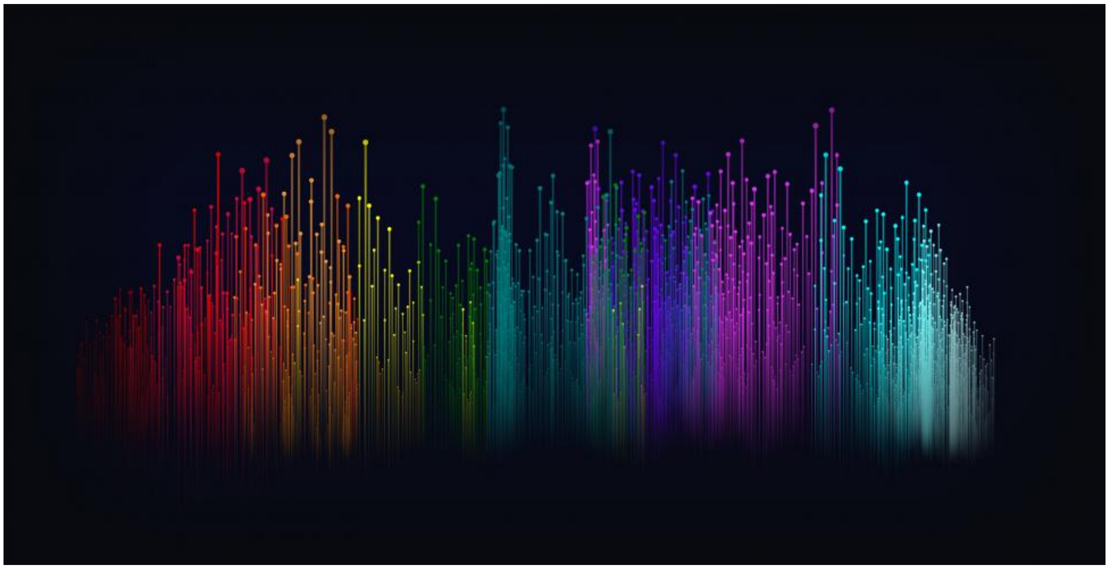
Projects and Systems based in OpenIE and NER - Overview and functionality tests.

Emanuel Matos - Student PhD Informatics

2020-09-22



Supervisor:

- PhD António Teixeira - Universidade de Aveiro
- PhD Mario Rodrigues - Universidade de Aveiro

Contents

1 Introduction 4

2 Method 5
- 2.1 Installation . . . 5
- 2.2 Dataset . . . 5
- 2.3 Evaluation . . . 5

3 Selected Tools 6
- 3.1 NER . . . 6
- 3.2 OpenIE . . . 6
- 3.3 AllenNLP . . . 6
- 3.4 Linguakit . . . 7
- 3.5 DeptOIE . . . 7

4 Datasets 8

5 Results 9

6 Conclusion 9

A Apendice 10
- A.1 Projects . . . 10
  - A.1.1 Linguakit . . . 10
  - A.1.2 AllenNP . . . 11
  - A.1.3 Stanford CoreNLP . . . 12
  - A.1.4 KnowItAll . . . 13
  - A.1.5 NTLK - PyThon . . . 14
  - A.1.6 Others NLP Projects . . . 15
  - A.1.7 The Top 40 Information Extraction Open Source Projects . . . 15
  - A.1.8 Open Information Extraction - Softwares . . . 15
  - A.1.9 Top 26 Free Software Text analysis-Text Mining . . . 16

B Data Set 17
- B.1 Datasets for NER in English . . . 17
- B.2 Datasets for NER in Portuguese . . . 17
  - B.2.1 HAREM . . . 17
  - B.2.2 SIGARRA . . . 17
  - B.2.3 Comparative . . . 17

C Evaluation Propose 18

List of Tables

1 Systems in General 10
2 Systems Base KnowItAll 14
3 Projects NLP 15
4 The Top 40 15
5 Open Information Extraction - Softwares 16
6 Free Software Text analysis-Text Mining 16

Abstract

The aim of this study was to evaluate and understand metrics and comparisons of some of the main and most recent Projects or Software based on OpenIE and NER since its implementation until comparing them with traditional metrics and others. The main language in focus was the Portuguese language.

1 Introduction

The goals of study NER and OpenIE softwares are direct relationship with understand about the enviromnent develop , straightforward and easy to understand – but they aren't always easy to meet. This is because there are so many different ways to approach software engineering and so many outcomes that are possible. While we do have best practices and there are standards in place, every software engineer has a different approach and sometimes they don't always mesh well with other members of an IT team.

- NER
  The Named Entity Recognition (NER) is within the natural processing of language, being a process that from a sentence or part of it, this text is captured and verified its possibilities of finding entities that can be of the most varied categories such as names, organizations , places, quantities, monetary values, percentages, etc. NER algorithms can be trained on the basis of ML to extract the entities mentioned above. The NER for having a simple but effective approach, is also used as a kind of "Pre-Processing" for Open IE.

- OpenIE
  The Open Information Extraction (Open IE) is a way to extract information based on obtaining non-predefined and independent domain relations of a text, this field of study for not having rigid search rules makes this process rich in obtaining results where not expected, this is your greatest strength.

2 Method

In the Oriented study I had the opportunity to get to know various types of processes and systems for Open Information extraction (Open IE), old, traditional and modern, as well as Entity Recognition (NER).

From the Supervised Study I was able to list the possible processes and software so that we could start a research and a more in-depth study. As a result, we will have an individual and comparative performance evaluation between at least 3 softwares, in table 1.

The list in this table was designed to guide us in the search for the most recent software and / or that were available for free and accessible use. We also seek to orient ourselves so that the technology used is within the concepts of Open IE and NER.

Of course, we left out many that we could test, but we will evaluate the top 3 on the list, this does not mean that those we do not evaluate are the worst evaluation in terms of innovation or technology or accessibility, but because we have time and scope restrictions for this job.

2.1 Installation

The software was installed in different environments, tested on Windows, Mac OS and Ubuntu. In each of them we had different difficulties, from adapting the software to the operating system, to adjusting the environment variables within the base system. Each has a feature, one runs in Perl, one in Python and one in Java. In dealing with each assessment, we will detail the learning processes that were necessary and their difficulties for each OS and software.

2.2 Dataset

The choice of data sets were, in principle, those made available by each of the software for testing, and then we opted for one that has already been thoroughly tested and lastly a few phrases without training and the Tourism environment for effect and comparison " balanced "

2.3 Evaluation

The assessment went through qualitative and quantitative items such as statistical metrics of software adequacy and performance metrics for the environment. Each part of this evaluation will have its details in the evaluation chapter.

3 Selected Tools

The NER and OpenIE tools play an important role in modern technology and information systems. We have tools based on Java, Python, Perl and other languages. Each language is linked to the practicality that each group or person has with the tool. Currently, there is a tendency to develop without worrying about how to leave a general framework for the construction of a pipeline and this can be considered one of the challenges within NLP, the identification, structuring and adherence of a language / system to the use of OpenIE . The focus and result of OpenIE should not be in the system itself, but in what it can offer information and intelligence to be extracted from a text or from a pre-processed text. This study seeks to go through the work with software from different bases and with different operating systems to identify work difficulties and facilities. Because we believe that the main focus of these tools should be the delivery of Information and Intelligence regardless of the user's ability with the tools.

3.1 NER

Linguakit and AllenNLP were used as basic tools for NER. AllenNLP has Python as its programming base, while Linguakit was developed in Perl and its code in JAVA.

3.2 OpenIE

AllenNLP, DeptOIE were basic tools for Open IE in this study. AllenNLP for using Python proved to be more versatile, since DeptOIE has its development in JAVA and its code requires compilation.

3.3 AllenNLP

AllenNLP was tested using its internal guide, initially the idea was to run Python code in a pre-defined environment, but due to time and training limitations I chose to use the guide which is available at `https://guide.allennlp.org` .

The guide provides conceptual and technical explanations that can be used, but if you want to create paths different from the path proposed by the guide, it may not be as enlightening as staying on the original trail.

I want to come to believe that as a guide, I should be more friendly, I put myself in the position of a professional who wants to have an answer or information about a certain text, without having to know about Pythorc or Instances or any other term within the world of NLP and I realized that there is no possibility to get this information directly.

But considering everything that exists on both GIT and the AllenNLP website, there are excellent conditions for adaptation for the professional end user.

- Language = Portuguese, English, Spanish and more 13 languages https://spacy.io/models
- SO - Windows
- Dependencies - Pythorch / SpaCy
- Installation - Require several adjusts included CUDA.

3.4 Linguakit

Linguakit was one of the software or set of processes chosen for its diversity, as it works in natural language. This software is evaluated by Relationship Entities (NER), we have tagging, we can structure phrases or a text in order to capture entries to assemble answers to questions that are not answered directly.

- Languages - Portuguese, Spanish, English, and Galician
- SO - Windows/MacOS/Ubuntu
- Dependencies - CPAN Perl (some modules)
- Installation - Require Administration profile, follow README.md to installation

3.5 DeptOIE

DeptOIE was a software chosen because it is a very recent process and has the Portuguese language as the base language for the study of Relationship of Entities and Extraction of Open Information. It was recently presented at a conference on Extraction of Open Information.

- Language - Java
- SO - MacOS
- Dependencies - Java
- Installation

4 Datasets

In this chapter, we list which possible data sets we can work with, these data sets must be easily accessible and available to the general public. What we realized is that when searching these data sets, only the data sets that are part of the guides for each software were available and easy to process. Below is the list of possible data sets and their characteristics.

Drop AllenNLP 2019 Dataset AllenNLP https://allenai.org/data/drop

CONLL 2003 – Since 1999, CoNLL (Conference on Natural Language Learning), is an annual SIGNLL meeting. CoNLL 2003 stood out because 1,393 news items were made available in English and 909 in German. The English corpus is free, but the German corpus is not. You will have access a few days after sending the organizational and individual contract for free. The entities are noted with LOC (local), ORG (organization), PER (person) and MISC (miscellaneous).

https://www.clips.uantwerpen.be/conll2003/ner/

https://github.com/davidsbatista/NER-datasets/tree/master/CONLL2003

Ontonote-5.0 – Entities:Organization, Art Work, Numbers in word, Numbers, Quantity, Person, Location, Geopolitical Entity, Time, Date, Facility, Event, Law, Nationalities or religious or political groups, Language, Currency, Percentage, Product.

Ontonote 5.0 is a Dataset provide by LDC "The Linguistic Data Consortium is an open consortium of universities, libraries, corporations and government research laboratories. LDC was formed in 1992 to address the critical data shortage then facing language technology research and development. The Advanced Research Projects Agency provided seed funding for the Consortium and the National Science Foundation provided additional support via Grant IRI-9528587 from the Information and Intelligent Systems division. " - https://www.ldc.upenn.edu/about

At the beginning of July I signed up and waited for the email to be validated, it didn't arrive, I did the same procedure in the second week with my email from the University of Aveiro, and I also didn't get a response. At the beginning of the third week, I requested DataSets, LDC 2013T19 - Ontonotes 5.0, LDC99T42 TreeBank, LDC2008T19 The new York Times and LDC2006T06 ACE 2005 Multilingual with a Non-Member request.

GMB(Groningen Meaning Bank) – Entities: Natural Phenomenon, Person, Geographical, Organization, Art Work, Event, Time, Geopolitical.

NAACL 2019 – Entities: Organization, Person, Location, Geopolitical, Facility, Vehicles.

Wnut2017 – Entities: Location, Person, Product, Groups, Corporations, Creative.

5 Results

The results we expected were of 3 types, evaluation of the processes in the same dataset, evaluation of the processes with their own dataset and finally a text without any markup or annotation to compare the results of the NER in each dataset.

Due to the search for excellence, the time required for the adaptation between notebook and systems was longer than expected and that is why we only ran the software in its respective guides, which, already adapted, showed the simple operation of each one. 1 We then tried to obtain a standard data set, such as the Ontonote that the site said was available, just needing to register, so I did, the return was not fast and when we had access we were unable to download the data set.

As a final result, we installed only the 3 software, AllenNLP, Linguakit and DepOIE. They are working, but we have not made comparisons or developed any metrics.

6 Conclusion

As a conclusion of all the work, we had some learnings and some perceptions. Learning: The time spent learning how to install and process each software individually was underestimated, we did not take into account the learning curve for each environment. On the same physical equipment, the best alternative is to create virtual environments so that there are no installation / driver conflicts. Try to have a data set before choosing the software to be evaluated. Intuitions: We believe that based on the experience obtained in this learning opportunity, it was very relevant for the structuring and mapping of what we should look for in a future work, including the conception of the Doctoral Thesis script.

A Apendice

A.1 Projects

In the table 1 we have a first relation with some systems and / or groups of systems.

Table 1: Systems in General

System	HTTP	Access date
AllenNP	https://docs.allennlp.org/master/	07/Jun/2020
Linguakit	https://linguakit.com/en/full-analysis	10/Jun/2020
DptOIE	https://github.com/FORMAS/DptOIE	01/Ago/2020
BERT (Devlin et al., 2018) Relation Extraction (Soares et al., 2019)	https://tinyurl.com/yaen9y52	08/Jun/2020
CrossOIE	https://github.com/FORMAS/CrossOIE	01/Ago/2020
Stanford CoreNLP – Natural language software	https://tinyurl.com/yxq9sysp	07/Jun/2020
KnowItAll	https://github.com/knowitall	07/Jun/2020
Varios Projetos NLP	https://tinyurl.com/ybvv43jt	08/Jun/2020
The Top 40 Information Extraction Open Source Projects	https://tinyurl.com/yc76ewz7	08/Jun/2020
Open Information Extraction - Softwares	http://www.cse.iitd.ernet.in/~mausam/software.html	08/Jun/2020
Varios sistemas de inteligencia competitiva	https://tinyurl.com/yc5lrhmo	11/Jun/2020
Top 26 Free Software Text analysis-Text Mining	https://tinyurl.com/yb8sgjew	12/Jun/2020
NTLK - Python	https://www.nltk.org/book/ch07.html	07/Jun/2020

A.1.1 Linguakit

https://linguamatica.com/index.php/linguamatica/article/view/v9n1p2/391

In the link https://linguakit.com/en/full-analysis we can access the linguakit.

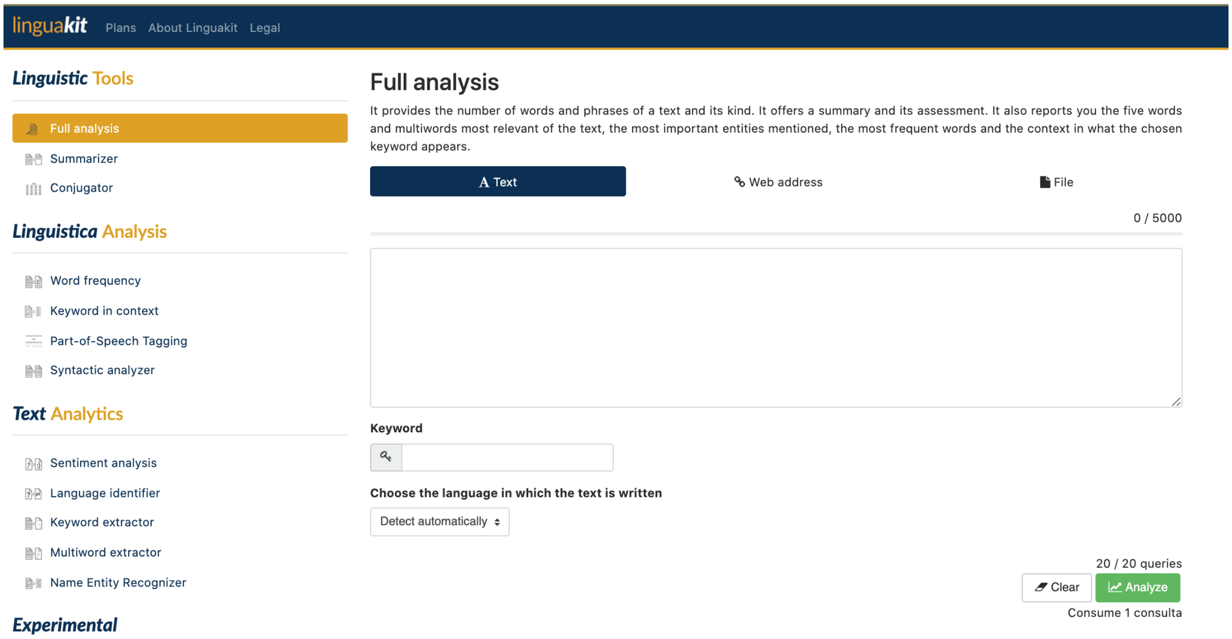
With Linguakit, it is possible to explore, analyze and obtain better information from texts and written documents.

This multilingual system, which includes, among other language tools, a summary, a sentiment analyzer or a keyword extractor that makes sense of a text, is intended for a wide range of users who make language a professional use, educational or general.

Linguakit was developed so that anyone with a linguistic interest can get the most out of written texts.

This platform presents its linguistic modules organized in four guiding sections: a first that addresses more generic aspects of the language with modules such as the conjugator or translator; a second, for a user profile more linked to the educational area, with modules such as the morphosyntactic marker or the analyzer; a third section aimed at communication and marketing professionals, such as the sentiment analyzer or the keyword extractor; and finally, an experimental section where Linguakit presents the new tools of the project.

Linguakit is an idea of Cilenis Language Technology that arises from the company's years of research in the area of Natural Language Processing (PLN).



Full analysis

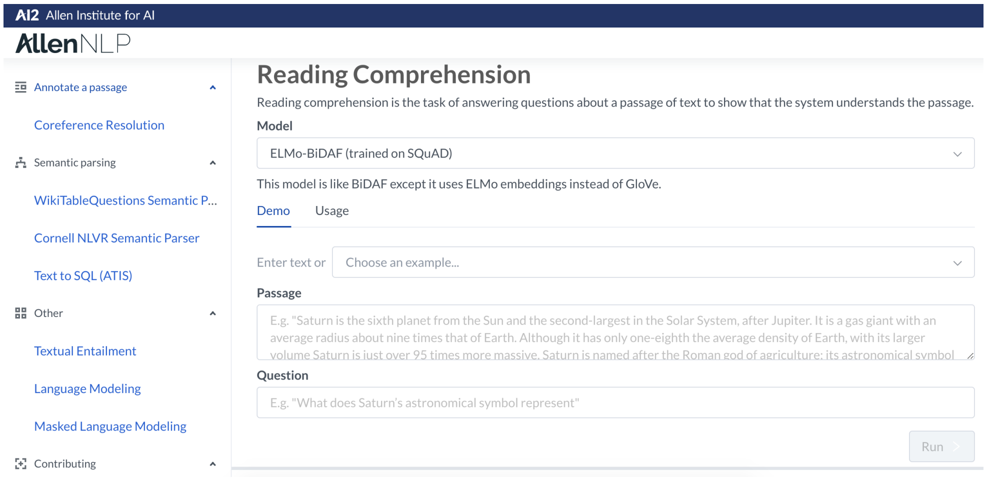
It provides the number of words and phrases of a text and its kind. It offers a summary and its assessment. It also reports you the five words and multiwords most relevant of the text, the most important entities mentioned, the most frequent words and the context in what the chosen keyword appears.

Choose the language in which the text is written: Detect automatically

Figure 1: Screen de (lin, 2020)

A.1.2 AllenNP

Temos o link de demo https://demo.allennlp.org/reading-comprehension e p link de entrada https://docs.allennlp.org/master/.



Reading Comprehension

Reading comprehension is the task of answering questions about a passage of text to show that the system understands the passage.

Model

ELMo-BiDAF (trained on SQuAD)

This model is like BiDAF except it uses ELMo embeddings instead of GloVe.

Enter text or Choose an example...

Passage

E.g. "Saturn is the sixth planet from the Sun and the second-largest in the Solar System, after Jupiter. It is a gas giant with an average radius about nine times that of Earth. Although it has only one-eighth the average density of Earth, with its larger volume Saturn is just over 95 times more massive. Saturn is named after the Roman god of agriculture; its astronomical symbol

Question

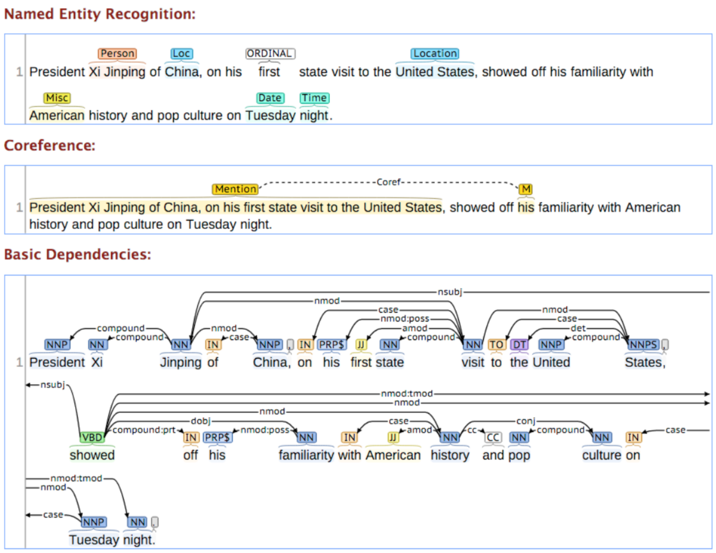
E.g. "What does Saturn's astronomical symbol represent"

Figure 2: Screen Demo de (Gardner et al., 2017)

A.1.3 Stanford CoreNLP

We have the link https://stanfordnlp.github.io/CoreNLP/ where we have the CoreNLP system.

Stanford CoreNLP provides a set of human language technology tools. It can provide the basic forms of words, their grammatical classes, be they company names, people, etc., normalize dates, times and numerical quantities, mark the sentence structure in terms of phrases and syntactic dependencies, indicate which substantive phrases refer to the same entities, to indicate feelings, to extract private or open class relations between mentions of entities, to obtain the quotes that people have said etc.



Named Entity Recognition:

1 President Xi Jinping of China, on his first state visit to the United States, showed off his familiarity with American history and pop culture on Tuesday night.

Coreference:

1 President Xi Jinping of China, on his first state visit to the United States, showed off his familiarity with American history and pop culture on Tuesday night.

Basic Dependencies:

1 President Xi Jinping of China, on his first state visit to the United States, showed off his familiarity with American history and pop culture on Tuesday night.

Figure 3: Areas (Manning et al., 2014)

A.1.4 KnowItAll

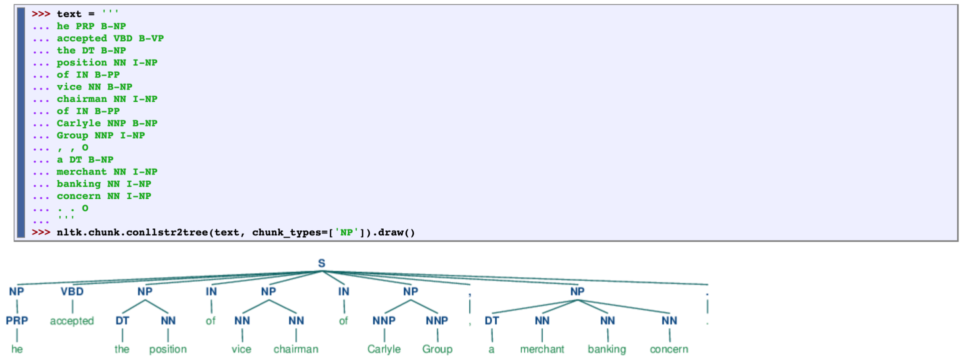
The link https://github.com/knowitall we have a several systems based into KnowItAll. In the table 2 we have some of systems.

Table 2: Systems Base KnowItAll

System	HTTP	Base	Data	Atualização
reverb	https://github.com/knowitall/reverb	Java	07/Jun/2020	update 2019
openie	https://github.com/knowitall/openie	Scala	07/Jun/2020	update 2018
implie	https://github.com/knowitall/implie	Scala	07/Jun/2020	update 2018
ollie	https://github.com/knowitall/ollie	Scala	07/Jun/2020	update 2018
srlie	https://github.com/knowitall/ollie	Scala	07/Jun/2020	update 2017
nlptools	https://github.com/knowitall/nlptools	Scala	07/Jun/2020	update 2017
chunkedextractor	hhttps://github.com/knowitall/chunkedextractor	Scala	07/Jun/2020	update 2017
common-scala	https://github.com/knowitall/common-scala	Scala	07/Jun/2020	update 2016
MultirFramework	https://github.com/knowitall/MultirFramework	Java	07/Jun/2020	update 2015

A.1.5 NTLK - PyThon

For Python, we have to work with NTLK, Multiparse and install some functions. In the link https://www.nltk.org/book/ch07.html there are explanations of how NTLK works. The http://www.nltk.org/ link also provides an overview of the documentation.



```
>>> text = '''
... he PRP B-NP
... accepted VBD B-VP
... the DT B-NP
... position NN I-NP
... of IN B-PP
... vice NN B-NP
... chairman NN I-NP
... of IN B-PP
... Carlyle NNP B-NP
... Group NNP I-NP
... , , O
... a DT B-NP
... merchant NN I-NP
... banking NN I-NP
... concern NN I-NP
... . . O
... '''
>>> nltk.chunk.conllstr2tree(text, chunk_types=['NP']).draw()
```

Figure 4: Example

A.1.6 Others NLP Projects

——-

We have a link, https://paperswithcode.com/task/relation-extraction/codeless some systems. In the table 3 some related projects.

Table 3: Projects NLP

System	HTTP	Base	Access date	Version
BERT-Relation-Extraction	https://tinyurl.com/yaen9y52	Python (3.6+), PyTorch (1.2.0), Spacy (2.1.8)	08/Jun/2020	update 2019
SpanBERT	https://tinyurl.com/y87k6wad	Python/PyTorc	08/Jun/2020	update 2020

A.1.7 The Top 40 Information Extraction Open Source Projects

——-

We have a link, https://awesomeopensource.com/projects/information-extraction, others projects in a table 4

Table 4: The Top 40

System	HTTP	Base	Data	Atualização
MitIE	https://github.com/mit-nlp/MITIE	C, C++, Java, R, or Python 2.7.	08/Jun/2020	update 2020
Annotated Semantic Relationships Datasets	https://tinyurl.com/yafd8mxs	Relações (Portugues e Ingles)	08/jun/2020	update 2020
Nlp Cube	https://tinyurl.com/ya79q8u6	Python3 (Multilinguas)	08/jun/2020	update 2019

A.1.8 Open Information Extraction - Softwares

——

In the link: http://www.cse.iitd.ernet.in/~mausam/software.html where we have several systems, we list some in the table 5

Table 5: Open Information Extraction - Softwares

System	HTTP	Base	Access date	Version
OpenIE5-STDAlone	https://tinyurl.com/yay4e9j7	Java/Python	09/Jun/2020	update 2018
OREO	https://tinyurl.com/yamtcvtr	?	09/Jun/2020	update 2015

A.1.9 Top 26 Free Software Text analysis-Text Mining

In the link https://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis the softwares.

Table 6: Free Software Text analysis-Text Mining

System	HTTP	Base	Access date	Version
Natural Language - Google	https://cloud.google.com/natural-language/	Rest API	12/Jun/2020	update 2018
OpenNLP - Apache	https://opennlp.apache.org/	?	12/Jun/2020	update 2015
GATE - General Architeture for Text Engineering	https://tinyurl.com/yaj3j3ph	Java	12/Jun/2020	update
KH Coder	https://tinyurl.com/yb3mr9j5	R, Mysql, Perl	12/Jun/2020	update 2016
Mahout	http://mahout.apache.org/	Scala	12/Jun/2020	?

B Data Set

B.1 Datasets for NER in English

We can access several data sets for the English language through the link below: https://github.com/juand-r/entity-recognition-datasets

B.2 Datasets for NER in Portuguese

We can access several data sets for the Portuguese language through the link below:

B.2.1 HAREM

https://www.linguateca.pt/aval_conjunta/HAREM/harem_ing.html

B.2.2 SIGARRA

https://tinyurl.com/ybwmyrdm

B.2.3 Comparative

From the article by Glauber (2019), we identified that the extraction for evaluation of the Portuguese language was 25 sentences and CETENFolha corpus, film reviews Adoro Cinema and Europarl corpus.

Each system was evaluated by a different team, as shown in the figure 5.

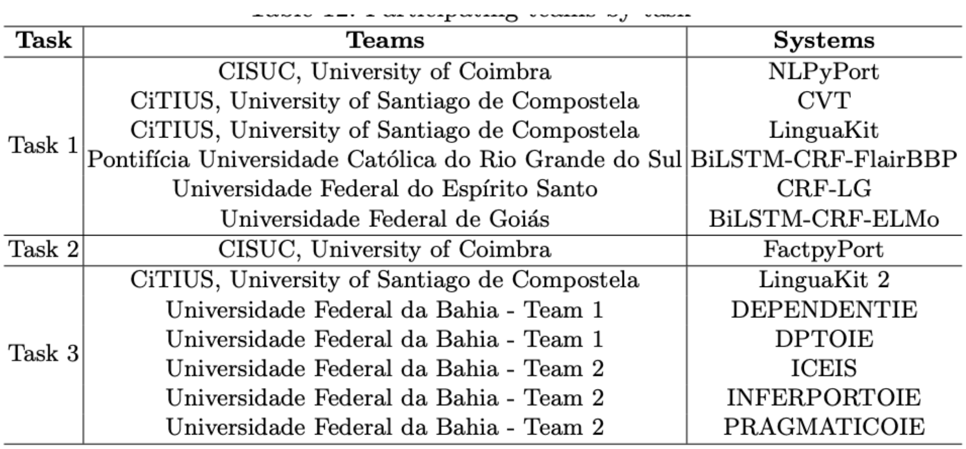


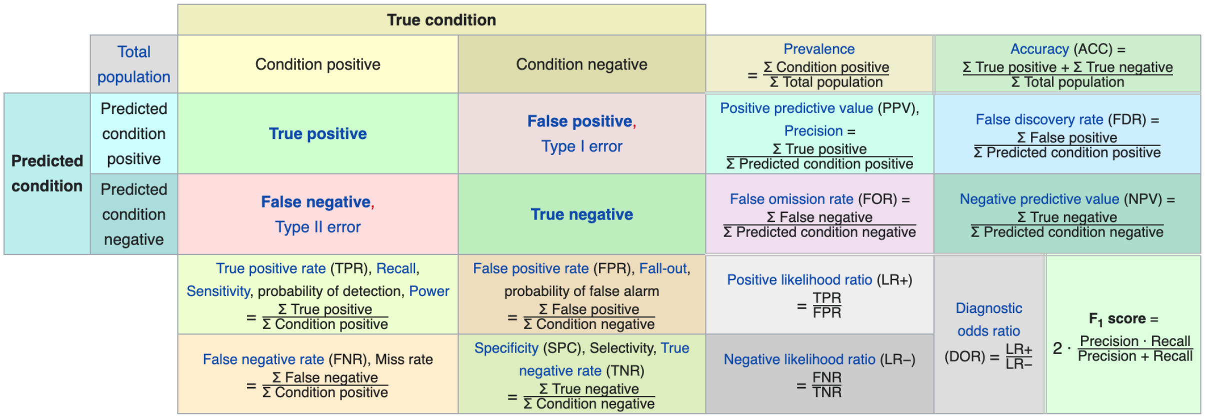
Table 12: Participating teams by task

Task	Teams	Systems
Task 1	CISUC, University of Coimbra	NLPyPort
	CiTIUS, University of Santiago de Compostela	CVT
	CiTIUS, University of Santiago de Compostela	LinguaKit
	Pontifícia Universidade Católica do Rio Grande do Sul	BiLSTM-CRF-FlairBBP
	Universidade Federal do Espírito Santo	CRF-LG
	Universidade Federal de Goiás	BiLSTM-CRF-ELMo
Task 2	CISUC, University of Coimbra	FactpyPort
Task 3	CiTIUS, University of Santiago de Compostela	LinguaKit 2
	Universidade Federal da Bahia - Team 1	DEPENDENTIE
	Universidade Federal da Bahia - Team 1	DPTOIE
	Universidade Federal da Bahia - Team 2	ICEIS
	Universidade Federal da Bahia - Team 2	INFERPORTOIE
	Universidade Federal da Bahia - Team 2	PRAGMATICOIE

Figure 5: Team and Systems from (Glauber, 2019)

C Evaluation Propose

Below in the figure 6 the main ways of calculating the indexes of evaluation metrics:



		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	True positive	False positive, Type I error	Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	False negative, Type II error	True negative	False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		

Figure 6: Index (wik, 2020)

To study the evaluation of methods, we can use some metrics such as:

- F1 ou F_Score

  The F score is often used in the field of information retrieval to measure **search performance, document classification and query classification**. The F score is also used in machine learning. However, these F measures do not take the real negatives into account and that measures such as the correlation coefficient of Matthews, Informedness or Cohen's kappa may be preferable to assess the performance of a binary classifier. The F score has been widely used in the literature on natural language processing (Derczynski, 2016), as an assessment of named entity recognition and word segmentation.

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

Figure 7: F1 (wik, 2020)

- PRECISION / RECALL

  "Accuracy and recall are suitable for assessing problems where the goal is to find a set of items from a larger set of items. In NLP, this may correspond to finding certain linguistic phenomena in a corpus"(Derczynski, 2016).

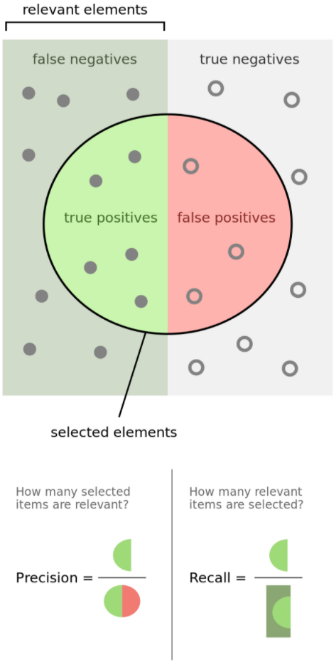


Figure 8: Precision / Recall (wik, 2020)

"*Precision* represents the proportion of items - in this case, entities - that the system returns with the correct precision. Rewards a careful selection and punishes overzealous systems that return many results: to obtain high accuracy, everything that may not be correct must be discarded. False positives - spurious entities - decrease accuracy."

$$P = \frac{|true\,positives|}{|true\,positives| + |false\,positives|}$$

Figure 9: Precision (wik, 2020)

Recall indicates how many of all items that should be found have been found. This metric rewards coverage: to get a high recall, it’s best to include entities that you’re not sure about. False negatives - lost entities - lead to low recall.”(Derczynski, 2016)

$$R = \frac{|true\,positives|}{|true\,positives| + |false\,negatives|}$$

Figure 10: Recall (wik, 2020)

- EXTRA TRAINING

  Extra data to train the model following some characteristic of the environment to be studied gives us a new basis for comparison with the original scope.

- ENVIROMENT / REAL WORLD

  An environment that brings the System a challenge for its evaluation, the concept of this idea is to have a distinct environment or environments little explored by the academy.

- TEST PERFORMANCE

  We can create metrics based on the volume of extraction, number of return times in 1 second.

- LANGUAGE

  Which and how many languages without adaptation the system can be used.

References

Linguakit full, Jun 2020. URL https://linguakit.com/en/full-analysis.

F1 score, Jun 2020. URL https://en.wikipedia.org/wiki/F1_score.

L. Derczynski. Complementarity, f-score, and NLP evaluation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 261–266, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA). URL https://www.aclweb.org/anthology/L16-1040.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

M. Gardner, J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N. F. Liu, M. Peters, M. Schmitz, and L. S. Zettlemoyer. Allennlp: A deep semantic natural language processing platform. 2017.

R. Glauber. Iberlef 2019 portuguese named entity recognition and relation extraction tasks. 2019.

C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014. URL http://www.aclweb.org/anthology/P/P14/P14-5010.

L. B. Soares, N. FitzGerald, J. Ling, and T. Kwiatkowski. Matching the blanks: Distributional similarity for relation learning. *arXiv preprint arXiv:1906.03158*, 2019.