

Processamento de Linguagem Natural e suas Aplicações Computacionais

João Mendes de Oliveira Neto, Sávio Duarte Tonin, Soraia Silva Prietch

Universidade Federal de Mato Grosso - Campus Universitário de Rondonópolis
Instituto de Ciências Exatas e Naturais - Curso de Licenciatura Plena em Informática
Rondonópolis – MT – Brasil

{lord.edy,savio.roo}@gmail.com, soraia@ufmt.br

***Abstract.** This article describes the steps of Natural Language Processing raised the key features necessary for an implementation of this mechanism, and exemplifies some of its applications, demonstrating the potential of this branch of artificial intelligence that is increasingly present in everyday users' computers, especially the Internet, have since said that search applications by a large majority is dedicated to the user's global network of computers.*

***Resumo.** Este artigo descreve as etapas do Processamento de Linguagem Natural levantando as principais características necessárias para uma implementação desse mecanismo, bem como exemplifica algumas de suas aplicações, demonstrando o potencial desse ramo da Inteligência Artificial que está cada dia mais presente no cotidiano dos usuários de computadores, em especial dos internautas, haja vista que pelas aplicações citadas nessa pesquisa, uma grande maioria é voltada para os usuários da rede mundial de computadores.*

Palavras chave: Linguagem Natural, Comunicação Homem-Máquina, *Chatterbots*.

1. Introdução

Atualmente, muitos usuários de computadores estão familiarizados com vários produtos comerciais, cuja função é processar textos escritos: editores de textos, sistemas de busca de páginas na *web*, sistemas de ajuda à tradução etc. Segundo Laporte (2000) *apud* Smarsaro (2004), esses exemplos de sistemas computacionais estão disponíveis e é fácil constatar que, embora sejam úteis, o seu desempenho ainda não é satisfatório. Os melhores editores de textos apontam erros em palavras corretas, propõem correções erradas e deixam de detectar certos tipos de erros ortográficos. Os sistemas de busca na *web* selecionam, às vezes, dezenas de páginas sem qualquer relação com o assunto pesquisado pelo usuário, mesmo que este expresse seu objetivo de forma suficientemente precisa. Até os textos produzidos pelos melhores sistemas de ajuda à tradução necessitam de um sistema para tradutores humanos, devido a erros de tradução que, aliás, tornam os resultados da tradução automática quase um gênero literário cômico.

A comunicação é um fenômeno tão generalizado que se torna difícil estabelecer uma definição exata. Em geral, comunicação é uma forma intencional de trocar informações através da produção e percepção de sinais a partir de um sistema

convencional. Os seres humanos desenvolveram um complexo e estruturado sistema de sinais conhecido como Linguagem Natural, é através desse sistema que os seres humanos se comunicam. A tarefa de processar uma linguagem natural permite que os seres humanos comuniquem-se com os computadores da forma mais "natural" possível, utilizando a linguagem com a qual mais estão acostumados. Eliminando-se, 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 [Russel e Norvig, 1995].

Esse artigo busca então mostrar através de um levantamento bibliográfico, quais são os mecanismos que possibilitam a comunicação entre seres humanos e computadores, utilizando-se de linguagem natural para esse processo. Por são apresentadas algumas aplicações que já se utilizam dessa conhecida técnica de inteligência artificial conhecida como, Processamento de Linguagem Natural.

1.1. Justificativa

Segundo Gariba [2005] *et al.*, o processamento de Linguagem Natural vem a facilitar a interação do software (através de sua interface) com o usuário, para que se torne mais fácil a comunicação e a passagem de conhecimento, assim quem fizer o uso de um software, possa compreender o que ele tem a oferecer e consiga saber o que o usuário está necessitando. Utilizando a linguagem natural torna-se mais simples o questionamento de uma determinada área, já que não há necessidade de se saber corretamente a implementação do sistema, o que ele irá buscar, como por exemplo, em uma consulta a um banco de dados, o usuário não precisa saber o que são tabelas e nem como elas buscam as informações, e nem o funcionamento de um banco de dados, ele apenas deseja que o resultado da pesquisa seja mostrado de forma simples e objetiva.

1.2. Objetivos

O trabalho tem como objetivo geral buscar e descrever algumas aplicações que utilizam em sua implementação a Linguagem Natural.

Os objetivos específicos definidos pelo trabalho são: mostrar de forma teórica como é utilizada a linguagem natural, através das análises: morfológica, sintática, semântica; fazer um apontamento de forma teórica e descrever algumas aplicações em que são utilizadas técnicas de linguagem natural.

2. Metodologia de Pesquisa

Este trabalho utiliza como metodologia de pesquisa quanto aos procedimentos técnicos a pesquisa bibliográfica, em que podemos fazer o levantamento de materiais impressos ou digitais na forma de livros e ou artigos, tendo uma abordagem qualitativa com um objetivo descritivo [Silva, 2004]. Sendo assim necessário para que possamos fazer o devido enriquecimento do trabalho e ter uma base sólida do tema, buscando incluir de forma clara e objetiva o assunto ao qual refere-se o trabalho.

3. Revisão de literatura

Segundo Oliveira [2002], as aplicações que tratam de Linguagem Natural consistem basicamente em dois segmentos:

- Aplicações baseadas em texto: são sistemas que procuram documentos específicos em uma base de dados (exemplo: encontrar livros relevantes em uma biblioteca), tradutores de documentos, e sistemas que resumem textos (exemplo: produzir 3 páginas resumidas de um livro de 100 páginas);
- Aplicações baseadas em diálogos: por serem baseadas em diálogos, referem-se às interfaces de linguagem natural para bancos de dados, os sistemas tutores e os sistemas que interpretam e respondem a comandos expressados em linguagem escrita ou falada.

No início da década de 50, começaram a surgir os primeiros trabalhos de pesquisa concernentes ao tratamento informatizado da língua natural [Santos [2003] *apud* Silva e Lima [2007]].

Feigenbaum [1981] *apud* Hübner [1992], afirma que quatro categorias históricas de programas para interpretação de linguagem natural podem ser identificadas, sendo elas:

- Alguns programas que tinham por objetivo a geração de um número reduzido de resultados em domínios específicos. A simplicidade do processo permitia que muitos dos problemas da linguagem natural pudessem ser ignorados;
- Em alguns dos primeiros sistemas, era armazenada uma representação do texto, recomendando-se a engenhos de indexação para auxiliar a recuperação de determinadas palavras ou frases. Como o texto armazenado podia cobrir qualquer assunto, os sistemas não eram restritos a um determinado domínio. Todavia, esses sistemas eram semanticamente fracos e não tinham poderes dedutivos;
- Sistemas de lógica limitada que tinham por objetivo traduzir frases de entrada para a notação formal usada na base de dados. Aqui a intenção era permitir que se fizessem deduções a partir da informação mantida na base de dados, mesmo que somente alguns dos processos utilizados na conversação do dia-a-dia pudessem ser explorados;
- Os sistemas com base de conhecimento, que se utilizavam da informação sobre um assunto específico para compreender as frases de entrada. Tais programas, alguns dos quais constituíam sistemas especialistas, exibiam vários poderes dedutivos.

Segundo Ranchhod [2001] *apud* Smarsaro [2004], nos últimos anos, tornou-se evidente que os recursos lingüísticos e, em particular os recursos lexicais, são a pedra de toque de qualquer sistema de processamento de linguagem natural. Na verdade, a crescente necessidade de aplicações da lingüística computacional fez ressaltar a carência de dados lingüísticos de dimensões reais, em particular, de léxicos e gramáticas de grande cobertura. Para atender a essas exigências de qualidade, o tratamento automático requer das línguas uma descrição sistemática e o mais completa possível, pois a insuficiência de dados lingüísticos pode gerar falhas no processamento automático. Ao falarmos de processamento de linguagem natural, surgem vários problemas, como a

grande variação morfológica e sintática das unidades lexicais ou a ambigüidade intrínseca da língua portuguesa.

Para resolver os problemas acima citados, destacam-se três níveis de análise lingüística: morfológico, sintático e semântico. Para cada nível associam-se descrições lexicais apropriadas. Se, em relação ao tratamento do léxico, os dicionários utilizados pelos sistemas de processamento não forem adequados, quer do ponto de vista da sua cobertura lexical, quer, do ponto de vista da formalização e sistematização da informação lingüística, isso afetará não só a análise lexical de um determinado texto, mas também todas as fases de processamento subseqüentes. Se, por exemplo, uma palavra não for reconhecida ou não for corretamente identificada, a análise sintática da frase ou da estrutura em que ela se encontra não poderá ser feita. O léxico surge, portanto, como um componente de grande importância em qualquer sistema de processamento automático da linguagem natural [Smarsaro, 2004].

Silva e Lima [2007], um sistema computacional interpreta uma sentença em linguagem natural, através da análise de informações morfológicas (léxicas), sintáticas (regras gramaticais) e semânticas (significados), armazenadas em um dicionário, juntamente com as palavras que o sistema compreende.

3.1. Análise Morfológica

Segundo Oliveira [2002], o analisador morfológico identifica palavras ou expressões isoladas em uma sentença, sendo este processo auxiliado por delimitadores (pontuação e espaços em branco). As palavras identificadas são classificadas de acordo com seu tipo de uso ou, em linguagem natural, categoria gramatical.

Neste contexto, uma instância de uma palavra em uma sentença gramaticalmente válida pode ser substituída por outra do mesmo tipo, configurando uma sentença ainda válida (exemplo: substantivos, pronomes, verbos, etc.). Dentro de um mesmo tipo de palavra, existem grupos de regras que caracterizam o comportamento de um subconjunto de vocábulos da linguagem (exemplo: formação do plural de substantivos terminados em "ão", flexões dos verbos regulares terminados em "ar", etc.). Assim, a morfologia trata as palavras quanto a sua estrutura, forma, flexão e classificação, no que se refere a cada um dos tipos de palavras [Oliveira, 2002].

Ainda Oliveira [2002], defende que o emprego do analisador morfológico é fundamental para a compreensão de uma frase, pois para formar uma estrutura coerente de uma sentença, é necessário compreender o significado de cada uma das palavras componentes. Ele ainda cita que um autômato finito é um exemplo de reconhecedor utilizado para a análise morfológica.

3.2. Análise Sintática

Enquanto o analisador léxico-morfológico lida com a estrutura das palavras e com a classificação das mesmas em diferentes categorias, o analisador sintático trabalha em nível de agrupamento de palavras, analisando a constituição das frases.

Segundo Gonzalez e Lima [2003], a análise sintática (*parsing*) é o procedimento que avalia os vários modos de como combinar regras gramaticais, com a finalidade de gerar uma estrutura de árvore que represente a estrutura sintática da sentença analisada.

Se a sentença for ambígua, o analisador sintático (*parser*) irá obter todas as possíveis estruturas sintáticas que a representam.

O papel do processamento sintático varia em importância [Nunes [1999] *apud* Gonzalez e Lima [2003]]. Ele, tradicionalmente, ocupa posição de destaque, com a semântica sendo considerada uma interpretação da sintaxe, mas, também, pode ser considerado em posição secundária, de acordo com os pesquisadores denominados semântico-gerativistas. Neste último caso, a sintaxe é uma projeção da semântica. Entretanto, qualquer que seja a visão adotada, o processamento sintático é uma etapa indispensável para viabilizar o processamento semântico que passamos a discutir a seguir.

3.3. Análise Semântica

Enquanto a sintaxe corresponde ao estudo de como as palavras agrupam-se para formar estruturas em nível de sentença, a semântica está relacionada ao significado, não só de cada palavra, mas também do conjunto resultante delas. O processamento semântico é considerado um dos maiores desafios do Processamento de Linguagem Natural, pois se vincula, de um lado, com a morfologia e a estrutura sintática e, de outro lado em alguns casos, com informações da pragmática [Saint-Dizier [1999] *apud* Gonzalez e Lima [2003]].

O analisador semântico analisa o sentido das estruturas das palavras que foram reagrupadas pelo analisador sintático, uma vez que o analisador morfológico permitiu identificar estas palavras individualmente.

Segundo Oliveira [2002], pode-se dizer que a semântica se desdobra em semântica léxica e em semântica gramatical. A semântica léxica busca uma representação conceitual para descrever o sentido, sendo que, para construir esta representação, pode ser feita a decomposição semântica das unidades léxicas (em primitivas ou em traços semânticos), ou podem ser utilizadas redes semânticas. Esta última forma de representação foi originada da psicologia e leva em conta a forma como os seres humanos categorizam e memorizam conceitos.

3.4. Pragmática

À medida que se avança no processamento da Linguagem Natural é necessário fazer uma interpretação do todo e não mais analisar o significado de suas partes, como ocorre na análise das informações morfológicas (léxicas), sintáticas (regras gramaticais) e semânticas (significados) [Smarsaro, 2004].

As palavras podem se associar através de dois tipos de relações: paradigmáticas e sintagmáticas. As relações paradigmáticas associam palavras através do significado, como “nadar” e “água”. As relações sintagmáticas conectam palavras que são freqüentemente encontradas no mesmo discurso, como “água” e “poça” [Gonzalez e Lima, 2003].

4. Aplicações

O Processamento de Linguagem Natural pode ser utilizado em diversas aplicações, tal uso justifica a compreensão desta ampla área. A seguir veremos algumas aplicações que se utilizam dessa técnica de Inteligência Artificial.

4.1. Consulta em Banco de Dados Utilizando Linguagem Natural

Podemos usar Linguagem Natural em consultas à Banco de Dados, de forma que o usuário não precisa conhecer especificamente a estrutura, funcionamento e nem a linguagem a ser utilizada para a realização de consulta neste Banco de Dados. Sem a utilização de Linguagem Natural o usuário precisa conhecer todos os dados referentes à implementação do Banco de Dados, ou seja, a sua estrutura e a linguagem que por ele possa ser usada, para realizar uma consulta entre os campos. Com o devido uso de Linguagem Natural, é apenas necessário conhecer a área em que o Banco de Dados foi desenvolvido, para se fazer as devidas formulações para a realização de pesquisas internas em sua base. Se alguma consulta através de Processamento Linguagem Natural for feita de maneira errada ou inexata, a Linguagem Natural busca por termos próximos ou semelhantes do correto [Silva e Lima, 2007].

Segundo Silva e Lima [2007], com o uso de um tradutor de Linguagem Natural, o usuário poderá criar suas próprias consultas em linguagem natural, assim o tradutor verificará a sentença e a transformará em uma consulta SQL. Após a devida definição da gramática a ser utilizada, bem como o vocabulário, a implementação na análise sintática é preciso ser realizada, o qual o PROLOG é o software utilizado para que isto possa ocorrer.

Silva e Lima [2007], afirmam ainda, que poderá ocorrer algum tipo de resistência ou até mesmo desconfiança do usuário quanto da utilização do Banco de Dados com a Linguagem Natural, já que pode-se colocar termos na pesquisa em que o resultado retornado não é o esperado, fazendo assim com que pensamos que a utilização é indevida ou mesma ineficiente.

4.2. O Robô “Ed”

Ed é um Robô/Virtual (Figura 1) desenvolvido pela empresa Insite para a Petrobras com base no software InBot, porém foram realizadas diversas melhorias como, por exemplo, a utilização de modernas técnicas de Processamento de Linguagem Natural [CONPET, 2008].

O intuito do projeto era o desenvolvimento de um personagem virtual/robô (voltado para o público infanto-juvenil), que pudesse se comunicar com as pessoas para esclarecer dúvidas sobre alguns temas afins, entreter, ensinar e interagir com as pessoas. Alguns dos temas abordados pelo robô virtual são: utilização racional dos derivados do petróleo e do gás natural, preservação de energia, meio ambiente, projetos e dicas de economia [Ibid].



Figura 1. Início da modelagem do personagem Ed em 3D.

4.3. ChatterBot Doroty

Segundo Laven [2008], *chatterbot* pode ser definido como um programa de computador que procura simular uma conversação, em linguagem natural, com a finalidade de tornar esta comunicação a mais próxima possível da conversação de um ser humano.

Leonhardt e Tarouco [2005] afirmam que os *chatterbots* podem ser utilizados em várias situações, uma delas é para o gerenciamento de redes. Com esse objetivo desenvolveram um *chatterbot*, o qual recebeu o nome de Doroty. Doroty tem o objetivo de interagir e comunicar com o gerenciador da rede e também com a própria rede em si, através do protocolo SNMP (*Simple Network Management Protocol*).

Doroty (Figura 2) possui em sua arquitetura alguns elementos necessários para fazer essa interação como a interface, a base de conhecimento, o banco de dados, módulo de atualização, módulo coletor e módulo central. A interação é feita da seguinte forma: o usuário faz a pergunta em linguagem natural, a entrada é processada e o *chatterbot* faz a pesquisa relacionada em sua base de dados, fazendo com que a explicação/resposta seja alcançada, mas se a resposta necessitar ser mais específica, um outro módulo é ativado, o chamado módulo central, que por sua vez busca em sua base de dados uma resposta mais dinâmica [Leonhardt e Tarouco, 2005].

Durante o desenvolvimento do *chatterbot* Doroty foi utilizada uma base de dados de conhecimentos chamada AIML (*Artificial Intelligence Markup Language*), a linguagem descreve um conjunto de padrões que o *chatterbot* pode reconhecer através de seus objetos. Essa linguagem foi desenvolvida baseada no XML. Toda a informação é recebida pelo módulo central, o qual irá receber essas informações e processá-las, de tal forma que ela se torne mais útil e completa. O módulo central também é responsável pelas aplicações de regras e o recebimento de parâmetros para cada combinação de entrada que se possa ter [Ibid].

A imagem a seguir mostra o funcionamento da arquitetura de Doroty, segundo Leonhardt e Tarouco [2005].

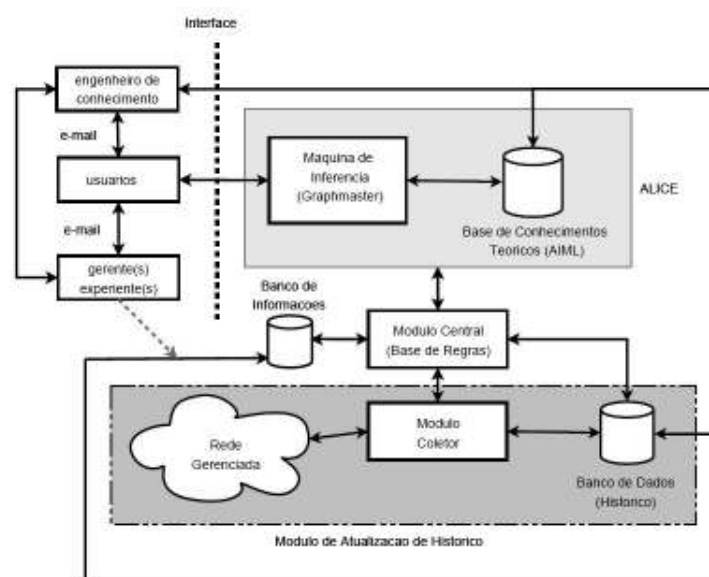


Figura 2. Diagrama da arquitetura de Doroty.

4.4. Sete Zoom

Sete Zoom é um *chatbot* desenvolvido pela empresa InSite para a Gessy Lever. O projeto consistiu na criação da inteligência e conhecimento da personagem que tem a função de interagir com os internautas que acessam o site da empresa. Por possuir uma personalidade atrevida e despojada, o *chatbot* Sete Zoom (Figura 3) se tornou uma espécie de modelo virtual [InSite, 2008].

Segundo os dados estatísticos dos seus desenvolvedores houveram:

- Mais de 1 milhão de frases conversadas em apenas 2 meses;
- Mais de 40 matérias publicadas na imprensa;
- A modelo virtual brasileira mais famosa;
- 7 escritores, 1 roteirista, 2 programadores;
- 2 meses de desenvolvimento;
- Respostas para mais de 95% das frases dos usuários.

Diversos conceitos existentes em uma conversa humana foram incorporados no programa como memória, contexto, aprendizado e aparente elaboração de idéias a partir de um assunto. As respostas dadas pelo programa são criadas a partir de uma personalidade (consciência sintética) e de uma base de conhecimento criada pela equipe de escritores e programadores [InSite, 2008].

De acordo com os desenvolvedores a linguagem utilizada no projeto foi desenvolvida pela própria empresa e se chama BDL (Bot Description Language) e utiliza conceitos de inteligência artificial, lingüística, análise sintática, semântica e morfológica, além de modelos de tomadas de decisão, processamento de linguagem natural, estatística, análise de padrões, Lógica Fuzzy e métodos de representação do conhecimento [Medeiros, 2002].



Figura 3. Janela de conversação do *chatbot* SeteZoom.

4.5 ChatterBot Elektra

Elektra é um *chatterbot* que foi criado na UFRGS com o intuito inicial de se ensinar Física para alunos do ensino médio, que teve no *chatterbot* Alice a base para a construção do projeto. Mas posteriormente ela recebeu em sua base de dados de conhecimento, conteúdos de Redes de Computador, para auxiliar uma disciplina do curso de especialização a distância em Informática [Leonhardt, Castro, Dutra e Tarouco, 2003].

Em sua implementação teve-se a idéia de colocar mais de uma resposta para uma devida possível pergunta de uma área de conhecimento específica, fazendo assim com que ao ser questionado por mais de uma vez, o *chatterbot* apresente uma resposta diferente, já que se a pergunta foi repetida, pode significar que o questionamento não foi sanado de forma satisfatória. No *chatterbot* também há a possibilidade de se clicar em *hiperlinks* que foram devidamente colocados nas respostas e ou imagens, fazendo com que, assim o usuário busque mais informações do assunto questionado [Leonhardt, Castro, Dutra e Tarouco, 2003].

Com o tempo os responsáveis pelo projeto perceberam que problemas estavam reduzindo a eficiência da aplicação, foi observado que as respostas trazidas pela *chatterbot* eram diretas, contudo ocorria a falta de estímulos para que o questionador continuasse com as perguntas. Com isso os desenvolvedores perceberam a necessidade de algumas melhorias em Elektra, devido ao fato de que as pessoas que a utilizavam eram de diferentes partes do Brasil, por isso os questionamentos sobre um mesmo assunto eram feitos de formas bem diferentes, assim o *chatter* devolvia como uma solução inexistente. Os resultados destes ajustes foram notados. Os diálogos com o *chatter* começarão a ter mais tempo, e a impressão de que Elektra os compreendia, fazendo com que assim eles compreendem-se e interessavam-se mais pelo assunto pedido, segundo Leonhardt, Castro, Dutra, Tarouco [2003].

5. Considerações Finais

A utilização de Linguagem Natural em conjunto com outras aplicações vem a nos auxiliar quando necessitamos alcançar uma determinada solução/resposta para um problema, trazendo assim benefícios onde a interação do software com o usuário será feita de forma mais simples onde um simples usuário que não possua um amplo domínio da linguagem utilizada em um banco de dados, por exemplo, possa interagir facilmente com o sistema. Outras aplicações também vêm a facilitar essa busca de conhecimento, como é o caso dos *chatterbots*, onde o programa busca entender o que o usuário necessita do tema referido no bate-papo.

Referências

- Castro, Daiane Dorneles de. Dutra, Renato Luís de Souza. Tarouco, Liane Margarida Rockenbach. Leonhardt, Michelle Denis. "ELEKTRA: Um *Chatterbot* para Uso em Ambiente Educacional".
- CONPET Programa Nacional de Racionalização do uso dos Derivados de Petróleo e do Gás Natural. Robô Virtual da Petrobrás. Disponível em <<http://www.inbot.com.br/ed/>>, Acessado em: 15 de junho 2008.

- Gariba, M. Jr., Schneider, M. C. K., Rosa, A. E., Casagrande, J. B., Santos, C. S. "Reconhecimento de Fala e Processamento da Linguagem Natural". Disponível em: <<http://www.cefetsc.edu.br/~gariba/VOZ.PRN.pdf>>. Acessado em: 05 de Julho 2008.
- Gonzalez, M. Lima, V. L. S. "Recuperação de Informação e Processamento da Linguagem Natural". XXIII Congresso da Sociedade Brasileira de Computação, Campinas, 2003. Anais do III Jornada de Mini-Cursos de Inteligência Artificial, Volume III, p.347-395.
- Hübner, Jomi Fred. "Interface em Linguagem Natural", In: Trabalho de Conclusão de Curso, Departamento de Sistemas e Computação da Universidade Regional de Blumenau.
- InSite, Soluções Internet. Disponível em: <<http://www.insite.com.br/>>, Acessado em: 15 de junho 2008.
- Laven, S.J. "*The Simon Laven Page*". Disponível em: <<http://www.simonlaven.com/>>, Acessado em: 03 de julho 2008.
- Leonhardt, M. D., Tarouco, L. M. R.. "Aplicando Linguagem Natural ao Gerenciamento de Redes de Computadores através do Chatterbot Doroty", In: V ENIA – Encontro Nacional de Inteligência Artificial, 2005.
- Leonhardt, M. D., Castro, D. de, Dutra, R. de S., Tarouco, L. M. R. "ELEKTRA: Um *Chatterbot* para Uso em Ambiente Educacional". In: 2º Ciclo de palestras Novas Tecnologias na Educação, 2003.
- Medeiros, E. "Elas estão sempre dispostas a bater papo". O Estadão: Jornal da Tarde. São Paulo. Ed. 13/06/2008. Disponível em: <<http://www.jt.com.br/>>. Acessado em: 05 de Junho 2008.
- Oliveira, Fabio Abreu Dias de. "Processamento de linguagem natural: princípios básicos e a implementação de um *analisador sintático* de sentenças da língua portuguesa". In: Revista de Ciência da Informação. Rio de Janeiro. n. 5. Maio 2002. Disponível em: <<http://www.inf.ufrgs.br/proccpar/disc/cmp135/trabs/992/Parser/parser.html>>, Acessado em: 15 de junho 2008.
- Russel, Stuart J., Norvig, Peter. "Artificial Intelligence: A Modern Approach". New Jersey EUA. Prentice Hall Series in Artificial Intelligence. 1995.
- Silva, Cassandra Ribeiro de O. "Metodologia e Organização do projeto de pesquisa". Centro Federal de Educação Tecnológica do Ceará (Cefetce). Disponível em: <http://www.cefetce.br/Pesquisa/dippg/metodologia/Metodologia%20e%20Organiza%20de%20pesquisa_apostila.pdf>. Acessado em: 05 de Julho 2008.
- Silva, R. R., Lima, S. M. B. "Consultas em Bancos de Dados Utilizando Linguagem Natural".
- Smarsaro, Aucione das Dores. (2004) "Descrição e formalização de palavras compostas do português do Brasil para elaboração de um dicionário eletrônico", In: Tese de Doutorado, Programa de Pós-graduação em Letras do Departamento de Letras da PUC-Rio.