

UNIVERSIDADE SAGRADO CORAÇÃO

MURILO LUIZETTO GONÇALVES DE AGUIAR

**IMPLEMENTAÇÃO DE UM ATENDENTE
VIRTUAL ESPECIALIZADO EM LICENCIAMENTO
AMBIENTAL UTILIZANDO AIML**

BAURU

2013

MURILO LUIZETTO GONÇALVES DE AGUIAR

**IMPLEMENTAÇÃO DE UM ATENDENTE
VIRTUAL ESPECIALIZADO EM LICENCIAMENTO
AMBIENTAL UTILIZANDO AIML**

Trabalho de Conclusão de Curso apresentado ao Centro de Ciências Exatas e Sociais Aplicadas como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação, sob orientação do Prof. Ms. Patrick Pedreira Silva.

BAURU

2013

A282i

Aguiar, Murilo Luizetto Gonçalves de

Implementação de um atendente virtual especializado em licenciamento ambiental utilizando AIML / Murilo Luizetto Gonçalves de Aguiar -- 2013.

83f. : il.

Orientador: Prof. Me. Patrick Pedreira Silva.

Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade do Sagrado Coração – Bauru – SP.

1. Chatterbot. 2. AIML. 3. Inteligência artificial. 4. Licenciamento ambiental. 5. Processamento linguagem natural. I. Silva, Patrick Pedreira da. II. Título.

MURILO LUIZETTO GONÇALVES DE AGUIAR

**IMPLEMENTAÇÃO DE UM ATENDENTE VIRTUAL
ESPECIALIZADO LICENCIAMENTO AMBIENTAL UTILIZANDO AIML**

Trabalho de Conclusão de Curso apresentado ao Centro de Ciências Exatas e Sociais Aplicadas como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação, sob orientação do Professor Ms. Patrick Pedreira Silva.

Banca examinadora:

Prof. Ms. Patrick Pedreira Silva
Universidade Sagrado Coração

Prof. Esp. Henrique Pachioni Martins
Universidade Sagrado Coração

Prof. Dr. Elvio Gilberto da Silva
Universidade Sagrado Coração

Bauru, dezembro de 2013.

AGRADECIMENTOS

Agradeço aos meus amigos de faculdade e professores que conviveram nesses quatro anos de faculdade dia a dia comigo.

Agradeço aos meus pais que sempre me orientaram muito bem e me proporcionaram a possibilidade de estudar.

Agradeço ao mestre Patrick Pedreira por acompanhar desde o início esse projeto e pelas informações fornecidas imprescindíveis para a realização desse trabalho.

RESUMO

A Companhia (Cia) Ambiental do Estado de São Paulo (CETESB) é a agência do governo do estado de São Paulo, responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição, com a preocupação fundamental de preservar e recuperar a qualidade da água, do ar e do solo. Essas Informações são disponibilizadas para os usuários, visando atender e orientar quem está ligado ao segmento ambiental. A página da CETESB possui também informações onde o usuário pode se cadastrar para receber os boletins diários por e-mail, consultar boletins anteriores, obter informações a respeito da instituição e sobre o que acontece dia a dia na CETESB. É neste cenário que é apresentada a proposta deste trabalho que consistiu em realizar uma investigação que culminou na proposta de implementação de um protótipo de sistema de Processamento de Linguagem Natural (especificamente um *chatbot*) que foi responsável por permitir a interação entre os usuários e as informações disponibilizadas pela CETESB. O *chatbot* foi implementado na linguagem AIML como um atendente virtual que forneceu uma interface amigável para o usuário que buscava informações ou dados relacionados ao Licenciamento Ambiental, atendendo as necessidades dos usuários. A avaliação do projeto se deu por meio da aplicação de um questionário avaliativo e os resultados mostram o seu potencial para a tarefa proposta nesta investigação.

Palavras-chave: *Chatbot*. Licenciamento Ambiental. Linguagem Natural.AIML.

ABSTRACT

The Environmental Company of São Paulo State is the government agency responsible for the control, supervision, monitoring and licensing of pollution - generating activities, with the primary concern of preserving and restoring the quality water, air and soil. Such information is provided to users to meet and guide who is linked to the environmental segment. The page also has information CETESB where the user can sign up to receive the daily newsletter by email, view past bulletins, information about the institution and what happens day to day in CETESB. Against this backdrop, the aim of this work was to conduct an investigation that culminated in the proposed implementation of a prototype Natural Language Processing (specifically a *chatterbot*) system that was responsible for allowing the interaction between users and the information is presented available by CETESB. The *chatterbot* was implemented in AIML language as a virtual attendant who provided a friendly interface for users seeking information or data related to environmental licensing, meeting the needs of users. The project evaluation was accomplished through the application of an evaluation questionnaire and the results show the potential for the proposed research in this task.

Keywords: *chatterbot*. environmental licensing. natural language. AIML.

LISTA DE FIGURAS

Figura 1: Fases do licenciamento ambiental no Estado de São Paulo	20
Figura 2: A medalha para o vencedor do <i>The Loebner Prize</i>	24
Figura 3: A medalha do <i>The Loebner Prize</i> com o rosto de Alan Turing	24
Figura 4: <i>Interface</i> AliceBot.....	38
Figura 5: <i>Interface</i> do chatterbot Robô ED.	39
Figura 6: <i>Interface</i> do <i>chatterbot</i> Sete Zoom	40
Figura 7: <i>Interface</i> do <i>chatterbot</i> iGod.	40
Figura 8: Estrutura de uma categoria em AIML	43
Figura 9: Exemplo de uso etiqueta <i>random</i>	44
Figura 10: Categoria utilizando etiqueta <i>That</i>	45
Figura 11: exemplo prático utilização predicado.....	46
Figura 12: Exemplo utilização etiqueta <i>Think</i>	46
Figura 13: Exemplo utilização etiqueta <i>Topic</i>	47
Figura 14: Botão <i>Teach</i> e demais ações disponíveis dentro da área administrativa do <i>chatterbot</i>	49
Figura 15: <i>Interface</i> de aprendizagem do <i>bot</i>	49
Figura 16: <i>Interface Clear AIML</i> - limpar categoria	50
Figura 17: <i>Interface Spellcheck</i> - Armazena palavras erradas e as suas correções.....	51
Figura 18: <i>Interface Word Censor</i> - substituir uma palavra por outra mais apropriada.....	51
Figura 19: <i>Interface Search/Edit AIML</i> - Substituir ou editar um AIML.....	52
Figura 20: <i>Interface</i> edição das características do <i>bot</i>	52
Figura 21: <i>Interface</i> de personalização da personalidade do <i>bot</i>	53
Figura 22: Demonstração do Chat.....	53
Figura 23 - <i>Interface</i> do Gaitobot.....	59
Figura 24 - Trecho da base de conhecimentos específica sobre a CETESB. 59	
Figura 25 - Exemplo de categorias para tratar perguntas relativas ao perfil do bot.....	60
Figura 26 - Exemplo de categorias para tratamento das saudações do usuário.....	60

Figura 27 - Categoria que fornece a resposta para perguntas relacionadas a informações relativas a CETESB (localizada no arquivo cetesb.aiml).....	61
Figura 28 - Categoria que trata uma saudação do usuário (localizada no arquivo apresentacoes.aiml).....	62
Figura 29 - Categoria que responde a perguntas relacionadas a aspectos que envolvem o processo de licenciamento ambiental (localizada no arquivo aspectoslic.aiml).	62
Figura30- Categoria que trata sobre itens relacionados a legislação ambiental(localizada no arquivo legislacaoambsp.aiml)	62
Figura 31 - Categorias do arquivo outras.aiml, que trata perguntas variadas, não relacionadas à licenciamento ambiental, à cetesb.....	63
Figura 32 - Categoria que responde a uma pergunta feita pelo usuário sobre os interesses do bot (localizada no arquivo perfil.aiml).....	63
Figura 33 - Categoria que responde a pergunta “Por que o licenciamento ambiental?” (localizada na categoria preliminares.aiml).	64
Figura 34 - Categorias que tratam as diferentes formas de entrada sobre a CETESB que retornam mesma resposta (localizadas no arquivo cetesb_var.aiml).	65
Figura 35 - Categorias (localizadas no arquivo legislacaoambsp_var.aiml) que tratam diferentes formas de entrada para a pergunta “LEI ESTADUAL Nº.997/76 ...”.....	66
Figura 36 - Categorias (localizadas no arquivo perfil_var.aiml) que tratam diferentes formas de entrada para a pergunta “Quem é você”.	66
Figura37 -Exemplo das categorias (localizadas no arquivo apresentacoes_var.aiml) que tratam diferentes entradas que geram a mesma resposta da entrada “Oi”.	67
Figura 38 - Categorias (localizadas no arquivo outras_var.aiml) que tratam diferentes entradas para a mesma resposta fornecida pela categoria “VOCE ESTA COM CALOR”.....	67
Figura 39 - Categorias (localizadas no arquivo preliminareslic_var.aiml) que tratam diferentes formas de entrada para a pergunta “O QUE E LICENCIAMENTO AMBIENTAL?”.....	68

Figura 40 - Categorias (localizadas no arquivo aspectoslic_var.aiml) que tratam diferentes formas de entrada para a pergunta “AREA DE PRESERVAÇÃO PERMANENTE APP – INTERVENÇÃO”.....	68
Figura 41 - Categoria que fornece uma saída padrão caso a resposta do usuário não seja tratada por nenhuma outra categoria.	69
Figura 42: Formulário avaliação do chatterbot.....	73
Figura 43 - Trecho de um registro de conversação entre usuário e chatterbot	74
Figura 44 - Pergunta não reconhecida pelo chatterbot.....	75
Figura 45 - Pergunta passa a ser reconhecida após atualização da BDC.....	75
Figura 46 - Bot não reconhece pergunta devido a erro de digitação do usuário.	76

LISTA DE SIGLAS

AIML - Artificial Intelligence Markup Language

ALICE - Artificial Linguistic Internet Computer Entity

APM - Área de Proteção aos Mananciais da Região Metropolitana de São Paulo

CETESB - Companhia Ambiental do Estado de São Paulo

DEPRN - Departamento Estadual Proteção Recursos Naturais

FAQs - Frequently Asked Questions

HTML - HyperText Markup Language

IA - Inteligência Artificial

PHP - Hypertext Preprocessor

PLN/NLP - Processamento Linguagem Natural

SEAQUA - Sistema Estadual de Administração da Qualidade Ambiental

SIGAM – Sistema Integrado de Gestão Ambiental

SQL - Structured Query Language

SMA – Secretaria do Meio Ambiente

TI - Tecnologia da Informação

TT - Teste de Turing

XML - Extensible Markup Language

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Objetivos	14
2 LICENCIAMENTO AMBIENTAL	15
2.1 A Licença Ambiental	18
2.1.1 Tipos de Licença.....	20
3 INTELIGÊNCIA ARTIFICIAL.....	22
4 PROCESSAMENTO DE LINGUAGEM NATURAL.....	28
4.1 Análise Morfológica.....	30
4.2 Análise Sintática	30
4.3 Análise Semântica	31
4.4 Análise Pragmática.....	32
5 CHATTERBOT	33
5.1 Exemplos DE CHATTERBOTS.....	37
6 AIML	41
6.1 CATEGORY.....	42
6.2 PATTERN	43
6.3 TEMPLATE	43
6.4 RANDOM.....	44
6.5 THAT	45
6.6 PREDICADOS	45
6.7 THINK	46
6.8 TOPIC.....	46
7 FERRAMENTAS PARA DESENVOLVIMENTO DE CHATTERBOTS	47
7.1 Plataforma Program-O e sua utilização no desenvolvimento de chatterbots	47
7.2 Funcionamento do motor de busca.....	48
8 METODOLOGIA.....	54
8.1 Locais de estudo e critérios de amostragem	55
8.2 Softwares e materiais UTILizados como instrumento de coleta de dados	55
8.3 Processo de construção da base de conhecimento.....	56
9 A BASE DE CONHECIMENTO SOBRE LICENCIAMENTO AMBIENTAL	58
10 TESTES COM OS USUÁRIOS.....	70
11 CONSIDERAÇÕES FINAIS	78
REFERÊNCIAS	80

1 INTRODUÇÃO

Segundo Levy (2004), o advento da Internet, prometeu provocar outra revolução, em escala superior à provocada pela imprensa. Agora o cidadão pode usar para as suas aprendizagens, informações que se renovam a cada segundo, providas por fontes dispersas em vários continentes. Por outro lado, sendo o volume muito grande, corremos o risco de sermos sufocados por tanta informação, sem descobrir onde obter os documentos que nos interessam e que sejam relevantes. Tal situação pode, inclusive, provocar um descrédito na tecnologia. Portanto, é fundamental que usuário de internet, principalmente o novato, disponha de facilidades de interação que facilitem a sua integração a este ambiente dinâmico (LEVY, 2004).

Para que a sociedade consiga acompanhar a dinamicidade do mundo, é necessário que as pessoas tenham acesso à informática e a tecnologia. Para que esse acesso seja equivalente a todos os seres, necessita-se de meio de inclusão digital que, segundo Becker (2009), é o nome dado ao processo de democratização do acesso às tecnologias da informação, de forma a permitir a inserção de todos na sociedade da informação.

A computação já faz, há algumas décadas, parte da vida de grande parte da população. Após o surgimento da internet, o computador tornou-se um dos meios principais de obtenção e aquisição de informações (CARVALHO, 2010). Hoje, com a invenção dos *tablets* e *smartphones*, essa informação pode ser acessada a partir de qualquer localização, a qualquer momento (SATO, 2011).

Através de um computador ou dispositivo conectado à internet pode-se obter informações sobre compra de produtos e contratação de serviços, trocar mensagens eletrônicas com amigos, familiares e colegas de trabalho, consultar serviços bancários, entre outras atividades (CARVALHO, 2010).

Carvalho (2010) afirma que embora a internet seja um grande repositório de conhecimento, obter informações sobre assuntos específicos pode ser uma tarefa custosa e ineficaz. Mesmo utilizando poderosas ferramentas de buscas, muitos assuntos possuem informações conflitantes ou até mesmo não são considerados, o

que se deve, em parte, a falta de uma padronização para disponibilizar informações na *Web*, uma vez que o volume de informações, principalmente textuais, é elevado.

A dificuldade que os usuários encontram para acessar os computadores, seja ela por falta de treinamento ou habilidade, é um fator limitativo tão importante que faz com que cientistas e técnicos foquem no desenvolvimento de projetos que facilitem o acesso aos computadores por pessoas com pré-requisitos cada vez menores (CARVALHO, 2003).

A possibilidade de dar a uma máquina a habilidade de interagir com o ser humano através da compreensão e simulação do seu comportamento tem sido alvo de pesquisas na área de Inteligência Artificial (mais especificamente com o estudo do Processamento de Linguagem Natural). A Inteligência Artificial desenvolveu diversas maneiras de simular a comunicação humana, acarretando, entre outras tecnologias, o surgimento dos *chatterbots*, programas de computador que simulam uma conversa com uma pessoa, utilizando, para isso, a linguagem natural (LEONHARDT, 2003). Com o objetivo de, pelo menos temporariamente, levar o ser humano a pensar que está conversando com outra pessoa (LAVEN, 2007). Os *chatterbots* vêm sendo amplamente utilizados na Internet para os mais diversos fins, como conversar em salas de bate papo sobre variados temas, fazer propagandas de produtos, auxiliar como suporte técnico, dentre outros.

Mesmo que os *chatterbots* existam há algumas décadas, existe um crescente aumento na utilização desta tecnologia desde a popularização da *Internet*. A *Internet* fez os *chatterbots* acessíveis ao público em geral e ao mesmo tempo criou para os websites a oportunidade de oferecerem uma *interface* mais amigável para fornecer informações de forma rápida e eficiente a várias pessoas simultaneamente, diminuindo a dificuldade que os usuários sentem para encontrar as informações que necessitam. A linguagem natural permite que o usuário se sinta mais à vontade na troca de mensagens com o sistema, passando a vê-lo não como um equipamento, mas como um “amigo”. (SGANDERLA, 2003). Segundo Primo (2000, p. 5) “robôs de conversação oferecem um forte apelo motivacional”. Os internautas demonstram grande interesse em interagir com esses robôs. Além disso, o fator “inovação” também chama a atenção do público, atraindo-o a utilizar o “sistema”. Neste contexto, o objetivo desta pesquisa será estudar formas de modelagem e desenvolvimento de *chatterbots* que possam oferecer ao usuário uma forma de

interação mais natural que um simples sistema de busca. Para tal, será considerado como foco desta investigação um sistema de conversação que lide especificamente com informações relacionadas à área do segmento açúcar/etanol/energia elétrica. A Companhia Ambiental do Estado de São Paulo (CETESB) é a agência do governo do estado de São Paulo, responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição, com a preocupação fundamental de preservar e recuperar a qualidade das águas, do ar e do solo. (CETESB, 2013).

A CETESB possui um *Website*, onde o usuário pode se cadastrar para receber os boletins diários por *e-mail*, consultar boletins anteriores, obter informações a respeito da instituição e sobre o que acontece dia a dia na CETESB. (CETESB, 2013).

De acordo com levantamento feito, mesmo com os dados disponíveis e centralizados no website da CETESB a qual faz parte da SMA (Secretaria Meio Ambiente), as pessoas têm dificuldades de encontrar as informações sendo, muito comum surgirem dúvidas quando se deseja implantar uma atividade (ou ampliar uma atividade existente) potencialmente poluidora. No Website existe uma série de informações/pré-requisitos e possíveis caminhos que o usuário deveria seguir, mas que, na prática, confundem quase todos os usuários que nunca tiveram contato com as ferramentas e situações do gênero. Esse fato leva o usuário comum a ligar no atendimento telefônico ou tentar sanar sua dúvida através da regional mais próxima. Além disso, por mais simples que seja, há regras para o Licenciamento Ambiental que devem ser seguidas e tal informação deve ficar clara para o usuário.

É neste cenário que foi apresentada a proposta deste trabalho, na qual consistiu em realizar a implementação de um protótipo de sistema de Processamento de Linguagem Natural (especificamente um *chatbot*) que foi responsável por permitir a interação entre os usuários e as informações disponibilizadas pela CETESB. O *chatbot* foi implementado como um atendente virtual que forneceu uma *interface* mais amigável para o usuário que busca informações ou dados. Ele esclareceu as dúvidas relacionadas ao Licenciamento Ambiental e como proceder dentro da lei, assim como prestar informações sobre como realizar o licenciamento de acordo com a atividade a ser realizada, o que contribuiu para diminuir o tempo de busca e possibilitou acesso a resultados que

atenderam as necessidades dos usuários, além de esclarecer dúvidas básicas sobre licenciamento.

Esta monografia está estruturada em 11 capítulos. O capítulo 2 apresenta informações sobre licenciamento ambiental, o capítulo 3 apresenta um breve histórico e os conceitos básicos de Inteligência Artificial. No capítulo 4, é tratada uma das subáreas da Inteligência Artificial, o Processamento de Linguagem Natural. No capítulo 5, a tecnologia dos *chatbots* é apresentada, com histórico e exemplos de *bots*. No capítulo 6 é apresentada a linguagem AIML (*Artificial Intelligence Markup Language*), utilizada na construção de *chatbots*. A plataforma *program-o*, utilizada para hospedagem de *chatbots* é apresentada no capítulo 7. No capítulo 8 é apresentado a metodologia para o desenvolvimento do trabalho. O capítulo 9 apresenta o processo de construção da base de conhecimento do *chatbot*. O capítulo 10 relata os testes com os usuários e o refinamento da base de conhecimento. Por fim, o capítulo 11 apresenta as considerações finais, indicando as limitações do protótipo implementado bem como as sugestões de trabalhos futuros.

1.1 OBJETIVOS

OBJETIVO GERAL:

- Desenvolver um sistema de Processamento de Linguagem Natural (*chatbot*) em AIML, que atuará como um atendente virtual, tirando dúvidas de questões relacionadas ao processo de licenciamento ambiental.

OBJETIVOS ESPECÍFICOS:

- Estudar o Processamento de Linguagem Natural, suas técnicas, ferramentas e limitações;
- Fazer um estudo das técnicas de aquisição e representação do conhecimento;
- Estudar o funcionamento e possíveis tarefas que podem ser realizadas utilizando o motor de busca PROGRAM-O, utilizado no processo de criação de *chatbot*.
- Utilizar a linguagem PHP e um banco de dados MySQL para implementar as perguntas, respostas e tratamento de palavras utilizadas pelo *bot*.

2 LICENCIAMENTO AMBIENTAL

O licenciamento ambiental é um importante instrumento de gestão da Política Nacional de Meio Ambiente. Por meio dele, a administração pública busca exercer o necessário controle sobre as atividades humanas que interferem nas condições ambientais. Desta forma tem, por princípio, a conciliação do desenvolvimento econômico com o uso dos recursos naturais, de modo a assegurar a sustentabilidade dos ecossistemas em suas variabilidades físicas, bióticas, socioculturais e econômicas. Deve, ainda, estar apoiado por outros instrumentos de planejamento de políticas ambientais como a avaliação ambiental estratégica; avaliação ambiental integrada; bem como por outros instrumentos de gestão - zoneamento ecológico econômico, planos de manejo de unidades de conservação, planos de bacia, etc. (Ministério Meio Ambiente, 2013).

O licenciamento é um poderoso mecanismo para incentivar o diálogo setorial, rompendo com a tendência de ações corretivas e individualizadas ao adotar uma postura preventiva, mas proativa, com os diferentes usuários dos recursos naturais. É um momento de aplicação da transversalidade nas políticas setoriais públicas e privadas que interlacom a questão ambiental. A política de transversalidade para o licenciamento é, por definição, uma política de compartilhamento da responsabilidade para a conservação ambiental por meio do desenvolvimento sustentável do país. Para sua efetividade, os preceitos de proteção ambiental devem ser definitivamente incorporados ao planejamento daqueles setores que fazem uso dos recursos naturais (Ministério Meio Ambiente, 2013).

O licenciamento ambiental é o procedimento administrativo pelo qual o órgão ambiental competente analisa a localização, instalação, ampliação e operação de empreendimentos e atividades utilizadoras de recursos ambientais consideradas efetiva ou potencialmente poluidoras, ou daquelas que, sob qualquer forma, possam causar degradação ambiental, considerando as disposições legais e regulamentares e as normas técnicas aplicáveis ao caso (Secretaria Meio Ambiente, 2013).

O licenciamento ambiental no Estado de São Paulo passou a ser obrigatório às atividades industriais após a promulgação do Regulamento da Lei Estadual nº 997/76 aprovado pelo Decreto Estadual nº 8468/76, que dispõe sobre a prevenção e o controle da poluição do meio ambiente. A obtenção das licenças ambientais, aliada

ao cumprimento das exigências técnicas, constitui a base para a conformidade ambiental. O controle da poluição ambiental contemplado nas licenças foca aspectos relativo ao ar, solo, águas, ruído e vibração (Secretaria Meio Ambiente, 2013).

A CETESB – Companhia Ambiental do Estado de São Paulo, subordinada à SMA – Secretaria do Meio Ambiente, é a agência do Governo do Estado responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição, desde a década de 70. A partir de 2009, com a entrada em vigor da lei estadual 13.542, passou a ser a única responsável pelo licenciamento ambiental que antes era realizado por quatro áreas diferentes do sistema estadual de meio ambiente (DEPRN, DUSM, DAIA e CETESB). Assim, hoje o licenciamento de fontes de poluição incluindo aqueles casos envolvendo a avaliação de impacto ambiental, as solicitações de autorização de supressão de vegetação nativa ou intervenção em áreas de preservação permanente e de alvarás para intervenções em APM – Área de Proteção aos Mananciais da Região Metropolitana de São Paulo é realizado pela CETESB. Parte das solicitações de licenciamento e de autorizações pode ser efetivada pela Internet (CETESB, 2013).

Criada em 1986, a Secretaria do Meio Ambiente do Estado de São Paulo – SMA – surgiu para promover a preservação, melhoria e recuperação da qualidade ambiental, coordenando e integrando atividades ligadas à defesa do meio ambiente. Três anos mais tarde, novas atribuições foram conferidas à pasta, que foi a responsável pela elaboração da Política Estadual de Meio Ambiente e pela sua implantação em 1997, que estabeleceu o Sistema Estadual de Administração da Qualidade Ambiental, Proteção, Controle e Desenvolvimento do Meio Ambiente e Uso Adequado dos Recursos Naturais – SEAQUA, do qual a SMA é o órgão central (Secretaria Meio Ambiente, 2013).

Em 2008, a SMA teve a sua estrutura reorganizada, conforme decreto estadual. Desde então, além de coordenar a formulação, aprovação, execução, avaliação e atualização da Política Estadual de Meio Ambiente, a secretaria também ficou responsável por analisar e acompanhar as políticas públicas setoriais que tenham impacto ao meio ambiente, bem como articular e coordenar os planos e ações relacionados à área ambiental. Dessa maneira, as questões ambientais deixaram de integrar apenas a pasta de Meio Ambiente, para estarem presentes em

diferentes órgãos e esferas públicas do Estado de São Paulo, que trabalham de maneira integrada com a SMA.

A secretaria também é responsável por executar as atividades relacionadas ao licenciamento e à fiscalização ambiental, além de promover ações de educação ambiental, normatização, controle, regularização, proteção, conservação e recuperação dos recursos naturais. Para isto, departamentos, coordenadorias e Fundações atuam vinculadas à SMA para exercer as atividades competentes à pasta (Secretaria Meio Ambiente, 2013).

Para agilizar a gestão ambiental no Estado, foram criados dez eixos temáticos que trabalham agendas ambientais em diferentes áreas, como água, solo, resíduos sólidos, economia verde e planejamento ambiental, biodiversidade e unidades de conservação, fiscalização ambiental, cuidado animal, ar, licença ambiental. Tais eixos resultam em mais de 65 projetos/programas com variadas ações. Com um tema universal e transversal, a SMA trabalha integrada a outros órgãos do governo do Estado, além de firmar parcerias com prefeituras, setor privado, organizações não governamentais e instituições de ensino e pesquisa.

A principal função do SIGAM – Sistema Integrado de Gestão Ambiental é registrar, controlar e fornecer informações (operacionais e gerenciais) sobre processos e documentos relacionados às atividades da SMA e seus órgãos vinculados.

Há dois grupos principais de processos e documentos gerenciados pelo SIGAM:

Técnicos: os processos técnicos referem-se às atividades de licenciamento, autorização, fiscalização e gestão ambiental atribuída à SMA e órgãos vinculados. Estes processos são detalhados de modo a permitir a caracterização dos interessados, empreendimentos, atividades e impactos ambientais envolvidos, bem como o registro e anexação de diversos tipos de documentos e informações. Além disso, o SIGAM permite o acompanhamento da movimentação (trâmite) e da situação dos processos, bem como das atividades relacionadas às análises técnicas. A emissão de documentos decorrentes dos processos técnicos também é automatizada através do SIGAM, sendo possível a emissão de boletos de cobrança, pareceres técnicos, notificações, termos, autorizações e licenças ambientais de diversas naturezas, de acordo com as atribuições da Secretaria e órgãos vinculados.

Administrativos: os processos administrativos referem-se eminentemente às atividades de apoio para a Secretaria e órgãos vinculados, como compras, recursos humanos, pagamentos etc. Estes processos são detalhados de modo a permitir a caracterização dos interessados, assunto e acompanhamento de movimentação (tramitação) e situação.

Em ambos os casos o SIGAM oferece apoio operacional e gerencial de acordo com permissões específicas para cada perfil de usuário, com objetivo de facilitar, agilizar e suprir de informações adequadas a todas as unidades (internas e/ou externas) relacionadas à SMA e seus órgãos vinculados.

2.1 A LICENÇA AMBIENTAL

É o ato administrativo pelo qual o órgão ambiental competente estabelece as regras, condições, restrições e medidas de controle ambiental que deverão ser obedecidas pelo empreendedor, pessoa física ou jurídica, para localizar, instalar, ampliar e operar empreendimentos ou atividades utilizadoras dos recursos ambientais, consideradas efetiva ou potencialmente poluidoras, ou aquelas que, sob qualquer forma, possam causar degradação ambiental.

A licença ambiental é uma ferramenta fundamental, pois permite ao empresário tomar conhecimento das possíveis fontes de poluição e de riscos existentes na sua atividade e de que forma estas podem ser controladas.

A licença permite o funcionamento da atividade de forma compatível com os padrões de qualidade ambiental, garantindo o desenvolvimento sustentável.

O controle da poluição ambiental contemplados nas licenças foca aspectos relativos ao ar, solo, águas, ruído e vibração (CETESB, 2013).

De acordo com CETESB (2013) uma pergunta que muitas vezes é feita é porque devemos licenciar uma atividade e alguns dos motivos são:

- **Obrigatoriedade legal**

No Estado de São Paulo, desde 8 de Setembro de 1976, é obrigatório o licenciamento ambiental das atividades industriais. Assim, as empresas instaladas a partir desta data e que funcionam sem a licença estão sujeitas às penas previstas

em lei, tais como: advertências, multas, paralisação temporária ou definitiva da atividade. Incluem-se também as punições relacionadas à Lei de Crimes Ambientais

- **Base estrutural do relacionamento com a sociedade**

A licença constitui uma forma de contrato entre a empresa e o poder público estadual. Por meio dela a empresa conhece seus direitos e obrigações, tornando-se referência para o relacionamento com o órgão ambiental e a sociedade.

Desta forma, o atendimento aos termos exigidos na licença torna-se o principal respaldo da empresa para o equacionamento de eventuais conflitos, como reclamações da comunidade, fiscalização dos órgãos competentes, denúncias de concorrentes e outros.

- **Melhora da imagem pública e acesso a novos mercados**

Estando em conformidade legal, as empresas aumentam sua competitividade e credibilidade junto ao mercado.

Cada vez mais a licença é requisito para obtenção de financiamentos, aprovação da empresa como fornecedora na cadeia produtiva e principalmente na certificação de produtos tanto para o mercado interno quanto para o externo.

- Atividades e empreendimentos sujeitos ao licenciamento ambiental: (CETESB, 2013).

- Indústrias e serviços
- Aquicultura
- Aterros de resíduos inertes e da construção civil
- Aterros Sanitários
- Assentamentos para Reforma Agrária
- Bases de armazenamento
- Cemitérios
- Central de reciclagem de caminhões e outros veículos
- Cogeração de energia
- Coleta, Transporte e Disposição de lodos
- Crematórios
- Depósito ou comércio atacadista de produtos químicos
- Dutos e linhas
- Estações de tratamento de água
- Extração Mineral

- Fabricação de biocombustível (exceto álcool)
- GRAPROHAB
- Hospitais e Similares
- Incineradores de resíduos de serviços de saúde
- Manejo de fauna exótica e criadouro de fauna silvestre
- Outros sistemas de tratamento de resíduos de serviço de saúde
- Parcelamento do solo e Condomínios
- Postos de Combustíveis
- Recebimento de embalagens vazias de agrotóxicos
- Sistemas de tratamento de esgotos sanitários
- Termoelétricas
- Transbordos de resíduos sólidos domiciliares
- Transbordo de resíduos de serviços de saúde
- Usina de açúcar e etanol
- Usina de Compostagem
- Usinas de reciclagem de resíduos da construção civil

2.1.1 TIPOS DE LICENÇA

Segundo CETESB (2013), a licença ambiental é concedida em etapas que podem ser vistas (Figura 1). Dependendo da atividade, o empreendedor obtém primeiramente a Licença Prévia, em separado da Licença de Instalação. Para a maioria das atividades, as Licenças Prévia e de Instalação são concedidas em conjunto, e posteriormente é obtida a Licença de Operação.

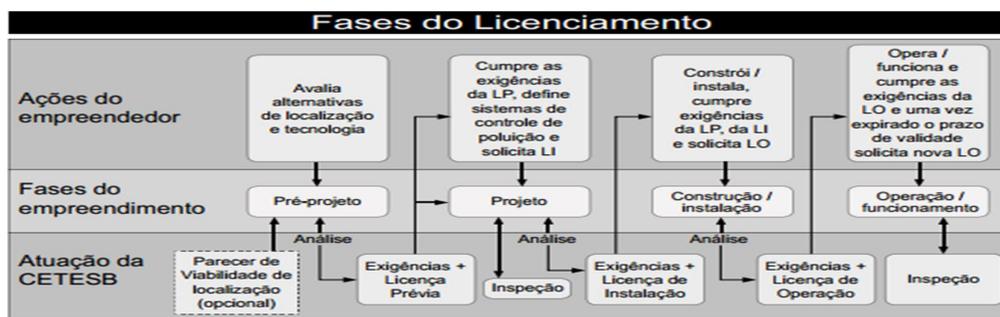


Figura 1: Fases do licenciamento ambiental no Estado de São Paulo
Fonte: CETESB (2013).

Licença Prévia – LP

É a licença dada na fase do planejamento do empreendimento ou atividade, aprovando sua localização e concepção, atestando a viabilidade ambiental e estabelecendo os requisitos básicos e exigências técnicas a serem atendidas nas próximas fases.

Licença de Instalação – LI

É a licença que autoriza a instalação do empreendimento ou de uma determinada atividade de acordo com as especificações constantes dos planos, programas e projetos aprovados, incluindo as medidas de controle ambiental e demais exigências técnicas necessárias.

Licença de Operação – LO

É a licença que autoriza o funcionamento da atividade mediante o cumprimento integral das exigências técnicas contidas na licença de instalação.

Poderá ser emitida Licença de Operação a Título Precário, cujo prazo de validade não poderá ser superior a 180 (cento e oitenta) dias, nos casos em que o funcionamento ou operação da fonte, for necessário para testar a eficiência dos sistemas de controle de poluição ambiental.

Segundo CETESB (2013), a obtenção das licenças é feita analisando primeiramente se a empresa se enquadra no SILIS - Sistema de Licenciamento Simplificado.

O SILIS é um sistema informatizado, baseado na certificação digital, que permite ao empreendimento de baixo potencial poluidor obter, via internet, o seu licenciamento ambiental, por meio de um procedimento simplificado, no qual os documentos Licença Prévia, Licença de Instalação e Licença de Operação são concedidos com a emissão de apenas um documento. Além disso, o SILIS também pode ser utilizado para a renovação da Licença de Operação.

Caso se enquadre no SILIS, a solicitação de licenciamento deve ser feita por meio do site da CETESB (www.cetesb.sp.gov.br/silis).

Caso não se enquadre, a solicitação deve ser feita na Agência Ambiental da CETESB enquadrada no CEP ou Município em que a empresa está instalada, para retirada dos formulários e orientação quanto ao preenchimento dos documentos.

Por conta de toda a complexidade envolvida no tema do licenciamento ambiental, as pessoas que precisam lidar com essas questões acabam por apresentar várias dúvidas. Neste sentido, o uso de tecnologias pode minimizar esse problema, provendo sistemas que sirvam de *interface* entre o conteúdo e o usuário, tornando o seu acesso e entendimento menos complexo. Assim, sistemas que possuem algum nível de inteligência poderão contribuir para esta tarefa. A construção de sistemas com tais características é estudada pela área de Inteligência Artificial que será descrita na seção seguinte.

3 INTELIGÊNCIA ARTIFICIAL

Segundo Winston (1992 apud RUSSEL; NORVIG, 2004, p. 5), a Inteligência Artificial é “o estudo das computações que tornam possível perceber, raciocinar e agir.” Para Russel e Norvig (2004, p. 3), o campo da Inteligência Artificial tenta “[...] não apenas compreender, mas também construir entidades inteligentes.”

A Inteligência Artificial é a arte de criar máquinas que executam tarefas que exigem inteligência quando executadas por pessoas. (KURZWEIL, 1990 citado por RUSSEL; NORVIG, 2004).

Para Luger (2002), a Inteligência Artificial pode ser caracterizada como o ramo da computação que se ocupa da automação do comportamento inteligente e se fundamenta em sólidos princípios teóricos e práticos

Mas a Inteligência Artificial nasceu oficialmente em 1956 com uma conferência em Dartmouth College, nos Estados Unidos. John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon escreveram na proposta dessa conferência, submetida à fundação Rockfeller, a intenção de realizar “um estudo durante dois meses, por dez homens, sobre o tópico inteligência artificial”. (MCCORDUCK, 1979 *apud* BITTENCOURT, 1999).

Segundo Russel e Norvig (2004, p. 19), “talvez o resultado mais duradouro do seminário tenha sido um acordo para adotar o nome sugerido por McCarthy para o campo: Inteligência Artificial”.

A Inteligência Artificial é a área da computação que tenta resolver os problemas computacionais das mais diversas áreas do conhecimento usando o princípio do comportamento inteligente.

O primeiro trabalho agora reconhecido como IA foi realizado por Warren McCulloch e Walter Pitts (1943). Eles se basearam em três fontes: o conhecimento da fisiologia básica e da função dos neurônios no cérebro, uma análise formal da lógica proposicional criada por Russell e Whitehead e a teoria da computação de Turing (RUSSEL, NORVIG, 2003).

A Inteligência Artificial é uma das áreas da computação que mais cresceu nas últimas décadas e podem ser encontradas aplicações em inúmeras áreas do conhecimento. Simulando o comportamento inteligente, essa área auxilia na resolução de problemas que não possuem um algoritmo como forma de resolução.

Segundo Rich (1994, p.3), a IA “permite que objetos tecnológicos tenham ações humanas como de raciocinar, perceber, tomar decisões e resolver problemas, enfim, a capacidade de ser inteligente”.

Segundo Luger (2002), “[...] a palavra “inteligência” vem do latim *inter* (entre) e *legere* (escolher). [...] A palavra “artificial” vem do latim *artificiale*, significa algo não natural, isto é, produzido pelo homem [...]”.

Portanto, pode-se definir Inteligência Artificial como “[...] um tipo de inteligência produzida pelo homem para dotar as máquinas de algum tipo de habilidade que simula a inteligência do homem”(FERNANDES, 2005, p.160).

O desenvolvimento da área começou logo após a Segunda Guerra Mundial, com o artigo "*Computing Machinery and Intelligence*", do matemático inglês Alan Turing, e o próprio nome foi cunhado em 1956. Seus principais idealizadores foram os cientistas Herbert Simon, Allen Newell, John McCarthy, Warren McCulloch, Walter Pitts e Marvin Minsky, entre outros.

Alan Turing foi um dos precursores da Inteligência Artificial e muitos o consideram o pai dessa área. Seus questionamentos a respeito da possibilidade de funções matemáticas expressarem conhecimento e inteligência impulsionaram o desenvolvimento da inteligência artificial. Turing desenvolveu um teste denominado teste de Turing, que consistia basicamente em colocar usuários para dialogar com computadores e com pessoas reais. Para passar no teste a máquina deveria dar respostas aos questionamentos dos humanos de forma que os mesmos não

soubessem distinguir quem é máquina de quem é humano (RUSSEL, NORVIG, 2003).

O teste de Turing foi a base para a criação de um concurso denominado “O prêmio Loebner” (The Loebner Prize), que em 1990, na sua primeira edição, ofereceu um prêmio de 100.000 dólares e uma medalha de ouro para o vencedor conforme pode ser visto na Figura 2 (*THE LOEBNER PRIZE*, s/d). A medalha oferecida pelo concurso possuía em uma de suas faces o rosto de seu idealizador, Hugh G. Loebner. A Figura 3 mostra o outro lado da medalha do *The Loebner Prize* com o rosto de Alan Turing e sua pergunta “Máquinas podem pensar?” (*Can Machines Think?*). Atualmente, o concurso ainda é realizado todos os anos, mas o prêmio é bem mais singelo, são cerca de dois mil dólares e uma medalha de bronze.

Na última edição do concurso em 2008 o vencedor foi Fred Roberts and Artificial Solutions.



Figura 2: A medalha para o vencedor do *The Loebner Prize*
Fonte: The Loebner Prize(2004).



Figura 3: A medalha do *The Loebner Prize* com o rosto de Alan Turing.
Fonte: The Loebner Prize(2004).

Simon (1988 *apud* FERNANDES, 2005) aponta duas abordagens para a Inteligência Artificial, sendo elas a abordagem cognitiva e a abordagem conexionista.

A abordagem cognitiva também chamada de “descendente” ou “simbolista” enfatiza a forma de raciocínio do ser humano, os processos cognitivos. Seu objetivo é explicar os comportamentos inteligentes, se baseando em aspectos psicológicos e processos algorítmicos. Nessa corrente, estão os pesquisadores John McCarthy, Marvin Minsky, Allen Newell e Herbert Simon. (FERNANDES, 2005).

A abordagem conexionista também conhecida por “biológica” ou “ascendente”, destaca o “modelo de funcionamento do cérebro, dos neurônios e das conexões neurais.” (FERNANDES, 2005 p. 3). Entre seus pesquisadores estão McCulloch, Pitts, Helder, Rosenblatt e Widrow. Essa abordagem sofreu um grande impacto quando, segundo Fernandes (2005), Minsky e Papert publicaram o livro *Perceptrons*, em 1969, criticando os modelos das redes neurais e afirmando que esses modelos não possuíam a sustentação matemática necessária para que fosse possível conferir a eles alguma confiabilidade. As pesquisas nessa corrente continuaram, mas só na década de 1980 que John Hopfield conseguiu recuperar a credibilidade no uso das redes neurais.

Segundo Helder Coelho (1994) há uma discussão em que se consideram duas vertentes básicas para a Inteligência Artificial: uma conhecida como "forte" e outra conhecida como "fraca".

-Inteligência artificial forte: Aborda a criação da forma de inteligência baseada em computador que consiga raciocinar e resolver problemas; uma forma de IA forte é classificada como autoconsciente. É um tema bastante controverso. Segundo John Pollock (1989), a tese da IA forte é a de que podemos arquitetar uma pessoa (um ser que literalmente pensa, sente e é consciente) por meio da construção de um sistema físico detentor de “inteligência artificial” adequada.

-Inteligência artificial fraca: Trata-se da noção de como lidar com problemas não determinísticos. Foca a sua investigação na criação de inteligência artificial que não é capaz de, genuinamente, raciocinar e resolver problemas. A máquina com esta característica de inteligência agiria como se fosse inteligente, mas não tem autoconsciência ou noção de si. Segundo Searle, (1986, p.112) “... ainda que uma máquina possa parecer falar chinês por meio de recursos de exame comparativo

com mostras e tabelas de referência, binárias, isso não implica que tal máquina fale e entenda efetivamente a língua”.

Atualmente a Inteligência Artificial tem aplicações em diversos campos. Russel e Norvig (2004) citam como exemplo a aplicação da IA em planejamento autônomo e escalonamento, jogos, controle autônomo, diagnósticos, planejamento logístico, robótica e reconhecimento de linguagem e resolução de problemas.

A seguir é apresentada uma explicação breve de algumas das subáreas da Inteligência Artificial:

- Sistemas Baseados em Conhecimento: Construídos com regras que reproduzem o conhecimento de um perito, são utilizados para solucionar problemas em domínios específicos (MENDES, 1997).

- Lógica Fuzzy: Consiste em aproximar a decisão computacional da decisão humana, tornando as máquinas mais capacitadas a seu trabalho. Isto é feito de forma que a decisão de uma máquina não se resume apenas a um “sim” ou um “não”, mas também tenha decisões “abstratas”, do tipo “um pouco mais”, “talvez sim”, e outras tantas variáveis que representem as decisões humanas. É um modo de interligar inerentemente processos analógicos que se deslocam através de uma faixa contínua para um computador digital que podem ver coisas com valores numéricos bem definidos (ANNES, 2011).

- Processamento de Linguagem Natural: Técnica que permite ao homem se comunicar com a máquina utilizando sua linguagem de conversação. (FÉ, 2005).

- Redes Neurais Artificiais: Sistemas computacionais que imitam as habilidades computacionais do sistema nervoso biológico usando vários neurônios artificiais simples ligados entre si (LOESCH, 1996 apud FERNANDES, 2005).

- Algoritmos Genéticos: “[...] métodos adaptativos que podem ser usados para resolver problemas de busca e otimização. Estão inspirados no processo genético e evolutivo dos organismos vivos.” (FERNANDES, 2005).

- Aprendizagem de Máquina: Desenvolvimento de algoritmos e técnicas que possibilitem ao computador aprender, aperfeiçoando seu desempenho em alguma tarefa (ANNES, 2011).

Simon (1983) define aprendizado como qualquer mudança num sistema que melhore o seu desempenho na segunda vez que ele repetir a mesma tarefa, ou outra tarefa da mesma população.

Aprendizado de Máquina (do inglês, *Machine Learning*) é a área de Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre processo de aprendizado, conclui Bishop (2007).

- Inteligência Artificial Distribuída: “[...] a preocupação no processo de busca de soluções de problemas parte da individualidade para a coletividade.” (FERNANDES, 2005). Quebra-se um problema complexo em problemas menores, a fim de simplificar e facilitar sua resolução.

-Agentes: Segundo Faraco (1998, apud FERNANDES, 2005), agentes são “entidades que se comunicam no ambiente distribuído”, trocando informações e interagindo umas com as outras.

Essa característica aumenta significativamente o poder computacional dos sistemas, uma vez que a linearidade da maioria dos sistemas convencionais é completamente superada pelo paradigma da Inteligência Artificial (RUSSEL & NORVIG, 2003).

Agentes são interativos. Isto Significa que eles formam uma coleção de indivíduos que cooperam numa tarefa particular. Neste sentido, eles podem ser vistos como uma “sociedade” e, como na sociedade humana , o conhecimento , as habilidades e as responsabilidades , mesmo quando vistas como coletiva, estão distribuídas pela população de indivíduos (LUGER, 2002).

Os agentes inteligentes, ou seja, sistemas computacionais que utilizam Inteligência Artificial na resolução dos problemas são sistemas que utilizam sensores para fazer a varredura no ambiente e obter informações sobre onde o agente está inserido e através de atuadores interagem com o meio e exteriorizar o processamento inteligente realizado pelo agente (RUSSEL, NORVIG, 2003).

Esses agentes estão presentes em várias ferramentas e muitas vezes passam despercebidas. Robôs para busca de informações na internet, sistemas baseados em conhecimentos e sistemas especialistas são apenas alguns exemplos de aplicações que utilizam a Inteligência Artificial. A área tende a alcançar um crescimento ainda maior, sobretudo com o incremento da computação ubíqua, ou seja, introdução de computação cada vez menos perceptível no dia a- dia das pessoas.

A visão da Inteligência Artificial, substituindo julgamento humano profissional, tem surgido muitas vezes na história, em Ficção Científica e, hoje em dia, em algumas áreas especializadas em que "Sistemas Especialistas" são usados para melhorar ou para substituir o julgamento humano. Autores, como Wittgenstein (citado por Shanker, 1987, p.28) e Piaget (1968, p. 121-152), sustentam que mesmo nesses termos não seria possível afirmar que as máquinas pensam, pois o pensamento seria uma característica humana, não se tratando de um processo mecânico estruturado por regras lógicas. Caracterizá-lo dessa forma seria transgredir sua própria natureza. Todavia, deve ser lembrando a definição de pensamento proposta por Turing (1950, p.59), "pensamento é o processo pelo qual se resolve problemas manipulando-se informações".

Assim, algo que pensa é algo que resolve problemas, não importando se o processo pelo qual se chega às respostas seja lógico-formal e mecanicista ou baseado em modelos mentais e estratégias. Dessa forma, há tanto homens quanto máquinas plenamente capazes de resolver problemas.

4 PROCESSAMENTO DE LINGUAGEM NATURAL

Rich (1988 p. 344) afirma que "os computadores não serão capazes de realizar muitas das tarefas cotidianas das pessoas antes que possam partilhar a capacidade de utilizar a linguagem".

Segundo Silva (2012), o significado de linguagem natural já está implícito em seu próprio nome. O processamento de linguagem natural (NLP, ou PLN, em português) é uma técnica que permite que o homem se comunique com uma máquina utilizando-se de sua linguagem de conversação, estabelecendo assim uma comunicação da maneira mais "natural" possível.

Esta comunicação consiste no processo de formularem-se frases, sendo estas elaboradas quase de maneira automática na mente humana para expressar seus sentimentos. Ao expor uma frase, não há a preocupação com toda a análise sintática e/ou morfológica inclusa nesta, saindo esta automaticamente da maneira correta para a comunicação no meio do qual se encontra, não necessariamente sendo plenamente culta (SILVA, 2012).

Como ferramenta fundamental no entendimento da linguagem realizada pelos *chatterbots*, temos uma técnica de interpretação e processamento das entradas do usuário, denominada Processamento de Linguagem Natural.

Segundo Othero (2006), a Linguística Computacional é a área responsável pela investigação do tratamento computacional da linguagem e das línguas naturais. Ela explora a relação entre linguística e informática, possibilitando a construção de sistemas que são capazes de reconhecer e produzir informação apresentada em linguagem natural (VIEIRA; LIMA, 2001).

A Linguística Computacional se divide em duas subáreas: a Linguística de Corpus e o Processamento de Linguagem Natural. (Othero, 2006).

Essa subárea da IA tem aplicações em diversas atividades. É aplicada na recuperação de informações, que trata da “indexação, busca e classificação de documentos (textuais), com o objetivo de satisfazer necessidades de informação de seus usuários, expressas através de consultas.” (GONZALEZ; LIMA, 2003, p. 1).

O PLN permite fazer a correção ortográfica de textos, possibilitando a detecção erros e oferece possíveis correções, bem como na geração automática de textos em linguagem natural por computadores. Outra aplicação possível é na tradução de textos, área em que, segundo Santos (2001) o PLN foi primeiramente empregado. Alguns exemplos de outras aplicações são o reconhecimento de voz, o reconhecimento de texto manuscrito e a segurança e identificação. (SANTOS, 2001).

É dito que um sistema trata linguagem natural se duas condições forem satisfeitas: A entrada e/ou a saída do sistema deve ser em linguagem natural, e o processamento da entrada e/ou a geração da saída deve ser baseada no conhecimento de aspectos sintáticos, semânticos e/ou pragmáticos de linguagem natural (FÉ, 2005).

Ainda segundo Fé (2005), para obedecer a essas condições, o sistema não pode tratar a linguagem simplesmente como *strings*. Ele deve interpretar as sentenças e armazenar informações morfológicas, sintáticas e semânticas em um dicionário, junto com as palavras que são compreendidas pelo sistema.

4.1 ANÁLISE MORFOLÓGICA

Segundo Oliveira (2002), a análise morfológica é essencial para que uma frase seja compreendida, já que para formar uma estrutura coerente da sentença precisa-se compreender o que cada uma das palavras que compõem essa sentença significa. O analisador morfológico reconhece palavras e expressões isoladas em uma determinada sentença. Para esse processo, considera os delimitadores contidos na sentença, como a pontuação e os espaços em branco. As palavras identificadas são classificadas de acordo com seu tipo de uso, ou categoria gramatical. Dentro de um mesmo tipo estão grupos de regras que caracterizam o comportamento de um subconjunto de vocábulos da linguagem (exemplo: formação do plural de substantivos terminados em “ão”, flexões de verbos regulares terminados em “ar”, entre outros). Podemos dizer então que 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).

Rich e Knight (1993 apud OLIVEIRA, 2002) ainda cita um exemplo de como o analisador morfológico trata a frase “Eu quero imprimir o arquivo .init do Mário”. Ele deve separar a expressão “do Mário”, ficando com o substantivo próprio “Mário”, a preposição “de” e o artigo “o”. “Também deve reconhecer a sequência “.init” como uma extensão de um arquivo, que funciona como um adjetivo. Todas as palavras são então classificadas de acordo com suas categorias sintáticas.

4.2 ANÁLISE SINTÁTICA

O analisador sintático denominado *parser*, recebe as informações do analisador morfológico. Essas informações juntamente com a gramática da linguagem serão usadas para a construção de árvores de derivação para cada sentença responsável por mostrar o relacionamento entre as palavras.

Para a construção da árvore de derivação é checado se as sequências de palavras se enquadram nas regras de construção estabelecidas pela linguagem para a construção de frases. Entre essas regras estão a concordância e regência nominal e verbal e o posicionamento de termos na frase. Um termo é um elemento de informação (palavra ou expressão) é tratado como unidade funcional da oração e

participa da estrutura como um de seus constituintes, os sintagmas. Para a análise sintática de uma oração na língua portuguesa devem-se considerar os seguintes sintagmas: termos nominais (sujeito e predicado), termos integrantes (complementos verbal e nominal) e termos acessórios (adjunto adverbial, adjunto adnominal e aposto). Já para a análise sintática de um período deve-se considerar o tipo desse período (simples ou composto), sua composição (por subordinação ou por coordenação) e a classificação das orações (absoluta, principal, coordenada ou subordinada). (OLIVEIRA, 2002).

Nos sistemas de processamento de linguagem natural, o maior problema é a transformação de uma frase potencialmente ambígua em uma não ambígua, a qual será utilizada pelo sistema. Essa transformação é conhecida como *parsing*. (OLIVEIRA, 2002).

Entre as abordagens de linguagens formais utilizadas no estudo da análise sintática, Oliveira (2002) cita:

- Gramáticas Regulares: São simples e reconhecidas com facilidade, mas possuem um poder limitado de expressão.

- Gramáticas Livres de Contexto: Permitem a representação de linguagens com certo grau de complexidade, sendo mais poderosas que as gramáticas regulares, mas ainda possuem uma dificuldade para expressar dependências simples, como concordância entre verbo e sintagma nominal.

- Gramáticas Sensíveis ao Contexto: Apesar de não apresentarem os problemas de dependência das gramáticas livres de contexto, não abordam satisfatoriamente o tratamento de restrições gramaticais. O que impede sua utilização é a questão do reconhecimento. “O problema de decidir se uma sentença pertence a uma gramática sensível ao contexto é uma função exponencial sobre o tamanho da sentença” (OLIVEIRA, 2002), o que, do ponto de vista computacional, torna essa implementação da verificação uma questão complexa.

4.3 ANÁLISE SEMÂNTICA

As estruturas criadas pelo analisador sintático possuem, também, um significado (semântica). A questão da representação do significado apresenta várias dificuldades, como ambiguidades (palavras que podem assumir vários significados,

como “tomar banho” ou “tomar algo de alguém”), ou a distinção entre significado e sentido. O analisador semântico analisa o sentido da estrutura de palavras que foram reagrupadas pelo analisador sintático, uma vez que o analisador morfológico permitiu identificar estas palavras individualmente.

Segundo Oliveira (2002), a análise semântica é decomposta em léxica e gramatical. A léxica procura uma representação conceitual para descrever o sentido, e para construir essa representação pode ser feita uma decomposição semântica das unidades léxicas ou utilizar de redes semânticas.

A compreensão da relação entre as palavras é tão importante quanto a compreensão das próprias palavras. Enfoques formais para a semântica gramatical tentam descrever o sentido de uma frase mediante a tradução de sua estrutura sintática para uma fórmula lógica-semântica. Como não existe uma correspondência imediata e biunívoca entre sintaxe e semântica, uma mesma estrutura sintática pode dar origem a diferentes representações semânticas. Por exemplo, uma professora de capoeira pernambucana pode referir-se a uma pessoa nascida em Pernambuco, a qual ensina capoeira, ou a uma pessoa que ensina capoeira no estilo em que esta é praticada em Pernambuco(OLIVEIRA, 2002).

4.4 ANÁLISE PRAGMÁTICA

A última, e mais difícil etapa a se solucionar é a pragmática. Além de verificar toda a parte gramatical e o sentido da frase, a mesma verifica a coerência desta. Uma frase pode satisfazer os três etapas, mas a frase: “A Alemanha é tricampeã da Copa América” é totalmente fora de nexos, pois um país europeu nunca pode disputar uma Copa América. Aliando-se este a um contexto, se exponencia este problema.

Segundo Fé (2005), no caso de frases em que o efeito desejado é claramente declarativo, o processo poderia ser encerrado após a etapa análise semântica. Entretanto algumas frases possuem um efeito almejado diferente, como frases irônicas e comentários sarcásticos. Assim, precisa-se fazer a interpretação do todo.

A análise pragmática irá buscar compreender o contexto da frase que está sendo analisada nas outras frases do texto ou diálogo, ou seja, não vai analisar apenas a estrutura de palavras ou uma frase específica (MÜLLER, 2003).

5 CHATTERBOT

Chatterbot é um programa de computador que tenta simular um ser humano na conversação com as pessoas. O objetivo é responder as perguntas de tal forma que as pessoas tenham a impressão de estar conversando com outra pessoa e não com um programa de computador (TEIXEIRA; MENEZES, 2003). Após o envio de perguntas em linguagem natural o programa consulta uma base de conhecimento e em seguida fornece uma resposta que tenta imitar o comportamento humano.

Essa possibilidade de se dar a uma máquina habilidade para interagir com o ser humano, através de uma compreensão e simulação do seu comportamento, tem sido, há muito tempo, alvo de pesquisas na área de Inteligência Artificial. Nesse contexto, foram surgindo, ao longo dos anos, diversas categorias de *chatterbots* são utilizados na *internet*: comerciais, de entretenimento, *FAQs*, de busca, educacionais, entre outros. (TEIXEIRA; MENEZES, 2003).

O conhecimento do personagem interativo é determinado pelo desenvolvedor, de forma que o personagem pode tornar-se especialista em qualquer assunto, desde que todos os tópicos correspondentes lhe tenham sido passados durante a elaboração de seu perfil. A aplicação depende apenas de qual conteúdo é ensinado ao personagem, e várias funcionalidades podem estar presentes em um mesmo bot. (TEIXEIRA; MENEZES, 2003).

Segundo Fé, (2005) algumas das principais aplicações dos *bots* atualmente são:

- Atendimento automático: Um atendente virtual pode resolver questões e dúvidas de usuários sem a necessidade de interação de humanos ou atendentes reais de plantão no site. Pode ser amplamente utilizado em comércio eletrônico para tirar dúvidas sobre produtos ou ajudar clientes com dúvidas sobre produtos já adquiridos.
- *E-learning*: Personagens virtuais podem se tornar assistentes treinados com conteúdo de aulas para dar suporte a professores e alunos, podendo conversar sobre os assuntos específicos das diversas matérias dos cursos. Unindo esta tecnologia com tecnologias de conversão de texto em áudio (*Text-to-Speech*) e de reconhecimento de voz, podem-se tornar ferramentas muito poderosas no treinamento de novos idiomas.

- *Call Center*: Aqui também se pode fazer bom uso da tecnologia *Text-to-Speech* e de reconhecimento de voz. Dessa forma, os *scripts* de atendimento de *Call Centers* podem ser convertidos para a linguagem do atendente virtual, fornecendo um atendimento adequado ao cliente sem a necessidade de atendentes reais.
- Salas de *chat*: Em salas de bate-papo, personagens virtuais podem interagir com outros usuários e com outros *bots*, de maneira que, trocando informações entre si, podem servir como uma forte ferramenta estatística através do armazenamento e análise dos diálogos e tópicos abordados.
- Entretenimento: Como podem ter qualquer tipo de perfil, personagens virtuais causam curiosidade, atraindo usuários para conversarem. Podem, portanto, tornar-se “amigos” do usuário e conversar sobre diversos assuntos.
- CRM: Em todas as aplicações citadas anteriormente, as informações coletadas sobre os usuários podem ser armazenadas em bases de dados, podendo também ser consolidadas na forma de relatórios de pesquisas. Assim podem-se realizar análises estatísticas sobre preferências dos clientes, sobre seus dados de compra ou mesmo sobre seus dados pessoais, sendo o *bot* capaz de reconhecer o cliente nas conversas seguintes.
- Sistemas de Busca: A aplicação dos personagens virtuais em sistemas de busca pode criar uma *interface* mais amigável, de maneira que, através de um diálogo, o usuário chegue à informação desejada sem precisar utilizar os sistemas de busca convencionais, que podem retornar milhares de opções ao usuário.
- Informação: Podem ser criados *bots* com a finalidade específica de fornecer notícias sobre atualidades, política, economia, esportes ou qualquer assunto específico. As notícias podem ser alimentadas diariamente no “cérebro” do *bot*, fazendo assim com que ele esteja sempre atualizado e possa informar o usuário sobre o assunto desejado.
- Jogos: Atualmente já são usados, em alguns jogos, *bots* que podem dialogar com o jogador, fazendo assim com que os personagens artificiais do jogo tenham uma interação mais realista e possam realizar várias tarefas durante a brincadeira, como fornecer instruções sobre o jogo, sobre o

enredo em que o jogo está contextualizado ou mesmo fazer provocações ao jogador no caso dele cometer erros ou ser derrotado.

Antes de se classificar os *chatbots* é importante esclarecer a diferença entre *chatbots* e agentes. Segundo Primo (2000), nem todo agente é um agente de conversação. *Software* agente é um sinônimo para *software robot* ou simplesmente “*bot*”. Existe agente que busca por ofertas, sugere produtos, etc., mas não entende linguagem natural.

Um *chatbot* é um tipo específico de *bot* capaz de entender linguagem natural. Talvez, o termo “entender” seja um tanto ambicioso, como diz Primo (2000), já que reconhecer signos não quer dizer propriamente que haja uma verdadeira compreensão por parte do robô.

De acordo com Franklin (1996) deve-se entender que um robô de conversação não compreende a fala, mas sim, reconhece a linguagem podendo reagir a ela. De forma geral, os *chatbots* podem ser analisados segundo uma série de características dentre as quais se destaca:

- Capacidade de aprender – capacidade do *chatbot* de aprender durante o diálogo com o usuário. O *chatbot* pode aprender sobre o domínio e sobre o usuário. Este tipo de *chatbot* tem uma maior probabilidade de manter diálogo por mais tempo sem ser percebido o fato de ser um programa.
- Memória – capacidade de lembrar os diálogos passados e sentenças mencionadas previamente dentro de um mesmo diálogo.
- Domínio – possíveis temas para diálogo. Alguns *chatbots* não fazem restrição de domínio, contudo podem ser especialistas em determinados assuntos. Em outros *chatbots* existe a opção do usuário escolher qual o domínio da conversa.
- Robustez – capacidade do robô de responder a sentenças do usuário não reconhecidas.
- Autoconhecimento – capacidade de o robô falar sobre si mesmo.

Os *chatbots* podem ser classificados segundo sua área de aplicação como de entretenimentos, de busca, acadêmicos, de comércio, dentre outros. Laven (2000) apresenta uma classificação que considera os recursos utilizados. Nesta classificação os *chatbots* podem ser encontrados nas seguintes categorias: clássicos, complexos, amigáveis e ensináveis.

-Clássicos: *chatterbots* criados com o objetivo inicial de estudar a complexidade na comunicação em linguagem natural entre homens e máquinas.

-Complexos: criados a partir de linguagens de programação mais sofisticadas. A ideia principal neste programa é avaliar a natureza da comunicação entre homem e computador. Seu funcionamento é baseado em padrões de mapeamentos das perguntas mais comuns de usuários nas respostas adequadas.

- Amigáveis: são capazes de reproduzir o comportamento de uma pessoa.

- Ensináveis: *chatterbots* que possuem capacidade de aprendizagem

Independente do tipo de classificação utilizada, os *chatterbots* apresentam uma série de vantagens citadas a seguir:

- Um robô não precisa de salário, vale alimentação ou vale transporte;

-O vocabulário e respostas do robô são totalmente personalizáveis;

-O robô fica 24 horas disponíveis;

- Todas as mensagens trocadas entre o robô e usuários são gravadas e podem ser visualizadas posteriormente;

- O custo de manutenção muito baixo.

O primeiro *chatterbot* de que se tem conhecimento é o ELIZA, desenvolvido no MIT (*Massachusetts Institute of Technology*) por Joseph Weizenbaum, entre os anos de 1964 e 1966. Este programa tinha como objetivo simular uma conversa entre uma psicóloga (*bot*) e um paciente (usuário). ELIZA possui uma personalidade amigável e estimula o usuário a contar mais detalhes sobre seus problemas através de perguntas.

Segundo Leonhardt (2003), ELIZA possui algumas limitações. Uma delas é a ausência de uma memória para o *bot*. Devido a isso, este *chatterbot* não consegue se lembrar de informações fornecidas anteriormente pelo usuário com quem está conversando. Outra limitação é quanto à construção de algumas sentenças. ELIZA usa partes da entrada fornecida pelo usuário para elaborar suas perguntas, na tentativa de dar uma maior naturalidade para o diálogo. Porém muitas vezes essa técnica pode gerar frases confusas, desanimando o usuário a continuar conversando.

Na seção seguinte serão mostrados alguns exemplos de *chatbots* e suas características.

5.1 EXEMPLOS DE CHATTERBOTS

Segundo Alcione (2010), em pesquisas realizadas, descobriu-se que já existem diversos *chatbots*, os quais possuem como função atender clientes de empresas. Cada vez mais as empresas investem nessa nova modalidade de atendimento. Os *chatbots* estão inclusos em diversas áreas, como: saúde, educação, comercial, governamental, entre outras. A seguir, destacam-se os principais agentes de conversação disponíveis na internet.

A.L.I.C.E.: O *chatbot* A.L.I.C.E. foi desenvolvido em 1995 por Richard Wallace, que também é responsável pelo desenvolvimento doo AIML (*Artificial Intelligence Markup Language*), uma linguagem baseada no XML (*eXtensible Markup Language*), criada com a finalidade para o desenvolvimento de robôs de chat. Foi vencedor do prêmio Loebner, uma competição anual onde é aplicado o Teste de Turing, nos anos de 2000 e 2001(WALLACE, 2009).

A.L.I.C.E. contém uma documentação extensa e conta com uma fundação para ajudar a promover o *software* gratuito A.L.I.C.E. e a linguagem AIML. Por este motivo existem muitos outros *chatbots* que utilizam o *software* do A.L.I.C.E., sendo necessária apenas a construção de uma nova base de conhecimento em AIML. (LEONHARDT, 2003). A Figura 4 mostra a *interface* de A.L.I.C.E

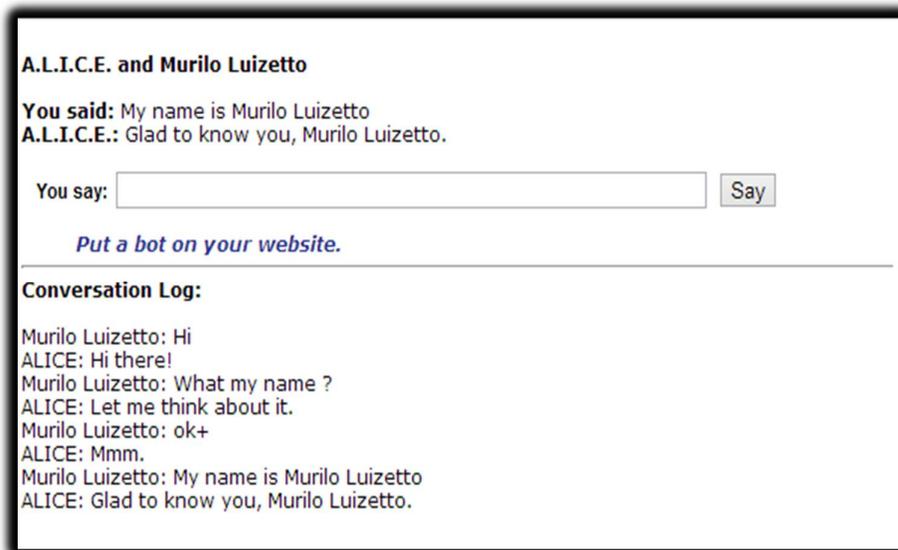


Figura 4: Interface AliceBot.
Fonte: Alicebot.org (2013).

ELIZA: Iniciando a primeira geração de *chatbot* em 1966 e desenvolvido por Joseph Weizenbaum, no *Massachusetts Institute of Technology* (Instituto de Tecnologia de Massachusetts), nasceu o primeiro *chatbot*, Eliza. Este tinha por objetivo fazer uma psicanálise e diagnosticar, por meio de diálogos amigáveis, problemas psicológicos em pacientes. Já naquela época obtinha-se êxito. Porém, Eliza não mantinha em base as conversas realizadas anteriormente, com isso deixava um pouco a desejar (LEONHARDT, 2003).

Cybelle (Nacional): Já no Brasil, o estudo de *chatbots* é mais recente. Criado por Alex Primo, em 2000, o primeiro *chatbot* brasileiro é chamado de Cybelle. Com poder de expressões emocionais, como tristeza, humor, entre outros, isso ajudou a quebrar as limitações com o humano em relação à comunicação, além, é claro, de propiciar um bom bate-papo para descontrair. Cybelle possui uma base de conhecimento composta por várias respostas pré-programadas, possibilitando maior abrangência e flexibilidade com o humano (LEONHARDT, 2003).

Ed: Este *chatbot* (Figura 5) foi desenvolvido pela Insite e lançado pela Petrobrás em 20 de outubro de 2004, surgiu para interagir com o meio social em um público-alvo infanto-juvenil, com a intenção de conscientizar quanto ao consumo racional energético brasileiro. Integrado ao *site* da Conpet (Programa Nacional de

Racionalização do Uso dos Derivados do Petróleo e do Gás Natural), fornece amplo material, como vídeos, apostilas e centenas de links para saciar as dúvidas em questão (OLIVEIRA, 2004).

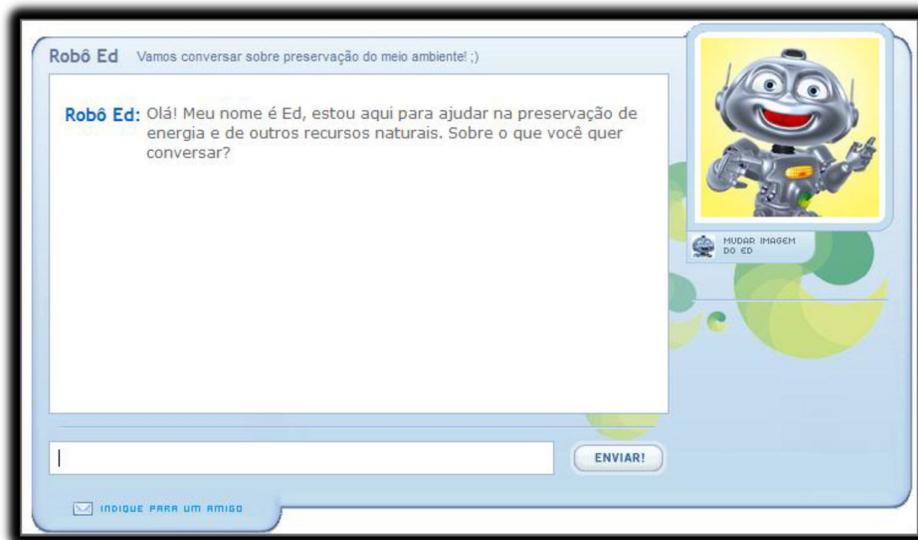


Figura 5: Interface do chatterbot Robô ED.
Fonte: Ed.conpet.gov.br.(2013).

Sete Zoom: Também desenvolvida pela Insite, para o *site* da Close Up, pela sua personalidade atrevida, é considerada a modelo virtual mais famosa do Brasil. Vários contextos e características de comunicação humana foram implementados como pode ser visto na (figura 6); possui capacidade de bolar ideias a partir de assuntos abordados durante a conversa. Suas respostas são baseadas em uma personalidade forte por meio de uma base de conhecimento pré-alimentada pelos programadores. Em sua implementação, destacam a habilidade de memorizar assuntos já conversados, como seu nome, por exemplo. Tem capacidade de direcionamento de assunto baseado na inserção do internauta, além de apresentar e elaborar ideias (FERREIRA, 2008).



Figura 6: Interface do *chatbot* Sete Zoom
 Fonte: SeteZoom.com. (2013).

Para finalizar os exemplos de *chatbot*, existe também o iGod, um *chatbot* que tem atribuído a ele o perfil de deus. O *bot* está disponível na *web* e pode ser acessado no endereço <http://www.titane.ca/main.html>. Na Figura 7 é exibida a imagem de sua *interface*. Essa *interface* é composta por um campo para a entrada da pergunta do usuário e o histórico da conversa.



Figura 7: Interface do *chatbot* iGod.
 Fonte: Titane.ca.(2013).

O processo de desenvolvimento dos *chatterbots* pode se dar de diversas maneiras, sendo uma das mais utilizadas aquela que faz uso de uma linguagem específica para modelagem de uma base de conhecimento: a AIML (*Artificial Intelligence Markup Language*). As características da linguagem AIML são descritas na seção seguinte.

6 AIML

A linguagem AIML (*Artificial Intelligence Markup Language*) é utilizada para o desenvolvimento das bases de conhecimento dos *chatterbots*. A AIML é uma especificação da linguagem XML para a construção de perfis de *chatterbots* (FÉ, 2005).

O XML é uma linguagem dinâmica com ampla liberdade de programação. O usuário pode definir etiquetas novas e atribuir valores a elas. A linguagem AIML é um exemplo do padrão XML que define uma linguagem especializada para a utilização em Inteligência Artificial. As vantagens da utilização do padrão XML são as diversas ferramentas existentes para edição e manipulação de arquivos XML, sua facilidade de aprendizado, e a conveniência quanto à programação de *chatterbots* via *web*. O grande trunfo do AIML é a simplicidade dessa linguagem(FÉ, 2005).

O AIML foi desenvolvido pela comunidade Alicebot durante o período de 1995-2000 com uma gramática proprietária também chamada AIML que formava a base para o primeiro Alicebot. Após diversas atualizações visando à padronização da gramática foi adotado o XML (*eXtensible Markup Language*) (LEONHARDT, 2003).

Para a criação do AIML, algumas metas e objetivos foram estabelecidos (BUSH, 2001):

- deve ser de fácil aprendizagem;
- deve ser compatível com o XML;
- deve codificar o conjunto de conceitos mínimos necessário para permitir um sistema de conhecimento do tipo estímulo-resposta modelado no *chatterbot* A.L.I.C.E. original;
- seu *design* deveria ser formal e conciso;

- o desenvolvimento de programas que processam documentos AIML deve ser fácil;
- objetos AIML devem ser legíveis para o ser humano e razoavelmente claros;
- não deve incorporar dependências de nenhuma linguagem.

No código AIML um *chatbot* é definido entre as *etiquetas* `<aimlversion="xxx">...</aiml>`. Entre essas etiquetas estarão suas categorias, padrões e modelos, necessárias para identificação e criação da base de conhecimento (RINGATE, 2001).

O AIML não é um processador de linguagem natural, por não formular ele próprio as suas frases completamente. As frases do AIML já vêm pré-moldadas, sofrendo poucas adaptações no decorrer de sua utilização, enquanto raramente um PLN repetirá uma frase já dita.

A vantagem de utilizar-se o AIML é de que a análise pragmática da frase é feita automaticamente pelo *Botmaster* durante sua programação, mas exige um esforço humano intenso para ensinar tudo o que o robô necessita falar. Entretanto, um PLN maduro pode responder uma infinidade de frases, mas o pode desviar o foco da conversação ou expor uma frase indesejável, como “Não gosto de meu mestre”. Fica provada a diferença de direções destas duas tecnologias: uma é focada na geração de frases, e outra em manter uma conversação.

A satisfação na resolução do teste de Turing só pode ser alcançada com uma união destas duas vertentes de resolução da linguagem, pois caso um *chatbot* AIML não possua a devida resposta, o mesmo pode recorrer a um PLN para resolvê-la.

Os principais elementos da linguagem AIML estão listados nas subseções seguintes.

6.1 CATEGORY

Na linguagem AIML a unidade básica de conhecimento é denominada categoria. Cada categoria contém uma pergunta de entrada (estímulo), uma resposta de saída, e um contexto opcional. A pergunta é chamada de padrão (*pattern*), e a resposta é conhecida como modelo (*template*). O contexto opcional pode ser de dois tipos, que são chamados de “*that*” e “*topic*”, cujas definições

dessas etiquetas serão descritas nos próximos tópicos(WALLACE, 2009). A Figura 8 a seguir mostra um exemplo de uso desta etiqueta (*category*):

```

1 <category>
2     <pattern>Pergunta</pattern>
3     <template>Resposta</template>
4 </category>
5

```

Figura 8: Estrutura de uma categoria em AIML
Fonte: Elaborada pelo autor (2013).

6.2 PATTERN

A etiqueta `<pattern>` só é permitida dentro das etiquetas `<category>`, que definem uma categoria AIML. É a primeira etiqueta a aparecer após o início da categoria, e só pode existir uma etiqueta *pattern* por categoria (RINGATE, 2001). Ainda segundo Ringate (2001), o *pattern* (padrão) é uma expressão em linguagem formal que consiste de:

- Palavras em linguagem natural, escritas em letras maiúsculas;
- O símbolo `*`, que é usado para representar uma sequência de uma ou mais palavras. Na documentação do AIML, temos um exemplo de uso desse símbolo. Considerando os *patterns* OLA e OLA `*`. Enquanto OLA irá só corresponder a frases contendo apenas a palavra “Hello”, a expressão OLA `*` irá corresponder a qualquer frase com duas ou mais palavras que inicie com a palavra “Olá”, como na frase “Olá como vai você?”. Por isso chama-se `*` de coringa, pois ele compreende maneiras diferentes de perguntar que envolvem o mesmo conceito ou uso da palavra no caso OLA.
- o símbolo `_`, que tem a mesma função que o `*`, porém vem após a letra Z na ordem lexicográfica;

6.3 TEMPLATE

A etiqueta *template* contém a resposta (saída) para o usuário, e é parte de uma categoria AIML. A linguagem oferece funções de marcação para que cada

resposta possa ser adequada a cada entrada individual ou usuário. O *template* também pode conter códigos HTML, Javascript e XML, que permitem que o navegador emita um som, mostre uma imagem ou rode um *applet* (programa na linguagem de programação Java que pode ser incluído e executado no navegador) (RINGATE, 2001).

O *template* pode incluir as etiquetas `<sr>` (sigla que significa *Stimulus-Response Artificial Intelligence*) e `<sr/>` (RINGATE, 2001).

Segundo Wallace (2009), a etiqueta `<sr>` é utilizada para lidar com sinônimos, ou seja, entradas de usuário que significam a mesma coisa, e com isso o *chatbot* deve retornar uma resposta em comum a qualquer uma dessas entradas.

Categorias onde *template* contém a etiqueta `<sr>` possuem somente a função de “mapear” essa entrada do usuário para outras entradas de outras categorias, seja para simplificar a linguagem ou para lidar com sinônimos. As categorias com tal funcionalidade são nomeadas recursivas (RINGATE, 2001).

6.4 RANDOM

Dentro do *template* podemos usar também a *etiqueta* `<random></random>`. Essa etiqueta permite que o *bot* escolha uma resposta aleatoriamente para retornar ao usuário.

Entre as etiquetas *random* estão as etiquetas ``, que se refere a cada uma das respostas que o *chatbot* pode escolher. Um *template* pode conter mais de um conjunto de etiquetas *random*, assim o *bot* pode combinar frases de diferentes conjuntos de respostas (RINGATE, 2001). A Figura 9 exibe um código exemplo de uso da etiqueta *random*.

```
1 <random>
2 <li>O ramo de atuação</li>
3 <li>A área de atuação</li>
4 <li>O setor de atuação</li>
5 </random>
6
7
```

Figura 9: Exemplo de uso etiqueta *random*
Fonte: Elaborada pelo autor (2013).

6.5 THAT

Esta é uma etiqueta opcional que pode estar contida dentro de uma categoria AIML. Refere-se a qualquer coisa que o *bot* tenha dito antes da última entrada do usuário. É comum o uso em casos de perguntas onde são usadas como resposta sim ou não (RINGATE, 2001).

Esta categoria lida com a entrada do usuário “SIM/NÃO” e checa se o usuário está respondendo a uma pergunta em específico. (RINGATE, 2001). O uso desta etiqueta pode ser visto na Figura 10.

```

1 <category>
2     <pattern>SIM</pattern>
3     <that>Você gosta de seriados</that>
4     <template>Qual é seu seriado preferido?</template>
5 </category>
6

```

Figura 10: Categoria utilizando etiqueta That
Fonte: Elaborada pelo autor (2013).

6.6 PREDICADOS

Predicados AIML podem ser declarados, recuperados e manipulados a qualquer momento da execução do *chatbot*. Um nome de predicado é composto por uma cadeia de caracteres contendo um ou mais caracteres normais. (RINGATE, 2001).

O elemento “*set*” é usado para atribuir um valor a um predicado, definido por um nome (RINGATE, 2001).

Caso o *chatbot* necessite retornar o valor armazenado em um predicado para o usuário, o elemento “*get*” será utilizado. Na Figura 11 podemos ver o resultado do uso dessa etiqueta

```

1 <category>
2     <pattern>Meu nome é *</pattern>
3     <template>
4         Prazer em te conhecer amigo, <set name="nome"><star/></set>
5     </template>
6 </category>

```

Figura 11: exemplo prático utilização predicado
Fonte: Ringate (2001).

6.7 THINK

O elemento “*think*” é utilizado em situações em que o *chatterbot* deve processar a entrada do usuário, mas sem retornar valor algum resultante desse processamento, a não serem os valores necessários que irão compor a saída. (RINGATE, 2001). O resultado do uso da etiqueta pode ser visto na Figura 12.

```

1 <category>
2     <pattern>EU SOU HOMEM</pattern>
3     <template>
4         Obrigado por informar.
5         <think><set name="genero">masculino</set></think>
6     </template>
7 </category>
8

```

Figura 12: Exemplo utilização etiqueta *Think*
Fonte: Ringate (2001).

6.8 TOPIC

A etiqueta “*topic*” é uma das etiquetas fundamentais para que exista coerência no diálogo entre *bot* e o usuário. Ela permite ao *bot* selecionar, preferencialmente, respostas que estejam relacionadas ao tópico que está sendo debatido no momento (FÉ, 2005).

As etiquetas “*topic*” são colocadas em volta de uma ou mais categorias. Estas categorias dentro entre as etiquetas “*topic*” são associadas a um determinado assunto. O nome do assunto é passado pela propriedade “*name*”, contida na etiqueta <topic name=“xxx”> (RINGATE, 2001).

A Figura 13 mostra um exemplo de utilização da etiqueta “*topic*”. No código, o tópico sendo discutido na conversa é “Secretaria Meio Ambiente”. Dentro deste tópico está a categoria que trata da entrada do usuário “Quando foi inaugurada?”, retornando a informação sobre a data de inauguração da Secretaria do Meio Ambiente. O exemplo mostra apenas uma categoria inclusa dentro da etiqueta “*topic*”, mas é comum que esta etiqueta englobe várias categorias.

```

1 <topic name="Secretaria Meio Ambiente">
2 <category>
3     <pattern> Quando foi inaugurada</pattern>
4     <template>
5         Criada em 1986 a Secretaria do Meio Ambiente do Estado de São Paulo
6         surgiu para promover a preservação, melhoria e recuperação
7         da qualidade ambiental, coordenando e integrando atividades ligadas
8         a defesa do meio ambiente
9     </template>
10 </category>
11 </topic>
12

```

Figura 13: Exemplo utilização etiqueta *Topic*
 Fonte: Elaborada pelo autor (2013).

7 FERRAMENTAS PARA DESENVOLVIMENTO DE CHATTERBOTS

7.1 PLATAFORMA PROGRAM-O E SUA UTILIZAÇÃO NO DESENVOLVIMENTO DE CHATTERBOTS

O motor de busca *program-o* é um interpretador de linguagem AIML atual e conciso, que utiliza a linguagem PHP e banco MySQL. Oferece através de *scripts* algumas funcionalidades na criação de *chatterbots*.

O motor de busca *program-o* foi desenvolvido por Elizabeth Perreau e teve código fonte aberto a partir de janeiro de 2009 e desde então vem recebendo atualizações de uma grande comunidade de desenvolvedores ao redor do mundo, o que torna justificável e plausível a escolha desse motor de busca para desenvolvimento do *chatterbot*.

A plataforma conta com os seguintes diretórios:

Pasta Admin: Onde se encontram todos os scripts para a parte administrativa do motor de busca.

Pasta: Bot: Contém os arquivos necessários para o *bot* funcionar, como *scripts* para buscar a resposta no bando de dados.

Arquivo Index.php: Página inicial do motor de busca, onde de fato poderemos conversar com nosso personagem.

Arquivo install_programo.php: *Script* para instalar o motor de busca no servidor *web*.

7.2 FUNCIONAMENTO DO MOTOR DE BUSCA

O *chatterbot* possui uma área administrativa que dá acesso a tudo que a plataforma oferece de opções e ações que podem ser utilizadas pelo *bot* (é possível escolher uma ação específica na lista de ações (*Actions*) existentes na plataforma).

O *program-o* funciona de acordo com a lógica da linguagem AIML, permitindo a criação de um *chatterbot* por meio da criação de sentenças de acordo com a linguagem AIML na área administrativa do *chatterbot*. Esta área administrativa oferece uma série de opções, tais como:

- Logs: Exibe o histórico das conversas do *chatterbot*;
- BotPersonality: Permite alterar informações relacionadas a personalidade do *chatterbot*, como religião, idade e filme favorito;
- Teach: Permite adicionar a “inteligência” do *chatterbot*, através de perguntas e respostas baseadas em AIML;
- Upload AIML: Permite adicionar arquivos AIML ao banco de dados. Isso feito através da linguagem SQL;
- Search/Edit AIML: Permite pesquisar e editar as sentenças em AIML do banco de dados;
- Demo Chat: Para visualizar uma demonstração do *chatterbot*;

Por meio da ferramenta é possível criar o *chatterbot* de acordo com um objetivo, adicionando palavras ou expressões comuns dentro de um cenário.

Essas opções de “ensinar um *bot*”, “ver os *logs*”, “limpar categoria”, e “definir uma personalidade” podem ser vistas nas Figuras 14 e 15.



Figura 14: Botão *Teach* e demais ações disponíveis dentro da área administrativa do *chatterbot*
 Fonte: Área administrativa do *chatterbot*(2013).

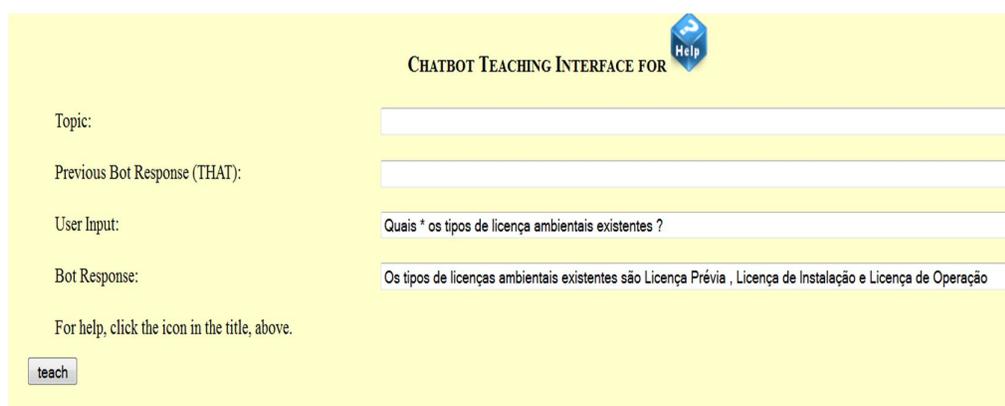


Figura 15: Interface de aprendizagem do *bot*.
 Fonte: Área Administrativa – Interface de aprendizagem do *bot* (*Teaching Interface*) (2013).

Na figura 15 os campos relacionados à *interface* de aprendizagem são:

- Topic: Indica a qual tópico ou categoria a pergunta pertence.
- Previous Bot Res: Última resposta do *chatterbot*.
- User Input: Pergunta digitada pelo usuário.
- Bot Response: Resposta do *chatterbot*

O preenchimento destas informações permite que a pergunta e resposta configuradas sejam adicionadas na tabela AIML

A *interface* oferece também a possibilidade de que links sejam associados às respostas fornecidas ao usuário. A seguir, um possível exemplo de pergunta/resposta que utiliza essa característica.

Pergunta do usuário: * licenciamento ambiental

Resposta do *chatterbot*. De uma olhada nessa página:

<http://www.cetesb.sp.gov.br/licenciamento/licenciamento-ambiental/1-pagina-inicial>

Esse exemplo, em particular, oferece a possibilidade do usuário clicar em um *link* e ser redirecionado a uma página que responda a sua dúvida. “A pergunta poderia ser: “Como é o processo de licenciamento ambiental?”, “Qual o foco do licenciamento ambiental?”, “Fale sobre o licenciamento ambiental”, etc. isso porque foi feito o uso do coringa “*”.

A seguir algumas imagens que mostram a *interface* de cada umas das ações possíveis dentro da plataforma *program-o*.

Na Figura 16 temos a ação *Clear AIML Categories* que permite que você apague uma categoria do *bot* selecionado em questão. Essa função é importante a partir do momento que você vai moldando o *bot* de acordo com sua funcionalidade e exigência de informações, e é necessário apagar alguma informação para outra ser colocada no lugar

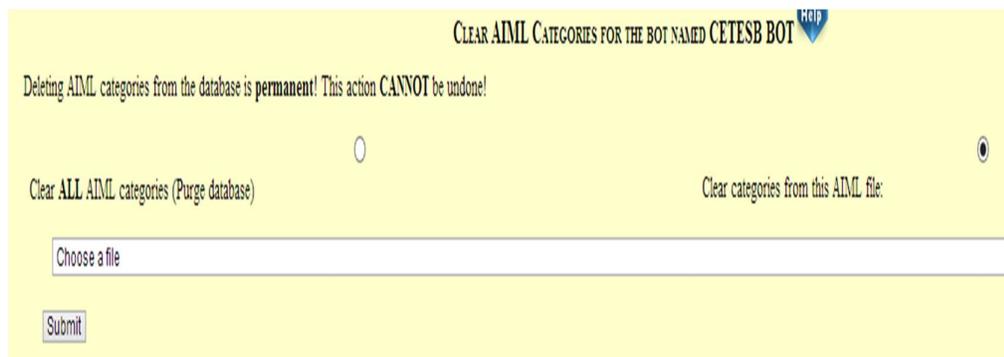


Figura 16: *Interface Clear AIML* - limpar categoria
Fonte: Área Administrativa do *Chatterbot*(2013).

Na Figura 17 temos o campo *Spellcheck* que armazena palavras erradas e as suas correções. Mesmo que o usuário digite errado, o *chatterbot* vai entender, e compreender o que foi dito para seguir adiante a conversação.

Isso é importante, pois como muitas pessoas tem costume de usar abreviações e maneiras distintas de escrever algumas palavras isso faz com que o *bot* tenha uma maior capacidade de entendimento do que está sendo dito a ele facilitando no processo de comunicação.

SPELLCHECK EDITOR

The Spell checker addon is **enabled**
Note: To enable or disable the spell checker, you will need to edit the config file. Please see the README file in the chatbot/addons/spell_checker folder for more details.

Search for a Misspelled word:

Add a Misspelled word and it's correction:

Misspell:

Correction:

Figura 17: Interface Spellcheck - Armazena palavras erradas e as suas correções.
 Fonte: Area Administrativa do Chatterbot(2013).

Na Figura 18 se encontra a *interface Word Censor* que tem como funcionalidade a função de substituir uma palavra por outra mais apropriada para o contexto e conversação.

WORD CENSOR EDITOR

Search for an improper word:

Add an improper word and it's replacement:

Bad word:

Replacement:

Figura 18: Interface Word Censor - substituir uma palavra por outra mais apropriada
 Fonte: Área Administrativa do Chatterbot(2013).

A Figura 19 ilustra a *interface Search/Edit AIML* que permite que procure/edite um comando AIML. Se você deseja procurar um *That*, *Template* ou *Topic* do seu *bot* aqui é onde você pode realizar esta função.

Figura 19: Interface Search/Edit AIML - Substituir ou editar um AIML
 Fonte: Área Administrativa do *Chatterbot*(2013).

A Figura 20 apresenta a *interface* para edição das características do *bot*, podendo alterar o nome do *bot*, sua descrição, tipo de formato das respostas, *email* para contato, mensagem de erro quando algo for digitado e não possuir resposta, etc.

Figura 20: Interface edição das características do *bot*
 Fonte: Área Administrativa do *Chatterbot*(2013).

A Figura 21 exibe a *interface* de personalização do *bot* onde podemos colocar informações relativas à personalidade como nome, idade, tipo, categoria dentre outras características que tornam o *bot* mais específico possível.

BOT PERSONALITY SETTINGS FOR CETESB BOT

Name:	Gender:	Botmaster:
Master:	Age:	Website:
Email:	Birthday:	Birthplace:
Size:	Version:	Build:
Language:	Feelings:	Etype:
Emotions:	Ethics:	Orientation:
Baseballteam:	Footballteam:	Hockeyteam:
Favoritesport:	Vocabulary:	Celebrities:
Celebrity:	Favoriteactor:	Favoriteactress:
Favoriteartist:	Favoritemusician:	Favoritesong:
Favoriteauthor:	Friend:	Nationality:
Religion:	President:	Party:
Kingdom:	Phylum:	Class:
Order:	Family:	Genus:
Species:	Boyfriend:	Favoritebook:
Favoriteband:	Favoritecolor:	Favoritefood:
Favoritemovie:	Forum:	Friends:
Girlfriend:	Kindmusic:	Location:
Friendship:	Character:	Room:

Figura 21: Interface de personalização da personalidade do *bot*
 Fonte: Área Administrativa do *Chatterbot*(2013).

A Figura 22 apresenta um exemplo simples através da *interface chat demo* que permite realizar um teste simples de seu *bot* onde podemos fazer uma pergunta e em seguida o *bot* responde de acordo com o que foi armazenado na base de dados

CHAT DEMO

Talk to CETESB BOT using the HTML interface:

Program O Example GUI Page - HTML

Say:

Seeker: What is the importance of the environmental licensing?
 CETESB BOT: The importance is to make the companies are within the expected environmental regulation to start a business

Figura 22: Demonstração do Chat
 Fonte: Área Administrativa do *Chatterbot*(2013).

8 METODOLOGIA

Segundo Gil (2010, p. 27), as pesquisas exploratórias têm como finalidade proporcionar maior familiaridade com o problema, com intuito torná-lo mais explícito ou permitir a construção de hipóteses. Seu planejamento tende a ser bastante flexível, pois interessa considerar os mais variados aspectos relativos ao fato ou fenômeno estudado. O autor também conclui que a maioria das pesquisas realizadas com propósitos acadêmicos, pelo menos num primeiro instante, assume o caráter de pesquisa exploratória, pois nesta ocasião é pouco provável que o pesquisador tenha uma definição clara do que irá investigar.

Portanto, este projeto é inicialmente uma pesquisa exploratória, pois visa estudar a integração de três linguagens (MySQL, PHP e AIML), na construção de um protótipo específico, no caso, um *chatbot* especializado no processo de licenciamento ambiental de uma atividade.

Este projeto foi desenvolvido em duas fases distintas: uma fase de averiguação dos aspectos teóricos, e uma etapa prática de implementação do sistema.

Na primeira fase foi feito um estudo teórico dos diversos assuntos relacionados com o escopo do trabalho, envolvendo áreas como Inteligência Artificial, Processamento de Linguagem Natural, AIML, *chatbot*, Licenciamento Ambiental. Como produto desta fase foi proposto um modelo, para o protótipo do *chatbot*. Este modelo consistiu em definir quais as etiquetas AIML deveriam ser utilizadas para a implementação do protótipo de forma a manter a consistência da base de conhecimentos, bem como a coerência das respostas dadas pelo *chatbot*. As etiquetas (*tags*) utilizadas foram aquelas descritas no capítulo referente a AIML

Na segunda fase foi realizada a implementação do protótipo proposto (escrita da base de conhecimento em AIML) e os testes necessários.

8.1 LOCAIS DE ESTUDOS E CRITÉRIOS DE AMOSTRAGEM

O levantamento das informações inseridas na base de conhecimento do *chatterbot* foi feito através de consulta bibliográfica, pesquisa no *Website* da CETESB e entrevistas com profissionais que atuam na área ambiental.

8.2 SOFTWARES E MATERIAIS UTILIZADOS COMO INSTRUMENTO DE COLETA DE DADOS

O protótipo do *chatterbot* foi desenvolvido utilizando o motor de busca *program-o*. As bases de conhecimento foram escritas na linguagem AIML, que é baseada na linguagem XML, e desenvolvida com a finalidade de criar conversações em linguagem natural. A plataforma Program-o foi escolhida, pois a construção de uma máquina de inferência não traria nenhum benefício adicional ao projeto, uma vez que a plataforma fornece todas as funcionalidades necessárias para construção do agente com rapidez e controle, deixando o construtor livre para enfatizar o trabalho no que de fato é importante em um agente de diálogo, ou seja, sua base de conhecimento.

O *software* utilizado para criar o *chatterbot* especializado no processo de licenciamento ambiental foi o PHP EDITOR 2.22, que é um editor gratuito e totalmente em português, destinado à edição de páginas PHP e HTML. Também foram utilizadas outras duas ferramentas: o Apache (servidor HTTP *Web*), que foi utilizado para armazenar os dados do projeto, devido à reconhecida qualidade desta ferramenta que é a mais utilizada no mundo; e o PHPMyAdmin, que é um gerenciador de banco de dados.

Para utilização do mesmo foi necessária a instalação de um servidor *Web* que, neste caso, o WampServer 2.1e-x64 (*Wamp*) que inclui um conjunto de programas que instala automaticamente o Apache 2.4.2, MySQL 5.5.24, PHP 5.4.3, phpMyAdmin 3.4.10.1, SQLBuddy 1.3.3, webgrind 1.0. Como banco de dados vinculado ao servidor, foi utilizado o MySQL.

8.3 PROCESSO DE CONSTRUÇÃO DA BASE DE CONHECIMENTO

Para a construção da base de conhecimento proposta neste projeto, foi necessário primeiramente instalar o motor de busca. Para isso, é necessário que o computador tenha um servidor Web instalado. O servidor utilizado é o programa WampServer 2.1e-x64 (*Wamp*), que é um pacote de programas que instala automaticamente o Apache 2.4.2, PHP5, MySQL 5.5.24 *Database*, phpMyAdmin 3.4.10.1 e SQLiteManager. Sendo que cada um dos programas instalados tem a sua função durante o desenvolvimento de um *chatbot*.

Depois de instalado o servidor *web*, o mesmo deve ser acessado para configuração de todos os *scripts* da parte administrativa como, por exemplo, os arquivos necessários para o *bot* funcionar (*scripts* para buscar a resposta no banco de dados). Adicionalmente, foi criada a página inicial do motor de busca, que funciona como a *interface* de conversação com o *chatbot*, permitindo assim, a interação com o usuário. Todo esse processo de instalação do motor de busca envolve a criação de um banco de dados para armazenar as suas informações, para este estudo implementou-se um banco de dados chamado “banco” usando o aplicativo phpMyAdmin. Posteriormente, foram adicionados alguns *scripts* AIML ao banco de dados para simular uma inteligência básica no *chatbot*.

Para que o *chatbot* desenvolvido seja capaz de responder as principais dúvidas de profissionais e demais pessoas que necessitam dessas informações presentes no site da Companhia Ambiental do Estado de São Paulo (CETESB), foi necessário inseri-las em sua base de dados. Tudo isso usando uma linguagem não muito formal, a fim de deixar o *chatbot* o mais “humano” possível. Foram selecionadas e incluídas as dúvidas mais comuns no banco de dados, conforme o levantamento feito durante a fase de coleta de dados sobre licenciamento ambiental.

Portanto, para a criação da base de conhecimentos houve a necessidade de uma análise minuciosa do que seria disponibilizado nesta base. O estudo dos documentos referentes ao processo de licenciamento ambiental foi fundamental para elaboração de perguntas e eventuais respostas do sistema. Cada tópico sobre o qual o *chatbot* poderia conversar foi definido como uma categoria na linguagem AIML, sendo que dentro desta categoria existem várias possibilidades de entrada e saída, correspondentes às perguntas feitas por usuários e às respostas que o

chatterbot pode fornecer a estas perguntas. A seguir, um exemplo de uma pergunta que consta na base de conhecimento e a resposta associada.

1) Pergunta do usuário: Quais * os tipos de licença ambientais existentes ?

Resposta do *chatterbot*: Os tipos de licenças ambientais existentes são Licença Prévia, Licença de Instalação e Licença de Operação.

9 A BASE DE CONHECIMENTO SOBRE LICENCIAMENTO AMBIENTAL

Primeiramente o chatterbot foi criado sem nenhum conhecimento já que a plataforma program-o não fornece nenhuma base de conhecimento padrão em português. A base de conhecimento específico do chatterbot foi construída a partir de pesquisa bibliográfica e informações fornecidas por profissionais da área ambiental. Assim, foi possível fazer com que o bot respondesse as perguntas mais frequentes e requisitadas sobre a CETESB e o licenciamento ambiental.

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 que fazem uso de interfaces gráficas. Um software conhecido para a criação desses motores robóticos na linguagem de marcação AIML é o GaitoBot que foi utilizado juntamente do Notepad++ para criação e edição das bases de conhecimento (BDC) .

O software GaitoBot (Figura 23) 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.

Além da base de conhecimentos específicos sobre preliminares do licenciamento, licenciamento ambiental e sobre a CETESB (exemplo visto na Figura 24), também foi incluído o tratamento para entradas como apresentação do usuário (saudações) e perguntas relacionadas ao perfil do chatterbot, como idade, localização, nome, etc; conforme podem ser vistas nas Figuras 25 e 26. Essas entradas permitem dar à conversa entre o usuário e o *chatterbot* um tom menos formal e mais “humanizado”.

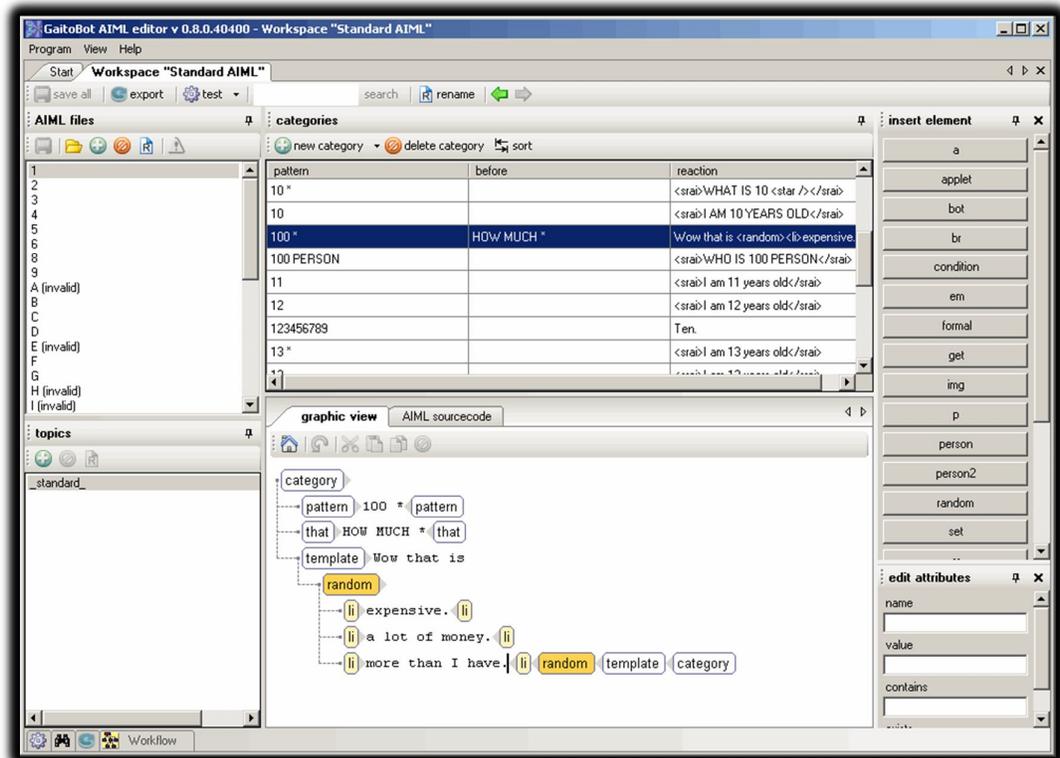


Figura 23 - Interface do Gaitobot

Fonte: Gaitobot.de/gaitobot/. Capturado em 21 novembro 2013.

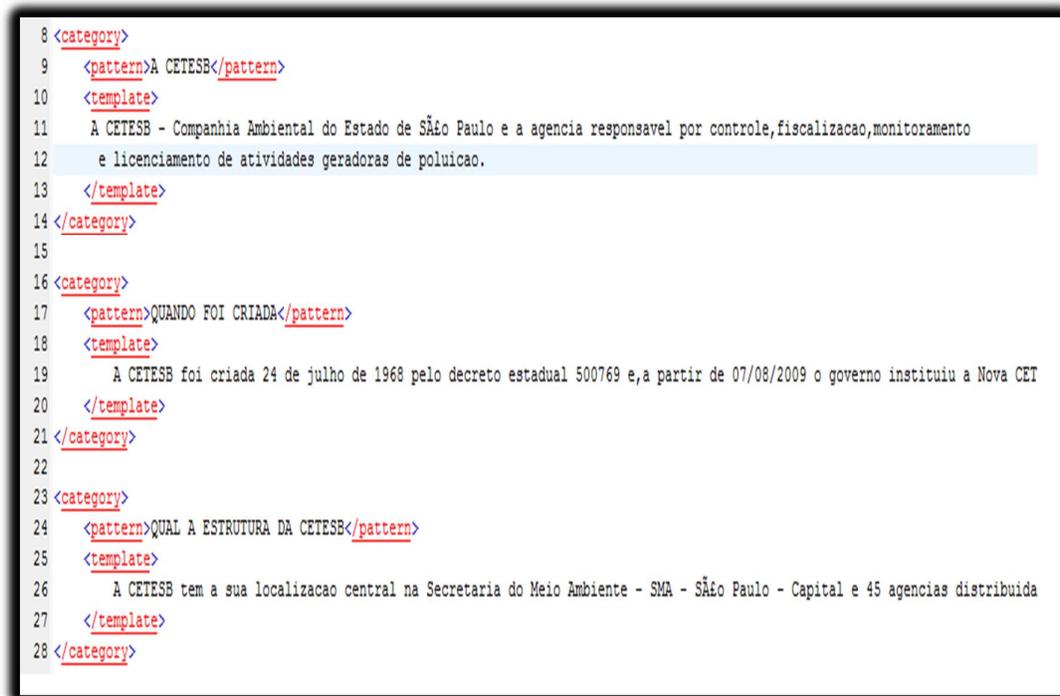


Figura 24 - Trecho da base de conhecimentos específica sobre a CETESB.

Fonte: Elaborada pelo autor (2013).

```

8 <category>
9   <pattern>INTERESSES</pattern>
10  <template>Gosto de várias coisas, mas meu interesse principal é o meio ambiente.</template>
11 </category>
12
13 <category>
14   <pattern>TELEFONE</pattern>
15   <template>Você pode contatar a CETESB pelo telefone <bot name="telefone"/> ou pelo email <bot name="email"/>.
16   </template>
17 </category>
18
19 <category>
20   <pattern>RELIGIAO</pattern>
21   <template>Robô's não tem religião.</template>
22 </category>
23
24 <category>
25   <pattern>GENERO</pattern>
26   <template>
27     <random>
28       <li>Eu sou um <bot name="genero"/>.</li>
29       <li>Meu gênero é <bot name="genero"/>.</li>
30     </random>
31   </template>
32 </category>

```

Figura 25 - Exemplo de categorias para tratar perguntas relativas ao perfil do bot.
Fonte: Elaborada pelo autor (2013).

```

8 <category>
9   <pattern>OI</pattern>
10  <template>
11    <random>
12      <li>Olá; humano!</li>
13      <li>Olá;! Seja bem vindo.</li>
14    </random>
15  </template>
16 </category>
17
18 <category>
19   <pattern>TUDO CERTO</pattern>
20   <template>Sim, tudo conforme.</template>
21 </category>
22
23 <category>
24   <pattern>MEU NOME É: *</pattern>
25   <template>
26     <random>
27       <li>
28         Prazer em te conhecer, <set name="nome_usuario"><star index="1"/></set>
29       </li>
30       <li>
31         Seja bem vindo, <set name="nome_usuario"><star index="1"/></set>
32       </li>
33     </random>
34   </template>
35 </category>

```

Figura 26 - Exemplo de categorias para tratamento das saudações do usuário.
Fonte: Elaborada pelo autor (2013).

As categorias contendo a base de conhecimento foram divididas de acordo com sua função e tópico em quinze arquivos AIML.

Sete dos quinze arquivos que compõe a base são arquivos AIML com a forma mais simples de categoria AIML, ou seja, que não possuem recursões ou os símbolos '*' ou '_'.

Essas categorias contêm as saídas que o bot fornecerá ao usuário, sendo esses arquivos AIML: cetesb.aiml, apresentacoes.aiml, aspectoslic.aiml , legislacaoambsp.aiml, outras.aiml , perfil.aiml, e preliminareslic.aiml. As Figuras 27, 28, 29, 30, 31, 32 e 33 mostram respectivamente exemplos de trechos dos arquivos contendo esse conhecimento específico.

Na figura 27 é apresentada uma categoria que fornece respostas sobre à Companhia Ambiental do Estado de São Paulo (CETESB) com algumas das perguntas e respostas mais solicitadas e relacionadas a mesma.

```

38 <category>
39   <pattern>RELACAO CETESB PREFEITURAS</pattern>
40   <template>
41     Com a disponibilidade de 45 agencias ambientais no estado,
42     a CETESB permite fortalecer a celebracao de convenios com varias prefeituras
43     para descentralizar licenciamento de atividades e empreendimentos de pequeno impacto local.
44   </template>
45 </category>
46
47 <category>
48   <pattern>RELACOES CETESB COM CORTE DE VEGETACAO</pattern>
49   <template>
50     A CETESB passou a licenciar atividades que impliquem no corte de vegetacao
51     e intervencoes em areas consideradas de preservacao permanente e ambientalmente protegida
52   </template>
53 </category>
54
55 <category>
56   <pattern>CONSULTAR DADOS AMBIENTAIS</pattern>
57   <template>
58     Os dados ficam disponiveis no site da CETESB.
59     O endereço e <a href="http://www.cetesb.sp.gov.br/" target="_blank">aqui</a>.
60   </template>
61 </category>
62
63 <category>
64   <pattern>QUAL A DOCUMENTACAO NECESSARIA PARA REALIZAR O LICENCIAMENTO NA CETESB</pattern>
65   <template>
66     A documentacao necessaria pode ser encontrada no link abaixo:
67     <a href="http://licenciamento.cetesb.sp.gov.br/cetesb/lic_previa_solicitacao.asp" target="_blank">aqui</a>
68   </template>
69 </category>

```

Figura 27 - Categoria que fornece a resposta para perguntas relacionadas a informações relativas a CETESB (localizada no arquivo cetesb.aiml).

Fonte: Elaborada pelo autor (2013).

Na Figura 28 é apresentada uma categoria chamada apresentação utilizada no processo de saudação e diálogo entre as partes.

```

8 <category>
9   <pattern>OI</pattern>
10  <template>
11    <random>
12      <li>Olá! humano!</li>
13      <li>Olá! Seja bem vindo.</li>
14    </random>
15  </template>
16 </category>

```

Figura 28 - Categoria que trata uma saudação do usuário (localizada no arquivo apresentacoes.aiml).

Fonte: Elaborada pelo autor (2013)

A Figura 29 ilustra um exemplo categoria que responde questões relacionadas a aspectos que envolvem o processo de licenciamento ambiental.

```

9 <category>
10 <pattern>FONTES DE POLUIÇÃO</pattern>
11 <template>
12   São consideradas fontes de poluição todas as obras, atividades, instalações,
13   empreendimentos, processos, dispositivos, móveis ou imóveis,
14   ou meios de transportes que, direta ou indiretamente,
15   causem ou possam causar poluição ao meio ambiente.
16   Entende-se como fontes móveis todos os veículos automotores, embarcações e assemelhados,
17   e como fontes estacionárias, todas as demais.
18 </template>
19 </category>

```

Figura 29 - Categoria que responde a perguntas relacionadas a aspectos que envolvem o processo de licenciamento ambiental (localizada no arquivo aspectoslic.aiml).

Fonte: Elaborada pelo autor (2013).

A Figura 30 mostra uma categoria que trata sobre as principais leis no âmbito estadual e federal, permitindo saber a partir destas variações as demais leis.

```

18 <category>
19 <pattern>LEGISLAÇÃO ESTADUAL</pattern>
20 <template>
21   A Legislação Estadual tem a sua essência também lastreada na constituição federal, acrescida da Constituição Estadual - SP
22   em especial no capítulo IV, pela Emenda Constitucional n°.23 ano 2007
23   e na sequência de Leis, Decretos, Resoluções, Portarias, Deliberações e Outras (Comunicados, Decisões de Diretoria, etc).
24   Devido à dinâmica da legislação, as consultas devem ser permanentemente atualizadas sob pena de cair em caducidade.
25 </template>
26 </category>

```

Figura 30 - Categoria que trata sobre itens relacionados a legislação ambiental (localizada no arquivo legislacaoambsp.aiml)

Fonte: Elaborada pelo autor (2013)

A Figura 31 mostra um exemplo da categoria “outras” que fornece respostas a perguntas variadas não relacionadas a CETESB ou ao processo de licenciamento ambiental em si.

```

101 <category>
102   <pattern>MENINO OU MENINA</pattern>
103   <template>Nenhum dos dois. Sou um robo'.</template>
104 </category>
105
106 <category>
107   <pattern>QUER SER MEU AMIGO</pattern>
108   <template>Vou ser seu melhor amigo!</template>
109 </category>
110
111 <category>
112   <pattern>VOCE ESTA COM CALOR</pattern>
113   <template>Robos não sentem calor.</template>
114 </category>

```

Figura 31 - Categorias do arquivo outras.aiml, que trata perguntas variadas, não relacionadas à licenciamento ambiental, à cetesb.
Fonte: Elaborada pelo autor (2013)

A Figura 32 mostra uma categoria que fornece dados relacionados ao perfil e interesses do bot.

```

8 <category>
9   <pattern>INTERESSES</pattern>
10  <template>Gosto de varias coisas, mas meu interesse principal é meio ambiente.</template>
11 </category>

```

Figura 32 - Categoria que responde a uma pergunta feita pelo usuário sobre os interesses do bot (localizada no arquivo perfil.aiml).
Fonte: Elaborada pelo autor (2013)

Na Figura 33 é apresentado um exemplo de questionamentos preliminares ao processo de licenciamento ambiental.

```
16 <category>
17   <pattern>POR QUE O LICENCIAMENTO AMBIENTAL</pattern>
18   <template>
19     Não basta que o empreendimento esteja regularizado perante a Prefeitura.
20     Dependendo do tipo de atividade a ser desenvolvida, o empreendimento deverá
21     também estar regularmente licenciado perante o orgao ambiental.
22   </template>
23 </category>
```

Figura 33 - Categoria que responde a pergunta “Por que o licenciamento ambiental?” (localizada na categoria preliminares.aiml).

Fonte: Elaborada pelo autor (2013)

Outros sete arquivos AIML possuem categorias recursivas, responsáveis por tratar diferentes entradas que fornecem uma mesma informação como saída, informação esta que é fornecida pelas categorias dos arquivos citados anteriormente. Os arquivos AIML com as categorias recursivas são: cetesb_var.aiml, legislacaoambsp_var.aiml, perfil_var.aiml, apresentacoes_var.aiml, outras_var.aiml, preliminareslic_var, aspectoslic_var. Essas categorias possuem no seu *pattern* recursões para as categorias mais simples citadas anteriormente. Dessa forma, esses arquivos AIML tratam as variações para as entradas que são tratadas nos arquivos AIML contendo as categorias mais simples. As Figuras 34, 35, 36, 37, 38, 39 e 40 mostram respectivamente exemplos de trechos dos arquivos contendo as categorias citadas.

A Figura 34 mostra tratamento de variações, ou seja, sinônimos que podem ser utilizados para realizar perguntas relacionadas à CETESB.

```

36 <category>
37   <pattern>* INAUGURACAO</pattern>
38   <template><srai>QUANDO FOI CRIADA</srai></template>
39 </category>
40 <category>
41   <pattern>INAUGURACAO *</pattern>
42   <template><srai>QUANDO FOI CRIADA</srai></template>
43 </category>
44 <category>
45   <pattern>* INAUGURACAO *</pattern>
46   <template><srai>QUANDO FOI CRIADA</srai></template>
47 </category>
48 <category>
49   <pattern>* SURGIMENTO</pattern>
50   <template><srai>QUANDO FOI CRIADA</srai></template>
51 </category>
52 <category>
53   <pattern>* SURGIMENTO *</pattern>
54   <template><srai>QUANDO FOI CRIADA</srai></template>
55 </category>
56 <!-- FIM entrada QUANDO FOI CRIADA -->
57
58 <!-- INICIO entrada QUAL A ESTRUTURA DA CETESB -->
59 <category>
60   <pattern>* ESTRUTURA</pattern>
61   <template><srai>QUAL A ESTRUTURA DA CETESB</srai></template>
62 </category>
63 <category>
64   <pattern>ESTRUTURA *</pattern>
65   <template><srai>QUAL A ESTRUTURA DA CETESB</srai></template>
66 </category>
67 <category>
68   <pattern>* ESTRUTURA *</pattern>
69   <template><srai>QUAL A ESTRUTURA DA CETESB</srai></template>
70 </category>
71 <category>
72   <pattern>* COMPOSICAO</pattern>
73   <template><srai>QUAL A ESTRUTURA DA CETESB</srai></template>
74 </category>
75 <category>
76   <pattern>* COMPOSICAO *</pattern>
77   <template><srai>QUAL A ESTRUTURA DA CETESB</srai></template>
78 </category>

```

Figura 34 - Categorias que tratam as diferentes formas de entrada sobre a CETESB que retornam mesma resposta (localizadas no arquivo cetesb_var.aiml).

Fonte: Elaborada pelo autor (2013)

Na Figura 35 apresenta-se um exemplo de categorias que tratam diferentes formas maneiras de se perguntar sobre a questão das leis estadual e federal que envolve o processo de licenciamento ambiental

```

36 <category>
37   <pattern>* LEI ESTADUAL</pattern>
38   <template><srai>LEGISLAÇÃO ESTADUAL</srai></template>
39 </category>
40 <category>
41   <pattern>LEI ESTADUAL *</pattern>
42   <template><srai>LEGISLAÇÃO ESTADUAL</srai></template>
43 </category>
44 <category>
45   <pattern>* LEI ESTADUAL *</pattern>
46   <template><srai>LEGISLAÇÃO ESTADUAL</srai></template>
47 </category>
48 <category>
49   <pattern>* ESTADUAL</pattern>
50   <template><srai>LEGISLAÇÃO ESTADUAL</srai></template>
51 </category>
52 <category>
53   <pattern>* ESTADUAL *</pattern>
54   <template><srai>LEGISLAÇÃO ESTADUAL</srai></template>
55 </category>

```

Figura 35 - Categorias (localizadas no arquivo legislacaoambsp_var.aiml) que tratam diferentes formas de entrada para a pergunta "LEI ESTADUAL Nº.997/76 ...".
Fonte: Elaborada pelo autor (2013)

A Figura 36 mostra um exemplo de categoria que trata maneiras distintas de escrita para se perguntar sobre dados relacionados ao perfil

```

120 <category>
121   <pattern>QUEM E VOCÊ</pattern>
122   <template><srai>QUEM E VOCE</srai></template>
123 </category>
124 <category>
125   <pattern>QUEM É VOCÊ</pattern>
126   <template><srai>QUEM E VOCE</srai></template>
127 </category>
128 <category>
129   <pattern>QUEM É VOCE</pattern>
130   <template><srai>QUEM E VOCE</srai></template>
131 </category>

```

Figura 36 - Categorias (localizadas no arquivo perfil_var.aiml) que tratam diferentes formas de entrada para a pergunta "Quem é você".
Fonte: Elaborada pelo autor (2013)

Na Figura 37 é apresentado um exemplo de categoria relacionada a diálogo de apresentação que trata sinônimos para realizar o mesmo tipo de diálogo

```

9 <category>
10   <pattern>OI *</pattern>
11   <template><srai>OI</srai><sr/></template>
12 </category>
13 <category>
14   <pattern>OLÁ *</pattern>
15   <template><srai>OI</srai><sr/></template>
16 </category>
17 <category>
18   <pattern>OLÁ </pattern>
19   <template><srai>OI</srai></template>
20 </category>
21 <category>
22   <pattern>OLA *</pattern>
23   <template><srai>OI</srai><sr/></template>
24 </category>

```

Figura 37 - Exemplo das categorias (localizadas no arquivo apresentacoes_var.aiml) que tratam diferentes entradas que geram a mesma resposta da entrada "OI".
Fonte: Elaborada pelo autor (2013)

Na Figura 38 é apresentado um exemplo de variações possíveis para perguntas que não tem relação com a CETESB ou ao processo de licenciamento ambiental em si.

```

144 <category>
145   <pattern>VOCÊ * COM CALOR</pattern>
146   <template><srai>VOCE ESTA COM CALOR</srai></template>
147 </category>
148 <category>
149   <pattern>VOCE * COM CALOR</pattern>
150   <template><srai>VOCE ESTA COM CALOR</srai></template>
151 </category>

```

Figura 38 - Categorias (localizadas no arquivo outras_var.aiml) que tratam diferentes entradas para a mesma resposta fornecida pela categoria "VOCE ESTA COM CALOR".
Fonte: Elaborada pelo autor (2013)

A Figura 39 mostra um exemplo de possíveis variações para perguntas relacionadas às preliminares do processo de licenciamento ambiental

```

9 <category>
10   <pattern>* SIGNIFICADO DE LICENCIAMENTO AMBIENTAL</pattern>
11   <template><srai>O QUE É LICENCIAMENTO AMBIENTAL</srai></template>
12 </category>
13 <category>
14   <pattern>SIGNIFICADO DE LICENCIAMENTO AMBIENTAL *</pattern>
15   <template><srai>O QUE É LICENCIAMENTO AMBIENTAL</srai></template>
16 </category>
17 <category>
18   <pattern>* SIGNIFICADO DE LICENCIAMENTO AMBIENTAL *</pattern>
19   <template><srai>O QUE É LICENCIAMENTO AMBIENTAL</srai></template>
20 </category>
21 <category>
22   <pattern>* DEFINIÃ+Ã,O LICENCIAMENTO</pattern>
23   <template><srai>O QUE É LICENCIAMENTO AMBIENTAL</srai></template>
24 </category>
25 <category>
26   <pattern>* DEFINIÃ+Ã,O LICENCIAMENTO *</pattern>
27   <template><srai>O QUE É LICENCIAMENTO AMBIENTAL</srai></template>
28 </category>

```

Figura 39 - Categorias (localizadas no arquivo preliminareslic_var.aiml) que tratam diferentes formas de entrada para a pergunta "O QUE É LICENCIAMENTO AMBIENTAL?".

Fonte: Elaborada pelo autor (2013)

Na Figura 40 é mostrado um exemplo de categorias com variações de perguntas relacionadas aos aspectos de licenciamento ambiental

```

55 <category>
56   <pattern>* APP</pattern>
57   <template><srai>AREA DE PRESERVAÇÃO PERMANENTE APP - INTERVENÇÃO </srai></template>
58 </category>
59 <category>
60   <pattern>APP *</pattern>
61   <template><srai>AREA DE PRESERVAÇÃO PERMANENTE APP - INTERVENÇÃO </srai></template>
62 </category>
63 <category>
64   <pattern>* APP *</pattern>
65   <template><srai>AREA DE PRESERVAÇÃO PERMANENTE APP - INTERVENÇÃO </srai></template>
66 </category>
67 <category>
68   <pattern>*INTERVENÇÃO </pattern>
69   <template><srai>AREA DE PRESERVAÇÃO PERMANENTE APP - INTERVENÇÃO </srai></template>
70 </category>
71 <category>
72   <pattern>* INTERVENÇÃO *</pattern>
73   <template><srai>AREA DE PRESERVAÇÃO PERMANENTE APP - INTERVENÇÃO </srai></template>
74 </category>
75 <category>
76   <pattern>PRESERVAÇÃO *</pattern>
77   <template><srai>AREA DE PRESERVAÇÃO PERMANENTE APP - INTERVENÇÃO </srai></template>
78 </category>
79 <category>
80   <pattern>* PRESERVAÇÃO *</pattern>
81   <template><srai>AREA DE PRESERVAÇÃO PERMANENTE APP - INTERVENÇÃO </srai></template>
82 </category>

```

Figura 40 - Categorias (localizadas no arquivo aspectoslic_var.aiml) que tratam diferentes formas de entrada para a pergunta "AREA DE PRESERVAÇÃO PERMANENTE APP – INTERVENÇÃO"

Fonte: Elaborada pelo autor (2013)

Um último arquivo AIML, o arquivo *padrao.aiml*, possui apenas uma categoria que trata qualquer outra entrada que não for reconhecida por nenhuma das categorias especificadas nos outros arquivos AIML. Essa categoria tem a função de fornecer uma saída padrão para o usuário, caso nenhuma das outras categorias da base de conhecimentos possua um modelo para a entrada do usuário.

A categoria padrão então fornece a saída “Não entendi sua frase, pode reformular para mim?”, conforme pode ser vista na figura 41. Essa saída pede que o usuário escreva sua pergunta de uma forma diferente, que então tem a chance de ser reconhecida por uma categoria AIML, fornecendo a resposta corretamente.

```
9 <category>
10   <pattern>*</pattern>
11   <template>Não entendi sua pergunta, pode reformular a frase para mim?</template>
12 </category>
```

Figura 41 - Categoria que fornece uma saída padrão caso a resposta do usuário não seja tratada por nenhuma outra categoria.

Fonte: Elaborada pelo autor (2013)

O capítulo 10 traz uma discussão referente aos testes realizados com usuários e sua relação com o processo de refinamento da base de conhecimento.

10 TESTES COM OS USUÁRIOS

A base de conhecimento foi disponibilizada para teste entre usuários da CETESB e profissionais do ramo ambiental da Usina Açucareira São Manoel. O público-alvo englobou, dessa forma, o pessoal do setor sucroalcooleiro e agrícola, sendo analisada por um grupo em torno de dez pessoas. Esses usuários da CETESB possuem atividades e empreendimentos sujeitos ao licenciamento ambiental, que no caso são usinas de açúcar e álcool, usinas de reciclagem, dentre outras que no geral envolvem a modificação da água, ar, solo, resíduo sólido e que podem gerar riscos ambientais e que não visam a sustentabilidade.

Essa disponibilização diz respeito apenas à visualização das bases através de *screenshots* das linhas de programação e de *screenshots* dessas mesmas linhas de código em execução, explicando o processo de funcionamento das mesmas visto que a ferramenta é local e necessita do interpretador instalado na máquina. Devido a essa dependência, existe dificuldade de acesso e teste diretamente do bot em qualquer máquina de maneira independente, sendo possível apenas em casos específicos onde se teve acesso para instalação do program-o. Uma vez que um usuário realize uma conversa com o *chatbot*, a plataforma mantém o registro desse diálogo (através de uma ferramenta de registro de conversas mais conhecida por *log*), o que permite que a base de conhecimentos seja examinada em termos de fragilidades. Apenas para teste de registro e funcionamento do *log*, foram analisados um *log* de cada base existente e sua respectiva base de variação.

Os testes com os usuários se deram de forma incremental, ou seja, à medida que os usuários foram interagindo ou tendo contato com *chatbot*, seus registros ou observações foram sendo acompanhados e anotados, para posterior refinamento da base de conhecimento. Testes envolvendo a percepção do usuário também foram importantes, para isso foi coletado sugestões do pessoal que trabalha no setor sucroalcooleiro e de pessoas envolvidas com o tema sustentabilidade em geral e que são os principais utilizadores potenciais do *chatbot* desenvolvido. A ideia desse teste foi de coletar, ainda que de forma simples, a percepção dos usuários sobre a utilidade de uma ferramenta como a desenvolvida nesta investigação. Assim, foi possível coletar as sensações dos usuários com relação à ferramenta e seu desempenho, além de dicas para melhoramentos. Quando se tinha a opção e permissão para demonstrar de maneira prática o funcionamento, o teste foi realizado

colocando a ferramenta em execução e analisando quais os termos-chave que usuários utilizavam para obterem as respostas desejadas. A sensação dos usuários foi de satisfação a partir do momento que o *chatterbot* respondeu de maneira bem embasada e concisa e de modo instantâneo o que era solicitado. Além disso, foi destacado também o fato de a base de conhecimento estar separada por categorias de tópicos, já que isso facilitava a compreensão e o relacionamento dos dados, principalmente se comparado ao uso do site para obtenção de informações. Quando essa disponibilidade de execução da ferramenta não era possível foi fornecido apenas um questionário de avaliação (mostrado na Figura 42) que continha *screenshots* da programação das bases e *screenshots* da mesma em modo de execução para anotarem sugestões, dicas e escreverem pontos positivos e negativos do *bot* com base na explicação dada de seu funcionamento.

Com relação a pergunta número um que diz respeito se o *chatterbot* respondeu a todas as perguntas realizadas, a resposta dos questionários mostrou que na maioria dos casos o *chatterbot* respondeu tudo que foi requisitado, o que mostra que a base está bem enriquecida e que apenas com alguns refinamentos os resultados podem ser ainda mais eficazes. Já a pergunta número dois que diz respeito ao cumprimento de seu papel e se foi eficiente, a resposta foi unanime que sim e que realmente cumpre com o papel, respondendo de maneira rápida e de uma forma bem completa os dados ou questionamentos solicitados. A resposta à pergunta três mostrou que os usuários consideraram que a *interface* é bem limpa e não cansativa durante sua utilização, porém alguns questionaram a possibilidade de uma interface mais customizável de acordo com suas preferências o que pode ser algo a ser pensado como trabalho futuro. Para pergunta quatro referente às dúvidas sanadas pelo *chatterbot*, a resposta foi unanime que havia esclarecimento das dúvidas, também houve destaque o fato de que o *chatterbot* além de orientar, muitas vezes forneceu links que levam diretamente a página que desejavam, o que evitou desperdício de tempo para obter informação relevante. A pergunta cinco sobre a avaliação geral do *chatterbot* indica que os usuários consideraram a ferramenta como um recurso a mais de suporte à busca de informações, caracterizado por ser mais informal e rápido. A pergunta seis aberta para críticas, sugestões ou elogios contou com elogios em relação a grande quantidade e organização de informações, sugestões de design da interface e comentários apenas sobre a acentuação (que

ainda é um recurso que precisa ser melhorado no program-o para a língua portuguesa).

Considerando o ainda *feedback* dado durante os testes, destaca-se o fato de os os usuários apontarem como interessante a ideia de um *chatterbot* para esse ramo de atividade , pois a maioria das coisas que já viram na *internet* relacionadas a esse tipo de ferramenta servem apenas para conversar sobre assuntos familiares, lazer , ou respondem coisas muito simples , etc. e não ligadas ao ambiente profissional ou que seja relacionada a CETESB. Muitos destacaram que foi, portanto, a primeira vez que viram algo do gênero e com esse nível de detalhes de informação. A única ressalva feita foi com relação à limitação no uso de acentuação, o que pode dificultar a obtenção de uma resposta. Desse modo, é preciso sempre manter avisado quem for utilizar a ferramenta de que as perguntas não necessitam do uso de caracteres de acentuação.

Questionário Chatterbot CETESB

O Chatterbot CETESB é um trabalho de conclusão de curso de um aluno de Ciências da Computação. Responda essas 5 perguntas do Formulário de avaliação chatterbot que tem como funcionalidade responder perguntas relacionados ao processo de licenciamento ambiental da CETESB.

Codificação da base de apresentação do chatterbot

```

8 <category>
9   <pattern>OI</pattern>
10  <template>
11    <random>
12      <li>Olá! humano!</li>
13      <li>Olá! Seja bem vindo.</li>
14    </random>
15  </template>
16 </category>

```

Codificação de apresentação em modo de execução

Program O Example GUI Page - HTML

Say:

```

Seeker: Oi
CETESB BOT: Olá! Seja bem vindo.

Seeker: tudo bem
CETESB BOT: Sim,tudo conforme.

Seeker: MEU NOME E MURILO
CETESB BOT: Prazer em te conhecer.Murilo

```

O chatterbot respondeu a todas as perguntas realizadas ?

Justifique sua resposta

O chatterbot cumpriu com seu papel e foi eficiente ?

Justifique sua resposta

Como você avalia a interface do chatterbot ?

Justifique sua resposta

O chatterbot sanou suas dúvidas ?

Justifique sua resposta

De uma maneira geral , como você avalia o Chatterbot ?

Justifique sua resposta

Caso vocês desejem , façam algum comentário , crítica , sugestão ou elogio

Justifique sua resposta

Figura 42: Questionário avaliação do chatterbot
Fonte: Próprio autor

A partir dos registros de conversas (como o do exemplo mostrado pela Figura 43) gerados por estes testes foi possível identificar novas entradas de usuários, cujas saídas foram adicionadas à base de conhecimento.

A Figura 44 mostra uma pergunta que não era, a princípio, reconhecida pelo *chatbot*, especificamente a questão “O que é SILIS?”. Posteriormente, a base de conhecimento foi atualizada para incluir uma resposta para esta pergunta (bem como as diferentes formas de se perguntar por essa mesma informação), resultando em uma resposta com a informação solicitada, como é mostrado na Figura 45.

Também foi possível identificar fatores que dificultam uma resposta precisa do *chatbot*, como erros de digitação do usuário, conforme exemplifica a Figura 46. Além disso, a própria língua portuguesa é um fator de dificuldade devido à acentuação, característica não presente na língua inglesa. Devido a isso, várias categorias para variações de uma mesma entrada tiveram que ser elaboradas, pois a falta de acentuação em uma determinada palavra poderia acarretar a não interpretação da entrada do usuário pelo *chatbot*.

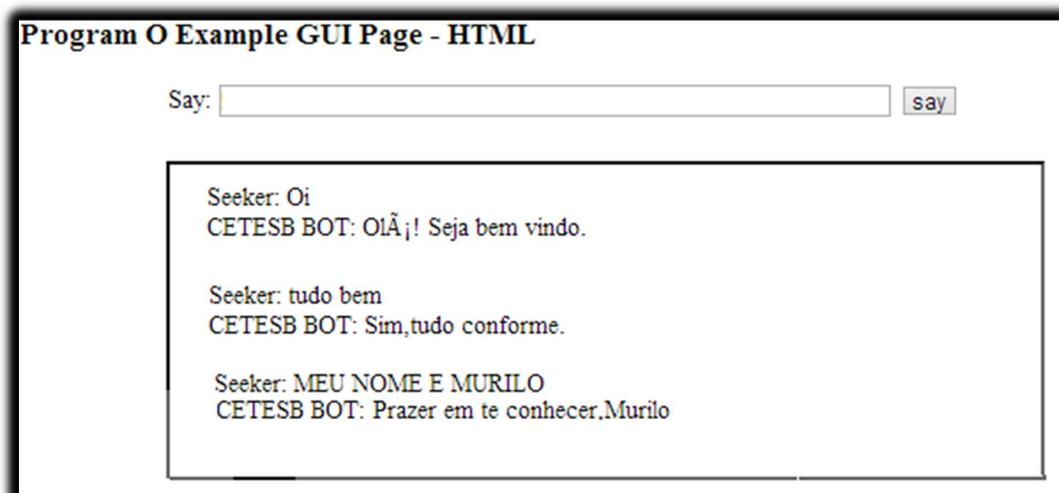


Figura 43 - Trecho de um registro de conversação entre usuário e chatbot.
Fonte: <http://blog.program-o.com/>. Capturado em 22 novembro 2013.

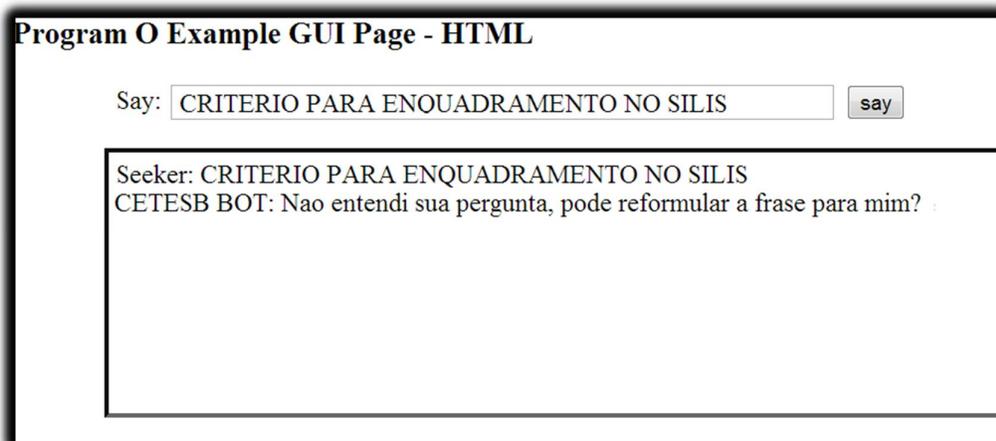


Figura 44 - Pergunta não reconhecida pelo chatterbot.
 Fonte: <http://blog.program-o.com/>. Capturado em 22 novembro 2013.

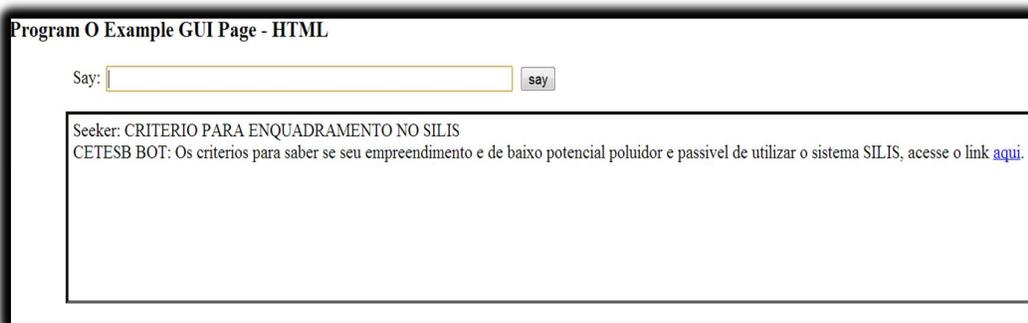


Figura 45 - Pergunta passa a ser reconhecida após atualização da BDC
 Fonte: <http://blog.program-o.com/>. Capturado em 22 novembro 2013.

Outra dificuldade encontrada foram os diferentes modos de se perguntar sobre o mesmo assunto. Uma mesma pergunta pode ser feita de várias maneiras, sendo impossível antecipar todas elas na base de conhecimento inicial. Por isso, a análise dos registros de conversação é fundamental. Nela são identificadas as diferentes maneiras que o usuário pergunta pela mesma informação, permitindo que a base de conhecimento seja atualizada.

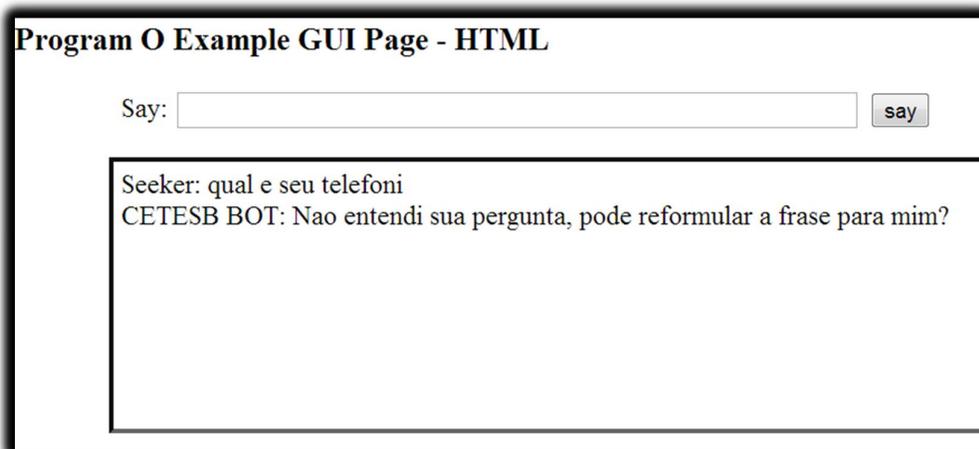


Figura 46 - Bot não reconhece pergunta devido a erro de digitação do usuário.
Fonte: <http://blog.program-o.com/>. Capturado em 22 novembro 2013.

Obviamente nem todo o conteúdo das conversas pôde ser aproveitado no processo de aprimoramento da base de conhecimento, pois algumas perguntas não faziam parte do escopo do trabalho. Como forma de lidar com essa situação referente às perguntas para qual o *bot* não tinha conhecimento associado foi criada na base de conhecimento uma categoria específica para informar o usuário a impossibilidade de responder certos assuntos.

Essa etapa de testes com os usuários foi de fundamental importância para o desenvolvimento do protótipo, pois por mais robusta que seja à base de conhecimento, sempre é possível refiná-la de modo a melhorar a qualidade das respostas fornecidas.

Outra análise interessante fazendo um comparativo entre o uso do *chatterbot* versus *site* da CETESB foi que, utilizando o *chatterbot* apenas usando termos-chave, foi possível obter as informações que são relevantes ao assunto, além da possibilidade do *chatterbot* fornecer como resposta links que levam diretamente ao conteúdo que a pessoa deseja analisar. Por outro lado, no *site* muitas vezes é necessário visualizar uma infinidade de abas ou uma série categorias dentro de categorias tornando-se algo cansativo, maçante, gastando de tempo em demasia para a com informações que não são relevantes ou que não estão ligadas ao que está sendo buscado.

Desse modo, o *chatterbot* é algo mais incisivo, direto e por ser um chat torna o processo de busca e comunicação mais amigável e natural se comparado à formalidade existente no *site*.

11 CONSIDERAÇÕES FINAIS

A dificuldade que os usuários encontram para acessar os computadores, seja ela por falta de treinamento ou habilidade, é um fator limitativo tão importante que justifica o desenvolvimento de projetos que facilitem o acesso aos computadores. Neste contexto, ferramentas como os *chatterbots* apresentam-se como alternativas ao acesso à informação.

O componente deste trabalho pode ser utilizado como uma ferramenta de disseminação de conhecimentos, uma vez que as informações disponibilizadas pelo *chatterbot* ao público podem ser facilmente acessadas, por meio de uma *interface* que faz uso de linguagem natural, utilizando técnicas de Inteligência Artificial para desenvolvimento de agentes de conversação. Tal ferramenta apresenta adicionalmente como benefício o baixo custo devido à utilização de *softwares* gratuitos. A parte mais custosa da ferramenta refere-se à construção da base de conhecimentos, que demanda bastante tempo e exige um estudo apurado de conceitos de Inteligência Artificial, sobretudo relacionados ao Processamento das Linguagens Naturais, de forma a se criar um agente que simule o comportamento humano em um diálogo.

Embora a ferramenta construída tenha alcançado os objetivos propostos neste trabalho, por se tratar de um protótipo elaborado com a finalidade de demonstrar o potencial da tecnologia aplicada, a ferramenta não supre, obviamente, todas as questões relacionadas ao tema proposto e, por isso, não responde plenamente a todos os questionamentos possíveis de serem encontradas em situações reais. Outro fator limitativo é que a língua portuguesa ainda está passando por melhorias, e correções que estão sendo feitas nas versões utilizadas pelo program-o e no momento não é possível o tratamento de acentuação como o uso de acento agudo, circunflexo, cedilha. Para se atingir um maior grau de amadurecimento e precisão, a base de conhecimentos deve ser submetida a um processo contínuo de aprimoramento e refinamento, o que exige tempo e estudo.

Outra limitação associada ao *chatterbot* refere-se à coerência das respostas. A coerência do diálogo envolvendo uma máquina é, seguramente, um dos maiores desafios do desenvolvimento de um agente de conversação, já que em situações reais uma conversa tende a ser bastante ampla, envolvendo vários temas.

Quanto a aplicação do questionário avaliativo analisando-se as respostas dos participantes da pesquisa tivemos alguns resultados importantes analisando qualitativamente o *bot* e mostrando o seu potencial para a tarefa proposta nesta investigação. Como trabalhos futuros podem citar a ampliação e o tratamento de acentuação das palavras da base de conhecimento, incrementando as categorias manipuladas pelo protótipo. Além disso, a divulgação e a disponibilização do agente em larga escala, permitirá a coleta de um número expressivo de arquivos de registro de conversa, refletindo diretamente na qualidade do processo de aprimoramento da base de conhecimentos.

REFERÊNCIAS

ANNES, R. Inteligência Artificial: Aula 2. **Inteligência Artificial**, 2011. Disponível em: <<http://www.profs.iffca.edu.br/~ricardo/ia2.html>>. Acesso em 25 abr. 2013.

BITTENCOURT, G. **Breve História da Inteligência Artificial**. Florianópolis, 1999. Disponível em: <<http://www2.dem.inpe.br/ijar/AIBreveHist.doc>>. Acesso em 23 abr. 2013.

BUSH, N. Artificial Intelligence Markup Language (AIML) Version 1.0.1. **Alicebot.org**. 2001. Disponível em: <<http://www.alicebot.org/TR/2001/WD-aiml/>>. Acesso em: 12 mar. 2013.

CARVALHO, J. O. F. O papel da interação humano-computador na inclusão digital. In: **Revista Transinformação**, V. 15, n. 3, edição especial, p. 75-89, Campinas, 2003.

CARVALHO, L. R. De. **F.L.A.VIA SES**: ferramenta linguística de assistência virtual e automática em seguridade social. 2010. 73 f. Trabalho de Conclusão de Curso (Ciência da Computação) – Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, 2010.

CETESB: Companhia de Tecnologia de Saneamento Ambiental, 2013. O licenciamento ambiental e suas cartilhas. Disponível em <<http://www.cetesb.sp.gov.br/licenciamento/licenciamento-ambiental/1-pagina-inicial>>. Acesso em 27 abril. 2013.

COELHO, Helder. **Inteligência Artificial em 25 Lições**. Lisboa: Fundação Calouste Gulbenkian, 1994. 532 p. p. 341

FERNANDES, A. M. R. **Inteligência Artificial**: Noções gerais. Florianópolis: Visual Books, 2005.

FÉ, I. M. M. **AIML aplicada à criação de personagens interativos**, 2005. 67f. Trabalho de Conclusão de Curso (Graduação) – Instituto Tecnológico de Aeronáutica, São José dos Campos. 57

GAITOBOT. **AIML Editor**, 2013. Disponível em < <http://www.gaitobot.de/gaitobot/>>. Acesso em: 22 novembro 2013.

GIL, A. C. **Como Elaborar Projetos de Pesquisa**. 5ª ed. São Paulo: Atlas, 2010.

GONZALEZ, M.; LIMA, V. L. S. **Recuperação de Informação e Processamento de Linguagem Natural**, 2003. Disponível em: <<http://www.inf.pucrs.br/~gonzalez/docs/minicurso-jaia2003.pdf>>. Acesso em: 24 abr. 2013.

INSITE. **CASE ROBÔ ED – CONPET – PETROBRÁS**. 2011. 12 slides. Disponível em <<http://www.slideshare.net/camilacanonici/case-rob-ed>>. Acesso em: 14 abril 2013.

LAUREANO, E.A. G. C. **ConsulBot**– Um ChatterbotConsultor para Ambientes Virtuais de Estudo na Internet. 1999. Disponível em: <<http://www.di.ufpe.br/~tg/1999-1/eagcl.doc>>. Acesso em: 27 abril 2013.

LAVEN, S. The Simon Laven Page, c1996. Disponível em: <<http://www.simonlaven.com>>. Acesso em: 01 maio 2013.

LEONHARDT, M. D.; CASTRO, D. D.; DUTRA, R. L. S.; TAROUCO, L. M. R., 2003

ELEKTRA: Um Chatterbot para Uso em Ambiente Educacional. In: Renote. Porto Alegre. Disponível em: <<http://penta3.ufrgs.br/~elektra/info/artigos/chatterbot-Elektra%5B1%5D.PDF>>. Acesso em: 05 mar. 2013.

LUGER, G. F **Estrutura e Estratégias para a solução de problemas complexos**. 4ª ed. Bookman, 2002
Disponível em: <<http://books.google.com.br/books?id=h2GJLP6jQ4kC&printsec=frontcover#v=onepage&q&f=false>>

MENDES, R. D. Inteligência artificial: sistemas especialistas no gerenciamento da informação. **Ci. Inf.**, Brasília, v. 26, n. 1, Jan. 1997. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651997000100006&lng=en&nrm=iso>. Acesso em: 20 abril. 2013.

MMA: Ministério Meio Ambiente, 2013. **O que é licenciamento ambiental e sua importância**. Disponível em <<http://www.mma.gov.br/governanca-ambiental/portal-nacional-de-licenciamento-ambiental/licenciamento-ambiental>>. Acesso em 22 maio. 2013.

MÜLLER, D. N.; Processamento de linguagem natural. Porto Alegre: UFRGS, 2003. 11 p.

OLIVEIRA, F. A. D. De. **Processamento de linguagem natural**: princípios básicos e a implementação de um analisador sintático de sentenças da língua portuguesa, 2002. Disponível em: <<http://www.inf.ufrgs.br/gppd/disc/cmp135/trabs/992/Parser/parser.html>>. Acesso em: 22 abr. 2013.

OTHERO, G. A. Linguística Computacional: uma breve introdução. **Letras de Hoje**. Porto Alegre, jun. 2006. Disponível em: <<http://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/viewFile/605/436>>. Acesso em: 10 abr. 2013.

PRIMO, A. F. T. **O uso de chatterbots na educação à distância**, 2000. Disponível em: <http://www.nied.unicamp.br/oea/mat/chatterbots_lec.pdf>. Acesso em: 26 abril. 2013.

PROGRAM-O – **um interpretador de linguagem AIML para PHP e MySQL** – Disponível em <<http://www.program-o.com/>>;

RABELLO, R. S.; MARCHI, A. C B. De; FOSSATTI, M. **AGEBOT**: Um chatterbot em AIML voltado a responder questões sobre epilepsia. In: XI Workshop de informática médica, 2011, Natal. Anais do WIM, 2011. Disponível em:

<http://www.dimap.ufrn.br/csbc2011/anais/eventos/contents/WIM/WIM_Sessao_4_Artigo_2_Fossatti.pdf>. Acesso em: 23 abril 2013.

RICH, E. **Inteligência artificial**. São Paulo: McGraw-Hill, 1988.

RINGATE, T. AIML Reference Manual. **Alicebot.org**, 2001. Disponível em <<http://www.alicebot.org/documentation/aiml-reference.html>>. Acesso em: 14 abr. 2013.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. 2 ed. Rio de Janeiro: Elsevier, 2004. 59

SANTOS, D. **Introdução ao processamento de linguagem natural através das aplicações**, 2001. Disponível em: <<http://193.136.2.105/Diana/download/Santos2001Aplicacoes.pdf>>. Acesso em: 29 abr. 2013.

SATO, S. K. A Estética Publicitária da Inovação: smartphones e tablets. **Pensamento e Realidade**. [S. l.], n 3, 2011. Disponível em: <[http://revistas.pucsp.br/index.php/pensamento e realidade/article/view/8078](http://revistas.pucsp.br/index.php/pensamento_e_realidade/article/view/8078)>. Acesso em 10 abril. 2013.

SMA: Secretaria Meio Ambiente, 2013. **Quem somos e o que fazemos**. Disponível em <<http://www.ambiente.sp.gov.br/a-secretaria/quem-somos/>>. Acesso em 22. maio 2013.

SGANDERLA, R. B.; FERRARI, D. N.; GEYER, C. F. R. **BonoBOT**: Um chatterbot para interação com usuários em um sistema tutor inteligente. In:

SILVA, Jeferson Luís da - **Chatterbots – Simuladores de Diálogo** – 2ª revisão – Bookess Editora – Impresso e em PDF-;

SIMON, H.A. **Why should machines learn?** In Michalski et al.,1983.

SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 14, 2003, Rio de Janeiro. Disponível em: <<http://www.nce.ufrj.br/sbie2003/publicacoes/paper46.pdf>>. Acesso em: 26 abril. 2013.

TALKING about aiml predicates. **Alicebot**, c2012. Disponível em: <http://www.alicebot.org/video_transcripts/predicates.txt>. Acesso em: 4 abril. 2013.

THE LOEBNER PRIZE. **The Loebner Prize - The first Turing test**. s/d. Acesso em 13/05/2013, disponível em THE LOEBNER PRIZE - The first Turing test: <http://www.loebner.net/Prizef/loebner-prize.html>;

TURING, A. **Computing machinery and intelligence**, 1950. Disponível em: <<http://www.loebner.net/Prizef/TuringArticle.html>>. Acesso em: 14 abril. 2013.

VIEIRA, R.; LIMA, V. L. S. **Lingüística computacional: princípios e aplicações**. In: IX Escola de Informática da SBC-Sul. Luciana Nedel (Ed.) Passo Fundo, Maringá, São José. SBC-Sul, 2001. Disponível em: <<http://www.inf.unioeste.br/~jorge/MESTRADOS/LETRAS%20-%20MECANISMOS%20DO%20FUNCIONAMENTO%20DA%20LINGUAGEM%20-%20PROCESSAMENTO%20DA%20LINGUAGEM%20NATURAL/ARTIGOS%20INTERESSANTES/lingu%EDstica%20computacional.pdf>>. Acesso em: 10 abr. 2013.

WALLACE, R. **Be Your Own Botmaster**. 2005. 2ª ed. ALICE A.I. Foundation, Inc.;

WALLACE, R. S. The Anatomy of A.L.I.C.E. **Alicebot.org**, 2009. Disponível em: <<http://www.alicebot.org/anatomy.html>>. Acesso em: 18 abr. 2013.

Implementação de um Atendente Virtual Especializado em Licenciamento Ambiental utilizando AIML

Murilo L.G de Aguiar, Patrick P. Silva, Elvio G da Silva, Henrique P. Martins.

Centro de Ciências Exatas e Sociais Aplicadas – Universidade Sagrado Coração (USC)
Caixa Postal 15.064 – 91.501-970 – Bauru – SP – Brasil

murilo_lilooo@hotmail.com, {patrick.silva, egsilva, henrique.martins }@usc.br

Abstract. *The Environmental Company of São Paulo State is the government agency responsible for the control, supervision, monitoring and licensing of pollution-generating activities. It's accessible to the population and is used by professionals in the environmental sector. The purpose of this work is an investigation that culminates in the proposal and implementation of a chatterbot responsible for the interaction between users and information provided by CETESB in natural language. This system provides clarifying questions related to CETESB, available data and informing about where to find the data, reducing the search time and providing access to results that meet user needs.*

Resumo. *A Companhia Ambiental do Estado de São Paulo é a agência do governo, responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição. É acessível a população e é utilizada por profissionais do setor ambiental. A proposta deste trabalho consiste em uma investigação que culmine na proposta e implementação de um chatterbot responsável pela interação entre usuários e informações disponibilizadas pela CETESB em linguagem natural. Esse sistema permite esclarecer questões relacionadas à CETESB e aos dados disponibilizados e informando sobre onde encontrar esses dados, diminuindo o tempo de busca e possibilitando acesso a resultados que atendam as necessidades do usuário.*

1. Introdução

A dificuldade que os usuários encontram para acessar os computadores, seja ela por falta de treinamento ou habilidade, é um fator restritivo tão importante que faz com que cientistas e técnicos se dediquem ao desenvolvimento de projetos que facilitem o acesso aos computadores por pessoas com pré-requisitos cada vez menores. (CARVALHO, 2003).

A possibilidade de dar a uma máquina a habilidade de interagir com o ser humano através da compreensão e simulação do seu comportamento tem sido alvo de pesquisas na área de Inteligência Artificial (mais especificamente com o estudo do Processamento de Linguagem Natural). A Inteligência Artificial desenvolveu diversas maneiras de simular a comunicação humana, acarretando, entre outras tecnologias, o surgimento dos *chatterbots*, programas de computador que simulam uma conversa com uma pessoa, utilizando, para isso, a linguagem natural. (LEONHARDT, 2003).

Segundo Primo (2000, p. 5) “robôs de conversação oferecem um forte apelo motivacional”. Os internautas demonstram grande interesse em interagir com esses robôs. Além disso, o fator “novidade” também chama a atenção do público, atraindo-o a

utilizar o “sistema”..

A Companhia Ambiental do Estado de São Paulo (CETESB) é a agência do governo do estado de São Paulo, responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição, com a preocupação fundamental de preservar e recuperar a qualidade das águas, do ar e do solo (CETESB, 2013).

A CETESB possui um *Website*, onde o usuário pode se cadastrar para receber os boletins diários por *e-mail*, consultar boletins anteriores, obter informações a respeito da instituição e sobre o que acontece dia a dia na CETESB (CETESB, 2013).

De acordo com levantamento feito, mesmo com os dados disponíveis e centralizados no website da CETESB a qual faz parte da SMA (Secretaria Meio Ambiente), as pessoas têm dificuldades de encontrar as informações sendo, muito comum surgirem dúvidas quando se deseja implantar uma atividade (ou ampliar uma atividade existente) potencialmente poluidora. No Website existe uma série de informações/pré-requisitos e possíveis caminhos que o usuário deveria seguir, mas que, na prática, confundem quase todos os usuários que nunca tiveram contato com as ferramentas e situações do gênero. Esse fato leva o usuário comum a ligar no atendimento telefônico ou tentar sanar sua dúvida através da regional mais próxima. Além disso, por mais simples que seja, há regras para o Licenciamento Ambiental que devem ser seguidas e tal informação deve ficar clara para o usuário.

O *chatbot* proposto nesta investigação foi implementado como um atendente virtual que fornece uma *interface* mais amigável para o usuário que busca informações ou dados. Ele esclarece as dúvidas relacionadas ao Licenciamento Ambiental e informa como proceder dentro da lei, assim como, presta informações sobre como realizar o licenciamento de acordo com a atividade a ser realizada. O objetivo foi, portanto, contribuir para diminuir o tempo de busca, possibilitando acesso a resultados que atendam as necessidades dos usuários, além de esclarecer dúvidas básicas sobre licenciamento.

2. Referencial Teórico

Segundo Winston (1992 apud RUSSEL; NORVIG, 2004, p. 5), a Inteligência Artificial é “o estudo das computações que tornam possível perceber, raciocinar e agir.” Para Russel e Norvig (2004, p. 3), o campo da Inteligência Artificial tenta “[...] não apenas compreender, mas também construir entidades inteligentes.”.

Com a crescente evolução da IA, as tarefas de percepção começam a ganhar mais destaque. Tais tarefas podem ser representadas pela técnica da Inteligência Artificial conhecida como Processamento de Linguagem Natural (RABELLO, 2011 apud SBC, 2009).

Segundo Silva (2012), o significado de linguagem natural já está implícito em seu próprio nome. O processamento de linguagem natural (NLP, ou PLN, em português) é uma técnica que permite que o homem se comunique com uma máquina utilizando-se de sua linguagem de conversação, estabelecendo assim uma comunicação da maneira mais “natural” possível. No PLN é feita a análise sintática (escrita), semântica (significado), morfológica (compreensão frase) e pragmática (coerência expressão).

O Processamento de Linguagem Natural é uma técnica de interpretação e processamento das entradas do usuário fundamental para o entendimento da linguagem

realizada pelos *chatterbots* (FÉ, 2005).

Segundo Laven (1996), um *chatterbot* é um programa de computador que tem o objetivo de simular uma conversa fazendo com que o usuário pense, ao menos momentaneamente, que está realmente falando com outra pessoa.

Estes programas de Inteligência Artificial possuem um grande poder de interação com o usuário, motivando-o a continuar na conversa. É isso que o torna uma ferramenta interessante para a educação já que desperta o interesse dos alunos. (FOSSATTI; RABELLO; MARCHI, 2011).

Uma das principais linguagens utilizadas para o desenvolvimento das bases de conhecimento do *chatterbot* é a AIML (*Artificial Intelligence Markup Language*), que é uma especificação da linguagem XML para a construção de perfis de *chatterbots*. (FÉ, 2005).

A AIML foi desenvolvido pela comunidade Alicebot durante o período de 1995-2000 com uma gramática proprietária, também chamada AIML, que formava a base para o primeiro Alicebot. Após diversas atualizações visando à padronização da gramática foi adotado o XML (*eXtensible Markup Language*). (LEONHARDT, 2003).

A unidade básica de conhecimento na linguagem AIML é chamada de categoria, definida pelas etiquetas `<category>` e `</category>`. Cada categoria contém uma pergunta de entrada (estímulo), uma resposta de saída, e um contexto opcional. A entrada é chamada de padrão (*pattern*) e é definida entre as etiquetas `<pattern>` e `</pattern>`, e a resposta é conhecida como modelo (*template*), definido entre as etiquetas `<template>` e `</template>` (WALLACE, 2009).

O modelo ainda pode conter as etiquetas `<srail>` e `<sr/>` que são utilizadas na redução simbólica de entradas, e as etiquetas `<random></random>` que permitem que o *bot* escolha uma resposta aleatoriamente para retornar ao usuário (RINGATE, 2001).

O contexto opcional incluso numa categoria pode ser de dois tipos, que são chamados de “*that*”, que utiliza as etiquetas `<that>` e `</that>` para se referir a qualquer coisa que o *bot* tenha dito antes da última entrada do usuário (RINGATE, 2001) e “*topic*” (etiquetas `<topic>` e `</topic>`), que permite ao *bot* selecionar, preferencialmente, respostas que estejam relacionadas ao tópico que está sendo discutido no momento (FÉ, 2005).

O AIML ainda possui predicados que podem ser declarados, recuperados e manipulados a qualquer momento da execução do *chatterbot*. O elemento “*set*” (etiquetas `<set name=”xxxx”>` e `</set>`) é usado para atribuir um valor a um predicado, definido por um nome (RINGATE, 2001). Caso o *chatterbot* necessite retornar o valor armazenado em um predicado para o usuário, o elemento “*get*” (etiqueta `<get name=”xxxx”/>`) será utilizado (TALKING..., c2012).

O licenciamento ambiental é o procedimento administrativo pelo qual o órgão ambiental competente analisa a localização, instalação, ampliação e operação de empreendimentos e atividades utilizadoras de recursos ambientais consideradas efetiva ou potencialmente poluidoras, ou daquelas que, sob qualquer forma, possam causar degradação ambiental, considerando as disposições legais e regulamentares e as normas técnicas aplicáveis ao caso (Secretaria Meio Ambiente, 2013).

3. Metodologia

Este trabalho foi desenvolvido em duas fases distintas: uma fase de investigação dos aspectos teóricos e uma etapa prática de implementação do sistema. Na primeira fase foi feito um estudo teórico dos diversos assuntos relacionados com o escopo do trabalho, envolvendo áreas como envolvendo áreas como Inteligência Artificial, Processamento de Linguagem Natural, AIML, *chatterbot*, Licenciamento Ambiental. Como produto desta fase foi proposto um modelo para o protótipo do *chatterbot*. Este modelo consistiu em definir quais as etiquetas AIML deveriam ser utilizadas para a implementação do protótipo de forma a manter a consistência da base de conhecimentos, bem como a coerência das respostas dadas pelo *chatterbot*. Na segunda fase foi realizada a implementação do protótipo proposto (escrita da base de conhecimento em AIML) e os testes necessários. As bases foram escritas e editadas utilizando os softwares Gaitobot e o Notepad++.

O protótipo foi desenvolvido utilizando o motor de busca *program-o* e as bases de conhecimento foram escritas na linguagem AIML que é baseada na linguagem XML. A plataforma do *program-o* foi escolhida, pois a construção de uma máquina de inferência não traria nenhum benefício adicional ao projeto, uma vez que a plataforma fornece todas as funcionalidades necessárias para construção do agente com rapidez e controle, deixando o construtor livre para enfatizar o trabalho no que de fato é importante em um agente de diálogo, ou seja, sua base de conhecimento.

Cada tópico sobre o qual o *chatterbot* pode conversar foi definido como uma categoria na linguagem AIML, sendo que dentro desta categoria existem várias possibilidades de entrada e saída, correspondentes às perguntas feitas por usuários e às respostas que o *chatterbot* pode fornecer a estas perguntas. As figuras 1 e 2 exibem o exemplo de uma categoria AIML da base de conhecimentos e o respectivo resultado retornado pelo *chatterbot*.

```
1 <category>
2   <pattern>CRITERIO PARA ENQUADRAMENTO NO SILIS</pattern>
3   <template>
4     Os criterios para saber se seu empreendimento e de baixo potencial poluidor e passivel de utilizar o sistema SILIS,
5     acesse o link <a href="http://silis.cetesb.sp.gov.br/pdf_criterio/criterios_para_classificacao_SILIS.pdf" target="_blank">aqui</a>.
6   </template>
7 </category>
```

Figura 1. Categoria que trata a entrada “Critério para enquadramento no SILIS?”.

O levantamento das informações que foram inseridas na base de conhecimento do *chatterbot* foi feito através de consulta bibliográfica, pesquisa no *Website* da CETESB e entrevistas com profissionais que atuam na área ambiental.

Após o desenvolvimento do protótipo, este foi testado por usuários da CETESB e profissionais da área ambiental, gerando arquivos de log, o que permitiu fazer uma análise qualitativa do sistema. O histórico das conversas foi analisado e então a base de conhecimento foi refinada com o intuito de abranger o máximo possível do escopo do conhecimento proposto, num processo contínuo de atualização, aprimorando sua base de conhecimento e a qualidade das respostas fornecidas.

Program O Example GUI Page - HTML

Say:

Seeker: CRITERIO PARA ENQUADRAMENTO NO SILIS
CETESB BOT: Os criterios para saber se seu empreendimento e de baixo potencial poluidor e passivel de utilizar o sistema SILIS, acesse o link [aqui](#).

Figura 2. Resposta do bot para a pergunta “Critério para enquadramento no SILIS?”.

Além da base de conhecimentos específicos sobre a CETESB e sobre licenciamento ambiental (aspectoslic, preliminareslic, legislacaoambsp) também foi incluído o tratamento para entradas como apresentação do usuário e perguntas relacionadas ao perfil do *chatbot*, como idade, localização, nome, etc. As categorias contendo a base de conhecimento foram divididas, de acordo com sua função e tópico, em quinze arquivos AIML.

Sete dos 15 arquivos que compõe a base são arquivos AIML com a forma mais simples de categoria AIML, ou seja, que não possuem recursões ou os símbolos ‘*’ ou ‘_’. Essas categorias contêm as saídas que o *bot* fornecerá ao usuário, sendo esses arquivos AIML: cetesb.aiml, aspectoslic.aiml, apresentacoes.aiml, outras.aiml, perfil.aiml, preliminareslic.aiml e legislacaoambsp.aiml.

Outros sete arquivos AIML possuem categorias responsáveis por tratar sinônimos, ou seja, diferentes entradas que fornecem uma mesma informação como saída, que é fornecida pelas categorias dos arquivos citados anteriormente. Os arquivos AIML com as categorias recursivas são: cetesb_var.aiml, preliminares_var.aiml, legislacaoambsp_var.aiml, perfil_var.aiml, aspectoslic_var.aiml, apresentacoes_var.aiml e outras_var.aiml. Essas categorias possuem no seu *template* chamadas para as categorias mais simples citadas anteriormente. Dessa forma, esses arquivos AIML tratam as variações para as entradas que são tratadas nos arquivos AIML contendo as categorias mais simples.

Um último arquivo AIML, o default.aiml, possui apenas uma categoria que trata qualquer outra entrada que não for reconhecida por nenhuma das categorias especificadas nos outros arquivos AIML. Essa categoria tem a função de fornecer uma saída padrão para o usuário, caso nenhuma das outras categorias da base de conhecimentos possua um modelo para a entrada do usuário.

A categoria *default* (padrão) então fornece a saída “*Não entendi sua frase, pode reformular para mim?*”, que solicita que o usuário escreva sua pergunta de uma forma diferente, que então pode ser reconhecida por uma categoria AIML, retornando uma resposta coerente.

4. Resultados

A base de conhecimento foi incluída na plataforma Program-O (figura 3) e disponibilizada para teste entre funcionários da CETESB e profissionais do ramo ambiental da Usina Açucareira São Manoel. Uma vez que um usuário realize uma

conversa com o *chatterbot*, a plataforma mantém o registro desse diálogo (por meio de uma ferramenta log), o que permite que a base de conhecimentos seja verificada em termos de fragilidades. Os testes com os usuários se deram de forma incremental, ou seja, à medida que os usuários iam interagindo com o *chatterbot*, seus registros iam sendo acompanhados e, assim, a base de conhecimento ia sendo refinada.

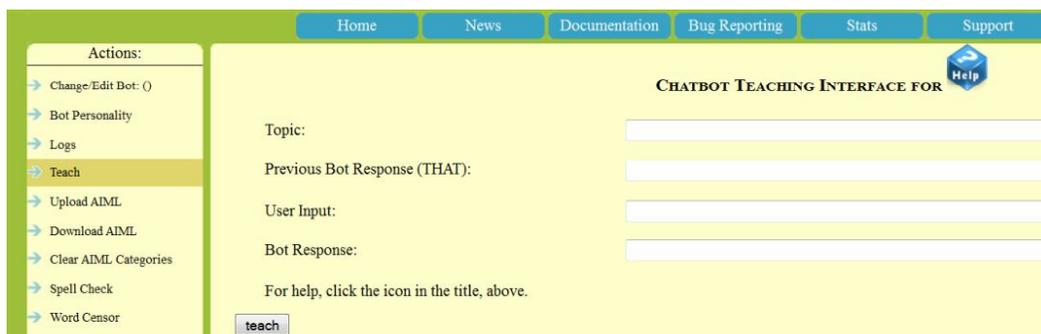


Figura 3. Página inicial da plataforma Program-O.

A partir dos registros de conversas gerados por estes testes foi possível identificar novas entradas de usuários, cujas saídas foram adicionadas à base de conhecimento. A figura 4 mostra um trecho do arquivo de log.

Program O Example GUI Page - HTML

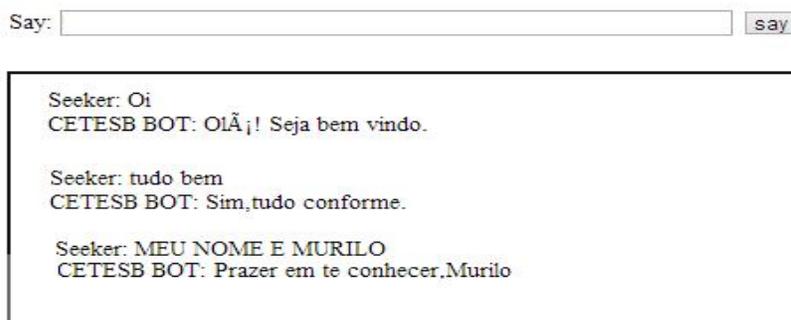


Figura 4. Trecho de um arquivo de log.

Também foi possível identificar fatores que dificultam uma resposta precisa do *chatterbot*, como erros de digitação do usuário. Além disso, outro fator limitativo é que a versão do programa para o processamento da língua portuguesa ainda está passando por melhorias e correções que estão sendo feitas nas versões utilizadas pelo program-o e no momento não é possível o tratamento de acentuação como o uso de acento agudo, circunflexo, cedilha. Para se atingir um maior grau de amadurecimento e precisão, a base de conhecimentos deve ser submetida a um processo contínuo de aprimoramento e refinamento, o que exige tempo e estudo.

Outra dificuldade encontrada foram as diferentes maneiras de se perguntar sobre o mesmo assunto. Uma mesma pergunta pode ser feita de várias maneiras, sendo impossível prever todas elas na base de conhecimento inicial. Por isso, a análise dos registros de conversação é fundamental. Nela são identificadas as diferentes formas que o usuário pergunta pela mesma informação, possibilitando que a base de conhecimento seja atualizada.

Uma alternativa de avaliação da base quando não se tinha a possibilidade de instalação do program-o, foi a aplicação de um questionário de avaliação que permitiu através de screenshots das bases e de seis perguntas relacionada ao funcionamento, eficiência e design fazer uma avaliação e analisar sugestões, críticas, elogios ao chatterbot.

Essa etapa de testes com os usuários é de fundamental importância para o desenvolvimento do protótipo, pois por mais robusta que seja a base de conhecimento, sempre é possível refiná-la de modo a melhorar a qualidade das respostas fornecidas.

5. Conclusão

A dificuldade que os usuários encontram para acessar os computadores, seja ela por falta de treinamento ou habilidade, é um fator limitativo tão importante que justifica o desenvolvimento de projetos que facilitem o acesso aos computadores. Neste contexto, ferramentas como os *chatterbots* apresentam-se como alternativas ao acesso à informação.

O componente deste trabalho pode ser utilizado como uma ferramenta de disseminação de conhecimentos, uma vez que as informações disponibilizadas pelo *chatterbot* ao público podem ser facilmente acessadas, por meio de uma *interface* que faz uso de linguagem natural, utilizando técnicas de Inteligência Artificial para desenvolvimento de agentes de conversação. Tal ferramenta apresenta adicionalmente como benefício o baixo custo devido à utilização de softwares gratuitos. A parte mais custosa da ferramenta refere-se à construção da base de conhecimentos, que demanda bastante tempo e exige um estudo apurado de conceitos de Inteligência Artificial, sobretudo relacionados ao Processamento das Linguagens Naturais, de forma a se criar um agente que simule o comportamento humano em um diálogo.

Embora a ferramenta construída tenha alcançado os objetivos propostos neste trabalho, por se tratar de um protótipo elaborado com a finalidade de demonstrar o potencial da tecnologia aplicada, a ferramenta não supre, obviamente, todas as questões relacionadas ao tema proposto e, por isso, não responde plenamente a todos os questionamentos possíveis de serem encontradas em situações reais. Outro fator limitativo é que a língua portuguesa ainda está passando por melhorias e correções que estão sendo feitas nas versões utilizadas pelo program-o e no momento não é possível o tratamento de acentuação como o uso de acento agudo, circunflexo, cedilha. Para se atingir um maior grau de amadurecimento e precisão, a base de conhecimentos deve ser submetida a um processo contínuo de aprimoramento e refinamento, o que exige tempo e estudo.

Outra limitação associada ao *chatterbot* refere-se à coerência das respostas. A coerência do diálogo envolvendo uma máquina é, seguramente, um dos maiores desafios do desenvolvimento de um agente de conversação, já que em situações reais uma conversa tende a ser bastante ampla, envolvendo vários temas.

Quanto a aplicação do questionário avaliativo analisando-se as respostas dos participantes da pesquisa forma obtidos alguns resultados importantes analisando qualitativamente o bot e mostrando o seu potencial para a tarefa proposta nesta investigação. Como trabalhos futuros podem citar a ampliação e o tratamento de acentuação das palavras da base de conhecimento, incrementando as categorias manipuladas pelo protótipo. Além disso, a divulgação e a disponibilização do agente em

larga escala, permitirá a coleta de um número expressivo de arquivos de registro de conversa, refletindo diretamente na qualidade do processo de aprimoramento da base de conhecimentos.

Referências

- CARVALHO, J. O. F. O papel da interação humano-computador na inclusão digital. In: **Revista Transinformação**, V. 15, n. 3, edição especial, p. 75-89, Campinas, 2003.
- CETESB: Companhia de Tecnologia de Saneamento Ambiental, 2013. O licenciamento ambiental e suas cartilhas. Disponível em <<http://www.cetesb.sp.gov.br/licenciamento/licenciamento-ambiental/1-pagina-inicial>>. Acesso em 27 abril. 2013.
- FÉ, I. M. M. **AIML aplicada à criação de personagens interativos**, 2005. 67f. Trabalho de Conclusão de Curso (Graduação) – Instituto Tecnológico de Aeronáutica, São José dos Campos.
- LAVEN, S. The Simon Laven Page. Disponível em: <<http://www.simonlaven.com>>. Acesso em: 01 maio 2013.
- LEONHARDT, M. D.; CASTRO, D. D.; DUTRA, R. L. S.; TAROUÇO, L. M. R., 2003 **ELEKTRA**: Um Chatterbot para Uso em Ambiente Educacional. In: *Renote*. Porto Alegre. Disponível em: <<http://penta3.ufrgs.br/~elektra/info/artigos/chatterbot-Elektra%5B1%5D.PDF>>. Acesso em: 05 mar. 2013.
- PRIMO, A. F. T. **O uso de chatterbots na educação à distância**, 2000. Disponível em: <http://www.nied.unicamp.br/oea/mat/chatterbots_lec.pdf>. Acesso em: 26 ago. 2013.
- RABELLO, R. S.; MARCHI, A. C. B. De; FOSSATTI, M. **AGEBOT**: Um *chatterbot* em AIML voltado a responder questões sobre epilepsia. In: XI Workshop de informática médica, 2011, Natal. Anais do WIM, 2011. Disponível em: <http://www.dimap.ufrn.br/csbc2011/anais/eventos/contents/WIM/WIM_Sessao_4_Artigo_2_Fossatti.pdf>. Acesso em: 23 maio 2013.
- RINGATE, T. AIML Reference Manual. **Alicebot.org**, 2001. Disponível em <<http://www.alicebot.org/documentation/aiml-reference.html>>. Acesso em: 14 abr. 2013.
- RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. 2 ed. Rio de Janeiro: Elsevier, 2004.
- SILVA, Jeferson Luís da - **Chatterbots – Simuladores de Diálogo** – 2ª revisão – Bookess Editora – Impresso e em PDF-;
- SMA: Secretaria Meio Ambiente, 2013. **Quem somos e o que fazemos**. Disponível em <<http://www.ambiente.sp.gov.br/a-secretaria/quem-somos/>>. Acesso em 22. maio 2013.
- TALKING about aiml predicates. **Alicebot**, c2012. Disponível em: <http://www.alicebot.org/video_transcripts/predicates.txt>. Acesso em: 4 jul. 2013.
- WALLACE, R. S. The Anatomy of A.L.I.C.E. **Alicebot.org**, 2009. Disponível em: <<http://www.alicebot.org/anatomy.html>>. Acesso em: 18 abr. 2013.
- WALLACE, R. S. **Be Your Own Botmaster**. 2005. 2ª ed. ALICE A.I. Foundation, Inc.;