

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA**

**JOSÉ ANTONIO BUIAR**

**MODELO PARA ESTRUTURAÇÃO E REPRESENTAÇÃO DE DIÁLOGOS**  
**EM FÓRUM DE DISCUSSÃO**

**DISSERTAÇÃO**

**CURITIBA**  
**2012**

JOSÉ ANTONIO BUIAR

**MODELO PARA ESTRUTURAÇÃO E REPRESENTAÇÃO DE DIÁLOGOS  
EM FÓRUM DE DISCUSSÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do título de “Mestre em Computação Aplicada” – Área: Engenharia de Sistemas Computacionais – Linha de Pesquisa: Sistemas de Informação.

Orientador: Robinson Vida Noronha

CURITIBA  
2012

Dados Internacionais de Catalogação na Publicação

---

B932 Buiar, José Antonio  
Modelo para estruturação e representação de diálogos em fórum de discussão / José Antonio Buiar. - 2012.  
87 f. : il. ; 30 cm

Orientador: Robinson Vida Noronha.  
Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Computação Aplicada, Curitiba, 2012.  
Bibliografia: f. 83-85.

1. Ensino à distância. 2. Representação do conhecimento (Teoria da informação). 3. Fórum (Debates) - Avaliação. 4. Tecnologia educacional. 5. Multimídia interativa. 6. Análise de interação em educação. 7. Software - Desenvolvimento. 8. Computação - Dissertações. I. Noronha, Robinson Vida, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Computação Aplicada. III. Título.

CDD (22. ed.) 004

---

**Título da Dissertação N°: 05**

**“Modelo para Estruturação e Representação de Diálogos em Fórum de Discussão”**

por

**José Antonio Buiar**

Esta dissertação foi apresentada como requisito parcial à obtenção do grau de **MESTRE EM COMPUTAÇÃO APLICADA** - Área de Concentração: Engenharia de Sistemas Computacionais, pelo PPGCA - Programa de Pós-Graduação em Computação Aplicada - Mestrado Profissional - da Universidade Tecnológica do Paraná - UTFPR - Câmpus Curitiba, às 10 horas do dia 16 de outubro de 2012. O trabalho foi aprovado pela Banca Examinadora, composta pelos professores:

**PPGCA**

---

Prof. **Robinson Vida Noronha, Dr.**  
presidente - (UTFPR - CT)

---

**Programa de Pós-Graduação  
em Computação Aplicada**

Prof. **Gilson Yukio Sato, Dr.**  
(UTFPR - CT)

---

Prof. **Alexandre Ibrahim Direne, Dr.**  
(UFPR)

---

Prof. **Andrey Ricardo Pimentel, Dr.**  
suplente - (UFPR)

---

Prof. **João Alberto Fabro, Dr.**  
(UTFPR - CT)

## **AGRADECIMENTOS**

À Denise, Analiz, Gabriel e Rafael por toda a paciência, compreensão e apoio.

Aos colegas do DAINF por todo o incentivo e colaboração.

Ao meu orientador Robinson pela maestria na condução de seu mister.

Aos professores doutores Direne, Fabro e Sato pelas valiosas contribuições.

## RESUMO

BUIAR, José Antonio. Modelo para Estruturação e Representação de Diálogos em Fórum de Discussão. 87 f. Dissertação – Programa de Pós-Graduação em Computação Aplicada, Universidade Tecnológica Federal do Paraná. Curitiba, 2012.

A adaptação dos sistemas tradicionais de ensino presencial para o ambiente de ensino a distância introduz diversas mudanças na práxis escolar. Com a ausência do contato direto entre educador e educandos, surge a necessidade de utilização de artefatos tecnológicos que substituam a interação direta. O fórum de discussão é um desses artefatos tecnológicos. Ele possui a característica de ser um elemento catalisador da comunicação entre os envolvidos e pode ser um importante instrumento no processo educacional. Contudo, a natureza não estruturada das mensagens textuais de um fórum dificulta o seu uso como instrumento na avaliação individual do aluno. A análise e qualificação do conteúdo das mensagens armazenadas em um fórum representa um grande desafio para o instrutor. A ausência de uma estrutura formal de representação dos conceitos, crenças e idéias dos alunos poderia ser apontado como um dos elementos que contribuem para esse desafio. A proposta desta pesquisa é o desenvolvimento de um modelo que permita a estruturação e representação das mensagens de um fórum. Essa estruturação considera três aspectos da mensagem: i) os conceitos apresentados, ii) quem os apresentou e finalmente iii) quando esses conceitos foram apresentados. Para validar esse modelo, um programa de computador foi desenvolvido e testado em um fórum do ambiente virtual Moodle. Os conceitos desenvolvidos para o Modelo de Estruturação e Representação das Mensagens do Fórum foram utilizados no desenvolvimento desse programa de computador. Por meio desse modelo de estruturação e representação das mensagens, um mapa ou guia é gerado. Esse mapa ou guia pode ser acessado pelo professor ou instrutor. Esse novo recurso desenvolvido, pode ser utilizado como uma ferramenta de apoio à análise ou avaliação do fórum do ambiente Moodle como um todo ou de cada participação individual do aluno.

**Palavras-chave:** Representação do Conhecimento, Fórum de Discussão

## ABSTRACT

BUIAR, José Antonio. Model to Structuring and Representation of Discussion Forum Dialogs . 87 f. Dissertação – Programa de Pós-Graduação em Computação Aplicada, Universidade Tecnológica Federal do Paraná. Curitiba, 2012.

The traditional learning practices adaptation to the distance learning introduces several changes in school practice. Since in distance learning the direct contact between educators and students does not exist, new technological artifacts become necessary in order to replace direct interaction. One of these artifacts is the discussion forum, which works as a catalyzer element of the communication between involved ones and can be an important tool in the educational process. Nevertheless, non-structured nature of text messages on a forum hampers its use as a tool in individual student assessment. Analysis and qualification of message contents stored on a forum represents an important challenge for instructors. The absence of a formal representation of concepts, ideas and beliefs from students could be designated as one of the factors that make this challenge even harder. This research proposes the development of a model that allows the messages on a forum can be structured and represented. This structuration considers three message aspects: i) presented concepts, ii) who has presented it, and iii) when concepts have been presented. As a means to validate this model, a computer program was developed and tested in a Moodle virtual environment forum. The concepts developed to the Structuration and Representation of the Forum Messages Model were used on this computer program development. Through the use of this model a map or guide is generated. This map or guide can be accessed by the professor or instructor. This new feature can be used as a support tool to analysis or evaluation of a Moodle forum environment.

**Keywords:** Knowledge Representation, Discussion Forum

## LISTA DE FIGURAS

FIGURA 1 – EXEMPLO DE MAPA CONCEITUAL .....	20
FIGURA 2 – ESTADO DA ARTE .....	24
FIGURA 3 – SEQUÊNCIA DE MENSAGENS DE UMA DISCUSSÃO .....	27
FIGURA 4 – CONCEITOS DA MENSAGEM 1 .....	29
FIGURA 5 – CONCEITOS DA MENSAGEM 2 .....	30
FIGURA 6 – REPRESENTAÇÃO DA MENSAGEM 1 .....	32
FIGURA 7 – REPRESENTAÇÃO DA MENSAGEM 2 .....	32
FIGURA 8 – CONCEITOS DA DISCUSSÃO E SUAS RELAÇÕES .....	33
FIGURA 9 – DIAGRAMA USUÁRIO X TEMPO .....	34
FIGURA 10– DIAGRAMA CONCEITO X TEMPO .....	35
FIGURA 11– DIAGRAMA CONCEITO X USUÁRIO .....	36
FIGURA 12– MODELO DE REPRESENTAÇÃO TRIDIMENSIONAL .....	37
FIGURA 13– MODELO COMPUTACIONAL DO FORUM3D .....	39
FIGURA 14– MÓDULO EXTRATOR .....	40
FIGURA 15– MÓDULO ESTRUTURADOR .....	45
FIGURA 16– SEPARADOR DE SENTENÇAS .....	46
FIGURA 17– IDENTIFICADOR DE VOCÁBULOS .....	47
FIGURA 18– ETIQUETADOR DE VOCÁBULOS .....	49
FIGURA 19– AGRUPADOR DE SINTAGMAS .....	50
FIGURA 20– MÓDULO CLASSIFICADOR .....	53
FIGURA 21– DIAGRAMA DO MÓDULO VISUALIZADOR .....	56
FIGURA 22– SAÍDA DO PROGRAMA GRAPHVIZ .....	61
FIGURA 23– ARQUIVO DISCUSSAO.XML OBTIDO .....	65
FIGURA 24– ARQUIVO POST.XML OBTIDO .....	66
FIGURA 25– ARQUIVO ESTRUTURADO.XML OBTIDO .....	68
FIGURA 26– ARQUIVO CONCEITUAL.XML OBTIDO .....	69
FIGURA 27– TELA DE SELEÇÃO DE CURSO .....	70
FIGURA 28– TELA DE SELEÇÃO DE FÓRUM .....	71
FIGURA 29– PAINEL DE RELATÓRIOS .....	71
FIGURA 30– TELA DE SELEÇÃO DE VISÃO .....	72
FIGURA 31– TELA DE FILTRO DE TEMPO .....	73
FIGURA 32– TELA DE FILTRO DE USUÁRIO .....	74
FIGURA 33– TELA DE FILTRO DE CONHECIMENTO .....	75
FIGURA 34– TELA DO GRENCIADOR DE DICIONÁRIOS .....	75
FIGURA 35– RELATÓRIO DE USUÁRIO X TEMPO .....	76
FIGURA 36– RELATÓRIO DE CONCEITO X TEMPO .....	77
FIGURA 37– RELATÓRIO DE CONCEITO X USUÁRIO .....	77
FIGURA 38– GERAÇÃO DE MAPA CONCEITUAL .....	78



## LISTA DE QUADROS

QUADRO 1 –	MAPEAMENTO DO BANCO DE DADOS .....	42
QUADRO 2 –	ESTRUTURA DO ARQUIVO DISCUSSAO.XML .....	43
QUADRO 3 –	ESTRUTURA DO ARQUIVO MENSAGEM.XML .....	51
QUADRO 4 –	ESTRUTURA DO ARQUIVO ESTRUTURADO.XML .....	52
QUADRO 5 –	ELEMENTOS DE RECONHECIMENTO .....	54
QUADRO 6 –	ESTRUTURA DO ARQUIVO CONCEITUAL.XML .....	55

## LISTA DE SIGLAS

AVA	Ambiente Virtual de Aprendizagem
NLP	<i>Natural Language Processing</i>
SNA	<i>Social Network Analysis</i>
XML	<i>Extensible Markup Language</i>
API	<i>Application Programming Interface</i>
CentOS	<i>Community ENTERprise Operating System</i>
PHP	<i>Hypertext Preprocessor</i>
HTML	<i>HyperText Markup Language</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
1.1	APRESENTAÇÃO	13
1.2	OBJETIVOS	15
1.3	ESTRUTURA	16
<b>2</b>	<b>REVISÃO DA LITERATURA</b>	<b>17</b>
<b>3</b>	<b>MODELO CONCEITUAL</b>	<b>26</b>
3.1	EXEMPLO DA DINÂMICA DE INSERÇÃO DE CONCEITOS	26
3.2	VISUALIZAÇÃO DA DISCUSSÃO	32
3.2.1	Visão Usuário x Tempo	34
3.2.2	Visão Conceito x Tempo	34
3.2.3	Visão Conceito x Usuário	35
3.3	MODELO DE REPRESENTAÇÃO TRIDIMENSIONAL	36
3.4	ANÁLISE DO MODELO	37
<b>4</b>	<b>MODELO COMPUTACIONAL</b>	<b>38</b>
4.1	MODELO PROPOSTO	38
4.2	MÓDULO EXTRATOR	39
4.2.1	Componente Conector com Banco de Dados	40
4.2.2	Componente Seletor da Discussão	41
4.2.3	Componente Gerador XML	42
4.3	MÓDULO ESTRUTURADOR	43
4.3.1	Componente Controlador	44
4.3.2	Componente Separador	45
4.3.3	Componente Identificador	47
4.3.4	Componente Etiquetador	48
4.3.5	Componente Agrupador	48
4.3.6	Componente LeitorXml	51
4.3.7	Componente Gerador XML	51
4.4	MÓDULO CLASSIFICADOR	53
4.4.1	Componente LeitorXml	53
4.4.2	Componente Processador	54
4.4.3	Componente Gerador XML	55
4.5	MÓDULO VISUALIZADOR	55
4.5.1	Componente Controlador	56
4.5.2	Componente Leitor XML	57
4.5.3	Componente Seletor	57
4.5.4	Componente Filtro	57
4.5.5	Componente Gerenciador de Dicionários	58
4.5.6	Componente Gerador Usuário Tempo	58
4.5.7	Componente Gerador Conceito Tempo	59
4.5.8	Componente Gerador Conceito Usuário	59
4.5.9	Componente Gerador Mapa Conceitual	60
4.6	CONCLUSÕES	61

<b>5 ESTUDO DE CASO</b> .....	<b>63</b>
5.1 CARACTERÍSTICAS DE IMPLEMENTAÇÃO .....	63
5.2 BASE DE TESTES .....	64
5.3 ARQUIVOS XML .....	64
5.4 UTILIZAÇÃO DO SISTEMA .....	70
5.5 OBSERVAÇÕES .....	78
<b>6 CONCLUSÃO</b> .....	<b>80</b>
<b>REFERÊNCIAS</b> .....	<b>83</b>
<b>GLOSSÁRIO</b> .....	<b>86</b>

# 1 INTRODUÇÃO

Este capítulo está dividido em 3 seções. A Seção 1.1 apresenta uma visão geral da pesquisa realizada, descreve o tema abordado, o contexto no qual o tema está inserido e os desafios ou problemas que motivaram a realização desta pesquisa. Por meio da Seção 1.2 são apresentados os objetivos gerais e específicos a serem alcançados. Finalmente na Seção 1.3 está apresentada a organização utilizada neste documento.

## 1.1 APRESENTAÇÃO

O Ensino a Distância é uma modalidade de educação que tem sido utilizada por Instituições de Ensino. Nesse contexto, essas instituições fazem uso de ferramentas computacionais conhecidas como Ambientes Virtuais de Aprendizagem, ou simplesmente AVA.

Os AVA também têm sido utilizados como ferramentas de apoio ao ensino presencial. Nesse contexto, o ambiente permite gerenciar as atividades definidas pelo professor e que devem ser executadas pelos alunos cadastrados nesse ambiente.

No AVA, o professor tem a sua disposição uma série de artefatos tecnológicos que podem ser utilizados na elaboração e condução de treinamentos. Dentre estes artefatos, destaca-se o Fórum de Discussão Online ou simplesmente FÓRUM. O fórum também pode ser considerado como um repositório dos conhecimentos construídos pelos alunos. Em contraste com a riqueza de informações que podem ser depositadas livremente pelos alunos, a sua característica de representação textual não estruturada representa uma dificuldade significativa para o processo avaliativo. Ou seja, pode-se afirmar que um fórum é um arcabouço rico de informações e de difícil análise e avaliação pelo professor. Se um determinado

professor, por exemplo, desejar analisar, qualificar ou avaliar o conhecimento apresentado pelos alunos em uma discussão de um fórum, ele deverá acessar essas informações, lê-las e tirar as suas próprias conclusões.

Esse tipo de manipulação torna-se um desafio quando a quantidade de mensagens é elevada. De acordo com Gerosa, Fuks e Lucena (2004), isso poderia desmotivar o uso do fórum como ferramenta de apoio à avaliação dos alunos e avaliação do fórum como um todo.

No caso de uma discussão de um fórum que envolva 100 alunos, se cada aluno apresentar suas ideias com uma média de 20 palavras, o volume total de 2.000 palavras dificultaria uma avaliação qualitativa manual das contribuições individuais dos alunos. Haveria algum modelo que permitisse auxiliar o professor a analisar um fórum e:

- Identificar as contribuições de cada aluno no fórum?
- Identificar quais elementos do domínio foram discutidos com a profundidade desejada pelo professor?
- Identificar quais elementos do domínio não foram discutidos pelos alunos?
- Identificar classes ou grupos de alunos de acordo com os seus pontos de vista?

A utilização do fórum de discussão em Ambiente de Aprendizado Virtual, pelo professor, é um recurso que facilita a comunicação entre as pessoas envolvidas no processo de aprendizado.

Cabe aos professores, a verificação de que o direcionamento deste fluxo de comunicação ocorra em torno do conteúdo a ser ministrado no processo de ensino.

Qualquer tentativa de desenvolvimento de mecanismos computacionais para auxiliar ao professor poderá ser facilitada se os dados a serem manipulados e analisados estiverem estruturados. Seria possível o desenvolvimento de um modelo de estruturação das mensagens armazenadas em um fórum? Se sim, como essa estrutura poderia ser utilizada em um mecanismo para auxiliar o professor na tarefa de avaliar qualitativamente um fórum?

Essas questões nortearam este projeto de pesquisa. O desenvolvimento de um modelo que permita estruturar as informações das discussões de um fórum e apresentá-las ao professor é o objetivo desta pesquisa. Nesse contexto, busca-se também definir alguns mecanismos que resumam essas informações para esse professor.

A questão de pesquisa é apresentada a seguir:

Como estruturar o conteúdo das mensagens das discussões de um fórum de discussão a fim de permitir a visualização dos elementos do domínio apresentados pelos alunos?

Espera-se que o modelo de estruturação definido e investigado neste trabalho possa auxiliar o professor na tarefa de realizar avaliações qualitativas das mensagens do fórum ou que sirva de elemento norteador de futuros modelos de automatização dessa tarefa.

Por meio da qualificação das mensagens registradas em um fórum, o instrutor poderia obter alguns indicadores da participação individual de cada aluno. Esta informação poderia permitir ao instrutor, realizar alguns ajustes necessários na condução do treinamento. Por exemplo, para os alunos que não estão contribuindo com conceitos e conhecimentos significativos ao fórum de discussão, o professor, com o auxílio de algumas informações estruturadas, poderia definir ações que modificassem as atitudes desse hipotético aluno.

Pode-se ainda vislumbrar que a definição de um modelo que permita estruturar as mensagens de um fórum poderia ser útil em um outro contexto. Por exemplo, as informações geradas por esse modelo poderiam ser utilizadas para a definição de perfis ou estereótipos da participação de alunos.

## 1.2 OBJETIVOS

Para investigar a questão de pesquisa apresentada anteriormente, alguns objetivos foram definidos, a saber:

- Desenvolver uma proposta de Modelo Conceitual que permita a estruturação das mensagens de um Fórum;

- Desenvolver uma proposta de Modelo Computacional para implementação do Modelo Conceitual;
- Realizar um estudo de caso para validar o modelo de estruturação proposto.

### 1.3 ESTRUTURA

Este trabalho está estruturado em seis capítulos, sendo que neste capítulo foi apresentado o contexto no qual a pesquisa foi elaborada. Ele contempla também a motivação da pesquisa e os objetivos a serem alcançados.

No Capítulo 2 analisa-se alguns dos trabalhos e relatos de pesquisa referentes ao tema deste trabalho. No Capítulo 3 é apresentado o modelo conceitual desenvolvido para estruturar as mensagens de um fórum.

O Capítulo 4 descreve o modelo computacional especificado a partir do modelo conceitual descrito no Capítulo 3.

No Capítulo 5 implementa-se um estudo de caso. Nesse estudo de caso, alguns testes do modelo foram realizados com uma base de dados obtida a partir do AVA Moodle.

Finalizando este documento, o Capítulo 6 apresenta algumas conclusões obtidas com esta pesquisa e apresenta sugestões para trabalhos futuros.



## 2 REVISÃO DA LITERATURA

Este capítulo apresenta as principais referências bibliográficas utilizadas para o desenvolvimento deste trabalho de pesquisa.

Em uma comunidade *online*, os participantes podem compartilhar os seus conhecimentos, ideias, experiências e entendimentos através de conversas. Os participantes negociam significados, aprendendo um com os outros, e construindo novos conhecimentos durante o processo de discussão. Esse processo de interação é mediado pelo discurso, conforme observado por Heo e Breuleux (2009). Os elementos compartilhados nas conversas são tratados, no escopo do trabalho de pesquisa apresentado, genericamente como conhecimento.

A ferramenta de fórum de discussão com finalidades educacionais em ambiente *online*, que é tratado como Fórum nesta pesquisa, é um dos recursos tecnológicos oferecidos pelos AVA. Conforme Sánchez (2005), o fórum é um espaço de comunicação formado por caixas de diálogos que podem ser classificados de forma temática. Conforme observado por Brito (2010), o Fórum tem como característica a realização do discurso textual de forma assíncrona. A característica assíncrona do fórum permite que o aluno formule suas contribuições em um momento que lhe for mais adequado e propiciando uma oportunidade de revisá-las. As contribuições realizadas pelos diversos alunos fica disponível para consulta na forma de tópicos.

Conforme a pesquisa realizada por Landis (2001), os registros contínuos que são realizados durante a atividade do fórum permitem documentar a comunicação realizada. Esses registros documentados auxiliam na organização dos conceitos e distribuição de responsabilidades. O fórum possibilita ao professor a condução de uma discussão a ser realizada pelos alunos, durante o processo de aprendizagem. Um fórum é composto por diversas seções chamadas de discussão. Através de uma discussão, o professor realiza uma interação com os alunos em torno de um tema específico.

Com o objetivo de normalização da terminologia usada na pesquisa, os indivíduos que participam dos processos de fórum em EaD estão caracterizados como :

**Professor:** é quem realiza a proposta da atividade de fórum e realiza a sua avaliação, engloba o professor propriamente dito, mas pode também incluir tutores ou outros colaboradores. Também citado na literatura como instrutor ou educador.

**Aluno:** indivíduo incumbido de efetuar contribuições no fórum através de mensagens, que serão avaliadas pelo professor. Também citado como aprendiz, educando ou estudante na literatura.

**Usuário:** o termo usuário compreende a qualquer indivíduo com participação no fórum, seja professor ou aluno.

O processo de comunicação ocorrido na atividade de discussão é realizado através da troca de mensagens entre os usuários. Cada mensagem gerada por um usuário, também conhecida por *post*, será referida neste documento como Mensagem. O texto apresentado pela mensagem apresenta o registro do domínio do conhecimento que o usuário desejou registrar na discussão.

O Processamento de Linguagem Natural, NLP (*Natural Language Processing*), é uma abordagem computacional que permite a análise de informações textuais. Conforme Liddy (2001) e Chowdhury (2003), a abordagem está baseada em um conjunto de teorias e um conjunto de tecnologias.

De acordo com Liddy (2001), as abordagens utilizadas pelas técnicas de NLP podem ser classificadas como simbólica, estatística e conexionista. A abordagem simbólica, também conhecida como linguística, realiza uma análise profunda dos fenômenos linguísticos sendo baseada na representação explícita dos fatos. Utiliza como referência os padrões da linguagem, extrações baseadas nestes padrões e relatividade semântica, conforme verificado em Gómez-Pérez e Manzano-Macho (2004).

A abordagem estatística, baseada na utilização de diversas técnicas matemáticas e grandes bases de textos (chamados *corpus*), utiliza dados observáveis como a fonte primária de evidência. Um estudo conduzido por Ng

e Zelle (1997) descreve as vantagens da utilização da interpretação semântica estatística baseada em *corpus* em relação às outras abordagens.

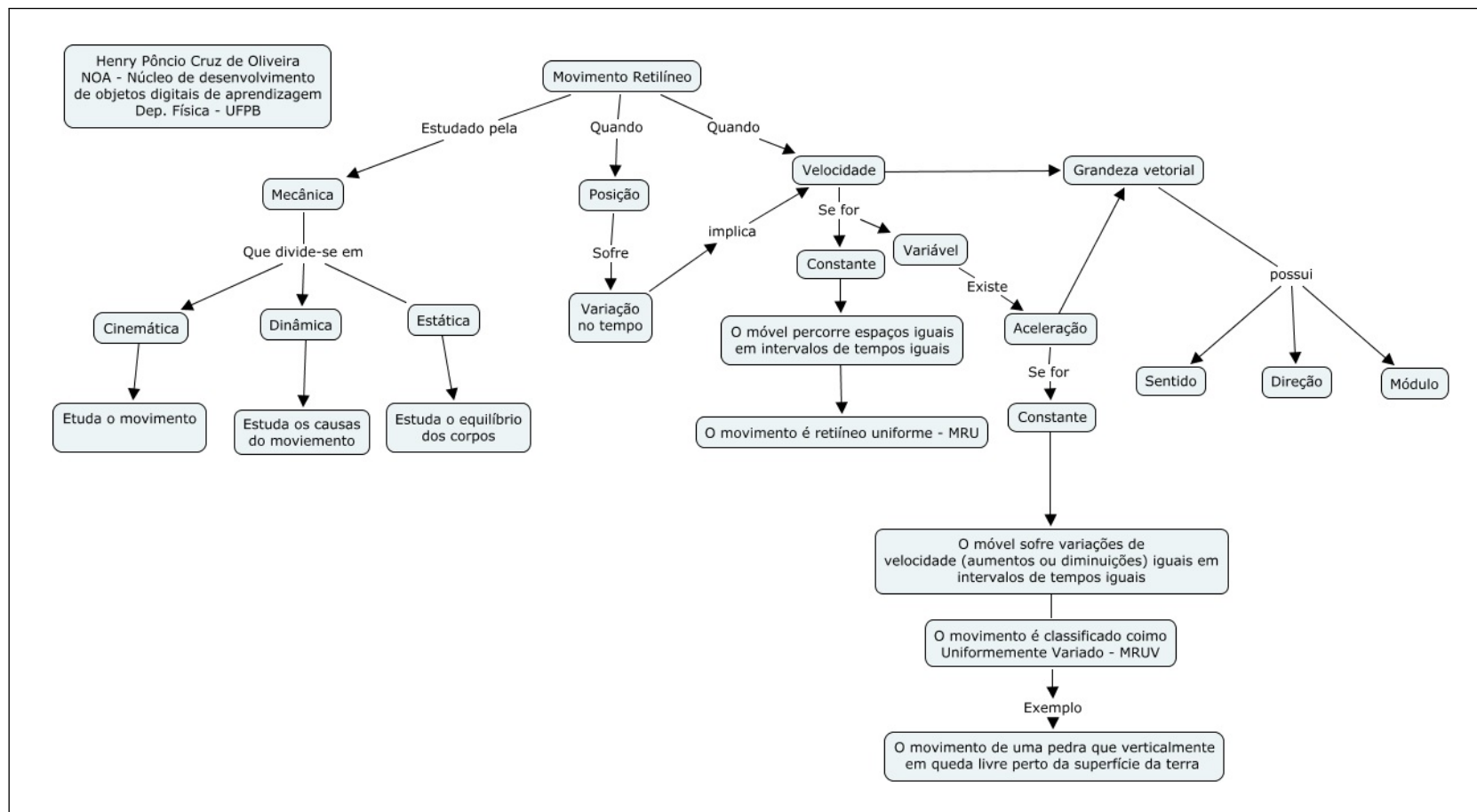
A abordagem conexionista, baseada na abordagem estatística, combina o aprendizado estatístico com várias teorias de representação, permitindo transformação, inferência e manipulação de fórmulas lógicas (LIDDY, 2001).

Foi verificado na literatura pesquisada que, apesar da grande evolução das técnicas de NLP, ainda existe uma carência de maiores referências voltadas para a língua portuguesa.

A análise da interação dos elementos metodológicos e conceituais do processo de construção do conhecimento e a abordagem de Mapas Conceituais é realizada por Novak e Gowin (1984). Nesse trabalho é mostrado que a escolha das palavras é muito importante para rotular conceitos e relacionamentos. A estruturação hierárquica dos conceitos que são apresentados por meio de diferenciação progressiva e reconciliação integrativa é a base desta proposta. Os mapas gerados são estruturados de acordo com a Teoria de Aprendizagem Significativa de Ausubel, Novak e Hanesian (1983).

Na Figura 1 tem-se um exemplo de um Mapa Conceitual mostrando um conceito de Física, no qual pode ser observada a forma de representação dos conceitos e relações. Em um Mapa Conceitual, a representação dos Conceitos é feita por meio de quadros. Já a relação entre esses conceitos é representada por uma linha terminada em seta.

Uma metodologia para extração manual de Mapas Conceituais a partir de textos em linguagem natural é desenvolvida por Oliveira, Pereira e Cardoso (2001). A apresentação de um protótipo para geração de um mapa conceitual a partir de textos é apresentada por Pérez e Vieira (2005). A geração manual de um Mapa Conceitual e sua utilização como estratégia de ensino e aprendizagem está tratada em Souza e Boruchovitch (2010) e Tavares (2007).



**Figura 1 - Exemplo de Mapa Conceitual**  
 Fonte: Oliveira (2011)

Na pesquisa desenvolvida por Kowata (2010), é realizada uma Abordagem Computacional para Construção de Mapas Conceituais tendo como referência textos na língua portuguesa do Brasil. Nesse trabalho também é realizado um levantamento sobre as abordagens semiautomáticas utilizadas para construção de Mapas Conceituais com o uso de recursos computacionais. A autora descreve que as abordagens disponíveis ainda estão longe de serem totalmente automatizadas, necessitando ainda da ação humana nos processos de geração.

Na proposta de um Extrator de Termos para Criação de Mapas Conceituais elaborada por Dalmolin (2010) é desenvolvida a especificação de uma ferramenta para geração de Mapas Conceituais tendo como referência um conjunto de documentos texto. Através das técnicas descritas na metodologia proposta, os elementos que representam os conceitos e relações são identificados nos textos oferecidos como referência, sendo possibilitado aos professores realizarem os ajustes nos Mapas Conceituais obtidos. Os Mapas Conceituais obtidos podem ser utilizados para o desenvolvimento de atividades educacionais.

Estudos sobre a qualificação das mensagens, como o apresentado por Schrire (2006), podem ser encontrados na literatura. Algumas iniciativas no sentido de automatização dessa qualificação são apresentadas. Nelas as informações do fórum são analisadas de maneira quantitativa ou estatística. Poucas pesquisas têm focado uma possível análise qualitativa das mensagens do fórum, como em Lau et al. (2009).

No trabalho de Gerosa, Fuks e Lucena (2004) é apresentada uma proposta de um sistema de estruturação e categorização manual de mensagens em fórum, denominado AulaNet. Na proposta apresentada, as mensagens são registradas no próprio sistema permitindo ao professor definir as categorias em que as mensagens podem ser categorizadas. Os alunos definem as categorias no qual a mensagem se enquadra. O professor pode realizar a alteração das categorias a que uma determinada mensagem já registrada pelo aluno foi previamente catalogada.

Um estudo sobre comunidades virtuais de aprendizagem em EaD é apresentado por Bassani (2009). Este trabalho realiza uma análise do conteúdo de mensagens, o mapeamento de interações e a análise da relação entre eles, concluindo que a continuidade das trocas interindividuais está relacionada com o texto das mensagens.

As respostas encaminhadas para um fórum em geral são compostas por uma frase curta. Uma parcela destas comunicações, porém, é de mensagens mais elaboradas, o que deixa mais complexa a análise destas informações. Neste caminho, Erlin, Yusof e Rahman (2010) realizaram um estudo para segmentação da análise das mensagens em sentenças unitárias, para simplificar o processo de análise e avaliação no nível de participação dos alunos, utilizando ferramentas de análise de redes sociais SNA (*Social Network Analysis*). A proposta do trabalho é a utilização de sentenças como unidade de interação ao invés de mensagens. Com este enfoque, os autores demonstram que a avaliação do nível de participação é mais eficiente quando é realizada a quebra das mensagens em sentenças.

Uma proposta para o desenvolvimento de uma técnica de levantamento de requisitos para uma ferramenta integrada ao AVA, com objetivo de avaliação de fórum, foi apresentado por Dias Junior e Ferreira (2008). Além da proposta da pesquisa dos requisitos, o trabalho define detalhes de projeto para construção futura de uma ferramenta de avaliação de docentes para integração com fóruns virtuais.

Uma proposta de análise qualitativa das mensagens de um fórum, utilizando técnicas de mineração de dados, é apresentado por Azevedo (2011). Nessa proposta é apresentado o desenvolvimento de uma ferramenta, o Minerafórum, que realiza o cálculo da relevância de cada mensagem do fórum baseado em um texto ou lista de palavras utilizada como referência.

Uma arquitetura para avaliação educacional em fórum de discussão é apresentado por Gonçalves e Elia (2008), incluindo a especificação de um protótipo WEB para avaliação manual das mensagens.

Um trabalho sobre qualificação de mensagens em Fóruns Online é apresentado por Wanas et al. (2008). Nesta proposta é desenvolvido um método para classificação automática das mensagens com a realização do agrupamento em cinco categorias com o objetivo de informar o valor de um *post* para os usuários. O trabalho foi realizado em uma base pública de um fórum, no qual um critério inicial de *ranking* é aplicado a partir da informação de que o usuário é cadastrado no site ou não. Em seguida o *ranking* é ajustado em função do nível de contribuição do usuário no fórum medido pelo nível de inserções e respostas (credibilidade) e em função de palavras-chave obtidas na pesquisa do fórum. Este trabalho contribui na etapa de qualificação de mensagens com a análise sobre utilização de elementos

linguísticos computacionais, ao invés de análise do ponto de vista da frequência de ocorrência de palavras.

Uma proposta de ferramenta para avaliação de fórum, a ser desenvolvida em ambiente *Open Source* e em ambiente Web, foi realizada por Shaul (2007). Nesta proposta, a ferramenta realiza o fornecimento de um indicador numérico em função da participação de cada aluno no fórum.

O trabalho de Rajaraman e Tan (2002) descreve uma técnica de descoberta de conhecimento por meio de busca textual, denominada *Concept Frame Graph*. Nesta técnica, a base de conhecimento conceitual é representada por meio de *Concept Frames* que são obtidos por meio da extração de tuplas do texto, compostas por Substantivo, Verbo, Substantivo. É aplicada então uma técnica para agrupamento dos elementos obtidos.

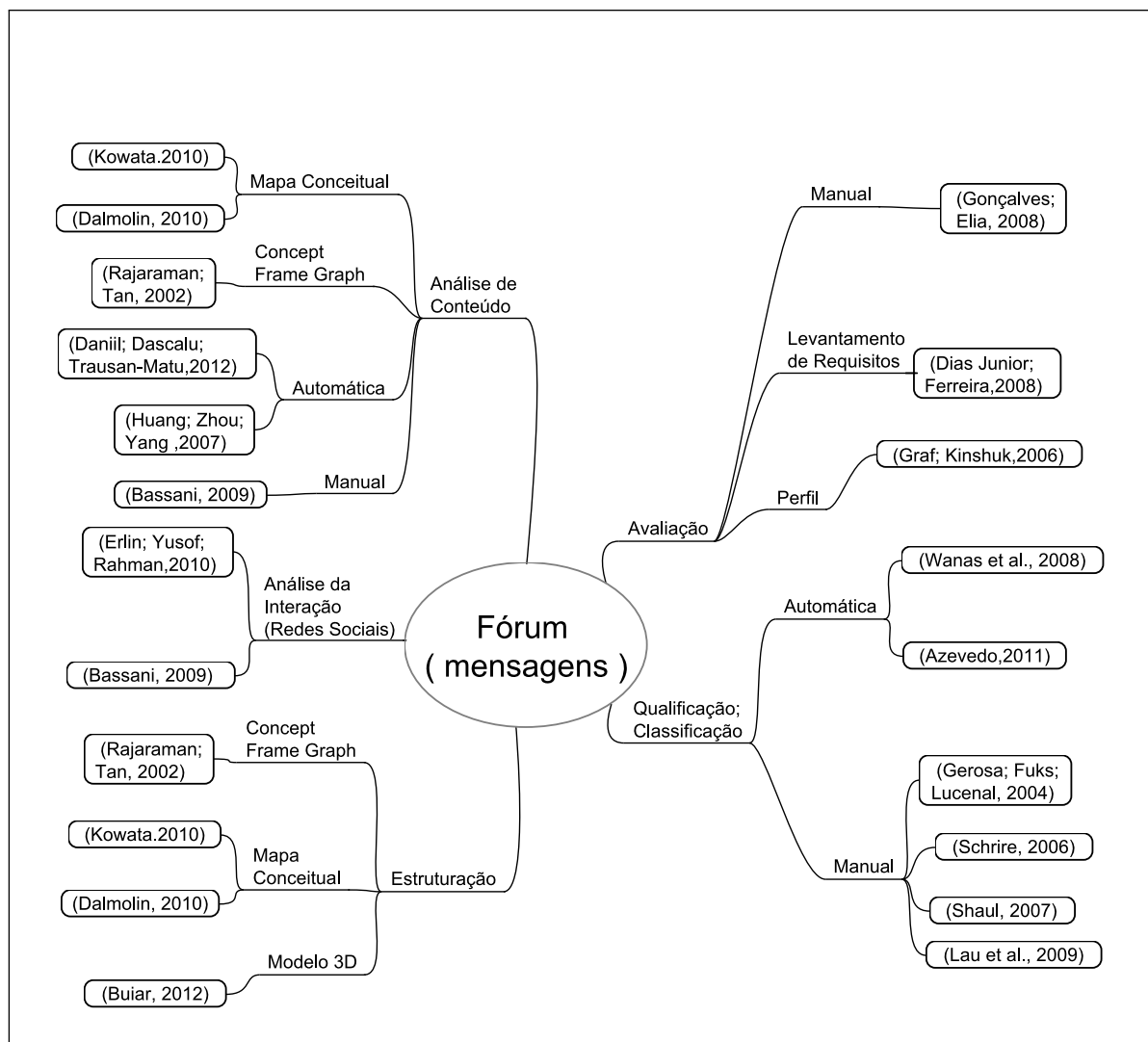
Uma proposta para análise automática de fórum é apresentada por Daniil, Dascalu e Trausan-Matu (2012). Essa proposta está baseada em uma estrutura de múltiplas camadas, compreendendo a extração dos dados de fórum, armazenamento dos dados em arquivos XML, aplicação de processos NLP, análises de conteúdo e geração de graduação e recomendação do conteúdo. A proposta desses autores está baseada na obtenção de dados de fóruns públicos e não voltada a fórum em EaD.

O trabalho realizado por Huang, Zhou e Yang (2007) apresenta um método para extração de conhecimento de fórum de discussão que utilizam agentes automatizados para geração de respostas, conhecidos como *chatbots*. O objetivo da proposta é a alimentação de uma base de conhecimento que será utilizada pelos agentes para a geração das respostas automáticas.

Conforme verificado neste capítulo, diversos trabalhos foram desenvolvidos no sentido de realizar a qualificação das mensagens registradas em um fórum de discussão.

A Figura 2 ilustra o estado da arte apresentado neste capítulo. Nessa figura, as diversas áreas de conhecimento relacionadas ao tratamento de mensagens de um fórum de discussão estão representadas pelos ramos em torno do elemento central da figura, caracterizado pela palavra “fórum”.

O ramo superior esquerdo dessa figura ilustra os estudos de avaliação, que por



**Figura 2 - Estado da Arte**  
**Fonte: Autor**

sua vez, estão desmembrados nas linhas de avaliação manual, levantamento de requisitos e perfil.

Os trabalhos na área de qualificação e classificação das mensagens estão apresentadas no ramo inferior direito dessa figura, relacionando os trabalhos realizados com os enfoques manual e automático.

No ramo superior esquerdo dessa figura, as linhas de trabalho relacionadas com análise de conteúdo são apresentadas, relacionando as formas de análise manual e automática, bem como as abordagens com mapa conceitual e *concept frame graph*.

Logo abaixo desse ramo, estão ilustrados os trabalhos que abordam a análise



da interação.

Finalmente no ramo inferior esquerdo da figura estão ilustrados os trabalhos na área de estruturação de mensagens, desmembrados nas linhas de *concept frame graph*, mapa conceitual e o modelo 3D, que é apresentado nesta pesquisa.

Apesar de todo o desenvolvimento apresentado nesta área, a análise qualitativa do conhecimento armazenado não conta com uma metodologia apropriada. Tanto a avaliação individual quanto a avaliação coletiva dos alunos, carecem de modelos, técnicas e artefatos computacionais que auxiliem o professor nessa atividade.

### 3 MODELO CONCEITUAL

Este capítulo, composto por 3 seções, apresenta e descreve o Modelo Conceitual da solução investigada. A Seção 3.1 apresenta um exemplo real de mensagens de um fórum. A partir desse exemplo, o Modelo de Representação Estruturada das Mensagens de um Fórum ou simplesmente MOREM é descrito.

Na Seção 3.2, apresenta-se como a mensagem estruturada pelo MOREM pode ser visualizada de acordo com três aspectos importantes da mensagem:

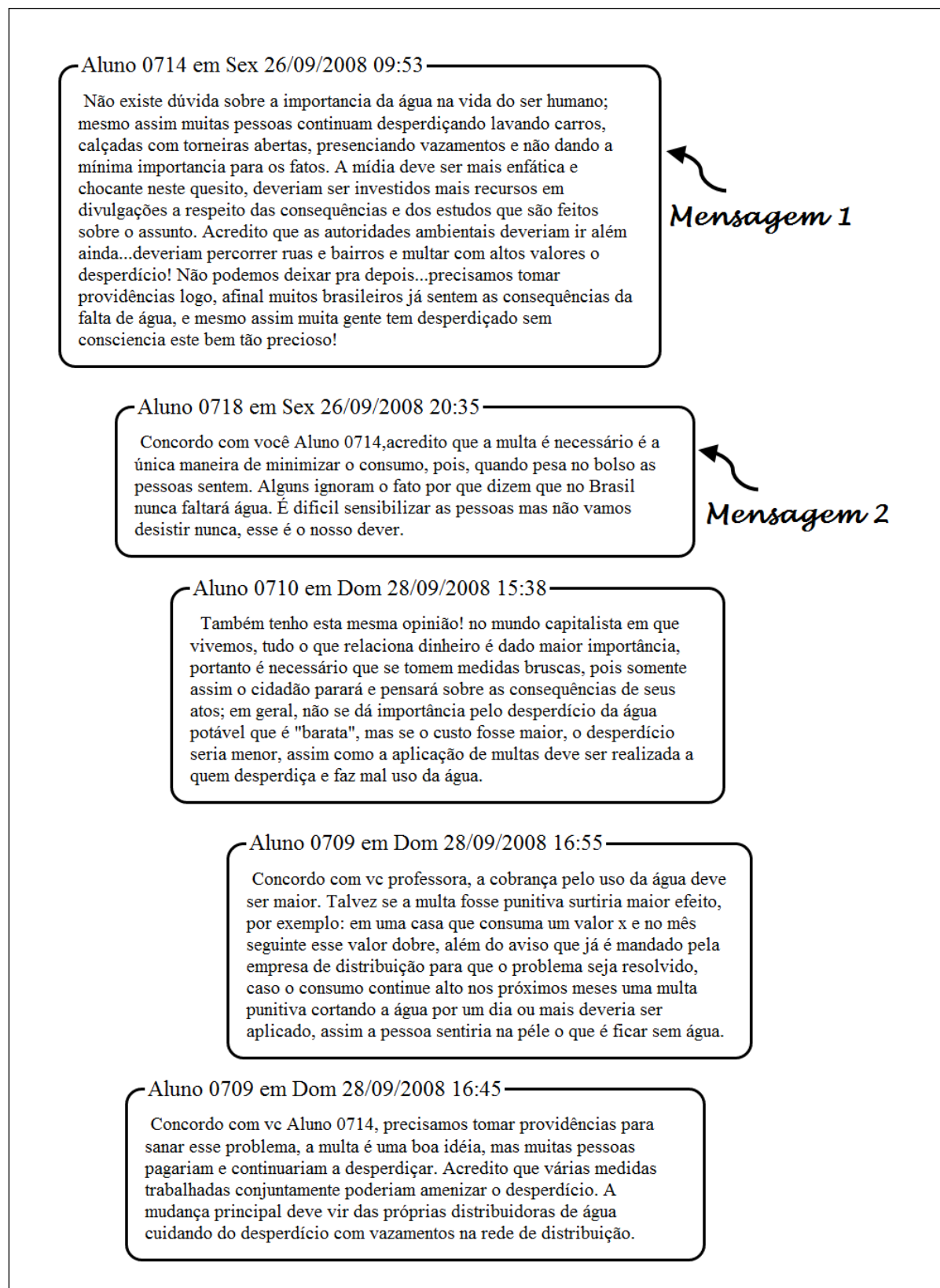
- Conceitos do Domínio constantes da mensagem;
- Aluno que apresentou a mensagem;
- Quando isso ocorreu.

Na Seção 3.3, por sua vez, esses três aspectos são reunidos formando um modelo de representação tridimensional das mensagens do fórum. Finalizando este capítulo, na Seção 3.4, o MOREM é analisado.

#### 3.1 EXEMPLO DA DINÂMICA DE INSERÇÃO DE CONCEITOS EM UM FÓRUM

Um exemplo é utilizado para demonstrar a dinâmica do processo de estruturação das mensagens registradas em um fórum. A Figura 3 ilustra uma sequência de mensagens componentes de uma discussão. Apenas para definição do domínio selecionado, esse trecho de discussão foi aleatoriamente extraído de uma discussão de um fórum realizado em um curso de pós graduação lato sensu em Gestão Ambiental da Universidade Tecnológica Federal do Paraná, no ambiente Moodle.

O texto das mensagens é apresentado na íntegra, sem correções ortográficas. Os nomes dos alunos do curso foram omitidos e um novo nome foi utilizado



**Figura 3 - Sequência de Mensagens de uma Discussão**  
**Fonte: Fórum de discussão do repositório EaD da UTFPR**

para representar a contribuição desse aluno. Esse novo nome consiste no rótulo “ALUNO” seguido de um número sequencial.

Na Figura 3, as cinco caixas de borda arredondada correspondem às cinco mensagens extraídas do ambiente. Cada caixa, representa o autor da mensagem, quando ela foi inserida e o seu conteúdo. Por exemplo, na primeira mensagem da Figura 3 está identificado que o ALUNO 0714 registrou em 26/09/2008 às 09h53 a mensagem com o conteúdo :

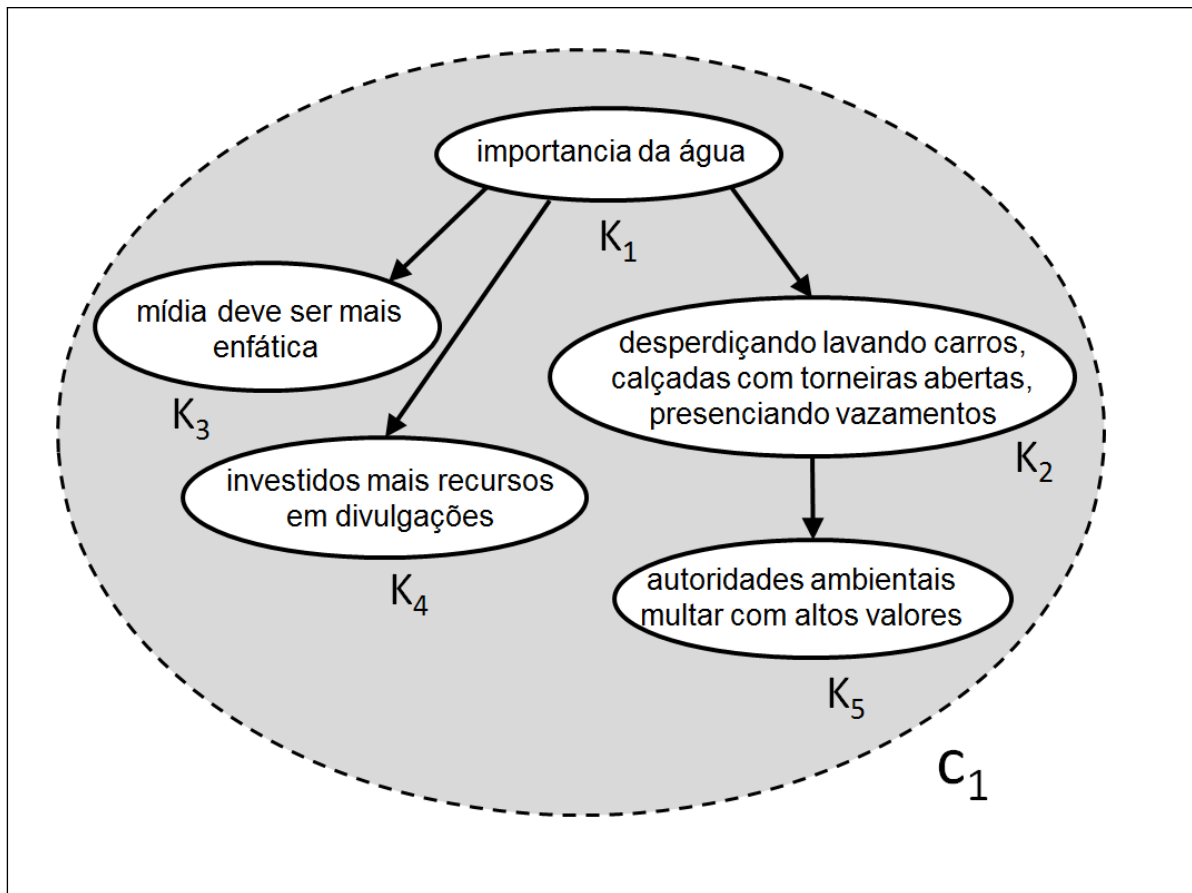
“ Não existe dúvida sobre a importancia da água na vida do ser humano; mesmo assim muitas pessoas continuam desperdiçando lavando carros, calçadas com torneiras abertas, presenciando vazamentos e não dando a mínima importancia para os fatos. A mídia deve ser mais enfática e chocante neste quesito, deveriam ser investidos mais recursos em divulgações a respeito das consequências e dos estudos que são feitos sobre o assunto. Acredito que as autoridades ambientais deveriam ir além ainda... deveriam percorrer ruas e bairros e multar com altos valores o desperdício! Não podemos deixar pra depois...precisamos tomar providências logo, afinal muitos brasileiros já sentem as consequências da falta de água, e mesmo assim muita gente tem desperdiçado sem consciencia este bem tão precioso!”

Este trabalho de pesquisa considera que o texto da mensagem representa parcialmente o domínio de conhecimento, denominado  $C_1$ , que cada aluno possui ou adquire durante a execução da discussão no fórum.

Uma análise preliminar da primeira mensagem ilustrada na Figura 3 permite identificar a presença de alguns elementos do domínio de conhecimento  $C_1$  que poderiam ser considerados como relevantes para o curso de Gestão Ambiental, representados pelas caixas em torno das palavras.

Cada um destes elementos constitui um conceito  $K$  e recebe um índice. Pode-se destacar cinco elementos relevantes:

$K_1$ : importância da água;



**Figura 4 - Conceitos da Mensagem 1**  
**Fonte: Autor**

$K_2$ : desperdiçando lavando carros, calçadas com torneiras abertas, presenciando vazamentos;

$K_3$ : mídia deve ser mais enfática;

$K_4$ : investidos mais recursos;

$K_5$ : autoridades ambientais multar com altos valores.

A proximidade conceitual desses elementos no texto da mensagem são representados por setas, conforme ilustrado na Figura 4. Logo o conhecimento  $C_1$  é representado por  $C_1 \rightarrow \{ K_1, K_2, K_3, K_4, K_5 \}$ .

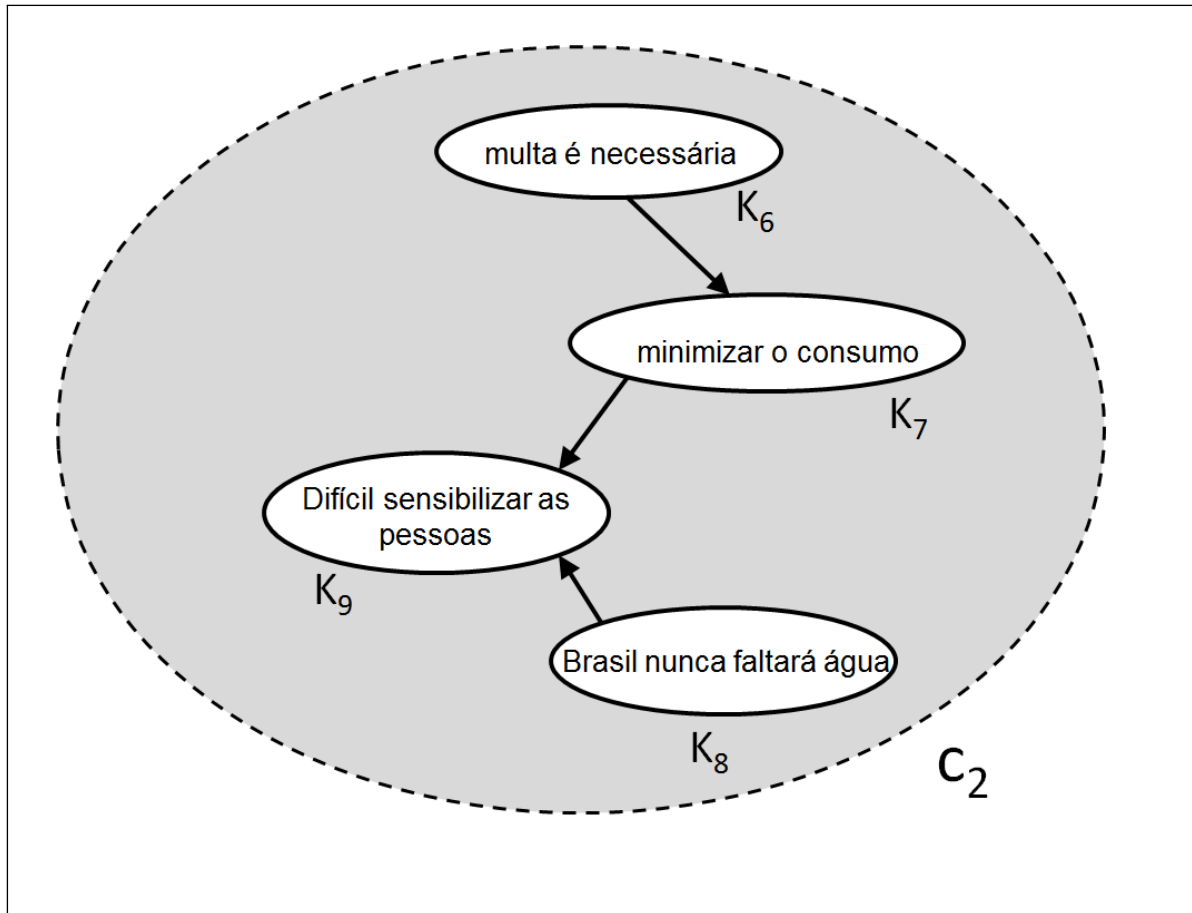
Do mesmo modo, conforme ilustra a Figura 5, a MENSAGEM 2 da Figura 3 pode ser representada através dos seguintes conceitos:

$K_6$ : multa é necessária;

$K_7$ : minimizar o consumo;

$K_8$ : difícil sensibilizar as pessoas;

$K_9$ : nunca faltará água.



**Figura 5 - Conceitos da Mensagem 2**  
Fonte: Autor

A “MENSAGEM 2” teria sido influenciada pela “MENSAGEM 1” ? Uma análise dos conceitos presentes em cada mensagem permite identificar um elemento comum, nesse exemplo. O conceito “muita” está presente em duas células. A célula  $K_5$  contém: “autoridades ambientais multar com altos valores”. A célula  $K_6$ , por sua vez, contém: “muita é necessária”. Nesse caso, considera-se que o elemento “muita” ou “multar” é um elemento comum às duas mensagens e pode-se considerar que a “MENSAGEM 1” influenciou ou fomentou a presença da “MENSAGEM 2”.

Da mesma forma apresentada anteriormente, a identificação dos conceitos relevantes nas MENSAGENS 3, 4 e 5 da Figura 3, podem ser representada como :

### **MENSAGEM 3**

$K_{10}$ : dinheiro é dado maior importância;

$K_{11}$ : tomem medidas bruscas;

$K_{12}$ : importância pelo desperdício;

$K_{13}$ : água potável;

$K_{14}$ : aplicação de multas.

### **MENSAGEM 4**

$K_{15}$ : multa;

$K_{16}$ : consumo continue alto.

### **MENSAGEM 5**

$K_{17}$ : multa;

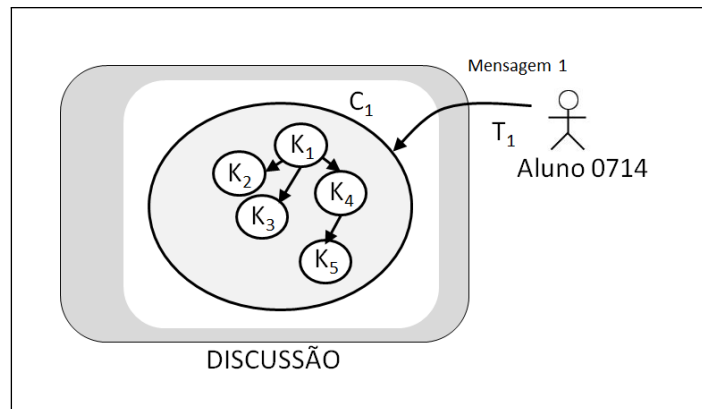
$K_{18}$ : desperdício com vazamentos.

A “MENSAGEM 3” também apresenta o elemento “multa” em seu texto, conforme é possível visualizar na Figura 3. A “MENSAGEM 4” por sua vez, também apresenta o conceito “multa” no trecho “... se a multa fosse punitiva ...”. A “MENSAGEM 5” também apresenta o conceito “multa”.

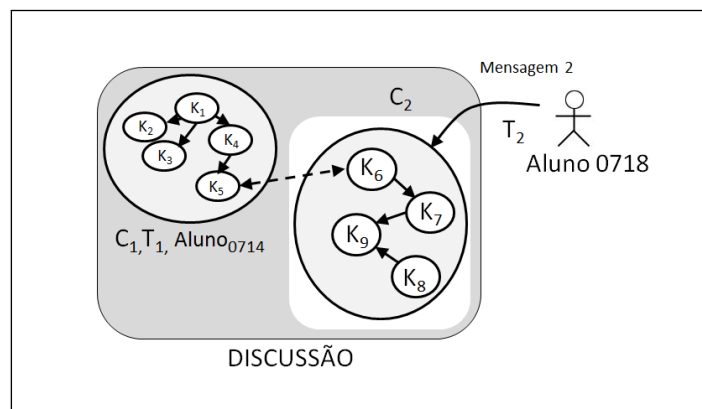
Além do elemento “multa” outro elemento está presente nas mensagens, o termo “água”. De maneira resumida, esse conjunto de cinco mensagens tratou dos elementos “multa” e “água”.

Na Figura 6 ilustra-se os conceitos identificados a partir da MENSAGEM 1 da Figura 3. O momento do registro da mensagem na discussão está rotulado como  $T_1$ .

A Figura 7 ilustra a contribuição do Aluno 0718 ao Fórum por meio da MENSAGEM 2. Pode-se afirmar que os conceitos apresentados pelo ALUNO 0718 na MENSAGEM 2, ilustrados na Figura 7, foram influenciados pelos conceitos apresentados anteriormente. Na Figura 7, essa “influência” é representada pela seta tracejada que liga as células  $K_5$  e  $K_6$ , por meio do elemento comum “multa” ou “multar”. Por questão de simplificação, o termo “água” que também é um termo comum não está sendo analisado na Figura 7.



**Figura 6 - Representação da Mensagem 1**  
**Fonte: Autor**



**Figura 7 - Representação da Mensagem 2**  
**Fonte: Autor**

Prosseguindo com essa análise para as outras mensagens ilustradas na Figura 3, obtém-se a representação gráfica da Figura 8, que representa os conceitos apresentados nessas mensagens, sendo indicadas por setas a existência de conceitos comuns a mais de uma mensagem.

### 3.2 VISUALIZAÇÃO DA DISCUSSÃO

Pode-se verificar que a quantidade de conceitos e suas relações em uma discussão, cresce com o número de mensagens analisadas.

A partir da forma de representação ilustrada na Figura 8, este trabalho de pesquisa define três tipos de visualização, a saber :



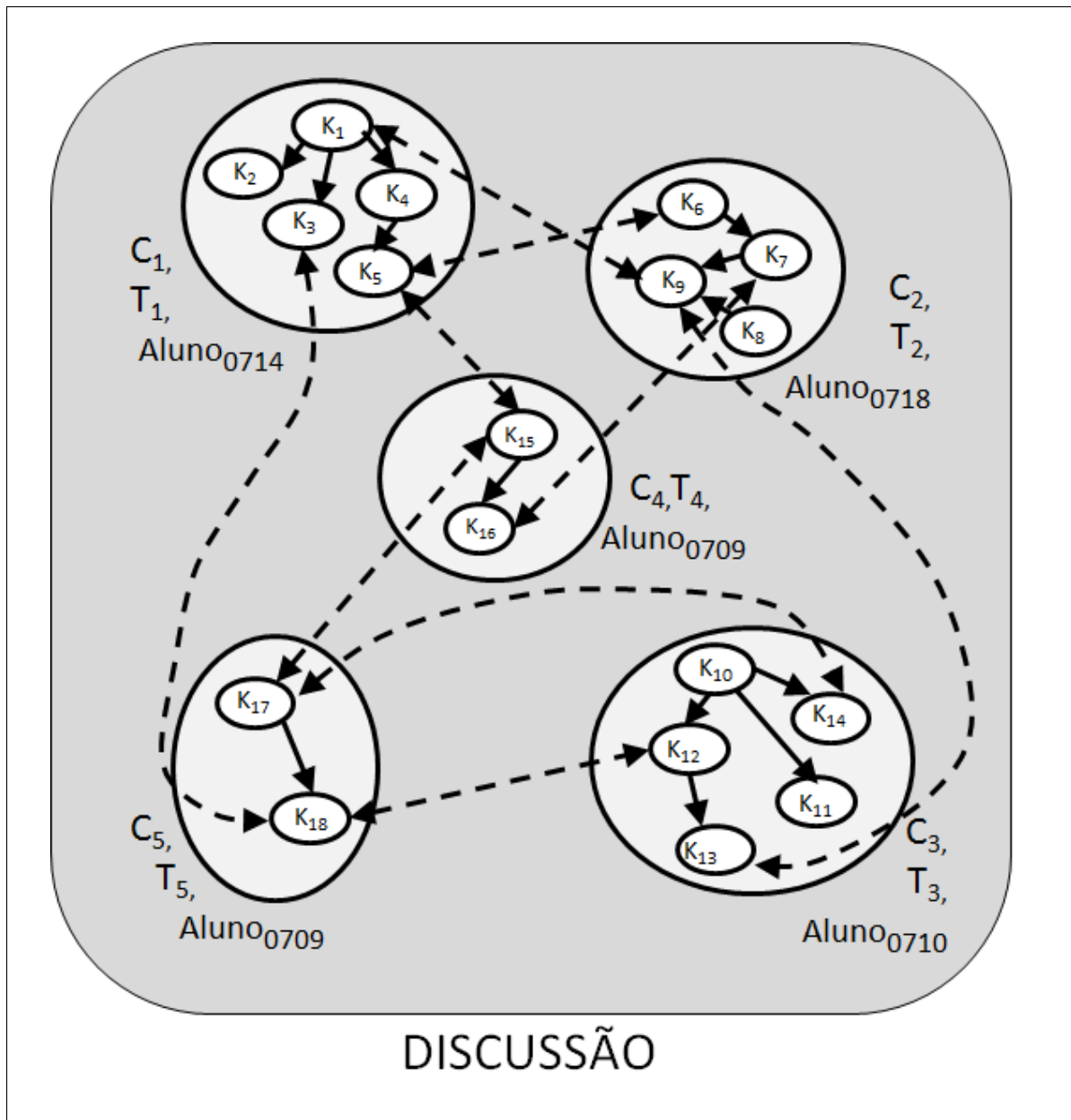


Figura 8 - Conceitos da Discussão e sua Relações  
Fonte: Autor

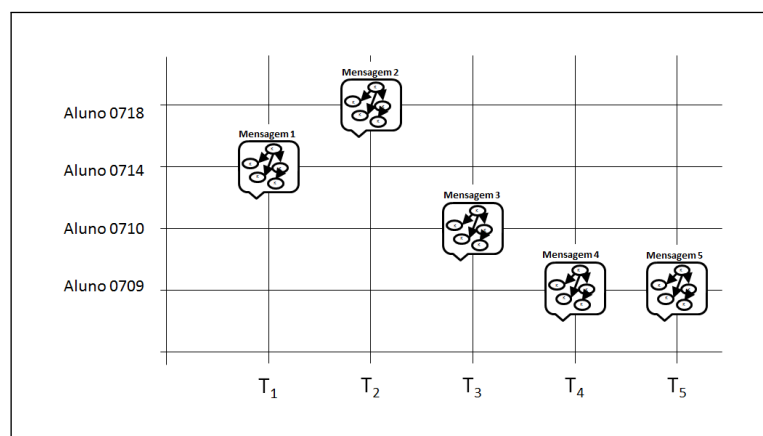
- Visão Usuário x Tempo
- Visão Conceito x Tempo
- Visão Conceito x Usuário

A justificativa da necessidade dessas 3 visões é apresentada nas Seções 3.2.1 a 3.2.3.

### 3.2.1 Visão Usuário x Tempo

A Figura 9, apresenta a forma de visualização escolhida para a visualização da contribuição de cada aluno e quando isso aconteceu.

Espera-se que essa forma de representação permita identificar a partir de que momento um aluno não esteja participando do fórum. Além disso, seria possível identificar a existência de algum padrão comportamental desses alunos, como por exemplo, um aluno inserir uma mensagem sempre que um outro aluno tenha inserido previamente.



**Figura 9 - Diagrama Usuário x Tempo**  
Fonte: Autor

### 3.2.2 Visão Conceito x Tempo

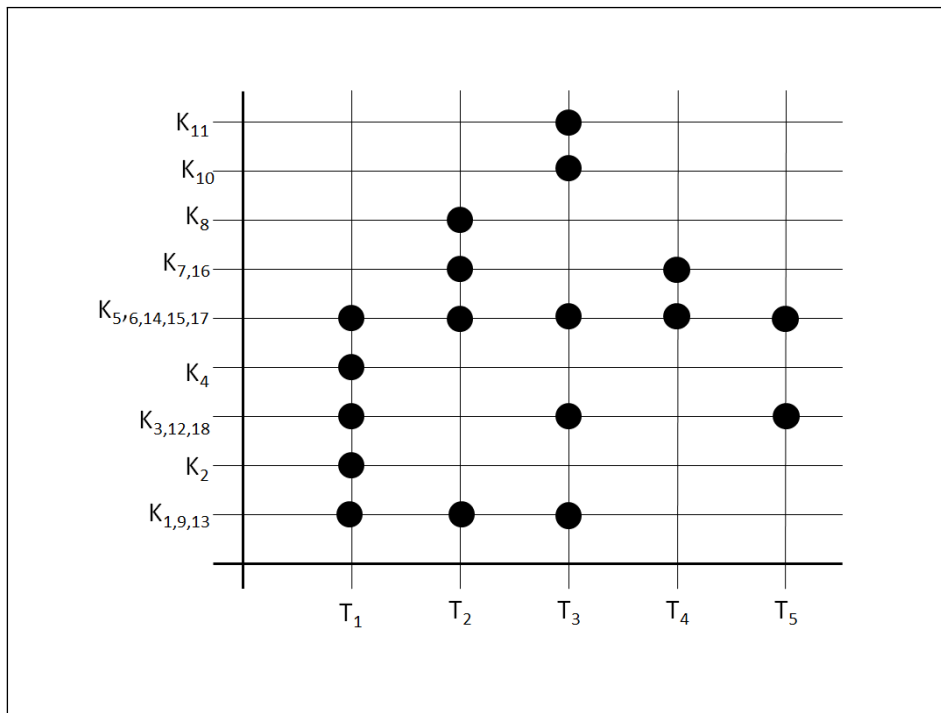
A partir da observação da Dinâmica de Inserção de Conceitos em um Fórum descrita na Seção 3.1, é verificada a ocorrência de conceitos que são apresentados em mais de uma mensagem da discussão, como por exemplo os conceitos “multa” e “água”. Nessa visão, pode-se identificar, o instante de tempo em que ocorreu a primeira ocorrência de um conceito. Caracteriza-se este determinado instante de tempo como sendo o instante de “criação” do conceito, no escopo da discussão.

A partir do instante da “criação” pode ocorrer a repetição desse conceito em outras mensagens, condição essa denominada de “crescimento” do conceito.

O instante de tempo em que ocorre a última ocorrência de um determinado conceito pode ser chamado de “desaparecimento”.

A sequência dos eventos de criação, crescimento e desaparecimento de um determinado conceito ao longo de uma discussão, pode-se chamar de Evolução dos Conceitos em Função do Tempo.

É possível a representação da Evolução dos Conceitos em Função do Tempo, através de um diagrama denominado VISÃO CONCEITO X TEMPO, apresentado na Figura 10.



**Figura 10 - Diagrama Conceito x Tempo**  
Fonte: Autor

### 3.2.3 Visão Conceito x Usuário

A VISÃO CONCEITO x USUÁRIO relaciona cada conceito presente nas mensagens e qual usuário que inseriu esse conceito.

O diagrama que representa esta visão está ilustrado na Figura 11.

Com esse diagrama, pode-se visualizar quais conceitos foram apresentados por quais alunos, professores e tutores. Complementarmente é possível verificar que aluno deixou de abordar quais conceitos. Além disso, pode-se identificar quais usuários apresentaram uma menor ou maior quantidade de conceitos.

		Usuário			
		Aluno 0709	Aluno 0710	Aluno 0714	Aluno 0718
Conceito	$K_{11}$		X		
	$K_{10}$		X		
	$K_8$				X
	$K_{7,16}$	X			X
	$K_{5,6,14,15,17}$	X	X	X	X
	$K_4$			X	
	$K_{3,12,18}$	X	X	X	
	$K_2$			X	
	$K_{1,9,13}$		X	X	X

**Figura 11 - Diagrama Conceito x Usuário**  
**Fonte: Autor**

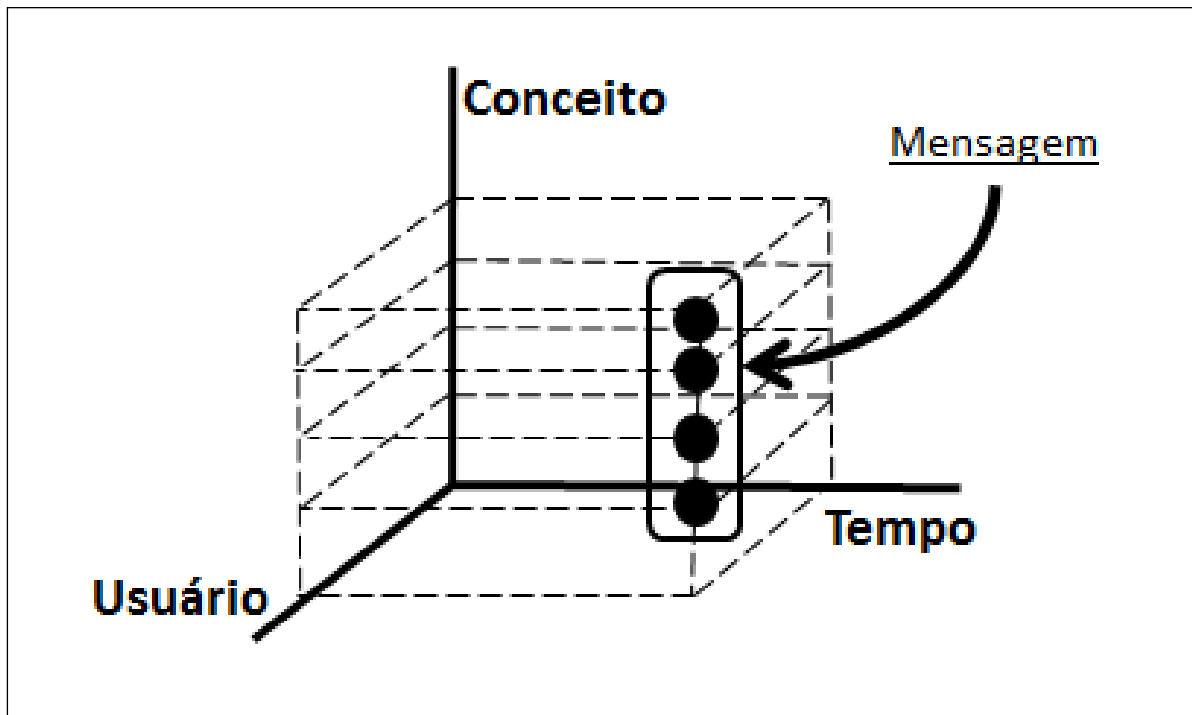
Espera-se que essa visão facilite uma possível avaliação ou classificação dos participantes do fórum.

### 3.3 MODELO DE REPRESENTAÇÃO TRIDIMENSIONAL

Tendo como referência as visões descritas na Seção 3.2 pode-se imaginar uma forma de representação que agregue essas 3 dimensões.

As dimensões TEMPO, USUÁRIO e CONCEITO, podem ser unidas em uma VISÃO TRIDIMENSIONAL resultante, conforme ilustrado de maneira simbólica na Figura 12.

Na Figura 12 uma mensagem qualquer é posicionada em um espaço tridimensional. Logo, essa mensagem do fórum é representada de maneira estruturada em três dimensões e ela pode ser visualizada ou interpretada de acordo



**Figura 12 - Modelo de representação tridimensional da mensagem**  
**Fonte: Autor**

com cada par de eixos, conforme apresentado nas Seções 3.2.1 a 3.2.3.

### 3.4 ANÁLISE DO MODELO

Através da identificação dos conceitos que compõem uma mensagem, o MOREM permite o desenvolvimento de um Modelo Computacional que simplifica o processo de consulta ao conhecimento registrado nas mensagens do fórum, em virtude da natureza estruturada de representação do conhecimento, proposta nas três dimensões apresentadas.

Como características principais do MOREM pode-se destacar :

- Identificação dos conceitos que compõem as mensagens;
- Agregação do usuário e tempo em cada conceito;
- Especificação de uma estrutura de armazenamento tridimensional das informações.

## 4 MODELO COMPUTACIONAL

Este capítulo descreve o Modelo Computacional desenvolvido a partir do Modelo Conceitual descrito no Capítulo 3.

A Seção 4.1 apresenta o Modelo Computacional proposto e seus componentes. O Módulo Extrator é apresentado na Seção 4.2. A Seção 4.3 descreve o Módulo Estruturador. O Módulo Classificador é abordado na Seção 4.4 e a Seção 4.5 descreve o Módulo Visualizador. Na Seção 4.6 tem-se a apresentação das conclusões obtidas com este capítulo.

### 4.1 MODELO PROPOSTO

O modelo conceitual apresentado no Capítulo 3 define um modelo de estruturação das mensagens registradas em um fórum de um AVA. Esse modelo é utilizado como referência para a especificação da arquitetura e funcionalidades necessárias ao desenvolvimento de um sistema computacional. O sistema computacional especificado deve permitir a realização da simulação do modelo de estruturação proposto com dados reais.

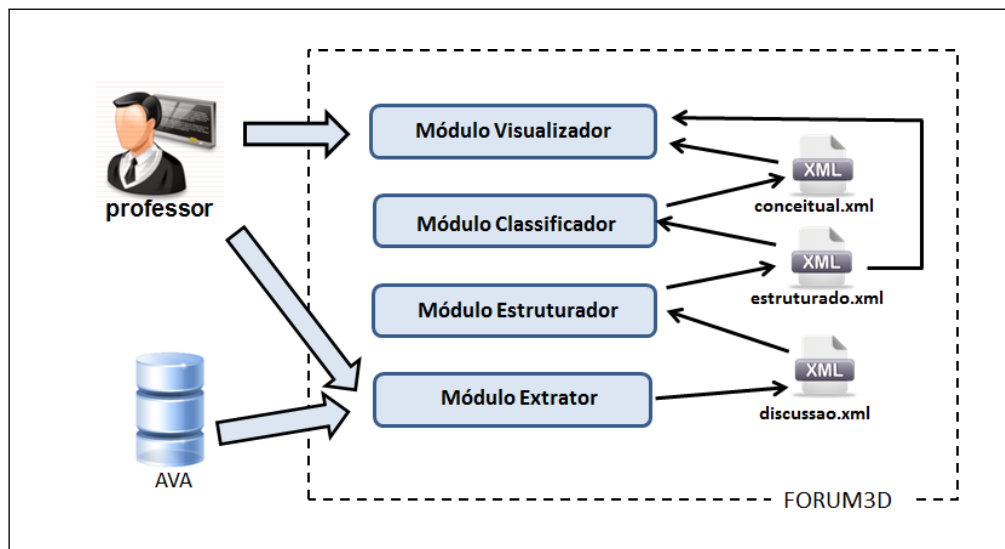
O sistema computacional concebido, denominado FORUM3D, permite a extração das informações de um fórum de discussão. Esse sistema também realiza a estruturação dessas informações conforme o modelo de estruturação de mensagens, ou MOREM, proposto nesta pesquisa. Finalmente, o FORUM3D realiza a apresentação das visualizações previstas no modelo conceitual concebido.

O modelo computacional desenvolvido é composto pelas seguintes partes :

- Módulo Extrator
- Módulo Estruturador

- Módulo Classificador
- Módulo Visualizador

A Figura 13 ilustra o relacionamento entre as partes componentes, bem como ilustra o fluxo de informações entre os componentes e os arquivos XML gerados.



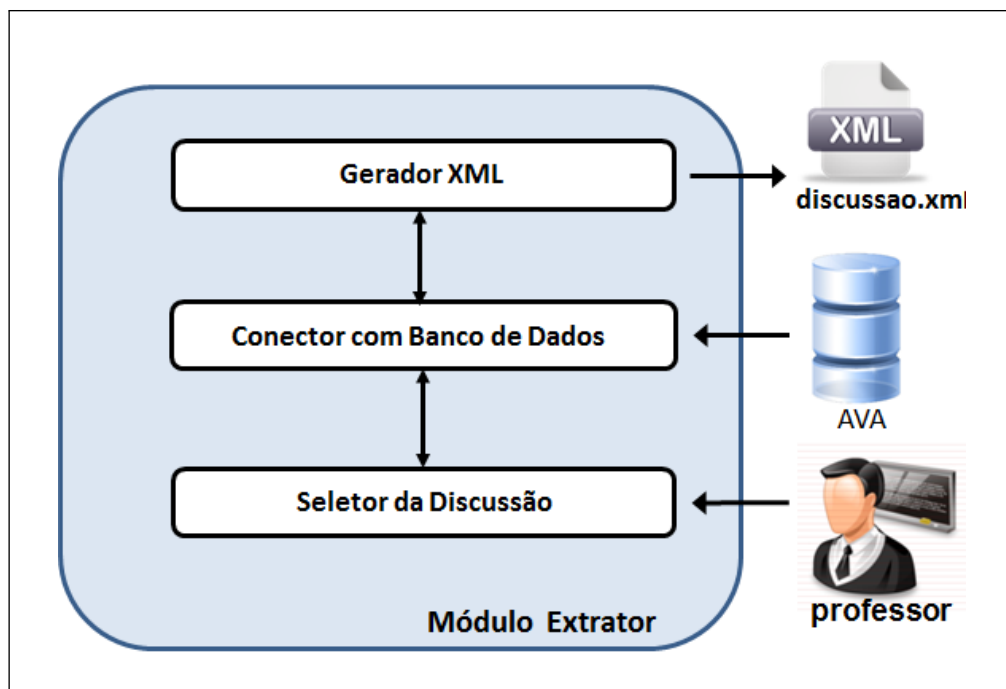
**Figura 13 - Modelo Computacional do FORUM3D**  
Fonte: Autor

#### 4.2 MÓDULO EXTRATOR

Em termos gerais, o Módulo Extrator é responsável pela obtenção das informações armazenadas no AVA. As funcionalidades necessárias para este módulo são :

- Gerenciar as informações para conexão ao banco de dados;
- Obter as informações referentes aos cursos, fóruns e discussões;
- Obter as informações referentes aos usuários;
- Selecionar a discussão a ser analisada;
- Obter as informações referentes às mensagens registradas na discussão selecionada.

A Figura 14 apresenta os componentes que formam o Módulo Extrator.



**Figura 14 - Módulo Extrator**  
**Fonte: Autor**

#### 4.2.1 Componente Conector com Banco de Dados

As mensagens do fórum são armazenadas em um gerenciador de banco de dados utilizado pelo AVA. Neste modelo computacional está previsto o acesso direto a esse gerenciador. A opção de acesso direto a esse gerenciador foi escolhida como uma simplificação para a implementação de um protótipo de validação do modelo.

No Componente Conector com Banco de Dados são armazenadas as informações referentes ao endereço do servidor do banco de dados, usuário, senha e nome do banco de dados utilizados pelo AVA.

O acesso realizado ao banco de dados do AVA não realiza operações de escrita, não sendo, portanto, invasivo ao AVA.

O usuário utilizado para o acesso ao gerenciador de banco de dados, pelo sistema computacional especificado neste capítulo, necessita possuir somente privilégios de leitura ao banco de dados e somente para as tabelas acessadas, que são :

- *mdl\_course*, contém informações sobre os Cursos;



- *mdl\_forum*, contém informações sobre os Fóruns;
- *mdl\_forum\_discussions*, contém informações sobre as Discussões;
- *mdl\_forum\_posts*, contém informações sobre as Mensagens;
- *mdl\_user*, contém informações sobre os Usuários.

Este módulo é responsável pela realização do acesso às informações das listas de cursos, fóruns e discussões para envio ao Componente Seletor da Discussão.

Este módulo também realiza a leitura das informações das mensagens de uma discussão específica, conforme seleção realizada pelo usuário no Componente Seletor da Discussão.

Através de uma tabela de mapeamento, é realizada a correspondência entre as entidades físicas do gerenciador de banco de dados utilizadas, ou seja, as tabelas e colunas, e os correspondentes ELEMENTOS utilizados no modelo desenvolvido. Como elemento, define-se os campos (*tags*) utilizados nos arquivos XML, especificados neste modelo, para armazenamento das informações estruturadas. O Quadro 1 apresenta o mapeamento definido neste modelo.

Nesse quadro estão apresentadas as colunas utilizadas em cada tabela do banco de dados e o correspondente elemento que será utilizado no arquivo XML. Utilizando como referência a coluna "*mdl\_forum\_discussions.name*", as informações extraídas dessa coluna, do banco de dados, serão armazenadas no elemento "DISCUSSÃO" com a denominação "NOME".

#### 4.2.2 Componente Seletor da Discussão

O objetivo deste componente é oferecer uma interface gráfica que permita a seleção da discussão para a qual o professor deseja realizar a análise. Essa interface é do tipo hierárquica, possibilitando a seleção do curso, logo após dos fóruns deste curso e em seguida das discussões, dentre as disponíveis. A obtenção da relação de cursos, fóruns e discussões é realizada a partir do componente Conector com Banco de Dados, descrito na Seção 4.2.1.

Com a obtenção do identificador único da discussão selecionada, as informações de todas as mensagens dessa discussão, correspondentes aos

<b>Tabela</b>	<b>Coluna</b>	<b>Elemento</b>	<b>Descrição</b>
<b>CURSO</b>			
mdl_course	id	ID	Identificador único do curso
mdl_course	fullname	NOME	Nome do curso
<b>FÓRUM</b>			
mdl_forum	id	ID	Identificador único do fórum
mdl_forum	name	NOME	Nome do fórum
<b>DISCUSSÃO</b>			
mdl_forum_discussions	id	ID	Identificador único da discussão
mdl_forum_discussions	name	NOME	Nome da discussão
<b>MENSAGEM</b>			
mdl_forum_posts	id	ID	Identificador único da mensagem
mdl_forum_posts	created	DATAHORA	Horário da geração da mensagem
mdl_forum_posts	userid	USUÁRIO	Identificador do usuário
mdl_forum_posts	parent	ANTERIOR	Identificador da mensagem antecessora
mdl_forum_posts	message	TEXTO	Texto da mensagem
<b>USUÁRIO</b>			
mdl_user	id	ID	Identificador único do Usuário
mdl_user	firstname	NOME	Nome do usuário
mdl_user	lastname		Sobrenome do usuário

**Quadro 1 - Mapeamento do banco de dados**  
**Fonte: Autor**

elementos relacionados no Quadro 1, são extraídas do AVA através do Componente Conector com Banco de Dados, e enviadas ao Componente Gerador XML, descrito na Seção 4.2.3, para obtenção do arquivo XML que contém as informações da discussão.

#### 4.2.3 Componente Gerador XML

A partir das informações da discussão selecionada no Componente Seletor da Discussão, descrito na Seção 4.2.2, o componente Gerador XML realiza o armazenamento dos dados em um arquivo XML de acordo com a estrutura ilustrada no Quadro 2.

Esse arquivo, que ainda registra os dados das mensagens armazenadas de forma textual, será utilizado com entrada para o Módulo Estruturador, descrito na Seção 4.3.

Elemento	Descrição		
ID	Identificador único da Discussão		
GERAÇÃO	Horário ( <i>timestamp</i> ) em que o arquivo foi gerado		
NOME	Nome da Discussão		
CURSO	Contém as informações do Curso a que está vinculado o Fórum que contém a Discussão		
	ID	Identificador único do Curso	
	NOME	Nome do Curso	
FÓRUM	Contém as informações do Fórum a que a Discussão está vinculada		
	ID	Identificador único do Fórum	
	NOME	Nome do Fórum	
POSTS	Sequência de Mensagens ( <i>posts</i> ) que formam a Discussão		
	POST	Informações de cada Mensagem	
		ID	Identificador único da Mensagem
		DATAHORA	Horário de geração da Mensagem
		USUÁRIO	Identificador único do Usuário que inseriu a Mensagem
		ANTERIOR	Identificador Único da Mensagem antecessora
TEXTO	Texto inserido na Mensagem		
USUÁRIOS	Informações de todos os Usuários que inseriram Mensagens na Discussão		
	USUÁRIO	Informações de cada Usuário	
		ID	Identificador único do Usuário
		NOME	Nome do Usuário (Nome + Sobrenome)

**Quadro 2 - Estrutura do arquivo discussao.xml**

**Fonte: Autor**

#### 4.3 MÓDULO ESTRUTURADOR

O MOREM representa o conhecimento, registrado na mensagem da discussão, como um conjunto de elementos sintáticos que compõem essa mensagem. Cada um desses elementos, formado por uma ou mais palavras, constitui uma unidade sintática que está desempenhando uma função na mensagem. Essa unidade sintática será referida como sintagma.

O Módulo Estruturador é responsável pela estruturação do texto livre, presente nas mensagens, através do desmembramento do texto em um conjunto de sintagmas equivalente. Para a realização do processo de estruturação desse texto são utilizadas ferramentas de Processamento de Linguagem Natural (NLP).

Dentre as diversas ferramentas de NLP disponíveis no mercado, durante os experimentos realizados neste projeto de pesquisa, foi utilizado o *Apache OpenNLP Toolkit* (BALDRIDGE, 2005). Os principais fatores que nortearam esta escolha

foram :

- Suporte para as tarefas básicas de NLP;
- Fácil obtenção;
- Documentação *online* de fácil acesso;
- Suporte de modelos para diversos idiomas, incluindo português do Brasil;
- Possibilidade de chamada das ferramentas através de linha de comando;
- Disponibilidade de API bem documentada, para utilização em conjunto com programas desenvolvidos pelo usuário;
- Distribuída para utilização dentro da modalidade de licença *Apache License, Version 2.0*, que permite a utilização sem necessidade de pagamento de taxas ou *royalties*.

A arquitetura definida para o Módulo Estruturador está baseada na chamada aos programas de linha de comando disponibilizados pelo *Apache OpenNLP Toolkit*, realizada a partir do Componente Estruturador, descrito nesta seção, com os modelos de treinamento para a língua portuguesa baseado no *corpus* “bosque” (FERNANDES; SANTOS; MILIDIU, 2010).

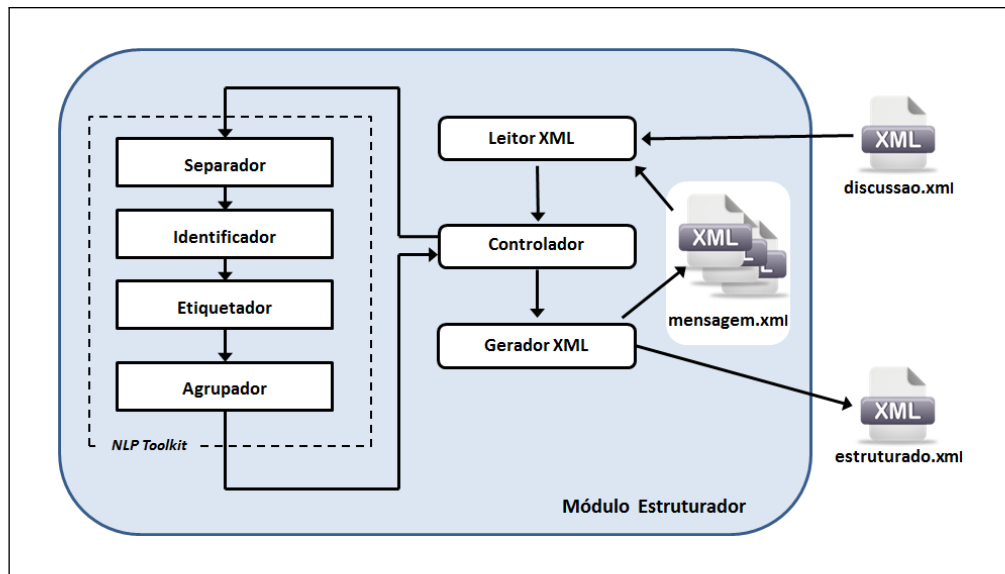
A Figura 15 apresenta os componentes que formam o Módulo Estruturador. A descrição das funcionalidades desses componentes está apresentada nas Seções 4.3.1 a 4.3.7.

#### 4.3.1 Componente Controlador

O Componente Controlador é responsável pelo gerenciamento dos processos internos do Módulo Estruturador.

Os processos realizados nesse módulo são :

- Recuperar as informações do arquivo “discussao.xml” através do Componente Leitor XML;
- Submeter as informações armazenadas no elemento TEXTO de cada mensagem, obtidas na etapa anterior, ao processo de estruturação NLP;



**Figura 15 - Módulo Estruturador**  
Fonte: Autor

- Armazenar os dados estruturados de cada mensagem em um arquivo XML específico, identificado genericamente como “mensagem.xml”, através do Componente Gerador XML;
- Consolidar os dados estruturados de todas as mensagens em um arquivo “estruturado.xml” através do Componente Gerador XML.

O processo de estruturação NLP é realizado através de uma chamada encadeada de processos, para as ferramentas disponibilizadas pelo *Apache OpenNLP Toolkit*, via sistema operacional. Essa técnica é conhecida como *software pipeline* (RITCHIE; THOMPSON, 1974, p. 371). Os processamentos realizados nesse processo de estruturação estão apresentados nas Seções 4.3.2 a 4.3.5.

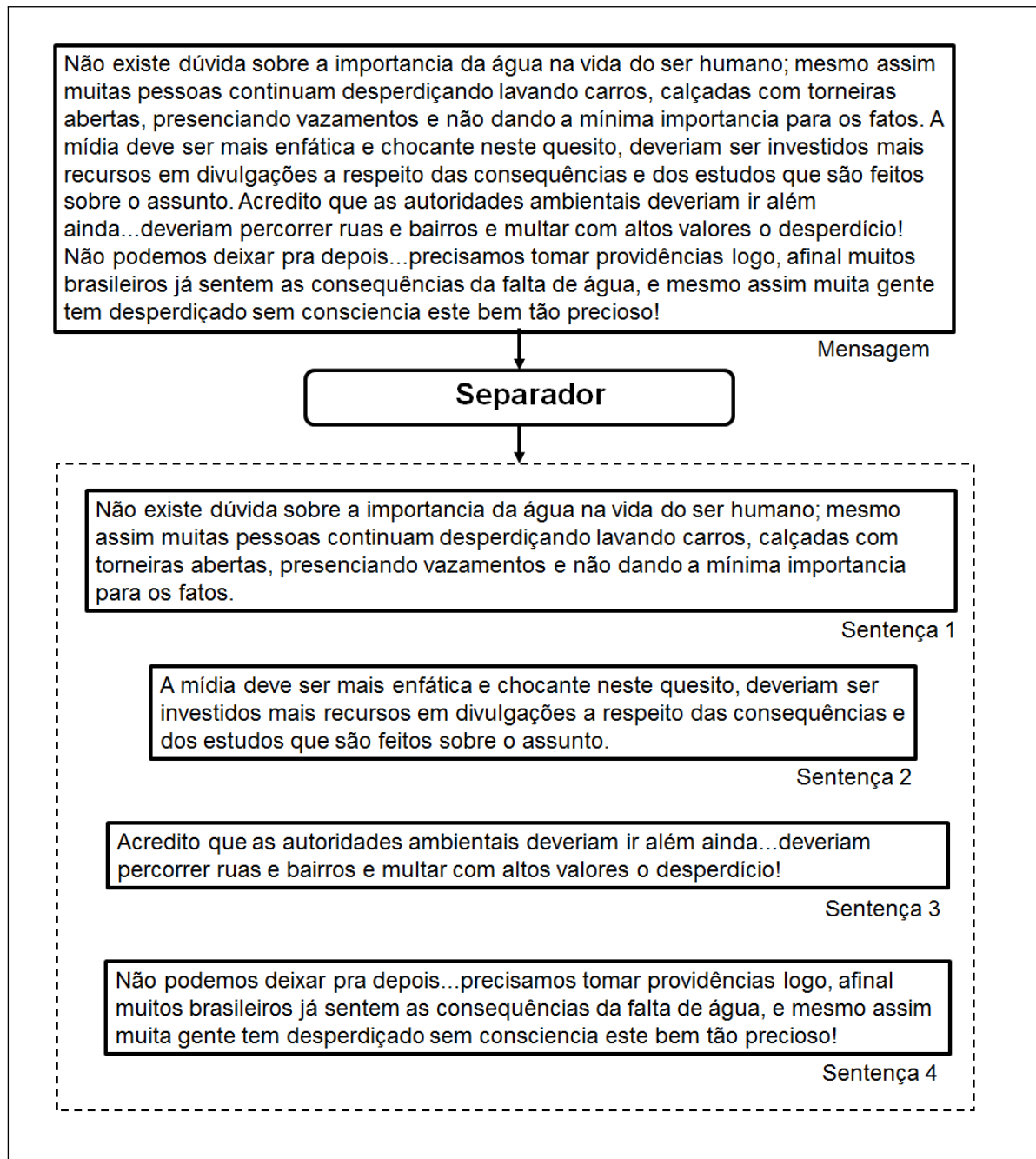
Após o processo de obtenção dos sintagmas que representam cada mensagem, ocorre a geração dos arquivos XML de saída, detalhados na Seção 4.3.7.

#### 4.3.2 Componente Separador

Esse componente realiza a identificação das sentenças presentes na mensagem, através da localização dos sinais de pontuação presentes. Os dados recebidos do Componente Controlador, correspondentes ao texto de uma

mensagem, passam pelo processo de identificação de sentenças, gerando como saída um texto que contém uma sentença por linha.

A Figura 16 ilustra o processo de identificação de sentenças. Nesse exemplo foi utilizada a Mensagem 1 do mesmo trecho da discussão apresentado no Modelo Conceitual (Figura 3).



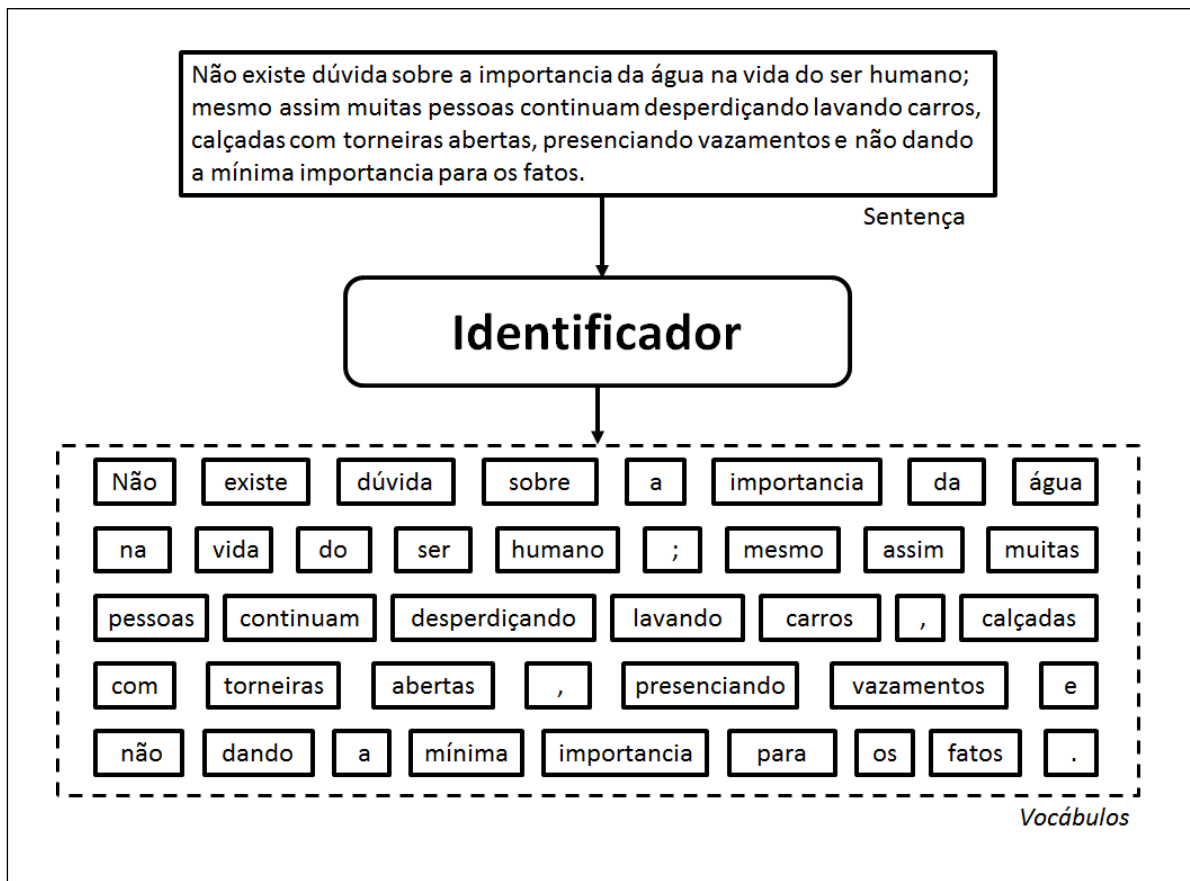
**Figura 16 - Separador de Sentenças**  
Fonte: Autor

Para a implementação deste componente é utilizada a ferramenta *Sentence Detector* do *Apache OpenNLP Toolkit*.

O fluxo de dados de saída do Componente Separador é encaminhado como entrada para o Componente Identificador, descrito na Seção 4.3.3.

#### 4.3.3 Componente Identificador

Este componente é responsável pela identificação dos vocábulos presentes em cada sentença. A Figura 17 ilustra o processo de identificação dos vocábulos aplicadas na primeira sentença da mensagem utilizada como exemplo na Seção 4.3.2.



**Figura 17 - Identificador de Vocábulos**  
Fonte: Autor

Para a implementação deste componente é utilizada a ferramenta *Tokenizer* do *Apache OpenNLP Toolkit*, com a opção de detecção de limites baseado no modelo probabilístico treinado no *corpus* "bosque". A partir da identificação dos limites

sintáticos dos vocábulos, o que inclui também os sinais de pontuação e valores numéricos, é gerada uma sequência de vocábulos delimitadas por espaços em branco como saída deste componente.

O fluxo de dados de saída do Componente Separador é encaminhado como entrada para o Componente Etiquetador, descrito na Seção 4.3.4.

#### 4.3.4 Componente Etiquetador

O processo de etiquetagem consiste da categorização dos vocábulos recebidos e da inserção de uma identificação, ou etiqueta, em cada vocábulo, para indicar o seu tipo. A saída desse processo consiste de uma lista contendo os vocábulos etiquetados, conforme a Figura 18, na qual está ilustrado o processo de etiquetagem realizado nos vocábulos utilizados como exemplo na Seção 4.3.3.

Para a implementação deste componente é utilizada a ferramenta *POS Tagger (Part of Speech)* do *Apache OpenNLP Toolkit*. Essa ferramenta utiliza o modelo probabilístico para categorização dos vocábulos, treinado no *corpus* “bosque”.

O *POS Tagger* insere a identificação anexando ao final de cada vocábulo o caractere *underline* seguido de uma sequência de caracteres, ou etiqueta, que identifica sua categoria. Dessa forma, após a realização do processo de etiquetagem no vocábulo “humano”, por exemplo, será obtido como saída o vocábulo etiquetado “humano\_adj”, indicando a categoria “adjetivo”.

O fluxo de dados de saída, contendo a lista dos vocábulos etiquetados, é direcionada para o componente descrito na Seção 4.3.5.

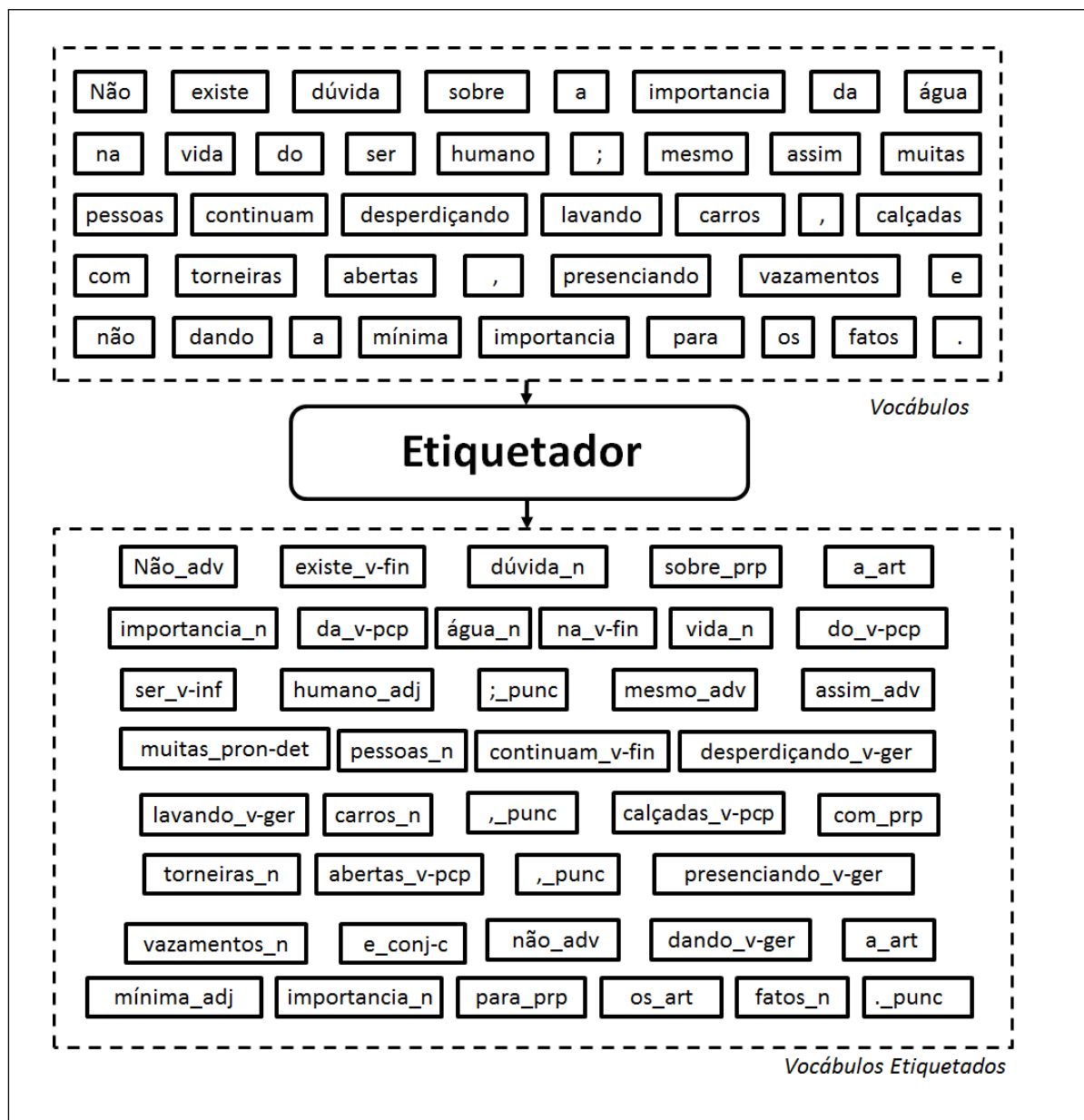
#### 4.3.5 Componente Agrupador

A última etapa do processamento de linguagem natural compreende o agrupamento dos vocábulos etiquetados em sintagmas. Esse processo é realizado através do agrupamento dos vocábulos vizinhos que estejam sintaticamente correlacionados na frase.

A Figura 19 ilustra o processo de agrupamento de vocábulos aplicado na mensagem utilizada como exemplo nas seções anteriores deste capítulo.

O Componente Agrupador utiliza a ferramenta *ChunkerME* do *Apache OpenNLP*





**Figura 18 - Etiquetador de Vocábulos**  
 Fonte: Autor

*Toolkit*. Essa ferramenta utiliza o modelo probabilístico para categorização dos vocábulos, treinado no *corpus* "bosque".

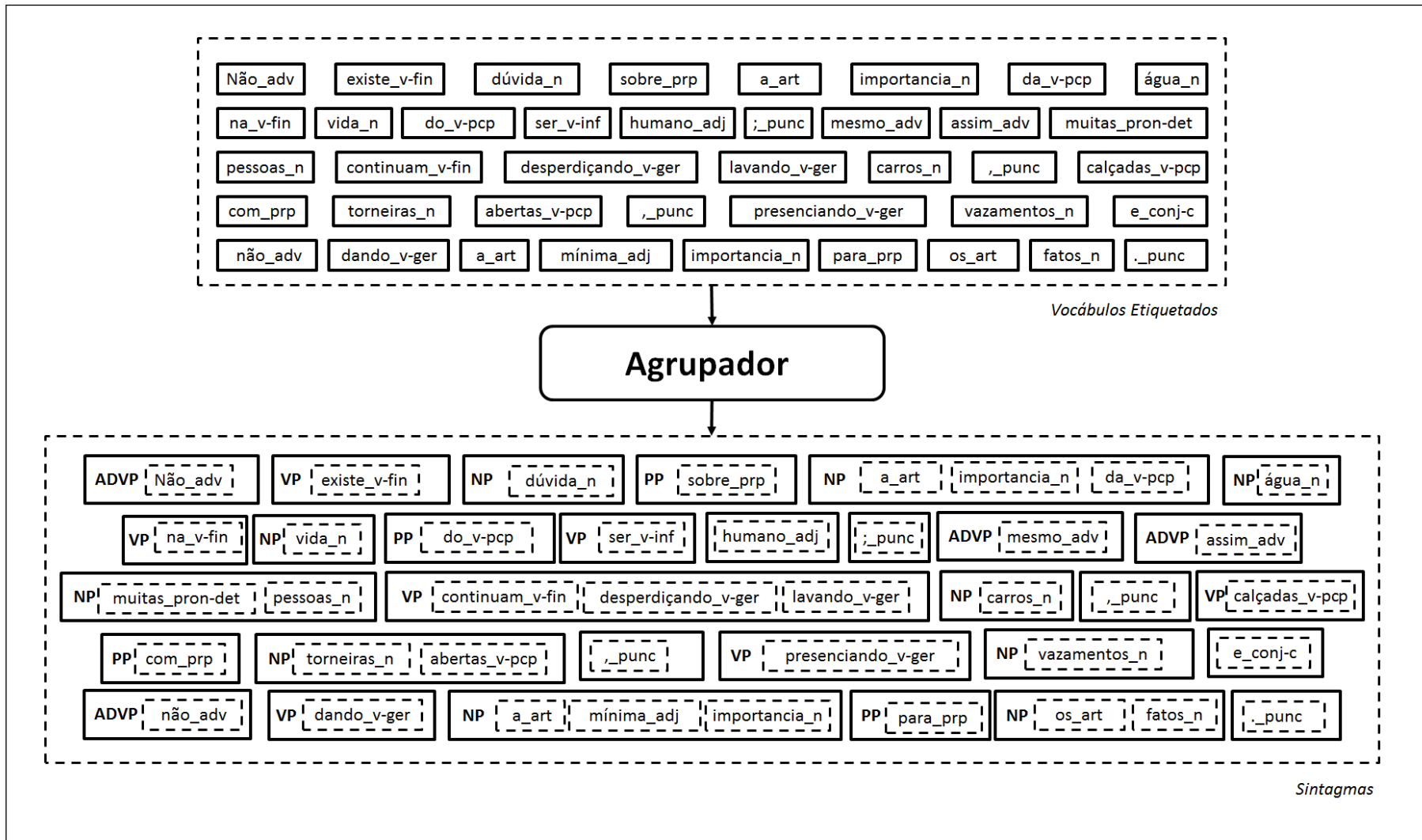


Figura 19 - Agrupador de Sintagmas  
 Fonte: Autor

#### 4.3.6 Componente LeitorXml

O Componente Leitor XML é responsável por realizar a leitura das informações da discussão previamente armazenada no arquivo "discussao.xml".

#### 4.3.7 Componente Gerador XML

Este componente realiza a gravação dos arquivos XML utilizados no processo de estruturação, sendo acionado pelo Componente Controlador.

Após o processo de estruturação, é gerado um arquivo XML correspondente para cada mensagem (*post*) presente na discussão selecionada. Esses arquivos são identificados como "mensagem\_nnnn.xml", onde "nnnn" corresponde ao identificador único da mensagem, obtido do arquivo "discussao.xml" (Quadro 2). A estrutura do arquivo XML obtido está ilustrada no Quadro 3.

Elemento	Descrição		
ID	Identificador único da Mensagem ( <i>Post</i> )		
GERACAO	Horário ( <i>timestamp</i> ) em que o arquivo foi gerado		
DATAHORA	Horário ( <i>timestamp</i> ) em que a Mensagem foi gerada		
ANTERIOR	Identificador único da Mensagem antecessora		
USUARIO	Contém as informações do Usuário		
	ID	Identificador único do Usuário	
	NOME	Nome do Usuário	
SINTAGMAS	Sequência de Sintagmas obtidos da estruturação da mensagem da Mensagem		
	S	Sintagma	
		VALOR	Conteúdo (texto) do Sintagma
		TIPO	Tipo do Sintagma
		SUFIXO	Sufixo do Sintagma

**Quadro 3 - Estrutura do arquivo mensagem.xml**  
Fonte: Autor

A partir da realização da estruturação de todas as mensagens da discussão, ocorre o armazenamento das informações das mensagens da discussão de forma consolidada, em um arquivo denominado "estruturado.xml". A estrutura desse arquivo está ilustrada no Quadro 4.

O arquivo "estruturado.xml" contém o registro do conhecimento armazenado na discussão do fórum, através dos sintagmas identificados, dos usuários e do tempo

Elemento	Descrição		
ID	Identificador único da Discussão		
GERAÇÃO	Horário ( <i>timestamp</i> ) em que o arquivo foi gerado		
NOME	Nome da Discussão		
CURSO	Contém as informações do Curso a que a Discussão está vinculada		
	ID	Identificador único do Curso	
	NOME	Nome do Curso	
FÓRUM	Contém as informações do Fórum ao qual a Discussão está vinculada		
	ID	Identificador único do Fórum	
	NOME	Nome do Curso	
POSTS	Sequência de Mensagens registradas na Discussão		
	POST	Informações de cada Mensagem	
		ID	Identificador único da Mensagem
		DATAHORA	Horário de geração da Mensagem
		USUÁRIO	Identificador único do Usuário que inseriu a Mensagem
		ANTERIOR	Identificador Único da Mensagem antecessora
		SINTAGMAS	Sequência de Sintagmas
S	Identificador único do Sintagma		
USUÁRIOS	Informações de todos os Usuários que inseriram Mensagens neste fórum		
	USUÁRIO	Informações de cada Usuário	
		ID	Identificador único do Usuário
		NOME	Nome do Usuário
SINTAGMAS	Informações de todos os Sintagmas		
	S	Informações de cada Sintagma	
		ID	Identificador único do Sintagma
		VALOR	Conteúdo (texto) do Sintagma
		POST	Identificador único do <i>post</i> a que pertence o Sintagma
		TIPO	Tipo do Sintagma
SUFIXO	Sufixo do Sintagma		

**Quadro 4 - Estrutura do arquivo estruturado.xml**

Fonte: Autor

em que ocorreram as contribuições individuais. Esse arquivo contém uma estrutura similar ao do arquivo XML obtido com o Módulo Extrator, descrito na Seção 4.2. A diferença está no fato de que este novo arquivo contém as informações das mensagens de forma estruturada.

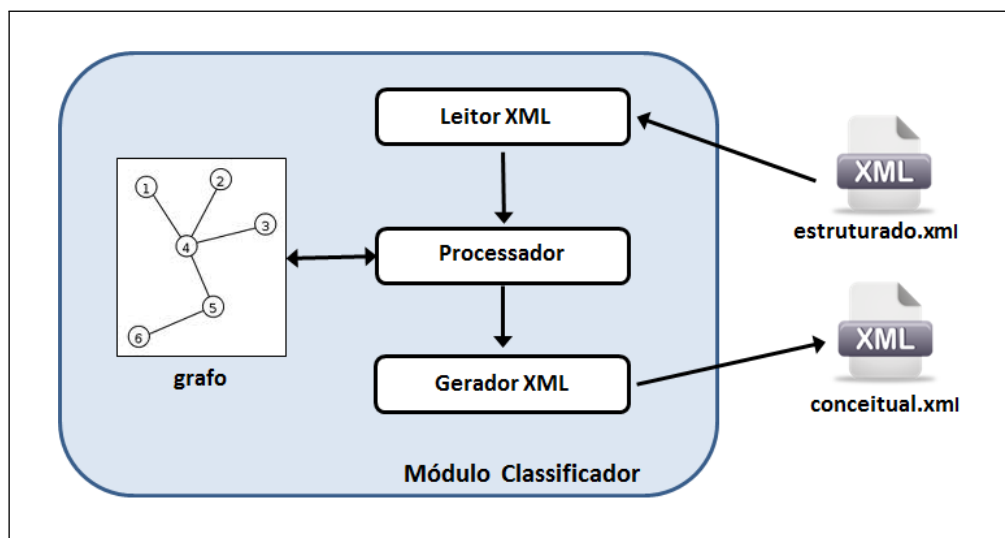
Os elementos do tipo “POST” não contém mais o elemento “TEXTO” e em seu lugar está presente o elemento “SINTAGMAS”. Este elemento “SINTAGMAS” contém uma sequência de elementos “S” que armazenam somente a referência para os sintagmas obtidos dos arquivos intermediários que representam cada mensagem.

Um outro elemento “SINTAGMAS” é inserido ao final deste arquivo, contendo as informações detalhadas de cada sintagma presente no fórum. Estas informações representam o identificador único gerado para cada sintagma, o texto do sintagma, a referência para a mensagem (*post*) a que pertence este sintagma, o tipo do sintagma e o sufixo do sintagma, caso exista.

#### 4.4 MÓDULO CLASSIFICADOR

Por meio da identificação dos sintagmas presentes nas mensagens do fórum, é possível a classificação destes em conceitos e relações. A partir dessa classificação é oferecido ao professor um Mapa Conceitual que permite a visualização dos conceitos e relações registrados no fórum.

A Figura 20 apresenta os componentes que formam o Módulo Classificador.



**Figura 20 - Módulo Classificador**  
Fonte: Autor

##### 4.4.1 Componente LeitorXml

O Componente Leitor XML é responsável por realizar a leitura das informações estruturadas, da discussão, armazenada no arquivo “estruturado.xml”.

#### 4.4.2 Componente Processador

Este componente realiza o processo de análise dos sintagmas das diversas mensagens da discussão, obtidos através do Componente Leitor XML. Essa análise consiste da distinção entre os sintagmas nominais e verbais, que representam conceitos e as palavras de ligação, ou relações (NOVAK; GOWIN, 1984). Para a realização desse processo, é utilizada uma estrutura de dados na forma de grafo.

Os sintagmas verbais, que representam as relações, são mapeados como arestas, ao passo que os sintagmas nominais, que representam os conceitos, são mapeados como nós. Os sintagmas que não forem classificados como verbais ou nominais, como o caso dos separadores e sinais de pontuação são descartados.

Esse processo deve identificar os elementos de reconhecimento da língua portuguesa, tendo como base o estudo já realizado por Kowata (2010). Os elementos de reconhecimento utilizados no desenvolvimento do Módulo Classificador estão descritos no quadro 5.

Elemento	Descrição	Exemplo
NP	Sintagma Nominal	"o secretário"...
CCNP	Sintagma Nominal precedido de conjunção coordenativa	"e o secretário"...
PRNP	Sintagma Nominal precedido de pronome relativo	"que a nova lei"
CSNP	Sintagma Nominal precedido de conjunção subordinativa	[disseram] "que o grupo"
QLNP	Sintagma Nominal precedido de preposição "de"	"do ministro"
SPNP	Sintagma Nominal precedido de preposição	"para a fabricação"
VP	Sintagma Verbal	"foi selecionada"
SPVP	Sintagma Verbal precedido de preposição	"para cobrar"
CCVP	Sintagma Verbal precedido de conjunção coordenativa	"e alterava"
PRVP	Sintagma Verbal precedido de pronome relativo	[ele] "que se aposentou"
CSVP	Sintagma Verbal precedido de conjunção subordinativa	[dizem] "que a Portaria 1.275 revelou"
RP	Advérbio	"assim"
VRG	Vírgula	","
PNT	Ponto final, ponto de interrogação, ponto de exclamação	"."
DPT	Dois pontos	".:"
PTV	Ponto e vírgula	".;"

**Quadro 5 - Elementos de reconhecimento**

Fonte: Kowata (2010, p. 73)

Após a finalização do processo de classificação, as informações classificadas são enviadas para o Componente Gerador XML.

#### 4.4.3 Componente Gerador XML

Este componente realiza a gravação do arquivos XML de saída, sendo acionado pelo Componente Processador.

A partir da realização da classificação das mensagens estruturadas, ocorre o armazenamento dessas informações classificadas em um arquivo denominado “conceitual.xml”. A estrutura desse arquivo está ilustrada no Quadro 6.

<b>Elemento</b>	<b>Descrição</b>		
SINTAGMAS	Sequência de Sintagmas obtidos da estruturação da Discussão		
	S	Sintagma	
		ID	Identificador único do Sintagma
		CATEGORIA	C=Conceito, R=Relação
		DADO	Conteúdo (texto) do sintagma
		U	Identificador único do Usuário
T	Data e Hora da inserção da Mensagem da qual foi extraída o Sintagma		
USUÁRIOS	Sequência indicando usuários da Discussão		
	USUÁRIO	Contém as informações dos Usuários que inseriram Mensagens na Discussão	
		ID	Identificador único do Usuário
NOME	Nome do Usuário		
TEMPOS	Sequência relacionando todos as informações de data e hora em que foram inseridas Mensagens na Discussão		
	T	Data e Hora	
CONCEITOS	Sequência indicando todos os Sintagmas que são Conceitos		
	C	Identificador único do Sintagma	
RELAÇÕES	Sequência indicando todos os Sintagmas que são Relações		
	R	Contém as informações das Relações	
		S	Identificador único do Sintagma
		ORIGEM	Identificador único do Sintagma de origem
DESTINO	Identificador único do Sintagma de destino		

**Quadro 6 - Estrutura do arquivo conceitual.xml**

**Fonte: Autor**

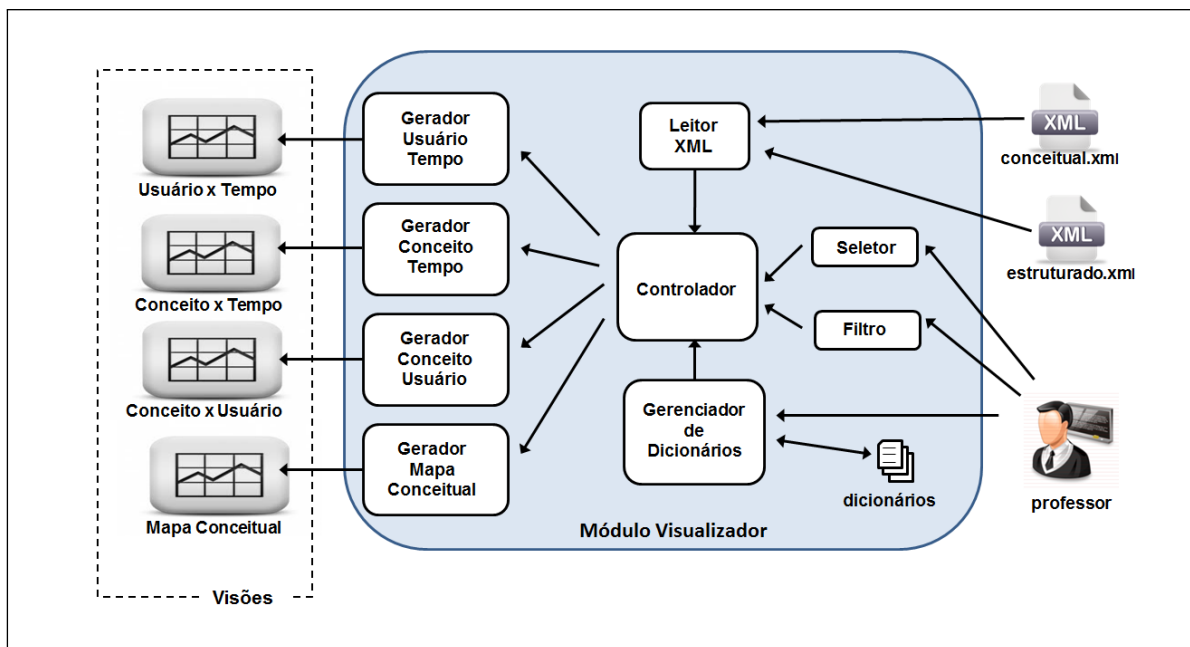
#### 4.5 MÓDULO VISUALIZADOR

O Módulo Visualizador é responsável pela apresentação de relatórios gráficos que representem o conhecimento registrado na discussão, selecionada pelo professor, de acordo com os critérios escolhidos.

As funcionalidades apresentadas por este módulo são:

- Recuperar as informações da discussão estruturada e classificada dos arquivos gerados nos outros módulos;
- Selecionar o escopo desejado para apresentação dos resultados;
- Gerenciar um conjunto de conhecimentos definidos pelo professor através de dicionários;
- Selecionar a forma de apresentação dos resultados, ou visões.

A Figura 21 apresenta os componentes que formam o Módulo Classificador.



**Figura 21 - Diagrama do Módulo Visualizador**  
Fonte: Autor

#### 4.5.1 Componente Controlador

Este componente realiza o controle do fluxo dos processos que permitem a realização da visualização escolhida.

A partir das opções selecionadas pelo professor por meio do Componente Filtro e do Componente Seletor, o Componente Controlador recebe os dados enviados pelo Componente Leitor XML, realiza a filtragem selecionada e envia os



dados preparados para o Componente Gerador específico para a realização da apresentação da visualização escolhida.

#### 4.5.2 Componente Leitor XML

Este componente realiza a leitura das informações armazenadas nos arquivos “conceitual.xml” e “estruturado.xml” para envio ao Componente Controlador.

#### 4.5.3 Componente Seletor

Este componente oferece a apresentação de uma interface gráfica para a seleção do tipo de visão desejada, conforme especificado nas Seções 4.5.6 a 4.5.9.

#### 4.5.4 Componente Filtro

Este componente permite a seleção do escopo desejado para a apresentação realizada, de acordo com três tipo de filtro :

- Usuário
- Tempo
- Conhecimento

O filtro de usuário oferece uma interface gráfica para seleção múltipla, apresentando a relação de usuários que participaram da discussão selecionada. Caso um ou mais usuários sejam selecionados, somente as mensagens que forem apresentadas por esses usuários serão enviados para o gerador da visão selecionada. Desta forma é oferecido ao professor o recurso de realização da análise individual ou de um determinado grupo dentro da discussão do fórum.

O filtro de tempo oferece uma interface gráfica para permitir a seleção do tempo de início e do tempo final, inicialmente selecionados com os valores correspondentes ao horário em que a primeira mensagem e a última foram registradas, respectivamente, na discussão.

Somente serão enviados para o gerador de visão selecionado as informações das mensagens que estiverem dentro do intervalo selecionado. Desta forma, o

professor tem o recurso de restringir a pesquisa dentro de um intervalo de tempo específico.

O filtro de conhecimento oferece ao professor a possibilidade de realizar a apresentação de uma visão contendo somente, os conhecimentos que pertençam a um domínio por ele pré-estabelecido. Esse domínio é denominado DICIONÁRIO e está detalhado na Seção 4.5.5. Com a seleção de um dicionário, o professor pode verificar quais alunos realizaram contribuições, dentro de um domínio de conhecimento relacionado ao assunto tratado, pela atividade de fórum proposto, por exemplo.

#### 4.5.5 Componente Gerenciador de Dicionários

Este componente oferece uma interface gráfica para administração do conjunto de dicionários utilizados com o FORUM3D. A partir dessa interface, o professor pode criar um novo dicionário, inserindo inicialmente o nome do dicionário que está sendo criado.

A partir da criação de um dicionário, uma lista de palavras pode ser inserida através dessa interface, bem como pode ocorrer a alteração ou remoção de uma determinada palavra.

Caso necessário, um dicionário pode ser removido, utilizando a interface oferecida por este componente.

Cada dicionário é formado por um arquivo texto que contém uma lista de vocábulos que compõem o domínio de conhecimento de interesse.

#### 4.5.6 Componente Gerador Usuário Tempo

O Componente Gerador Usuário Tempo permite a geração de um diagrama que apresenta a visualização da Visão Usuário x Tempo, definida na Seção 3.2.1.

A partir desse diagrama, estima-se que o professor possa realizar o acompanhamento da frequência de colaborações realizadas por um determinado aluno, de um grupo de alunos específico ou mesmo de todo o grupo.

Essas consultas podem ser refinadas através da aplicação dos filtros definidos na Seção 4.5.4.

#### 4.5.7 Componente Gerador Conceito Tempo

O Componente Gerador Conceito Tempo oferece a visualização da Visão Conceito x Tempo, definida na Seção 3.2.2.

Um agrupamento dos conceitos similares deve ser realizado com o objetivo de facilitar a visualização dos resultados a serem obtidos. Para a realização desse agrupamento é utilizado um algoritmo que verifica a similaridade entre os conceitos.

Diversas técnicas para verificação da similaridade de texto estão disponíveis na literatura. Para este projeto foi realizada a opção pela utilização da similaridade fonética utilizando o algoritmo Soundex (KNUTH, 2006, p. 391) para a detecção de similaridade entre dois conceitos. A opção utilizada levou em conta a simplicidade da implementação desse algoritmo, visto que o objetivo do experimento a ser realizado é a verificação do modelo de estruturação proposto, estando fora de seu escopo a definição das melhores técnicas de comparação fonética.

A partir da visualização oferecida por este componente estima-se oferecer ao professor o recurso para acompanhamento da Dinâmica de Inserção de Conceitos, descrita na Seção 3.2.2. Através desse acompanhamento o professor pode verificar a partir de que momento um determinado conceito, ou grupo de conceitos, começa a ser discutido no fórum, quando ocorre essa discussão e finalmente quando se encerra a discussão dos conceitos selecionados.

#### 4.5.8 Componente Gerador Conceito Usuário

O Componente Gerador Conceito Usuário oferece a visualização da Visão Conceito x Usuário, definida na Seção 3.2.3.

O agrupamento de conceitos também ocorre quando da geração dessa visualização, utilizando o Algoritmo Soundex.

Esta visão pode permitir ao professor relacionar quais alunos estão realizando contribuições no fórum relacionadas ao conjunto de conceitos definidos na pesquisa. Estima-se com isto identificar os grupos de alunos que apresentam maior interesse, ou participação, dentro do conjunto de assuntos de interesse do professor.

#### 4.5.9 Componente Gerador Mapa Conceitual

O Componente Gerador Mapa Conceitual realiza a apresentação de um gráfico que representa o mapa conceitual correspondente à discussão selecionada. Busca-se com a utilização dos mapas conceituais, oferecer ao professor uma forma de visualização dos conhecimentos registrados no fórum, através dos conceitos e relações. A representação visual desses conhecimentos pode ser um importante instrumento para o professor obter uma realimentação do conteúdo gerado pelos alunos pela atividade do fórum.

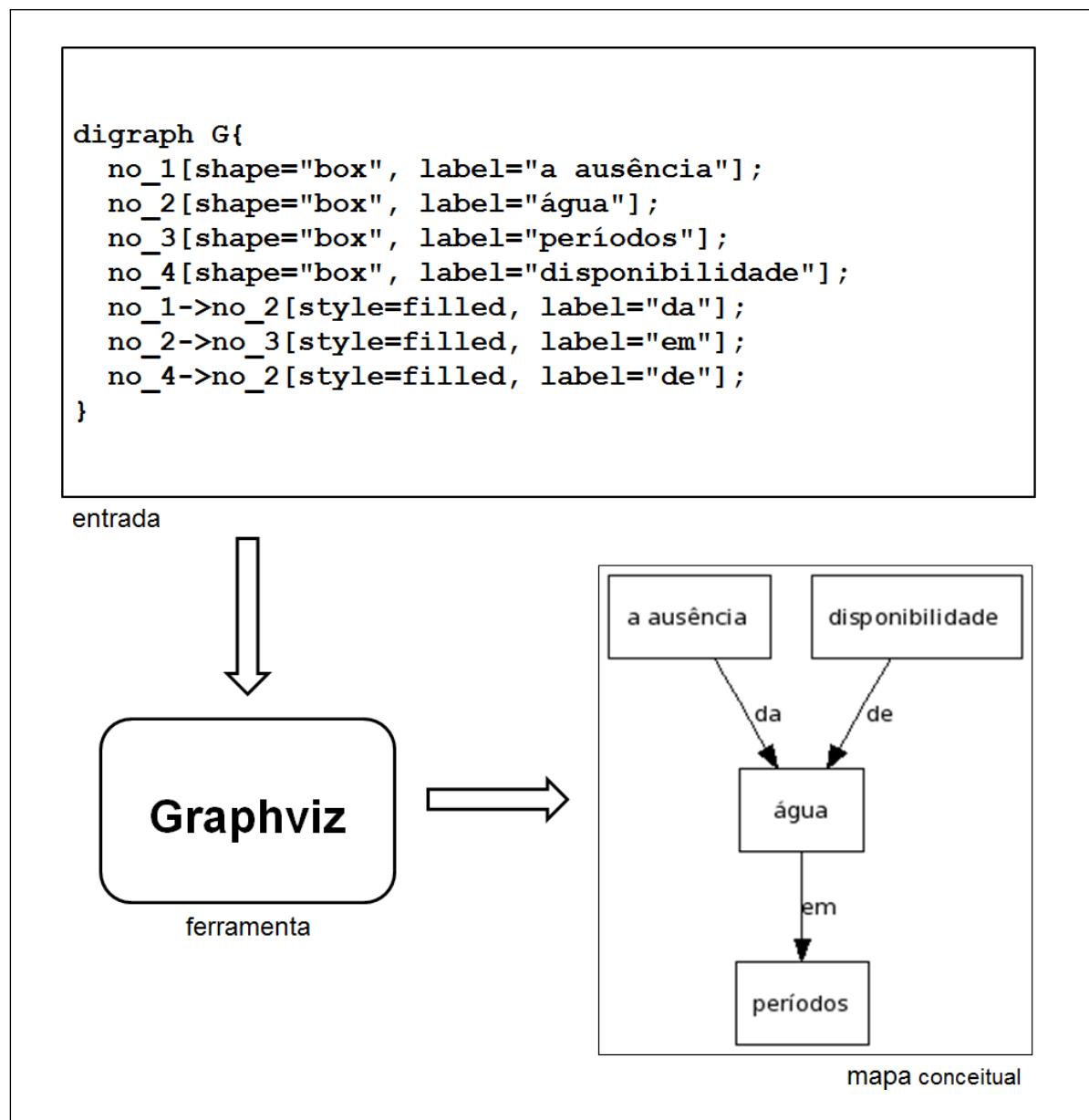
Utilizando as informações já classificadas em conceitos e relações, disponíveis no arquivo “conceitual.xml”, este componente gera uma lista sequencial dos conceitos, seguido de uma lista sequencial de informações.

As duas listas obtidas são geradas com uma sintaxe específica, para a utilização por uma ferramenta de visualização de grafos. A ferramenta utilizada neste projeto é o *Graphviz* (ELLSON et al., 2002). Essa ferramenta oferece uma saída na forma de um arquivo de imagem que pode ser visualizado diretamente no navegador internet. Foi realizada a opção pela utilização dessa ferramenta, pela simplicidade de uso, oferecimento de uma interface adequada ao objetivo do projeto e simples obtenção e instalação.

A Figura 22 apresenta a imagem fornecida por essa ferramenta a partir de um exemplo especificando um mapa conceitual selecionado de uma amostra aleatória obtida do fórum mencionado no Capítulo 3.

Na Figura 22, o quadro “entrada” ilustra a sintaxe utilizada pela ferramenta *Graphviz*. A primeira linha define o objeto gráfico. As quatro linhas seguintes descrevem os nós apresentados, que correspondem aos conceitos extraídos do arquivo “conceitual.xml”. Nas três linhas seguintes estão descritas as relações, extraídas do arquivo “conceitual.xml”. Finalmente, a última linha encerra a definição do objeto gráfico.

Quando da visualização de um mapa conceitual, através do FORUM3D, o Componente Gerador Mapa Conceitual realiza a extração da lista de conceitos e relações, por meio do Componente Leitor XML, gera uma saída de dados conforme a sintaxe descrita nesta seção e encaminha essa saída de dados para a ferramenta *Graphviz*. Por sua vez, essa ferramenta gera um arquivo de imagem que representa



**Figura 22 - Saída do programa Graphviz**  
**Fonte: Autor**

o mapa conceitual, conforme ilustrado na Figura 22 pelo quadro “mapa conceitual”.

#### 4.6 CONCLUSÕES

Para a verificação da aplicação do modelo de estruturação proposto, o MOREM, foi realizada a especificação de um modelo computacional que permitiu a construção do correspondente sistema computacional.

O objetivo atingido com o modelo computacional proposto foi a elaboração dos requisitos necessários fornecimento das especificações necessários para a etapa de construção desse sistema computacional. Além dessa especificação, o MOREM oferece uma base que pode ser utilizada na construção de ferramentas computacionais mais abrangentes.

A partir da concepção modular do modelo computacional obtido, espera-se que as naturais adaptações e expansões do sistema concebido, possam ser executadas de forma mais simples e funcional.

Como o objetivo do sistema gerado a partir do modelo computacional concebido foi realizar a verificação da viabilidade computacional da aplicação do MOREM, a otimização dos módulos utilizados não foi realizada. Mas, a partir da documentação apresentada pelo modelo computacional, está delineado o caminho para o aprimoramento do sistema para uma aplicação para utilização em ambientes de produção.

A definição de intercâmbio das informações entre os módulos do FORUM3D através da utilização de arquivo XML possibilita a utilização das informações estruturadas, obtidas de um fórum e processadas pelo FORUM3D, por aplicações externas de uma forma padrão e bem documentada.

## 5 ESTUDO DE CASO

Com o objetivo de validar a proposta de estruturação apresentada nesta pesquisa, foi desenvolvida uma aplicação *Web*, o FORUM3D, baseado no modelo computacional descrito no Capítulo 4.

A Seção 5.1 descreve as características de implementação utilizadas no desenvolvimento do FORUM3D. A base de dados utilizada é apresentada na Seção 5.2. A Seção 5.3 ilustra os arquivos XML utilizados pelo FORUM3D. Na Seção 5.4 é descrita a utilização do FORUM3D, apresentando a interface do usuário desenvolvida. Finalmente a Seção 5.5 relaciona algumas observações obtidas no desenvolvimento do estudo de caso.

### 5.1 CARACTERÍSTICAS DE IMPLEMENTAÇÃO

Baseado no modelo computacional especificado no Capítulo 4 foram definidas as características de implementação utilizadas no desenvolvimento do FORUM3D. Essas características são:

**Topologia:** Aplicação *Web*;

**Servidores:** Servidor único contendo servidor de aplicação e servidor de banco de dados;

**Sistema Operacional:** Linux CentOS;

**Banco de Dados:** MySQL ;

**Linguagem de Programação:** PHP;

**Interface Gráfica:** Padrão HTML.

## 5.2 BASE DE TESTES

Durante o desenvolvimento desta pesquisa foi utilizada uma base de testes disponibilizada pela UTFPR, que corresponde a uma cópia do ambiente AVA de produção. Os dados pessoais dos usuários foram substituídos por dados alterados, como por exemplo, “ALUNO 001”, “ALUNO 002” e “PROFESSOR 001” em sua origem, sendo utilizada uma base de dados que não continha identificação dos usuários reais.

Em termos quantitativos, a base de teste utilizada armazena :

- 55 cursos;
- 969 fóruns;
- 1.620 usuários;
- cerca de 8 milhões de palavras;
- na faixa de 400 mil palavras diferentes, sem considerar aplicação de técnicas de similaridade e limpeza;

## 5.3 ARQUIVOS XML

A partir das especificação do modelo de armazenamento de informações em arquivos XML, descrito no Capítulo 4, foi definida a sintaxe para geração dos arquivos XML pelos módulos do FORUM3D.

A Figura 23 ilustra o arquivo “discussao.xml” obtido a partir da extração dos dados de um dos fóruns utilizados durante o projeto. Com o objetivo de facilitar a visualização, os textos mais longos foram truncados, e as sequências de elementos de mesmo tipo foram omitidas. Esta condição pode ser observada nos elementos “post” e “usuario” em que somente os primeiros elementos de cada série estão apresentados, sendo os demais omitidos e substituídos pela utilização de reticências. Essa forma de representação dos elementos repetidos será utilizada nas demais figuras que representam arquivos XML nesta seção. A estrutura desse arquivo corresponde a especificação apresentada no Quadro 2 da página 43.



```

<discussao>
  <id>1402</id>
  <geracao>28/07/2012 17:48:45</geracao>
  <nome>
    Fórum III - Importância da água para ...
  </nome>
  <curso>
    <id>57</id>
    <nome>
      III Curso de Especialização em ...
    </nome>
  </curso>
  <forum>
    <id>847</id>
    <nome>Fórum III - Legislação</nome>
  </forum>
  <post>
    <id>24210</id>
    <datahora>1218743909</datahora>
    <usuario>25</usuario>
    <anterior>0</anterior>
    <texto>
      Este fórum está aberto para ...
    </texto>
  </post>
  <post>
    ...
  </post>
  <usuario>
    <id>25</id>
    <nome>Professor 025</nome>
  </usuario>
  <usuario>
    <id>705</id>
    <nome>Aluno 0705</nome>
  </usuario>
  <usuario>
    ...
  </usuario>
</discussao>

```

**Figura 23 - Arquivo discussao.xml obtido**  
**Fonte: Autor**

A Figura 24 ilustra um dos arquivos “post.xml” gerados durante o processo NLP realizado pelo Módulo Estruturador. Cada mensagem estruturada acarreta a geração de um arquivo desse tipo, recebendo a nomenclatura correspondente a palavra “post”, seguido do caractere “\_”, da representação de um inteiro correspondente ao identificador único da discussão e a extensão “.xml”. Portanto

```

<post>
  <id>24210</id>
  <geracao>28/07/2012 17:49:05</geracao>
  <datahora>1218743909</datahora>
  <anterior>0</anterior>
  <usuario>
    <id>25</id>
    <nome>Professor 025</nome>
  </usuario>
  <sintagmas>
    <s>
      <valor>Este fórum</valor>
      <tipo>NP</tipo>
      <sufixo>n</sufixo>
    </s>
    <s>
      <valor>está aberto</valor>
      <tipo>VP</tipo>
      <sufixo>v-pcp</sufixo>
    </s>
    <s>
      .....
    </s>
  </sintagmas>
</post>

```

**Figura 24 - Arquivo post.xml obtido**  
**Fonte: Autor**

o arquivo que armazena as informações estruturadas da mensagem utilizada na ilustração da Figura 24 é armazenado com o nome físico “post\_24210.xml”. A estrutura desse arquivo corresponde a especificação apresentada no Quadro 3 da página 51.

Os arquivos intermediários “post.xml” são mantidos após a realização do processo de estruturação, não sendo apagados. Essa condição foi utilizada

para evitar a realização do processo de estruturação novamente, quando de futuras execuções do FORUM3D em um mesmo fórum. Quando ocorrer a geração de novas mensagens em um fórum em que foi realizado um processo de estruturação anterior, somente é necessária a estruturação das novas mensagens. Foi comprovado nos experimentos realizados que esta técnica reduz muito o tempo necessário para o processamento futuro em fóruns contendo um grande número de mensagens.

A figura 25 ilustra o arquivo “estruturado.xml” obtido. A estrutura desse arquivo corresponde a especificação apresentada no Quadro 4.

A figura 26 ilustra o arquivo “conceitual.xml” gerado com o sistema. A estrutura desse arquivo corresponde a especificação apresentada no Quadro 6.

```

<discussao>
  <id>1402</id>
  <geracao>28/07/2012 17:48:45</geracao>
  <nome>
    Fórum III - Importância da água para ...
  </nome>
  <curso>
    <id>57</id>
    <nome>
      III Curso de Especialização em ...
    </nome>
  </curso>
  <forum>
    <id>847</id>
    <nome>Fórum III - Legislação </nome>
  </forum>
  <posts>
    <post>
      <id>24210</id>
      <datahora>1218743909</datahora>
      <usuario>25</usuario>
      <anterior>0</anterior>
      <sintagmas>
        <s>1</s>
        ...
      </sintagmas>
    </post>
    ....
  </posts>
  <usuarios>
    <usuario>
      <id>25</id>
      <nome>Aluno 705</nome>
    </usuario>
    ...
  </usuarios>
  <sintagmas>
    <s>
      <id>1</id>
      <valor>Este fórum</valor>
      <post>24210</post>
      <tipo>NP</tipo>
      <sufixo>n</sufixo>
    </s>
    ...
  </sintagmas>
</discussao>

```

**Figura 25 - Arquivo estruturado.xml obtido**  
**Fonte: Autor**

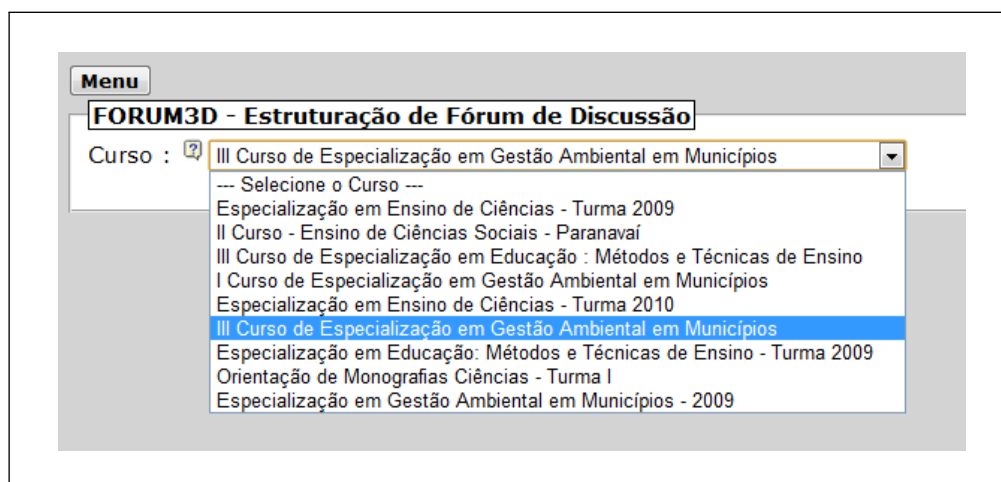


Figura 26 - Arquivo conceitual.xml obtido  
Fonte: Autor

#### 5.4 UTILIZAÇÃO DO SISTEMA

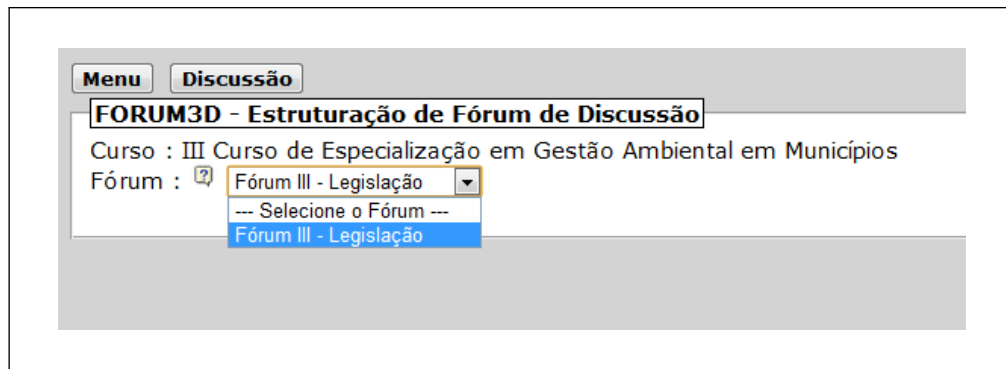
Para a utilização do sistema, o professor inicialmente seleciona a Discussão que deseja estruturar e analisar. Essa seleção é feita pela interface gráfica do Componente Seletor da Discussão, do Módulo Extrator do Forum3D, descrito na Seção 4.2.2.

A Figura 27 ilustra a interface gráfica de seleção de curso, realizada através da apresentação de um componente do tipo menu “*drop-down*”. Nessa interface é realizada a seleção do curso que contém o fórum em que será realizada a pesquisa. A ilustração da Figura 27 demonstra a seleção do curso “III Curso de Especialização em Gestão Ambiental em Municípios”. A partir dessa escolha o sistema realiza a apresentação da próxima tela da interface do usuário, que corresponde à seleção de fórum.



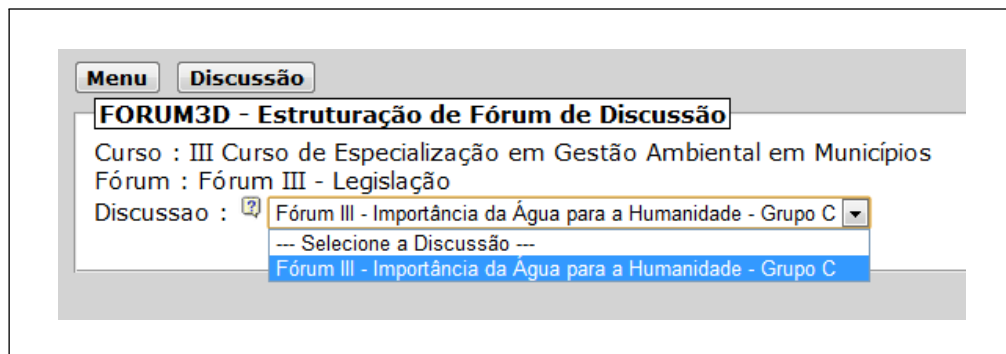
**Figura 27 - Tela de seleção de curso**  
**Fonte: Autor**

Na Figura 28 é visualizada a tela de seleção do fórum, na qual ocorre também a disponibilização de um componente menu “*drop-down*”. Esse componente disponibiliza para o professor uma lista contendo os fóruns registrados no curso selecionado. No exemplo apresentado na Figura 28 ocorre a apresentação de um fórum, que é o “Fórum III - Legislação”. A partir da seleção do fórum, pelo professor, ocorre a apresentação da última tela de seleção, que corresponde à seleção da discussão.



**Figura 28 - Tela de seleção de fórum**  
**Fonte: Autor**

A Figura 29 ilustra a interface gráfica de seleção da discussão, que oferece um componente menu “*drop-down*” para a escolha da discussão. Esse menu apresenta a relação das discussões que foram registradas no fórum selecionado. No exemplo ilustrado na Figura 29, ocorre a apresentação de somente uma discussão, que corresponde ao “Fórum III - Importância da Água para a Humanidade - Grupo C”. A partir da seleção de discussão é apresentada a próxima tela que oferece um painel para apresentação dos relatórios, ou visões.



**Figura 29 - Painel de relatórios**  
**Fonte: Autor**

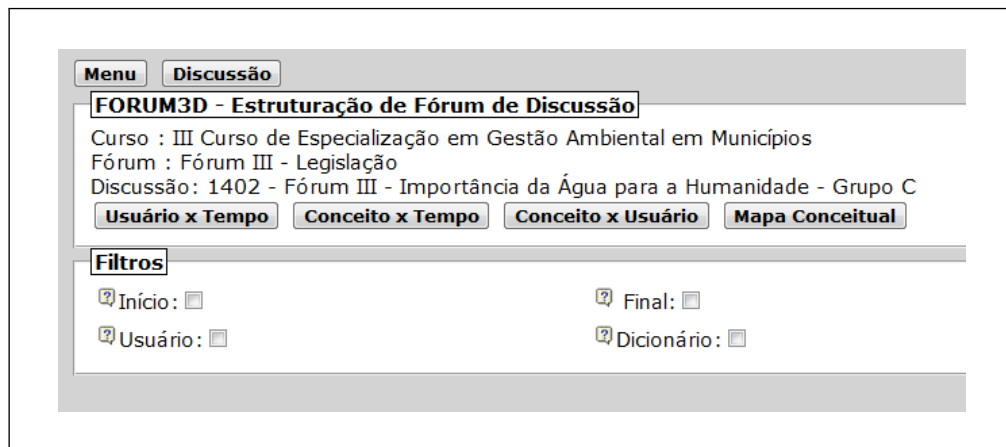
Na Figura 30 está ilustrada a interface que apresenta o painel de relatórios. Um componente “Discussão”, do tipo botão, é oferecido para possibilitar ao professor a seleção de uma outra discussão, sendo direcionado para a tela de seleção de curso. Por meio da interface do painel de relatórios, o professor pode selecionar qual visualização, dentre as definidas para o FORUM3D, será utilizada para apresentação dos dados estruturados da discussão selecionada. Diversas

visualizações consecutivas, de diversos tipos, podem ser realizadas, sem sair dessa interface.

As opções de relatório oferecidas, baseadas nas visões descritas na Seção 3.2, são :

- Diagrama de Usuário x Tempo;
- Diagrama de Conceito x Tempo;
- Diagrama de Conceito x Usuário.

Além desses diagramas, também está disponível a apresentação do mapa conceitual da discussão selecionada.



**Figura 30 - Tela de seleção de Visão**  
Fonte: Autor

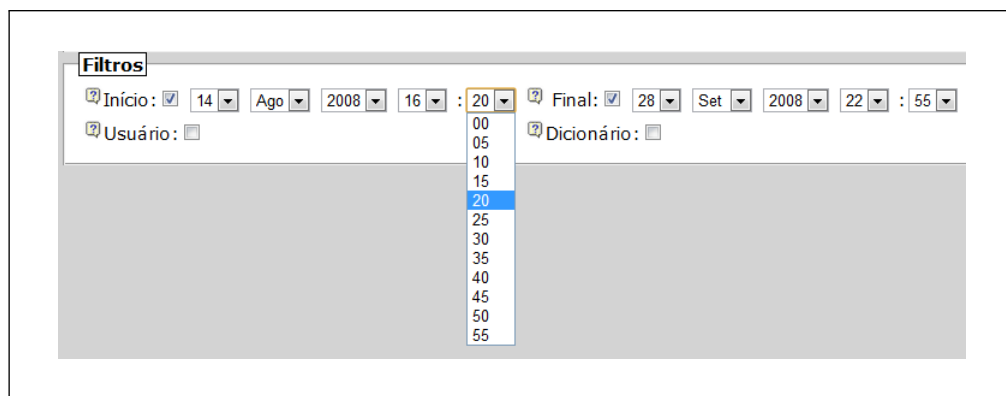
Além da seleção do relatório a ser gerado, por meio da interface visualizada na Figura 30, é realizada a seleção do escopo de pesquisa desejado, através da utilização dos filtros, conforme especificado na Seção 4.5.4. A seleção é realizada através dos componentes do tipo “*check-box*” ilustrados abaixo do rótulo “Filtros”, na Figura 30.

Com o recurso de filtro de tempo, o professor pode selecionar um conjunto de mensagens que foram registradas em um intervalo de tempo específico. A Figura 31 ilustra a utilização do filtro de tempo. A partir da seleção do “*check-box*” “Início”, o professor tem a sua disposição um conjunto de menu “*drop-down*” que permitem a seleção a partir de qual data e hora ocorrerá a filtragem das mensagens. Quando da



primeira ativação do recurso de filtro de tempo, após a seleção de uma discussão, os valores inicialmente apresentados por esse filtro correspondem ao horário de inserção da primeira mensagem na discussão.

Além do filtro “Início”, também é oferecida a opção de utilização do filtro “Final” que permite a seleção da data e horário na qual será encerrada a seleção de mensagens pesquisadas, desenvolvido de forma similar à do filtro “Início”. Quando da primeira seleção do recurso do filtro de tempo “Final”, os componentes do menu “*drop-down*” apresentam o horário correspondente à última mensagem registrada na discussão.

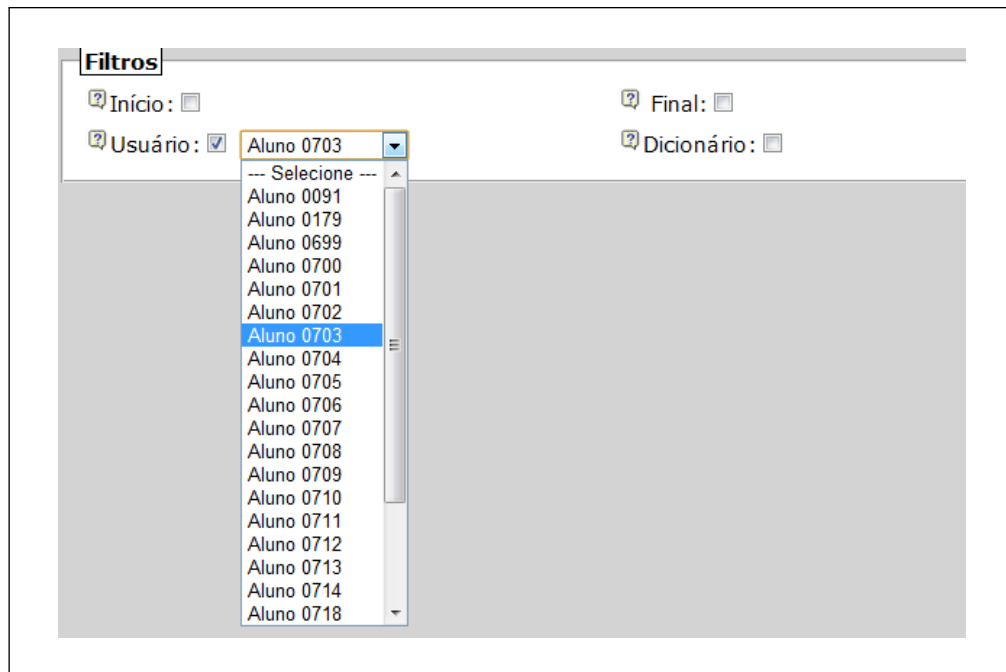


**Figura 31 - Tela de filtro de tempo**  
**Fonte: Autor**

A partir da utilização do filtro de usuário, o professor pode restringir o escopo de pesquisa para um usuário, realizando a avaliação das contribuições individuais. Também pode ocorrer a seleção de um conjunto de usuários para realizar a análise focada em um grupo específico. Caso não ocorra a ativação desse filtro, as mensagens correspondentes a todos os usuários serão apresentadas nos relatórios solicitados.

A Figura 32 ilustra a utilização desse filtro. A partir da ativação do “*check-box*” “Usuário”, é realizada a apresentação de uma interface de seleção múltipla, do tipo menu “*drop-down*”, que apresenta a relação de todos os usuários, em ordem alfabética, que realizaram contribuições na discussão selecionada. O professor pode então selecionar um ou mais usuários para os quais queira restringir o universo de pesquisa.

O terceiro filtro disponibilizado corresponde à seleção da dimensão de

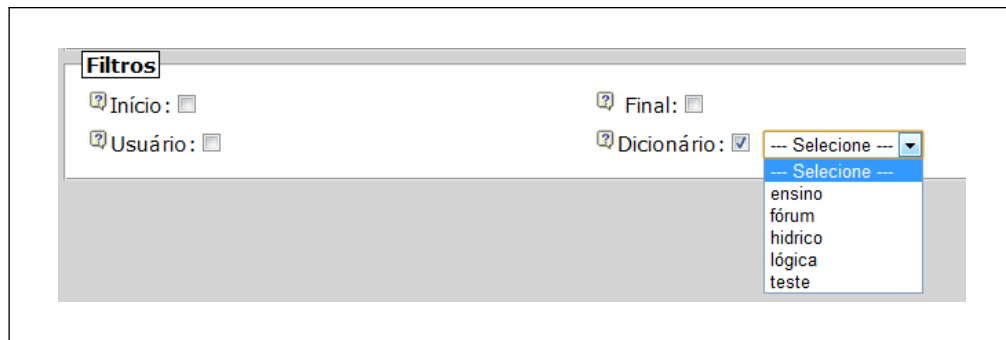


**Figura 32 - Tela de filtro de usuário**  
**Fonte: Autor**

conhecimento. Com esse filtro, a partir de um dicionário que é composto por um conjunto de palavras, somente os conceitos que estiverem contidos no universo de conceitos do dicionário escolhido serão utilizados na apresentação dos relatórios. A partir da ativação do “*check-box*” “Dicionário”, conforme ilustrado na Figura 33, ocorre a apresentação de um menu “*drop-down*” que contém a lista, em ordem alfabética, dos dicionários disponíveis para utilização. A partir da seleção de um dicionário, somente os sintagmas que contém conceitos similares às palavras registradas nesse dicionário, é que serão utilizados na geração dos relatórios. Na apresentação do mapa conceitual, as relações e conceitos vizinhos aos conceitos selecionados pelos filtros, também são apresentados para oferecer uma maior clareza a esse diagrama.

A aplicação desenvolvida permite o gerenciamento dos dicionários através de uma interface específica, ilustrada na Figura 34. Essa interface permite a criação de um novo dicionário a exclusão de um dicionário existente, bem como a inserção e remoção de palavras dentro de um dicionário específico.

A partir da seleção de um ou mais filtros, ou ainda, mesmo sem selecionar algum filtro, o professor pode selecionar um dos diagramas que combinam duas das três



**Figura 33 - Tela de filtro de conhecimento**  
**Fonte: Autor**

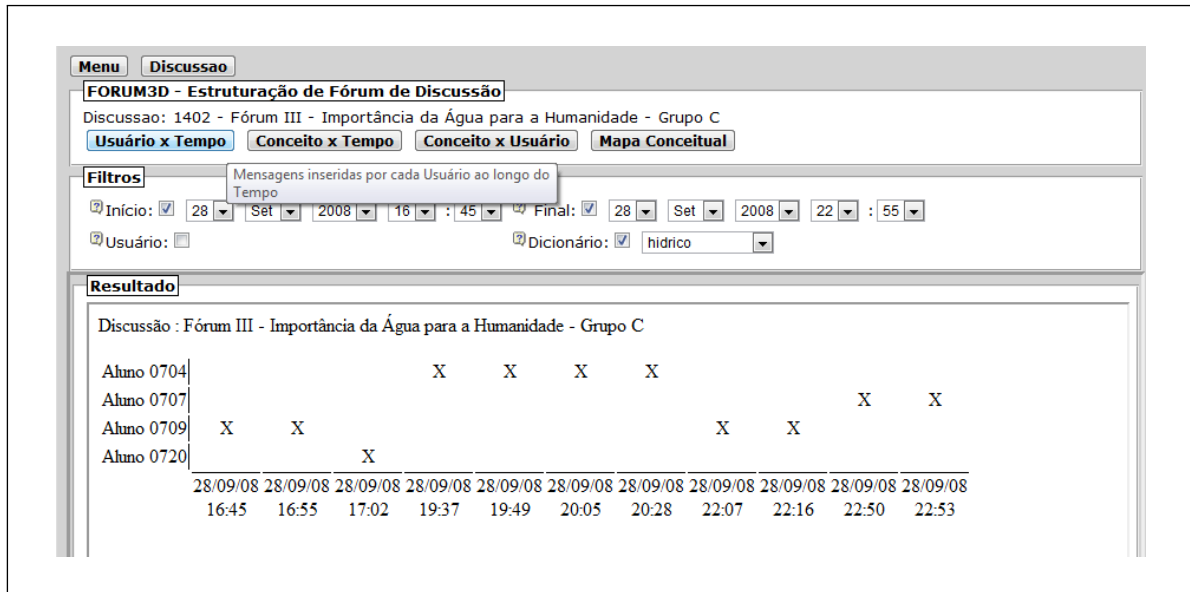


**Figura 34 - Tela do gerenciador de dicionários**  
**Fonte: Autor**

dimensões definidas no MOREM, para realizar suas consultas sobre a dinâmica apresentada no fórum.

Na Figura 35 é apresentada a geração do diagrama correspondente à Visão Usuário x Tempo, aplicado em um período selecionado no filtro de Tempo, bem como a utilização de um dicionário de conceitos, denominado “hídrico”. A partir dessa visão o professor pode verificar a dinâmica de inserção de mensagens dos usuários, conforme ilustrado no exemplo apresentado na Figura 35. Nesse gráfico é possível ao professor verificar que os usuários “Aluno 0704” e “Aluno 0709” tiveram uma frequência de inserção de mensagens mais significativa no fórum, dentro dos

critérios de filtragem de pesquisa, selecionados.



**Figura 35 - Relatório de Usuário x Tempo**  
Fonte: Autor

A Figura 36 apresenta a Visão Conceito x Tempo com a aplicação dos filtros de tempo e conceitos. Com essa visão o professor pode obter uma visualização da dinâmica da evolução dos conceitos ao longo do tempo. Pode-se observar nessa visualização, a concentração da discussão em torno do conceito “água”, ocorrida em torno das 19h37 do dia 28/09/2008.

Um exemplo da Visão Conceito x Usuário utilizando o dicionário “hídrico” é visualizada na Figura 37. A partir dessa visão o professor pode verificar qual aluno ou grupo realizou a inserção de determinados conceitos na discussão. Na implementação realizada, a quantidade de conceitos individualmente inserida é registrada no diagrama. É observado nesse exemplo, a concentração de contribuições em torno do conceito “água” ocorrendo com maior intensidade pelos usuários “Aluno 0703”, “Aluno 0704”, “Aluno 0708” e “Aluno 0711”.

Menu **Discussao**

**FORUM3D - Estruturação de Fórum de Discussão**

Discussao: 1402 - Fórum III - Importância da Água para a Humanidade - Grupo C

**Usuário x Tempo** **Conceito x Tempo** **Conceito x Usuário** **Mapa Conceitual**

**Filtros** Conceitos registrados ao longo do Tempo

Início:  28 Set 2008 16:45  Final:  28 Set 2008 22:55

Usuário:  Dicionário:  hidrico

**Resultado**

Discussão : Fórum III - Importância da Água para a Humanidade - Grupo C

	28/09/08 16:45	28/09/08 16:55	28/09/08 17:02	28/09/08 19:37	28/09/08 19:49	28/09/08 20:05	28/09/08 20:28	28/09/08 22:07	28/09/08 22:16	28/09/08 22:50	28/09/08 22:53
água	1	2	3	4	2	1	3	1			3
carros				1							
conscientize											
calçadas deve-se				1							
multar					1						
multa		1									

**Figura 36 - Relatório de Conceito x Tempo**  
Fonte: Autor

Menu **Discussao**

**FORUM3D - Estruturação de Fórum de Discussão**

Discussao: 1402 - Fórum III - Importância da Água para a Humanidade - Grupo C

**Usuário x Tempo** **Conceito x Tempo** **Conceito x Usuário** **Mapa Conceitual**

**Filtros** Conceitos registrados por cada Usuário

Início:  Final:

Usuário:  Dicionário:  hidrico

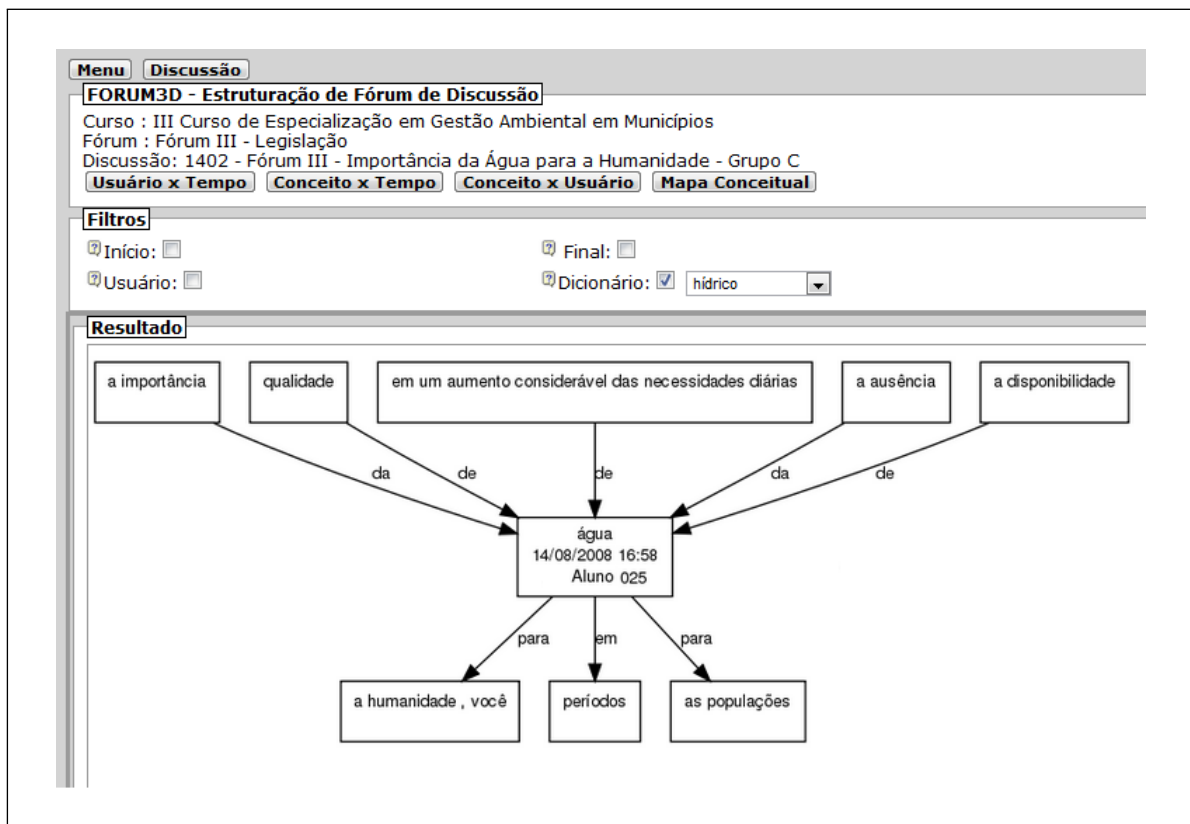
**Resultado**

Discussão : Fórum III - Importância da Água para a Humanidade - Grupo C

	Aluno 0091	Aluno 0179	Aluno 0699	Aluno 0700	Aluno 0701	Aluno 0702	Aluno 0703	Aluno 0704	Aluno 0705	Aluno 0706	Aluno 0707	Aluno 0708	Aluno 0709	Aluno 0710	Aluno 0711	Aluno 0712	Aluno 0713	Aluno 0714	Aluno 0718	Aluno 0719	Aluno 0720	Aluno 0721
água	1		2	4	4	2	11	10		5	3	9	5	4	10	3	1	2	4	4	3	7
carros								1										1				
conscientize		1																				
calçadas deve-se								1														
multar								1										1				
multa													1						1			

**Figura 37 - Relatório de Conceito x Usuário**  
Fonte: Autor

A Figura 38 ilustra a apresentação do mapa conceitual obtido com o FORUM3D. A partir da obtenção dos Mapas Conceituais do fórum, é oferecida uma apresentação dos conceitos inseridos, e suas relações, de uma forma gráfica e de fácil interpretação. O mapa conceitual ilustrado na Figura 38, corresponde à representação do conhecimento em torno do conceito “água”, sendo registrado que a primeira contribuição com este conceito ocorreu pelo “Aluno 025” no dia 14/08/2008 às 16h58. A visualização da Figura 38 apresenta somente um trecho escolhido a partir do mapa conceitual gerado pelo FORUM3D.



**Figura 38 - Geração de Mapa Conceitual**  
Fonte: Autor

## 5.5 OBSERVAÇÕES

A especificação e utilização dos filtros para realização da qualificação e apresentação dos conceitos estruturados do fórum foi de fundamental importância para a verificação da aplicabilidade do modelo proposto pela pesquisa. Quando da geração das visões e dos mapas conceituais sobre os fóruns utilizados no

experimento, sem a aplicação de filtros, ocorreu a geração de diagramas com uma densidade muito elevada de informações o que acarretava uma difícil interpretação. A aplicação dos filtros combinados permitiu a obtenção de diagramas e mapas conceituais claros e bem apresentáveis.

A utilização do filtro de conceitos, com a definição dos arquivos de dicionário, permitiu isolar inúmeros conceitos nos processos de análise e apresentação dos resultados, resultando em relatórios mais concisos e apresentáveis.

## 6 CONCLUSÃO

Os Ambientes Virtuais de Aprendizagem continuam sendo utilizados como ferramentas de apoio ao ensino presencial. Nesse contexto, esses ambientes possibilitam que professores ou instrutores disponibilizem aos seus alunos alguns materiais instrucionais e também definam momentos de discussão e interação entre os alunos. Esses momentos de discussão e interação ocorrem na ferramenta Fórum do Ambiente Virtual de Aprendizagem.

O Fórum é um ambiente no qual os alunos podem expressar os seus conhecimentos por meio da criação de textos que são acessíveis a todos os participantes do curso. Logo, o fórum pode ser considerado como um repositório dos conhecimentos construídos pelos alunos. Apesar disso, a sua característica de representação textual não estruturada representa uma dificuldade significativa para o processo avaliativo do professor ou instrutor.

Este projeto de pesquisa apresenta uma solução para a questão da falta de estrutura das mensagens armazenadas no Fórum. Com essa solução, o professor possui um recurso que antes não havia no ambiente virtual. Além dele poder inspecionar as mensagens armazenadas no fórum, ele conta agora com um novo recurso associado ao fórum. Um recurso que tem a função de apresentar uma estrutura de resumo ao professor.

Com esse novo recurso, as mensagens do fórum são analisadas, processadas, estruturadas e disponibilizadas ao professor. Nessa estrutura, os conceitos do domínio são extraídos e destacados do corpo de mensagens do fórum. Essa estrutura também armazena o autor de cada mensagem e quando ela foi apresentada. Logo, o professor pode identificar quais elementos do domínio foram tratados nas mensagens, quando eles foram tratados, quais alunos participaram desse processo, quando isso aconteceu e quais alunos não fizeram contribuição alguma.



Vislumbra-se, assim, que um hipotético professor ao propor um fórum de discussão a respeito do tema “Violência e Segurança na Escola”, por exemplo, possa rapidamente inspecionar se alguns conceitos que ele julga serem importantes foram tratados pelos alunos. Nesse exemplo hipotético, o processo de estruturação do Fórum3D poderia identificar quais alunos apresentaram os seguintes conceitos, por exemplo: “pessoas que fumam maconha”, “alunos que entram armados no colégio”, “falta de um policiamento preventivo”, entre outros. Além disso, o Fórum3D permite apresentar quando esses conceitos foram tratados, por quem e em qual frequência.

Espera-se, futuramente, que essa ferramenta possa vir a ser utilizada pelo professor como um mecanismo de auxílio ao processo avaliativo. Diversos outros trabalhos da literatura fazem uma análise quantitativa das informações do fórum. Informações tais como quantidade de mensagens, números de acessos, tempo de acessos, entre outras informações qualitativas tem sido alvo de estudos. Com este trabalho, os Ambientes Virtuais como por exemplo o Moodle, podem contar com uma ferramenta que permite a realização a posteriori de uma análise quantitativa das mensagens do Fórum pelo professor.

Apesar dos esforços realizados por esta pesquisa, alguns aspectos ainda precisam ser melhorados no ambiente. Desses aspectos, destaca-se:

- Melhorar o processo de Geração de Mapas Conceituais. O processo de geração de Mapas Conceituais utilizado neste trabalho não leva em consideração a opinião do professor. Ou seja, se um determinado professor considera que determinados elementos do domínio são mais importantes que outros, isso não se reflete na construção dos Mapas Conceituais;
- A possibilidade de realizar uma análise parcial das mensagens do fórum. No ambiente desenvolvido, todas as mensagens são analisadas. O professor não tem ao seu dispor recursos que lhe permitam especificar algumas condições de análise. Por exemplo, se um professor desejar analisar apenas as mensagens que possuem elementos comuns apresentados por três ou mais alunos. Como exemplo de outra situação, em um determinado fórum, um aluno fornece uma mensagem sempre que um outro aluno fornece uma mensagem. Esse tipo de recurso de reconhecimento de padrão ainda não está disponível na ferramenta.

Como trabalho futuro, destaca-se ainda a realização de estudos que permitam, a partir dos dados estruturados pelo Fórum3D, a realização de inferências e prognósticos a respeito dos alunos e também do desenvolvimento de novas formas de visualização dessas informações.

Vislumbra-se também como trabalho futuro, disponibilizar o Fórum3D para usuários do ambiente Moodle para a sua avaliação. Nesse trabalho futuro, buscar-se-á identificar quais as necessidades de informações que os professores poderiam necessitar e que o Fórum3D ainda não apresenta.

## REFERÊNCIAS

AUSUBEL, David Paul; NOVAK, Joseph D.; HANESIAN, Helen. **Psicología educativa : un punto de vista cognoscitivo**. México: Trillas, 1983.

AZEVEDO, Breno Fabrício Terra. **Mineratórum: Um recurso de apoio para análise qualitativa em fóruns de discussão**. 2011. 204 f. Dissertação (Doutorado) — Programa de Pós-graduação em Informática na Educação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

BALDRIDGE, Jason. **The opennlp project**. 2005.

BASSANI, Patrícia B. Scherer. **Trocas interindividuais no fórum de discussão: um estudo sobre as comunidades de aprendizagem em espaços de educação à distância**. 2009. XX Simpósio Brasileiro de Informática na Educação - Florianópolis - SC.

BRITO, Josilene Almeida. **Engajamento em atividades assíncronas na modalidade de ensino a distância: requisitos de interfaces colaborativas**. Tese (Doutorado) — Dissertação de mestrado em computação. Universidade Federal de Pernambuco (UFPE). Cin. Ciência da Computação (UFPE), 2010.

CHOWDHURY, Gobinda G. Natural language processing. **Annual review of information science and technology**, Wiley Online Library, v. 37, n. 1, p. 51–89, 2003.

DALMOLIN, Luiz Claudio Duarte. **Extrator de termos para criação de mapas conceituais**. 2010. 121 f. Dissertação (Mestrado) — Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2010.

DANIIL, Constantin; DASCALU, Mihai; TRAUSAN-MATU, Stefan. Automatic forum analysis: a thorough method of assessing the importance of posts, discussion threads and of users' involvement. In: **Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics**. New York, NY, USA: ACM, 2012. (WIMS '12), p. 37:1–37:9.

DIAS JUNIOR, Luiz Dourado; FERREIRA, Benedito de Jesus Pinheiro. **Avaliação Docente Dialógica: requisitos para uma ferramenta integrada de apoio à avaliação e reflexão sobre a prática docente**. 2008.

ELLSON, John; GANSNER, Emden; KOUTSOFIOS, Lefteris; NORTH, Stephen C.; WOODHULL, Gordon. Graphviz—open source graph drawing tools. In: SPRINGER. **Graph Drawing**. [S.l.], 2002. p. 594–597.

ERLIN; YUSOF, Norazah; RAHMAN, Azizah Abdul. Messages segmentation of asynchronous text-based discussion in social network analysis. In: **Education Technology and Computer (ICETC), 2010 2nd International Conference on**. [S.l.: s.n.], 2010. v. 3, p. V3–8 –V3–12.

FERNANDES, Eraldo; SANTOS, Cícero dos; MILIDIU, Ruy. A machine learning approach to portuguese clause identification. In: PARDO, Thiago; BRANCO, Antonio; KLAUTAU, Aldebaro; VIEIRA, Renata; LIMA, Vera de (Ed.). **Computational Processing of the Portuguese Language**. [S.l.]: Springer Berlin / Heidelberg, 2010, (Lecture Notes in Computer Science, v. 6001). p. 55–64.

GEROSA, Marco Aurélio; FUKS, Hugo; LUCENA, Carlos José Pereira de. Estruturação e categorização de mensagens em ferramentas de comunicação textuais assíncronas. In: **Electronic Proceedings of the World Congress on Engineering and Technology Education-WCETE**. [S.l.: s.n.], 2004. p. 14–17.

GÓMEZ-PÉREZ, Asunción; MANZANO-MACHO, David. An overview of methods and tools for ontology learning from texts. **The Knowledge Engineering Review**, v. 19, n. 03, p. 187–212, 2004.

GONÇALVES, Marcus Vinicius Ferreira; ELIA, Marcos da Fonseca. Arquitetura de avaliação educacional em fórum de discussão. In: **Anais do Workshop de Informática na Escola**. [S.l.: s.n.], 2008. v. 1, n. 1.

HEO, Gyeong Mi; BREULEUX, Alain. Roles of initiators and interaction patterns: exploring an informal online community at the interpersonal plane. In: **Proceedings of the 9th international conference on Computer supported collaborative learning - Volume 1**. [S.l.]: International Society of the Learning Sciences, 2009. (CSCL'09), p. 13–17.

HUANG, Jizhou; ZHOU, Ming; YANG, Dan. Extracting chatbot knowledge from online discussion forums. In: **Proceedings of IJCAI**. [S.l.: s.n.], 2007. p. 423–428.

KNUTH, Donald E. **The art of computer programming: Generating all trees: history of combinatorial generation**. [S.l.]: addison-Wesley, 2006.

KOWATA, Juliana Hiroko. **Uma abordagem computacional para construção de mapas conceituais a partir de textos em língua portuguesa do Brasil**. 2010. 104 f. Dissertação (Mestrado) — Programa de Pós-graduação em Informática, Universidade Federal do Espírito Santo, Vitória, 2010.

LANDIS, Melodee. A comparison of interaction in av-based and internet-based distance courses. **Educational Technology & Society**, v. 4, n. 2, p. 120–132, 2001.

LAU, Raymond Y. K.; SONG, Dawei; LI, Yuefeng; CHEUNG, Terence C. H.; HAO, Jin-Xing. Toward a fuzzy domain ontology extraction method for adaptive e-learning. **IEEE Trans. on Knowl. and Data Eng.**, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 21, p. 800–813, June 2009.

LIDDY, Elizabeth D. **Natural Language Processing for Information Retrieval**. [S.l.: s.n.], 2001. 3864-3873 p.

MOREIRA, M.A. **A teoria da aprendizagem significativa e sua implementação em sala de aula**. [S.l.]: Editora Universidade de Brasília, 2006.

NG, Hwee Tou; ZELLE, John. Corpus-based approaches to semantic interpretation in nlp. **AI magazine**, v. 18, n. 4, p. 45, 1997.

NOVAK, Joseph Donald; GOWIN, D. B. **Learning how to learn**. Cambridge, MA, USA: Cambridge University Press, 1984.

OLIVEIRA, A.; PEREIRA, F.C.; CARDOSO, A. Automatic reading and learning from text. In: **Proceedings of the International Symposium on Artificial Intelligence (ISAI)**. [S.l.: s.n.], 2001.

OLIVEIRA, Henry Poncio Cruz de. Mapa conceitual do movimento retilíneo. set 2011. Disponível em: <<http://www.fisica.ufpb.br/romero/objetosaprendizagem/Rived/01Cinematica/materiais/mapa.html>>.

PÉREZ, Cláudia Camerini Corrêa; VIEIRA, Renata. Mapas conceituais: geração e avaliação. In: **Anais do XXV Congresso da SBC - Sociedade Brasileira de Computação**. [S.l.: s.n.], 2005. p. 2158 – 2167.

RAJARAMAN, Kanagasabai; TAN, Ah-Hwee. Knowledge discovery from texts: a concept frame graph approach. In: **Proceedings of the eleventh international conference on Information and knowledge management**. New York, NY, USA: ACM, 2002. (CIKM '02), p. 669–671.

RITCHIE, Dennis M.; THOMPSON, Ken. The unix time-sharing system. **Communications of the ACM**, ACM, New York, NY, USA, v. 17, n. 7, p. 365–375, jul. 1974.

SÁNCHEZ, Lourdes Pérez. El foro virtual como espacio educativo: propuestas didácticas para su uso. **Verista Quaderns Digitals Net**, n. 40, p. 1–18, 2005.

SCHRIRE, Sarah. Knowledge building in asynchronous discussion groups: going beyond quantitative analysis. **Computers & Education**, Elsevier Science Ltd., Oxford, UK, UK, v. 46, p. 49–70, jan 2006.

SHAUL, M. Assessing online discussion forum participation. **International Journal of Information and Communication Technology Education**, IGI Publishing, 701 E. Chocolate Ave, Suite 200, Hershey, PA, 17033-1240, USA., v. 3, n. 3, p. 39–46, 2007.

SOUZA, Nadia Aparecida de; BORUCHOVITCH, Evely. Mapas conceituais: estratégia de ensino/aprendizagem e ferramenta avaliativa. **Educação em Revista**, SciELO Brasil, v. 26, p. 195 – 217, 12 2010.

TAVARES, Romero. Construindo mapas conceituais. **Ciências Cognição**, v. 12, p. 72–85, 2007.

WANAS, Nayer; EL-SABAN, Motaz; ASHOUR, Heba; AMMAR, Waleed. Automatic scoring of online discussion posts. In: **Proceeding of the 2nd ACM Workshop on Information Credibility on the Web**. New York, NY, USA: ACM, 2008. (WICOW '08), p. 19–26.

## GLOSSÁRIO

**Conceito:** Representação de um objeto pelo pensamento, relativamente a apresentação de suas características gerais, sendo expressado por palavras que os descrevem e particularizam (NOVAK; GOWIN, 1984).

**Discussão:** Um conjunto finito de mensagens relacionadas a um tema específico, que representa o registro das informações geradas pelos usuários na atividade de fórum de discussão em um ambiente virtual de aprendizagem.

**Fórum:** O Fórum de Discussão Online é uma atividade composta por um conjunto finito de discussões.

**Mapa Conceitual:** Diagrama que indica graficamente um conjunto de conceitos e suas relações.

**Mensagem:** Um conjunto finito de sentenças inseridas no Fórum em um determinado tempo por um usuário.

**Proposição:** Uma proposição é uma unidade semântica formado por dois ou mais conceitos, ligados por relações (MOREIRA, 2006).

**Relação:** Elemento semântico que realiza a ligação entre dos conceitos.

**Sentença:** Conjunto finito de elementos (gramaticais) pertencentes ao domínio escrito na língua portuguesa.

**Sintagma:** Unidade sintática, composta por um ou mais vocábulos que forma orações.

**Tempo:** Representação numérica que expressa uma relação de sequencialidade entre as mensagens do fórum. Pode ser expressa pelo uso de um rótulo de tempo representativo da data e hora (*timestamp*) da inserção do diálogo ou uma numeração sequencial.

**Usuário:** Indivíduo que está inserindo algum diálogo no fórum. Pode ser tanto um professor quanto um aluno.

**Vocábulo:** Elemento gramatical unitário, formado por palavras, números ou sinais de pontuação. Um conjunto de mais de uma palavra pode formar um elemento gramatical, como em um nome próprio por exemplo.