes e discentes, e entre os alunos. Um AVA também oferece recursos para auxiliar o professor a organizar o material e acompanhar as atividades dos estudantes.

A seção seguinte apresenta algumas ferramentas disponíveis em três AVAs que foram pesquisados pelo autor deste trabalho, utilizados como exemplos de ilustração.

3.2.1 FERRAMENTAS DISPONÍVEIS NOS AMBIENTES VIRTUAIS DE APRENDIZAGEM

Atualmente, existem diversos Ambientes Virtuais de Aprendizagem que podem ser utilizados pelo docente. Eles se diferenciam uns dos outros em vários aspectos, permitindo ao professor/instituição escolhas que promovam maior adequação ao modelo pedagógico⁵ proposto.

Os AVAs possuem diversas ferramentas. Dentre as mais comuns, podem ser destacadas: a existência de conteúdos interativos; comunicação síncrona e assíncrona entre os participantes (alunos, professores, tutores); área de registro/avaliação de atividades, entre outras. Estes disponibilizam ferramentas para facilitar a comunicação entre os atores envolvidos na EAD. O bate-papo possibilita a realização da comunicação síncrona, na qual os participantes devem interagir ao mesmo tempo. O fórum de discussão permite a comunicação assíncrona, na qual a interação pode ocorrer em momentos distintos. O correio permite que os participantes enviem e recebam mensagens entre si. O mural possibilita aos alunos visualizar recados encaminhados pelo professor ou tutor.

A título de exemplificação, serão citadas algumas das ferramentas disponíveis nos ambientes Moodle, Rooda e ETC.

3.2.1.1 MOODLE

O Moodle (Modular Object-Oriented Dynamic Learning Environment), apresentado na figura 2, foi criado em 2001 por Martin Dougiamas. Ele é um software livre desenvolvido

⁵ Modelo pedagógico pode ser definido como um sistema de premissas teóricas que representa, explica e orienta a forma como se aborda o currículo, concretizando-se nas práticas pedagógicas, assim como nas interações professor-aluno-objeto de conhecimento (BEHAR, 2009a).

colaborativamente por uma comunidade virtual formada por programadores, desenvolvedores de softwares livres, administradores de sistemas, professores, designers e usuários.

Algumas das ferramentas disponíveis no Moodle são: agenda, avaliações, atividades, material de apoio, enquetes, mural, correio, grupos, fórum, chat, perfil, diário, portfólio, relatórios, notas, livro, glossário, blog, usuários, wiki, ajuda (FRANCISCATO et al., 2008).

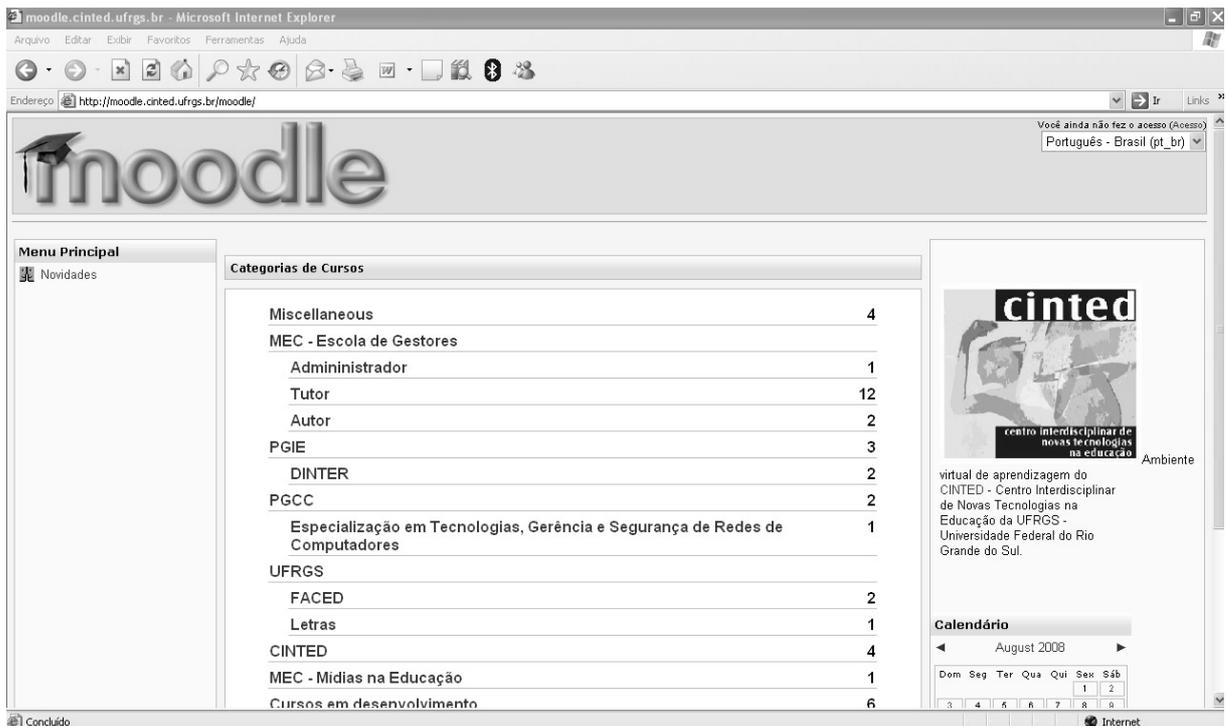


Figura 2 - Moodle
Fonte: <http://moodle.org>

O Moodle oferece opções para gerar relatórios sobre a participação dos alunos dentro de um curso. As principais ferramentas para esta finalidade são: Relatórios e Perfil.

Os relatórios oferecem resultados quantitativos, onde é possível visualizar o acesso e a atuação de um aluno (ou todos os alunos) no ambiente. As participações dos discentes podem ser verificadas em um dia específico (ou em todos os dias) de realização de um curso, ou em uma atividade específica (ou em todas as atividades). As ações realizadas pelos alunos também podem ser observadas.

A integração do MineraFórum ao Moodle é interessante, pois o AVA não possui um recurso que auxilie na análise qualitativa das postagens dos fóruns.

3.2.1.2 ROODA

O ROODA⁶ - Rede cOOperativa De Aprendizagem (BEHAR, 2005a, 2005c, 2006, 2007a, 2007c, 2009a), apresentado na figura 3, foi desenvolvido em 2000, e disponibilizado pela Universidade Federal do Rio Grande do Sul (UFRGS) em 2003.

De acordo com Behar (2005b), o ROODA é um ambiente que possibilita a interação, aprendizagem, cooperação.

Assim, o ROODA pode ser considerado um ambiente com ênfase na aprendizagem, um lugar rico em recursos, onde os sujeitos podem efetivar trocas e construir conhecimentos. Este espaço virtual caracteriza-se pela possibilidade de: interação, aprendizagem enquanto processo individual/coletivo, comunicação assíncrona e síncrona, interdisciplinaridade, hipertextualidade e trabalho em equipe (BEHAR, 2005b, p. 7).

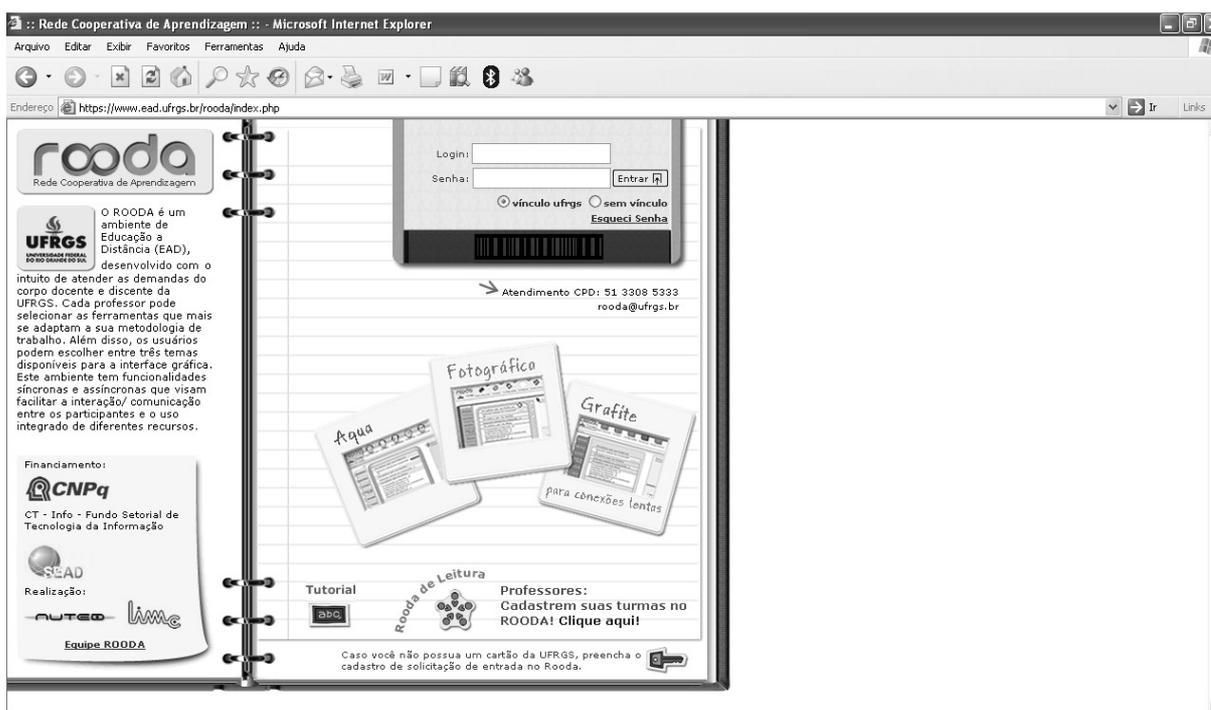


Figura 3 - Interface principal do ROODA
Fonte: <https://www.ead.ufrgs.br/rooda>

De acordo com Behar (2005a), a interface do ROODA oferece uma agradável experiência estética, e permite uma navegação intuitiva e rápida. Observando-se questões de adaptabilidade e usabilidade, o ambiente pode ser configurado pelo docente segundo seus objetivos educacionais. Existem três temas de interface disponíveis: Fotográfica, Aqua e

⁶ Disponível em <https://www.ead.ufrgs.br/rooda>

Grafite.

Da mesma forma que os AVAs citados anteriormente, o ROODA é um software livre. Behar (2007a) apresenta as funcionalidades existentes no ambiente:

- A2: apresenta a lista dos usuários conectados ao ambiente.
- Aulas: possibilita a montagem das aulas e o acesso às mesmas pelos alunos.
- Exercícios: oferece a criação de exercícios.
- Bate-Papo: permite a comunicação síncrona em grupo.
- Biblioteca: possibilita a publicação e organização de materiais e links.
- Compromissos: organiza os compromissos em uma agenda.
- Conceitos: possibilita que o professor registre os conceitos/notas e, que o aluno visualize os conceitos/notas publicados.
- Configurações: configura o ambiente.
- Contatos: apresenta a lista de nome e e-mail dos usuários, e disponibiliza o envio de mensagens.
- Dados Pessoais: oferece o cadastro de informações pessoais.
- Diário de Bordo: é o local para o aluno registrar seu processo, com a possibilidade do professor e outros alunos postarem comentários.
- Disciplina: permite acesso às disciplinas que o aluno participa, apresenta o mural geral e lembretes. Representa a área de trabalho do ambiente.
- Enquete: possibilita a realização de pesquisas.
- Fórum: é o espaço de interações assíncronas entre os participantes, dividido em Fórum Geral e Fórum Específico.
- Gerência da Disciplina: permite que o professor ative, registre dados e habilite funcionalidades para as suas disciplinas.
- Lembretes: possibilita anotações pessoais.
- Lista de Discussão: apresenta a lista dos endereços eletrônicos dos participantes de uma disciplina.

- Mural: divide-se em Mural Geral, que disponibiliza as mensagens enviadas pelo ambiente, e Mural da Disciplina, que apresenta as que foram encaminhadas pelo professor.
- Grupos: possibilita o uso coletivo de recursos. Divide-se em Gerais e Específicos da disciplina.
- Webfólio: permite que o aluno envie arquivos, divide-se em webfólio geral, webfólio das disciplinas e webfólio dos grupos.
- Ajuda: apresenta instruções para auxiliar na utilização do ambiente, podendo ser feito através do tutorial, do glossário e por funcionalidades.
- RoodaExata: é uma meta-funcionalidade do ROODA, integrada aos recursos de interação e comunicação, como o bate-papo e o fórum. Permite a escrita de fórmulas matemáticas, expressando de forma objetiva os conceitos que fazem uso da linguagem simbólica (NOTARE, 2009).
- InterROODA: é uma ferramenta que permite o acompanhamento de acessos e frequência e o mapeamento das trocas interindividuais (BASSANI, 2006).

3.2.1.3 ETC

O ETC⁷ - Editor de Texto Coletivo (BEHAR et al., 2005d, 2005e, 2006b, 2007d, 2009b, 2010) foi desenvolvido por um grupo interdisciplinar do NUTED/UFRGS (Núcleo de Tecnologia aplicada à Educação/ Universidade Federal do Rio Grande do Sul). O ambiente oferece suporte à escrita cooperativa/colaborativa através da *Web*. O ETC representa um espaço virtual para que pessoas a distância possam elaborar textos de forma síncrona e/ou assíncrona. A figura 4 apresenta a interface principal.

⁷ Disponível em <http://www.nuted.ufrgs.br/etc2/>

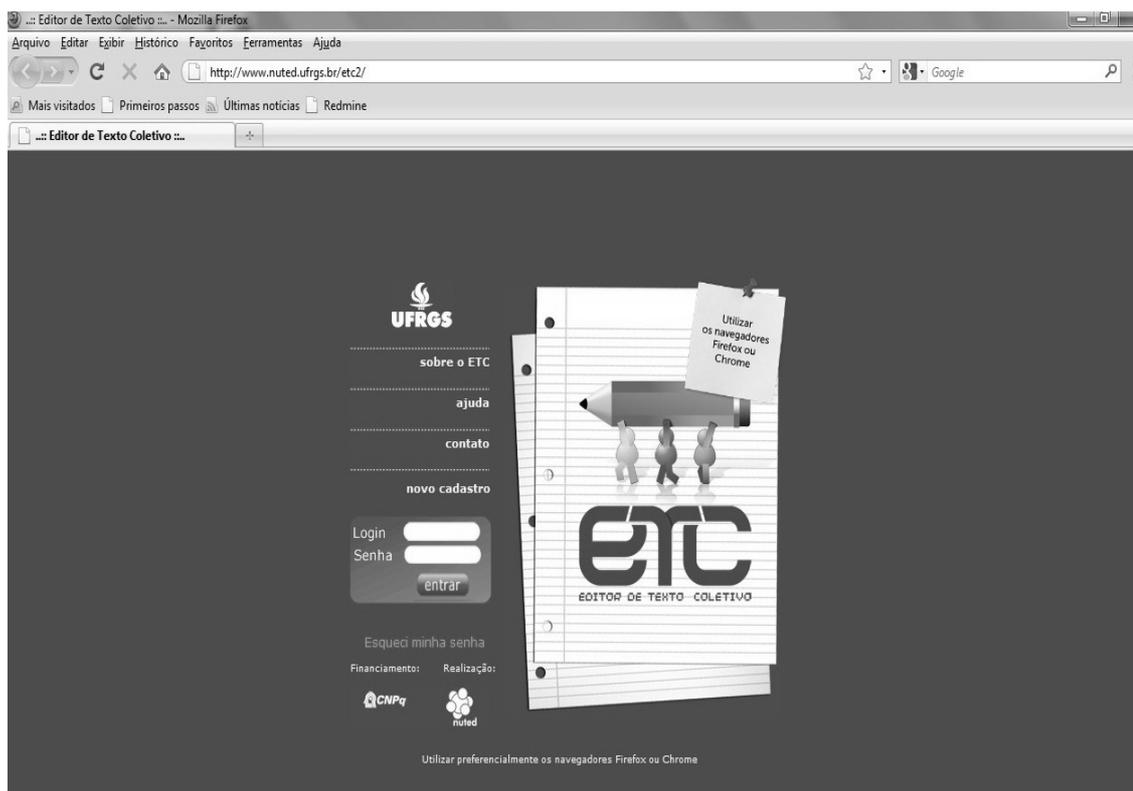


Figura 4 - Interface principal do ETC
 Fonte: <http://www.nuted.ufrgs.br/etc2/>

No ETC, o usuário constrói seu conhecimento a partir da interação com outros sujeitos, com os recursos oferecidos pelo ambiente, com os materiais publicados, suas próprias produções e dos demais usuários. O ambiente oferece ferramentas destinadas à interação entre as pessoas envolvidas, como o fórum de discussão e o bate-papo.

O ambiente disponibiliza ferramentas para armazenar e consultar materiais para o estudo dos participantes. Há ainda o espaço para a edição dos textos coletivos, que possui funcionalidades de formatação e edição. O ETC possibilita (Behar et al., 2005e): organização dos participantes em comunidades; criação de textos por qualquer sujeito cadastrado; cadastro direto através da *Web*; inserção, edição e exclusão de textos; inclusão de comentários em um texto; verificação do histórico do texto; visualização da versão final a qualquer momento; e visualização da descrição dos textos da comunidade que participa.

Behar et al. (2010) apresentam a reestruturação realizada no ETC, a partir de avaliações feitas com os usuários em contextos educacionais. A reformulação ocorreu nos seguintes aspectos: na administração de dados e cadastros, permitindo que a criação, edição e o gerenciamento sejam iguais para todos os participantes; na forma de edição de texto, onde pode ser editado o conteúdo por completo, de forma simultânea entre os usuários; e na

reformulação integral da interface.

Observando as ferramentas disponibilizadas nos três AVAs apresentados, identifica-se que muitas são comuns entre eles. Considerando que o fórum de discussão é a ferramenta a ser enfatizada nesta pesquisa, a próxima seção descreve as suas características.

3.2.2 FÓRUMS DE DISCUSSÃO

O fórum de discussão é um recurso interessante disponível nos AVAs pois permite a comunicação assíncrona entre professores, tutores e alunos. Além disto, este permite que os sujeitos participantes discutam um determinado assunto e troquem ideias. Ele auxilia na elucidação de dúvidas, na expressão de opiniões, na discussão de tópicos, e na reflexão sobre os conceitos debatidos. Esta seção apresenta as características desta ferramenta.

De acordo com Dornelles (2001), o recurso possibilita a um grupo de pessoas de interesse comum, debater e compartilhar determinadas informações, dúvidas e opiniões.

Segundo Sánchez (2005), o fórum para fins educacionais em um ambiente *online* é definido como um espaço de comunicação composto por quadros de diálogo, onde as mensagens redigidas podem ser classificadas tematicamente. Nestes espaços os alunos podem realizar contribuições, refutar outras, esclarecer dúvidas. A comunicação é efetuada de forma assíncrona e as mensagens digitadas permanecem à disposição dos participantes.

Carlini (2010) indica que devem ser realizadas discussões envolvendo troca de saberes, perspectivas, experiências, além das opiniões. Para a autora, é importante que o professor motive as trocas entre os alunos, a comunicação entre eles e a reflexão.

Antes de iniciar sua participação no debate é importante que o aluno realize leituras, pesquisas, ou outras atividades envolvendo o assunto a ser tratado. Para Oliveira (2005), o fórum é um espaço de discussões e reflexões.

No contexto dos cursos virtuais, a participação no espaço criado pelo fórum pede preparo, geralmente provido por leituras adequadas, pesquisas, resgates ao background próprio a cada participante, entre outras formas de busca. Trata-se de organizar o pensamento, enriquecendo-o com pertinentes referências, permitindo o uso do espaço de discussões e reflexões proporcionado pelo fórum para gerar colaborações, para agregar ideias (OLIVEIRA, 2005, p. 5).

As interações entre aluno e professor, ou entre os discentes, ocorrem através da troca

de mensagens escritas. Geralmente, há separação por temas, e dentro dos temas pode haver uma organização por tópicos. Dependendo do AVA, a visualização das mensagens postadas pode acontecer de forma linear e/ou hierárquica, cronológica ou por assunto. No Rooda (figura 5), as postagens são agrupadas por tópicos. Dentro de cada tópico as contribuições textuais são exibidas em ordem cronológica.

The screenshot shows a web browser window displaying the Rooda forum. The browser's address bar shows the URL <https://www.ead.ufrgs.br/rooda/rooda.php>. The forum page is titled "Rooda > Tópicos Especiais - Z1 > Fórum". On the left side, there is a navigation menu with options: Fórum, Aulas, Bate-Papo, Biblioteca, Diário de Bordo, interRooda, Grupos, and Webfolio. The main content area displays a table of discussion topics under the heading "Tópicos para Tópicos Especiais - Z1".

Tópico	Autor	Respostas	Funcionalidades Vinculadas	Mensagens Novas
Sugestões para o ROODA	ANDRÉ FERNANDO UÉBE MANSUR	15	-	0
teste	Patricia Alejandra Behar	0	-	0
Links com materiais afins	ANDRÉ FERNANDO UÉBE MANSUR	10	-	0
Sugestão de leitura para o trabalho final de OAs	ARILISE MORAES DE ALMEIDA LOPES	2	Aula:Final:Oficinas Virtuais de Aprendizagem	0
Dicas para análise de software educacional	ADELSON SIQUEIRA DE CARVALHO	0	-	0
Discussão sobre Comunidades Virtuais do Grupo: Rodney, Maurício, Helvia e Maria	RODNEY CEZAR DE ALBUQUERQUE	6	-	0
As CVA	RODNEY CEZAR DE ALBUQUERQUE	4	-	1

Figura 5 - Fórum de discussão do AVA Rooda
Fonte: <https://www.ead.ufrgs.br/rooda/rooda.php>

No ETC (figura 6), as mensagens são agrupadas por tópicos. Eles podem ser encontrados através do título, do conteúdo, ou do autor. Dentro de cada tópico as contribuições textuais são exibidas em ordem cronológica. É possível encontrar as postagens através do título, do seu conteúdo, ou do autor da mesma. O recurso permite que o usuário possa responder a uma determinada mensagem. Também se pode visualizar qual foi a postagem colocada anteriormente a outra.

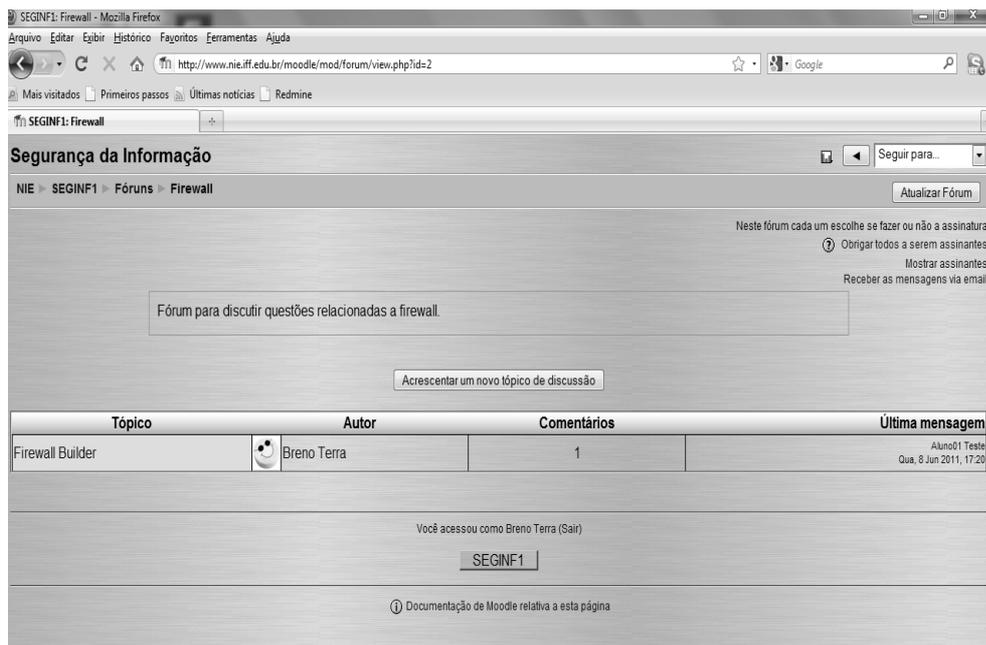


The screenshot shows a web browser window displaying the forum page for ETC. The page title is "EPT e a Competência para o Tr". The forum interface includes a search bar, a "Criar tópico" (Create topic) button, and a list of topics with columns for "Tópico", "Autor", "Mensagens", and "Última Mensagem".

Tópico	Autor	Mensagens	Última Mensagem
Cafezinho	[Redacted]	59	29.01.2011 às 03:44:04 por [Redacted]
Trabalho em Equipe: como, por quê e para quê ou para quem?	[Redacted]	101	29.01.2011 às 03:23:29 por [Redacted]
Desenvolvimento de competências	[Redacted]	55	29.01.2011 às 03:10:23 por [Redacted]
Grupotencia	[Redacted]	55	27.01.2011 às 10:47:44 por [Redacted]
Grupo "Alfa"	[Redacted]	2	26.01.2011 às 16:37:39 por [Redacted]
Dúvidas sobre o Trabalho Final do Curso	[Redacted]	8	26.01.2011 às 19:27:51 por [Redacted]
Clube da Luzuzinha	[Redacted]	22	26.01.2011 às 15:45:00 por [Redacted]
Grupo "Aprendendo Sempre"	[Redacted]	12	26.01.2011 às 14:59:50 por [Redacted]
Dúvidas sobre o ETC	[Redacted]	37	27.01.2011 às 13:16:12 por [Redacted]
Grupo "Da Praia"	[Redacted]	23	26.01.2011 às 23:55:01 por [Redacted]

Figura 6 - Fórum de discussão do ETC
 Fonte: <http://www.nuted.ufrgs.br/etc2/forum/index.php>

No Moodle (figura 7), as mensagens são agrupadas por tópicos. Dentro de cada tópico, as contribuições textuais podem ser exibidas de forma aninhada. Também é possível visualizar as postagens a partir daquela que foi mais recentemente colocada, ou da que foi escrita há mais tempo.



The screenshot shows a Moodle forum page for the topic "Firewall". The page title is "Segurança da Informação". The forum interface includes a search bar, a "Seguir para..." button, and a list of topics with columns for "Tópico", "Autor", "Comentários", and "Última mensagem".

Tópico	Autor	Comentários	Última mensagem
Firewall Builder	Breno Terra	1	Aktm01 Teste Qua, 8 Jun 2011, 17:20

Additional details from the screenshot: "Fórum para discutir questões relacionadas a firewall.", "Acrescentar um novo tópico de discussão", "Atualizar Fórum", "Neste fórum cada um escolhe se fazer ou não a assinatura", "Obrigado todos a serem assinantes", "Mostrar assinantes", "Receber as mensagens via email", "Você acessou como Breno Terra (Sair)", "SEGINF1", "Documentação de Moodle relativa a esta página".

Figura 7 - Fórum de discussão do Moodle
 Fonte: <http://www.nie.iff.edu.br/moodle/mod/forum/view.php?id=2>

Em um fórum, dependendo do tema abordado, do perfil dos discentes, do interesse despertado pelo tópico, podem-se ter centenas de contribuições textuais. Algumas delas apresentam argumentos favoráveis ao tema proposto e outras relatam posicionamentos contrários. Alguns textos redigidos pelos alunos possuem relevância com relação ao assunto tratado, e outros não. A próxima seção apresenta a importância do uso pedagógico dos fóruns de discussão.

3.3 USO PEDAGÓGICO DE FÓRUMS DE DISCUSSÃO

De acordo com Gilbert e Dabbagh (2005), um benefício pedagógico importante da comunicação assíncrona é o seu potencial para apoiar a co-construção do conhecimento por meio do discurso.

No processo de aprendizagem, uma forma de interação entre os sujeitos é a discussão. De acordo com Macedo (2010), a discussão pode auxiliar a construção de conhecimento, favorecer a socialização, desenvolver o pensamento crítico, e promover a criatividade.

[...] as trocas sociais apoiadas na discussão, no confronto de diferentes pontos de vista, podem contribuir com a constituição de novos conhecimentos, novas perspectivas e apontar para o progresso da socialização entre os pares. Assim, a troca social funciona como uma ferramenta que incita a criatividade e o espírito crítico (MACEDO, 2010, p.35).

Para Palloff e Pratt (2004), a discussão assíncrona deve ser incentivada pelo professor, pois é a melhor forma de estabelecer e manter a interação com os alunos.

Uma vez que os alunos determinem um ritmo e comecem a interagir ativamente, eles assumirão a responsabilidade de sustentar esse contato, seja pela interação social, seja como uma resposta às perguntas para discussão enviadas pelo professor (PALLOFF e PRATT, 2004, p. 47).

De acordo com estes mesmos autores, as interações dos discentes nas discussões proporcionam um momento de reflexão sobre os conteúdos educacionais abordados. A capacidade de refletir é fundamental para o aluno virtual, esta deve ser motivada. O fórum é um espaço propício para realizar este tipo de ação. O fato dos estudantes responderem às questões de uma discussão e às mensagens indica uma possível reflexão.

Os debates entre os discentes, propiciados através da utilização do fórum, motivam o

desenvolvimento do pensamento crítico⁸. De acordo com Palloff e Pratt (2004), o professor pode utilizar técnicas como estudo de caso, simulações, debates, para trabalhar o pensamento crítico dos estudantes.

Para Kratochwill e Silva (2008), o fórum favorece a construção do conhecimento e possibilita uma avaliação sob a perspectiva dialógica e colaborativa, desde que os alunos e professores percebam este propósito.

[...] o fórum pode ser usado para promover a avaliação a partir das possibilidades interativas e dialogais que suscita, pois propicia uma ampla rede conversacional, polifônica, híbrida e motivadora da colaboração. As interações entrelaçam-se numa multiplicidade de textos que se complementam, contrapõem-se e permitem o acompanhamento e as interferências do educador-mediador (KRATOCHWILL e SILVA, 2008, p. 455).

De acordo com Palloff e Pratt (2004), o papel do professor nos fóruns de discussão é muito importante. Além de escrever mensagens de apoio e motivação para os alunos, responder às perguntas, ele deve observar o nível de participação de cada discente. Caso seja detectado que um aluno não está participando ou está desviando do tema da discussão, o docente deve tentar ajudá-lo a suplantar as dificuldades e a solucionar os problemas.

O papel do professor-facilitador nas discussões online é ajudar os alunos a sintetizarem o material estudado, a enxergarem paralelos e a resumirem os pontos primordiais, auxiliando-os, assim, a irem em frente. Quando o professor intervém em demasia, há, de fato, uma interrupção do diálogo, pois ele se torna o centro das atenções. Por outro lado, a ausência da intervenção do professor pode também interromper a discussão, pois os alunos talvez se sintam confusos e incertos sobre o que se espera deles. Atingir o equilíbrio é, então, fundamental (PALLOFF e PRATT, 2004, p. 155).

Dennen (2008) destaca que a discussão é frequentemente considerada como um componente necessário em aulas a distância, mesmo que seja somente para proporcionar uma sensação de conexão entre os participantes. Alguns professores sugerem que os alunos utilizem um fórum, em um sentido geral, para discutir leituras ou conceitos, enquanto outros definem atividades com regras claras. Estas atividades demonstram a importância do contexto ao analisar elementos de uma experiência de aprendizagem utilizando discussão assíncrona. O ato de utilizar um fórum para apoiar o processo de aprendizagem sugere que a interação social é valorizada. Um fórum de discussão é capaz de suportar um tipo de engajamento que seja

⁸ A expressão “pensamento crítico” é aqui utilizada como sinônimo de “espírito crítico” definido por Piaget. O sentido da experiência e necessidade de coerência lógica que se colocam ao serviço de uma razão autônoma, comum a todos os indivíduos e que não depende de nenhuma autoridade externa (PIAGET, 1998).

orientado para o processo (com foco na discussão e contando com a experiência anterior dos participantes) e orientado para o produto (com foco em reafirmar o que os alunos já compreenderam e ajustar os seus equívocos). O propósito da discussão em uma aula a distância pode se estender além do apoio à aprendizagem. Independentemente da orientação do produto ou processo, a interação inerente ao discurso pode ser motivadora para os alunos.

O estudo apresentado por Garrison, Anderson, e Archer (2000) sugere que a comunicação baseada em texto oferece tempo para reflexão. Estes autores relatam que algumas literaturas indicam que a comunicação escrita está intimamente relacionada com o pensamento cuidadoso e crítico. O uso da escrita pode ser crucial quando o objetivo é facilitar o pensamento sobre questões complexas, a aprendizagem significativa e profunda.

Segundo Xin e Feenberg (2006), a discussão é a principal forma de interação humana mediada por computador na educação a distância. A discussão possui um papel vital, ainda mais significativo para a aprendizagem do que na sala de aula face-a-face. Os benefícios identificados incluem debates mais reflexivos e uma maior participação dos alunos. Os professores muitas vezes preferem liderar uma discussão por escrito, com tempo para refletir sobre as ideias dos discentes antes de preparar as respostas. Para estes autores, o discurso colaborativo engajado representa o melhor uso dos fóruns para fins educacionais. Ele constitui o diálogo do grupo em busca de entendimento compartilhado e de convergência. O termo engajado é utilizado para enfatizar que os participantes interagem uns com os outros em torno de questões substantivas. Os autores descrevem que a realização deste tipo de discurso é uma atividade difícil e demorada para professores e alunos.

De acordo com Klisc, McGill e Hobbs (2009), a interatividade e colaboração entre os alunos, o pensamento crítico e a aprendizagem colaborativa, têm sido identificados em diferentes estudos como importantes indicadores de sucesso dos fóruns de discussão.

O professor e/ou tutor possuem um papel fundamental para promover o sucesso dos fóruns de discussão. A presença pedagógica nas discussões é de extrema importância. Anderson et al. (2001) definem a presença pedagógica em uma conferência por computador com três características: projeto e administração, facilitação do discurso e instrução direta. Facilitar o discurso é fundamental para manter o interesse, motivação e engajamento dos alunos na aprendizagem ativa. O professor divide a responsabilidade com cada aluno para a realização dos objetivos de aprendizagem acordados. O docente apóia e incentiva a participação, comentando e incentivando as respostas dos discentes, e atraindo a participação

menos ativa. Os professores podem ser obrigados a ajudar os alunos a encontrar soluções quando duas opiniões aparentemente contrárias estão sendo expressas. Da mesma forma, ajudar os estudantes a articular o consenso e a compreensão compartilhada, quando estes já estão implícitos na discussão, também é útil.

Mazzolini e Maddison (2003) relatam que, dependendo da finalidade do fórum, os professores podem atuar de maneiras distintas. As discussões podem ser limitadas a um ou mais temas iniciados pelo docente, conduzir discussões mais gerais, assumir o papel de responder a maioria das perguntas dos alunos, moderar o debate mantendo uma pequena interação, ou até mesmo ser completamente ausente das discussões. O grau ideal de visibilidade do docente nas discussões depende muito do propósito dos fóruns.

Lim e Cheah (2003) efetuaram uma revisão da literatura sobre os papéis do professor no fórum (antes, durante e após o debate). Antes do início da discussão, o professor desenvolve os papéis:

- a) Gerencial: definindo quem serão os participantes do fórum, o tempo de duração do fórum, os tópicos a serem debatidos.
- b) Facilitação: explicando o uso e as características do fórum, delimitando expectativas, regras de conduta, diretrizes para a participação, estabelecendo regras processuais do fórum e definindo os objetivos da discussão.
- c) Pedagógico: uma das tarefas mais importantes do tutor é garantir que o tópico da discussão seja interessante, relevante e significativo. Boas perguntas promovem a participação ativa e estimulam diversos níveis de pensamento.

Durante a discussão, o professor desenvolve os papéis (LIM e CHEAH, 2003):

- a) Facilitação: propondo questões de pesquisa ou contradizendo pontos de vista, para fazer os discentes pensarem. O tutor chama a atenção para perspectivas opostas, direções diferentes ou opiniões conflitantes que podem levar a debates e críticas. O professor pode responder às perguntas que são colocadas pelos alunos, auxiliando ou reconhecendo as contribuições. O docente pode requerer que as opiniões e contribuições enviadas pelos alunos sejam suportadas com dados e argumentos racionais, e até mesmo reexaminados à luz do que os outros no grupo estão pensando. Além de solicitar elaboração e raciocínio, o professor pode incentivar os participantes a se engajarem na formulação das sentenças. Quando há diversos

pontos de vista na discussão, o docente pode resumir as postagens, e conceituar as contribuições e diferenças de opinião que surgiram. Quando há falta de contribuições dos alunos, o professor precisa incentivar a participação. No entanto, o docente não precisa desempenhar o papel do facilitador para todos os fóruns de discussão. Os alunos podem ser convidados a liderar e facilitar uma discussão sobre um tema específico relacionado a uma leitura prévia.

- b) Gerencial: no caso de surgirem conflitos entre os participantes, o professor deverá mediar. Este papel gerencial inclui a promoção de relações humanas, o desenvolvimento da coesão do grupo, e ajuda aos participantes para trabalharem juntos em uma causa comum. O docente deve elogiar os alunos que estão contribuindo de forma relevante no fórum, e solicitar mudança no comportamento dos alunos que não estão interagindo. O professor também precisa reagir e lembrar aos participantes, através de e-mail ou outro mecanismo, quando houver violação na etiqueta ou uso de uma linguagem imprópria.

Após a discussão, o professor desenvolve os papéis (LIM e CHEAH, 2003):

- a) Facilitação: uma atividade de conclusão bem planejada, útil e apropriada, oferece uma finalização adequada para a discussão e permite que os alunos avaliem a experiência. O professor pode resumir o debate no final e efetuar as devidas conclusões dos vários segmentos da discussão. Os alunos devem se engajar ativamente nas reflexões após o encerramento, a fim de internalizar os conceitos e questões relativas ao tema debatido.

Nesta seção foi destacado o uso pedagógico do fórum de discussão. Na próxima serão relatados alguns aspectos teóricos relacionados com a mineração de textos.

3.4 MINERAÇÃO DE TEXTOS

Esta seção apresenta o referencial teórico relacionado à área de mineração de textos, cujas técnicas serão utilizadas para desenvolver o MineraFórum. A seção 3.4.1 apresenta os princípios básicos da área. A seção 3.4.2 relata algumas pesquisas que demonstram como a utilização de informações adicionais pode auxiliar as técnicas de mineração de textos. A seção 3.4.3 descreve aplicações em algumas áreas do conhecimento. A seção 3.4.4 apresenta

especificamente algumas pesquisas na área de educação.

3.4.1 PRINCÍPIOS BÁSICOS

Esta seção descreve os principais conceitos, técnicas, e abordagens relacionadas à área de Mineração de Textos (MT).

Muitas definições para a MT são propostas como:

- De acordo com Tan (1999), a MT trata do processo de extrair padrões interessantes e não-triviais de conhecimento a partir de textos.
- Para Hearst (1999), a MT deve utilizar grandes conjuntos de textos para descobrir novos fatos e tendências sobre um determinado assunto.
- Segundo Dorre et al. (1999), a MT trata da extração de características a partir de documentos simples, bem como a análise da distribuição de características em grandes coleções de documentos, para detectar fenômenos interessantes, padrões, ou tendências.
- Segundo Feldman e Sanger (2007), a MT pode ser definida como um processo intensivo de conhecimento no qual um usuário interage com uma grande quantidade de documentos utilizando ferramentas para análise dos mesmos. O objetivo é extrair informações úteis a partir de coleções de documentos. Estas informações são identificadas em padrões interessantes nos dados textuais não estruturados.

Os sistemas de mineração de textos baseiam-se em rotinas de pré-processamento, algoritmos para descoberta de padrões, e elementos para apresentação dos resultados. As etapas que compõe a arquitetura de um sistema para mineração de textos são apresentadas na figura 8. O usuário do sistema interage com a etapa de pré-processamento, com o núcleo de mineração e com a apresentação dos resultados.

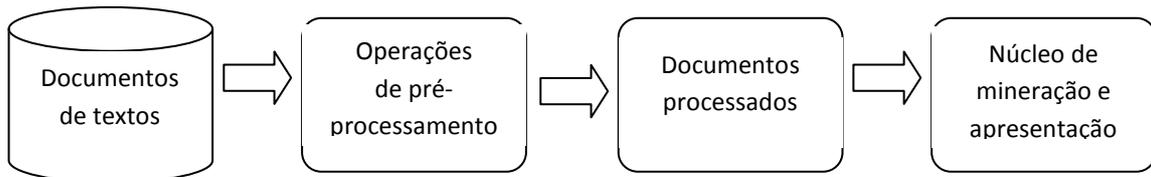


Figura 8 - Arquitetura de um sistema para mineração de textos
 Fonte: o autor, baseado em Feldman e Sanger (2007)

As operações de pré-processamento baseiam-se na identificação e extração de características representativas dos documentos em linguagem natural. Essas operações são responsáveis por transformar dados não estruturados, armazenados em coleções de documentos, em uma estrutura expressa em um modelo intermediário (FELDMAN e SANGER, 2007; TAN,1999).

Segundo Torre et al. (2005), os modelos intermediários são baseados na escolha da unidade mínima do texto. O quadro 3 apresenta a relação entre as unidades e algumas das possíveis representações. Destaca-se que podem ser realizadas abordagens híbridas e outros tipos de modelos.

Quadro 3 - Relação entre a unidade mínima do texto e algumas representações intermediárias

Unidade do texto	Representação intermediária
Palavra	Bag-of-words
	N-grams
	Hierarquia de conceitos
Conceito	Grafos conceituais
	Grafos semânticos
	Dependência conceitual
	N-frases
Frase	Frases textuais multi-termos
	Tendências
	Parágrafo
Parágrafo	N-frases
	Frases textuais multi-termos
	Tendências
	Documento
Documento	Documento

Fonte: Torre et al. (2005)

As operações que fazem parte do núcleo de mineração, também chamadas de processos de destilação do conhecimento, representam o coração de um sistema de mineração de texto, e incluem: descoberta de padrões, análise de tendências, e algoritmos incrementais para descoberta de conhecimento. Os mecanismos mais utilizados são as distribuições e as proporções, conjunto de conceitos frequentes, e associações. As atividades também podem estar relacionadas a comparações, e à identificação de níveis de interesse, com alguns padrões. Sistemas orientados a domínio também podem melhorar a qualidade das operações a

partir de consultas realizadas em bases de conhecimento (FELDMAN e SANGER, 2007).

Os componentes que fazem parte da apresentação dos resultados representam a interface do sistema, com funcionalidades para navegação, e acesso à linguagem utilizada para consultas (FELDMAN e SANGER, 2007; PURETSKIY et al., 2010; TAN, 1999).

A mineração de textos explora técnicas e metodologias provenientes das áreas de recuperação de informação, extração de informação, e linguística computacional de corpus. Para extrair informações úteis, são descobertas características relevantes nos documentos, cujas mais utilizadas são: caracteres, palavras, termos e conceitos. Os caracteres são as letras individuais, números, caracteres especiais e espaços. As palavras são representadas por agrupamentos de caracteres. Termos são palavras únicas ou conjunto de palavras selecionadas diretamente do texto. Conceitos são características geradas para um documento através de uma metodologia. Podem ser realizadas abordagens híbridas para a geração de representações de documentos baseadas em características. Por exemplo, pode-se primeiro extrair termos do texto, e então adaptá-los, comparando-os a uma lista de tópicos relevantes (conceitos), obtidos através de uma categorização (FELDMAN e SANGER, 2007).

Observando-se as quatro características descritas anteriormente (caracteres, palavras, termos e conceitos), os termos e conceitos refletem as que possuem melhores níveis de valor semântico. Existem muitas vantagens na utilização delas para representar documentos na mineração de textos. As representações utilizando termos podem ser mais facilmente geradas a partir do texto original se comparadas aos conceitos. No entanto, as representações com conceitos são melhores do que qualquer outra, podendo inclusive fazer uso de sinonímia⁹ e polissemia¹⁰, e também trabalhar com hipônimos e hiperônimos¹¹. Elas também podem ser processadas para suportar hierarquias muito sofisticadas, utilizando o conhecimento do domínio proporcionado por ontologias e bases de conhecimento (FELDMAN e SANGER, 2007).

As técnicas empregadas na mineração de textos incluem extração de informação, rastreamento de tópicos, produção de sumários, categorização de textos, agrupamento de textos, links conceituais, visualização de informação, análise de perguntas e respostas (FAN et

⁹ Sinonímia é o fato de haver mais de um vocábulo com a mesma ou quase a mesma significação (BECHARA, 2001).

¹⁰ Polissemia é o fato de um vocábulo ter mais de uma significação (BECHARA, 2001).

¹¹ Hiperônimo é uma palavra cujo significado é mais abrangente do que o do seu hipônimo. Por exemplo, veículo é hiperônimo de carro porque em seu significado está contido o de carro. Carro é um hipônimo de veículo (CIPRO NETO e INFANTE, 1998).

al., 2006; GUPTA e LEHAL, 2009).

A extração de informação identifica as frases principais e os relacionamentos dentro de um texto. O procedimento é realizado através da busca de sequências previamente definidas, sendo este um processo chamado de casamento de padrões. A técnica infere os relacionamentos entre pessoas, lugares, datas, para fornecer informação significativa ao usuário (FAN et al., 2006; MOONEY e NAHM, 2003).

O rastreamento de tópicos mantém armazenados os perfis dos usuários, e baseado nos documentos que o usuário visualiza, prevê outros que possam interessá-lo. O rastreamento pode ser utilizado, por exemplo, para descobrir as referências de uma área de pesquisa (GUPTA e LEHAL, 2009; MAHALAKSHMI e SENDHILKUMAR, 2009).

Os links conceituais conectam documentos relacionados a partir da identificação de conceitos comumente compartilhados. Esta técnica ajuda os usuários a encontrar informações que talvez não fossem descobertas com métodos tradicionais de busca. Os links conceituais favorecem a navegação na informação ao invés de auxiliar a sua busca (FAN et al., 2006; GUPTA e LEHAL, 2009).

A visualização de informação apresenta grandes quantidades de documentos em uma hierarquia visual, ou em um mapa, fornecendo opções de navegação, além da realização de pesquisas (FELDMAN e SANGER, 2007; PURETSKIY et al., 2010; TAN, 1999).

A análise de perguntas e respostas é proveniente da área de processamento de linguagem natural, que trata de como encontrar a melhor resposta para uma respectiva pergunta (GUPTA e LEHAL, 2009; RAMAKRISHNAN e BHATTACHARYYA, 2009).

A produção de sumários (sumarização) é útil para tentar descobrir se um documento extenso atende ou não as necessidades do usuário. O objetivo da sumarização é reduzir o tamanho e o nível de detalhamento de um texto, retendo os pontos principais e o significado geral. A grande dificuldade para o software ainda é fazer a análise semântica e a interpretação de significado. Uma das estratégias mais utilizadas pela técnica é a extração de sentenças importantes do texto, a partir do peso estatístico. Algumas heurísticas adicionais, tais como informações sobre posicionamento também são empregadas. Os métodos de sumarização podem ser classificados em dois grandes grupos: abordagens superficiais, que são restritas ao nível sintático da representação; e as abordagens profundas, que envolvem um nível de representação semântico do texto original e utilizam processamento lingüístico (CASTELLANOS, 2004; GUPTA e LEHAL, 2009).

A categorização (classificação) envolve a identificação dos principais temas de um documento, colocando o mesmo em um conjunto pré-definido de tópicos (GUPTA e LEHAL, 2009; HOWLAND e PARK, 2004; LIU, 2009). Neste processo, o texto é tratado como um conjunto de palavras, ao invés de realizar o processamento da informação real, como na extração de informação. A categorização conta as palavras que aparecem no documento, e a partir desta contagem, identifica os principais tópicos. O processo pode utilizar um *thesaurus*¹² no qual os tópicos são pré-definidos, e os relacionamentos podem ser identificados procurando termos gerais, próximos, sinônimos. A categorização geralmente utiliza um método de classificação dos documentos, para ordenar aqueles que possuem mais conteúdo sobre um tema específico.

O agrupamento de textos (*clustering*) é uma técnica utilizada para agrupar documentos semelhantes. Ela é diferente da categorização, pois ao invés do uso de tópicos pré-definidos, os textos são agrupados em tempo real. No agrupamento, os documentos podem aparecer em vários subtópicos. Um algoritmo de *clustering* básico cria um vetor de tópicos para cada texto, e calcula os pesos para identificar em qual grupo ele deve fazer parte (GUPTA e LEHAL, 2009; KOBAYASHI e AONO, 2004).

Um modelo de representação de documento bastante utilizado nas técnicas de categorização e agrupamento de textos é o modelo de espaço vetorial (*vector space model*), desenvolvido por Salton (1971). Ele também é chamado de abordagem *bag-of-words* (SALTON e MCGILL, 1983). Nesta abordagem, um documento é representado por um vetor de termos ponderados. Os pesos podem ser calculados pela fórmula TF-IDF indicada abaixo (SALTON e BUCKLEY, 1988).

$$TF\text{-}IDF\text{ Weight}(w,d) = TermFreq(w,d) \cdot \log(N / DocFreq(w))$$

TF-IDF Weight(w,d): peso da palavra w, no documento d

TermFreq(w, d): frequência da palavra no documento

N: quantidade total de documentos

DocFreq(w): número de documentos contendo a palavra w

De acordo com Feldman e Sanger (2007), os algoritmos de *clustering* devem reunir textos semelhantes. O agrupamento geralmente é especificado de acordo com uma função de

¹² Um *thesaurus* (tesauro) é um vocabulário controlado e dinâmico de descritores relacionados semântica e genericamente, que cobre de forma extensiva um ramo específico do conhecimento (FERREIRA, 1999).

similaridade entre os objetos. No modelo de espaço vetorial, esta função geralmente é calculada com base na distância entre os vetores. Dois métodos de cálculo são a distância Euclidiana e a medida de similaridade cosseno. As fórmulas dos mesmos são apresentadas abaixo, onde x_i e x_j são vetores no espaço de características dimensionais, e x' é o vetor normalizado $x/|x|$.

$$D(x_i, x_j) = \sqrt{\sum (x_{ik} - x_{jk})^2}$$

$$Sim(x_i, x_j) = (x'_i \cdot x'_j) = \sum x'_{ik} \cdot x'_{jk}$$

3.4.2 UTILIZAÇÃO DE INFORMAÇÕES ADICIONAIS PARA AUXILIAR A MINERAÇÃO DE TEXTOS

Esta seção apresenta algumas pesquisas que demonstram como a utilização de informações lingüísticas, e informações sobre o domínio de conhecimento a ser analisado, podem auxiliar as técnicas de mineração de textos. Neste contexto, um domínio pode ser definido como uma área específica com ontologias¹³ e taxonomias¹⁴.

De acordo com Feldman e Sanger (2007), nos sistemas de mineração de textos, os conceitos não possuem somente atributos descritivos de um documento, mas também pertencem a domínios. A partir das ontologias, os sistemas podem criar representações mais completas das coleções de documentos e permitir buscas mais refinadas. O conhecimento do domínio pode ser utilizado nas operações de pré-processamento para melhorar a extração de conceitos e as atividades de validação.

Hotho et al. (2003) integraram ontologias no processo de agrupamento de textos. As avaliações experimentais dos autores compararam técnicas de agrupamento baseadas na pré-categorização de textos do Reuters-21578¹⁵ e de um domínio menor de um curso sobre Java. Nas experiências com o grande conjunto de documentos foi utilizada a ontologia da WordNet¹⁶. Nos testes com os textos do curso sobre Java, foi utilizada a WordNet em alguns experimentos, e em outros, uma ontologia específica. Os experimentos demonstram que o uso de ontologias produz melhores resultados no processo de agrupamento de textos.

¹³ Segundo Gruber (1993), uma ontologia é uma especificação explícita de uma conceituação dos objetos, conceitos e outras entidades, em alguma área de interesse e as relações entre eles.

¹⁴ Teoria das classificações (FERREIRA, 1999).

¹⁵ <http://www.daviddlewis.com/resources/testcollections/reuters21578/>

¹⁶ WordNet é um grande banco de dados léxico da língua inglesa, disponível em <http://wordnet.princeton.edu>

Bloehdorn et al. (2005) descrevem em seu artigo alguns experimentos realizados com técnicas de mineração de textos na área médica. No estudo, os autores exploraram o conhecimento prévio oferecido por ontologias para estender o modelo *bag-of-words* com características conceituais em um nível semântico mais elevado. Os resultados do trabalho indicam que a utilização de ontologias melhora os processos de classificação e agrupamento de textos.

Na pesquisa apresentada por Shehata et al. (2007) é demonstrado como um modelo baseado em conceitos do domínio pode melhorar o processo de categorização de textos. O modelo proposto pode discriminar termos importantes com relação à semântica das sentenças. O modelo consiste de um analisador estatístico baseado em conceitos, uma representação na forma de um grafo ontológico conceitual, e um extrator de conceitos.

No trabalho de Loh et al. (2009) é apresentada uma abordagem baseada em conceitos para realizar a mineração de textos. Segundo os autores, minerar conceitos ao invés de palavras-chaves minimiza os problemas de vocabulário que ocorre no uso da linguagem natural. Os conceitos são características textuais de alto nível e, portanto, são melhores do que simples palavras, para representar objetos do mundo real, eventos, ou fenômenos. No estudo, os conceitos são utilizados para representar o conteúdo de um texto ou uma coleção de documentos. Para os pesquisadores, a eficiência do uso de conceitos para representar o conteúdo de um texto é maior porque mais e diferentes termos são cobertos.

O trabalho de Rigo et al. (2007) descreve como uma classificação de textos pode ser efetuada com o uso de informações lingüísticas. O trabalho utiliza uma ontologia de domínio para descrever as características necessárias para identificar a estrutura e os conceitos dos textos, relacionados a uma classe específica.

Com base nas pesquisas apresentadas nesta seção, observa-se que a utilização de ontologias melhora o processo de mineração de textos. De acordo com Feldman e Sanger (2007), um *thesaurus* também pode ser usado como um conhecimento de domínio generalizado para auxiliar na mineração. O MineraFórum utiliza um *thesaurus* para identificar os sinônimos das palavras com maior ocorrência no texto.

3.4.3 ÁREAS DE APLICAÇÃO

As técnicas de mineração de textos podem ser utilizadas em diversas áreas do

conhecimento. Esta seção apresenta aplicações em algumas áreas.

Feldman e Sanger (2007) citam algumas áreas nas quais a mineração de textos tem sido utilizada: nas finanças corporativas, na pesquisa de patentes, e nas ciências da vida. Nas finanças corporativas, o objetivo é filtrar grandes quantidades de dados textuais para criar formas de utilização de *business intelligence*, anotando tendências, identificação de correlações, e pesquisando referências a operações específicas, pessoas jurídicas ou físicas. Na pesquisa de patentes, o objetivo é investigar as estratégias de desenvolvimento de patentes e encontrar formas para explorar melhor ativos de patentes corporativas existentes. Como exemplo de utilização nas ciências da vida, alguns pesquisadores estão explorando grandes coleções de relatos de pesquisa biomédica para identificar padrões complexos de interatividade entre proteínas.

Tsuruoka (2008) apresenta um sistema de busca textual denominado FACTA, apresentado na figura 9. O sistema pesquisa vários documentos da área biomédica para encontrar conceitos que possuem associação entre si.

The screenshot shows the FACTA search interface. At the top, there is a search bar containing 'apoptosis AND blood' and buttons for 'Find Relevant Concepts' and 'Search MEDLINE'. Below the search bar, there are checkboxes for various categories: Gene/Protein, Disease, Symptom, Drug, Enzyme, and Compound, all of which are checked. The search results are displayed in a table format, showing the frequency of various concepts found in the documents.

Human Gene/Protein	Disease	Symptom	Drug	Enzyme	Compound
Fas	tumor	anesthesia	IL 2	caspase 3	DNA
cytokine	cancer	collage	IFN-gamma	mitogen-activated protein kinase	oxygen
Bcl-2	HIV	functional recovery	CSF	protein kinase	iodine
CD4	ischemia	splenomegaly	G-CSF	protein kinase	dUTP
caspase-3	leukemia	pain	Angiotensin II	caspase-8	glucose
TNF-alpha	immunodeficiency	diarrhea	vitamin E	ERK	nitric oxide
Fas ligand	tumor growth	hypothermia	Arsenic trioxide	reverse transcriptase	oside
apoptin V	diabetes	sedation	MAD	TdT	calcium

Figura 9 - Resultados de uma busca no sistema FACTA

Fonte: Tsuruoka (2008)

Qi (2009) cita a possibilidade da utilização da mineração de textos na Bioinformática. De acordo com o autor, como muitos documentos da literatura de estudo da Biologia estão armazenados como modelos semi-estruturados e não estruturados, a mineração pode desempenhar um papel importante para ajudar a encontrar as informações necessárias e os relacionamentos dos dados biológicos.

Dreweke et al. (2009) apresentam como a mineração de textos pode ser utilizada em código de programação, para encontrar trechos duplicados em grandes projetos de software.

Desta forma, os softwares ficam mais compreensíveis e ajudam a evitar erros em caso de ser realizada uma reengenharia dos programas.

Xu e Luo (2009) explicam que a mineração de textos também é uma ferramenta importante na área dos sistemas de segurança da informação. Em seu artigo, os autores discutem várias aplicações tecnológicas empregadas na área de segurança. As técnicas são classificadas de acordo com o tipo de conhecimento a ser descoberto e os formatos de texto a serem analisados.

O artigo de Yoo et al. (2009) apresenta uma análise de redes sociais pessoais para capturar grupos e obter características importantes que representem o perfil de um usuário específico. Os autores também desenvolveram um algoritmo de aprendizagem semi-supervisionado que propaga rótulos de importância para testar exemplos de mensagens e nós de usuários em uma rede de e-mail pessoal. Foi utilizada uma representação de vetor para cada mensagem de e-mail. Cada vetor possui os dados da mensagem e as características sociais do emitente e do destinatário. Esta representação foi usada como a entrada de classificadores SVM (*Support Vector Machine*) para prever o nível de importância de cada mensagem.

No trabalho de Glance et al. (2005) é apresentado um sistema para obter inteligência de marketing, que reúne e anota discussões relacionadas com produtos de consumo. O sistema utiliza diversas técnicas, incluindo rastreamento, acondicionamento, pesquisa, classificação de textos e lingüística computacional. O trabalho mostra que um sistema de análise interativa pode ser utilizado para obter informações sobre inteligência de marketing a partir de grandes quantidades de relatos.

3.4.4 APLICAÇÕES NA EDUCAÇÃO

As técnicas de mineração de textos também podem ser utilizadas para auxiliar a área de educação. Esta seção cita algumas pesquisas realizadas.

Chen N. et al. (2008) apresentam uma pesquisa para construir mapas conceituais utilizando técnicas de mineração de textos. Para construir um mapa, foi realizada uma análise de diversos artigos acadêmicos relacionados a um domínio específico. As palavras-chaves listadas nos documentos foram colocadas nos nós, representando os conceitos. As ligações entre os conceitos foram obtidas a partir de uma medida de distância entre as palavras-chaves

nos textos.

Romero e Ventura (2007) analisam diversos trabalhos que aplicaram técnicas de mineração de dados e textos em cursos baseados na *Web*, sistemas de gestão de conteúdos de aprendizagem, e sistemas educacionais inteligentes e adaptativos baseados na *Web*. Algumas pesquisas citadas pelos autores são:

- Ueno (2004) utilizou a mineração de textos para efetuar uma análise de correspondência expandida em fóruns de discussão. Os alunos selecionam a categoria que representa sua mensagem e o sistema fornece avaliações para os comentários do aluno entre os colegas.
- Chen et al. (2005) propõem a construção automática de um livro eletrônico através de mineração de conteúdo *Web*. Eles empregaram uma estratégia de classificação para avaliar a adequação de páginas *Web*, extraíram características, e estabeleceram hierarquias de conceitos.
- Tane et al. (2004) apresentam uma ferramenta baseada em ontologia para aproveitar recursos disponíveis na *Web*. Eles utilizaram técnicas de mineração de textos para construir ontologias a partir de conjuntos de documentos.
- Tang et al. (2000) desenvolveram um tutor *Web* personalizado, através da mineração do contexto e estrutura de um curso. Eles usaram um algoritmo de mineração de textos direcionado a palavras-chaves, que seleciona artigos para alunos de educação a distância.

Mochizuki et al. (2005) descrevem um método de auto-avaliação para os alunos em uma discussão colaborativa. Os autores propõem uma forma para visualizar a discussão dos discentes em um BBS (*Bulletin Board System*), que utiliza uma técnica de mineração de textos para avaliar a conversação. O método extrai palavras-chaves da discussão e usa análise de correspondência para visualizar a relação delas com o assunto tratado. O procedimento também produz um mapeamento que indica se o aluno fez referência a cada palavra-chave em suas mensagens. No trabalho, os autores desenvolveram um software chamado i-Bee (*Bulletin board Enrollee Envisioner*).

Dalmolin et al. (2009) demonstram uma ferramenta para auxiliar no processo de elaboração de mapas conceituais. A ferramenta foi desenvolvida para organizar conteúdos a partir dos módulos temáticos de um curso. Na ferramenta, os professores podem construir

seus próprios mapas baseados em documentos textuais. Os docentes também podem utilizar um editor visual para criar mapas e vincular objetos de aprendizagem a conceitos. Os alunos podem acessar objetos de aprendizagem e fazer anotações em seus mapas, melhorando o processo de ensino-aprendizagem. O extrator de mapa conceitual é um algoritmo baseado em técnicas de mineração de texto, que extrai os termos relevantes considerados conceitos ou links.

Ferreira et al. (2009) apresentam uma pesquisa sobre a realização da sumarização de textos provenientes de repositórios de arquivos textuais e blogs. De acordo com os autores, os sumários auxiliam as pessoas a identificar mais rapidamente, se o conteúdo do texto original é útil aos seus interesses. O trabalho cita que a sumarização de textos é uma alternativa viável a ser incorporada em Ambientes Educacionais na *Web*, pois pode ajudar alunos e professores a diagnosticar, com maior velocidade, textos que podem atender as suas necessidades.

Barbosa et al. (2009) apresentam um estudo comparativo entre duas ferramentas para mineração de textos, Sobek e Tag Clouds. A mineração foi executada em textos selecionados em blogs. Uma das contribuições do artigo, baseando-se em um contexto educacional, foi que o Sobek possui uma capacidade maior para extrair informações relevantes, uma vez que relaciona conceitos e exibe resultados mais completos.

3.5 SOFTWARES PARA MINERAÇÃO DE TEXTOS

Esta seção apresenta alguns softwares de mineração que foram investigados. O intuito foi identificar qual deles atenderia ao problema desta pesquisa, que é analisar como um recurso de mineração de textos, incorporado a um ambiente virtual de aprendizagem, pode auxiliar o professor na análise qualitativa das contribuições em um fórum de discussão.

Destaca-se que, além de solucionar o problema desta pesquisa, o recurso de mineração de textos a ser utilizado nas mensagens, deve apresentar indicadores que auxiliem o professor na análise dos fóruns.

Alguns dos softwares analisados foram observados no relatório técnico¹⁷ preparado por Jan van Gemert. Ele está nas séries dos relatórios técnicos do *Intelligent Sensory*

¹⁷ <http://staff.science.uva.nl/~jvgemert/pub/textminingtools.pdf>

Information Systems, do Departamento de Ciência da Computação da Universidade de Amsterdam. Neste, o autor faz uma análise de 71 softwares para mineração de textos disponíveis na internet. No relatório é apresentada uma breve descrição de cada software e da empresa que o criou, bem como o preço, a disponibilidade de uma versão de demonstração, e o site do mesmo. Ao final do relatório é apresentada uma tabela com as características de todos os softwares analisados.

Os softwares pesquisados foram:

a) Bow

O Bow (McCALLUM, 1996) é uma biblioteca escrita na linguagem de programação C que pode ser utilizada na mineração de textos. Alguns softwares foram construídos com a Bow, como o Rainbow (para realizar a classificação de documentos), o Arrow (recuperação de documentos) e o Crossbow (agrupamento de documentos). O Rainbow é capaz de realizar classificação de documentos utilizando vários métodos, entre eles, Naive Bayes, TFIDF, K-nearest neighbor. A execução do Rainbow consiste em analisar um conjunto de textos e criar um modelo indexado das palavras relevantes constantes nos mesmos.

b) Eureka

O Eureka¹⁸ foi proposto por Wives (1999). O Eureka efetua agrupamento de documentos. No momento, ele é disponibilizado na modalidade freeware, para fins educativos. A figura 10 apresenta a interface do Eureka.

¹⁸ <http://www.leandro.wives.nom.br/index.htm>

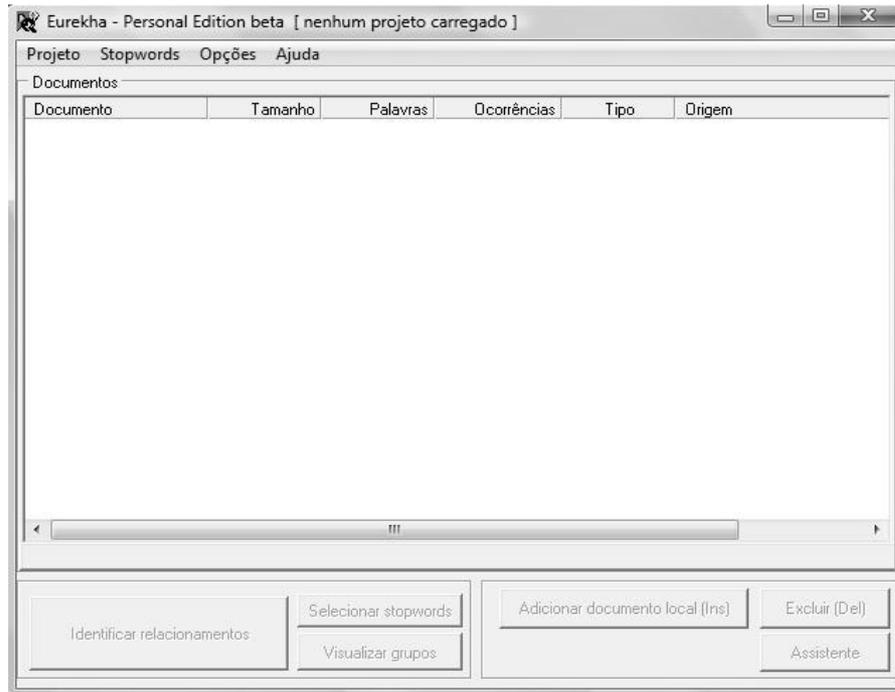


Figura 10 - Interface do Eureka
Fonte: Wives (1999)

c) Text-Garden - Text Mining Software Tools

O Text-Mining Software Tools¹⁹ permite trabalhar com documentos textuais e páginas *Web* para realizar análises como: classificação, agrupamento, visualização. O software foi desenvolvido em C++ por Marko Grobelnik, Dunja Mladenic e outros colaboradores.

A figura 11 apresenta a interface do Bow2NNbrs, um software que faz parte destas ferramentas, cujo objetivo é classificar documentos que estão no formato *bag-of-words*, utilizando o algoritmo *nearest-neighbour*.

```
C:\a1>Bow2NNbrs
Bag-Of-Words To Nearest-Neighbours [Sep 30 2004]
=====
Input-BagOfWords-FileName (-i:)=
Output-NNeighbours-Xml-File (-oxml:)=NNbrs.Xml
Output-NNeighbours-Txt-File (-otxt:)=NNbrs.Txt
Output-To-Screen (-oscr:)=Yes
Output-Top-NNeighbour-Documents (-topdocs:)=10
Output-Top-NNeighbour-Words (-topwords:)=10
Query-DocumentId (-qid:)=-1
Query-All-DocumentIds (-qall:)=No
Cut-Word-Weight-Sum-Percentage (-cutww:)=0
Minimal-Word-Frequency (-mnwfg:)=0
=====
```

Figura 11 - Interface do Bow2NNbrs
Fonte: <http://kt.ijs.si/Dunja/textgarden/>

¹⁹ <http://kt.ijs.si/Dunja/textgarden/>

d) TML - Text Mining Library

TML²⁰ é uma biblioteca de múltiplos propósitos para mineração de textos, completamente integrada ao software Apache Lucene, oferecendo recursos aos desenvolvedores que necessitem integrar funcionalidades de mineração de textos nas suas aplicações. Algumas tarefas disponibilizadas são: indexação de documentos, redução de dimensionalidade, anotações de metadados, agrupamento de tópicos, extração de palavras-chaves, entre outras.

e) TagCrowd

O Tagcrowd²¹ é uma aplicação *Web* cujo objetivo é criar uma *tag cloud* (nuvem de palavras) de um texto, permitindo visualizar a frequência das palavras. Ele foi desenvolvido por Daniel Steinbock na Universidade Stanford. A figura 12 apresenta a interface do TagCrowd.

Figura 12 - Interface do TagCrowd

Fonte: <http://tagcrowd.com/>

²⁰ <http://sourceforge.net/projects/tml-java/>

²¹ <http://tagcrowd.com/>

f) Sobek

O software Sobek, desenvolvido por Lorenzatti (2007), apresentado na figura 13, é um exemplo de utilização de mineração de textos baseado em métodos estatísticos. Ele permite construir um grafo a partir de informações estatísticas obtidas do texto. Neste grafo, os vértices e as arestas possuem informações referentes ao número absoluto e relativo de ocorrência dos termos (vértices) e relacionamentos (arestas) em determinado documento. O grafo obtido representa uma rede dos conceitos que foram trabalhados dentro do texto.



Figura 13 - Interface do Sobek
Fonte: Lorenzatti (2007)

A primeira etapa do processamento realizada pelo Sobek compreende a análise léxica, onde o texto é desmembrado palavra por palavra. Depois deste passo, as palavras são submetidas à análise estatística. Com base nos dados estatísticos, é criada uma “base de conceitos” que auxiliará na construção do grafo. No Sobek, uma palavra ou um conjunto de palavras pode ser considerado um conceito. O passo seguinte consiste na remoção de palavras

que não agregam significado ao texto (*stop words*²²). Feito isso, e em posse da análise estatística, o grafo é gerado. Neste estágio, parâmetros como frequência absoluta e tamanho do conceito (se simples ou composto) são analisados.

O pseudo-algoritmo realizado pelo Sobek para o processamento e análise dos textos é apresentado abaixo (LORENZATTI, 2007). O valor de N indica a quantidade de termos que podem ser utilizados para formar uma palavra composta. Por exemplo, se N=2 pode-se compor as palavras compostas: "motor elétrico", "eclipse solar", entre outras.

1. Carga dos perfis de configuração.
2. Seleção do perfil desejado (definição dos parâmetros, partes do texto, e outros).
3. Entrada de um texto puro (sem quaisquer formatações).
4. Extração das linhas do texto para dar início ao processo.
5. Carga da lista de *stopwords*.
6. Extração das palavras contidas no texto (tokens).
7. Se $N > 1$
 8. Criação dos conceitos combinados.
9. Fim Se
10. Remoção dos conceitos de acordo com seu tamanho e número de ocorrências.
11. Remoção dos caracteres de pontuação.
12. Remoção das *stopwords*.
13. Construção dos relacionamentos entre os termos (grafo).
14. Desenho do grafo representando o texto.

Após a investigação dos softwares apresentados nesta seção, verificou-se que nenhum deles oferece funcionalidades para analisar qualitativamente as mensagens de um fórum de discussão, e apresentar indicadores que auxiliem o professor. Além disso, eles não estão integrados a ambientes virtuais de aprendizagem.

Cabe ressaltar que o Sobek realiza a mineração de textos gerando um grafo que

²² As *stopwords* são palavras que podem ser descartadas no processo de mineração, como: advérbios, artigos, preposições, entre outras.

representa os conceitos que mais foram citados. Desta forma, a programação referente à técnica de mineração do Sobek será utilizada no MineraFórum.

A seção 4.2 explica quais características do Sobek foram utilizadas pelo MineraFórum e apresenta os motivos pelo qual foi realizada a escolha deste software.

3.6 TRABALHOS CORRELACIONADOS

Esta seção apresenta algumas pesquisas que têm sido realizadas com o intuito de auxiliar na análise de fóruns de discussão. A seção 3.6.1 apresenta alguns trabalhos realizados que não utilizaram técnicas de mineração de textos. A seção 3.6.2 relata algumas pesquisas que efetuaram a análise de fóruns empregando as referidas técnicas e realiza uma discussão com as características do MineraFórum.

3.6.1 PESQUISAS SOBRE ANÁLISE DE FÓRUNS DE DISCUSSÃO

Rourke et al. (2001) explicam como avaliar a presença social, que é um dos elementos do modelo de comunidade de investigação proposto por Garrison et al. (2000). De acordo com os autores, a presença social é definida como a capacidade dos alunos para projetar-se socialmente e afetivamente. Um modelo para avaliar a presença social em conferências por computador baseadas em texto é apresentado através da análise de conteúdo das transcrições. A expressão da emoção, de sentimentos e humor é uma característica definidora da presença social, como descrito por Garrison et al. (2000). O artigo conclui com uma discussão sobre as implicações e benefícios sociais para avaliar a presença de professores, moderadores e pesquisadores.

Edelstein e Edwards (2002) destacam que é possível avaliar fóruns de discussão com critérios (*rubrics*) bem planejados. Os critérios podem fornecer *feedback* da participação dos estudantes e oferecer mecanismos para medir e documentar o progresso dos mesmos. Na pesquisa, os autores aplicaram um critério que possui diversas categorias. Foram escolhidas aquelas que refletem os itens que devem ser considerados pelos tutores na estruturação de comunidades de aprendizagem. As categorias utilizadas foram: iniciativa, distribuição de postagens, relevância das mensagens, expressão do aluno nas postagens, contribuição para a

comunidade de aprendizagem. A avaliação de cada categoria é realizada com uma pontuação de um a quatro.

Fahy (2003) analisou transcrições de fóruns de discussão para investigar a existência de estratégias de apoio entre os participantes. O autor relata a ocorrência de 13 estratégias, de acordo com as categorias encontradas em um instrumento chamado TAT (*Transcript Analysis Tool*). O estudo foi realizado com três grupos de transcrições produzidas por alunos em cursos *online*. As estratégias de apoio mais utilizadas foram: as afirmativas referenciais (aquelas que fazem referência a comentários anteriores), assinaturas, cumprimentos e questões horizontais (questões abertas que convidam o debate de uma resposta plausível). A codificação e análise das transcrições foi realizada com um software de análise qualitativa (ATLAS.ti) e com um de estatística (SPSS-PC).

Anderson (2004) assinala que é importante existir um sistema que auxilie na organização e visualização das postagens em um fórum de discussão. O sistema deve facilitar a exibição das contribuições textuais realizadas por cada aluno. É interessante que existam recursos para: analisar as mensagens dos estudantes para identificar erros ortográficos e gramaticais, contabilizar a quantidade de palavras escritas, permitir a exibição de postagens anteriores ou posteriores, possibilitar a visualização das datas das mensagens, permitir a avaliação das postagens, automaticamente alertar os estudantes quando uma avaliação for realizada. Com relação à análise de fóruns de discussão, o autor cita a referência bibliográfica de um critério para avaliar o desempenho dos estudantes. Este critério sugere a investigação de algumas características como: a quantidade das postagens e as respostas redigidas com demonstração de conhecimento.

Meyer (2004) utilizou quatro modelos para analisar dezessete fóruns de discussão: julgamento reflexivo de King e Kitchener (1994), modelo de Perry de desenvolvimento ético e intelectual (PERRY, 1999), pensamento crítico de quatro estágios (GARRISON et al., 2001), taxonomia de Bloom dos objetivos educacionais (BLOOM, 1956). King e Kitchener (1994) elaboraram um modelo de sete estágios de julgamento reflexivo, que captura as habilidades de raciocínio e desenvolvimento dos estudantes. O processo de avaliação é baseado na apresentação de situações complexas para os alunos e na análise de como eles pensam sobre os problemas. Perry (1999) criou um modelo de nove estágios que rastreia o desenvolvimento dos alunos. Garrison et al. (2001) desenvolveram um modelo de processamento cognitivo de quatro estágios para avaliar habilidades de pensamento crítico em fóruns de discussão. O modelo estabelece indicadores e exemplos para ajudar a classificar as

respostas dos discentes em: ativação, exploração, integração e resolução. A taxonomia de Bloom (1956) foi empregada para avaliar as postagens das discussões, de acordo com as categorias previstas. O estudo conclui que cada modelo é útil e focaliza a atenção sobre diferentes aspectos do pensamento do aluno. Alguns modelos são mais difíceis de empregar do que outros, necessitando de um treinamento específico e/ou a adaptação do tema das discussões.

O trabalho de Nisbet (2004) apresenta uma ferramenta para medir a interação de grupos de discussão, chamada DIGIT (*Discussion Group Interaction Tool*). Ela foi utilizada em diferentes casos para avaliar a quantidade e a qualidade da interação. A ferramenta foi aplicada para avaliar o crescimento (ou não) da interação longitudinal dentro de um grupo e com grupos em diferentes estágios de progressão. Embora os resultados apresentados pelo software sejam claros, o processo de avaliação é demorado. A ferramenta utiliza o modelo de cinco estágios de Salmon (2000) para medir a qualidade da interação (acesso e motivação, socialização *online*, troca de informação, construção de conhecimento, desenvolvimento). As métricas quantitativas referem-se à quantidade e tamanho das contribuições textuais.

O estudo de Marra et al. (2004) compara a aplicação de dois protocolos de análise de conteúdo para fóruns de discussão. A pesquisa fornece uma descrição detalhada de como usar os protocolos e avalia as potencialidades e fraquezas de ambos. O primeiro modelo de análise de conteúdo utilizado no estudo foi o IAM (*Interaction Analysis Model*), desenvolvido por Gunawardena et al. (1997). O IAM foi idealizado com o objetivo de compreender melhor os processos de negociação e construção de conhecimento em um ambiente de discussão colaborativa *online*. O segundo protocolo utilizado no estudo é o de Newman et al. (1995, 1997). Ele é baseado no modelo de pensamento crítico de cinco estágios de Garrison (1992): identificação, definição, exploração, avaliação, integração. Garrison (1992) e Newman et al. (1995, 1997) instanciaram indicadores do pensamento crítico através de aproximadamente 40 códigos, estruturados em categorias como: relevância, justificativa, novidade, e ambiguidades. Cada código possui um sinal positivo ou negativo para indicar se a instrução codificada contribui ou diminui o desenvolvimento do pensamento crítico.

O artigo de Saltz et al. (2004) emprega redes sociais para realizar uma análise das interações de cada aluno. O trabalho propõe a criação de um grafo social do estudante, que é uma rede social. A partir do grafo, o professor pode visualizar melhor as interações do discente com outros colegas. Os autores citam um estudo piloto realizado com a aplicação do grafo social.

Pelz (2004) utiliza um processo de dois estágios para avaliar as postagens dos alunos. Primeiramente, o conteúdo de cada mensagem é analisado. Depois, as postagens são avaliadas entre si. Com este processo, uma mensagem será considerada excelente se ela for precisa, original, relevante, e bem escrita.

Ware e Kramsch (2005) discutiram os desafios da educação a distância para professores de línguas. O artigo apresenta em detalhes um episódio que ocorreu entre dois alunos relatando suas versões de uma história durante a aula. O debate foi realizado em um projeto colaborativo assíncrono entre estudantes de alemão nos Estados Unidos e discentes de inglês na Alemanha. O estudo chama atenção para questões que geralmente são invisíveis na comunicação intercultural: a natureza do assunto, as condições das trocas interlinguísticas, a linguagem do discurso, e os objetivos da educação de língua estrangeira. O trabalho ressalta que os professores são muito importantes para auxiliar os alunos na comunicação intercultural.

A pesquisa de Ho e Swan (2007) investigou as participações efetivas e dinâmicas que ocorrem em fóruns de discussão e sua relação com os resultados da aprendizagem dos alunos. A abordagem de análise de conteúdo foi utilizada para investigar processos sócio-cognitivos dos discentes em uma aula de gramática de Inglês. As mensagens dos estudantes foram avaliadas utilizando o critério de pontuação do Princípio Cooperativo de Grice (1989) para avaliar a participação dos alunos. Em um estudo de Ho (2004), os elementos da teoria de Grice foram adaptados para ambientes *online*, da seguinte forma: quantidade (a quantidade de informação é suficiente para indicar o objetivo da mensagem); qualidade (a postagem é uma contribuição nova, reflexo da opinião do aluno, e é apoiada por exemplos precisos); relevância (a mensagem refere-se ao tema do debate); estrutura (a postagem é logicamente organizada, não possui erros de pontuação, ortográficos, gramaticais, e o significado da mensagem é claramente apresentado). Um foco específico do estudo foi explorar as relações entre os elementos de Grice nas postagens das discussões.

Chen e Chiu (2008) realizaram um estudo do fluxo das discussões em fóruns, examinando como as mensagens postadas inicialmente afetam as que foram redigidas posteriormente. A análise envolveu cinco dimensões: (1) avaliações (concordância, discordância, ou ações sem resposta); (2) conteúdo de conhecimento (contribuição, repetição, ou conteúdo nulo); (3) sinais sociais (positivo, negativo ou nenhum); (4) informações pessoais (número de visitas) e (5) elicitación (obtenção de respostas ou não). O estudo analisou 131 mensagens em sete tópicos nos fóruns de matemática de um sistema universitário. A pesquisa

relata que a discordância, contribuição, sinal social, e as visitas feitas a mensagens anteriores podem afetar as propriedades de uma mensagem subsequente. Depois de discutir os resultados, os autores consideraram as suas implicações para melhorar as discussões. O estudo indica que os professores podem gerenciar as discussões no nível das mensagens, para promover o pensamento crítico e facilitar o debate de temas controversos.

O artigo de Han e Hill (2006) descreve um estudo que investiga como a aprendizagem dos alunos se reflete nos fóruns. Ao examinar discussões utilizando análise do discurso, cinco diferentes tipos de discurso foram identificados: estabelecimento de metas, reflexão, conexão, reformulação original, e reorientação. Com os diferentes tipos de discurso, os autores explicaram como cada mensagem (ou conjunto de mensagens) facilitou a discussão, particularmente em termos de esforços colaborativos para atingir as metas. O estudo também descreve as questões metodológicas relacionadas com a análise do discurso *online* e discute as implicações para a pesquisa e prática.

O estudo de Wever et al. (2006) apresentou as características de 15 instrumentos diferentes de análise de conteúdo, que é uma técnica utilizada para analisar transcrições de discussões assíncronas. Os autores investigaram a utilização da análise de conteúdo, com o objetivo de apoiar a pesquisa na área da aprendizagem colaborativa suportada por computador. Os instrumentos aplicados refletem uma grande variedade de abordagens e diferem-se no nível de detalhamento e no tipo de categoria de análise utilizada. Outras diferenças estão relacionadas com a diversidade das bases teóricas, a quantidade de informação sobre validade e confiabilidade, e a escolha da unidade de análise.

Dennen (2008) analisa em seu artigo as práticas de avaliação e os métodos analíticos mais comuns para investigar fóruns de discussão. Os méritos de várias abordagens são discutidos com relação à forma como elas fazem ou não uma investigação da aprendizagem. O estudo destaca a importância de avaliar os debates, podendo ser empregadas medidas de participação, de qualidade e conteúdo, e holísticas. O trabalho apresenta um resumo de algumas formas de análise utilizadas para examinar conversações, descrevendo-as e indicando as suas vantagens e desvantagens. Os métodos discutidos foram: medidas de participação, análise de conteúdo, análise estrutural, microetnografia, análise de diálogo. De acordo com a autora, muitos aspectos da discussão *online*, no que se refere à aprendizagem, permanecem pouco estudados. Segundo a pesquisadora, é necessário estabelecer padrões e métodos validados. Este desafio metodológico é complicado pela necessidade de operacionalizar de forma mais clara os indicadores de aprendizagem.

Gerosa et al. (2003) apresentam uma análise utilizando o ambiente AulaNet, para acompanhamento das discussões através de aspectos relacionados à forma de estruturação do discurso e à categorização das mensagens. A estruturação permite observar o encadeamento hierárquico do debate para ter indícios do aprofundamento da discussão. A categorização ajuda na identificação das características de cada tipo de texto. O trabalho proposto pelos autores é interessante por fornecer dados quantitativos ao professor sobre as discussões no fórum. Estes dados envolvem: a identificação rápida da profundidade dos debates, a quantidade de mensagens por nível de interação, o número de caracteres dos textos de cada categoria. Os pesquisadores sinalizam a importância do conteúdo da mensagem ser relevante com o tema em debate. Cabe destacar que, no trabalho, este conteúdo deve ser analisado manualmente.

Sales et al. (2008) apresentam uma ferramenta que utiliza os indicadores de aprendizagem *Learning Vectors* para avaliar as mensagens postadas pelos alunos em um fórum de discussão do ambiente Moodle. Para cada texto postado pelo aluno, o professor ou tutor analisa a mensagem e a categoriza segundo uma escala de valores. Com o resultado desta categorização, a ferramenta realiza a avaliação de cada discente. O trabalho dos autores é interessante, pois os *Learning Vectors* representam um auxílio ao docente, reunindo aspectos qualitativos e quantitativos relativos ao acompanhamento do desempenho do aluno. No entanto, cabe ressaltar que o conteúdo de cada mensagem deve ser analisado manualmente pelo professor ou tutor, para que seja atribuída uma categoria.

Barbosa (2006) relembra em seus estudos que, no fórum de discussão, a aprendizagem ocorre de forma colaborativa e cooperativa. A autora realizou uma investigação em um fórum de discussão para verificar de que maneira o pensamento crítico do aluno é desenvolvido. As mensagens foram investigadas para detectar características referentes a presença social, presença cognitiva, e presença de ensino. O estudo da autora é interessante, pois pesquisa o desenvolvimento do pensamento crítico do aluno a partir da análise dos textos. No entanto, cabe destacar que o conteúdo das mensagens foi analisado manualmente pelo docente. Os indicadores quantitativos relatados no estudo também foram construídos de forma manual.

O trabalho de Bassani e Behar (2006) apresenta a ferramenta interROODA. No módulo de “trocas interindividuais”, o professor tem acesso a todas as mensagens de um aluno em determinado tópico de um fórum de discussão. Cada mensagem é classificada em enunciado (as mensagens que abrem uma nova discussão) ou citação (resposta a um enunciado ou a outra citação). O interROODA apresenta o número de interações vinculadas a

uma mensagem, como forma de indicar o valor desta. O interROODA é capaz de fornecer dados quantitativos ao professor para auxiliá-lo na avaliação de aprendizagem dos discentes. As autoras reforçam que um processo de avaliação da aprendizagem pautado também em aspectos qualitativos, deve considerar o conteúdo das mensagens. No trabalho das autoras, o conteúdo das mensagens deve ser analisado manualmente.

No trabalho desenvolvido por Chen Y. et al. (2008), são apresentadas algumas características que podem ser consideradas para medir tópicos de qualidade em fóruns de discussão na *Web*. Baseado nestas características foi proposto um modelo para reconhecê-las que envolve a extração de sinais de séries temporais. O diagnóstico de características destes sinais foi realizado com a transformada *wavelet*²³, e em seguida o reconhecimento dos tópicos de qualidade foi efetuado com uma rede neural *back-propagation*²⁴. Na pesquisa, a qualidade de um tópico é analisada considerando-se a quantidade de respostas a uma mensagem, o número de participantes do debate, e a profundidade das discussões. Cabe ressaltar que, o conteúdo das mensagens não foi considerado na análise. Apesar do artigo não abordar um enfoque pedagógico na utilização dos fóruns, as técnicas e metodologias apresentadas podem ser direcionadas com esta finalidade.

Penny e Murphy (2009) realizaram um estudo com critérios (*rubrics*) projetados para analisar a aprendizagem em fóruns de discussão. A pesquisa investigou os elementos de desempenho e de avaliações descritos pelos critérios. O estudo selecionou 50 critérios coletados a partir de fontes da Internet. Foram identificados 153 elementos de desempenho em 19 categorias e 831 avaliações em 40 categorias. Segundo a pesquisa, as avaliações para classificar aspectos mecânicos da escrita das mensagens podem ser úteis para aumentar a clareza da redação nos fóruns. No entanto, é necessária mais investigação para determinar se essa ênfase na mecânica aumenta (ou não) a capacidade do aluno para contribuir em profundidade nas análises e reflexões. Critérios de desempenho e análises que investigam os elementos processuais e de gestão da discussão podem ser melhorados com a inclusão de avaliações do comprimento das postagens, número de sentenças, número mínimo de mensagens ou palavras. As autoras relatam que a maioria dos critérios examinados avalia os estudantes em interação, participação, comportamento colaborativo e social, desenvolvimento

²³ A transformada *wavelet* é uma técnica de processamento que possibilita analisar sinais em diferentes momentos e frequências simultaneamente (MAKAREMI et al., 2011).

²⁴ Uma rede neural *back-propagation* utiliza o algoritmo de retropropagação de erro em seu treinamento supervisionado (HAYKIN, 2001).

cognitivo. De acordo com a pesquisa, é possível criar e utilizar critérios para acompanhar fóruns de discussão, e em seguida, analisar as postagens para tentar determinar se os alunos conseguiram benefícios.

3.6.2 PESQUISAS SOBRE ANÁLISE DE FÓRUNS DE DISCUSSÃO REALIZADAS COM TÉCNICAS DE MINERAÇÃO DE TEXTOS

Dringus e Ellis (2005) conduziram um estudo para avaliar a utilização de mineração de dados e de textos para analisar fóruns de discussão. No artigo, os autores discutem os problemas que contribuem para dificultar a avaliação destes ambientes. Eles sugerem indicadores de participação que o professor pode extrair das discussões e utilizar para avaliar o progresso e o desempenho dos alunos. Os autores recomendam que a mineração de dados seja utilizada para descobrir a distribuição e frequência das contribuições do participante ao longo do tempo de duração do fórum. Ela pode ser empregada também para identificar iniciativas ou respostas, e descobrir quanto tempo após a postagem inicial, a resposta foi emitida. A pesquisa indica que a mineração de textos seja utilizada para avaliar se uma contribuição está relacionada ao tópico do fórum, e se os alunos compartilham recursos, como referências da *Web* ou citações da literatura. De acordo com o estudo, as técnicas de mineração de textos e dados podem ser parte da solução para auxiliar o professor a analisar fóruns de discussão e obter *feedback* dos alunos. O MineraFórum está em consonância com as sugestões indicadas, pois utiliza mineração de textos para avaliar se uma mensagem está relacionada ao tema em debate, e auxilia o docente a obter informações sobre os discentes.

Rebedea et al. (2008) apresentam uma análise de chats que pode ser utilizada em fóruns de discussão. A pesquisa sugere a extração de dados sócio-semânticos a partir das conversações produzidas pelos participantes, utilizando técnicas de mineração de textos baseadas em ontologias. O método proposto utiliza uma combinação de perspectiva sócio-cultural, dialógica, e técnicas de processamento de texto. O estudo também apresenta o software que foi desenvolvido para descobrir os tópicos mais importantes do debate, a contribuição de cada participante na conversação, e descobrir novas referências implícitas entre os enunciados, para oferecer uma representação com múltiplas vozes da conversação. Estas vozes provêm da teoria de Bakhtin (1973), que afirma que em cada diálogo existem comunidades de vozes, onde a totalidade das ideias, pensamentos e palavras passam através de inúmeras vozes mescladas. A mineração de textos foi utilizada para: analisar se o conteúdo

das mensagens do chat tem relação com o tema em debate, e determinar o momento que um novo tópico é introduzido na discussão. A WordNet foi utilizada para detecção de sinônimos nos textos analisados. O trabalho desenvolvido assemelha-se ao MineraFórum, no sentido de desenvolver uma ferramenta que utiliza técnicas de mineração de textos para analisar a relevância de mensagens textuais em relação ao tema em debate. No entanto, existem diferenças entre o MineraFórum e a pesquisa de Rebedea et al. (2008), pois esta última analisa mensagens de chats para fornecer indicadores relacionados a teoria polifônica de Bakhtin, e a ferramenta desenvolvida não está integrada a um AVA.

O trabalho de Ravi e Kim (2007) apresenta uma abordagem para identificar automaticamente perfis de interações de alunos em fóruns de discussões. Foram utilizados recursos de seqüência de palavras e algoritmos SVM (*Support Vector Machine*), para desenvolver classificadores de “ato de discurso” que identificam os papéis das mensagens individuais, como: pergunta, resposta, elaboração, correção. Os classificadores foram utilizados na busca de mensagens que contêm perguntas ou respostas. Foi utilizado um conjunto de regras para análise dos tópicos para descobrir aqueles que poderiam ter perguntas sem resposta e necessitar da atenção do professor. Este projeto assemelha-se ao MineraFórum, no sentido de analisar, de forma automática, o conteúdo de mensagens de fóruns para auxiliar o docente. No entanto, existem diferenças, pois os classificadores desenvolvidos por Ravi e Kim (2007) analisam os textos através da identificação de “atos de discurso”. Cabe ressaltar que os indicadores fornecidos pelos classificadores são distintos dos apresentados pelo MineraFórum.

Kim et al. (2007) apresentou em seu artigo um agente inteligente, que foi implementado dentro de um fórum de discussão (figura 14), para fornecer automaticamente respostas às perguntas dos alunos. O trabalho mostra como os tópicos de discussão foram modelados utilizando “atos de discurso”. Cada postagem foi classificada de acordo com categorias de “atos de discurso”, como: pergunta, resposta, elaboração e correção. Ao classificar as contribuições da discussão de acordo com as categorias, os autores foram capazes de identificar os papéis dos alunos e do professor nas discussões. Os autores desenvolveram um conjunto de padrões para analisar as interações dos discentes nas discussões. Alguns destes padrões foram utilizados para descobrir mensagens onde os alunos poderiam ter perguntas não respondidas. O agente inteligente utiliza técnicas de mineração de texto para extrair palavras e suas freqüências da pergunta do aluno, dos documentos do curso, e das discussões anteriores. Os pesos das palavras são calculados pela fórmula TF-IDF. A

medida de similaridade cosseno é utilizada para descobrir quais destes documentos e/ou discussões passadas possuem semelhança com o texto envolvido na pergunta do aluno. O agente inteligente assemelha-se ao MineraFórum, no sentido de analisar automaticamente o conteúdo de mensagens de fóruns. No entanto, as técnicas de mineração de textos utilizadas em cada ferramenta são distintas. Destaca-se que o objetivo do agente é responder automaticamente as perguntas dos alunos, e o intuito do MineraFórum é apresentar indicadores para auxiliar o docente na análise de fóruns.



Figura 14 - Exemplo do agente inteligente dentro do fórum de discussão
 Fonte: Kim et al. (2007)

Outra pesquisa para analisar fóruns de discussão é apresentada por Lin et al. (2009). O trabalho utiliza mineração de textos e propõe um sistema de classificação dos gêneros das contribuições textuais, como: anúncio, pergunta, explicação, interpretação, conflito, afirmação, entre outros. Este sistema pode ser utilizado para facilitar o processo de codificação da análise do conteúdo de um fórum. Foram coletados dados de um fórum de discussão do Moodle para realizar os experimentos. Para os pesquisadores, o professor pode participar de um debate para contribuir com o processo de aprendizagem dos alunos. A maior questão de pesquisa no estudo foi validar a coerência dos resultados codificados por um sistema automático de classificação de gêneros e a análise realizada por juízes humanos. O artigo conclui que o modelo em cascata, embutido no sistema desenvolvido, pode facilitar o processo de codificação da análise de conteúdo de fóruns. O sistema de classificação de gêneros desenvolvido no estudo assemelha-se ao MineraFórum, em relação a execução de uma análise automatizada para investigar o conteúdo das mensagens. A ferramenta desenvolvida pelos autores também possui o objetivo de auxiliar o professor. No entanto, as

técnicas utilizadas para analisar o conteúdo das mensagens e os resultados fornecidos, são distintos dos apresentados pelo MineraFórum. Cabe ressaltar também que a ferramenta desenvolvida não está integrada a um AVA.

Li e Huang (2008a) apresentam uma pesquisa para fornecer um retrato mais completo das interações entre os sujeitos, na aprendizagem colaborativa suportada por computador. Os autores propõem um modelo de análise multidimensional para investigar as interações, baseado em técnicas de análise de conteúdo, mineração de texto, e análise de redes sociais. A análise de conteúdo é empregada para pesquisar como os alunos interagem, descobrindo os possíveis padrões de processo (a intenção do discurso) dentro da conversação. A mineração de texto é utilizada para descobrir os tópicos que apareceram nos debates. As mensagens de um conjunto de discussões são agrupadas em um documento, que é representado por um vetor de termos ponderados com a função TF-IDF. O método do cosseno é utilizado para calcular a similaridade entre o vetor do documento e o vetor do tema, definido pelo professor. O estudo também descreve o projeto e implementação de uma ferramenta para análise inteligente de conteúdo, denominada VINCA (*Visual Intelligent Content Analyzer*). Foi realizado um experimento com a ferramenta para analisar um conjunto de discussões, com a finalidade de desvendar a interação dos alunos em termos de padrão de processo, espaço de tópicos, e rede social. A pesquisa desenvolvida assemelha-se ao MineraFórum, no sentido de utilizar técnicas de mineração de textos para analisar a relevância das mensagens em relação ao tema em debate. No entanto, existem diferenças entre o MineraFórum e o trabalho de Li e Huang (2008a), pois este último representa os textos com o modelo de espaço vetorial e realiza a comparação dos mesmos com a medida de similaridade cosseno. Observa-se que a ferramenta desenvolvida não está integrada a um AVA.

Um fórum de discussão com características tecnológicas avançadas é apresentado por Li et al.(2008b). O projeto utiliza ontologia de domínio e técnicas de mineração de textos. No trabalho desenvolvido, as transcrições das discussões do fórum são automaticamente transformadas em uma modelagem estrutural em três etapas: o reconhecimento do tópico, a identificação do tipo de cada uma, e a associação semântica entre elas. A primeira etapa agrupa em um documento as mensagens de um mesmo conjunto de discussões. Cada documento é representado por um vetor de termos ponderados com a função TF-IDF. O método do cosseno é utilizado para calcular a similaridade entre o vetor do documento e o vetor de conceitos da ontologia do domínio. A segunda etapa identifica seis tipos de mensagens: pergunta, opinião, sugestão, recomendação, pedido, e citação. A terceira etapa

utilizada o modelo SLN (*Semantic Link Network*) para organizar os textos com associação semântica. O fórum proposto oferece três funcionalidades aos discentes: busca de informações úteis às suas necessidades, navegação temática pelas mensagens, e recomendação de outros alunos que possuem interesse em realizar comunicação e colaboração. Um estudo experimental foi realizado para apresentar que a abordagem é eficaz para descobrir parceiros de aprendizagem com os mesmos interesses e a busca de mensagens com guia de navegação temático. Este estudo assemelha-se ao MineraFórum, no sentido de utilizar técnicas de mineração de textos para analisar a relevância de mensagens textuais em relação ao tema em debate. No entanto, o MineraFórum apresenta diferenças em relação ao trabalho de Li et al.(2008b). Este último representa os textos com o modelo de espaço vetorial, e detecta se os mesmos são semelhantes com a medida de similaridade cosseno. O sistema desenvolvido pelos autores oferece auxílio aos alunos, enquanto o MineraFórum apresenta informações para auxiliar os docentes na análise de fóruns.

4. METODOLOGIA

Considerando o objetivo geral da presente pesquisa a metodologia adotada para o desenvolvimento do mesmo pautou-se nas seguintes etapas:

1. Escolha de uma técnica de mineração de textos para realizar a análise qualitativa das mensagens de um fórum.
2. Verificação da existência de algum software de mineração de textos que auxiliasse o professor a analisar qualitativamente as contribuições em um fórum de discussão.
3. Realização de estudos preliminares para avaliar a técnica de mineração de textos escolhida.
4. Definição dos indicadores de relevância das mensagens postadas em fóruns de discussão que foram implementados no MineraFórum.
5. Construção do MineraFórum.
6. Integração do MineraFórum ao AVAs: Rooda, ETC, Moodle.
7. Realização de experimentos com o MineraFórum.

Este capítulo descreve na seção 4.1 a técnica de mineração de textos escolhida para realizar a análise qualitativa das mensagens de um fórum de discussão. A seção 4.2 relata a seleção do software de mineração de textos. A seção 4.3 explica os estudos preliminares que foram efetuados para analisar se a técnica de mineração de textos utilizando grafos é adequada para analisar as mensagens de fóruns de discussão. A seção 4.4 define os indicadores que podem ser utilizados para identificar contribuições relevantes em um fórum. A seção 4.5 relata a construção do MineraFórum.

4.1 ESCOLHA DE UMA TÉCNICA DE MINERAÇÃO DE TEXTOS

Na mineração de textos, uma técnica bastante comum utilizada para representar um documento é identificar as características do mesmo através de um modelo de espaço vetorial. Nesta técnica, cada termo do documento torna-se uma característica dimensional. O valor de cada dimensão pode indicar o número de vezes que o termo aparece no texto, ou pode indicar o peso a ser considerado, como por exemplo, a quantidade de documentos nos quais o termo aparece. No entanto, essa técnica descarta informações importantes, como por exemplo, a ordem nos quais os termos aparecem, onde os mesmos aparecem, e a proximidade entre eles (SCHENKER, 2003).

Os grafos são construções matemáticas importantes e efetivas para realizar a modelagem de relacionamentos e de informação estrutural. Eles são utilizados em diversos tipos de problemas, incluindo ordenação, compressão, análise de tráfego, alocação de recursos, entre outros. Como os grafos retêm mais informações do que vetores de características atômicas simples, eles podem ser utilizados para a representação de textos (SCHENKER, 2003).

A técnica de mineração de textos utilizando grafos identifica as palavras com maior ocorrência no texto, e ainda se elas encontram-se próximas. O grafo obtido a partir da mineração apresenta em seus nós as palavras que mais ocorreram. As associações entre os nós indicam a proximidade.

Segundo Bakhtin (1997), na comunicação verbal, a palavra não deve ser tratada de forma isolada, nem como unidade da língua, nem somente com a significação dessa palavra. O mais importante é o enunciado acabado e com um sentido concreto, que revela o conteúdo do mesmo. De acordo com o mesmo autor, para compreender um discurso, além da identificação das palavras utilizadas, é de fundamental importância analisar o contexto.

A significação da palavra se refere à realidade efetiva nas condições reais da comunicação verbal. É por esta razão que não só compreendemos a significação da palavra enquanto palavra da língua, mas também adotamos para com ela uma atitude responsiva ativa (simpatia, concordância, discordância, estímulo à ação). Quando construímos nosso discurso, sempre conservamos na mente o todo do nosso enunciado, tanto em forma de um esquema correspondente a um gênero definido como em forma de uma intenção discursiva individual. Não alinhavamos as palavras, passando de uma para outra. Pelo contrário, é como se preenchêssemos um todo com as palavras úteis (BAKHTIN, 1997, p. 310).

Conforme cita Bakhtin (1997), para analisar um texto, deve-se observar dois itens: as palavras e o contexto. Em um contexto existem palavras que são pertinentes ao mesmo. As palavras relevantes de um contexto se relacionam umas com as outras, pois aparecem próximas no decorrer de um discurso que as mencione. Como a mineração de textos utilizando grafos apresenta as palavras, e as proximidades entre elas dentro de um texto, a técnica é capaz de analisar os dois itens citados por Bakhtin (1997).

Em um fórum, o tema do debate indica o contexto. Dentro do tópico em discussão pode-se identificar as palavras que são importantes. Um texto pode ser avaliado como pertinente ao tema, caso ele utilize palavras relativas ao mesmo. Destaca-se que a proximidade entre as palavras relevantes também demarcam o contexto.

A partir do estudo das técnicas de mineração de textos, observou-se que a mineração utilizando grafos é uma opção viável para a análise das mensagens de um fórum de discussão. Considerando-se o grafo gerado a partir de uma contribuição textual, pode-se identificar se ela refere-se ao contexto, e avaliar se a mesma é relevante ao tema em debate. Ao analisar uma mensagem, quanto mais palavras relativas ao tópico forem citadas, e quanto mais próximas elas estiverem umas das outras, maior será a relevância temática.

4.2 SELEÇÃO DO SOFTWARE PARA MINERAÇÃO DE TEXTOS

Os softwares apresentados na seção 3.5 foram analisados para identificar se algum deles atenderia as necessidades desta pesquisa, que são: analisar as postagens de fóruns de discussão, calcular a relevância das contribuições textuais dos alunos, apresentar um relatório com as informações desejadas, integração com ambientes virtuais de aprendizagem. No entanto, nenhum dos softwares foi considerado adequado para a realização das tarefas.

No entanto, identificou-se que a programação do Sobek, para executar a mineração de textos, poderia ser utilizada no MineraFórum. Desta forma, o Sobek foi selecionado pelos seguintes motivos:

- É um software livre.
- Executa a mineração de textos utilizando grafos.

- Foi desenvolvido por um grupo de pesquisa envolvendo professores do PPGEDU e PPGIE (UFRGS), o que possibilita a participação em seu desenvolvimento e a solicitação de mudanças e ajustes considerados importantes.
- Pode ser executado em computadores com diferentes sistemas operacionais.
- O software é capaz de efetuar a leitura de textos nos formatos “txt”, “doc”, ou “pdf”.

O MineraFórum utilizou a programação de algumas funcionalidades implementadas no Sobek: processo de mineração de textos utilizando grafos; definição do valor mínimo da frequência dos termos a serem identificados em um texto; utilização de uma lista de *stopwords*²⁵ no processo de mineração; leitura de arquivos nos formatos “txt”, “doc”, ou “pdf”; possibilidade de criar e/ou alterar uma base de conceitos; gravação de uma base de conceitos em arquivo; leitura de uma base de conceitos gravada em arquivo; exibição do grafo gerado com a mineração de um texto; mineração de textos em língua portuguesa; exibição dos recursos disponíveis no sistema em língua portuguesa ou inglesa.

4.3 ESTUDOS PRELIMINARES

Foram realizados estudos preliminares, com o objetivo de analisar se a técnica de mineração de textos utilizando grafos é adequada para analisar as mensagens de fóruns de discussão. O intuito de minerar as mensagens é identificar se os alunos estão debatendo os conceitos relacionados ao tema em questão.

Para efetuar a mineração dos textos provenientes do fórum de discussão, foi escolhido o software Sobek, descrito na seção 3.5. A partir dos grafos gerados com os textos, foi realizado manualmente um estudo dos mesmos, para analisar quais contribuições foram relevantes de acordo com o tema proposto.

Para direcionar as atividades deste trabalho definiu-se o coeficiente de relevância temática (CRT) (Azevedo et al., 2009) de uma contribuição textual, cujo objetivo é analisar o

²⁵ Palavras que podem ser descartadas no processo de mineração, como: advérbios, artigos, preposições, entre outras.

quanto um texto é relevante a um determinado tópico de discussão. O CRT indica o grau de relevância de uma mensagem de acordo com o tema do fórum.

O coeficiente de relevância temática é calculado a partir da seguinte fórmula:

$$CRT = NC + NA$$

NC: número de conceitos relevantes utilizados no texto

NA: número de associações entre os conceitos relevantes utilizados no texto

Para o cálculo do NC, os termos considerados como equivalentes semânticos também são considerados como conceitos relevantes. Os prefixos, sufixos, e plurais dos conceitos relevantes citados no texto são convertidos, para também serem considerados. Com o CRT, os grafos que possuem mais conceitos importantes relativos ao tema, com mais associações entre eles, possuem maior coeficiente de relevância temática.

A metodologia empregada nos estudos preliminares envolveu:

- a) A escolha de um fórum de discussão.
- b) A definição dos conceitos importantes relativos ao tema do fórum, bem como a definição da associação entre estes conceitos.
- c) A definição de possíveis termos que podem ser considerados como equivalentes semânticos aos conceitos envolvidos.
- d) A definição do valor mínimo do coeficiente de relevância temática que será considerado para análise.
- e) A coleta das contribuições textuais produzidas pelos alunos no fórum.
- f) Para cada contribuição textual, foi gerado o seu grafo conceitual com o Sobek.
- g) A partir de cada grafo gerado, foi calculado o coeficiente de relevância temática da contribuição textual.
- h) A partir da organização realizada com os grafos gerados, as contribuições textuais relativas aos mesmos foram organizadas.
- i) A partir dos grafos gerados foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno no fórum.

Para avaliar a metodologia proposta e analisar os resultados, foram realizados três experimentos, com fóruns de discussão distintos.

O primeiro experimento foi realizado em um fórum de discussão do Ambiente Rooda, na disciplina “Tópicos Especiais Z1” da turma do Doutorado em Informática na Educação, que ocorreu no primeiro semestre de 2008. O tema foi “Comunidades Virtuais”. O fórum contou com a participação de 05 alunos, que realizaram 25 contribuições textuais.

Para este fórum, o professor definiu como importantes, os seguintes conceitos: “comunidades”, “virtuais”, “comunidades virtuais”, “aprendizagem”, “comunidades virtuais de aprendizagem”, “educação”, “relações sociais”. As associações relevantes entre os conceitos ficaram estabelecidas da seguinte maneira: o conceito “comunidades virtuais de aprendizagem” está relacionado com os conceitos “comunidades”, “comunidades virtuais”, “aprendizagem”, “educação”. Os termos “CVs” e “CVAs” foram definidos como equivalentes semânticos aos termos “comunidades virtuais” e “comunidades virtuais de aprendizagem”, respectivamente.

Foi definido o valor mínimo um (1) para o coeficiente de relevância temática (CRT) a ser considerado na análise²⁶. O coeficiente de relevância temática foi calculado a partir do grafo gerado em cada contribuição textual. Como um exemplo de cálculo do CRT, pode-se observar a mensagem abaixo extraída do fórum, postada pelo aluno C.

Mensagem: “Acredito que as CVAs podem oferecer um recurso valioso à educação, pois elas possibilitam um mecanismo de construção coletiva de conhecimento, através das discussões realizadas pelos membros da comunidade.”

O grafo gerado a partir da mensagem anterior é apresentado na figura 15.

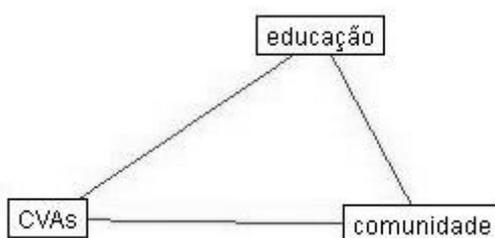


Figura 15 - Grafo gerado a partir de uma mensagem do fórum do experimento 1
Fonte: o autor

Para o grafo anterior, o CRT é calculado da seguinte maneira:

$$NC = 3, NA = 3$$

²⁶ O valor mínimo indica que toda mensagem com CRT acima deste resultado será considerada como relevante. O valor um (1) é o menor número que pode ser calculado pela fórmula do CRT, situação obtida quando $NC = 1$ e $NA = 0$.

$$\text{CRT} = \text{NC} + \text{NA} = 6$$

Foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno no fórum. Os resultados são apresentados na tabela 1.

Tabela 1 - Quantidade de contribuições relevantes do primeiro experimento

Aluno	Quantidade de contribuições relevantes
01	32%
02	24%
03	16%
04	12%
05	4%
Contribuições não relevantes	12%

Fonte: o autor

Através da tabela 1, percebeu-se que a utilização da metodologia proposta permitiu verificar que a maior parte das contribuições feitas pelos estudantes era relevante.

O segundo e terceiro experimentos foram realizados em um fórum de discussão do Ambiente Rooda, na disciplina “Seminário Integrador VII – B” de uma turma do Curso de Graduação em Pedagogia, que ocorreu no segundo semestre de 2009. O tema foi “Aprender com os outros”. A turma foi dividida em dois grupos, e cada grupo contribuiu em um fórum de discussão. O primeiro contou com a participação de 27 alunos, que realizaram 45 contribuições textuais. O segundo contou com a participação de 29 alunos, que postaram 67 mensagens.

Para estes fóruns, o professor definiu como importantes, os seguintes conceitos: “aprender”, “aprendizagem”, “conhecimento”, “construção do conhecimento”, “Piaget”, “afeto”, “afetividade”, “estruturas cognitivas”, “descentração”, “relações afetivas”, “experiências afetivas”, “laços afetivos”, “cooperação”, “colaboração”. Foram estabelecidas as associações relevantes entre estes conceitos.

Foi definido o valor mínimo 01 para o coeficiente de relevância temática (CRT) a ser considerado na análise. O coeficiente de relevância temática foi calculado a partir do grafo gerado em cada contribuição textual. Foi realizada uma análise da quantidade de contribuições relevantes colocadas por cada aluno do primeiro e segundo grupos. Os resultados são apresentados nas tabelas 2 e 3, respectivamente.

Tabela 2 - Quantidade de contribuições relevantes do primeiro grupo

Aluno	Quantidade de contribuições relevantes
01	4,4%
02	4,4%
03	8,8%
04	6,6%
05	8,8%
06	2,2%
07	8,8%
08	2,2%
09	2,2%
10	2,2%
11	2,2%
12	2,2%
13	2,2%
14	2,2%
15	4,4%
16	4,4%
17	2,2%
18	2,2%
19	2,2%
20	4,4%
21	4,4%
22	4,4%
23	2,2%
24	2,2%
25	2,2%
26	2,2%
27	2,2%
Contribuições não relevantes	0%

Fonte: o autor

Tabela 3 - Quantidade de contribuições relevantes do segundo grupo

Aluno	Quantidade de contribuições relevantes
01	5,97%
02	5,97%
03	1,49%
04	2,98%
05	5,97%
06	1,49%
07	8,95%
08	2,98%
09	2,98%
10	2,98%
11	2,98%
12	2,98%
13	2,98%
14	1,49%
15	1,49%
16	2,98%
17	1,49%

18	1,49%
19	1,49%
20	1,49%
21	1,49%
22	2,98%
23	1,49%
24	1,49%
25	1,49%
26	1,49%
27	2,98%
28	1,49%
29	1,49%
Contribuições não relevantes	20,89%

Fonte: o autor

Observando as tabelas 1, 2, 3, pode-se verificar a quantidade de contribuições relativas ao tema do fórum produzidas por cada aluno, e a porcentagem de mensagens não relevantes.

Destaca-se que a versão do Sobek, utilizada nos estudos preliminares, realiza a mineração de um texto e a geração do grafo representativo dos conceitos encontrados no mesmo. No entanto, como o Sobek não foi desenvolvido com o intuito de minerar as mensagens de fóruns, o cálculo do CRT, a organização das postagens e a análise da quantidade de contribuições relevantes colocadas por cada aluno, foram atividades executadas manualmente.

A partir dos resultados apresentados nas tabelas 1, 2, 3, percebeu-se que a técnica de mineração de textos utilizando grafos é adequada para realizar a análise qualitativa das mensagens de um fórum de discussão. Isto é possível, pois a técnica verifica se conceitos importantes de serem tratados estão efetivamente sendo considerados nos textos dos estudantes.

4.4 DEFINIÇÃO DOS INDICADORES DE RELEVÂNCIA A SEREM DIAGNOSTICADOS PELO MINERAFÓRUM

Com o intuito de atender ao primeiro objetivo específico desta pesquisa, um questionário foi aplicado com dezoito (18) professores, que atuam no nível superior, na graduação ou na pós-graduação. Estes docentes já utilizaram fóruns de discussão com seus alunos, em disciplinas que lecionaram. Entre os dezoito professores, sete (07) ministram aula

no IFF²⁷ e quatro (04) na UFRGS. Dois (02) docentes trabalharam com EAD na FSMA²⁸. Cinco (05) entrevistados são doutorandos em Informática na Educação, que já empregaram fórum de discussão com seus discentes. O questionário completo encontra-se no Apêndice B. O formulário apresentou dez (10) questões, onde cada uma corresponde a uma ação que o professor pode executar para realizar um diagnóstico da atuação de cada aluno em uma discussão. Cada uma destas ações corresponde a um indicador, que pode ser utilizado para mapear a relevância das mensagens postadas em fóruns.

As questões apresentadas no questionário foram:

1. Analisar indicadores quantitativos, como a quantidade de mensagens postadas e a quantidade de palavras digitadas em cada mensagem.
2. Investigar se as mensagens postadas referem-se ao tema do debate.
3. Verificar se as mensagens postadas não possuem erros conceituais.
4. Verificar se as mensagens postadas não possuem erros gramaticais.
5. Investigar se existem mensagens cujos textos são muito semelhantes.
6. Analisar a categoria de cada mensagem. Se ela é uma pergunta, resposta, afirmativa, explicação, interpretação, conflito, concordância, elaboração, correção, agradecimento, motivação.
7. Verificar a estruturação das mensagens postadas. Isto permite descobrir se uma mensagem originou outras, e a sequência das postagens, para saber qual mensagem foi colocada em função de outra.
8. Descobrir as pessoas com quem o aluno interagiu dentro do debate.
9. Considerar aspectos culturais dos alunos envolvidos no debate.
10. Investigar fatores afetivos nas mensagens postadas.

As ações descritas nas perguntas do questionário foram baseadas em pesquisas sobre análise de fóruns de discussão, descritas na seção 3.6.1. O quadro 4 apresenta as referências bibliográficas dos trabalhos que citaram cada um dos indicadores mencionados em cada questão.

²⁷ Instituto Federal de Educação, Ciência e Tecnologia Fluminense

²⁸ Faculdade Salesiana de Macaé

Quadro 4 - Referências bibliográficas que citaram os indicadores descritos no questionário

Pergunta do questionário	Indicadores	Referências bibliográficas
1	Analisar indicadores quantitativos, como a quantidade de mensagens postadas e a quantidade de palavras digitadas em cada mensagem.	Nisbet (2004), Penny e Murphy (2009).
2	Investigar se as mensagens postadas referem-se ao tema do debate.	Pelz (2004), Ho e Swan (2007).
3	Verificar se as mensagens postadas não possuem erros conceituais.	Pelz (2004), Ho e Swan (2007).
4	Verificar se as mensagens postadas não possuem erros gramaticais.	Pelz (2004), Ho e Swan (2007).
5	Investigar se existem mensagens cujos textos são muito semelhantes.	Pelz (2004), Ho e Swan (2007).
6	Analisar a categoria de cada mensagem. Se ela é uma pergunta, resposta, afirmativa, explicação, interpretação, conflito, concordância, elaboração, correção, agradecimento, motivação.	Ravi e Kim (2007), Lin et al. (2009).
7	Verificar a estruturação das mensagens postadas. Isto permite descobrir se uma mensagem originou outras, e a sequência das postagens, para saber qual mensagem foi colocada em função de outra.	Gerosa et al. (2003), Chen e Chiu (2008).
8	Descobrir as pessoas com quem o aluno interagiu dentro do debate.	Bassani e Behar (2006), Saltz et al. (2004).
9	Considerar aspectos culturais dos alunos envolvidos no debate.	Ware e Kramsch (2005).
10	Investigar fatores afetivos nas mensagens postadas.	Rourke et al. (2001).

Fonte: o autor

No final do formulário, o docente também poderia sugerir ações a serem realizadas. Cabe ressaltar que nenhum professor sugeriu a realização de outras ações além das mencionadas no questionário.

Para cada ação foi solicitada a indicação de um peso entre zero (0) e dez (10), onde zero (0) significa que a mesma não tem relação com o tema do debate, e dez (10) indica que ela possui relevância máxima.

As respostas informadas pelos docentes no formulário foram analisadas. A tabela 4 apresenta os resultados do questionário aplicado.

Tabela 4 - Respostas informadas pelos professores

Prof.	Quest. 1	Quest. 2	Quest. 3	Quest. 4	Quest. 5	Quest. 6	Quest. 7	Quest. 8	Quest. 9	Quest. 10
1	2	10	10	4	8	5	10	7	3	1
2	5	9	8	3	7	6	10	6	7	5
3	2	9	10	9	9	10	10	10	8	8
4	4	7	8	3	3	8	9	8	8	4
5	10	10	1	1	1	1	10	5	1	1
6	5	10	9	3	8	7	8	3	10	6
7	6	10	10	6	8	8	8	7	7	7
8	6	10	10	5	10	7	10	10	3	10
9	6	10	10	6	10	10	10	10	8	10
10	6	10	10	6	5	8	9	9	8	7
11	2	10	9	6	9	10	7	9	8	7
12	1	9	10	7	8	10	10	9	8	8
13	10	9	9	6	10	10	8	8	7	10
14	7	10	10	7	7	7	10	5	7	5
15	7	10	10	8	10	10	10	10	10	8
16	7	10	10	6	5	6	10	5	9	5
17	8	10	10	3	9	3	5	5	3	5
18	8	10	10	8	10	10	10	9	7	7
Média	5,67	9,61	9,11	5,39	7,61	7,56	9,11	7,50	6,78	6,33

Fonte: o autor

Para a construção do MineraFórum²⁹, foram escolhidos os indicadores que possuíam o maior peso, que poderiam ser implementados com a técnica de mineração de textos utilizando grafos, ou com informações fornecidas pelo Ambiente Virtual de Aprendizagem. Cada um destes indicadores será mapeado pelo MineraFórum. As ações citadas no questionário, selecionadas para implementação, foram:

- 1) Investigar se as mensagens postadas referem-se ao tema do debate (questão 2 do formulário).
- 2) Verificar a estruturação das mensagens postadas. Isto permite descobrir se uma mensagem originou outras e a sequência das postagens, para saber qual mensagem foi colocada em função de outra (questão 7 do formulário).
- 3) Investigar se existem mensagens cujos textos são muito semelhantes (questão 5 do formulário).

²⁹ É importante destacar que os indicadores diagnosticados no questionário foram implementados na versão 3.0 do MineraFórum (apresentada na próxima seção).

O MineraFórum executará as ações acima para diagnosticar a relevância das mensagens redigidas em um fórum. A primeira e a terceira ação serão executadas com a técnica de mineração de textos utilizando grafos. A segunda ação será efetuada com o auxílio de informações obtidas do AVA.

Cabe ressaltar que a ação “Verificar se as mensagens postadas não possuem erros conceituais” (questão 3 do formulário) também foi mensurada com alta relevância pelos docentes. No entanto, neste trabalho não será proposto o mapeamento deste indicador pelo MineraFórum. A implementação da verificação de erros conceituais nas mensagens requer pesquisas aprofundadas, que não foram realizadas neste trabalho.

A próxima seção descreve a construção das três versões do MineraFórum. A primeira e a segunda versão calculam a relevância temática das mensagens do fórum. Já a versão 3.0 executa o cálculo da relevância das postagens a partir de três indicadores: análise temática das contribuições textuais, proporção de citações da mensagem, verificação da similaridade da mensagem com outra do fórum.

4.5 CONSTRUÇÃO DO MINERAFÓRUM

Após a realização dos estudos descritos na seção 4.3, concluiu-se que a técnica de mineração de textos utilizando grafos é capaz de analisar qualitativamente as mensagens de fóruns de discussão.

Com o intuito de resolver a questão de pesquisa, foi planejada a construção do MineraFórum. Conforme citado na seção 4.2, não foi encontrado um software capaz de realizar a análise qualitativa das postagens em fóruns que apresentasse resultados importantes para o docente. Desta forma, optou-se pelo desenvolvimento do MineraFórum. O software foi planejado para atender as seguintes necessidades:

- Analisar qualitativamente as contribuições textuais redigidas pelos alunos em fóruns de discussão.
- Calcular a relevância de cada mensagem postada.
- Apresentar um relatório da análise efetuada, com informações úteis para auxiliar o professor.

- Integração com ambientes virtuais de aprendizagem.

Alguns pontos importantes também foram considerados no planejamento do MineraFórum:

- O software foi concebido para ser utilizado em fóruns de discussão baseados preferencialmente em referências teóricas. Desta forma, é necessário que o professor informe um texto de referência sobre o tema debatido no fórum. O docente pode digitar os conceitos relevantes (e suas associações) referentes ao tópico do fórum, ao invés de indicar um texto de referência.
- Os usuários do sistema serão somente professores ou tutores.
- O MineraFórum pode ser utilizado com fóruns e textos de referência escritos em língua portuguesa.

A construção do MineraFórum foi conduzida com três versões. De forma geral, o funcionamento das versões foi elaborado da seguinte forma:

- a) O usuário escolhe um fórum de discussão dentro de um AVA.
- b) O usuário ativa a execução do MineraFórum, a partir de um link ou botão, dentro do AVA.
- c) O MineraFórum captura as postagens do fórum selecionado, na ordem que foram redigidas.
- d) O usuário informa um texto de referência sobre o tema debatido no fórum. O texto pode ser digitado, ou lido de um arquivo nos formatos “txt”, “doc”, ou “pdf”. Ao invés de indicar um texto, é possível digitar os conceitos relevantes (e suas associações) referentes ao tópico da discussão.
- e) Caso deseje, o usuário pode informar uma lista de *stopwords*, um dicionário de sinônimos, e uma lista de palavras com equivalência semântica. Caso o usuário não indique estas informações, o software utiliza as *stopwords* e o dicionário de sinônimos definidos internamente no sistema.
- f) Caso seja informado um texto de referência, o MineraFórum executa a mineração do mesmo. O processo ocorre nas seguintes etapas:
 - O usuário pode escolher a quantidade mínima de ocorrência das palavras, para que elas sejam consideradas como relevantes na mineração. Por

exemplo, se o valor escolhido for 15, somente as palavras com frequência maior ou igual a 15, serão consideradas como relevantes. Caso esta frequência não seja informada, o MineraFórum a calcula automaticamente, baseado na quantidade de palavras do texto.

- A quantidade de ocorrências de cada palavra é calculada, desde que não seja uma *stopword*.
 - As palavras com maior frequência são consideradas como conceitos relevantes ao tema da discussão. Cabe ressaltar que os conceitos relevantes detectados pelo MineraFórum podem ser simples ou compostos. Um conceito simples possui somente uma palavra, enquanto um conceito composto possui várias palavras. Um exemplo de conceito simples é “computação”. Exemplos de conceitos compostos são “computação gráfica”, “computação em nuvem”.
 - As *stopwords* são removidas do texto.
 - O MineraFórum gera um grafo com os conceitos relevantes do texto. Cada vértice do grafo é um conceito. As arestas interligam os vértices que representam conceitos localizados próximos uns aos outros dentro do texto.
- g) Caso os conceitos relevantes (e suas associações) do tema em debate sejam digitados, o software gera um grafo com os conceitos.
- h) O usuário solicita a mineração das postagens do fórum selecionado.
- i) O MineraFórum executa a mineração do texto de cada mensagem.
- j) O MineraFórum calcula a relevância de cada postagem. Para desempenhar este cálculo, é necessário que o software analise a correspondência entre o grafo da mensagem e o grafo do texto de referência (ou digitação dos conceitos). Para realizar esta tarefa, a ferramenta efetua uma análise de similaridade entre os dois grafos, descobrindo os vértices equivalentes³⁰ em ambos. A fórmula utilizada para o cálculo da relevância foi modificada em cada uma das versões do sistema. A próxima seção explica as fórmulas empregadas.

³⁰ Dois vértices são equivalentes se eles possuem conteúdo equivalente, isto é, se possuem as mesmas palavras, ou se as palavras podem ser reduzidas ao mesmo radical, ou se possuem sinônimos, ou se possuem palavras com equivalência semântica.

- k) O software exibe um relatório com informações sobre a análise qualitativa de todas as postagens do fórum.

4.5.1 CÁLCULO DA RELEVÂNCIA DAS MENSAGENS

As fórmulas utilizadas pelo MineraFórum para calcular a relevância das mensagens foi aprimorada ao longo do desenvolvimento das versões do sistema.

A primeira e a segunda versão do sistema empregaram uma fórmula para calcular a relevância temática de cada postagem. No contexto desta pesquisa, uma mensagem é relevante ao tema do debate se ela cita conceitos importantes detectados no texto de referência, ou digitados pelo professor.

Na primeira e segunda versão, o usuário pode definir um valor mínimo para que uma mensagem seja considerada relevante ao tópico da discussão. Desta forma, uma contribuição textual é considerada relevante ao tema do debate, se o seu valor de relevância é maior ou igual ao mínimo estabelecido pelo usuário.

A fórmula utilizada na primeira versão foi:

$$CRT = NC + NA \quad (1)$$

CRT: Coeficiente de relevância temática da mensagem

NC: Número de vértices equivalentes no grafo da mensagem e no grafo do texto de referência

NA: Número de arestas entre os vértices equivalentes no grafo da mensagem e no grafo do texto de referência

Na segunda versão do sistema, a fórmula para o cálculo da relevância temática foi aprimorada em dois aspectos:

- a) A fórmula da primeira versão considera o número de arestas entre os vértices equivalentes nos dois grafos. Ao invés de identificar a quantidade de arestas, analisar a distância entre os vértices produz informações mais detalhadas. Com esta análise é possível saber qual a proximidade entre dois vértices equivalentes nos grafos. Outro fator importante a ser observado é a quantidade de nós vizinhos

de um vértice equivalente. Desta forma, o termo NA, utilizado na fórmula da primeira versão, foi substituído por DC e PC.

- b) Para identificar se duas palavras são similares, o MineraFórum consulta um dicionário de sinônimos e uma lista de equivalentes semânticos. Mesmo assim, alguns conceitos importantes relacionados ao tema do debate podem não estar contemplados. A união de todas as mensagens redigidas no fórum constitui um texto no qual os alunos podem mencionar conceitos significativos ao tópico da discussão. Este texto é minerado pelo software para identificar conceitos relevantes, que não foram citados no texto de referência, nem no dicionário de sinônimos, nem na lista de equivalentes semânticos. Desta forma, na segunda versão, o software gera um grafo a partir do texto construído com a união de todas as postagens, denominado grafo do fórum. O cálculo da relevância temática de cada mensagem também é realizado de acordo com a análise da correspondência entre os grafos da mensagem e do fórum.

A fórmula utilizada na segunda versão foi:

$$RT = \frac{RT_1 + RT_2}{2} \quad (2)$$

RT: Relevância temática da mensagem

$$RT_1 = NC + DC + PC \quad (3)$$

RT₁: Relevância temática da mensagem em relação ao texto de referência

NC: Número de conceitos relevantes da mensagem

$$DC = \frac{1}{1 + \sum_{i=1}^n \sum_{j=1}^n |d_{min}^{\alpha}(C_i, C_j) - d_{min}^{\beta}(C_i, C_j)|} \quad (4)$$

DC: Distância relativa entre os conceitos relevantes da mensagem e o texto de referência

n= Total de nós equivalentes

α= Grafo do texto de referência

β= Grafo da mensagem

d_{min}^α(C_i, C_j)= menor distância entre os nós equivalentes C_i e C_j no grafo do texto de

referência

$d_{min}^{\beta}(C_i, C_j)$ = menor distância entre os nós equivalentes C_i e C_j no grafo da mensagem

$$PC = \frac{1}{1 + \sum_{i=1}^n \left| \frac{v_i^{\alpha}}{N_{\alpha}} - \frac{v_i^{\beta}}{N_{\beta}} \right|} \quad (5)$$

PC : Peso relativo dos conceitos relevantes da mensagem e o texto de referência

n = Total de nós equivalentes

α = Grafo do texto de referência

β = Grafo da mensagem

v_i^{α} = Total de vizinhos do nó i em α

v_i^{β} = Total de vizinhos do nó i em β

N_{α} = Total de nós em α

N_{β} = Total de nós em β

$$RT_2 = NC + DC + PC \quad (6)$$

RT_2 : Relevância temática da mensagem em relação ao texto do fórum

NC : Número de conceitos relevantes da mensagem

$$DC = \frac{1}{1 + \sum_{i=1}^n \sum_{j=1}^n \left| d_{min}^{\alpha}(C_i, C_j) - d_{min}^{\beta}(C_i, C_j) \right|} \quad (7)$$

DC : Distância relativa entre os conceitos relevantes da mensagem e o texto do fórum

n = Total de nós equivalentes

α = Grafo do texto do fórum

β = Grafo da mensagem

$d_{min}^{\alpha}(C_i, C_j)$ = menor distância entre os nós equivalentes C_i e C_j no grafo do texto do fórum

$d_{min}^{\beta}(C_i, C_j)$ = menor distância entre os nós equivalentes C_i e C_j no grafo da mensagem

$$PC = \frac{1}{1 + \sum_{i=1}^n \left| \frac{v_i^\alpha}{N_\alpha} - \frac{v_i^\beta}{N_\beta} \right|} \quad (8)$$

PC: Peso relativo dos conceitos relevantes da mensagem e o texto do fórum

n= Total de nós equivalentes

α = Grafo do texto do fórum

β = Grafo da mensagem

v_i^α = Total de vizinhos do nó i em α

v_i^β = Total de vizinhos do nó i em β

N_α = Total de nós em α

N_β = Total de nós em β

Na terceira versão do sistema, as alterações ocorridas foram:

- Foi observado que na fórmula para o cálculo de RT da segunda versão, o valor do termo NC apresentava um impacto muito maior que DC e PC. Quanto mais conceitos relevantes uma mensagem possuísse, maior era o valor de NC. No entanto, os valores de DC e PC variavam somente entre zero (0) e um (1).
- Observou-se que a fórmula para cálculo de NC poderia ser aprimorada. Além de contabilizar a quantidade de conceitos relevantes da mensagem, também é interessante considerar a frequência de cada conceito. É importante calcular a proporção dos conceitos relevantes com relação ao total de palavras da mensagem. Desta forma, a fórmula de NC foi alterada e o termo foi denominado RC.
- Optou-se por modificar a fórmula de RT, para que o valor máximo de relevância a ser calculado fosse um (1). A nova fórmula utiliza RC, DC, PC. O termo RC é o mais importante, porém seu impacto é menor no resultado final, em relação ao NC da fórmula da segunda versão.
- As fórmulas para calcular DC e PC foram mantidas como na segunda versão.
- Uma inovação importante da terceira versão ocorreu no cálculo da relevância de uma postagem. Além de considerar a relevância temática, também foi calculada a relevância de citações da mensagem e a similaridade da mensagem com outra do

fórum. A fórmula da relevância de uma postagem calcula a média ponderada destes três fatores. Estes indicadores e seus respectivos pesos foram analisados na seção 4.4.

- Para calcular a relevância de citações de uma mensagem, foi considerada a quantidade de citações da postagem e o total de contribuições textuais do fórum. Estes dados foram obtidos a partir do AVA.
- O cálculo da similaridade de uma mensagem com outra do fórum foi realizado com a técnica de mineração de textos utilizando grafos. O procedimento para calcular a relevância temática foi empregado. Os grafos das mensagens são comparados entre si, para verificar se as contribuições textuais são semelhantes.
- Com o intuito de facilitar a visualização do resultado da relevância das postagens, o valor calculado foi mapeado para uma escala de valores inteiros entre zero (0) e cinco (5). O valor zero (0) indica que a mensagem não é relevante, enquanto o valor cinco (5) significa que a postagem possui alta relevância para a discussão.

As fórmulas utilizadas na terceira versão foram:

$$RF = \frac{(P1.RT)+(P2.RM)-(P3.SM)}{P1+P2-P3} \quad (9)$$

RF = Relevância da mensagem

Pesos: P1 = 9.61, P2 = 9.11, P3 = 7.47

Os pesos P1, P2, P3 foram obtidos a partir das respostas dos formulários preenchidos pelos professores (tabela 4, seção 4.4).

$$RT = \frac{RT_1+RT_2}{2} \quad (10)$$

RT: relevância temática da mensagem

$$RT_1 = (RC + (DC * (1 - RC))) + (PC * (1 - (RC + (DC * (1 - RC))))) \quad (11)$$

RT₁: Relevância temática da mensagem em relação ao texto de referência

$$RC = \frac{\sum_{i=1}^n f_i \cdot C_i}{\sum_{i=1}^N f_i \cdot C_i} \quad (12)$$

RC: Relevância dos conceitos da mensagem

C_i = Nó equivalente

f_i = Frequência do conceito C_i

n = Total de nós equivalentes

N = Total de nós do grafo da mensagem

DC: Distância relativa entre os conceitos relevantes da mensagem e o texto de referência

PC: Peso relativo dos conceitos relevantes da mensagem e o texto de referência

$$RT_2 = (RC + (DC * (1 - RC))) + (PC * (1 - (RC + (DC * (1 - RC))))) \quad (13)$$

RT₂: Relevância temática da mensagem em relação ao texto do fórum

$$RC = \frac{\sum_{i=1}^n f_i \cdot C_i}{\sum_{i=1}^N f_i \cdot C_i} \quad (14)$$

RC: Relevância dos conceitos da mensagem

C_i = Nó equivalente

f_i = Frequência do conceito C_i

n = Total de nós equivalentes

N = Total de nós do grafo da mensagem

DC: Distância relativa entre os conceitos relevantes da mensagem e o texto do fórum

PC: Peso relativo dos conceitos relevantes da mensagem e o texto do fórum

$$RM = \frac{m}{M} \quad (15)$$

RM: Relevância de citações da mensagem

m = quantidade de citações da mensagem m

M = total de mensagens postadas no fórum

$$SM = \begin{cases} 0, & \text{se a mensagem não possuir similaridade com outra do fórum} \\ RT, & \text{se a mensagem possuir similaridade com outra do fórum} \end{cases}$$

SM: Similaridade da mensagem com outra do fórum

O próximo capítulo descreve o desenvolvimento do MineraFórum, a análise efetuada em cada versão do sistema e as respectivas interfaces. A integração do MineraFórum com os AVAs Rooda, ETC, Moodle, é explicada. Os experimentos realizados com a ferramenta são apresentados e discutidos.

5. MINERAFÓRUM

Este capítulo apresenta o desenvolvimento das versões do MineraFórum, envolvendo o projeto e implementação das mesmas. A seção 5.1 demonstra o processo de desenvolvimento de software utilizado. A seção 5.2 apresenta a especificação dos requisitos da primeira versão do MineraFórum, o diagrama de casos de uso, o modelo de classes, a interface, o algoritmo utilizado para calcular a relevância temática de uma mensagem, os experimentos realizados e a análise e discussão dos resultados. As seções 5.3 e 5.4 relatam as alterações ocorridas na segunda e terceira versão do sistema, as experiências efetuadas com as versões e a discussão dos resultados obtidos. A seção 5.5 explica a integração do MineraFórum com os AVAs Rooda, ETC, Moodle.

5.1 PROCESSO DE DESENVOLVIMENTO

Esta seção apresenta o processo de software escolhido para realizar o desenvolvimento do MineraFórum.

A Engenharia de Software é a área da Ciência da Computação cujo objetivo é o desenvolvimento de softwares de qualidade, dentro de custos adequados. Um processo de software é um conjunto de atividades que conduz à produção de um produto de software (SOMMERVILLE, 2007).

Existem diversos modelos de processos de software. O modelo em cascata apresenta os fluxos de trabalho como fases separadas do processo, como especificação de requisitos, análise, projeto, implementação, testes. O desenvolvimento somente prossegue para o próximo estágio após o término de cada fase (SOMMERVILLE, 2007).

De acordo com Larman (2004), a análise preocupa-se com a investigação do problema e dos requisitos, enquanto o projeto preocupa-se com uma solução conceitual que atenda aos requisitos. Durante a análise orientada a objetos, o objetivo é identificar e descrever os objetos (conceitos) no domínio do problema. Durante o projeto orientado a objetos, o intuito é definir os objetos de software, e como eles interagem para atender aos requisitos.

No MineraFórum, optou-se por utilizar a análise e o projeto orientado a objetos, bem como a implementação utilizando uma linguagem orientada a objetos. Desta forma, a linguagem de programação escolhida foi Java.

O Processo Unificado (JACOBSON et al., 1999) é um modelo de processo de software. As características deste processo são:

- Dirigido a caso de uso³¹: significa que o processo de desenvolvimento prossegue através de diversos fluxos de trabalho que derivam a partir dos casos de uso, que são especificados, projetados, e testados.
- Centrado em arquitetura³²: significa que as funções do software e as formas com que elas são realizadas devem estar balanceadas. Cada caso de uso deve cuidar de uma função a ser realizada e a arquitetura da forma como será feito. O caso de uso deve adequar-se a arquitetura, enquanto esta deve permitir espaço para as necessidades do primeiro.
- Iterativo, incremental: o desenvolvimento do software é subdividido em mini-projetos, onde cada um representa uma iteração que resulta em um incremento. As iterações são os passos dos fluxos de trabalho e o incremento é o crescimento do produto. As iterações devem ser selecionadas e realizadas de forma planejada.
- Baseado em componentes: significa que o software a ser construído possui componentes interligados através de interfaces bem definidas.
- Utiliza a UML³³ (*Unified Modeling Language*).

O Processo Unificado consiste na repetição de uma série de ciclos que compõem a vida de um software. O resultado após o final de cada ciclo é a entrega de uma versão do software. Cada ciclo possui quatro fases (concepção, elaboração, construção, transição), e cada uma é subdividida em iterações. Os principais fluxos de trabalho do desenvolvimento de um software (especificação de requisitos, análise, projeto, implementação, testes) perpassam as quatro fases (JACOBSON et al., 1999).

³¹ Um caso de uso é uma parte da funcionalidade do software, que fornece um resultado ao usuário (JACOBSON et al., 1999).

³² A arquitetura de um software representa as diferentes visões do mesmo, envolvendo os aspectos estáticos e dinâmicos mais importantes (JACOBSON et al., 1999).

³³ A UML é uma linguagem que pode ser utilizada para a especificação, construção, e documentação de artefatos de software (BOOCH et al., 2000).

Na Engenharia de Software, uma técnica muito utilizada para documentar a especificação de um software é utilizar um conjunto de modelos (SOMMERVILLE, 2007). Eles são usados para descrever processos de negócio, o problema a ser resolvido, e o software a ser desenvolvido. Como será utilizado neste trabalho o Processo Unificado, os modelos gerados serão construídos com a UML.

A UML utiliza diagramas para representar graficamente um conjunto de elementos e permitir a visualização do software sob diferentes perspectivas (BOOCH et al., 2000). Neste trabalho, serão empregados o diagrama de casos de uso e o diagrama de classes. O primeiro exibe um conjunto de casos de uso, os atores³⁴, e os relacionamentos entre eles. O segundo organiza um conjunto de classes³⁵ e suas associações.

Para construir o MineraFórum foi definida a metodologia do Processo Unificado para executar o processo de desenvolvimento do software. Esta escolha deve-se ao fato de que, como o Processo Unificado é iterativo e incremental, pode-se construir e testar, gradualmente, as funcionalidades desejadas. Desta forma, é possível apresentar ao usuário final a construção parcial do trabalho, para que o mesmo possa avaliar sob sua ótica, as opções disponibilizadas pelo software. É importante ressaltar também outra característica do Processo Unificado que auxiliou na escolha do mesmo, que é a importância da criação de artefatos de documentação, como os diversos diagramas da UML.

5.2 VERSÃO 1.0

Esta seção descreve a análise efetuada para desenvolver a primeira versão do MineraFórum. A interface do sistema e o algoritmo utilizado para calcular a relevância temática também são apresentados.

³⁴ Um ator representa um papel que um ser humano, um dispositivo de hardware, ou outro software, desempenham na interação com os casos de uso (BOOCH et al., 2000).

³⁵ Uma classe é uma descrição de um conjunto de objetos que possuem as mesmas características (BOOCH et al., 2000).

5.2.1 ANÁLISE DO SISTEMA

O MineraFórum efetua uma análise qualitativa das contribuições textuais redigidas pelos alunos em um fórum de discussão. A ferramenta calcula a relevância de cada postagem. Após analisar as mensagens, o MineraFórum apresenta as seguintes informações:

- a) A quantidade total de mensagens postadas no fórum por cada aluno.
- b) A quantidade de contribuições relevantes ao tema feitas por cada aluno.
- c) A quantidade de contribuições não relevantes ao tema feitas por cada aluno.
- d) Os conceitos que foram utilizados nas postagens relevantes.
- e) A relevância temática de cada mensagem.

O desenvolvimento do MineraFórum foi iniciado a partir das ações definidas para serem realizadas pelo mesmo:

- a) Executar a leitura automática das postagens de um fórum de discussão.

Todas as mensagens digitadas pelos alunos são extraídas do fórum.

- b) Permitir a leitura de um texto de referência indicado pelo professor.

O professor pode informar ao MineraFórum um texto de referência sobre o assunto, a partir do qual serão extraídos os conceitos relevantes. O texto pode ser digitado, ou lido de um arquivo nos formatos “txt”, “doc”, ou “pdf”.

- c) Permitir que o professor digite os conceitos relevantes de um tema.

O docente pode informar ao MineraFórum quais são os conceitos importantes sobre o tópico de um fórum de discussão.

- d) Permitir a leitura e gravação dos conceitos relevantes em arquivos texto.

O professor pode gravar os conceitos relevantes extraídos do texto de referência em um arquivo texto. Também existe a possibilidade da leitura dos conceitos gravados em arquivo.

- e) Exibir um grafo dos conceitos relevantes encontrados na mineração do texto de referência, ou dos conceitos digitados pelo professor.

O MineraFórum permite a visualização do grafo gerado com os conceitos relevantes.

- f) Realizar o *stemming*³⁶ de palavras.

As palavras de um texto podem aparecer sob diversas formas, como por exemplo: no plural, no gerúndio, com sufixo, entre outras. O processo de *stemming* consiste em remover as variações das palavras, conduzindo a palavra resultante a um *stem* (raiz). Este processo é importante na etapa de pré-processamento de um texto, pois reduz a quantidade de palavras a serem analisadas. O algoritmo de *stemming* utilizado foi proposto por Orenge e Huyck (2001).

- g) Executar a leitura de uma lista de *stopwords*.

O professor pode informar um arquivo contendo uma lista de *stopwords*, que será utilizada durante o processo de mineração. Caso o docente não informe, o MineraFórum utiliza a lista de *stopwords* definida internamente no sistema.

- h) Executar a leitura de palavras que possuam equivalência semântica.

O docente pode indicar ao MineraFórum quais são as palavras que possuem equivalência semântica. Esta equivalência semântica é importante quando o software compara os conceitos redigidos pelo aluno na sua mensagem com os conceitos relevantes do texto de referência. Caso o aluno digite uma palavra que possua equivalência semântica, ela será considerada como relevante ao tema.

- i) Realizar a identificação de sinônimos a partir de um *thesaurus*.

O MineraFórum é capaz de identificar os sinônimos das palavras do texto, baseando-se num *thesaurus*. Os sinônimos são importantes quando o software compara os conceitos citados nas mensagens com os conceitos relevantes do texto de referência. Caso o aluno digite uma palavra que seja sinônima de um conceito relevante, ela será considerada. O professor pode indicar um *thesaurus* a ser utilizado pelo MineraFórum. Caso o docente não indique, o *thesaurus* a ser empregado pelo sistema utilizará a base de dados TEP 2.0³⁷, proposta por Maziero et al. (2008), Dias-da-Silva et al. (2000), Dias-da-Silva e Moraes (2003).

³⁶ *Stemming* é o processo de fundir as formas variantes de uma palavra em uma representação comum, o *stem* (ORENGO e HUYCK, 2001).

³⁷ O TeP 2.0 é uma extensão do TeP (*Thesaurus* Eletrônico para o Português do Brasil), que foi desenvolvido segundo os pressupostos da WordNet de Princeton (MAZIERO et al., 2008). Disponível para download em <http://www.nilc.icmc.usp.br/tep2/download.htm>.

- j) Permitir a definição do valor mínimo da frequência das palavras para a mineração do texto de referência.

O docente pode estabelecer a quantidade mínima de ocorrências que uma palavra deve ter no texto, para que seja considerada como um conceito relevante no processo de mineração.

- k) Realizar a mineração do texto de referência indicado pelo professor.

- l) Executar a mineração das mensagens do fórum.

- m) Permitir a definição do valor mínimo do Coeficiente de Relevância Temática (CRT) que será considerado na análise das mensagens dos alunos.

O professor pode definir o valor mínimo do CRT que deve ser considerado na análise de uma mensagem.

- n) Executar o cálculo do Coeficiente de Relevância Temática de cada mensagem.

O MineraFórum calcula o CRT de cada contribuição textual. As mensagens que possuem valor de CRT maior ou igual ao mínimo estabelecido pelo professor são consideradas como relevantes.

- o) Descobrir a quantidade total de mensagens colocadas por cada aluno no fórum.

O MineraFórum descobre a quantidade total de mensagens digitadas por cada aluno, para permitir que o professor saiba quantas vezes o discente realizou interações no fórum.

- p) Realizar o cálculo da quantidade de contribuições relevantes colocadas por cada aluno no fórum.

O MineraFórum calcula a quantidade de mensagens relevantes digitadas por cada aluno. A partir deste cálculo, é possível identificar quantas postagens relevantes o discente escreveu, dentro da quantidade total de mensagens redigidas por ele.

- q) Gerar um relatório visual para o professor, contendo informações sobre as mensagens postadas pelos alunos no fórum de discussão.

As informações disponibilizadas indicam para cada aluno: o nome do discente, a quantidade total de mensagens postadas, a quantidade de mensagens relevantes ao tema da discussão, a quantidade de mensagens que não envolveram conceitos sobre o tópico do debate, e os conceitos utilizados nas contribuições relevantes.

A primeira etapa no processo do desenvolvimento do MineraFórum foi o levantamento dos requisitos do mesmo. Segundo Sommerville (2007), os requisitos de um software representam as descrições dos serviços fornecidos pelo mesmo e as suas restrições operacionais. A partir das ações realizadas pelo MineraFórum, definidas anteriormente nesta seção, foram especificados os requisitos funcionais³⁸ necessários a versão 1.0 (quadro 5).

Quadro 5 - Requisitos funcionais da versão 1.0 do MineraFórum

Requisito	Descrição
REQ.01	O MineraFórum deve executar a leitura automática das postagens de um fórum de discussão.
REQ.02	O MineraFórum deve ler um texto de referência indicado pelo professor. O texto pode ser digitado, ou lido de um arquivo nos formatos “txt”, “doc”, ou “pdf”.
REQ.03	O MineraFórum deve admitir a digitação dos conceitos relevantes do tema em debate.
REQ.04	O MineraFórum deve possibilitar a gravação dos conceitos relevantes em um arquivo texto.
REQ.05	O MineraFórum deve permitir a leitura dos conceitos relevantes gravados em arquivo texto.
REQ.06	O MineraFórum deve exibir um grafo dos conceitos relevantes encontrados na mineração do texto de referência, ou dos conceitos digitados pelo professor.
REQ.07	O MineraFórum deve realizar o <i>stemming</i> de palavras.
REQ.08	O MineraFórum deve executar a leitura de uma lista de <i>stopwords</i> .
REQ.09	O MineraFórum deve efetuar a leitura de palavras que possuam equivalência semântica.
REQ.10	O MineraFórum deve realizar a identificação de sinônimos a partir de um <i>thesaurus</i> .
REQ.11	O MineraFórum deve permitir a definição do valor mínimo da frequência das palavras para a mineração do texto de referência.
REQ.12	O MineraFórum deve realizar a mineração do texto de referência indicado pelo professor.
REQ.13	O MineraFórum deve executar a mineração das mensagens do fórum.
REQ.14	O MineraFórum deve permitir a definição do valor mínimo do CRT que será considerado na análise das mensagens dos alunos.
REQ.15	O MineraFórum deve calcular o CRT de cada mensagem.
REQ.16	O MineraFórum deve descobrir a quantidade total de mensagens colocadas por cada aluno no fórum.
REQ.17	O MineraFórum deve calcular a quantidade de contribuições relevantes colocadas por cada aluno no fórum.
REQ.18	O MineraFórum deve gerar um relatório visual, contendo informações sobre as mensagens postadas pelos alunos no fórum. Para cada aluno, devem ser exibidos: o nome do discente, a quantidade total de mensagens postadas, a quantidade de mensagens relevantes ao tema da discussão, a quantidade de mensagens que não envolveram conceitos sobre o tópico do debate, e os conceitos utilizados nas contribuições relevantes.

Fonte: o autor

³⁸ Os requisitos funcionais de um software descrevem o que o mesmo deve fazer (SOMMERVILLE, 2007).

A partir da especificação dos requisitos funcionais, foi elaborado o diagrama dos casos de uso do MineraFórum (figura 16).

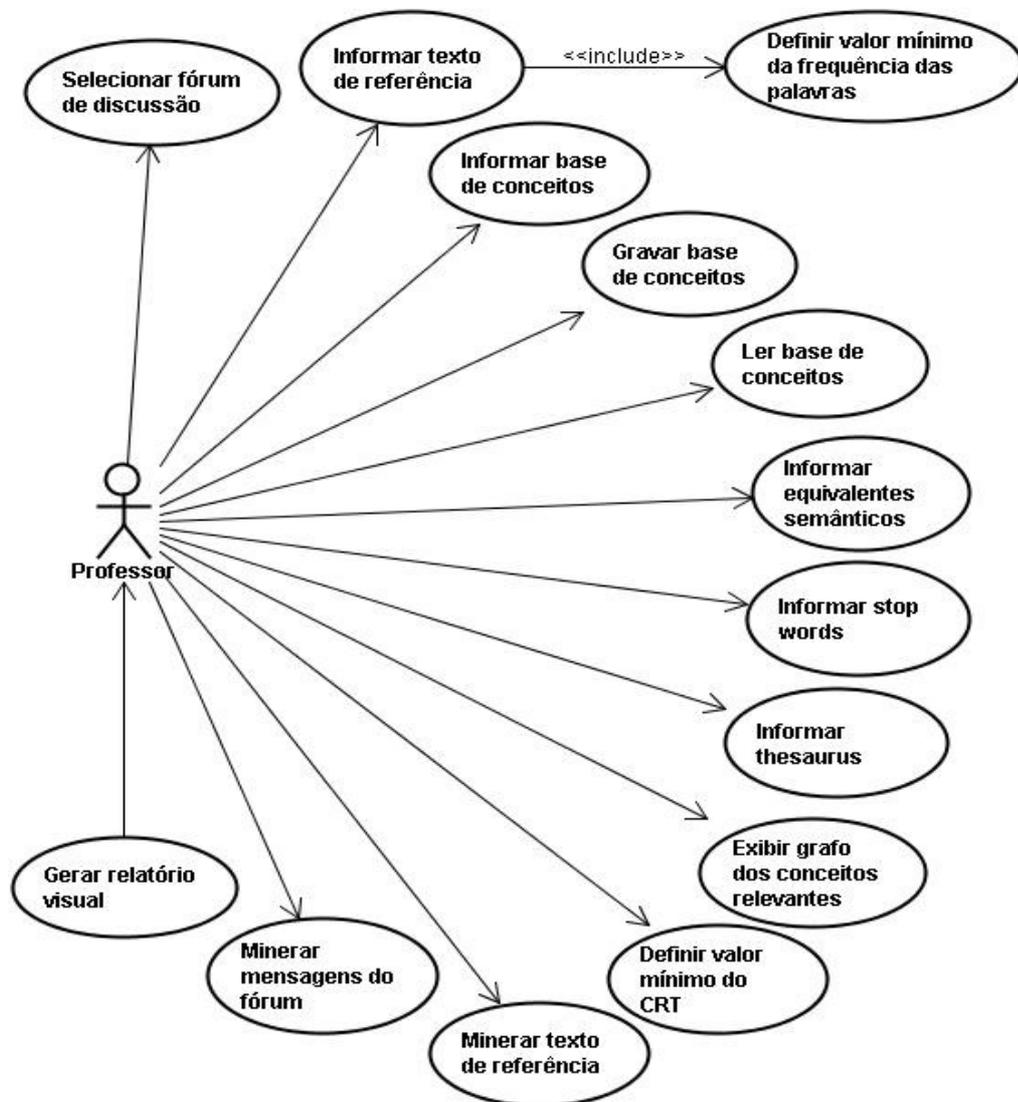


Figura 16 - Diagrama de Casos de Uso da versão 1.0 do MineraFórum

Fonte: o autor

O quadro 6 apresenta a descrição dos casos de uso citados no diagrama da figura 16. A partir da explicação de cada um é possível compreender os recursos oferecidos pela versão 1.0 do MineraFórum.

Quadro 6 - Descrição dos casos de uso da versão 1.0 do MineraFórum

Caso de Uso (CDU)	Descrição
CDU01 - Selecionar fórum de discussão	Permite a escolha de um fórum de discussão.
CDU02 - Informar texto de referência	Possibilita que o docente informe um texto de referência sobre o tema do debate.
CDU03 - Definir valor mínimo da frequência das palavras	Permite que o professor indique a quantidade mínima de ocorrências que uma palavra deve ter no texto, para que seja considerada como um conceito relevante no processo de mineração.
CDU04 - Informar base de conceitos	Admite a digitação dos conceitos relevantes do tema da discussão.
CDU05 - Gravar base de conceitos	Permite a gravação dos conceitos relevantes em um arquivo texto.
CDU06 - Ler base de conceitos	Possibilita a leitura dos conceitos relevantes gravados em arquivo texto.
CDU07 - Informar equivalentes semânticos	Permite que o docente informe um arquivo que contenha palavras com equivalência semântica.
CDU08 - Informar <i>stopwords</i>	Possibilita que o professor informe um arquivo que relacione uma lista de <i>stopwords</i> .
CDU09 - Informar <i>thesaurus</i>	Permite que o docente indique um arquivo que contenha um <i>thesaurus</i> .
CDU10 - Exibir grafo dos conceitos relevantes	Exibe um grafo dos conceitos relevantes encontrados na mineração do texto de referência, ou dos conceitos digitados pelo professor.
CDU11 - Definir valor mínimo do CRT	Permite a definição do valor mínimo do CRT que será considerado na análise das mensagens dos alunos.
CDU12 - Minerar texto de referência	Executa a mineração do texto de referência indicado pelo professor.
CDU13 - Minerar mensagens do fórum	Executa a mineração das mensagens do fórum.
CDU14 - Gerar relatório visual	Exibe um relatório para o professor, contendo informações sobre as mensagens postadas pelos alunos em um fórum de discussão. Para cada aluno, devem ser exibidos: o nome do discente, a quantidade total de mensagens postadas, a quantidade de mensagens relevantes ao tema da discussão, a quantidade de mensagens que não envolveram conceitos sobre o tópico do debate, e os conceitos utilizados nas contribuições relevantes.

Fonte: o autor

A figura 17 apresenta o diagrama das classes que constituem a versão 1.0 do MineraFórum. Este diagrama apresenta os nomes e atributos das classes, bem como a associação entre elas.

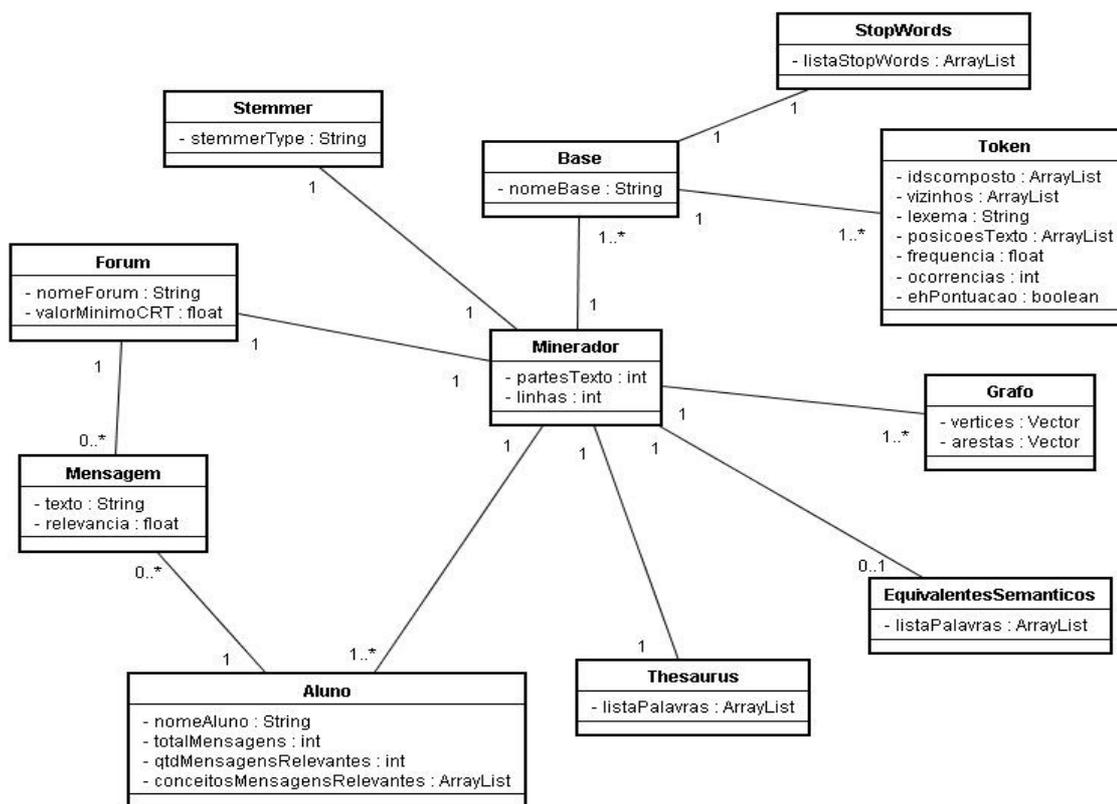


Figura 17 - Diagrama de Classes da versão 1.0 do MineraFórum

Fonte: o autor

O quadro 7 relata a descrição de cada uma das classes do diagrama da figura 17.

Quadro 7 - Descrição das classes da versão 1.0 do MineraFórum

Classe	Descrição
Aluno	Representa a coletânea das informações sobre as mensagens postadas por um aluno no fórum. Estas informações são: o nome do aluno, a quantidade total de mensagens postadas, a quantidade de mensagens relevantes ao tema da discussão, e os conceitos relativos ao tema que foram abordados nas mensagens.
Base	Representa o conjunto dos conceitos extraídos de um texto.
EquivalentesSemanticos	Representa uma lista das palavras que possuem equivalência semântica com os conceitos importantes de um tema.
Forum	Representa o fórum de discussão.
Grafo	Constrói o grafo dos conceitos extraídos de um texto.
Mensagem	Representa uma mensagem postada pelo aluno.
Minerador	Executa o processo de mineração em um fórum de discussão.
Stemmer	Executa o processo de <i>stemming</i> nas palavras de um texto.
StopWords	Lista de palavras que podem ser descartadas no processo de mineração, como: advérbios, artigos, preposições, entre outras.
Thesaurus	Representa o <i>thesaurus</i> a ser utilizado no processo de mineração.
Token	Representa um símbolo extraído do texto (token), que pode ser uma palavra, um número, um sinal de pontuação. Armazena informações de cada token, como: a quantidade de ocorrências no texto, a frequência no texto, as posições no texto onde ele aparece, os tokens vizinhos.

Fonte: o autor

Para executar o processo de mineração do fórum de discussão, o MineraFórum executa cinco etapas, apresentadas na figura 18. Estas etapas podem ser descritas como:

- Extração das mensagens do fórum: para obter todas as mensagens do fórum escolhido pelo docente.
- Mineração do texto de referência.
- Cálculo da relevância das mensagens.
- Organização das mensagens: para organizar e agrupar as mensagens de cada aluno.
- Geração do relatório da mineração: para apresentar os resultados da mineração das postagens do fórum de discussão.

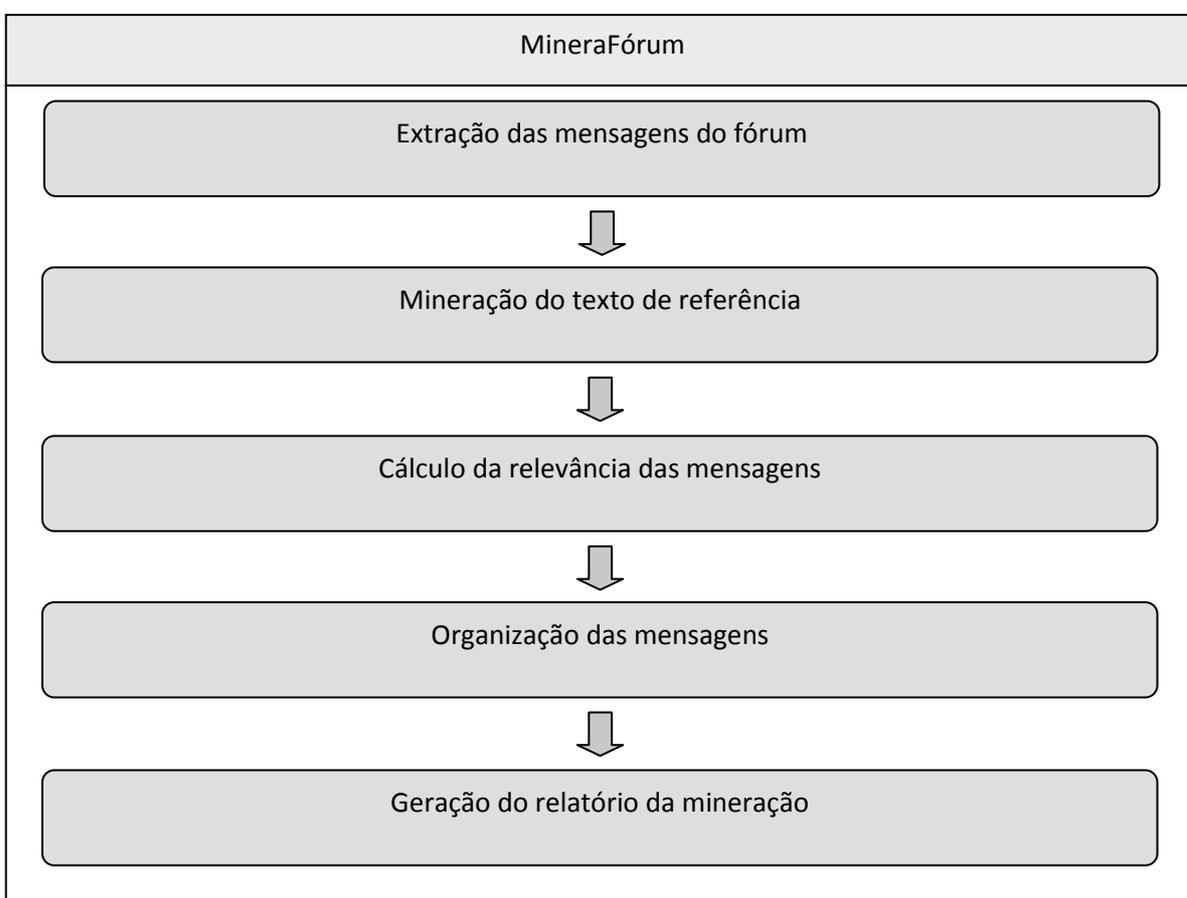


Figura 18 - Etapas do MineraFórum
Fonte: o autor

5.2.2 INTERFACE DO SISTEMA

Esta seção descreve os recursos disponibilizados na versão 1.0 do MineraFórum. Nas figuras 19 a 24 são apresentadas a interface do MineraFórum.

A figura 19 apresenta a tela principal do sistema. Nela observam-se os itens de menu:

- Arquivo: permite o acesso a alguns recursos da ferramenta, que serão explicados na figura 22.
- Ajuda: apresenta um tutorial sobre a utilização da ferramenta e informações sobre a versão do sistema.
- Idioma/Language: permite a escolha da língua portuguesa ou inglesa. A língua selecionada será utilizada para exibir as mensagens do sistema, as opções dos menus e as legendas dos botões.

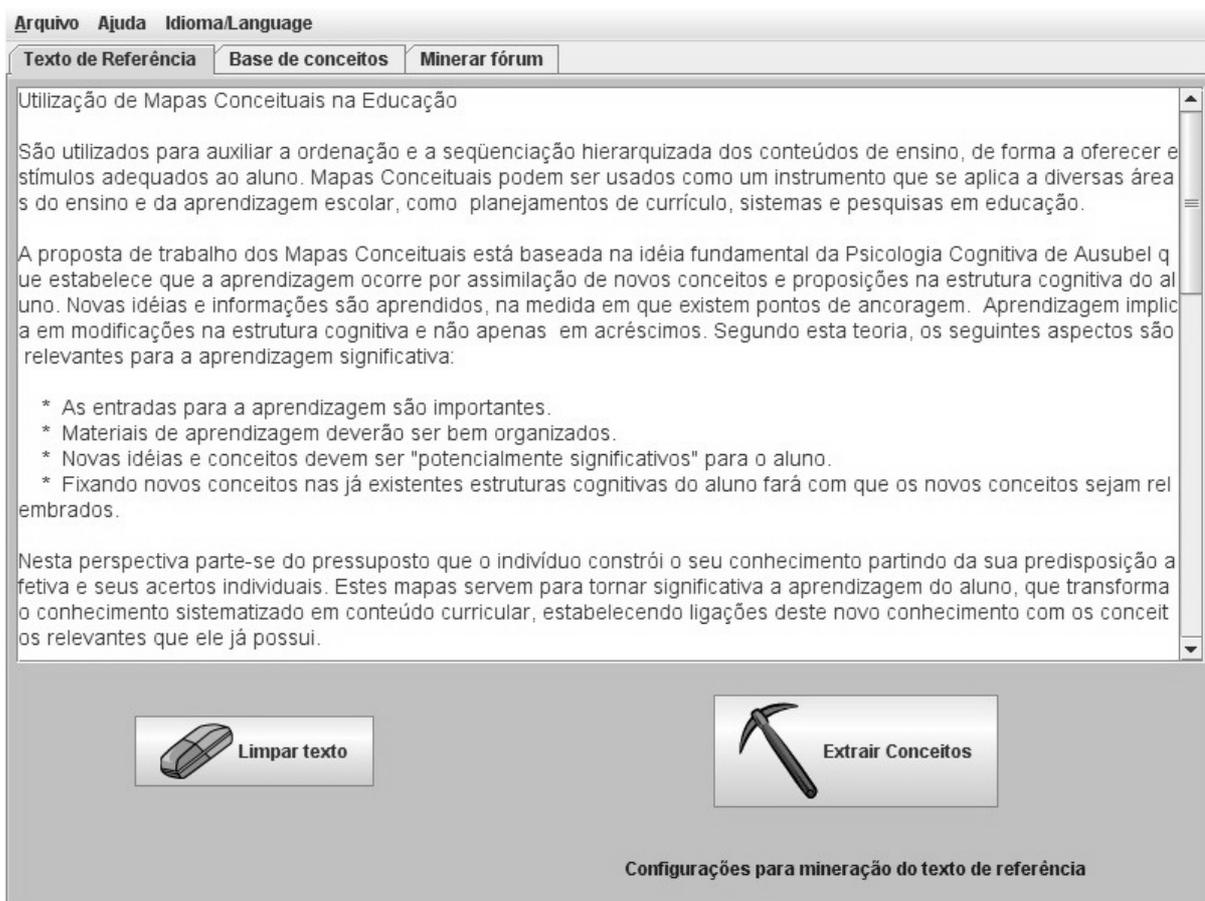


Figura 19 - Tela principal da versão 1.0 do MineraFórum
Fonte: o autor

Na aba “Texto de Referência” da tela principal visualiza-se os seguintes recursos:

- Na área branca é possível digitar um texto de referência sobre o tema debatido no fórum de discussão.
- O botão “Limpar texto” apaga o texto digitado.
- O botão “Extrair Conceitos” realiza a mineração do texto de referência e apresenta a tela da figura 21, na qual é possível visualizar os conceitos relevantes encontrados no texto.

Ao selecionar a mensagem “Configurações para mineração do texto de referência” é exibida a tela da figura 20. Nela é possível indicar a quantidade mínima de ocorrências que uma palavra deve ter no texto de referência, para ser considerada como um conceito relevante no processo de mineração. Caso não seja informada a frequência mínima, o próprio sistema definirá este valor.

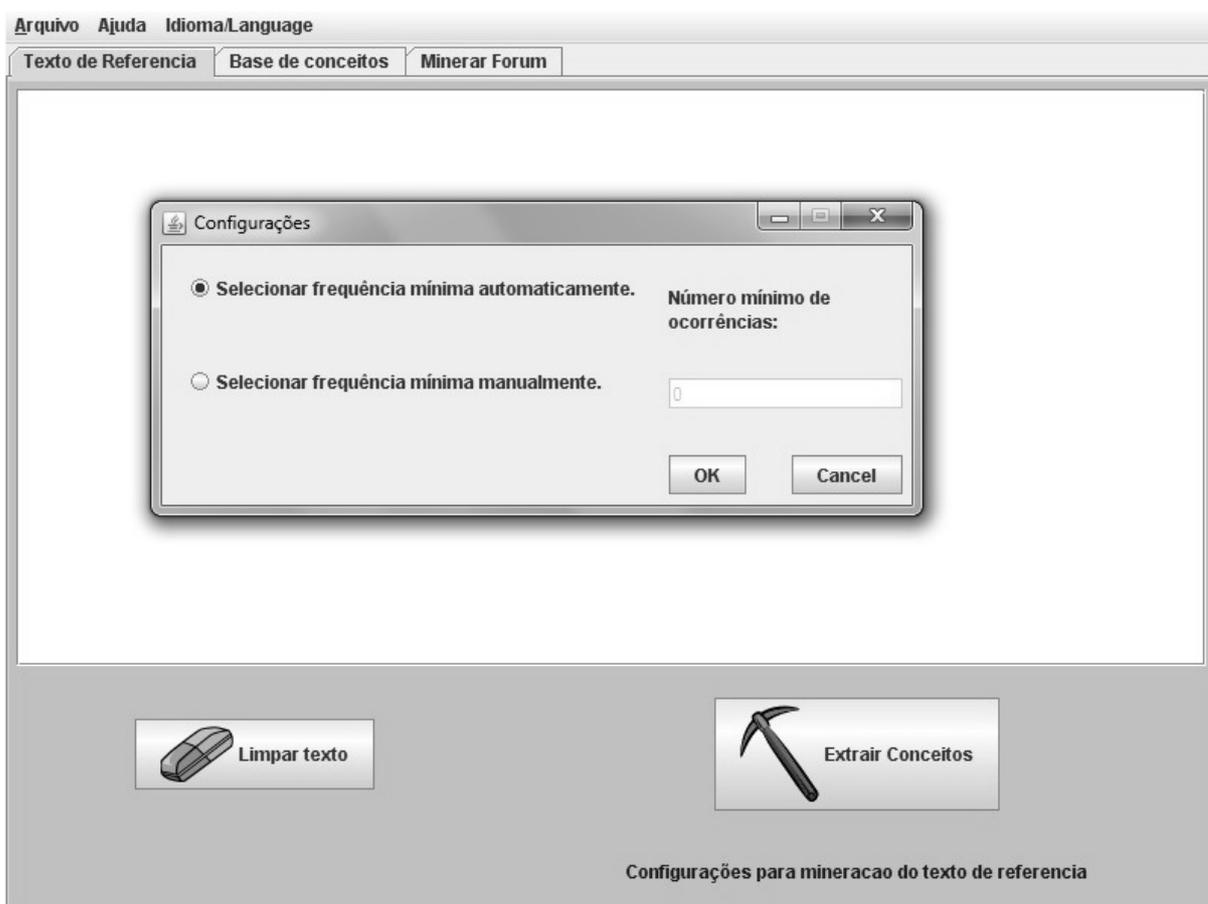


Figura 20 - Tela da versão 1.0 para indicar a frequência mínima no texto de referência
Fonte: o autor

A figura 21 apresenta a tela principal com a aba “Base de conceitos” selecionada. Nesta aba há uma coluna com todos os conceitos extraídos do texto de referência (coluna

“Conceitos”). Pode-se apagar um conceito desta coluna selecionando o mesmo e pressionando a tecla “Delete”.

A coluna denominada “Conceitos Relacionados” exibe os conceitos relacionados aos citados na coluna “Conceitos”. A coluna “Conceitos Não Relacionados” apresenta os conceitos que não estão relacionados aos mencionados na coluna “Conceitos”. Caso seja necessário relacionar um conceito ou retirar o relacionamento, basta utilizar os botões:

- “Adicionar conceito relacionado”: seleciona-se um conceito que está na coluna “Conceitos Não Relacionados” e depois se pressiona este botão. Desta forma, o conceito selecionado irá aparecer na coluna “Conceitos Relacionados”.
- “Remover conceito relacionado”: seleciona-se um conceito que está na coluna “Conceitos Relacionados” e depois se pressiona este botão. Desta forma, o conceito selecionado irá aparecer na coluna “Conceitos Não Relacionados”.

Também há a opção de inserir um novo conceito que não esteja na base. Basta digitá-lo no campo de texto “Novo Conceito” e pressionar o botão “Enviar”. Com o botão “Gerar grafo” é possível exibir um grafo dos conceitos.

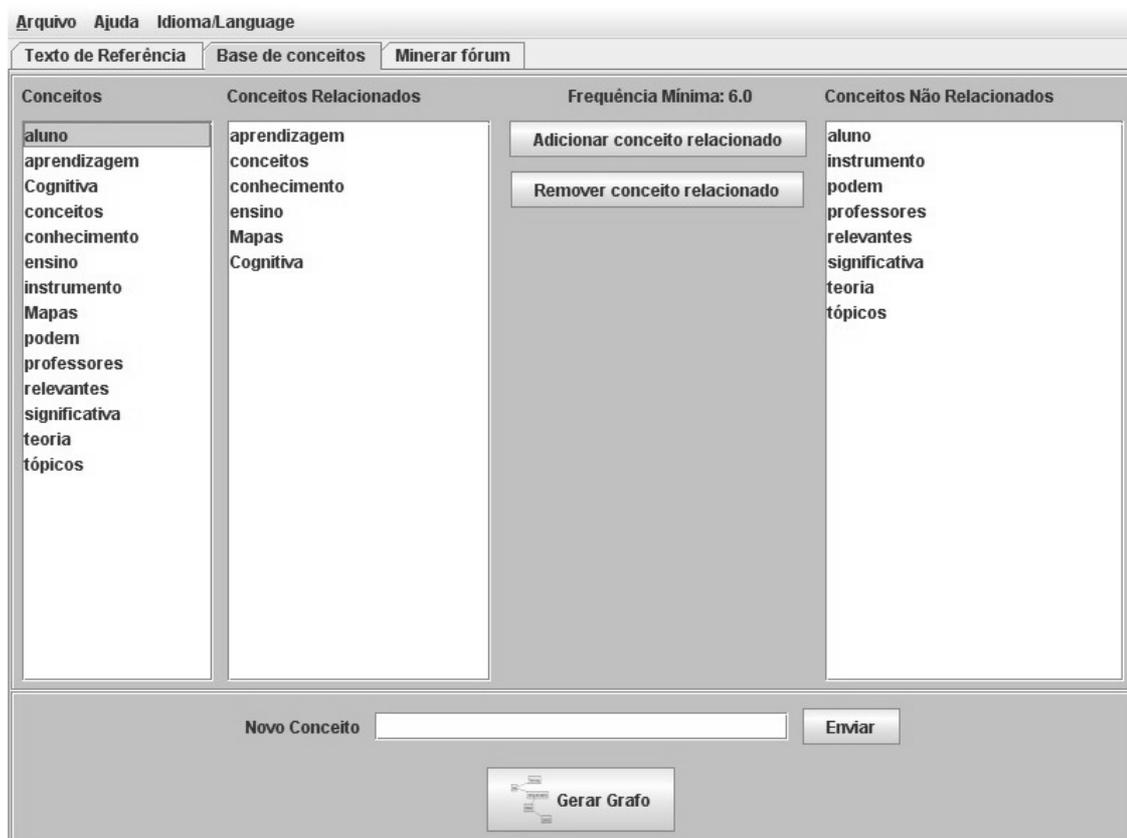


Figura 21 - Tela principal da versão 1.0 com a aba “Base de conceitos” selecionada
Fonte: o autor

A figura 22 apresenta a tela principal com o menu “Arquivo” selecionado. As opções disponíveis são:

- “Importar texto de referência”: ao selecionar esta opção, aparece uma tela na qual é possível carregar um arquivo com o texto de referência. Este arquivo pode estar nos formatos “txt”, “doc”, “pdf”.
- “Criar nova base de conceitos”: ao escolher esta opção é apresentada a tela da figura 21.
- “Salvar base de conceitos”: com esta opção pode-se gravar a base de conceitos apresentada na tela da figura 21. A base de conceitos será gravada em um arquivo no formato “txt”.
- “Carregar base de conceitos”: com esta opção pode-se carregar um arquivo que contenha uma base de conceitos.
- “Limpar base de conceitos”: esta opção permite excluir os conceitos da base que está sendo utilizada.
- “Equivalentes semânticos”: com esta opção pode-se carregar um arquivo contendo uma lista de palavras com equivalência semântica.
- “*Stopwords*”: ao selecionar esta opção, aparece uma janela, para carregar um arquivo contendo uma lista de *stopwords*.
- “*Thesaurus*”: com esta opção pode-se selecionar um arquivo que contenha um *thesaurus*.

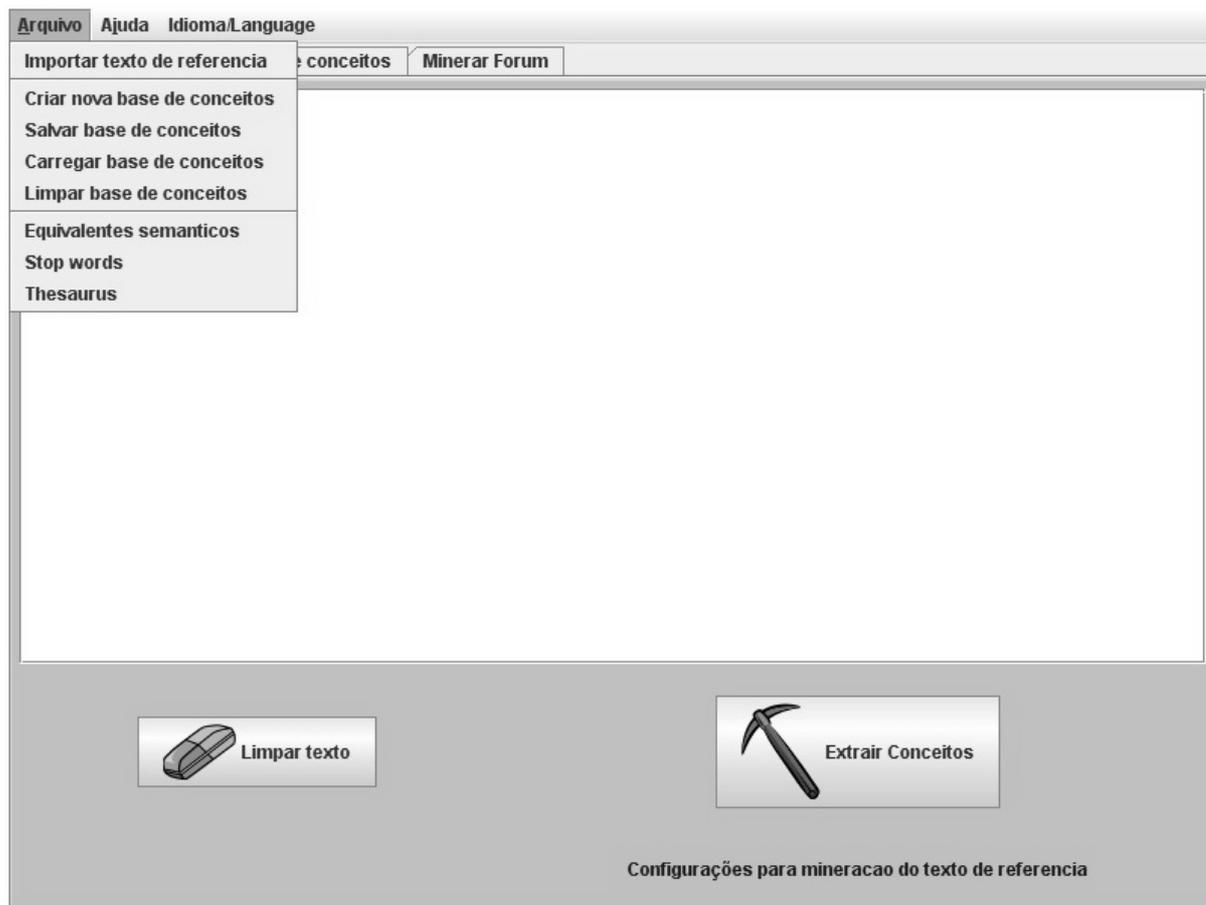


Figura 22 - Tela principal da versão 1.0 com o menu “Arquivo” selecionado
Fonte: o autor

A figura 23 apresenta a tela principal com a aba “Minerar Fórum” selecionada. O botão “Minerar Fórum” realiza a mineração das mensagens redigidas pelos alunos no fórum de discussão. O resultado é apresentado na área branca. As informações sobre cada aluno são: o nome do discente, o total de postagens, a quantidade de mensagens relevantes ao tema da discussão, a quantidade de mensagens que não envolveram conceitos sobre o tópico do debate, e os conceitos utilizados nas contribuições relevantes.

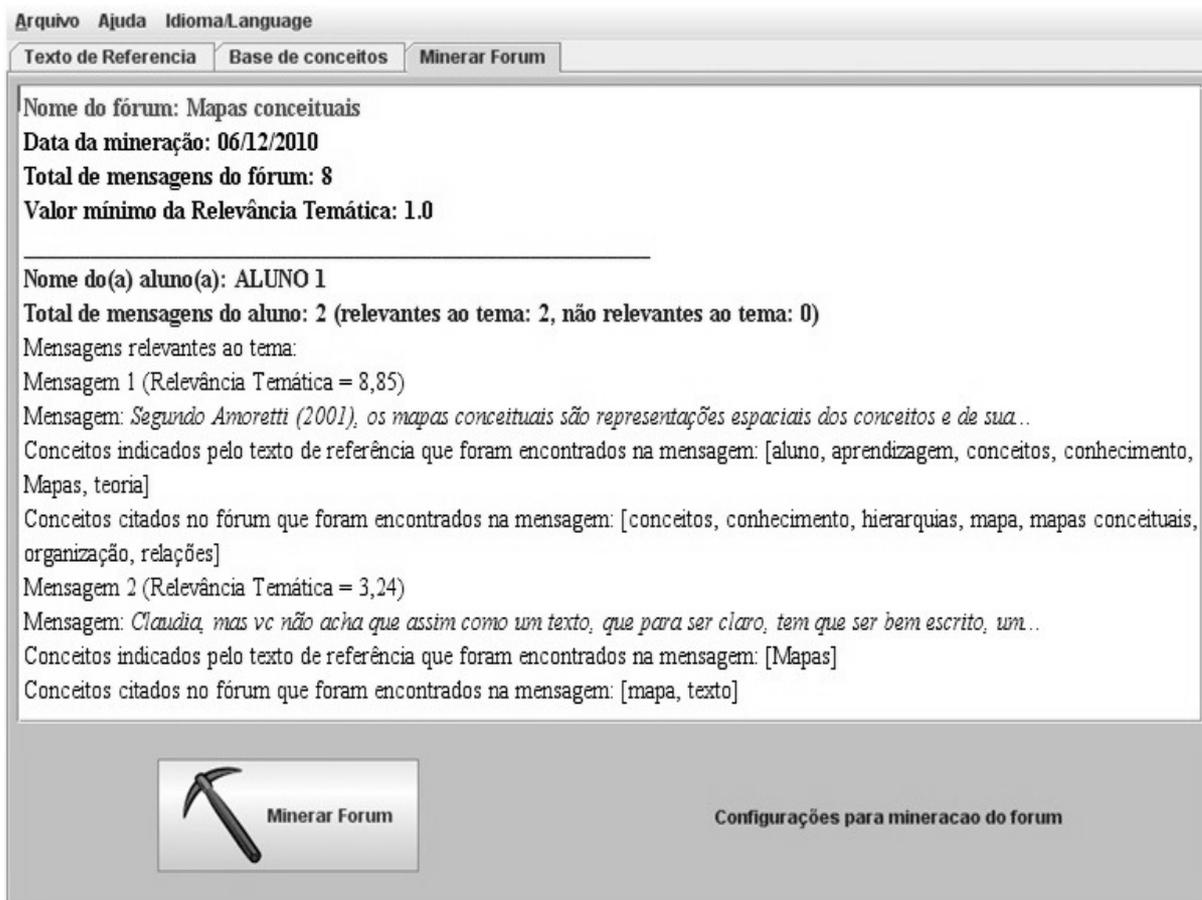


Figura 23 - Tela principal da versão 1.0 com a aba “Minerar Fórum” selecionada

Fonte: o autor

Ao pressionar a mensagem “Configurações para mineração do fórum” é apresentada a tela da figura 24. Nesta tela, é possível definir o valor mínimo do CRT que deve ser considerado na análise de uma mensagem. Caso não seja informado, o valor mínimo acatado será um (1).

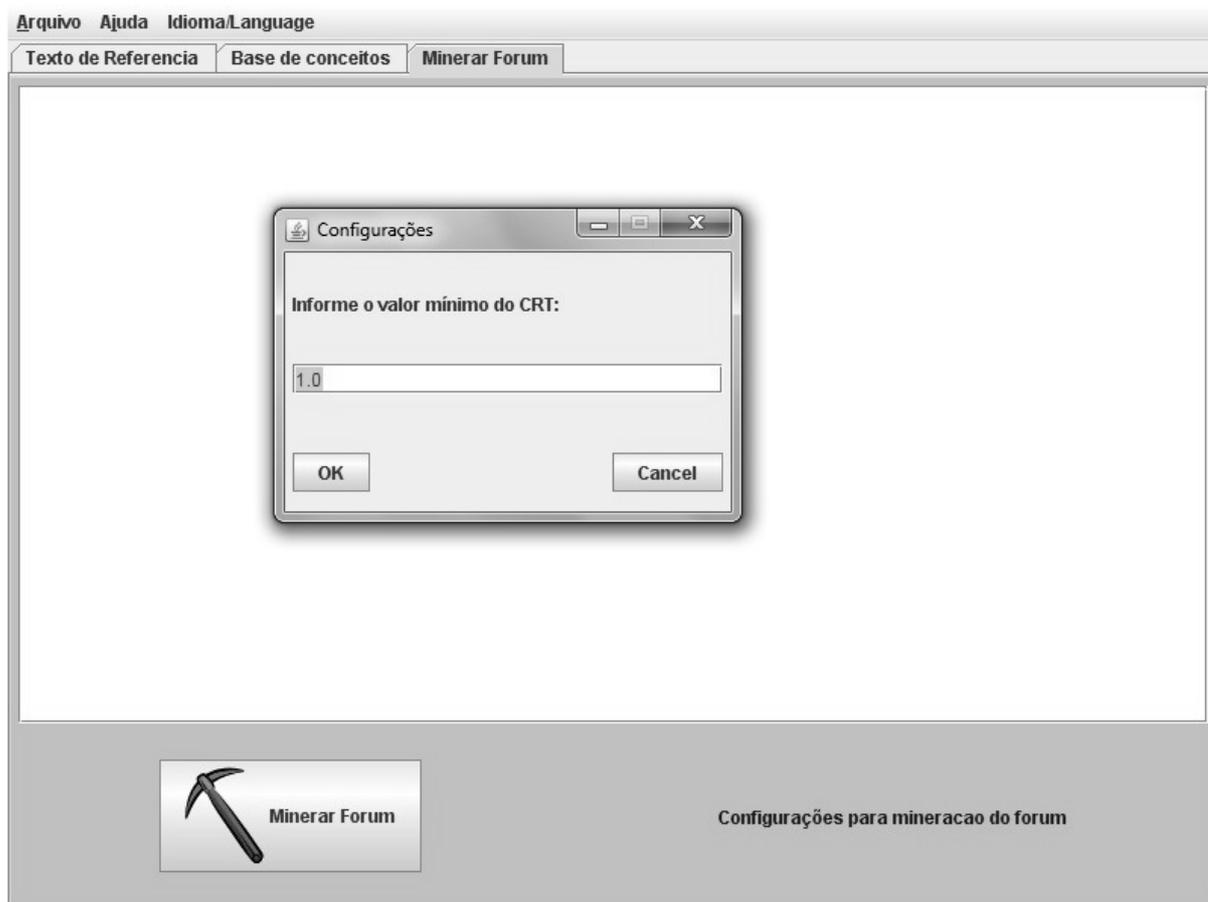


Figura 24 - Tela da versão 1.0 para informar o valor mínimo do CRT

Fonte: o autor

A próxima seção descreve o algoritmo utilizado na versão 1.0 para calcular a relevância temática de uma mensagem redigida no fórum.

5.2.3 ALGORITMO PARA CALCULAR A RELEVÂNCIA TEMÁTICA DE UMA MENSAGEM

Conforme explicado anteriormente, o MineraFórum calculará a relevância temática de contribuições textuais. Antes de descrever como este procedimento é realizado, faz-se necessário apresentar alguns conceitos sobre a teoria dos grafos.

Um grafo G é uma tripla ordenada $(V(G), E(G), \varphi_G)$, onde $V(G)$ é um conjunto não vazio de vértices, $E(G)$ é um conjunto de arestas, e φ_G é uma função de incidência que associa cada aresta de G a um par ordenado de vértices de G . Se e é uma aresta e u e v são vértices, então diz-se que e une u e v ; os vértices u e v são as extremidades de e . Um grafo H é um subgrafo de G ($H \subseteq G$) se $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$, e φ_H é a restrição de φ_G para $E(H)$

(BONDY e MURTY, 1982).

Os vértices e arestas de um grafo também podem conter informação. Quando esta informação é um rótulo simples, então o grafo é chamado de grafo rotulado (BENGOETXEA, 2002).

Dois grafos G e H são isomórficos se existem bijeções $\theta:V(G)\rightarrow V(H)$ e $\phi:E(G)\rightarrow E(H)$ tal que $\phi_G(e)=uv$ se e somente se $\phi_H(\phi(e))=\theta(u)\theta(v)$. Este par de mapeamentos é chamado de um isomorfismo entre G e H (BONDY e MURTY, 1982). Este tipo de problema também é conhecido como correspondência exata de grafos (BENGOETXEA, 2002).

Segundo Bengoetxea (2002), em alguns problemas de correspondência de grafos não é possível encontrar um isomorfismo entre dois grafos analisados. Nestes casos, o problema em questão não consiste em procurar a forma exata de corresponder os vértices de um grafo com os vértices do outro, mas em encontrar a melhor correspondência entre eles. Isto leva a uma classe de problemas conhecida como correspondência inexata de grafos.

Segundo Bunke (2000) existem diversos algoritmos para encontrar correspondência entre grafos. Existem métodos que encontram a melhor solução, mas necessitam de muito tempo, por se tratar de um problema de complexidade NP^{39} . Por outro lado, existem métodos aproximados que são limitados polinomialmente no número de passos de computação, mas podem não conseguir encontrar a solução ótima.

O algoritmo padrão para detectar isomorfismo de grafos e subgrafos foi proposto por Ullman (1976). Este é um algoritmo de enumeração projetado para encontrar todos os isomorfismos entre um grafo e os subgrafos de outro grafo.

Outra forma de analisar a correspondência entre grafos é através de medidas de similaridade entre eles. Chartrand et al. (1998) definem um método para calcular a distância entre dois grafos conectados, que pode ser utilizada para analisar a similaridade. Bunke e Shearer (1998) propõem outra forma para computar a distância, que é baseada no máximo subgrafo comum de dois grafos. Blondel e Van Dooren (2003) introduziram um conceito de similaridade entre vértices de dois grafos direcionados, propondo uma matriz na qual os seus elementos indicam uma pontuação de semelhança entre os vértices.

O MineraFórum constrói um grafo a partir do texto de referência informado pelo

³⁹ A classe de complexidade NP é o conjunto de problemas de decisão que podem ser aceitos não deterministicamente em tempo polinomial (GOODRICH e TAMASSIA, 2002).

professor, ou a partir dos conceitos relevantes indicados pelo docente na base de conceitos. Um grafo também é gerado a partir de cada contribuição textual extraída do fórum de discussão. Para calcular a relevância temática de uma mensagem, o MineraFórum deve analisar a correspondência entre o grafo gerado a partir da indicação do professor e o grafo construído a partir da mensagem.

Para o caso do MineraFórum, onde todas as mensagens postadas pelos alunos devem ser analisadas, o objetivo não é verificar se uma mensagem possui exatamente todos os conceitos e associações indicadas pelo professor. Isto significa que o grafo da mensagem do discente não é isomórfico em relação ao grafo criado a partir da indicação do docente. Desta forma, uma análise de similaridade entre os dois grafos é capaz de apresentar melhores resultados.

Para calcular a similaridade entre o grafo de uma mensagem do fórum e o grafo gerado a partir da indicação do professor, poderia ter sido utilizado um dos métodos expostos anteriormente (BUNKE e SHEARER, 1998; CHARTRAND et al., 1998; BLONDEL e VAN DOOREN, 2003).

Considerando que o grafo de uma mensagem possui N vértices, e dentre estes N vértices existem S ($S \leq N$) que são equivalentes⁴⁰ aos vértices do grafo do professor, o autor desta pesquisa optou por executar um algoritmo que analisasse somente estes vértices nos dois grafos. Observando a fórmula para o cálculo do CRT na versão 1.0 (citada na seção 4.5.1) e considerando que:

- G_M é o grafo de uma mensagem com N vértices;
- G_P é o grafo criado a partir da indicação do professor com T vértices;
- $\text{equiv}(V_a, V_b)$ verifica se os vértices V_a e V_b são equivalentes, indicando o resultado 1 caso eles sejam, e indicando 0 caso contrário;
- dois vértices de um grafo são vizinhos se possuem uma aresta que os une.

O algoritmo para o cálculo do CRT é indicado abaixo:

1. $NC = 0$

⁴⁰ No contexto desta pesquisa, um vértice de um grafo gerado a partir da indicação do professor é equivalente a um vértice de um grafo gerado a partir de uma mensagem, se eles possuem conteúdo equivalente, isto é, se possuem as mesmas palavras, ou se as palavras podem ser reduzidas ao mesmo radical, ou se possuem sinônimos, ou se possuem palavras com equivalência semântica.

2. $NA = 0$
3. Para todo vértice V_i ($i \leq N$) de G_M
 4. Para todo vértice V_j ($j \leq T$) de G_P
 5. $NC = NC + \text{equiv}(V_i, V_j)$
 6. Se $\text{equiv}(V_i, V_j) = 1$ Então
 7. Para todo vizinho do vértice V_i
 8. Para todo vizinho do vértice V_j
 9. $NA = NA + \text{equiv}(\text{vizinho de } V_i, \text{vizinho de } V_j)$
 10. Retorne ao passo 8
 11. Retorne ao passo 7
 12. FimSe
 13. Retorne ao passo 4
14. Retorne ao passo 3
15. $CRT = NC + NA$

5.2.4 EXPERIMENTOS COM A VERSÃO 1.0

Esta seção apresenta os experimentos realizados com a versão 1.0. Cabe destacar alguns pontos importantes:

- As postagens foram analisadas pelo MineraFórum de acordo com a ordem de inclusão nos fóruns.
- O objetivo dos experimentos foi investigar a relevância temática das mensagens dos fóruns.
- Os experimentos foram realizados com fóruns de discussão do Rooda.
- O termo de consentimento para a análise das mensagens dos fóruns encontra-se no Apêndice C.
- Os nomes dos alunos, professores e tutores foram preservados.

Com a primeira versão do MineraFórum foram realizados três experimentos, com fóruns distintos. Em todas as experiências, o valor mínimo do Coeficiente de Relevância Temática (CRT) foi definido como 1.0. O primeiro experimento foi realizado na disciplina “Tópicos Especiais Z1” da turma do Doutorado em Informática na Educação, que ocorreu no primeiro semestre de 2008. O tema do fórum foi “Comunidades Virtuais”. O fórum contou com a participação de 05 alunos, que escreveram 25 mensagens. A figura 25 apresenta a quantidade de contribuições relevantes identificadas.

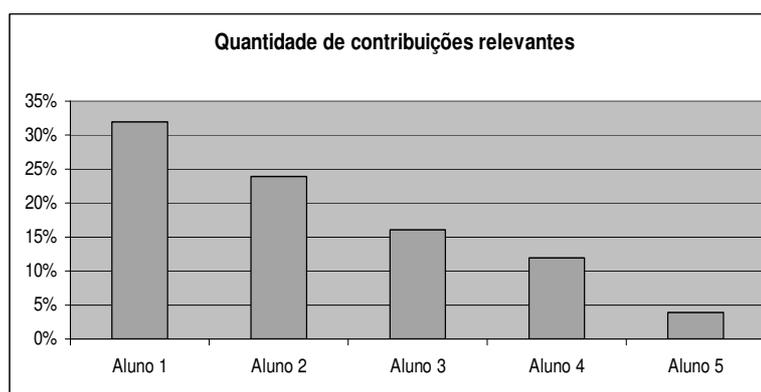


Figura 25 - Quantidade de contribuições relevantes do primeiro experimento
Fonte: o autor

O segundo e terceiro experimentos foram realizados em um fórum de discussão da disciplina “Seminário Integrador VII – B” de uma turma do Curso de Graduação em Pedagogia, que ocorreu no segundo semestre de 2009. O tema foi “Aprender com os outros”. A turma foi dividida em dois grupos, e cada grupo contribuiu em um fórum. O primeiro fórum contou com a participação de 27 alunos, que realizaram 45 contribuições textuais. O segundo fórum contou com a participação de 29 alunos, que escreveram 67 mensagens. As figuras 26 e 27 apresentam a quantidade de contribuições relevantes identificadas.

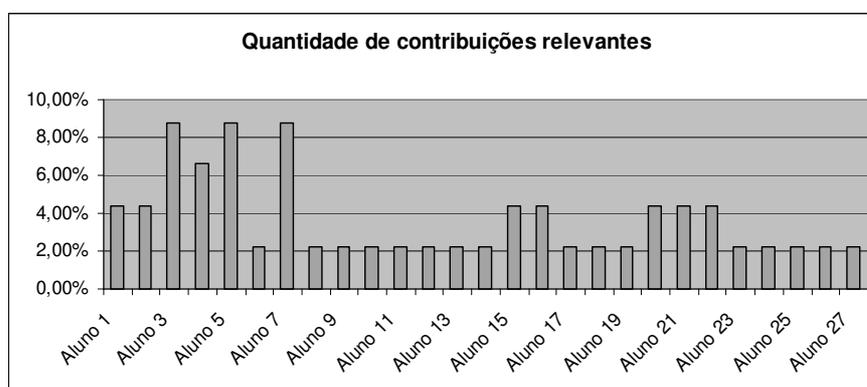


Figura 26 - Quantidade de contribuições relevantes do primeiro grupo
Fonte: o autor

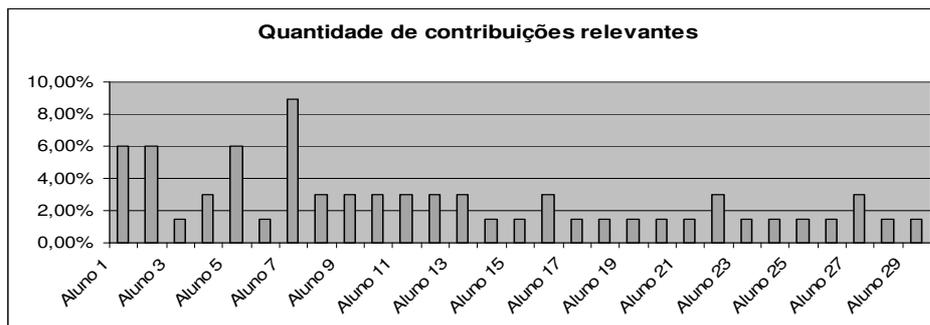


Figura 27 - Quantidade de contribuições relevantes do segundo grupo
Fonte: o autor

5.2.5 ANÁLISE E DISCUSSÃO DOS RESULTADOS DA VERSÃO 1.0

Ao iniciar esta seção, cabe lembrar a metodologia para o cálculo da relevância de cada postagem, realizada pelo MineraFórum. Na versão 1.0 a relevância é diagnosticada a partir da análise temática da mensagem.

A partir dos resultados apresentados nas figuras 25, 26, 27, percebeu-se que a primeira versão do MineraFórum foi capaz de calcular a relevância temática das mensagens. A quantidade de contribuições relevantes ao tema, redigidas por cada aluno, foi exibida nos gráficos.

Os resultados obtidos nos experimentos possibilitaram a observação de aspectos que foram aperfeiçoados na versão 2.0, como:

- a) A fórmula para o cálculo da relevância temática foi aprimorada. O termo NA foi substituído por DC e PC. Eles calculam a distância entre dois vértices equivalentes nos grafos e a quantidade de nós vizinhos de um vértice equivalente.
- b) Algumas mensagens citaram conceitos importantes relacionados ao tema do debate que não estavam no texto de referência. Estes conceitos não foram identificados como sinônimos e não estavam na lista de equivalentes semânticos. Desta forma, verificou-se a importância de construir um grafo com a união de todas as postagens, denominado grafo do fórum. O cálculo da relevância temática de cada mensagem também deve considerar a análise da correspondência entre os grafos da mensagem e do fórum.

- c) Alguns alunos escreveram mensagens com alta relevância, mas que apresentavam muita semelhança com outra já postada no fórum. Desta forma, verificou-se a necessidade de implementar no sistema a identificação desta situação.

A próxima seção descreve as alterações ocorridas na versão 1.0 do sistema para o desenvolvimento da segunda versão.

5.3 VERSÃO 2.0

Esta seção descreve as modificações efetuadas para desenvolver a segunda versão do MineraFórum. Os experimentos realizados são apresentados. Os resultados obtidos são analisados e discutidos.

5.3.1 MODIFICAÇÕES NO SISTEMA

Na segunda versão, após efetuar a análise qualitativa das mensagens, o MineraFórum apresenta as seguintes informações:

- a) A quantidade total de mensagens postadas no fórum por cada aluno.
- b) A quantidade de contribuições relevantes ao tema feitas por cada aluno.
- c) A quantidade de contribuições não relevantes ao tema feitas por cada aluno.
- d) Os conceitos que foram utilizados nas postagens relevantes.
- e) A relevância temática de cada mensagem.
- f) Se a mensagem é similar (ou não) a alguma outra do fórum.

O desenvolvimento do MineraFórum foi continuado a partir dos recursos existentes na primeira versão. Todas as ações previstas na versão 1.0 do sistema foram mantidas. A ação para geração do relatório foi ampliada para permitir a indicação da similaridade entre mensagens do fórum. Foram acrescentadas duas novas ações:

- a) Exibir um grafo dos conceitos relevantes encontrados na mineração do texto do fórum.

O MineraFórum permite a visualização do grafo gerado com os conceitos relevantes diagnosticados no texto do fórum. O texto do fórum é composto por todas as mensagens, colocadas uma após a outra, na ordem em que foram postadas.

b) Gravar o resultado da mineração do fórum.

As informações obtidas com o resultado da mineração podem ser gravadas em um arquivo no formato “html”.

Os requisitos funcionais de 1 a 17 da primeira versão do sistema foram mantidos. O requisito 18 foi alterado. Foram adicionados os requisitos 19 e 20 (quadro 8).

Quadro 8 - Requisitos funcionais da versão 2.0 do MineraFórum

Requisito	Descrição
REQ.01 REQ.17	a Os requisitos são os mesmos da versão 1.0.
REQ.18	O MineraFórum deve gerar um relatório visual, contendo informações sobre as mensagens postadas pelos alunos no fórum. Para cada aluno, devem ser exibidos: o nome do discente, a quantidade total de mensagens postadas, a quantidade de mensagens relevantes ao tema da discussão, a quantidade de mensagens que não envolveram conceitos sobre o tópico do debate, os conceitos utilizados nas contribuições relevantes, a indicação se cada mensagem é similar a outra do fórum.
REQ.19	O MineraFórum deve permitir a exibição do grafo dos conceitos relevantes encontrados na mineração do texto do fórum.
REQ.20	O MineraFórum deve permitir a gravação do resultado da mineração das postagens do fórum em um arquivo no formato “html”.

Fonte: o autor

A partir da especificação dos requisitos funcionais, foi elaborado o diagrama dos casos de uso da versão 2.0 do MineraFórum (figura 28).

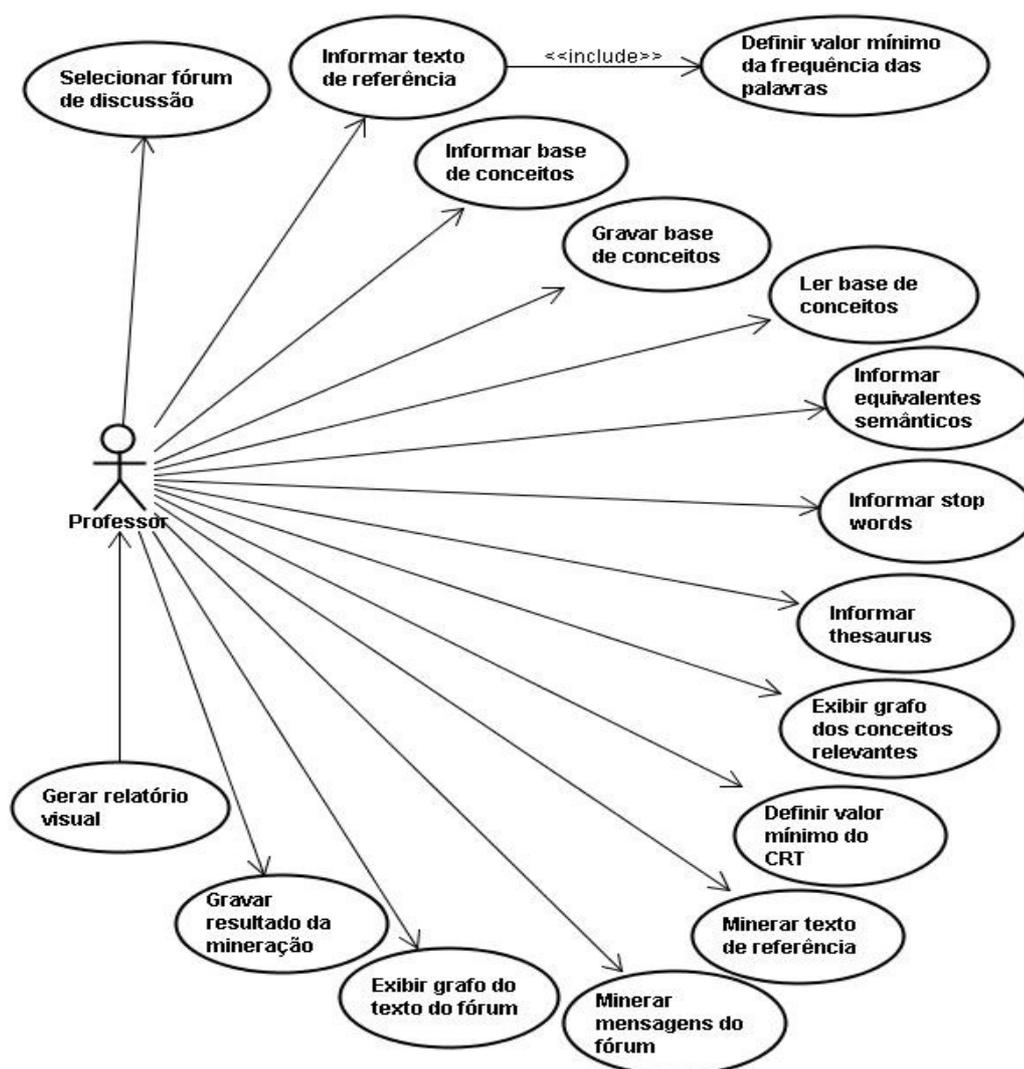


Figura 28 - Diagrama de Casos de Uso da versão 2.0 do MineraFórum
Fonte: o autor

O quadro 9 apresenta a descrição dos casos de uso citados no diagrama da figura 28. Os CDUs de 1 a 3 existentes na primeira versão, foram mantidos. O CDU14 foi alterado. Foram acrescentados os CDUs 15 e 16.

Quadro 9 - Descrição dos casos de uso da versão 2.0 do MineraFórum

Caso de Uso (CDU)	Descrição
CDU01 a CDU13	Igual a versão 1.0.
CDU14 - Gerar relatório visual	Exibe um relatório contendo informações sobre as mensagens postadas pelos alunos. Para cada aluno, devem ser exibidos: o nome do discente, o total de postagens, a quantidade de mensagens relevantes ao tema, a quantidade de mensagens que não envolveram conceitos sobre o tópico, os conceitos utilizados nas contribuições relevantes, a indicação se cada mensagem é similar a outra do fórum.
CDU15 - Exibir grafo do texto do fórum	Permite a exibição do grafo dos conceitos relevantes encontrados na mineração do texto do fórum.
CDU16 - Gravar resultado da mineração	Permite a gravação do resultado da mineração do fórum em um arquivo no formato “html”.

Fonte: o autor

A figura 29 apresenta o diagrama das classes que constituem a versão 2.0 do MineraFórum. Todas as classes da primeira versão foram mantidas. A única alteração foi a adição do atributo “similar” na classe “Mensagem”. Este atributo representa a identificação se uma mensagem é similar, ou não, a outra do fórum.

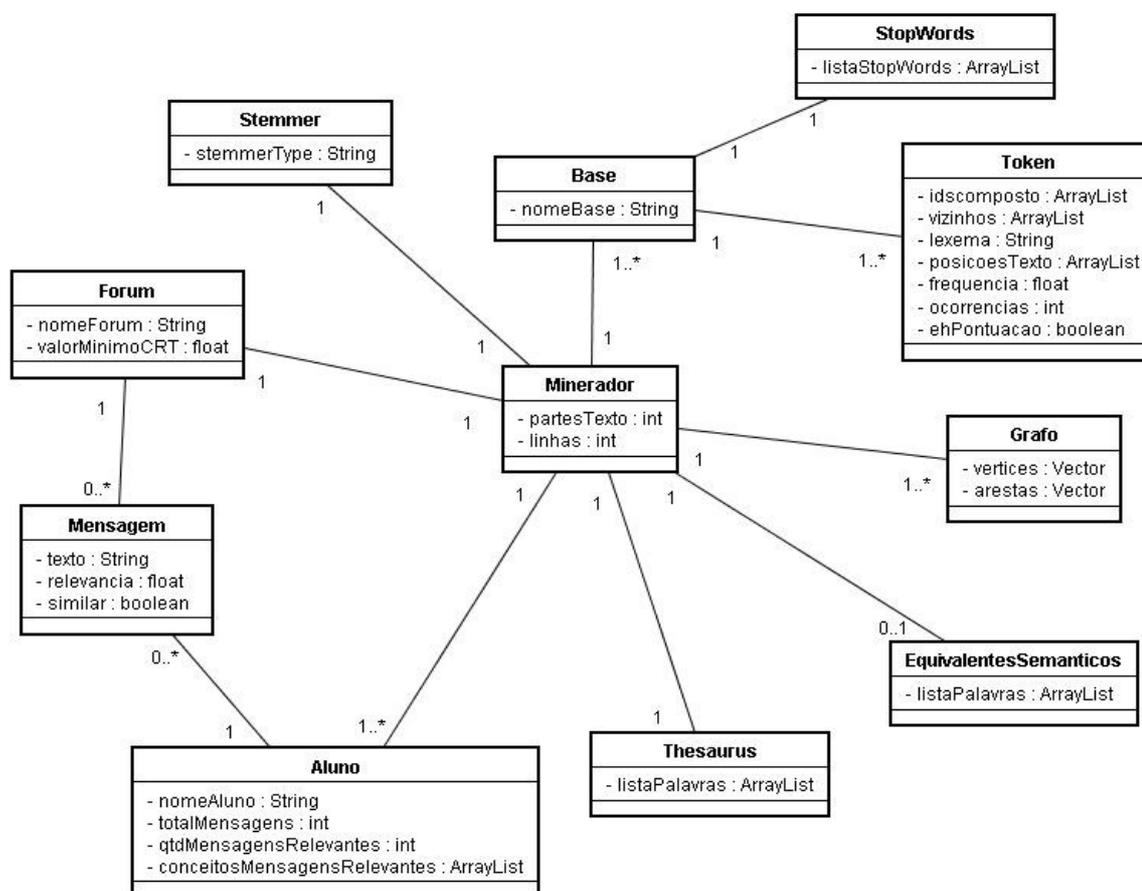


Figura 29 - Diagrama de Classes da versão 2.0 do MineraFórum

Fonte: o autor

Com relação a interface do sistema, as telas apresentadas nas figuras 19 a 21 foram mantidas. Os recursos disponíveis na figura da tela 22 foram mantidos. Foi acrescentada a opção “Salvar resultado da mineração” para permitir a gravação do resultado da mineração das mensagens do fórum (figura 30).

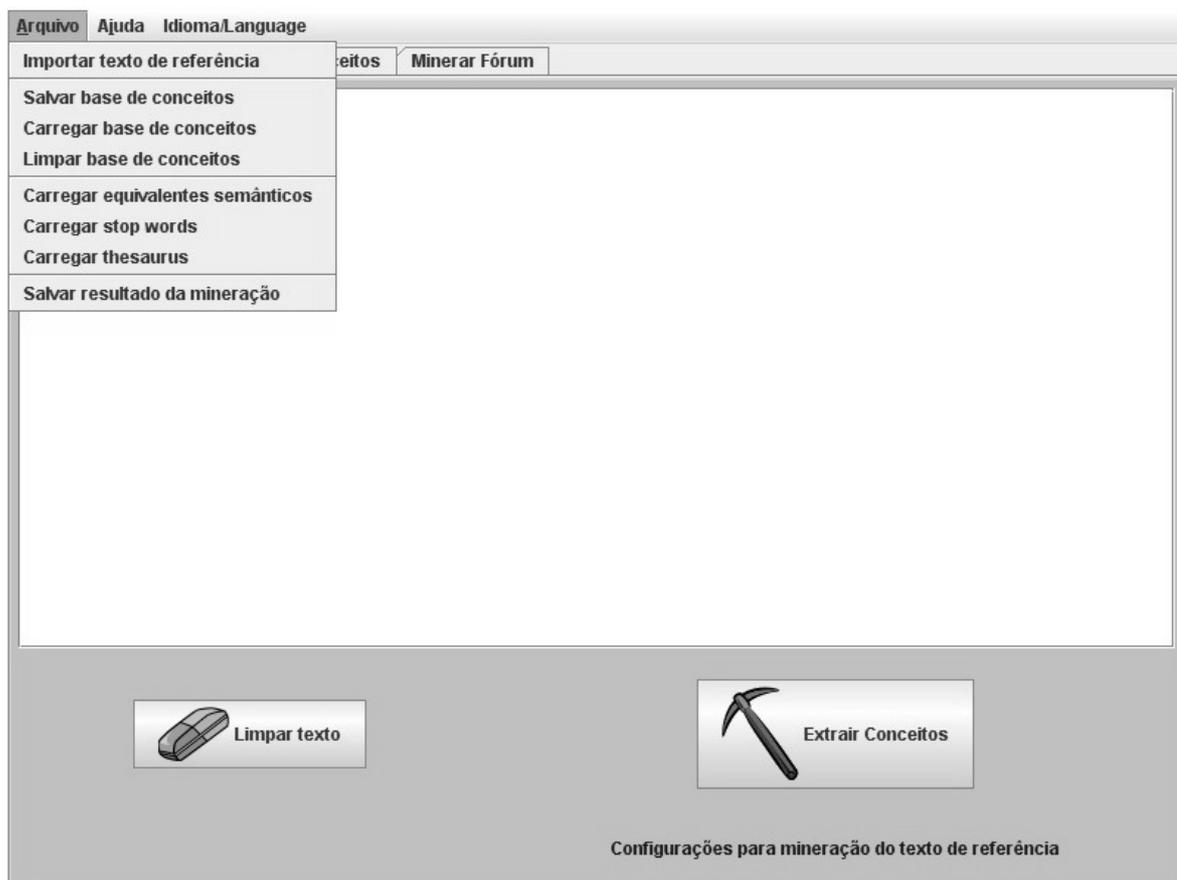


Figura 30 - Tela principal da versão 2.0 com o menu “Arquivo” selecionado
Fonte: o autor

Os recursos disponíveis na figura da tela 23 permanecem na versão 2.0. Foram adicionados os botões abaixo (figura 31):

- “Gerar grafo do fórum”: para exibir o grafo dos conceitos relevantes encontrados na mineração do texto do fórum.
- “Similaridades”: para apresentar todas as mensagens do fórum que são similares.

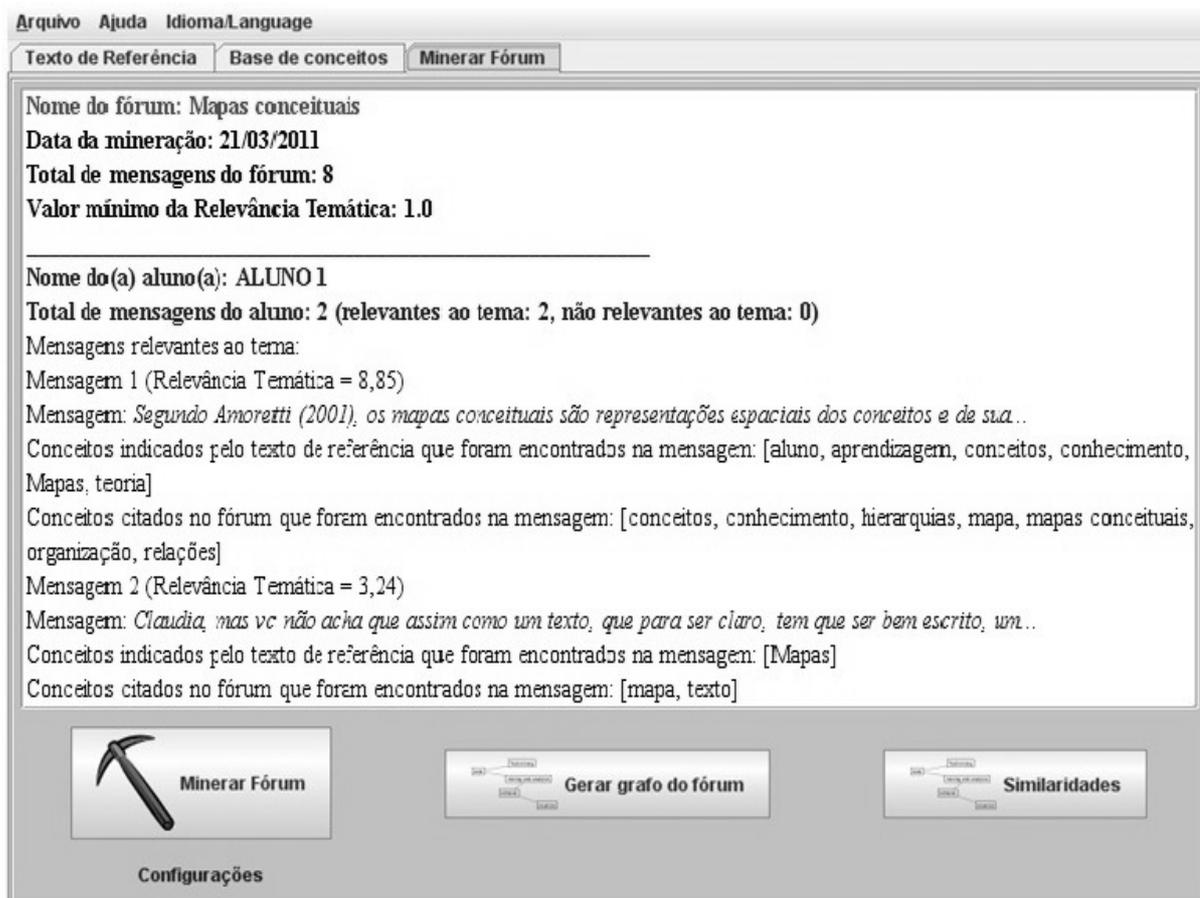


Figura 31 - Tela principal da versão 2.0 com a aba “Minerar Fórum” selecionada
 Fonte: o autor

Na tela da figura 24, todos os recursos foram mantidos. Apenas a frase “Informe o valor mínimo do CRT” foi alterada para “Informe o valor mínimo da Relevância Temática” (figura 32).



Figura 32 - Tela da versão 2.0 para informar o valor mínimo da relevância temática
Fonte: o autor

Observando a fórmula para o cálculo da relevância temática (RT) de uma mensagem na versão 2.0 (explicada na seção 4.5.1) e considerando que:

- G_α é o grafo do texto de referência com T_α vértices;
- G_β é o grafo de uma mensagem com T_β vértices;
- G_Ω é o grafo do texto do fórum com T_Ω vértices;
- $\text{equiv}(V_a, V_b)$ verifica se os vértices V_a e V_b são equivalentes, indicando o resultado 1 caso eles sejam, e indicando 0 caso contrário;
- $\text{menordist}(G, V_a, V_b)$ descobre a menor distância entre os vértices V_a e V_b no grafo G ;
- $\text{totalvizinhos}(G, V_a)$ descobre a quantidade de vizinhos do vértice V_a no grafo G ;
- dois vértices são vizinhos se existe uma aresta que os une.

O algoritmo para o cálculo da RT foi elaborado da seguinte forma:

1. $NC = 0; DC = 0; PC = 0; \text{Somadist} = 0; \text{Somapesos} = 0$
2. Para todo vértice V_i de G_α
 3. Para todo vértice V_j de G_β
 4. $NC = NC + \text{equiv}(V_i, V_j)$
 5. Se $\text{equiv}(V_i, V_j) = 1$ Então
 6. $\text{Somadist} = \text{Somadist} + (\text{menordist}(\alpha, V_i, V_j) - \text{menordist}(\beta, V_i, V_j))$
 7. $\text{Somapesos} = \text{Somapesos} + ((\text{totalvizinhos}(\alpha, V_i) / T_\alpha) + (\text{totalvizinhos}(\beta, V_i) / T_\beta))$
 8. FimSe
 9. Retorne ao passo 3
10. Retorne ao passo 2
11. $DC = 1 / 1 + (\text{Somadist})$
12. $PC = 1 / 1 + (\text{Somapesos})$
13. $RT_1 = NC + DC + PC$
14. $NC = 0; DC = 0; PC = 0; \text{Somadist} = 0; \text{Somapesos} = 0$
15. Para todo vértice V_i de G_Ω
 16. Para todo vértice V_j de G_β
 17. $NC = NC + \text{equiv}(V_i, V_j)$
 18. Se $\text{equiv}(V_i, V_j) = 1$ Então
 19. $\text{Somadist} = \text{Somadist} + (\text{menordist}(\Omega, V_i, V_j) - \text{menordist}(\beta, V_i, V_j))$
 20. $\text{Somapesos} = \text{Somapesos} + ((\text{totalvizinhos}(\Omega, V_i) / T_\Omega) + (\text{totalvizinhos}(\beta, V_i) / T_\beta))$
 21. FimSe
 22. Retorne ao passo 16
23. Retorne ao passo 15
24. $DC = 1 / 1 + (\text{Somadist})$

$$25. PC = 1 / 1 + (\text{Somapesos})$$

$$26. RT_2 = NC + DC + PC$$

$$27. RT = (RT_1 + RT_2) / 2$$

5.3.2 EXPERIMENTOS COM A VERSÃO 2.0

Esta seção relata os experimentos efetuados com a segunda versão do MineraFórum. Cabe ressaltar alguns pontos importantes:

- As postagens foram analisadas pelo MineraFórum de acordo com a ordem de inclusão nos fóruns.
- O objetivo dos experimentos foi investigar a relevância temática das mensagens.
- Os experimentos foram realizados com fóruns de discussão do Rooda.
- O termo de consentimento para a análise das postagens é apresentado no Apêndice C.
- Os nomes dos discentes, docentes e tutores foram preservados.

Com a versão 2.0 do MineraFórum foram realizados cinco experimentos, com fóruns distintos. Em todas as experiências, foi definido o valor mínimo 01 para a relevância temática (RT) a ser considerada na análise dos textos. Isto significa que após o cálculo da relevância temática de cada mensagem, aquelas cuja RT foi maior ou igual a 01 foram consideradas como relevantes ao tema do debate.

Os três primeiros experimentos foram realizados em fóruns de discussão da disciplina “Tópicos Especiais Z1” da turma do Doutorado em Informática na Educação, que ocorreu no primeiro semestre de 2008. O tema do primeiro fórum foi “Mapas conceituais”, que apresentou 16 contribuições textuais. No segundo fórum, o tema foi “TICs e Educação”, com 13 contribuições textuais. O tema do terceiro fórum foi “Comunidades virtuais”, que apresentou 26 contribuições textuais.

O quarto e quinto experimentos utilizaram fóruns da disciplina “Seminário Integrador VII – B” de uma turma do Curso de Graduação em Pedagogia, que ocorreu no segundo semestre de 2009. O tema do fórum foi “Aprender com os outros”. A turma foi dividida em dois grupos, e cada grupo contribuiu em um fórum de discussão. O primeiro fórum contou

com a participação de 27 alunos, que postaram 42 mensagens. O segundo fórum contou com a participação de 28 alunos, que realizaram 67 contribuições textuais.

Nos três primeiros experimentos, o professor informou ao MineraFórum um texto de referência sobre o assunto “Mapas conceituais”. Na quarta e quinta experiência, o texto indicado pelo docente referiu-se ao tema “Aprender com os outros”.

Os resultados das experiências são apresentados nas figuras 33 a 37. Os gráficos indicam o valor da relevância temática das mensagens postadas por cada aluno.

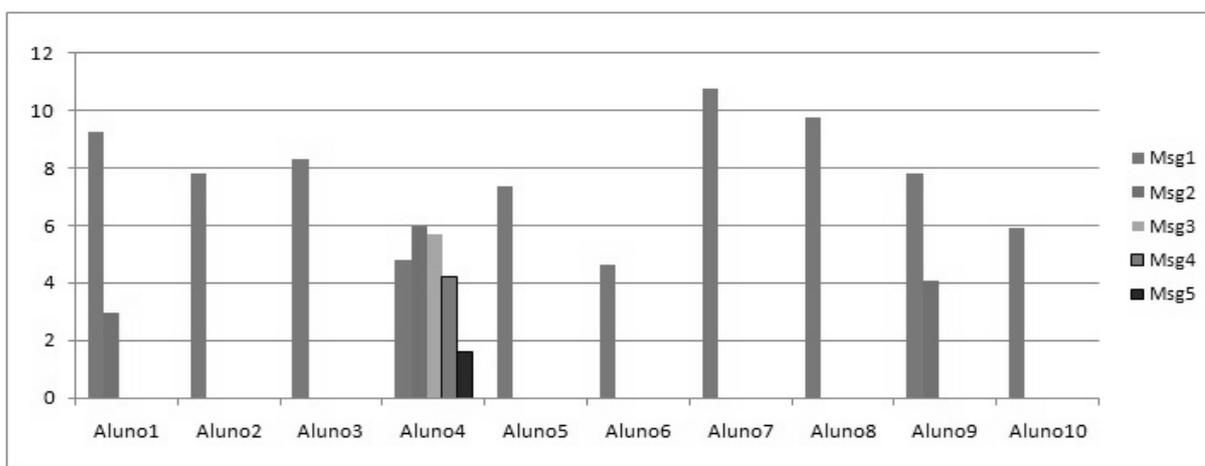


Figura 33 - Relevância temática das mensagens postadas pelos alunos no primeiro fórum analisado
Fonte: o autor

Pode-se observar na figura 33 que o Aluno7 redigiu apenas uma mensagem, porém com um alto valor de relevância temática. Verifica-se que o Aluno4 redigiu cinco mensagens. No entanto, a última mensagem possui um baixo valor de relevância.

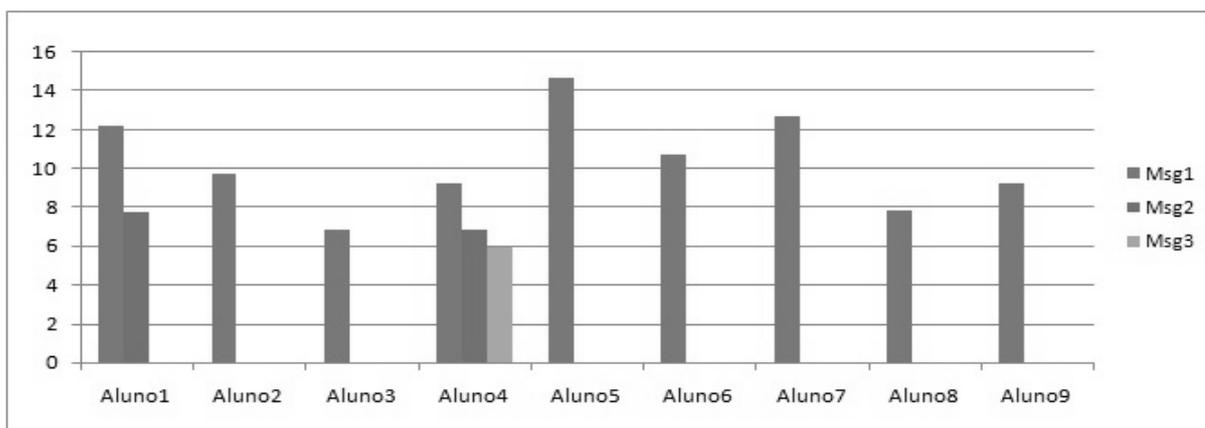


Figura 34 - Relevância temática das mensagens postadas pelos alunos no segundo fórum analisado
Fonte: o autor

Na figura 34, observa-se que o Aluno1, Aluno5 e Aluno7 foram os que redigiram

mensagens com maior relevância temática. Neste fórum, verifica-se que todas as mensagens escritas pelos alunos foram relevantes ao tema do debate.

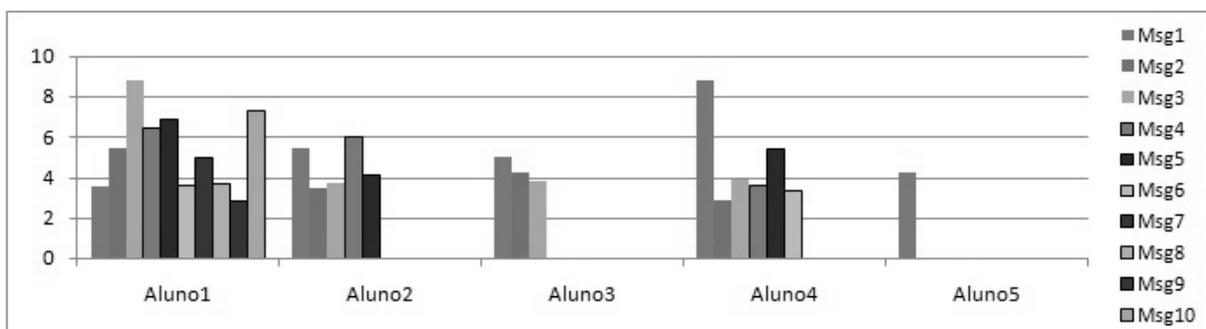


Figura 35 - Relevância temática das mensagens postadas pelos alunos no terceiro fórum analisado
Fonte: o autor

Na figura 35 observa-se que o Aluno1 participou bastante do fórum, redigindo várias mensagens relevantes ao tema do debate. No entanto, o Aluno5 realizou apenas uma postagem.

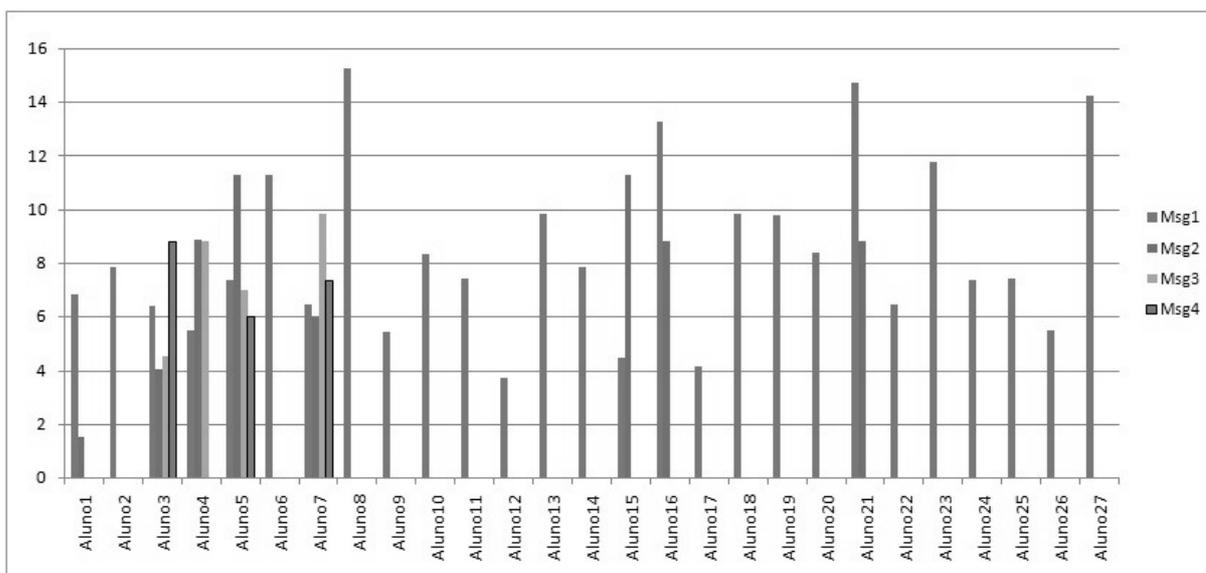


Figura 36 - Relevância temática das mensagens postadas pelos alunos no quarto fórum analisado
Fonte: o autor

Na figura 36, observa-se que o Aluno3, Aluno4, Aluno5 e Aluno7 foram os que mais participaram do fórum, redigindo mensagens relevantes. Verifica-se que Aluno9, Aluno12, Aluno17, Aluno26 postaram apenas uma mensagem com valor de relevância entre 3.0 e 5.0.

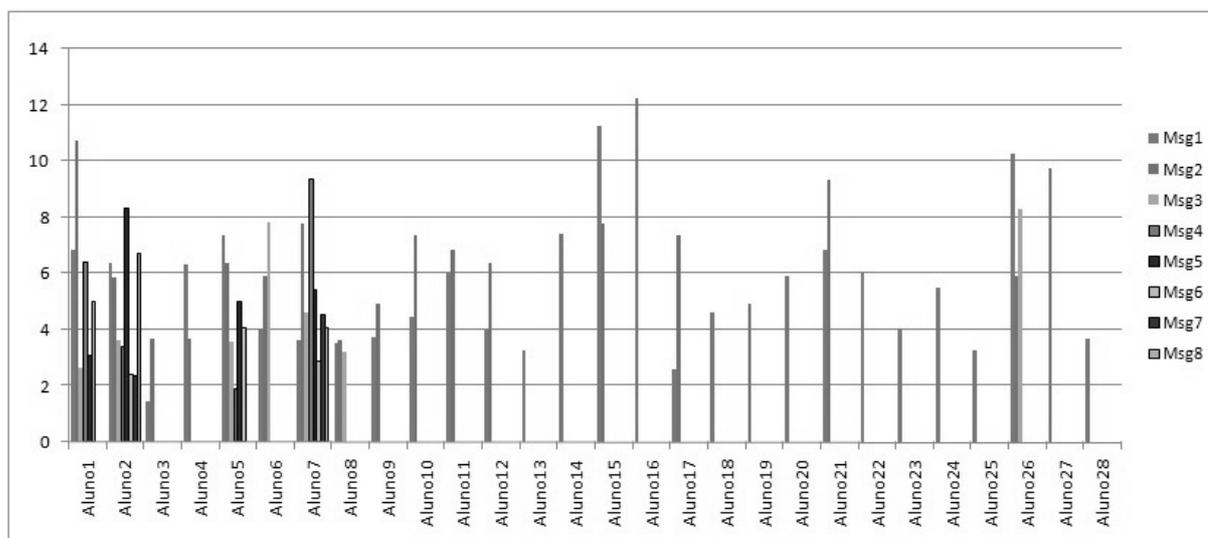


Figura 37 - Relevância temática das mensagens postadas pelos alunos no quarto fórum analisado
Fonte: o autor

Na figura 37, observa-se que o Aluno1, Aluno2, Aluno5 e Aluno7 foram os que mais participaram do fórum, redigindo mensagens relevantes. Verifica-se que o Aluno13, Aluno23, Aluno25, Aluno28 postaram apenas uma mensagem com valor de relevância entre 3.0 e 4.0.

5.3.3 ANÁLISE E DISCUSSÃO DOS RESULTADOS DA VERSÃO 2.0

Neste momento cabe relembrar a metodologia para o cálculo da relevância de cada postagem. Na segunda versão a relevância é diagnosticada a partir da análise temática da mensagem.

Com os resultados descritos nas figuras 33 a 37, verificou-se que a versão 2.0 do MineraFórum calculou a relevância temática das postagens. Foi possível observar quais alunos redigiram mais contribuições textuais relativas ao tópico da discussão, e quais escreveram poucas mensagens relevantes. Além disso, identificaram-se quais foram os discentes que postaram as mensagens com maior relevância ao tema em debate.

A análise da relevância temática das mensagens auxilia o professor a realizar um diagnóstico sobre os alunos. Nos gráficos das figuras 33 a 37, foi possível observar que alguns discentes escreveram várias mensagens com relevância temática mediana. Por outro lado, alguns alunos redigiram somente uma ou duas contribuições textuais com alta relevância.

Os resultados dos experimentos possibilitaram a observação de aspectos que foram

aperfeiçoados na versão 3.0, como:

- a) A fórmula da relevância temática foi alterada:
 - Verificou-se que é importante calcular a proporção dos conceitos relevantes com relação ao total de palavras da mensagem. Desta forma, o cálculo do termo NC foi modificado.
 - Identificou-se a necessidade de estabelecer um limite máximo para a relevância temática, para que fosse possível comparar os resultados de cada postagem. Definiu-se como limite máximo o valor um (1).
- b) Verificou-se que, além da importância temática, outros critérios devem ser considerados para analisar uma postagem. Um questionário foi aplicado com professores (seção 4.4) para mensurar o peso de indicadores de relevância. Com isto, a fórmula para o cálculo da relevância de uma postagem foi aprimorada. Além de considerar a relevância temática, também foi calculada a proporção de citações da mensagem e a similaridade da mensagem com outra do fórum.
- c) Para facilitar a visualização do resultado da relevância das postagens, o valor calculado foi mapeado para uma escala de valores inteiros entre zero (0) e cinco (5).

A próxima seção descreve as alterações ocorridas na versão 2.0 do sistema para o desenvolvimento da terceira versão.

5.4 VERSÃO 3.0

Esta seção descreve o desenvolvimento da terceira e última versão do MineraFórum. Os experimentos efetuados e seus resultados são relatados e discutidos.

5.4.1 DESCRIÇÃO DAS ALTERAÇÕES

Na terceira versão, após realizar a análise qualitativa das mensagens, o MineraFórum apresenta as seguintes informações:

- a) A quantidade total de mensagens postadas no fórum por cada aluno.
- b) A quantidade de contribuições relevantes feitas por cada aluno.
- c) A quantidade de contribuições não relevantes feitas por cada aluno.
- d) Os conceitos que foram utilizados nas postagens relevantes.
- e) A relevância de cada mensagem.
- f) Se a mensagem é similar (ou não) a alguma outra do fórum.
- g) A média das relevâncias das mensagens de cada aluno.
- h) A quantidade de vezes que cada mensagem foi citada dentro do fórum.

O desenvolvimento do MineraFórum foi continuado a partir da segunda versão. As ações previstas na versão 2.0 do sistema foram mantidas, com exceção da ação para definir o valor mínimo da relevância temática, que foi excluída. Esta ação foi removida na versão 3.0, pois o sistema calcula a relevância da mensagem dentro do fórum com outra fórmula (seção 4.5.1).

Cabe ressaltar que a ação para geração do relatório foi ampliada. O relatório da terceira versão também exibe a média das relevâncias das postagens de cada aluno e a quantidade de vezes que cada mensagem foi citada dentro do fórum. Foi adicionada uma nova ação na versão 3.0 para exibir um gráfico com a média das relevâncias das mensagens de cada aluno.

Os requisitos funcionais de 1 a 13, 16, 17, 19, 20, da segunda versão do sistema foram mantidos. O requisito 14 foi excluído, o 15 foi modificado e o 18 foi ampliado. Foi adicionado o requisito 21 (quadro 10).

Quadro 10 - Requisitos funcionais da versão 3.0 do MineraFórum

Requisito	Descrição
REQ.01 a	Igual a versão 2.0.
REQ.13	
REQ.14	Excluído.
REQ.15	O MineraFórum deve calcular a relevância de cada mensagem.
REQ.16	Igual a versão 2.0.
REQ.17	Igual a versão 2.0.
REQ.18	O MineraFórum deve gerar um relatório visual, contendo informações sobre as mensagens postadas pelos alunos no fórum. Para cada aluno, devem ser exibidos: o nome do discente, a quantidade total de mensagens postadas, a quantidade de mensagens relevantes ao tema da discussão, a quantidade de mensagens que não envolveram conceitos sobre o tópico do debate, os conceitos utilizados nas contribuições relevantes, a indicação se cada mensagem é similar a outra do fórum, a média das relevâncias das postagens, a quantidade de vezes que cada mensagem foi citada dentro do fórum.
REQ.19	Igual a versão 2.0.
REQ.20	Igual a versão 2.0.
REQ.21	O MineraFórum deve permitir a exibição de um gráfico com a média das relevâncias das mensagens de cada aluno.

Fonte: o autor

A partir da especificação dos requisitos funcionais, foi elaborado o diagrama dos casos de uso da versão 3.0 do MineraFórum (figura 38).

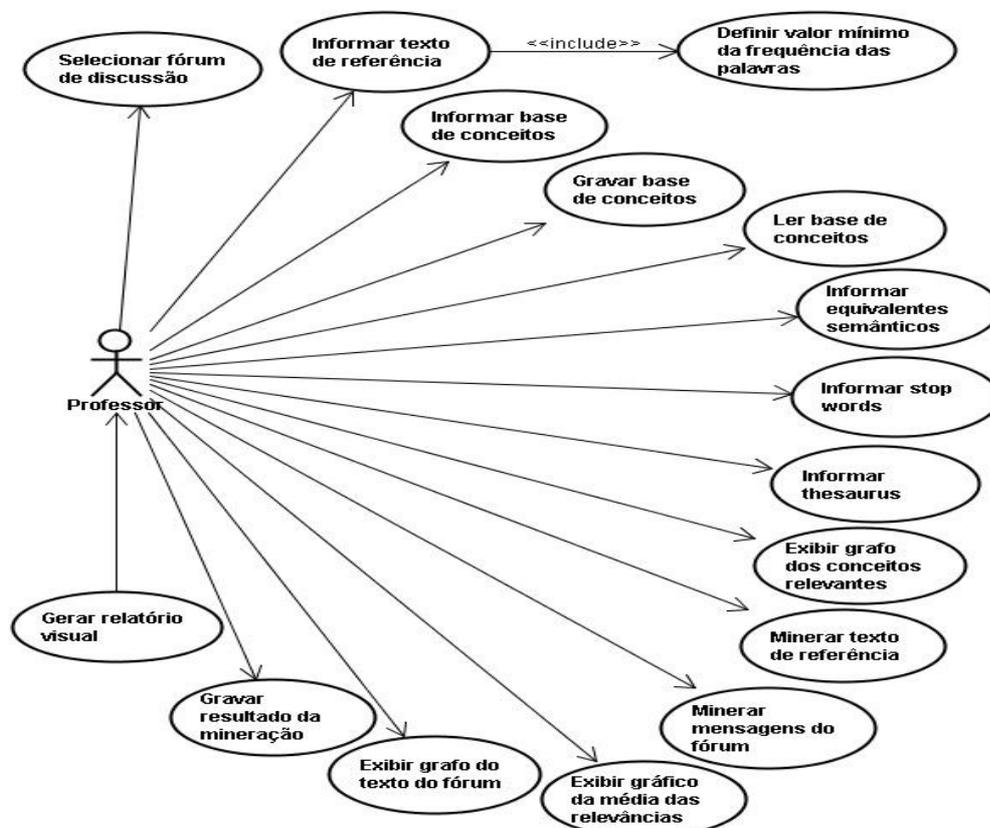


Figura 38 - Diagrama de Casos de Uso da versão 3.0 do MineraFórum

Fonte: o autor

O quadro 11 apresenta a descrição dos casos de uso citados no diagrama da figura 38. Os CDUs de 1 a 10, 12, 13, 15, 16, existentes na segunda versão, foram mantidos. O CDU11 foi excluído e o CDU14 foi alterado. Foi acrescentado o CDU 17.

Quadro 11 - Descrição dos casos de uso da versão 3.0 do MineraFórum

Caso de Uso (CDU)	Descrição
CDU01 a CDU10	Igual a versão 2.0.
CDU11 - Definir valor mínimo do CRT	Excluído.
CDU12 - Minerar texto de referência	Igual a versão 2.0.
CDU13 - Minerar mensagens do fórum	Igual a versão 2.0.
CDU14 - Gerar relatório visual	Exibe um relatório contendo informações sobre as mensagens postadas pelos alunos. Para cada aluno, devem ser exibidos: o nome do discente, o total de postagens, a quantidade de mensagens relevantes ao tema, a quantidade de mensagens que não envolveram conceitos sobre o tópico, os conceitos utilizados nas contribuições relevantes, a indicação se cada mensagem é similar a outra do fórum, a média das relevâncias das postagens, a quantidade de vezes que cada mensagem foi citada dentro do fórum.
CDU15 - Exibir grafo do texto do fórum	Igual a versão 2.0.
CDU16 - Gravar resultado da mineração	Igual a versão 2.0.
CDU17 - Exibir gráfico da média das relevâncias	Exibe um gráfico com a média das relevâncias das mensagens de cada aluno.

Fonte: o autor

A figura 39 apresenta o diagrama das classes que constituem a versão 3.0 do MineraFórum. As classes da segunda versão foram mantidas. As alterações ocorridas foram:

- Adição do atributo “qtdCitacoes” na classe “Mensagem”. Este atributo representa a quantidade de vezes que a mensagem foi citada dentro do fórum.
- O atributo “valorMinimoCRT” foi excluído da classe “Forum”.

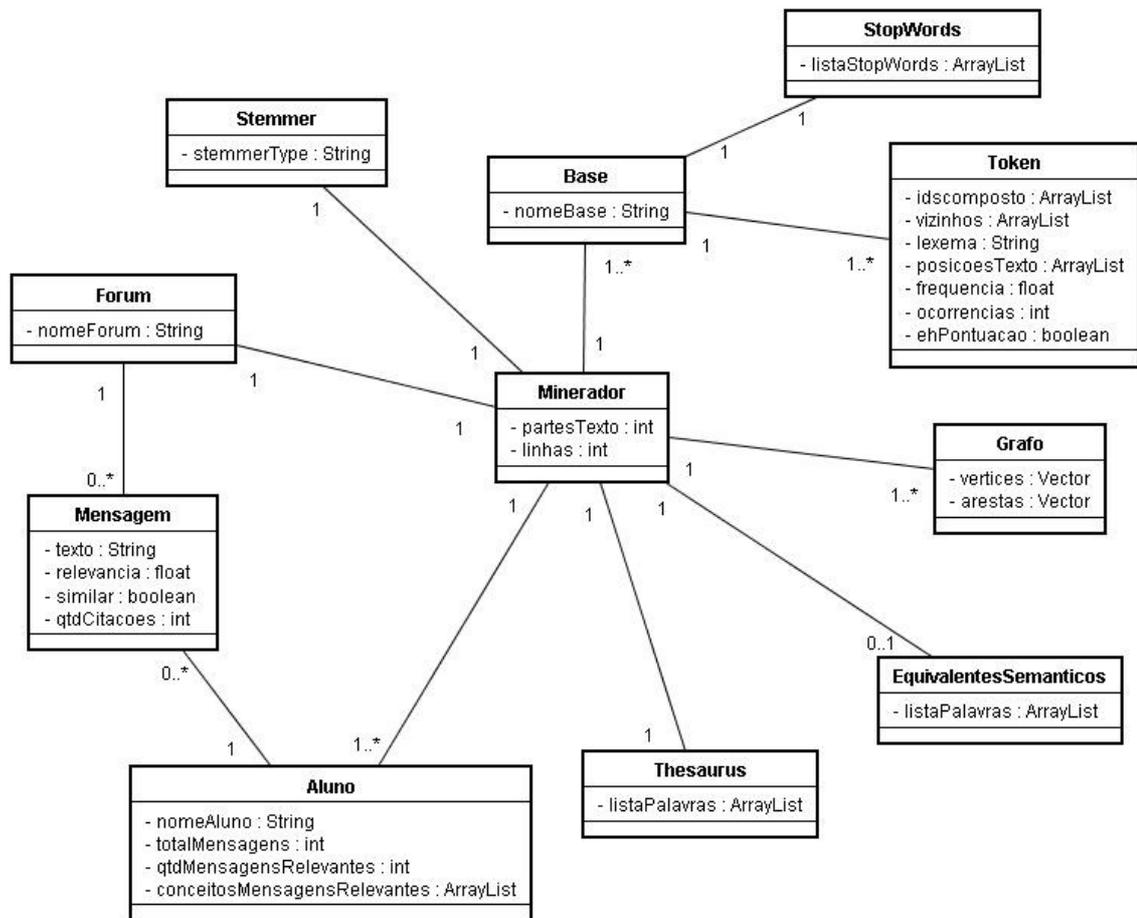


Figura 39 - Diagrama de Classes da versão 3.0 do MineraFórum
Fonte: o autor

Com relação a interface do sistema, as telas apresentadas nas figuras 19 a 21 foram mantidas. A tela da figura 30 foi alterada. Foi retirada a opção “Salvar resultado da mineração” (figura 40).

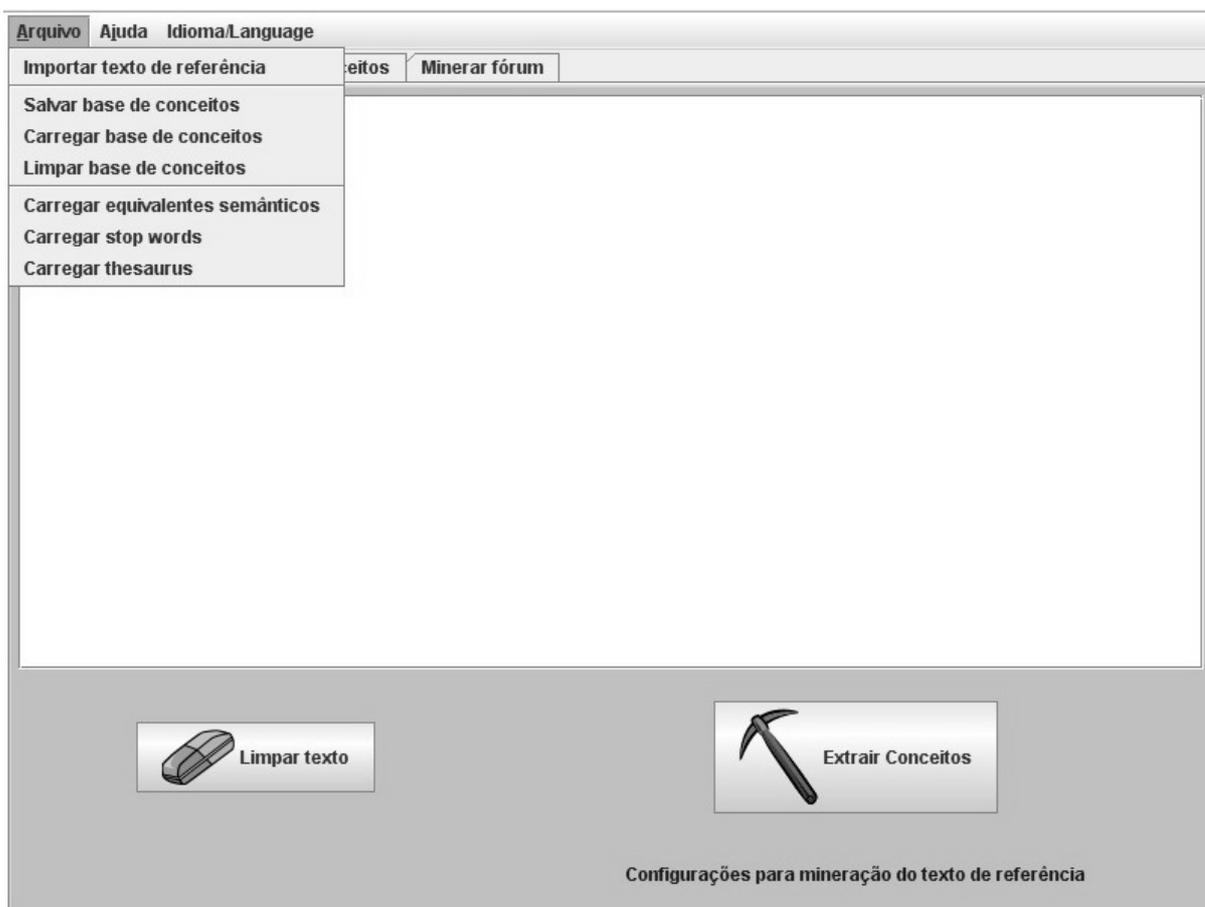


Figura 40 - Tela principal da versão 3.0 com o menu “Arquivo” selecionado
 Fonte: o autor

A tela da figura 31 foi alterada. Os botões disponíveis na versão 3.0 são (figura 41):

- “Minerar fórum”: executa a mineração das postagens do fórum.
- “Grafo do fórum”: exhibe o grafo dos conceitos relevantes encontrados na mineração do texto do fórum.
- “Ver similaridades”: apresenta todas as mensagens do fórum que são similares.
- “Exibir conceitos”: exhibe os conceitos relevantes mencionados nas postagens.
- “Exibir gráfico”: exhibe o gráfico com a média das relevâncias das mensagens de cada aluno.
- “Salvar resultados”: grava os resultados da mineração das postagens em um arquivo no formato “html”.

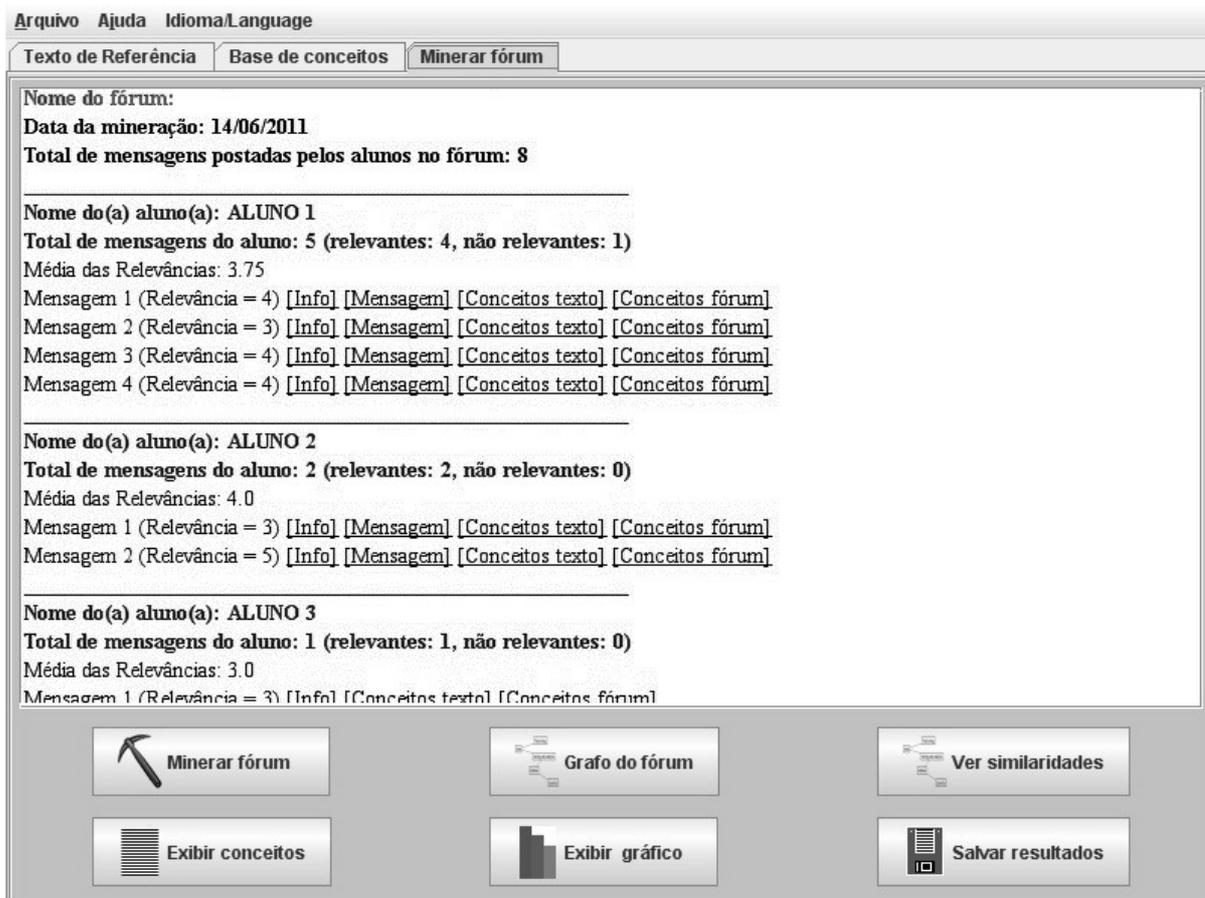


Figura 41 - Tela principal da versão 3.0 com a aba “Minerar Fórum” selecionada

Fonte: o autor

Cabe destacar que a tela da figura 32 foi excluída, pois não há definição do valor mínimo para a relevância temática das mensagens.

Um tutorial completo da versão 3.0, com a explicação detalhada dos recursos disponíveis, é apresentado no Apêndice D. O Apêndice E relata os procedimentos para efetuar a instalação do MineraFórum.

Observando a fórmula para o cálculo da relevância de uma mensagem (RF) na versão 3.0 (explicada na seção 4.5.1) e considerando que:

- G_α é o grafo do texto de referência com T_α vértices;
- G_β é o grafo de uma mensagem com T_β vértices;
- G_Ω é o grafo do texto do fórum com T_Ω vértices;
- M é a quantidade total de mensagens do fórum;
- $\text{equiv}(V_a, V_b)$ verifica se os vértices V_a e V_b são equivalentes, indicando o

resultado 1 caso eles sejam, e indicando 0 caso contrário;

- $\text{menordist}(G, V_a, V_b)$ descobre a menor distância entre os vértices V_a e V_b no grafo G ;
- $\text{totalvizinhos}(G, V_a)$ descobre a quantidade de vizinhos do vértice V_a no grafo G ;
- dois vértices são vizinhos se existe uma aresta que os une;
- $\text{ocorrencias}(V_a)$ descobre a quantidade de ocorrências do conceito correspondente ao vértice V_a ;
- $\text{qtdcitacoes}(m)$ descobre a quantidade de citações que a mensagem m obteve no fórum;
- $\text{similar}(m)$ descobre se a mensagem m é similar a outra do fórum.

O algoritmo para o cálculo da RF foi elaborado da seguinte forma:

1. $RC = 0$; $DC = 0$; $PC = 0$; $\text{Somadist} = 0$; $\text{Somapesos} = 0$; $\text{Somafreqconc} = 0$;
 $\text{Somafreqtodosconc} = 0$
2. Para todo vértice V_i de G_α
3. Para todo vértice V_j de G_β
 4. $\text{Somafreqtodosconc} = \text{Somafreqtodosconc} + \text{ocorrencias}(V_j)$
 5. Se $\text{equiv}(V_i, V_j) = 1$ Então
 6. $\text{Somafreqconc} = \text{Somafreqconc} + \text{ocorrencias}(V_j)$
 7. $\text{Somadist} = \text{Somadist} + (\text{menordist}(\alpha, V_i, V_j) - \text{menordist}(\beta, V_i, V_j))$
 8. $\text{Somapesos} = \text{Somapesos} + ((\text{totalvizinhos}(\alpha, V_i) / T_\alpha) + (\text{totalvizinhos}(\beta, V_i) / T_\beta))$
 9. FimSe
10. Retorne ao passo 3
11. Retorne ao passo 2
12. $RC = \text{Somafreqconc} / \text{Somafreqtodosconc}$
13. $DC = 1 / 1 + (\text{Somadist})$
14. $PC = 1 / 1 + (\text{Somapesos})$

15. $RT_1 = (RC + (DC * (1 - RC))) + (PC * (1 - (RC + (DC * (1 - RC)))))$
16. $RC = 0$; $DC = 0$; $PC = 0$; $Somadist = 0$; $Somapesos = 0$; $Somafreqconc = 0$;
 $Somafreqtodosconc = 0$
17. Para todo vértice V_i de G_Ω
 18. Para todo vértice V_j de G_β
 19. $Somafreqtodosconc = Somafreqtodosconc + ocorrencias(V_j)$
 20. Se $equiv(V_i, V_j) = 1$ Então
 21. $Somafreqconc = Somafreqconc + ocorrencias(V_j)$
 22. $Somadist = Somadist + (menordist(\Omega, V_i, V_j) - menordist(\beta, V_i, V_j))$
 23. $Somapesos = Somapesos + ((totalvizinhos(\Omega, V_i) / T_\Omega) + (totalvizinhos(\beta, V_i) / T_\beta))$
 24. FimSe
 25. Retorne ao passo 18
26. Retorne ao passo 17
27. $RC = Somafreqconc / Somafreqtodosconc$
28. $DC = 1 / 1 + (Somadist)$
29. $PC = 1 / 1 + (Somapesos)$
30. $RT_2 = (RC + (DC * (1 - RC))) + (PC * (1 - (RC + (DC * (1 - RC)))))$
31. $RT = (RT_1 + RT_2) / 2$
32. $RM = qtdcitacoes(m) / M$
33. $SM = 0$
34. Se $similar(m)$ Então
 35. $SM = RT$
36. FimSe
37. $P1 = 9.61$; $P2 = 9.11$; $P3 = 7.47$
38. $RF = ((P1 * RT) + (P2 * RM) - (P3 * SM)) / (P1 + P2 - P3)$

5.4.2 EXPERIMENTOS COM A VERSÃO 3.0

Esta seção destaca os experimentos realizados com a terceira versão do MineraFórum. Cabe enfatizar alguns pontos importantes:

- As postagens foram analisadas pelo MineraFórum de acordo com a ordem de inclusão nos fóruns.
- O objetivo dos experimentos foi calcular a relevância das postagens e comparar os resultados obtidos com a análise executada pelos professores.
- Nos experimentos foram utilizados os fóruns do Rooda, ETC, Moodle.
- O termo de consentimento para a análise das mensagens dos fóruns encontra-se no Apêndice C.
- O sigilo dos nomes das pessoas foi preservado.

Com a versão 3.0 do MineraFórum foram realizados vinte (20) experimentos, com fóruns distintos. Cada experiência foi conduzida da seguinte forma:

- Um fórum de discussão foi selecionado.
- O MineraFórum calculou a relevância de cada postagem do fórum.
- Solicitou-se que dois professores realizassem a análise das postagens do fórum, atribuindo um valor inteiro entre zero (0) e cinco (5). O valor zero (0) indica que a mensagem não é relevante para o debate, enquanto o valor cinco (5) significa que a postagem possui alta relevância para a discussão.
- Foi calculada a média das relevâncias (e o desvio padrão) das postagens. O cálculo foi efetuado com os valores indicados pelo MineraFórum e pelos dois professores.
- Calculou-se o grau de similaridade entre a média das relevâncias indicadas pelo MineraFórum com a média das relevâncias indicadas pelos dois professores.

Dez (10) professores realizaram a análise das postagens dos fóruns. Estes docentes atuam no nível superior, na graduação ou na pós-graduação. Eles já utilizaram fóruns de discussão com seus alunos, em disciplinas que lecionaram. Entre os professores, seis (06) ministram aula no IFF. Quatro (04) são doutorandos em Informática na Educação, que já empregaram fórum com seus discentes.

O quadro 12 apresenta os nomes e temas dos fóruns utilizados nos experimentos. O AVA no qual cada fórum foi concebido também é relatado. O quadro 13 descreve o nível de ensino, a modalidade e o curso dos alunos que participaram dos fóruns.

Quadro 12 - Nomes e temas dos fóruns utilizados nos experimentos da versão 3.0

Fórum	AVA	Nome do fórum	Tema do fórum
1	Rooda	Temática 2 - parte 2	Mapas conceituais
2	Rooda	Temática 2 - parte 1	TICs na Educação
3	Rooda	Discussão sobre Comunidades Virtuais de Aprendizagem	Comunidades Virtuais de Aprendizagem
4	Rooda	Seminário Integrador - alunos de A a L	Aprender com os outros
5	Rooda	Seminário Integrador - alunos de L a Z	Aprender com os outros
6	Rooda	Seminário Integrador - alunos de A a L	Aprender como transformação
7	Rooda	Seminário Integrador - alunos de L a Z	Aprender como transformação
8	ETC	Trabalho em Equipe como, por quê e para quê ou para quem	Trabalho em equipe
9	ETC	Desenvolvimento de competências	Desenvolvimento de competências
10	Moodle	Derivadas - Origens Históricas	Derivadas - Origens Históricas
11	Moodle	Teste da Derivada Segunda para Máximos e Mínimos Relativos	Teste da Derivada Segunda para Máximos e Mínimos Relativos
12	Moodle	Integrais - Origens Históricas	Integrais - Origens Históricas
13	Moodle	Certificação Digital	Certificação Digital
14	Moodle	Certificação Digital	Certificação Digital
15	Moodle	Política de Segurança da Informação	Política de Segurança da Informação
16	Moodle	Aparente verdade	Aparente verdade
17	Moodle	Atividade 03 - 1º Fórum de Perguntas e Respostas - Polo Açú	Educação a distância
18	Moodle	Atividade 04 - 2º Fórum de Perguntas e Respostas - Polo Açú	EAD e educação aberta
19	Moodle	Atividade 03 - 1º Fórum de Perguntas e Respostas - Polo SJB	Educação a distância
20	Moodle	Atividade 04 - 2º Fórum de Perguntas e Respostas - Polo SJB	EAD e educação aberta

Fonte: o autor

Quadro 13 - Nível de ensino, modalidade e curso dos alunos que participaram dos fóruns

Fórum	Curso	Nível de ensino	Modalidade
1	Doutorado em Informática na Educação	Doutorado	Presencial
2	Doutorado em Informática na Educação	Doutorado	Presencial
3	Doutorado em Informática na Educação	Doutorado	Presencial
4	Pedagogia	Graduação	EAD
5	Pedagogia	Graduação	EAD
6	Pedagogia	Graduação	EAD
7	Pedagogia	Graduação	EAD
8	Extensão	Extensão	Presencial
9	Extensão	Extensão	Presencial
10	Sistemas de Informação	Graduação	Presencial
11	Sistemas de Informação	Graduação	Presencial
12	Sistemas de Informação	Graduação	Presencial
13	Sistemas de Informação	Graduação	Presencial
14	Tecnologia em Análise e Desenvolvimento de Sistemas	Graduação	Presencial
15	Sistemas de Informação	Graduação	Presencial
16	Especialização em Gestão, Design e Marketing	Pós-graduação Lato-Sensu	Presencial
17	Técnico em Segurança do Trabalho	Ensino Médio	EAD
18	Técnico em Segurança do Trabalho	Ensino Médio	EAD
19	Técnico em Segurança do Trabalho	Ensino Médio	EAD
20	Técnico em Segurança do Trabalho	Ensino Médio	EAD

Fonte: o autor

Os experimentos foram realizados para comparar o resultado do MineraFórum com a análise de dois professores. As tabelas 5 a 8 apresentam informações sobre os experimentos realizados com o software e um professor. As tabelas 9 a 12 envolvem o MineraFórum e outro professor. As informações das tabelas são: docente que analisou as postagens, quantidade de alunos, quantidade de mensagens, média da análise do MineraFórum, média da análise do professor, grau de similaridade entre a média da análise do software e a do docente, média da diferença entre as análises, desvio padrão da análise do sistema, desvio padrão da análise do professor, desvio padrão da diferença entre as análises.

Tabela 5 - Análise do MineraFórum e do professor 1 nos fóruns 1 a 5

Fórum	1	2	3	4	5
Professor que analisou as postagens	A	B	B	A	A
Quantidade de alunos	10	9	5	27	28
Quantidade de mensagens	16	13	26	42	67
Média da análise do MineraFórum	3,19	2,54	3,88	2,45	3,00
Média da análise do professor	3,63	3,54	3,58	2,98	2,25
Grau de similaridade entre a média da análise do MineraFórum e a do professor	87,93%	71,74%	92,08%	82,40%	75,12%
Média da diferença entre as análises	1,06	1,15	1,23	1,00	1,37
Desvio padrão da análise do MineraFórum	1,11	0,97	1,11	0,74	1,36
Desvio padrão da análise do professor	1,41	1,39	1,39	0,87	1,11
Desvio padrão da diferença entre as análises	0,85	1,28	1,07	0,88	1,27

Fonte: o autor

Tabela 6 - Análise do MineraFórum e do professor 1 nos fóruns 6 a 10

Fórum	6	7	8	9	10
Professor que analisou as postagens	A	A	C	C	D
Quantidade de alunos	28	31	18	11	20
Quantidade de mensagens	48	73	76	42	22
Média da análise do MineraFórum	2,92	2,88	3,00	3,21	3,68
Média da análise do professor	2,79	2,32	3,61	3,90	3,82
Grau de similaridade entre a média da análise do MineraFórum e a do professor	95,71%	80,48%	83,21%	82,32%	96,43%
Média da diferença entre as análises	1,04	1,33	1,29	1,50	1,23
Desvio padrão da análise do MineraFórum	1,11	1,19	1,63	1,39	0,72
Desvio padrão da análise do professor	0,80	1,22	1,59	1,32	1,10
Desvio padrão da diferença entre as análises	1,05	1,34	0,99	1,02	0,75

Fonte: o autor

Tabela 7 - Análise do MineraFórum e do professor 1 nos fóruns 11 a 15

Fórum	11	12	13	14	15
Professor que analisou as postagens	D	D	E	E	E
Quantidade de alunos	23	28	12	9	18
Quantidade de mensagens	29	28	12	11	18
Média da análise do MineraFórum	2,66	2,93	3,42	3,55	2,67
Média da análise do professor	3,34	4,14	2,67	3,82	3,56
Grau de similaridade entre a média da análise do MineraFórum e a do professor	79,38%	70,69%	78,05%	92,86%	75,00%
Média da diferença entre as análises	2,07	1,57	1,25	1,36	1,33
Desvio padrão da análise do MineraFórum	1,29	0,86	0,67	0,69	0,59
Desvio padrão da análise do professor	2,02	1,04	1,44	1,25	1,15
Desvio padrão da diferença entre as análises	1,36	0,69	1,54	0,92	1,03

Fonte: o autor

Tabela 8 - Análise do MineraFórum e do professor 1 nos fóruns 16 a 20

Fórum	16	17	18	19	20
Professor que analisou as postagens	F	G	G	G	G
Quantidade de alunos	13	23	24	46	46
Quantidade de mensagens	16	23	24	49	47
Média da análise do MineraFórum	4,00	3,09	2,88	2,98	3,13
Média da análise do professor	3,88	4,39	4,08	4,55	4,34
Grau de similaridade entre a média da análise do MineraFórum e a do professor	96,88%	70,30%	70,41%	65,47%	72,06%
Média da diferença entre as análises	1,00	1,83	1,38	1,73	1,34
Desvio padrão da análise do MineraFórum	1,15	0,95	0,34	0,83	0,45
Desvio padrão da análise do professor	1,15	0,78	1,06	1,12	0,84
Desvio padrão da diferença entre as análises	0,89	0,78	0,97	0,70	0,81

Fonte: o autor

Tabela 9 - Análise do MineraFórum e do professor 2 nos fóruns 1 a 5

Fórum	1	2	3	4	5
Professor que analisou as postagens	H	H	H	B	B
Quantidade de alunos	10	9	5	27	28
Quantidade de mensagens	16	13	26	42	67
Média da análise do MineraFórum	3,19	2,54	3,88	2,45	3,00
Média da análise do professor	4,25	4,23	4,27	3,48	3,31
Grau de similaridade entre a média da análise do MineraFórum e a do professor	75,00%	60,00%	90,99%	70,55%	90,54%
Média da diferença entre as análises	1,44	1,69	1,23	1,12	1,18
Desvio padrão da análise do MineraFórum	1,11	0,97	1,11	0,74	1,36
Desvio padrão da análise do professor	1,24	1,42	1,34	0,63	0,78
Desvio padrão da diferença entre as análises	0,89	1,11	0,91	0,86	1,09

Fonte: o autor

Tabela 10 - Análise do MineraFórum e do professor 2 nos fóruns 6 a 10

Fórum	6	7	8	9	10
Professor que analisou as postagens	B	B	D	D	B
Quantidade de alunos	28	31	18	11	20
Quantidade de mensagens	48	73	76	42	22
Média da análise do MineraFórum	2,92	2,88	3,00	3,21	3,68
Média da análise do professor	3,29	2,97	3,95	4,12	3,09
Grau de similaridade entre a média da análise do MineraFórum e a do professor	88,61%	96,77%	76,00%	78,03%	83,95%
Média da diferença entre as análises	0,96	1,00	1,32	1,38	0,86
Desvio padrão da análise do MineraFórum	1,11	1,19	1,63	1,39	0,72
Desvio padrão da análise do professor	0,50	0,58	1,69	1,19	0,87
Desvio padrão da diferença entre as análises	0,94	0,85	1,07	1,19	0,71

Fonte: o autor

Tabela 11 - Análise do MineraFórum e do professor 2 nos fóruns 11 a 15

Fórum	11	12	13	14	15
Professor que analisou as postagens	B	B	I	I	I
Quantidade de alunos	23	28	12	9	18
Quantidade de mensagens	29	28	12	11	18
Média da análise do MineraFórum	2,66	2,93	3,42	3,55	2,67
Média da análise do professor	3,21	3,00	2,92	4,09	3,78
Grau de similaridade entre a média da análise do MineraFórum e a do professor	82,80%	97,62%	85,37%	86,67%	70,59%
Média da diferença entre as análises	1,24	1,00	1,17	1,45	1,33
Desvio padrão da análise do MineraFórum	1,29	0,86	0,67	0,69	0,59
Desvio padrão da análise do professor	0,98	0,94	1,16	1,14	1,06
Desvio padrão da diferença entre as análises	1,48	0,86	1,19	0,69	0,84

Fonte: o autor

Tabela 12 - Análise do MineraFórum e do professor 2 nos fóruns 16 a 20

Fórum	16	17	18	19	20
Professor que analisou as postagens	J	F	F	F	F
Quantidade de alunos	13	23	24	46	46
Quantidade de mensagens	16	23	24	49	47
Média da análise do MineraFórum	4,00	3,09	2,88	2,98	3,13
Média da análise do professor	3,69	3,35	4,04	3,92	3,47
Grau de similaridade entre a média da análise do MineraFórum e a do professor	92,19%	92,21%	71,13%	76,04%	90,18%
Média da diferença entre as análises	0,69	1,30	1,33	1,27	1,28
Desvio padrão da análise do MineraFórum	1,15	0,95	0,34	0,83	0,45
Desvio padrão da análise do professor	1,08	1,03	0,91	1,22	1,33
Desvio padrão da diferença entre as análises	0,60	1,26	0,70	0,86	0,77

Fonte: o autor

Observando os resultados apresentados nas tabelas 5 a 12, é possível verificar que a média das relevâncias calculadas pelo MineraFórum foi similar à média indicada pelos professores.

A tabela 13 exibe o grau de similaridade entre a média dos professores e a média do MineraFórum. Em relação ao professor 1, a variação deste grau ocorreu de 65,47% a 96,88%. Em comparação ao docente 2, foi de 60,00% a 97,62%.

Tabela 13 - Grau de similaridade entre a média dos professores e a média do MineraFórum

Fórum	Média da Análise do MineraFórum	Média da Análise do Professor 1	Média da Análise do Professor 2	Grau de similaridade entre a média do Professor 1 e a média do MineraFórum	Grau de similaridade entre a média do Professor 2 e a média do MineraFórum
1	3,19	3,63	4,25	87,93%	75,00%
2	2,54	3,54	4,23	71,74%	60,00%
3	3,88	3,58	4,27	92,08%	90,99%
4	2,45	2,98	3,48	82,40%	70,55%
5	3,00	2,25	3,31	75,12%	90,54%
6	2,92	2,79	3,29	95,71%	88,61%
7	2,88	2,32	2,97	80,48%	96,77%
8	3,00	3,61	3,95	83,21%	76,00%
9	3,21	3,90	4,12	82,32%	78,03%
10	3,68	3,82	3,09	96,43%	83,95%
11	2,66	3,34	3,21	79,38%	82,80%
12	2,93	4,14	3,00	70,69%	97,62%
13	3,42	2,67	2,92	78,05%	85,37%
14	3,55	3,82	4,09	92,86%	86,67%
15	2,67	3,56	3,78	75,00%	70,59%
16	4,00	3,88	3,69	96,88%	92,19%
17	3,09	4,39	3,35	70,30%	92,21%
18	2,88	4,08	4,04	70,41%	71,13%
19	2,98	4,55	3,92	65,47%	76,04%
20	3,13	4,34	3,47	72,06%	90,18%

Fonte: o autor

Cabe ressaltar que o valor do grau de similaridade foi menor que 70% apenas no experimento 19, na comparação com o professor 1. Em relação ao docente 2, este fato ocorreu somente no experimento 2.

De acordo com a tabela 13, a variação do grau de similaridade ocorreu da seguinte forma:

a) Entre a média do professor 1 e a média do MineraFórum:

- 70% a 80% em nove (9) experimentos;
- 80% a 90% em cinco (5) experimentos;
- 90% a 100% em cinco (5) experimentos.

b) Entre a média do docente 2 e a média do MineraFórum:

- 70% a 80% em sete (7) experimentos;
- 80% a 90% em cinco (5) experimentos;
- 90% a 100% em sete (7) experimentos.

A figura 42 apresenta a variação do grau de similaridade no decorrer dos experimentos. O gráfico “Similaridade1” corresponde ao grau entre a média do professor 1 e a média do MineraFórum. O gráfico “Similaridade2” refere-se à comparação com o docente 2.

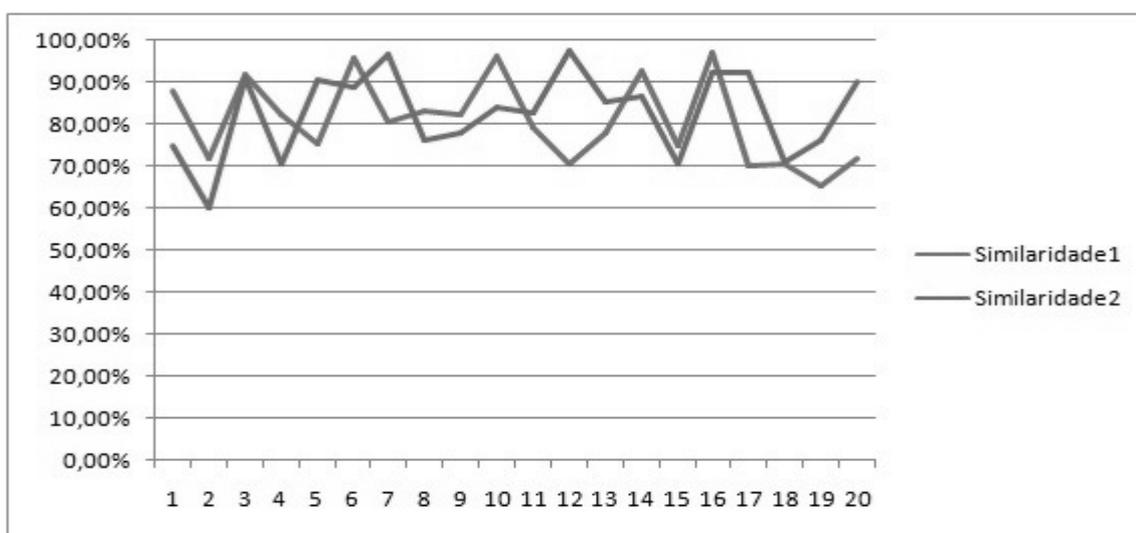


Figura 42 - Variação do grau de similaridade nos experimentos
Fonte: o autor

5.4.3 ANÁLISE E DISCUSSÃO DOS RESULTADOS DA VERSÃO 3.0

Ao iniciar esta seção, cabe lembrar a metodologia para o cálculo da relevância de cada postagem. Na versão 3.0, a relevância é calculada considerando a pertinência temática da contribuição textual, a proporção de citações da postagem e a verificação de similaridade da mensagem com outra do fórum.

A partir dos dados relatados na tabela 13 e no gráfico da figura 42, foi possível observar que o grau de similaridade entre a média dos professores e a média do MineraFórum apresentou valores próximos. Desta forma, pode-se concluir que a análise qualitativa das postagens, efetuada pela ferramenta, é semelhante ao diagnóstico realizado pelos docentes.

Cabe ressaltar que os resultados exibidos na tabela 13 apresentam a similaridade da média da relevância de todas as postagens. Esta pesquisa não avaliou a semelhança entre os

valores individuais obtidos com o cálculo da relevância de cada mensagem. Este ponto não foi considerado, pois cada professor utilizou seus próprios critérios para analisar as postagens. Como estes fatores podem ser distintos dos que são empregados pelo MineraFórum, o valor da relevância de cada postagem pode ser diferente. Destaca-se que quanto maior for a proximidade dos critérios de análise utilizados pelo docente com os do software, maior será a similaridade entre os resultados.

Na tabela 13, observa-se que o grau de similaridade no experimento 19 (com o professor 1) foi 65,47%. No experimento 2 (com o professor 2), o grau foi 60,00%. Os fatores que podem ter contribuído para estes valores são:

- Os critérios de análise empregados pelo professor 1 no fórum 19 e pelo docente 2 no fórum 2, foram distintos do MineraFórum.
- No experimento 2, a quantidade de postagens do fórum foi pequena.

Considerando que foi atribuído um valor inteiro entre 0 e 5 para a relevância de cada postagem, verificou-se que em algumas situações, os valores calculados pelo MineraFórum foram muito distintos⁴¹ dos indicados pelos professores. Este fato ocorreu nas seguintes circunstâncias:

- a) A mensagem citou conceitos relevantes ao tema, mas não agregou valor ao debate.

Estes casos ocorreram em mensagens onde não havia coerência e coesão no texto digitado. Como o MineraFórum não analisa estes parâmetros, o software calculou a relevância das mensagens de acordo com os conceitos importantes citados. O professor atribuiu um valor baixo para estas postagens.

- b) A contribuição textual mencionou conceitos relevantes, que não foram citados no texto de referência ou no fórum.

Nesta situação, o professor atribuiu um valor alto para a postagem. Como não havia condições para o software identificar estes conceitos, a relevância calculada foi baixa.

- c) A postagem não citou conceitos relevantes à discussão, mas apresentou um exemplo importante ou uma experiência pessoal interessante.

⁴¹ Considerou-se como valores muito distintos, aqueles nos quais houve uma diferença superior a 2 pontos entre o valor calculado pelo software e o indicado pelo docente.

A ferramenta atribuiu importância nula para as mensagens que não abordaram conceitos relevantes. O docente analisou com alta relevância as postagens com exemplos interessantes ou com experiências significantes.

- d) A contribuição textual não abordou conceitos relevantes à discussão, mas indicou uma referência bibliográfica importante.

O MineraFórum não analisa a importância de referências bibliográficas. Desta forma, o software calculou relevância nula para a postagem e o professor indicou um valor alto.

- e) A postagem não citou conceitos relevantes ao tema, mas mencionou um *site* interessante.

O MineraFórum não verifica o conteúdo de *sites*. As mensagens deste tipo receberam relevância nula, atribuída pelo software. O docente diagnosticou com alta significância estas contribuições textuais.

- f) A mensagem não mencionou conceitos relevantes, mas anexou uma imagem ou arquivo importante.

A ferramenta não analisa o conteúdo de uma imagem ou arquivo anexado. Desta forma, o software calculou relevância nula para estas postagens e o docente indicou um valor alto.

A próxima seção descreve a integração do MineraFórum em três Ambientes Virtuais de Aprendizagem.

5.5 COMPARAÇÃO ENTRE O MINERAFÓRUM E OUTROS SISTEMAS/SOLUÇÕES

A seção 3.6.2 apresentou diversos trabalhos sobre análise de fóruns de discussão realizadas com técnicas de mineração de textos. As pesquisas relataram as características dos sistemas/soluções utilizados. A tabela 14 apresenta uma comparação entre estes recursos e o MineraFórum.

Tabela 14 - Comparação entre sistemas/soluções e o MineraFórum

Autores	Semelhanças dos sistemas/soluções utilizados e o MineraFórum	Diferenças dos sistemas/soluções utilizados e o MineraFórum
Rebedea et al. (2008)	Ferramenta que utiliza técnicas de mineração de textos para analisar a relevância de mensagens textuais em relação ao tema em debate.	Analisa mensagens de chats para fornecer indicadores relacionados à teoria polifônica de Bakhtin. A ferramenta não está integrada a um AVA.
Ravi e Kim (2007)	Classificadores para analisar o conteúdo de mensagens de fóruns para auxiliar o docente.	Os classificadores analisam os textos através da identificação de “atos de discurso”. Os indicadores fornecidos pelos classificadores são distintos dos apresentados pelo MineraFórum.
Kim et al. (2007)	Agente inteligente que analisa o conteúdo de mensagens de fóruns.	As técnicas de mineração de textos utilizadas são distintas do MineraFórum. O objetivo do agente é responder automaticamente as perguntas dos alunos, e o intuito do MineraFórum é apresentar indicadores para auxiliar o docente na análise de fóruns.
Lin et al. (2009)	Sistema de classificação de gêneros para investigar o conteúdo das mensagens de fóruns. A ferramenta auxilia o professor.	As técnicas utilizadas para analisar o conteúdo das mensagens e os resultados fornecidos são distintos do MineraFórum. Não há integração com AVA.
Li e Huang (2008a)	Ferramenta que utiliza técnicas de mineração de textos para analisar a relevância das mensagens em relação ao tema em debate.	O trabalho representa os textos com o modelo de espaço vetorial e realiza a comparação dos mesmos com a medida de similaridade cosseno. A ferramenta desenvolvida não está integrada a um AVA.
Li et al.(2008b)	Ferramenta que utiliza técnicas de mineração de textos para analisar a relevância das postagens em relação ao tópico da discussão.	A ferramenta representa os textos com o modelo de espaço vetorial, e detecta se os mesmos são semelhantes com a medida de similaridade cosseno. O sistema desenvolvido pelos autores oferece auxílio aos alunos, enquanto o MineraFórum apresenta informações para auxiliar os docentes na análise de fóruns.

Fonte: o autor

5.6 INTEGRAÇÃO DO MINERAFÓRUM AOS AMBIENTES VIRTUAIS DE APRENDIZAGEM

Esta seção apresenta a integração do MineraFórum aos AVAs Rooda, ETC e Moodle. Observando a pesquisa das ferramentas disponíveis nestes ambientes, apresentada na seção 3.2.1, verificou-se que nenhum deles possui um recurso para analisar qualitativamente os textos das mensagens de um fórum de discussão. Desta forma, destaca-se a importância do MineraFórum em oferecer esta análise e proporcionar indicadores que possam auxiliar o docente.

A integração do MineraFórum a um AVA é fundamental para que o professor possa utilizar o recurso dentro do ambiente. O docente poderá escolher os fóruns de discussão disponíveis, para que as mensagens sejam mineradas.

5.6.1 ROODA

O primeiro AVA adotado para integração com o MineraFórum foi o Rooda, descrito na seção 3.2.1, cuja escolha ocorreu devido aos seguintes critérios:

- a) É um ambiente que possibilita a interação, comunicação, colaboração/cooperação.
- b) No fórum do ambiente, as mensagens são agrupadas por tópicos. Dentro de cada um, as contribuições textuais são exibidas em ordem cronológica.
- c) É um software livre.
- d) Foi desenvolvido na UFRGS, o que possibilita acesso ao seu código fonte e suporte tecnológico.

Cabe ressaltar que foi realizada a integração das três versões do MineraFórum com o Rooda. A figura 43 apresenta a interface do Rooda com o *link* que aciona a execução do MineraFórum.

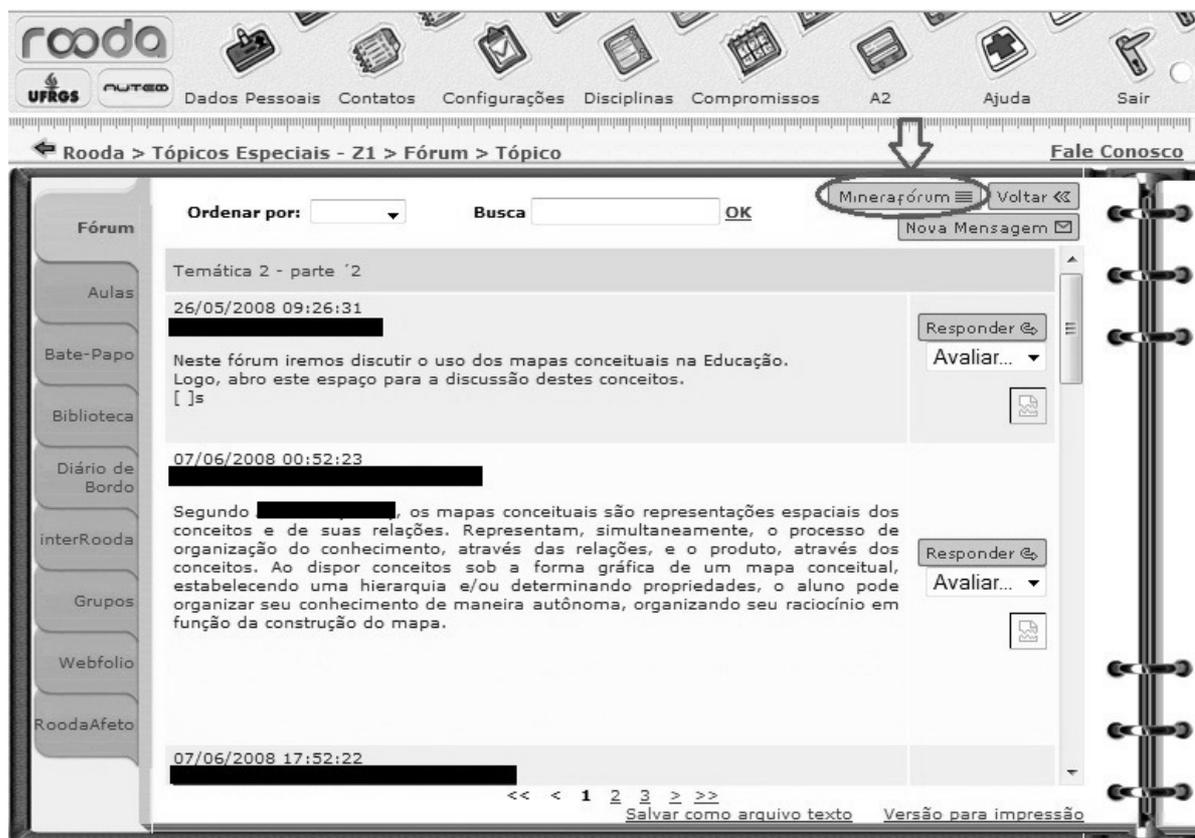


Figura 43 - Interface do Rooda com *link* para executar o MineraFórum

Fonte: <https://rooda-breno.dev.ufrgs.br>

5.6.2 ETC

O segundo AVA adotado para integração com o MineraFórum foi o ETC, descrito na seção 3.2.1, cuja escolha ocorreu devido aos seguintes critérios:

- a) É um ambiente que possibilita a interação, comunicação, colaboração/cooperação.
- b) No fórum do ambiente, as mensagens são agrupadas por tópicos.
- c) É um software livre.
- d) Foi desenvolvido na UFRGS, o que possibilita acesso ao seu código fonte e suporte tecnológico.

Cabe observar que foi realizada somente a integração da versão 3.0 do MineraFórum com o ETC. A figura 44 apresenta a interface do ETC com o *link* que permite a execução do MineraFórum.

Figura 44 - Interface do ETC com *link* para executar o MineraFórum
 Fonte: <http://www.nuted.ufrgs.br/etc2/>

5.6.3 MOODLE

O terceiro AVA onde ocorreu a integração do MineraFórum foi o Moodle, descrito na seção 3.2.1, cuja escolha ocorreu devido aos seguintes critérios:

- a) É um ambiente que possibilita a interação, comunicação, colaboração/cooperação.
- b) No fórum do ambiente, as mensagens são agrupadas por tópicos.
- c) É um software livre.
- d) O código fonte pode ser acessado. Existem diversas listas de discussão na *Web*, fóruns, sites, que auxiliam no suporte ao código fonte.

Cabe observar que foi realizada somente a integração da versão 3.0 do MineraFórum com o Moodle. A figura 45 apresenta a interface do Moodle com o *link* que inicia a execução do MineraFórum.

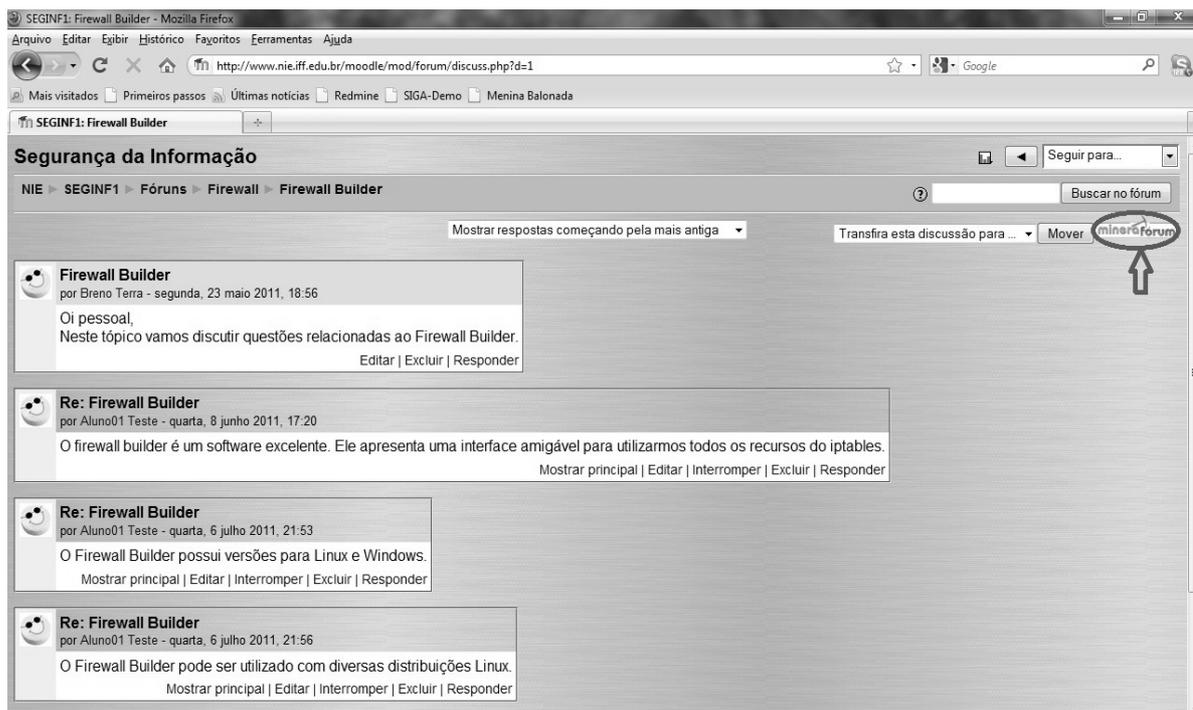


Figura 45 - Interface do Moodle com *link* para executar o MineraFórum
Fonte: <http://www.nie.iff.edu.br/moodle/>

O próximo capítulo descreve as considerações finais desta pesquisa. As principais contribuições são explicadas. As características inovadoras do MineraFórum e suas limitações são relatadas. Os recursos da nova versão do software e as perspectivas futuras são apresentados.

6. CONSIDERAÇÕES FINAIS

De acordo com Palloff e Pratt (2004), o papel do professor nos fóruns de discussão é muito importante. Além de escrever mensagens de apoio e motivação para os alunos, responder às perguntas, ele deve observar o nível de participação de cada discente. Caso seja detectado que um aluno não está participando ou está desviando do tema da discussão, o docente deve tentar ajudá-lo a suplantar as dificuldades e a solucionar os problemas.

A análise do fórum permite que o professor possa diagnosticar informações sobre os alunos. No entanto, se houver uma grande quantidade de alunos, o tempo necessário para o docente analisar as discussões será grande. Um software que realize uma análise qualitativa das postagens pode auxiliar o professor, no sentido de otimizar o seu tempo. Isto permitirá que o docente identifique os alunos que estão (e aqueles que não estão) discutindo os conceitos do tema em debate. Desta forma, o docente poderá dedicar um tempo maior para descobrir o motivo pelo qual alguns discentes não discutiram determinados conceitos. Caso o professor identifique alunos com dificuldade para compreender os conceitos importantes, ele poderá oferecer ajuda. Se for necessário, o docente também pode motivar os discentes a participar mais do fórum.

6.1 CAMINHOS PERCORRIDOS

A ideia que originou este estudo foi pesquisar uma metodologia para realizar a análise qualitativa das mensagens de fóruns de discussão. Por ser uma área da Ciência da Computação que envolve a investigação de conteúdos textuais, a mineração de textos foi escolhida como ponto de partida. Os experimentos realizados na seção 4.3 demonstraram que a mineração de textos utilizando a representação através de grafos pode ser utilizada para diagnosticar a relevância das postagens em fóruns.

Neste momento, cabe lembrar a questão desta pesquisa: “Como um recurso de mineração de textos, incorporado a um ambiente virtual de aprendizagem, pode auxiliar na análise qualitativa das contribuições em um fórum de discussão?”. Considerando esta

questão, foi elaborado o objetivo geral: “Analisar de que forma o MineraFórum poderia contribuir na análise qualitativa das mensagens postadas em um fórum de discussão”.

A partir do objetivo geral, foram derivados os objetivos específicos abaixo:

- a) Definir indicadores que possam ser utilizados para identificar contribuições relevantes em um fórum de discussão.

Nas primeiras etapas desta pesquisa utilizou-se somente o indicador da relevância temática para caracterizar mensagens importantes. Este indicador foi implementado na primeira e segunda versão do MineraFórum. Posteriormente, a partir do questionário aplicado com os professores (seção 4.4), foram diagnosticados outros fatores significativos. Escolheu-se os indicadores que possuíam o maior peso, que poderiam ser implementados com a técnica de mineração de textos utilizando grafos, ou com informações fornecidas pelo AVA. Os indicadores selecionados foram: análise temática das contribuições textuais, proporção de citações da mensagem, verificação da similaridade da mensagem com outra do fórum. Estes foram mapeados na versão 3.0 do MineraFórum.

- b) Construir o MineraFórum, para auxiliar no acompanhamento qualitativo e quantitativo das contribuições em um fórum de discussão.

O desenvolvimento do sistema ocorreu ao longo de três versões. A metodologia para construção do software foi apresentada na seção 4.5. O projeto e implementação das mesmas foi relatado nas seções 5.1 a 5.4. Os experimentos realizados com as versões foram analisados e discutidos.

- c) Integrar o MineraFórum aos Ambientes Virtuais de Aprendizagem Rooda, ETC e Moodle.

A integração do MineraFórum com os três AVAs foi realizada com sucesso.

As produções científicas relacionadas a esta pesquisa estão relatadas no Apêndice A. Cabe destacar que o MineraFórum faz parte dos projetos abaixo:

1. Título do Projeto: MineraROODA: Ferramentas de mineração de conteúdo cognitivo e de subjetividade afetiva no Ambiente Virtual de Aprendizagem ROODA. Referência da Chamada: Edital MCT/CNPq Nº 014/2010 - Universal. Faixa de Financiamento da proposta: Faixa A. Término: 2012.

2. ROODA: novas ferramentas para incorporação no ambiente virtual de aprendizagem. Referência da Chamada: Programa Pesquisador Gaúcho (PqG). Edital FAPERGS 006/2010. Faixa de Financiamento da proposta: Faixa A. Término: 2013.
3. Ampliando possibilidades pedagógicas através da tecnologia de mineração de textos integrada à escrita coletiva a distância. Edital MEC/CAPES N° 029/2010. Término: 2013.

6.2 CONTRIBUIÇÕES DA TESE

A principal contribuição desta pesquisa foi o desenvolvimento do MineraFórum. O software é capaz de apresentar uma visão sobre as contribuições escritas pelos alunos, organizando e agrupando as mensagens de cada um. A ferramenta disponibiliza informações que podem auxiliar o docente na análise de fóruns.

Cabe destacar outras inovações desta pesquisa:

- a) A utilização da técnica de mineração de textos com representação através de grafos, para calcular a relevância temática de contribuições textuais.
- b) O questionário aplicado com professores para mapear os pesos de dez indicadores de relevância das postagens de fóruns (seção 4.4).
- c) As fórmulas para o cálculo da relevância das mensagens (seção 4.5.1).
- d) Os algoritmos das fórmulas para o cálculo da relevância das mensagens (seções 5.2.3, 5.3.1, 5.4.1).

6.3 DESAFIOS E LIMITAÇÕES

Destaca-se que o uso do MineraFórum auxilia a análise da relevância das mensagens de um fórum. No entanto, cabe ressaltar que o tempo necessário para que o software realize o processo depende de alguns aspectos. Os fatores que influenciam no tempo de processamento

são: hardware do computador no qual o sistema é executado, quantidade de postagens, quantidade de caracteres das mensagens, quantidade de caracteres do texto de referência.

As limitações do MineraFórum referem-se às ações que não são executadas pelo software na análise das postagens:

- Detecção de erros ortográficos e gramaticais.
- Identificação de abreviações.
- Avaliação de erros conceituais.
- Diagnóstico de coerência e coesão textual.
- Verificação se houve cópia do texto de um site ou referência bibliográfica.
- Apreciação de exemplos ou experiências pessoais relacionados ao tópico da discussão, nos quais não houve citação de conceitos relevantes.
- Análise de imagens e arquivos anexados.

Cabe ressaltar que a interface do MineraFórum não foi redesenhada e nem adaptada ao design dos AVAs onde foi integrado. Esta limitação ocasiona a falta de semelhança nos modelos das telas. A tecnologia utilizada para implementar o MineraFórum (applet⁴² Java) impossibilitou que a interface fosse construída de forma idêntica aos AVAs.

Ao utilizar o MineraFórum, é importante observar:

- Dois fatores influenciam diretamente o resultado da mineração dos conceitos relevantes do texto de referência: a quantidade de palavras e a qualidade técnica do texto.
- Se não for utilizado um texto de referência, a base de conceitos digitada pelo professor deve possuir boa qualidade técnica.
- Caso as limitações do sistema relatadas nesta seção desfavoreçam a análise da maioria das postagens de um fórum, o resultado apresentado pelo MineraFórum não será útil.

⁴² Um applet é um programa Java que pode ser incorporado a páginas Web (DEITEL e DEITEL, 2005).

6.4 TRABALHOS FUTUROS

Atualmente está em desenvolvimento uma nova versão do MineraFórum, que contemplará as seguintes funcionalidades:

- a) Definição do grau de flexibilidade na análise.

Com este recurso o professor poderá ajustar o rigor da análise da relevância das mensagens.

- b) Exposição de um grafo que reúna os conceitos relevantes extraídos do texto de referência e das postagens. Os vértices serão exibidos com cores diferentes.

Este grafo permitirá que o docente verifique quais conceitos relevantes do texto de referência foram abordados no fórum.

- c) Criação de uma “aba” na tela principal do sistema para permitir a inclusão/exclusão de *stopwords*.

Este recurso possibilitará que o professor adicione ou remova *stopwords* em uma tela apropriada.

- d) Apresentação de uma “barra de progresso” durante o processo de mineração.

Com a “barra de progresso” será possível visualizar a duração da mineração.

No futuro, outros recursos serão disponibilizados no MineraFórum:

- a) Possibilidade de edição dos grafos apresentados, para que seja possível aumentar o tamanho dos vértices e alterar a cor.
- b) Serão considerados outros indicadores de relevância, além dos atuais. Os indicadores estão listados no questionário aplicado com os professores (seção 4.4).

6.5 A TÍTULO DE CONCLUSÃO

Esta pesquisa descreveu como um recurso de mineração de textos, integrado a um Ambiente Virtual de Aprendizagem, pode auxiliar na análise qualitativa das postagens de fóruns de discussão. O software permite executar, de forma automatizada, o referido processo.

A ferramenta apresenta uma visão sobre as contribuições escritas pelos discentes, organizando e agrupando as postagens de cada um. O sistema disponibiliza indicadores que podem auxiliar o docente.

Os experimentos realizados demonstraram que o MineraFórum ofereceu bons resultados. Na versão 3.0 do sistema, a média da análise realizada pelos professores e a média calculada pelo software apresentaram semelhanças.

A partir das informações apresentadas pelo MineraFórum, o professor poderá direcionar seu apoio aos alunos que colocaram poucas contribuições relevantes no fórum. O docente também poderá motivar a interação entre os discentes que mais registraram textos relevantes com os que redigiram poucos.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, T.; ROURKE, L.; GARRISON, R.; ARCHER, W. Assessing teaching presence in a computer conferencing context. **Journal of Asynchronous Learning Networks**, 5(2), p. 1-17, 2001.

ANDERSON, T. Teaching in an online learning context. In: ANDERSON, T.; ELLOUMI F. (Eds.). **Theory and practice of online learning**, 2004. p. 273-294.

ARAÚJO JÚNIOR, C. F.; MARQUESI, S. C. Atividades em ambientes virtuais de aprendizagem: parâmetros de qualidade. In: LITTO, F. M.; FORMIGA, M. M. M. (Org.). **Educação a distância: o estado da arte**. São Paulo: Pearson Education do Brasil, 2009. p. 358-368.

AZEVEDO, B. F. T.; BEHAR, P. A.; REATEGUI, E. B. Estudo de análise qualitativa em fórum de discussão. **Revista Novas Tecnologias na Educação (RENOTE)**, v. 7, n. 2, 2009.

BAKHTIN, M. **Problems of Dostoevsky's Poetics**. Ardis, 1973.

BAKHTIN, M. **Estética da criação verbal**. São Paulo: Martins Fontes, 1997.

BARBOSA, A. E. T. Ambientes de aprendizagem de curso online e desenvolvimento de pensamento crítico. **Revista Estudos Lingüísticos XXXV**, p. 392-401, 2006.

BARBOSA, M. L.; SEVERO, C. E. P.; REATEGUI, E. B. Mineração de padrões no gênero textual blog. **Revista Novas Tecnologias na Educação (RENOTE)**, v. 7, n. 2, 2009.

BARRETO, R. G. et al. As tecnologias da informação e da comunicação na formação de professores. **Revista Brasileira de Educação**, Rio de Janeiro, v. 11, n. 31, p. 31-42, 2006.

BASSANI, P. B. S. **Mapeamento das interações em ambiente virtual de aprendizagem: uma possibilidade para avaliação em educação a distância**. 2006. Tese (Doutorado em Informática na Educação) - PPGIE, UFRGS, Porto Alegre, 2006.

BASSANI, P. B. S.; BEHAR, P. A. Análise das interações em ambientes virtuais de aprendizagem: uma possibilidade para avaliação da aprendizagem em EAD. **Revista Novas Tecnologias na Educação (RENOTE)**, v. 4, n. 1, 2006.

BEHAR, P. A.; COLOMBO, M. C.; SOUZA, L. B.; AMARAL, C. A categorização das funcionalidades do ambiente virtual ROODA. **Revista Novas Tecnologias na Educação (RENOTE)**, v.3, n.1, 2005a.

BEHAR, P. A.; LEITE, S. M. Criando novos espaços pedagógicos na Internet: o ambiente ROODA. In: WWW/INTERNET, 2005, Lisboa. **Anais...** Lisboa: IADIS, 2005b, v. 1, p. 3-10.

BEHAR, P. et al. ROODA/UFRGS: uma articulação técnica, metodológica e epistemológica. In: BARBOSA, R. M. (Org.). **Ambientes Virtuais de Aprendizagem**. Porto Alegre: Artmed, 2005c. p. 51-70.

BEHAR, P. A.; MACEDO, A. L.; BITENCOURT, J. B.; MAZZOCATO, S. B. Um editor de escrita coletiva na Web. In: I ENCONTRO INTERNACIONAL DE EDUCAÇÃO A DISTÂNCIA: POLÍTICAS PÚBLICAS E NOVAS POSSIBILIDADES DE SE FAZER EAD, 2005, Taquara. **Anais...** Taquara: I Encontro Internacional de Educação a Distância, 2005d.

BEHAR, P. A.; et al. Construção e Aplicação do ETC - editor de texto coletivo. In: BARBOSA, R. M. (Org.). **Ambientes Virtuais de Aprendizagem**. Porto Alegre, 2005e, p. 17-28.

BEHAR, P. A.; LEITE, S. M. The Virtual Learning Environment ROODA: An Institutional Project of Long Distance Education. **Journal of science education and technology**, v. 15, n. 2, p. 159-167, 2006a.

BEHAR, P. A.; MACEDO, A. L.; BITENCOURT, J. B.; MAZZOCATO, S. B. Escrita Coletiva: o potencial de um Groupware via Web. **Revista Novas Tecnologias na Educação (RENTE)**, v.4, n.1, 2006b.

BEHAR, P. A.; BERNARDI, M.; SOUZA, A. P. F. de Castro; KELLEN, K. ROODA: desenvolvimento, implementação e validação de um AVA para UFRGS. In: TALLER INTERNACIONAL DE SOFTWARE EDUCATIVO, XII, 2007, Santiago. **Anais...** Santiago: LOM Ediciones S.A., 2007a, v. 1. p. 321-338.

BEHAR, P. A.; LEITE, S. M.; BORDINI, S.; SOUZA, L. B.; SIQUEIRA, L. G. Avaliação de Ambientes Virtuais de Aprendizagem: O Caso do ROODA na UFRGS. **Revista Avances en Sistemas e Informática**, Medellín, v. 4, n. 1, 2007c.

BEHAR, P. A.; MACEDO, A. L.; MAZZOCATO, S. B. Tecnologias de Suporte ao Trabalho Coletivo: planejamento e aplicação de um objeto de aprendizagem. In: XVIII SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2007, São Paulo. **Anais...** Porto Alegre: Sociedade Brasileira de Computação, 2007d, p. 90-99.

BEHAR, P. A. e colaboradores. **Modelos Pedagógicos em Educação a Distância**. Porto Alegre: ArtMed, 2009a.

BEHAR, P. A.; MACEDO, A. L.; PASSOS, J.; PASSOS, P. C. S. J. Collective Text Editor: a new interface focused on interaction design. In: TATNALL, A.; JONES, A. (Org.). **Education and Technology for a better world**. Springer, 2009b, 1 ed., v. 1, p. 331-339.

BEHAR, P. A.; MACEDO, A. L.; ZANK, C.; VINADE, F. ETC: o que mudou e por quê. **RENTE. Revista Novas Tecnologias na Educação**, v. 8, n. 3, 2010.

BELLONI, M. L. **Educação a distância**. 5. ed. Campinas: Autores Associados, 2008.

BENGOETXEA, E. **Inexact Graph Matching Using Estimation of Distribution Algorithms**. 2002. PhD Thesis, Ecole Nationale Supérieure des Télécommunications, Paris, 2002.

BERGE, Z. L. Facilitating computer conferencing: recommendations from the field. **Educational Technology**, n. 35, p. 22-30, 1995.

BERGE, Z. L. Example case studies in post-secondary, online teaching. In: "THE VIRTUAL UNIVERSITY?", 1996, Melbourne. **Proceedings...** HART, G.; MASON J. (eds.), 1996. p. 99-105.

BLOEHDORN, S.; CIMIANO, P.; HOTH, A. Learning ontologies to improve text clustering and classification. In: ANNUAL CONFERENCE OF THE GERMAN CLASSIFICATION SOCIETY (GfKI), 29, 2005, Magdeburg. **Proceedings...** 2005, p. 334-341.

BLONDEL, V. D.; VAN DOOREN, P. Similarity matrices for pairs of graphs. In: ICALP CONFERENCE, 2003, Eindhoven. **Proceedings...** New York: BAETEN, J. C. M. et al. (eds.), LNCS, Springer-Verlag, 2003. p. 739-550, n. 2719.

BLOOM, B. S.; KRATHWOHL, D. R. **Taxonomy of Educational Objectives: The Classification of Educational Goals.** New York: Longmans, Green, 1956.

BONDY, J. A.; MURTY, U. S. R. **Graph theory with applications.** New York: Elsevier Science Publishing Co., 1982.

BOOCH, G.; RUMBAUGH, J.; JACOBSON, I. **UML - guia do usuário.** Rio de Janeiro: Campus, 2000.

BOYD, R.; APPS, J. **Redefining the discipline of adult education.** San Francisco: Jossey-Bass, 1980.

BUNKE, H.; SHEARER, K. A graph distance metric based on maximal common subgraph. **Pattern Recognition Letters**, v. 19, n. 3-4, p. 255-259, 1998.

BUNKE, H. Graph matching: Theoretical foundations, algorithms, and applications. **Vision Interface**, p. 82-88, 2000.

BRASIL. **Decreto nº 5.622**, de 19 de dezembro de 2005. Disponível em: <http://portal.mec.gov.br/seed/arquivos/pdf/dec_5622.pdf>. Acesso em: 18/09/2009.

BRASIL. **Referenciais de Qualidade para Educação Superior a Distância.** Ministério da Educação e Cultura, 2007. Disponível em: <<http://portal.mec.gov.br/seed/arquivos/pdf/legislacao/refead1.pdf>>. Acesso em: 18/09/2009.

CARLINI, A. **20% a distância: e agora ?**: orientações práticas para o uso de tecnologia de educação a distância. São Paulo: Pearson Education do Brasil, 2010.

CASTELLANOS, M. HotMiner: Discovering Hot Topics from Dirty Text. In: BERRY, M. W. (ed.). **Survey of text mining: clustering, classification, and retrieval.** New York: Springer-Verlag, 2004. p. 123-157.

CHARTRAND, G.; KUBICKI, G.; SCHULTZ, M. Graph similarity and distance in graphs. **Aequationes Mathematicae**, v. 55, p. 129-145, 1998.

CHEN, G.; CHIU M. M. Online discussion processes: Effects of earlier messages' evaluations, knowledge content, social cues and personal information on later messages. **Computers & Education**, 50, p. 678-692, 2008.

CHEN, J.; LI, Q.; WANG, L.; JIA, W. Automatically generating an e-textbook on the web. **World Wide Web**, v. 8, n. 4, p. 377-394, 2005.

CHEN, Y.; CHENG, X.; HUANG, Y. A Wavelet-Based Model to Recognize High-Quality Topics on Web Forum. In: IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY, 2008, Sydney. **Proceedings...** 2008. p.343-351, v. 1.

CHEN, N.; KINSHUK; WEI, C.; CHEN, H. Mining e-Learning domain concept map from academic articles. **Computers & Education**, v. 50, p. 1009-1021, 2008.

COLL, C; MONEREO, C. Educação e aprendizagem no século XXI: novas ferramentas, novos cenários, novas finalidades. In: COLL, C; MONEREO, C. e colaboradores. **Psicologia da educação virtual: aprender e ensinar com as tecnologias da informação e da comunicação**. Porto Alegre: Artmed, 2010a. p. 15-46.

COLL, C; MAURI, T.; ONRUBIA, J. A incorporação das tecnologias da informação e da comunicação na educação. In: COLL, C; MONEREO, C. e colaboradores. **Psicologia da educação virtual: aprender e ensinar com as tecnologias da informação e da comunicação**. Porto Alegre: Artmed, 2010b. p. 66-93.

DALMOLIN, L. C. D.; NASSAR, S. M.; BASTOS, R. C.; MATEUS, G. P. A Concept Map Extractor Tool for Teaching and Learning. In: IEEE INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 9., 2009, Riga. **Proceedings...** IEEE Computer Society, 2009. p. 18-20.

DEITEL, H. M.; DEITEL, P. J. **Java: como programar**. 6. ed. São Paulo: Pearson Education do Brasil, 2005.

DENNEN, V. P. Looking for evidence of learning: Assessment and analysis methods for online discourse. **Computers in Human Behavior**, n. 24, p. 205-219, 2008.

DIAS-DA-SILVA, B. C.; MORAES, H. R.; OLIVEIRA, M. F.; HASEGAWA, R.; AMORIM, D. A.; PASCHOALINO, C.; NASCIMENTO, A. C. Construção de um thesaurus eletrônico para o português do Brasil. **Processamento Computacional do Português Escrito e Falado (PROPOR)**, v. 4, p. 1-10, 2000.

DIAS-DA-SILVA, B. C.; MORAES, H. R. A construção de um thesaurus eletrônico para o português do Brasil. **ALFA**, v. 47, n. 2, p. 101-115, 2003.

DILLENBOURG, P.; SCHNEIDER, D. K.; SYNTETA, P. Virtual Learning Environments. In: HELLENIC CONFERENCE INFORMATION & COMMUNICATION TECHNOLOGIES IN EDUCATION, 3., 2002, Greece. **Proceedings...** Greece: DIMITRACOPOULOU, A. (ed.), Kastaniotis Editions, 2002. p. 3-18.

DORNELLES, R. J. **A utilização de tecnologias de Internet na educação a distância: o caso de uma disciplina de graduação da Escola de Administração da Universidade Federal do Rio Grande do Sul**. 2001. Dissertação (Mestrado em Administração), UFRGS, Porto Alegre, 2001. Disponível em: < http://www.ea.ufrgs.br/professores/hfreitas/files/orientacao/mestrado/defesa/pdf/28_dissertacao_dornelles.pdf>. Acesso em: 28/08/2009.

DORRE, J., GERSTL, P., SEIFFERT, R. Text mining: Finding nuggets in mountains of textual data. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 5., 1999, San Diego. **Proceedings...** ACM, 1999. p. 398-401.

DREWEKE, A.; FISCHER, I.; WERTH, T.; WÖRLEIN, M. Text Mining in Program Code. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 626-645.

DRINGUS, L. P.; ELLIS, T. Using data mining as a strategy for assessing asynchronous discussion forums. **Computers & Education**, n. 45(1), p. 141-160, 2005.

DURAN, D.; VIDAL, V. **Tutoria: aprendizagem entre iguais: da teoria à prática**. Porto Alegre: Artmed, 2007.

EDELSTEIN, S.; EDWARDS, J. If You Build It, They Will Come: Building Learning Communities Through Threaded Discussions. **The Online Journal of Distance Learning Administration**, 5(1), 2002.

FAHY, P. J. Indicators of Support in Online Interaction. **International Review of Research in Open and Distance Learning**, v. 4, n. 1, 2003.

FAN, W.; WALLACE, L.; RICH, S.; ZHANG, Z. Tapping the power of text mining. **Communications of ACM**, n. 49(9), p. 76-82, 2006.

FELDMAN, R.; SANGER, J. **The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data**. Cambridge, MA: Cambridge University Press, 2007.

FERREIRA, A. B. H. **Novo Aurélio Século XXI: o dicionário da língua portuguesa**. 3. ed. Rio de Janeiro: Nova Fronteira, 1999.

FERREIRA, F. J. de M., et al. Sumarização de Texto em Ambientes Educacionais na Web. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, XX, 2009, Florianópolis. **Anais...** Florianópolis: SBIE, 2009.

FLEURY, A.; FLEURY, M. T. **Construindo o conceito de competência**. 2001. Disponível em: <<http://www.scielo.br/pdf/rac/v5nspe/v5nspea10.pdf>>. Acesso em: 05 de outubro de 2011.

FORMIGA, M. M. M. A terminologia da EAD. In: LITTO, F. M.; FORMIGA, M. M. M. (Org.). **Educação a distância: o estado da arte**. São Paulo: Pearson Education do Brasil, 2009. p. 39-46.

FRANCISCATO, F. T., et al. Avaliação dos Ambientes Virtuais de Aprendizagem Moodle, TelEduc e Tidia-Ae: um estudo comparativo. **Revista Novas Tecnologias na Educação (RENOTE)**, v. 6, n. 2, 2008.

GARRISON, D. R. Critical thinking and selfdirected learning in adult education: an analysis of responsibility and control issues. **Adult Education Quarterly**, 42(3), p. 136-148, 1992.

GARRISON, D. R.; ANDERSON, T.; ARCHER, W. Critical inquiry in a text-based environment: computer conferencing in higher education. **The Internet and Higher**

Education, 2(2-3), p. 87-105, 2000.

GARRISON, D. R.; ANDERSON, T.; ARCHER, W. Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. **The American Journal of Distance Education**, 15(1), p. 7-23, 2001.

GEROSA, M. A.; PIMENTEL, M. G.; FUKS, H.; LUCENA, C. J. P. Coordenação de Fóruns Educacionais: Encadeamento e Categorização de Mensagens. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, XIV, 2003, Rio de Janeiro. **Anais...** Rio de Janeiro: SBIE, 2003.

GILBERT, P. K.; DABBAGH, N. How to structure online discussions for meaningful discourse: a case study. **British Journal of Educational Technology**, v. 36, n. 1, p. 5-18, 2005.

GLANCE, N. et al. Deriving Marketing Intelligence from Online Discussion. In: KDD, 2005, Chicago. **Proceedings...** ACM, 2005. p. 419-428.

GOODRICH, M. T.; TAMASSIA, R. **Algorithm Design: Foundations, Analysis, and Internet Examples**. New York: John Wiley & Sons, Inc. 2002.

GRICE, H. P. **Studies in the way of words**. Cambridge, MA: Harvard University Press, 1989.

GRUBER, T. R. A Translation Approach to Portable Ontology Specifications. **Knowledge Acquisition**, 5, 2, p. 199-220, 1993.

GUNAWARDENA, C. N.; LOWE, C. A.; ANDERSON, T. Analysis of global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferencing. **Journal of Educational Computing Research**, 17(4), p. 397-431, 1997.

GUNAWARDENA, C. N.; McISAAC, M. S. Distance education. In: D. H. Jonassen (ed.). **Handbook of research for educational communications and technology**. 2. ed. Mahwah, NJ: Erlbaum, 2004. p. 355-396.

GUPTA, V.; LEHAL, G. S. A Survey of Text Mining Techniques and Applications. **Journal of Emerging Technologies in Web Intelligence**, v. 1, n. 1, 2009.

HAN, S.; HILL, J. R. Building understanding in asynchronous discussions: Examining types of online discourse. **Journal of Asynchronous Learning Networks**, 10(4), p. 29-50, 2006.

HARASIM, L. A History of E-learning: Shift Happened. In: WEISS, J. et al. (eds.). **The International Handbook of Virtual Learning Environments**. Netherlands: Springer, 2006. p. 59-94.

HAYKIN, S. **Redes Neurais: princípios e prática**. Porto Alegre: Artmed, 2001.

HEARST, M. Untangling Text Data Mining. In: ANNUAL MEETING OF THE ASSOCIATION OF COMPUTATIONAL LINGUISTICS, 37, 1999, Morristown. **Proceedings...** Morristown: Association of Computational Linguistics, 1999. p. 3-10.

HO, C. H. Assessing electronic discourse: A case study in developing evaluation rubrics. In: 14TH ANNUAL MEETING OF THE SOCIETY FOR TEXT AND DISCOURSE (ST and D), 2004, Chicago. **Proceedings...**, 2004.

HO, C. H.; SWAN, K. Evaluating online conversation in an asynchronous learning environment: An application of Grice's cooperative principle. **Internet and Higher Education**, 10, p. 3-14, 2007.

HOTH0, A.; STAAB, S.; STUMME, G. Ontologies Improve Text Document Clustering. In: INTERNATIONAL CONFERENCE ON DATA MINING, 2003, Melbourne. **Proceedings...** IEEE Computer Society, 2003.

HOWLAND, P.; PARK, H. Cluster-Preserving Dimension Reduction Methods for Efficient Classification of Text Data. In: BERRY, M. W. (ed.). **Survey of text mining: clustering, classification, and retrieval**. New York: Springer-Verlag, 2004. p. 103-122.

JACOBSON, I.; BOOCH, G.; RUMBAUGH, J. **The Unified Software Development Process**. Reading, MA: Addison-Wesley, 1999.

KELLNER, D. Technological Transformation, Multiple Literacies, and the Re-visioning of Education. In: WEISS, J. et al. (eds.). **The International Handbook of Virtual Learning Environments**. Netherlands: Springer, 2006. p. 241-268.

KIM, J., et al. An Intelligent Discussion-Bot for Guiding Student Interactions in Threaded Discussions. In: AAAI SPRING SYMPOSIUM ON INTERACTION CHALLENGES FOR INTELLIGENT ASSISTANTS, 2007, Stanford University. **Proceedings...** AAAI Press, 2007.

KING, P. M.; KITCHENER, K. S. **Developing Reflective Judgment**. San Francisco: Jossey-Bass, 1994.

KLISC, C.; MCGILL, T.; HOBBS, V. The effect of assessment on the outcomes of asynchronous online discussion as perceived by instructors. **Australasian Journal of Educational Technology**, 25(5), p. 666-682, 2009.

KOBAYASHI, M.; AONO, M. Vector Space Models for Search and Cluster Mining. In: BERRY, M. W. (ed.). **Survey of text mining: clustering, classification, and retrieval**. New York: Springer-Verlag, 2004. p. 103-122.

KONRATH, M. L. P.; TAROUCO, L. M. R., BEHAR, P. A. Competências: desafios para alunos, tutores e professores da EaD. **Revista Novas Tecnologias na Educação (RENTE)**, v. 7, n. 1, 2009.

KRATOCHWILL, S.; SILVA, M. Avaliação da aprendizagem online: contribuições específicas da interface fórum. **Revista Diálogo Educacional**, Curitiba, v. 8, n. 24, p. 445-458, 2008.

LARMAN, C. **Utilizando UML e padrões: uma introdução à análise e ao projeto orientados a objetos e ao Processo Unificado**. 2. ed. Porto Alegre: Bookman, 2004.

LÉVY, P. **Cibercultura**. Rio de Janeiro: Editora 34, 1999.

LI, Y; HUANG, R.; Analyzing Peer Interactions in Computer-Supported Collaborative Learning: Model, Method and Tool. **LNCS**, Springer-Verlag Berlin Heidelberg, n. 5169, p.125-136, 2008a.

LI, Y; DONG, M.; HUANG, R. Semantic Organization of Online Discussion Transcripts for Active Collaborative Learning. In: IEEE INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 8., 2008b, Santander. **Proceedings...** IEEE Computer Society Press, 2008b. p. 756-760.

LIM, C. P.; CHEAH, P. T. The role of the tutor in asynchronous discussion boards: A case study of a preservice teacher course. **Educational Media International**, v. 40, n. 2, p. 33-47, 2003.

LIN, F.; HSIEH, L.; CHUANG, F. Discovering genres of online discussion threads via text mining. **Computers & Education**, n. 52(2), p. 481-495, 2009.

LIU, Y. On Document Representation and Term Weights in Text Classification. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 1-22.

LITTO, F. M O atual cenário internacional da EAD. In: LITTO, F. M.; FORMIGA, M. M. M. (Org.). **Educação a distância: o estado da arte**. São Paulo: Pearson Education do Brasil, 2009. p. 14-20.

LOH, S.; WIVES, L. K.; LICHTNOW, D.; OLIVEIRA, J. P. M. de; Concept-Based Text Mining. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 346-358.

LORENZATTI, A. **SOBEK**: uma Ferramenta de Mineração de Textos. 2007. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Departamento de Informática, Universidade de Caxias do Sul, Caxias do Sul, 2007.

MACEDO, A. L. **Rede de Conceitos**: uma ferramenta para contribuir com a prática pedagógica no acompanhamento da produção textual coletiva. 2010. Tese (Doutorado em Informática na Educação) - PPGIE, UFRGS, Porto Alegre, 2010.

MAKAREMI, I.; LEOEUF, K.; AHMADI, M. Wavelet domain blur invariants for 1D discrete signals. In: KAMEL, M.; CAMPILHO, A. (Org.). ICIAR 2011, **LNCS**, n. 6753, p. 69-79, 2011.

MARRA, R. M.; MOORE, J. L.; KLIMCZAK, A. K. Content Analysis of Online Discussion Forums: A Comparative Analysis of Protocols. **ETR&D**, v. 52, n. 2, p. 23-40, 2004.

MAHALAKSHMI, G. S.; SENDHILKUMAR, S. Automatic Reference Tracking. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 483-499.

MATTAR, J. Interatividade e aprendizagem. In: LITTO, F. M.; FORMIGA, M. M. M. (Org.). **Educação a distância: o estado da arte**. São Paulo: Pearson Education do Brasil, 2009. p. 112-120.

MAURI, T.; ONRUBIA, J. O professor em ambientes virtuais: perfil, condições e

competências. In: COLL, C; MONEREO, C. e colaboradores. *Psicologia da educação virtual: aprender e ensinar com as tecnologias da informação e da comunicação*. Porto Alegre: Artmed, 2010. p. 118-135.

MAZIERO, E. G.; PARDO, T. A. S.; DI FELIPPO, A.; DIAS-DA-SILVA, B. C. A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. In: *Workshop em Tecnologia da Informação e da Linguagem Humana, VI*, 2008, Vila Velha. *Anais...* Vila Velha: XIV Simpósio Brasileiro de Sistemas Multimídia e Web, 2008. p. 390-392.

MAZZOLINI, M.; MADDISON, S. Sage, guide or ghost? The effect of instructor intervention on student participation in online discussion forums. **Computers & Education**, 40, p. 237–253, 2003.

MAZZOLINI, M. When to jump in: The role of the instructor in online discussion forums. **Computers & Education**, n. 49(2), p. 193-213, 2007.

MCCALLUM, A. K. **Bow**: a toolkit for statistical language modeling, text retrieval, classification and clustering. 1996. Disponível em: <<http://www.cs.cmu.edu/~mccallum/bow>>. Acesso em 30/04/2010.

MEYER, K. Evaluating online discussions: four different frames of analysis. **Journal of Asynchronous Learning Networks**, 8(2), p. 101-114, 2004.

MOCHIZUKI, T. et al. Promotion of self-assessment for learners in online discussion using the visualization software. In: *CONFERENCE ON COMPUTER SUPPORT FOR COLLABORATIVE LEARNING*, 2005, Taipei. **Proceedings...** 2005. p. 440-449.

MOISEY, S. D.; HUGHES, J. A. Supporting the Online Learner. In: ANDERSON, T. (Eds.). **Theory and practice of online learning**, Second Edition, 2008. p. 419-439.

MOONEY, R. J.; NAHM, U. Y. Text Mining with Information Extraction. In: *INTERNATIONAL MIDP COLLOQUIUM*, 4., 2003, Bloemfontein. **Proceedings...** Bloemfontein, 2003, p. 141-160.

MOORE, M. G.; KEARSLEY, G. **Educação a distância: uma visão integrada**. São Paulo: Cengage Learning, 2008.

NEWMAN, D. R.; WEBB, B.; COCHRANE, C. A content analysis method to measure critical thinking in face-to-face and computer supported group learning. **Interpersonal Computing and Technology**, 3, p. 56–77, 1995.

NEWMAN, D. R.; JOHNSON, C.; WEBB, B.; COCHRANE, C. Evaluating the quality of learning in computer supported co-operative learning. **Journal of the American Society for Information Science**, 48(6), p. 484-495, 1997.

NISBET, D. Measuring the Quantity and Quality of Online Discussion Group Interaction. **Journal of eLiteracy**, v. 1, p. 122-139, 2004.

NISKIER, A. **Educação a Distância: a tecnologia da esperança**. São Paulo: Loyola, 1999.

NOTARE, M. L. **Comunicação e Aprendizagem Matemática Online**: um estudo com o Editor Científico ROODA Exata. 2009. Tese (Doutorado em Informática na Educação) - PPGIE, UFRGS, Porto Alegre, 2009.

OLIVEIRA, G. P. O fórum em um ambiente virtual de aprendizado colaborativo. São Paulo. PUC-SP. **Revista Digital de Tecnologia Educacional e Educação a Distância**, v. 2, n.1, 2005.

ONRUBIA, J.; COLOMINA, R.; ENGEL, A. Os ambientes virtuais de aprendizagem baseados no trabalho em grupo e na aprendizagem colaborativa. In: COLL, C; MONEREO, C. e colaboradores. **Psicologia da educação virtual**: aprender e ensinar com as tecnologias da informação e da comunicação. Porto Alegre: Artmed, 2010, p. 208-225.

ORENGO, V. M.; HUYCK, C. R. A Stemming Algorithm for the Portuguese Language. In: INTERNATIONAL SYMPOSIUM ON STRING PROCESSING AND INFORMATION RETRIEVAL (SPIRE), 8., 2001, Laguna de San Raphael. **Proceedings...** 2001. p. 183-193.

PALLOFF, R. M.; PRATT, K. **Construindo comunidades de aprendizagem no ciberespaço**. Porto Alegre: Artmed, 2002.

PALLOFF, R. M.; PRATT, K. **O aluno virtual**: um guia para trabalhar com estudantes online. Porto Alegre: Artmed, 2004.

PELZ, B. (My) 3 principles of effective online pedagogy. **Journal of Asynchronous Learning Networks**, 8(3), p. 33-46, 2004.

PENNY, L.; MURPHY, E. Rubrics for designing and evaluating online asynchronous discussions. **British Journal of Educational Technology**, v. 40, n. 5, p. 804-820, 2009.

PERRAUDEAU, M. **Estratégias de aprendizagem**: como acompanhar os alunos na aquisição dos saberes. Porto Alegre: Artmed, 2009.

PERRY, W. G., Jr. **Forms of Ethical and Intellectual Development in the College Years**: A Scheme. San Francisco: Jossey-Bass, 1999.

PURETSKIY, A. A.; SHUTT, G. L.; BERRY, M. W. Survey of text visualization techniques. In: BERRY, M. W.; KOGAN, J. **Text mining**: applications and theory. John Wiley & Sons Ltd, 2010, p. 107-127.

QI, Y. Text Mining in Bioinformatics: Research and Application. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 748-757.

RAMAKRISHNAN, G.; BHATTACHARYYA, P. Question Answering Using Word Associations. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 571-603.

RAMAL, A. C. Avaliar na cibercultura. **Revista Pátio**, Porto Alegre, fev., 2000.

RAMSEY, C. Using virtual learning environments to facilitate new learning relationships. **International Journal of Management Education**, 3(2), p. 31-41, 2003.

RAVI, S.; KIM, J. Profiling Student Interactions in Threaded Discussions with Speech Act Classifiers. In: AI IN EDUCATION CONFERENCE (AIED), 2007, Los Angeles. **Proceedings...**, 2007.

REBEDEA, T.; TRAUSAN-MATU, S.; CHIRU, C. Extraction of Socio-semantic Data from Chat Conversations in Collaborative Learning Communities. DILLENBOURG, P.; SPECHT, M. (eds.). EC-TEL 2008. **LNCS**, Springer-Verlag Berlin Heidelberg, n. 5192, p. 366-377, 2008.

RIGO, S. J., OLIVEIRA, J. P. M. de, BARBIERI, C. Classificação de textos baseada em ontologias de domínio. In: V WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA, 2007, Rio de Janeiro. **Anais...** XXVII Congresso da SBC, 2007.

ROMERO, C.; VENTURA, S. Educational data mining: A survey from 1995 to 2005. **Expert Systems with Applications**, v. 33, p. 135-146, 2007.

ROURKE, L.; ANDERSON, T.; GARRISON, D. R.; ARCHER, W. Assessing social presence in asynchronous text-based computer conferencing. **Journal of Distance Education**, 14(2), 50, 2001.

SALES et al. Indicadores de Aprendizagem Learning Vectors: Uma Aplicação em Fóruns do Ambiente Virtual Moodle. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, XIX, 2008, Fortaleza. **Anais...** Fortaleza: SBIE, 2008.

SALMON, G. **E-Moderating - the key to teaching and learning online**. London: Kogan Page, 2000.

SALTON, G. **The SMART Retrieval System**. Englewood Cliffs, NJ: Prentice-Hall, 1971.

SALTON, G.; MCGILL, M. **Introduction to Modern Information Retrieval**. New York: McGraw-Hill, 1983.

SALTON, G.; BUCKLEY, C. Term Weighting Approaches in Automatic Text Retrieval. **Information Processing and Management**, 24(5), p. 513-523, 1988.

SALTZ, J.; HILTZ, S. R.; TUROFF, M. Student Social Graphs: Visualizing a Student's Online Social Network. In: CSCW, 2004, Chicago, Illinois. **Proceedings...** ACM, 2004. p. 596-599.

SÁNCHEZ, L. P. El foro virtual como espacio educativo: propuestas didácticas para su uso. **Verista Quaderns Digitals Net**, n. 40, p. 1-18, 2005. Disponível em: <http://www.quadernsdigitals.net/datos_web/hemeroteca/r_1/nr_662/a_8878/8878.html>. Acesso em: 27/08/2009.

SCHENKER, A. **Graph-Theoretic Techniques for Web Content Mining**. 2003. PhD Thesis, University of South Florida, Florida, 2003.

SCHLEMMER, E.; FAGUNDES, L. da C. Uma proposta para avaliação de ambientes virtuais de aprendizagem na sociedade em rede. **Informática na educação: teoria e prática**, Porto alegre, v. 4, n. 2, p. 25-36, 2001.

SCHLEMMER, E. **AVA: Um Ambiente Virtual de convivência Interacionista Sistêmico**

para Comunidades Virtuais na Cultura da Aprendizagem. 2002. Tese (Doutorado em Informática na Educação) - PPGIE, UFRGS, Porto Alegre, 2002.

SCHLEMMER, E. Metodologias para educação à distância no contexto da formação de comunidades virtuais de aprendizagem. In: BARBOSA, R. M. (Org.). **Ambientes virtuais de aprendizagem.** Porto Alegre: Artmed, 2005. p. 29-49.

SCHLEMMER, E. Ambiente Virtual de Aprendizagem (AVA): uma proposta para a sociedade em rede na cultura da aprendizagem. In: VALENTINI, C. B.; SOARES, E. M. do S. (Org.). **Aprendizagem em ambientes virtuais:** compartilhando ideias e construindo cenários. Caxias do Sul: Editora da UCS, 2005.

SHEHATA, S.; KARRAY, F.; KAMEL, M. A Concept-based Model for Enhancing Text Categorization. In: KDD, 2007, San Jose. **Proceedings...** ACM, 2007. p. 629-637.

SOMMERVILLE, I. **Engenharia de Software.** 8. ed. São Paulo: Pearson Addison-Wesley, 2007.

SUOMINEN, et al. Performance Evaluation Measures for Text Mining. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies.** Hershey: Information Science Reference, 2009. p. 724-747.

TAN, A. Text Mining: The State of the Art and the Challenges. In: WORKSHOP ON KNOWLEDGE DISCOVERY FROM ADVANCED DATABASES, 1999, Beijing. **Proceedings...** 1999. p. 71-76.

TANE, J.; SCHMITZ, C.; STUMME, G. Semantic resource management for the web: An e-learning application. In: WWW CONFERENCE, 2004, New York. **Proceedings...** 2004. p. 1-10.

TANG, C.; YIN, H.; LI, T.; LAU, R.; LI, Q.; KILIS, D. Personalized courseware construction based on web data mining. In: INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS ENGINEERING, 1., 2000, Washington-DC. **Proceedings...** 2000. p. 204-211.

TELES, L. A aprendizagem por e-learning. In: LITTO, F. M.; FORMIGA, M. M. M. (Org.). **Educação a distância:** o estado da arte. São Paulo: Pearson Education do Brasil, 2009. p. 72-80.

TORRE, C. J. de la; MARTÍN-BAUTISTA, M. J.; SÁNCHEZ, D.; VILA, M. A. Text mining: Intermediate forms for knowledge representation. In: JOINT EUSFLAT/LFA CONFERENCE, 2005, Barcelona. **Proceedings...** MONTSENY, E.; SOBREVILLA, P. (eds.), European Society for Fuzzy Logic and Technology, 2005. p. 1082-1087.

UENO, M. Data mining and text mining technologies for collaborative learning in an ILMS "Samurai". In: IEEE INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 2004, Joensuu. **Proceedings...** IEEE Computer Society, 2004.

ULLMAN, J. R. An algorithm for subgraph isomorphism. **Journal of the Association for Computing Machinery**, v. 23, n. 1, p. 31-42, 1976.

VERGARA, S. Estreitando relacionamentos na educação a distância. **Cadernos EBAPe.br**, edição especial, Rio de Janeiro, v. 5, p. 61-65, 2007.

WARE, P.; KRAMSCH, C. Toward an intercultural stance: Teaching German and English through telecollaboration. **Modern Language Journal**, 89(2), p. 190-205, 2005.

WEVER, B. De; SCHELLENS, T.; VALCKE, M; VAN KEER, H. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. **Computers & Education**, 46, p. 6-28, 2006.

WIVES, L. K. **Um estudo sobre agrupamento de documentos textuais em processamento de informações não estruturadas usando técnica de clustering**. 1999. Dissertação de Mestrado, CPGCC, UFRGS, Porto Alegre, 1999.

XIN, C.; FEENBERG, A. Pedagogy in Cyberspace: The Dynamics of Online Discourse. **Journal of Distance Education**, v. 21, n. 2, p. 1-25, 2006.

XU, S.; LUO, X. Current Issues and Future Analysis in Text Mining for Information Security Applications. In: SONG, M.; WU, Y. (Org.). **Handbook of research on text and web mining technologies**. Hershey: Information Science Reference, 2009. p. 694-707.

YOO, S. et al. Mining Social Networks for Personalized Email Prioritization. In: ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 15., 2009, Paris. **Proceedings...** ACM, 2009. p. 967-975.

APÊNDICE A - PRODUÇÃO CIENTÍFICA

Produções acadêmicas publicadas:

1. AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Análise temática das mensagens de discussões online. *Cadernos de Informática (UFRGS)*, v. 6, p. 47-54, 2011.
2. AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Qualitative Analysis of Discussion Forums. *International Journal of Computer Information Systems and Industrial Management Applications*, v. 3, p. 671-678, 2011.
3. AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Aplicação da mineração de textos em fóruns de discussão. *RENOTE. Revista Novas Tecnologias na Educação*, v. 9, p. 1-10, 2010.
4. AZEVEDO, Breno Fabrício Terra; REATEGUI, E. B.; BEHAR, P. A. Estudo de análise qualitativa em fórum de discussão. *RENOTE. Revista Novas Tecnologias na Educação*, v. 7, p. 1-10, 2009.
5. AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Qualitative Analysis of Discussion Forums. In: *IADIS Multi Conference on Computer Science and Information Systems, 2010, Freiburg. Proceedings of the IADIS International Conference e-Learning 2010, 2010. v. I. p. 251-258.*
6. AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Um software para análise das mensagens de fóruns de discussão. In: *IADIS Ibero-Americana WWW/Internet (CIAWI) Conference, 2011, Rio de Janeiro. Proceedings of the IADIS Ibero-Americana WWW/Internet (CIAWI) Conference, 2011.*
7. AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Automatic analysis of messages in discussion forums. In: *14th International Conference on Interactive Collaborative Learning, 2011, Piešťany. Proceedings of the 14th International Conference on Interactive Collaborative Learning, 2011.*
8. AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Aplicação da mineração de textos na área de educação. In: *Adelson Siqueira Carvalho; Hélvia Pereira Pinto Bastos; Maria Letícia Felicori Tonelli e Teixeira Leite; Rodrigo Garrett da Costa. (Org.). Educação e Tecnologia: um caminho interinstitucional. Campos dos Goytacazes: Essentia Editora, 2011, v. 1, p. 59-73.*
9. AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Análise das mensagens de fóruns de discussão através de um software para mineração de textos. In: *SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, XXII, 2011, Aracaju. Anais... Aracaju: SBIE, 2011.*

Produções acadêmicas aceitas para publicação:

1. MACEDO, A. L.; AZEVEDO, Breno Fabrício Terra; BEHAR, P. A.; REATEGUI, E. B. Acompanhamento da interação e produção textual coletiva através de mineração de textos. *Revista Informática na Educação: teoria & prática*, 2011.

APÊNDICE B – QUESTIONÁRIO APLICADO COM PROFESSORES PARA DIAGNÓSTICO DE INDICADORES DE RELEVÂNCIA

Prezado(a) Professor(a),

Visando desenvolver uma pesquisa que é parte da minha tese de doutorado, gostaria de contar com a sua colaboração para responder o questionário em anexo. Para tanto, é importante assinar abaixo desta mensagem, tomando ciência de que as informações que você fornecer serão tratadas somente para fins de pesquisa e que seu nome, como sujeito da pesquisa, será mantido em sigilo.

Agradeço a sua colaboração e colocamo-me à disposição para quaisquer esclarecimentos.

Atenciosamente,

Breno F. Terra Azevedo

Doutorando em Informática na Educação -
UFRGS

Termo de ciência e concordância

Eu, _____, aceito participar da pesquisa sobre a análise de relevância de fórum de discussão em ambiente online, no processo de ensino e aprendizagem, exclusivamente para fins científicos e acadêmicos.

Campos dos Goytacazes, ____ de _____ de 2010.

Ciente: _____

Em uma discussão realizada em um fórum, no qual participam alunos, tutores e professores, existem alguns fatores que contribuem de forma positiva para um bom debate. Sob a ótica do professor, um bom debate seria aquele no qual os alunos participam, redigindo mensagens que são relevantes para a discussão. Para o docente é importante diagnosticar como, quando, porque, e de que forma cada aluno interagiu no fórum.

Este questionário apresenta dez ações que podem ser realizadas pelo professor para efetuar um diagnóstico da atuação de cada aluno em uma discussão. Cada uma destas ações possui um peso de 0 a 10, onde 0 significa que a mesma não tem importância, e 10 indica que ela possui relevância máxima.

De acordo com sua análise pessoal, gostaria de solicitar que V.Sa. assinalasse um peso entre 0 e 10 para cada uma das ações listadas abaixo.

1. Analisar indicadores quantitativos, como a quantidade de mensagens postadas e a quantidade de palavras digitadas em cada mensagem.

- (0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
2. Investigar se as mensagens postadas referem-se ao tema do debate.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
3. Verificar se as mensagens postadas não possuem erros conceituais.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
4. Verificar se as mensagens postadas não possuem erros gramaticais.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
5. Investigar se existem mensagens cujos textos são muito semelhantes.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
6. Analisar a categoria de cada mensagem. Se ela é uma pergunta, resposta, afirmativa, explicação, interpretação, conflito, concordância, elaboração, correção, agradecimento, motivação.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
7. Verificar a estruturação das mensagens postadas. Isto permite descobrir se uma mensagem originou outras, e a sequência das postagens, para saber qual mensagem foi colocada em função de outra.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
8. Descobrir as pessoas com quem o aluno interagiu dentro do debate.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
9. Considerar aspectos culturais dos alunos envolvidos no debate.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)
10. Investigar fatores afetivos nas mensagens postadas.
(0) (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

Caso exista algum comentário que V.Sa. gostaria de mencionar sobre as ações acima, por favor, descreva-o abaixo.

Caso exista alguma ação que V.Sa. gostaria de citar, além das mencionadas acima, por favor, descreva-a abaixo.

APÊNDICE C - TERMO DE CONSENTIMENTO PARA A ANÁLISE DAS MENSAGENS DOS FÓRUNS

O Núcleo de Tecnologias Digitais Aplicada à Educação (NUTED) coordenado pela Profa. Dra. Patricia Alejandra Behar está realizando estudos sobre a análise qualitativa das mensagens de fóruns de discussão, pesquisa realizada pelo doutorando e pesquisador Breno Fabrício Terra Azevedo.

O MineraFórum é um software que efetua a análise qualitativa das mensagens redigidas em fóruns de discussão. Para a realização de experimentos com a ferramenta será necessário coletar diversas mensagens postadas em fóruns.

As mensagens redigidas pelos alunos no fórum de discussão _____, realizado na disciplina de _____, no Ambiente Virtual de Aprendizagem _____, serão analisadas pelo MineraFórum.

O sigilo dos nomes dos alunos será preservado nos relatórios, com a substituição dos mesmos por siglas ou nomes fictícios. O professor da disciplina não terá acesso aos dados registrados para efeitos de avaliação do aluno.

Agradecemos pela colaboração.

Profa. Dra. Patricia Alejandra Behar - Coordenadora do NUTED

Breno Fabrício Terra Azevedo - Pesquisador

Eu, _____, número de matrícula _____, autorizo a coleta das mensagens que redigi no fórum de discussão _____, realizado na disciplina de _____, no Ambiente Virtual de Aprendizagem _____.

APÊNDICE D – TUTORIAL DA VERSÃO 3.0 DO MINERAFÓRUM

I. Introdução

O MineraFórum é um minerador de textos para fóruns de discussão. A ferramenta realiza um mapeamento dos principais conceitos relacionados ao tema em debate, calculando a relevância de cada mensagem postada no fórum. Através do MineraFórum o professor pode analisar mais rapidamente um fórum de discussão. Com esta análise, o docente pode identificar quais alunos estão discutindo os conceitos relacionados ao tema do debate, quais discentes ainda não apresentaram textos abordando os conceitos e quais apresentaram as contribuições mais relevantes dentro do tópico proposto.

Após acessar o tópico de uma discussão dentro de um Ambiente Virtual de Aprendizagem, e acionar o MineraFórum, os passos para utilização do software são:

- 1.1. Acessar a aba “Texto de referência”.
- 1.2. Escolher um texto que aborde os principais conceitos relacionados ao tópico da discussão do fórum. Este texto de referência poderá ser informado ao MineraFórum de duas formas:
 - a) O texto pode ser digitado na área branca.
 - b) O texto pode ser lido pelo MineraFórum, escolhendo-se o menu “Arquivo”, opção “Importar texto de referência”.
- 1.3. Na aba “Texto de referência”, pode-se pressionar o botão “Extrair conceitos”. Esta opção faz com que o MineraFórum crie uma base dos conceitos mais importantes do texto de referência.
- 1.4. Ao invés de informar um texto de referência para o MineraFórum, pode-se construir manualmente uma base de conceitos. Para construí-lo é necessário:
 - a) Acessar a aba “Base de conceitos”.
 - b) Digitar um conceito na área branca em frente à frase “Novo conceito”, e pressionar o botão “Enviar”.

- c) É possível digitar vários conceitos, um a um.
- d) Os conceitos digitados podem ser relacionados uns aos outros. Para fazer o relacionamento, primeiro escolhe-se um conceito na área “Conceitos”, depois escolhe-se um conceito na área “Conceitos não relacionados”, e então pressiona-se o botão “Adicionar conceito relacionado”. Para desfazer um relacionamento, escolhe-se um conceito na área “Conceitos relacionados”, e pressiona-se o botão “Remover conceito relacionado”.

1.5. Na aba “Base de conceitos”, pode-se pressionar o botão “Gerar grafo” para visualizar o grafo dos conceitos relevantes do texto de referência.

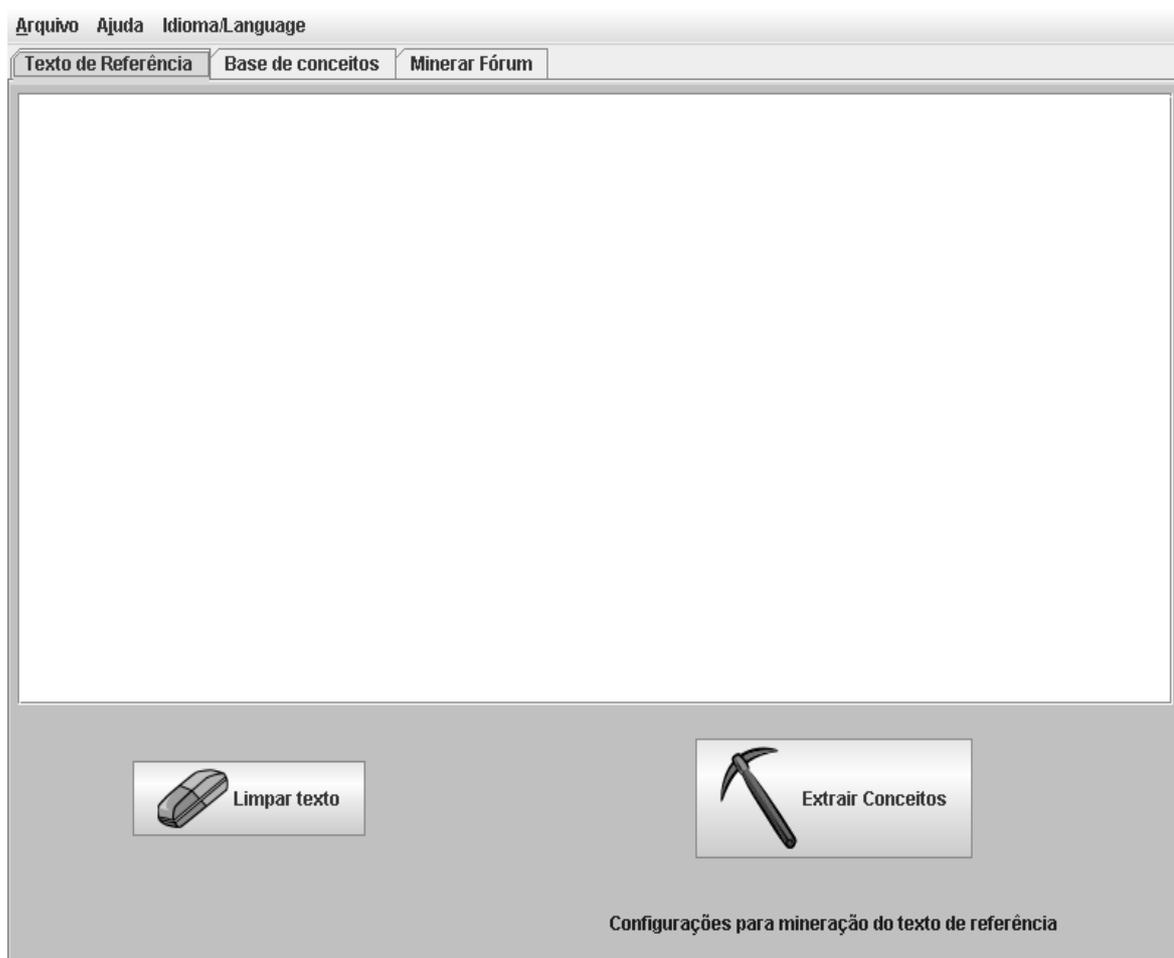
1.6. Na aba “Minerar fórum”, os botões possibilitam os seguintes recursos:

- a) O botão “Minerar fórum” realiza a análise das mensagens postadas pelos alunos no fórum. As informações exibidas são: nome do fórum; data da mineração; total de mensagens redigidas por todos os alunos. Para cada aluno, são exibidos: nome completo; total de mensagens postadas; média da relevância das mensagens. Para cada mensagem, são exibidos o valor da relevância e quatro links ([Info] [Mensagem] [Conceitos texto] [Conceitos fórum]).
 - ➔ Link [Info]: informa a quantidade de vezes que a mensagem foi citada no fórum e se a mensagem é similar a alguma outra.
 - ➔ Link [Mensagem]: exibe os primeiros caracteres da mensagem.
 - ➔ Link [Conceitos texto]: apresenta os conceitos indicados pelo texto de referência que foram encontrados na mensagem.
 - ➔ Link [Conceitos fórum]: exibe os principais conceitos citados no fórum que foram encontrados na mensagem.
- b) O botão “Grafo do fórum” possibilita a visualização do grafo gerado com os principais conceitos citados nas mensagens postadas no fórum.
- c) O botão “Ver similaridades” apresenta as mensagens similares que foram redigidas no fórum.
- d) O botão “Exibir conceitos” permite visualizar quais foram os alunos que citaram cada um dos principais conceitos relacionados ao tema do debate.

- e) O botão “Exibir gráfico” apresenta um gráfico com a média das relevâncias das postagens dos alunos.
- f) O botão “Salvar resultados” possibilita que os resultados da análise do MineraFórum sejam gravados em um arquivo no formato “html”.

II. O MineraFórum possui três abas principais:

1. Texto de referência.



Nesta aba pode-se digitar um texto de referência sobre o tema debatido no fórum de discussão. Este texto pode estar nos formatos: “.txt”, “.pdf” e “.doc”.

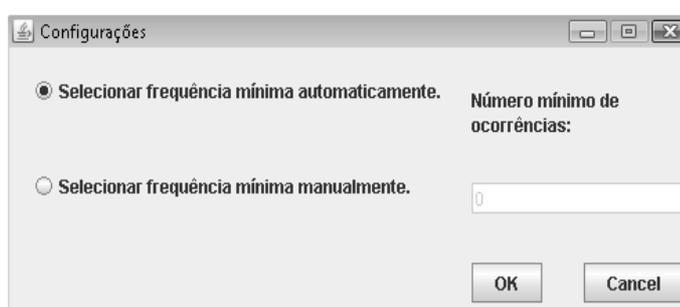
Depois de indicar o texto de referência, pode-se apagá-lo ao pressionar o botão “Limpar texto”.

O MineraFórum constrói uma base de conceitos através do número de ocorrências de cada palavra (que não sejam *stopwords* (palavras consideradas irrelevantes)). Para configurar

o número mínimo de ocorrências das palavras do texto de referência, basta pressionar a mensagem “Configurações para mineração do texto de referência”.

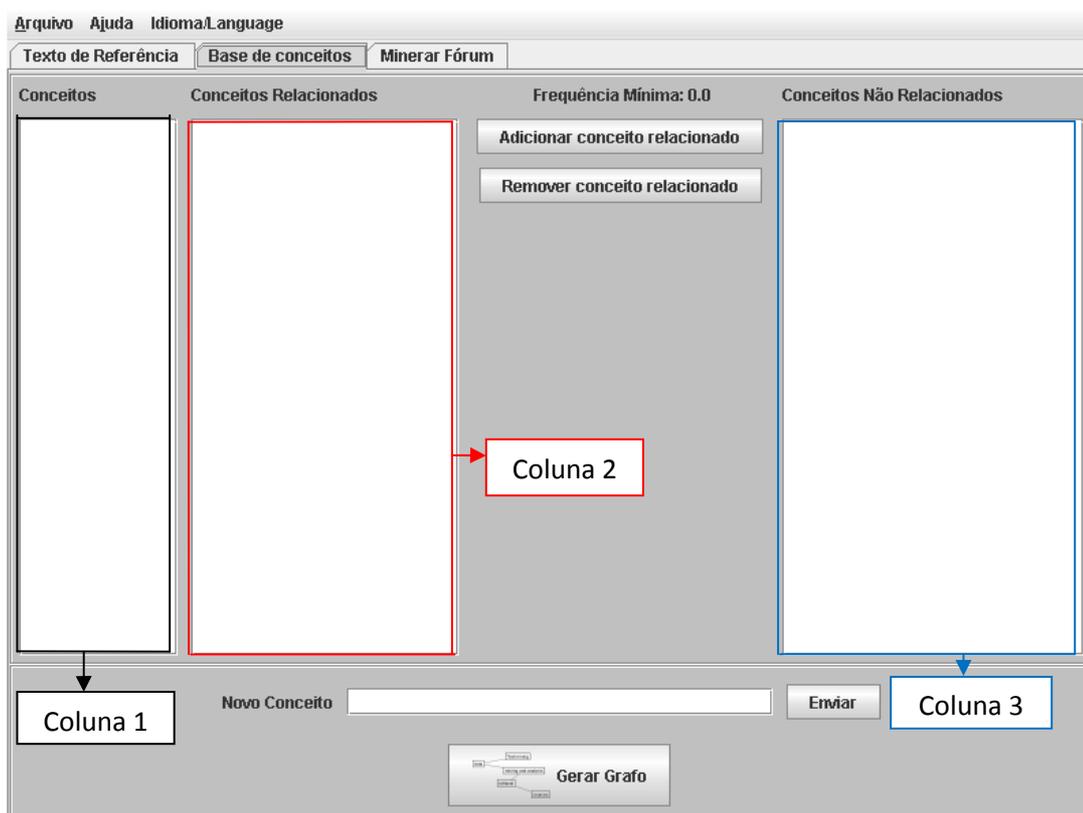


Após pressionar a mensagem, aparece uma nova tela onde é possível manter a frequência mínima de ocorrências no modo automático, ou digitar outro valor.



Ao pressionar o botão “Extrair conceitos”, o MineraFórum constrói uma base dos conceitos que mais ocorreram dentro do texto de referência.

2. Base de conceitos.

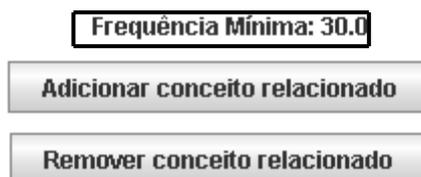


Nesta aba há uma coluna com todos os conceitos extraídos do texto de referência (coluna 1), uma outra coluna ao lado com todos os conceitos relacionados à coluna 1, e outra com todos os conceitos não relacionados à primeira coluna. Pode-se apagar um conceito, selecionando-o e pressionando a tecla “Delete”.

É possível alterar a 2ª ou a 3ª coluna para acrescentar um conceito ou retirar um conceito de uma coluna para outra. Para isto, basta utilizar os seguintes botões:

- “Adicionar conceito relacionado”: pode-se selecionar alguma palavra que está na coluna dos conceitos não-relacionados, e depois pressionar este botão. Assim esta palavra irá aparecer na coluna dos conceitos relacionados.
- “Remover conceito relacionado”: pode-se selecionar alguma palavra que está na coluna dos conceitos relacionados, e depois pressionar este botão. Assim esta palavra irá aparecer na coluna de conceitos não relacionados.

Acima destes botões aparece a frequência mínima dos conceitos relevantes do texto de referência.



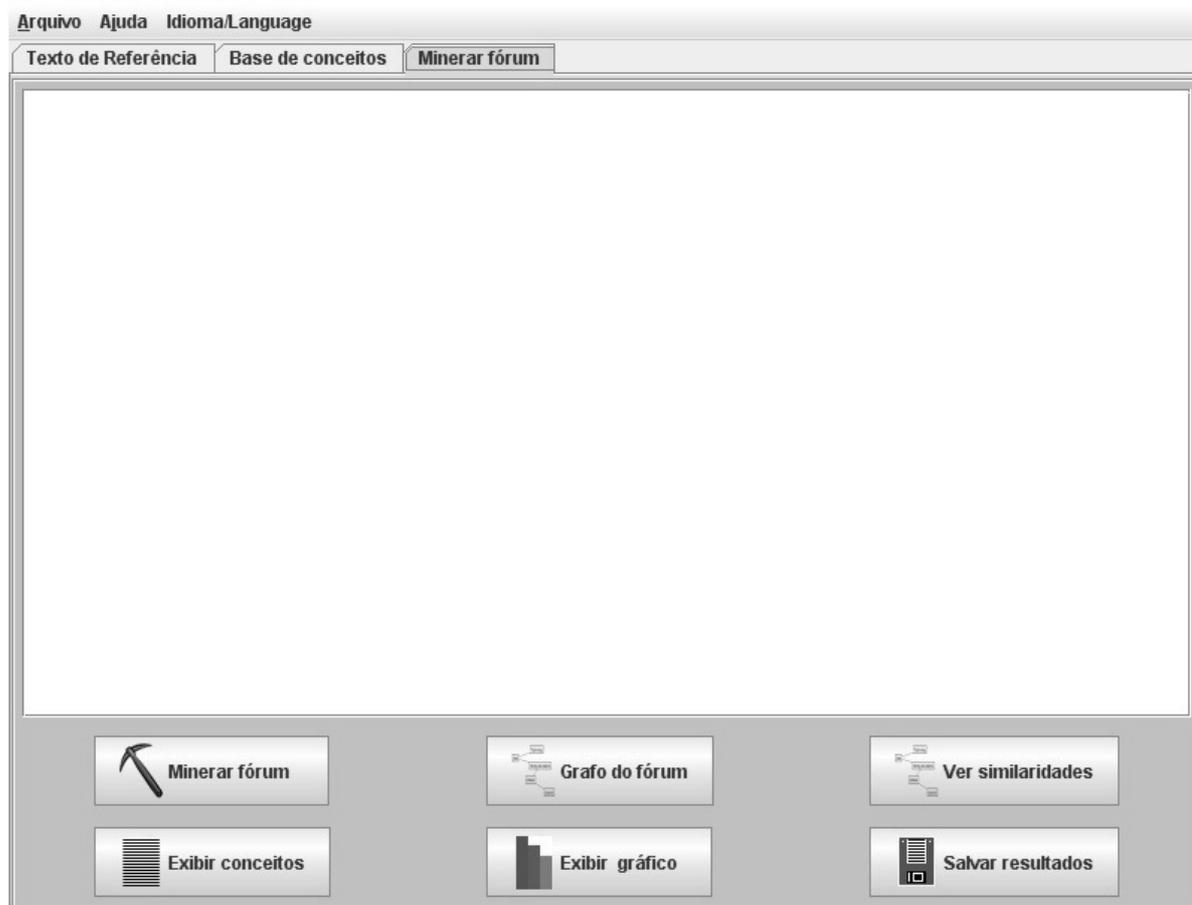
Também há a opção de inserir um novo conceito que não esteja na base. Para isto, basta digitá-lo no campo de texto ao lado da frase “Novo Conceito” e pressionar o botão “Enviar”.



O botão “Gerar Grafo” exibe o grafo com os conceitos relevantes do texto de referência.



3. Minerar fórum.



Esta aba possibilita as seguintes ações:

- O botão “Minerar fórum” realiza a análise das mensagens postadas pelos alunos no fórum.



As informações exibidas são: nome do fórum; data da mineração; total de mensagens redigidas por todos os alunos. Para cada aluno, são exibidos: nome completo; total de mensagens postadas; média da relevância das mensagens. Para cada mensagem, são exibidos o valor da relevância e quatro links ([Info] [Mensagem] [Conceitos texto] [Conceitos fórum]).

- ➔ Link [Info]: informa a quantidade de vezes que a mensagem foi citada no fórum e se a mensagem é similar à outra.
- ➔ Link [Mensagem]: exhibe os primeiros caracteres da mensagem.

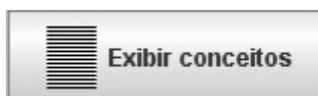
- ➔ Link [Conceitos texto]: apresenta os conceitos indicados pelo texto de referência que foram encontrados na mensagem.
- ➔ Link [Conceitos fórum]: exibe os principais conceitos citados no fórum que foram encontrados na mensagem.
- O botão “Grafo do fórum” possibilita a visualização do grafo gerado com os principais conceitos citados nas postagens do fórum.



- O botão “Ver similaridades” apresenta as mensagens similares que foram redigidas no fórum.



- O botão “Exibir conceitos” permite visualizar quais foram os alunos que citaram cada um dos principais conceitos relacionados ao tema do debate.



- O botão “Exibir gráfico” apresenta um gráfico com a média das relevâncias das postagens dos alunos.



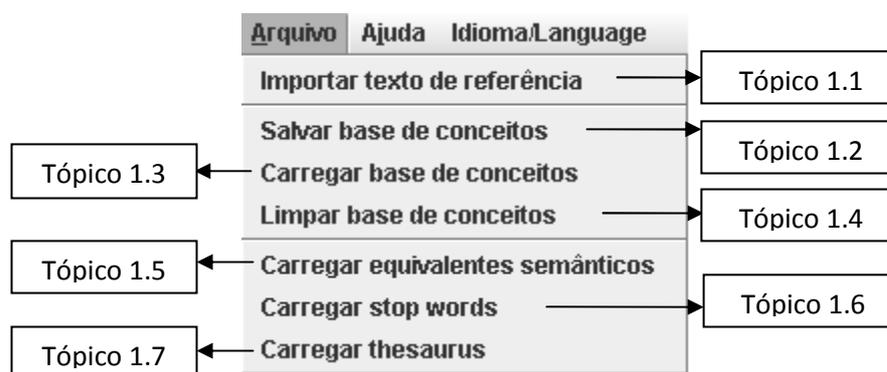
- O botão “Salvar resultados” possibilita que os resultados da análise do MineraFórum sejam gravados em um arquivo no formato “html”.



Além disso, independente destas abas, há opções no menu principal relacionadas à ferramenta, que são:

Arquivo Ajuda Idioma/Language

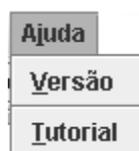
1. “Arquivo”: neste item há opções relacionadas ao texto de referência, à base de conceitos, e alguns itens importantes para a mineração. As opções são:



- 1.1. “Importar texto de referência”: ao selecionar esta opção, aparece uma tela, na qual é possível carregar o texto de referência.
- 1.2. “Salvar base de conceitos”: esta opção permite gravar a base de conceitos gerada pelo MineraFórum (a partir da extração feita do texto de referência), ou que foi digitada. A base de conceitos será armazenada em um arquivo no formato “txt”.
- 1.3. “Carregar base de conceitos”: com esta opção pode-se carregar uma base de conceitos que foi previamente gravada.
- 1.4. “Limpar base de conceitos”: ao selecionar esta opção, os conceitos serão excluídos da base que está sendo utilizada.
- 1.5. “Carregar equivalentes semânticos”: com esta opção pode-se carregar um arquivo contendo uma lista de palavras que possuem equivalência semântica. Este recurso é opcional.
- 1.6. “Carregar *stopwords*”: com esta opção pode-se carregar um arquivo contendo uma lista de *stopwords*, isto é, palavras que são irrelevantes para a mineração de textos (palavras que não são tão importantes para determinar o conteúdo e o sentido do texto). Este recurso é opcional, uma vez que o MineraFórum já possui uma lista de *stopwords* interna.
- 1.7. “Carregar thesaurus”: com esta opção pode-se carregar um arquivo contendo um *thesaurus*. Este recurso é opcional, uma vez que o MineraFórum já possui um *thesaurus* interno. Caso o usuário não indique, o *thesaurus* a ser empregado pelo MineraFórum utilizará a base de dados TEP 2.0, proposta por Maziero et al. (2008), Dias-da-Silva et al. (2000), Dias-da-Silva e Moraes (2003). O TeP 2.0 é uma extensão do TeP (Thesaurus Eletrônico para o Português do Brasil), que foi desenvolvido segundo os pressupostos da WordNet de Princeton (MAZIERO et

al., 2008). Disponível para *download* em <http://www.nilc.icmc.usp.br/tep2/download.htm>

2. “Ajuda”: neste item há outras opções relacionadas às informações sobre a ferramenta, como a sua versão e este tutorial. Estas opções são:



- “Versão”: apresenta informações sobre a versão do software.
- “Tutorial”: exhibe este tutorial.

3. “Idioma/Language”: nesta opção é possível escolher o idioma para utilizar a ferramenta.



III. Alguns arquivos utilizados pelo MineraFórum necessitam de uma sintaxe específica, como:

1. O arquivo que contém as mensagens do fórum será passado como parâmetro para o applet do MineraFórum. Este arquivo deve possuir a seguinte sintaxe:

#NomeForum#[50]

Nome do fórum de discussão

#Aluno#

Nome do aluno que escreveu a mensagem abaixo.

#Mensagem#[1,5]

Mensagem redigida pelo aluno, cujo nome foi citado acima.

#Aluno#

Nome do aluno que escreveu a mensagem abaixo.

#Mensagem#[2,0]

Mensagem redigida pelo aluno, cujo nome foi citado acima.

Obs.:

#NomeForum#[50]

Significa que 50 mensagens foram postadas no fórum, incluindo alunos, professores e tutores.

#Mensagem#[1,5]

Significa que a mensagem possui código 1 e foi citada 5 vezes no fórum

#Mensagem#[2,0]

Significa que a mensagem possui código 2 e foi citada 0 vezes no fórum

Obs.:

1) Caso não exista a possibilidade de descobrir a quantidade de mensagens postadas no fórum, deve-se colocar apenas os colchetes, da seguinte forma:

#NomeForum#[]

2) Caso não exista a possibilidade de descobrir o código da mensagem nem do número de vezes que a mensagem foi citada no fórum, deve-se colocar apenas os colchetes e a vírgula, da seguinte forma:

#Mensagem#[,]

2. O arquivo que contém as *stopwords* pode ter várias linhas, onde cada uma indica uma palavra que é considerada como *stopword*.
3. O arquivo que contém o thesaurus deve possuir a seguinte sintaxe:
numero_linha. [tipo do termo] {termo1, termo2, ... , termoN}

Exemplo:

1. [Verbo] {exagerar, exceder, quinta-essenciar, rebuscar, refinar, requintar}
4. O arquivo que contém os equivalentes semânticos pode ter várias linhas, onde cada uma indica os termos que possuem equivalência semântica. Os termos são separados por vírgula e colocados entre chaves. O formato de cada linha do arquivo é:

{termo1, termo2, ... , termoN}

5. O MineraFórum está configurado para trabalhar com “arquivos texto” no formato UTF-8, mas esta configuração pode ser alterada. Todos os “arquivos texto” a serem carregados no MineraFórum devem estar no formato que foi configurado para a ferramenta.
6. Para executar o applet do MineraFórum, a página html deve informar quatro parâmetros:
 - forum: localização do arquivo que contém as mensagens extraídas do fórum
 - tutorialportugues: localização do arquivo do tutorial em português
 - tutorialingles: localização do arquivo do tutorial em inglês
 - charsetArquivos: codificação de caracteres dos arquivos que serão utilizados pelo MineraFórum

Exemplo:

```
<applet code="gui/Web" archive="S_TextMining.jar" width=800 height=600>
```

Your browser doesn't support applets! Try another one.

```
<param name=forum value='<?=$linkArquivoForum?>'>
```

```
<param name=tutorialportugues value='tutorial_mineraforum_portugues.pdf'>
```

```
<param name=tutorialingles value='tutorial_mineraforum_ingles.pdf'>
```

```
<param name=charsetArquivos value='UTF-8'>
```

```
</applet>
```

APÊNDICE E – INSTALAÇÃO DA VERSÃO 3.0 DO MINERAFÓRUM

Os arquivos do MineraFórum estão disponíveis para *download* em www.nie.iff.edu.br/mineraforum.

1. Procedimentos para instalação no ROODA.

a) Criar um diretório chamado `mineraforum` (com permissões `rxwxrwx`) dentro de `/home/www/rooda/forum`

Dentro deste diretório serão gravados arquivos texto. Cada arquivo contém as mensagens de um tópico de um fórum.

b) Copiar o arquivo `mineraforum.gif` (ícone do MineraFórum) para o diretório `/home/www/rooda/images/meta/botoes`

c) No arquivo `topico.php` (diretório `/home/www/rooda/forum`), antes da linha:

```
<a href='forum.php?f=<?=$codForum?>'><img src='<?=$pagina-
>skin_loc?>/botoes/voltar.gif' border=0></a>
```

Adicionar as linhas abaixo:

```
<?
$select1 = "SELECT * FROM Forum WHERE codForum = '$codForum'";
$result1 = db_busca($select1);
if (count($result1) != 0){
    foreach ($result1 as $r1){
        $codTurma = $r1['codTurma'];
    }
    if ($codTurma != NULL){
        $select1 = "SELECT associacao as tipousuario FROM TurmaUsuario
WHERE codTurma = $codTurma AND codUsuario = $codUsuario";
        $mensagem1=db_busca($select1);
        if (count($mensagem1) != 0){
```

```

        foreach ($mensagem1 as $aux1){
            $tipoUsuario=$aux1[tipousuario];
        }
        if (($tipoUsuario == "P") || ($tipoUsuario == "M")){
            echo "<a
href='mineraforum.php?t=".$codTopico."'><img src='".$pagina-
>skin_loc."/botoes/mineraforum.gif' border=0></a>";
        }
    }
}
else {
    $select1 = "SELECT associacao as tipousuario FROM TurmaUsuario
WHERE codUsuario = $codUsuario";
    $mensagem1=db_busca($select1);
    if (count($mensagem1) != 0){
        foreach ($mensagem1 as $aux1){
            $tipoUsuario=$aux1[tipousuario];
        }
        if (($tipoUsuario == "P") || ($tipoUsuario == "M")){
            echo "<a
href='mineraforum.php?t=".$codTopico."'><img src='".$pagina-
>skin_loc."/botoes/mineraforum.gif' border=0></a>";
        }
    }
}
?>

```

d) Copiar os arquivos `mineraforum.php`, `S_TextMining.jar`, `applet.policy`, e `texto_apresentacao_mineraforum.html` para o diretório `/home/www/rooda/forum`

e) Copiar os arquivos `tutorial_mineraforum_portugues.pdf` e `tutorial_mineraforum_ingles.pdf` para o diretório `/home/www/rooda/forum/mineraforum`

2. Procedimentos para instalação no ETC.

a) Criar um diretório chamado mineraforum (com permissões rwxrwxrwx) dentro de /home/www/nuted/etc2/forum

Dentro deste diretório serão gravados arquivos texto. Cada arquivo contém as mensagens de um tópico de um fórum.

b) No arquivo mensagens.php (diretório /home/www/nuted/etc2/forum), após as linhas abaixo:

```
<div id="criar" name="criar" style="display:none;">
<form name="cria_mensagem" action="cria_mensagem.php" method="post">
    <textarea name="conteudo" style="width: 760px; height:
100px;"></textarea>
    <input type="hidden" name="topico_id" value="<?=$topico_id?">
    <input type="submit" value="Criar">
    <input type="button" name="Cancelar" value="Cancelar"
onclick="esconde_div('criar') ">
</form>
</div>
</div>
```

Acrescentar a linha:

```
<?php echo "<a href='mineraforum.php?topico_id=" . $topico_id . "'
target='_blank'>MineraForum</a>"?>
```

c) Copiar os arquivos mineraforum.php, S_TextMining.jar, applet.policy, texto_apresentacao_mineraforum.html para o diretório /home/www/nuted/etc2/forum

d) Copiar os arquivos tutorial_mineraforum_portugues.pdf e tutorial_mineraforum_ingles.pdf para o diretório /home/www/nuted/etc2/forum/mineraforum

3. Procedimentos para instalação no MOODLE.

a) Criar um diretório chamado mineraforum (com permissões rwxrwxrwx) dentro de /var/www/moodle/mod/forum

Dentro deste diretório serão gravados arquivos texto. Cada arquivo contém as mensagens de um tópico de um fórum.

b) No arquivo discuss.php (diretório /var/www/moodle/mod/forum):

Após a linha abaixo:

```
require_once('../../config.php');
```

Adicionar as linhas:

```
// ***** Used by mineraforum *****
require_once($CFG->dirroot.'/user/profile/lib.php');
require_once($CFG->dirroot.'/tag/lib.php');
// *****
```

Após a linha abaixo:

```
require_course_login($course, true, $cm);
```

Adicionar as linhas:

```
// ***** Used by mineraforum *****
// Find the roles of the user in the course
if ($course->id == SITEID) {
    $coursecontext = get_context_instance(CONTEXT_SYSTEM); //
SYSTEM context
} else {
    $coursecontext = get_context_instance(CONTEXT_COURSE,
$course->id); // Course context
}
$showmineraforum = false;
if ($rolestring = get_user_roles_in_context($discussion->userid,
$coursecontext)) {
    $existrole = $rolestring;
    while (strpos($existrole,"roleid=")) {
```

```

$existrole = strstr($existrole,"roleid=");
$existrole = substr($existrole,7);
$posaspas = strpos($existrole,'');
$coderole = substr($existrole,0,$posaspas);

if (($coderole == 1) || ($coderole == 2) || ($coderole ==
3) || ($coderole == 4))

    $showmineraforum = true;

    }

}

// *****

```

Após a linha abaixo:

```

echo popup_form("$CFG->wwwroot/mod/forum/", $forummenu, "forummenu",
"", get_string("movethisdiscussionto", "forum"), "", "", true,'self','',
NULL, get_string('move'));

```

Adicionar as linhas:

```

// ***** Used by mineraforum *****

if ($showmineraforum) {

    echo "<a href='mineraforum.php?d=".$d."'><img
src='mineraforum.png'></a>";

}

// *****

```

c) Copiar os arquivos `mineraforum.php`, `mineraforum.png`, `S_TextMining.jar`, `applet.policy`, `texto_apresentacao_mineraforum.html` para o diretório `/var/www/moodle/mod/forum`

d) Copiar os arquivos `tutorial_mineraforum_portugues.pdf` e `tutorial_mineraforum_ingles.pdf` para o diretório `/var/www/moodle/mod/forum/mineraforum`