

DEPARTAMENTO DE COMPUTAÇÃO
CURSO DE CIÊNCIA DA COMPUTAÇÃO

Eduardo Gabriel Côrtes

**PROCESSAMENTO DE LINGUAGEM NATURAL PARA INTERAÇÃO
ENTRE USUÁRIOS E AGENTES PEDAGÓGICOS EM AMBIENTES
VIRTUAIS DE APRENDIZAGEM**

Santa Cruz do Sul
2016

Eduardo Gabriel Côrtes

**PROCESSAMENTO DE LINGUAGEM NATURAL PARA INTERAÇÃO ENTRE
USUÁRIOS E AGENTES PEDAGÓGICOS EM AMBIENTES VIRTUAIS DE
APRENDIZAGEM**

Trabalho de Conclusão II apresentado ao Curso de
Ciência da Computação da Universidade de Santa Cruz
do Sul para obtenção parcial do título de Bacharel em
Ciência da Computação.

Orientadora: Prof^a. Me. Daniela Duarte da Silva Bagatini

Coorientadora: Prof^a. Dra. Rejane Frozza

Santa Cruz do Sul
2016

RESUMO

Com o objetivo de melhorar o processo de aquisição de conhecimento em ambientes virtuais de aprendizagem, buscaram-se metodologias de interação mais naturais entre o estudante e o ambiente. Como a linguagem é o principal meio de comunicação utilizado pelos humanos, um ambiente que permita ao estudante expressar-se através da sua linguagem natural, pode agregar conforto no processo de interação. Além disso, é comum que estes ambientes virtuais disponham de agentes pedagógicos para auxiliar o estudante no processo de aprendizagem. O propósito deste trabalho consiste em desenvolver a capacidade de *chatterbot* do agente pedagógico Dóris, para interagir com o estudante através da linguagem natural escrita. Assim, através da pesquisa e desenvolvimento de técnicas e métodos eficientes de processamento de linguagem natural, busca-se acrescentar a característica de *chatterbot* ao agente Dóris. Os resultados obtidos permitiram concluir que métodos e técnicas de processamento da linguagem natural escrita podem beneficiar o processo de interação de um agente pedagógico com um usuário.

Palavras chaves: Ambiente Virtual de Aprendizagem, *Chatterbot Dóris*, Processamento de Linguagem Natural.

ABSTRACT

In order to improve the process of knowledge acquisition in virtual learning environments, are sought more natural interaction methodologies between the student and the environment. As language is the main means of communication used by humans, an environment that allows the student to express themselves through their natural language can add comfort to the interaction process. In addition, it is common for these virtual environments to have pedagogical agents to assist the student in the learning process. The purpose of this work is to develop the chatterbot capacity of the pedagogic agent Dóris, to interact with the student through written natural language. Thus, through the research and development of efficient techniques and methods of natural language processing, wanted to add the chatterbot characteristic to Agent Dóris. The results obtained allowed to conclude that methods and techniques of natural written language processing can benefit the interaction process of a pedagogical agent with a user.

Keywords: Virtual Learning Environment, Dóris Chatterbot, Natural Language Processing.

LISTA DE ILUSTRAÇÕES

FIGURA 1. AMBIENTE VIRTUAL UNISC.....	16
FIGURA 2. ÁRVORE DE PARSER COM ANEXO SEMÂNTICO.....	22
FIGURA 3. ETAPAS DO PROCESSAMENTO DA SENTENÇA.....	37
FIGURA 4. FUNÇÃO PARA ADICIONAR ESPAÇO ENTRE PONTUAÇÕES.....	38
FIGURA 5. EXEMPLO DE ETIQUETAGEM DE UMA SENTENÇA.....	40
FIGURA 6. DICIONÁRIO DE SINÔNIMOS.....	41
FIGURA 7. CÓDIGO AIML PARA PROCESSAMENTO PRAGMÁTICO.....	43
FIGURA 8. FUNÇÃO PARA COMPARAR SENTENÇA COM A BASE AIML.....	44
FIGURA 9. ETAPAS DO PROCESSAMENTO DA SENTENÇA EM CÓDIGO FONTE.....	45
FIGURA 10. INTERAÇÃO COM O <i>CHATTERBOT</i> DÓRIS NO AMBIENTE VIRTUAL DE APRENDIZAGEM.....	47
FIGURA 11. INTERAÇÃO COM O <i>CHATTERBOT</i> DÓRIS DURANTE A AULA “DIREÇÃO PREVENTIVA”.....	48

LISTA DE TABELAS

TABELA 1. TRABALHOS REALIZADOS NO AVA UNISC.....	17
TABELA 2. QUADRO COMPARATIVO ENTRE OS TRABALHOS RELACIONADOS.....	28
TABELA 3. QUADRO DE RESULTADO DA BIBLIOMETRIA QUANTITATIVA.....	32
TABELA 4. QUADRO COM OS RESULTADOS DO ESTUDO DE CASO.	50
TABELA 5. COMPARAÇÃO DESTE TRABALHO COM OS RELACIONADOS	55

LISTA DE ABREVIATURAS

3D	Três dimensões
AIML	<i>Artificial Intelligence Markup Language</i>
ALICE	<i>Artificial Linguistic Internet Computer Entity</i>
API	<i>Application Program Interface</i>
AVA	Ambiente Virtual de Aprendizagem
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
JSE	<i>Java Standard Edition</i>
JVM	<i>Java Virtual Machine</i>
OS	<i>Operating system</i>
PLN	Processamento de Linguagem Natural
SGBD	Sistema Gerenciador de Banco de Dados
UFRGS	Universidade Federal do Rio Grande do Sul
UNISC	Universidade de Santa Cruz do Sul
XML	<i>Extensible Markup Language</i>

SUMÁRIO

SUMÁRIO	8
1 INTRODUÇÃO	9
1.1 Justificativas e Problema de Pesquisa	11
1.2 Objetivos.....	12
2 REFERENCIAL TEÓRICO	13
2.1 Ambientes Virtuais de Aprendizagem.....	13
2.2 Ambiente Virtual de Aprendizagem – Estudo de Caso com o Agente Dóris	15
2.3 Agentes.....	17
2.4 Processamento de Linguagem Natural	20
2.5 Trabalhos Relacionados	23
2.5.1 A pattern based approach for the derivation of base forms of verbs from participles and tenses for flexible NLP (RAJ; ABDUL-KAREEM, 2011).....	23
2.5.2 A Proposal of Topic Map Based Chatterbot for Non-English Natural Language Input (KIMURA, 2015).....	25
2.5.3 Building a Hybrid: Chatterbot – Dialog System (DINGLI; SCERRI, 2013). 26	
2.5.4 Quadro comparativo dos trabalhos relacionados estudados.....	28
3 METODOLOGIA	31
3.1 Estudo de referenciais teóricos e levantamento de trabalhos relacionados	31
3.2 Criação e estudo de técnicas de processamento de linguagem natural	33
3.3 Estudo de Caso com o agente Dóris	34
4 CHATTERBOT DÓRIS	36
4.1 Métodos e Técnicas de PLN.....	36
4.1.1 Pré-processamento da sentença do usuário.....	37
4.1.2 Etiquetador Morfossintático.....	39
4.1.3 Processamento Semântico	41
4.1.4 Processamento Pragmático	42
4.1.5 Consulta na Base de Conhecimento	43
4.2 Estudo de Caso com o <i>chatterbot</i> Dóris.....	46
4.2.1 Resultados do Estudo de Caso com <i>chatterbot</i> Dóris.....	49
4.3 Análise dos Resultados	51
5 CONCLUSÃO	56

1 INTRODUÇÃO

A computação está cada vez mais presente no cotidiano, mas ainda encontram-se áreas de aplicação que apresentam desafios, seja por motivos de dificuldade, complexidade ou outros. A capacidade de uma pessoa se comunicar de forma espontânea com um computador por meio da linguagem natural humana é um desafio que ainda deve ser explorado, já que o sucesso do mesmo pode trazer benefícios como: facilitar a interação entre o homem e o computador; permitir converter o conhecimento em linguagem natural para uma linguagem mais formal, o que admite que programas de computadores manipulem essas informações; também, permitir que o computador consiga expressar seu conhecimento aos usuários por meio da linguagem natural.

A comunicação por meio da linguagem é uma das características que diferenciam o homem do animal, mesmo levando em consideração que animais, como chimpanzés e golfinhos, utilizam vocabulários com centenas de sinais. Somente os seres humanos têm a capacidade de se comunicar de forma confiável com um número ilimitado de mensagens qualitativamente diferentes sobre qualquer tema utilizando sinais discretos (RUSSELL; NORVIG, 2013).

Neste trabalho, buscam-se técnicas de processamento de linguagem natural escrita que permitam que o usuário interaja com o agente pedagógico Dóris por meio de uma interface de texto. O agente Dóris é um agente pedagógico tutor que tem a função de guiar o usuário dentro do ambiente virtual de aprendizagem a fim de auxiliá-lo no processo de aprendizagem. Este ambiente foi desenvolvido pelo grupo de pesquisa de professores e estudantes do Departamento de Computação e do Programa de Pós-Graduação em Sistemas e Processos Industriais da UNISC (FROZZA, *et al.*, 2011), (HOMRICH; FROZZA; SCHREIBER, 2013), (FROZZA, *et al.*, 2009) e (BORIN *et al.*, 2012).

Segundo Rezende (2003), a linguagem natural é a maneira mais espontânea das pessoas descreverem o conhecimento, mas, por outro lado, ela apresenta uma série de características que dificultam o seu processamento computacional. As sentenças em linguagens naturais podem ser ambíguas, inconsistentes, dependentes de contexto, imprecisas e incompletas, além de serem combinatoriamente explosivas, uma vez que o número de vocábulos é grande, o número de possíveis combinações de vocábulos para a formação de sentenças é infinitamente grande.

Em (BANDA; MENEZES, 2012), ao propor a utilização de técnicas de processamento de linguagem natural na construção de agentes para apoiar no esclarecimento de dúvidas de usuários, foi concluído que a utilização dessas técnicas aliadas à pesquisa com padrões morfossintáticos permite identificar estruturas de conhecimento em textos.

Conforme Russell e Norvig (2013), um agente computacional que deseja adquirir conhecimento precisa entender, no mínimo e parcialmente, a ambígua e confusa linguagem que os seres humanos usam. Em (WOOLDRIDGE, 2001), define que um agente é um sistema computacional que está situado em algum ambiente, e que é cabível de ações autônomas neste ambiente a fim de satisfazer seus objetivos designados.

No trabalho de Wooldridge (2001) sugere-se três recursos que podem ser esperados de um agente inteligente: proatividade, que é apresentar um comportamento dirigido a um objetivo e tomar a iniciativa a fim de satisfazê-lo; reatividade, que é estar apto a perceber o seu ambiente e responder em tempo hábil às mudanças para satisfazer seus objetivos designados; habilidade social, que é a capacidade de interação com outros agentes (e possíveis humanos) a fim de satisfazer seus objetivos.

Os *chatterbots* são uma categoria de agentes inteligentes programados para simulação de conversas semelhantes aos bate-papos virtuais através de troca de mensagens de texto (TEIXEIRA *et al.*, 2005). São baseados na ideia básica da interação entre pessoas através da conversação e, tomam como base, o trabalho desenvolvido por Turing (1950) que propôs um jogo chamado "The Imitation Game", cujo o propósito é fazer com que um ser humano acredite estar conversando com outra pessoa. Logo, para uma conversação convincente, os *chatterbots* geralmente utilizam técnicas de processamento de linguagem natural nas suas soluções para a criação da base de conhecimento, assim, eles podem atender as expectativas e necessidades dos usuários. Essa categoria de agentes apresenta um grande potencial como agentes pedagógicos, pois possuem autonomia e desenvoltura para direcionar o assunto do estudo de forma natural, o que permite ao usuário sentir-se mais à vontade na troca de informação com o sistema (SGANDERLA; FERRARI; GEYER, 2003).

Outra categoria de agentes inteligentes é o agente pedagógico, que procura agir de modo a apoiar o processo de ensino e aprendizagem, levando em consideração os meios do usuário, e intervir diretamente com o mesmo, evidenciando como um tutor de conteúdo ou métodos mais adaptados ao perfil do usuário (KAMPFF *et al.*, 2005).

De acordo com (KERLY; ELLIS; BULL, 2009), um agente que permita a interação por meio de uma interface de texto pode oferecer melhorias na aprendizagem em ambientes virtuais, mas há uma vasta gama de questões e desafios, desde o processo de conversa, até questões técnicas relativas à execução da tecnologia de conversação.

O agente pedagógico deste trabalho está inserido em um ambiente virtual de Aprendizagem (AVA). Em (BEHAR *et al.*, 2007), entende-se que ambiente virtual de Aprendizagem é uma plataforma de *software* (infraestrutura tecnológica composta pelas funcionalidades e interfaces gráficas), que busca dar suporte à aprendizagem do usuário via *web*.

1.1 Justificativas e Problema de Pesquisa

Tendo em vista a justificativa social para o desenvolvimento deste tema, ainda se encontra certa resistência por parte de alguns usuários em relação ao computador, já que a computação é algo novo para muitos deles e o medo do desconhecido é comum. Logo, uma interface que permita ao usuário interagir com a máquina pela mesma linguagem que ele utiliza para se comunicar com outros humanos, poderá agregar certo conforto ao usuário em relação a sua interação com o computador.

A justificativa científica para o desenvolvimento desse tema tem foco na área de processamento de linguagem natural que contém diversos desafios, logo, este trabalho poderá trazer contribuições para a área, já que serão buscadas e desenvolvidas técnicas e métodos baseados em processamento de linguagem natural para auxiliar a interação entre usuários e agentes pedagógicos.

Como justificativa empresarial para o desenvolvimento desse tema, um ambiente virtual de aprendizagem pode ser utilizado em organizações para capacitação e treinamento de seus colaboradores com o diferencial de permitir que os usuários fiquem livres para determinar os momentos de interação com o ambiente.

Define-se o problema de pesquisa deste trabalho: Como o processamento de linguagem natural pode contribuir na interação usuário e agente pedagógico?

1.2 Objetivos

O objetivo principal deste trabalho foi desenvolver a característica de *chatbot* ao agente Dóris através da pesquisa e desenvolvimento de técnicas e métodos eficientes de processamento de linguagem natural, buscando um processo de interação mais natural entre o agente e o usuário no ambiente virtual de aprendizagem já desenvolvido.

Sendo assim, foi possível atingir os seguintes objetivos específicos:

- Estudar sobre a interação entre homem e computador, por meio da linguagem natural.
- Pesquisar técnicas e métodos eficientes de processamento de linguagem natural.
- Analisar quais dessas técnicas e métodos podem ser utilizadas para compreender as sentenças de entrada informadas pelo usuário.
- Comparar as técnicas, a fim de selecionar quais serão implementadas no ambiente virtual de aprendizagem.
- Implementar as técnicas escolhidas no ambiente e permitir que o usuário expresse o conhecimento através de uma interface de texto semelhante a um bate-papo virtual, cujo o agente Dóris buscará compreender a informação de entrada e gerar uma resposta coerente à sentença de entrada, dando sequência ao processo de interação.

Além da seção introdutória, o segundo capítulo apresenta o referencial teórico. Os temas abordados são os ambientes virtuais de aprendizagem, contendo uma seção para o ambiente virtual UNISC já desenvolvido, Agentes computacionais, incluindo agentes do tipo pedagógico e *chatbots*, Processamento de Linguagem Natural e os Trabalhos relacionados aos temas de pesquisa deste trabalho.

O terceiro capítulo apresenta a metodologia utilizada para alcançar os objetivos do trabalho. O capítulo apresenta as partes fundamentais na metodologia que estão divididas em três seções. A primeira refere-se ao estudo de referências teóricas e levantamento de trabalhos relacionados, a segunda seção faz referência ao estudo de técnicas de processamento de linguagem natural e suas aplicações, e a terceira refere-se ao estudo de caso com o agente Dóris a fim de analisar os resultados no processo de interação com o usuário.

No quarto capítulo apresenta-se o desenvolvimento do *chatbot* Dóris, especificando cada método e técnica utilizados, também o estudo de caso realizado e os dados obtidos no mesmo. Por fim, apresenta-se a análise dos resultados obtidos no estudo de caso. O quinto capítulo conclui o trabalho através das conclusões e trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta a fundamentação teórica referente aos assuntos envolvidos na pesquisa, conceituando e caracterizando os ambientes virtuais de aprendizagem, agentes e as suas categorias pedagógicas como os *chatterbots*, o ambiente virtual utilizado como estudo de caso, processamento de linguagem natural e trabalhos relacionados.

2.1 Ambientes Virtuais de Aprendizagem

A possibilidade de ter acesso a uma quantidade relativamente grande de informação dentro de casa a qualquer hora é uma realidade mais do que comum no século XXI, já que a *internet*, a maior causadora desse fato, possibilita acessar um grande número de informação e, conseqüentemente, proporciona um espaço de busca por experiências compartilhadas e conhecimento.

É comum que estudantes, profissionais ou curiosos utilizem a *internet* para estudo ou treinamento, logo, segundo (MORAN, 2000), é necessária uma infraestrutura tecnológica que permite interações espaço-temporais mais livres, adaptação a ritmos diferentes dos alunos, maior liberdade de expressão a distância e novos contatos com pessoas de interesses semelhantes, fisicamente distantes.

A *internet* possibilita a existência de redes de aprendizagem, que são definidas por (HARASIM, *et al.*, 2005) como grupos de pessoas que se comunicam com o mesmo propósito de aprendizagem, proporcionando que cada uma possua o seu ritmo particular de aprendizagem, horário e local próprio para a realização das mesmas. Com o auxílio das redes de aprendizagem, podem ser criados ambientes nos quais professores e alunos, em locais diferentes, construam juntos o entendimento relacionado a um assunto particular. Esses ambientes são conhecidos como ambientes virtuais de aprendizagem (AVAs). Segundo Valentini e Soares (2008, p. 16):

[...] a expressão “ambiente virtual de aprendizagem” está relacionada ao desenvolvimento de condições, estratégias e intervenções de aprendizagem num espaço virtual na Web, organizado de tal forma que propicie a construção de conceitos, por meio da interação entre alunos, professores e objeto de conhecimento.

(BEHAR, *et al.*, 2007) define que um AVA é constituído por uma infraestrutura tecnológica e por todas as relações, sejam elas afetivas, cognitivas, simbólicas, entre outras, estabelecidas pelos sujeitos participantes, tendo como foco principal a aprendizagem. Assim, um AVA não necessariamente estará presente apenas em processos de aprendizagem a distância, mas também pode ser utilizado como suporte na aprendizagem presencial.

Diferentes autores justificam a utilização de ambientes virtuais de aprendizagem. (FERNANDES, 2007) comenta que os AVAs podem juntar em um único lugar diversas oportunidades de aprendizado como, atividades interdisciplinares, tarefas *on-line*, conferências, projetos, cursos ou novas ideias. Já para (FROZZA, *et al.*, 2011), esses ambientes vão além da ideia de um conjunto de páginas educacionais, mas sim um espaço social constituído de interações cognitivo-sociais sobre um objeto de conhecimento, que pode ser considerado como o assunto ou tema de estudo. (SILVA; HECK, 2008) complementa que as utilizações de recursos tecnológicos nas atividades de ensino têm sido relatadas por estudantes e professores como experiências agradáveis.

Atualmente, diversos pesquisadores desenvolvem trabalhos na área de ambientes virtuais de aprendizagem. (FRADE *et al.*, 2014), propuseram um ambiente virtual multiagentes em três dimensões para apoiar a recomendação de Objetos de Aprendizagem, a fim de melhorar os processos de ensino a distância. (SILVA *et al.*, 2015) em seu artigo, utiliza mineração de dados educacionais para descrever as análises dos dados oriundos do banco de dados do ambiente virtual *Moodle*. Busca-se, aqui, destacar o AVA desenvolvido no Projeto de Pesquisa de professores e estudantes do Departamento de Computação e do Programa de Pós-Graduação em Sistemas e Processos Industriais (PPGSPi) da UNISC, com pesquisas em andamento desde o trabalho de (SANTOS *et al.*, 2001). Este será o AVA estudo de caso deste trabalho e será descrito na próxima seção.

2.2 Ambiente Virtual de Aprendizagem – Estudo de Caso com o Agente Dóris

O ambiente virtual de aprendizagem utilizado neste trabalho foi desenvolvido pelo grupo de pesquisa de professores e estudantes do Departamento de Computação e do Programa de Pós-Graduação em Sistemas e Processos Industriais da UNISC, no qual foi cadastrado no *site* dos Grupos de Pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

O primeiro trabalho realizado no ambiente foi em 1998 (SANTOS *et al.*, 2001), onde o propósito era desenvolver o agente de acompanhamento pedagógico Dóris para Sistemas Tutores Inteligentes. Este agente tem como tarefa auxiliar na modelagem do perfil do estudante e guiar e monitorar o estudante dentro do ambiente. Este trabalho concluiu que a utilização do agente Dóris permitiu uma interação mais agradável do aluno com o processo de ensino-aprendizagem, tornando mais atraente o ambiente e motivando o aluno a participar deste processo. Em 2002, foi desenvolvido o trabalho de (SILVA, 2002) com o propósito de inserir o novo agente Dimi no ambiente. Esse novo agente tem o propósito de ser um colega do aluno no ambiente. Assim, além de ajudar o aluno nas atividades do ambiente, ele também é responsável por obter suas características cognitivas.

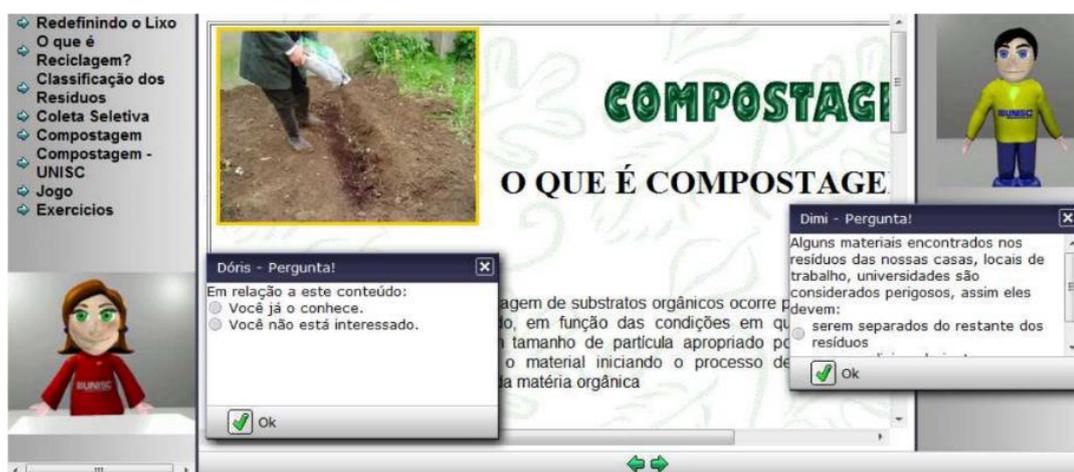
No trabalho de (FROZZA *et al.*, 2007) é apresentado um ambiente educacional baseado em estilos cognitivos de raciocínio utilizando os agentes Dóris e Dimi que desempenham diferentes funções, o agente tutor e o agente companheiro. As ações dos agentes foram desenvolvidas com o uso de regras de produção, a fim de coletar informações para a detecção do perfil do estudante através de redes *bayesianas* e teste de Ross. Em (FROZZA *et al.*, 2011), o intuito foi de apresentar o desenvolvimento e a atuação dos agentes Dóris e Dimi animados, expressando emoções, integrados no ambiente virtual de aprendizagem, a fim de interagir com o estudante.

Em (BORIN, 2010), o agente Dóris passou a expressar sete emoções (alegria, tristeza, expectativa, dúvida, atenção, surpresa e indignação) na interação do ambiente com o usuário, além de ser modelada em 3D (três dimensões). Este trabalho concluiu que a utilização do ambiente virtual de aprendizagem com o uso do agente Dóris expressando emoções é um recurso didático que traz contribuições ao processo de

ensino e aprendizado, tornando o assunto atraente e prazeroso ao usuário. Em 2010, o agente companheiro Dimi foi desenhado com uma caricatura no formato 3D.

Um dos problemas encontrados em AVAs multiagentes é a comunicação simultânea dos agentes com o aluno. Um exemplo disto pode ser visto na figura 1. Em (GRIESANG; FROZZA. 2013), foi desenvolvido um mecanismo de comunicação entre os agentes Dóris e Dimi, com objetivo de permitir a interação organizada e coordenada dos agentes pedagógicos com o usuário. Assim, através de um protocolo de comunicação, foi possível evitar a comunicação simultânea dos agentes pedagógicos com o aluno.

Figura 1. Ambiente Virtual UNISC.



Fonte: (GRIESANG; FROZZA, 2013).

O ambiente virtual de aprendizagem deste trabalho está sendo estudado e aperfeiçoado constantemente desde seus primeiros trabalhos realizados até o presente momento (SANTOS *et al.*, 2001), (SILVA, 2002), (MERGEN e SCHREIBER, 2005), (MAINIERI *et al.*, 2005), (FROZZA *et al.*, 2011), (SILVA *et al.*, 2010), (BORIN, 2010), (GRIESANG; FROZZA, 2013), (HOMRICH, FROZZA e SCHREIBER, 2013), (SCHAEFER, FROZZA, SILVEIRA, RUTSATZ, 2015). Porém, o ambiente não contém uma interação entre o usuário e a agente Dóris através de uma interface de texto em linguagem natural, objetivo principal deste trabalho.

Para muitos usuários de um ambiente virtual de aprendizagem, a interação com o agente tutor pode trazer certo desconforto ao exercer o papel de aluno, já que os costumes dele é desempenhar esta atividade em uma sala de aula com um professor. De acordo com (KERLY; ELLIS; BULL, 2009), um agente que permita a interação por

meio de uma interface de texto pode oferecer melhorias na aprendizagem em ambientes virtuais. Logo, propor ao usuário uma comunicação através de uma interface de texto com a agente Dóris no AVA, pode proporcionar um processo de aprendizagem mais natural, já que o usuário estará usando sua própria língua natural para interagir com o ambiente.

Várias pesquisas já foram realizadas neste ambiente, e muitos resultados já foram atingidos, sempre agregando ao sistema funcionalidades no decorrer dos anos a nas evoluções das pesquisas, como mostra a Tabela 1.

Tabela 1. Trabalhos realizados no AVA UNISC

Evolução do Ambiente Virtual de Aprendizagem da UNISC Característica	2000	2002	2005	2007	2009	2010	2011	2012	2014
Agente Dóris	Sim	Sim	Sim						
Agente Dimi	-	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Guiar Estudante	Sim	Sim	Sim						
Coletar Informações	-	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Rede Bayesiana	-	-	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Teste de Ross	-	-	-	Sim	Sim	Sim	Sim	Sim	Sim
Emoções Dóris	3	3	3	3	7	7	7	7	7
Emoções Dimi	-	Fixa	Fixa	Fixa	Fixa	2	2	2	Fixa
Domínio de Conhecimento Dóris	Geral e específico	Geral e específico	Geral e específico						
Domínio de Conhecimento Dimi	-	Geral	Geral	Geral	Geral	Geral	Geral	Geral	Geral
Aparência Dóris	2D	2D	2D	2D	3D	3D	3D	3D	3D
Aparência Dimi	-	2D	2D	2D	2D	3D	3D	3D	3D
Interação entre agentes pedagógicos	-	-	-	-	-	-	-	Sim	-
Consideração da afetividade e estilo de aprendizagem do estudante	-	-	-	-	-	-	-	Afetividade	-
Consideração da animação e estilo de aprendizagem do estudante	-	-	-	-	-	-	-	-	Animação

Fonte: (SCHAEFER, FROZZA, 2015).

Conforme a figura 1, a agente Dóris interage com o usuário através de caixas de perguntas onde o mesmo deve escolher uma das respostas já definidas pela agente. O passo seguinte e proposta deste trabalho é possibilitar que o ambiente consiga interagir com o usuário de forma natural. Para isso, é proposto que a agente Dóris se comunique com o usuário através de uma interface de texto semelhante a bate-papos virtuais. Através da pesquisa e desenvolvimento de métodos e técnicas eficientes de processamento de linguagem natural, busca-se acrescentar a característica de *chatterbot* à agente Dóris. A próxima seção apresenta os principais conceitos de agentes.

2.3 Agentes

Um agente é um sistema computacional que está em um ambiente percebendo suas informações através de sensores (entrada), sejam eles câmeras, localizadores, sensores infravermelho, e reage de forma autônoma através de atuadores, a fim de

atender seus objetivos estabelecidos (WOOLDRIDGE, 2001), (RUSSELL; NORVING, 2013). Ferber (1999) define que um agente pode ser uma entidade física ou virtual que age e se comunica através de percepções de seu ambiente de forma autônoma e tem habilidades para atingir os seus objetivos.

Já segundo Bradshaw (1997), um agente é uma entidade de *software* que trabalha de forma contínua e independente sem a necessidade de orientação ou intervenção humana. O mesmo está inserido em um ambiente que, muitas vezes, é habitado por outros agentes e processos, nos quais o agente é sensível às mudanças do ambiente interagindo de forma flexível e inteligente com outros agentes, além de aprender com base em suas experiências.

De acordo com suas habilidades cognitivas, os agentes podem ser classificados em reativos e cognitivos (FERBER, 1999):

- Agentes reativos são baseados nos modelos de organização biológica ou etológica, em que um único agente não pode ser considerado inteligente, mas o comportamento em grupo desses agentes sim. Além disso, os comportamentos desses agentes são baseados em estímulo e resposta, no qual não consegue consultar seu histórico de experiências já que não possui memória.
- Agentes cognitivos são baseados em organizações sociais dos seres humanos e adquirem seu conhecimento através das percepções de mudanças do ambiente, raciocínio e comunicação direta com outros agentes. Assim, conseguem consultar o histórico de experiências passadas e raciocinar e planejar suas ações.

Wooldridge e Jennings (1995) definem um agente como uma entidade capaz de resolver problemas encapsulados, no qual é possível encontrar as seguintes propriedades:

- **Autonomia:** Capacidade de executar ações sem interferência de agentes computacionais ou humanos, contendo algum tipo de controle sobre suas ações e estado interno.
- **Habilidades Sociais:** Capacidade de interagir com outros agentes através de uma linguagem em comum.
- **Reatividade:** Capacidade de perceber o ambiente e responder em tempo hábil às mudanças que ocorrem nele.

- Proatividade: Capacidade de tomar iniciativa e exibir um comportamento direcionado por objetivos e não ficar limitado a tomar decisões quando houver mudanças no ambiente.

Uma categoria de agentes inteligentes são os agentes pedagógicos que, segundo Girafa (1999), são utilizados para fins educacionais podendo atuar como estudantes virtuais, companheiros virtuais de aprendizagem ou ainda tutores virtuais, tendo como objetivo auxiliar os estudantes no processo de ensino-aprendizagem. Esses agentes podem ser divididos em:

- *Goal-driven* (guiados por objetivos): Tutores, mentores e assistentes. Esses agentes possuem mobilidade e operam em diferentes contextos, sejam eles textuais, hipermídia, ou realidade virtual, por exemplo. A interação acontece no processo cooperativo ou competitivo entre os agentes artificiais e agentes humanos que se comunicam e realizam atividades.
- *Utility-driven* (guiados pela utilidade): São utilizados como agentes que realizam tarefas auxiliares ligadas às atividades pedagógicas, como: ajudar os estudantes a encontrar programas específicos, arquivos, diretórios, agendar tarefas, recordar atividades, entre outras.

Os *chatbots* são uma categoria de agentes inteligentes programados para simulação de conversas semelhantes aos bate-papos virtuais através de troca de mensagens de texto (TEIXEIRA *et al.*, 2005). São baseados na ideia básica da interação entre pessoas através da conversação e tomam como base o trabalho desenvolvido por Turing (1950), que propôs um jogo chamado *The Imitation Game*, cujo propósito é fazer com que um ser humano acredite estar conversando com outra pessoa.

(MIKIC *et al.*, 2009) define um *chatbot* como um tipo de agente de conversação entendido como um programa de computador feito para simular um diálogo inteligente. Ele processa as entradas dos usuários em uma linguagem natural e realiza uma consulta à sua base de dados de conhecimento para formular uma resposta semelhante à de um ser humano.

Essa categoria de agentes apresenta um grande potencial como agentes pedagógicos, pois possui autonomia e desenvoltura para direcionar o assunto do estudo de forma natural e permite ao usuário sentir-se mais à vontade na troca de informação

com o sistema (SGANDERLA; FERRARI; GEYER, 2003). A utilização de *chatterbots* para responder perguntas sobre um determinado assunto tornou-se bastante popular em sistemas que tentam prover uma interação mais natural com os usuários (BANDA; MENEZES, 2012).

Para simular uma conversação convincente, os *chatterbots* precisam fornecer capacidade de processamento da linguagem para envolver-se em uma conversação produtiva com usuário. Isto requer que o agente entenda perguntas e declarações dos usuários e retorne uma resposta cabível. Assim, é preciso empregar técnicas de gestão de diálogo eficazes, onde programadores de *chatterbots* utilizam uma técnica de inteligência artificial chamada Processamento de Linguagem Natural (LESTER; BRANTING; MOTT, 2004), (RUSSELL; NORVING, 2013). A próxima seção apresenta os principais conceitos desta técnica.

2.4 Processamento de Linguagem Natural

Processamento de Linguagem Natural (PLN) é uma área de pesquisa que explora como computadores podem ser usados para entender e manipular a linguagem natural de textos ou da fala. A PLN procura conhecer como os humanos conseguem compreender e usar a linguagem natural, no intuito de permitir que sistemas computacionais manipulem a linguagem para executar tarefas desejadas (CHOWDHURY, 2003). Para Liddy (1998), PLN é uma gama de técnicas computacionais para análise e representação de textos em linguagem natural em um ou mais níveis de análise linguística, com a finalidade de alcançar o processamento da linguagem humana para uma variedade de tarefas ou aplicações.

Segundo (LIDDY, 1998) o PLN faz uso de seis níveis da análise linguística que estão interligados e são classificados nos seguintes níveis:

- **Fonológico:** Refere-se à forma como as palavras são pronunciadas. Este nível é importante para o entendimento na linguagem falada e em sistemas de reconhecimento de voz. O nível fonológico não é importante para aplicações que irão fazer uso exclusivo de informações em textos escritos.
- **Léxico-Morfológico:** Analisa a estrutura, formação e classificação das palavras. O léxico define os conjuntos de palavras de uma língua

enquanto que os morfemas são responsáveis por informar o gênero, tempo e modo verbal, o número (singular ou plural) e a classe gramatical de uma palavra. Este nível é interessante em motores de buscas que consideram palavras derivadas de outras.

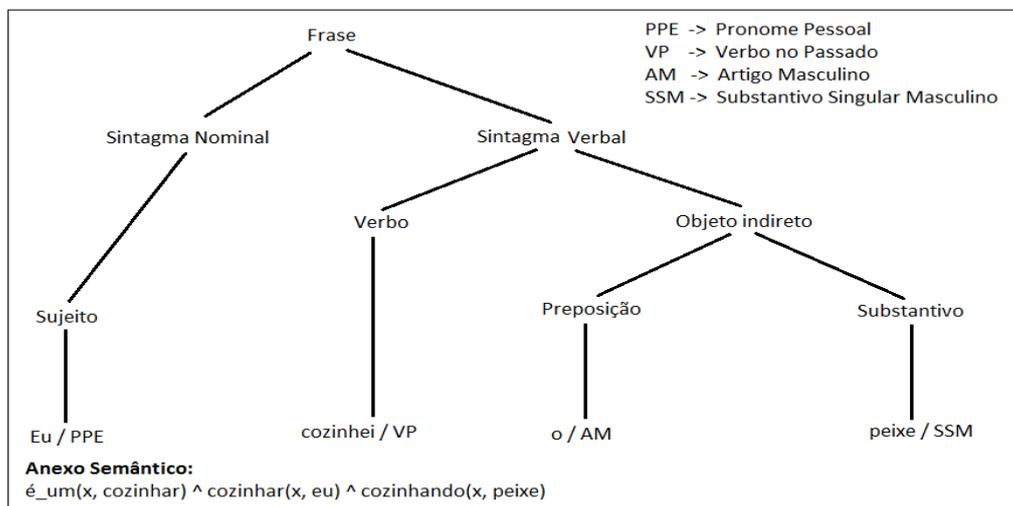
- **Sintático:** Analisa as palavras em uma frase a fim de descobrir a estrutura gramatical da sentença. Tem a função de identificar o significado de uma palavra em uma frase, além da sua relação com as outras palavras. Sistemas de PLN fazem uso dessa estrutura a fim de saber se uma palavra é um verbo ou sujeito, ou até mesmo o tipo de verbo que ela pode ser. Assim, quando a estrutura gramatical de uma sentença é armazenada, por exemplo, é possível que o sistema saiba a idade de João e a cor de sua casa na frase “João é um menino de 10 anos que mora em uma casa azul”.
- **Semântico:** Determina os possíveis significados de uma frase, já que uma palavra pode ter mais de um significado, este nível tem a função de identificar o significado mais adequado olhando o contexto da frase.
- **Discursivo:** Examina a própria estrutura de um texto para extrair um significado adicional. Por exemplo, em um artigo de jornal, normalmente, o primeiro parágrafo tem os fatos mais importantes da notícia deixando os impactos do evento para o final. Já em um romance de mistério, os segredos são revelados apenas no fim. Assim, o PLN usa essa estrutura previsível para entender o significado de um pedaço de informação que está em um documento.
- **Pragmático:** Abrange o uso intencional de uma língua em situações que requerem conhecimento do mundo. Para utilizar este nível de compreensão, é necessário que a aplicação tenha conhecimento amplo sobre o mundo e usá-lo como um guia de referência ou base de conhecimento.

Existem diversas técnicas de PLN que buscam solucionar os diversos desafios encontrados em cada nível da análise linguística. Para este trabalho, serão utilizadas técnicas dos níveis léxico-morfológico, sintático, semântico e pragmático.

Com o intuito de reduzir o tamanho do dicionário de palavras do sistema, no nível léxico-morfológico, busca-se armazenar apenas os radicais das palavras e seus afixos. Para isso, são utilizadas regras que analisam as palavras e as classificam segundo tabelas de afixos (SIMÕES; ALMEIDA, 2001). Por exemplo, a entrada *inho* em uma tabela de sufixos está associada ao diminutivo de substantivos, logo, *passarinho* é o diminutivo da palavra *pássaro*.

Na análise sintática, busca-se identificar a classes das palavras, para isso, são implementados *parsers* que utilizam técnicas de etiquetagem para indicar as classes das palavras que compõem uma frase (Figura 2). Segundo (BONFANTE; NUNES, 2002), o aprendizado estatístico, através do modelo *n-gramas* é muito utilizado para etiquetagem morfossintática que consiste em estabelecer uma estatística de probabilidade a cada *n* palavras. Geralmente *n* é igual a 3, onde são analisadas 2 palavras para a previsão de uma terceira. Assim, é definida uma estatística de construção de uma frase, admitindo encontrar qual é a palavra de combinação mais provável, além de permitir a verificação e correção da construção frasal (MULLER, 2002).

Figura 2. Árvore de *parser* com anexo semântico.



Fonte: ilustração elaborada pelos autores a partir de informações recolhidas na obra de Muller (2002).

O processo de análise semântica pode ocorrer durante o processo de análise sintática, onde é possível acrescentar anexos semânticos nas árvores de *parser* enquanto que a mesma é completada, o que permite dar precisão no objetivo das palavras e

distinguir certas categorias (MULLER, 2002). A figura 2 apresenta um exemplo de árvore de *parser* com o anexo semântico que define o verbo *cozinhar*, no qual necessita de um agente (sujeito que cozinha) e um paciente (algo que o agente cozinha).

A análise pragmática não analisa apenas uma frase, mas procura nas demais frases o entendimento do contexto que falta na análise de uma única frase. Para (VIEIRA; LIMA, 2001), sistemas que trabalham no nível pragmático costumam considerar o contexto linguístico, que é mais fácil de tratar computacionalmente, já que se refere ao que é especificado no texto. O mais difícil de se tratar é o contexto imediato de uma expressão, já que o conhecimento compartilhado entre participantes de uma conversa é difícil de se representar adequadamente.

A próxima seção apresenta os trabalhos relacionados selecionados, com temas em processamento de linguagem natural e *chatbots*.

2.5 Trabalhos Relacionados

Esta seção destina-se a apresentar os trabalhos relacionados aos temas de PLN e *chatbots*, com o propósito de contribuir para a realização do presente trabalho. Além disso, é apresentado um quadro comparativo e considerações.

2.5.1 A pattern based approach for the derivation of base forms of verbs from participles and tenses for flexible NLP (RAJ; ABDUL-KAREEM, 2011)

O artigo, apresentado por (RAJ; ABDUL-KAREEM, 2011), tem como objetivo principal desenvolver um algoritmo capaz de derivar verbos no particípio ou em diferentes tempos verbais em seus estados infinitivos. A implementação deste algoritmo é utilizada no desenvolvimento de um agente *chatbot* chamado RONE, que tem a característica de aprender com suas experiências. A derivação de verbos em sua forma permite obter os sinônimos e antônimos de verbos quando os mesmos estão no infinitivo. Durante o desenvolvimento desse agente, foi necessário criar um esquema de representação do conhecimento capaz de designar classes lexicais a partir de um conjunto de dados, definidas manualmente durante o desenvolvimento do sistema. Ao propor a utilização de PLN para o agente RONE, foi descoberto que é mais eficiente que o sistema acrescente seus próprios vocabulários, do que criar todos eles de uma só vez. Assim, foram desenvolvidas etapas para a inserção de vocabulários em sua base de

conhecimento. As etapas se baseiam nas regras gramaticas da língua inglesa e se dividem em:

- Utilizar a ajuda das gramáticas da linguagem para compreender sentenças, para isso, se aproveita a relação entre sujeito, verbo e objeto para formar o entendimento do que o usuário quer dizer.
- Depois de analisar cada palavra individualmente, é preciso combinar essas palavras para compreender seu significado na frase. Quando existe uma entrada com um grande número de palavras e apenas uma parte desta entrada está na base de conhecimento do sistema, o agente RONE não encontra uma palavra que corresponde ao restante desta entrada. Assim, quando uma palavra desconhecida aparece em uma sentença, RONE utiliza um algoritmo de predição, que consiste no reconhecimento de padrões muito usados por seres humanos. Este reconhecimento de padrão envolve a estrutura da sentença em termos de parte da frase por palavra. Por exemplo, se uma palavra vem depois dos artigos ‘a’, ‘o’, ‘um’ e ‘uma’, certamente a palavra desconhecida será um substantivo ou adjetivo.
- Na próxima etapa, quando RONE se depara com um novo verbo, o mesmo pode estar em diferentes tempos verbais ou na forma de particípio, logo, é necessário transformar o verbo para sua forma infinitiva antes de armazená-lo na base de conhecimento do sistema. Este novo verbo será encontrado através do algoritmo de predição e transformado para forma infinitiva com base nos seus padrões finais. Por exemplo, o verbo *Walk*, do inglês, poderá estar em diferentes formas como *Walks*, *Walked* ou *Walking*, mas é possível derivá-lo para o infinitivo apenas realizando algumas transformações na sua parte final. Um dos desafios desta etapa está ligado aos verbos irregulares que não seguem um padrão em sua estrutura, portanto, é preciso definir exceções para estas irregularidades no algoritmo desta etapa.

Em seguida, os verbos poderão ser analisados em suas denotações de passado, presente e futuro, onde é possível analisar que geralmente o primeiro verbo de uma oração na língua inglesa determina o tempo verbal da oração, mas algumas exceções podem ser encontradas. Para realizar os testes, o autor desenvolveu um algoritmo em linguagem de programação Java, que testa uma entrada de 100.000 verbos na língua

inglesa em padrões estruturados através de condições “if else” (se então) e determina a forma infinitiva dos verbos. Os testes indicaram que qualquer verbo pode ser decomposto corretamente em seu estado infinitivo. Mesmo que um verbo seja decomposto incorretamente, é possível adicionar outra condição em sua tabela de padrões. A representação da tabela de padrões é em lógica de primeira ordem. Uma vez que a forma infinitiva do verbo é adquirida, é possível obter o particípio e os tempos verbais do verbo utilizando técnicas semelhantes a de padrões finais.

2.5.2 A Proposal of Topic Map Based Chatterbot for Non-English Natural Language Input (KIMURA, 2015)

No trabalho de (KIMURA, 2015), foi proposto um *chatterbot* que busca dar respostas relevantes às questões feitas por um usuário, cujo foco é extrair informações essenciais contidas na pergunta de entrada do usuário e pesquisar por uma resposta adequada correspondente às informações de uma base de dados. O estilo da língua falada, em consultas de usuários, utiliza uma estrutura menos rígida, especialmente em línguas diferentes da língua inglesa, cuja estrutura é mais flexível. Logo, é sugerido um método que seleciona respostas para perguntas de usuários que permitem diferentes expressões para uma mesma pergunta. Para isso, é aplicada uma análise de dependência de frases de entrada para extrair suas partes principais. Esta extração consiste em dividir a frase da pergunta do usuário em três partes, onde cada parte será uma combinação de substantivo, sujeito e predicado. Assim, para obter respostas inteligentes para as perguntas, são introduzidos tópicos mapeados para associar as relações de palavras para respostas apropriadas. Para os testes são utilizados um conjunto de perguntas do *Yahoo! Answers*, no qual o texto contém diversas informações como contexto, premissas e conteúdo da questão, o que causa várias expressões e dificulta encontrar padrões de descrição para extrair a informação. Assim, o primeiro passo é assumir a pergunta de entrada para o *chatterbot* com uma sentença simples. Também, é importante considerar a intenção do usuário para que seja possível associar corretamente as frases de entrada com as palavras-chaves na mente dos usuários durante a consideração da sentença, pois se for ignorada a intenção, um *chatterbot* pode não prover uma resposta significativa. Além disso, as sentenças de entrada não necessariamente precisam de uma sentença de saída apropriada e, por isso, foram utilizadas frases preparadas com antecedência nos testes do trabalho. Ao dividir a frase da pergunta do usuário em três partes (substantivo,

sujeito e predicado) arranja-se uma associação de uma palavra (palavra modificada), onde seu significado é modificado por outra palavra (palavra modificando), o que permite definir o tipo da palavra. Por exemplo, se a palavra “modificada” é um verbo ou adjetivo, a palavra “modificando” será um substantivo ou um advérbio e, neste caso, o sujeito será nulo. Para simplificar, o trabalho limita-se em substantivo, sujeito, verbo/adjetivo. Os testes utilizaram uma base de conhecimento com 8.838 questões baixadas do *Yahoo! Questions* no *Yahoo! Japão*. Participaram 8 estudantes de graduação e 3 estudantes de pós-graduação para realizar 10 perguntas cada um. O método estima a validade da resposta resultante para a entrada de questões por assunto e relevância. A validade é definida pelo usuário que define se a saída é uma resposta apropriada para sua entrada. A relevância é definida pela quantidade de informações que a saída tem relacionada com a entrada. Depois de realizar as perguntas, os autores conduziram um questionário no qual os participantes são questionados sobre a validade e relevância de cada resposta para cada pergunta realizada para o *chatbot*. Ao todo, foram realizadas 100 frases de perguntas para o *chatbot*, onde um dos participantes foi excluído por fornecer respostas inconsistentes. Os resultados mostraram que o método utilizado retornou uma resposta para aproximadamente 88% das questões sendo que cerca de 40% das perguntas tiveram respostas válidas. Como trabalhos futuros, os pesquisadores desejam melhorar o método para prevenir a representação de respostas irrelevantes, além de introduzir o peso atribuído aos tipos de palavras e combinação da divisão em três partes das frases para tomar conta da sua influência sobre a pertinência das respostas.

2.5.3 Building a Hybrid: Chatterbot – Dialog System (DINGLI; SCERRI, 2013)

Em (DINGLI; SCERRI, 2013), foi proposto o desenvolvimento de um *framework* e um protótipo de um sistema capaz de analisar as entradas de texto de um usuário em um nível mais profundo de abstração, o que resultará em um sistema capaz de fornecer respostas mais adequadas, que por sua vez, resultam em uma melhor experiência para o usuário. O objetivo do trabalho é fornecer um protótipo que pode ser classificado como sendo um híbrido entre um *chatbot* e um sistema de diálogo, capaz de fornecer um entendimento sintático, semântico e pragmático da entrada do usuário. A principal diferença entre *chatbots* e um sistema de diálogo é que o primeiro busca simular uma conversa em seu sentido básico com a intenção de enganar o usuário de

que está se comunicando com uma entidade inteligente. Já os sistemas de diálogo tentam modelar o processo de diálogo real que incorpora também a tarefa de analisar e compreender as entradas do usuário, o que por sua vez, ajuda na geração de uma resposta dinâmica adequada. Para obter estes resultados, o *framework* faz uso de ferramentas e tecnologias de PLN para analisar as entradas de diálogos do usuário, usando informações de fontes externas para compreender melhor estas entradas e gerar respostas adequadas para o usuário. As fontes externas incluem conteúdo e informações sobre diferentes conceitos, conhecimento de senso comum, conhecimento sobre pessoas específicas e entidades do mundo. Assim, busca-se mesclar estas fontes externas para criar um único local que será a base de conhecimento do sistema, o que permitirá acompanhar a relação e interação do usuário com várias entidades. O sistema proposto é personalizável, flexível e modular, já que busca um esforço mínimo na atualização e adaptação do sistema para manipular as diversidades, complexidades e variedades de temas da entrada de dados dos usuários. O desenvolvimento do protótipo está dividido em três fases principais: o entendimento da linguagem natural; o processamento intermediário; e a geração da saída de dados. A implementação do protótipo inclui diversos módulos de ação, que são responsáveis por executar várias tarefas, que incluem a consulta e gestão das bases de conhecimento locais e externas, normalização da entrada de dados, recuperador de tópicos, gerenciador de caixa de diálogo e gerenciador de dados de saída. A fase de entendimento da linguagem natural consiste em utilizar o *engine ChatScript*, desenvolvido para *chatterbots* para mapear a entrada em linguagem natural em uma representação mais formal em XML. Na fase de processamento intermediário são realizadas consultas na base de conhecimento locais e externas para permitir melhor compreensão da entrada de dados, onde é utilizada uma série de módulos de processamentos que são desenvolvidos para lidar com tipos específicos de entrada. Os testes foram divididos em duas partes. Na primeira parte, cada usuário realizou uma conversação com o *chatterbot*. Na segunda parte do teste, cada usuário respondeu um questionário onde deviam classificar a interação com o sistema em uma escala de cinco pontos, a fim de concluir se o sistema realmente atingiu seus objetivos. Os resultados indicaram uma interação positiva, já que 65,27 % dos participantes classificaram com nota 4 ou cinco. Assim, os autores concluíram que o trabalho propôs um *framework* com potencial para dar suporte em implementação de agentes de conversação dinâmica. Como possíveis trabalhos futuros, podem ser realizadas

melhorias na integração de técnicas de gestão de diálogo mais refinadas que irá permitir que o sistema seja usado na prática. Também, podem ser implementados novos módulos para alcançar resultados de mais qualidade em termos de compreensão de entrada.

2.5.4 Quadro comparativo dos trabalhos relacionados estudados

O estudo individual de cada trabalho relacionado permite realizar uma comparação dos mesmos, a fim de avaliar e comparar os objetivos, métodos de PLN, técnicas de PLN, níveis da análise linguística e os resultados obtidos. Assim, com o objetivo de facilitar a comparação destes pontos, a tabela 2 apresenta o quadro comparativo dos trabalhos relacionados selecionados.

Tabela 2. Quadro comparativo entre os trabalhos relacionados.

Trabalhos Relacionados	Objetivos	Métodos de PLN	Técnicas de PLN	Níveis da análise linguística	Resultados Obtidos
(RAJ; ABDUL-KAREEM, 2011)	Desenvolver um algoritmo capaz de derivar verbos no particípio ou em diferentes tempos verbais em seus estados infinitivos	Usar gramáticas para compreender as sentenças / Combinar palavras para entender seu significado / Derivar a forma infinitiva do verbo	Usar condições "IF ELSE" para testar verbos em padrões estruturados / Adicionar novas condições em sua tabela de padrões estruturados / Representar os padrões em lógica de primeira ordem / Usar a relação sujeito, verbo e objeto / Algoritmo de predição	Lexical / Morfológico / Sintático	Qualquer verbo pode ser decomposto corretamente em seu estado infinitivo. Se um verbo não puder ser decomposto naquele momento, então é adicionada uma nova condição na tabela de padrões.
(KIMURA, 2015)	Extrair informações essenciais contidas na pergunta de entrada do usuário e pesquisar por uma resposta adequada correspondente às informações de uma base de dados	Dividir a frase da pergunta do usuário em três partes (substantivo, sujeito e predicado)	Tópicos mapeados para associar as relações de palavras para respostas apropriadas / Análise de dependência nas sentenças de entrada para extrair suas partes principais	Sintático / Semântico	O método utilizado retornou uma resposta para aproximadamente 88% das questões sendo que cerca de 40% das perguntas tiveram respostas válidas

(DINGLI; SCERRI, 2013)	Fornecer um protótipo capaz de prover um entendimento sintático, semântico e pragmático da entrada do usuário	Entendimento da Linguagem Natural / Processamento intermediário / Geração de resposta / Usar técnicas e ferramentas modernas de PLN / Unir diversas bases de informações em uma única base	<i>Engine ChatScript</i> para <i>chatterbots</i> / Mapear a entrada em linguagem natural em uma representação mais formal em XML / Mapear pronomes para entidades nomeadas / Verificação ortográfica	Morfológico / Sintático / Semântico / Pragmático	O trabalho propôs um <i>framework</i> com potencial para dar suporte em implementação de agentes de conversação dinâmica
------------------------	---	--	--	--	--

Fonte: dos autores (2016)

Com o estudo do trabalho (RAJ; ABDUL-KAREEM, 2011), é possível compreender a importância de derivar verbos para a forma de infinitivo, já que os mesmos podem estar em diferentes formas. Ao realizar o processamento da linguagem natural é preciso reconhecer que estas formas pertencem a um único verbo ao invés de analisá-las como palavras com significados distintos. Além disso, a derivação do verbo pode ser justificada pela facilidade de obter sinônimos e antônimos do verbo, como também, é possível obter as diferentes formas do verbo utilizando técnicas semelhantes à utilizada para derivá-lo.

O estudo do trabalho (KIMURA, 2015), permite entender como são extraídas as informações essenciais contidas nas perguntas de entradas dos usuários e como é criada uma resposta adequada correspondente. Além disso, é destacada a importância de analisar o significado de um conjunto de palavras ao invés de cada uma delas separadamente, já que uma palavra pode ter seu significado alterado ao ser utilizada com uma segunda palavra, característica comum em línguas não inglesas.

Ao estudar o trabalho de (DINGLI; SCERRI, 2013), são analisadas as técnicas e métodos de processamento de significados de uma sentença através da análise semântica e pragmática. Assim, o desenvolvimento de um *chatterbot* capaz de analisar as entradas de texto em um nível menor de abstração, resulta em um sistema com capacidade de fornecer respostas mais adequadas às perguntas do usuário o que torna a sua experiência mais agradável.

É possível analisar características distintas no processamento de diferentes línguas, por exemplo, na língua portuguesa, é comum encontrar substantivos seguidos de adjetivos enquanto que, na língua inglesa, é comum encontrar os mesmos na ordem

inversa. Esta é apenas uma das particularidades que compõe uma língua, onde estas particularidades irão servir de vantagem ou desvantagem no PLN.

Ao analisar os trabalhos relacionados como um todo é possível notar a dependência dos níveis da análise linguística, onde os níveis mais altos, como a análise semântica e pragmática, irão depender dos resultados dos níveis mais baixos, como a análise morfológica e lexical. Logo, para um *chatbot* receber uma sentença de entrada do usuário em forma de texto e processar esta sentença a fim de compreendê-la, aprendê-la e retornar uma resposta cabível à pergunta do usuário, tornando a interação mais natural, é necessário que o agente tenha certo domínio nas análises lexical, morfológica, sintática, semântica e pragmática.

3 METODOLOGIA

A pesquisa se caracteriza como descritiva experimental (CERVO, BERVIAN e SILVA, 2007), pois busca descrever e identificar resultados obtidos na utilização de métodos e técnicas de PLN para a interação entre o usuário e o agente pedagógico. Um estudo de caso experimental (YIN, 2015) foi realizado junto ao ambiente virtual de aprendizagem, já desenvolvido, com o agente pedagógico Dóris sendo utilizado para análise dos resultados referentes ao uso de linguagem natural no processo de interação com usuários.

A pesquisa é bibliográfica e consiste em análises e estudos de trabalhos relacionados ao assunto proposto, ou desenvolvidos na área. É uma pesquisa qualitativa, pois envolve desenvolver e utilizar métodos e técnicas de PLN, bem como, analisar seus resultados com a intenção de obter melhor qualidade no processo de interação dos usuários com o agente Dóris do ambiente virtual de aprendizagem.

Portanto, esta pesquisa possui três partes fundamentais que podem ser entendidas como: o estudo de referenciais teóricos e levantamento de trabalhos relacionados para sustentar os argumentos considerados neste trabalho; desenvolvimento e estudo de técnicas de PLN, que consiste no estudo e aplicação das técnicas; estudo de caso experimental com o agente Dóris, a fim de analisar os resultados no processo de interação com o usuário. Estas três partes fundamentais serão aprofundadas nas próximas seções.

3.1 Estudo de referenciais teóricos e levantamento de trabalhos relacionados

Com o intuito de construir conhecimento nos principais temas deste trabalho e sustentar argumentos descritos, foi realizado um estudo utilizando autores dos domínios de cada assunto. Assim, para o estudo do ambiente virtual de aprendizagem já desenvolvido foram utilizados como referência os trabalhos de (SANTOS *et al.*, 2001) até (SCHAEFER *et al.*, 2015). O estudo de agentes utiliza como principais referências autores como Wooldridge, Russell e Norving, além de autores específicos da área de *chatbots*, como Lester, Branting, Mott e Mikic. Para o estudo de PLN são utilizados como referências autores como Chowdhury e Liddy.

Com a finalidade de medir a difusão do conhecimento científico e o fluxo da informação sob enfoques diversos, a bibliometria é um conjunto de métodos de pesquisa da área das Ciências da Informação que permite mapear a estrutura de um campo científico (VANTI, 2002). Assim, realizou-se neste trabalho uma bibliometria quantitativa utilizando como termos de pesquisa “*Chatterbot*”, “*Natural Language*” e “*Human-computer interaction*”.

A tabela 3 apresenta o quadro da bibliometria quantitativa dos temas *chatterbots*, linguagem natural e interação homem-computador. As bases de dados utilizadas foram a *Scopus* e *Web of Science*. Foram buscados trabalhos no período de 2011 a 2016 nas áreas de computação, engenharias e educação. Nenhum trabalho foi retornado ao realizar a busca com os três termos de pesquisa juntos para ambas as bases de dados.

Tabela 3. Quadro de resultado da bibliometria quantitativa

<i>Scopus e Web of Science</i> – Critério de pesquisa: “Linha” e “Coluna”	<i>Chatterbot</i>		<i>Natural Language</i>		<i>Human-computer interaction</i>	
	<i>Scopus</i>	<i>Web of Science</i>	<i>Scopus</i>	<i>Web of Science</i>	<i>Scopus</i>	<i>Web of Science</i>
<i>Chatterbot</i>	30	15	6	4	4	0
<i>Natural Language</i>			16252	4896	369	72
<i>Human-computer interaction</i>					22491	2895

Fonte: dos autores (2016)

Para os trabalhos relacionados, foram selecionados dos resultantes na bibliometria, os que mais se assemelharam com os objetivos deste. Assim, foi selecionado o artigo de (RAJ; ABDUL-KAREEM, 2011), que busca desenvolver um algoritmo capaz de derivar verbos no particípio ou em diferentes tempos verbais em seus estados infinitivos. O artigo de (KIMURA, 2015), que procura extrair informações essenciais contidas na pergunta de entrada do usuário e pesquisar por uma resposta adequada correspondente às informações de uma base de dados. Finalmente, o artigo de (DINGLI; SCERRI, 2013), que busca fornecer um protótipo capaz de prover um entendimento sintático, semântico e pragmático da entrada do usuário. Depois de selecionar e estudar cada trabalho, foi realizada a sua comparação com o objetivo de selecionar os métodos e técnicas de PLN utilizadas neste trabalho.

3.2 Desenvolvimento e estudo de métodos e técnicas de PLN

Como trata-se de uma pesquisa aplicada, a solução dos problemas específicos ocorreu por meio de aplicação prática, que tem como objetivo proporcionar uma interação mais natural entre os usuários de um ambiente virtual de aprendizagem e um agente pedagógico. Para isso, buscou-se através de métodos e técnicas de PLN acrescentar a característica de *chatbot* ao agente pedagógico Dóris.

As técnicas de PLN fazem uso de um ou mais níveis da análise linguística, com a finalidade de alcançar o processamento da linguagem humana. Já que a interação entre o agente e os usuários deste trabalho é feita através de textos em linguagem natural, é preciso considerar os níveis léxico-morfológico, sintático, semântico e pragmático da linguística, para que o agente tenha a capacidade de responder de forma convincente a uma entrada de texto feita pelo usuário.

Uma das etapas de desenvolvimento é a análise léxica-morfológica, que tem como principais objetivos analisar a estrutura, formação e classificação das palavras. O léxico define os conjuntos de palavras de uma língua, enquanto que os morfemas serão responsáveis por informar o gênero, tempo e modo verbal, o número (singular ou plural) e a classe gramatical de uma palavra.

A linguagem natural tem um dicionário de palavras extremamente grande, se for levado em conta todas as derivações que uma mesma palavra pode assumir. Por exemplo, o verbo “Comprar” pode estar derivado em diferentes formas, como “Comprei”, “Comprou”, “Compravam”, entre outras. Para que o dicionário de palavras de um sistema computacional não fique extremamente grande, busca-se derivar as diferentes formas que uma palavra pode estar em uma forma padrão. Assim, nesta etapa de desenvolvimento, são implementadas técnicas que derivem palavras em diferentes formas para uma forma padrão, além de derivar a palavra na forma padrão em suas diferentes formas.

As palavras sozinhas não constroem significados em uma sentença, assim, em uma linguagem natural a relação entre palavras em uma frase é o que constrói o seu significado. Para que um sistema computacional consiga compreender o significado de

uma frase, é preciso que ele analise a relação das palavras. Na análise sintática, numa das etapas do desenvolvimento procura-se identificar o significado de uma palavra em uma sentença, além da sua relação com as outras palavras, ou seja, pretende-se descobrir a estrutura gramatical da oração. Nesta etapa buscam-se técnicas para identificar sujeitos, verbos e objetos de uma oração e suas variantes, permitindo, assim, que o agente, ao receber uma sentença de entrada, identifique, por exemplo, que o usuário deseja saber o ano que Tiradentes morreu na frase de entrada “Em que ano Tiradentes morreu? ”.

O contexto em uma sentença, muitas vezes, determina o seu significado, já que uma frase pode assumir diferentes significados dependendo do seu contexto. A análise semântica tem o objetivo de identificar os possíveis significados de uma frase e determinar o significado mais adequado com base no contexto. Esta etapa do desenvolvimento acontece durante a geração da árvore de *parse* na análise sintática e semântica.

Na etapa de análise pragmática, busca-se o entendimento do contexto externo que falta no diagnóstico de uma única oração nas demais sentenças. A importância desta etapa está no fato de que em uma conversação com o agente, o usuário pode enviar uma sentença onde o agente deve, através de conhecimento externo, enviar a resposta que o usuário espera. Por exemplo, quando o usuário pergunta ao agente “Você tem horas? ”, se o agente não tiver o conhecimento externo necessário ele irá responder “Sim, tenho horas”, ao invés de fornecer a hora atual ao usuário.

O desenvolvimento de cada etapa é importante, já que uma depende, respectivamente, da outra para que o agente consiga um melhor entendimento da sentença de entrada e, assim, apresentar uma resposta com sentido de coesão e coerência ao usuário.

3.3 Estudo de Caso com o agente Dóris

Para validar a proposta apresentada neste trabalho, um estudo de caso experimental foi realizado com o agente Dóris, a fim de analisar os resultados referentes ao uso de linguagem natural no processo de interação com o usuário. Os autores do trabalho realizaram os testes formulando sentenças conforme o contexto do cenário de estudo, o ambiente virtual de aprendizagem.

Dóris é um agente pedagógico tutor inserido no ambiente virtual de aprendizagem desenvolvido pelo grupo de pesquisa de professores e estudantes do Departamento de Computação e do Programa de Pós-Graduação em Sistemas e Processos Industriais da UNISC. Este agente tem o objetivo de permitir uma interação mais agradável do estudante com o processo de ensino-aprendizagem, para isso, ele auxilia na modelagem do perfil do estudante e guia e monitora o estudante dentro do ambiente (SANTOS *et al.*, 2001).

A interação entre usuário e agente acontece por meio de uma interface de texto semelhante ao bate-papo virtual, onde o usuário fica livre para formular perguntas à Dóris, como também responde a perguntas realizadas pela agente. É importante que a agente Dóris formule respostas coerentes às perguntas realizadas para o usuário, como também realize perguntas coesas ao momento de estudo e, assim, proporcionar uma interação mais natural entre usuário e agente.

A seção de metodologia apresentou os métodos para alcançar os objetivos deste trabalho. O próximo capítulo apresenta o desenvolvimento do *chatbot* Dóris.

4 CHATTERBOT DÓRIS

Este capítulo apresenta os métodos e as técnicas pesquisadas para acrescentar a característica de *chatterbot* ao agente Dóris e como eles foram implementados, o estudo de caso realizado e os resultados obtidos e, por fim, a análise dos resultados com base no comportamento em cada técnica e método pesquisado e implementado.

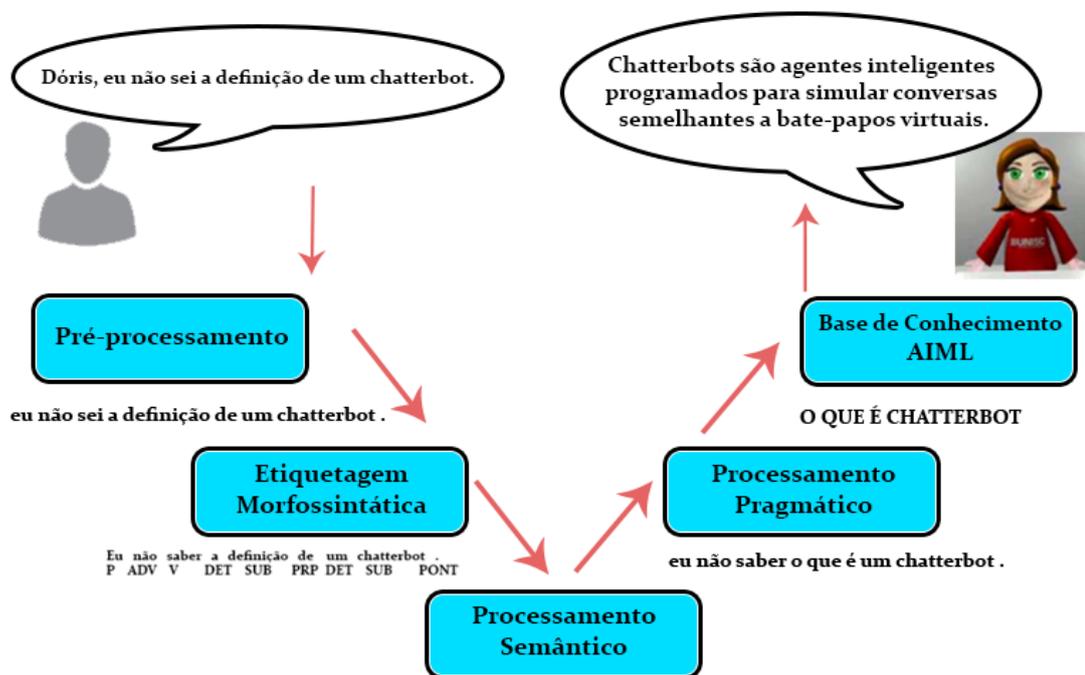
Ao analisar a arquitetura do agente Dóris, a linguagem de programação Java foi escolhida para o desenvolvimento dos métodos e técnicas de PLN, já que a linguagem foi utilizada para o desenvolvimento dos módulos do agente. A linguagem de programação e plataforma computacional Java foi lançada pela Sun Microsystems em 1995. Muitos *sites* e aplicações funcionam com Java, já que a linguagem proporciona velocidade, segurança e confiabilidade para as aplicações.

4.1 Métodos e Técnicas de PLN

Os métodos e técnicas pesquisados e utilizados, juntamente com os desenvolvidos, foram selecionados com base nos objetivos deste trabalho, principalmente pela necessidade de padronizar a sentença de entrada enviada pelo usuário ao agente, de forma que a sentença corresponda ao mesmo padrão associado a ela na base de conhecimento, possibilitando, assim, associar diferentes sentenças com o mesmo significado, mas estruturadas de forma diferente a uma mesma resposta. Logo, foram selecionados métodos e técnicas que representam etapas de processamento de uma sentença, modificando-a até o momento que ela possa ser associada a um padrão na base de conhecimento do agente *chatterbot*.

Uma sentença que o usuário envia ao agente *chatterbot* passa por diversas etapas com papéis específicos. Estas etapas podem ser resumidas e ordenadas da seguinte maneira, conforme a Figura 3: A sentença original do usuário passa primeiramente por um pré-processamento, a fim de corrigir características que possam causar erros nas próximas etapas; Em seguida, ela passa pelo etiquetador morfossintático, que classifica cada palavra em *tokens*, o que possibilita classificar a classe gramatical e padronizar para um mesmo padrão cada palavra; A próxima etapa é o processamento semântico, que tem o objetivo de traduzir termos da sentença para o mesmo padrão da base de conhecimento respeitando seu significado semântico; A etapa seguinte é o processamento pragmático, que traduz a sentença para o sentido que o usuário refere-se, ou seja, nesta etapa o agente deve ter um conhecimento externo, além da própria sentença. Por fim, a sentença processada é comparada a padrões em uma base de conhecimento, e quando este padrão é encontrado, a resposta correspondente a ele é enviada ao usuário. As próximas subseções detalham cada uma destas etapas.

Figura 3. Etapas do processamento da sentença.



Fonte: dos autores (2016)

4.1.1 Pré-processamento da sentença do usuário

Ao considerar que a sentença enviada ao agente é formulada por usuários, deve-se assumir que a mesma pode conter diferentes tipos de erros, como de ortografia ou gramática. Além disso, mesmo que a sentença enviada pelo usuário esteja correta, ela pode conter características que dificultam o processamento das etapas seguintes, como a falta de espaço entre pontuação, além de termos irrelevantes da sentença. Essas características podem ocasionar erros nas etapas seguintes, por isso é importante tratá-los antes das próximas etapas de processamento.

Quando uma pontuação, como uma vírgula ou um ponto de interrogação está junto com uma palavra, o etiquetador morfosintático irá considerar a pontuação como parte da palavra, o que acaba classificando-a erroneamente. Para este problema, implementa-se uma função em Java que recebe uma sentença de entrada e retorna a mesma sentença, mas com espaço entre as pontuações e as palavras, como mostra a Figura 4 com o código fonte da função implementada.

Figura 4. Função para adicionar espaço entre pontuações.

```
public static String addEspacoCaracteresPontuacao(String sentenca) {
    for (int i = 0; i < sentenca.length(); i++) {
        if (isCaracterePontuacao(sentenca.charAt(i))) {
            if (i > 0) {
                if (sentenca.charAt(i - 1) != ' ') {
                    sentenca = sentenca.substring(0, i)
                        + " " + sentenca.substring(i);
                    i++;
                }
            } else {
                sentenca = " " + sentenca;
                i++;
            }
        }
        if (i < sentenca.length() - 1) {
            if (sentenca.charAt(i + 1) != ' ') {
                sentenca = sentenca.substring(0, i + 1)
                    + " " + sentenca.substring(i + 1);
                i++;
            }
        } else {
            sentenca = sentenca + " ";
        }
    }
    return sentenca;
}
```

Fonte: dos autores (2016)

Outra característica que se busca tratar nesta etapa são termos irrelevantes da sentença do usuário, como por exemplo o termo “Dóris”, na frase “Dóris, o que é *chatterbot*?”. Esta característica dificulta o processamento das etapas seguintes, pois as mesmas podem considerar o termo como significante na sentença. Para este problema, utiliza-se um arquivo onde cada linha representa um termo que deve ser desconsiderado na sentença caso ela o contenha. Assim, sempre que um termo da sentença está contido neste arquivo, o mesmo é removido da sentença.

Um tratamento de erros ortográficos e gramaticais de forma automatizada, sem a interferência do usuário, pode forçar uma modificação na sentença mesmo que o usuário não esteja de acordo. Sendo assim, aconselha-se que o próprio usuário visualize e corrija estes erros antes de enviar a sentença ao agente *chatterbot*. Este mecanismo de visualizar e corrigir erros ortográficos e gramaticais pelo usuário pode estar contido na própria interface de bate-papo.

Após o pré-processamento, a sentença do usuário estará pronta para o etiquetador morfossintático. Esta próxima etapa é responsável por classificar a classe gramatical de cada palavra, além de padronizar a mesma para o mesmo formato da base de conhecimento.

4.1.2 Etiketador Morfossintático

Para determinar quais são os termos importantes da sentença, é importante classificar cada palavra conforme sua classe gramatical. Isso permite, por exemplo, descobrir quais são os verbos da frase e como os mesmos estão conjugados. Este tipo de classificação é alcançado através de um etiquetador morfossintático.

O etiquetador morfossintático tem a função de classificar as palavras da sentença em *tokens*. Um *token*, além de armazenar o formato original da palavra, armazena o *lemma* e a *tag*. O *lemma* é a forma padrão da palavra, já que verbos conjugados são passados para o infinitivo, palavras no plural são passadas para o singular e o gênero delas é passado para o masculino. Por fim, a *tag* determina a classe gramatical da palavra.

Para este trabalho, foram pesquisados etiquetadores morfossintáticos para classificar a sentença enviada do usuário para o agente *chatterbot* no ambiente virtual de aprendizagem. Assim, os critérios de seleção para o etiquetador foram que o mesmo classificasse sentenças em português do Brasil, fosse gratuito e disponibilizasse uma API de implementação Java. Desta forma, foi selecionado o analisador morfossintático *Tree-Tagger* (SCHMID, 1995), já

que o mesmo oferece uma API em Java, tem licença livre para pesquisa e classifica sentenças no Português Brasileiro.

O analisador morfossintático *Tree-Tagger* é implementado no projeto através da sua API Java, o que permite abstrair a implementação de um analisador morfossintático. A configuração do etiquetador consiste em definir o arquivo de modelo de linguagem em português. Após a configuração, o analisador será capaz de classificar a classe gramatical de palavras de uma sentença em português. A Figura 5 demonstra a etiquetagem de uma frase, classificando as palavras em *tokens* que contém o *lemma* e a *tag* da mesma.

Figura 5. Exemplo de etiquetagem de uma sentença.

O	que	é	chatterbot
Tag: D	Tag: C	Tag: V	Tag: N
Lemma: o	Lemma: que	Lemma: ser	Lemma: chatterbot

Fonte: dos autores (2016)

Uma palavra de uma sentença, através do *Tree-Tagger*, pode ser classificada nas seguintes classes gramaticas:

- *Tag A*: Adjetivo. Pode ser dividido em tipo Qualificativo ou Ordinal
- *Tag R*: Advérbio. Pode ser dividido em tipo Geral ou Negação
- *Tag D*: Determinante. Pode ser dividido em tipo Demonstrativo, Possessivo, Interrogativo, Exclamativo, Indefinido ou Artigo.
- *Tag N*: Substantivo. Pode ser dividido em tipo Comum ou Próprio. Também pode ser classificado por gênero Masculino, Feminino ou Comum. Por fim, pode ser classificado em Singular, Plural ou Invariável.
- *Tag V*: Verbo. Pode ser dividido em Modo Indicativo, Subjuntivo, Imperativo, Infinitivo, Gerúndio ou Particípio.
- *Tag P*: Pronome. Pode ser dividido em tipo Pessoal, Demonstrativo, Possessivo, Indefinido, Interrogativo, Relativo ou Exclamativo.
- *Tag C*: Conjunção. Pode ser dividido em tipo Coordenada ou Subordinada.
- *Tag I*: Interjeição.
- *Tag S*: Preposição. Pode ser dividido na forma Simples ou Contraída.
- *Tag F*: Pontuação.

- *Tag Z*: Cifras e numerais.

Após a sentença ser pré-processada e analisada morfossintaticamente para classificar suas palavras em *tokens*, a próxima etapa é o processamento semântico que tem o objetivo traduzir termos da sentença para o mesmo padrão da base de conhecimento, respeitando seu significado semântico.

4.1.3 Processamento Semântico

Duas sentenças, mesmo que com termos diferentes podem ter o mesmo significado semântico, por exemplo, as sentenças “Devo ajudar as pessoas que não enxergam?” e “ Devo ajudar os cegos? ” contém o mesmo significado. Assim, para que uma sentença enviada pelo usuário ao agente *chatbot* possa ser comparada aos mesmos padrões na base de conhecimento do agente, é preciso traduzir termos da sentença para o mesmo padrão da base acatando seu significado.

Neste trabalho, busca-se resolver este problema através de um dicionário de sinônimos, em que termos da sentença do usuário são comparados a uma lista de termos com o mesmo significado em um arquivo de texto e, ao encontrar seu semelhante na lista, o termo é substituído pelo mesmo padrão contido na base de conhecimento. O arquivo é estruturado da seguinte maneira: entre colchetes é definido o tipo dos termos (Conjunto, Verbo, Substantivo, Adjetivo) e entre chaves estão os termos com o mesmo significado semântico. A Figura 6 que demonstra como é estruturado o dicionário de sinônimos. Assim, quando um termo da sentença do usuário é encontrado em um dos conjuntos da lista de sinônimos, o termo é traduzido para o primeiro termo do conjunto entre chaves.

Figura 6. Dicionário de sinônimos.

```
[Conjunto] {cego, não enxergar}
[Verbo] {ser, definir}
[Verbo] {adequar, apropriar, ajustar, acondicionar, adaptar, aplicar, apropositar, justar, correta, regular}
[Substantivo] {cinto, cinta}
[Substantivo] {encosto, assento, almofada, escora}
[Conjunto] {peito, peito de motorista, torso, torso de motorista, tórax, tórax de motorista}
[Substantivo] {motivo, causa, razão, explicação, fonte}
[Substantivo] {veículo, carro, moto, caminhão}
[Verbo] {usar, utilizar, empregar, utilização, uso}
[Substantivo] {fatalidade, acidente}
[Substantivo] {troca, mudança}
[Conjunto] {motivo, ocasionar}
```

Fonte: dos autores (2016)

Os termos no dicionário de sinônimos estão na sua forma simplificada, ou seja, se o termo passasse pelo etiquetador morfossintático, a sua forma simplificada seria os *lemmas* de

seus *tokens*. Também os termos da sentença do usuário estarão simplificados, pois o mesmo acabara de passar pelo etiquetador morfossintático. Assim, é permitido comparar palavras da sentença do usuário em diferentes conjugações com a lista de sinônimos do arquivo.

Após traduzir termos da sentença do usuário para o mesmo padrão da base de conhecimento, realiza-se o processamento pragmático da sentença, em que busca definir o sentido real da sentença, ou seja, o conhecimento além da sentença do usuário.

4.1.4 Processamento Pragmático

Muitas vezes o usuário envia uma sentença esperando uma resposta diferente do sentido literal que a sentença original propõe, logo o *chatterbot* deve estar preparado para tratar este tipo de situação. Por exemplo, quando o usuário envia a pergunta ao agente “Você tem horas? ”, se o agente *chatterbot* não tem o conhecimento externo e interpreta apenas a sentença enviada, ele irá retornar a resposta “Sim, eu tenho horas! ”, porém o usuário espera que o agente forneça a hora atual. Logo, o processamento pragmático busca conhecimento além da sentença do usuário para retornar uma resposta com coerência e coesão ao usuário. Para tratar esta característica, são utilizados mecanismos do AIML.

O AIML (*Artificial Intelligence Markup Language*), que segundo (A.L.I.C.E, 2016), é uma linguagem desenvolvida para criar diálogos em linguagem natural por meio de software. Assim, a base de conhecimento do *chatterbot* deste trabalho é um arquivo em AIML, onde o conteúdo dentro da *tag* <pattern> representa o padrão da sentença de entrada. Quando um padrão é associado, sua resposta é buscada dentro da *tag* <template> e retornada ao usuário.

As *tags* <pattern> e <template> ficam dentro da *tag* <category>, que identificam uma unidade de conhecimento do arquivo AIML. Outra característica importante do AIML é a utilização de coringas, como o “_” (caracteres *underline*) e o “*” (caractere asterisco). Esses caracteres são utilizados juntos com o padrão de sentença dentro da *tag* <pattern>, o que permite maior flexibilidade do padrão da sentença, já que representa uma ou mais palavras. Por exemplo, na sentença “O que é*”, é aceitável qualquer sentença que comece com “O que é”. Além das *tags* *pattern* e *template* o AIML permite utilizar as *tags* <srail> e <star>. A *tag* <srail>, é utilizada dentro das *tags* <template> e permite, de forma recursiva, direcionar a resposta para outro padrão de sentença. Já a *tag* <star>, é utilizada dentro das *tags* <template> representa o conteúdo dos coringas “*” e “_” nas *tags* <pattern>.

As características do processamento pragmático são tratadas utilizando padrões que direcionam o conteúdo da sentença para outros padrões. Para isso, parte da sentença é direcionada para outra *tag* <pattern> através das *tag* <srai> e <start>. Por exemplo, o usuário envia a sentença “Eu não entendo o que é *chatterbot*”. Nesta sentença, o usuário espera que o *chatterbot* responda o que é um *chatterbot*, então a sentença é reconhecida pelo padrão “<pattern>E não entendo o que é */</pattern>” e direcionada para o padrão “<pattern>O que é *chatterbot*</pattern>” e retorna a resposta desejada ao usuário, conforme a Figura 7 com o código AIML do exemplo.

Figura 7. Código AIML para processamento pragmático.

```
<category>
  <pattern>Não entendo o que é */</pattern>
  <template><srai>O que é <star/></srai></template>
</category>

<category>
  <pattern>O que é chatterbot</pattern>
  <template>Chatterbots são agentes inteligentes
  programado para simular conversas semelhantes a
  bate-papos virtuais.</template>
</category>
```

Fonte: dos autores (2016)

Através do AIML, busca-se tratar problemas contidos na análise pragmática, direcionando a sentença do usuário para o padrão de pergunta que o usuário pretende obter a resposta. Após o processamento pragmático, a sentença processada é comparada com a base de conhecimento.

4.1.5 Consulta na Base de Conhecimento

Após a sentença do usuário passar pelas etapas de pré-processamento, etiquetagem morfossintática, processamento semântico e, por fim, processamento pragmático, a sentença é comparada com os padrões da base de conhecimento em AIML.

Como a base de conhecimento do *chatterbot* Dóris está em AIML, é necessário utilizar um interpretador que associe a sentença do usuário com a base de conhecimento em AIML. Para este trabalho, é utilizado o interpretador *Program-AB* (PROGRAM AB, 2016), que tem licença livre e é desenvolvido em Java, o que permite anexar ao código do projeto do sistema virtual de aprendizagem já desenvolvido.

A comparação da sentença do usuário processada com a base de conhecimento consiste em remover palavras da sentença conforme a importância de sua classe gramatical até que a sentença tenha um padrão de associação na base de conhecimento. A remoção de palavras é feita devido à falta de importância da classe gramatical das mesmas na sentença. Como por exemplo, preposições e artigos têm menos importância do que verbos e substantivos.

Para a técnica de remoção de palavras, desenvolve-se uma função Java que as remove em ordem conforme uma lista de filtro de classes gramaticais. Como mostra a Figura 8, sempre ao remover as palavras de um tipo de classe gramatical, verifica-se se a sentença tem um padrão associado na base de conhecimento, caso encontre, retorna-se a resposta deste padrão ao usuário; caso não encontre, são removidas as palavras conforme o próximo filtro de classes gramaticais da lista de filtros e testado novamente. Se após percorrer todo o filtro não tiver um padrão de associação à sentença, a função retorna que não encontrou uma resposta. O filtro de classes gramaticais foi construído conforme a importância de cada classe, de forma que ele inicie com as classes gramaticais menos importantes e termine com as mais importantes.

Figura 8. Função para comparar sentença com a base AIML.

```
private String getRespostaPorRemocaoPalavra(List<TokenSTI> tokens, String[] filtros) {
    String ret;

    List<TokenSTI> listAux = new ArrayList<TokenSTI>(tokens);

    for (String filtro : filtros) {
        AuxMorSin.removerTipo(listAux, filtro);
        ret = chatSession.multisentenceRespond(tokensToSentencaAIML(listAux));

        if (!ret.trim().equals(SEM_RESPOSTA)) {
            return ret;
        }
    }
    return SEM_RESPOSTA;
}
```

Fonte: dos autores (2016)

Mesmo após a remoção de palavras de classes gramaticais menos importantes, pode haver casos onde a sentença apresenta palavra de classes importantes, mas com significados menos significativos em seu contexto. Além disso, observa-se que em grande parte destes casos, o último verbo da sentença e os substantivos seguintes a ele, tem maior significância na sentença. Por exemplo, se o usuário envia a sentença “O que pode acontecer se eu dirigir em alta velocidade? ”, neste caso, ele irá considerar o verbo “dirigir” como mais significante.

Sendo assim, caso não encontre padrão que corresponda à sentença, como último recurso, é utilizada a técnica criada para selecionar apenas o último verbo e os substantivos seguintes a ele na sentença, para então fazer a comparação com a base de conhecimento. Caso a sentença encontre um padrão associado na base de conhecimento AIML, retorna-se a resposta deste padrão ao usuário. Caso não encontre, o *chatterbot* irá retornar ao usuário que não entendeu a sentença enviada a ele.

Figura 9. Etapas do processamento da sentença em código fonte.

```
public String interacaoTexto(String sentencaAluno) {
    configurar();
    /**
     * Pré-processamentos na sentença de entrada do aluno.
     */
    sentencaAluno = PreProcessamento.addEspacoCaracteresPontuacao(sentencaAluno);
    sentencaAluno = PreProcessamento.removerPalavrasConfiguradas(sentencaAluno);

    List<TokenSTI> tokens;
    //Enquanto houver alteração da sentença pelo processamento semântico.
    Boolean alteracao;
    do {
        /**
         * Analisador morfossintático irá analisar e retornar em uma lista
         * de tokens.
         */
        tokens = this.anlMorfsint.analisarSentenca(sentencaAluno);

        /**
         * Processamento semântico e pragmático a partir do retorno da
         * análise morfossintática. O resultado é aplicado no lemma dos
         * tokens.
         */
        alteracao = processamentoSemantico(tokens, 4);
        if (alteracao) {
            sentencaAluno = tokensToSentencaOriginal(tokens);
        }
    } while (alteracao);

    /**
     * Filtro de classe gramaticais que serão removidas até encontrar uma
     * resposta.
     */
    String[] filtros = ORDEM_PADRAO_FILTRO;
    String retorno = getRespostaPorRemocaoPalavra(tokens, filtros);

    /**
     * Caso não encontre nenhuma resposta, é pego o último verbo e
     * seus substantivos.
     */
    if (retorno.equals(SEM_RESPOSTA)) {
        retorno = chatSession.multisentenceRespond(anlMorfsint.getVerboSubstantivos(tokens));
    }

    return retorno;
}
```

Fonte: dos autores (2016)

Com os métodos e técnicas utilizados e implementados, é possível que: a sentença enviada ao *chatterbot* Dóris seja pré-processada, o que evita problemas nas etapas seguintes; classifique cada palavra através de um etiquetador morfossintático; processe a sentença

semanticamente a fim de traduzir seus termos para o mesmo padrão da base de conhecimento, respeitando seus significados; através do processamento pragmático, associe a sentença do usuário para o padrão na base de conhecimento que o usuário refere-se; e por fim, compara a sentença processada a um padrão na base de conhecimento, retornando sua resposta. A Figura 9 mostra estes passos no código-fonte implementado. Após a implementação e uso dos métodos e técnicas de PLN, realiza-se um estudo de caso com o *chatbot* Dóris, descrito na próxima seção.

4.2 Estudo de Caso com o *chatbot* Dóris

Com o intuito de avaliar os resultados dos métodos e técnicas de PLN utilizados e desenvolvidos, realiza-se um estudo de caso experimental com o *chatbot* Dóris. Segundo (SCHRAMM, 1971, p. 5), “a essência de um estudo de caso é tentar esclarecer uma decisão ou um conjunto de decisões: o motivo pelo qual foram tomadas, como foram implementadas e com quais resultados”. Sendo assim, este estudo de caso busca explicar como e porque métodos e técnicas de PLN podem contribuir no processo de interação de um usuário com um agente *chatbot* em um ambiente virtual de aprendizagem. Também, como cada método e técnica se comporta com diferentes sentenças de entrada e por quê.

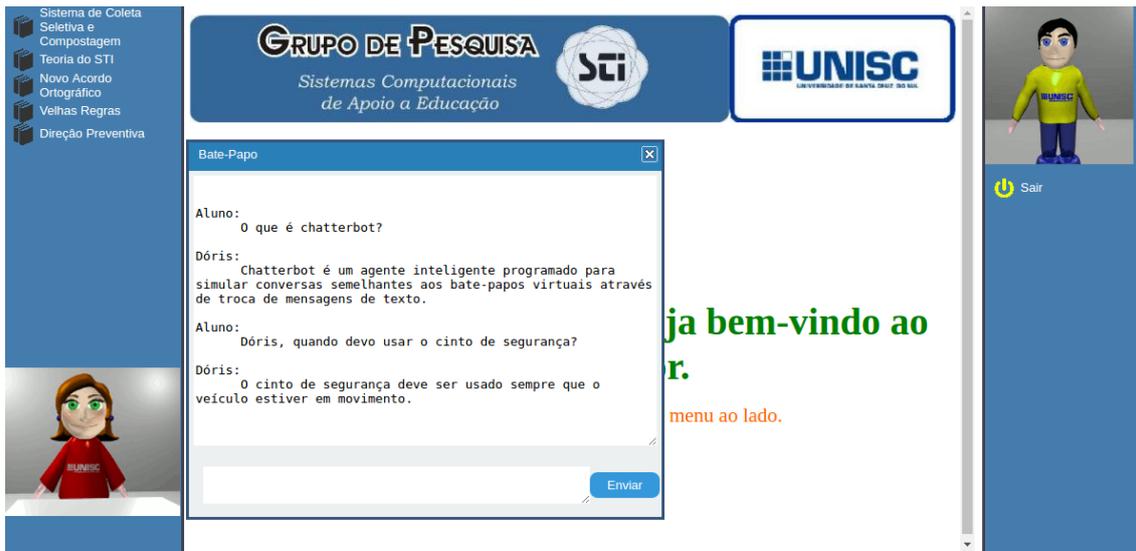
O objetivo não é verificar se a base conhecimento do agente *chatbot* está completa, mas sim, se ele consegue associar uma ou mais sentenças de entrada com o mesmo significado, a um padrão pré-estabelecido na base de conhecimento. Assume-se assim que as sentenças de entrada do usuário e base de conhecimento AIML estão corretas, cabendo aos métodos e técnicas de PLN fazerem a associação delas corretamente.

Os dados relevantes para o estudo são os comportamentos das etapas de processamento da sentença do usuário. Logo, os dados coletados são as constatações do comportamento dos métodos e técnicas de PLN quando uma sentença é enviada ao *chatbot*. Mais especificamente, o comportamento representa se o método ou técnica está processando a sentença corretamente. Sendo assim, contabiliza-se a quantidade de acertos e erros de cada etapa para um conjunto de sentenças de entradas enviadas ao *chatbot* Dóris.

Como o *chatbot* Dóris está inserido em um ambiente virtual de aprendizagem que apresenta páginas de conteúdo de estudo, para o estudo de caso são simuladas perguntas de dúvidas de usuários a respeito do conteúdo da aula em que o estudante está no ambiente. Estas perguntas estão relacionadas ao tema da aula, para o estudo de caso foi escolhido o tema

“Direção Preventiva”, que está previamente desenvolvido no ambiente, além de estar atualizado até o momento do estudo de caso. A Figura 10 demonstra o processo de interação com o *chat*bot Dóris no ambiente virtual de aprendizagem, onde o usuário realiza duas perguntas e o agente retorna as repostas.

Figura 10. Interação com o *chat*bot Dóris no ambiente virtual de aprendizagem.



Fonte: dos autores (2016)

A base de conhecimento do arquivo AIML do *chat*bot Dóris foi desenvolvida com base no tema de aula. Uma mesma resposta pode ser associada corretamente a diferentes perguntas, pois uma pergunta pode ser formulada de diversas maneiras. Assim, existe na base de conhecimento uma resposta associada a cada pergunta criada para a simulação de conversação com o *chat*bot no estudo.

O cenário do estudo de caso se passa após o estudante do ambiente virtual visualizar o material da aula, pois neste momento poderá haver dúvidas que o usuário gostaria de sanar antes de responder as questões de exercício. A Figura 10 demonstra a interação com o *chat*bot Dóris durante uma aula “Direção Preventiva”. Assim, as perguntas elaboradas ao *chat*bot são baseadas nas questões do exercício da aula. Também, elabora-se diferentes perguntas para uma mesma dúvida a fim de verificar se o *chat*bot é capaz de associar estas perguntas a mesma resposta. Para a variação de pergunta, as sentenças são estruturadas de diversas formas, utilizando diferentes termos–com o mesmo significado, porém variando a complexidade da sentença.

Para o estudo de caso, foram criadas 40 perguntas para 10 exercícios da aula no ambiente virtual de aprendizagem, ou seja, para cada exercício, foram variadas 4 perguntas. Para a base de conhecimento AIML, utiliza-se 10 padrões de associação, onde cada padrão corresponde a uma dúvida de cada exercício. Além dos padrões das dúvidas, são utilizados mais 7 padrões na base de conhecimento AIML, referente ao processamento pragmático, que direcionam a sentença para outros padrões. Assim, cada padrão tem 4 perguntas, onde os métodos e técnicas de PLN buscam associar durante o estudo de caso. O arquivo de termos ignorados, utilizados no pré-processamento, contém 12 termos que devem ser ignorados na sentença. O arquivo de sinônimos, utilizado no processamento semântico, contém 12 termos que devem ser traduzidos na sentença. Todos os dados dos arquivos foram criados com base nas perguntas elaboradas no estudo de caso.

Figura 11. Interação com o *chatbot* Dóris durante a aula “Direção Preventiva”.

The screenshot displays a virtual learning environment interface. On the left, a navigation menu includes 'Conteúdo', 'Animação', and 'Exercícios'. The main content area is titled 'Cintos de segurança e airbags' and contains text stating: 'Os encostos de cabeça deverão estar posicionado de 5 a 10 cm e nivelados com a cabeça.' Below this text is a diagram of a car seat with a yellow crash test dummy. Labels on the diagram indicate: 'Alinhamento com cabeça' (5 - 10 cm), 'Se possível ajuste o centro do volante com o centro do seu peito' (25 cm), 'Distância confortável', and 'Recline levemente o banco e ajuste a altura se possível'. A chat window titled 'Bate-Papo' is overlaid on the left, showing a student's question: 'Qual a distância do encosto da cabeça?' and the chatbot's response: 'A distância correta do encosto de cabeça é de 5 à 10 cm.' On the right side, there is a menu with options: 'Menu Disciplinas', 'Menu Aulas', 'Dúvida', and 'Sair', along with a small avatar of a character in a yellow shirt.

Fonte: dos autores (2016)

Durante o estudo de caso, envia-se cada uma das 40 sentenças, referente às dúvidas dos 10 exercícios, ao *chatbot* Dóris. A Figura 11 mostra um exemplo de interação realizado

durante o estudo de caso. Para cada sentença, são escritos em um arquivo de saída os detalhes do processamento de cada método e técnica de PLN. Através deste arquivo de saída, são avaliados se os seguintes critérios atingiram seus objetivos no processamento na sentença de entrada do usuário:

- Pré-processamento: verifica se a técnica criada de adicionar espaços entre as pontuações e o método de remoção de termos menos significativos da sentença atingiram seus objetivos.
- Etiquetagem Morfossintática: verifica se o etiquetador morfossintático *Tree-Tagger* classificou corretamente a classe gramatical das palavras da sentença e se as mesmas foram padronizadas corretamente.
- Processamento Semântico: verifica se a técnica de traduzir os termos da sentença para o mesmo padrão da base de conhecimento, respeitando o significado semântico, atingiram seus objetivos esperados.
- Processamento Pragmático: verifica se no método de processamento pragmático, a sentença de entrada é direcionada corretamente para o padrão na base de conhecimento que o usuário se refere ao enviá-la.
- Remoção de palavras com classes gramaticais menos significativas: verifica se o método de comparação da sentença do usuário com a base de conhecimento através da remoção de palavras menos significativas, com base em sua classe gramatical, atingiu o objetivo esperado.
- Comparação do último verbo e substantivos seguintes: verifica se o método de comparação da sentença do usuário com a base de conhecimento, através da técnica de comparar o último verbo com os substantivos seguintes ao verbo, atingiram os objetivos esperados.
- Resposta do *chatbot* Dóris: verifica se a resposta do agente *chatbot* tem sentido de coesão e coerência à sentença enviada pelo usuário.

4.2.1 Resultados do Estudo de Caso com *chatbot* Dóris

Os dados analisados no arquivo de saída após a realização do estudo de caso estão dispostos na tabela 4, com uma análise quantitativa dos acertos e erros em cada etapa do processamento da sentença do usuário. Cada linha da tabela representa uma sentença enviada ao *chatbot*. As colunas representam a sentença e, respectivamente, os resultados dos métodos e técnicas no processamento.

Tabela 4. Quadro com os resultados do Estudo de Caso.

Nº	Sentença de entrada	Pré-processamento	Etiquetagem	Processamento Semântico	Processamento Pragmático	Compara palavras por classe gramatical	Compara verbo e substantivos	Coerência da Resposta
1	Qual o ajuste adequado do cinto de segurança?	✓	✓	✓	✓	✗	✓	✓
2	Como ajustar adequadamente o cinto de segurança?	✓	✓	✓	✓	✓	-	✓
3	Não sei como regular o cinto de segurança.	✓	✓	✓	✓	✓	-	✓
4	Qual a melhor forma de ajustar o cinto de segurança do veículo?	✓	✓	✓	✓	✗	✗	✗
5	Qual a distância correta da cabeça no encosto de banco do veículo?	✓	✗	✓	✓	✗	✗	✗
6	Como distanciar adequadamente a cabeça do encosto do veículo?	✓	✗	✗	✓	✗	✗	✗
7	Qual a distância do encosto de cabeça?	✓	✓	✓	✓	✓	-	✓
8	Não sei qual a distância do encosto de cabeça.	✓	✓	✓	✓	✓	-	✓
9	Dóris, qual a distância ideal do peito do motorista do volante?	✓	✓	✓	✓	✓	-	✓
10	Por favor, qual a distância correta do peito do motorista e do volante?	✓	✗	✓	✗	✗	✗	✗
11	Qual a distância do tórax do motorista do volante?	✓	✓	✓	✓	✓	-	✓
12	Sabe qual a distância corretamente o tórax do volante?	✓	✓	✓	✓	✓	-	✓
13	Qual a redução de fatalidades ao utilizar o cinto de segurança?	✓	✓	✓	✓	✓	-	✓
14	Qual a redução de acidentes com o uso do cinto de segurança?	✓	✓	✓	✓	✓	-	✓
15	Qual a redução do cinto de segurança nas fatalidades em caso de acidentes?	✓	✓	✓	✓	✗	✗	✗
16	Qual a redução de fatalidades na utilização do cinto de segurança?	✓	✓	✓	✓	✓	-	✓
17	Quando usar o cinto de segurança?	✓	✓	✓	✓	✓	-	✓
18	Quando devemos usar o cinto de segurança?	✓	✓	✓	✓	✗	✓	✓
19	Qual o momento de usar o cinto de segurança?	✓	✓	✓	✓	✗	✓	✓
20	Dóris, quando devo utilizar o cinto de segurança?	✓	✓	✓	✓	✓	-	✓
21	Qual o principal motivo de colisões em troca de faixas?	✓	✓	✓	✓	✓	-	✓
22	Me diga, quando ocorrem colisões em mudança de faixas?	✗	✓	✗	✓	✗	✗	✗
23	Qual o motivo das colisões em troca de faixas?	✓	✓	✓	✓	✓	-	✓
24	Quando são ocasionadas colisões em troca de faixas?	✓	✗	✓	✓	✗	✗	✗
25	Qual a regra para distância segura entre o veículo da frente?	✓	✓	✓	✓	✓	-	✓
26	Como funciona a regra de distância correta entre o veículo da frente?	✓	✗	✓	✓	✗	✗	✗
27	Qual a regra de distância do veículo da frente?	✓	✓	✓	✓	✓	-	✓
28	Dóris, alguma regra importante de distância para o veículo da frente?	✓	✓	✓	✓	✓	-	✓
29	Qual é a distância ideal para parar no cruzamento?	✓	✓	✓	✓	✓	-	✓
30	Como são as distâncias ideais para quando eu for parar em cruzamentos?	✓	✓	✓	✓	✗	✗	✗
31	Qual é a distância certa para se parar atrás de outro carro durante um cruzamento?	✓	✗	✗	✓	✗	✗	✗
32	Não sei qual a distância ideal que devo parar atrás de um cruzamento.	✓	✗	✓	✗	✗	✗	✗
33	Não sei qual o principal motivo de acidentes em cruzamentos.	✓	✓	✓	✓	✓	-	✓
34	Dóris, sabe me dizer qual o principal motivo de acidentes em cruzamentos?	✓	✓	✓	✓	✓	-	✓
35	Não entendo qual o principal motivo de fatalidades em cruzamentos.	✓	✓	✓	✓	✓	-	✓
36	Por que acontecem colisões em cruzamentos?	✓	✓	✓	✓	✗	✗	✗
37	O que acontece se dirigir em alta velocidade?	✓	✓	✓	✓	✗	✓	✓
38	O que pode acontecer se eu dirigir em alta velocidade?	✓	✓	✓	✓	✗	✓	✓
39	Sabe quais são as coequências com quem dirigir em velocidade acima da permitida?	✓	✓	✗	✗	✗	✗	✗
40	O que pode acontecer se dirigir em alta velocidade?	✓	✓	✓	✓	✗	✓	✓
Porcentagem de Acerto		97,5%	82,5%	90,0%	92,5%	52,5%	31,6%	67,5%

Fonte: dos autores (2016)

O símbolo verde de correto “✓” indica que a etapa ocorreu corretamente, enquanto o símbolo vermelho de um “✗”, indica que houve processamento errado da etapa, e por fim, o símbolo de traço “-”, indica que a etapa não foi utilizada para aquela sentença. Ao fim da tabela, é possível ver a porcentagem de acerto de cada etapa do processamento da sentença.

Os resultados obtidos no estudo de caso mostram que 67.5% das sentenças enviadas ao *chatbot* foram associadas aos padrões pré-estabelecidos na base de conhecimento e 32,5% não tiveram padrões associados. Também, que a etapa de pré-processamento processou 97,5% das sentenças corretamente. O etiquetador morfossintático *Tree-Tagger* classificou

corretamente 82,5% das sentenças. A etapa de processamento semântico, processou corretamente 90% das sentenças. Na etapa de processamento pragmático, 92,5% das sentenças foram processadas corretamente. Na comparação com a sentença com a base de conhecimento, o método comparou corretamente 52,5% das sentenças, enquanto que o método de comparação do último verbo e substantivos seguintes comparou 31,6% das sentenças.

A elaboração deste estudo de caso busca demonstrar confiabilidade nas operações realizadas e procedimentos de coleta de dados, onde os mesmos podem ser repetidos, apresentando resultados semelhantes. Depois das coletas de dados realiza-se a análise dos resultados obtidos buscando esclarecer características observadas durante o estudo de caso com o *chatterbot* Dóris.

4.3 Análise dos Resultados

Com a finalidade de análise dos métodos e técnicas de PLN usados e implementados, a análise dos resultados busca esclarecer características observadas durante o estudo de caso com o *chatterbot* Dóris.

Os resultados na etapa de pré-processamento da sentença de entrada, obtidos no estudo de caso, mostram que 97,5% das sentenças enviadas ao *chatterbot* foram pré-processadas corretamente. Durante esta etapa, observa-se que a alta porcentagem de acerto se deve à técnica de adicionar espaço entre caracteres de pontuação, já que para todas as sentenças enviadas, a função desenvolvida em Java sempre irá adicionar os espaços corretamente na sentença. O método de ignorar termos da sentença apresentou uma alta taxa de acertos, contudo, os 2,5% de erros desta etapa são causados por falta de termos que deveriam ser ignorados na sentença através do arquivo de configuração deste método. Porém, já que se trata de falta de termos no arquivo de configuração, a adição de novos termos pode evitar inexistências futuras.

O etiquetador morfossintático *Tree-Tagger* mostrou-se satisfatório, já que etiquetou corretamente 82,5% das sentenças de entrada. Os resultados mostraram que a grande maioria das palavras das sentenças tiveram sua classe gramatical classificada corretamente, além da derivação de sua forma padrão. Também, esta etapa é de grande importância para o restante do processamento da sentença, já que as próximas etapas fazem uso das informações obtidas

nesta. Além disso, o etiquetador morfossintático obteve 17,5% de erros, onde são observadas duas características principais:

1. Em casos de sentenças que contém palavras que são derivadas de outra palavra contida na base de conhecimento do agente, porém com classe gramatical diferente, a palavra terá sua classe gramatical etiquetada de forma diferente, o que acaba dificultando o processo de comparação. Por exemplo, o substantivo "Distância" e o verbo "Distanciar".
2. O etiquetador morfossintático comete alguns erros de etiquetagem por falta de termos no arquivo de configuração que comprometem o restante do processamento da sentença. Por exemplo, na sentença "Qual a distância correta do peito do motorista e do volante?", ele classifica a palavra distância como verbo, logo a palavra é passada para o seu estado infinitivo que é "distanciar", fazendo o *chatbot* não reconhecer o padrão da sentença.

O processamento semântico apresentou resultados promissores, já que processou corretamente 90% das sentenças enviadas ao *chatbot*. Através de um arquivo de dicionário de sinônimos é possível traduzir termos da sentença para outros termos com o mesmo padrão da base de conhecimento, respeitando seu significado. Os 10% de erros da etapa, se deve ao caso de termos que não foram traduzidos para o padrão da base de conhecimento, pois os mesmos não estavam especificados no arquivo de sinônimos. Porém, estas inexistências podem ser corrigidas adicionando novos termos no arquivo de dicionário, evitando que as mesmas inexistências ocorram novamente.

Os resultados do processamento pragmático mostram que 92,5% das sentenças foram processadas corretamente. Nesta etapa, pode-se observar que através do AIML é possível direcionar sentenças, que dependem de um conhecimento externo do agente *chatbot*, além da sentença de entrada, para direcioná-la a uma resposta que o usuário espera. Observa-se que os 7,5% de erros da etapa se devem ao caso de padrões que não estão especificados na base de conhecimento AIML. Porém, estas inexistências de termos podem ser corrigidas adicionando novos padrões na base de conhecimento para direcionar a sentença para outros padrões que representam a resposta esperada do usuário, evitando que as mesmas inexistências ocorram novamente.

O primeiro método de comparação da sentença processada com a base de conhecimento é a remoção de palavras menos significantes com base em sua classe gramatical. Neste método, 52,5% das sentenças foram comparadas corretamente, assim,

observa-se que a seleção de palavras pela classe gramatical é um método promissor, desde que a sentença processada contenha exatamente as mesmas palavras do padrão da base de conhecimento. A taxa de erro de 47,5% deste método deve-se à sentença processada não apresentar exatamente a mesma quantidade de palavras da base de conhecimento em AIML. Este problema acontece devido às diferentes formas que uma sentença pode ser processada.

O segundo método de comparação da sentença processada com a base de conhecimento é seleção do último verbo da sentença e seus substantivos seguintes ao verbo. Este método é utilizado quando o primeiro método de comparação não encontra um padrão associado à sentença na base de conhecimento. No estudo de caso, este método foi utilizado 19 vezes, sendo que 31,6% das sentenças tiveram um padrão associado através deste, mas 68,4% das sentenças não tiveram um padrão associado na base de conhecimento AIML. É possível apontar que a principal causa desta taxa de erro é que mesmo que sentenças formuladas apresentem este padrão em sua estrutura sintática, também poderão apresentar outros padrões. Assim, uma possível solução para este problema é criar outros padrões para comparação da sentença, e não apenas este.

É retornada uma resposta a 67,5% das sentenças submetidas ao *chatterbot* Dóris durante o estudo de caso. Através das taxas de acertos de cada etapa, é possível analisar que as etapas de pré-processamento, etiquetagem morfosintática, processamento semântico e processamento pragmático tiveram resultados promissores para o processo de interação do *chatterbot* com o usuário. A etapa de remoção de palavras com base na sua classe gramatical e a seleção do último verbo mais os substantivos seguintes, ambos métodos de comparação da sentença processada com a base de conhecimento AIML, mostram resultados abaixo do esperado, sendo o principal fator na taxa de 32,5% de sentenças que não tiveram respostas.

Através da análise dos resultados do estudo de caso, pode-se observar que quanto maior a complexidade da sentença, maior é a probabilidade de erro no processamento da mesma, pois a base de conhecimento pode apresentar um padrão menos complexo. Também podem haver situações contrárias a esta, pois se o padrão da base de conhecimento é complexo, e a sentença enviada pelo usuário é mais simples, a sentença não será associada ao padrão, já que a sentença não terá informações suficientes.

Pode-se observar que as principais características que causaram problemas no retorno de uma sentença ao usuário acontecem no processo de comparação da sentença processada com a base de conhecimento em AIML. Assim, sugere-se como solução, métodos e técnicas

mais flexíveis no processo de comparação da sentença processada com a base de conhecimento do agente *chatterbot*.

A implementação de métodos e técnicas de PLN para *chatterbots* se mostram promissoras no processo de interação com o usuário. A realização do estudo de caso possibilita observar aspectos específicos de cada método e técnica e, portanto, é possível analisar seus resultados e definir quais métodos e técnicas apresentam resultados satisfatórios e quais devem ser ajustados.

É apresentado na tabela 5 os trabalhos relacionados pesquisados neste trabalho juntamente com este, comparando os objetivos, métodos de PLN, técnicas de PLN, níveis da análise linguística e resultados obtidos. Assim, através da relação dos trabalhos pode-se analisar que este trabalho, juntamente com os demais, se objetiva em padronizar a sentença de entrada do usuário a um mesmo padrão que corresponda a sua base de conhecimento. Porém, a maioria dos outros trabalhos foca a pesquisa em níveis específicos da análise linguística, enquanto que este busca nos quatro principais níveis da análise linguística escrita métodos e técnicas de PLN que beneficiam o processo de interação entre o usuário e o agente *chatterbot*.

Este capítulo destinou-se a descrever o desenvolvimento do *chatterbot* Dóris através de métodos e técnicas de PLN, juntamente com o estudo de caso experimental no ambiente virtual de aprendizagem e, por fim, a análise dos resultados obtidos no estudo.

Tabela 5. Comparação deste trabalho com os relacionados

Trabalhos Relacionados	Objetivos	Métodos de PLN	Técnicas de PLN	Níveis da análise linguística	Resultados Obtidos
(CÓRTEZ; BAGATINI; FROZZA, 2016)	Desenvolver a capacidade de <i>chatbot</i> do agente tutor Dóris a fim de beneficiar o processo de interação com o usuário.	Analisador morfossintático / Remover termos menos significativos / Direcionar sentença para resposta que o usuário espera. / Selecionar palavras conforme a sua classe gramatical.	Etiquetador <i>Tree-Tagger</i> / Adicionar espaço entre pontuações / Traduzir termos através de dicionário semântico / Base de conhecimento em AIML.	Morfológico / Sintático / Semântico / Pragmático	Métodos e técnicas podem beneficiar o processo de interação de um agente <i>chatbot</i> com um usuário em um ambiente virtual de aprendizagem.
(RAJ; ABDUL-KAREEM, 2011)	Desenvolver um algoritmo capaz de derivar verbos no participio ou em diferentes tempos verbais em seus estados infinitivos	Usar gramáticas para compreender as sentenças / Combinar palavras para entender seu significado / Derivar a forma infinitiva do verbo	Usar condições "IF ELSE" para testar verbos em padrões estruturados / Adicionar novas condições em sua tabela de padrões estruturados / Representar os padrões em lógica de primeira ordem / Usar a relação sujeito, verbo e objeto / Algoritmo de predição	Lexical / Morfológico / Sintático	Qualquer verbo pode ser decomposto corretamente em seu estado infinitivo. Se um verbo não puder ser decomposto naquele momento, então é adicionada uma nova condição na tabela de padrões.
(KIMURA, 2015)	Extrair informações essenciais contidas na pergunta de entrada do usuário e pesquisar por uma resposta adequada correspondente às informações de uma base de dados	Dividir a frase da pergunta do usuário em três partes (substantivo, sujeito e predicado)	Tópicos mapeados para associar as relações de palavras para respostas apropriadas / Análise de dependência nas sentenças de entrada para extrair suas partes principais	Sintático / Semântico	O método utilizado retornou uma resposta para aproximadamente 88% das questões sendo que cerca de 40% das perguntas tiveram respostas válidas
(DINGLI; SCERRI, 2013)	Fornecer um protótipo capaz de prover um entendimento sintático, semântico e pragmático da entrada do usuário	Entendimento da Linguagem Natural / Processamento intermediário / Geração de resposta / Usar técnicas e ferramentas modernas de PLN / Unir diversas bases de informações em uma única base	<i>Engine ChatScript</i> para <i>chatbots</i> / Mapear a entrada em linguagem natural em uma representação mais formal em XML / Mapear pronomes para entidades nomeadas / Verificação ortográfica	Morfológico / Sintático / Semântico / Pragmático	O trabalho propôs um <i>framework</i> com potencial para dar suporte em implementação de agentes de conversação dinâmica

Fonte: dos autores (2016)

5 CONCLUSÃO

Este trabalho de conclusão propõe, através do desenvolvimento de métodos e técnicas de PLN acrescentar a característica de *chatterbot* ao agente Dóris, de forma a contribuir no processo de interação com o usuário em um ambiente virtual de aprendizagem já desenvolvido. Para atingir este objetivo, são utilizados e implementados métodos e técnicas de PLN e através de um estudo de caso com o *chatterbot* Dóris, analisa-se os resultados obtidos no processamento da sentença enviada pelo usuário ao *chatterbot*.

Os métodos e técnicas de PLN, são utilizados para padronizar a sentença de entrada do usuário, de forma que a mesma corresponda ao mesmo padrão na base de conhecimento AIML. Assim, faz necessário etapas no processamento da sentença que tratam problemas na linguagem natural escrita, mais especificamente nos níveis da análise linguística. Na análise morfosintática, busca-se definir a classe gramatical das palavras, definir o seu gênero, tempo e modo verbal e o número (singular ou plural). A análise semântica busca identificar os possíveis significados de uma frase e determina qual o significado mais adequado olhando o seu contexto. Por fim, na análise pragmática, procura-se o entendimento através do conhecimento externo que falta no diagnóstico de uma única sentença.

Com o estudo destes métodos e técnicas, pode-se notar que a área de pesquisa de PLN possui diversos desafios na compreensão automática da língua humana, que consiste em representar a linguagem natural em um formato mais formal, o que permite que programas de computadores manipulem estas informações. Os principais desafios na área de PLN estão relacionados às ambiguidades que a linguagem natural proporciona, dentre elas se destaca as ambiguidades sintáticas e semânticas, inconsistência, dependência de contexto, erros gramaticais, sentenças imprecisas e incompletas e sentenças combinatoriamente explosivas, já que uma sentença pode ser formulada de diversas formas.

Através da análise dos resultados obtidos no estudo de caso com o *chatterbot* Dóris pode-se concluir que os métodos e técnicas de PLN pesquisados e implementados mostraram-se promissores no processo de interação de um *chatterbot* com o usuário. As etapas do processamento da sentença se dividem em: realizar o pré-processamento da sentença, a fim de tratar características que dificultem o processamento nas etapas seguintes; classificar a classe gramatical das palavras, como também derivá-las em sua forma padrão através de um analisador morfosintático; traduzir termos da sentença para o mesmo padrão da base de

conhecimento através do processamento semântico; direcionar a sentença através do processamento pragmático para uma resposta que o usuário espera na base de conhecimento; e por fim, comparar a sentença processada com a base de conhecimento que é representada por um arquivo AIML.

Diante as taxas de acertos de cada etapa do processamento da sentença obtida no estudo de caso, é possível analisar que as etapas de pré-processamento, etiquetagem morfosintática, processamento semântico e processamento pragmático tiveram resultados promissores para o processo de interação do *chatbot* com o usuário. Nas etapas de comparação da sentença processada com a base de conhecimento foram obtidos resultados abaixo do esperado, sendo o principal fator na taxa de sentenças que não tiveram respostas.

As principais características que causam problemas no retorno de uma sentença de resposta ao usuário neste trabalho, acontecem no processo de comparação da sentença processada com a base de conhecimento em AIML, pois os métodos de comparação precisam que a sentença processada e o padrão correspondente na base de conhecimento contenham exatamente as mesmas palavras. Assim, sugere-se como solução, métodos e técnicas mais flexíveis no processo de comparação, de forma que permitam a associação da sentença processada com o padrão da base de conhecimento, sem que os mesmos tenham exatamente os mesmos termos e na mesma ordem.

Antes deste trabalho, a interação do usuário com o agente pedagógico Dóris no ambiente virtual de aprendizagem acontecia através de caixas de diálogo, onde o usuário deveria expressar-se através da escolha de opções pré-definidas que mais se assemelhavam com sua resposta. Ou seja, o usuário não tinha a liberdade de expressar-se de maneira natural. Este trabalho desenvolveu a característica de *chatbot* ao agente Dóris e com isso, permite que o usuário interaja e expresse seu conhecimento através da linguagem natural, o que beneficia o processo de aprendizagem, segundo (KERLY; ELLIS; BULL, 2009). Assim, conclui-se que o desenvolvimento de métodos e técnicas de PLN escrita podem beneficiar o processo de interação de um agente pedagógico com um usuário em um ambiente virtual de aprendizagem.

Para trabalhos futuros, sugere-se a pesquisa e desenvolvimento de um método que compare de forma flexível a sentença processada com a base de conhecimento, já que os métodos de comparação utilizados neste trabalho não obtiveram resultados satisfatórios. Também, mesmo que o etiquetador morfosintático *Tree-Tagger* apresente uma taxa de

82,5% de etiquetagem correta, a etapa é considerada de extrema importância por fornecer dados para as etapas seguintes do processamento da sentença. Logo sugere-se um analisador morfossintático que apresente resultados superiores no processo de etiquetagem das palavras da sentença do usuário.

REFERÊNCIAS

A.L.I.C.E. Artificial Intelligence Foundation. [Online]. Disponível em: <http://www.alicebot.org/>. Acessado em maio de 2016.

BANDA, Everton Moschen; MENEZES, Crediné Silva. *Uma Proposta para extração de perguntas e respostas de textos*. Nuevas Ideas en Informática Educativa, TISE. Vol.8, 2012. p.44-49.

BEHAR, P. A; MEIRELLES L., S; MAZZOCATO, S. B; SOUZA, L. B; SIQUEIRA, L. G. *Avaliação de Ambientes Virtuais de Aprendizagem: O Caso do ROODA na UFRGS*. Revista Avances en Sistemas e Informática. Vol. 4, No. 1, 2007. p.87-99.

BONFANTE, Andreia Gentil; NUNES, Maria das Graças Volpe. *Parsing probabilístico para o Português do Brasil*. In: Brazilian Symposium on Artificial Intelligence, 16. 2002.

BORIN, M; FROZZA, R.; KIPPER, L.; SCHREIBER, J. N. C. *Virtual Learning Environments with Emotional Pedagogical Agents for Training in Organizations*. In: International Conference on Industrial Engineering and Operations Management (ICIEOM), Guimarães, Portugal, 2012,

BORIN, P. M.; *O uso de ambientes virtuais de aprendizagem com agentes pedagógicos emocionais para capacitação em organizações*. Dissertação (Mestrado em Sistemas e Processos Industriais) - UNISC, Programa de Pós-Graduação em Sistemas e Processos Industriais, 2010.

CERVO, Amado Luiz; BERVIAN, Pedro Alcino; SILVA, Roberto da. *Metodologia científica*. 6. ed. São Paulo: Pearson, 2007. xii, 162 p.

CHOWDHURY, Gobinda G. *Natural language processing. Annual review of information science and technology*, v. 37, n. 1, 2003. p. 51-89.

DINGLI, Alexiei; SCERRI, Darren. *Building a Hybrid: Chatterbot – Dialog System*. Springer Berlin Heidelberg, 2013. P.145-152.

FERNANDES, W. L.; AGROREDE: *Estudo exploratório acerca da implementação de uma rede virtual de aprendizagem colaborativa envolvendo as escolas agro técnicas federais. Mestrado profissional em tecnologia da informação e comunicação na formação em educação a distância*. Universidade Federal do Ceará (UFC) e Universidade Norte do Paraná (UNOPAR). Salvador, BA – Brasil, 2007.

FRADE, Rodrigo Valença C.; NETO, Francisco Milton M.; LIMA, Rommel W.; LIMA, Rodrigo M.; SILVA, Luiz Claudio N.; SOUZA, Rafael C. *Um Ambiente Virtual 3D Multiagente com Recomendação Personalizada de Objetos de Aprendizagem*. III Congresso Brasileiro de Informática na Educação, SBIE. 2014.

FROZZA, R.; KONZEN, A., MAINIERI, A. G., SCHREIBER, J., MOLZ, K., TAUTZ, J., PEDÓ, R., DRESCH, j.; *Agentes tutor e companheiro em um ambiente educacional baseado em estilos cognitivos*. In: Workshop - Escola de Sistemas de Agentes para ambientes Colaborativos - WESAAC 2007. UCPEL, Pelotas, 2007.

FROZZA, R.; SILVA, A. A. K. da; SCHREIBER, J. N. C.; LUX, B.; MOLZ, K. W.; KIPPER, L. M.; BORIN, M. P.; CARVALHO, A. B. de; BAIERLE, J. L.; SAMPAIO, L. *Agentes Pedagógicos Emocionais atuando em um Ambiente Virtual de Aprendizagem*. Centro Interdisciplinar de Tecnologia Educacional da Universidade Federal do Rio Grande do Sul (CINTED-UFRGS). RENOTE - Revista Novas Tecnologias na Educação, 2011.

FROZZA, R.; SILVA, A. K. da; LUX, B.; CRUZ, M. E. J. K da; BORIN, M. Dóris 3D: *Agente Pedagógico baseado em Emoções*. In: XX SBIE - Simpósio Brasileiro de Informática na Educação, 2009.

GIRAFFA, Lúcia M. M. *Uma arquitetura de tutor utilizando estados mentais*. Tese (Doutorado em Ciências da Computação) – Instituto de Informática, UFRGS, Porto Alegre, 1999.

GRIESANG, G.; FROZZA, R. *Desenvolvimento de um Mecanismo de Comunicação entre Agentes Pedagógicos em um Ambiente Virtual de Aprendizagem*. Dissertação de mestrado. Universidade de Santa Cruz Sul – UNISC, 2013.

HARASIM, L.; HILTZ, R.; TELES, L.; TUROFF, M.; *Redes de Aprendizagem: Um guia para o Ensino e Aprendizagem Online*. São Paulo, SENAC, 2005.

HOMRICH, N. C.; FROZZA, R.; SCHREIBER, Jacques Nelson Corleta. *The use of emotions in student assessment in a virtual learning environment*. In: International Conference on Industrial Engineering and Operations Management, 2013, Valladolid. XIX International Conference on Industrial Engineering and Operations Management, ICIEOM 2013. Rio de Janeiro: ABEPRO, v. único, 2013. p. 1-10.

JAVA; Site oficial Java. Disponível em: <http://www.java.com>. Acessado em: maio de 2016.

KAMPFF, A. J. C; LIRA, A. F; REITZ, D. S; GOMES, F. J. L; FONSECA, L. C. C; MACHADO, N. C. F; BERCHT, M. *Relação entre o Perfil do Usuário e a Escolha do Perfil do Tutor*. In: Novas Tecnologias da Educação. Vol. 3, No. 1, 2005.

KERLY, Alice; ELLIS, Richard; BULL, Susan. *Conversational Agents in E-Learning*. Applications and Innovations in Intelligent Systems XVI, Springer, 2009. p. 169-182.

KIMURA, Masaomi. *A Proposal of Topic Map Based Chatterbot for Non-English Natural Language Input*. Procedia Computer Science, Volume 60, 2015, p. 841-849.

LESTER, James; BRANTING, Karl; MOTT, Bradford. *Conversational agents*. The Practical Handbook of Internet Computing, 2004. p. 220-240.

LIDDY, Elizabeth D. *Enhanced text retrieval using natural language processing*. Bulletin of the American Society for Information Science and Technology, v. 24, n. 4, 1998. p. 14-16.

MAINIERI, A. G.; FROZZA, R.; Schreiber, J.; Molz, K.; *Educational System based on Cognitive styles and/or Learning styles*. In: XXVII Annual Conference of the Cognitive Science Society. Stresa, 2005.

MERGEN, C. H. S., SCHREIBER, J. N. C.; *Agentes Pedagógicos Bayesianos*. In: SBIE 2005 - XII Simpósio Brasileiro de Informática na Educação, 2005, Juiz de Fora. Anais do SBIE, 2005.

MIKIC, Fernando A.; BURGUILLO, Juan C.; LLAMAS, Martín; RODRÍGUEZ, Daniel A.; RODRÍGUEZ, Eduardo. CHARLIE: *An AIML-based Chatterbot which Works as an Interface among INES and Humans*. EAEEIE Annual Conference, 2009.

MORAN, José M. *Ensino e Aprendizagem Inovadores com Tecnologias*. Revista Informática na Educação: Teoria & Prática – PGIE – UFRGS, 2000. p.137 - 144

MULLER, D. N. *Compreensão da Linguagem Falada*. Porto Alegre: PPGC-UFRGS. Exame de Qualificação EQ-72. 2002.

POSTGRESQL; *Sistema gerenciador de banco de dados objeto-relacional de código aberto*. Comunidade internacional PostgreSQL disponível em: www.postgresql.org. Acessado em maio de 2016.

PROGRAM AB; *Program AB*. Alice A.I Foundation. Disponível em: <https://code.google.com/archive/p/program-ab/>. Acessado em novembro de 2016.

RAJ, Ram Gopal. *A pattern based approach for the derivation of base forms of verbs from participles and tenses for flexible NLP*. Malaysian Journal of Computer Science, v. 24, n. 2, 2011. p. 63-72.

REZENDE, Solange Oliveira (Org.). *Sistemas inteligentes: fundamentos e aplicações*. Barueri: Manole, 2003. 525 p.

RUSSELL, Stuart; NORVING, Peter. *Inteligência Artificial*. Rio de Janeiro: Elsevier, 2013. 1016 p.

SANTOS, C. T. dos; FROZZA, R.; DAHMER A.; GASPARY, L. P.; *Dóris – Um Agente de Acompanhamento Pedagógico em Sistemas Tutores Inteligentes*. In: XII Simpósio Brasileiro de Informática na Educação – SBIE – UFES, 2001.

SCHAEFER, M.; FROZZA, REJANE; SILVEIRA, R. A.; RUTSATZ, R. F. *Treinamentos E-Learning com estratégia de ensino baseada em animações*. In: Congreso Internacional de Informática Educativa - TISE, 2015, Santiago - CHILE. XX Congreso Internacional de Informática Educativa - TISE. Santiago: Universidad de Chile, v. 11, 2015. p. 786-791.

SCHAEFER, M.; FROZZA, REJANE. *Treinamentos E-Learning com estratégia de ensino baseada em animações*. Programa de Pós-Graduação em Sistemas e Processos Industriais (PPGPSI): Universidade de Santa cruz do Sul, Santa Cruz do Sul, 2015.

SCHRAMM, W. *Notes on case studies of instructional media projects*. Working paper, the Academy for Educational Development, Washington, DC, 1971.

SGANDERLA, Rachele. B.; FERRARI, Débora. N.; GEYER, Cláudio. F. R. *BonoBOT: Um Chatterbot para Interação com Usuários em um Sistema Tutor Inteligente*. XIV Simpósio Brasileiro de Informática na Educação - NCE - IM/UFRJ, 2003.

SILVA, A. K.; Frozza, R; Borin, M. P.; Molz, K.; Schreiber, J.; Lux, B.; Carvalho, A. B.; Baierle, J.; Kipper, L.; *O Método Clínico de Piaget como forma de Avaliação da Aprendizagem em um Sistema Tutor Inteligente com Agente Pedagógico*. XXI Brazilian Symposium on Computer in Education (SBIE). João Pessoa/PB, 2010.

SILVA, J. M. C. da; HECK G. *Utilizando o questionário MSLQ para identificação de estados afetivos em um ambiente virtual de aprendizagem*. Itajaí, 2008.

SILVA, L. F. R. da; *DIMI: um agente selecionador de estratégias de ensino para sistemas tutores inteligentes*. Monografia (graduação) - Universidade de Santa Cruz do Sul, 2002.

SILVA, Ricardo E. D.; RAMOS, Jorge L. C.; RODRIGUES, Rodrigo L.; GOMES, Alex Sandro; FONSÊCA, José Alexandre V. *Mineração de dados educacionais na análise das interações dos alunos em um Ambiente Virtual de Aprendizagem*. Anais do XXVI Simpósio Brasileiro de Informática na Educação, 2015.

SIMÕES, Alberto; ALMEIDA, José João. *Jspell.pm – um módulo de análise morfológica para uso em processamento de linguagem natural*. In Actas do XVII Encontro da Associação Portuguesa de Linguística, Lisboa 2001. p. 485–495.

SCHMID, Helmut. *"Treetagger a language independent part-of-speech tagger."* Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart 43. 1995.

TAGNIN, Stella Esther; VALE, Oto Araújo. *Avanços da Linguística de Corpus no Brasil*. Humanitas, São Paulo – SP, 2008. 437p.

TEIXEIRA, Sérgio; RAMIRO, Thiago B.; OLIVEIRA, Elias; MENEZES, Crediné S. *Chatterbots em ambientes de aprendizagem – uma proposta para a construção de bases de conhecimento*. XI Workshop de Informática na Escola. São Leopoldo, 2005.

TURING, Alan M. *Computer Machinery and Intelligence*. *Mind.*, Vol. 9, No. 236, 1950. p. 433-460.

VALENTINI, C. B.; SOARES, E. M. do S. *Aprendizagem em Ambientes Virtuais: compartilhando idéias e construindo cenários*. 2008. Disponível em: <http://www.ucs.br/etc/revistas/index.php/aprendizagem-ambientes-virtuais/article/viewFile/393/323>, acessado em abril de 2016.

VANTI, Nadia Aurora Peres. *Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento*. *Ciência da informação*, v. 31, n. 2, 2002. p. 152-162.

VIEIRA, Renata; LIMA, Vera LS. *Linguística computacional: princípios e aplicações*. In: Anais do XXI Congresso da SBC. I Jornada de Atualização em Inteligência Artificial, 2001. p. 47-86

WOOLDRIDGE, Michael J. *An Introduction to Multiagent Systems*. Chichester: John Wiley & Sons, 2001. 348 p.

YIN, Robert K. Estudo de caso: planejamento e métodos. 5. ed. Porto Alegre: Bookman, 2015. 290 p.