

Fake News Detection

Clément Loeuillet, Samson Dubuy, Cheng Yang, Yi Yang, TingJun Yuan

February 10, 2025

1 Introduction

The application that this project creates is the Fake News Detection System which specializes in sports news. The entire process involves analyzing the information of articles given by the user, author, and source to ensure the credibility of particular news in this system. The application is powered by natural language processing (NLP) techniques to identify entities within the text, as well as a variety of data sources to compute a credibility score and combine the ontology definitions to present a visual representation of the relationships within the ontology. The score also assists readers in evaluating the legitimacy of sports news.

Disinformation can be contagious in social media and other online platforms, creating a pervasive sense of confusion and distrust. To tackle this issue, the current application enables users to assess the credibility of sports news, which can assist them in making decisions and enhancing media literacy.

2 Application Description

2.1 Application System

The fundamental structure of the application system consists of a front-end and a back-end. The front-end is built using React, allowing users to input text, author, and source information. The back-end is developed using Flask, which processes the input data, performs entity extraction, and calculates the credibility score. The front-end and back-end communicate via HTTP requests. The front-end sends user input to the back-end, which processes the data and returns the results.

2.2 Technical Details

The application utilizes several key technologies and libraries:

- **Ontologies:** The application models relationships between concepts like sources, authors, and entities. The ontology helps in structuring the extracted information and visualizing the relationships. This allows the system to better understand the context of the information and provide more accurate credibility assessments by defining these relationships.
- **Methods/Techniques:** The system employs a specialized sports news ontology to model semantic relationships between key concepts such as news sources, authors, and entities. This ontology provides a structured framework for organizing extracted information while enabling interactive relationship visualization. By explicitly defining hierarchical classes (e.g., *Athlete* as a subclass of *Person*) and property constraints (e.g., *hasAuthor* domain-range restrictions), the ontology enhances contextual understanding and delivers more precise credibility assessments.

2.3 Application Usage

The application provides the following functionalities:

- **Input Text Analysis:** Users can input text about sports news, author name, and source. The system then processes this input to extract relevant entities and information.
- **Ontology Construction with RDFLib:** Defined classes (*SportsArticle*, *Athlete*, *NewsBrand*) and properties (*participatesIn*, *mentionsEntity*) using RDFLib.
- **Entity Validation via SPARQL:** Developed SPARQL queries to verify entities against Wikidata.
- **Visualization:** The system produces a graph visualization to present the interaction between entities and displays these graphics on screen. This representation helps users understand how various elements are connected, making it easier to evaluate the credibility of the news article.

Use Case 1: Checking Credibility of a Sport News

1. User inputs the news article text, author name, and source.
2. User submits the form.

3. The system processes the input, extracts entities, and calculates the credibility score.
4. The system displays the credibility score.

Use Case 2: Visualizing Entity Relationships

1. User inputs the news article text, author name, and source.
2. User submits the form.
3. The system processes the input and extracts entities.
4. The system generates a graph visualization of the entities and their relationships.
5. The user views the graph to understand the connections between different pieces of information.

3 Discussion

Throughout the development of this project, several lessons were learned:

- **Ontology Definition and Modeling:** In this project, we leveraged RDFLib to construct a domain-specific ontology that classifies entities into distinct types (e.g., authors, sources, athletes, teams, and events) and defines semantic relationships (e.g., *hasAuthor*, *participatesIn*). By introducing hierarchical class structures (e.g., *Athlete* as a subclass of *Person*) and linking entities to Wikidata via *owl:sameAs*, we enhanced the system's ability to organize and visualize complex relationships within news articles. This structured approach improved contextual understanding and enabled precise credibility assessments.
- **SPARQL Queries and Entity Validation:** We implemented SPARQL to validate entities against trusted knowledge bases like Wikidata. For example, queries such as:

```
SELECT ?item WHERE { ?item rdfs:label "BBC Sport"@en; wdt:P31/wdt:P279*
wd:Q1173904 }
```

ensured accurate verification of news sources and athletes.

- **Rule-Based Reasoning:** The system employs rule-based reasoning to dynamically calculate credibility scores based on ontology relationships. For example:
 - If a source (e.g., *BBC Sport*) is verified as authoritative (*wd:Q552359*), add +10 points.

- If an athlete (e.g., *Cristiano Ronaldo*) participates in an event (e.g., *Champions League*), add +5 points.
- Use SPARQL to check if `?athlete ex:participatesIn ?event` exists in the RDF graph and trigger scoring rules based on inferred relationships.

This approach replaces static scoring with logic grounded in semantic relationships, improving transparency and accuracy.

Overall, this project provided a comprehensive understanding of the theories and technologies related to natural language processing, knowledge representation, and ontology modeling. It demonstrated the potential of using these techniques to assess the credibility of news articles and highlighted the importance of continuous improvement and adaptation to address emerging challenges.