

Técnicas de Processamento de Linguagem Natural para Auxiliar o Estudante na Identificação das Pragas da Soja

Carolinne Roque e Faria¹, Cinthyan Renata Sachs C de Barbosa¹

¹Programa de Pós-Graduação em Ciência da Computação
Universidade Estadual de Londrina (UEL) – Londrina, PR - Brasil

carolinne.rf@outlook.com, cinthyan@uel.br

Abstract. *The world is going through technological revolutions, mainly in the agronomic field. These have also become interesting for academic purposes in this area, aiming to provide better solutions for the development and increase the productivity of the crop. Therefore, the present work aims to build dialogues and provide rich consultations through the use of Natural Language Processing techniques for the development of a conversational system in the identification of the main pests and their characteristics of the soybean crop stored in a repository for assisting Agricultural Science students in learning.*

Resumo. *O mundo está passando por revoluções tecnológicas, principalmente no campo agrônomo. Essas têm se tornado interessante também para fins acadêmicos nessa área, tendo como objetivo proporcionar melhores soluções para o desenvolvimento e aumentar a produtividade da cultivar. Assim, o presente trabalho tem o intuito de construir diálogos e proporcionar consultas ricas por meio da utilização de técnicas de Processamento de Linguagem Natural para o desenvolvimento de um sistema conversacional na identificação das principais pragas e suas características da cultura da soja armazenados em um repositório para auxiliar os estudantes de Ciências Agrárias no aprendizado.*

1. Introdução

Atualmente, existem inúmeros invasores que prejudicam a cultura da soja e que estão distribuídos em regiões produtoras, que em decorrência da falta de preparo em fazer uma análise do solo antes de realizar o plantio ou por fazer o uso de agrotóxicos de maneira desequilibrada ou aração incorreta, não rotacionam culturas e usam de forma irracional os produtos para o manejo das pragas.

De acordo com Chowdhury (2003), fazer uma análise visual da cultivar para detectar o vilão é ineficiente e difícil para plantações de grandes proporções, visto que o ataque acontece de forma inesperada e agressiva, além do procedimento fornecer uma acurácia relativamente baixa que requer um profissional qualificado e bem treinado para a realização de tal função [Pereira *et al.* 2018].

O gerenciamento de dados agrícolas depende de informações adquiridas de diversas tecnologias e sistemas que sejam capazes de auxiliar na tomada de decisão. Portanto, é fundamental mapear os dados (principais características de identificação de pragas e doenças na cultura da soja) e padronizá-los para colaborar com profissionais e estudantes da agronomia ou agronegócios, além de diminuir o impacto negativo no meio ambiente e aumentar a eficiência no controle dos patógenos.

Com o enorme fluxo de informações, essas tecnologias necessitam elaborar soluções para entender a linguagem e uma alternativa é o uso de Processamento de Linguagem Natural (PLN) aplicado para grandes volumes de dados. Essas técnicas pretendem analisar e extrair as informações de uma determinada cultura para proporcionar um diagnóstico com a finalidade de potencializar a produção dos agricultores a partir do aprendizado sobre as pragas da cultura da soja.

Para Jurafsky e Martin (2000), PLN tem como objetivo extrair textos em linguagem natural (LN) e executar tarefas relevantes, permitindo o diálogo entre homem e máquina, melhorando a comunicação humano-computador ou fazer processamento de texto ou fala (discurso).

Visto a complexidade das linguagens naturais, este trabalho tem o intuito de desenvolver um software inteligente e significativo aos estudantes e profissionais da área de ciências agrárias por meio do PLN para identificar as ameaças que afetam a cultura da soja analisada e auxiliá-los na tomada de decisão, como a aplicação de defensivos agrícolas sem prejudicar a qualidade do solo, contaminar o produto e fazer análise de custo. Assim, não ter que aprender uma linguagem de consulta específica para acesso a um repositório de banco de dados e poder dialogar em suas consultas permitindo o aprendizado é fundamental.

Este artigo está organizado da seguinte forma: a seção 2 apresenta os trabalhos relacionados; a seção 3 tem a fundamentação teórica sobre Processamento em Linguagem Natural; a seção 4 aborda a arquitetura geral do sistema; o desenvolvimento deste trabalho é mostrado na seção 5; por fim, a seção 6 traz as considerações finais.

2. Trabalhos relacionados

Devido à grande quantidade de informações armazenadas e disponibilizadas em rede aberta, será apresentada a seguir uma pesquisa focada na aplicação de métodos computacionais na soja e ontologias no campo. Dentre os trabalhos nessa área, conta-se com o de Lacerda (2017) para a detecção e predição da ferrugem asiática da soja, associados às técnicas de processamento digital de imagens e de mineração de dados, visando a obtenção de modelos preditivos de severidade nos diferentes estágios de desenvolvimento da soja.

Foram realizados estudos por Rosa (2018) para a classificação das pragas mais comuns do estado do Paraná, que são a *Anticarsia Gemmatalis*, *Helicoverpa armigera* e *Spodoptera Cosmioide*, e utiliza Rede Neural Profunda para prover a classificação de pragas mediante serviço, baseada na arquitetura *Inception V3*. Foi obtido um índice de acerto de 92,5%.

Um estudo da cultura da soja [Ferreira 2017] foi realizado pela elaboração de um banco de imagens com mais de 15 mil imagens do solo, soja e ervas daninhas de folhas largas e gramíneas. Com base nas imagens, uma Rede Neural Convolutiva foi instruída para detectar as ervas daninhas e os resultados foram comparados aos Algoritmos de Máquina de Vetor de Suporte, dos quais, as Redes Neurais Convolutivas foram capazes de alcançar uma precisão de mais de 98% na detecção das ervas daninhas de folhas largas e gramíneas.

PlantAI é um aplicativo desenvolvido [Bento 2018] para a detecção e identificação de doenças em plantas que utiliza técnicas de *Deep Learning* (DL) para auxiliar na tomada

de decisão do profissional, de forma a fornecer um diagnóstico e sugestão de tratamento, caso necessite.

Dentre os trabalhos da área de ontologias de domínio no campo da agricultura pode-se destacar o AgroPortal (2016), que reutiliza as ferramentas e *insights* semânticos do domínio biomédico para atender algumas áreas, dentre elas o domínio agrônomo. O portal apresenta hospedagem, pesquisa, versionamento, visualização, comentário, recomendação de ontologias, além de permitir anotações semânticas.

O CGIAR (*Consultative Group for International Agricultural Research*) (2017) sugeriu o desenvolvimento de um *software* de *big data* para a agricultura, com o uso da ontologia de domínio para a agricultura Crop Ontology (*Crop and Agronomy Ontology Community*) e o projeto AGROVOC (2013) também utiliza um vocabulário voltado para agricultura, compartilha palavras e faz o reuso de ontologias.

O Edubot [Lima 2017] é um projeto que possui um processo de aquisição de um modelo de domínio baseado em ontologias, modelado em lógica de descrição para a extração de conhecimento em linguagem natural, por meio da construção e representação de diálogos entre pessoas e máquinas, capaz de aprender e raciocinar sobre os fatos dialogados.

As pesquisas sobre o desenvolvimento de ferramentas e metodologias para facilitar e contribuir ao campo agrônomo visando melhorias crescem gradativamente. Entretanto, não foram encontrados trabalhos que utilizam LN para auxiliar os estudantes de ciências agrárias na identificação dos patógenos da soja e que possuem um grande repositório capaz de ser ampliada conforme o diálogo acontece.

3. Fundamentação teórica

Processamento de Linguagem Natural, conhecido como PLN, é o campo da Inteligência Artificial (IA) que pesquisa como os computadores podem ser utilizados para entender e manipular texto ou fala em linguagem natural para fazer coisas úteis [Sidhu 2013]. Isto é, desenvolve modelos computacionais para analisar e gerar interações entre humanos e computadores, os quais ocorrem por meio de língua natural [Lago 1999] segundo a estrutura morfológica, sintática, semântica e pragmática de um texto.

Mais precisamente, PLN é um conjunto de técnicas computacionais para a análise de textos em um ou mais níveis linguísticos, com o propósito de simular o processamento humano da língua [Barbosa 2004] e pode ser utilizado em diferentes abordagens para tratar das infindas aplicações pertinentes para o mundo. Assim, essas tecnologias necessitam elaborar soluções para entender a linguagem humana e gerar diálogos semelhantes aos que são reproduzidos naturalmente por pessoas.

Para Miura (2019) é imprescindível o desenvolvimento de aplicações automatizadas para analisar e interpretar textos em linguagem natural para que ações sejam executadas como resposta, fornecendo informações.

Visto a complexidade do assunto, o PLN tem o intuito de desempenhar um papel fundamental para a comunicação com os alunos, de maneira que esses se sintam mais confortáveis ao fazerem suas consultas em banco de dados com sua própria língua de comunicação. Assim, para que o sistema possa compreender e interpretar os dados gerados, Dale (2010) apresenta as cinco fases utilizadas na análise no PLN e serão detalhadas a seguir.

Com a **tokenização** é possível fazer o reconhecimento das letras maiúsculas e minúsculas, palavras compostas e abreviadas, bem como quebra de caracteres.

A **Análise Léxica** é responsável por manipular o léxico, que é composto por palavras que armazenam os seus significados e categorias lexicais. Essa etapa é responsável por fazer a verificação ortográfica, podendo ser classificada como: substantivo, verbo, advérbio, pronome, numeral, preposição, conjunção, interjeição, artigo e adjetivo; segundo sua estrutura e formação, identificando as partes: radical, tema, vogal temática, dentre outras [Domingues 2011].

A LN é uma linguagem de comando na qual as regras sintáticas são as da linguagem natural utilizadas pelo usuário e, portanto, dispensa a aprendizagem de sintaxes muito específicas e inflexíveis, como as de linguagens de programação. A tarefa da **Análise Sintática** (*parsing*) é de extrair informações de uma frase representada por meio de uma gramática e árvores sintáticas. Para Domingues (2011), a finalidade do *parsing* é analisar e gerar sentenças corretas de acordo com a estrutura de cada palavra. Maiores detalhes dessas no uso da agricultura serão exibidos na próxima seção.

A **Análise Semântica** consiste em analisar os significados das palavras, ou seja, interpretar as expressões fixadas, sentenças inteiras e enunciados no contexto [Goddard and Schalley 2010], pois as frases podem ser ambíguas. Oliveira (2002) afirma que a semântica pode ser dividida em léxica e gramatical, onde a primeira busca uma representação conceitual para descrever o sentido, sendo que, para construir essa representação, pode ser feita a decomposição semântica das unidades léxicas (em primitivas ou em traços semânticos), ou podem ser utilizadas redes semânticas. A semântica gramatical (também chamada composicional) procura identificar o sentido por meio de uma fórmula lógico-semântica. Porém, pode ocorrer ambiguidade.

Em relação à **Análise Pragmática**, o estudo fundamenta-se em reconhecimento de palavras dentro de um contexto [Muller 2003]. A estrutura não parte de apenas uma frase, visto que faz buscas nas frases do texto para compreender a frase analisada. Chomsky (1980) exemplifica essa fase com a fala em um diálogo, como em “*Já é muito tarde?*”, podendo o autor da frase se referir ao tempo/às horas do dia ou à ausência de pontualidade. A aplicação desse conhecimento em PLN é indispensável, principalmente quando o destaque do processamento interpreta as informações do usuário, permitindo o acesso à base de dados e aos textos esquemáticos, com uma parte variável que só é determinada no decorrer do discurso (ex.: essa, isso, aquela etc.).

4. Arquitetura Geral de um Sistema de Interpretação e Geração de Linguagem Natural

A arquitetura de um sistema computacional [Silva *et al.* 2007] que executa a língua natural pode variar de acordo com as especificidades da aplicação. O tradutor automático é um possível exemplo em um sistema mais completo que deverá ser capaz de: **a)** identificar, ou seja, extrair cada uma das palavras da sentença; **b)** analisar sintaticamente a sentença, isto é, relacionar a cada palavra suas propriedades e funções sintáticas; **c)** construir uma nova sentença para retomar o sentido das informações levantadas anteriormente, ou seja, extrair um significado absoluto da mesma, a partir dos significados das palavras e das relações entre elas; **d)** associar o significado extraído em uma representação adequada. Essa pode ser independente da língua destino (interlíngua); **e)** transformar a representação anterior em

uma sentença na língua destino, isto é, traduzir ou associar as palavras da língua origem para a língua destino, desde que sejam equivalentes.

A partir disso, é possível determinar as inúmeras maneiras de escrever ou expressar certo material informacional, fazendo com que o sistema controle automaticamente a geração da tarefa e que seu processo seja mais simples do que a de interpretação, como sistemas que têm a função específica de transmitir informações constantes em uma base de dados em PLN. A Geração de Linguagem Natural (GLN), de acordo com Araujo (2020), é o processo de produzir frases, sentenças e parágrafos que são significativos a partir de uma representação interna.

Baseado na GLN, foi elaborada para este projeto uma arquitetura geral de um sistema de interpretação, como em Silva *et al.* (2007), que é exibida na Figura 1, a qual está sendo pautada este trabalho.

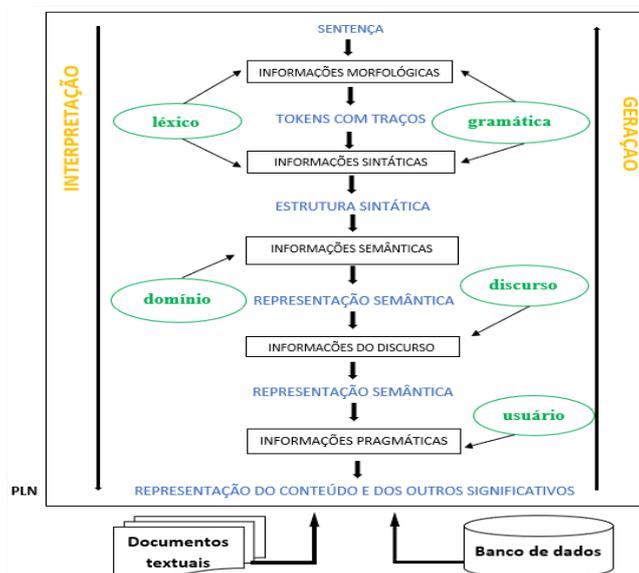


Figura 1. Arquitetura de um Sistema de Interpretação e Geração de Linguagem Natural (Fonte: Autores 2020)

Os indicadores de processamento são representados por retângulos, enquanto os recursos necessários ao processamento, de ordem linguística (gramática, léxico) ou não (modelos do domínio e do usuário), aparecem representados por círculos. Esses recursos são essenciais durante a fase de interpretação, bem como na de geração. Destacam-se as aplicações que não envolvam a interpretação de uma sentença que têm sua arquitetura simplificada, eliminando-se alguns dos módulos e/ou bases de conhecimento.

Os recursos linguísticos presentes na arquitetura de interpretação e geração de linguagem natural, fundamentada em Silva *et al.* (2020) são divididos em: **Léxico** - conjunto de palavras ou expressões da língua associadas a um conjunto de atributos, ou traços morfossintáticos e traços semânticos (opcionais). Durante a interpretação, o léxico é acessado pelos analisadores léxico, sintático e semântico, onde cada um deles visam funções específicas, sendo que suas principais tarefas, respectivamente, são: reconhecer *tokens* da sentença de entrada e recuperar seus principais traços; reconhecer ou atribuir categorias sintáticas aos tokens para a obtenção da estrutura profunda da sentença e verificar a validade do relacionamento semântico do *token* sob análise em função do contexto em que ela ocorre na sentença em relação aos demais *tokens* obtidos durante a análise dos demais componentes sentenciais. **Gramática** - conjunto de regras gramaticais

que definem as cadeias de palavras válidas em uma sentença em linguagem natural. **Modelo de Domínio** - fornece informações sobre o domínio específico da aplicação, como de senso comum sobre as entidades do discurso em foco, e padrões ontológicos sobre tal modelo. **Modelo de Usuário** - permite reconhecer características do significado textual a partir do contexto do discurso.

5. Desenvolvimento do Sistema de Identificação das Pragas da Soja

Perante os problemas levantados, a elaboração deste trabalho permite contribuir com a área agrícola, principalmente no que diz respeito à viabilidade de um assistente virtual, utilizando técnicas de PLN na identificação de pragas e doenças nas plantas da soja.

Os principais aspectos que identificam as ameaças da sojicultura são classificados conforme suas características e são detalhadas na Tabela 1, respectivamente.

Tabela 1. Classificação e características para identificação das pragas e doenças

Classificação	Características
Praga	Nome comum e nome científico
Descrição	Descreve a morfologia e ciclo
Biologia	Características biológicas
Danos	Como se comportam e os danos causados
Ocorrência na planta	Localização do ataque da planta
Distribuição geográfica	Localização geográfica de onde essas pragas são encontradas
Controle	Métodos de controle
Categoria	Pragas que atacam: plântulas/raízes/caules/folhas/vagens; Pragas subterrâneas/aérea/solo/mastigadoras/sugadoras; outros

O dicionário utilizado é uma estrutura de dados, nas quais as palavras são armazenadas e associadas a elas algumas de suas informações [Barbosa 2004]. O repertório das palavras registradas no dicionário conta com 101 pragas e doenças nas plantas da soja [Santos 1995], [Moreira and Aragão 2009], [Sosa-Gomez *et al.* 2010] e [Ávila e Grigolli 2014].

A partir dos estudos realizados acerca do presente tema, foi feita uma modelagem para BDs NoSQL do tipo grafo (Neo4j¹), que permite consultas e a extração das informações de um conjunto de vértices e arestas, como é apresentada na Figura 2.

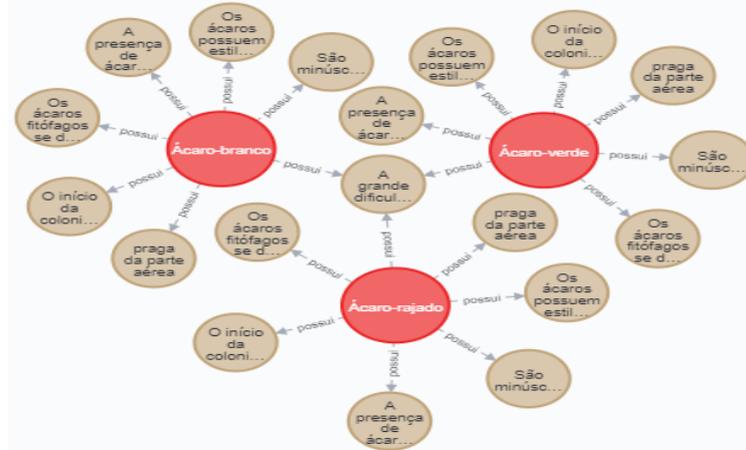


Figura 2. Dados dos Ácaros Fitófagos no Banco de Dados NoSQL Neo4j

¹ <https://neo4j.com/>

O processo descrito permite a extração das informações que são efetuadas com a finalidade de identificar a qual rótulo específico uma determinada praga ou relacionamento está atrelado. Este projeto tem o objetivo não só de colaborar com o estudante da área agrônômica, mas também ganhar vivência no campo, aprender sobre as pragas e conhecer os tipos de danos diretos e indiretos e aplicar na prática os devidos métodos de controles. Visando isso, foi possível realizar perguntas para realizar o diálogo e possibilitar consultas, como: *“Tem um verme provocando uma lesão na raiz da soja.”*; *“A minha semente está com uma mancha-púrpura.”*; *“A soja armazenada está infestada de insetos.”*; *“O broto não está se desenvolvendo.”*; *“Tem uma praga esverdeada na plantação.”*; *“Como posso controlar os nematoides?”*; *“Tem uma praga atacando a haste da soja.”*; *“O que é um bichinho dourado na soja?”*; *“O que posso fazer para controlar a Antracnose?”*; *“Os percevejos atacam que parte da planta?”*; *“Qual o dano causado pela Formiga-cortadeira?”*; *“Os ácaros atacam as folhas da planta?”*; *“Como prevenir a presença de pragas na minha lavoura?”*; *“Como controlar a Antracnose?”*; *“A haste da soja está apodrecendo.”*

As perguntas vão surgindo conforme o diálogo entre o usuário e o sistema acontece (pergunta-resposta) e é esperado que o usuário obtenha informações relevantes para tomada de decisão. As palavras para esas perguntas foram catalogadas separadamente de acordo com a sua categoria morfológica para formar o dicionário utilizado na análise léxico-morfológica das palavras. A partir disso, está sendo desenvolvido um sistema para dialogar com alunos sobre as principais características das principais pragas e doenças na cultura da soja. Esse sistema recebe o texto e analisa as palavras de uma sentença isoladamente, como pode ser visto na Figura 3, para poder interpretá-lo como um todo.

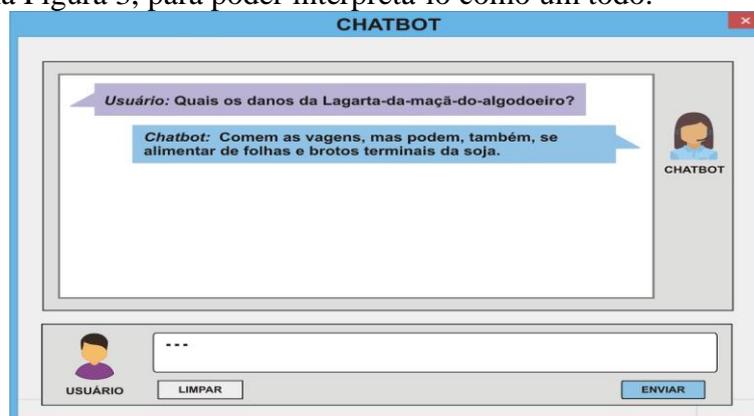


Figura 3. Protótipo de Interface de Consulta à Base de Dados das Pragas da Soja

Identificar as partes da frase é essencial porque ajuda a entender as frases de entrada e constrói com mais exatidão as frases de saída (resposta). Assim, a primeira etapa é tokenizar os textos e extrair as etiquetas morfológicas para classificação de cada palavra pelo framework spaCy².

Posteriormente, essa ferramenta rotula todos os tokens a partir da análise para prever qual tag provavelmente se aplica nesse contexto. Para isso, conta-se com as seguintes tarefas: **Text** (texto puro), **Lemma** (reduz as palavras em seu formato base/raiz), **POS** (tags simples que determinam as categorias gramaticais de um token), **Dep** (Dependência sintática é a relação entre os tokens presentes dentro de uma sentença para entender o seu significado), **Shape** (classificação da palavra em maiúscula ou minúscula),

² <https://spacy.io>

Alpha (especificação das palavras em alfanuméricas ou não), **Stop** (indica se as palavras são consideradas *stopwords*), como é detalhada na Figura 4.

TEXT	LEMMA	POS	DEP	SHAPE	ALPHA	STOP
As	as	DET	det	Xx	True	True
folhas	folhar	NOUN	nsubj	xxxx	True	False
atacadas	atacar	VERB	acl	xxxx	True	False
ficam	ficar	VERB	root	xxxx	True	False
com	com	ADP	case	xxxx	True	True
grandes	grande	ADJ	amod	xxxx	True	True
áreas	área	NOUN	obl	xxxx	True	False
recortadas	recortar	VERB	Acl	xxxx	True	False
ou	ou	CCONJ	Cc	xxxx	True	True
são	ser	VERB	aux:pass	xxxx	True	True
completamente	completamente	ADV	advmod	xxxx	True	False
consumidas	consumir	VERB	conj	xxxx	True	False
?	?	PUNCT	punct	?	False	False

Figura 4. Parsing para “As folhas atacadas ficam com grandes áreas recortadas ou são completamente consumidas”

O spaCy também possibilita descrever a relação sintática das palavras que se conectam na formação da árvore. Isso permite percorrer toda a árvore e retornar uma sequência ordenada de tokens e verificar os atributos e domínios das palavras. Nessa fase, além do que foi descrito anteriormente, conta-se com o **Head Text** (relação entre as palavras nos tokens), **Head Pos** (rotula as palavras em categorias) e **Children** (dependentes sintáticos do token) e são apresentados na Figura 5.

TEXT	DEP	HEAD TEXT	HEAD POS	CHILDREN
As	DET	folhas	NOUN	[]
folhas	NSUBJ	ficam	VERB	[As, atacadas]
atacadas	ACL	folhas	NOUN	[]
ficam	ROOT	ficam	VERB	[folhas, áreas, consumidas]
com	CASE	áreas	NOUN	[]
grandes	AMOD	áreas	NOUN	[]
áreas	OBL	ficam	VERB	[com, grandes, recortadas]
recortadas	ACL	áreas	NOUN	[]
ou	CC	consumidas	VERB	[]
são	AUX:PASS	consumidas	VERB	[]
completamente	ADVMOD	consumidas	VERB	[]
consumidas	CONJ	ficam	VERB	[ou, são, completamente]
?	PUNCT	ficam	VERB	[]

Figura 5. Navegação pela Árvore de Análise sintática em “As folhas atacadas ficam com grandes áreas recortadas ou são completamente consumidas”

Com isso, a biblioteca spaCy foi capaz de processar grandes volumes de textos e extrair informações para realizar tarefas relacionadas ao PLN. Visto isso, o elemento principal do trabalho foi a elaboração de um sistema para identificação das pragas da soja utilizando técnicas de PLN, uma vez que há um aumento de interesse em utilizar os sistemas de computador como auxílio no aprendizado do estudante.

A principal dificuldade no desenvolvimento está sendo dialogar com o usuário e reconhecer as suas intenções a partir de uma frase e respondê-lo automaticamente. Dessa forma, o sistema inteligente de pré-atendimento aos discentes tem o intuito de ser um canal alternativo de comunicação, para facilitar o acesso às informações e auxiliar no ensino para identificar o patógeno.

6. Considerações finais

Este trabalho concentrou as informações sobre as principais pragas e doenças da sojicultura em um repositório com o intuito de auxiliar o aluno no ensino-aprendizagem por meio de

uma agente conversacional, com bom desempenho de análise de *parsing*, a partir da linguagem própria desse público.

Um modelo foi criado utilizando o formato JSON *Schema* para extrair os textos do banco de dados Neo4j. O foco da presente proposta é a análise sintática, responsável por organizar as estruturas gramaticais. Assim, foi possível construir um modelo por meio do framework spaCy para aplicar regras gramaticais à sentença, reconhecer a estrutura e extrair seus significados, além de permitir treinar o modelo para refinar seus dados.

O tema proposto indicou algumas possibilidades de efeitos positivos e que pode auxiliar o estudante da área que precisa se envolver com um amplo volume de informações na identificação das ameaças que atingem a lavoura.

Referências

- Agroportal (2016) “AgroPortal LIRMM”, <http://agroportal.lirmm.fr/>. Junho 2020.
- Araujo, E. F. S. (2020) Solução Chatbot no Ambiente acadêmico da UFRJ, Departamento de Eletrônica e de Computação da Universidade Federal do Rio de Janeiro. Trabalho de Conclusão de Curso. Rio de Janeiro, 77 f.
- Ávila, C. J. e Grigolli, J. F. J. (2014) “Pragas da soja e seu controle”, Embrapa Agropecuária Oeste, p.109-168.
- Barbosa, C. R. S. C. (2004) Técnicas de Parsing para Gramática Livre de Contexto Lexicalizada da Língua Portuguesa, Departamento de Engenharia Eletrônica e Computação do Instituto Tecnológico da Aeronáutica, Tese de Doutorado. São José dos Campos, 171 f.
- Bento, D. C. P. G. C. (2019) Detecção e identificação de doenças em plantas utilizando Deep Learning, Instituto Superior de Engenharia do Porto, Dissertação de Mestrado. Porto, Portugal. 125 f.
- Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbahndari, S., Jaques, Y. and Keizer, J. (2013) The AGROVOC Linked Open Dataset, *Semantic Web*, v.4, n.3, p.341-348. <http://aims.fao.org/aos/agrovoc/>, Junho 2020.
- CGIAR. (2017) “Plataform for Big Data in Agriculture”, <http://bigdata.cgiar.org/ontology>, Maio 2020.
- Chomsky, N. (1980) Reflexões Sobre a Linguagem. São Paulo: Cultrix.
- Chowdhury, G. G. (2003) Natural Language Processing. In: *Annual Review of Information Science and Technology*, v.37, pages 51-89, Glasgow.
- Dale, R. (2010) “Classical approaches to natural language processing”, In: N. Indurkha & F. J. Damerau (Eds.). *Handbook of Natural Language Processing*, 2ª edição, p.3-7.
- Domingues, M. L. C. S. (2011) Abordagem para o Desenvolvimento de um Etiquetador de Alta Acurácia para o Português do Brasil”, Departamento de Engenharia Elétrica da Universidade Federal do Pará. Tese de Doutorado. Belém, 154 f.
- Ferreira, A. S. (2017) Redes Neurais Convolucionais Profundas na Detecção de Plantas Daninhas em Lavoura de Soja, Faculdade de Computação da Universidade Federal do Mato Grosso do Sul. Dissertação de Mestrado. Campo Grande, 80 f.
- Goddard, C. and Schalley, C. A. (2010) *Semantic Analysis: a Practical Introduction*, Oxford: Chapman and Hall/CRC, 2ª edição.

- Gomez, D. R. S., Campo, C. B. H., Correa-Ferreira, B. S., Corso, I. C., Oliveira, L. J., Moscardi, F., Panizzi, A. R., Bueno, A. F., Hirose E. (2014) “Manual de identificação de insetos e outros invertebrados da cultura da soja”, Embrapa Soja-Documents (Infoteca-E), Londrina.
- Jurafsky, D. and Martin, J. H. (2000) *Speech and Language Processing: an introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, New Jersey, USA: Prentice Hall/PTR, 1ª edição.
- Lacerda, V. S. (2017) Estimativa do índice de severidade de ferrugem asiática na cultura da soja por meio de imagens obtidas com aeronave remotamente pilotada, Programa de Pós-Graduação em Computação Aplicada da Universidade Estadual de Ponta Grossa. Dissertação de Mestrado. Ponta Grossa, 62 f.
- Lago, S. (1999) “Processamento de Linguagem Natural”, www.ime.usp.br/~slago/IA-pln.pdf. Maio 2020.
- Lima, C. E. T. (2017) Um Chatterbot para criação e desenvolvimento de ontologias com lógica de descrição, Departamento de Ciência da Computação da Universidade Federal de Pernambuco. Dissertação de Mestrado. Recife, 101 f.
- Miura, N. K. (2019) Geração incremental de parsers dependentes de contexto para o português brasileiro, Departamento de Engenharia de Computação e Sistemas Digitais. Tese de Doutorado. São Paulo, 132 f.
- Moreira, H. J. C. e Aragão, F. D. (2009) *Manual de Pragas da Soja*, Campinas: FMC Agricultural Products, 144f.
- Muller, D. N. (2003) “Processamento de Linguagem Natural”, <http://www.inf.ufrgs.br/~danielnm/docs/pln.pdf>, Setembro 2020.
- Oliveira, F. A. D. (2002) Processamento de linguagem natural: princípios básicos e a implementação de um analisador sintático de sentenças da língua portuguesa, In: *Revista de Ciência da Informação*, n.5. Rio de Janeiro, pages 6-14.
- Pereira, P. H. S., Pereira, P. H. S., Oliveira, G. V., Yokoo, R. S. S., Rodrigues, M. M. S. e Coletta, L. F. S. (2018) Análise de descritores de imagens na classificação de folhas de soja visando o diagnóstico de doenças, In: *X Simpósio Nacional de Tecnologia em Agronegócio*, Presidente Prudente, p.89-100.
- Rosa, R. P. (2018) Método de classificação de pragas por meio de rede neural convolucional profunda, Departamento de Computação Aplicada da Universidade Estadual de Ponta Grossa. Dissertação de Mestrado. Ponta Grossa, 99 f.
- Santos, O. S. (1995) *A Cultura da Soja 1 Rio Grande do Sul-Santa Catarina-Paraná*. São Paulo: Globo, 2ª edição.
- Sidhu, B. K. (2013) Natural language processing, In *International Journal of Computer Technology and Applications*, pages 751-758.
- Silva, B. C. D., Montilha, G., Rino, L. H. M., Specia, L., Nunes, M. G. V., Oliveira Junior, O. N., Martins, R. T., Pardo, T. A. S. (2007) *Introdução ao Processamento das Línguas Naturais e suas Aplicações*, Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional da Universidade de São Paulo. São Carlos, 121 f.