M2 Software Project

Online App for Knowledge Substantiation

Anna MOSOLOVA
Elisa LUBRINI

OAKS

UNIVERSITÉ DE LORRAINE

IDMC Institut des sciences du Digital Management & Cognition

October 12, 2021

Contents

1. Product

2. System Modules

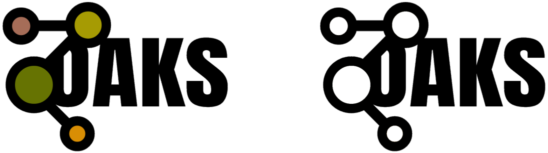
3. Dataset

4. Next Steps

Table of Contents

1. Product
2. System Modules
3. Dataset
4. Next Steps

Product
Brand



The oak is a symbol of **strength**, **morale**, **resistance** and **knowledge**.

Our Objective

Creating a **reliable** app that double-checks the **veridicity** of claims proposed by news pieces.

Product
Development

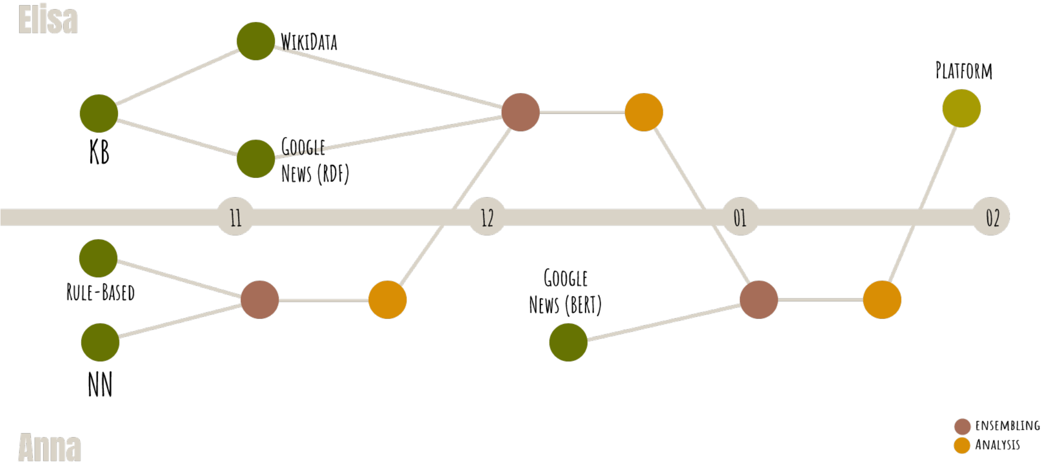


Figure: Timeline

Table of Contents

1. Product
2. System Modules
3. Dataset
4. Next Steps

Knowledge Base

Module Structure

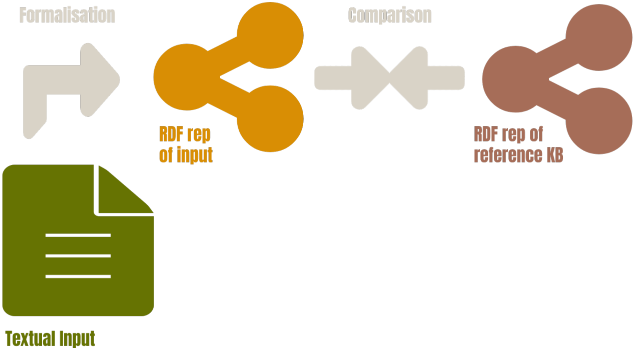


Figure: The module takes text as input, converts it to an RDF graph, and compares it to an RDF representation of a reference Knowledge Base.

Knowledge Base

Text to RDF

Input: *colourless ideas sleep furiously*

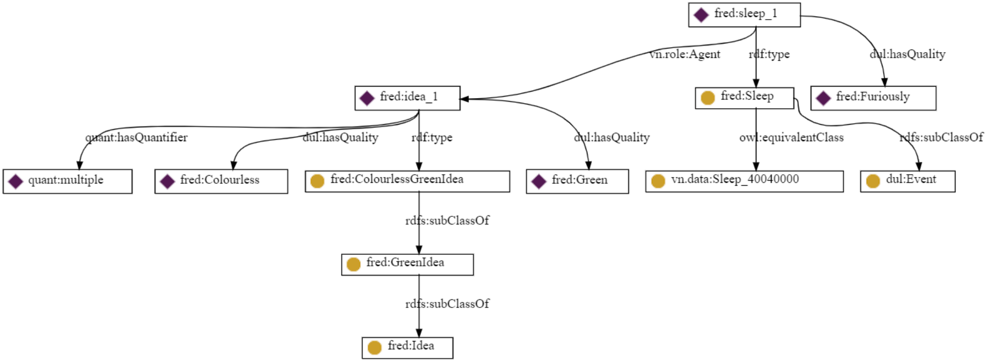


Figure: Example of text to RDF conversion using FRED.[3]

Knowledge Base

RDF Comparison [4]

- Performing **entity comparison**
- SPARQL **similarity** queries
  - Query Reverse Engineering [1]
  - Least Common Subsumers in Description Logic ontologies [2]
- SPARQL **difference** queries (coNP-complete)

Rule-based Approach

Fake News Features

- ✓ spelling mistakes
- ✓ lexical repetitions
- ✓ plural forms
- ✓ loaded language
- × misrepresentation of facts
- × links to an unreliable source
- × rhetorical questions
- × excessive usage of exclamation marks and imperative mood
- × punctuation errors
- × contrast of the "we" and "they" concepts
- × throwbacks

Table of Contents

1. Product
2. System Modules
3. Dataset
4. Next Steps

Dataset

LIAR-PLUS dataset

- **Training set** 10,269
- **Validation set** 1,284
- **Testing set** 1,283
- **Main topic** politics (economy, healthcare, taxes, federal budget, education, jobs, state budget, candidates biography, elections, and immigration)

- **Average sentence length** 20
- Similar vocabulary for all classes
- **Classes**
  - half-true: 2114
  - false: 1995
  - mostly-true: 1962
  - true: 1676
  - barely-true: 1654
  - pants-fire: 839

Table of Contents

1. Product
2. System Modules
3. Dataset
4. Next Steps

Coming Up...

For next month...

- Implement **text-to-RDF** conversion
- Implement **NN module** method
- Research **RDF coparisons** algorithms
- Start **ensembling**

Bibliography I

[1] Marcelo Arenas, Gonzalo I. Diaz, and Egor V. Kostylev. “Reverse Engineering SPARQL Queries”. In: *Proceedings of the 25th International Conference on World Wide Web*. WWW ’16. Montréal, Québec, Canada: International World Wide Web Conferences Steering Committee, 2016, pp. 239–249. ISBN: 9781450341431. DOI: 10.1145/2872427.2882989. URL: https://doi.org/10.1145/2872427.2882989.

[2] Franz Baader, Baris Sertkaya, and Anni-Yasmin Turhan. “Computing the least common subsumer w.r.t. a background terminology”. In: *Journal of Applied Logic* 5.3 (2007). Selected papers from the 9th European Conference on Logics in Artificial Intelligence (JELIA ’04), pp. 392–420. ISSN: 1570-8683. DOI: https://doi.org/10.1016/j.jal.2006.03.002. URL: https://www.sciencedirect.com/science/article/pii/S1570868306000036.

[3] 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.

Bibliography II

[4] Alina Petrova et al. “Entity Comparison in RDF Graphs”. In: *The Semantic Web – ISWC 2017*. Ed. by Claudia d’Amato et al. Cham: Springer International Publishing, 2017, pp. 526–541. ISBN: 978-3-319-68288-4.