

Ordenador Semântico de Mensagens do Correio Eletrônico

Alextian Bartholomeu Liberato², Vanderson Jose Ildefonso Silva², Lucinéia Barbosa da Costa¹, Orivaldo de Lira Tavares¹, Crediné Silva de Menezes¹

¹Universidade Federal do Espírito Santo (UFES)
PPGI – Vitória – ES

²Instituto Federal do Espírito Santo (IFES)
Coordenadoria de Informática – Campus Colatina – ES.

{abliberato, vanderson.ildefonso, cliklucineia}@gmail.com,
{credine,tavares}@inf.ufes.br,

***Abstract.** This paper presents experiences gained from the use of Natural Language Processing for ordering messages of e-mail semantically. These experiences were recorded and tested in order to assess the ordering of each message to the areas educational, professional and entertainment. It also presents a working prototype of a digital procedure to sort semantically messages to the user in the selected area.*

***Resumo.** Esse artigo apresenta experiências adquiridas com o uso de Processamento de Linguagem Natural para a ordenação semântica das mensagens de correio eletrônico. Essas experiências foram registradas e testadas de modo a avaliar a ordenação de cada mensagem para as áreas educacionais, profissionais e de entretenimento. É apresentado também o protótipo funcional de um procedimento digital que ordena semanticamente as mensagens para o usuário na área selecionada.*

1. Introdução

O correio eletrônico (*e-mail*) tem sido usado como um importante serviço no ambiente educacional, uma vez que a interação é o procedimento indispensável para a aprendizagem. Em qualquer modalidade de ensino encontramos esse serviço aplicado nos ambientes acadêmicos, quer seja na interação aluno-aluno, aluno-professor ou aluno-monitor, nos dois sentidos.

Fácil de usar e disponível gratuitamente, o correio eletrônico tornou-se um grande aliado dos docentes que interagem com seus alunos via esse serviço, principalmente, por permitir uma comunicação assíncrona, além do registro das dúvidas, questionamentos, sugestões e críticas, sem a necessidade de conhecimento avançado de computação.

O e-mail ainda é uma das ferramentas de comunicação mais usadas na web, mesmo quando comparadas com as redes sociais, devido à sua facilidade de uso. Pode ser empregado para: entrega de trabalhos; oferecimento de informações; orientações; atividades escolares; incentivo ao trabalho colaborativo (Montovani et al. 2011).

Os campos de teoria da aprendizagem e design instrucional estão no meio de uma revolução científica em que seus fundamentos filosóficos objetivos estão sendo substituídos por uma epistemologia construtivista. Jonassena et al. (2009) descrevem os pressupostos de uma epistemologia construtivista, contrastando-os com os pressupostos

objetivistas, e, em seguida, descrevem os sistemas de ensino que podem apoiar a aprendizagem construtiva à distância. Por exemplo, com troca de mensagens e informações curriculares.

Apesar do crescente uso de sofisticados sistemas de *webmail* com recursos de automação para troca de mensagens, localização de contatos e diversos outros, o processo de ordenação das mensagens para leitura ainda é realizado com base na ordem cronológica de recebimento, o que impõe uma série de desafios para o proprietário do e-mail. Afinal, a identificação da importância de cada mensagem recebida por *e-mail* não é algo trivial ou facilmente automatizável.

Neste contexto, propomos o desenvolvimento de um procedimento computacional inteligente, que utiliza um conjunto de técnicas e ferramentas para auxiliar o proprietário do e-mail na ordenação semântica de suas mensagens. Cada técnica tem como objetivo criar uma ordenação semântica automática do e-mail, conforme seu contexto e não apenas por palavras-chave como o utilizado na maioria dos sistemas convencionais. A ordenação semântica faz a classificação das mensagens e prioriza as classes relacionadas com atividades educacionais.

O uso de Processamento de Linguagem Natural (*PLN*) é proposto em diversas áreas do conhecimento, com aplicações em tradução, interpretação, auto resumo, busca de informação em texto, interface homem-máquina, “*chatterbots*”, ensino-aprendizagem, entre outras. Em Fontana et al. (2007) encontramos o uso de *PLN* em um sistema com interface em linguagem natural, denominado *Hades*, onde o médico pode consultar informações do paciente usufruindo da própria linguagem utilizada em um diálogo.

Em Araujo e Carvalho (2006) é proposto o uso de *PLN* para análise de frases, obtendo informações sobre cada palavra, de forma a encontrar as sentenças, a organização, título, árvore sintática, entre outros.

Borges e Lima (2008) propõem o uso de *PLN* como um método de expansão de consulta, o que representa um ganho de precisão e abrangência na recuperação dos documentos.

Nosso objetivo é descrever neste artigo, as técnicas e ferramentas utilizadas para o desenvolvimento de um procedimento computacional inteligente que utiliza *PLN* para ordenação semântica de mensagens eletrônicas, permitindo aos docentes maior rapidez no retorno das mensagens de contexto educacional.

Este artigo está estruturado de acordo com a seguinte ordem: na seção 2, temos a definição do problema. Na seção 3, apresentamos a proposta de solução. Na seção 4, mostramos o procedimento computacional inteligente desenvolvido e os experimentos. Na seção 5 apontamos as dificuldades encontradas. Na seção 6, trazemos as considerações finais, seguidas das referências.

2. Definição do Problema

A ordenação semântica das mensagens recebidas no e-mail, ainda é um desafio. A cada dia recebemos por e-mail dezenas de mensagens que são organizadas por ordem de chegada.

Grande parte dessas mensagens são propagandas ou mesmo informações não relevantes. Com isso nos deparamos com dificuldades originadas pela necessidade de organizar, priorizar, reutilizar e localizar nossas mensagens.

Neste artigo abordaremos apenas uma dessas dificuldades, a de distinguir os e-mails que são de fato importantes e que devem ser tratados com prioridade. Muitos usuários, por exemplo, utilizam uma conta de e-mail para assuntos pessoais e outra para assuntos profissionais, o que aumenta o tempo necessário para esses usuários gerenciarem essas contas.

Nos últimos anos, o uso do e-mail ocupou um lugar de destaque na comunicação humana, pelo baixo custo e simplicidade. Paiva (2002) afirma que as principais vantagens do e-mail são: rapidez, confiabilidade na recepção, envio e encaminhamento de mensagens, facilidade em anexar arquivos e integração com ambientes que compõem um sistema de informação.

Todavia existe uma grande preocupação com relação à priorização das mensagens recebidas. No procedimento tradicional, as mensagens são recebidas por ordem de chegada, o que dificulta identificar as mensagens mais importantes para um usuário do e-mail.

Caberá ao procedimento computacional inteligente encontrar, nos elementos componentes das mensagens, características que apontem tendência da mensagem para uma determinada área. Na abordagem usada aqui, identificaremos em cada mensagem a tendência dela para assuntos educacionais, profissionais e de entretenimento.

3. Proposta de solução

Para desenvolvimento da proposta de solução aplicamos técnica de Processamento de Linguagem Natural, baseada em uma ontologia linguística (*Wordnet*) e em ferramentas como *Lucene*, *Prolog* e no ambiente *JAVA*, com o uso do *IDE Netbeans*.

Também foi aplicado como instrumento de levantamento de dados, um questionário constituído por três questões objetivas. O principal objetivo das questões era descobrir os critérios adotados pelos usuários para selecionar uma mensagem a ser lida primeiro. Observamos que a maioria tomava essa decisão após a leitura de um pequeno trecho da mensagem, analisando a importância de seu conteúdo. Este trecho aparece do lado direito do remetente na maioria dos provedores de webmail e aplicativos de e-mail. O que descobrimos nas respostas desse questionário foi que vários usuários observavam primeiro um trecho que contém parte do conteúdo do e-mail, antes de verificar o remetente do mesmo.

Esse questionário foi aplicado a 20 pessoas com faixa etária de 15 a 50 anos, embora a amostra seja mínima, consideramos de grande utilidade para a validação desta pesquisa, sendo recomendado aumentar a amostra e evoluir as questões analisadas para trabalhos futuros.

A tarefa de processar uma linguagem natural permite que os seres humanos se comuniquem com os computadores da forma mais “natural” possível, usando a linguagem com a qual estão habituados. Elimina-se, dessa maneira, a necessidade de adaptação a formas inusitadas de interação, ou mesmo o aprendizado de uma linguagem artificial, cuja sintaxe costuma ser de difícil aprendizado e domínio (Oliveira 2013).

Oliveira (2013) destaca ainda que o caráter interdisciplinar do Processamento de Linguagem Natural apresenta um nível de complexidade normalmente requerido às abordagens desenvolvidas, as quais usam conceitos de Inteligência Artificial, teoria da computação, compiladores e linguística computacional, dentre outras disciplinas.

Para o Ordenador Semântico foi usada a *API* em Java desenvolvida pelo *The Stanford Natural Language Processing Group*. O trabalho desse grupo da Universidade de *Stanford* favorece a análise de sentenças no idioma inglês, especificamente os substantivos extraídos de mensagens para análise semântica.

Basicamente, as aplicações dos sistemas que tratam a língua natural podem ser divididas em duas classes: aplicações baseadas em texto e aplicações baseadas em diálogos (Oliveira, 2013). Para o procedimento proposto foi usada a aplicação baseada em diálogo, ou seja, na troca de mensagens entre os usuários do e-mail.

A área de Processamento da Linguagem Natural (*PLN*), também denominada Linguística Computacional ou, ainda, Processamento de Línguas Naturais, lida com problemas relacionados à automação da interpretação e da geração da língua humana em aplicações como tradução automática, sumarização automática de textos, ferramentas de auxílio à escrita, perguntas e respostas, categorização textual, recuperação e extração de informação, entre muitas outras (SBC 2013).

Em Bateman et al. (2010) é apresentada uma semântica detalhada para expressões linguísticas de apoio de processamento computacional, aplicada substancialmente sobre os princípios e ferramentas da engenharia ontológica e ontologia formal.

O objetivo de uso do Processamento de Linguagem Natural é fornecer aos aplicativos computacionais a capacidade de entender o texto. Neste sentido, o termo “entender” um texto significa reconhecer o contexto, fazer análise sintática, semântica, léxica e morfológica, criar resumos, extrair informação, interpretar os sentidos e até aprender conceitos com os textos processados.

No procedimento computacional inteligente proposto, o uso desta técnica se restringe à análise semântica.

Nesta proposta de solução, também foi utilizado conforme comentado uma biblioteca em Java denominada *Lucene*. Essa biblioteca é um mecanismo de procura de texto altamente escalável e de *software* livre, disponibilizada pela *Apache Software Foundation*. Sua utilização pode ser feita em aplicativos comerciais e de *software* livre.

O *Lucene* foi utilizado com dois objetivos: o da remoção das *stop words* e da radicalização ou *Stemming* das palavras.

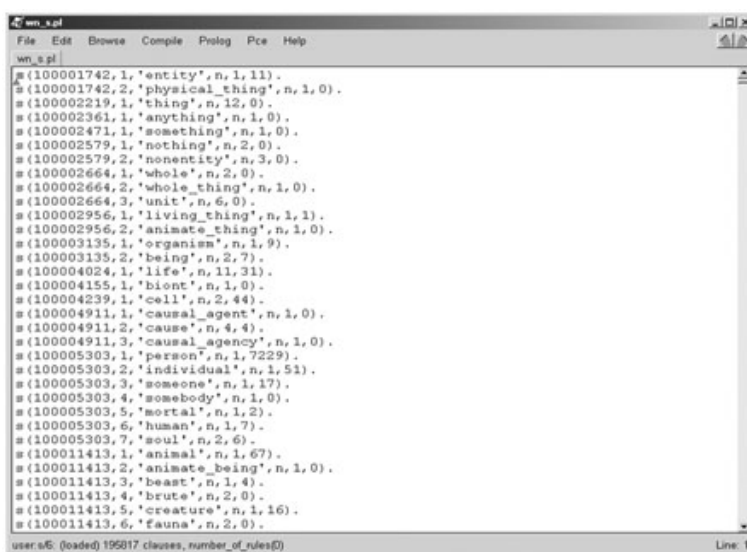
As *stop words* são palavras frequentes em um texto e que não representam nenhuma informação de maior relevância para a extração de palavras-chave. Por exemplo: advérbios, artigos, conjunções, preposições e pronomes.

Sua remoção tem como objetivo eliminar palavras que não são representativas e com isso diminuir o número de palavras a serem analisadas (Foly 2010). Outra técnica utilizada no *Lucene* é o processo de combinar as formas diferentes de uma palavra em uma representação comum, o radical (*stem*).

O radical é o conjunto de caracteres resultante de um processo de radicalização. Este não é necessariamente igual à raiz linguística, mas permite tratar variações diferentes de uma palavra da mesma forma (Foly 2010).

Neste processo de radicalização das palavras é possível combinar a solução *Lucene* + *PLN* com o *Prolog*, e, mais especificamente, com a base de informação disponível na *WordNet*.

O *Prolog* é uma linguagem declarativa que usa lógica de 1ª ordem para representar o conhecimento sobre um dado problema. Um programa em *Prolog* é formado por um conjunto de fatos e regras de inferência (definindo relações entre objetos ou entre objetos e seus atributos). A Figura 1 apresenta a base da *WordNet* de sinônimos.



```
wn.pl
File Edit Browse Compile Prolog Pce Help
wn.pl
#(100001742,1,'entity',n,1,11).
#(100001742,2,'physical_thing',n,1,0).
#(100002219,1,'thing',n,12,0).
#(100002361,1,'anything',n,1,0).
#(100002471,1,'something',n,1,0).
#(100002579,1,'nothing',n,2,0).
#(100002579,2,'nonentity',n,3,0).
#(100002664,1,'whole',n,2,0).
#(100002664,2,'whole_thing',n,1,0).
#(100002664,3,'unit',n,6,0).
#(100002956,1,'living_thing',n,1,1).
#(100002956,2,'animate_thing',n,1,0).
#(100003135,1,'organism',n,1,9).
#(100003135,2,'being',n,2,7).
#(100004024,1,'life',n,11,31).
#(100004155,1,'biota',n,1,0).
#(100004239,1,'cell',n,2,44).
#(100004911,1,'causal_agent',n,1,0).
#(100004911,2,'cause',n,4,4).
#(100004911,3,'causal_agency',n,1,0).
#(100005303,1,'person',n,1,7229).
#(100005303,2,'individual',n,1,51).
#(100005303,3,'someone',n,1,17).
#(100005303,4,'somebody',n,1,0).
#(100005303,5,'mortal',n,1,2).
#(100005303,6,'human',n,1,7).
#(100005303,7,'soul',n,2,6).
#(100011413,1,'animal',n,1,67).
#(100011413,2,'animate_being',n,1,0).
#(100011413,3,'beast',n,1,4).
#(100011413,4,'brute',n,2,0).
#(100011413,5,'creature',n,1,16).
#(100011413,6,'fauna',n,2,0).
user:s6: (loaded) 195817 clauses, number_of_rules(0) Line: 1
```

Figura 1. Base de sinônimos da *WordNet*.

Esse conjunto, chamado de base de conhecimento, descreve o domínio de um dado problema. A execução de um programa em *Prolog* consiste na dedução de consequências lógicas da base de conhecimento.

No procedimento desenvolvido usamos uma *API*, em *Java*, do *SWI-Prolog*, para buscar os sinônimos relacionados aos substantivos, com o objetivo de encontrar a relação da palavra pesquisada com o contexto (educacional, profissional e entretenimento).

A *WordNet* é uma base de dados de conhecimento *lexical*. É um grande banco de dados *léxico* do inglês. Substantivos, verbos, adjetivos e advérbios são agrupados em conjuntos de sinônimos cognitivos (*synsets*), cada um expressando um conceito distinto (*WordNet* 2013).

É importante reforçar que os *Synsets* estão interligados por meio de relações semânticas e conceitual-lexical e que a *WordNet* também é livre e publicamente disponível para *Download*.

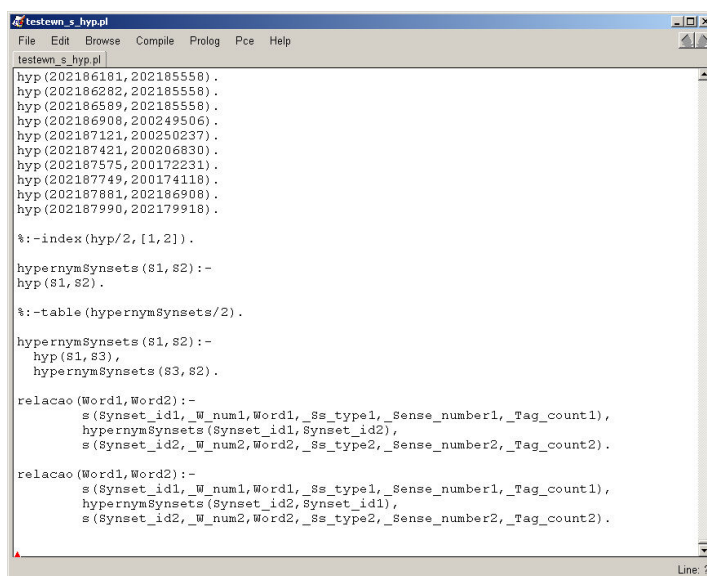
Para Budanitsky e Hirst (2006) a quantificação de parentesco semântico dos itens léxicos tem muitas aplicações em *PLN*. Eles avaliaram e mediram cinco

aplicações, que usaram a *WordNet* como seu recurso central, comparando o seu desempenho na detecção e correção de erros de ortografia de palavras em um texto.

A estrutura da *WordNet* torna-se uma ferramenta útil para o Processamento de Linguagem Natural. A *WordNet*, superficialmente, se assemelha a um dicionário, onde estão agrupadas palavras de acordo com seus significados.

No entanto, há algumas diferenças importantes: 1) articula não só a palavra, mas sentidos específicos das palavras. Como resultado, as palavras que se encontram na proximidade uma da outra na rede são semanticamente diferenciadas; 2) rotula as relações semânticas entre as palavras, enquanto que os agrupamentos de palavras em um dicionário de sinônimos não segue nenhum padrão explícito que não signifique semelhança.

Para melhorar a qualidade da classificação, também foi utilizado o hiperônimo além dos sinônimos. Hiperônimo é uma palavra que pertence ao mesmo campo semântico de outra, mas com o sentido mais abrangente, podendo ter várias possibilidades para um único hiperônimo. Por exemplo, a palavra flor está associada a todos os tipos de flores: rosa, dália, violeta, etc., conforme mostra a Figura 2.



```
testewn_s_hyp.pl
File Edit Browse Compile Prolog Pce Help
testewn_s_hyp.pl
hyp(202186181,202185558).
hyp(202186282,202185558).
hyp(202186589,202185558).
hyp(202186908,200249506).
hyp(202187121,200250237).
hyp(202187421,200206830).
hyp(202187575,200172231).
hyp(202187749,200174118).
hyp(202187881,202186908).
hyp(202187990,202179918).

%:-index(hyp/2,[1,2]).

hypernymSynsets(S1,S2):-
hyp(S1,S2).

%:-table(hypernymSynsets/2).

hypernymSynsets(S1,S2):-
hyp(S1,S3),
hypernymSynsets(S3,S2).

relacao(Word1,Word2):-
s(Synset_id1,_W_num1,Word1,_Ss_type1,_Sense_number1,_Tag_count1),
hypernymSynsets(Synset_id1,Synset_id2),
s(Synset_id2,_W_num2,Word2,_Ss_type2,_Sense_number2,_Tag_count2).

relacao(Word1,Word2):-
s(Synset_id1,_W_num1,Word1,_Ss_type1,_Sense_number1,_Tag_count1),
hypernymSynsets(Synset_id2,Synset_id1),
s(Synset_id2,_W_num2,Word2,_Ss_type2,_Sense_number2,_Tag_count2).

Line: ?
```

Figura 2. Regras Hiperônimos.

Também na Figura 2 é ilustrada a definição das regras para buscar em profundidade os substantivos, com o intuito de encontrar a relação da palavra com os contextos abordados no procedimento computacional inteligente.

4. Protótipo Computacional

O procedimento computacional foi desenvolvido em *Java*, usando o *Netbeans IDE 7.2.1*, além das *APIs* descritas anteriormente (*PLN da Universidade de Stanford, Lucene, Prolog e WordNet*).

O procedimento computacional inteligente apresentado segue a estrutura ilustrada na Figura 3.

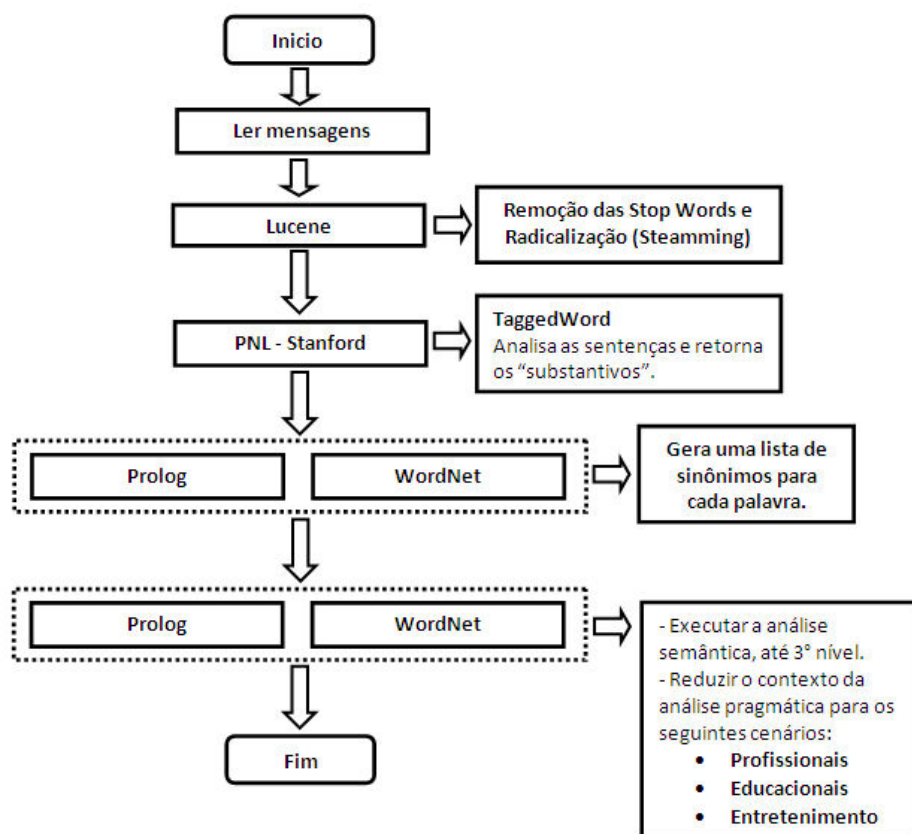


Figura 3. Diagrama do Procedimento computacional inteligente.

O procedimento computacional inteligente recebe uma mensagem texto cujo conteúdo é apresentado na Figura 4. Abordaremos uma mensagem de contexto educacional neste experimento apresentado.

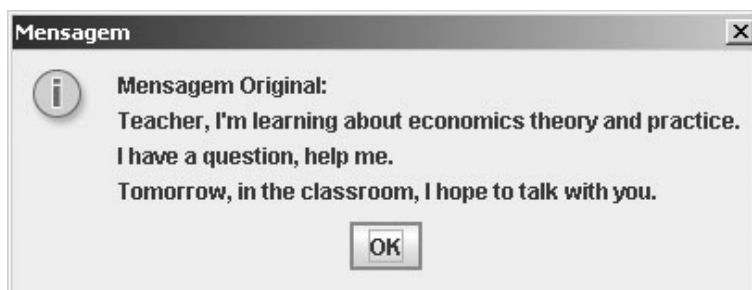


Figura 4. Mensagem original retirada do e-mail recebido.

Em seguida retira as *stop words* e executa o processo de *stemming*, conforme ilustra a Figura 5.

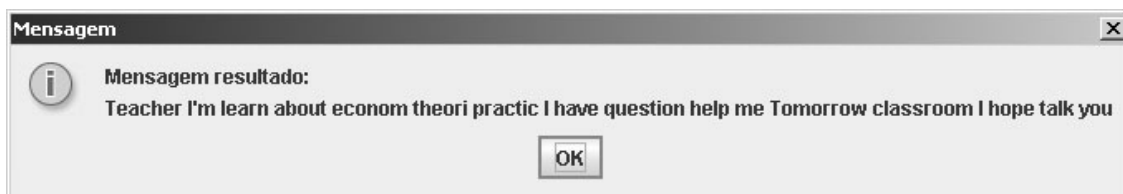


Figura 5. Remoção das *stop words* e *stemming*.

Logo após, são retiradas do texto da mensagem somente os substantivos utilizando as *tags* NN – Substantivo no singular e NNS – Substantivo no plural, PNN – nome próprio no singular e NNPS – nome próprio no plural, conforme ilustra a Figura 6, o restante das palavras que não estão neste conjunto não serão analisadas.

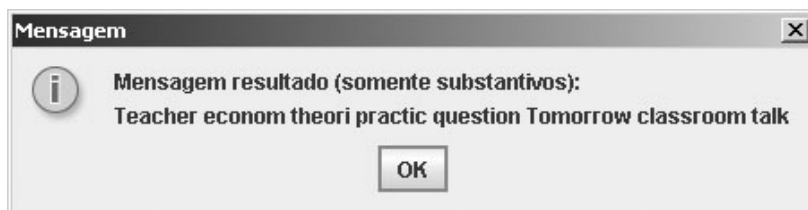


Figura 6. Somente os substantivos.

Note que o processo de radicalização do *Lucene* trouxe algumas palavras com erros, é o caso de *economics*, *theori* e *practical* que retornaram com falhas. Consequente estas palavras não serão encontradas na base de conhecimento da *Wordnet*.

Em seguida, conforme ilustrado na Figura 7, o procedimento computacional apresenta os substantivos, retirados anteriormente, e os sinônimos coletados na base de dados da *WordNet* usando o *Prolog*.

Palavra: Teacher	Palavra: question	Palavra: Tomorrow
Palavra: econom	sinonimo: doubt	Palavra: classroom
Palavra: theori	sinonimo: doubtfulness	sinonimo: schoolroom
Palavra: practic	sinonimo: dubiousness	Palavra: talk
Palavra: question	sinonimo: enquiry	sinonimo: babble
Palavra: Tomorrow	sinonimo: head	sinonimo: blab
Palavra: classroom	sinonimo: inquiry	sinonimo: lecture
Palavra: talk	sinonimo: interrogate	sinonimo: mouth
	sinonimo: interrogation	sinonimo: peach
	sinonimo: interrogative	sinonimo: sing
	sinonimo: interview	sinonimo: speak
	sinonimo: motion	sinonimo: talking
	sinonimo: oppugn	sinonimo: tattle
	sinonimo: query	sinonimo: utter
	sinonimo: wonder	sinonimo: verbalise
		sinonimo: verbalize

Figura 7. Substantivos e sinônimos – Wordnet e Prolog.

Ao final do processamento, o procedimento apresenta uma mensagem informando o contexto do e-mail, neste exemplo tem contexto educacional, se o usuário parametrizar o protótipo para priorizar os e-mails deste contexto, este e-mail teria maior prioridade para a leitura, ou seja, entraria no grupo de mensagens prioritárias de forma automática, com base no conteúdo do corpo da mensagem, conforme mostra a Figura 8.

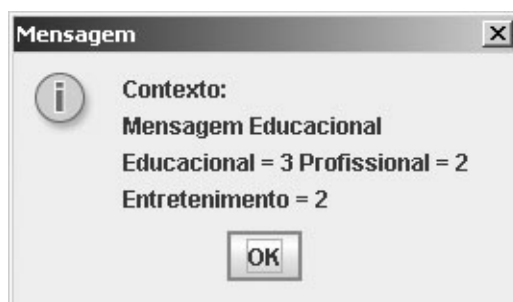


Figura 8. Resultado para esta mensagem – Contexto Educacional.

Outros experimentos semelhantes foram realizados para validar a eficácia do procedimento computacional inteligente para identificar diferentes mensagens recebidas como pertencentes aos contextos educacionais, profissionais e de entretenimento.

5. Dificuldades Encontradas

Durante o período em que se deu a pesquisa, as seguintes dificuldades foram encontradas:

- tempo excessivo de processamento para algumas mensagens: algumas mensagens demoraram cerca de 10 minutos para serem processadas em um micro-computador *Celeron* com 1 GB de memória *RAM* e processador 1.6 GHz;
- palavras compostas não analisadas: cada palavra foi analisada separadamente e as palavras compostas não foram analisadas, por exemplo: inteligência artificial seria processada de forma separada o que não é adequado;
- problema com a radicalização de algumas palavras: encontramos algumas palavras, apresentadas neste artigo, que o *Lucene* retornou de forma incorreta, porém, uma grande quantidade de palavras apresentou resultado correto, como no exemplo, *learning* que resultou em *learn*;
- no contexto educacional: observamos que diversos termos comumente utilizados por brasileiros, que estariam relacionados na área de educação, por exemplo, escola, professor, disciplina, entre outros, na base da *WordNet* estão relacionados na área profissional e não educacional.

6. Considerações Finais

A ordenação semântica de mensagens, em um correio eletrônico, tem se tornado um grande desafio, para os pesquisadores e profissionais da área, isso porque cada indivíduo possui um perfil diferenciado para a ordenação de mensagens.

Nos testes realizados observamos que várias mensagens obtiveram uma classificação indefinida, com tendência para assuntos profissionais e educacionais empatados. Esta incerteza exige o apoio do usuário para desempatar.

Para aplicar nas mensagens os testes realizados em língua portuguesa, é necessário aguardarmos uma maior consistência da base de dados da *WordNet BR*, que está em crescente expansão. Por isso, o procedimento computacional inteligente foi parametrizado para dar suporte nesses futuros testes.

Deve-se destacar que o questionário utilizado na pesquisa possui caráter orientativo, resultando em indicadores básicos para o uso das mensagens eletrônicas e que os mesmos precisam ser melhorados.

Este artigo apresentou a descrição de um procedimento computacional inteligente que permite a classificação e priorização de mensagens utilizando Processamento de Linguagem Natural. Os próximos passos são: recurso de interação com o usuário para resolver situações de incerteza (por exemplo, empate no peso semântico de assuntos); a implementação deste procedimento em um servidor de mensagens (*pop3* e *smtp*), além do incremento da base *WordNet* (para domínios de conhecimento específicos) para facilitar a contextualização dos experimentos ampliando a base de testes, além de experimentos usando a base da *WordNet BR*.

7. Referências

- Araujo, Vanessa Leonardo. (2006) “Um Agente de Interface em Linguagem Natural para Recuperação de Informações na Web Semântica”, in: Congresso de Pesquisa, Ensino e Extensão da UFG - Conpeex, 2, Goiânia.
- Bateman, John A.; Hois, Joana; Ross, Robert; Tenbrink, Thora. (2010) “A linguistic ontology of space for natural language processing”. *Artificial Intelligence*, Volume 174, Issue 14, September, Pages 1027-1071.
- Borges, Thyago Bohrer e Lima, Vera Lúcia Strube de. (2008) “Aplicação de uma Técnica Tradicional de Expansão de Consulta ao Modelo TR+”, Evento Integração PLN. PUC - Rio Grande do Sul.
- Budanitsky, Alexander; Hirst, Graeme. (2006) “Evaluating WordNet-based Measures of Lexical Semantic Relatedness”, *Massachusetts Institute of Technology*. March, Vol. 32, No. 1, Pages 13-47.
- Foly, L. S. (2010) “Uma Arquitetura para Aferir o Conhecimento do Aluno em um Ambiente de Aprendizagem”, *Dissertação de Mestrado*. Universidade Candido Mendes – Campos dos Goytacazes – RJ.
- Fontana, Adas Pavei; Mattos, Merisandra Côrtes de; Simões, Priscyla Waleska Targino de Azevedo; Ramos, Rozenir; Cechinel, Cristian. (2007) “Aplicação de Processamento de Linguagem Natural em Interface Médica”, *Grupo de Pesquisa em Inteligência Computacional Aplicada*. UNESC e UNIPAMPA/Bagé.
- Jonassena, David; Davidsonb, Mark; Collinsc, Mauri, Campbelle John; Haagc, Brenda Bannan. (2009) “Constructivism and computer-mediated communication in distance education”, *American Journal of Distance Education*.
- Mantovani, Daielly Melina Nassif; Viana, Adriana Backx Noronha; Gouvêa, Maria Aparecida. (2011) “Métodos para implementação e acompanhamento de atividades on-line em disciplinas de estatística aplicada à administração”, *Revista Eletrônica de Educação*. São Carlos, SP: UFSCar, v. 5, no. 2, p. 236-255.
- Oliveira, F. A. D. (2013) “Processamento de linguagem natural”, <http://www.inf.ufrgs.br/gppd/disc/cmp135/trabs/992/Parser/parser.html>. Janeiro.
- Paiva, M. A. P. (2002) “O monitoramento do correio eletrônico no ambiente de trabalho”, in *conferencia proferida no Congresso Internacional de Direito e Tecnologias da Informação*, Brasília.
- SBC, Sociedade Brasileira de Computação. (2013) “Processamento de Linguagem Natural”, http://www.sbc.org.br/index.php?option=com_content&view=category&layout=blog&id=50&Itemid=71. Abril.
- WordNet. (2013) “About WordNet”, <http://wordnet.princeton.edu/>. Janeiro.