

Pesquisa e Desenvolvimento de um Ambiente para a Construção de *Chatterbots* Através de Redes Semânticas

Giulia Faustini Milan¹, Bárbara Santos Munhão¹, Antonio Miguel Faustini Zarth¹,
Isaias Leonidio Farias¹

¹Instituto Federal de Mato Grosso do Sul (IFMS)
Campus Campo Grande – Avenida Júlio de Castilho, 4960 – 79113-000 – Campo
Grande MS - Brasil

{giulia-milan, barbara.munhao}@hotmail.com, {antonio.zarth,
isaias.farias}@ifms.edu.br

Abstract. *Identified several factors that impede the keeping, creation and update of chatterbots when the AIML language (Artificial Intelligence Markup Language) is used, the objective of this work is a research and creation of a development environment for automated construction of chatterbots through semantic networks.*

Resumo. *Identificados alguns fatores que dificultam a manutenção, criação e atualização de chatterbots ao se utilizar a linguagem de formatação AIML (Artificial Intelligence Markup Language), objetiva-se com este trabalho a pesquisa e criação de um ambiente de desenvolvimento para a construção automatizada de chatterbots através de redes semânticas.*

1. Introdução

O objetivo deste trabalho é a pesquisa e criação de um ambiente de desenvolvimento automatizado para a construção de *chatterbots*¹ através de redes semânticas². Esse ambiente visa facilitar a criação de robôs virtuais em virtude da dificuldade quanto ao uso do AIML [Milan et al. 2013] para a construção destes.

Notou-se que esses robôs de conversação possuíam diversas limitações, principalmente quanto a parte semântica de um texto. Assim como dificuldades na manutenção, criação e atualização do robô. A linguagem AIML trata as frases apenas a nível sintático, tornando-se necessária uma base de dados muito grande, referente a um domínio pequeno, para englobar as frases que possuem ambiguidade, inúmeros sentidos ou mesmo sentido.

Com este novo sistema, a necessidade de codificação da base de conhecimentos é menor, já que a automatização é um dos focos da ferramenta proposta, além das informações serem tratadas a nível sintático-semântico. Em meio a esses processos existem etapas menores, como o tratamento de vícios de linguagem, sinônimos, dentre outras.

A construção do ambiente possui o propósito de facilitar na criação e na manutenção de *chatterbots* com um acréscimo de funcionalidades em sua estrutura, melhorando também sua eficiência.

1 Robôs de conversação que simulam conversas com humanos através de linguagem natural [Russel; Norving, 2004].

2 Forma de representação de conhecimento que pode ligar relações semânticas entre conceitos.

2. Construtor de *Chatterbots* Através de Redes Semânticas e Casamento de Padrões

Identificados os problemas quanto ao uso do AIML [Milan et al. 2013], foi iniciado um processo de estudo da sintaxe aliada à semântica nas frases, e a busca por ferramentas que tratam a semântica de textos. Com esse embasamento, modelos de sistemas foram montados para sua implementação.

Esse ambiente de desenvolvimento é dividido em quatro grandes partes: tratamento de sinônimos, casamento de padrões, rede semântica e construtor de frases. O esquema do sistema completo é mostrado na figura 1.

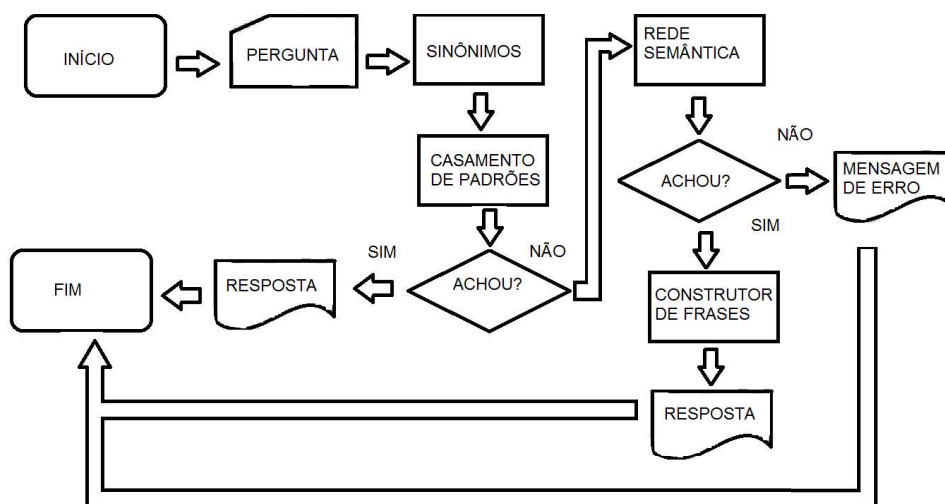


Figura 1: Fluxograma do ambiente de desenvolvimento de *Chatterbots*

2.1. Dicionário de Sinônimos e Casamento de Padrões

A integração de um dicionário de sinônimos e um sistema de casamento de padrões, no ambiente de desenvolvimento proposto, visa auxiliar e ampliar a eficiência do processo de criação de *chatterbots*. Para que um *bot* possa desenvolver uma conversa com um humano, é necessário não somente que ele domine um determinado assunto, mas também que sejam conhecidos alguns fatores de uma conversa humana.

Pensando nas características antropomórficas a serem dadas aos *chatterbots*, viabilizou-se a agregação de um sistema de padrões que trata, principalmente, questões ligadas aos cumprimentos e outras informações simples, que não seriam suficientemente precisas para se armazenar na forma de redes semânticas. Portanto, foi criado um sistema de casamento de padrões que se assemelha ao funcionamento do AIML (*Artificial Intelligence Markup Language*).

O casamento de padrões se assemelha ao AIML pelo fato de conter funções como: asterisco (*) - representação genérica de uma ou mais palavras -, funções aleatórias de respostas e funções recursivas, que definem uma resposta de acordo com o diálogo anterior.

A implementação de algumas funções, baseadas nos *patterns* e outras aplicações do AIML, é voltada para partes específicas do sistema, tornando-o mais sólido e preciso. Porém quando se trata de informações mais complexas ou de uma facilidade no armazenamento de diversos dados, o AIML não é eficaz o suficiente, pois trata apenas da sintaxe, perdendo o sentido das sentenças. Por isso, viabilizou-se o uso de uma rede semântica.

2.2. Criação e Busca na Rede Semântica

A rede semântica é construída por meio da inserção de um texto que, após a passagem por métodos de identificação e extração de informações, é combinado com regras previamente definidas, descritas e codificadas em PHP. Os objetos e suas relações identificadas são armazenados dentro de um banco de dados, onde formam uma rede de informações que se definem por um sentido, automaticamente classificado. Dessa forma, as frases constituídas no texto, anteriormente inserido, serão tratadas de forma semântica e automática.

Com a rede semântica criada, é iniciado o processo de recuperação de informações que irão compor a resposta do usuário. Para que o robô possa estabelecer uma conversa, é necessário que haja consultas à base de conhecimento para que ele possa fazer uma espécie de *matching*, seja de sentido ou de palavras, e, em sequência, devolver sentenças da própria rede semântica ao usuário.

A primeira parte do sistema de recuperação de informações funciona como o de extração destas. Ele recebe a pergunta do usuário e essa frase passa por um processo de divisão e classificação de termos e sentido. Após a primeira etapa, o sentido e a palavra que estabelece a relação são procurados dentro da rede semântica. Quando encontrados, a informação da rede será a resposta para o usuário.

2.2. Construtor de Frases

Quando o usuário inserir uma pergunta ao sistema, esta passará por diversos estágios e tratamentos que farão com que ela combine com a informação que deve ser devolvida.

Primeiro, a frase de entrada no sistema passará pelo banco de sinônimos, fazendo as alterações necessárias. Após isso, passa pelo casamento de padrões em busca de um *matching* correto; se esse *matching* acontecer, responderá ao usuário. Se ele não acontecer, entrará na rede semântica. Antes, a frase será dividida, os objetos e palavra que fazem a relação serão identificados e um sentido será classificado. Na rede, buscará por uma mesma relação (palavra que relaciona os objetos) da pergunta do usuário, como também procurará pelo mesmo sentido da frase, classificado anteriormente. Quando for encontrada, a própria relação que está na rede semântica será a resposta de saída ao usuário. Se não for encontrada, emitirá uma mensagem de erro. Como mostra o pseudocódigo 1.

3. Resultados

Mediante a construção de um ambiente de desenvolvimento automatizado de *chatterbots* através de redes semânticas, obtiveram-se resultados qualitativos na construção das redes, que são a base de conhecimento dos robôs virtuais de conversação, e no processo de formação das sentenças que servirão de respostas para o usuário.

Com a utilização das redes semânticas e outras funcionalidades do sistema, é possível visualizar uma maior facilidade de manutenção e criação de robôs de conversação, comparando à criação através da linguagem de formatação AIML [Milan et al. 2013].

Início

Recebe pergunta de usuário;
Entra no banco de sinônimos e faz as alterações necessárias;
Entra no casamento de padrões;

Se fizer matching

Responde ao usuário;

Senão

Divide a pergunta do usuário e classifica sentido;
Entra na rede semântica;
Busca a informação através da relação e sentido da entrada e das informações da rede;

Se encontrar resposta

Constrói a frase de saída ao usuário;

Senão

Mostra uma mensagem de erro;

Fim-se

Fim-se

Fim

PseudoCódigo 1: Construtor de frases

Essa maior facilidade é comprovada com o fato do sistema ser automatizado quanto à criação da base de conhecimento, além de ser armazenado em um banco de dados, que facilita a manutenção.

Quanto às respostas aos usuários, há um aumento na eficiência dos robôs nos *matchings*, já que estão baseados na análise sintático-semântica das sentenças. Gera-se então um maior acerto em relação ao assunto referente à entrada do usuário, podendo compará-lo aos conhecimentos da base com maior propriedade. Além disso, funcionalidades de substituições e *patterns* foram implementados, deixando o sistema mais sólido.

Agradecimentos

Os autores agradecem ao CNPq pelo apoio através da concessão de bolsas de pesquisa PIBIC-EM para a execução deste trabalho.

Referências

- MILAN, Giulia; MUNHÃO, Bárbara; ZARTH, Antonio. Chatterbot Tira-Dúvidas do Curso Técnico de Informática do IFMS. In: Anais do Computer on the Beach, Florianópolis, SC, 2013.
- RUSSEL, S.; NORVIG, P.; Inteligência Artificial; 2 ed.; Rio de Janeiro; editora: CAMPUS; 2004.
- OTHERO, Gabriel; MENUZZI, Sérgio. Linguística Computacional: teoria e prática. São Paulo: Parábola Editorial, 2005.
- CUNHA, Celso & CINTRA, Lindley. Nova Gramática do Português Contemporâneo. Rio de Janeiro: Nova Fronteira, 3a ed..2001.
- SOWA, John. Semantic Networks. Disponível em: <<http://www.jfsowa.com/pubs/semnet.htm>>. Acesso em junho de 2013.