

# Ferramenta para o auxílio no ensino-aprendizagem de pragas na cultura da soja

**Carolinne Roque e Faria**  
Universidade Estadual de Londrina  
Londrina, Brasil  
carolinne.rf@outlook.com

**Maurilio Martins Campano Junior**  
Universidade Estadual de Londrina  
Londrina, Brasil  
maurilio.campanojr@gmail.com

**Cinthyan Renata Sachs Camerlengo de Barbosa**  
Universidade Estadual de Londrina  
Londrina, Brasil  
cinthyan@uel.br

## ABSTRACT

The volume of data available in the field of agronomy is increasing, especially regarding agriculture. One of the challenges in this context is to approach natural language described data targeting educational purposes. This paper proposes the development of a mobile application, consisting of a support system that uses data on pests in soybean, such as behavior, damage, location and control to aid the decision making process in agriculture. Such data will be stored in a database, thus making it possible to build dialogues and acquire more information. Through the developed of a prototype, a pre-evaluation was made with the proposed theme, which indicated possibilities of positive effects that can facilitate the work of professionals who need to be involved with a large volume of information, as well as be useful for students in process of individual study.

## Keywords

Learning environment; Informatics in Agriculture; Natural Language Processing.

## RESUMO

Os volumes de dados disponíveis no campo da agronomia, principalmente quando falamos sobre agricultura é cada vez maior. Um dos obstáculos nesse contexto é abordar os dados em linguagem natural para fins educacionais. Este artigo propõe o desenvolvimento de um sistema de apoio à decisão para a agricultura em um aplicativo móvel,

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Times New Roman 8-point font. Please do not change or modify the size of this text box.

Each submission will be assigned a DOI string to be included here.

utilizando os dados sobre as pragas na cultura da soja, como comportamento, danos, localização e o controle que serão armazenados em um banco de dados, possibilitando assim consultas para construir diálogos e obter diagnósticos precisos às informações do banco de dados. Através de um protótipo desenvolvido foi feita uma pré-avaliação com o tema proposto, o que indicou possibilidades de efeitos positivos e que pode facilitar o trabalho de profissionais que precisam se envolver com um amplo volume de informações, bem como pode ser útil para estudantes em processo de estudo individual.

## Palavras-chave

Ambiente de aprendizado; Informática na Agricultura; Processamento de Linguagem Natural.

## ACM Classification Keywords

[Applied Computing]: Interactive Learning Environments

## INTRODUÇÃO

Atualmente, a importância da agricultura é inquestionável, já que é um segmento que impacta de maneira significativa o desenvolvimento de todo o mundo e no Brasil [1], como pode ser visualizado na Figura 1, exibida em [2].

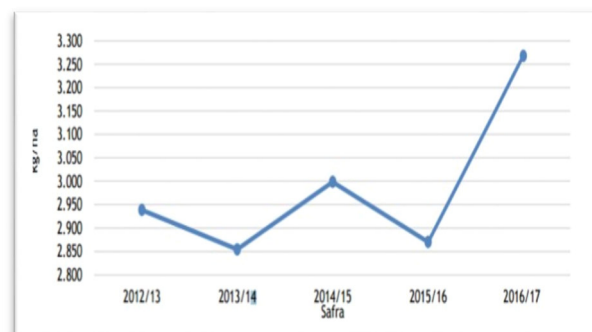


Figura 1 - Produtividade Média da Cultura da Soja, Brasil [2, 2012/13 a 2016/17]

As práticas agrícolas consistem em sistemas avançados, devido à ciência e à produção de conhecimento. Este artigo pretende ser útil ao incentivo à pesquisa na agricultura, já que há um aumento de interesse em utilizar os sistemas de computador como auxílio, visando otimizar a análise dos dados de uma cultura e também melhorar a produção do agricultor.

A área de Processamento de Linguagem Natural (PLN) [3,4] apresenta soluções computacionais para a compreensão da língua natural (escrita ou falada), como, por exemplo, a Análise Léxico-Morfológica (tokenização), Análise Sintática (análise), Análise Semântica e Análise Pragmática.

O sucesso de um sistema de PNL depende de seu conhecimento do domínio da aplicação, ou seja, da integridade relativa do modelo de linguagem natural que ele encapsula e da adequação ou eficiência de seus algoritmos [5].

Uma das razões que torna o PLN tão importante [6] é que modifica as técnicas de processamento dos computadores o mais próximo possível dos mecanismos do raciocínio humano.

É necessária uma compreensão automática da linguagem da agricultura, de modo que possa ser manipulada por computadores e também que analise as questões em linguagem natural para que essas contribuam com os estudantes de agronomia e até mesmo com os agricultores nas tomadas de decisões, de acordo com o diagnóstico feito por meio de diálogos com a finalidade de proporcionar consultas produtivas sobre as descrições de diagnósticos e fazer escolhas de acordo com as alternativas geradas.

## TRABALHOS RELACIONADOS

Hoje em dia outras áreas de conhecimento estão utilizando da informática para automatizar e melhorar o desempenho de suas tarefas, como é o caso da agricultura. Dentre os trabalhos nessa área por meio de análise dos segmentos do DNA pode-se utilizar algoritmos baseados em [7]: sequências promotoras de genes co-regulados, palavras, predição probabilística e rastro filogenéticos que comparam sequências promotoras de genes ortólogos de múltiplas espécies.

O projeto AgroPortal [8] reutiliza as ferramentas e *insights* semânticos do domínio biomédico para atender não apenas a agronomia, mas também às ciências da alimentação, das plantas e da biodiversidade. Foi oferecido um portal que apresenta hospedagem, pesquisa, versionamento, visualização, comentário e recomendação de ontologias. O AgroPortal satisfaz especificamente os requisitos da comunidade de agronomia em termos de formatos de ontologia (por exemplo, vocabulários e dicionários de traços do SKOS (*Simple Knowledge Organization System*)

[9], que é uma área de trabalho que desenvolve especificações e padrões para apoiar o uso de sistemas de organização do conhecimento, como esquemas de classificação, sistemas de títulos de assuntos e taxonomias dentro da estrutura da *Web Semântica*) e recursos suportados (oferecendo, por exemplo, metadados e anotações avançadas). Nesse projeto foi apresentado o conteúdo e os recursos da plataforma, incluindo os acréscimos à tecnologia original, bem como os resultados preliminares de cinco casos de uso agrônomo que eram as principais fontes de ontologias e vocabulários. Com base na experiência e na tecnologia existente adquirida no domínio biomédico, o AgroPortal apresentou um repositório robusto e rico em recursos e de grande valor para o domínio agrônomo.

O trabalho [10] alia a *visão computacional* à análise de folhas de soja, com o objetivo de encontrar e extrair características que permitam a detecção de doenças foliares.

Tal trabalho faz uso de pré-processamento da imagem de folhas de soja, utilizando-se de filtros de média, mediana e métodos de detecção de bordas e linhas para identificar a folha da soja. O trabalho também faz uso de segmentação de imagens, porém com uma dificuldade na obtenção das veias nas folhas de soja mais escuras.

Para [11], a computação se faz presente na viticultura (cultura da uva). Nesse trabalho utiliza-se de redes de sensores sem fio para monitoramento ambiental permitindo a análise e controle de quantidade de chuva, da temperatura e a incidência solar em fruteiras, parâmetros nos quais influenciam a qualidade da fruta e conseqüentemente a qualidade do cultivo.

Em [12] foram realizados estudos de cultura de soja, construindo um banco de imagens com mais de 15 mil imagens do solo, soja e ervas daninhas de folhas largas e gramíneas. A partir dessas fotos uma *Rede Neural Convolutiva* foi treinada para detectar as ervas daninhas, sendo os resultados comparados com os Algoritmos de Máquina de Vetor de Suporte, sendo que as Redes Neurais Convolutivas foram capazes de obter uma precisão de mais de 98% na detecção das ervas daninhas de folhas largas e gramíneas.

Já em [13] também foram utilizados de *processamento de imagens*, para controle e identificação do fungo *Cercospora kikuchii*, causador da mancha púrpura na folha da soja, que pode implicar em prejuízos na produtividade e também na produção dos derivados. O sistema computacional foi desenvolvido utilizando a biblioteca de visão computacional *OpenCV*, em conjunto com a linguagem Java e integrando com a ferramenta *Weka*. A análise de 150 sementes sadias e 150 sementes contaminadas se mostrou eficiente, com um índice 86% de precisão.

O trabalho de [14] utiliza-se de um *ambiente virtual* para disponibilizar um meio de estudo, de experimentação, análises e avaliações de situações da cultura da soja. No

ambiente virtual, que pode ser visualizado na Figura 2, por meio de um formalismo matemático, uma *gramática L-Systems* e o desenvolvimento de sua representação gráfica no ambiente virtual foi possível simular o comportamento da cultura da soja com alterações dos macronutrientes do solo, tendo como resultados um melhor acompanhamento da cultura da soja e manejo do solo.

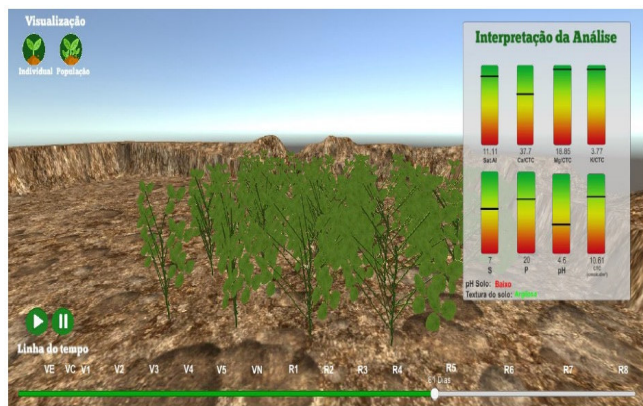


Figura 2 - População de plantas em ambiente virtual [14, 2017]

## FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica deste trabalho é baseada em técnicas de PLN para o desenvolvimento de uma ferramenta educativa que auxilie no processo de ensino de agronomia, para que haja compreensão dessa área no estudo de pragas no cultivo da soja.

A soja é um dos principais produtos agrícolas no contexto econômico mundial e nacional. O sucesso e a competitividade do sojicultor no mercado mundial de grãos são devido às condições edafoclimáticas brasileiras e a utilização de tecnologias para o aumento da produtividade, embora não haja comprovação científica do benefício das tecnologias [15].

No entanto, a cultura da soja pode ser atacada por vários tipos de pragas [16, 17, 18, 19]. A habilidade de identificação das pragas em seus diversos estágios de desenvolvimento e os desafios enfrentados a cada safra foram o que motivou o desenvolvimento desta ferramenta com o uso de técnicas de Processamento de Linguagem Natural.

Um sistema de PLN pode possuir módulos autômatos, que realizam tarefas específicas e especializadas e módulos que armazenam um modelo de conhecimento proposicional, que visa a criar simulacros de parcelas de mundo que lhe servem de referencial para interpretar os enunciados linguísticos [20].

A análise em PLN é dividida em vários segmentos, considerando os traços linguísticos. As cinco fases

utilizadas na análise de PLN [21] são descritas abaixo como mostra a Figura 3.

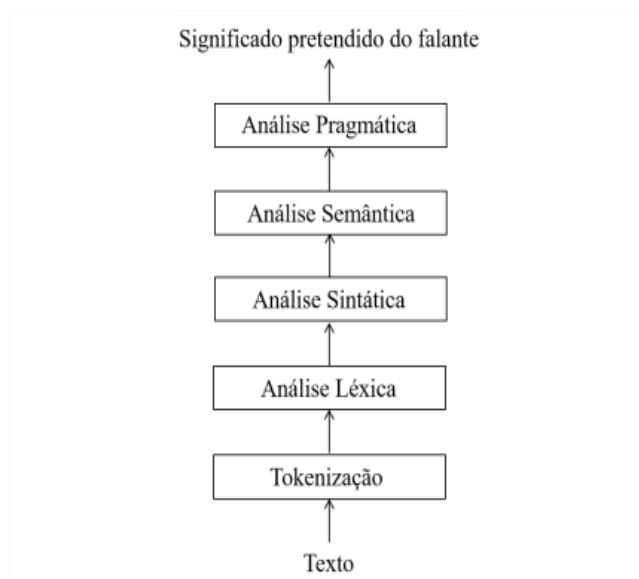


Figura 3 – Estágios de análise no PLN [21, 2010]

A *tokenização* é um processo conhecido como segmentação de palavras, reconhecendo letras maiúsculas e minúsculas, palavras compostas e abreviadas, quebra de caracteres, identificando ambiguidades lexicais e atenção à verificação ortográfica.

O *léxico* é composto por palavras que armazena seu significado e categoria lexical e pode conter seu radical, que a partir dele podem ser formadas outras palavras.

A tarefa da *análise sintática (parsing)* é extrair informações de uma frase representada por meio de gramáticas e árvores sintática como é apresentada na frase Figura 4. Nessa imagem é exibida uma árvore de uma Gramática Livre de Contexto baseada no trabalho de [4], a qual poderia ser também usada para a linguagem agrônoma. No referido trabalho é utilizada uma Gramática Livre de Contexto Lexicalizada, uma vez que a lexicalização é importante [22] do ponto de vista computacional, porque as gramáticas lexicalizadas podem muitas vezes ser analisadas mais eficientemente que as não-lexicalizados. Operações de substituição e adjunção como mostradas em [4] também podem ser usadas no domínio da radiologia.

A análise semântica consiste em analisar os significados das palavras, ou seja, interpretar as expressões fixadas, sentenças inteiras e enunciados no contexto [23], pois as palavras podem ser ambíguas.

Em relação à análise pragmática, o estudo fundamenta-se em reconhecimento de palavras dentro de um contexto [24]. A estrutura não parte de apenas uma frase, visto que busca nas frases do texto para compreender a frase analisada.

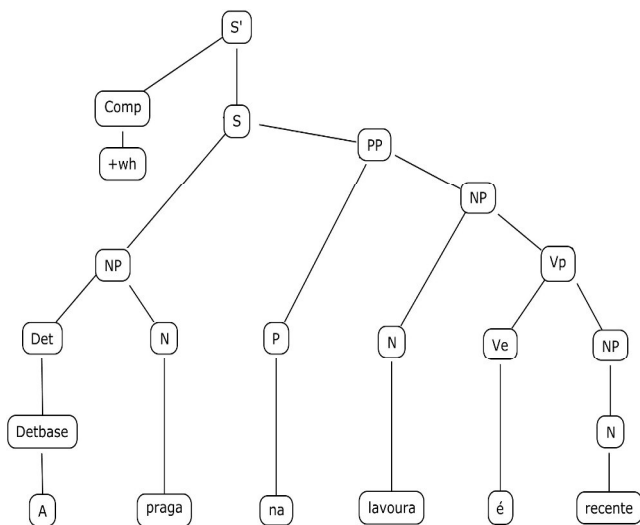


Figura 4 – Substituição em *A praga na lavoura é recente*

## METODOLOGIA E DESENVOLVIMENTO

O recurso utilizado na seleção de palavras foi realizado em três passos: reunião com agrônomos; leitura e indicações bibliográficas sobre as pragas; consultas nos site EMBRAPA (*Empresa Brasileira de Pesquisas Agropecuária*) [25] e demais manuais de pragas da soja [16, 17, 18, 19].

A partir da análise das características das pragas extraídas dos documentos acima e o discurso informal das entrevistas, tornou-se possível construir uma ontologia de palavras sobre o assunto, proporcionando várias formas de se fazer a mesma pergunta relacionada às características das pragas que atacam a produção de soja. Essas perguntas estão sendo validadas por meio de um analisador léxico de domínio da agronomia para ser usado no ensino de pragas na cultura de soja.

Tal analisador dará, por exemplo, além da categoria morfológica da praga, seu nome científico, suas características biológicas, seu comportamento, os danos causados, sua localização na soja, os métodos de controle de danos e uma literatura disponível sobre cada uma das pragas. Para isso está sendo utilizada a tabela de símbolos recomendada por [26].

Tendo em vista que o objetivo principal do projeto é auxiliar estudantes da área da agronomia sobre o diagnóstico de pragas na cultura da soja, permitiu-se a elaboração de questões possíveis de serem feitas pelo agricultor para promover consultas ricas, as quais serão descritas abaixo.

- Qual o dano da praga Percevejo-marrom-da-Soja?
- Qual estado tem maior incidência da praga Formiga cortadeira?

- Qual praga ataca mais no estado do Paraná?
  - Há quanto tempo surgiu a praga Larva-Angorá na sua lavoura?
  - Alguém da mesma região teve a lavoura atingida pela mesma praga?
  - Como controlar a praga Cochonilha-da-Raiz?
  - Qual produto posso utilizar para controlar a praga Lagarta-cabeça-de-fósforo?
  - Como identifico a praga Coró-da-Soja em minha lavoura?
  - A praga na lavoura é recente?
  - Está utilizando algum inseticida?
  - É a primeira vez que a Mosca-branca aparece na lavoura?
  - Qual a diferença entre o Percevejo-verde e o Percevejo-de-barriga-verde?
  - Qual é o ciclo biológico da praga Mosca-Branca?
  - Qual a origem da praga Tamanduá-da-Soja?
- Que danos são causados nas plantas de soja?

### I) Pergunta analisada

- Frase: “Qual é o ciclo biológico da praga Tamanduá-da-Soja?”
- Padrão: Pronome relativo + verbo [ser] + artigo + substantivo comum + adjetivo + preposição + substantivo comum + substantivo composto.

### II) Resposta completa

**a. Frase:** “Os ácaros fitófagos da soja podem ser naturalmente controlados em situações de chuvas intensas e períodos prolongados de umidade relativa elevada, especialmente porque essas condições propiciam o desenvolvimento de fungos entomopatogênicos e outros agentes de controle biológico natural no agroecossistema. Inseticidas organofosforados, como metamidofós, profenofós, dimetoato, endosulfam e clorpirifós são sugeridos para o controle de ácaros na cultura da soja.”

Essas questões foram catalogadas e cada palavra individualmente foi classificada no dicionário de acordo com a categoria morfológica para que se possa formar um

dicionário de bases, o qual será utilizado na análise léxico-morfológica das palavras.

Um protótipo de consulta por meio do *framework* pode ser visualizado na Figura 5, no qual é possível inserir uma pergunta em linguagem natural e obter uma resposta com base nos dados das pragas.

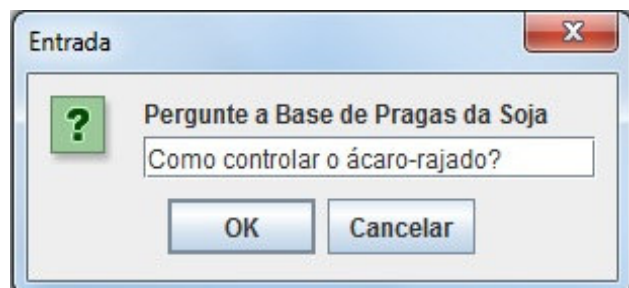


Figura 5 – Protótipo de Interface de Consulta na base de dados das pragas

Respostas para essas consultas podem ser visualizadas, como na Figura 6.

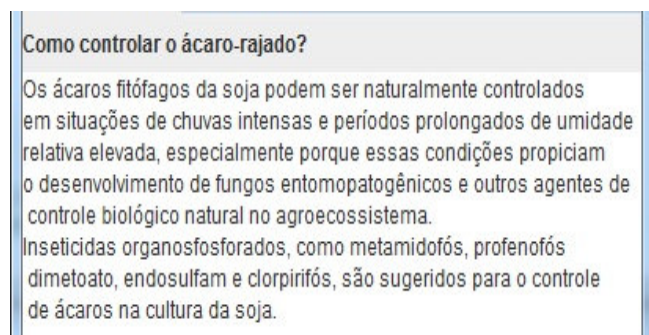


Figura 6 - Resposta de Consulta a uma pergunta sobre uma praga

## CONSIDERAÇÕES FINAIS

Embora ainda não esteja completo o dicionário, boa parte das pragas mais conhecidas na soja foi cadastrada totalizando 78 pragas. Para este trabalho, o próximo passo é a validação da gramática da linguagem de consulta apresentada em [4] para que possam ser realizadas as fases de análise sintática e semântica das frases da agronomia em linguagem natural. Sugere-se também o uso de famílias de árvores, como mostradas em [4], para reduzir a busca de frases válidas semanticamente dizendo.

Pretende-se ainda descrever mais consultas na base de dados da agricultura, mas para isso faz-se necessário previamente mais entrevistas com os agricultores para verificar se há mais jargões nesse domínio para que sejam abarcadas em nosso dicionário. É preciso validar se as

regras gramaticais desse público alvo são as mesmas da norma culta ou não para saber se há a necessidade de acrescentar mais regras à referida gramática.

## REFERÊNCIAS

1. Hirakuri, M. H. and Lazzarotto, J. J. *O agronegócio da soja nos contextos mundial e brasileiro*. Londrina. Embrapa Soja, 2014.
2. Barbosa, M. Z. and Sampaio, R. M. Soja: *alta produtividade e tecnologia*. Disponível em: <<http://maissoja.com.br/soja-alta-produtividade-e-tecnologia/>>. Acesso em: 14 set 2018.
3. Barbosa, C. R. S. C. *Interfaces em Linguagem Natural para Banco de Dados*. Porto Alegre. CPGCC da UFRGS. 1997. 165p. (TI-640)
4. ————. *Gramática para Consultas Radiológicas em Língua Portuguesa*. Porto Alegre. CPGCC da UFRGS. 1998. 143p. Dissertação de Mestrado.
5. Manaris, B. Z. and Slator, B. M.: Interactive Natural Language Processing: Building on Success. *IEEE Computer*, New York, 29(7), 28-32, July (1996).
6. Agosti, C. *Interface em Linguagem Natural para Banco de Dados: uma abordagem prática*. Florianópolis. PPGCC da UFSC, 2003. 203p. Dissertação de Mestrado.
7. Silla, P. R. *Identificação de miRNAs e seus alvos no genoma da soja com o uso de abordagens computacionais*. Londrina. UEL, 2011. 85p. Dissertação de Mestrado.
8. Agroportal. *AgroPortal LIRMM*. Disponível em: <<http://agroportal.lirmm.fr/>>. Acesso em: 21 set 2018.
9. Alcts. *SKOS: a Guide for Information Professionals*. Disponível em: <<http://www.ala.org/alcts/resources/z687/skos/>>. Acesso em 24 set 2018.
10. Rocha, I. A. A. and Sartin, M. A. 2018. Pré processamento e segmentação de imagens de folhas de soja com base na visão computacional. In *Workshop de Tecnologias Emergentes em Computação*.
11. Verona, A. B. *Simulação de Redes de Sensores Sem Fio Aplicadas à Viticultura*. PCC da UEM, 2010. 119p. Dissertação de Mestrado.
12. Ferreira, A. S. *Redes Neurais Convolucionais Profundas na Detecção de Plantas Daninhas em Lavoura de Soja*. Campo Grande. FACOM da UFMS. 2017. 80f. Dissertação de Mestrado.
13. Franco, J. R. *Método computacional para identificação do fungo Cercospora Kikuchii em sementes de soja*. PPGCA da UEPG. Universidade Estadual de Ponta

- Grossa, Ponta Grossa. Dissertação de Mestrado. 2017. 111f.
14. Camargo, V. X.; Nogueira, E. A.; Ribeiro, M. W. S. and Teixeira, D. S. 2017. ManejoSoja3D: Ambiente Virtual para Aprendizado de Manejo da Cultura da Soja. In *VI Congresso Brasileiro de Informática na Educação (CBIE'17)*, 776-786.
  15. Carvalho, M. M. et al. Potencial do controle biológico para o controle de *Pseudoplusia includens* (Walker, 1857) e *Anticarsia gemmatalis* Hubner, 1818 (Lepidoptera: Noctuidae) em soja. *Enciclopédia Biosfera*, v. 9, n. 17, p. 2049-2063, 2013. Disponível em: <<http://hdl.handle.net/11449/137603>>. Acesso em: 14 set. 2018.
  16. Santos, O. S. *A Cultura da Soja, 1: Rio Grande do Sul-Santa Catarina-Paraná*. 2. ed. – São Paulo: Globo, 1995.
  17. Moreira, H. J. C. and Aragão, F. D. *Manual de Pragas da Soja*. Campinas: FMC Agricultural Products, 2009. 144p.
  18. Sosa-Gomez, D. R., Corrêa-Ferreira, B. S., Hoffman-Campo, C. B., Corso, I. C., Oliveira, L. J., Moscardi, F., Panizzi, A. R., Bueno, A. F., Hirose, E. and Roggia, S. *Manual de identificação de insetos e outros invertebrados da cultura da soja*. Londrina, PR: Embrapa Soja, 2010 (Documentos 269).
  19. Ávila, C. J. *Pragas da soja e seu controle*. 2017. Disponível em: <<https://pragas.cpao.embrapa.br/>>. Acesso em: 12 set 2018.
  20. Di Felippo, A. and Dias-da-Silva, B. C. *Dos olhares sobre o léxico: diferenças e semelhanças*. Disponível em: <[http://www.geterm.ufscar.br/ariani/Dos\\_olhares\\_sobre\\_o\\_lexico.pdf](http://www.geterm.ufscar.br/ariani/Dos_olhares_sobre_o_lexico.pdf)>. Acesso em: 21 set 2018.
  21. Dale, R. Classical approaches to natural language processing. In *Handbook of natural language processing*. 2010. Chapman and Hall/CRC, 2nd edition.
  22. Schabes, Y. and Waters, R. C. *Lexicalized Context-Free Grammar: A Cubic-Time Parsable, Lexicalized Normal Form For Context Free Grammar That Preserves Tree Structure*. Broadway, Cambridge: Mitsubishi Electric Research Laboratories. Technical Report 93- 04. June. 30p. (1993).
  23. Goddard, C. and Schalley, A. C. Semantic analysis. In *Handbook of natural language processing*. Chapman and Hall/CRC, 2nd edition, 2010.
  24. Muller, D. N. *Processamento de Linguagem Natural*. Universidade Federal do Rio Grande do Sul. Technical Report, 2003.
  25. EMBRAPA. *Soja*. Disponível em: <<https://www.embrapa.br/soja/publicacoes/>>. Acesso em: 19 set 2018.
  26. Moreno, F. C. *Visual Tahs: Ferramenta para analisar a eficácia de buscas das funções hash em um Léxico para Língua Natural*. PPGCC da UEL. Dissertação de Mestrado. Universidade Estadual de Londrina, 2017. 202f.