M2 Software Project

Online App for Knowledge Substantiation

Anna MOSOLOVA
Elisa LUBRINI

UNIVERSITÉ DE LORRAINE
IDMC Institut des sciences du Digital Management & Cognition

December 10, 2021

Contents

1. Overview
2. Knowledge Base Module
3. Ensemble
4. Next Steps

Table of Contents

1 Overview

2 Knowledge Base Module

3 Ensemble

4 Next Steps

Overview

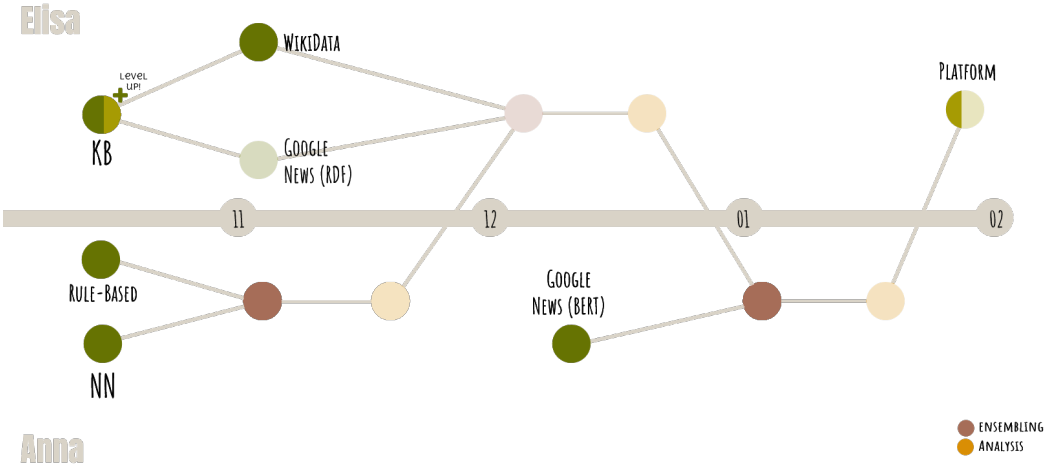

Figure: Current Timeline

Table of Contents

1 Overview

2 Knowledge Base Module

3 Ensemble

4 Next Steps

Knowledge Base

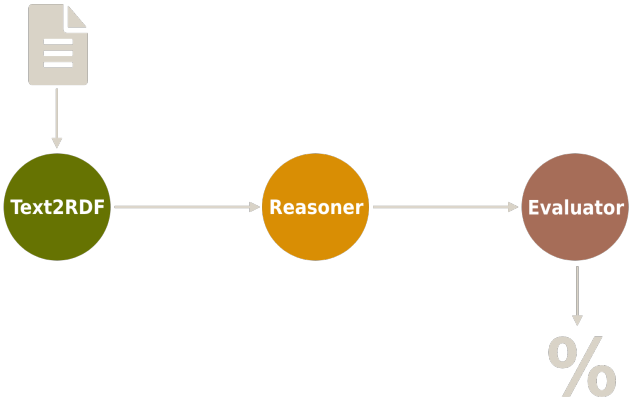

Figure: KB Pipeline

Knowledge Base

Text2RDF Possibilities

end-to-end tools **vs.**

- × FRED → not accessible
- × OpenIE → very limited
- × Protegé *Addons + API* → ?

text to triples...

Nocino is a cat → (Nocino; is; cat)

deepex → SotA, Sep 2021

...and triples to RDF

NER + string matching

Knowledge Base

Text2RDF

Previously:

- × FRED → missing resource
- × Basic Custom-made Script → underperforming
- ✓ OpenIE[1] [1] → currently working, very basic

Currently:

Dual system:

- ✓ Neural Approaches → implemented
  +
- … Ontology population via NER and string matching → implementation in progress

[1] OpenIE python wrapper [2]: https://github.com/philipperemy/Stanford-OpenIE-Python

Knowledge Base

Reasoner

- *Output:* number of inconsistent classes
- Complexity: SPARQL is PSPACE-complete

→ import .ttl entities via Virtuoso[2]
→ use only classes without entities

[2]high-performance object-relational SQL database [3]

Knowledge Base

Evaluator

- *Output:*
  - % of inconsistent classes/entities over total
  - name of inconsistent classes
  - [other info for algo transparency]

Table of Contents

1 Overview

2 Knowledge Base Module

3 Ensemble

4 Next Steps

Ensemble architecture

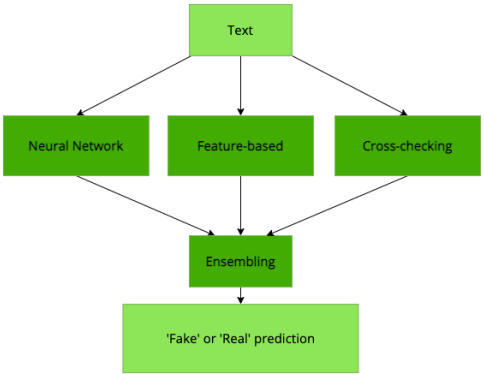

Figure: Ensemble architecture

Feature-based Module

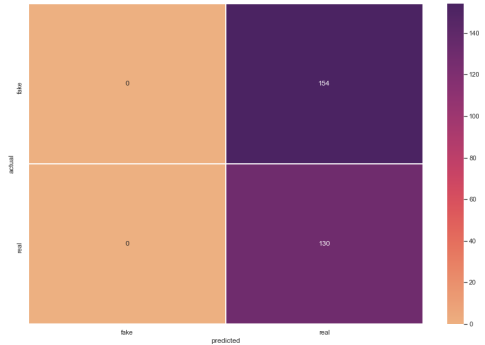

Figure: Confusion matrix for Feature-based Module

Cross-checking Module

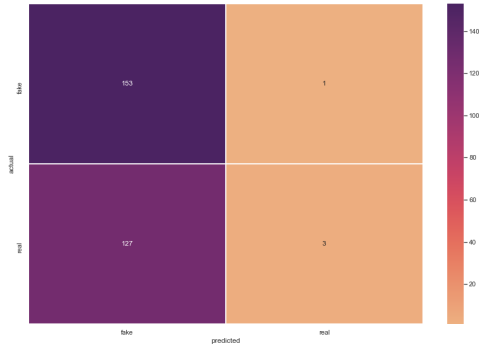

Figure: Confusion matrix for Cross-checking Module

Neural Network Module

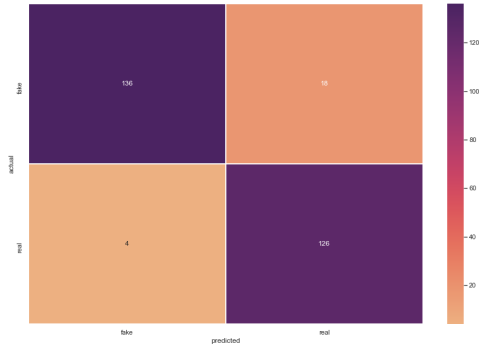

Figure: Confusion matrix for Neural Network Module

Ensembling

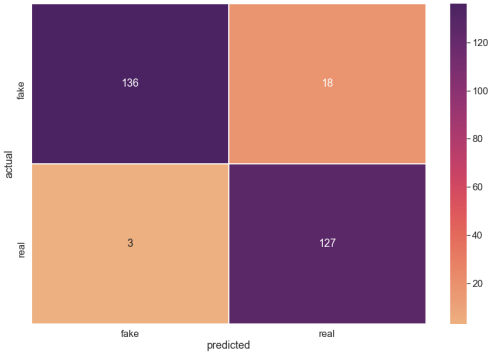

Figure: Confusion matrix for Ensembling

Table of Contents

1 Overview

2 Knowledge Base Module

3 Ensemble

4 Next Steps

Next steps

Fine-tune the **thresholds** for features and cross-checking

Try the algorithm on a new **dataset**

Implement other **ensemble** methods

Finish the implementation of the **ontology population** script

Load **entities from .ttl** to a graph

Connect **back-end to front-end**

Bibliography I

[1] Lei Cui, Furu Wei, and Ming Zhou. “Neural Open Information Extraction”. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne, Australia: Association for Computational Linguistics, July 2018, pp. 407–413. DOI: 10.18653/v1/P18-2065. URL: https://aclanthology.org/P18-2065.

[2] Philippe Remy. *Python wrapper for Stanford OpenIE*. https://github.com/philipperemy/Stanford-OpenIE-Python. 2020.

[3] OpenLink Software. *Virtuoso SPARQL Query Editor*. 2009. URL: http://dbpedia.org/sparql (visited on 03/24/2014).