

Um Agente de Interface em Linguagem Natural para Recuperação de Informações na Web Semântica

Araujo, Vanessa Leonardo¹; **Carvalho, Cedric Luiz de**².

Palavras-chave: processamento de linguagem natural, ferramenta GATE.

1. INTRODUÇÃO (justificativa e objetivos)

A linguagem natural é a linguagem mais usada na comunicação entre seres humanos tanto na forma oral como escrita, como o português, o inglês, o espanhol, entre outros. Esta linguagem é um dos meios de comunicação mais utilizados em todo o mundo. O processamento de linguagem natural é uma técnica muito importante para “aproximar” o homem e o computador – visto que a linguagem de máquina – entendida por computadores - é complexa e, conseqüentemente, dificulta a comunicação entre eles. [1]

Alguns pesquisadores acham que o processamento de linguagem natural é um dos objetivos mais importantes da Inteligência Artificial (IA), pois quando este for alcançado, o computador entenderá a linguagem humana e, assim, será desnecessário que engenheiros de software o programem. Na verdade, a importância desta área foi observada há muitos anos atrás por Allan Turing, que relacionava a inteligência dos computadores com a capacidade destes em lidar com a linguagem natural. [2] [3]

Para que o computador possa realizar algumas tarefas humanas, é necessário que ele entenda e utilize a linguagem natural. Ainda não existe um software com desempenho lingüístico suficiente para competir com pessoas. Isto se deve ao fato de que a compreensão da linguagem natural é difícil, já que toda língua humana é complexa e ambígua; e é difícil fazer com que o computador entenda informações contextuais presentes em todas as situações. Esta compreensão requer não somente o conhecimento lingüístico de uma determinada língua – seja ela o português, o inglês, o espanhol, etc. – mas também o conhecimento do mundo relacionado ao assunto utilizado, isto é, o domínio. [4]

Os computadores entendem e executam comandos escritos em linguagens de programação, como Java, Basic, C, Prolog, etc.; mas eles não conseguem compreender instruções escritas em linguagem natural. Isto ocorre porque as linguagens de programação são bastante precisas, e possuem regras fixas e estruturas lógicas bem definidas de modo que o computador entenda exatamente o que deve ser feito. Já as linguagens naturais, são ambíguas e complexas; e as interpretações das frases dependem do contexto da situação no momento, do conhecimento do mundo, das regras gramaticais e culturais.

A proposta inicial deste projeto era a construção de um agente que fosse capaz de “dialogar” com o usuário, utilizando linguagem natural, e identificar o contexto em que o mesmo deseja realizar sua pesquisa. Ao longo do projeto estes objetivos foram modificados, sendo o foco principal a construção de um software que processasse uma frase ou um texto em português, isto é, obtendo informações como tokens (cada palavra do texto), sentenças, organização, título, árvore sintática, entre outras, utilizando as bibliotecas disponibilizadas pelo arcabouço GATE de Processamento de Linguagem Natural.

2. METODOLOGIA

A metodologia adotada para realização deste projeto foi constituída basicamente de duas fases. A primeira envolveu o estudo teórico de assuntos relacionados com o processamento de linguagem natural como as etapas necessárias de um processamento de linguagem natural, aplicações, ferramentas mais utilizadas nesta área (principalmente o GATE que foi usado neste projeto), entre outros. A segunda fase foi constituída da implementação de um software que processa uma frase em português usando a ferramenta GATE.

3. RESULTADOS E DISCUSSÃO

Este projeto tem como resultado um estudo sobre o GATE (encontrado em [5]), descrições de seu ambiente gráfico e aplicações da ferramenta, bem como suas principais funcionalidades e alguns conceitos importantes.

Foi construída uma implementação que usa a API GATE, que recebe um arquivo como entrada e realiza várias funções de processamento de linguagem natural como a obtenção de tokens, sentenças, entre outras.

O GATE é considerado como uma infra-estrutura para a Engenharia da Língua. Esta ferramenta tem sido usada com sucesso como um ambiente que ensina o processamento de linguagem natural, e também como um veículo para a construção de aplicações de processamento de linguagem natural e componentes de software reusáveis.

O GATE possui um ambiente de desenvolvimento gráfico (ou GUI) que facilita aos estudantes inspecionarem resultados do processamento de linguagem e eliminarem erros dos módulos. O ambiente tem facilidades na visualização de documentos, corpora, ontologias (incluindo o editor Protegè), e dados lingüísticos expressados como anotações.

A figura a seguir mostra alguns recursos oferecidos pelo ambiente gráfico do GATE, ela mostra os os tokens (cada palavra, símbolo, números, etc.) destacados no texto que está sendo processado.

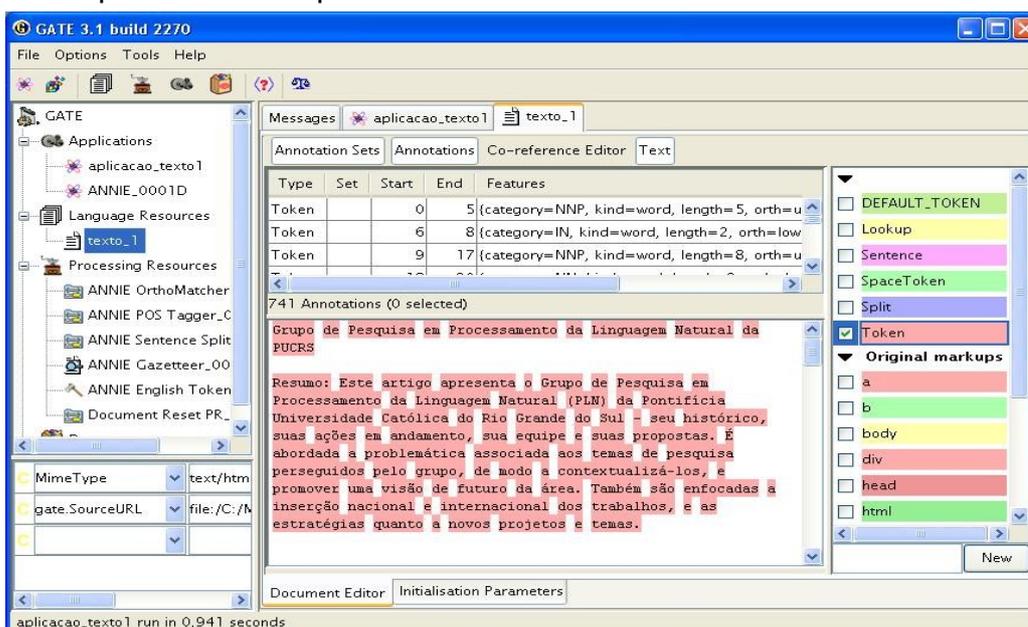


Figura 1: Ambiente de Desenvolvimento Gráfico do GATE – mostra os tokens do texto que está sendo processado

4. CONCLUSÃO

O processamento de linguagem natural é muito complexo, devido à linguagem natural ser inerentemente ambígua. O português é mais difícil de ser analisado que a língua inglesa devido alguns aspectos, por exemplo, a conjugação dos verbos.

Existem vários projetos que “tentam” processar a língua humana. Infelizmente ainda não se tem um software capaz de realizar tal operação. Mas existem vários softwares que realizam algum tipo de processamento, por exemplo para realizar uma busca feita em linguagem natural como Google; robôs que conversam com humanos, ressaltando que essas conversas são difíceis porque os robôs erram muito; entre outras aplicações.

A área de processamento de linguagem natural é bastante inovadora e revolucionária, pois facilitaria a inclusão digital no mundo todo, já que não seria necessário uma pessoa entender alguma linguagem de programação.

A ferramenta GATE estudada neste projeto é muito utilizada em vários projetos em todo o mundo. Ela permite o processamento de várias linguagens como o português, o inglês, o chinês, entre várias outras. Apesar da documentação desta ferramentas não ser muito simples e organizada, ela realiza funções fundamentais no processamento de linguagem natural. Utilizando-a é possível se preocupar com outros aspectos relevantes de um projeto, já que não será necessário construir funções que já existem e que podem ser usadas facilmente devido o Gate disponibilizar seu código.

5. REFERÊNCIAS BIBLIOGRÁFICAS

[1] <http://www.di.uminho.pt/~jas/teaching>. , João Saraiva <jas@di.uminho.pt>, acessado em setembro de 2005.

[2] Schildt H. *Inteligência Artificial – Utilizando a linguagem C*, tradução: Cláudio Gaiger Silveira e Mônica Soares Rufino, revisão técnica: José Carlos Barbosa Damski e André de Souza Mello, Editora Mc-Graw-Hill, 1989.

[3] Vieira R. e Strube de Lima V. L. *Linguística Computacional: Princípios e Aplicações*. In: *Anais do XXI Congresso da Sociedade Brasileira de Computação*. Fortaleza, 30 de julho a 03 de agosto, 2001.

[4] Rich E. *Inteligência Artificial*, Tradução: Newton Vasconcellos, revisão técnica: Nizam Omar. São Paulo, Editora McGraw-Hill, 1988.

[5] <http://gate.ac.uk>, página principal da ferramenta GATE, acessado em setembro de 2005.

[6] Russel S. e Norvig P. *Inteligência Artificial – tradução da segunda edição*, Rio de Janeiro, Elsevier. Editora Ltda., 2004.

¹ Bolsista de iniciação científica. Instituto de Informática – vanessae@hotmail.com

² Orientador/Instituto de Informática/UFG, cedric@inf.ufg.br