

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**

Luiz Claudio Duarte Dalmolin

**EXTRATOR DE TERMOS PARA CRIAÇÃO DE MAPAS
CONCEITUAIS**

**Florianópolis
2010**

LUIZ CLAUDIO DUARTE DALMOLIN

**EXTRATOR DE TERMOS PARA CRIAÇÃO DE MAPAS
CONCEITUAIS**

Dissertação submetida à Universidade Federal de Santa Catarina como requisito parcial exigido pelo Programa de Pós-Graduação em Ciência da Computação para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Rogério Cid Bastos.

Co-orientadora: Prof.^a Dra. Sílvia Modesto Nassar.

Florianópolis, fevereiro de 2010.

Catálogo na fonte pela Biblioteca Universitária da
Universidade Federal de Santa Catarina

D148e Dalmolin, Luiz Claudio Duarte
Extrator de termos para criação de mapas
conceituais
[dissertação] / Luiz Claudio Duarte Dalmolin ;
orientador,
Rogério Cid Bastos. - Florianópolis, SC, 2010.
118 p.: il., graf., tabs., mapas
Dissertação (mestrado) - Universidade Federal de
Santa
Catarina, Centro Tecnológico. Programa de Pós-

LUIZ CLAUDIO DUARTE DALMOLIN

**EXTRATOR DE TERMOS PARA CRIAÇÃO DE MAPAS
CONCEITUAIS**

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, Área de Concentração Engenharia de *Softwares* e Banco de Dados, e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Mário Dantas (Coordenador do PPGCC)

Banca Examinadora

Dr. Rogério Cid Bastos (orientador),UFSC

Dra. Sílvia Modesto Nassar (co-orientadora),UFSC

Dr. Fernando Álvaro Ostuni Gauthier,UFSC

Dr. Mauro Roisenberg,UFSC

Este trabalho é dedicado aos meus pais, que sempre me incentivaram a buscar conhecimento, fornecendo a estrutura necessária para que conseguisse atingir meus principais objetivos.

Agradeço meus primos Márcio e Fabiano pela morada e companhia desde o primeiro momento que vim para Florianópolis.

À minha esposa que durante os meus estudos sempre me apoiou e entendeu minhas ausências, obrigado pela companhia nas diversas horas de trabalho.

Agradeço a meus orientadores Rogério e Silvia pelo conhecimento compartilhado e a ajuda que me deram sempre que foi necessário.

Agradeço a toda a equipe do LABTRANS que me acolheu e me aturou durante todo o tempo em que estive cursando a Pós-Graduação, grato pelas colaborações e as oportunidades que me ofereceram, inclusive pelo apoio financeiro durante o período.

Agradeço aos colegas de mestrado que em conjunto pudemos identificar rumos e resolver dúvidas comuns em nossos trabalhos.

RESUMO

Este trabalho consiste no desenvolvimento de um método para extração de termos que podem ser utilizados para a criação de mapas conceituais. O método é fundamentado em técnicas linguísticas e estatísticas para a extração de palavras-chave que podem ser usadas para originar esses tipos de mapas. Parte do método compreende um algoritmo de mineração de texto que, a partir de um conjunto de documentos selecionados pelo usuário, é capaz de extrair termos que podem ser qualificados como conceitos. Esse algoritmo pode também obter conectores semânticos entre conceitos, formando uma proposição válida sobre o tema dos documentos cadastrados. Para aplicar as funcionalidades propostas no método, foi concebida a ferramenta MapXtractor. Esta ferramenta implementa o método de extração, permite a edição de mapas conceituais, a vinculação de objetos de aprendizagem aos conceitos e o controle de acesso aos mapas. Os resultados obtidos com a utilização desse método demonstram quantitativamente e qualitativamente a superioridade da utilização de técnicas híbridas de extração de termos em relação às técnicas estatísticas para extração de termos, quando se trata da criação de mapas conceituais.

Palavras-chave: mapas conceituais; mineração de textos; extração de termos

ABSTRACT

This work consists in a development of a term extraction method to create conceptual maps. This method is based in linguistic and statistics techniques to extract keywords that can be used to build conceptual maps. A part of this method is compose by a textmining algorithm which is able to extract concepts from a set of documents, selected or created by an user. This algorithm also can extract semantic connectors between two concepts becoming a valid proposition in the documents' subject . To apply the features propose on the method a tool called MapXtractor was developed. This tool implements the extraction method, allows editing concept maps and link learning objects to concepts, and the access controls to the concept maps. The results show the qualitative and quantitative superiority of hybrid techniques over statistical techniques to term extraction when applied to concept map building.

Key-words: concept maps; textmining; term extraction

LISTA DE FIGURAS

Figura 1- Mapa Conceitual da Organização do Trabalho.....	32
Figura 2 - Etapas de mineração de texto	35
Figura 3- Servidores e clientes do cmap tools espalhados pelo planeta	52
Figura 4- Modelo de arquivo CXL.....	53
Figura 5- Exemplo de mapa conceitual (IHMC, 2009).....	55
Figura 6- Interface do CMAP-tools.....	62
Figura 7- Interface do <i>software</i> inspiration	63
Figura 8- Interface do Verified Concept Map (VCM)	64
Figura 9- Processo de extração de conceitos e relações semânticas.....	68
Figura 10- Cadastro de Documento no MAPXTRACTOR.....	70
Figura 11- Interface para edição de parâmetros	73
Figura 12- Trecho de texto para exemplo	76
Figura 13- Procedimento para encontrar conexões semânticas.....	78
Figura 14- Tabela de conexões semânticas	80
Figura 15- Atividades dos usuários no sistema	81
Figura 16- Caso de uso para o professor ao utilizar a ferramenta	82
Figura 17- Diagrama de seqüência das ações realizadas pro professores e os módulos do sistema	83
Figura 18- Diagrama de seqüência das ações realizadas pelos professores ao utilizar o extrator de termos.....	84
Figura 19- Interface de Login.....	85
Figura 20- Interface para Criação e Edição de Cursos	85
Figura 21- Interface para criação e edição de informações sobre mapas conceituais.....	86
Figura 22- Interface para edição e visualização de mapas conceituais .	87

Figura 23- Interface para edição de conceito	88
Figura 24- Interface de apresentação de resultados do extrator de conceitos	89
Figura 25- Interface de exibição e seleção de conexões semânticas extraídas	90
Figura 26- Comparação de <i>Rankings</i> para Unigramas	99
Figura 27- Comparação de <i>Rankings</i> para Bigramas	100
Figura 28- Comparação de <i>Rankings</i> para Trigramas	100
Figura 29- Mapa extraído automaticamente pelo MapXtractor	103

LISTA DE QUADROS

Quadro 1 - Relações de Medidas Estatísticas e as Dimensões envolvidas na Extração de Termos.....	40
Quadro 2 - Lista de Trabalhos Correlatos	48
Quadro 3 - Dados comparativos de trabalhos analisados	49
Quadro 4 - Comparação entre ferramentas para edição de mapas conceituais.....	66
Quadro 5 - Lista de Verbos de Ligação.....	71
Quadro 6 - Lista de Verbos de Ligação Adicionais	72
Quadro 7 - Lista de N-Gramas Extraídos	76
Quadro 8 - Lista filtrada de N-gramas.....	77
Quadro 9 - Documentos selecionados a partir do GOOGLE	93
Quadro 10 - Documentos selecionados a partir do BING	93
Quadro 11 - Documentos selecionados a partir do Yahoo <i>Search</i>	93

LISTA DE TABELAS

Tabela 1- Lista de unigramas extraídos no cenário de teste.....	96
Tabela 2- Lista de bigramas extraídos no cenário de teste.....	97
Tabela 3- Lista de Trigramas obtidos no cenário de teste.....	98
Tabela 4- Proposições extraídas pelo MapXtractor.....	101

SUMÁRIO

1 INTRODUÇÃO	23
1.1 Objetivo Geral	25
1.2 Objetivos Específicos	25
1.3 Justificativa	26
1.4 Escopo do Trabalho	29
1.5 Organização do Trabalho	30
2 TÉCNICAS PARA EXTRAÇÃO E IDENTIFICAÇÃO DE CONCEITOS A PARTIR DE DOCUMENTOS	31
2.1 Data Mining e Text Mining	31
<i>2.1.1 Aplicações práticas de Data Minings e Text Minings</i>	32
2.2 Extração de Termos e Palavras-Chave	34
<i>2.2.1 Abordagem Linguística</i>	36
<i>2.2.2 Abordagem estatística</i>	37
<i>2.2.3 Abordagem de aprendizagem de máquina</i>	38
<i>2.2.4 Abordagem híbrida</i>	39
2.3 Ranking de Termos	39
2.4 Pré-processamento	40
<i>2.4.1 Tokenização</i>	40
<i>2.4.2 Capitalização</i>	41
<i>2.4.3 Tratamento de Stopwords</i>	42
<i>2.4.4 Identificação de Expressões e Termos Compostos</i>	43
<i>2.4.5 Tratamento de Plurais e Gênero</i>	43
<i>2.4.6 Separação de Sentenças</i>	44
2.5 Análise Comparativa de Ferramentas Similares	44
<i>2.5.1 Lista de Trabalhos Analisados</i>	46

2.6 Concept map mining	48
2.7 Modelos de Armazenamento de Mapas Conceituais – CXL	50
3 MAPAS CONCEITUAIS	52
3.1 Definições	52
3.1.1 <i>Vantagens na utilização de Mapas Conceituais no Processo de Ensino-Aprendizagem</i>	56
3.1.2 <i>Desvantagens na utilização de Mapas Conceituais no Processo de Ensino-Aprendizagem</i>	56
3.2 Mapas Conceituais como Objeto de Aprendizagem	57
3.3 Análise Comparativa de Ferramentas Similares	58
3.4 Ferramentas para construção de mapas conceituais	58
3.4.1 <i>CmapTools</i>	59
3.4.2 <i>Inspiration</i>	60
3.4.3 <i>Verified Concept Map</i>	63
3.4.4 <i>Quadro comparativo entre sistemas de construção de mapas conceituais</i>	63
4 EXTRATOR DE TERMOS PARA CRIAÇÃO DE MAPAS CONCEITUAIS EDUCACIONAIS	65
4.1 Método para Extração de Termos Relevantes e Conectores	65
4.1.1 <i>Formalização do problema</i>	66
4.1.2 <i>Parâmetros de Entrada</i>	67
4.2 Método de extração de conceitos	72
4.2.1 <i>Pré-processamento dos textos</i>	72
4.2.2 <i>Parsing dos Documentos</i>	73
4.2.3 <i>Ranking de Conceitos</i>	75
4.3 O extrator de conexões semânticas	76
5 MAPXTRACTOR	78

5.1 Projeto do Sistema	79
5.2 Interfaces do Sistema e Funcionalidades	82
5.2.1 <i>Criação e controles de cursos</i>	83
5.2.2 <i>Edição de Objetos do Mapa</i>	85
5.2.3 <i>Vinculando objetos de aprendizagem a um conceito</i>	86
5.2.4 <i>Módulo de extração de termos e a apresentação dos resultados</i>	86
6 TESTES E VALIDAÇÃO DO TRABALHO	89
6.1 Metodologia de Avaliação e Resultados Esperados	89
6.2 Cenário de teste para modelo proposto	90
6.2.1 <i>Análise Quantitativa dos Resultados</i>	92
6.2.2 <i>Análise Qualitativa</i>	101
6.3 Discussões sobre os Resultados	102
6.4 Limitações	104
7 CONCLUSÕES	105
7.1 Trabalhos Futuros	106
REFERÊNCIAS BIBLIOGRÁFICAS	107
APÊNDICE A – MODELAGEM SIMPLIFICADA DO SISTEMA	116
APÊNDICE B - MODELO E-R	119

1 INTRODUÇÃO

O processo de ensino-aprendizagem vem sofrendo grandes transformações desde que se passou a utilizar o computador como ferramenta didática. O uso de computadores em salas de aula criou novas perspectivas para a evolução de métodos aplicados nesse processo, como por exemplo, os mapas conceituais.

O mapa conceitual é um recurso que tem sido utilizado com eficácia no ensino presencial e à distância, ampliando as possibilidades para a colaboração entre professores e estudantes na construção do conhecimento (TSENG et al, 2007).

Os mapas conceituais organizam de maneira concisa os conceitos-chaves que precisam ser trabalhados em um curso. A organização hierárquica dos mapas sugere um encadeamento lógico dos conceitos, evidenciando a contribuição que um mapa conceitual pode trazer se for utilizado como representação nesse contexto (NOVAK; CAÑAS, 2006).

Conforme Tseng et al (2007), o uso de mapas conceituais em ambientes de aprendizagem pode demonstrar como o aprendizado de um conceito é influenciado pelo conhecimento prévio no domínio. Esses autores destacam a dificuldade e o tempo necessário para a construção de mapas conceituais adequados para serem utilizados em um curso.

Mapas conceituais surgem como uma forma de representação de conhecimento que busca um aprimoramento dos métodos a fim de organizar conteúdos para a criação de cursos para aulas presenciais ou à distância. O uso deste tipo de objeto de aprendizagem no dia-a-dia dos professores e estudantes torna-se mais fácil quando é utilizada uma ferramenta computacional específica para sua criação.

Mesmo tendo fundamentos na área pedagógica, os mapas conceituais podem ser usados para diferentes aplicações que sugerem uma forma de representar o conhecimento. Esse tipo de representação pode ser utilizado para demonstrar relações semânticas entre termos-chave de um domínio.

A definição de um conceito pode ser apoiada por objetos de aprendizagem vinculados aos conceitos inseridos nos mapas. Estes objetos se tornam uma fonte para a expansão do conteúdo de referência do conceito relacionado. O uso destas representações permite que professores estimulem significativamente a aprendizagem por parte dos

estudantes através de materiais de referência indicados pelos próprios professores.

Com o objetivo de facilitar a criação de mapas conceituais, Tseng et al (2007) propõem abordagens para construí-los automaticamente, utilizando ferramentas computacionais associadas a técnicas lingüísticas e estatísticas. Um dos recursos que pode auxiliar o processo de criação de mapas conceituais é a automatização da extração de conceitos.

A extração automática de conceitos pode ser basear na ocorrência dos termos no texto, bem como na noção de proximidade entre eles. Utilizando-se de princípios de mineração de textos e processamento de linguagem natural, é possível identificar quais conceitos são relevantes - conceitos compostos por mais de uma palavra - e encontrar ambigüidades em diferentes contextos (VALERIO; LEAKE, 2006).

Dessa forma, a utilização de uma lista de conceitos e conexões semânticas, extraídos automaticamente, com a supervisão do usuário, a partir de um conjunto de documentos de referência, facilita o processo de construção de um mapa conceitual.

De posse de ferramentas automatizadas para a construção de mapas conceituais, professores podem utilizar, com maior frequência em suas aulas, subsídios didáticos oferecidos por esses mapas.

Enquanto as ferramentas de busca na *web* evoluem, a quantidade de conteúdos disponibilizados na rede aumenta exponencialmente. Ao realizar uma consulta sobre um tema a ser abordado em um curso, é possível obter uma infinidade de recursos que podem ou não servir como referência para a sua elaboração.

Ao lidar com um conjunto de documentos textuais de diferentes fontes, o professor realiza um esforço intelectual para identificar quais conceitos têm maior relevância para serem trabalhados em uma aula. Entre outros fatores, o esforço empregado depende do quanto o usuário conhece sobre o assunto e quais fontes de conteúdo foram pesquisadas.

Considerando as limitações humanas para processar e armazenar informações de um texto sobre um dado assunto, observa-se que a tarefa de criação de um mapa conceitual é complexa e demanda tempo, esforço intelectual do professor e o domínio de uma ferramenta computacional.

Com o intuito de facilitar o trabalho dos docentes na elaboração de seus cursos e promover uma forma de aprendizagem expressiva por parte dos estudantes, este trabalho propõe um método para automatizar o processo de extração de termos relevantes de um determinado domínio do conhecimento.

A partir de um conjunto de documentos textuais conhecidos pelo professor, uma ferramenta computacional é capaz de identificar relações semânticas entre termos relevantes.

De posse do resultado do processo de extração de termos e relações semânticas, é possível construir um mapa conceitual. Esse mapa pode representar o currículo de um curso, um tema a ser abordado em aula, ou ainda, o conhecimento prévio de um conjunto de estudantes.

1.1 Objetivo Geral

O presente trabalho tem o objetivo de desenvolver um método para realizar a construção de mapas conceituais tendo como referência um conjunto de documentos de texto.

1.2 Objetivos Específicos

Os objetivos específicos deste estudo são:

- incorporar ao método a possibilidade de inserir um conjunto de textos de referência avaliados pelo usuário;
- criar um algoritmo para definir a relevância de conceitos e conectores semânticos, considerando um valor de relevância para os documentos usados no processo de extração;
- criar uma ferramenta (*software* protótipo) para executar o método proposto adicionando funcionalidades que facilitem a colaboração e controle de cursos;
- avaliar o método desenvolvido de forma qualitativa e quantitativa;
- apresentar um modelo de boas práticas de uso da ferramenta.

1.3 Justificativa

Tanto para a elaboração de um curso quanto para a fixação de um conteúdo ou representação de idéias, é necessário definir quais conceitos são mais relevantes para proporcionar uma compreensão significativa do tema abordado ou estudado.

Sendo assim, é interessante buscar formas de se mensurar a relevância de um conceito em textos de referência indicados em uma ementa de um curso ou sugerido para uma aula. Segundo Ausubel et al (1980), é esperado que, ao trabalhar conceitos relevantes e previamente conhecidos, seja possível ancorar novas apreciações à estrutura cognitiva do estudante.

Conforme Valerio e Leake (2006), os documentos *online* disponibilizam uma vasta gama de documentos que podem ser utilizados como referencial teórico para a criação de cursos. Todavia, a análise desses documentos consome um tempo considerável dos professores durante o processo de concepção de cursos.

Nesse contexto, existe a possibilidade de se construir métodos que permitam extrair automaticamente os conceitos relevantes de um conjunto de documentos. De posse desses conceitos, os professores podem trabalhar o conhecimento contido nesses documentos representando-o sob a forma de um mapa conceitual.

Pesquisas recentes como as de Villalon e Calvo (2009) mostram que técnicas de extração de termos vêm sendo utilizadas no âmbito educacional tanto para identificação de conhecimento prévio de estudantes quanto para representação de domínios através de mapas conceituais.

Diversas técnicas podem ser utilizadas para extrair termos de documentos. Segundo Abramovitz (2009), elas podem ser estatísticas, linguísticas, por aprendizagem de máquina ou híbridas. Essas técnicas, de forma geral, buscam, a partir de um conjunto de documentos, extrair uma lista de palavras-chave ou conceitos que ao final definirão o contexto e o domínio dos documentos analisados.

O trabalho de Cañas e Saxe (2005) ressalta que os planos de aula e cursos são organizados de maneira linear e, em alguns casos, contêm conceitos desconexos. Estas características podem prejudicar a forma como os estudantes organizam e representam conhecimento, ferindo assim, os preceitos da aprendizagem significativa propostos por Ausubel et al (1980).

No Programa “e-Tec Brasil” (Escola Técnica Aberta do Brasil), desenvolvido pelo Ministério da Educação e Cultura, foram estabelecidos alguns referenciais para elaboração de material didático para o Ensino à Distância profissional e tecnológico.

Os referenciais do programa supramencionado propõem para os professores que expliquem aos estudantes os objetivos de aprendizagem gerais e específicos a serem trabalhados em cada bloco temático. Nos materiais do e-Tec (unidades, módulos, aulas, etc.) também se devem articular os objetivos propostos em cada bloco, utilizando, preferencialmente, mapas conceituais (ETEC, 2007).

Baseando-se nos referenciais propostos pelo programa e-Tec, se observa que os mapas conceituais estão presentes no meio educacional e podem ser utilizados principalmente no que diz respeito à organização de cursos e conteúdos. Seguindo este raciocínio, pode-se inferir que os mapas conceituais promovem a aprendizagem significativa ao permitir que a estrutura cognitiva do estudante seja representada em um diagrama composto por conceitos e conectores semânticos.

De acordo com Ausubel et al (1980), a aprendizagem significativa acontece quando um novo material, uma nova idéia ou uma nova informação são logicamente estruturados e interagem com conceitos relevantes e inclusivos disponíveis na estrutura cognitiva dos estudantes. Essa interação resulta na diferenciação, elaboração e estabilidade do conhecimento.

Novak e Ausubel (1980) propõem o uso de mapas conceituais para representar conceitos e suas relações semânticas por meio de um diagrama simples onde os primeiros são representados por nós e as últimas são notadas por arestas. Sugere-se que os conceitos já conhecidos ou os pré-requisitos para o curso sejam apresentados na forma de um mapa conceitual criado pelo professor. Nesse mapa, devem constar os conceitos relevantes para a assimilação dos conteúdos a serem trabalhados em um módulo temático.

Seguindo os preceitos das novas tecnologias de informação e comunicação, é cabível a aquisição de uma ferramenta computacional para edição, armazenamento e compartilhamento de mapas conceituais. A partir disso, professores e estudantes tornam-se aptos a criar e compartilhar conhecimento organizado construído de forma colaborativa.

A organização dos conteúdos e a elaboração do currículo são consideradas peças-chave para que o uso do computador seja mais vantajoso para professores e estudantes tanto em sala de aula quanto no Ambiente Virtual de Ensino e Aprendizagem (AVEA). Acompanhar o

aprendizado e identificar conhecimentos prévios também facilitam o trabalho dos professores na mediação do processo de construção de conhecimento (WU et al, 2005).

A utilização de módulos temáticos representados por mapas conceituais elaborados por um professor referencia o “princípio de diferenciação sucessiva” em que os conceitos mais gerais vão sendo sucessivamente especificados pelo professor (NOVAK; CAÑAS, 2006).

Para Novak e Gowin (1984) e Wafula (2006), quando o objetivo é promover o aprendizado dos estudantes, é imprescindível que estes realizem algum esforço para explicitar o seu conhecimento prévio sobre um dado assunto, que então poderá ser representado por meio de um mapa conceitual.

O método proposto considera que o estudante pode e deve associar novos conceitos a um mapa criado pelo professor, além de navegar por objetos de aprendizagem vinculados aos conceitos dispostos em um mapa conceitual.

De acordo com Wafula (2006), um professor, ao criar um mapa conceitual, pode encontrar dificuldades em identificar quais os recursos relevantes para o módulo temático ou até mesmo esquecer-se de conceitos importantes sobre um determinado assunto. O extrator pode identificar conceitos esquecidos, e estudantes e colaboradores podem complementar as proposições existentes em um mapa.

É sabido que nem sempre os estudantes dominam todos os conceitos propostos em um curso. É desta constatação que surge o termo “Organizadores Prévios” que, segundo Ostermann e Moreira (1999), são materiais que auxiliam os estudantes a dominar conceitos considerados pré-requisitos para um melhor aproveitamento do curso. Mostrando um mapa conceitual inicial, o professor lista os organizadores prévios necessários para alcançar esse objetivo.

Em diversas atividades de aprendizagem, estudantes são instigados a escrever textos dissertativos ou a elaborar respostas discursivas. A partir desses textos, aplicando técnicas de mineração, é possível supor o conhecimento prévio e/ou adquirido de um estudante sobre o tema que está sendo trabalhado em aula.

O professor, ao construir um curso, poderá associar diferentes recursos digitais a um mapa conceitual, que podem ser reusados para apoiar o aprendizado. Esses recursos, chamados objetos de aprendizagem, foram descritos detalhadamente por Willey (2000).

No processo de aprendizagem amparado pelo computador, cabe aos estudantes utilizar o mapa conceitual como objeto de aprendizagem

e acessar os conteúdos de outros objetos vinculados aos conceitos contidos no mapa.

1.4 Escopo do Trabalho

A idéia que norteia este trabalho é a facilitação do processo de construção de um mapa conceitual por **professores**. Neste sentido, o método desenvolvido realiza a extração de termos relevantes de um módulo temático a partir de um conjunto de documentos de texto conhecidos pelo docente.

A extração de termos tem a finalidade de auxiliar o professor a identificar e a relacionar os conceitos que podem compor o mapa de um curso ou módulo temático de um curso.

Para comprovar a efetividade do método proposto foi concebida a ferramenta computacional denominada **MAPXTRACTOR**. A essa ferramenta, além de um módulo que faz a extração de termos, foram acopladas as funcionalidades que permitem organizar os conteúdos em mapas conceituais, os quais podem ser editados de diversas formas (fontes, cores, figuras).

Complementares à ferramenta mencionada, existem funcionalidades que permitem consultas rápidas de conceitos na *web* e principalmente o acesso a objetos de aprendizagem vinculados aos conceitos.

Com vistas a tornar viável a pesquisa e a implementação do método, foram estabelecidos alguns limites para construção da ferramenta, a saber:

- sugere-se que os textos usados como referência para o extrator não tenham mais de vinte páginas (por questões de desempenho e memória utilizada);
- o mapa gerado automaticamente não tem a opção *auto-layout*, comum em ferramentas de diagramas como esta.

1.5 Organização do Trabalho

Considerando que este trabalho está fundamentado em técnicas de representação de conhecimento baseadas em mapas conceituais, optou-se por representar a sua organização por meio de um mapa conceitual. Tal mapa é apresentado na Figura 1.

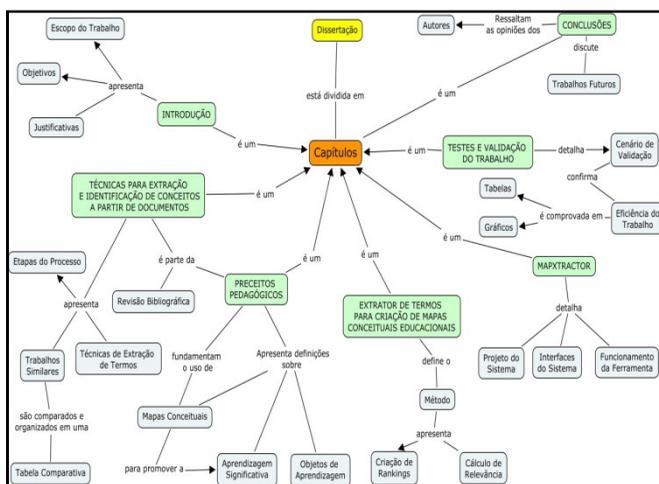


Figura 1- Mapa Conceitual da Organização do Trabalho

2 TÉCNICAS PARA EXTRAÇÃO E IDENTIFICAÇÃO DE CONCEITOS A PARTIR DE DOCUMENTOS

O objetivo deste capítulo é apresentar a fundamentação teórica que garante a aplicabilidade de técnicas de extração de termos para o desenvolvimento de um método que oriente a construção de mapas conceituais com fins didáticos.

Neste capítulo serão apresentadas técnicas de extração de informação, mineração de textos, processamento de linguagem natural e análise de discurso, que tornam possível a automatização do processo de construção de mapas conceituais a partir de um conjunto de documentos de texto.

Em um segundo momento, serão apresentadas algumas ferramentas que trabalham com a construção de mapas conceituais de forma automática e/ou semi-automática, considerando diferentes abordagens para extração de termos e propostas de formas de avaliação de ferramentas para extração de mapas conceituais.

2.1 Data Mining e Text Mining

Datamining (mineração de dados) é a busca por informações novas, preciosas e não triviais, em grandes volumes de dados. Os resultados são obtidos equilibrando o conhecimento de especialistas humanos ao descrever problemas e metas a partir das capacidades de processamento e busca dos computadores (KANTARDZIC, 2002).

O *textmining* (mineração de textos) é uma especialização do *datamining* onde a análise está centrada em textos em linguagem natural, estruturados ou não.

Feldman e Sanger (2007) definem *textmining*, de forma geral, como um processo de organização de conhecimento em que o usuário interage com uma coleção de documentos ao longo do tempo, utilizando ferramentas de análise similares a um *datamining*.

A sua principal diferença, conforme relatam esses autores, está no fato de que ao invés de utilizar tuplas de um banco de dados formal, as

informações são extraídas a partir de informações não estruturadas, em linguagem natural, descritas em uma coleção de documentos.

A coleção de documentos é um elemento chave em *textmining*, e pode ser agrupada de acordo com objetivo da mineração. Ao se extraírem termos para a construção de mapas conceituais a partir de uma coleção de documentos não há a intenção de agrupar ou anotar documentos de acordo com o seu conteúdo, mas sim buscar relações semânticas relevantes que compõem essa coleção.

Sob a ótica de Feldman e Sanger (2007), que considera pequena uma coleção com 15.000 documentos, pode-se concluir que aplicações de mineração de textos trabalham com uma massa extremamente grande de dados.

Em *datamining* de texto (ou *textmining*), a utilidade da técnica reside na habilidade de considerar uma maior quantidade de dados. Como exemplo, tem-se o trabalho de Swanson e Smalheiser (1997), onde é possível avaliar a associação entre enxaquecas e deficiência de magnésio através de uma análise semi-automática dos títulos de documentos científicos nos campos médicos e biológicos.

No presente trabalho, a proposta é a extração de termos de uma quantidade pequena de documentos, não mais do que 100, os quais são previamente selecionados pelos usuários durante o processo de criação de um mapa conceitual. Cabe ressaltar que embora o método proposto siga princípios de *textmining*, ele não pode ser considerado um minerador de textos.

2.1.1 Aplicações práticas de Data Minings e Text Minings

Conforme Gelfand et al (1998) e Gupta et al (2006), os mineradores de texto visam classificar, agrupar, indexar termos e encontrar padrões em documentos. Tanto organizações quanto usuários comuns podem encontrar benefícios ao fazer uso desta técnica.

Em suma, técnicas de mineração de textos são capazes de identificar peculiaridades em documentos de texto. Tais peculiaridades referem-se à descoberta de conhecimento através de uma análise na estrutura do texto. O processo de análise envolve as seguintes etapas: *parsing* dos textos, reconhecimento de padrões, análise semântica e

tokenização. A Figura 2 esboça as etapas para execução da mineração de textos.

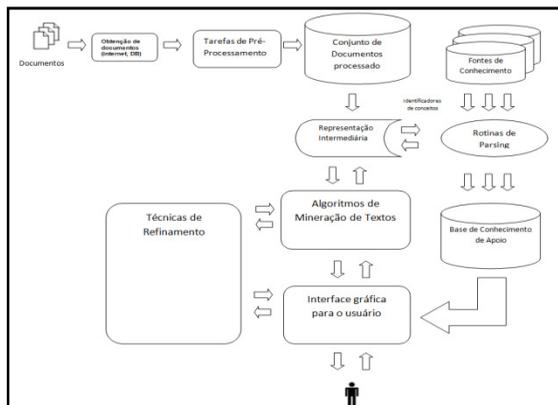


Figura 2 - Etapas de mineração de texto

Fonte: Adaptado de Feldman e Sanger (2007)

Apesar de compartilhar de preceitos de *textmining*, o foco deste trabalho transcende alguns preceitos dessa técnica, já que o propósito é encontrar termos que sejam capazes de representar o domínio do conhecimento correspondente a um tema que será abordado em um curso.

O método proposto neste trabalho utiliza um algoritmo que obtém listas de conceitos formados por um ou mais termos. Esses termos são classificados de acordo com a sua relevância dentro do conjunto de documentos utilizados como referência no processo de extração. Ademais, o método é capaz de identificar termos considerados conectores semânticos que relacionam os conceitos extraídos.

A extração de termos baseada em documentos sem uma estrutura previamente definida é um procedimento complexo. Os algoritmos utilizados para solucionar este problema demandam um grande processamento computacional e heurísticas otimizadas, de acordo com Gaudio e Branco (2007) e Bendersky e Croft (2008).

Visando aprimorar o processo de extração de conceitos, Gupta et al (2006) propõem um método para organização conceitual a partir de textos semi-estruturados, onde o conceito é representado por substantivos presentes em textos de um determinado domínio. Outros

trabalhos, como o de Chin et al (2006), aliam técnicas de mineração de textos, sumarização e categorização com a finalidade de descobrir conceitos relevantes em textos.

2.2 Extração de Termos e Palavras-Chave

A Extração de Informação (EI) tem por objetivo extrair, a partir de um conjunto de documentos textuais, apenas os dados relevantes para o usuário (KUSHMERICK; THOMAS, 2003). Extrair informações sugere também extrair elementos semânticos de um texto para povoar uma nova base de dados, como por exemplo, identificar datas, telefones, endereços.

EI pode ser vista como um processo que parte de uma entrada de texto livre ou um conteúdo semi-estruturado, gerando saídas que são dados estruturados de interesse, representados, por exemplo, por *k*-tuplas ou objetos complexos (CHANG et al, 2006).

Os autores Nakagawa e Mori (2002) afirmam que o reconhecimento e a extração automática de termos consistem na extração de termos de um domínio específico a partir de um *corpus* (conjunto de documentos) com características técnicas ou acadêmicas. Em outras palavras, são processos automáticos de reconhecimento e extração de candidatos a conceitos relevantes presentes em um determinado *corpus* de um domínio.

Segundo Giarlo (2005), a extração de palavras-chave pode ser limitada a substantivos, já que esses itens léxicos são mais comumente encontrados em textos. Assim, após análise sintática do texto e a marcação dos termos correspondentes a essa classe gramatical, é possível afirmar que os termos selecionados são palavras-chave em potencial.

Para contextualizar o processo de extração de termos é preciso definir algumas premissas que embasam as técnicas utilizadas nos algoritmos que buscam solucionar este problema.

Primeiramente, deve-se considerar a definição de termo como uma palavra ou expressão usadas em certo domínio com um significado específico, ou seja, é um elemento capaz de expressar e transferir o conhecimento de caráter especializado (PEREZ-MARIN et al, 2006).

Em segundo lugar, ressalta-se que um documento pode abranger um ou mais tópicos conceituais. Tais tópicos podem ser descritos como conceitos que são representados pela combinação de alguns significados particulares de um termo. Se comparado à definição de palavra-chave, um conceito expressa os significados de forma mais concreta e precisa (ZANG et al, 2003).

Thanopoulos et al (2000) comentam que, quando a extração ocorre em textos de um domínio específico, geralmente existe um vocabulário restrito inerente ao tema abordado. Surge, então, uma característica que favorece a aquisição de conhecimento a partir de textos previamente selecionados e classificados de acordo com um domínio.

Soluções propostas para a extração de termos usualmente analisam um conjunto de documentos de um domínio com o objetivo de encontrar palavras ou expressões relevantes para o domínio analisado.

Reduzindo-se o escopo de análise, ganha-se tempo no processo de extração e menos ruído (falsos positivos) nos resultados.

Para o português brasileiro, salienta-se, ainda não existe uma ferramenta consolidada que funcione eficazmente resolvendo problemas de extração de conceitos.

Os trabalhos de Kowata et al (2009) e Abramowicz e Wisniewski (2008) citam quatro abordagens para a extração de termos a partir de documentos, a saber:

- linguística;
- estatística;
- aprendizagem de máquina e
- híbrida.

As seguintes métricas podem ser utilizadas para avaliar os resultados: precisão (P), abrangência (A) e *f-measure* (F).

A precisão ($P = (\text{Termos Referência} \cap \text{Termos Extraídos}) / \text{Termos Extraídos}$) indica a capacidade do método de identificar os termos corretos, considerando a lista de referência.

A abrangência ($A = (\text{Termos Referência} \cap \text{Termos Extraídos}) / \text{Termos Referência}$) avalia a quantidade de termos corretos extraídos pelo método.

A *f-measure* ($F = (2 * P * A) / (P + A)$) é a medida harmônica entre a precisão e a abrangência.

2.2.1 Abordagem Linguística

Esta abordagem é baseada em informações linguísticas organizadas em conjuntos de anotações. As informações necessárias são obtidas de acordo com algumas regras linguísticas. Tais regras geralmente são criadas por especialistas no domínio e em gramática.

Para definir as regras de extração são usadas características morfosintáticas do texto (ABRAMOWICZ; WISNIEWSKI, 2008).

As palavras que possuem comportamento sintático-semântico similar podem ser agrupadas em uma mesma classe, originando as chamadas categorias sintáticas ou gramaticais, mais comumente denominadas *Parts of Speech* (PoS). As três principais são substantivo, verbo e adjetivo. Os substantivos referem-se a pessoas, animais, conceitos e coisas; o verbo é usado para expressar a ação numa sentença e os adjetivos, por sua vez, são propriedades dos substantivos (MANNING; SHÜTZE, 1999).

Na abordagem linguística são utilizadas técnicas de análise de discurso, processamento de linguagem natural e reconhecimento de padrões previamente definidos por projetistas que geralmente são especialistas no domínio.

Conforme Pazienza (2005), o processo de extração de termos com abordagem linguística deve ser capaz de:

- analisar o *corpus* de um domínio e identificar as PoS;
- identificar e extrair termos candidatos conforme as regras criadas;
- preservar os significados de acordo com o termo original e
- implementar filtros linguísticos para refinar a terminologia.

Um dos módulos mais importantes de um extrator com bases linguísticas é o *parser*. O *parser* deve classificar os termos de acordo com a sua classe gramatical ou função sintática para permitir a aplicação das regras linguísticas. Por exemplo: uma proposição válida deve ser composta por 3 termos que sigam a ordem substantivo > verbo > substantivo.

Outra forma de se aplicar um *parser* é utilizando *tags* existentes em um documento semi-estruturado, onde é possível, por exemplo, extrair termos em negrito ou escritos todos em letras maiúsculas.

Ao se analisar trabalhos que utilizam uma abordagem estritamente linguística, observa-se que esta técnica de extração de

termos demanda um tempo maior de projeto para que sejam elaboradas regras e formado o conjunto de anotações (ZOUAQ; NKAMBOU, 2009).

A abordagem linguística se baseia em técnicas de processamento de linguagem natural, tendo um alto custo de projeto. Os extrativos são fortemente baseados no conjunto de palavras-chave associadas aos textos e em algumas outras "pistas" simples para a seleção da informação relevante.

Os resultados obtidos com esse tipo de abordagem mostram-se ligeiramente mais corretos do que os obtidos a partir das abordagens estatísticas, porém, a aplicação efetiva da técnica linguística esbarra no esforço necessário tanto na etapa de projeto, quanto na implementação do modelo.

2.2.2 Abordagem estatística

As soluções que propõem o uso de técnicas estatísticas avaliam a frequência com que os termos ocorrem em um texto. As palavras com maior número de ocorrências em um documento tendem a representar de forma mais expressiva o conteúdo deste.

A abordagem estatística para extração de termos de um conjunto de textos, ao contrário da abordagem linguística, emprega métodos puramente estatísticos ou heurísticas, fazendo uso de medidas estatísticas para a seleção de um subconjunto de sentenças do texto original que expressem a idéia central do mesmo.

De acordo com Pazienza (2005), existem mais medidas estatísticas que podem denotar relações entre as que podem ser utilizadas na extração de termos, que representam a dimensão linguística dos termos extraídos. O Quadro 1 mostra como se dão tais relações.

		Dimensão Estatística		
		Grau de associação	Significância da Associação	Heurística
Dimensão Linguística	<i>Unithood</i>	MI	z-score, T-score	IM ² IM ³
		Fator Dice	X ² , Log Likelihood Ratio	
<i>Termhood</i>				Frequência
				C-Value
				Co-Ocorrência

Quadro 1 - Relações de Medidas Estatísticas e as Dimensões envolvidas na Extração de Termos

Fonte: Adaptado de PAZIENZA (2005).

Unithood significa o quanto um termo extraído pode ser considerado válido como termo (considera o problema de termos compostos, também chamados de *collocations*), enquanto *Termhood* representa o quanto um termo extraído pertence ao domínio do conjunto de documentos.

2.2.3 Abordagem de aprendizagem de máquina

É possível identificar padrões de escrita e relações entre termos a partir de um conjunto de exemplos que podem ser utilizados como referência para treinamento de algoritmos capazes de aprender a identificar padrões (como por exemplo, redes neurais, algoritmos baseados em técnicas bayesianas) (TURNEY, 2003).

Zhou et al (2008) sugerem o uso de modelos de Markov para extrair palavras formadas por mais de um termo.

A frequência das ocorrências de termos no conjunto de documentos analisados e a distância entre os termos podem ser úteis para a definição de parâmetros de organização léxico-sintática dos

termos e frases principais de um texto (FRANK et al, 1999). Estas variáveis são usadas para treinar redes neurais que são capazes de identificar padrões e outras informações pertinentes que podem ser interessantes para o processo de extração.

2.2.4 Abordagem híbrida

O procedimento para extração de termos pode ser tratado por abordagens híbridas combinando modelos estatísticos, linguísticos e de aprendizagem de máquina.

É utilizado quando se pretende obter os benefícios de diferentes abordagens para melhorar a qualidade dos resultados, resolvendo problemas que podem eventualmente surgir devido às deficiências existentes em cada uma das abordagens.

Segundo Pazienza (2005), esse tipo de abordagem tende a alcançar resultados mais corretos se comparado às abordagens puramente estatística ou puramente linguística.

2.3 Ranking de Termos

De acordo com Nakagawa e Mori (2002), em documentos técnicos de domínios específicos, a maioria dos termos que representam estes domínios são compostos por locuções ou substantivos compostos constituídos por mais de uma palavra.

Os autores supramencionados comentam sobre a importância de criar *rankings* diferentes para termos simples e termos compostos, e apresentam um modelo onde o *ranking* de termos compostos é influenciado pelo *ranking* de um termo simples. Foi nesse tipo de *ranking* que o método proposto neste trabalho se baseou.

Ao elaborar um *ranking* é necessário encontrar atributos que possam ser mensurados para determinar a posição de um termo no *ranking*. Exemplos de atributos de um termo podem ser: quantidade de documentos em que o termo aparece; quantidade de vezes em que o

termo aparece em um documento; relação de distância entre palavras de um conjunto de termos a serem pesquisados.

No entanto, ao definir um *ranking* de termos é imperioso que o usuário de um sistema possa identificar o que este *ranking* representa. Depois de criado o *ranking*, pode ser necessário filtrar os termos ranqueados de acordo com características léxico-sintáticas pré-determinadas no método de extração.

2.4 Pré-processamento

As fases de pré-processamento são responsáveis por preparar o conjunto de documentos para a realização da extração de termos. Esta é uma etapa do algoritmo que demanda processamento computacional pesado e deve ser bem planejada no projeto de um extrator de termos.

Nesta etapa, operações entre *strings* são realizadas consumindo tempo do processador. Várias listas são criadas e consultadas durante o processo, requisitando espaço na memória principal do computador.

2.4.1 Tokenização

Um dos princípios de um algoritmo de extração de termos é identificar *tokens* que possam ser classificados e ranqueados de acordo com alguma métrica ou peso que definem a relevância do termo no conjunto analisado.

Tokens podem ser palavras, sinais de pontuação, números e outros símbolos contidos em um documento de texto. Os *tokens* podem ser úteis para identificar notações semânticas como datas, valores de moeda ou unidades de medida. Pode-se, como exemplo, definir a seguinte regra: caso seja captado um texto situado entre dois espaços em branco ou um espaço em branco e um ponto final, que possua dois caracteres “/”, entende-se que esse texto refere-se a uma data.

Um elemento importante na etapa de *tokenização* é a definição do separador de *tokens*, que pode ser: espaços em branco, sinais de

pontuação, parágrafos e outros. O separador de *tokens* influencia diretamente nos resultados, uma vez que permite a formação de diferentes *n-gramas* conforme o separador escolhido.

N-gramas podem ser classificados como uma seqüência de caracteres de um texto agrupados conforme a aplicação. Existem nomações que representam o número de palavras que compõem um *n-grama*, a saber: unigrama, bigrama e trigrana, que respectivamente são formados por um, dois e três termos.

Seguem os exemplos para ilustrar o conceito explicitado: o termo “Mapa” corresponde a um unigrama; os termos “Mapa Conceitual”, juntos, formam um bigrama; enquanto os termos “Mapa de Conceitos”, também juntos, por sua vez compõem um trigrana.

2.4.2 Capitalização

A capitalização de termos é uma etapa que pode influenciar nos resultados do processo de extração. Em alguns momentos, é necessário identificar termos que definem inícios de períodos, nomes próprios ou siglas em meio ao texto analisado.

Para que essas características sejam identificadas, pode-se, por exemplo, verificar se a primeira letra de uma palavra é maiúscula e se as outras são minúsculas, o que diferencia um nome próprio.

Em outro exemplo: se todas as letras forem maiúsculas, estas podem ser consideradas siglas. Dessa forma, nota-se que, se houver a capitalização de todos os elementos de um texto, algumas informações morfossintáticas podem ser perdidas.

Durante o processo de extração de termos são criadas listas onde são adicionados os termos. Quando se faz uso de técnicas estatísticas para a extração de termos, é necessário obter a quantidade de vezes que um termo se repete no texto. Para isso, é preciso encontrar formas de verificar se o termo já está na lista de termos extraídos e incrementar uma variável que representa a quantidade de ocorrências.

Ao se trabalhar com listas de palavras extraídas de um texto, deve-se atentar para o fato de que, conforme o tipo de comparação realizada, o resultado pode ser FALSO, como na seguinte comparação: “Casa” = “casa”. Novas linguagens já trazem funções para realizar a

verificação desconsiderando a capitalização dos termos, embora o tempo de processamento aumente quando são utilizadas estas funções.

Considerando que os métodos estatísticos para extração de termos realizam operações de comparação de *string* muitas vezes durante o processo, pode-se otimizar o tempo de resposta fazendo comparação com os dados na (mesma) forma “maiúsculas ou minúsculas”. Sugere-se, então, mudar a capitalização de todo o texto antes de iniciar o processo de extração.

A vantagem de se colocar todas as letras em formato maiúsculo se justifica pelo fato de que a extração de termos repetidos é evitada, facilitando a comparação entre todos os termos.

Se fosse proposto, por exemplo, um método que não versasse a capitalização dos termos, ele passaria a ser sensível a letras maiúsculas e minúsculas, o que permitiria que na lista de termos existissem os termos “casa” e “Casa”, algo que pode ser inviável para algumas aplicações e prejudicial à resposta fornecida pelo extrator.

2.4.3 Tratamento de *Stopwords*

Stopwords são palavras como preposições, artigos e conjunções, que podem ser descartadas em uma fase de indexação ou em consultas por termos (LO, 2005). A eliminação de *stopwords* faz parte do processo de seleção de palavras que podem ser definidas como indexadores ou de seleção de termos que representem um domínio.

Essa eliminação resulta em uma redução do espaço de descritores, ou seja, diminui a quantidade de termos que podem ser considerados conceitos. Por outro lado, ao serem descartadas as preposições, perde-se em representatividade (GONZALES; LIMA, 2006).

Dependendo da aplicação, é interessante utilizar *stopwords* para identificar conceitos formados por uma composição de termos. Por exemplo: “caixa de vidro” e “caixa para vidro” têm, evidentemente, significados diferentes e esta diferença não pode ser representada, nesse caso, sem as preposições (DALMOLIN, 2009).

Conforme estas preposições, ora faz sentido descartar as *stopwords*, ora é necessário mantê-las para dar sentido a determinados conceitos existentes nos textos.

2.4.4 Identificação de Expressões e Termos Compostos

A extração de conceitos compostos por mais de um termo é comumente utilizada em aplicações de extração de informação e frases-chave de textos (TURNEY, 2003).

Para se ter noção de como os algoritmos podem classificar e ranquear conceitos compostos, o seguinte exemplo demonstra algumas classificações:

- Considerando-se o trecho “... têm preocupado os pesquisadores ...”, um extrator pode classificar os termos através dos seguintes tipos de relacionamentos:

- o par modificado-modificador “pesquisador-preocupado”;
- o bigrama “(pesquisador preocupado)”;
- o sintagma nominal “pesquisador preocupado” ou algum outro formato, como a expressão ternária “preocupação-de-pesquisador” e
- a relação binária “(preocupação,pesquisador)”.

Particularmente neste trabalho, o conceito de bigrama será utilizado devido à sua simplicidade de obtenção através de uma abordagem estatística para extração de termos. Para a obtenção de bigramas basta combinar palavras duas a duas e iterar o procedimento para todos os termos de um texto.

2.4.5 Tratamento de Plurais e Gênero

O tratamento de plurais consiste na identificação do termo sempre na sua forma singular, visando evitar a identificação de termos duplicados. Uma forma de se realizar o tratamento de plurais e gêneros é analisando a morfologia das palavras. São utilizadas listas de terminações para extração de radicais e definição de sufixos para padronizar a entrada de um termo na lista de termos extraídos.

Em uma aplicação podem haver algumas inconsistências causadas por este tipo de tratamento, como pode ser observado na seguinte regra:

“se uma palavra terminar em ‘as’, então modifique o valor do sufixo para ‘a’”.

Ex.: Casas > casa. No entanto, palavras como ‘atrás’ poderiam ficar em um estado inválido. Ex.: atrás > atrás.

O tratamento de plurais e gêneros é realizado por algoritmos conhecidos como *stemmers*, que conseguem identificar os radicais que formam as palavras.

2.4.6 Separação de Sentenças

A separação de sentenças é essencial quando se faz análise com *n-gramas*. Uma vez que um conceito pode ser composto por um ou mais termos, é importante que esses termos estejam na mesma frase.

Para obter melhores resultados e evitar processamento desnecessário, são feitas combinações respeitando parágrafos, pontos, vírgulas e outras marcações de períodos no texto.

Exemplo: “*O mercado fica próximo a rua da praça central. Você deve ir à pé!*”. Se o algoritmo estiver analisando este trecho, ele não precisa montar os seguintes *n-gramas*:

- “praça central Você”;
- “central Você deve” ou
- “central deve”.

Os termos, como se pode notar, não possuem conexão semântica bem definida. Assim, não precisam ficar em memória no momento da extração, o que agiliza o processo de ranqueamento, poupando memória e tempo de processador.

2.5 Análise Comparativa de Ferramentas Similares

Para Valerio e Leake (2006), a análise de um documento visando extrair informações para a criação de um mapa conceitual apóia-se em duas etapas: a primeira é a identificação dos termos que formam os conceitos no texto (objetos diretos e indiretos nas sentenças ou sujeitos);

a segunda etapa é a identificação e o registro dos termos de ligação entre os conceitos (conjunções e verbos).

Diferentemente dos preceitos de recuperação de informação, onde se pretende localizar termos que raramente ocorrem em um conjunto de dados, o processo de extração de termos relevantes busca ressaltar aqueles que mais se repetem. Isto porque se um termo é citado com frequência em um ou em vários documentos, infere-se que este seja relevante para o contexto analisado.

Nesse caso, não são consideradas as *stopwords* porque embora elas ocorram com alta frequência em todos os textos, quando analisadas individualmente são irrelevantes.

Segundo Wafula (2006), mapas conceituais podem ser gerados a partir de textos, onde uma aplicação pode analisar os parágrafos um a um e extrair termos considerados conceitos. Além disso, utilizando-se de heurísticas, podem ser criadas relações entre esses conceitos, possibilitando que outros termos sejam usados como conectores. Portanto, conceitos e conectores extraídos automaticamente de um texto podem gerar mapas conceituais.

No trabalho de Kowata et al (2009) é apresentada uma revisão bibliográfica que propõe uma forma para avaliar as funcionalidades e características de ferramentas de extração de termos para a construção de mapas conceituais. Segundo os autores, as características foram agrupadas em perspectivas relacionadas aos **objetivos estabelecidos**, à **fonte de dados utilizada** pelas pesquisas e aos **mapas conceituais construídos**. Os quesitos utilizados para a análise foram os seguintes:

Objetivo: neste quesito, os mapas podem ser classificados como *educacionais* quando permitem que tanto educadores quanto estudantes utilizem os mapas conceituais construídos como ferramentas complementares no processo de ensino e aprendizagem; *analíticos* quando permitem que especialistas de domínio utilizem mapas conceituais para exploração, análise e definição de padrões de comportamento minerados a partir de grandes quantidades de dados; e *computacionais* quando permitem que *softwares* manipulem os mapas conceituais produzidos como fontes para aquisição de conhecimento.

Fonte de Dados: no que se refere à fonte dados, a análise proposta pelos autores considera a estrutura do texto utilizado como entrada no extrator, a necessidade da existência de um domínio e o método de extração e manipulação dos textos.

Mapas Construídos: quanto à completude, um mapa conceitual pode ser analisado em relação à aderência aos elementos definidos na Teoria de Mapas Conceituais, caracterizando-se como um mapa

completo ou incompleto. Em mapas conceituais ditos incompletos, proposições não podem ser identificadas, pois não há rótulos para as relações ou sua semântica é insuficiente para o propósito de gerar proposições, ao contrário dos mapas completos nos quais as proposições são claramente definidas.

Kowata et al (2009) ressaltam características sobre o mapa construído por cada uma das ferramentas e demonstram como este fora gerado (se automaticamente ou não). Os autores analisam os resultados avaliando a forma de representação e armazenamento da informação, fazendo também reflexões sobre como avaliar ferramentas de extração de mapas conceituais.

2.5.1 Lista de Trabalhos Analisados

O trabalho de Kowata et al (2009) analisa os seguintes trabalhos (Quadro 2):

Referência	Título do trabalho
Alves et al (2001)	Automatic Reading and Learning from Text
Bai et al (2008)	Automatically constructing concept maps based on fuzzy rules for adapting learning systems
Chang et al (2008)	Automatic Concept Map Constructing using top-specific training corpus
Chen et al (2008)	Mining e-Learning domain concept map from academic articles
Clariana e Koul (2004)	A Computer-Based Approach for Translating Text into Concept Map-Like Representations.
Gaines e Shaw (1994)	Using Knowledge Acquisition and Representation Tools to Support Scientific Communities
Graudina e Grundspenkis (2008)	Concept Map Generation from OWL Ontologies
Kumazawa et al (2009)	Toward knowledge structuring of sustainability science based on ontology engineering
Lau et al (2008)	Towards Fuzzy Domain Ontology Based Concept Map Generation for E-Learning
Lee et al (2009)	Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning
Pérez e Vieira (2004)(2005)	Aquisição de Conhecimento a partir de Textos para Construção de Mapas Conceituais
Richardson e Fox (2007)	Using Concept Maps in NDLTD as a Cross-Language
Tseng et al (2007)	A new approach for constructing the concept map
Valerio e Leake (2006)	Jump-Starting Concept Map Construction with Knowledge Extracted from Documents
Zouaq e Nkambou (2008)(2009)	Building Domain Ontologies from Text for Educational Purposes

Quadro 2 - Lista de Trabalhos Correlatos

Fonte: Kowata et al (2009)

Analisando o Quadro 3 observa-se que somente um dos trabalhos lida especificamente com a Língua portuguesa. Percebe-se, então, a demanda para se trabalhar a extração de termos para o português brasileiro.

Outra conclusão diz respeito às avaliações, que na sua maioria se dão de forma subjetiva dada a inexistência de formalismo e métricas bem definidas para avaliar a representatividade de um mapa conceitual extraído semi ou automaticamente.

Abordagem		Fonte de Dados			Mapa Construído		
Referência	Objetivo	Domínio	Estrutura	Método	Construção	Leitave Completada	Avaliação
Alves et al (2001)	Educacional	Independente de Domínio	Não estruturada	Método Híbrido	Semi	Grafo Completo	Subjetiva
Bai et al (2008)	Analtico	Específica de Domínio	Estruturada	Recuperação de Dados	Automática	Grafo Incompleto	não disponível
Chang et al (2008)	Analtico	Específica de Domínio	Não estruturada	Método Estatístico	Automática	Árvore Incompleto	não disponível
Chen et al (2008)	Analtico	Específica de Domínio	Não estruturada	Método Estatístico	Automática	Grafo Incompleto	Subjetiva
Clariana e Koul (2004)	Educacional	Específica de Domínio	Não estruturada	Método Estatístico	Semi	Grafo Incompleto	Objetiva
Craines e Shaw (1994)	Analtico	Específica de Domínio	Não estruturada	Método Estatístico	Semi	Grafo Incompleto	Subjetiva
Graudma e Grundspenkis (2008)	Educacional	Específica de Domínio	Estruturada	Mapeamento de Elementos	Semi	Grafo Incompleto	Subjetiva
Kumazawa et al. (2009)	Analtico	Específica de Domínio	Estruturada	Mapeamento de Elementos	Automática	Grafo Incompleto	Subjetiva
Lau et al (2008)	Educacional	Específica de Domínio	Não estruturada	Método Híbrido	Automática	Grafo Incompleto	Subjetiva
Lee et al. (2009)	Analtico	Específica de Domínio	Estruturada	Recuperação de Dados	Automática	Grafo Incompleto	Subjetiva
Pérez e Vieira (2004)(2005)	Educacional	Independente de Domínio	Não estruturada	Método Lingüístico	Informação não disponível	Grafo Completo	Subjetiva
Richardson e Fox (2007)	Educacional	Específica de Domínio	Não estruturada	Método Lingüístico	Semi	Grafo Completo	Subjetiva
Tseng et al (2007)	Educacional	Específica de Domínio	Estruturada	Recuperação de Dados	Automática	Grafo Incompleto	Subjetiva
Valerio e Leake (2006)	Analtico	Independente de Domínio	Não estruturada	Método Híbrido	Manual	Grafo Completo	Subjetiva
Zouaq e Nkambou (2008)(2009)	Computacional	Específica de Domínio	Não estruturada	Método Híbrido	Automática	Grafo Completo	Subjetiva
MapXtractor	Educacional	Específica de Domínio	Não estruturada	Método Híbrido	Semi / Automática	Grafo Incompleto	Subjetiva

Quadro 3 - Dados comparativos de trabalhos analisados

Fonte: Kowata et al (2009)

Sobre técnicas de avaliação de mapa, Novak e Gowin (1984) propõem analisar a estrutura hierárquica dos mapas e de proposições, considerando o número de ramificações, a existência de *cross-links* e exemplos específicos. Todavia, nenhum dos trabalhos citados avaliou os

grafos dessa forma, e não foram encontradas referências de trabalhos que utilizem este tipo de técnica para avaliar mapas conceituais.

2.6 Concept map mining

A expressão *concept map mining* foi concebida por Vilallon e Calvo (2008), quando apresentaram a viabilidade da extração de mapas conceituais a partir de ensaios (textos dissertativos e descritivos elaborados em atividades de aprendizagem) criados por estudantes.

Seguindo abordagens cognitivistas, os autores firmam-se no princípio de que dissertar sobre um assunto é uma tarefa que ativa funções cognitivas altamente especializadas como síntese e análise.

De acordo com Wafula (2006), entre outras finalidades, os mapas conceituais podem ser utilizados para representar informações de livros-texto, laboratórios, estúdios, artigos de jornais e artigos científicos e representações na fase planejamento de um artigo ou exposição.

Em busca de uma aprendizagem adaptativa, que respeite o ritmo e o nível de conhecimento de um grupo, sugere-se utilizar um mapa conceitual pré-definido que disponibilize meios para o professor analisar e refinar suas estratégias pedagógicas.

No entanto, a criação de um mapa por um usuário pode se apresentar um tanto complicada e lhe exigir certo esforço. Sabendo que este procedimento pode ser substituído, com grandes vantagens, por um método automatizado, é válido descobrir métodos para a construção automática de mapas conceituais (TSENG et al, 2007).

O objetivo da extração automática de conceitos é conceder um suporte interativo para o usuário a fim de que este comece a desenvolver o mapa conceitual a partir de uma lista de conceitos e conectores semânticos gerada automaticamente. Em uma abordagem semi-automática o usuário pode aceitar ou não os conceitos extraídos, Porém, a eficiência do método está intimamente ligada à qualidade dos texto usados como referência para a extração.

A utilidade educacional de um mapa é evidenciada quando o mapa gerado representa de forma organizada e concisa o conhecimento sobre um dado domínio. De posse deste mapa, o estudante consegue adquirir e construir novos conhecimentos por meio de relações

semânticas entre conceitos, fixando o conhecimento construído pela análise das ligações ali representadas.

Segundo Vilallon e Calvo (2008), os aspectos-chaves para a mineração de mapas conceituais são: a utilidade educacional, a simplicidade, a semi-formalidade e a subjetividade. Resumidamente, esses aspectos podem ser definidos da seguinte forma:

- a simplicidade de um mapa é definida pela quantidade de elementos interligados, onde as relações semânticas representam o conhecimento a ser produzido durante o estudo de um módulo temático. Um diagrama simples, poucos cruzamentos entre linhas e uma estrutura hierárquica facilitam a leitura do mapa;
- a semi-formalidade serve para estruturar os dados inerentes à construção de um mapa conceitual, fornecendo relações semânticas que podem ser representadas por grafos e propiciando a aplicação de algoritmos que analisam relações de distância e similaridade entre mapas;
- a subjetividade é inerente ao processo de modelagem de conhecimento prévio de um estudante a partir de um texto por ele criado. Esta técnica pode esbarrar no conhecimento linguístico e na forma de expressão escrita do indivíduo. Todavia, essa técnica pode servir para mostrar o nível de conhecimento dos estudantes sobre um tema e fornecer referências para o professor trabalhar de forma intensiva conceitos deficitários de uma turma, ou ainda, analisar o conhecimento adquirido ao final de um curso.

Conclui-se que, através do uso de ferramentas de extração de termos, é possível extrair mapas conceituais a partir de textos criados por um conjunto de estudantes e modelar o conhecimento de um grupo, com o intuito de determinar o nível de conhecimento que a turma detém sobre um tema a ser discutido (DALMOLIN et al, 2009).

O procedimento de extração de mapas conceituais a partir de documentos proposto por Vilallon e Calvo (2008) consiste em 3 etapas: 1) extração de conceitos (lista de termos que são conceitos em um mapa); 2) extração de relacionamentos (lista termos que relacionam dois conceitos); e 3) extração de topologia (forma como o mapa será organizado conforme a relevância dos conceitos e suas respectivas relações semânticas).

2.7 Modelos de Armazenamento de Mapas Conceituais – CXL

O Institute for Human and Machine Cognition (IHMC) há mais de dez anos vem desenvolvendo a ferramenta CMAP-TOOLS, uma das ferramentas mais disseminadas entre os profissionais que adotam o uso de mapas conceituais para fins educativos (NOVAK, 2008).

O *site* do IHMC permite observar que em diversos locais do mundo existem servidores de mapas conceituais onde são armazenados mapas conceituais criados com a ferramenta CMAP-TOOLS. Os locais onde existem servidores e clientes que usam este *software* podem ser vistos na Figura 3.



Figura 3- Servidores e clientes do cmap tools espalhados pelo planeta

A quantidade de locais onde o CMAP é utilizado no Brasil demonstra o interesse da comunidade brasileira na utilização de ferramentas que trabalham com mapas conceituais.

Visando distribuir mapas conceituais e o *software cliente* de forma acessível aos navegadores de internet, a equipe de desenvolvimento propôs uma notação específica para a representação e armazenamento de mapas conceituais em arquivos de marcação de dados (XML).

A extensão proposta tem nomenclatura específica para a criação de mapas conceituais e contém elementos de conexão, formatação e referenciais para armazenamento de dados sobre um mapa conceitual.

Na Figura 4 é mostrado um modelo de um arquivo que contém a estrutura de atributos com extensão CXL.

```

<xs:element name="map">
  <xs:complexType>
    <xs:all>
      <xs:element ref="concept-list" minOccurs="0"/>
      <xs:element ref="linking-phrase-list" minOccurs="0"/>
      <xs:element ref="connection-list" minOccurs="0"/>
      <xs:element ref="resource-group-list" minOccurs="0"/>
      <xs:element ref="proposition-list" minOccurs="0"/>
      <xs:element ref="not-a-proposition-list" minOccurs="0"/>
      <xs:element ref="concept-appearance-list" minOccurs="0"/>
      <xs:element ref="linking-phrase-appearance-list" minOccurs="0"/>
      <xs:element ref="connection-appearance-list" minOccurs="0"/>
      <xs:element ref="resource-appearance-list" minOccurs="0"/>
      <xs:element ref="style-sheet-list" minOccurs="0"/>
      <xs:element ref="extra-properties-list" minOccurs="0"/>
      <xs:element ref="extra-graphical-properties-list" minOccurs="0"/>
      <xs:element ref="image-list" minOccurs="0"/>
    </xs:all>
    <xs:attribute name="root-id" type="xs:string" use="optional" />
    <xs:attribute name="header" type="xs:string" use="optional" />
    <xs:attribute name="footer" type="xs:string" use="optional" />
    <xs:attribute name="width" type="xs:int" use="optional" />
    <xs:attribute name="height" type="xs:int" use="optional" />
    <xs:attribute name="default-stylesheet-id" type="xs:string"
use="optional" />
  </xs:complexType>
</xs:element>

```

Figura 4- Modelo de arquivo CXL

O formato CXL permite integrar a ferramenta CMAP-Tools com outras ferramentas por meio de *webservices* e de importação de dados seguindo esta especificação.

2 MAPAS CONCEITUAIS

Neste capítulo será apresentado o embasamento teórico das correntes pedagógicas que norteiam o desenvolvimento da ferramenta apresentada neste trabalho. Fundamentado em estudos de pesquisadores como David Ausubel, John Novak, Alfred Cañas e Helen Hanesian, a teoria da aprendizagem significativa é apresentada como uma metodologia para aprimorar a forma como professores organizam os conceitos de um curso.

A proposta deste capítulo é discutir aspectos sobre o uso de mapas conceituais como ferramenta para a promoção da aprendizagem significativa. Por fim, são apresentadas definições sobre objetos de aprendizagem e suas aplicações no processo de ensino-aprendizagem apoiado por computador.

2.1 Definições

Um mapa conceitual é um dispositivo esquemático (diagrama) para representar um conjunto de conceitos e significados embutidos em um ambiente de proposições. Em outras palavras, mapas conceituais são criados com a função de representar relações significativas entre conceitos na forma de proposições (NOVAK; GOWIN, 1984). Tais proposições são compostas por dois ou mais conceitos ligados por termos em uma unidade semântica.

A idéia para a criação desta forma de representar relações entre conceitos surgiu a partir da Teoria da Aprendizagem Significativa proposta por Ausubel et al (1980). A teoria da aprendizagem descrita por esses autores lança as bases para a compreensão de como o ser humano constrói significados. Essa teoria indica caminhos para a elaboração de estratégias pedagógicas que suscitem a aprendizagem significativa.

A teoria da aprendizagem significativa propõe uma aprendizagem facilitada por meio da associação/interconexão do novo conteúdo com os conhecimentos pré-existentes na estrutura cognitiva do estudante, tornando-o significativo para este (RISSOLI et al, 2006).

Apesar de ter sido proposta primeiramente para fins educacionais, onde a utilização desta técnica seria mediada pelo professor, a aplicação de mapas conceituais não se dá necessariamente para fins didáticos. Mapas conceituais podem ser úteis para organizar e representar o conhecimento em diversas atividades, como na preparação de uma apresentação para uma reunião ou para relacionar um conjunto de tarefas de um processo.

Considerando que a proposta inicial para a utilização de mapas conceituais foi para fins didáticos, este trabalho enfatiza o uso de mapas conceituais no processo de ensino-aprendizagem, mas não suprime o fato de que as técnicas utilizadas para construir mapas conceituais podem ser generalizadas.

A Figura 5 apresenta um exemplo de mapa conceitual cujo tema engloba conceitos sobre o uso e características desse tipo de mapas.

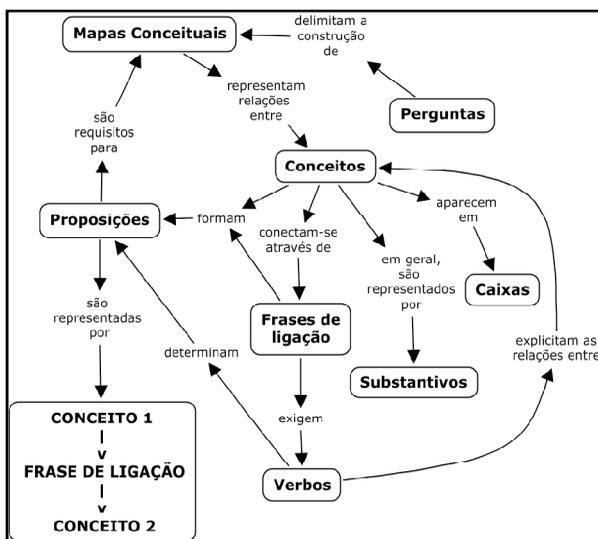


Figura 5- Exemplo de mapa conceitual (IHMC, 2009)

Um conceito é definido como uma regularidade que pode ser observada em eventos ou objetos, descrita por um rótulo, enquanto proposições ou regras são afirmações sobre um objeto ou evento de um domínio, ambos naturalmente constituídos ou construídos (NOVAK, 2008).

Os mapas conceituais induzem os usuários a identificar, organizar graficamente e “linkar” conceitos analisando informações sobre um tema. Eles constroem e explicitam a estrutura de conhecimento visualmente de uma forma conceitualmente coerente (WANG et al, 2008).

A criação de um mapa conceitual requer que o indivíduo identifique os conceitos de um domínio e os organize espacialmente de tal forma que seja possível identificar os *links* e os relacionamentos entre esses conceitos. Para Miertschin e Willis (2007), a representação de um mapa conceitual tende a mostrar o que o indivíduo conhece sobre o tema.

O mapa conceitual elaborado por um professor que deseja representar os conteúdos de um curso pode ser considerado um **organizador prévio**, que tem por finalidade facilitar e promover a aprendizagem significativa. Os conceitos e conteúdos apresentados nesses mapas têm como objetivo a criação de uma ponte cognitiva capaz de transpor o que o estudante sabe e o que o ele deverá aprender.

Mapas conceituais podem ser utilizados na compreensão de texto, cujo processo pode ser definido como a transformação de uma seqüência de palavras, sentenças e parágrafos em uma estrutura conceitual coerente, ou seja, conhecimento sintetizado.

Professores podem utilizar mapas conceituais como organizadores avançados em suas aulas. Nos trabalhos de Willerman e Harg (1991), foi constatado que os estudantes foram favorecidos pela representação visual dos relacionamentos entre os organizadores avançados durante a explanação de um tema. O mapa conceitual auxiliou os professores a descrever relacionamentos importantes entre as idéias centrais do tema juntamente com os conhecimentos prévios dos alunos.

Um mapa conceitual pode servir para mostrar o conhecimento anterior, bem como para facilitar a ligação daquilo que já é conhecido com os novos conceitos a serem trabalhados, promovendo a aprendizagem significativa. A organização dos conceitos consegue modelar a estrutura cognitiva do indivíduo no que diz respeito ao tema apresentado no mapa.

Ao dispor sob a forma gráfica de um mapa conceitual os conceitos conhecidos, estabelecendo uma hierarquia e/ou determinando propriedades, o estudante pode organizar o seu conhecimento de maneira autônoma, retificando seu próprio raciocínio em função da construção do mapa (AMORETTI, 2001).

Um mapa conceitual deve ser visto como uma das possíveis representações de uma estrutura conceitual. É uma forma de representar a organização dos conceitos de um assunto em um dado instante que é totalmente dependente da estrutura cognitiva de quem o construiu (MOREIRA; MASINI, 1982).

No entanto, avaliar e mapear a estrutura cognitiva de alguém é apenas uma das possíveis utilidades desta ferramenta pedagógica. Analisar um mapa conceitual construído por um especialista sobre determinado tema é uma forma interessante de conhecer e obter informações sobre tal tema; na medida em que são explicitadas as conexões relevantes entre os conceitos importantes, uma visão global sobre o tema é concedida pelo mapa (TAVARES, 2007).

Mapas conceituais podem ser usados para diversos fins, como por exemplo:

- representar idéias;
- descrever estruturas complexas;
- descrever relacionamentos entre diferentes componentes e processos;
- auxiliar na aprendizagem integrando o novo e o velho conhecimento;
- auxiliar especialistas a mapear um domínio;
- diagnosticar dificuldades na aprendizagem e possíveis confusões.

De acordo com Eskridge et al (2006), usuários de todos os níveis de ensino têm utilizado mapas conceituais para desenvolver novas idéias, organizar informações e formalizar o conhecimento de diversas áreas do conhecimento.

Conforme Amin (2005), o uso de mapas conceituais no ensino ajuda os professores a repensar os relacionamentos entre os conceitos-chaves de um tema. Professores devem preocupar-se em criar um diagrama onde os conceitos do tema estarão organizados coerentemente, favorecendo a aprendizagem significativa e o esclarecimento de dúvidas, evidenciando os termos mais relevantes.

3.1.1 Vantagens na utilização de Mapas Conceituais no Processo de Ensino-Aprendizagem

Diversos trabalhos já citados no atual estudo comprovam a eficiência do uso de mapas conceituais em diferentes aplicações. Enumerando as vantagens obtidas com uso desta técnica na área da educação, tem-se as seguintes características:

- enfatiza a estrutura sistematizada dos conceitos de uma disciplina e o papel dos sistemas conceituais em seu desenvolvimento;
- demonstra que os conceitos de uma disciplina diferem quanto ao grau de inclusividade e generalidade, e apresenta esses conceitos numa ordem hierárquica de inclusividade que facilita a aprendizagem e a retenção dos mesmos (MOREIRA, 1980);
- permite visão integrada do assunto e um tipo de listagem dos conteúdos abordados;
- como ferramenta de estudo facilita a revisão dos conteúdos de forma simplificada, mostrando as relações significativas dos conceitos abordados em um tema;
- permite que o mediador do processo de ensino-aprendizagem identifique deficiências conceituais nos estudantes, auxiliando na preparação de material didático para o nível de conhecimento da turma.

3.1.2 Desvantagens na utilização de Mapas Conceituais no Processo de Ensino-Aprendizagem

Todavia, esta técnica pode apresentar aspectos que podem dificultar sua aplicação, em caso de mau uso dos recursos disponibilizados por mapas conceituais. Possíveis desvantagens podem ser enumeradas da seguinte forma:

- se o mapa não fizer sentido para o estudante, pode ser encarado como algo a mais para se memorizar;

- os mapas podem ser muito complexos e confusos, o que dificultaria a aprendizagem e retenção do conteúdo, sendo que o seu papel é justamente facilitar esse processo;
- pode ocorrer inibição dos alunos em relação às suas capacidades de construir suas próprias hierarquias conceituais, uma vez que o mapa já é apresentado completamente pronto pelo professor.

As desvantagens podem ser controladas se forem seguidos alguns princípios da teoria da estrutura cognitiva proposta por Ausubel (1963). Segundo esse autor, é preciso enfatizar a finalidade do mapa no tema abordado. O professor precisa criar mapas que sejam claros e completos.

Ainda segundo Ausubel (1963), os mapas devem chamar a atenção para o fato de que podem ser traçados de várias maneiras e devem encorajar os estudantes a traçar os seus próprios mapas, ou ainda, a complementar os mapas propostos pelos professores.

3.2 Mapas Conceituais como Objeto de Aprendizagem

O *Learning Technology Standards Committee* (LTSC) do *Institute of Electrical and Electronics Engineers* (IEEE) define objetos de aprendizagem como qualquer objeto, digital ou não, que possa ser utilizado, reutilizado ou referenciado no aprendizado apoiado por tecnologia. Objetos de Aprendizagem (OA) podem ser conteúdos de multimídia, conteúdos instrucionais, *softwares* instrucionais e ferramentas de *software* (WILLEY, 2000).

OA's são entidades digitais que podem ser implementadas utilizando uma vasta gama de tecnologia de *software* (LIMA et al, 2007). Um OA é uma peça de *software* que proporciona algum tipo de experiência de aprendizagem, e pode ser aplicado ao aprendizado em diversos momentos e em diferentes cursos e situações (SILVA et al, 2008).

Um mapa conceitual, por si só, pode ser considerado um objeto de aprendizagem. Os conceitos apresentados e as conexões semânticas proporcionam uma experiência de aprendizado direta, principalmente quando os mapas elaborados pelo professor estão incluídos no contexto do tema a ser trabalhado.

A forma como um conteúdo é organizado em um mapa conceitual deve facilitar o estudante a associar novas idéias à sua estrutura cognitiva.

A ferramenta apresentada neste trabalho propõe que os estudantes, além de visualizar a relação entre os conceitos dispostos em um mapa conceitual, possam navegar por objetos de aprendizagem vinculados aos conceitos existentes no mapa. Assim, o estudante tem em mãos recursos úteis para o esclarecimento de dúvidas, construção do conhecimento e referências para pesquisas relacionadas ao conteúdo que está sendo trabalhado.

3.3 Análise Comparativa de Ferramentas Similares

Neste item serão apresentados trabalhos que têm características similares ao trabalho descrito neste documento, isto com o intuito de ressaltar aspectos que valorizam as funcionalidades apresentadas até o momento. A pesquisa por ferramentas similares concentrou-se basicamente em analisar ferramentas e pesquisas que se preocupam com o processo de construção de mapas conceituais e suas nuances.

3.4 Ferramentas para construção de mapas conceituais

Para este tipo de ferramenta foram analisadas características que facilitam a construção e a colaboração entre usuários, visando compartilhar e construir conhecimentos durante um curso presencial ou à distância.

Foram avaliadas as seguintes funcionalidades:

- interface de construção (edição de formas, *auto-layout*, idiomas disponíveis, anotações);
- forma de compartilhamento (extensões de arquivos, compartilhamento na internet, controle de acesso para edição de mapas);

- vinculação de itens multimídia (vinculação de *hyperlinks* da *web*, associação de documentos de som, vídeo, imagem);
- exportação de imagens;
- sugestão de conceitos;
- organização de mapas em pastas e anotações sobre o grupo;
- tipos de licença disponíveis, custos para aquisição da ferramenta;
- importação de diferentes formatos;
- utilização de banco de dados.

As ferramentas discutidas neste capítulo são: CMAPTOOLS, Inspiration e Verified Concept Map.

3.4.1 *CmapTools*

CmapTools é uma ferramenta para elaborar esquemas conceituais e representá-los graficamente; é um *software* que auxilia a desenhar mapas conceituais. O CMap Tools é desenvolvido pelo IHMC da Universidade da Flórida, nos Estados Unidos. É uma ferramenta gratuita disponibilizada na internet, bastando apenas que o usuário se cadastre no site dos desenvolvedores para adquiri-la.

O IHMC possui membros que desenvolveram e aprimoraram as técnicas didáticas que utilizam mapas conceituais. Dentre esses membros estão Joseph Novak e Alfred Cañas, cujos trabalhos foram citados ao longo deste estudo.

De acordo com Amoretti (2001), o CMap Tools integra o conceito de hipermídia com a agregação de mídias distintas em uma única mídia com um só objetivo, transformando cada mapa criado em um hiperdocumento, permitindo a navegação através de ilimitados links que podem ser utilizados para associar as informações, respeitando o estilo de aprendizagem do usuário.

Esse *software* vem sendo desenvolvido há cerca de uma década e já está maduro perante a comunidade científica e a diferentes usuários não acadêmicos. É possível confirmar estes dados ao verificar as bases de mapas compartilhadas que estão disponíveis na rede.

O CMapTools possui uma arquitetura cliente-servidor e permite que cada organização armazene seus mapas em um servidor próprio.

Suas principais funcionalidades residem na facilidade de transformar um mapa em uma página *web* que pode ser publicada para acesso externo, incluindo *hiperlinks* para outras páginas.

Como possui uma grande quantidade de mapas em suas bases públicas, o CMap Tools também dispõe de uma funcionalidade que sugere conceitos de acordo com os conceitos já existentes no mapa. A Figura 6 apresenta a interface de criação de mapas do CMap Tools.

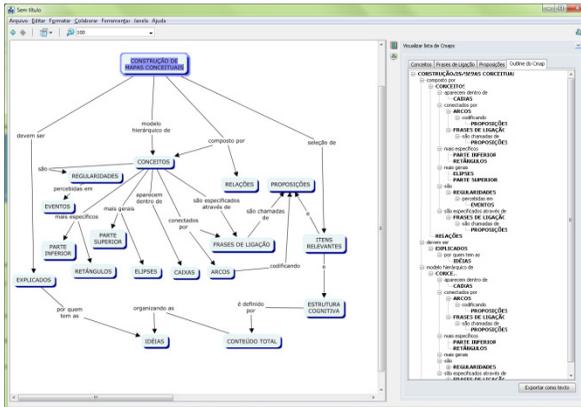


Figura 6- Interface do CMAP-tools

3.4.2 Inspiration

O *Inspiration* é um ambiente computacional criado por Don Helfgott e Mona Westhaver que proporciona ao usuário ferramentas que permitem a criação de figuras que representam idéias e concepções na forma de diagramas ou mapas.

Possui diferentes versões para aplicações, a saber: educação, negócios, organização de dados. Essa, ao contrário da ferramenta citada anteriormente, é uma ferramenta paga, cuja licença mais comum pode ser adquirida no site do fabricante sob custo de 69 dólares.

O ponto forte dessa ferramenta é a qualidade dos objetos de desenho e a variedade de opções de design que garantem gráficos bem

acabados utilizando desenhos e imagens de alta resolução nos mapas conceituais. A Figura 7 ilustra a *interface* deste *software*.

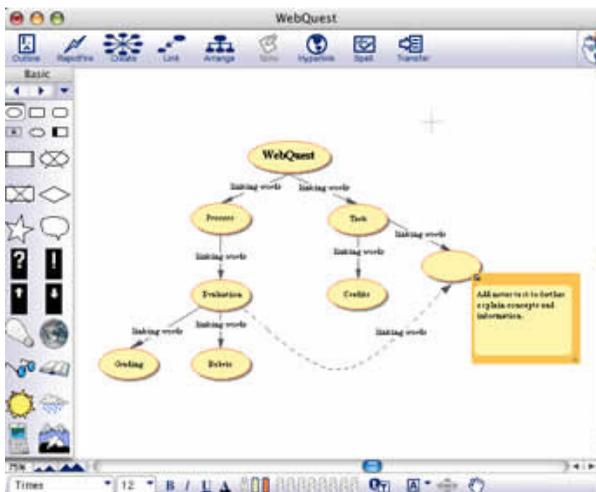


Figura 7- Interface do *software* inspiration

3.4.3 *Verified Concept Map*

Segundo Cimolino e Kay (2002; 2003), da Universidade de Sydney, Austrália, o *software* Verified Concept Map (VCM) tem por intenção disponibilizar uma ferramenta gráfica para construção de mapas conceituais de forma colaborativa.

A Figura 8, a seguir, ilustra a interface do *software* VCM.

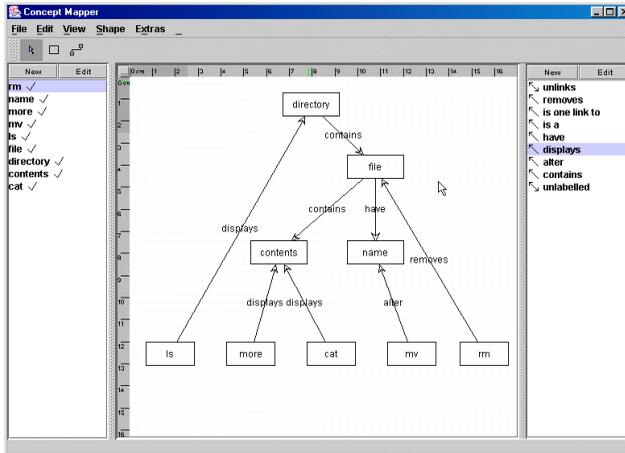


Figura 8- Interface do Verified Concept Map (VCM)

Professores e alunos podem construir seus mapas conceituais de forma iterativa. Nesses casos, o professor é capaz de criar um mapa conceitual completo sobre um tema e, ainda, disponibilizar outro mapa com menos conceitos para que os alunos o complementem conforme a evolução do aprendizado no curso.

A ferramenta disponibiliza uma funcionalidade que avalia as conexões inseridas para verificar as coerências das proposições de acordo com os conceitos previamente definidos pelo professor. Ao terminar a construção de um mapa conceitual, os estudantes são questionados sobre conexões faltantes e proposições incorretas.

Pode-se dizer que o VCM guia o estudante na construção de mapas conceituais, auxiliando-o a relacionar conceitos conforme um conjunto de proposições previamente elaborado pelo professor.

Embora a funcionalidade de conferência de mapas conceituais seja o diferencial desta ferramenta, o processo de criação do mapa conceitual de referência exige um trabalho considerável do professor, já que é necessário ter domínio completo do conteúdo.

O VCM não utiliza bancos de dados para armazenar informações, fato este que pode dificultar a forma como são persistidos os dados das entidades envolvidas no sistema.

3.4.4 Quadro comparativo entre sistemas de construção de mapas conceituais

Visando analisar as características funcionais das ferramentas para edição de mapas conceituais, foi idealizado um quadro comparativo onde estas são expostas (Quadro 4). Na parte inferior do quadro, nas duas últimas linhas, estão as contribuições propostas por este trabalho. A ferramenta MAPTRACTOR, indicada na quarta coluna, será detalhada adiante, no Capítulo 5.

Ferram. Caracterist.	CMAP TOOLS	INSPIRATION	VCM	MAPXTRACTOR
Plataforma	Java Multiplataforma	Windows	Java Multiplataforma	Windows
Interface do usuário	Intuitiva e de fácil manuseio para usuários iniciantes	Intuitiva e de fácil manuseio para usuários iniciantes	Sem muitos detalhes, funcionalidades extras complicadas	Simplificada tanto a edição de mapa quanto gerenciamento de curso
Cooperação na autoria	SIM	NÃO	Sim (professor + aluno)	SIM
Controle de Edição por tipo de usuário	SIM	NÃO	SIM	SIM
Vinculação de Objetos de Aprendizagem	SIM	NÃO	NÃO	SIM
Conectividade com BD	Sistema de arquivos próprio	NÃO	NÃO	SIM (postgreSQL)
Importação	txt, pdf, html	Txt	NÃO	NÃO
Exportação	Múltiplas	Múltiplas	NÃO	CXL (formato cmap tools)
Conectividade com a Internet	Disponibiliza mapas conectados através de servidores <i>web</i>	somente detecta url's	NÃO	Disponibiliza mapas conectados através de servidores <i>web</i>
Preço	grátis	69,00 (dólares)	Open source	Open source
Fabricante	IHMC - University of West Florida	Inspiration <i>Software</i>	Cimolino and Kay Universidade de Sidney	INE – UFSC
Organização de mapas para Cursos	NÃO	NÃO	NÃO	SIM
Extração de Termos	NÃO	NÃO	NÃO	SIM

Quadro 4 - Comparação entre ferramentas para edição de mapas conceituais

4 EXTRATOR DE TERMOS PARA CRIAÇÃO DE MAPAS CONCEITUAIS EDUCACIONAIS

O quinto Capítulo consiste na descrição detalhada da ferramenta que implementa o algoritmo para extração de conceitos e conectores semânticos. Os tópicos deste Capítulo apresentam as heurísticas utilizadas para a elaboração do método de extração de termos. Por fim, são expostas as funcionalidades da ferramenta bem como as melhores práticas para sua utilização em busca de resultados mais precisos.

Observando a forma como professores elaboram seus conteúdos, concluiu-se que, *a priori*, os docentes baseiam-se num conjunto de documentos de referência para trabalhar um tema.

É comum que os textos de referência sejam descritos formalmente em documentos que fundamentam um curso, tais como: ementas e programas de disciplina, artigos científicos e demais recursos textuais que podem ser úteis para identificar os conceitos relevantes sobre um tema.

4.1 Método para Extração de Termos Relevantes e Conectores

Fundamentalmente, o método consiste na aplicação de algoritmo em um conjunto de documentos (ementas, programas de cursos, avaliações, lista de exercícios) relacionados a um tema. é possível identificar conceitos e conectores que podem compor um mapa conceitual de um curso ou módulo temático.

A abordagem escolhida para resolver o problema de extração de conceitos é híbrida, baseada em técnicas estatísticas e linguísticas.

O objetivo do método é facilitar o trabalho do professor na criação de mapas conceituais para um curso. O algoritmo encontra relações semânticas entre conceitos e monta um mapa conceitual automaticamente a partir de textos em linguagem natural ou semi-estruturados.

O processo de extração resulta em uma lista de conceitos, uma lista de conectores e uma matriz para representar as conexões entre conceitos e conectores. Entre dois conceitos podem existir n conectores.

A ferramenta permite ao usuário adicionar e/ou excluir conceitos e relações dessa matriz. Parte-se então, de um mapa, uma vez que os conceitos mais relevantes e suas respectivas relações já estão inseridos na janela de edição de mapas (DALMOLIN et al, 2009).

A Figura 9 mostra, de forma resumida, as etapas do processo de extração de conceitos e relações semânticas entre termos de um conjunto de documentos.

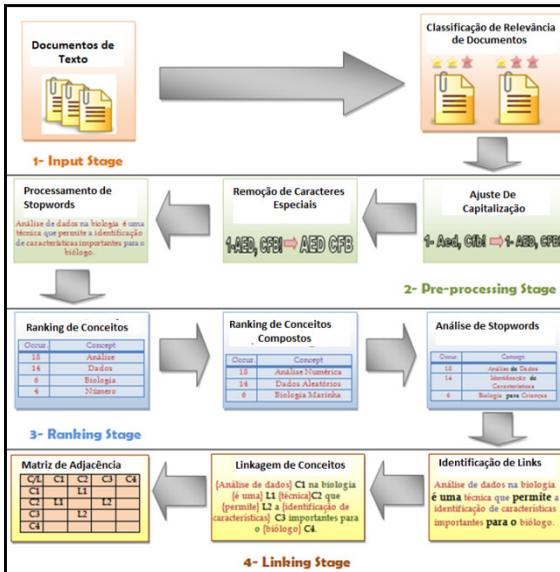


Figura 9- Processo de extração de conceitos e relações semânticas

4.1.1 Formalização do problema

O método desenvolvido tem as seguintes premissas: um **termo** é toda e qualquer palavra separada por um espaço em branco; **conceitos** podem ser formados por um ou mais termos; **conectores** são termos que estão entre dois conceitos e que geralmente possuem um verbo de ligação; e uma **conexão semântica ou proposição** é composta por dois conceitos e um conector.

Formalmente, o problema pode ser descrito da seguinte maneira: sejam D, C e L, respectivamente, um conjunto de documentos, um conjunto de conceitos e um conjunto de conectores semânticos (1):

$$D = \{d1, d2, d3..dm\}$$

$$C = \{c1, c2, c3, ci, cj..cn\} \quad (1)$$

$$L = \{l1, l2, l3, lk..lo\}$$

O algoritmo deve ser capaz de identificar um conjunto S que denota relações semânticas formadas por dois conceitos e um conector, conforme a estrutura (2):

$$S = \{\{ci, lk, cj\}, \dots\} \quad (2)$$

Com base em S, o usuário poderá criar um mapa M com as relações descritas em S. Após isto, o usuário pode editar M adicionando novos conceitos e conectores, estabelecer novas conexões semânticas e ajustar o *layout* de M para enfatizar conceitos proeminentes ao tema, propondo uma hierarquia de conceitos.

4.1.2 *Parâmetros de Entrada*

São definidos como parâmetros de entrada do extrator uma série de atributos e arquivos que servem de referência para o procedimento de extração.

O processo de extração propriamente dito inicia-se com a inclusão dos textos de referência no sistema. Esses textos são considerados o segundo parâmetro de entrada.

Para cada texto adicionado é solicitado que sejam associados um título e a fonte de onde o texto foi obtido. Estes dados permitem a indexação e referência dos conteúdos que serão trabalhados no mapa que será criado. Tais textos de referência podem ser considerados parte do material didático que serão disponibilizados para os estudantes que irão fazer um curso.

Conforme mencionado anteriormente, foi observado que, ao elaborar o material de um curso, o professor seleciona um conjunto de documentos de referência para embasar o tema a ser discutido. Esta ferramenta assume que os professores podem mensurar a **relevância** de um documento no contexto do tema a ser trabalhado.

O valor de relevância representa as seguintes características: qualidade do conteúdo, confiabilidade da fonte, adequação ao tema, linguagem clara e objetiva e formatação adequada.

Após inclusão de um texto no sistema, o usuário (professor) deve informar a relevância do texto que está inserindo. Essa relevância pode assumir um valor de 1 a 5, representado por estrelas, conforme a Figura 10. A relevância de um texto é um dos parâmetros a ser considerado e influencia diretamente na elaboração do *ranking* de conceitos.

Figura 10- Cadastro de Documento no MAPXTRACTOR.

O campo “Tipo de Documento” (Figura 10) agrega mais um detalhe sobre os documentos de referência utilizados na construção de um curso.

Tendo em vista que os cursos podem ser elaborados de forma colaborativa por dois ou mais professores, classificar o documento conforme o seu tipo pode auxiliar outros professores a entender as origens dos conceitos inseridos no mapa. O usuário pode escolher se deseja compartilhar o documento com outros usuários para permitir a colaboração na construção de um mapa.

Quando compartilhado, um documento torna-se mais uma referência para professores e estudantes utilizarem durante o processo de ensino-aprendizagem do tema.

Outros parâmetros surgem das aplicações clássicas de mineração de textos; entre estes podem ser citadas a lista de *stopwords* e a lista de termos conectores.

A lista de *stopwords* é um dos parâmetros mais importantes para o extrator de conceitos. Para que os resultados da extração de conceitos

sejam coerentes, é necessário uma lista de *stopwords* que contemple um conjunto de palavras realmente irrelevantes, cujos termos não podem ser considerados conceitos de um mapa conceitual.

A lista de *stopwords* utilizada nesta ferramenta foi adaptada a partir da lista elaborada por um conjunto de pesquisadores da Universidade do Minho (Portugal) e da Universidade de Neuchâtel (Suíça) (ROCHA, 2006). Nessa lista estão contidos termos como “este”, “aquela”, “alguns” e “antes”.

Alguns termos que não estavam contemplados na lista inicial foram adicionados (entre os quais outros pronomes e advérbios), bem como caracteres especiais. Essa lista é incluída no sistema a partir de um arquivo que pode ser editado pelo usuário. Se desejar, o usuário pode utilizar outras listas de *stopwords* que podem levar a diferentes resultados no processo de extração.

A lista padrão de *stopwords* utilizada nesta ferramenta possui 438 termos. Sugere-se que o usuário tenha pleno conhecimento das implicações acarretadas por uma troca de lista de *stopwords*.

Ao contrário do que os algoritmos de mineração de dados clássicos propõem, neste modelo as *stopwords* influenciam diretamente na identificação de conexões semânticas. A seção 4.3 demonstra como as *stopwords* influenciam nos resultados.

Para identificar as conexões semânticas, foram utilizados alguns fundamentos de análise de discurso propostos no trabalho de Travaglia (2004), onde o autor apresenta 17 verbos de ligação amplamente utilizados nas construções textuais em Língua portuguesa.

Segundo esse mesmo autor, os verbos de ligação são verbos simples que correlacionam um atributo, uma característica ou um estado a um ser ou coisa, fornecendo nuanças sobre o modo como esse atributo é percebido pelo produtor do texto sem indicar uma situação. A lista desses verbos é apresentada no Quadro 5.

Acabar	Deixar	Parecer	Tratar-se
Achar	Estar	Passar	Viver
Andar	Fazer	Permanecer	
Apresentar	Ficar	Ser	
Continuar	Mostrar	Tornar	

Quadro 5 - Lista de Verbos de Ligação

Fonte: Travaglia (2004)

Também com o intuito de aprimorar o extrator de relações semânticas, melhorando o desempenho e simplificando o modelo, optou-se por usar uma lista que contém esses verbos conjugados em todas as suas formas. Juntamente a esses verbos, em número de 17, foram adicionados 6 verbos que se provaram interessantes nos testes de validação da ferramenta. O Quadro 6 mostra os verbos adicionados à lista mencionada.

A lista de verbos de ligação contém 1.189 termos. Tal lista pode ser editada pelo usuário da mesma forma que a lista de *stopwords*. A recomendação é a mesma dada a respeito da lista de *stopwords*, ou seja, o usuário precisa conhecer o conjunto de texto para alterar os parâmetros e obter resultados mais precisos.

Consistir	Possibilitar
Permitir	Possuir
Poder	Ter

Quadro 6 - Lista de Verbos de Ligação Adicionais

Além da lista de verbos de ligação e de *stopwords*, o sistema também faz uso de uma lista de conjunções que também são consideradas no processo de análise semântica dos textos cadastrados.

Conjunções são definidas como palavras invariáveis que possuem o papel de conectar orações ou dois termos de uma mesma função sintática, estabelecendo entre eles uma relação de dependência ou de simples coordenação. São exemplos de conjunções: *portanto, logo, pois, como, mas, e, embora, porque, entretanto, nem, quando, ora, que, porém, todavia, quer, contudo, seja, conforme*.

Quando duas ou mais palavras exercem função de conjunção, dá-se-lhes o nome de locução conjuntiva. São exemplos de locuções conjuntivas: *à medida que, apesar de, a fim de que*. O objetivo do algoritmo proposto está intimamente ligado ao descobrimento de conjunções nos textos e na distância de termos por elas conectados.

Cabe ao modelo ponderar a proximidade entre conceitos, bem como a existência de termos sintáticos importantes que representam relações semânticas entre conceitos que podem ser extraídos de textos.

No item 4.3 descreveu-se a equação que pondera os valores de uma relação semântica.

A Figura 11 mostra como podem ser editadas as listas de *stopwords*, verbos de ligação e conjunções acopladas ao sistema.



Figura 11- Interface para edição de parâmetros

Outro parâmetro de entrada é a quantidade de termos que deverá ser selecionada automaticamente pelo sistema. A ferramenta seleciona os termos mais relevantes, preocupando-se em não repetir termos que estão referenciados nas composições com dois e três elementos.

Cita-se o exemplo: se o termo “ENSINO À DISTÂNCIA” for bem ranqueado na lista de termos compostos por 3 elementos e os termos “ENSINO” e “DISTÂNCIA” forem bem ranqueados na lista de termos compostos por um elemento somente, o termo “ENSINO À DISTÂNCIA” será selecionado. Esta funcionalidade seleciona o máximo de termos relevantes sem que haja repetições de um termo que está presente em mais de uma lista.

Para a elaboração da lista de relações semânticas, são utilizados os conceitos selecionados na primeira etapa do procedimento de extração. O usuário tem a possibilidade de avançar diretamente para a etapa de extração de conexões semânticas utilizando os conceitos automaticamente selecionados. Se preferir, ele pode manualmente adicionar ou excluir conceitos com um, dois ou três termos para executar a busca de relações semânticas e adicioná-las ao mapa.

Para que o usuário possa obter resultados com maior precisão, foram definidos alguns parâmetros de calibração que modificam a forma como o algoritmo analisa os textos usados na extração de conceitos.

A distância entre termos a ser analisada no processo de extração é um dos parâmetros a ser considerado. Tendo em vista que o objetivo do

algoritmo é encontrar relações semânticas entre termos de um texto, o método proposto determina que esses termos estejam separados por um número máximo caracteres.

O valor máximo de conceitos a serem extraídos é um parâmetro que pode ser configurado pelo usuário. Contudo, ele precisa estar ciente do quanto alterações no parâmetro que define a distância entre termos influenciam no resultado, ou seja, na qualidade da resposta fornecida pelo extrator.

O sistema pode ser calibrado para extrair conceitos de acordo com as características dos textos a serem analisados, como por exemplo: tamanho médio de cada texto, tamanho máximo dos agrupamentos semânticos, confiabilidade da fonte de onde o texto foi obtido.

4.2 Método de extração de conceitos

Neste tópico são detalhados os passos para a extração dos conceitos, assim como o método que determina a relevância de um conceito no conjunto de documentos de referência.

4.2.1 Pré-processamento dos textos

A etapa de pré-processamento prepara os dados de entrada para que sejam processados pela função que calcula o valor de relevância de conceitos.

Nessa etapa, todas as palavras são capitalizadas para maiúsculo com vistas a evitar ambiguidades e repetição desnecessária de termos. Os testes do sistema demonstraram que não é interessante diferenciar um termo se ele aparecer com capitalização diferente em mais de um local no texto.

Problemas com a capitalização estão principalmente relacionados a termos que iniciam frases ou a parágrafos, que são obrigatoriamente iniciados por letras maiúsculas (caixa-alta). Sendo assim, nas listas de termos extraídos, todos os termos são modificados para caixa alta. Caso

contrário, “Célula”, por exemplo, seria considerado diferente de “célula”, o que não é interessante para o extrator.

Em outro momento, são removidos os caracteres especiais presentes no texto (i.e. números, pontuação, aspas, travessão). O motivo para esta ação é que o algoritmo utiliza o caractere espaço para definir o início e o fim de um termo, portanto, se um ponto ou uma vírgula forem encontrados imediatamente após um termo, isso pode influenciar no resultado. Ex.: “termox” seria diferente de “termo x,”.

Caracteres especiais também podem ser utilizados para marcar tópicos no texto, fato que é muito comum nos materiais utilizados como textos de referência e nos experimentos realizados.

Em busca de diminuir redundâncias, outra etapa do pré-processamento é responsável por extrair sufixos de plurais nos termos extraídos. Para a extração dos plurais foi desenvolvido um artifício baseado nas técnicas de extração de radicais amplamente utilizadas em ferramentas de mineração de textos.

O último passo do pré-processamento consiste na extração e registro das *stopwords*. Nessa fase, o procedimento difere dos fundamentos de mineração de textos, uma vez que ao invés de simplesmente ignorar as *stopwords*, estas são registradas para posteriormente serem utilizadas na função de ranqueamento dos conceitos.

4.2.2 *Parsing dos Documentos*

A etapa de *parsing* é responsável por identificar todos os termos que poderão ser considerados conceitos. O sistema é capaz de identificar conceitos formados por vários termos. Para permitir a identificação de termos compostos foi utilizada uma técnica que separa os termos N-GRAMAS montando três listas que contêm elementos formados por um, dois e três termos.

Para a criação das listas de N-GRAMAS o algoritmo agrupa conceitos um a um, dois a dois e três a três, de forma incremental e respeitando trocas de linhas. Observando o texto mostrado na Figura 12 são obtidos os termos apresentados no Quadro 7.

De um modo geral, mapas conceituais, ou mapas de conceitos, são diagramas indicando relações entre conceitos, ou entre palavras que usamos

Figura 12- Trecho de texto para exemplo

unigramas	bigramas	trigramas
DE	CONCEITO SÃO	CONCEITO SÃO DIAGRAMA
MAPA	CONCEITUAL OU	CONCEITUAL OU MAPA
CONCEITO	DE	DE CONCEITO SÃO
CONCEITUAL	DE CONCEITO	DE UM
DIAGRAMA	DE UM	DE UM MODO
GERAL	DIAGRAMA INDICANDO	DIAGRAMA INDICANDO RELAÇÃO
INDICANDO	GERAL MAPA	GERAL MAPA CONCEITUAL
MODO	INDICANDO RELAÇÃO	MAPA CONCEITUAL OU
OU	MAPA CONCEITUAL	MAPA DE CONCEITO
RELAÇÃO	MAPA DE	MODO GERAL MAPA
SÃO	MODO GERAL	OU MAPA DE
UM	OU MAPA	SÃO DIAGRAMA INDICANDO
	SÃO DIAGRAMA	UM MODO GERAL
	UM MODO	

Quadro 7 - Lista de N-Gramas Extraídos

Para evitar processamento desnecessário e economizar memória de trabalho do computador, o algoritmo extrai conceitos que são considerados mal formados conforme as seguintes regras:

- dentre os unigramas, todas as *stopwords* contidas na lista de *stopwords* de referência do sistema são excluídas da lista;
- dentre os bigramas, são removidos todos os candidatos a conceitos que possuem *stopwords*;
- dentre os trigramas, são removidos todos os candidatos a conceitos que possuem uma ou mais *stopwords* que não estejam na posição intermediária do trigrama.

Ao final do processo de remoção de termos mal formados, o resultado para o texto citado anteriormente no Quadro 7 é o seguinte.

CONCEITOS
INDICANDO
PALAVRA
USAMO
DIAGRAMA INDICANDO
USAMO PARA
PALAVRA QUE USAMO
QUE USAMO PARA
RELAÇÃO ENTRE CONCEITO

Quadro 8 - Lista filtrada de N-gramas

4.2.3 *Ranking de Conceitos*

O *ranking* de conceitos tem a finalidade de listar termos organizados conforme sua relevância dentro do conjunto de textos selecionados pelo usuário. A relevância de um termo representa a tendência que determinado termo tem para ser considerado um conceito.

O *ranking* é obtido a partir de uma função aplicada para todos os termos candidatos a conceitos extraídos dos textos.

Foi elaborada uma estrutura de dados específica para o processo de extração de conceitos. Essa estrutura permite identificar atributos que representam a relevância do conceito $R(ci)$, em quantos documentos um termo candidato (Ti) está inserido; qual a relevância associada a cada um destes documentos (DR); o total de ocorrências em todos os textos ($Ti(occ)$) e o total de termos de um documento (T).

Essas informações são utilizadas para compor a função que determina o valor de relevância de um conceito no conjunto de documentos. A função é descrita na equação 3.

$$R(ci) = \frac{1}{DC_{Ti}} * DR_{Ti} * Ti_{occ} * \ln(T) \quad (3)$$

4.3 O extrator de conexões semânticas

A extração de conexões semânticas significativas de um conjunto de textos consiste em encontrar dois conceitos interligados por uma ou mais palavras que podem ser consideradas conectores semânticos.

Após identificar quais são os conceitos mais relevantes de um conjunto de textos, o algoritmo inicia uma segunda etapa onde tenta encontrar proposições formadas por conceitos relevantes que representem o domínio dos textos cadastrados.

O procedimento para encontrar conexões semânticas entre conceitos foi estabelecido conforme o fluxograma apresentado na Figura 13.

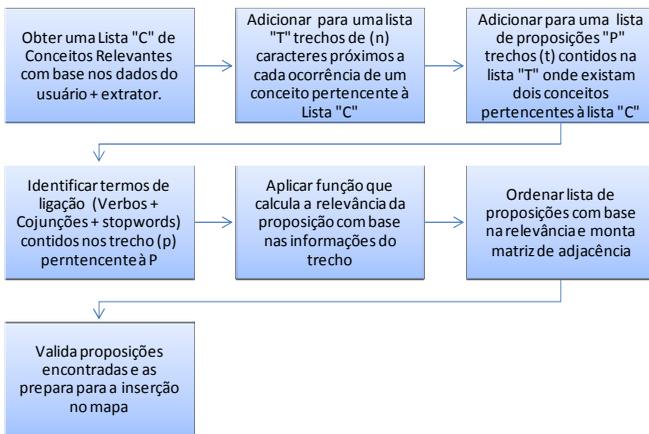


Figura 13- Procedimento para encontrar conexões semânticas

A função utilizada para calcular a relevância de um trecho candidato a uma proposição é descrita pelas equações 4 e 5:

$$\alpha = (T_c * T_v) / (s^2 + 1) * \sqrt{tam(ti)} \quad (4)$$

$$R(ti) = \frac{R_{ca} * R_{ci}}{1 - (\alpha)} \quad (5)$$

Analisando a equação 5 se observa o seguinte: $R(ca)$ e $R(ci)$ são dividendos, desta forma, quanto maior a relevância dos conceitos, mais bem ranqueada será a proposição. No divisor, o algoritmo pondera primeiro a existência de conjunções T_c e verbos de ligação T_v ; s representa a quantidade de stopwords e o tamanho da *string* $tam(ti)$.

O modelo parte do princípio que, se conceitos relevantes estiverem próximos entre si em um texto e vinculados por um verbo de ligação, a proposição tende a ser altamente relevante para o contexto dos documentos. O algoritmo utiliza um parâmetro que define a proximidade de termos a serem considerados vizinhos.

No modelo ideal concebido pelo sistema, a melhor forma para se compor é: conceito X (bem ranqueado) + verbo de ligação + conceito Y (bem ranqueado). Exemplos de proposições: (Computador possibilita Processamento) (Aluno possui Dificuldade). Todavia, nem sempre é possível extrair trechos de textos de maneira tão simplória.

Se houver mais de 1 (uma) conjunção e mais de um verbo de ligação, ou se a *string* for muito grande, o sistema está preparado para baixar a relevância desses itens na lista, isso porque o objetivo é encontrar proposições simples e objetivas com poucos termos para colocar no mapa. Este fato justifica a utilização do tamanho da *string* na fórmula.

Após a realização dos cálculos é montada uma matriz de adjacência, conforme a Figura 14, contendo os conceitos selecionados nas bordas e os *links* semânticos encontrados (que representam interseção entre estes conceitos) posicionados nas células da tabela.

Tabela de Conexões Semânticas

Clique em uma célula para exibir os conectores disponíveis para os links conectados

Lista	ALUNO	ANÁLISE	AVALIAÇÃO	CONCEITO	CONSTRUÇÃO	CONTEÚDO	DIFEREN
ALUNO		"COMPLEMENTE A"	"SERVIRÁ POSTER"	"VE PARA UM DAC"		"QUE ESTUDOU O"	
ANÁLISE	"INDIVIDUAL D"			"DE QUAIS SÃO D"		"CONCEITUAL DO"	"E VÁR"
AVALIAÇÃO	"GERAL DE TO"	"DA APRENDIZAGI"		"COMPARATIVA N"		"O MAPA CONCEIT"	
CONCEITO	"NAS DIVERSA"						
CONSTRUÇÃO	"DE CONHECIM"	"DO MAPA CONCE"		"DE UM MAPA DE"		"DE PROJETOS E I"	"DE VÁ"
CONTEÚDO	"E INCORPORA"			"DE ACORDO COM"			"FOI AE"
DIFERENTE							
ESTUDANTE				"ADQUIRE"; "NÃO I"	"COLABORA COM I"		"CONS"
FIGURA							
IDÉIA	"E A DE AVALI"		"PRINCIPAL DO US"		"DO PROTÓTIPO F"		
IMPORTANTE	"LEMBRAR QU"		"PARA ESTIMULAF"	"E QUE ELE EVIDE"	"QUE DEVE SER D"		"SALIE"
INFORMAÇÃO	"DADA PELO:"						
ORGANIZAÇÃO				"E DE EXPRESSÃO"			
PROFESSOR	"GUIE O"; "NA T"	"PRECISA FAZER I"		"NÃO PODE AFIRM"	"PODE TANTO AC"	"DU "REFORÇAR"	"SÃO M"
REPRESENTAÇÃO	"PRECISA E CC"			"MAIS COMPLETA"			
TEXTO				"QUE TRATA DA A"		"E CONSTRUINDO"	

Marque o campo "check" se deseja usar esta conexão no mapa conceitual

Check	Conceito Origem	Conector	Conceito Destino	Relevância
<input type="checkbox"/>	ALUNO	TIVESSE	CONCEITO	3,49184596
<input type="checkbox"/>	CONTEÚDO	A SER APRESENTADO AOS	ALUNO	3,25992548
<input type="checkbox"/>	AVALIAÇÃO	TEM COMO OBJETIVO POSSIBILITAR QUE O	PROFESSOR	2,82738100
<input type="checkbox"/>	IDÉIA	E A DE AVALIAR O QUE O	ALUNO	2,51986033

Mostrar Todas as Relações Inserir Relações no Mapa Sair

Figura 14- Tabela de conexões semânticas

5 MAPXTRACTOR

O MAPXTRACTOR é o nome da ferramenta que implementa a técnica proposta neste trabalho. Essa ferramenta foi desenvolvida por completo utilizando a linguagem de programação DELPHI na sua versão 7.0, em conjunto com o banco de dados Postgree SQL.

Além de um algoritmo para extração de conceitos e conectores semânticos, apresenta uma ferramenta para auxiliar no procedimento de edição e criação de mapas conceituais.

A principal aplicação do MAPXTRACTOR dá-se no processo de ensino-aprendizagem. Os casos de uso relacionam-se às atividades que o professor realiza ao organizar os conteúdos de um curso e nos objetos de aprendizagens vinculados aos conceitos que são disponibilizados aos estudantes que acessam os mapas criados pelos professores.

As atividades de cada usuário são apresentadas conforme o esquema da Figura 15. Em verde estão as atividades realizadas exclusivamente pelos professores; em vermelho, as realizadas exclusivamente pelos estudantes; em azul, as atividades permitidas para ambos.

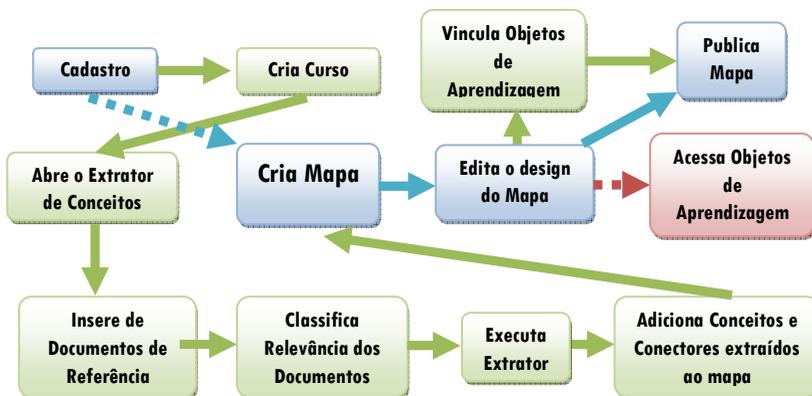


Figura 15- Atividades dos usuários no sistema

Considerando aspectos de uma atividade colaborativa de ensino-aprendizagem, a ferramenta controla o acesso de usuários para que estes possam editar recursos compartilhados. Assim, vários professores podem editar um mesmo mapa contribuindo para a construção de módulos temáticos.

As responsabilidades podem ser resumidas da seguinte forma: o professor elabora os cursos e usa a ferramenta de construção de mapas conceituais, podendo também, avaliar o andamento das atividades; os estudantes visualizam os mapas e acessam os objetos de aprendizagem vinculados aos conceitos.

5.1 Projeto do Sistema

Para representar como a ferramenta foi concebida foram utilizados alguns diagramas baseados na UML. Os diagramas apresentados neste documento descrevem o funcionamento da ferramenta e suas delimitações no que diz respeito às funções utilizadas pelo professor. Em anexo estão incluídos outros diagramas que complementam o modelo proposto.

Na Figura 16 é mostrado o diagrama de caso de uso para professores. Nesse caso de uso estão listadas todas as funcionalidades

que podem ser utilizadas por um professor ao utilizar o extrator de termos.

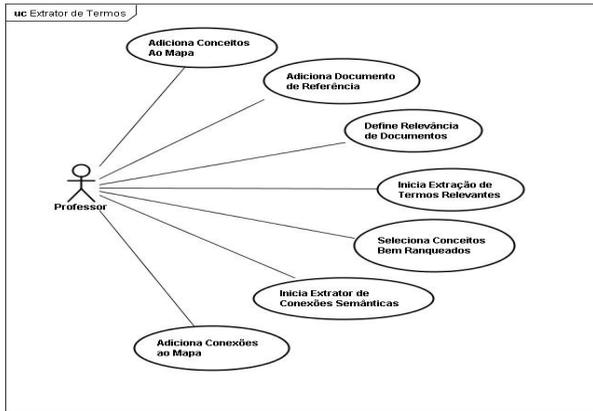


Figura 16- Caso de uso para o professor ao utilizar a ferramenta

No diagrama da Figura 17, em sequência, é definido o processo para a extração de conceitos, onde são ordenadas as ações realizadas pelo professor e pelo sistema durante a execução desse processo.

O diagrama permite identificar como ocorre o funcionamento da ferramenta, em que momento existem interações do usuário com o sistema durante o processo de extração e como o usuário, ao final do procedimento, obtém o resultado desejado.

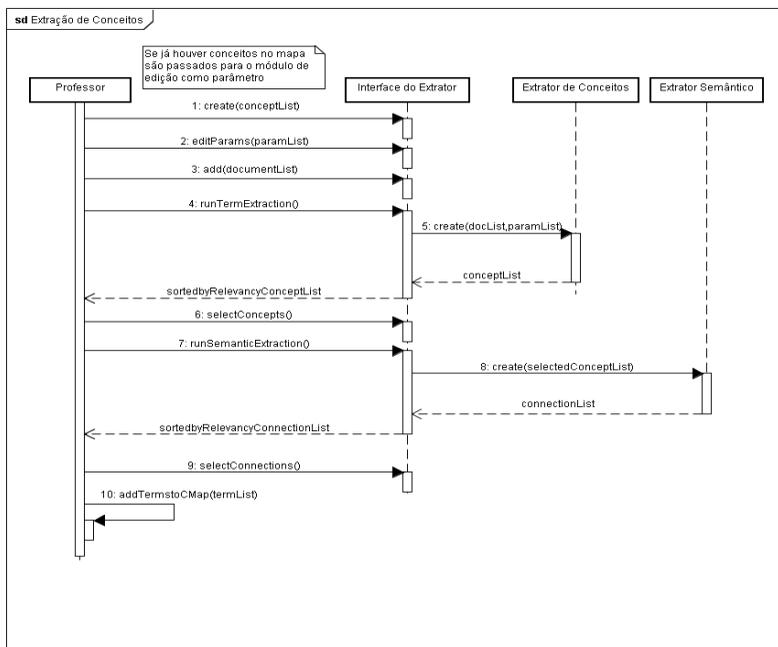


Figura 17- Diagrama de seqüência das ações realizadas pro professores e os módulos do sistema

Conforme citado anteriormente, um professor tem a responsabilidade de controlar o acesso aos cursos que cria utilizando o sistema. Então, no diagrama da Figura 18 são demonstrados os passos para a criação de um curso até a validação e publicação no ambiente colaborativo de visualização e edição de mapas conceituais.

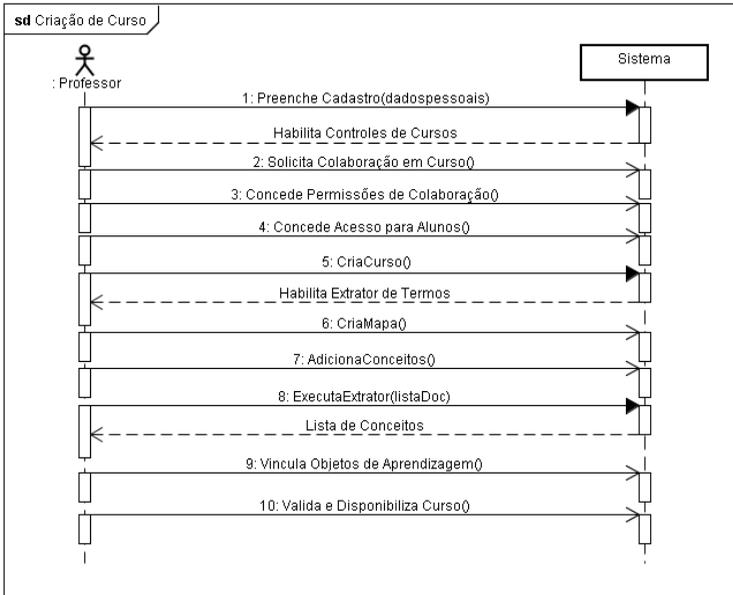


Figura 18- Diagrama de seqüência das ações realizadas pelos professores ao utilizar o extrator de termos

Outros digramas estão incluídos na seção de apêndices deste documento.

5.2 Interfaces do Sistema e Funcionalidades

Para apresentar as funcionalidades implementadas na ferramenta MAPXTRACTOR foram selecionadas algumas telas do sistema para ilustrar como acontece a interação do usuário com o sistema e como o sistema fornece suas respostas ao ser exigido.

5.2.1 Criação e controles de cursos

Ao efetuar o cadastro, o usuário pode entrar imediatamente no sistema. A *interface login* é semelhante à apresentada na Figura 19.



Figura 19- Interface de Login

Ao executar o sistema como usuário comum, o primeiro automaticamente lista quais cursos o usuário está autorizado a abrir. Na Figura 20 é apresentada a *interface* para a criação e edição de cursos.

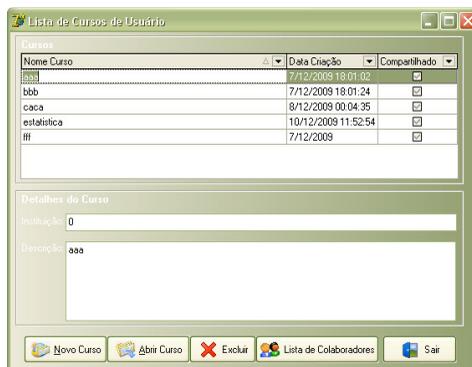


Figura 20- Interface para Criação e Edição de Cursos

Ao entrar no sistema com um usuário do tipo “outros”, o usuário consegue exibir todos os mapas criados por todos os usuários de um dado curso. Se entrar como professor, o usuário só pode acessar os

cursos que participa como colaborador. Se acessar o sistema como estudante, o usuário pode acessar os cursos para os quais recebeu permissão para ser incluído. A Figura 21 mostra a interface de seleção e edição de informações sobre mapas conceituais.

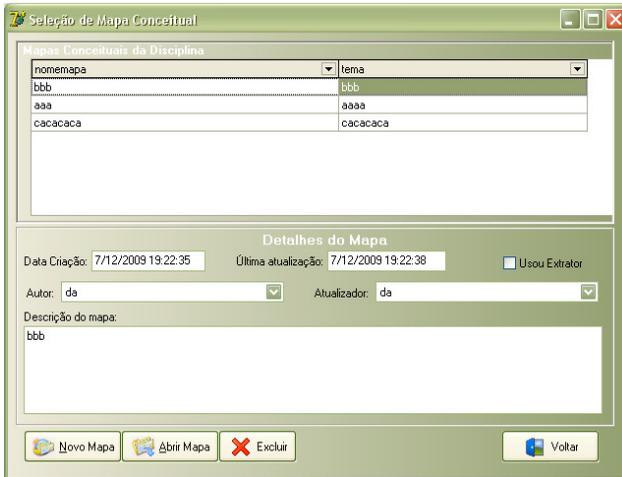


Figura 21- Interface para criação e edição de informações sobre mapas conceituais

Após definir as informações sobre o mapa conceitual ou acessar um mapa já construído, é possível exibir os conceitos e conectores semânticos contidos no mapa. Ao professor, é permitido adicionar e editar a posição dos conceitos; aos estudantes, cabe acessar os objetos de aprendizagem; aos usuários do tipo “outros”, somente visualizar os mapas. A Figura 22 apresenta a interface para edição e visualização de mapas conceituais

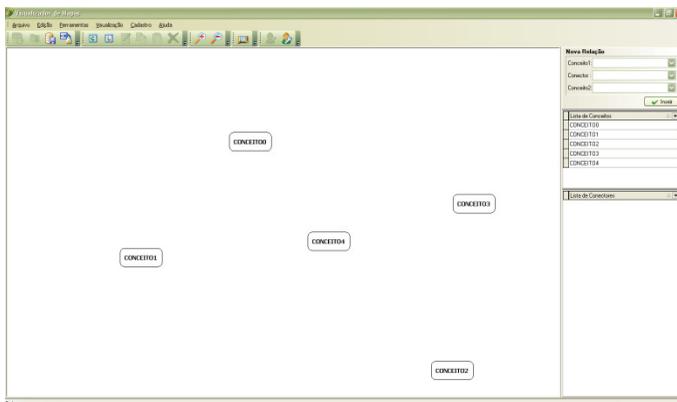


Figura 22- Interface para edição e visualização de mapas conceituais

5.2.2 Edição de Objetos do Mapa

Conceitos são objetos que podem ser inseridos no mapa por meio do extrator com um clique duplo sobre a área de desenho, ou ainda usando a tela para criação de novas proposições.

No editor é possível editar individualmente cada conceito inserido no mapa. Para cada conceito é permitido vincular n diferentes objetos de aprendizagem, classificados por tipo (vídeo, página de internet, objeto de alguma biblioteca e outros).

Quando o usuário entra no sistema com um usuário do tipo “professor”, ele pode acessar o módulo de cadastro de objetos de aprendizagem. Por outro lado, quando se conecta como estudante, ele simplesmente poderá acessar os objetos cadastrados pelo professor.



Figura 23- Interface para edição de conceito

5.2.3 Vinculando objetos de aprendizagem a um conceito

Para adicionar um objeto de aprendizagem a um conceito, o usuário precisa indicar a URL onde o objeto está publicado. Sugere-se que o usuário preencha alguns campos que indicam os objetivos da utilização de tal objeto e como ele se relaciona com os outros conceitos constantes no mapa.

Os professores podem vincular objetos disponibilizados em diferentes bibliotecas de objetos de aprendizagem ou criar novos objetos de acordo com o tema a ser trabalhado no mapa.

Cabe ainda ressaltar que o próprio mapa conceitual do curso é um objeto de aprendizagem.

5.2.4 Módulo de extração de termos e a apresentação dos resultados

Em um primeiro momento são apresentados os resultados da extração de conceitos formados por um, dois e três termos

imediatamente após o ranqueamento dos termos extraídos dos documentos de referência. Nesta tela é possível utilizar uma ferramenta que marca automaticamente termos que não se repetem nos conjuntos com um, dois e três termos. É possível observar os valores de relevância, a contagem de ocorrências nos textos adicionados e a quantidade de documentos em que o termo aparece.

Os conceitos podem ser selecionados para duas outras etapas. A primeira é a inserção direta desses conceitos no mapa conceitual sem nenhum *link* semântico, como pode ser visto na Figura 24. Outra opção seria inicializar o módulo de identificação de conexões semânticas para verificar ligações entre os termos nos textos já adicionados (ver Figura 25).

The screenshot shows the 'Extrator de Conceitos e Termos-Conectores' application. The main window is divided into several sections:

- Textos Utilizados na Extração de Conceitos:** A list of text files (e.g., texto01.TXT, texto10.TXT) and a summary of the extraction process, including the number of terms extracted with 2 or 3 words.
- Termos Extraídos:** A table listing extracted terms with their relevance, frequency, and document count.

Termos	Rele	Ocorrências	Documentos
<input checked="" type="checkbox"/> MAPA CONCEITUAL	5221,4136	498	14
<input type="checkbox"/> ESTRUTURA COGNITIVA	595,73643	61	13
<input type="checkbox"/> APRENDIZAGEM SIGNIFICATIVA	562,25242	62	12
<input type="checkbox"/> CONHECIMENTO PREVIU	209,27509	25	11
<input type="checkbox"/> PALAVRA CHAVE	163,23457	18	12
<input type="checkbox"/> CONCEITUAL CONSTRUÍDO	150,67806	36	5
<input type="checkbox"/> MAPEAMENTO CONCEITUAL	138,12156	33	5
<input type="checkbox"/> ESTRUTURA CONCEITUAL	102,54479	21	6
<input type="checkbox"/> FIGURA -	69,75835	20	4
<input type="checkbox"/> DIFERENCIAÇÃO PROGRESSIVA	58,597026	14	5
<input type="checkbox"/> FIGURA APRESENTA	58,597026	14	5
<input type="checkbox"/> ENSINO APRENDIZAGEM	50,226022	9	7
<input type="checkbox"/> REPRESENTAÇÃO GRÁFICA	50,226022	9	7
<input type="checkbox"/> REDE SEMÂNTICA	45,342937	13	4
<input type="checkbox"/> MAPA CONSTRUÍDO	44,645353	16	3
<input type="checkbox"/> INTELIGÊNCIA ARTIFICIAL	37,669517	9	5
<input type="checkbox"/> SETOR FLORESTAL	37,669517	27	1
<input type="checkbox"/> DETERMINADO ASSUNTO	33,484015	8	5

Figura 24- Interface de apresentação de resultados do extrator de conceitos

Tabela de Conexões Semânticas

Clique em uma célula para exibir os conectores disponíveis para os links conectados

Lista	ALUNO	ANÁLISE	AValiação	CONCEITO	CONSTRUÇÃO	CONTEÚDO	DIFERENTE
ALUNO		"COMPLEMENTE A"	"SERVIRÁ POSTER"	"VE PARA UM DAC"		"QUE ESTUDOU O"	"E VÁR"
ANÁLISE	"INDIVIDUAL D"			"DE QUAIS SÃO D"		"CONCEITUAL DO"	"E VÁR"
AValiação	"GERAL DE TO"	"DA APRENDIZAGI"		"COMPARATIVA N"		"O MAPA CONCEI"	
CONCEITO	"NAS DIVERSA"						
CONSTRUÇÃO	"DE CONHECIN"	"DO MAPA CONCE"		"DE UM MAPA DE"		"DE PROJETOS E I"	"DE VÁ"
CONTEÚDO	"E INCORPORA"			"DE ACORDO COM"			"FOI AE"
DIFERENTE							
ESTUDANTE				"ADQUIRE"	"NÃO I"	"COLABORA COM I"	"CONS"
FIGURA							
IDÉIA	"E A DE AVAlIA"		"PRINCIPAL DO US"		"DO PROTÓTIPO F"		
IMPORTANTE	"LEMBRAR QU"		"PARA ESTIMULAF"	"E QUE ELE EVIDE"	"QUE DEVE SER D"		"SALIE"
INFORMAÇÃO	"DADA PELO"						
ORGANIZAÇÃO				"E DE EXPRESSÃO"			
PROFESSOR	"GUIE O"	"NA T"	"PRECISA FAZER I"	"NÃO PODE AFIRM"	"PODE TANTO ACI"	"OU "REFORÇAR"	"SÃO N"
REPRESENTAÇÃO	"PRECISA E CC"			"MAIS COMPLETA"			
TEXTO				"QUE TRATA DA A"		"E CONSTRUINDO I"	

Marque o campo "check" se deseja usar esta conexão no mapa conceitual

Check	Conceito Origem	Conector	Conceito Destino	Relevância
<input type="checkbox"/>	ALUNO	TIVESSE	CONCEITO	9,49184596
<input type="checkbox"/>	CONTEÚDO	IA SER APRESENTADO AOS	ALUNO	3,25992548
<input type="checkbox"/>	AVAlIAÇÃO	TEM COMO OBJETIVO POSSIBILITAR QUE O	PROFESSOR	2,82738100
<input type="checkbox"/>	IDÉIA	F A DE AVAlIAÇÃO QUE O	ALUNO	1,51999999

Mostrar Todas as Relações Inserir Relações no Mapa Sair

Figura 25- Interface de exibição e seleção de conexões semânticas extraídas

6 TESTES E VALIDAÇÃO DO TRABALHO

Neste capítulo serão demonstrados os testes realizados no presente trabalho com a finalidade de avaliar a aplicabilidade do método, que visa implementar uma ferramenta para utilização efetiva na construção e representação de temas que podem ser trabalhados em disciplinas de cursos de qualquer área do conhecimento.

6.1 Metodologia de Avaliação e Resultados Esperados

O modelo proposto deve possibilitar ao usuário a obtenção de um mapa conceitual extraído automaticamente de um conjunto de textos. O resultado advém de uma lista de conceitos relevantes e conectores semânticos obtidos por meio de um algoritmo de extração de termos. A validação do trabalho busca avaliar os resultados quantitativa e qualitativamente.

Na avaliação qualitativa, o objetivo é identificar se os conceitos extraídos realmente apresentam o domínio do tema proposto pelo usuário. Tal tema deve estar em conformidade com os textos utilizados como entrada no extrator de termos. Por fim, é esperado que o mapa resultante do processo represente o conhecimento extraído dos textos analisados.

Devido à característica subjetiva das relações propostas em um mapa, a idéia é tomar como base afirmações de autores que confirmam as proposições encontradas pelo método.

Na avaliação quantitativa da ferramenta foi analisada a eficiência do método ao avaliar e ranquear os termos. Foram analisados o cálculo de relevância de conceitos e a relevância das conexões semânticas extraídas pela ferramenta.

Quantitativamente, os resultados serão comparados com a técnica de extração estatística que considera apenas a quantidade de ocorrências do termo no conjunto de documentos.

Conforme detalhado anteriormente, ao invés de usar simplesmente o número de ocorrências de um termo no conjunto de documentos, o modelo proposto pressupõe que nem sempre um termo

que não é uma *stopword* e ocorre muitas vezes é mais relevante do que outro termo que ocorre ($n - m$) vezes.

Espera-se que pela utilização dessa ferramenta, com base no método proposto, seja possível obter respostas mais interessantes do que os métodos de extração de termos baseados em ocorrências.

Seguindo a premissa de que o professor pode definir um valor de relevância para um texto de entrada, é esperado, ainda, que esse valor seja capaz de mostrar os interesses do professor e a importância de se utilizar textos conhecidos na busca de representação do conhecimento.

6.2 Cenário de teste para modelo proposto

Na preparação deste cenário optou-se por mostrar como a ferramenta pode ser útil na preparação de um curso extraindo os conceitos que tendem a ser relevantes a partir de documentos de texto obtidos na *internet*.

O cenário tem como premissa o fato de que o professor, ao preparar o conteúdo para um curso, utiliza *sites* de busca para encontrar documentos que sirvam de referencial teórico para o tema a ser lecionado. Levando isto em conta, a coleta de dados se baseou em três buscadores disponíveis na *internet*: Google, Yahoo e Bing.

Em cada um dos buscadores foi realizada a busca para a seguinte frase: “Mapas Conceituais e Aprendizagem Significativa”. Foram selecionados os primeiros resultados válidos obtidos pelos buscadores e considerados válidos documentos onde o texto poderia ser extraído para um arquivo de texto simples para posteriormente ser utilizado no MapXtractor.

Google: 18900 resultados
http://www.cienciasecognicao.org/pdf/v14_2/m006_09.pdf
http://www.bomjesus.com.br/publicacoes/pdf/revista_PEC/teoria_da_aprendizagem.pdf
http://pt.wikipedia.org/wiki/Mapa_conceitual
http://omnis.if.ufrj.br/~marta/aprendizagememfisica/mapasconceituais.pdf
http://www.cienciasecognicao.org/pdf/v13_2/m318261.pdf
http://ambiente.aied.com.br/UploadFiles/1_T0028-2.pdf
http://rived.mec.gov.br/artigos/2003-ICIPC.pdf
http://www.cienciamao.if.usp.br/tudo/exibir.php?media=eneq&cod=_utilizandomapasconceitua
http://www.quimica.ufpr.br/eduquim/eneq2008/resumos/R0373-1.pdf
http://juvenalneto.blogspot.com/2009/05/mapas-conceituais-e-aprendizagem.html

Quadro 9 - Documentos selecionados a partir do GOOGLE

Bing: 6.860 resultados
http://www.fae.ufmg.br/abrapec/viempec/CR2/p251.pdf
http://www.if.ufrgs.br/~moreira/mapasport.pdf
http://mapasconceituais.cap.ufrgs.br/mapas.php
http://www.cienciasecognicao.org/revista/index.php/ce/article/viewArticle/117
http://www.cienciamao.if.usp.br/tudo/exibir.php?media=eneq&cod=_utilizandomapasconceitua
http://www.tvebrasil.com.br/salto/bolefins2005/nfa/teetxt5.htm
http://pt.wikipedia.org/wiki/Mapa_conceitual
http://cmapspublic3.ihmc.us/rid=1178300225125_629576139_11663/mapasport.pdf
http://aveb.univap.br/opencms/sites/ve2007neo/pt-BR/imagens/27-06-07/Escola/trabalho_70_eloiza_anais.pdf
http://penta2.ufrgs.br/edutools/mapasconceituais/utilizamapasconceituais.html

Quadro 10 - Documentos selecionados a partir do BING

Yahoo: 18.800 resultados
http://penta.ufrgs.br/~luis/Ativ2/mapas2.html
http://www.cienciamao.if.usp.br/tudo/exibir.php?media=eneq&cod=_utilizandomapasconceitua
http://www.cienciamao.if.usp.br/tudo/exibir.php?media=eneci&cod=_utilizacaodemapasconcei
http://www.feg.unesp.br/~teiadosaber/AprendizagemSignificativa.doc
http://penta.ufrgs.br/~luis/Ativ2/mapas_mara.html
http://ticaki.blogspot.com/2009/11/mapa-conceitual_03.html
http://www.cienciasecognicao.org/pdf/v12/m347181.pdf
http://www.caxias.rs.gov.br/geemac/_upload/encontro_72.pdf
http://www.lsc.ufsc.br/~edla/ine5624/materiais/MapasConceituaisCardoso.doc
http://www.fisica.ufpb.br/~romero/pdf/2008MapasPrincipia.pdf

Quadro 11 - Documentos selecionados a partir do Yahoo Search

Após a extração dos textos das páginas que estavam disponíveis para o público, foi verificado que alguns trechos precisariam ser

retirados dos textos porque poderiam influenciar negativamente nos resultados.

Uma vez que a maioria dos documentos eram artigos científicos publicados em eventos e revistas, houve a necessidade de extrair *abstracts* e cabeçalhos de páginas, considerados trechos irrelevantes para o processo. Outros trechos retirados foram: bibliografias, identificação de autores, datas, tabelas e figuras, restando somente os trechos de texto “puro”, com informações sobre o tema trabalhado.

Como o modelo de extração de termos busca priorizar conceitos existentes em documentos marcados como relevantes, optou-se por definir a relevância dos documentos extraídos do Google com valor 4 estrelas, do Yahoo com 3 estrelas e do BING com 2 estrelas. Esta abordagem baseou-se na quantidade de resultados que cada um dos buscadores retornou.

6.2.1 *Análise Quantitativa dos Resultados*

O primeiro quesito avaliado foi o tempo de reposta da ferramenta para retornar os resultados da extração. Apesar de não ser o foco da análise, o tempo de resposta permite comprovar a viabilidade computacional do método.

O processo de extração foi realizado em computador AMD PHENON (2.8Ghz de clock) com Windows 7, e para esta máquina o sistema apresentou as seguintes informações:

- 30 documentos analisados;
- 5.173 termos com 1 palavra (unigramas);
- 23.413 termos com 2 palavras (bigramas);
- 32.794 termos com 3 palavras (trigramas);
- Total de termos extraídos = 61.380;
- 441 Stopwords Carregadas;
- 87 Conjunções Carregadas;
- 1.189 Termos de verbos Conjugados;
- 44.395 termos foram retirados das Listas por formação incorreta;
- Total de termos restantes = 16.985;
- Termos extraídos após tratamento de plural = 220.

Ao final do processo de extração de termos restaram 16.765 termos, e destes:

- 4.760 termos com 1 palavra;
- 6.242 termos com 2 palavras;
- 5.763 termos com 3 palavras;
- Tempo Total: 00:06:409ms.

De acordo com Tavares (2006), um mapa conceitual não deve conter muitos conceitos para que não se torne muito confuso. Dessa forma, a ferramenta seleciona automaticamente 25 conceitos melhor ranqueados. Todavia, como são extraídos conceitos uni, bi e trigramas, uma heurística é utilizada para selecionar os conceitos sem repetir os termos, sendo 20% trigramas, 30% bigramas, 50% unigramas.

Preocupado com a possibilidade de personalização, o sistema permite que o usuário selecione outros conceitos antes de iniciar o procedimento de identificação de conexões semânticas. As tabelas 1, 2 e 3 mostram os resultados obtidos automaticamente. Os termos selecionados estão marcados com um X na coluna “Seleção” de cada tabela.

Tabela 1- Lista de unigramas extraídos no cenário de teste

Seleção	Conceito	Relevância	Ocorrências	Documentos
0	MAPA	2819,957888	892	27
X	CONCEITO	2696,663765	853	28
0	CONCEITUAL	2140,259518	677	27
0	APRENDIZAGEM	1817,120526	619	27
X	ALUNO	1316,943818	432	27
0	SIGNIFICATIVA	1141,260984	361	28
X	CONHECIMENTO	804,3473733	274	26
0	ESTRUTURA	716,9575829	254	26
X	SIGNIFICADO	514,177139	198	21
0	COGNITIVA	536,3068533	190	26
X	ENSINO	443,1588209	157	23
0	PROCESSO	392,1249898	151	23
X	CONTEÚDO	408,0448354	139	24
X	INFORMAÇÃO	364,1241267	129	25
X	AUSUBEL	327,8810741	121	26
X	FIGURA	252,9110214	112	18
X	PROFESSOR	280,4602576	108	21
0	TEORIA	270,0728407	104	22
X	CONSTRUÇÃO	165,9728578	98	16
0	AULA	219,0390096	97	19
X	IDÉIA	225,2488784	95	21
0	PALAVRA	244,1042983	94	21
0	MOREIRA	123,2941229	91	14
X	TEMA	170,7149394	84	17
0	ESTUDANTE	153,5531201	80	16
0	APRENDIZ	176,1344613	78	16
X	AVALIAÇÃO	217,345409	77	23
0	DIFERENCIAÇÃO	127,0200442	75	13
0	DIFERENTE	144,2947702	71	15
0	PROGRESSIVA	120,2456419	71	13
0	THE	55,32428592	70	8
0	INTERAÇÃO	138,1978081	68	16
0	MCF	29,80737037	66	2

Tabela 2- Lista de bigramas extraídos no cenário de teste

Seleção	Conceito	Relevância	Ocorrências	Documentos
X	MAPA CONCEITUAL	2819,958	559	27
X	APRENDIZAGEM SIGNIFICATIVA	1320,977	282	27
X	ESTRUTURA COGNITIVA	743,1847	165	26
X	DIFERENCIAÇÃO PROGRESSIVA	133,6832	53	12
X	PALAVRA CHAVE	166,4734	44	17
0	RECONCILIAÇÃO INTEGRATIVA	108,4599	43	11
0	CONCEITO BRUTO	22,34058	31	2
0	CONCEITO DIFERENTE	20,89925	29	2
0	SIGNIFICATIVA CRÍTICA	20,89925	29	3
0	CONHECIMENTO PRÉVIO	100,8929	28	17
0	ANIMAÇÃO INTERATIVA	16,57527	23	1
0	APRENDIZAGEM MECÂNICA	35,67286	22	8
0	CIÊNCIA COGNIÇÃO	7,927303	22	2
0	DAVID AUSUBEL	54,77046	19	16
0	MAPEAMENTO CONCEITUAL	30,80838	19	6
0	CERTO CONTEÚDO	30,62822	17	6
0	FIGURA MAPA	23,06125	16	7
0	POTENCIALMENTE SIGNIFICATIVO	29,72739	15	12
0	HIERARQUIA CONCEITUAL	16,21494	15	6
0	CONCEITO RELEVANTE	25,22324	14	10
0	ESTRUTURA CONCEITUAL	25,22324	14	9
0	FASE INICIAL	10,08929	14	2
0	ENSINO APRENDIZAGEM	35,13237	13	12
0	APRENDER SIGNIFICATIVAMENTE	30,44805	13	9
0	CONCEITO CHAVE	16,3951	13	7
0	ASPECTO RELEVANTE	21,80008	11	8
0	ORGANIZADOR PRÉVIO	15,85461	11	8
0	MODELO MENTAL	11,89095	11	2
0	DIFERENCIAÇÃO PROGRESSIVA	3,963652	11	1
0	PRE TESTE	3,963652	11	1
0	JOSEPH NOVAK	18,0166	10	9
0	CONCEITUAL “CORRETO”	14,41328	10	5
0	MENO INCLUSIVO	14,41328	10	8
0	MUITA VEZE	14,41328	10	5
0	ESTRUTURA HIERÁRQUICA	10,80996	10	7

Tabela 3- Lista de Trigramas obtidos no cenário de teste

Seleção	Conceito	Relevância	Ocorrências	Documentos
0	APRENDIZAGEM SIGNIFICATIVA CRÍTICA	1090,38	29	3
X	TEORIA DA APRENDIZAGEM	2368,76	21	11
X	PROCESSO DE APRENDIZAGEM	2819,96	20	13
X	SALA DE AULA	2819,96	20	12
X	USO DE MAPA	2819,96	20	14
0	TEORIA DE AUSUBEL	1203,18	16	7
0	HUMANIZAÇÃO DE ATITUDE	601,59	16	2
X	COGNITIVA DO APRENDIZ	1409,98	15	9
0	CONCEITO MAIS INCLUSIVO	986,99	15	7
0	FIGURA MAPA CONCEITUAL	986,99	15	6
0	CONSTRUCAO DOS MAPA	263,2	14	1
0	CONJUNTO DE CONCEITO	977,59	13	8
0	PROGRESSIVA E RECONCILIAÇÃO	902,39	12	7
0	TEORIA DE APRENDIZAGEM	676,79	12	10
0	CURSO DA APRENDIZAGEM	563,99	12	3
0	AVALIACAO DA APRENDIZAGEM	188	10	1
0	ENSINO DE CIENCIA	188	10	1
0	PROFESSOR E ALUNO	761,39	9	8
0	CONCEITO MAIS ABRANGENTE	676,79	9	7
0	MAPA CONCEITUAL “CORRETO”	676,79	9	5
0	AVALIAÇÃO DA APRENDIZAGEM	592,19	9	6
0	CONTEXTO DA MATÉRIA	601,59	8	5
0	CORPO DE CONHECIMENTO	601,59	8	6
0	ESTABILIDADE E DIFERENCIAÇÃO	526,39	8	5
0	GRAU DE CLAREZA	526,39	8	5
0	PROCESSO DE ENSINO	451,19	8	7
0	BÁSICO DA TEORIA	375,99	8	4
0	TÉCNICA DE CONSTRUÇÃO	375,99	8	5
0	PALAVRA DE LIGAÇÃO	300,8	8	6
0	MAPA DA FIGURA	225,6	8	3
0	MAPA CONCEITUAL CONSTRUIDO	150,4	8	1
0	DIAGRAMA DE FLUXO	789,59	7	7
0	DURANTE A APRENDIZAGEM	592,19	7	7
0	COGNITIVA ESTÁ CONSTANTEMENTE	526,39	7	5
0	CONCEITO E PROPOSIÇÃO	526,39	7	6

A análise quantitativa se propõe a avaliar a eficiência do método ao definir o *ranking* dos termos de acordo com a relevância a eles calculada. O método é comparado ao *ranking* de termos que considera somente as ocorrências dos termos em todos os documentos.

Para este fim, foram elaborados gráficos que comparam o *ranking* de termos onde são consideradas a quantidade de ocorrências destes e o *ranking* de termos onde são considerados os valores de relevância calculados pelo MapXtractor.

As figuras 26, 27 e 28 mostram o comportamento das variáveis que determinam o *ranking* de ocorrências comparando-o ao *ranking* obtido com o valores de relevância calculado pelo MapXtractor.

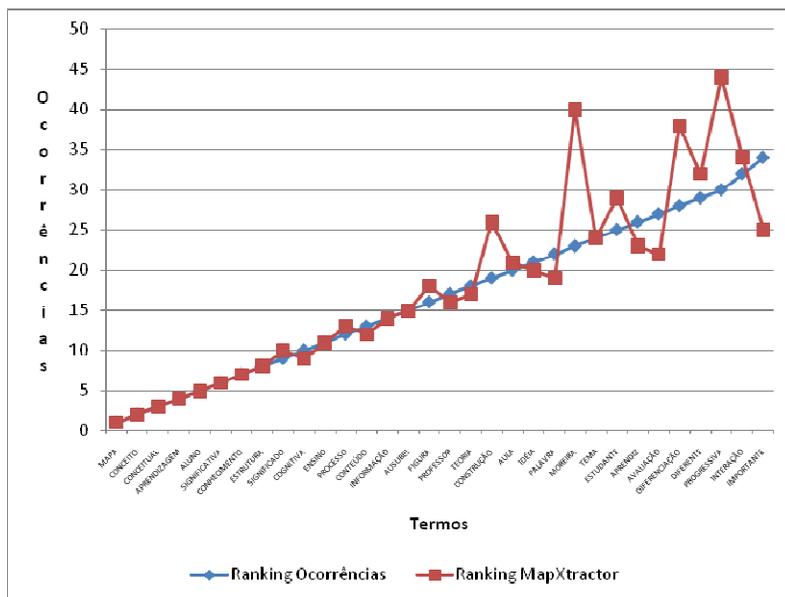


Figura 26- Comparação de *Rankings* para Unigramas

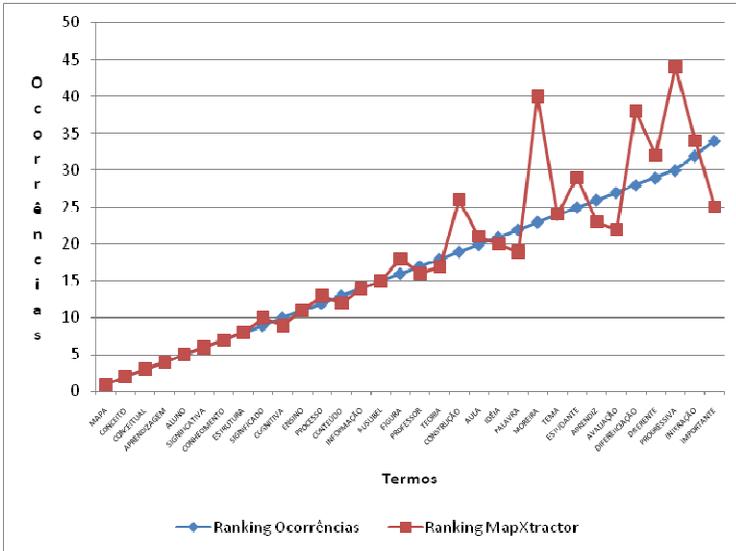


Figura 27- Comparação de Rankings para Bigramas

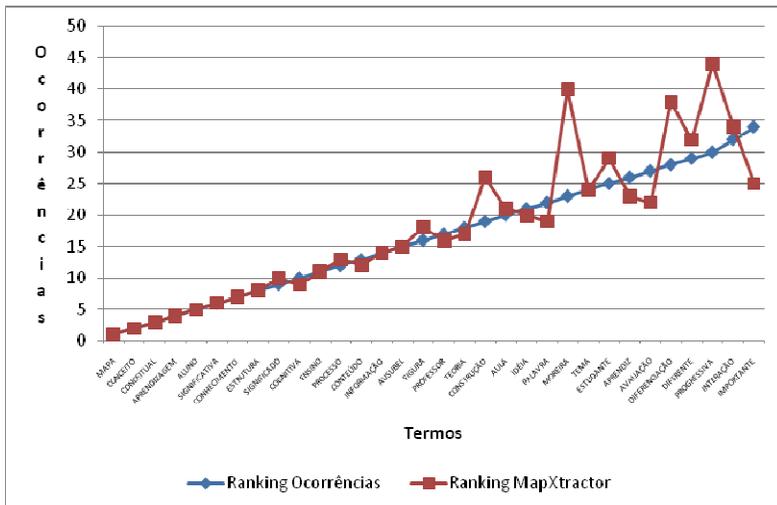


Figura 28- Comparação de Rankings para Trigramas

Os gráficos comparativos entre os n-gramas extraídos das duas formas propostas no teste (relevância x ocorrências) mostram-se similares quando são analisados os termos do topo do *ranking*. Entretanto, os mesmos resultados vão se tornando discrepantes quando o número de ocorrências começa a diminuir.

Os resultados evidenciam que termos que aparecem em poucos documentos vão recebendo valores inferiores. Nota-se isto para o conceito (APRENDIZAGEM SIGNIFICATIVA CRÍTICA) que apesar de aparecer 29 vezes no conjunto de textos, é citado em apenas três documentos. Esse fato acaba fazendo com que sua relevância seja consideravelmente reduzida.

Uma das premissas do trabalho é que o usuário, ao construir seus mapas, poderá utilizar documentos conhecidos ou não e quando desejar pode priorizar termos que constem em documentos que considere mais relevante.

Após o ranqueamento de termos e a seleção dos conceitos, foi realizado o processo de identificação de termos e/ou trechos considerados conexões semânticas. Esta etapa depende totalmente dos conceitos selecionados no passo anterior.

Como entrada para o método de extração de conexões semânticas, foram utilizados os conceitos marcados com um “X” nas tabelas 1, 2 e 3. O sistema automaticamente seleciona os conceitos de acordo com o *ranking* e a coexistência de um termo nas listas de unigramas, bigramas e trigramas.

Tabela 4- Proposições extraídas pelo MapXtractor

Conceito Origem	Conector Semântico	Conceito Destino	Relevância da Proposição	Classificação
ALUNO	TIVESSE	CONCEITO	10,14	Válida
CONHECIMENTO	A QUE SE RELACIONA O	MAPA CONCEITUAL	5,1	Válida
ALUNO	APRESENTE NA	AVALIAÇÃO	4,65	Válida
ALUNO	APRESENTA É O SEU MAPA E O	IMPORTANTE	3,98	Errônea
ENSINO	BUSCA FAZER COM QUE O	ALUNO	3,36	Incompleta
IMPORTANTE	QUE INFLUENCIA A APRENDIZAGEM É O QUE O	ALUNO	3,21	Errônea
APRENDIZAGEM SIGNIFICATIVA	É INERENTE À	ESTRUTURA COGNITIVA	2,71	Válida
PROFESSOR	AO ENSINAR TEM A INTENÇÃO DE FAZER COM QUE O	ALUNO	2,6	Incompleta
CONTEÚDO	ESTÃO LIGADOS A	CONCEITO	2,45	Incompleta
AVALIAÇÃO	QUE PREMIAM A	APRENDIZAGEM SIGNIFICATIVA	2,1	Válida
CONHECIMENTO	É ARMAZENADO NA	ESTRUTURA COGNITIVA	1,73	Válida
MAPA CONCEITUAL	É UM ESTRUTURADOR DO	CONHECIMENTO	1,6	Válida
IMPORTANTE	É QUE O MAPA SEJA UM INSTRUMENTO CAPAZ DE EVIDENCIAR	SIGNIFICADO	1,52	Válida
PROFESSOR	APRESENTAR PARA O	ALUNO	1,34	Errônea
PROFESSOR	NUNCA DEVE APRESENTAR AOS	ALUNO	1,34	Incompleta
MAPA CONCEITUAL	DE MODO QUE ORDENASSEM DOS	CONCEITO	1,32	Errônea
TEMA	PODERÁ FACILITAR AO	ALUNO	1,3	Válida
ESTRUTURA COGNITIVA	ESTÁ CONSTANTEMENTE SE REORGANIZANDO POR	DIFERENCIAÇÃO PROGRESSIVA	1,03	Válida
ALUNO	PROVAVELMENTE TERÁ AQUELE	CONHECIMENTO	0,99	Errônea
APRENDIZAGEM SIGNIFICATIVA	OCORRE QUANDO UMA NOVA	INFORMAÇÃO	0,88	Incompleta
APRENDIZAGEM SIGNIFICATIVA	ATÉ QUE UMA ESTRUTURA DE	CONHECIMENTO	0,87	Errônea
APRENDIZAGEM SIGNIFICATIVA	DE DAVID	AUSUBEL	0,87	Válida
MAPA CONCEITUAL	DE UM CERTO	CONTEÚDO	0,83	Válida
AUSUBEL	DESTACA COMO SÃO	IMPORTANTE	0,82	Errônea
ALUNO	É CAPAZ DE ATRIBUIR NOVOS	SIGNIFICADO	0,81	Válida
ALUNO	ESTÁ APRENDENDO SIGNIFICATIVAMENTE O	CONTEÚDO	0,81	Válida
SIGNIFICADO	DA RELAÇÃO QUE VÊ ENTRE ESSES	CONCEITO	0,76	Errônea
PROFESSOR	É AMBOS DETECTAREM ONDE ESTÃO AS DEFICIÊNCIAS DE	ENSINO	0,69	Válida
AUSUBEL	CHAMA DE	CONCEITO	0,69	Errônea
ALUNO	DEVE SER CAPAZ DE APLICAR ESSE	CONHECIMENTO	0,62	Válida
TEMA	DE ESTUDO A SER ABORDADO TRAZ COM	CONCEITO	0,62	Errônea
INFORMAÇÃO	ADQUIRE	SIGNIFICADO	0,56	Válida
ESTRUTURA COGNITIVA	CHAMASE	DIFERENCIAÇÃO PROGRESSIVA	0,56	Errônea
MAPA CONCEITUAL	PODE ATUAR COMO ESTRUTURADOR GLOBAL DO	CONHECIMENTO	0,55	Válida
INFORMAÇÃO	DADA PELO	ALUNO	0,55	Incompleta

Avaliando as proposições extraídas, foram selecionadas as válidas e incompletas para serem adicionadas ao mapa. O mapa da Figura 29 resultou desse procedimento.

usuário. As incompletas são proposições que podem ser inseridas no mapa, mas que dependem de uma nova ligação para fechar uma idéia.

As proposições errôneas são aquelas que não correspondem a uma afirmação interessante que caiba no mapa conceitual.

Observando as classificações verificou-se que a maioria delas (67%) são válidas ou incompletas, e de acordo com os propósitos do cenário de validação, o mapa conceitual gerado pela ferramenta é válido, já que as proposições extraídas refletem o domínio do problema analisado.

6.3 Discussões sobre os Resultados

O Capítulo 6 apresentou os resultados obtidos pelo método de extração de conceitos e conectores semânticos e como os seus resultados podem auxiliar na produção de cursos utilizando princípios da aprendizagem significativa.

Primeiramente, cabe citar Kozareva et al (2009), que em seu trabalho ressaltam que não existem bases de dados de referência para comparar efetivamente os resultados de algoritmos para extração de termos, principalmente no que diz respeito à extração de mapas conceituais aplicados na educação. Tal afirmação confirma a complexidade do problema.

Na literatura são encontradas várias metodologias aplicadas para comprovar a eficácia de algoritmos para extração de termos. O Capítulo 2 abordou alguns trabalhos que relacionam algumas dessas metodologias. Porém, nenhuma delas pode ser reproduzida para comparação dos experimentos apresentados neste trabalho.

Aplicações específicas e extração de conceitos em um domínio são abordagens amplamente utilizadas nas ferramentas apresentadas em diversos trabalhos publicados em diversos eventos e periódicos científicos.

Constantemente vêem-se aplicações onde o uso de ontologias tem aprimorado o nível de acertos obtidos por ferramentas de extração de termos. Todavia, a criação de ontologias demanda engenheiros com conhecimento e envolvidos em trabalhos multidisciplinares para definir as entidades que representam o domínio a ser avaliado.

Quando se fala em *web* semântica ou computação semântica fica evidenciado o fato de que são exigidas notações formais e estruturas de dados padronizadas para a representação das informações de um domínio. Por outro lado, a proposta do método de extração de mapas conceituais a partir de documentos é algo mais subjetivo, onde existe um acoplamento fraco das entidades. Mas, apesar de ser um acoplamento fraco, ele é suficientemente completo para representar o domínio por meio de um mapa conceitual.

Aplicações que utilizam técnicas híbridas (estatísticas e linguísticas) permitem extrair conceitos com um nível interessante de acertos sem tantas complicações na etapa de projeto e configuração de regras para a utilização do algoritmo. Com isso, é possível criar ferramentas de software com interfaces simplificadas (já que não é preciso configurar muitos parâmetros), e o tempo de resposta passa a ser aceitável, uma vez que a análise estatística, *a priori*, demanda menos processamento do que a análise linguística.

Cabe salientar que a análise gramatical e a marcação de termos são procedimentos que necessitam de um grande *corpus* para treinamento do algoritmo. Isso exige capacidade de armazenamento em memória, a qual é constantemente consultada no processo de extração de conceitos. A etapa de projeto é mais complexa e demanda muito mais tempo e recursos para ser realizada.

Quanto ao tempo de processamento, observou-se que o processo de extração de conceitos é computacionalmente viável quando se trata da construção de mapas conceituais. A resposta foi obtida em menos de dez segundos em todos os experimentos com cerca de 30 textos de referência. Dessa forma, pode-se considerar que o tempo de processamento não interfere de forma considerável no processo de criação de cursos e mapas conceituais.

No MapXtractor é possível associar objetos de aprendizagem aos conceitos existentes no mapa para que o estudante possa construir seu conhecimento utilizando outros recursos disponibilizados na *web*. É possível também obter, de forma rápida, consultas nas ferramentas de busca disponíveis na *internet* simplesmente clicando sobre os conceitos.

Mesmo não utilizando mapas conceituais para organizar o currículo e os módulos temáticos de um curso por completo, esta é uma forma interessante de organizar conteúdos e promover a aprendizagem significativa, podendo ser usada com frequência por professores em suas aulas à distância ou presenciais.

É possível afirmar que esta é uma nova alternativa para organizar conhecimento, compor um repositório de objetos de aprendizagem e

fomentar a construção colaborativa do conhecimento, visto que professores e estudantes unem-se para organizar o conhecimento de seus domínios.

Confirmando afirmações de Ausubel et al (1980), observou-se que mapas conceituais compostos por no máximo 25 conceitos são mais compreensíveis. Quando é extraída uma quantidade muito grande de conceitos o mapa resultante torna-se confuso e inadequado para representar o conhecimento do domínio; ao invés de organizar o conhecimento, o conteúdo torna-se complexo e custoso para se compreender.

Outra vantagem na utilização do MapXtractor é que os textos adicionados pertencem a um determinado domínio pré-definido pelo usuário. Ao criar um curso, supõe-se que o usuário utilize documentos de referência que pertençam ao domínio desse curso. Sendo assim, o *corpus* analisado tende a ser composto por poucos documentos de um domínio especializado. Dessa maneira, o MapXtractor torna-se uma fonte de consulta onde documentos estão vinculados a conceitos e não a palavras, caracterizando uma premissa da computação semântica.

6.4 Limitações

Sobre as limitações do trabalho é preciso salientar que não foram realizados testes massivos com diferentes conjuntos de dados. Em contrapartida, os resultados sempre se mostraram aceitáveis se considerado o tempo de resposta da ferramenta. Antevendo a realidade dos professores ao elaborarem seus cursos, não foram consideradas aplicações com mais de 50 textos.

Quanto às limitações do protótipo, não foram implementadas funcionalidades de *auto-layout* para organizar automaticamente o grafo composto pelos conceitos e suas respectivas conexões semânticas.

Opções de impressão e exportação de imagens não foram implementadas; para tanto, sugere-se utilizar o arquivo exportado para o CMAP Tools.

Quanto aos testes realizados, não foram feitas avaliações específicas quanto à usabilidade e aceitação de usuários, fato que poderia dar ainda mais respaldo ao MapXtractor.

7 CONCLUSÕES

A utilização da ferramenta favorece a transposição de cursos com matrizes curriculares de aprendizado sequencial para mapas conceituais, estimulando a aprendizagem significativa e criando condições para atingir a interdisciplinaridade do conhecimento.

Considerando o objetivo geral do trabalho, o método para a extração automática de mapas conceituais foi desenvolvido por completo e obteve resultados satisfatórios.

Considerando o cenário de avaliação descrito no Capítulo 6, ficou evidente que o método é capaz de extrair proposições válidas a partir de um conjunto de documentos. Ressalta-se, ainda, que as proposições de fato extraídas representam o contexto e o domínio dos documentos a que se referem.

Os resultados puderam ser obtidos através de um algoritmo capaz de mensurar a relevância de termos e de conectores semânticos. Esse algoritmo, um dos objetivos específicos do trabalho, se mostrou eficiente ao realizar sua tarefa.

A proposta do algoritmo desenvolvido no método era aprimorar as técnicas de extração de texto. O modelo proposto passou a considerar a relevância do documento para determinar a relevância de um termo, o que se mostrou interessante quando se deseja criar mapas conceituais de forma automática.

E como objeto final deste trabalho, foi criada a ferramenta MapXtractor, que concebida como protótipo de um *software* para edição e gerenciamento de mapas conceituais, implementou o método proposto. Para os estudantes, o MapXtractor funciona como um facilitador do aprendizado, uma vez que ao possibilitar o uso de objetos de aprendizagem e a realização atividades, lhes permite identificar conceitos deficitários e focar os estudos em conceitos relevantes para o curso.

Analisando o trabalho apresentado e outros trabalhos correlatos, ficou evidente que as técnicas de ensino-aprendizagem que se valem do uso de mapas conceituais obtêm sucesso na sua aplicação. O MapXtractor pode facilitar a construção de mapas conceituais e popularizar essa técnica de ensino-aprendizagem.

O MapXtractor pode auxiliar professores a identificar conceitos relevantes e suas respectivas conexões semânticas dentro do domínio

estudado. A lista de conceitos surge a partir de processamento do computador e não de seu esforço cognitivo.

De posse dos resultados do MapXtractor, o professor pode identificar conexões que não havia notado em momentos anteriores e agregá-las ao mapa, tornando-o mais completo e representando melhor o domínio.

As medidas de relevância são as grandes contribuições do método implementado na ferramenta, já que, ao invés de recuperar documentos relevantes conforme o termo consultado, obtêm-se termos relevantes dentro de um conjunto de documentos, a partir dos quais é possível gerar um mapa conceitual que representa o conhecimento ali contido.

7.1 Trabalhos Futuros

Como trabalhos futuros que poderão ser incorporados à ferramenta MapXtractor e ao método de extração de termos, podem ser citados como por exemplo:

- módulo de sugestão de conceitos com base no conjunto de conceitos existentes na base de dados central comparados aos conceitos inseridos no mapa;
- utilização de novas metodologias para identificação de *links* semânticos utilizando análises linguísticas mais complexas;
- módulos para leitura de documentos em formatos diferentes da extensão .txt (utilizar diretamente arquivos .pdf e .doc, entre outras ferramentas de edição de texto);
- módulo de consulta a objetos de aprendizagem utilizando conceitos ao invés de palavras;
- disponibilizar *webservices* para disponibilizar os mapas construídos sob formato do CmapTools, aumentando a compatibilidade entre os dois *softwares*;
- utilizar diferentes arquivos de *stopwords*, conjunções e listas de verbos de ligação para verificar se é possível melhorar os resultados da extração;
- preparar um *corpus* para testes e validações de ferramentas similares;

- adaptar a ferramenta para uso geral, que independa das funcionalidades inerentes ao processo de ensino-aprendizagem.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRAMOWICZ, W.; WISNIEWSKI, M. Proximity Window Context Method for Term Extraction in Ontology Learning from Text. In: INTERNATIONAL CONFERENCE ON DATABASE AND EXPERT SYSTEMS APPLICATION, 19., 2008, Washington, DC.

Proceedings... Washington, DC: IEEE Computer Society, 2009. p. 215-219.

AMIN, A. B. M. Using Graphic Organizers to Promote Active e-learning. In: WORLD CONFERENCE IN MULTIMEDIA, HYPERMEDIA AND TELECOMMUNICATION, 2005, Montreal, Canada. **Proceedings...** Chesapeake, VA: AACE, 2009. p. 4010-4015.

AMORETTI, M. S. M. Protótipos e Estereótipos: Aprendizagem de Conceitos. Mapas Conceituais: Experiência em Educação à Distância. **Revista Informática na Educação: Teoria & Prática**, Porto Alegre, v. 4, n. 2, 2001.

AUSUBEL, D. P.; NOVAK J.; HANESIAN, H. **Psicologia Educacional**. Rio de Janeiro: Ed. Interamericana, 1980.

BENDERSKY, M.; CROFT, W. B. Discovering key concepts in verbose queries. In: SPECIAL INTEREST GROUP INFORMATION RETRIEVAL, 2008, New York, NY. **Proceedings...** New York, NY: ACM, 2008. p. 491-498.

CAÑAS, A. J.; SAXE, E. B. Pensum No Lineal: Una propuesta innovadora para el diseño de planes de estudio. **Revista Electrónica actualidades investigativas en educación**, San José, Cosa Rica, v. 5, 2005. Disponível em:
<<http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=44759902>>. Acesso em: 10 maio 2009.

CHIN, O. S.; KULATHURAMAIYER, N.; YEO, A. W. Automatic Discovery of Concepts from Text. In: INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, 2006, Washington, DC. **Proceedings...** Washington, DC: IEEE Computer Society, 2006. p. 1046-1049.

CIMOLINO, L.; KAY, J. Verified concept mapping for eliciting conceptual understanding. In: INTERNATIONAL CONFERENCE ON COMPUTER SUPPORTED EDUCATION, 2002, Auckland, New Zealand. **Proceedings...** Auckland, New Zealand: ICCE Computer Society, 2002. p. 1561-1562.

CIMOLINO, L.; KAY, J.; MILLER, A. Incremental student modelling and reflection by verified concept-mapping. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE IN EDUCATION, 11., 2003, Sydney, Australia. **Proceedings...** Sydney, Australia : AIED, 2003. p. 219-227.

DALMOLIN, L. C. D.; NASSAR, S. M.; BASTOS, R. C. A Concept Map Extractor Tool for Teaching and Learning. In: THE IEEE INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 19., 2009, Los Alamitos, California, USA. **Proceedings...** Los Alamitos, California, USA: IEEE Computer Society Publications, 2009.

ESKRIDGE, T. C.; GRANADOS, A.; CAÑAS, A. J. *Ranking* Concept Map Retrieval in the CmapTools Network. In: INTERNATIONAL CONFERENCE ON CONCEPT MAPPING, 2., 2006, San José, Costa Rica. **Proceedings...** San José, Costa Rica: Universidad de Costa Rica, 2006. p. 477-484.

ETEC BRASIL. **Referenciais para elaboração de material didático para EAD no Ensino profissional e tecnológico.** 2007. Disponível em: <http://www.etcbrasil.mec.gov.br/gCon/recursos/upload/file/ref_materialdidatico.pdf>. Acesso em: 26 maio 2009.

FELDMAN, R.; SANGER, J. **The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.** New York, NY: Cambridge University Press, 2007. 422 p.

FRANK, E. et al. Domain-specific keyphrase extraction. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 16., 1999, Stockholm, Sweden. **Proceedings...** San Francisco, CA: Morgan Kaufmann, 1999. p. 668-673.

GAUDIO, R.; BRANCO, A. Automatic Extraction of Definitions in Portuguese: A Rule-Based Approach. In: PORTUGUESE CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2007, [S.l.]. **Anais...** [S.l.: s.n.], 2007.

GELFAND, B.; WULFEKULER., M.; PUNCH III, W. Automated concept extraction from plain text. In: WORKSHOP ON TEXT CATEGORIZATION, 1998, [S.l.]. **Anais...** [S.l.: s.n.], 1998.

GIARLO, M. J. **A comparative analysis of keyword extraction techniques.** 2005. Disponível em <<http://lackoftalent.org/michael/papers/596.pdf>>. Acesso em 20 abril 2009.

GONZALEZ, M.; LIMA, V. L. S. de; LIMA, J. V. de. **Termos, Relacionamentos e Representatividade na Indexação de Texto para Recuperação de Informação.** Porto Alegre: EDIPUCRS, 2006. (Letras de Hoje, v. 41) ISSN 0101-3335

GUPTA, S. et al. **Knowledge Discovery from Semi-Structured Data for Conceptual Organization.** In: INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE AND IEEE/WIC/ACM INTERNATIONAL CONFERENCE INTELLIGENT AGENT TECHNOLOGY, 2006, Hong Kong. **Resumos...** Washington, DC: IEEE Computer Society, 2006. p.291-294.

KANTARDZIC, M. M. **Data Mining: Concepts, Models, Methods, and Algorithms**. [S.l]: Wiley-IEEE Press, 2002.

KOWATA, J. H.; CURY, D.; BOERES, M. C. S. Caracterização das Abordagens para Construção (Semi) Automática de Mapas Conceituais. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 20., 2009, Florianópolis, SC. **Anais...** Florianópolis, SC: [s.n], 2006.

KOZAREVA, Z.; RILOFF, E.; HOVY, E. Toward Completeness in Concept Extraction and Classification. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2009, Singapore. **Resumos...** Moristown, NJ, USA: ACL end AFNLT, 2009. p. 948-957.

KUSHMERICK, N.; THOMAS, B. **Adaptive information extraction: Core technologies for information agents**. In: INTELLIGENT Information Agents R&D in Europe: An AgentLink perspective in Computer Science. [S.l]: Berlin/Heidelberg, 2003.

LIMA, I. S. L. et al. Criando Interfaces para Objetos de Aprendizagem. Em: PRATA, C.L.; NASCIMENTO, C. A. (orgs). **Objetos de aprendizagem: uma proposta de recurso pedagógico**. Brasília: MEC, SEED, 2007. p. 39-48.

LO R.T.-W.; HE, B.; OUNIS, I. Automatically building a stopword list for an information retrieval system. **J Digital Informat Manage**, v. 3, 2005.

MANNING, C. D.; SCHÜTZE, H. **Foundations of statistical natural language processing**. Cambridge: MIT Press, 1999.

MOREIRA M. A.; MASINI E. F. S. **Aprendizagem Significativa – A teoria de David Ausubel**. São Paulo: Ed. Moraes, 1982.

NAKAGAWA, H.; MORI, T. **A simple but powerful automatic term extraction method.** In: INTERNATIONAL WORKSHOP ON COMPUTATIONAL TERMINOLOGY, 2., 2002, Morristown, Estados Unidos. **Anais...** Morristown, Estados Unidos: Association for Computational Linguistics, 2002.

NOVAK, J. D.; GOWIN, D. B. **Learning How to Learn.** New York, NY: Cambridge University Press, 1984.

NOVAK, J. D.; CAÑAS, A. J. **The theory underlying concept maps and how to construct them.** Institute for Human and Machine Cognition. 2006. Disponível em: <http://cmap.ihmc.us/publications/researchpapers/theorycmaps/theoryunderlyingconceptmaps.htm>. Acesso em: 20 maio 2009.

_____. The Origins of the Concept Mapping Tool and the Continuing Evolution of the Tool. **Information Visualization**, [S.l.], v. 5, p. 175-184.

PAZIENZA, M. T.; PENNACCHIOTTI, M.; ZANZOTTO, F. M. Terminology Extraction: An Analysis of Linguistic and Statistical Approaches. **Studies In Fuzziness And Soft Computing**, v. 185, Alemanha, 2005. p. 255-280 ISSN: 1434-9922

PEREZ-MARIN, D. et al. Automatic Generation of Students' Conceptual Models underpinned by Free-Text Adaptive Computer Assisted Assessment. In: INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, Kerkrade, The Netherlands, 2006. **Proceedings...** Netherlands: [s.n.] 2006.

RISSOLI, V. R. V. et al. Sistema Tutor Inteligente baseado na Teoria da Aprendizagem Significativa com acompanhamento Fuzzy. **Revista Informática na Educação**, v. 9, [S.l.], p. 37-47, 2006.

SILVA, J. M. et al. Projeto e Desenvolvimento de um Sistema Multi-agentes para Objetos Inteligentes de Aprendizagem baseado em padrão SCORM. **Revista Brasileira de Informática na Educação**, v. 16, p. 1, 2008.

SOUZA, M. F. C. et al. Desenvolvimento e Habilidades em Tecnologias da Informação e Comunicação (TIC) Através de Objetos de Aprendizagem. **Objetos de Aprendizagem** - Uma proposta de Recurso Pedagógico. Brasília: MEC, SEED - Ministério da Educação. Secretaria de Educação a Distância, 2007, v. 01, p. 59-69.

SWANSON, D. R.; SMALHEISER, N. R. An interactive system for finding complementary literatures: a stimulus to scientific discovery, **Artificial Intelligence**, n. 91, p. 183-203, 1997.

T. H. CHANG et al. Automatic Concept Map Constructing using top-specific training corpus. In: Asia-Pacific Educational Research Association Board Meeting (APER'A'2008), Singapore, 2008. **Proceedings...** [S.l.: s.n.], 2008.

TAVARES R. Aprendizagem significativa em um ambiente multimídia. In: Encuentro Internacional sobre Aprendizaje Significativo, 2006, Madrid, Espanha. **Resumos...** [S.l.: s.n.], 2006.

TAVARES, R.. **Ambiente Colaborativo On-Line E A Aprendizagem Significativa de Física**. In: Congresso Internacional ABED de Educação a Distância, 13., 2007, Curitiba. **Anais...** Curitiba: [s.n.], 2007.

THANOPOULOS, A.; FAKOTAKIS, N.; KOKKINAKIS, G. Automatic extraction of semantic relations from specialized corpora. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, 18., 2000, Saarbrücken. Morristown, NJ: Association for Computational Linguistics, 2000. p. 836-842.

TRAVAGLIA, L. C, **Verbos de Ligação: itens lexicais ou gramaticais?** Campinas, SP: Grupo de Estudos Lingüísticos do Estado de São Paulo, UNICAMP, 2004. Publicado em CD-ROM. ISSN 14130939.

TSENG S. et al. A new approach for constructing the concept map. **Computers & Education**, v. 49, p. 691-707, 2007.

TURNEY, P. D. Coherent keyphrase extraction via *web* mining. In: International Joint Conference on Artificial intelligence, 18., 2003, Acapulco, Mexico. **Proceedings...** San Francisco, CA: Morgan Kaufmann Publishers, p. 434-439, 2003.

VALERIO, A.; LEAKE, D. Jump-Starting Concept Map Construction with Knowledge Extracted From Documents. In: INTERNATIONAL CONFERENCE ON CONCEPT MAPPING, 2., 2006, [S.I.]. **Proceedings...** [S.l.: s.n.], 2006. p. 296-303.

VILLALON, J.; CALVO, R.A. Concept Extraction from Student Essays, Towards Concept Map Mining. In: IEEE INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 2009, [S.I.]. **Proceedings...** [S.l.: s.n.], 2009. p. 221-225.

VILLALON, J.; CALVO, R.A. Concept Map Mining: A Definition and a Framework for Its Evaluation. In: International CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY, 2008, [S.I.]. **Proceedings...** [S.l.: s.n.], 2008. p. 357-360.

WAFULA, B. N. **Automatic Construction of Concept Maps**. 2006. Disponível em <<http://www.cs.joensuu.fi/pages/whamalai/sciwri/belinda.pdf>>. Acesso em: 23 maio 2009.

WANG, W. M. et al. Self-associated concept mapping for representation, elicitation and inference of knowledge. **Knowledge-Based Systems**, [S.l.], v. 21 n. 1, p. 52-61, 2008.

WILLERMAN, M.; MAC HARG, R. A. The concept map as an advance organizer. *Journal of Research in Science Teaching*. **Journal of Research in Science Teaching**, Wiley Periodicals, Chicago, Estados Unidos, 1991.

WILLEY A. **Connecting learning objects to instructional design theory : A definition, a metaphor, and a taxonomy** . 2000. Disponível em: < <http://reusability.org/read/> >. Acesso em: 02 junho 2009.

Y. WU et al. ENGINEERING, Curriculum Container System: A System to Support Curriculum and Learning Activity Management. In: INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 50., 2005, [S.l.]. **Proceedings...** [S.l.: s.n.], 2005. p. 287-289.

ZANG, X., LI; X.i, ZHAO, K.; GUAN, X. **A Data Mining Algorithm Based on Grid**. In: _____. **Grid and Cooperative Computing**. Lecture Notes in Computer Science. [S.l.; s.n], 2003. p 807-810.

ZILI, Z.; YANNA, W.; JUNZHONG, G.U. **Markov-Based Automatic Term Extraction**. In: INTERNATIONAL SEMINAR ON FUTURE BIOMEDICAL INFORMATION ENGINEERING, 2008 [S.l.]. [S.l.: s.n], 2008. DOI 10.1109/FBIE.2008.84.

ZOUAQ, A., NKAMBOU, R. Building Domain Ontologies from Text for Educational Purposes. In: **TRANSACTIONS on Learning Technologies**. [S.l.]: Berlin/Heidelberg, 2008. p. 49-62.

APÊNDICE A - MODELAGEM SIMPLIFICADA DO SISTEMA

