



# A modular framework for ontology learning from text in Portuguese

Norton Coelho Guimarães <sup>1\*</sup>, Cedric Luiz de Carvalho <sup>2</sup>

<sup>1</sup> Instituto Federal Goiano,

<sup>2</sup> Instituto de Informática, Universidade Federal de Goiás.

\*Corresponding author. E-mail: norton.guimaraes@ifgoiano.edu.br

**ABSTRACT.** Research on ontology learning has been carried out in many knowledge areas, especially in Artificial Intelligence. Semi-automatic or automatic ontology learning can contribute to the field of knowledge representation. Many semi-automatic approaches to ontology learning from texts have been proposed. Most of these proposals use natural language processing techniques. This paper describes a computational framework construction for semi-automated ontology learning from texts in Portuguese. Axioms are not treated in this paper. The work described here originated from the Philipp Cimiano's proposal along with text standardization mechanisms, natural language processing, identification of taxonomic relations and techniques for structuring ontologies. In this work, a case study on public security domain was also done, showing the benefits of the developed computational framework. The result of this case study is an ontology for this area.

**Keywords:** Semiautomatic Ontology Learning, Public Security Ontology, Natural Language Processing, Taxonomic relations.

**DOI:** <https://doi.org/10.33837/msj.v3i3.899>

Received August 15, 2018. Accepted October 15, 2020.

Associate editor: Anderson Rodrigo da Silva

## INTRODUCTION

One of the main challenges of Computer Science is to turn computers into machines that can learn by themselves. To do so, it will be required computers to have capacities that allow them to simulate, in some way, the human learning process. Some researchers have conducted studies on the field of computational learning with ontologies. Ontologies can store knowledge in their structure. They can be visualized as graphs of concepts and relations, where concepts are the nodes and relations are edges between nodes (Wong et al., 2012). One of the biggest problems faced by ontology engineers is precisely the ontology learning problem (Basegio & de Lima, 2006).

In Computer Science, ontologies serve as metadata schema, providing a concept-controlled vocabulary (Maedche & Staab, 2001). They can be composed of concepts, relations, concept and assertion instances and should be comprehensible to agents and other computational entities (Gruber, 1995). The effective man/machine communication through ontologies requires use of a specific language (Wong et

al., 2012).

Ontology structuring is an important step in knowledge-based system development, as it allows knowledge formalization and sharing between humans and computer systems (Lopes et al., 2012). Ontology use is important to reduce the ambiguity problem<sup>1</sup> existing in texts and serve as a concept dictionary within a given domain (Maedche & Staab, 2001).

The first stage of ontology learning processes is the choice of the corpus<sup>2</sup> to be used as a data source. The corpus quality and richness are fundamental for the performance of the extraction of information process (Lopes et al., 2009).

Generally, in the ontology learning process, three proposals are accepted: totally automated, not automated (manual) and semi-automated, the first one can be considered utopian, the second one is inefficient and the third one seems to be the best option (Cao et al., 2012). To semi-automate the ontology learning process, it is necessary to define a source of information where knowledge will be extracted from Ghisi et al. (2012). According to Cimiano (2006), the term ontology learning was originally used by Maedche & Staab (2001) to describe the process of acquiring knowledge from data. Here, different

Copyright © The Author(s).

This is an open-access paper published by the Instituto Federal Goiano, Urutai - GO, Brazil. All rights reserved. It is distributed under the terms of the Creative Commons Attribution 4.0 International License.



<sup>1</sup> Meaning uncertainty or inexactness in languages (Press, 2010).

<sup>2</sup> A corpus can consist of unstructured text (free format), with any format (HTML, XML, ...), or even with some kind of morphosyntactic or semantic annotation (Moraes & Lima, 2012).

subjects act in a complementary way, working with different types of data (unstructured, semi-structured or totally structured) (Maedche & Staab, 2001). However, there are gaps in providing integrated solutions to support ontology learning from texts. Philipp Cimiano proposed to organize the ontology learning process in many steps (Cimiano, 2006).

This article presents a computational framework<sup>3</sup> to help software engineers to introduce ontology learning capabilities in their applications. We do not propose new approaches to learn ontologies. Our purpose is to integrate third-party solutions in just one tool, following Cimiano's process. So, our framework works like a production line, where in each step a suitable solution can be used. If a better solution for a step is developed, it can be easily added to the framework to replace a previous solution in use.

To show the framework usefulness, we conducted an experiment on the public security context. In this work, we only considered ontology learning from text. So, for simplicity, when it is written "ontology learning", we mean "ontology learning from text".

The rest of this paper is organized as follows: Section 2 presents some related works; Section 3 presents the developed framework; Section 4 describes the framework evaluation and the results obtained. Finally, Section 5 presents conclusions and suggestions for future work.

## RELATED WORK

Baségio & de Lima (2006) proposed a semi-automatized approach to ontology learning from texts. He works specially on the phases of concept and taxonomic relation extraction for the Brazilian Portuguese. In his work, each step (relevant terms, compound terms and hearst-pattern-based relations) was validated manually by specialists.

Junior (2008) proposed a Java plug-in to help in semi-automatized ontology learning from text in Brazilian Portuguese, integrated to *Protégé*. Motta (2009) proposed a semi-automated process to ontology learning from text in Portuguese language. Zahra (2009) proposed a web application to semiautomatic extraction of ontologic structures from text in Portuguese language.

Gonçalves et al. (2011), proposed an application to ontology learning from text, focused on identification of concepts and relations. They presented an approach based on graphs to identify concepts and a concept analysis to get relations.

Lopes et al. (2012), proposed an application to extract relevant terms from a corpus in Portuguese language. They developed a Java tool that incorporates mechanisms for identifying relevant terms,

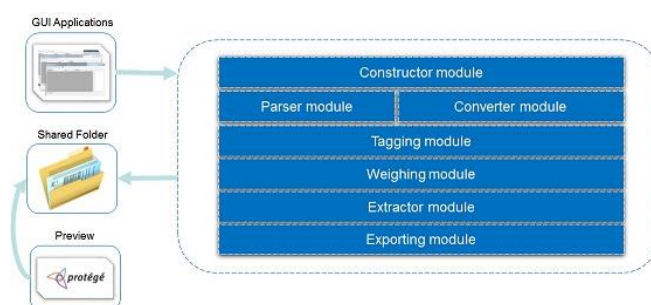
morphosyntactic labeling, agreement, term frequency, conceptual markings and hierarchy of concepts.

## THE PROPOSED FRAMEWORK

A framework is a set of cooperating classes that makes up a software project reusable Gamma et al. (1995). Thus, one can use object orientation in its structure targeting software development artifacts reuse. Being an incomplete system, it can be adapted to implement complete applications in a given domain, reducing the effort in deploying applications (Fayad et al., 1999).

In this section, we detail the framework design stages and architecture. The framework takes its input (textual documents), performs the tasks required for ontology learning and produces a textual document, encoded in OWL, as output. To view the ontological structure generated, the *Protégé* tool can be used.

The framework developed in this work was called *Sabença*<sup>4</sup>. It has a modular architecture, as shown in Figure 1. Third-party components have also been integrated to it. Each module is described in the following.



**Figure 1.** Framework modular architecture. Each module has its function based on a stage of the ontology learning process.

1. **Constructor module:** the constructor module performs the facade function of the framework. It also performs the following functions: controls communication flow between modules, receives calls from external applications, controls used directory mapping, controls parameterization and customization, and controls all document importation.
2. **Parser module:** the parser module preprocesses the unstructured textual documents. It has the following functions: lists all documents in the shared directory, selects and forwards documents to text conversion according to their file type and identifies the password protected documents.
3. **Converter module:** The converter module converts textual document into "pure text document"<sup>5</sup>. This standardization requires identifying the textual structures without format

<sup>3</sup> We will use simply "framework" in the rest of this paper.

<sup>4</sup> A synonym for wisdom in Portuguese.

<sup>5</sup> .txt format - in ASCII encoding.

markers. Text formatting removal can cause loss of valuable information because formatting can contain structures that may indicate relations between terms (Baségio & de Lima, 2006). This module generates a converted file for each document processed. For this, we have used third-party components that perform this type of conversion. To convert .doc and .docx document type, we have used Apache POI 3.10 (Surhone et al., 2010), and to convert the .pdf document type, we have used iText 5.5.2 (Lowagie, 2010).

4. **Tagger module:** the tagger module performs morphosyntactic labeling<sup>6</sup>. The morphosyntactic labeling requires marking terms with annotations. It depends on the language. In this module, it is used the models available in the OpenNLP site for Portuguese Opennlp (2010). A labeled file is generated for each document. In this work, we have used the Apache OpenNLP 1.5.3 and its models. According to our experiments, this is the tool best suited to the purpose of this work. Apache OpenNLP uses models that have multi-language support. For Portuguese, it has the models Tokenizer, Sentence Detector and POS Tagger. The models Chunker, Name Finder are not available for Portuguese. So, to have all OpenNLP features in Portuguese, we had to generate the models that were missing. To do this, we followed the directions in the OpenNLP site and trained models “pt-chunker.bin” and “ptner-person.bin”. The next step was to create the least relevant term list. At this stage, it was used a stopwords list. This list contains the domain-irrelevant semantic meaning words, such as articles, pronouns, prepositions, conjunctions, numerals, adverbs, non-alphabetic characters (numbers and symbols) and undefined terms (Lopes et al., 2012). A single stopwords file is generated as output of this module.
5. **Weigher Module:** the weigher module performs selected term weighing. In this module, the method Term Frequency-Inverse Document Frequency (TF-IDF) is used. It is based on the absolute term frequency for each term. The lemma<sup>7</sup> is used to avoid repeating terms. A single file to store this module results is created. TF-IDF determines the frequency of terms in a document. Terms with high occurrence frequency in a set of documents are considered relevant to the domain.
6. **Extractor module:** the extractor module extracts the taxonomic relations. For this, the Hearst (1992) method and complex terms<sup>8</sup> are used. A priori,

the hearst patterns are prepared for use in English. An adaptation was required for Portuguese. In this work, it was chosen the adaptation proposed by Baségio & de Lima (2006). After applying the hearst method, hyponymy and hypernymy relations between terms were discovered. A single file containing the found relations was generated. For this, we have used the Apache JENA 2.12 component (Jena, 2014). Complex terms are constructed from n-gram generation. We do not found any scientific study that used 5 or higher level n-grams. So, we considered that 4-level n-grams are sufficient to find complex terms that are valuable in ontology learning process. Then, a single file, containing the found 4-grams, was generated. After n-gram generation, complex terms are found. After that, a single file, containing the found compound terms is generated. To find taxonomic relations, relevant terms, compound terms and the hearst method relations on a single taxonomic structure were used. The found relations are stored in a single file.

7. **Exporter Module:** the exporter module builds ontological structures in OWL language. In this module, the taxonomic relations generated in the previous module are used. An ontological structure is generated containing simple terms, compound terms and hearst pattern based relations. Finally, two files are generated to save the ontology produced by the framework: “ontology.owl” e “ontology.rdf”.
8. **Replacing Modules in the Framework:** the *Sabença* framework uses dependency injection to maintain low coupling and high cohesion between classes and components and allow flexible modification in modules. So, modules can be replaced by another ones or new modules can be added easily. Dependency injection is a design pattern that allows to run a class within another class. Using dependency injection facilitates replacing framework existing resources by others resources. The framework is dependent on the property file “sabença.properties”. The configuration file should be properly modified in order to use other components.

We understand that some phases of ontology learning process require human intervention. Table 1 shows all the stages implemented, based on the Philipp Cimiano’s proposal (Cimiano, 2006).

**Table 1.** Steps of ontology learning implemented in this work.

Stage	Module (automatic)
Pre-processing	Constructor, Analyser and Converter
Extraction of terms	Weigher
Identification of concepts	Extractor
Extraction of relations	Extractor
Structuration of ontologies	Exporter
Validation	-
Visualization	-

<sup>6</sup> A morphosyntactic labeling is a task of NLP (Natural Language Processing) that labels words, phrases and punctuation in a sentence as nouns, adjectives, verbs, adverbs, etc.

<sup>7</sup> A lemma is a term, ignoring the tense if it is a verb, the word gender, plural, etc.

<sup>8</sup> Complex terms are found from n-gram generation.

## EVALUATION AND RESULTS

According to Wong et al. (2012), validation can be grouped into some categories: contextual approach, coverage approach, comparison-based approach, evaluative approach, structural approach and functional approach. An evaluative approach, where domain experts evaluate the ontology layers [terms, concepts, relations and axioms (optionally)], can also be used. In this work, we have evaluated simple terms, compound terms and taxonomic relations. We have followed the Janez Brank's recommendation to evaluate the different layers of ontology separately, rather than trying to directly evaluate the ontology as a whole (Brank et al., 2005). With this, we can determine the term relevance for the domain and if they are conceptually correct (Wong et al., 2012).

To validate the developed framework, a case study was conducted on public security area. A manual validation was done by experts in the area. Each expert chose terms he considered as correct. Those terms marked as correct by all experts were selected. After that, the values generated in the framework were manually replaced by those selected by experts. Then, these terms were used in the ontological structure generation.

The evaluation process took place by sending an email with a worksheet attached for each expert. The evaluation worksheet contained three tabs (hears pattern, relevant terms and compounds terms) with two columns (terms and correct). Each expert should mark with an "X" in the column "correct" for each corresponding value only if he understands that the term or pattern belongs to the public security domain. In the body of the email, there were instructions on how to complete the evaluation worksheet.

The validated data served as input to the ontology structuring process. For this, we analyzed the information marked by each expert and made the selection of the terms that were common in all evaluated worksheet. Later, we manually replaced the values generated in the framework by the results selected in the evaluation process. After that, the ontological structure was generated.

### Textual Document Acquisitions

To conduct experiments on our case study, textual documents (.PDF, .DOC and .DOCX) were selected from the library of the military police academy of the Goiás/Brazil state and from *Revista Brasileira de Estudos em Segurança Pública* - REBESP9 (the Brazilian Journal of Studies on Public Security), composing a 152-document corpus, totaling 3958 pages. In this collection, there were scientific articles and monographs on public

security. Librarians of this police academy manually selected the documents.

In the case study, we observed the selected documents contained terms from other areas, not related to public security, such as education, marketing, management and others.

### Data analysis

The framework only converts unprotected documents. During importation, 8 protected files were found. Thus only 144 documents were converted.

### Term identification

After conversion, the morphosyntactic labeling task is started. Morphosyntactic labeling is the ontology learning process main stage and its results influence in subsequent modules. 1283064 terms were found in this phase. To eliminate non-relevant domain terms, a list of stopwords was used. The stopword list was automatically generated after labeling terms. At the end, 5934 total non-relevant terms for the domain were found. To define relevant terms, all existing terms in stopword list were removed. Then, terms were grouped by nouns. After that, terms were weighed using the TF-IDF method. The weighing resulted in 20076 extracted terms.

The TF-IDF measure returns a very large amount of non-relevant terms to be presented to the ontology engineer. There is no a standard pruning rule for selecting them Basegio & de LIMA (2006). Therefore, the list of terms was analysed and classified in TF-IDF index descending order. Terms with index  $TF-IDF = 0.0$  were identified. These terms were characterized as not relevant to the domain. To select relevant terms,  $TF-IDF^{10}$

0.1<sup>11</sup> was used as the minimum frequency. This resulted in a list reduction. Thus, it was obtained 314 relevant terms for the domain.

### Compound term identification

The Markov method was used to identify compound terms. In this paper, only the compound terms contained in the relevant term list extracted in the previous section were selected. 1516 compound relevant term were extracted, out of the 18678 compound terms generated.

Table 2 shows the compound relevant term amount, classified by rules. The column "Rules" shows the rules used to define compound terms, e.g., "*polícia de primeiro mundo*" ("first-world police") will be selected by the rule "sub + prp + sub + adj", where

<sup>9</sup> Available on the address

<http://revista.ssp.go.gov.br/index.php/rebsp>, last seen in May 2017.

<sup>10</sup> The TF-IDF minimum frequency value can be changed in the property file, according to the ontology engineer needs.

<sup>11</sup> TF-IDF index = 0.1 means that there are at least two word repetitions in the corpus.

“sub”= noun, “prp” = preposition and “adj”= adjective. The column “Qty” shows the amount of each term found for the given rule.

**Table 2.** Amount of compound terms found by rule.

Rules					Qty
SUB	ADJ	PRP	SUB	ADJ	0
SUB	ADJ	PRP	SUB		235
SUB	PRP	ADJ	SUB		129
SUB	PRP	SUB	ADJ		10
SUB	PRP	SUB			550
SUB	ADJ				592

### Hearst Patterns

Hearst patterns were used to find hypernymy and hyponymy relations between terms. In this work, the pattern rules were applied. This resulted in 104 taxonomic relations. The amounts found by each hearst pattern can be seen in Table 3. The column “Pattern” shows the hearst patterns adapted by Baségio & de Lima (2006). The column “Qty” shows the amount found from the hearst pattern application in the corpus documents.

### Domain ontology structuring

To identify taxonomic relations from the phrase core of the compound terms, the relations between each compound term and the relevant term are used (Baségio & de Lima, 2006). Therefore, the relations between compound terms and relevant terms are first searched. The compound term found receives a <TYPE> annotation, indicating its taxonomic relation.

Next, the hearst pattern results are added to the existing taxonomic relations. Finally, the taxonomic relations are exported to the OWL Lite and to the RDF Triple-n pattern. The exporting result is stored into a OWL file, which can be viewed using the *Protégé* tool. The resulting ontology is composed of 2 levels and have 194 entities. We have limited the number of levels of the ontology just to get it generated faster in the computer we performed the experiments. This is not a limitation of the framework.

**Table 3.** Taxonomic relation amount found by the *hearst patterns*.

Pattern	Qty
SUB as (SUB,)*(ou-e) SUB	33
SUB such as (SUB,)*(ou-e) SUB	3
such SUB as (SUB,)*(ou-e) SUB	0
SUB , SUB *, or another(s) SUB	6
SUB , SUB *, and another(s) SUB	59
SUB, including SUB, * (ou-e) SUB	0
SUB , especially SUB, * (ou-e) SUB	0
SUB , mainly SUB, * (ou-e) SUB	1
SUB , particularly SUB, * (ou-e) SUB	0
SUB , in particular SUB, * (ou-e) SUB	2
SUB , in particular SUB, * (ou-e) SUB	0
SUB , in a especial way SUB, * (ou-e) SUB	0
SUB , especially SUB, * (ou-e) SUB	0

### CONCLUSION

An annotated corpus was generated from textual documents manually selected by librarians. The corpus used in this work has more than 1 million words extracted from textual documents. In this work, various frameworks available for text conversion, labeling, stemming, proper name identification and writing ontological language, were tested. Most of these frameworks were designed to operate just in English. Only tools that produce satisfactory results for the Portuguese language were chosen. The used tools impacted in the result quality.

It was found that identification of taxonomic relations from compound terms with n-grams of 4-levels produced better results in comparison with the hearst patterns. The five stages of ontology learning implementation allowed us to structure an ontology in Portuguese. However, we understand that ontologies are representations of shared realities and, as the knowledge acquired from the reality changes, the structure of ontologies should be changed as well. Therefore, there are no complete ontologies.

The *Sabença* framework allows the inclusion of third-party tools and customization of developed classes. In this version 0.1, we use third-party tools that perform the necessary steps to implement learning ontologies from texts in Portuguese. We were careful to choose the tools that were more suited to this objective. But, we did not use third-party tools in all the stages of ontology learning process. Some stages had to be implemented.

The main contribution of this work is to facilitate the development of applications that require the use of ontologies. The framework was developed mainly from the Philipp Cimiano's work.

The case study we conducted produced results indicating that the developed framework can facilitate the process of building domain ontologies from texts. The *Sabença* framework can be improved by constructing new modules to support extracting text from audio and video, new modules to handle other idioms and to non-taxonomic relation recognition. It could be interesting too to build a new module to merge ontologies.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### REFERENCES

- Baségio, T. L. & de Lima, V. L. S. (2006). Semi-automatically building ontological structures from portuguese written texts. In Vieira, R., Quaresma, P., Nunes, M. d. G. V., Mamede, N. J., Oliveira, C., and Dias, M. C., editors, *Computational Processing of the Portuguese Language*, volume 3960 of *Lecture Notes in Computer Science*, 208–211. Springer Berlin Heidelberg.
- Brank, J., Grobelnik, M., and Mladenic, D. (2005). A survey of

- ontology evaluation techniques. In Proc. of 8th Int. multi-conf. Information Society, 166-169.
- Cao, Y., Wang, X., Zhang, F., and Yang, W. (2012). Ontology-based domain knowledge acquisition technology. In Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on, 2, 487-490.
- Cimiano, P. (2006). *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Fayad, M. E., Schmidt, D. C., and Johnson, R. E. (1999). *Implementing Application Frameworks: Object-oriented Frameworks at Work*. John Wiley & Sons, Inc., New York, NY, USA.
- Gamma, E., Helm, R., Johnson, R., and Vlissides, J. (1995). *Design Patterns: Elements of Reusable Object-oriented Software*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Ghisi, F. B., Fachin, G. R. B., Santos, M. H. d., Sell, D., and Rados, G. J. V. (2012). A reference ontology for digital scientific journals applied to systematic literature review processes. *Transinformação*, 24, 91-101.
- Gonçalves, G., Wilkens, R., and Villavicencio, A. (2011). Sistema de aquisição semi-automática de ontologias. In Vieira, R., Guizzardi, G., and Fiorini, S. R., editors, *ONTOBRAS-MOST*, of CEUR Workshop Proceedings, 776, 189-194. CEUR-WS.org.
- Gruber, T. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal Human-Computer Studies*, 43(5-6), 907-928.
- Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In Proceedings of the 14th Conference on Computational Linguistics - Volume 2, COLING'92, 539-545, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jena (2014). The apache jena project. version 2.12, 2014. Available in: <http://jena.apache.org/>. Last seen in December 2014.
- Junior, L. C. R. (2008). *Ontolp: Construção semi-automática de ontologias a partir de textos da língua portuguesa*. (Dissertação de Mestrado em Computação Aplicada) - Universidade do Vale do Rio dos Sinos, São Leopoldo. Available in: <http://goo.gl/LFutcl>. Last seen in April 2013.
- Lopes, L., Fernandes, P., Vieira, R., and Fedrizzi, G. (2009). Exatolp - an automatic tool for term extraction from portuguese language corpora. In Proceedings of the LTC'09, Pozna, Poland.
- Lopes, L., Vieira, R., Fernandes, P., and Couto, G. (2012). Exatolp - an automatic tool for term extraction from portuguese language corpora. In International Conference on Computational Processing of the Portuguese Language - PROPOR, 45-47.
- Lowagie, B. (2010). *iText in Action*. Manning Publications Co., Greenwich, CT, USA.
- Maedche, A. & Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2), 72-79.
- Moraes, S., and Lima, V. (2012). Combining formal concept analysis and semantic information for building ontological structures from texts : an exploratory study. In Calzolari, N., Choukri, K., Declerck, T., Dog'an, M. U., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., editors, Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), pages 3653-3660, Istanbul, Turkey. European Language Resources Association (ELRA). ACL Anthology Identifier: L12-1556.
- Motta, E. N. (2009). *Preenchimento semi-automático de ontologias de domínio a partir de textos em língua portuguesa*. (Dissertação de Mestrado em Informática) - Centro de Ciências exatas em tecnologia, Universidade Estadual do Rio de Janeiro, Rio de Janeiro. Available in: <http://www2.uniriotec.br/ppgi/banco-de-dissertacoes/ppgi-unirio/ano-2009/preenchimento-semi-automatico-de-ontologias-de-dominio-a-partir-de-textos-em-lingua-portuguesa/view>. Last seen in April 2014.
- OpenNlp (2010). Toolkit for the processing of natural language text. version 1.5, 2010. Available in: <http://opennlp.apache.org>. Last seen in November 2014.
- Press, O. U. (2010). *Oxford American Desk Dictionary & Thesaurus*. Oxford University Press, USA.
- Surhone, L., Tennoe, M., & Henssonow, S. (2010). *Apache Poi*. Betascript Publishing.
- Wong, W., Liu, W., & Bennamoun, M. (2012). Ontology learning from text: A look back and into the future. *ACM Comput. Surv.*, 44(4), 20:1-20:36.
- Zahra, F. M. (2009). *Poronto - ferramenta para construção semiautomática de ontologias em português*. (Dissertação de Mestrado em Tecnologia em Saúde) - CCBS, PUCPR, Curitiba. Available in: <http://www.dominiopublico.gov.br/>. Last seen in April 2014.

---

To cite this paper:

Guimarães, N. C., & de Carvalho, C. L. (2020). A modular framework for ontology learning from text in Portuguese. *Multi-Science Journal*, 3(3), 37-42. DOI: <https://doi.org/10.33837/msj.v3i3.899>