Interpretador AIML alimentado com tags HTML5

Rafael Luiz de Macedo (Ciência da Computação - UNIVEM - Marília/SP) Email: rafaelldemacedo@gmail.com

Orientador: Prof. Dr. Elvis Fusco (UNIVEM - Marília/SP)

RESUMO

O AIML (Artificial Intelligence Markup Language) é um conjunto de tags XML (eXtensible Markup Language) capaz de representar e relacionar expressões em linguagem natural, permitindo a criação de motores robóticos capazes de manter um diálogo simples. Este artigo tem o objetivo de estender o AIML na incorporação de elementos HTML5 padrões e personalizados, utilizando um interpretador Java. Por meio destes elementos propostos e com o auxílio de um software analisador de código-fonte desenvolvido, é possível realizar buscas semânticas na Web, utilizando os códigos-fontes das páginas e, a partir delas, os elementos são analisados e extraídos para que ocorra um tratamento semântico nas informações extraídas, gerando, assim, informações tratadas para o AIML.

PALAVRA-CHAVE

AIML; Inteligência Artificial; Motores Robóticos; Web Semântica.

1. INTRODUÇÃO

As tecnologias atuais têm permitido, cada vez mais, a percepção que é possível fazer uma máquina ser inteligente a ponto de responder perguntas feitas por meio de motores robóticos construídos com inteligência artificial.

A linguagem destina-se à comunicação sobre o mundo. Quando se estuda a linguagem, passa-se a conhecer mais o mundo, podem-se testar teorias sobre o mundo se conhecem que sustentação elas dão à tentativa de compreender a linguagem. E, se conseguirmos criar um modelo computacional de linguagem, pode-se ter uma ferramenta poderosa para a comunicação sobre o mundo [RICH, Elaine e KNIGHT, Kevin, 1993, p.433].

Uma das linguagens que possibilita a criação e o desenvolvimento desses motores robóticos inteligentes é a linguagem de marcação AIML (*Artificial Intelligence Markup Languages*). Mas, a linguagem AIML possui uma limitação no desenvolvimento desses motores, fazendo com que seja limitado o número de perguntas que possa ser respondido por esses motores robóticos.

Com o surgimento da Internet, é possível criar novas tags AIML interpretadas com tags HTML5 padrões e customizadas para que sejam realizadas buscas semânticas na web. Essas buscas serão feitos por meio das perguntas que os motores robóticos não forem possíveis respondê-las. E, com um software analisador de código-fonte que faz o tratamento extraindo informações de Web sites encontrados na Internet e transformando essas novas informações em novas perguntas e respostas para potencializar e abranger o conhecimento desses motores robóticos inteligentes.

O objetivo desse artigo é estender a linguagem de marcação AIML com novos elementos AIML, a partir da incorporação de tags HTML5 e do desenvolvimento de um software analisador capaz de fazer o tratamento das informações extraídas do código-fonte de web site, para, assim, gerando novas informações na linguagem AIML.

Para que seja feita a busca e interpretação de informações na WEB, é necessária a utilização de conceitos de buscas e interpretações de linguagem natural. Em particular, a Ontologia e o Processamento da Linguagem Natural impõem essas regras.

2. AIML

A linguagem AIML possibilita a criação de motores robóticos de conversação chamados de *Chatterbots* (*Chat* = Conversa e *bot* = Robô), ou seja, um robô com Inteligência Artificial que interage com usuários, por meio de conversas, fazendo com que um humano pense que esteja "conversando" com outro humano.

Um dos softwares mais conhecidos em processamento de linguagem natural é a A.L.I.C.E. (*Artificial Linguistic Internet Computer Entity*). Com o desenvolvimento da linguagem A.L.I.C.E., resultou a linguagem de marcação AIML. A licença do AIML está sob a GNU GPL e existem interpretadores AIML disponíveis em Java, Ruby, Python, C++, C#, Pascal e outras linguagens.

A linguagem AIML é um conjunto de tags XML (eXtensible Markup Language) capaz de representar e relacionar expressões em linguagem natural, permitindo a criação de motores capazes de manter um diálogo simples. Cada conjunto de tags AIML possui uma ou mais tags chamadas de categoria, sendo que as categorias são desenvolvidas em cima de um contexto; uma categoria é formada pelo conjunto das tags template e pattern, que são as tags responsáveis por interpretar a mensagem digitada pelo usuário e enviar uma mensagem de resposta ao usuário.

No desenvolvimento de um arquivo AIML, é obrigatória a utilização de algumas tags que se referem aos direitos autorais da linguagem de marcação AIML. Logo em seguida, está sendo mostrado como é um conjunto de tags AIML.

<?xml version='1.0' encoding='ISO8859-1'?>

<aiml version='1.0.1' xmlns='http://
alicebot.org/2001/AIML-1.0.1'xmlns:html='http://
www.w3.org/1999/xhtml'xmlns:xsi='http://www.
w3.org/2001/XMLSchemainstance'xsi:schemaLoca
tion='http://alicebot.org/2001/AIML-1.0.1 http://
aitools.org/aiml/schema/AIML.xsd'>

<category>

<pattern>O que é AIML *

pattern>

<template> O AIML (Artificial Intelligence Markup Language) é um conjunto de tags XML (Linguagem de Marcação Extensível) capaz de representar e relacionar expressões em linguagem

natural, permitindo a criação de motores capazes de manter um diálogo simples.

</template>

<category>

<pattern>O que é GaitoBot

*</pattern>

<template> É um software para desenvolver chatterbot em AIML; o GaitoBot facilita no desenvolvimento de chatterbot por possuir uma interface gráfica que disponibiliza o usuário criar vários conjuntos de TAGs apenas com o click de um botão.

</template>

</aiml>

Nas tags AIML citadas anteriormente são elementos padrões na criação de um arquivo AIML. Existem outros elementos com diferentes funcionalidades. Uma desses elementos é a star, que pode ser representada de duas maneiras, como um elemento <star> ou como o símbolo * e tem a funcionalidade de reconhecer o conjunto de caracteres que não forem definidos dentro do elemento pattern ou template.

O AIML é dividido em dois módulos, linguagem e motor. A linguagem é onde fica todo o conhecimento da linguagem natural e as informações desenvolvidas na linguagem AIML, e o motor faz a comunicação entre a linguagem natural e a linguagem de marcação AIML, porém existe uma limitação em seus módulos. O desenvolvimento do AIML está relacionado a um contexto, fazendo com que seja limitada sua aplicação. Assim, em um sistema de processamento de linguagem natural utilizando o AIML, a conversação não se mantém por muito tempo em um diálogo devido ao fato dele se perder em uma pergunta fora do contexto que foi desenvolvido.

Alan Turing (1912-1954), em seu famoso ensaio "Computing Machinery and Intelligence" (Turing 1950), sugeriu que, em vez de perguntar se máquinas podem pensar, devemos perguntar se máquinas podem passar por um teste de inteligência comportamental, que veio a ser chamado teste de Turing [Stuart Russell e Peter Norvig, 2004, p.916].

A linguagem de marcação AIML é uma linguagem livre (open source), assim possibilitando o uso da linguagem na pesquisa para melhorias e cria-

ção de motores robóticos inteligentes de conversação. Alan Turing propôs que, se não é possível distinguir respostas de uma máquina daquelas fornecidas por uma pessoa, a máquina poderia ser considerada inteligente e, até os dias atuais, não existem máquinas capazes de passar por esse teste de Turing.

Para criarem motores robóticos, pode ser usada uma plataforma específica e, em uma linguagem de marcação, existem várias plataformas e linguagens disponíveis para a criação desses motores robóticos.

Neste artigo, está sendo usado um interpretador na linguagem de programação Java e a plataforma ProgramD, que é uma plataforma livre (Open Source).

A plataforma ProgramD é a plataforma mais utilizada no desenvolvimento desses motores robóticos inteligentes na linguagem de marcação AIML e a mais completa em recurso da linguagem AIML, além de ser a melhor implementação já testada do AIML.

No desenvolvimento de motores robóticos inteligentes na linguagem de marcação AIML, é possível encontrar softwares que possibilitam a criação desses motores robóticos de uma forma mais prática e usando interfaces gráficas. Um software conhecido para a criação desses motores robóticos na linguagem de marcação AIML é o GaitoBot.

O software GaitoBot é uma ferramenta que facilita o desenvolvimento de motores robóticos na linguagem de marcação AIML, possui uma interface gráfica que facilita a criação de motores robóticos inteligentes e, com apenas alguns cliques, permite criar um novo motor robótico.

3. HTML5

Com o avanço da Internet e dos navegadores, o HTML5 (Hypertext Markup Language) é uma tentativa válida de padronizar a maneira como os navegadores interpretam, classificam e apresentam as informações recebidas. Enquanto a W3C (World Wide Web Consortium) focava suas atenções para a criação da segunda versão do XHTML, um grupo chamado Web Hypertext Application Technology Working Group ou WHATWG trabalhava em uma versão do HTML que trazia mais flexibilidade para a produção de sites e sistemas baseados na Web.

O HTML5 é uma linguagem de marcação

que facilita o desenvolvimento de sites, pois, com poucas linhas, é possível construir um site simples e além de possibilitar *plugins* que facilitam o desenvolvimento de interfaces. Aproveitando esse novo padrão de organização das informações na Web é possível aumentar a abrangência e a capacidade de linguagens como o AIML. Atualmente, a maioria dos Web sites é desenvolvida em cima da linguagem HTML5 e os novos elementos criados na linguagem AIML farão a interpretação dessas tags HTML5, realizando busca semântica na Web para encontrar sites desenvolvimentos na linguagem HTML5.

4. ONTOLOGIA

A ontologia trata sobre o conhecimento do ser, estuda a compreensão do ser. Ela cria vários tipos de domínios diferentes, que, a cada domínio, é conceituado sobre um determinado assunto, como numa busca semântica na Web.

Segundo [Gruber, T.B., 1993], ontologia é uma especificação de uma conceitualização, isto é, uma ontologia é uma descrição (como uma especificação formal de um programa) dos conceitos e relacionamentos que podem existir para um agente ou uma comunidade de agentes.

Na Web, a ontologia é utilizada para organizar um domínio de conhecimento, para, assim poder modelar um conhecimento. Exemplo de utilização da ontologia é uma busca semântica na Web, pois se obtém uma busca mais precisa, sabendo do conhecimento relacionado ao assunto ou objeto que está sendo buscado e, assim, obtendo resultados mais precisos.

A integração da Web semântica com a linguagem AIML tem o principal objetivo de fazer o uso da Web para potencializar e abranger a linguagem AIML e os motores robóticos que forem criados em cima dessa mesma linguagem, fazendo com que aumente o número de perguntas que possam ser realizadas a esses motores robóticos.

Para que a linguagem AIML possa realizar essas buscas semânticas, serão desenvolvidos novos elementos AIML interpretados com tags HTML5 padrões e customizadas. Para a realização dessas buscas, porém, é necessário obter informações claras e precisas sobre o contexto das perguntas que for realizada a esses motores robóticos. Para isso serão

feitas perguntas ao humano e, assim, obter informações sobre o contexto que a pergunta pertence.

5. PROCESSAMENTO DE LINGUAGEM NATURAL

O processamento de linguagem natural é o estudo da linguagem, sendo linguagem falada ou linguagem de texto escrito. O estudo da linguagem nos possibilita uma melhor compreensão sobre a linguagem e, se as máquinas compreendem a linguagem, é possível criar motores robóticos inteligentes.

A linguagem escrita faz uso de conhecimentos léxicos, sintáticos e semânticos da linguagem e também de todas as informações necessárias sobre o mundo real [RICH, Elaine e KNIGHT, Kevin, 1993, p.434].

O processamento de linguagem natural é dividido em vários componentes, sendo que cada componente tem sua função no reconhecimento da linguagem natural. Os componentes, Análise Morfológica, Análise Sintática, Análise Semântica, Integração de Discurso e Análise Pragmática, são o grupo de componentes que compõem o corpo do processamento de linguagem natural.

1. Análise Morfológica

Cada palavra é analisada separada e é feito o reconhecimento da mesma; as palavras são analisados em termos de seus componentes, e os sinais, como a pontuação, são separados das palavras [RICH, Elaine e KNIGHT, Kevin, 1993, p.436].

2. Análise Sintática

A análise sintática faz o reconhecimento de um conjunto de palavras para ver se elas estão corretamente relacionadas entre si. Algumas sequências de palavras podem ser rejeitadas se violarem as regras da linguagem sobre como as palavras podem ser combinadas. Por exemplo, um analisador sintático do português rejeitaria a frase: "Menino o vai loja à" [RICH, Elaine e KNIGHT, Kevin, 1993, p.436].

3. Análise Semântica

É feito um mapeamento nos conjuntos de palavras para fazer o reconhecimento e ver se o significado do conjunto está correto. Cada conjunto de palavras tem um significado semântico, pertence a algum contexto específico. A análise semântica faz o reconhecimento desse contexto a que o conjunto de palavras pertence e faz um mapeamento semântico para ver se está correto seu significado. Um conjunto de palavras em que não é possível ser feito o mapeamento semântico é rejeitado pelo analisador semântico.

4. Integração de Discurso

É o significado de uma frase isolada que depende das frases anteriores para obter seu significado, podendo influenciar em frases posteriores. Por exemplo, a palavra "aquilo", na frase "João queria aquilo", depende do contexto do discurso anterior; enquanto a palavra "João" pode influenciar o significado de frases posteriores, como: "Ele sempre quis" [RICH, Elaine e KNIGHT, Kevin, 1993, p.436].

5. Análise Pragmática

Uma análise pragmática faz a interpretação da frase, reconhecendo o que ela quis dizer. A frase pode representar uma solicitação, uma afirmação, um aviso e outras representações contidas na linguagem natural. A frase "Você sabe que horas são?" deve ser interpretada como uma solicitação para que as horas sejam informadas [RICH, Elaine e KNIGHT, Kevin, 1993, p.437].

6. MOTORES ROBÓTICOS COMBINADOS COM PROCESSAMENTO DE LINGUAGEM NATURAL

Os motores robóticos fazem o reconhecimento de um conjunto de caracteres de uma linguagem natural, ou seja, reconhecem palavras e frases digitadas pelos usuários; Contudo esses motores robóticos não conseguem compreender o significado das palavras ou frases digitadas.

Os motores robóticos criados na linguagem de marcação AIML, fazem o reconhecimento de perguntas e respostas escritas em uma linguagem natural, mas possuí em uma limitação no reconhecimento dessas perguntas; apenas é possível realizar perguntas que esses motores possuem em sua linguagem natural.

Comas regras do processamento de linguagem natural e com um domínio específico da ontologia, é possível desenvolver um software analisador de código-fonte que faz uma melhor compreensão na linguagem natural, gerando novas perguntas e respostas para os motores robóticos.

7. METODOLOGIA E RESULTADOS

A metodologia utilizada para o desenvolvimento está dividida em fases que incluem: o levantamento bibliográfico e estudo do AIML e HTML5; pesquisa de correlatos; definição do modelo de integração; implementação de protótipo; teste e validação; comparação com correlatos. Para que seja feita a busca e interpretação de informações na Web, é necessário o uso de algumas regras de busca e interpretação, utilizando regras de Ontologia e Processamento da Linguagem Natural.

O estudo bibliográfico das linguagens AIML e HTML5 já foi concluído e criados os cenários de testes para o desenvolvimento do software analisador de código-fonte de web sites obtidos na busca semântica na Web. Para abranger o conhecimento sobre a linguagem AIML, foi escrito um tutorial na língua portuguesa sobre a linguagem de marcação AIML, que explica a funcionalidade de algumas tags e mostra como criar um conjunto de tags categorias.

Otutorial AIML, explica as tags mais utilizadas para o desenvolvimento desse artigo e mais algumas outras tags da linguagem.

Na etapa que se encontra a pesquisa é a de desenvolvimento do software analisador de códigofonte, que já realiza algumas análises e determinados tratamentos em cima dos cenários que foram desenvolvidos para esse artigo.

O funcionamento do software analisador de código-fonte é dividido por etapas. Na primeira etapa, é passado todo código-fonte do site encontrado em um arquivo já criado pelo software analisador, para que, assim, seja possível analisar e dar o tratamento correto ao código-fonte encontrado. Na figura 1, é mostrado como é um arquivo, após ser passado todo o código-fonte do site encontrado para o arquivo.

Figura 1. Arquivo carregado com o código-fonte do site.

```
<!DOCTYPE html>
<html>
<link href="css.css" rel="stylesheet" type="text/css"/>
<body>
<div id="total">
        <img src="banner.gif" alt="GaitoBot" width="800" height="160"/>
        <div id="conteudo">
                <article>
                <h2><b>Resumo</b></h2>
                <div id="paragrafo">
                        cp class="intro">0 que é AIML ? O AIML (Artificial Intelligence Markup Language) é um conjunto de tags XML
                        capaz de representar e relacionar expressões em linquagem natural permitindo a criação de motores capazes «
                        simples. Este projeto de pesquisa tem o intuito de potencializar o AIML incorporando TAGs HTML5 padrão e cu
                        um interpretador em java.
                        </article>
                </div>
        </div>
</div>
<div id="gaitobot">
        <h2>GaitoBot</h2>
               O que é GaitoBot ? É um software para desenvolver chatterbot em AIML, o GaitoBot facilita no desenvolvimento de cha
                gráfica que disponibiliza você criar vários conjuntos de TAGs apenas com um click de um botão.
```

Na segunda etapa, o software analisa todo o arquivo que contém o código-fonte do site e aplica o tratamento; o analisador faz o reconhecimento de conjuntos de palavras que não pertence a nenhuma linguagem de programação de Web sites e separa esses conjuntos de palavras em um novo arquivo na ordem em que foram encontrados os conjuntos de palavras. Todas as tags e comandos que pertence a alguma linguagem de programação de Web sites são descartados pelo software. Na figura 2, é mostrado o resultado da segunda etapa, que é o arquivo gerado com as informações encontradas no site.

Figura 2. Arquivo com as informações que foram extraídas do código-fonte do site.

Resumo O que é AIML ? O AIML (Artificial Intelligence Markup Language) é um conjunto de tags XML (Linguagem de Marcação Extensível) capaz de representar e relacionar expressões em linguagem natural permitindo a criação de motores capazes de manter um diálogo simples. Este projeto de pesquisa tem o intuito de potencializar o AIML incorporando TAGs HTML5 padrão e customizadas utilizand um interpretador em java. O que é GaitoBot ? É um software para desenvolver chatterbot em AIML, o GaitoBot facilita no desenvolvimento de chatterbot por gráfica que disponibiliza você criar vários conjuntos de TAGs apenas com um click de um botão. O GaitoBot é um software livre só para fins acadêmicos, esta disponível em duas línguas, Inglês e Alemão, possui uma interface gráfica fácil para a aprendizagem. AIML O AIML (Artificial Intelligence Markup Language, Linguagem de Marcação da Inteligência Artificial) é um conjunto de tags XML (eXtensible Markup Language, Linguagem de Marcação Extensível) capaz de representar e relacionar expressões em linguagem natura a criação de motores capazes de manter um diáloĝo simples em linguagem natural. Cada conjunto de tags AIML pode possui uma ou v chamada de categorias, pois as tags categorias possuem varias tipos de tags dentro de si que fazem a interpretação de uma mensa por um usuário e assim dar uma resposta através de funções que estão definidas em outras tags dentro da tag categoria, no tópic 4. TAGS ATM A linguagem AIML possibilita a criação de robôs de conversação, que esses robôs são chamados de Chatterbots (Chat = Conversa e ou seja, um robô com Inteligência Artificial que interage com usuário através de conversas, fazendo com que um humano pense que conversando com outro humano. Ontologia (em grego

Na terceira e última etapa, o software analisador carrega o arquivo que contém as informações tiradas do site analisado, que, com esse arquivo, será aplicada um tratamento de reconhecimento de perguntas e respostas. O software analisa todo o arquivo em busca de perguntas que possam estar, de forma direta ou indiretamente, escritas nessas informações.

As perguntas e respostas encontradas pelo

software analisador são escritas em um novo arquivo junto com as tags AIML, para que o motor robótico desenvolvido na linguagem AIML possa fazer uso desse arquivo e, assim, responder às perguntas que pertencem a esse arquivo gerado pelo software analisador de código-fonte. A figura 3 mostra como é um arquivo AIML gerado pelo software analisador de código-fonte.

Figura 3. Arquivo AIML gerado pelo software analisador de código-fonte.

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<aiml version='1.0.1' xmlns='http://alicebot.org/2001/AIML-1.0.1'xmlns:html='http://www.w3.org/1999/xhtml' xmlns:xsi='http://www.w3.org/1999/xhtml' xmln
```

O arquivo AIML que foi gerado pelo software analisador de código-fonte será utilizado pelo motor robótico construído na linguagem de marcação AIML, para abranger seu conhecimento com as novas perguntas e respostas que contém nesse novo arquivo. Está sendo utilizada a plataforma ProgramD, que é uma plataforma livre para testar os arquivos AIML gerados pelo software analisador.

8. CONCLUSÃO

A linguagem de marcação AIML possui uma limitação na criação de motores robóticos de conversação, fazendo com que seja limitado o número de perguntas que possam ser feitas a esses motores robóticos. Devido a essa limitação, os motores robóticos não conseguem manter uma conversa por muito tempo com um humano, pois perguntas realizadas fora do contexto de conhecimento em que esses motores foram desenvolvidos não conseguem ser respondidas.

Para tal problema, propõe-se a extensão da AIML com a incorporação de elementos interpretados com tag HTML5 padrão e customizados, para a realização de buscas semânticas na Web, usando conceitos de um domínio específico da ontologia para uma melhor compreensão nas informações que contém na pergunta realizada a esses motores robóticos. Com o uso dos conceitos de um domínio da ontologia, será possível fazer uma busca mais concreta e obter resultados mais precisos ao contexto da pergunta.

A integração dessas novas tags AIML com o software analisador de código-fonte fará o uso de técnicas de processamento de linguagem natural, para o tratamento dessas informações extraídas dos sites analisados, transformando-as em novas perguntas e respostas para abranger o conhecimento da linguagem AIML e aumentar o número de perguntas que possam ser realizadas para esses motores robóticos inteligentes.

O tutorial sobre a linguagem de marcação AIML na língua portuguesa tem o objetivo de facilitar

o funcionamento e a forma como são criados os motores robóticos inteligentes na linguagem AIML, explicando as tags que são necessárias na criação de um simples motor robótico inteligente.

No tutorial, não se encontram todas as tags disponíveis na linguagem de marcação AIML, este apenas explica as tags necessárias para a criação de um simples motor robótico inteligente.

A linguagem de marcação AIML contém várias outras tags, que se encontram em Web sites oficiais da A.L.I.C.E. e do AIML.

9. REFERÊNCIAS

ALICE AI Foundation. "AIML: The Artificial Intelligence Markup Language" http://www.alicebot.org/aiml.html, Maio.

ALVES, Rachel C. V. (2005) "Web Semântica: Uma Análise focada no uso de metadados", Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências – Universidade Estadual Paulista - UNESP, Marília/SP.

BUSH, Noel. "ProgramD", http://aitools.org/Programd, Novembro de 2012.

Gruber, T.R. (1993) "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", International Workshop on Formal Ontology, Padova, Itália.

MENEZES, Paulo, B. (1998), "Linguagens Formais e Autômatos" 2ª Edição. Porto Alegre.

O'Brien. P, Zhan C. (2000) "Domain Ontology Management Environment", Proceedings of 33rd Hawaii International Conference on Systems Sciences, Janeiro, Havaí.

RUSSELL, Stuart e NORVIG, Peter. (2004) "Inteligência Artificial" Tradução da Segunda Edição. Rio de Janeiro/RJ.

RICH, Elaine e KNIGHT, Kevin. (1993) "Inteligência Artificial" – Segunda Edição. São Paulo/SP.

SILVA, Maurício Samy. (2010) "HTML5 - A linguagem de marcação que revolucionou a Web" - Editora Novatec, São Paulo/SP.

WALLACE, Dr. Richard; TAYLOR, Anthony. "Manual de Referência AIML", http://www.alicebot.org/documentation/aiml-reference.html, Outubro de 2012.