

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
CURSO DE CIÊNCIA DA COMPUTAÇÃO

JÚLIA KIKUYE KAMBARA DA SILVA

**Automatização do Processo de Identificação  
de Presença Social em Fóruns e Chats**

Trabalho de Graduação.

Prof. Dr. Leandro Krug Wives  
Orientador

Porto Alegre, dezembro de 2011.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitora de Graduação: Profa. Valquiria Link Bassani

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do CIC: Prof. Raul Fernando Weber

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

## **AGRADECIMENTOS**

Agradeço à minha mãe, por ter me apoiado, por nunca me deixar desistir e ter estado sempre ao meu lado.

Agradeço ao restante da minha família e aos amigos queridos, pelo carinho, compreensão, amizade e companheirismo, por estarem sempre dispostos a me ouvir, ajudar e sorrir.

Agradeço ao meu namorado pelo apoio constante, que me acompanhou desde o começo do curso de graduação.

Por fim, agradeço ao meu orientador, Prof. Leandro Krug Wives, pelo apoio, orientação e compreensão, com que me guiou no desenvolvimento deste trabalho.

## SUMÁRIO

<b>LISTA DE ABREVIATURAS E SIGLAS .....</b>	<b>6</b>
<b>LISTA DE FIGURAS .....</b>	<b>7</b>
<b>LISTA DE TABELAS .....</b>	<b>8</b>
<b>RESUMO .....</b>	<b>9</b>
<b>ABSTRACT .....</b>	<b>10</b>
<b>1 INTRODUÇÃO .....</b>	<b>11</b>
1.1 Objetivos .....	11
1.2 Contribuição.....	12
1.3 Estrutura do Texto .....	12
<b>2 CONTEXTO .....</b>	<b>13</b>
2.1 Modelagem do Problema.....	13
2.2 Lexicometria.....	14
2.3 Softwares que utilizam lexicometria para análise de conteúdo .....	15
2.3.1 WordSmith.....	15
2.3.2 Eureka .....	15
<b>3 DESENVOLVIMENTO DA FERRAMENTA.....</b>	<b>16</b>
3.1 Conversor HTML dos fóruns/chats para XML .....	17
3.2 Construtor de Categorias .....	21
3.3 Analisador.....	23
<b>4 ALGORITMOS UTILIZADOS PARA ANÁLISE DAS CATEGORIAS CADASTRADAS.....</b>	<b>27</b>
4.1 Expressão Simples .....	29
4.2 Expressão Simples no Início da Frase .....	30
4.3 Expressão Simples no Meio da Frase .....	30
4.4 Expressão Simples no Final da Frase .....	30
4.5 Expressão Simples Início da Postagem .....	30
4.6 Expressão Simples Final da Postagem .....	31
4.7 Expressão Regular.....	31
4.8 Uso do Recurso Responder .....	31
4.9 Letras em Maiúsculo .....	31
4.10 Nomes Próprios.....	31
4.11 Nomes Genéricos .....	32
4.12 Pronome Oculto .....	32
4.13 Emoticons , Gifs e Imagens.....	32
4.14 Utilização do Recurso Gráfico Cor.....	33
4.15 Utilização do Recurso Gráfico Negrito .....	33
4.16 Utilização do Recurso Gráfico Itálico .....	33
4.17 Utilização do Recurso Gráfico Alteração do Tamanho da Fonte .....	33

<b>4.18 Utilização do Recurso Gráfico Aspas .....</b>	<b>34</b>
<b>4.19 Utilização de link .....</b>	<b>34</b>
<b>5 EXPERIMENTO .....</b>	<b>35</b>
<b>5.1 Experimento com o Fórum A .....</b>	<b>35</b>
5.1.1 Experimento 1 : Fórum A com pistas textuais específicas .....	35
5.1.2 Experimento 2 : Fórum A com pistas textuais genéricas .....	36
<b>5.2 Experimento com o Fórum B .....</b>	<b>37</b>
5.2.1 Experimento 1 : Fórum B com pistas textuais específicas do fórum A .....	38
5.2.2 Experimento 2 : Fórum B com pistas textuais genéricas .....	39
<b>5.3 Análise dos Resultados .....</b>	<b>39</b>
<b>6 CONCLUSÕES E TRABALHOS FUTUROS.....</b>	<b>41</b>
<b>REFERÊNCIAS .....</b>	<b>43</b>
<b>APÊNDICE A .....</b>	<b>45</b>
<b>APÊNDICE B .....</b>	<b>47</b>
<b>APÊNDICE C .....</b>	<b>49</b>
<b>APÊNDICE D .....</b>	<b>51</b>

## **LISTA DE ABREVIATURAS E SIGLAS**

AVEA	Ambiente Virtual de Ensino e Aprendizagem
CMS	Comunicação Mediada por Computador
HTML	HyperText Markup Language
PS	Presença Social
UFRGS	Universidade Federal do Rio Grande do Sul
XML	Extensible Markup Language

## LISTA DE FIGURAS

Figura 3.1: Esquema das ferramentas construídas .....	17
Figura 3.2: Formato padrão para arquivo XML com as postagens.....	18
Figura 3.3: Trecho de fórum extraído do Moodle .....	19
Figura 3.4: Arquivo XML gerado a partir do trecho extraído do Moodle.....	19
Figura 3.5 : Trecho de chat extraído de arquivo HTML.....	20
Figura 3.6: Construtor de categorias e pistas textuais.....	21
Figura 3.7: Formato padrão para o arquivo XML com as classes e pistas textuais .....	22
Figura 3.8 : Arquivo XML de classes com algumas classes e pistas textuais .....	22
Figura 3.9 : Interface do Analisador .....	23
Figura 3.10: Formato padrão para arquivo XML com o resultado da análise.....	23
Figura 3.11: Exemplo de arquivo XML com o resultado de uma análise .....	24
Figura 3.12 : Arquivo TXT resultante contendo informações sobre os alunos analisados .....	25
Figura 3.13: Visualização do número total de ocorrências de cada pista textual.....	25
Figura 3.14: Diagrama de Classes.....	26

## LISTA DE TABELAS

Tabela 2-1: Modelo de Presença Social proposto por Bastos .....	13
Tabela 3-1: Notação adotada no arquivo XML com as postagens.....	20
Tabela 4-1: Categorias que se deseja usar para análise .....	27
Tabela 4-2: Legenda dos algoritmos utilizados.....	28
Tabela 5-1: Fórum A - Análises automáticas com pistas textuais específicas .....	35
Tabela 5-2 : Fórum A - Análise automática com pistas textuais genéricas.....	37
Tabela 5-3: Fórum B - Comparação entre a Análise Manual e a Análise Manual Corrigida.....	37
Tabela 5-4: Fórum B - Análise automática com pistas textuais específica do fórum A.....	38
Tabela 5-5: Fórum B - Análise automática com pistas textuais genéricas .....	39
Tabela 5-6: Todos os resultados obtidos com os experimentos .....	40



## RESUMO

Seguindo o modelo proposto por Hélvia Pereira Pinto Bastos para identificação do nível de Presença Social de alunos em um Ambiente Virtual de Ensino e Aprendizagem (AVEA), desenvolveu-se três ferramentas.

O modelo é formado por categorias e subcategorias, e ao final de cada subcategoria existe pistas textuais, expressões lexicais, que são usadas para identificação das subcategorias num *corpus*, tratados aqui apenas para postagens em fóruns e chats.

A primeira ferramenta junta todos os arquivos com as postagens num só, neste processo é feito a padronização das *tags* de interesse e retira das demais. O segundo software permite criar classes, subclasses e ao final de cada subclasse cadastrar suas pistas textuais, desde modo, projeta-se o modelo de Bastos (2011) num arquivo de saída. Por fim, a última ferramenta analisa de acordo com o arquivo de categorias, o arquivo com o *corpus* criado pela primeira ferramenta, gerando um relatório de cada aluno com sua respectiva Presença Social.

Para testar a verificação do conjunto de ferramentas, foram feitos quatro experimentos ao todo, utilizando dois fóruns distintos e dois arquivos de classes com pistas textuais diferentes. Para ter uma comparação, pediu-se para um perito analisá-los também de acordo com o modelo de Bastos (2011). Um arquivo de classes foi povoado com pistas textuais encontradas num arquivo já analisado pelo perito, e o outro pediu-se para o perito povoar com pistas textuais que ele achava interessante.

Os resultados mostraram que o arquivo de classes com pistas textuais encontradas previamente num arquivo já analisado, obtiveram melhores resultados, chegaram mais perto do resultado encontrado pelo perito, em ambos os fóruns.

**Palavras-Chave:** Presença Social, AVEA, ferramenta, lexicometria.

# **Automating the Identification Process of Social Presence in Forums and Chat**

## **ABSTRACT**

Following the model proposed by H lvia Pereira Pinto Bastos to identify the level of Social Presence of students in a Virtual Environment for Teaching and Learning (VETL), was developed three tools.

The model consists of categories and subcategories, and the end of each subcategory there is textual clues, lexical expressions, which are used to identify a sub-corpus, treated here only for posts in forums and chats.

The first tool joins all files with the posts in only one, this process is done to standardize the tags of interest and withdraw from others. The second software allows you to create classes, subclasses and the end of each subclass register their textual clues, so long, projects the Bastos (2011) model in the output file. Finally, the last tool analyzes according to the categories file, the file with the corpus created by the first tool, generating a report of each student with their respective Social Presence.

To test the verification toolset, were made altogether four experiments using two different forums and two class files with different textual clues. For a comparison, we asked to an expert to analyze them also according to the model of Bastos (2011). A class file was populated with clues found in a text file already analyzed by the expert, and the other was asked to the expert populate with textual clues he found interesting.

The results showed that the class file with textual clues previously found a file already analyzed, achieved better results, closer to the results found by the expert in both forums.

**Keywords:** Social Presence, VETL, lexicometrics.

# 1 INTRODUÇÃO

Na Educação a Distância, em especial nos Ambientes Virtuais de Ensino e Aprendizagem (AVEAs), um problema importante é promover a participação e a retenção dos alunos nos cursos. Nesse sentido, segundo Bastos (2011), as interações entre os participantes são consideradas um fator decisivo para a retenção dos alunos nos cursos, visto que promovem a percepção de pertencimento destes alunos no grupo como um todo, diminuindo, assim, o sentimento de isolamento comum em experiências pedagógicas dessa natureza.

Ainda, conforme Bastos (2011), o pertencimento tem forte relação com a presença do aluno nas diferentes interações deste com o ambiente, e a observação de como os alunos se projetam nas interações via Comunicação Mediada por Computador (CMC) pode fornecer pistas da maior ou menor motivação e satisfação dos alunos em AVEAs.

Dentro desse contexto, a aluna Hélvia Pereira Pinto Bastos, doutoranda em Informática na Educação da Universidade Federal do Rio Grande do Sul (PPGIE-UFRGS), está desenvolvendo uma pesquisa em que o principal objetivo consiste justamente em identificar a Presença Social (PS) de alunos em AVEAs através de um modelo desenvolvido especificamente para fóruns e chats.

Contudo, conforme aponta Máximo (2009), a avaliação de um grande número de alunos de forma processual a partir de suas participações em conversações é ainda extremamente difícil e lenta, o que, frequentemente, não torna possível ao professor diagnosticar problemas em tempo hábil. Para realizar tal identificação, ferramentas computacionais de suporte são necessárias.

O trabalho aqui apresentado visa justamente descrever o desenvolvimento de ferramentas que auxiliem na identificação da presença social em fóruns e chats do ambiente Moodle, conforme o modelo e a abordagem propostos por tal aluna, conforme descritos em (BASTOS, 2011).

## 1.1 Objetivos

O objetivo geral é automatizar o processo de percepção do nível de Presença Social dos alunos de um curso no Moodle.

Como objetivos específicos, tem-se:

- Realizar a análise dos arquivos HTMLs contendo o conteúdo (postagens) dos fóruns/chats de um curso;
- Realizar o mapeamento e a uniformização dos dados analisados, gerando um único arquivo padronizado (em XML);

- Promover a definição e manipulação de categorias de presença social, incluindo suas pistas textuais (e armazenamento em XML);
- Realizar o processamento dos conteúdos dos fóruns e dos chats no formato padronizado, levando em conta as categorias e pistas textuais definidos em um arquivo de categorias, gerando um relatório (TXT) com os nomes dos alunos e sua correspondente presença social.

## 1.2 Contribuição

Automatizar o processo de identificação do nível de Presença Social de cada aluno em fóruns e chats do Moodle.

## 1.3 Estrutura do Texto

Este trabalho encontra-se organizado da seguinte forma:

- **Capítulo 2: Contexto** apresenta o problema e os softwares que poderiam resolvê-lo;
- **Capítulo 3: Desenvolvimento da ferramenta** detalha como foi desenvolvida a ferramenta para resolver o problema;
- **Capítulo 4: algoritmos utilizados para análise das categorias cadastradas** mostra como os algoritmos de busca funcionam;
- **Capítulo 5: experimento** apresenta os resultados obtidos com a ferramenta criada;
- **Capítulo 6: conclusões e trabalhos futuros** discute melhorias e novas funcionalidades que podem ser agregadas a ferramenta.

## 2 CONTEXTO

Neste Capítulo será relatado o problema e a abordagem usada para tratá-lo.

### 2.1 Modelagem do Problema

Para Albuquerque (apud MACHADO 2007), presença social é uma sensação subjetiva de estar num ambiente distinto do ambiente físico da pessoa, e a medida que aumenta o usuário de torna mais consciente e envolvido no ambiente remoto e menos consciente do ambiente que está fisicamente localizado.

Neste trabalho será adotada a noção de presença social de Bastos (2011), que define presença social (PS) como a manifestação verbal e percepção da afetividade e interatividade dos sujeitos em relação ao ambiente virtual de ensino e aprendizagem.

O objetivo de seu estudo é fornecer informação quanto ao envolvimento do aluno ao professor, tutor ou orientador. Tal informação pode ser obtida através do grau de PS dos alunos. Presença social é diferente do registro de login de alunos em AVEAs, tal apenas indica a “presença participativa” do aluno. Com base no grau de PS, o professor pode, se achar necessário, intervir de maneira a prevenir alguma eventual desistência de um aluno.

Bastos cria a partir dos trabalhos de Rourke et al. (apud BASTOS 2011) e de Martin e White (ibidem), o seguinte modelo para obtenção do grau de PS

Tabela 2-1: Modelo de Presença Social proposto por Bastos

CLASSE	SUBCLASSE	INDICADORES
AFETIVIDADE	AF1. Emoção	AF1a. Interjeições e expressões interjetivas
		AF1b. Onomatopéias
		AF1c. Emoticons, gifs e imagens
		AF1d. Pontuação repetida
		AF1e. Letras Repetidas
		AF1f. Realce
	AF2. Humor	AF2a. Expressão de humor
	AF3. Auto-revelação	AF3a. Vulnerabilidade
		AF3b. Vida pessoal
INTERATIVIDADE	IN1. Manutenção do diálogo	IN1a. Uso do recurso “responder”
		IN1b. Marcadores Conversacionais
	IN2. Referência a mensagens de outros	IN2a. Simples
		IN2b. Citação direta

CLASSE	SUBCLASSE	INDICADORES	
		IN2c. Elogio	
		IN2d. Concordância	
		IN2e. Discordância	
	IN3. Envolvimento	IN3a. Agradecimento	
		IN3b. Pedido de desculpa	
		IN3c. Oferta de Ajuda	
		IN3d. Incentivo	
		IN3e. Empatia	
	IN4. Perguntas	IN4a. Esclarecimento	
		IN4b. Solicitação / Sugestão	
		IN4c. Questionamento	
	COESÃO SOCIAL	CO1. Nomes Próprios	CO1a. Vocativo
CO1b. Sujeito ou objeto			
CO2. Nomes Genéricos		CO2a. Vocativo	
		CO2b. Sujeito ou objeto	
CO3. Pronomes		CO3a. Locutor	
		CO3b. Interlocutor	
		CO3c. Referente	
CO4. Expressões Fáticas		CO4a. Saudações	
		CO4b. Despedidas	
		CO4c. Termos Volitivos	
FORÇA		FO1. Intensidade	FO1a. Advérbios e expressões adverbiais
			FO1b. Adjetivos Comparativos
	FO2. Quantidade	FO2a. Pron. Indefinidos	
	FO3. Sufixos	FO3a. Aumentativos	
		FO3b. Diminutivos	
	FO4. Valoração	FO4a. Termos valorativos e superlativos	
	FO5. Figuras de linguagem	FO5a. Metáforas, analogias	
FO6. Ênfase	FO5a. Reforço		

FONTE: Bastos (2011, p. 50-51).

Segundo Bastos (2011), a ocorrência destes indicadores no corpus a ser analisado, seja em fóruns ou chats, gera o grau de PS.

## 2.2 Lexicometria

Para encontrar os indicadores, que formam as subclasses, utilizou-se a análise lexicométrica.

A lexicometria é um procedimento metodológico e tecnológico - objetivo, descritivo, indutivo e científico - para tratar estatisticamente dados qualitativos sob fundo quantitativo para a caracterização topológica e combinatória de elementos lexicais de um corpus dado e delimitado, a fim de que a trajetória do discurso, por operações conhecidas e controladas, seja

balizada, e a topologia e a combinatória dos elementos lexicais do corpus caracterizadas (DAMASCENO, 2008, p.3).

Complementando o conceito, conforme Ooi (1998), léxico ou dicionário é a estrutura de dados contendo uma lista de itens lexicais e as informações específicas correspondentes a estes itens, o todo sendo projetado para um fim específico.

Desta forma, cadastrando-se expressões léxicas, aqui chamadas de pistas textuais, e atribuindo um determinado tipo de busca à sua classe/subclasse que as agrupam, é possível encontrar ocorrência dos indicadores das classes.

Ressalta-se que, neste trabalho, não foi utilizado nenhum tipo de análise pragmática, que segundo Levison (1983), estuda os aspectos do significado no uso da linguagem, levando em consideração o contexto da declaração (como lugar, tempo, maneira, etc.), o conhecimento dos status envolvidos, e as ambiguidades linguísticas.

## 2.3 Softwares que utilizam lexicometria para análise de conteúdo

Atualmente existem alguns softwares que utilizam lexicometria para análise de conteúdo, entretanto, cada uma proporciona um enfoque a uma determinada área.

### 2.3.1 WordSmith

Com o WordSmith<sup>1</sup>, software pago, pode-se gerar uma lista de todas as palavras de um texto, em ordem alfabética ou de frequência, bem como suas co-relações.

Entretanto, embora seja uma análise lexicométrica o que precisa ser realizado no âmbito deste trabalho, seria necessário cadastrar em cada análise uma palavra, e, como um mesmo indicador pode ter dezenas ou centenas de palavras ou expressões, isto torna o seu uso, para este determinado fim, o de extração da PS, inconveniente.

### 2.3.2 Eureka

O software Eureka<sup>2</sup>, que está disponibilizado na modalidade freeware, para fins educativos, gera uma lista de todas as palavras de um texto, em ordem alfabética ou de frequência, agrupa diversos documentos por similaridade lexicografia, permitindo escolher o algoritmo para análise, como: algoritmo *String*, algoritmo *Cliques*, algoritmo *Star*, algoritmo *Best-Star*, algoritmo *Full Star*.

Entretanto, a utilização deste software não foi viável, pois, além de que teria de se separar o fórum/chat em um documento por aluno, o agrupamento destes documentos-alunos não nos informaria o PS, ou seja, não se consegue cadastrar palavras para serem procuradas e retornar a quantidade de ocorrências; o que o software nos informaria é que um determinado grupo debate sobre o mesmo tema.

<sup>1</sup> [HTTP://www.lexically.net/wordsmith/](http://www.lexically.net/wordsmith/).

<sup>2</sup> [HTTP://www.inf.ufrgs.br/~wives/wiki/doku.php?id=eureka/](http://www.inf.ufrgs.br/~wives/wiki/doku.php?id=eureka/).

### **3 DESENVOLVIMENTO DA FERRAMENTA**

A primeira abordagem idealizada consistiu numa ferramenta integrada ao Moodle, mas, além de se ter identificado que tal integração seria complexa, a ferramenta final ficaria muito atrelada ao ambiente Moodle. Optou-se, portanto, pelo desenvolvimento de uma ferramenta independente. Com isso, qualquer ambiente que promova diálogos/interações entre os participantes pode fazer uso de tal ferramenta. Para tanto, verificou-se que seria conveniente adotar um formato padrão e que ferramentas de extração e conversão seriam necessárias.

Uma vez que os dados estejam no padrão adotado, a ferramenta de análise de PS pode ser aplicada. Como as categorias de presença social podem variar, assim como suas pistas textuais, optou-se também por desenvolver uma ferramenta de gestão de categorias.

Diante do exposto, optou-se por fazer três ferramentas separadas, cada qual responsável por uma tarefa específica, sendo elas:

- Conversor dos arquivos HTMLs dos fóruns/chats para um único arquivo XML;
- Construtor de categorias, gerando um único arquivo XML;
- Analisador, que a partir dos arquivos XML contendo as postagens e as categorias, além de um arquivo TXT com os nomes dos alunos que devem ser observados, analisa a presença social dos alunos.

A Figura 3.1 apresenta uma visão geral do processo. Para resumir, o usuário deve manualmente salvar o conteúdo do fórum em arquivos HTML. Depois, utiliza o conversor, que transforma o conteúdo dos tais HTML em único arquivo XML. Com o Construtor de Categorias, o usuário monta as suas categorias e subcategorias, da forma que ele quiser, e cadastra as Pistas Textuais e tipo da pesquisa a ser usado, referentes a cada categoria. No final gera um arquivo XML com as categorias.



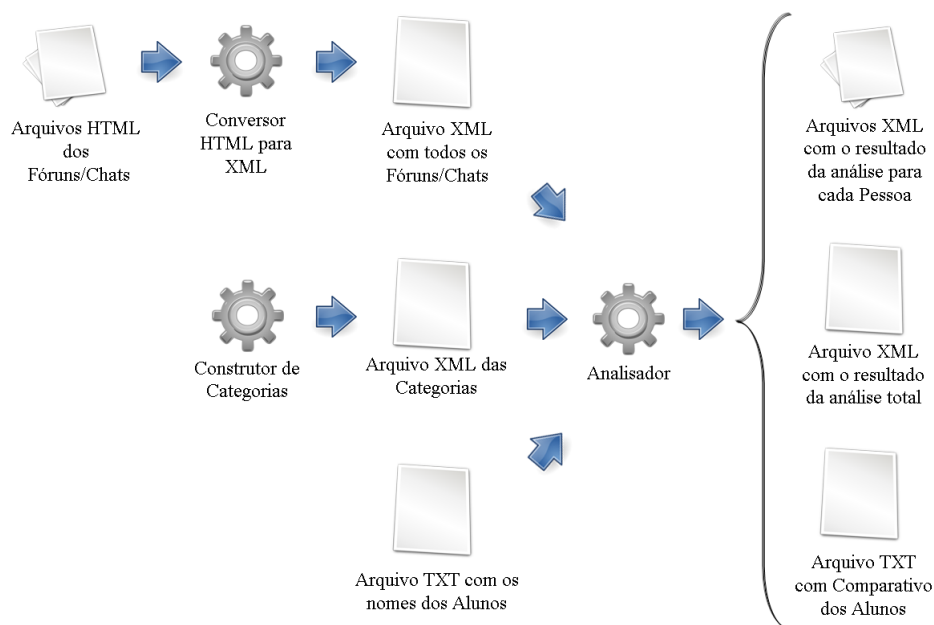


Figura 3.1: Esquema das ferramentas construídas

Após, com o arquivo XML do fórum, o arquivo XML das categorias e mais um arquivo TXT com o nome dos alunos, pegos de forma manual, o Analisador, irá gerar:

- para cada aluno cadastrado no arquivo TXT, um arquivo XML com o número de ocorrências das categoria encontrados em suas postagens;
- um arquivo XML, com a soma de todas as ocorrências encontradas fórum inteiro referente a cada categoria;
- um arquivo TXT, com a listagem dos alunos, sua presença social, e o número de postagens que ele escreveu.

O conversor HTML dos fóruns/chats para XML foi construído na linguagem Ruby e as outras ferramentas. Já o construtor de categorias e o analisador foram feitos na linguagem Java. Todos foram desenvolvidos no ambiente de desenvolvimento NetBeans IDE 6.9.1.

Como essas ferramentas são desacopladas, permite-se a extensão, modificação ou substituição de uma delas sem que isso afete as outras. Para tanto, foi definido um padrão de entrada e de saída para cada ferramenta. Qualquer ferramenta que siga o padrão poderá ser utilizada. Em seguida, tais ferramentas são detalhadas.

### 3.1 Conversor HTML dos fóruns/chats para XML

A fim de facilitar e padronizar o arquivo de entrada para o Analisador Final, optou-se por criar um conversor de arquivos HTML dos fóruns/chats para um único arquivo XML, com *tags* pré-definidas. O formato de tal arquivo é definido na Figura 3.2.

```
1 <topicos>
2   <topico_n>
3     <postagem_m>
4       <titulo> ... </titulo>
5       <autor> ... </autor>
6       <data> ... </data>
7       <conteudo> ... </conteudo>
8     </postagem_m>
9     <postagem_m+1>
10      ...
11    </postagem_m+1>
12  </topico_n>
13  <topico_n+1>
14    ...
15  </topico_n+1>
16 </topicos>
```

Figura 3.2: Formato padrão para arquivo XML com as postagens

Em tal formato:

- <topicos> é a tag raiz da árvore XML, pode conter vários elementos do tipo <topico\_n>;
- <topico\_n> encapsula um arquivo HTML referente a um fórum, com  $n$  variando de 1 até o número de arquivos HTML, um <topico\_n> pode conter várias *tags* <postagem\_m>;
- <postagem\_m> é referente a uma única postagem dentro do fórum, com  $m$  variando de 1 até o número de postagens;
- <titulo> é o título da postagem;
- <autor> é o autor da postagem;
- <data> é a data da postagem;
- <conteudo> é o que a pessoa escreveu dentro da postagem.

Esse formato visa guardar todas as informações pertinentes ao fórum. Pois um fórum possui vários tópicos, tópicos possuem várias postagens, e postagens são formadas por título, autor, data/hora, e conteúdo.

A seguir, na Figura 3.3, encontra-se um exemplo de conversão realizada pela ferramenta a partir de um trecho de fórum no Moodle.

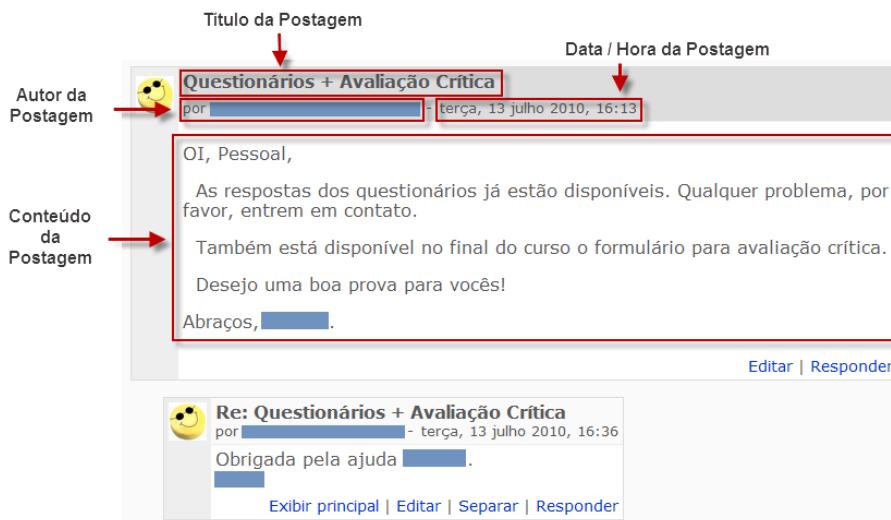


Figura 3.3: Trecho de fórum extraído do Moodle

Percebe-se, na Figura 3.3, que a postagem possui um título (“Questionário + Avaliação Crítica”), um autor (escondido para não revelar a identidade do aluno), uma data/hora e um conteúdo. Tal tópico, ao ser convertido, fica conforme apresentado na Figura 3.4.

```

1 <topicos>
2   <topico_1>
3     <nome>Questionários + Avaliação Crítica</nome>
4     <postagem_1>
5       <titulo>Questionários + Avaliação Crítica
6       </titulo>
7       <autor>Beltrana</autor>
8       <data>terça, 13 julho 2010, 16:13</data>
9       <conteudo>OI , Pessoal , <br /> As respostas
10      dos questionários já estão disponíveis .
11      Qualquer problema , por favor , entrem em
12      contato . <br /> Também está disponível no
13      final do curso o formulário para avaliação
14      crítica . <br /> Desejo uma boa prova para
15      vocês ! <br /> Abraços , Renata . <br />
16      </conteudo>
17    </postagem_1>
18    <postagem_2>
19      <titulo>Re: Questionários + Avaliação Crítica
20      </titulo>
21      <autor>Ciclana</autor>
22      <data>terça, 13 julho 2010, 16:36</data>
23      <conteudo>Obrigada pela ajuda Beltrana . <br />
24      Ciclana </conteudo>
25    </postagem_2>
26  </topico_1>
27 </topicos>

```

Figura 3.4: Arquivo XML gerado a partir do trecho extraído do Moodle

Após o estudo, para gerar este padrão em XML, pensou-se em estender o *corpus* para chats também. O que se fez foi tratar o chat como um fórum.

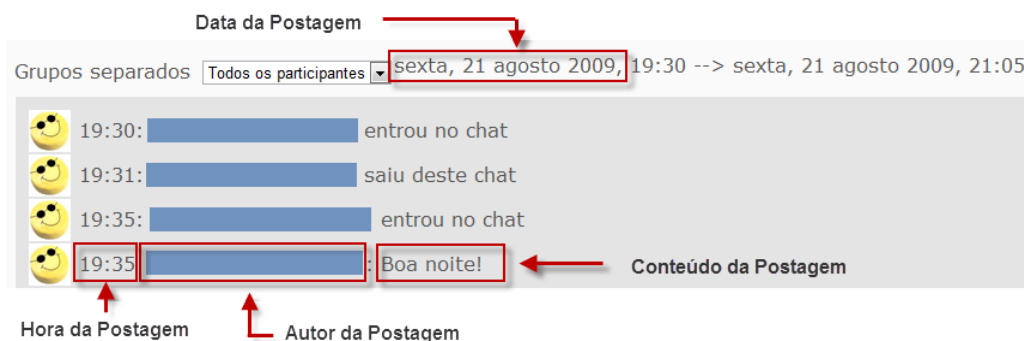


Figura 3.5 : Trecho de chat extraído de arquivo HTML

Como mostrado na Figura 3.5, um chat possui os mesmos elementos de um fórum, apenas não tem um Título referente a cada postagem, a solução adotada foi atribuir como título "Fala n", onde n varia de 1 até o número total de postagens naquela sessão de chat.

É importante salientar que, é feito uma filtragem no arquivo original, visto que nem todos os elementos HTML necessitam constar no arquivo XML. Por exemplo, definição de parágrafos (<p> </p>), seções (<div></div>) e demais elementos de formatação. Alguns elementos que são importantes para a identificação de etiquetas (e.x., cores, negrito, etc.), são mantidos, mas convertidos para uma notação específica.

Tabela 3-1: Notação adotada no arquivo XML com as postagens

ELEMENTO	NOTAÇÃO ADOTADA
Negrito	<b> </b>
Itálico	<i> </i>
Cor	<color> </color>
Alteração do Tamanho da Fonte	<size> </size>
Quebra de linha	 
Emoticon	<emoticon />
Gif	<gif />
Imagem	<image />
Link	<link />
Espaçamento contínuo	Espaçamento simples
Palavras e pontuação	Espaçamento simples entre uma palavra e pontuação
Elementos de formatação encadeados (e.x.: <size> <size> ... </size> </size>)	Foram deixados apenas um (e.x.: <size> ... </size>)
Elementos de formatação que não afetam nenhuma palavra (e.x.: <b> </b>)	Foram excluídos estes elementos

Ressalta-se ainda que, o que realizou-se, foi uma padronização das *tags* de interesse e filtragem das que não introduzem informação relevante. Entretanto, o seu uso indevido ou redundância não foram tratados.

Depois de convertidas, tal arquivo é utilizado pelo analisador de presença social. Cabe salientar que tais elementos são utilizados no construtor de categorias a fim de definir regras de extração. Elas estão, portanto, limitadas à notação adotada. Havendo

necessidade de analisar um elemento diferente, o conversor deve ser refinado, ficando tal tarefa fora do escopo deste trabalho.

## 3.2 Construtor de Categorias

Embora as categorias de presença social sejam relativamente estáticas, elas podem sofrer mudanças de acordo com a abordagem teórica utilizada. Da mesma forma, as pistas textuais, que são utilizadas para identificar cada categoria de presença social, não são fixas. Dependendo do contexto, elas podem variar. Logo, desenvolveu-se uma ferramenta para a definição de categorias de presença social que permite a gerência (inserção, remoção, alteração) de pistas textuais correspondentes. Isso dá flexibilidade ao usuário, permitindo-o testar diferentes opções e refinar a análise dos fóruns.

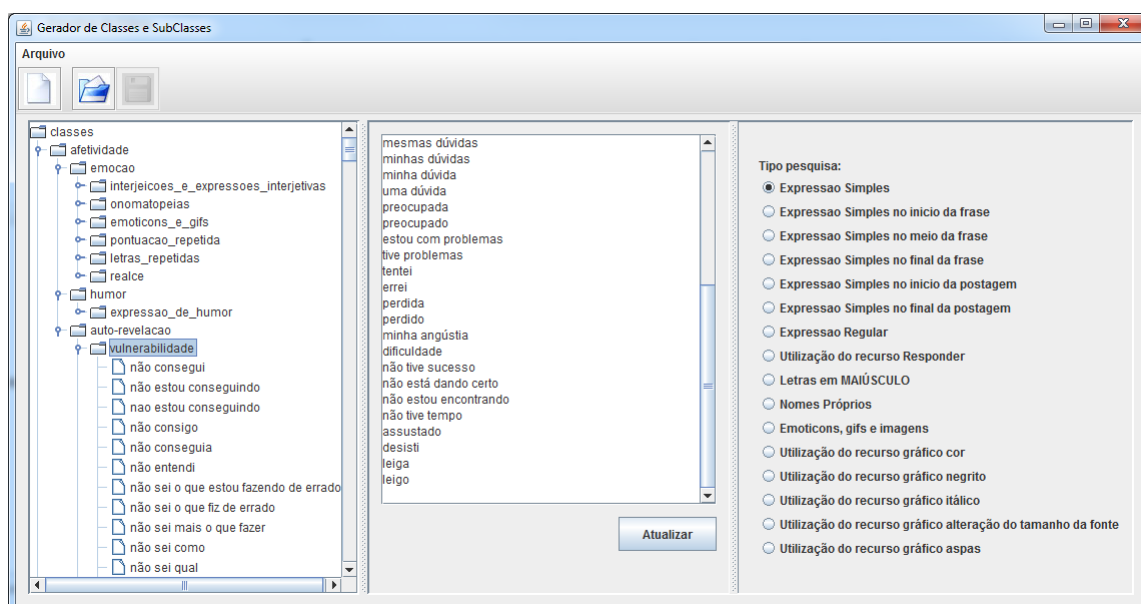


Figura 3.6: Construtor de categorias e pistas textuais

Como mostrado na Figura 3.6, a ferramenta possui três regiões principais. A região mais à esquerda apresenta as categorias de presença social. As categorias podem ter diferentes níveis, de acordo com o grau de profundidade ou granularidade em que se deseja a análise. Para cada categoria, há um conjunto de pistas textuais atreladas. Estas podem ser inseridas, editadas e/ou excluídas na região central.

Durante o desenvolvimento da ferramenta, percebeu-se que existem tipos ou padrões diferentes de pistas textuais, e, para cada um desses padrões, pode haver um algoritmo diferenciado de reconhecimento. Estes algoritmos são mostrados na região à direita. Assim, a cada categoria informa-se o algoritmo que deseja utilizar. Mais informações sobre os algoritmos na Seção 4.

A lista de categorias e suas respectivas pistas textuais também é armazenada em um arquivo XML, utilizando a estrutura descrita na Figura 3.7.

```

1 <classes algoritmo="x">
2   <classe algoritmo="x">
3     <classe algoritmo="x">
4       <pista_textual> ... </pista_textual>
5       <pista_textual> ... </pista_textual>
6     </classe>
7   </classe>
8   <classe algoritmo="x">
9     <pista_textual> ... </pista_textual>
10    <pista_textual> ... </pista_textual>
11    <pista_textual> ... </pista_textual>
12  </classe>
13 </classes>

```

Figura 3.7: Formato padrão para o arquivo XML com as classes e pistas textuais

Em tal formato:

- <classes> é a tag raiz da árvore XML;
- <classe> é o nome da categoria, podendo conter outras tags <classe> ou <pista\_textual>;
- algoritmo="x" é o atributo da tag <classe> e <classes>, x é um número que especifica qual algoritmo deverá ser usado, por simplificação toda tag <classe> o contém;
- <pista\_textual> especifica o que deve ser procurado no texto.

Na Figura 3.8 mostra-se como ficaria a categoria "afetividade", sua subcategoria "emocao", que por sua vez possui quatro subcategorias: "onomatopeias", "emoticons\_e\_gifs", "interjeicoes\_e\_expressoes\_interjetivas", "pontuacao\_repetida". Cabe ressaltar que o nível de aninhamento dentro de uma categoria não é fixo. Além disso, apenas a categoria ou subcategoria mais aninhada é que possui Pista Textual.

```

1 <classes algoritmo="1">
2   <afetividade algoritmo="1">
3     <emocao algoritmo="1">
4       <interjeicoes_e_expressoes_interjetivas
5         algoritmo="1">
6         <pista_textual>bah</pista_textual>
7         <pista_textual>ohhhhhh</pista_textual>
8         <pista_textual>ahhh</pista_textual>
9         <pista_textual>ah tah</pista_textual>
10        <pista_textual>aham</pista_textual>
11      </interjeicoes_e_expressoes_interjetivas>
12      <onomatopeias algoritmo="7">
13        <pista_textual>k(k+)</pista_textual>
14        <pista_textual>([he])([he]+)</pista_textual>
15        <pista_textual>rs(s+)</pista_textual>
16      </onomatopeias>
17      <emoticons_e_gifs algoritmo="11">
18        <pista_textual>xD</pista_textual>
19        <pista_textual>=P</pista_textual>
20        <pista_textual>:O</pista_textual>
21      </emoticons_e_gifs>
22      <pontuacao_repetida algoritmo="7">
23        <pista_textual>([!?!])([!?!]+)</pista_textual>
24      </pontuacao_repetida>
25    </emocao>
26  </afetividade>
27 </classes>

```

Figura 3.8 : Arquivo XML de classes com algumas classes e pistas textuais

### 3.3 Analisador

Esta ferramenta faz a análise de presença social considerando o arquivo XML gerado pelo conversor HTML dos fóruns para XML, que contém as postagens dos alunos, assim como o arquivo XML gerado pela ferramenta de definição de classes e pistas textuais.

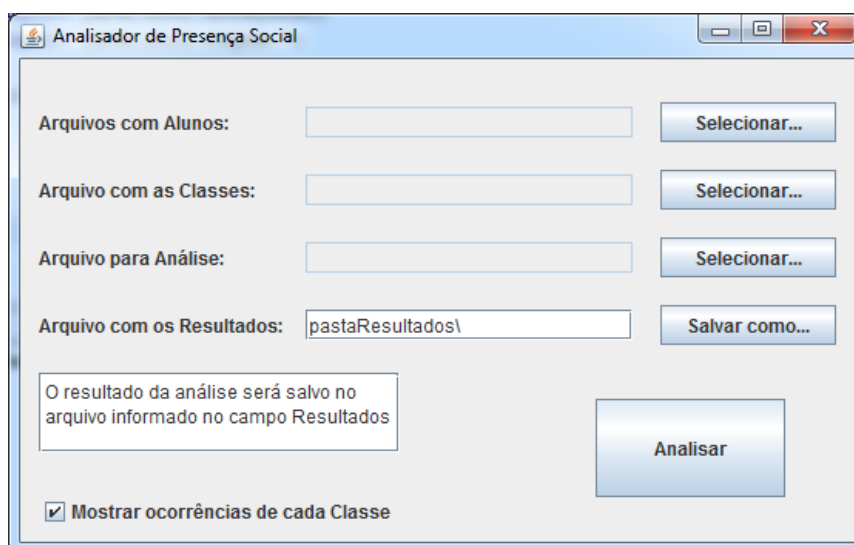


Figura 3.9 : Interface do Analisador

Como ilustrado na Figura 3.9, o Analisador precisa ser informado do arquivo TXT com os nomes dos alunos, o arquivo das categorias, o arquivo para análise e onde quer que ele salve os arquivos resultados.

Utilizando os algoritmos adequados para cada tipo de categoria que foi definido no construtor de categorias, o programa analisa os itens léxicos de cada postagem e contabiliza as categorias que encontra. Ao final, a ferramenta gera um relatório por aluno, um total com a soma de todos os alunos, ambos em XML, com a quantidade de pistas identificadas por categoria; e um arquivo com todos os alunos, sua presença social e a quantidade de postagens que escreveu.

Estes arquivos XML possuem o seguinte formato:

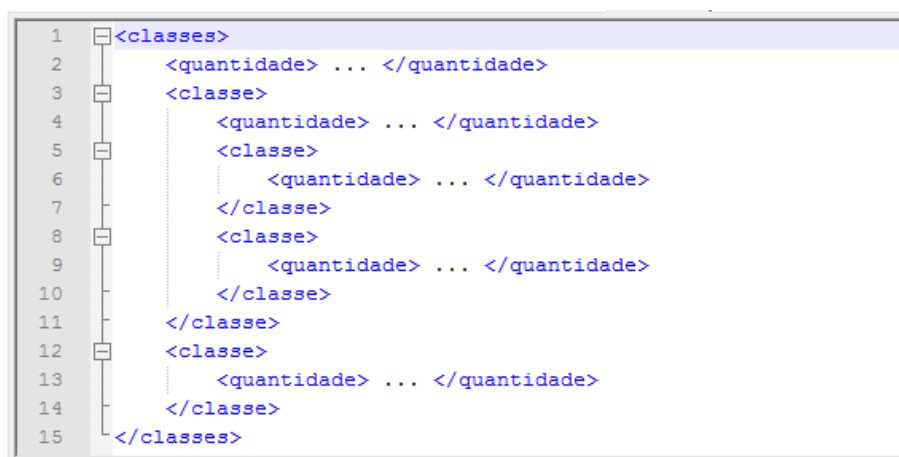



Figura 3.10: Formato padrão para arquivo XML com o resultado da análise

Nesse formato:

- <classes> é a *tag* raiz da árvore XML, e seu primeiro filho é a *tag* <quantidade>, que indica a quantidade total de itens léxicos encontrados no documento;
- <classe> é o nome da categoria, podendo conter outras tags <classe>, mas o seu primeiro filho é a *tag* <quantidade>, que indica a quantidade de itens léxicos desta categoria encontrados no documento, se esta *tag* conter outras *tags* <classe>, a *tag* <quantidade> indicará o somatório das quantidades das *tags* <classe> filhas.

Abaixo, na Figura 3.11, mostra-se um exemplo de arquivo XML com o resultado de uma análise. É importante ressaltar que o valor indicado em "quantidade", referente a uma categoria ou subcategoria, é formado pela soma das "quantidades" de suas subcategorias e assim consecutivamente, até chegar na subcategoria que só contém Pistas Textuais, e, nesta subcategoria, o valor em "quantidade" indica o número de ocorrências das Pistas Textuais.



```
1 <classes>
2   <quantidade>87</quantidade>
3   <afetividade>
4     <quantidade>87</quantidade>
5     <emocao>
6       <quantidade>87</quantidade>
7       <interjeicoes_e_expressoes_interjetivas>
8         <quantidade>3</quantidade>
9       </interjeicoes_e_expressoes_interjetivas>
10      <onomatopeias>
11        <quantidade>3</quantidade>
12      </onomatopeias>
13      <emoticons_e_gifs>
14        <quantidade>36</quantidade>
15      </emoticons_e_gifs>
16      <pontuacao_repetida>
17        <quantidade>45</quantidade>
18      </pontuacao_repetida>
19    </emocao>
20  </classes>
```

Figura 3.11: Exemplo de arquivo XML com o resultado de uma análise

O outro arquivo que o Analisador gera é um arquivo TXT com todos os alunos, sua presença social e a quantidade de postagens que escreveu, como mostrado na Figura 3.12.



	Nome	Presença Social	Qtd Postagens
1			
2	Aluno1	316	20
3	Aluno2	134	10
4	Aluno3	93	10
5	Aluno4	76	12
6	Aluno5	76	7
7	Aluno6	50	7
8	Aluno7	28	2
9	Aluno8	25	3
10	Aluno9	24	4
11	Aluno10	24	3
12	Aluno11	0	0
13	TOTAL	846	78

Figura 3.12 : Arquivo TXT resultante contendo informações sobre os alunos analisados

Também é possível visualizar a quantidade total de ocorrências de cada Pista Textual, conforme Figura 3.13.

Nº total de ocorrências da classe: 87

Pista Textual	Nº de Ocorrência
professora	37
professoras	0
professor	0
professores	2
profe	10
profes	0
prof	3
profs	1
colega	3
colegas	11
aluno	5
aluna	0
alunos	3
alunas	0
grupo	12

Figura 3.13: Visualização do número total de ocorrências de cada pista textual

Para a realização desta ferramenta foi utilizado o *Design Pattern* (Padrão de Projeto) *Strategy*, conforme definido por Gama et al. (2000), pois ele proporciona a alteração e a inserção de novos algoritmos de maneira mais natural e fácil. Para tanto, basta estender a classe *EstrategiaAlgoritmo*, adicionando o novo tipo de busca no Construtor de *Categorias*.

A Figura 3.14 apresenta o diagrama de classes dos principais algoritmos desenvolvidos. Tais algoritmos são detalhados na seção seguinte.

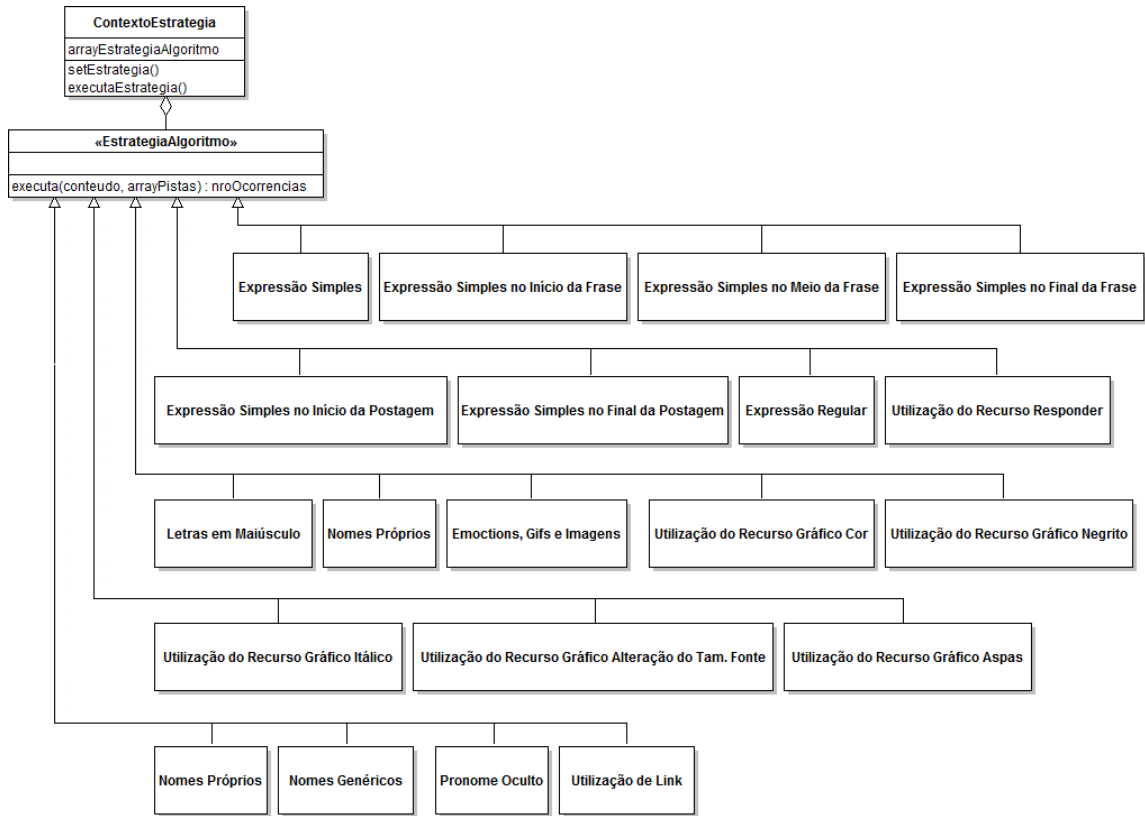


Figura 3.14: Diagrama de Classes

## 4 ALGORITMOS UTILIZADOS PARA ANÁLISE DAS CATEGORIAS CADASTRADAS

Esta seção descreve os algoritmos de detecção de pistas textuais nos fóruns e chats. Ao todo, foram desenvolvidos dezenove algoritmos para suprir as necessidades de busca diferenciadas. Todos fazem uso da lexicometria, entretanto, a maneira como irão procurar as expressões nas postagens varia. Por exemplo, podem procurar apenas no começo da frase, no final da frase, no começo da postagem, etc.

Tabela 4-1: Categorias que se deseja usar para análise

CLASSE	SUBCLASSE	INDICADORES		SUBINDICADORES AUXILIARES			
AFETIVIDADE	AF1. Emoção	1	AF1a. Interjeições e expressões interjetivas				
		7	AF1b. Onomatopéias				
		12	AF1c. Emoticons, gifs e imagens				
		7	AF1d. Pontuação repetida				
		7	AF1e. Letras Repetidas				
			AF1f. Realce	9	Maiúsculo		
				17	Tamanho da Fonte		
				15	Negrito		
				16	Itálico		
				14	Cor		
		18		Aspas			
		AF2. Humor	1	AF2a. Expressão de humor			
		AF3. Auto-revelação	1	AF3a. Vulnerabilidade			
	1		AF3b. Vida pessoal				
INTERATIVIDADE	IN1. Manutenção do diálogo		IN1a. Uso do recurso “responder”				
			IN1b. Marcadores Conversacionais	2	Início Frase		
				4	Final Frase		
	IN2. Referência a mensagens de outros	1	IN2a. Simples				
		1	IN2b. Citação direta				
		1	IN2c. Elogio				
			IN2d. Concordância				

CLASSE	SUBCLASSE	INDICADORES		SUBINDICADORES AUXILIARES	
			IN2e. Discordância		
	IN3. Envolvimento	1	IN3a. Agradecimento		
		1	IN3b. Pedido de desculpa		
			IN3c. Oferta de Ajuda	1 19	Expr Simples Links
		1	IN3d. Incentivo		
		1	IN3e. Empatia		
	IN4. Perguntas	1	IN4a. Esclarecimento		
		1	IN4b. Solicitação / Sugestão		
		1	IN4c. Questionamento		
COESÃO SOCIAL	CO1. Nomes Próprios	10	CO1a. Vocativo, sujeito ou objeto		
	CO2. Nomes Genéricos	11	CO2a. Vocativo, sujeito ou objeto		
	CO3. Pronomes		CO3a. Locutor, Interlocutor, Referente	1	Pronome Simples
				13	Pronome Oculto
CO4. Expressões Fáticas	1	CO4a. Saudações			
	1	CO4b. Despedidas			
	1	CO4c. Termos Volitivos			
FORÇA	FO1. Intensidade		FO1a. Advérbios e expressões adverbiais	1	Expr Simples
				3	Expr Meio da Frase
			FO1b. Adjetivos Comparativos	1	Expr Simples
				7	Expr Regular
	FO2. Quantidade	1	FO2a. Pron. Indefinidos		
	FO3. Sufixos	1	FO3a. Aumentativos		
		1	FO3b. Diminutivos		
	FO4. Valoração	1	FO4a. Termos valorativos e superlativos		
FO5. Figuras de linguagem	1	FO5a. Metáfora, analogias			
FO6. Ênfase	1	FO5a. Reforço			

Tabela 4-2: Legenda dos algoritmos utilizados

#	TIPO DE BUSCA
1	Expressão Simples
2	Expressão Simples no Início da Frase
3	Expressão Simples no Meio da Frase
4	Expressão Simples no Final da Frase
5	Expressão Simples Início da Postagem
6	Expressão Simples Final da Postagem
7	Expressão Regular

#	TIPO DE BUSCA
8	Uso do Recurso Responder
9	Expressões em Maiúsculo
10	Nomes Próprios
11	Nomes Genéricos
12	Emoticons, Gifs e Imagens
13	Pronome Oculoto
14	Utilização do Recurso Gráfico Cor
15	Utilização do Recurso Gráfico Negrito
16	Utilização do Recurso Gráfico Itálico
17	Utilização do Recurso Gráfico Alteração do Tamanho da Fonte
18	Utilização do Recurso Gráfico Aspas
19	Utilização de Link

Esta Tabela 4-1 é diferente da Tabela 2-1. A fim de simplificação, optou-se por colocar os identificadores CO1a.Vocativo e CO1b.Sujeito ou objeto da subclasse CO1.Nomes Próprios, como um único identificador, pois o importante é identificar os nomes próprios e não sua função na oração.

Da mesma maneira, optou-se por esta abordagem com os indicadores de CO2.Nomes Genéricos e CO3.Pronomes.

Entretanto, em outros indicadores, como indicador AF1f.Realce, que englobava diferentes tipos busca, preferiu-se acrescentar novas divisões no indicador. Desta forma, cada nova subdivisão diz respeito ao tipo de busca usado.

No decorrer do desenvolvimento dos tipos de buscas, percebeu-se a necessidade de se criar uma variável @nome para quando se quiser na pista textual indicar que pode ser qualquer nome. Sua utilidade é em expressões do tipo: "concordo com @nome", o que se quer nesta expressão é apenas os "concordo com" que sejam procedidos de algum nome, "concordo com o aumento do salário" não seria um exemplo válido. A variável @nome pode ser um nome de uma pessoa que postou no fórum/chat, ou um nome de um aluno passado ao programa, ou ainda o apelido cadastrado como pista textual no tipo de busca Nome Próprio.

Da mesma forma também se viu a necessidade da variável @nomeger, que pode ser qualquer nome genérico cadastrado como pista textual com o tipo de busca Nome Genérico.

A seguir para descrever o funcionamento de cada algoritmo utilizou-se a notação de expressões regulares da linguagem Java<sup>3</sup> e a idéia de variável.

## 4.1 Expressão Simples

Utiliza as Pistas Textuais cadastradas na categoria, e para cada uma delas executa a seguinte expressão regular:

```
ExpProcurada := ^pistaTextual$
                |^pistaTextual(\s)+
                |(\s)+pistaTextual(\s)+
```

<sup>3</sup> [HTTP://docs.oracle.com/javase/1.4.2/docs/api/java/util/regex/Pattern.html](http://docs.oracle.com/javase/1.4.2/docs/api/java/util/regex/Pattern.html)

$|(\backslash s)^+ \text{pistaTextual} \$$

Procura no conteúdo de cada postagem as pistas textuais cadastradas da maneira como elas foram escritas.

## 4.2 Expressão Simples no Início da Frase

Utiliza as Pistas Textuais cadastradas na categoria, e para cada uma delas executa a seguinte expressão regular:

ExpProcurada :=  $^{\wedge} \text{pistaTextual} \$$   
 $|^{\wedge} \text{pistaTextual} (\backslash s)^+$

Procura a pista textual no início de cada frase do conteúdo de cada postagem. Foram utilizados os caracteres ".?!" como marcadores de termino de frase.

## 4.3 Expressão Simples no Meio da Frase

Utiliza as Pistas Textuais cadastradas na categoria, e para cada uma delas executa a seguinte expressão regular:

ExpProcurada :=  $(\backslash s)^+ \text{pistaTextual} (\backslash s)^+$

Procura a pista textual no meio de cada frase, desconsidera a primeira e ultima palavra, do conteúdo de cada postagem. Foram utilizados os caracteres ".?!" como marcadores de termino de frase.

## 4.4 Expressão Simples no Final da Frase

Utiliza as Pistas Textuais cadastradas na categoria, e para cada uma delas executa a seguinte expressão regular:

ExpProcurada :=  $^{\wedge} \text{pistaTextual} \$$   
 $|(\backslash s)^+ \text{pistaTextual} \$$

Procura a pista textual no final de cada frase do conteúdo de cada postagem. Foram utilizados os caracteres ".?!" como marcadores de termino de frase.

## 4.5 Expressão Simples Início da Postagem

Utiliza as Pistas Textuais cadastradas na categoria, e para cada uma delas executa a seguinte expressão regular:

ExpProcurada :=  $^{\wedge} \text{pistaTextual} \$$   
 $|^{\wedge} \text{pistaTextual} (\backslash s)^+$   
 $|(\backslash s)^+ \text{pistaTextual} (\backslash s)^+$   
 $|(\backslash s)^+ \text{pistaTextual} \$$

Procura a pista textual nos três primeiros parágrafos do conteúdo de cada postagem.

## 4.6 Expressão Simples Final da Postagem

Utiliza as Pistas Textuais cadastradas na categoria, e para cada uma delas executa a seguinte expressão regular:

```
ExpProcurada := ^pistaTextual$
                |^pistaTextual(\s)+
                |(\s)+pistaTextual(\s)+
                |(\s)+pistaTextual$
```

Procura a pista textual nos três últimos parágrafos do conteúdo de cada postagem.

## 4.7 Expressão Regular

Utiliza as Pistas Textuais cadastradas na categoria, aqui Pista Textual é encarada como a expressão regular em si, e para cada uma delas executa a seguinte expressão regular:

```
ExpProcurada := pistaTextual
```

Procura a pista textual no conteúdo de cada postagem.

## 4.8 Uso do Recurso Responder

Não utiliza as Pistas Textuais.

Apenas verifica no título de cada postagem se os primeiros caracteres são "Re:".

## 4.9 Letras em Maiúsculo

Não utiliza as Pistas Textuais.

```
ExpProcurada := [^\p{javaLowerCase}](\p{javaUpperCase})(\p{javaUpperCase})
                (\s\p{javaUpperCase}|\p{Punct})+(s|$)?
```

Procura no conteúdo de cada postagem, seqüências de palavras em maiúsculas.

## 4.10 Nomes Próprios

Utiliza as Pistas Textuais (se quiser cadastrar um apelido ou diminutivo de alguns nomes), os nomes das pessoas que postaram no fórum, e o nome dos alunos passados ao programa.

```
nomeProprio := pistaTextual
                |pessoa.nome.primeiro
                |aluno.nome.primeiro
```

```
ExpProcurada := ^nomeProprio$
                |^nomeProprio(\s)+
                |(\s)+nomeProprio(\s)+
                |(\s)+nomeProprio$
```

Procura o nomeProprio no conteúdo de cada postagem, desde que nomeProprio seja diferente do autor da postagem.

É importante salientar que, como o programa não sabe a quem o apelido é proveniente, logo se um aluno assina com um apelido, o programa conta como outra pessoa.

#### 4.11 Nomes Genéricos

Utiliza as Pistas Textuais.

```
nomeGenerico = pistaTextual(?!( nome1| nome2| ...| nome_n))
```

```
ExpProcurada := ^nomeGenerico$  
                |^nomeGenerico(\s)+  
                |(\s)+nomeGenerico(\s)+  
                |(\s)+nomeGenerico$
```

Procura a pista textual, que tem a função de nome genérico (professor, professora, aluno, etc.) no conteúdo de cada postagem, desde que não seja procedido de um nome próprio.

#### 4.12 Pronome Oculto

Utiliza as Pistas Textuais.

```
nomeProprio = (nome1| nome2| ...| nome_n)  
nomeGenerico = (nomeGer1| nomeGer2| ... | nomeGer_n)  
pronomes = (eu|tu|ele|...|gente|alguém|ninguém|quem)  
ExpProcurada:= (?<!(nomeProprio|nomeGenerico|pronomes))(\s|^)^pistaTextual(\s|$)
```

Procura no conteúdo de cada postagem, desde que a pista textual, neste caso um verbo, não seja precedida por nome próprio, nome genérico ou pronome.

#### 4.13 Emoticons , Gifs e Imagens

Não utiliza as Pistas Textuais.

```
tag := <emoticon />  
      |<gif />  
      |<image />  
ExpProcurada := ^tag$  
                |^tag(\s)+  
                |(\s)+tag(\s)+  
                |(\s)+tag$
```

Procura a ocorrência de <emoticon />, <gif /> e <imagem /> no conteúdo de cada postagem.



#### 4.14 Utilização do Recurso Gráfico Cor

Não utiliza as Pistas Textuais.

tag := <color>

ExpProcurada := ^tag\$

|^tag(\s)+

|(\s)+tag(\s)+

|(\s)+tag\$

Procura a ocorrência de <color> no conteúdo de cada postagem. Note que, embora a tag <color> tenha a sua similar tag fechada </color>, só é procurada a tag aberta, pois o que se quer contabilizar é a quantidade de vezes que este recurso foi utilizado.

#### 4.15 Utilização do Recurso Gráfico Negrito

Não utiliza as Pistas Textuais.

tag := <n>

ExpProcurada := ^tag\$

|^tag(\s)+

|(\s)+tag(\s)+

|(\s)+tag\$

Procura a ocorrência de <n> no conteúdo de cada postagem. Da mesma forma que é descrito na Seção 4.15, só é procurado a tag aberta.

#### 4.16 Utilização do Recurso Gráfico Itálico

Não utiliza as Pistas Textuais.

tag := <i>

ExpProcurada := ^tag\$

|^tag(\s)+

|(\s)+tag(\s)+

|(\s)+tag\$

Procura a ocorrência de <i> no conteúdo de cada postagem. Da mesma forma que é descrito na Seção 4.15, só é procurado a tag aberta.

#### 4.17 Utilização do Recurso Gráfico Alteração do Tamanho da Fonte

Não utiliza as Pistas Textuais.

tag := <size>

ExpProcurada := ^tag\$

|^tag(\s)+

|(\s)+tag(\s)+

|(\s)+tag\$

Procura a ocorrência de <size> no conteúdo de cada postagem. Da mesma forma que é descrito na Seção 4.15, só é procurado a tag aberta.

#### **4.18 Utilização do Recurso Gráfico Aspas**

Não utiliza as Pistas Textuais.

Procura no conteúdo de cada postagem por " e o resultado divide por dois para, desta forma, contabilizar a ocorrência de seu uso.

#### **4.19 Utilização de link**

Não utiliza as Pistas Textuais.

tag := <link />

ExpProcurada := ^tag\$

^tag(\s)+

|(\s)+tag(\s)+

|(\s)+tag\$

Procura a ocorrência de <link /> no conteúdo de cada postagem.

## 5 EXPERIMENTO

Ao todo se fez quatro experimentos: dois relacionados ao fórum A, e dois com o fórum B. O que difere entre cada experimento em um determinado fórum são as pistas textuais cadastradas no arquivo das classes. Visto que as classes são apenas agrupamentos de pistas textuais com o mesmo algoritmo, o que se pretende aqui é mostrar como a escolha das pistas textuais e seus algoritmos impactam no resultado da análise.

### 5.1 Experimento com o Fórum A

Neste primeiro experimento, pediu-se para o perito fazer uma análise manual em um determinado fórum A com 66 tópicos diferentes e 142 postagens ao todo.

Estamos considerando aqui que a Análise Manual feita pelo perito é a ótima, qualquer resultado encontrado a cima dos valores ou abaixo dos encontrados na Análise Manual são considerados falhos.

É importante ressaltar ainda que, para se chegar a estes resultados na Análise Manual no fórum A, foram necessários várias iterações do perito, que possibilitou de um resultado total de 679 para 942, um aumento de 38,7%.

#### 5.1.1 Experimento 1 : Fórum A com pistas textuais específicas

Após o perito ter feito as marcações, cadastrou-se as pistas textuais no arquivo com as classes a partir das expressões marcadas pelo perito, ou seja, as pistas textuais do arquivo das classes são específicas a este fórum.

Os resultados são mostrados na tabela a seguir (APÊNDICE A para mais detalhes):

Tabela 5-1: Fórum A - Análises automáticas com pistas textuais específicas

CATEGORIA		ANÁLISE MANUAL	ANÁLISE AUTOMÁTICA
AFETIVIDADE	AF1. Emoção	138	167
	AF2. Humor	3	3
	AF3. Auto-revelação	48	48
INTERATIVIDADE	IN1. Manutenção do diálogo	94	94
	IN2. Referência a mensagens de outros	15	16
	IN3. Envolvimento	69	69
	IN4. Perguntas	72	72
COESÃO	CO1. Nomes Próprios	91	93

CATEGORIA		ANÁLISE MANUAL	ANÁLISE AUTOMÁTICA
SOCIAL	CO2. Nomes Genéricos	85	87
	CO3. Pronomes	108	126
	CO4. Expressões Fáticas	108	108
FORÇA	FO1. Intensidade	40	45
	FO2. Quantidade	48	48
	FO3. Sufixos	2	2
	FO4. Valoração	17	17
	FO5. Figuras de linguagem	3	3
	FO6. Ênfase	1	1
<b>TOTAL</b>		942	999
<b>PERCENTUAL DE ERRO</b>		-	6,05 %

Na Tabela 5-1 verifica-se alguns valores diferentes, num total de 57. Estas diferenças são provocadas por:

- 38,5% porque a ferramenta não diferencia siglas, e estas são marcadas como se tivesse sido o usado o recurso gráfico maiúsculo, o que não foi a intenção;
- 3,5% porque a ferramenta não identifica o apelido na assinatura no final da postagem como a pertencente a do autor da postagem, e marca como um nome próprio;
- 1,7% por problemas com o reconhecimento de aspas caso seja usado aspas em outra função que não seja a de destacar parte de um texto, como por exemplo na notação de segundo ("), a cada ocorrência par de aspas, seja em qualquer função, é marcado como um uso do recurso gráfico aspas;
- 56,1% porque a ferramenta não compreende o contexto onde a pista textual deve ser marcada, e como apenas busca lexical (com algumas restrições) encontra resultados que não são válidos. Por exemplo, a pista textual "ele", CO3.Pronomes, a ferramenta irá encontrar todas as suas ocorrências no fórum, entretanto, o desejado seria que apenas fossem encontradas a palavra "ele" que estivesse se referindo a pessoas.

### 5.1.2 Experimento 2 : Fórum A com pistas textuais genéricas

Em outro experimento, solicitou-se ao perito (no caso, Hélvia P. Bastos) para povoar as classes com pistas textuais mais genéricas, que possam ser usadas em qualquer fórum ou chat. Desta forma queremos verificar quanto de acerto obtemos no mesmo fórum A, mas com pistas textuais não específicas do fórum, ou seja, sem terem sido criadas, após análise manual. O resultado obtido é mostrado na Tabela 5-2.

Tabela 5-2 : Fórum A - Análise automática com pistas textuais genéricas

	<b>ANÁLISE MANUAL</b>	<b>ANÁLISE AUTOMÁTICA</b>
AFETIVIDADE	189	257
INTERATIVIDADE	250	276
COESÃO SOCIAL	392	534
FORÇA	111	263
<b>TOTAL</b>	<b>942</b>	<b>1330</b>
<b>PERCENTUAL DE ERRO</b>	<b>-</b>	<b>41,1%</b>

Como apresentado na Tabela 5-2, o resultado obtido com a Análise Automática com pistas textuais genéricas não foi tão bom quanto a Análise Automática com pistas textuais específicas do fórum A.

Isto já era esperado, visto que o arquivo com pistas textuais específicas do fórum A foi gerado a partir das ocorrências encontradas pelo perito em sua Análise Manual, enquanto o arquivo de pistas textuais genéricas foi gerado pensando-se em casos mais amplos. Contudo, ao que parece, este arquivo de pistas textuais genéricas não restringe devidamente as pistas textuais as suas classes, o que gerou uma contagem muito maior do que a esperada.

## 5.2 Experimento com o Fórum B

Para fazer esta análise pediu-se que o perito analisasse o fórum B, com 74 tópicos diferentes e 7309 postagens ao todo.

Entretanto, esta análise manual do perito não poderá ser considerada ótima, pois foram feitas apenas duas iterações. Por isto resolveu-se verificar rapidamente os valores encontrados na Análise Manual, a fim de se obter números mais próximos do ótimo.

Tabela 5-3: Fórum B - Comparação entre a Análise Manual e a Análise Manual Corrigida

	<b>ANÁLISE MANUAL</b>	<b>ANÁLISE MANUAL CORRIGIDA</b>
AFETIVIDADE	485	362
INTERATIVIDADE	937	680
COESÃO SOCIAL	1984	1488
FORÇA	1227	871
<b>TOTAL</b>	<b>4633</b>	<b>3401</b>
<b>PERCENTUAL DE ERRO</b>	<b>-</b>	<b>26,6 %</b>

A diferença de valores entre a Análise Manual e a Análise Manual Corrigida, apresentada na Tabela 5-3, é de 1232. Este número foi encontrado da seguinte forma:

- subtraindo as pistas textuais não cadastradas no arquivo de categorias mas marcadas na Análise Manual, um total de 95 % do valor da diferença;

- removendo as pistas textuais marcadas na Análise Manual que estavam sujeitas ao contexto, isto é, somente para o leitor é que fica evidente que é uma ocorrência de uma determinada categoria, o valor relevância é de 2,8 % do total da diferença;
- retirando as marcações duplas, feitas mais de uma vez numa a mesma marcação na mesma expressão, e as marcações que foram feitas sem terem sido especificadas no desenvolvimento da ferramenta, como considerar a alteração da cor do fundo das palavras, *background*, como um tipo de AF1f.Realce; isto gerou 3,3 % do valor total da diferença;
- subtraindo as marcações que foram feitas de maneira equivocadas, pois são referentes na verdade a outra categoria, possui um impacto de 7,2 % da diferença;
- adicionando as marcações que estavam trocadas nas suas devidas categorias, um total de 7,2 % da diferença;
- acrescentando marcações que são fáceis de perceber que foram esquecidas, tais como AF1c.Emoticons, gifs e imagens, AF1d.Pontuação repetida e IN1a. Uso do recurso “responder”, o que faz com que seja acrescentado 2,2%;
- desconsiderando as marcações realizadas em cima de expressões que apresentam erro de ortografia, um total de 1,05% da diferença.

Desta maneira, espera-se uma aproximação de uma Análise Ótima, entretanto salienta-se que apenas foram inseridas marcações de AF1c, AF1d e IN1a, por estas serem de fácil identificação, o que não ocorre com as demais categorias que precisariam de um especialista.

### 5.2.1 Experimento 1 : Fórum B com pistas textuais específicas do fórum A

No primeiro experimento com o fórum B utilizou-se o mesmo arquivo de pistas textuais específicas do fórum A.

Tabela 5-4: Fórum B - Análise automática com pistas textuais específica do fórum A

	<b>ANÁLISE MANUAL CORRIGIDA</b>	<b>ANÁLISE AUTOMÁTICA</b>
AFETIVIDADE	362	811
INTERATIVIDADE	680	649
COESÃO SOCIAL	1488	1453
FORÇA	871	872
<b>TOTAL</b>	<b>3401</b>	<b>3785</b>
<b>PERCENTUAL DE ERRO</b>	-	<b>11,2%</b>

Observando a Tabela 5-4 percebe-se que a maior diferença se encontra na classe Afetividade.

O erro em Afetividade acontece principalmente por ele contar, em Realce, a utilização dos recursos gráficos (cor, negrito, itálico, alteração do tamanho da fonte, expressões em maiúsculo). Quando se tentou filtrar e padronizar o arquivo XML com o

fórum/chat, não se pensou no caso que dentro de um conteúdo da postagem, cada parágrafo/frase pode estar formato com algum recurso gráfico, mas visualmente todo o conteúdo está com o mesmo recurso gráfico. Assim, onde na análise manual contaria uma única ocorrência, a automática conta várias.

A classes Interatividade e Coesão Social apresentaram menos ocorrências do que a Análise Manual Corrigida, isto pode ser devido de que como as pistas textuais foram retiradas apenas do fórum A, ele não deveria conter todas as pistas textuais que identificam estas categorias.

Já nas classes Força, a contagem a cima da Análise Manual Corrigida não é surpresa, pois nesta classe as pistas textuais são muito variadas e muitas vezes ocorrem de uma mesma pista textual ser cadastrada em mais de uma categoria, mas com tipo de busca diferente. Ao que parece, apenas as pistas textuais cadastradas por terem sido encontradas no fórum A, não foram capazes de restringir satisfatoriamente a categoria, o que permitiu que a ferramenta gerasse falsos positivos.

### 5.2.2 Experimento 2 : Fórum B com pistas textuais genéricas

Aqui utilizou para analisar o fórum B o arquivo de pistas textuais genéricas feito na Seção 5.1.2, segundo experimento no fórum A.

Tabela 5-5: Fórum B - Análise automática com pistas textuais genéricas

	<b>ANÁLISE MANUAL CORRIGIDA</b>	<b>ANÁLISE AUTOMÁTICA</b>
AFETIVIDADE	362	864
INTERATIVIDADE	680	953
COESÃO SOCIAL	1488	2429
FORÇA	871	2384
<b>TOTAL</b>	<b>3401</b>	<b>6630</b>
<b>PERCENTUAL DE ERRO</b>	-	<b>95%</b>

Neste experimento todos os resultados encontrados para as classes na Análise Automática são maiores que a da Análise Manual Corrigida.

A classe Afetividade apresenta os mesmos problemas relatados na Seção 5.2.1, primeiro experimento no fórum B, o que gera um percentual de 12,8% da diferença de 3229 entre as duas análises, Manual Corrigida e Automática.

Já nas demais classes, aparentemente, os resultados acima do esperado, cerca de 86,6% da diferença, são causados porque as pistas textuais cadastradas nas classes com seus tipos de buscas não são restritivas o suficiente ou porque o especialista pode ter esquecido de marcá-las no texto.

### 5.3 Análise dos Resultados

Na construção de uma tabela que apresente todos os resultados adquiridos com os experimentos, obtém-se o seguinte:

Tabela 5-6: Todos os resultados obtidos com os experimentos

	ANÁLISE MANUAL	ANÁLISE AUTOMÁTICA			
		Pistas textuais específicas do fórum A		Pistas textuais específicas genéricas	
		Total	% de erro	Total	% de erro
FÓRUM A	978	999	6,05 %	1330	41,1%
FÓRUM B	3401	4165	11,2%	6630	95,5%

Nota-se que, de acordo com a Tabela 5-6, os experimentos realizados com as pistas textuais específicas do fórum A, obtiveram melhores resultados do que as pistas textuais genéricas.

Entretanto, ressalta-se que, no fórum B, a Análise Manual não é a ótima.

Dessa forma, não se pode afirmar nada com certeza a respeito das pistas textuais genéricas, os resultados podem indicar tanto que elas estão muito amplas e não restringem devidamente a busca, como que as pistas textuais genéricas estão feitas de maneira apropriada e englobam a maioria dos casos.

O que se pode concluir é que embora não se tenha um analisador pragmático, para este problema, foi satisfatório o uso de pistas textuais e tipos de buscas diversificados para encontrar o grau de PS.

Contudo pistas textuais que levam em consideração o contexto são uma limitação da ferramenta. Assim como ter que cadastrar para cada expressão simples suas variações em grau, gênero, número, pessoa e tempo. O que pode levar a resultados falhos até se ter uma ampla listagem.



## 6 CONCLUSÕES E TRABALHOS FUTUROS

O trabalho aqui desenvolvido lidou com a automatização do processo de presença social conforme o modelo proposto por Bastos (2011). Para tanto, os objetivos do trabalho consistiam em realizar:

- a análise dos arquivos HTMLs com o conteúdo dos fóruns/chats, foi realizado de modo adequado;
- o mapeamento e padronização dos arquivos HTMLs o conteúdo dos fóruns/chats para um único arquivo em XML, embora tenha se conseguido satisfatoriamente mapear elementos do HTML para *tags* de interesse, não foi tratado a redundância, isto afeta principalmente as *tags* de formatação;
- a definição e manipulação de categorias e suas pistas textuais num arquivo em XML, foi um objetivo específico atingido;
- o processamento do conteúdo dos fóruns/chats levando em consideração as categorias e pistas textuais definidos no arquivo de categorias, gerando um relatório com os nomes dos alunos e sua correspondente presença social, foi efetuado de maneira apropriada.

Desta forma, alcançou-se o objetivo principal: automatizar o processo de percepção do nível de Presença Social dos alunos de um curso no Moodle.

Este trabalho teve como sua principal contribuição o conjunto de ferramentas que permitem automatizar a obtenção do nível de Presença Social, e o artigo aceito *Identifying Social Presence Indicators in Educational Computer Conferencing*, no IADIS CELDA 2011.

Contudo, uma das limitações do conjunto de ferramentas, são as pistas textuais que levam em consideração o contexto, pois não se fez uso de análise pragmática. Outra limitação é ter que cadastrar para cada expressão suas variações, em grau, gênero, número, pessoa e tempo, isto torna o trabalho cansativo e de fácil esquecimento de alguma variação.

Além disto, como as categorias não são estáticas, de modo que podem ser adicionadas novas, modificadas, ou excluídas, novos tipos de busca se farão necessários. Percebendo este ponto, para complementar este trabalho, o próximo passo seria a automatização da inserção de novos algoritmos já que a arquitetura proposta prevê isto, facilitando tal processo.

Seria também importante, no futuro, uma integralização maior com o Moodle, ou uma automatização mais facilitada do processo, por exemplo, um módulo do Moodle

para gerar o XML do fórum e encaminhá-lo para um serviço Web que o processe e retorne o arquivo de categorias de presença social identificadas.

A fim de complementar as pistas textuais cadastradas, seria interessante se trabalhar com alguma ferramenta de correlações entre as palavras. Pois, como visto, o quanto mais trabalhadas ou refinadas forem as pistas textuais maior a certeza no grau de PS.

Continuando a idéia de se criar variáveis para poderem colocar nas pistas textuais, poderia ser expandida para outros tipos, como @lugar (que englobaria casa, escola, parque, etc.), seria interessante que este recurso estivesse disponibilizado na interface, para assim, o usuário poder usar variáveis conforme a necessidade.

## REFERÊNCIAS

BASTOS, H. **Presença Plus: Um Modelo de Identificação de Presença Social como Ferramenta Auxiliar da Tutoria em Ambientes Virtuais de Ensino e Aprendizagem**. 2011. Tese de Doutorado em Informática na Educação da Universidade Federal do Rio Grande do Sul, PPGIE-UFRGS, Porto Alegre. (Em andamento).

BASTOS, H.; Bercht, M.; Wives, L. K.; Kambara-Silva, J. K. da. IDENTIFYING SOCIAL PRESENCE INDICATORS IN EDUCATIONAL COMPUTER CONFERENCING. In: **IADIS International Conference IADIS CELDA 2011**, 2011. (Artigo aceito).

DAMASCENO, E. A. Lexicometria, geração de descritores, construção de ontologias e ensino de línguas: implicações e perspectivas. In: **Magalhães, J. S. de & Travaglia, L. C.. (Org.). Múltiplas Perspectivas em Linguísticas**. 1a ed. Uberlândia: EDUFU, 2008, v. 01, p. 01-3037.

GAMMA, E. et al. **Padrões de Projeto : Soluções Reutilizáveis de Software Orientado a Objetos**. Porto Alegre: Bookman, 2000.

LEVINSON, Stephen C. **Pragmatics**. Cambridge, England: Cambridge University, 1983.

KAMPPFF, A. J. C. **Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio à Prática Docente**. 2009. 186 f. Tese de Doutorado em Informática na Educação, UFRGS, Porto Alegre. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/19032>>. Acesso em: dez. 2011.

MACHADO, G. J. C. **A Presença Social em Ambientes Virtuais de Aprendizagem**. 2007. Tese de Doutorado em Informática na Educação, UFRGS, Porto Alegre. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/14786>>. Acesso em: dez. 2011.

MÁXIMO, L. F. **A Efetividade de Feedbacks Informatizados Sobre a Autoregulação da Aprendizagem em Cursos a Distância : Um Estudo de Caso na Área da Computação**. 2009. 105 f. Tese de Doutorado em Informática na Educação, UFRGS, Porto Alegre. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/17432>>. Acesso em: dez. 2011.

OOI, V. **Computer corpus lexicography**. Edinburgh University Press, 1998

PEREIRA, F. C. N.; GROSZ, Barbara J. **Natural Language Processing**. Cambridge : Mit, 1994.

RICH, E.; KNIGHT, K. **Inteligência Artificial**. 2a ed. São Paulo: Makron Books, 1993.

WIVES, L. K. **Um estudo sobre Agrupamento de Documentos Textuais em Processamento de Informações não Estruturadas Usando Técnicas de "Clustering"**. 1999. 102 f. Dissertação de Mestrado em Ciência da Computação – Instituto de Informática, UFRGS, Porto Alegre. Disponível em: < <http://www.leandro.wives.nom.br/pt-br/publicacoes/dissertacao.pdf> >. Acesso em: ago. 2011.

WORDSMITH. **Step-by-step guide to WordSmith**. Disponível em: < [http://www.lexically.net/wordsmith/step\\_by\\_step/index.html](http://www.lexically.net/wordsmith/step_by_step/index.html) >. Acesso em: ago. 2011.

## APÊNDICE A - Comparação entre as análises do fórum A com pistas textuais específicas do fórum A

Classe	Subclasse	Indicador	Análise Manual	Análise Automática
AFETIVIDADE	AF1. Emoção	AF1a. Interjeições e locuções interjetivas	10	10
		AF1b. Onomatopéias	3	3
		AF1c. Emoticons, gifs e imagens	17	17
		AF1d. Pontuação repetida	42	42
		AF1e. Letras Repetidas	1	1
		AF1f. Realce	65	94
	AF2. Humor	AF2a. Expressão de humor	3	3
	AF3. Auto-revelação	AF3a. Vulnerabilidade	46	46
		AF3b. Vida pessoal	2	2
INTERATIVIDADE-DE	IN1. Manutenção do diálogo	IN1a. Uso do recurso “responder”	89	89
		IN1b. Marcadores Conversacionais	5	5
	IN2. Referência a mensagens de outros	IN2a. Citação Simples	14	15
		IN2b. Citação direta	0	0
		IN2c. Elogio	0	0
		IN2d. Concordância	1	1
		IN2e. Discordância	0	0
	IN3. Envolvimento	IN3a. Agradecimento	63	63
		IN3b. Pedido de desculpa	2	2
		IN3c. Of. de Ajuda / Sugestão	4	4
		IN3d. Incentivo	0	0
		IN3e. Empatia	0	0
	IN4. Perguntas	IN4a. Esclarecimento	72	72
		IN4b. Solicitação		
IN4c. Questionamento		0	0	
COESÃO SOCIAL	CO1. Nomes Próprios	CO1a. Vocativo	91	93
		CO1b. Sujeito ou objeto		
CO2. Nomes Genéricos	CO2a. Vocativo	85	87	
	CO2b. Sujeito ou objeto			

<b>Classe</b>	<b>Subclasse</b>	<b>Indicador</b>	<b>Análise Manual</b>	<b>Análise Automática</b>
	CO3. Pronomes	CO3a. Locutor CO3b. Interlocutor CO3c. Referente	108	126
	CO4. Expressões Fáticas	CO4a. Saudações	44	44
		CO4b. Despedidas	63	63
		CO4c. Expressões Volitivas	1	1
FORÇA	FO1. Intensidade	FO1a. Advérbios e expressões adverbiais	38	43
		FO1b. Adejtivos Comparativos	2	2
	FO2. Quantidade	FO2a. Pron. Indefinidos	48	48
	FO3. Sufixos	FO3a. Aumentativos	0	0
		FO3b. Diminutivos	2	2
	FO4. Valoração	FO4a. Termos Valorativos	17	17
	FO5. Figuras de linguagem	FO5a. Metáforas, Analogias	3	3
FO6. Ênfase	FO6a. Reforço	1	1	
<b>TOTAL</b>			<b>942</b>	<b>999</b>

## APÊNDICE B - Comparação entre as análises no fórum A com pistas textuais genéricas

Classe	Subclasse	Indicador	Análise Manual	Análise Automática
AFETIVIDADE	AF1. Emoção	AF1a. Interjeições e locuções interjetivas	10	36
		AF1b. Onomatopéias	3	3
		AF1c. Emoticons, gifs e imagens	17	17
		AF1d. Pontuação repetida	42	42
		AF1e. Letras Repetidas	1	1
		AF1f. Realce	65	94
	AF2. Humor	AF2a. Expressão de humor	3	3
	AF3. Auto-revelação	AF3a. Vulnerabilidade	46	60
		AF3b. Vida pessoal	2	1
INTERATIVIDADE	IN1. Manutenção do diálogo	IN1a. Uso do recurso “responder”	89	89
		IN1b. Marcadores Conversacionais	5	19
	IN2. Referência a mensagens de outros	IN2a. Citação Simples	14	15
		IN2b. Citação direta	0	0
		IN2c. Elogio	0	0
		IN2d. Concordância	1	6
		IN2e. Discordância	0	0
	IN3. Envolvimento	IN3a. Agradecimento	63	63
		IN3b. Pedido de desculpa	2	2
		IN3c. Of. de Ajuda / Sugestão	4	0
		IN3d. Incentivo	0	0
		IN3e. Empatia	0	0
	IN4. Perguntas	IN4a. Esclarecimento	72	82
		IN4b. Solicitação		
IN4c. Questionamento		0	0	
COESÃO SOCIAL	CO1. Nomes Próprios	CO1a. Vocativo CO1b. Sujeito ou objeto	91	93
	CO2. Nomes Genéricos	CO2a. Vocativo CO2b. Sujeito ou objeto	85	92

<b>Classe</b>	<b>Subclasse</b>	<b>Indicador</b>	<b>Análise Manual</b>	<b>Análise Automática</b>
	CO3. Pronomes	CO3a. Locutor CO3b. Interlocutor CO3c. Referente	108	230
	CO4. Expressões Fáticas	CO4a. Saudações	44	45
		CO4b. Despedidas	63	73
		CO4c. Expressões Volitivas	1	1
FORÇA	FO1. Intensidade	FO1a. Advérbios e expressões adverbiais	38	60
		FO1b. Adejtivos Comparativos	2	3
	FO2. Quantidade	FO2a. Pron. Indefinidos	48	169
	FO3. Sufixos	FO3a. Aumentativos	0	0
		FO3b. Diminutivos	2	0
	FO4. Valoração	FO4a. Termos Valorativos	17	28
	FO5. Figuras de linguagem	FO5a. Metáforas, Analogias	3	3
FO6. Ênfase	FO6a. Reforço	1	0	
<b>TOTAL</b>			<b>942</b>	<b>1330</b>



## APÊNDICE C - Comparação entre as análises no fórum B com pistas textuais específicas do fórum A

Classe	Subclasse	Indicador	Análise Manual	Análise Automática
AFETIVIDADE	AF1. Emoção	AF1a. Interjeições e locuções interjetivas	11	5
		AF1b. Onomatopéias	10	16
		AF1c. Emoticons, gifs e imagens	45	45
		AF1d. Pontuação repetida	65	65
		AF1e. Letras Repetidas	1	28
		AF1f. Realce	195	611
	AF2. Humor	AF2a. Expressão de humor	0	0
	AF3. Auto-revelação	AF3a. Vulnerabilidade	32	41
		AF3b. Vida pessoal	3	0
INTERATIVIDADE-DE	IN1. Manutenção do diálogo	IN1a. Uso do recurso “responder”	548	548
		IN1b. Marcadores Conversacionais	17	13
	IN2. Referência a mensagens de outros	IN2a. Citação Simples	5	9
		IN2b. Citação direta	0	0
		IN2c. Elogio	1	0
		IN2d. Concordância	11	1
		IN2e. Discordância	0	0
	IN3. Envolvimento	IN3a. Agradecimento	51	46
		IN3b. Pedido de desculpa	3	3
		IN3c. Of. de Ajuda / Sugestão	0	0
		IN3d. Incentivo	3	0
		IN3e. Empatia	1	0
	IN4. Perguntas	IN4a. Esclarecimento	39	29
		IN4b. Solicitação		
IN4c. Questionamento		1	0	
COESÃO SOCIAL	CO1. Nomes Próprios	CO1a. Vocativo CO1b. Sujeito ou objeto	160	166
	CO2. Nomes Genéricos	CO2a. Vocativo CO2b. Sujeito ou objeto	522	507

<b>Classe</b>	<b>Subclasse</b>	<b>Indicador</b>	<b>Análise Manual</b>	<b>Análise Automática</b>
	CO3. Pronomes	CO3a. Locutor	596	577
		CO3b. Interlocutor		
		CO3c. Referente		
	CO4. Expressões Fáticas	CO4a. Saudações	112	113
		CO4b. Despedidas	89	90
		CO4c. Expressões Volitivas	9	0
FORÇA	FO1. Intensidade	FO1a. Advérbios e expressões adverbiais	244	475
		FO1b. Adejtivos Comparativos	31	38
	FO2. Quantidade	FO2a. Pron. Indefinidos	323	293
	FO3. Sufixos	FO3a. Aumentativos	0	0
		FO3b. Diminutivos	3	1
	FO4. Valoração	FO4a. Termos Valorativos	262	65
	FO5. Figuras de linguagem	FO5a. Metáforas, Analogias	7	0
FO6. Ênfase	FO6a. Reforço	1	0	
<b>TOTAL</b>			<b>3401</b>	<b>3785</b>

## APÊNDICE D - Comparação entre as análises no fórum B com pistas textuais genéricas

Classe	Subclasse	Indicador	Análise Manual	Análise Automática
AFETIVIDADE	AF1. Emoção	AF1a. Interjeições e locuções interjetivas	11	28
		AF1b. Onomatopéias	10	10
		AF1c. Emoticons, gifs e imagens	45	45
		AF1d. Pontuação repetida	65	65
		AF1e. Letras Repetidas	1	28
		AF1f. Realce	195	611
	AF2. Humor	AF2a. Expressão de humor	0	0
	AF3. Auto-revelação	AF3a. Vulnerabilidade	32	71
		AF3b. Vida pessoal	3	6
INTERATIVIDADE-DE	IN1. Manutenção do diálogo	IN1a. Uso do recurso “responder”	548	548
		IN1b. Marcadores Conversacionais	17	67
	IN2. Referência a mensagens de outros	IN2a. Citação Simples	5	6
		IN2b. Citação direta	0	0
		IN2c. Elogio	1	1
		IN2d. Concordância	11	42
		IN2e. Discordância	0	4
	IN3. Envolvimento	IN3a. Agradecimento	51	53
		IN3b. Pedido de desculpa	3	4
		IN3c. Of. de Ajuda / Sugestão	0	6
		IN3d. Incentivo	3	3
		IN3e. Empatia	1	12
	IN4. Perguntas	IN4a. Esclarecimento	39	205
		IN4b. Solicitação		
IN4c. Questionamento		1	2	
COESÃO SOCIAL	CO1. Nomes Próprios	CO1a. Vocativo CO1b. Sujeito ou objeto	160	166
	CO2. Nomes Genéricos	CO2a. Vocativo CO2b. Sujeito ou objeto	522	628

<b>Classe</b>	<b>Subclasse</b>	<b>Indicador</b>	<b>Análise Manual</b>	<b>Análise Automática</b>
	CO3. Pronomes	CO3a. Locutor	596	1384
		CO3b. Interlocutor		
		CO3c. Referente		
	CO4. Expressões Fáticas	CO4a. Saudações	112	116
		CO4b. Despedidas	89	126
		CO4c. Expressões Volitivas	9	9
FORÇA	FO1. Intensidade	FO1a. Advérbios e expressões adverbiais	244	671
		FO1b. Adejtivos Comparativos	31	80
	FO2. Quantidade	FO2a. Pron. Indefinidos	323	1181
	FO3. Sufixos	FO3a. Aumentativos	0	0
		FO3b. Diminutivos	3	3
	FO4. Valoração	FO4a. Termos Valorativos	262	425
	FO5. Figuras de linguagem	FO5a. Metáforas, Analogias	7	21
FO6. Ênfase	FO6a. Reforço	1	3	
<b>TOTAL</b>			<b>3401</b>	<b>6630</b>