



Automated Event Extraction Model for Linked Portuguese Documents

By Kashyap Raiyani
University of Evora



Outline



- Motivation / Importance
- Related Background – Agatha
- System Overview
- SRL Parser for Portuguese Language
- SVO Extraction
- Knowledge Base Ontology
- Manual Annotation
- Evaluation Results
- Future Scope



Motivation



- High Information Present
- Video and Image
- Audio and Voice
- Biometric
- Multilingual text
- Classification and Semantic Segmentation - cross-referencing
- Intelligent system that resorts to open sources of information for surveillance and crime control.

Agatha



Intigration & Front-end



Voice



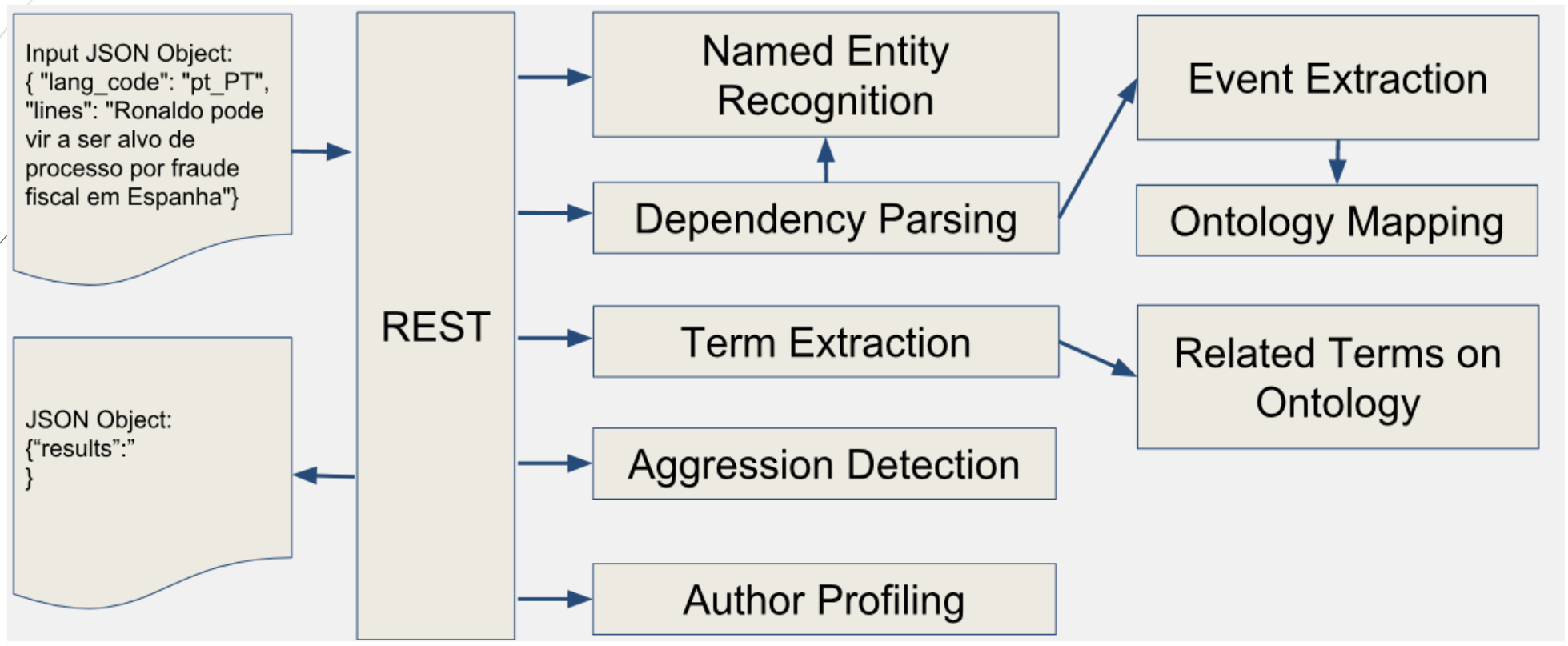
Video



Text

<http://www.agatha-osi.com/>

System Overview





SRL Parser



- Freeling[1] as the baseline tool.
- Supports SRL tagging over five languages namely, Catalan, Croatian, English, German and Spanish.
- Universal Dependencies (UD) Portuguese treebank dataset[2].
- Universal POS tags -> EAGLES tagset.
- 14 categories with 580 tags.

1. Xavier Carreras, Isaac Chao, Lluís Padró, and Muntsa Padro. Freeling: An open-source suite of language analyzers. Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04), 01 2004.

2. Alexandre Rademaker, Fabricio Chalub, Livy Real, Cláudia Freitas, Eckhard Bick, and Valeria de Paiva. Universal dependencies for portuguese. In Proceedings of the Fourth International Conference on Dependency Linguistics (Depling), pages 197–206, Pisa, Italy, September 2017



SRL Parser - 14 Categories

NOUN

VERB

PROPN

ADJ

DET

AUX

ADP

NUM

PUNCY

CCONJ

SCONJ

INTJ

ADV

SVO Extraction

Algorithm 1 SVO Extraction Algorithm

```
1: procedure SVO(sentence)
2:   SVO  $\leftarrow$  []
3:   while Predicates  $\neq$  NULL do ▷ For-each predicate in sentence
4:     Event  $\leftarrow$  Predicate
5:     while Arguments  $\neq$  NULL do ▷ For-each arguments associated with predicate
6:       switch Arguments.role() do
7:         case A0
8:           Actor  $\leftarrow$  Arguments
9:         case A1
10:          Object  $\leftarrow$  Arguments
11:        case AM-LOC
12:          Location  $\leftarrow$  Arguments
13:        case AM-TMP
14:          Time  $\leftarrow$  Arguments
15:      end while
16:      SVO.append(Event, Actor, Object, Location, Time)
17:    end while
18:  return SVO ▷ role() returns SRL tagging
19: end procedure
```



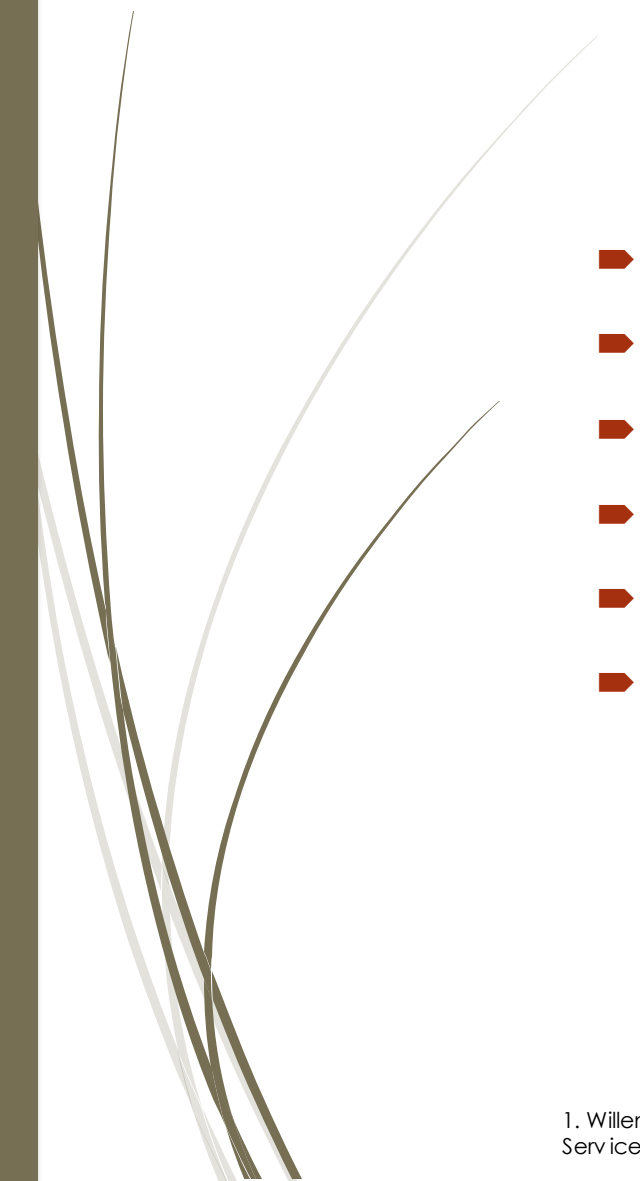
Knowledge Base Ontology

- Source - legal police documents.
- Case 1 has 100 documents and case 2 has 100 documents.
- Relation among all the 200 documents.
- GraphDB[1].
- Protégé[2].

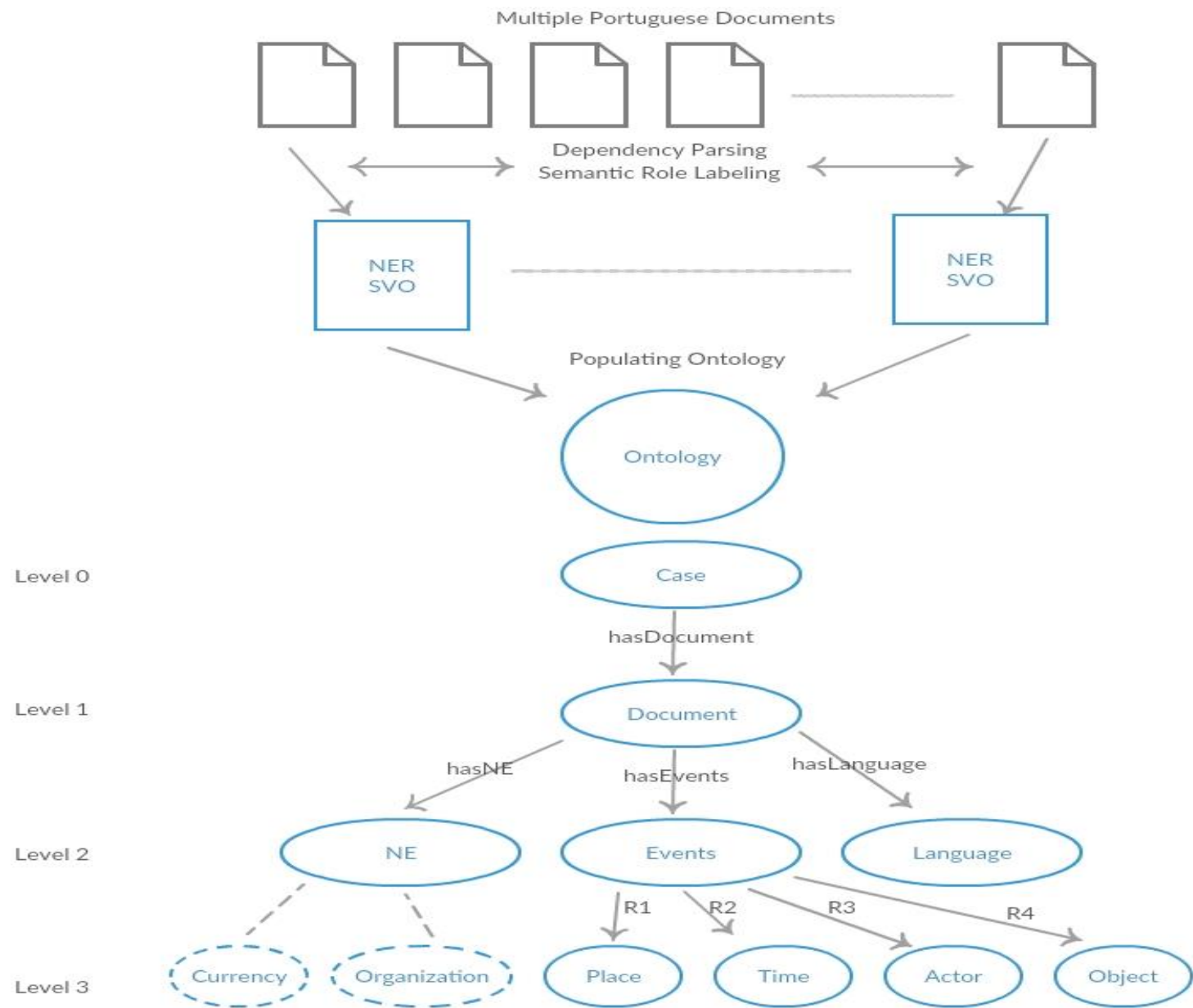
1. <https://protege.stanford.edu/>
2. <http://graphdb.ontotext.com/>



Knowledge Base Ontology Cont..

- 
- Simple Event Model – ontology[1].
 - Case - CaseID
 - Document - DocumentID foreach Case
 - Language - Language foreach Document
 - NER - NER foreach Document
 - Event - Events foreach Document

Ontology





Manual Annotation



- The manual annotation is being coordinated by Prof. Ana Leal, from the Portuguese Department of the University of Macau.
- Agreed to cooperate with us in this process (together with a group of 3 Master's students).
- Google Form -> Response



Evaluation & Results



Event Tagging	Number
System	6178
Manual	6471
Difference	296
Min. Missing	1
Max. Missing	39
Avg. Missing	5.75
Accuracy	95.00%

It is composed of 51 documents with 1221 sentences and 48914 words



Future Scope

- The Training, Development, Evaluation, and System Extracted Events datasets are accessible for further research and comparison purpose.
- System lacks discovering time categories associated with events.

Acknowledgements

- We would like to thank COMPETE 2020, PORTUGAL 2020 Programs, the European Union, and ALENTEJO 2020 for supporting this research as part of Agatha Project SI \& IDT number 18022 (Intelligent analysis system of open of sources information for surveillance/crime control).



UNIVERSIDADE
DE ÉVORA





Thank you for your kind attention
Any Questions?