

Uni-ANHANGUERA CENTRO UNIVERSITÁRIO DE GOIÁS
CURSO DE ENGENHARIA DA COMPUTAÇÃO

SOFTWARE EXPERIMENTAL DE COMBATE A FAKE NEWS
UTILIZANDO INTELIGÊNCIA ARTIFICIAL

DIMITRI MUNIZ JARDIM

GOIÂNIA
Novembro/2019

DIMITRI MUNIZ JARDIM

**SOFTWARE EXPERIMENTAL DE COMBATE A FAKE NEWS
UTILIZANDO INTELIGÊNCIA ARTIFICIAL**

Trabalho de conclusão de curso apresentado ao Centro Universitário de Goiás Uni-Anhanguera, sob orientação do Prof. Ms. Murilo Parreira Leal, como requisito parcial para obtenção do título de Bacharel em Engenharia da Computação.

GOIÂNIA
Novembro / 2019

FOLHA DE APROVAÇÃO

DIMITRI MUNIZ JARDIM

**SOFTWARE EXPERIMENTAL DE COMBATE A FAKE NEWS UTILIZANDO
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão de Curso apresentado à banca examinadora como requisito parcial para obtenção do título de bacharel em Engenharia de Computação do Centro Universitário de Goiás – Uni-ANHANGUERA, definido e aprovado em ____de____de____ pela banca examinadora constituída por:

Prof. Ms. Murilo Parreira Leal

Orientador

Prof. Dr. Vagner Reis da Silveira

Membro

Prof. Ms. Alisson Rodrigues Alves

Membro

Prof. Ms. Fabricio Schlag

Membro

Dedico este trabalho ao Grande Arquiteto do Universo, a minha família, minha filha, meus pais, colegas acadêmicos e professores, que estiveram sempre ao meu lado, compartilhando comigo suas histórias e me ajudando a construir a minha.

AGRADECIMENTOS

Agradeço acima de tudo ao meu senhor Jesus Cristo, por seu sacrifício que me permitiu viver. A minha mãe, que mesmo sem condições me apoiou, lutou para que eu tivesse a oportunidade de sonhar. A minha amada esposa, que caminha ao meu lado em todos os momentos e me mantém seguindo em frente. A minha filha, que é a luz da minha vida. A minha amiga Thais Oliveira, por todo o apoio e carinho. Aos membros do Clã Pioneiro Cruzeiro do Sul, meus amados amigos que se tornaram minha família e me afastaram da solidão. A esta instituição e seus professores, que com muito carinho me moldaram como cidadão e engenheiro.

"Nossa cultura vê as emoções como algo mais complexo e obscuro que o intelecto (...). Na verdade, acho que sabemos muito mais sobre a emoção hoje do que sobre a razão."

In Music, Mind and Meaning (1981)

Marvin Minsky

RESUMO

O presente trabalho expõe um problema social chamado “Fake News”, e demonstra uma abordagem usando inteligência artificial e suas tecnologias, para desenvolver um modelo experimental de software com o intuito de conter o fenômeno da desinformação causado por esse tipo de conteúdo. A inteligência artificial se mostra uma ferramenta capaz de auxiliar neste combate, através do uso de sistemas especialistas denominados de chatbots, que por sua vez, é um sistema capaz de conversar com uma pessoa em linguagem natural, de maneira similar a um diálogo entre seres humanos. Embutindo tecnologias de aprendizagem como o machine learning e o deep learning, o chatbot pode se tornar capaz de analisar uma notícia em formato de texto e comunicar ao leitor, se o teor daquela notícia é confiável ou não.

PALAVRAS-CHAVE: Chatbots. Inteligência artificial. Fakenews. Sistemas especialistas.

LISTA DE ABREVIATURAS E SIGLAS

IA	Inteligência Artificial
PNL	Processamento de Linguagem Natural
AWS	Amazon Web Services
ASR	Reconhecimento automático de fala
NLU	Compreensão de linguagem natural
EC2	Elastic Compute Cloud
API	Interface de Programação de Aplicativos
HTTP	Hyper Transfer Protocol

LISTA DE FIGURAS

Figura 1. Árvore Sintática	21
Figura 2. Transformações da sentença na estrutura sintática e na forma lógica	22
Figura 3. Estrutura da aplicação	24
Figura 4. Estrutura do crawler octaparse.	25
Figura 5. Funcionamento do PostgREST	25
Figura 6. Schema API	26
Figura 7. Tabela API	26
Figura 8. Função anon	27
Figura 9. Função authenticator	27
Figura 10. Função select fakenews	28
Figura 11. Instanciando o banco	29
Figura 12. Criando as tabelas	29
Figura 13. Inserção de Dados	29
Figura 14. Especificações da instancia	30
Figura 15. Bibliotecas usadas	31
Figura 16. Gerando token	
Figura 17. Criando e declarando a variável de ambiente	31
Figura 18. Configurando o webhook	32
Figura 19. Método POST	33
Figura 20. Função check_fakenews	33
Figura 21. Função send_text	34
Figura 22. Funcionamento do chatbot	35

LISTA DE TABELAS

Tabela 1. Formas de comunicação PLN

21

SUMÁRIO

1 INTRODUÇÃO	11
2 OBJETIVOS	13
2.1 <i>Objetivo Geral</i>	13
2.2 <i>Objetivos Específicos</i>	13
3 JUSTIFICATIVA	14
4 HIPOTESE	15
5 REFERENCIAL TEÓRICO	16
5.1 <i>Fake News</i>	16
5.3 <i>Machine Learning</i>	19
5.3.1 <i>Deep Learning</i>	19
5.4 <i>Processamento de Linguagem Natural</i>	20
5.4.1 <i>História PLN</i>	20
5.4.2 <i>Funcionamento</i>	21
5.4.3 <i>Forma Lógica</i>	22
5.5 <i>Sistemas Especialistas</i>	23
5.6 <i>Chatbots</i>	23
6 MATERIAS E MÉTODOS	24
6.1 <i>Aplicação</i>	24
6.2 <i>Coleta de Dados – Web Crawler</i>	24
6.3 <i>Bando de dados – Dataset e PostgREST API</i>	25
6.3.1 <i>API</i>	26
6.3.2 <i>Banco de Dados</i>	28
7 FACEBOOK CHAT BOT	31
RESULTADOS	35
CONCLUSÃO	36
REFERENCIAS BIBLIOGRÁFICAS	37

1 INTRODUÇÃO

Partiu-se da evidência que vivemos na Era do Conhecimento, na qual todo tipo de informação é considerado válida por muitos e tida como suprema verdade. Neste contexto, as tecnologias da informação se alicerçaram como uma realidade comum, conquistando progressivamente espaços entre a comunidade mundial. Comungando com essa concepção, Sanchez (2006, p. 45) registra que:

O livre acesso à Internet trouxe um desenvolvimento para o campo muito grande. Qualquer pessoa pode ter acesso a uma informação sobre outra, independentemente da localização de ambas. Isso traz rapidez e facilidade no processo relacional dos usuários. Porém, esse argumento somente é legítimo quando há o consentimento das partes sobre o compartilhamento de suas informações. O brilhantismo da internet é ofuscado por sua incapacidade de fornecer segurança a seus usuários, uma vez que uma informação pessoal pode cair nas mãos de pessoas mal-intencionadas, o que pode gerar danos financeiros, morais e imagéticos.

Observa-se que, com o aumento da velocidade da comunicação entre grupos de indivíduos, muitas questões importantes para a sociedade vem sendo discutidas por um maior número de pessoas, dentre estas, grupos que antes desse avanço tecnológico não se envolviam ou não se interessavam por certos assuntos, criando uma rede democrática onde todos tem opiniões e poder de voz.

Por um lado, essa maior abrangência acaba por incluir mais opiniões nas discussões, porém com a facilidade do acesso, a criteriosidade acerca dos conteúdos consumidos é cada vez menor, ou sequer existe, de modo que se propagam conteúdos de caráter duvidoso pelos meios digitais. Um exemplo são as recentes eleições Brasileiras e Americanas, ambas marcadas por disseminação desse tipo de notícia, fato este que coloca em xeque a legitimidade das eleições.

Com as novas redes, a disseminação desse tipo de notícia é cada vez mais veloz, e o usuário é bombardeado em seu *feed* com informações que o agradam. Este, por sua vez, acaba, pela forma flexível que o compartilhamento de *posts* se dá, sem nem ao menos buscar a fonte de certa notícia, repassando a outros grupos. Essas falsas informações tornam-se fatos importantes na formação de ideias pelos usuários, que tomam uma mentira por verdade, causando o que conhecemos hoje por polarização partidária. Sendo assim é imprescindível que medidas emergenciais sejam adotadas para o combate destas informações inverídicas, as quais podem impactar de forma negativa e, na pior das hipóteses, promover situações desastrosas, colocando inclusive a integridade dos envolvidos em risco.

Baseado nesse pressuposto, este estudo optou pela escolha da temática: Inteligência Artificial no combate a *Fake News* na Internet, considerando que este fenômeno vem se espalhando por todo o contexto mundial e por sua nocividade deve ser suprimido. Cumpre registrar que na atualidade a Inteligência Artificial tem sido empregada juntamente com a força humana no combate virtual a esses rumores através de ferramentas específicas, as quais tem se mostrado extremamente promissoras, especificamente empregando o processamento de linguagem natural, identificando e classificando notícias e informações como confiáveis ou não, as quais se espalham pelas redes sociais de forma vertiginosa.

Sendo assim, pretende-se mergulhar na busca de dados e informações que comprovem essa eficácia e respondam ao seguinte problema da pesquisa: como a Inteligência Artificial pode contribuir efetivamente para o combate das *Fake News* que se proliferam na internet?

2 OBJETIVOS

2.1 Objetivo Geral

Estudar como a Inteligência Artificial pode contribuir efetivamente para o combate das “*Fake News*” que se proliferam na internet e propor uma solução de software.

2.2 Objetivos Específicos

Conceituar o termo *Fake News* apresentando sua evolução histórica;

Apresentar o conceito de Inteligência Artificial, assim como de sua subárea Processamento de Linguagem Natural;

Descrever a importância da *Machine Learning* (Aprendizagem de Máquina), abordando como este campo de estudo combina aplicações de reconhecimento de modelos e Inteligência Artificial;

Conceituar o que é um Chatbot e sua evolução histórica;

Apresentar ferramentas oferecidas por grandes corporações para criação de chatbots;

Apresentar um protótipo de chatbot, utilizando as ferramentas apresentadas.

3 JUSTIFICATIVA

A relevância deste estudo justifica-se em decorrência do momento de transição que estamos perpassando, onde a internet e as informações que nela circulam, estão se tornando fatores determinantes nas tomadas de decisões no mundo real.

Por abordar estudos teóricos que apontem metodologias que demonstram como o uso da Inteligência Artificial pode combater a incidência de “*fake news*”. Tendo um forte apelo social, tal que uma vez implementadas as soluções aqui propostas, serão fortes ferramentas para que as pessoas tenham confiança nas informações que consomem.

4 HIPOTESE

A propagação de informação falsas através de meios de comunicação de acesso diário, tais como redes sociais, páginas de notícias e jornais online, necessita ser combatida e a Inteligência Artificial tem se mostrado uma ferramenta computacional capaz de fornecer segurança acerca da confiabilidade do conteúdo online, tendo em vista que confronta informações de milhares de fontes diferentes para se chegar a um veredito relativamente confiável.

5 REFERENCIAL TEÓRICO

5.1 Fake News

Segundo o Dicionário de Cambridge, o conceito *fake news* indica histórias falsas que, ao manterem a aparência de notícias jornalísticas, são disseminadas pela Internet (ou por outras mídias), sendo normalmente criadas para influenciar posições políticas, ou como piadas¹. Com efeito, as fakes news correspondem a uma espécie de “imprensa marrom” (ou *yellow journalism*), deliberadamente veiculando conteúdos falsos, sempre com a intenção de obter algum tipo de vantagem, seja financeira (mediante receitas oriundas de anúncios), política ou eleitoral.

A exemplo do poder exercido pelas Fake News na percepção da verdade, observou-se que, na semana que antecedeu a votação da abertura do processo de Impeachment do ex-presidente Dilma Rousseff, três das cinco notícias mais compartilhadas no nas redes sociais eram falsas (LAVARDA et al., 2017).

Os três Fake News convertidos em notícias foram:

- “Polícia Federal quer saber os motivos para Dilma doar R\$ 30 bilhões a Friboi” do site Pensa Brasil. Notícia com 90.150 mil compartilhamentos;
- “Presidente regional do PDT ordena que militância pró-Dilma vá armada no domingo: Atirar para matar” do site Diário do Brasil. Notícia com 65.737 mil compartilhamentos;
- “Lula deixa Brasília às pressas ao saber da nova fase da Lava-Jato. Seria um mandato de prisão?” do site Diário do Brasil. Notícia com 58.601 mil compartilhamentos.

As três notícias chamam atenção por seu cunho diretamente político, voltado a confundir a opinião pública em um momento marcante da história do país, direcionando a opinião popular para que se manifestasse a favor do impeachment da ex-chefe de estado.

¹ Cambridge University Press. Fake news. Disponível em: <https://dictionary.cambridge.org/english/fake-news>. Acesso em: 06 nov. 2018.

5.2 Inteligência Artificial

Inteligência Artificial é uma ciência jovem, que começou a ser explorada logo após a segunda guerra mundial. Porém suas raízes são profundamente fincadas na história, tendo em vista que psicólogos, filósofos, biólogos, cientistas e matemáticos já tentam decifrar o que é a inteligência, ou racionalidade:

“Poder-se-ia certamente definir inteligência pelas propriedades que ela exhibe: uma capacidade de lidar com novas situações; a capacidade de solucionar problemas, de responder a questões, de engendrar planos e assim por diante (COPPIN 2010)”

Uma vez descrito o conceito de inteligência, cabe definir o do que se trata IA, ciência esta que, conforme (RUSSEL e NORVING 2004. p. 3) “não só procura entender, mas também construir entidades inteligentes”. Para definir algo considerado mais complexo que a própria inteligência, é necessário misturar várias opiniões de autores renomados, segundo (JUNIOR e YONEYAMA 2000.p.2):

“IA é o estudo das faculdades mentais através do uso de modelos computacionais (apud Charniak e Mcdermott, 1985)”

“IA é o estudo de como fazer os computadores realizarem tarefas que, no momento, são feitas melhor por pessoas (apud Rich, 1983)”

“IA é o estudo das ideias que permitem habilitar os computadores a fazerem coisas que tornam as pessoas inteligentes (apud Winston, 1977)”

“IA é o campo de conhecimento onde se estuda sistemas capazes de produzir algumas das atividades mentais humanas (apud Nilsson, 1986)”

Observa-se que as definições de IA envolvem termos como *faculdades mentais*, *tarefas mais bem-feitas por pessoas*, *coisas que tornam as pessoas inteligentes* e *atividades mentais humanas*. Com tantas referências ao comportamento humano, torna-se interessante considerar os aspectos psicológicos da inteligência humana, ainda que o objetivo final não seja reproduzi-los.

Segundo (JUNIOR; YONEYAMA, 2000 p.8 apud Paim, 1986; Ey, Bernard e Brisset, 1985) as atividades humanas psíquicas fundamentais são:

Sensação: Fenômeno elementar resultante de estímulos mecânicos, físicos, químicos ou elétricos sobre o organismo. As sensações podem ser externas ou internas. Um

exemplo de alteração da sensação é a anestesia, ou seja, ausência da condução de sensações pelo feixe nervoso.

Percepção: é a tomada de conhecimento de um objeto exterior considerado real. As sensações são integradas de modo que as condições internas ou externas do organismo sejam conhecidas.

Representação: é a organização das imagens de objetos. Corresponde ao processo de modelamento em Engenharia. As imagens visuais poderiam ser comparadas a fotografias que representam, no papel, algumas características do objeto fotografado.

Conceituação: construção simbólica que busca captar a essência dos objetos, agrupando-os em classes. É uma das operações intelectuais básicas: conceituação, juízo e raciocínio. Por exemplo, o conceito de aves coloca na mesma classe: águia, canário, pato, papagaio, tucano, avestruz e muitos outros animais. Assim, canário é bem diferente de águia, mas existem características em comum entre eles que permitem agrupá-los na mesma classe.

Juízo: é a capacidade de exprimir os vínculos e as relações entre os fatos e os objetos da natureza. Em termos de engenharia corresponde à capacidade de qualificar os fenômenos e os objetos do sistema segundo as leis físicas específicas.

Raciocínio: é a concatenação disciplinada dos juízos, de modo a gerar novos juízos. Em Engenharia corresponde à capacidade de dedução: Por exemplo, das afirmações “A Terra é azul” e “Azul é a cor do mar”, o raciocínio permite obter: “A Terra é da cor do mar”.

Memória: é a capacidade de armazenamento de informações para utilização posterior.

Atenção: é a capacidade de concentrar a atividade psíquica sobre um dado estímulo que a solicita. Por exemplo, um Sistema de Visão Ativa por Computador pode concentrar a sua atenção sobre um determinado alvo a ser atacado.

Consciência: é um complexo de fenômenos psíquicos que permite conhecer a si próprio e ao mundo.

Orientação: é a capacidade do indivíduo de ter consciência de sua situação temporal e espacial em relação ao meio.

Afetividade: é a capacidade de experimentar sentimentos ou emoções, correspondendo à valorização qualitativa de determinados estados do indivíduo.

Volição: é o elemento psíquico que leva a iniciar atividades voluntárias. São exemplos de alterações, o estupor e os tiques;

Linguagem: é o mecanismo que permite a expressão simbólica. Tipicamente são verbais ou gráficas.

Sistemas que empregam AI tentam imitar e integrar diversas atividades psíquicas de forma simultânea. Os autores (JUNIOR e YONEYAMA 2000) exemplificam que um veículo autoguiado pode ser dotado de sistemas de sensopercepção, mecanismo para representação do ambiente, bem como bancos de conhecimentos e dados, de forma a realizar inferências de forma autônoma e navegar de um pronto a outro, sem colidir com obstáculos no ambiente onde se encontra.

Para que um sistema seja capaz de usar as combinações de habilidades supracitadas para demonstrar algum tipo de comportamento, é necessário que esse sistema aprenda como fazê-lo, para tal, o estudo do aprendizado deve ser explorado, para (JUNIOR e YONEYAMA 2000. p. 3) “o aprendizado é a aquisição de conceitos e de conhecimentos estruturados”.

5.3 Machine Learning

Machine Learning (em português Aprendizado de Máquina), é uma característica de um dispositivo tecnológico que possui Inteligência Artificial (IA). Inteligência Artificial pode ser categorizada como geral ou estreita. IA geral apresenta características de inteligência humana e capacidades de planejamento, reconhecimento de objetos, sons, aprendizado e resolução de problemas. IA estreita são os dispositivos que apresentam alguma inteligência na área de atuação para que foram desenvolvidos, porém em outras áreas apresentam deficiência. Nesse ponto entra o aprendizado de máquina, que tem como objetivo tornar a inteligência do dispositivo ou objeto tecnológico o mais humano possível, seja através da comunicação ou da tomada de decisão e resolução de problemas.

Como exemplo, podemos considerar o aprendizado de máquina através da observação de imagens ou vídeos. O ser humano poderia analisar, centenas de imagens que tem um cachorro ou não e saberia diferenciar cada uma delas afirmando qual imagem é de um cachorro e qual não é. Nesse mesmo caso, um algoritmo pode construir um modelo que pode indicar com precisão uma imagem que contém um cachorro ou não. Se o modelo for bem preciso, a máquina aprende como é um cachorro.

5.3.1 Deep Learning

Deep Learning (em português Aprendizado Profundo) é uma das várias abordagens do aprendizado de máquina. Essa tecnologia emergente diz respeito a um aprendizado profundo utilizando redes neurais para melhorar o reconhecimento de sons, imagens, visão computacional e processamento de linguagem natural. (SILVA, 2016) explica que:

“O diferencial tecnológico dessa abordagem está nos excelentes resultados obtidos na resolução de problemas, resultados esses que superam até mesmo o desempenho dos melhores especialistas em determinadas áreas de conhecimento, como, por exemplo, o reconhecimento de localidades e características semânticas em imagens, a vitória em jogos de estratégia mais complexos do que o xadrez e a superação de seres humanos em testes psicométricos de compreensão verbal.”

Comprovado a eficiência superior dessa ferramenta, aplicações como Google Translate e Cortana (Windows) utilizam o aprendizado profundo como base para seu funcionamento e aprendizado.

5.4 Processamento de Linguagem Natural

Processamento de Linguagem Natural ou PLN, é a tecnologia na área da Inteligência Artificial que permite a compreensão, interpretação, manipulação e realização de tarefas que estão dispostas em alguma língua natural. PLN vem sendo estudado e desenvolvido a muitos anos, pois sempre houve o interesse na comunicação homem-máquina. Mas atualmente, essa ciência tomou grandes proporções em sua área de estudo pois o avanço tecnológico no geral, está muito rápido, permitindo alto poder de processamento e algoritmos aprimorados, além de tecnologias como big data que se alia ao desenvolvimento do PLN.

5.4.1 História PLN

O primeiro registro que verificamos na história do PLN, se dá na década de 1940 quando foi desenvolvida a *Machine Translator* (Máquina de Tradução) que tinha o objetivo de diferenciar idiomas e encontrar problemas de sintaxe em textos. Essa máquina utilizava como base dicionários e conseguia fazer algumas traduções simples. Como cita (PERNA et al, 2010) em 1957 Chomsky desenvolveu trabalhos relevantes sobre o tema. Segundo (PERNA et al., 2010) “Um trabalho particularmente relevante desta época é o livro *Syntactic Structures* (CHOMSKY, 1957)” que introduziu a gramática gerativa, a partir daí ficou mais clara como a área de linguística poderia auxiliar a área de tradução automática.

Segundo (PERNA et al., 2010), a partir desse período, houve uma multiplicação dos estudos sobre PLN com o estabelecimento de diversas subáreas que vem sendo pesquisadas até hoje. Estas áreas se dedicam a assuntos tão variados quanto à categorização de textos e à extração de informações, passando pelos tradicionais temas de tradução automática e sistemas de diálogo. Os trabalhos desenvolvidos neste período podem, segundo os autores (JURAFSKY e MARTIN, 2009), ser agrupados em quatro grupos segundo os paradigmas utilizados: os métodos estocásticos, os métodos baseados em lógica, os métodos de entendimento de linguagem natural, e os métodos de modelagem de discurso.

5.4.2 Funcionamento

PLN busca tratar de modo computacional alguns aspectos humanos, como sons, palavras, discursos e sentenças. Nesse processo é necessário considerar estruturas e significados, formatos e referências, contextos e usos para que se possa estabelecer um resultado satisfatório. Sabendo que o PLN visa estabelecer essa comunicação via linguagem humana, vale ressaltar as várias formas que se pode estabelecer uma comunicação na interface homem-máquina. Como exemplificado na tabela 1, podemos classificar da seguinte forma:

Tabela 1 – Formas de comunicação PLN

Forma	Descrição
Fonético fonológico	Do relacionamento das palavras com os sons que produzem.
Morfológico	Da construção das palavras a partir unidades de significado primitivas e de como classificá-las em categorias morfológicas.
Sintático	Do relacionamento das palavras entre si, cada uma assumindo seu papel estrutural nas frases, e de como as frases podem ser partes de outras, constituindo sentenças.
Semântico	Do relacionamento das palavras com seus significados e de como eles são combinados para formar os significados das sentenças.
Pragmático	Do uso de frases e sentenças em diferentes contextos, afetando o significado.

Fonte: <https://www.ime.usp.br/~slago/IA-pln.pdf>, Acesso em: 14 mai. 2019.

Segundo Pereira, “a fonologia está relacionada ao reconhecimento dos sons que compõem as palavras de uma língua. A morfologia reconhece as palavras em termos das unidades primitivas que a compõem (e.g. *ca,cou* → *ca,c+ou*). A sintaxe define a estrutura de uma frase, com base na forma como as palavras se relacionam nessa frase (figura 1). A semântica associa significado a uma estrutura sintática, em termos dos significados das palavras que a compõem (e.g. à estrutura da figura 1, podemos associar o significado “um animal perseguiu/capturou outro animal”). Finalmente, a pragmática verifica se o significado associado à uma estrutura sintática é realmente o significado mais apropriado no contexto considerado (e.g. no contexto predador-presa, “perseguiu/capturou” → “comeu”).



Figura 1 – Árvore Sintática

Fonte: <https://www.ime.usp.br/~slago/IA-pln.pdf>, Acesso em: 14 mai. 2019.

5.4.3 Forma Lógica

Quando estabelecemos uma comunicação utilizamos a estrutura gramatical que padroniza nossa fala ou escrita. Como visto no tópico anterior, existe essa necessidade no PLN, já que o objetivo final é exatamente estabelecer essa comunicação em linguagem natural. Então depois de entender que as regras gramaticais se aplicam a esse meio, precisamos entender como essa linguagem humana entra em sua forma lógica. Segundo (Gonzales et al., 2003) “A representação do significado de uma sentença, independente de contexto, é obtida através de sua forma lógica (apud Allen1995, Franconi2001) (Figura 2). A forma lógica codifica os possíveis sentidos de cada palavra e identifica os relacionamentos semânticos entre palavras e frases. Uma vez que os relacionamentos semânticos são determinados, alguns sentidos para as palavras tornam-se inviáveis e, assim, podem ser desconsiderados.

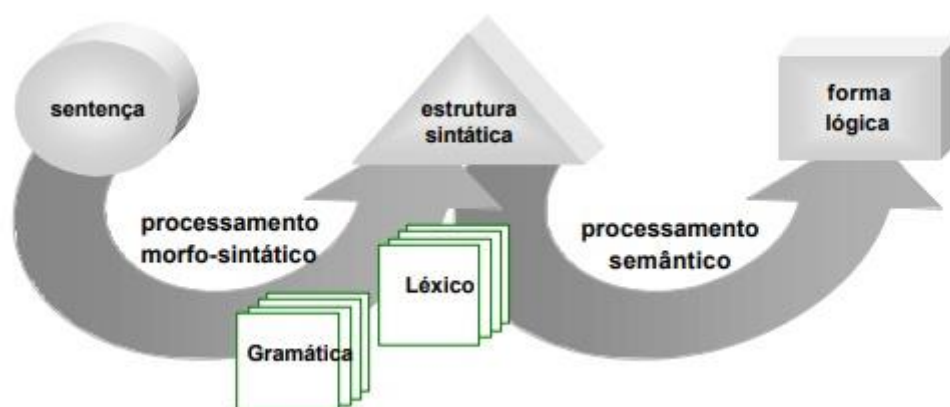


Figura 2 – Transformações da sentença na estrutura sintática e na forma lógica

Fonte: GONZALEZ, Marco; LIMA, Vera L. S. De. *Recuperação de informação e processamento da linguagem natural*.

Para transformação de uma sentença em forma lógica, a gramática e o léxico são recursos indispensáveis. A gramática é responsável por definir através de suas regras a estrutura sintática de uma sentença. O léxico é responsável por trazer informações como um dicionário acerca das palavras, não se resumindo apenas a um significado, mas todo um conteúdo acerca das palavras.

5.5 Sistemas Especialistas

Sistemas especialistas são criados a fim de servirem como consultores altamente qualificados, em uma determinada área do conhecimento. Devem, portanto, fornecer soluções, condutas, ações e sugestões ou outras informações úteis, obtidas a partir de dados fornecidos e de conhecimentos armazenados previamente.

Um sistema especialista pode ser utilizado para auxiliar médicos, vendedores, operadores de máquinas ou, como proposto nesta pesquisa, pessoas que desejam ter confiança de que a informação consumida é de fato verdade. Enfim, sistemas especialistas permitem que conhecimentos possam ser disseminados e utilizados por muitas pessoas, sem a necessidade da presença do especialista. Permite, ainda, que conhecimentos de vários especialistas possam ser reunidos em uma só base de dados.

5.6 Chatbots

Chatbots são software capazes de se comunicar com humanos, usando linguagem natural, idealmente, conforme proposto por Alan Turing em 1950 o humano não deve perceber que está falando com uma máquina também podem ser definidos como softwares que buscam interagir com as pessoas de uma forma humana e realizar tarefas dentro de uma determinada área do conhecimento (CALADO, 2016).

Essa tecnologia se popularizou e tem sido amplamente utilizada para automações de atendimentos bancários, logísticos e medicinais, permitindo que os usuários realizem tarefas como compras, ou marcação de consultas, diretamente pelo chatbot.

A gama de possibilidades é tamanha, que com o advento da inteligência artificial as suas aplicações se tornam ainda maiores.

6 MATERIAS E MÉTODOS

6.1 Aplicação

Nesta seção são apresentados os componentes da aplicação proposta, e descritos os processos coleta de dados, normatização, criação do banco de dados, criação da API e do Chatbot. A figura 3 mostra a estrutura da aplicação, as interações ocorrem em linguagem natural através do chatbot, são recebidas pela API e retornam ao usuário o resultado.

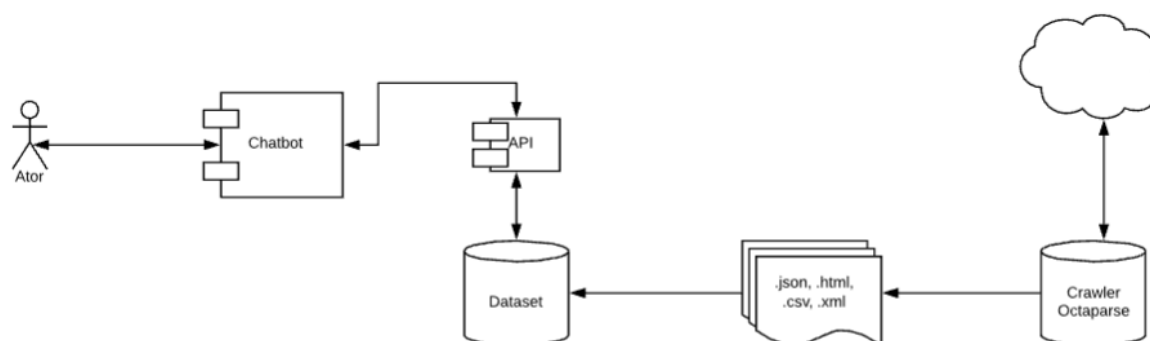


Figura 3 – estrutura da aplicação.

Fonte: o autor

6.2 Coleta de Dados – Web Crawler

A fim reunir os dados necessários para a análise, foi criado um mecanismo de *Crawler* para varrer o site “aosfatos.org” recolhendo conteúdo das notícias, e guardando de forma estruturada em arquivos (.json, .html, .csv e .xml.) para alimentar o dataset.

Após obter os arquivos, os dados precisam ser normalizados, retirando acentos, caracteres especiais, aspas e palavras reservadas. Após a normalização, os dados poderão ser incluídos no dataset.

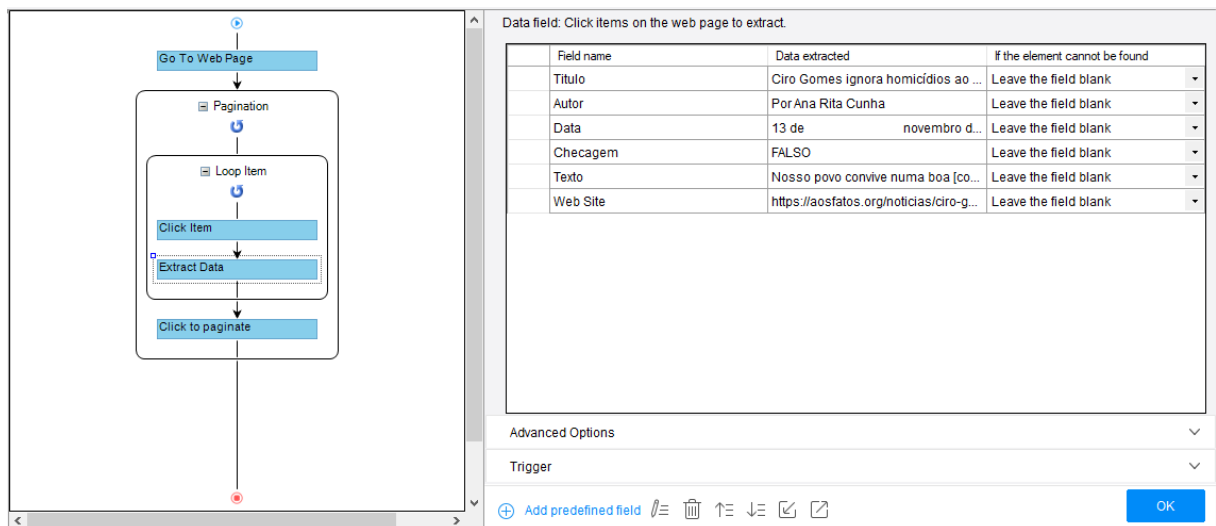


Figura 4 – Estrutura do crawler octoparse.

Fonte: <https://www.octoparse.com/>

6.3 Bando de dados – Dataset e PostgREST API

O PostgREST é um webservice independente que transforma o banco de dados PostgreSQL diretamente em uma API RESTful.

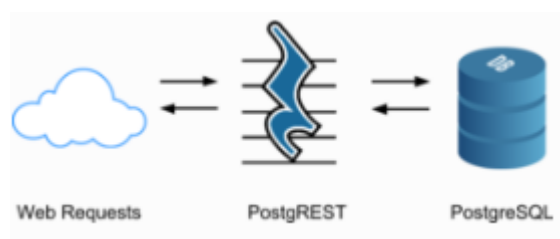


Figura 5 – Funcionamento do PostgREST.

Fonte: <https://postgrest.org/>

O PostgREST é distribuído com versões compiladas para as principais distribuições de Linux, BSD e Windows. Os binários pré-criados para download são arquivos compactados .tar.xz.

6.3.1 API

Uma API trata-se de um conjunto de rotinas e padrões estabelecidos e documentados por uma aplicação X, para que outras aplicações consigam utilizar as funcionalidades desta aplicação X. Desta forma entendendo que uma API permite a interoperabilidade entre aplicações, neste contexto a API será usada para que o chatbot possa fazer consultas no dataset de notícias.

Com o postgREST instalado podemos criar um schema no banco de dados, denominado API.

```
-- Schema: api

-- DROP SCHEMA api;

CREATE SCHEMA api
  AUTHORIZATION postgres;

GRANT ALL ON SCHEMA api TO postgres;
GRANT ALL ON SCHEMA api TO web_anon;
```

Figura 6 – Schema api.

Fonte: o autor

A API tem um por padrão um ponto de extremidade, /todos, que virá de uma tabela.

```
create table api . todos (
  id serial primary key ,
  done boolean not null default false ,
  task text not null ,
  due timestamptz
);

insert into api . todos ( task ) values
( 'finish tutorial 0' ), ( 'pat self on back' );
```

Figura 7 – Tabela api.

Fonte: o autor

Em seguida, será necessário criar uma função para solicitações anônimas da web. Quando uma solicitação é recebida, o PostgREST transforma essa solicitação em uma consulta no banco de dados.

```
create role web_anon nologin ;

grant usage on schema api to web_anon ;
grant select on api . todos to web_anon ;
```

Figura 8 – Função anon.

Fonte: o autor

A função web_anon tem permissão para acessar itens no schema api e ler linhas na tabela todos.

É uma boa prática criar uma função dedicada à conexão com o banco de dados, em vez usar a função do Postgres altamente privilegiada. Portanto, faremos isso, nomear o authenticator função e permitir que ele mude para a função web_anon:

```
create role authenticator noinherit login password 'mysecretpassword' ;
grant web_anon to authenticator ;
```

Figura 9 – Função authenticator.

Fonte: o autor

Com as configurações acima, a API está pronta para receber requisições da web. Criamos a função que irá consultar o dataset, e retornar os dados referente a notícia pesquisada.

```
-- Function: api.select_fakenews(json)
-- DROP FUNCTION api.select_fakenews(json);

CREATE OR REPLACE FUNCTION api.select_fakenews(param json)
  RETURNS json AS
$BODY$
--Função que recebe um parâmetro do tipo Json, realiza uma consulta e retorna os dados no formato json.
begin
  return (SELECT json_agg(r.*) as retorno
    from (
      select noticia_id,
             titulo,
             autor,
             data,
             checagem,
             texto,
             link
      from public.noticias
      where titulo = (param ->> 'titulo')::varchar
    ) as r);
end;
$BODY$
LANGUAGE plpgsql VOLATILE SECURITY DEFINER
COST 100;
ALTER FUNCTION api.select_fakenews(json)
  OWNER TO postgres;
```

Figura 10 – Função select_fakenews.

Fonte: o autor

6.3.2 Banco de Dados

O banco escolhido para receber o dataset foi o PostgreSQL versão 11 para sistemas Linux. O PostgreSQL é um sistema de gerenciamento de banco de dados objeto-relacional baseado no POSTGRES Versão 4.2 desenvolvido pelo Departamento de Ciência da Computação da Universidade da Califórnia em Berkeley. O POSTGRES foi pioneiro em vários conceitos que somente se tornaram disponíveis muito mais tarde em alguns sistemas de banco de dados comerciais.

```

-- Database: fakenews

-- DROP DATABASE fakenews;

CREATE DATABASE fakenews
  WITH OWNER = postgres
       ENCODING = 'UTF8'
       TABLESPACE = pg_default
       LC_COLLATE = 'C.UTF-8'
       LC_CTYPE = 'C.UTF-8'
       CONNECTION LIMIT = -1;

```

Figura 11 – Instanciando o banco.

Fonte: autor

```

-- Table: public.noticias

-- DROP TABLE public.noticias;

CREATE TABLE public.noticias
(
    noticia_id integer NOT NULL DEFAULT nextval('noticias_noticia_id_seq'::regclass),
    titulo character varying(1024),
    autor character varying(1024),
    data date,
    checagem character varying(255),
    texto character varying(1024),
    link character varying(1024),
    CONSTRAINT noticias_pk PRIMARY KEY (noticia_id)
)
WITH (
    OIDS=FALSE
);
ALTER TABLE public.noticias
  OWNER TO postgres;

-- Index: public.noticias_noticia_id_uindex

-- DROP INDEX public.noticias_noticia_id_uindex;

CREATE UNIQUE INDEX noticias_noticia_id_uindex
  ON public.noticias
  USING btree
  (noticia_id);

```

Figura 12 – Criando as tabelas.

Fonte: autor

```

INSERT INTO public.noticias(titulo, autor, data, checagem, texto, link )
VALUES ('Titulo da noticia','data','resultado da checagem','texto checado','link da fonte')

```

Figura 13 – Insert para povoar as colunas.

Fonte: autor

Para garantir a alta disponibilidade para a aplicação foi usada uma instancia t2.micro no serviço de EC2 da Amazon Web Services (AWS). O Amazon Elastic Compute Cloud (Amazon EC2) oferece uma capacidade de computação dimensionável na nuvem da Amazon Web Services (AWS). O uso do Amazon EC2 elimina a necessidade de investir em hardware inicialmente, portanto, é pode desenvolver e implantar aplicativos com mais rapidez. Usando o Amazon EC2 para executar o número de servidores virtuais que precisar, configurar a segurança e a rede, e gerenciar o armazenamento. O Amazon EC2 também permite a expansão ou a redução para gerenciar as alterações de requisitos ou picos de popularidade, reduzindo, assim, a sua necessidade de prever o tráfego do servidor.

Instance: i-0b7348d418cf3d558 Elastic IP:	
Description Status Checks Monitoring Tags	
Instance ID	i-0b7348d418cf3d558
Instance state	running
Instance type	t2.micro
Elastic IPs	
Availability zone	us-east-1d
Security groups	launch-wizard-2 , view inbound rules , view outbound rules
Scheduled events	No scheduled events
AMI ID	ubuntu/images/hvm-ssd/ubuntu-bionic-18.04-amd64-server-20191002 (ami-04b9e92b5572fa0d1)
Platform	-
IAM role	-
Key pair name	service_ubuntu
Owner	265326193946
Launch time	November 11, 2019 at 1:41:24 PM UTC-3 (26 hours)
Termination protection	False
Public DNS (IPv4)	ec2-3-233-169-170.compute
IPv4 Public IP	
IPv6 IPs	-
Private DNS	ip-172-31-95-3.ec2.internal
Private IPs	172.31.95.3
Secondary private IPs	
VPC ID	vpc-56e4fe2d
Subnet ID	subnet-65d9004b
Network interfaces	eth0
Source/dest. check	True 
T2/T3 Unlimited	Disabled
EBS-optimized	False
Root device type	ebs
Root device	/dev/sda1

Figura 14 – Especificações da instancia.

Fonte: <https://aws.amazon.com/pt/ec2/instance-types/t2/>

7 FACEBOOK CHAT

Optou-se pelo Facebook como ferramenta para o desenvolvimento do chatbot, a linguagem Python foi juntamente com as bibliotecas Flask, Requests e Gunicorn utilizada para programar o backend da aplicação.

```
chatbot_facebook-master > index.py > {} json
1 import os
2 import requests
3 import traceback
4 import json
5 from flask import Flask, request
```

Figura 15 – Bibliotecas usadas.

Fonte: o autor.

Para integração com o Facebook é necessário um token, este será a ponte entre o backend e a interface do chatbot.

Geração de token

É necessário o token da Página para começar a usar as APIs. O token desta Página terá todas as permissões do Messenger mesmo se o seu aplicativo ainda não estiver aprovado para usá-las, contudo, neste caso, você poderá enviar mensagens apenas para os administradores do aplicativo. Você também pode gerar tokens para Páginas que não são suas usando o Login do Facebook.

Página

Token de acesso da Página

Selecionar uma página ▼

You must select a Page to generate an access token.

[Criar uma nova Página](#)

Figura 16 – Gerando token.

Fonte: facebook.com

Uma vez gerado, o token será atribuído a uma variável local de ambiente, a mesma será declarada em uma função.

```
chatbot_facebook-master > .env
1  FB_ACCESS_TOKEN = EAANlejHhYFQBAJovkZBH
2  FB_VERIFY_TOKEN = verify

#Token de parametrização com a API do Facebook Messenger
token = os.environ.get('FB_ACCESS_TOKEN')

app = Flask(__name__)
```

Figura 17 – Criando e declarando a variável de ambiente.

Fonte: o autor

Para que o facebook saiba para onde enviar as requisições feitas no chat é necessário criar um webhook, que nada mais é que o retorno de uma chamada HTTP.

Webhooks

Para receber mensagens e outros eventos enviados pelos usuários do Messenger, o aplicativo precisa ativar a integração de webhooks.

URL de retorno de chamada	Verificar token
https://salty-lowlands-30789.herokuapp.com/	

As solicitações de validação e notificações de webhook para este objeto serão enviadas para esta URL.

Token que o Facebook enviará de volta para você como parte da verificação da URL de retorno de chamada.

Figura 17 – Configurando o webhook.

Fonte: facebook.com

Para ler o que o usuário enviou, foi verificado os dados enviados na requisição que o chat envia via POST.

```
def webhook():
    if request.method == 'POST':
        try:
            #Valores recebidos pela função POST (Enviada pelo usuário do chat no Messenger)
            data = json.loads(request.data.decode())

            sender = data['entry'][0]['messaging'][0]['sender']['id']

            print(data)
```

Figura 18 – Método POST.

Fonte: o autor

Com o `data = json.loads(request.data.decode())` é capturado o corpo da mensagem que foi enviado pelo chat.

Com o trecho `text = data['entry'][0]['messaging'][0]['message']['text']` é obtido o texto que o usuário enviou via chat.

Pra saber quem enviou, usamos o trecho `sender=data['entry'][0]['messaging'][0]['sender']['id']`. Essa informação é necessária para que possamos enviar uma resposta para o usuário correto.

Uma vez lida a mensagem do usuário a função `check_fakenews` faz a requisição na API do Dataset e e retornar os valores da notícia pesquisada.

```
def check_fakenews(sender, received_text):
    url_endpoint = 'http://3.233.169.170:3000/rpc/select_fakenews'

    headers = {
        'Content-Type': 'application/json',
        'Prefer': 'params=single-object'
    }

    data = {'titulo' : received_text}

    r = requests.post(url = url_endpoint, json = data, headers=headers)
    json_data = json.loads(r.text)
    return json_data[0]['checagem']
```

Figura 19 – Função `check_fakenews`.

Fonte: o autor

Com o retorno da função `check_fakenews` a função `send_text` monta o json de retorno para a API do facebook.

```
def send_text(sender, text):  
    return {  
        "recipient": {  
            "id": sender  
        },  
        "message": {  
            "text": text  
        }  
    }
```

Figura 20 – Função `send_text`.

Fonte: o autor

RESULTADOS

Após a escolha das ferramentas a serem utilizadas e criação da estrutura necessária para receber os títulos das notícias através do Facebook Messenger, iniciou-se a programação do backend da aplicação, após todo o cenário montado foram realizados os testes de consultas no dataset.

Embora poucos testes tenham sido realizados, os resultados demonstram que a estrutura de checagem é capaz de identificar trechos de notícias e retornar o seu título, autor, se é verdadeira ou falsa e o link para a notícia.

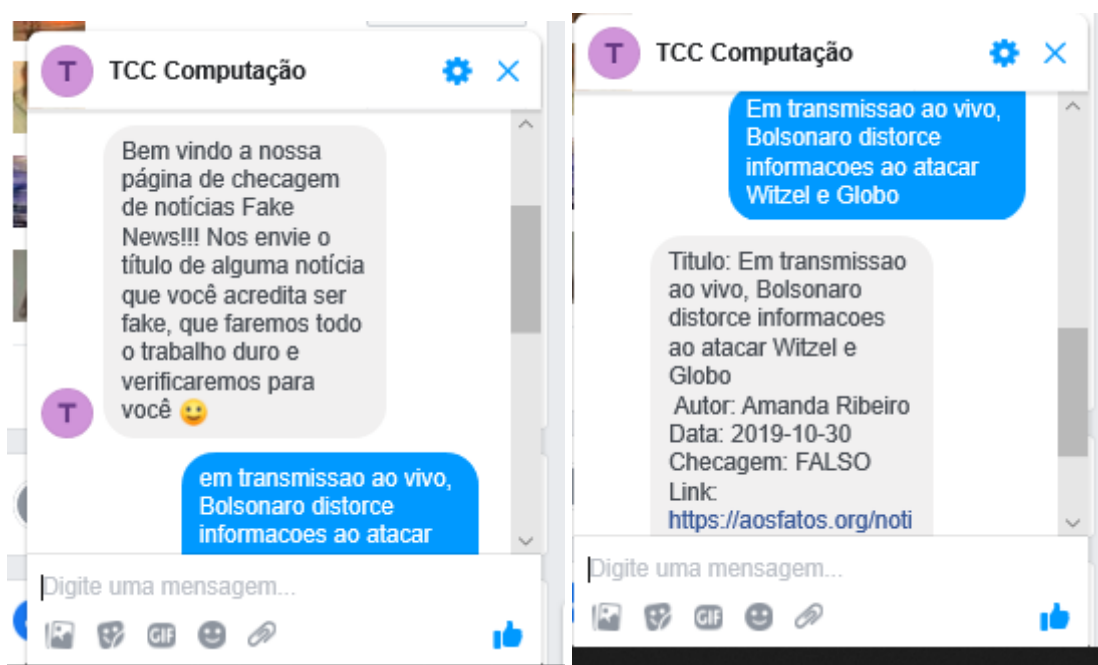


Figura 21 – Funcionamento do chatbot.

Fonte: <https://www.facebook.com/TCC-Computação>

CONCLUSÃO

Pode-se afirmar que todo fundamento teórico e a construção prática realizado neste trabalho, podem em muito contribuir para com a sociedade. Com intuito de desenvolver uma aplicação experimental capaz de conter o avanço das fake news na internet, levantou-se todo o material teórico e as ferramentas necessárias para iniciar a concepção da aplicação.

Utilizando ferramentas de SGBD, linguagem de programação Python e suas bibliotecas, foi possível criar um sistema especialista, que em linguagem natural verifica em um dataset pré-estabelecido se uma notícia é verídica.

O potencial de impacto na sociedade desse estudo é incalculável, pois uma vez implementado, e aperfeiçoado por estudos futuros e revisões, devolverá ao leitor a confiabilidade na informação por ele consumida, permitindo que suas tomadas de decisões na realidade sejam pautadas em fatos.

REFERENCIAS BIBLIOGRÁFICAS

ALLEN, J. **Natural Language Understanding**. Redwood City, CA: The Benjamin/Cummings Pub. Co., 1995. 654 p.

Cambridge University Press. **Fake news**. Disponível em: <https://dictionary.cambridge.org/english/fake-news>. Acesso em: 04 abr. 2019.

FRANCONI, Eurico. **Description Logics for Natural Language Processing**. In: Baader, F.; McGuinness, D. L.; Nardi, D.; Patel-Schneider, P. P. (editores) **Description Logics Handbook**. Cambridge: Cambridge University Press, 2001. Cap.18.

GONZALEZ, Marco; LIMA, Vera L. S. De. **Recuperação de informação e processamento da linguagem natural**. Google artigo acadêmicos, Disponível em: <https://s3.amazonaws.com/academia.edu.documents/43022679/minicurso-jaia2003.pdf?awsaccesskeyid=akiaiwowyygz2y53ul3a&expires=1557975816&signature=hsmarpo2ahyrj9o%2fj3swzan%2b5ck%3d&response-content-disposition=inline%3b%20filename%3drecuperacao_de_informacao_e_processament.pdf>. Acesso em: 14 mai. 2019.

JUNIOR, Cairo L. Nascimento; YONEYAMA, Takashi. **INTELIGENCIA ARTIFICIAL EM CONTROLE E AUTOMAÇÃO**. São Paulo: Blucher, 2000.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing**, 2nd ed. Prentice-Hall, 2009.

LULA deixa Brasília às pressas ao saber de nova fase da lava -jato: seria um mandado de prisão? 13 abr. 2016. Fanpage Diário do Brasil. Disponível em: <<https://www.facebook.com/DiariodoBrasil/posts/1191224050942217>>. Acesso em: 11 mai.2019.

PEREIRA, Silvio Do Lago. **Processamento de linguagem natural**. Disponível em: <<https://www.ime.usp.br/~slago/IA-pln.pdf>>. Acesso em: 14 mai. 2019.

PERNA, Cristina Lopes; DELGADO, Heloísa Koch; FINATTO, Maria José. **Linguagens especializadas em corpora: Modos de dizer e Interfaces de Pesquisa**. Porto Alegre: EdPUCRS, 2010.

POLÍCIA federal quer saber os motivos para Dilma doar R\$ 30 bilhões a Friboi. 13 abr. 2016.Fanpage pensa Brasil. Disponível em:

<<https://www.facebook.com/PensaBrasil/posts/1786948538202591>>. Acesso em: 11 maio 2019.

PRESIDENTE regional do PDT ordena que militância pró-Dilma vá armada no domingo: “atirar para matar”. 14 abr. 2016. Fanpage Diário do Brasil. Disponível em: <<https://www.facebook.com/DiariodoBrasil/posts/1191902627541026>>. Acesso em: 11 mai.2019.

SAINT-Dizier., P. **On the Polymorphic Behavior of Word-senses.** In: Linguística Computacional: Investigação Fundamental e Aplicações. Lisboa: Edições Colibri, 1999. p.29-56.

SANCHEZ, L. **Comunicação Organizacional Interna e o Ambiente Virtual: Novas Tendências. Tese (Doutorado) - Comunicação Social,** Universidade Metodista de São Paulo, São Bernardo do Campo, São Paulo, 2006.

SANCHOTENE, C.; MACHADO DA SILVEIRA, A. C.; DE LIMA LAVARDA, S. **Quando as notícias mais compartilhadas são falsas: a circulação de boatos durante a semana do impeachment no Facebook.** *Comunicação & Informação*, v. 20, n. 3, p. 99-112, 31 dez. 2017. Acesso em 11 mai. 2019

SILVA, Luís André Dutra E. **Utilização de Deep Learning em ações de controle.** *Revista do TCU*, [S.L], p. 135, jan./abr. 2016.

CALADO, Caio. Bots Brasil. O que é um chatbot? 2016. Disponível em: <<https://goo.gl/91wye4>>. Acesso em: 20 mai. 2019.