M2 Software Project

Online App for Knowledge Substantiation

Anna MOSOLOVA
Elisa LUBRINI

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

November 4, 2021

Contents

Table of Contents

Overview

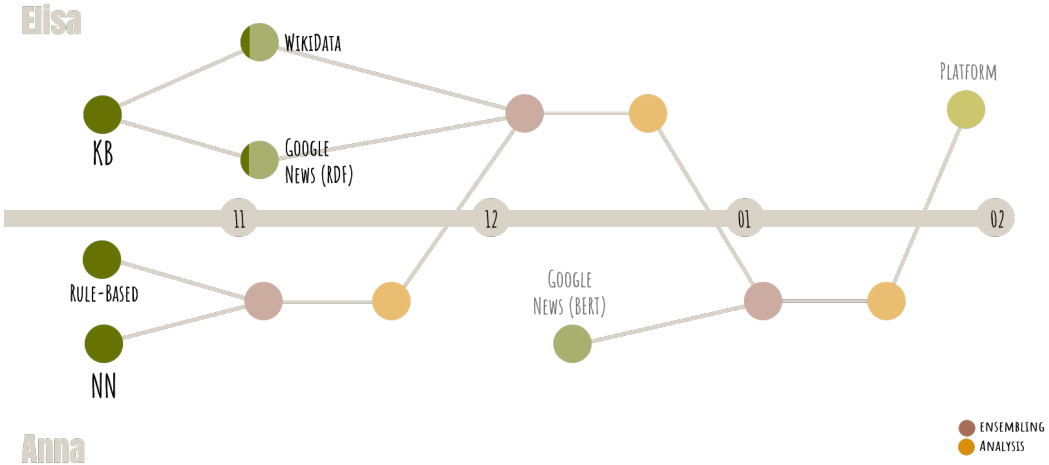


Figure: Expected Timeline

Overview

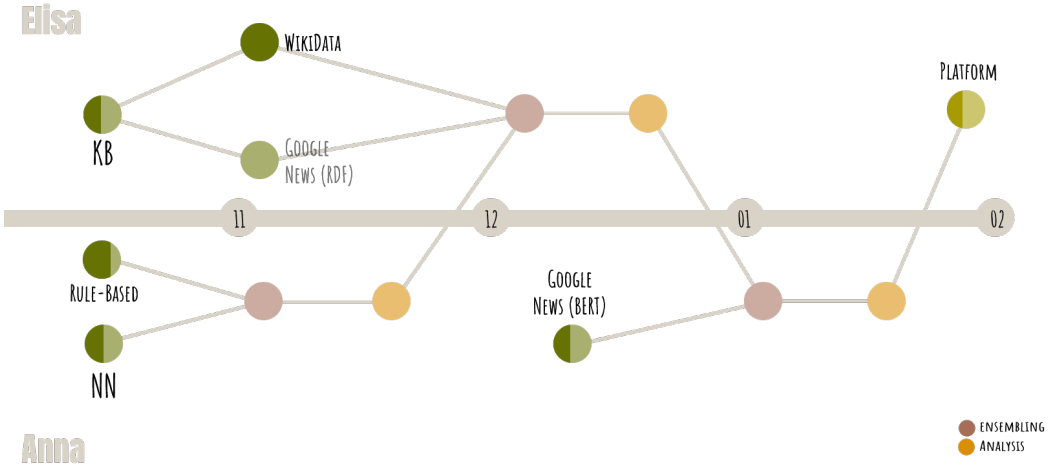
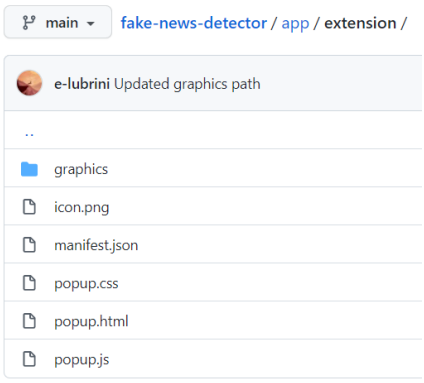


Figure: Current Timeline

Table of Contents

GUI Structure

Current



Planned

```
/extension
        /run.py
        /config.py
        /app
            /__init__.py
            /views.py
            /models.py
            /static/
                /main.css
                /main.js
                /graphics/
            /templates/
                /base.html
                /base.html
        /requirements.txt
        /env/
```

Table of Contents

Knowledge Base

- ✓ Retrieved and loaded DBpedia OWL ontology[1] (without instances)
- ✓ Retrieved DBpedia TTL entities[2]
- ✓ Implemented OWL reasoner for consistency checking
- × Text2RDF (key needed for Fred tool)
- × Loading subsection of entities

File	Size
OWL Ontology	3M
TTL Entities	5720M

Table: Size of knowledge base files

[1] http://dief.tools.dbpedia.org/server/ontology/dbpedia.owl

[2] https://databus.dbpedia.org/dbpedia/mappings/instance-types/2021.09.01/instance-types_lang=en_transitive.ttl.bz2

Table of Contents

Rule-based Approach

Fake News Features

- ✓ spelling mistakes
- ✓ lexical repetitions
- ✓ plural forms
- ✓ loaded language
- ✓ punctuation errors
- ✓ excessive usage of exclamation marks and imperative mood
- × misrepresentation of facts
- × links to an unreliable source
- × rhetorical questions
- × contrast of *we* VS *they* stances
- × references to past events

Table of Contents

Neural Network Module

Model	F1-weighted
GRU	60%
CNN	60%
GRU with BERT embeddings	53%
BERT	60%
roBERTa	61%
Fake news detector[2]	53% | 65% | 64% | 62%

Table: Observed results when implementing different architectures

Table of Contents

Next steps

Find **alternative Text2RDF** tool or **wait for key**
meanwhile use the FRED GUI [1] [3] instead of the API

Find URI of the elements in an article

Load only **relevant elements** from reference entities
mentioned elements + connected elements down to **depth** n according to optimisation

Try and implement **more neural networks** architectures in order to improve the model's f1-score

Implement the **remaining features** for the feature-based approach

Implement a module for news extraction using **Google News** python library

[3] http://wit.istc.cnr.it/stlab-tools/fred/demo

Bibliography I

[1] Francesco Draicchio et al. “FRED: From natural language text to RDF and OWL in one click”. In: vol. 7955. May 2013, pp. 263–267. ISBN: 978-3-642-38708-1. DOI: 10.1007/978-3-642-41242-4__36.

[2] Xiangyang Li et al. “Exploring text-transformers in aaai 2021 shared task: Covid-19 fake news detection in english”. In: *arXiv preprint arXiv:2101.02359* (2021).