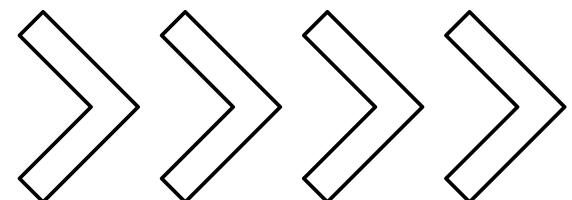


WORDIFY

CHECKPOINT 3

**WEB SEMANTICS AND
LINKED DATA**



Group D:

Fábio Araújo de Sá (up202007658@up.pt)
Marcos Rafael Peixoto Aires (up202006888@up.pt)
Pedro Pereira Ferreira (up202004986@up.pt)



Context

Wordify is a system that disambiguates existing entities, possibly in different languages, using Web Semantics and Linked Data principles, concepts, and technologies.

It will help a computer and humans to perceive which entity is being referred to, if it is ambiguous in a certain context, or to consult false friends between languages.



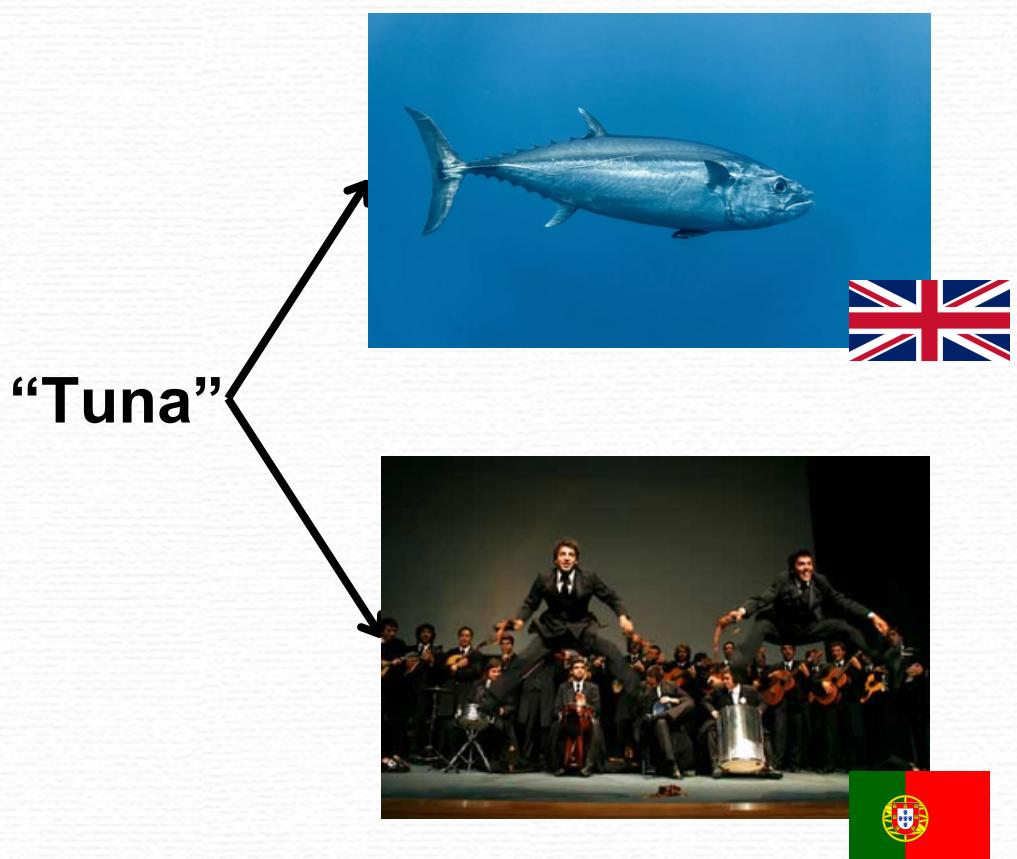
wordify

Motivation

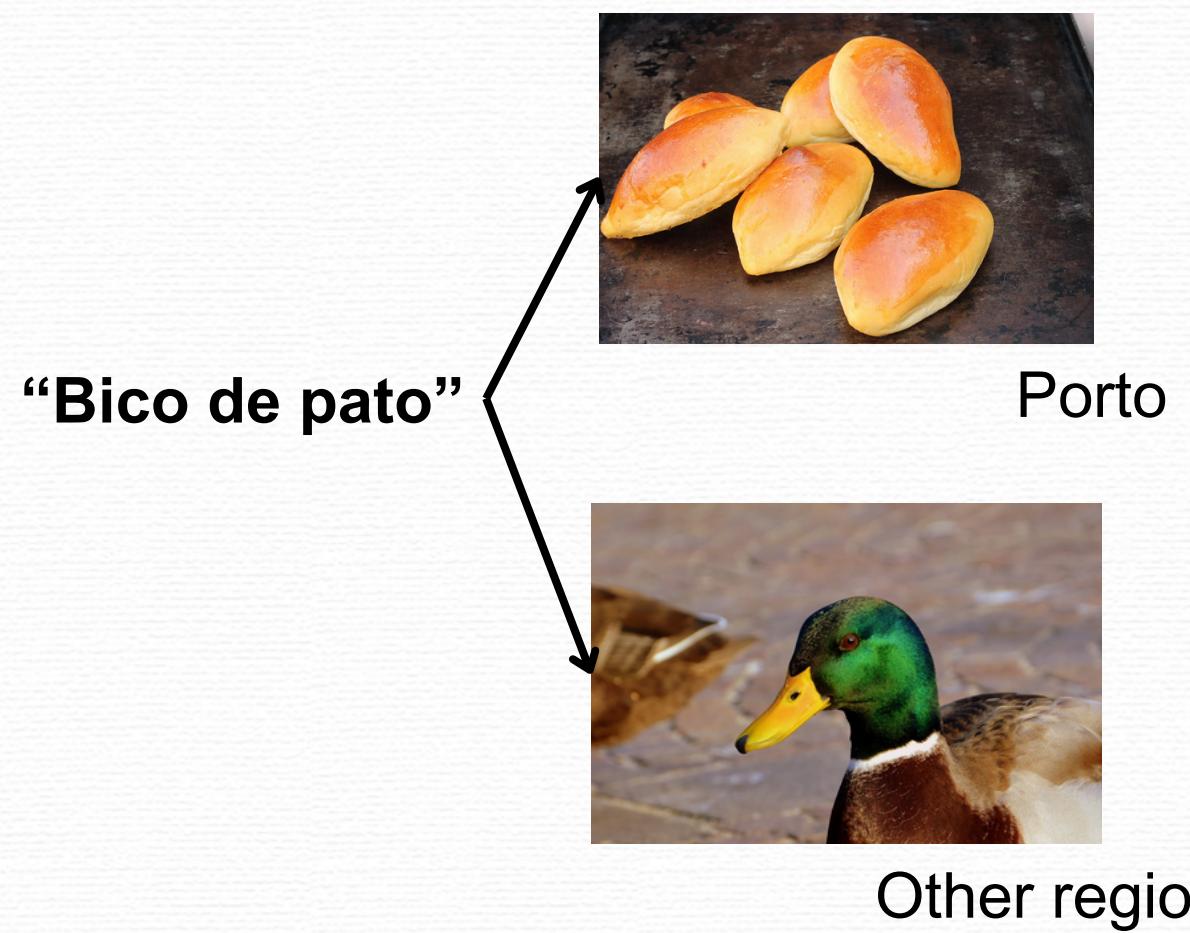
01 Disambiguate different contexts



02 Disambiguate different languages



03 Disambiguate different regionalisms



Requirements

Globally, the system should be capable of:

- Linking various **concepts** to the entities being searched;
- Linking various **definitions** to their corresponding concepts, as well as to the respective languages, if applicable;
- **Disambiguating concepts** within and across languages;
- Consulting **false friends** of the entity, if they exist.

Adhering to the Linked Data Principles / Tim Berners-Lee's 4 Principles:

- Use Uniform Resource Identifiers (URIs) to identify resources;
- Use HTTP URIs so people can look up data;
- Provide useful information in standard formats (RDF);
- Include links to other URIs to connect data

Challenges

- Vast number of potential entities that could be ambiguous, within and across different languages;
- There are instances where a word may exist in one language but cannot be automatically associated with another representation in another entity in a different language. For instance, "saudade" means missing someone in portuguese, but there is no direct english translation;

Existing Approaches

➤ **DBpedia
Spotlight**



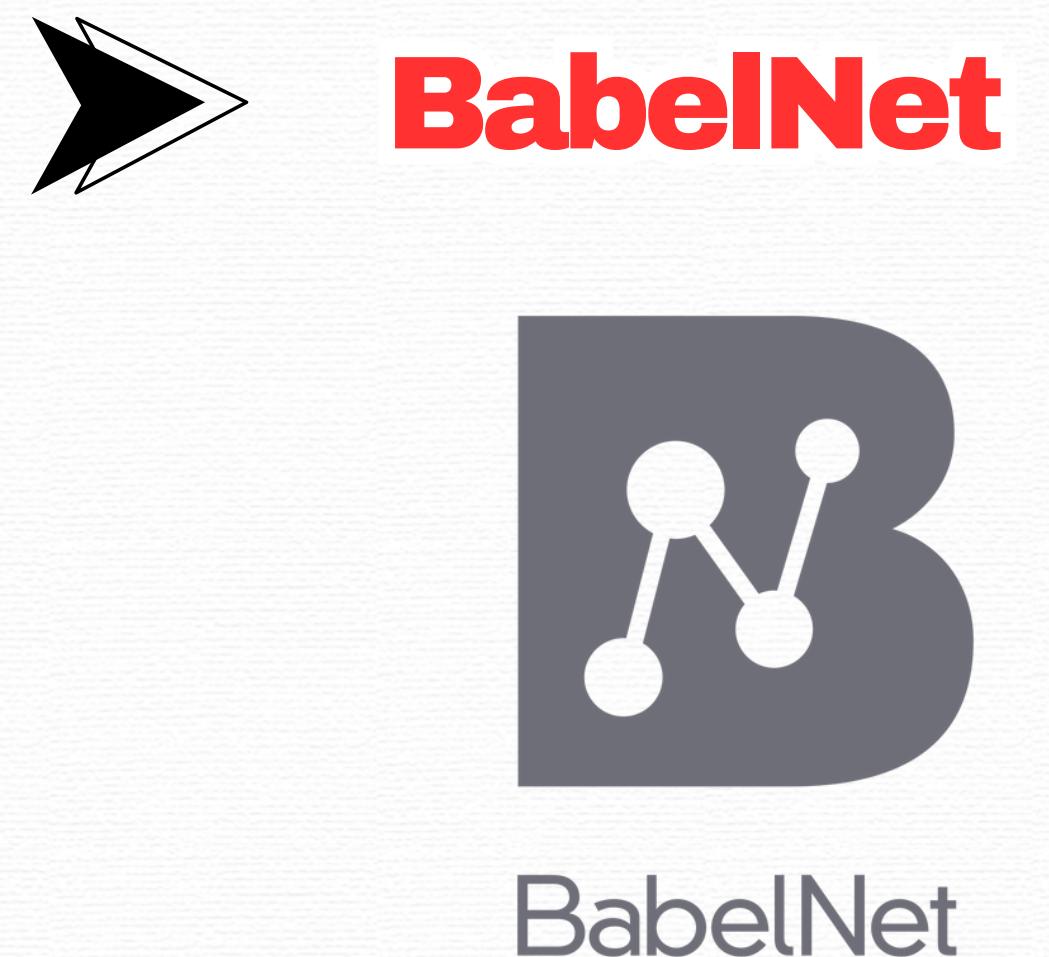
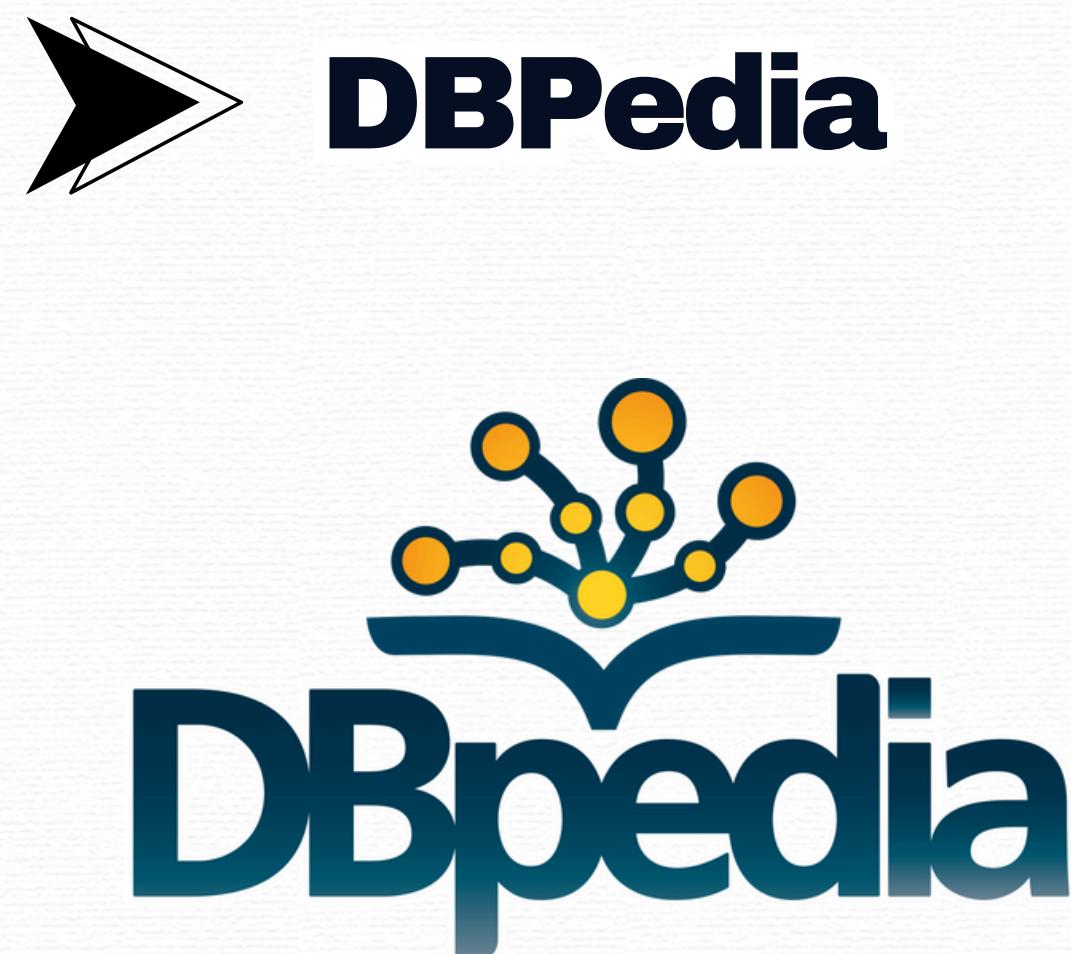
➤ **Wikidata Query
Service**



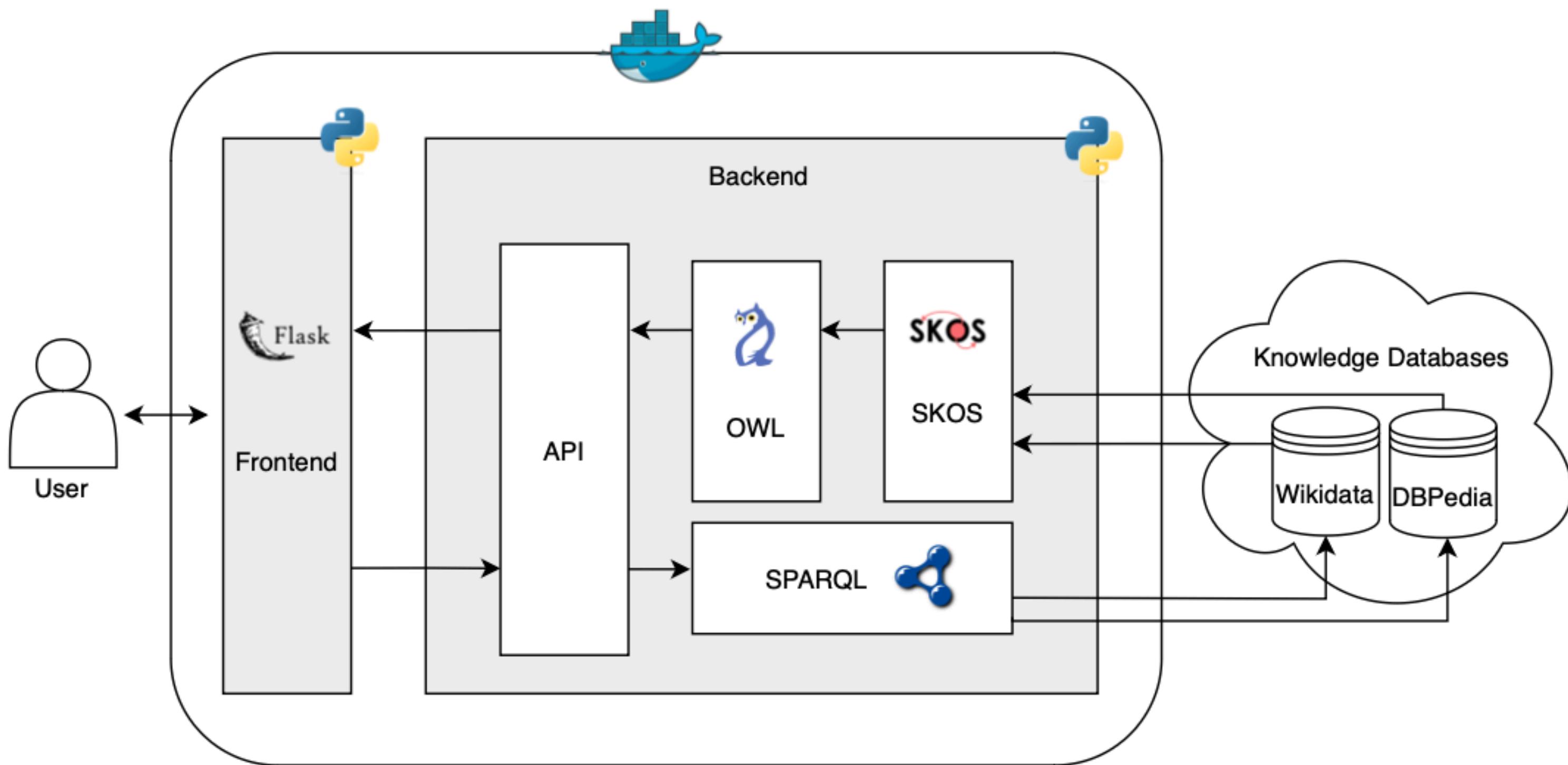
➤ **YAGO**



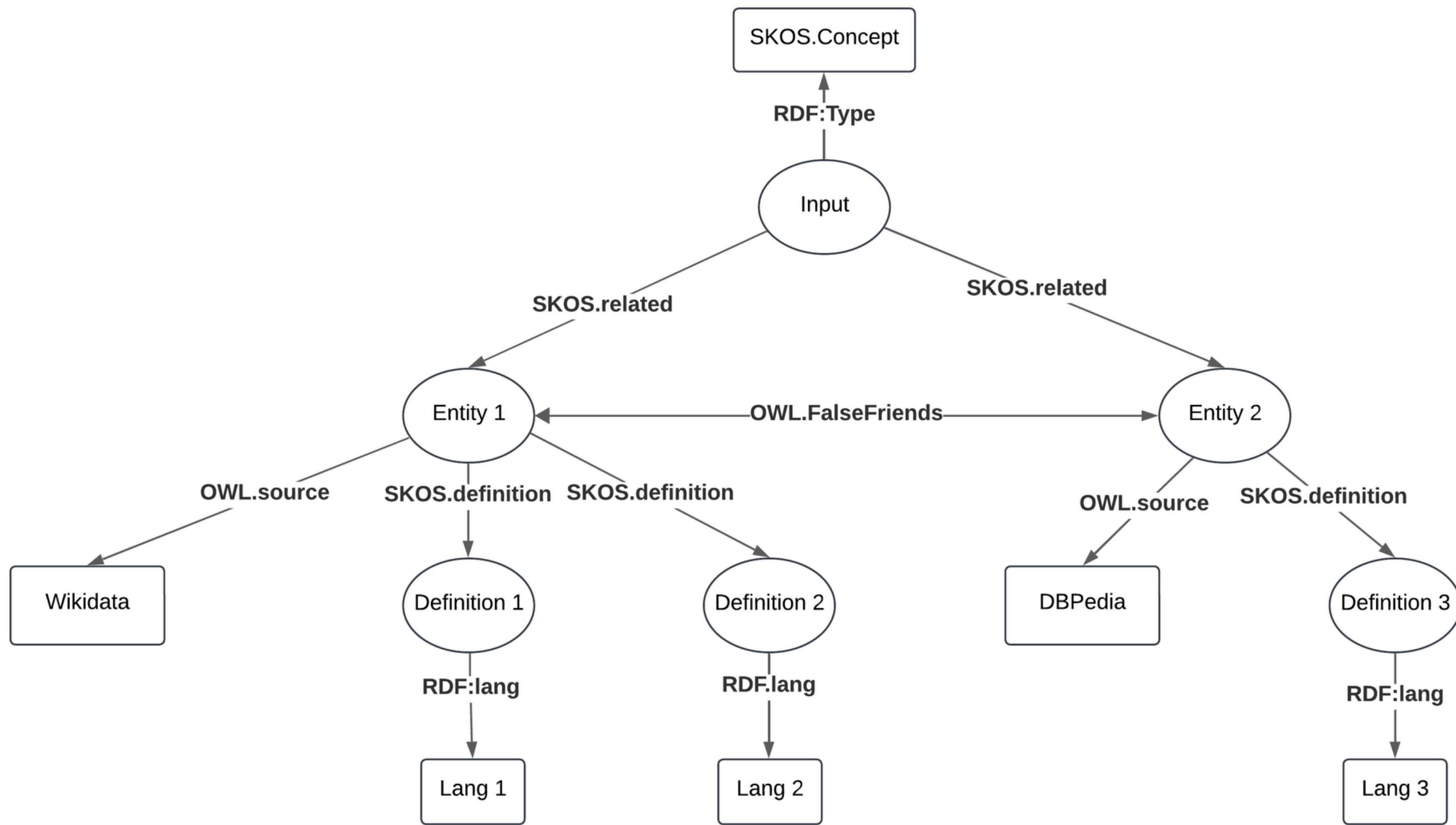
Knowledge base sources



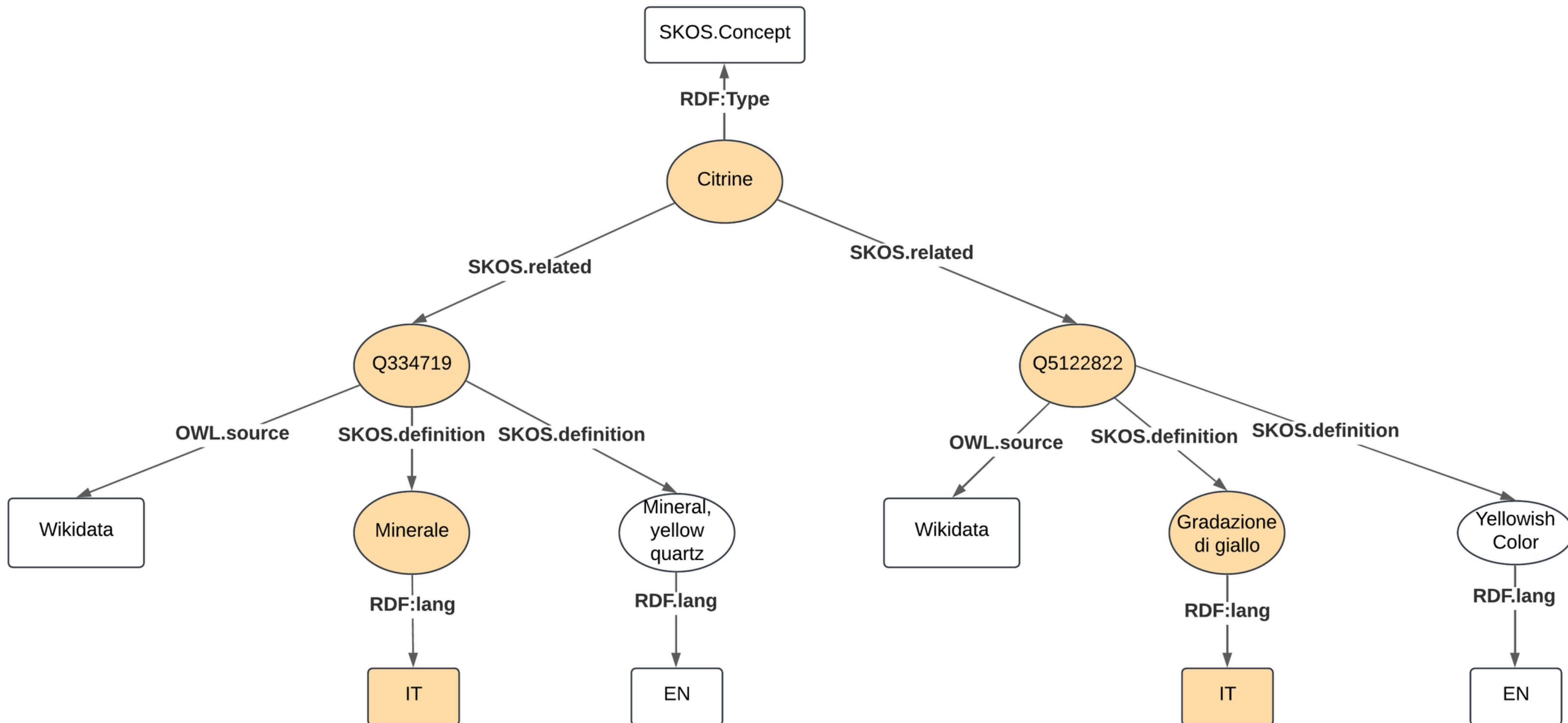
Architecture



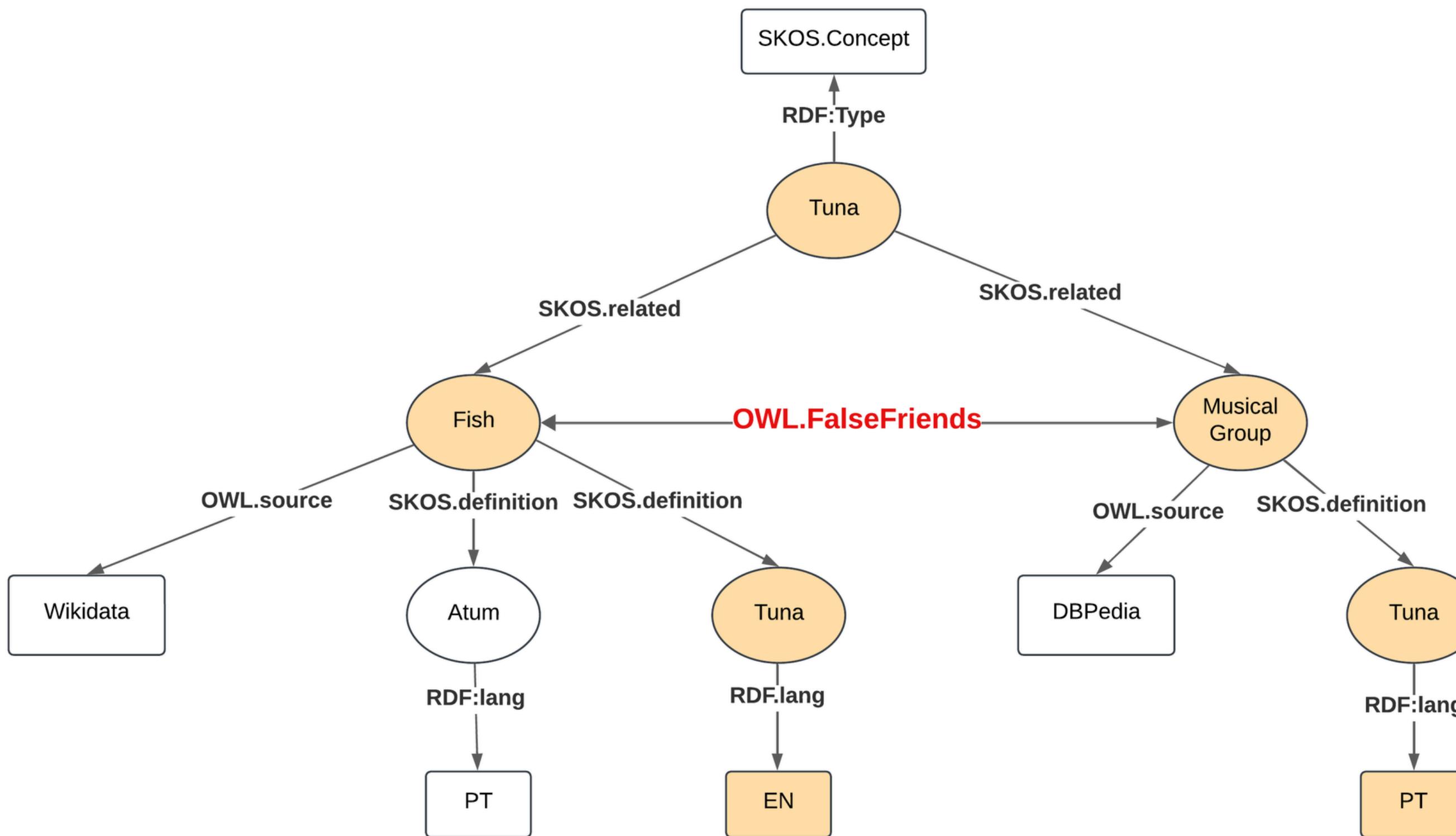
Architecture



Features - Ambiguities

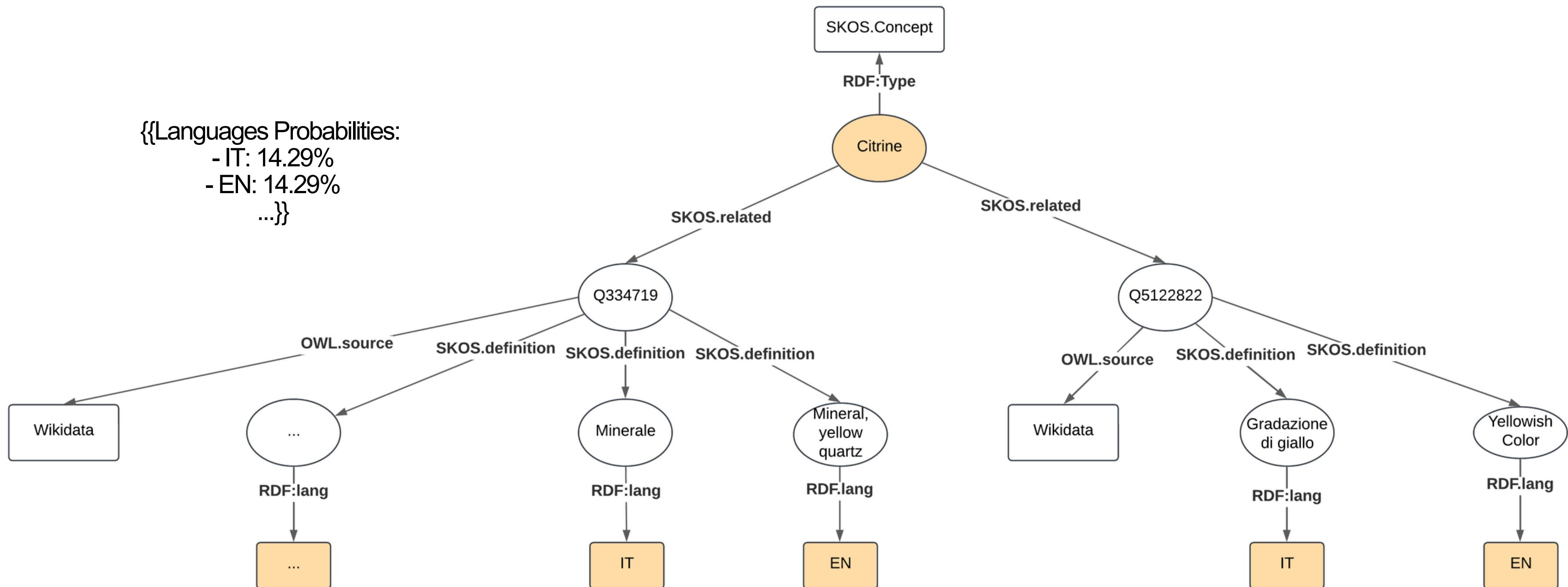


Features - False Friends

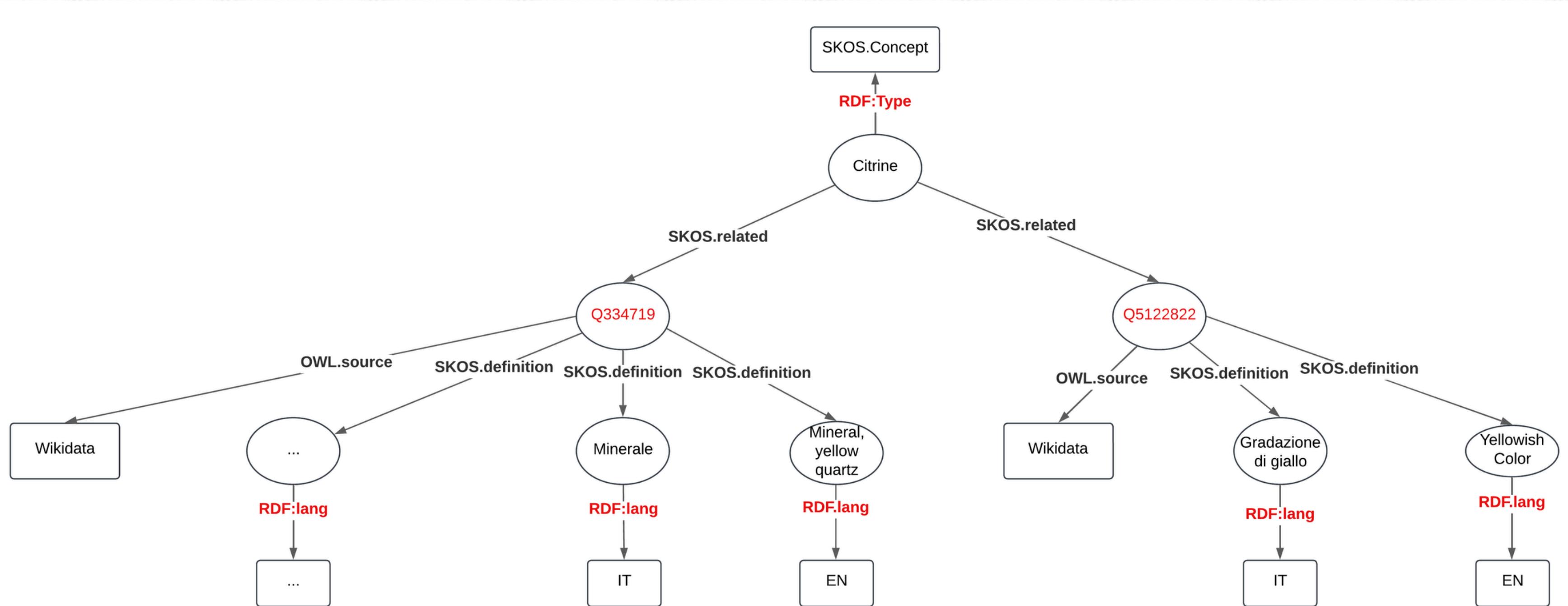


Features - Languages

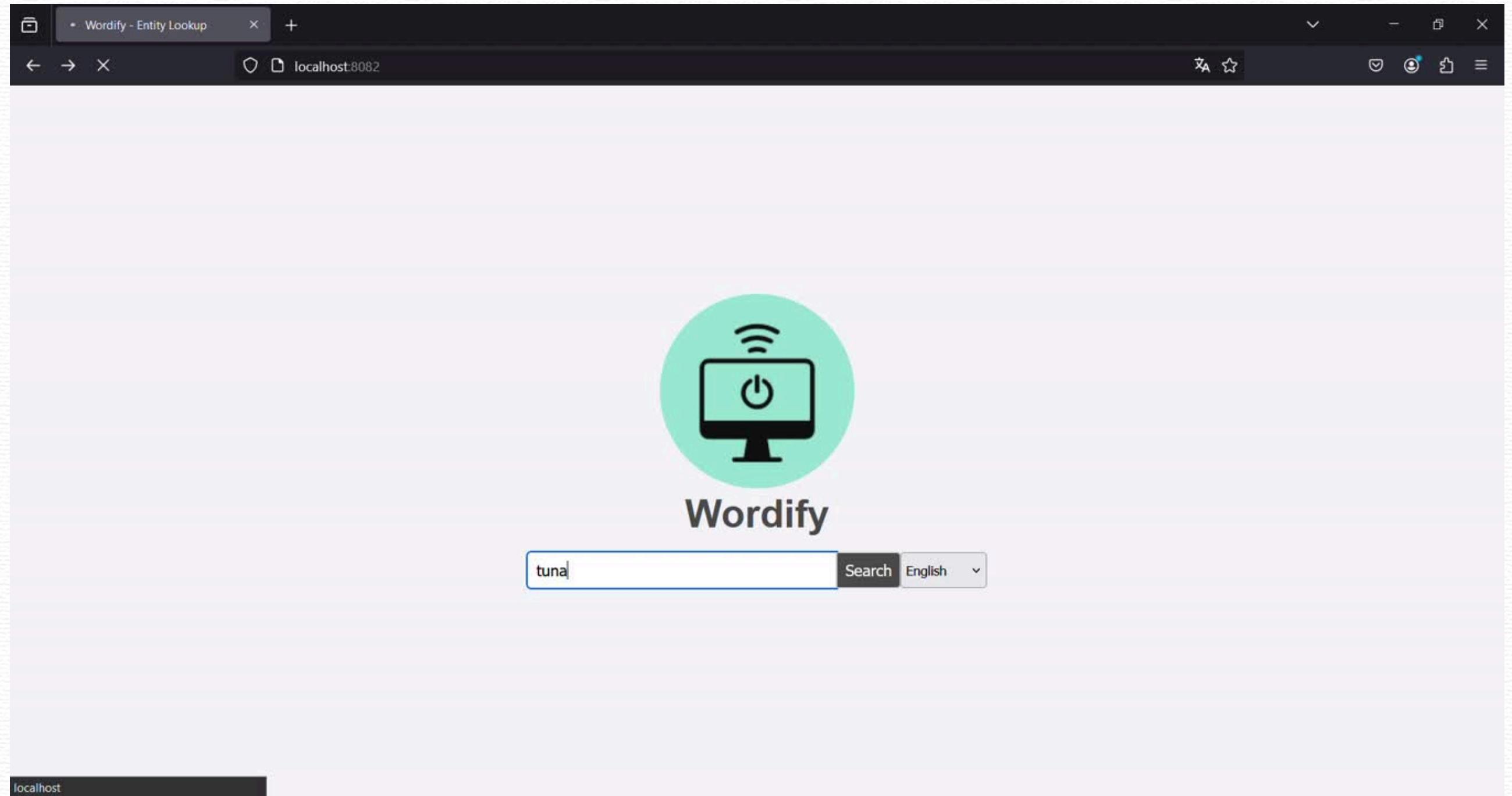
{Languages Probabilities:
- IT: 14.29%
- EN: 14.29%
...}



Linked Data principles

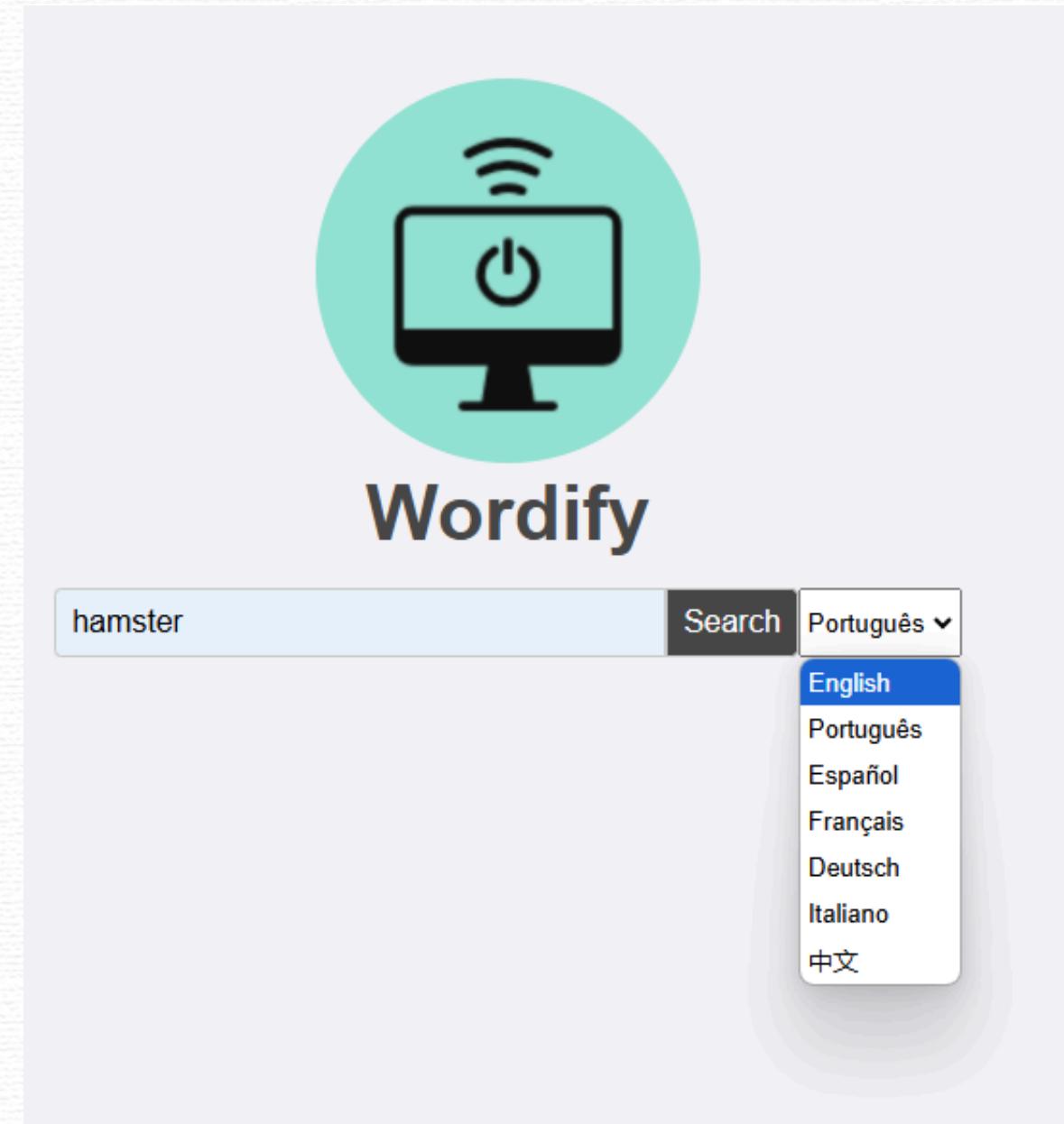


Video demo



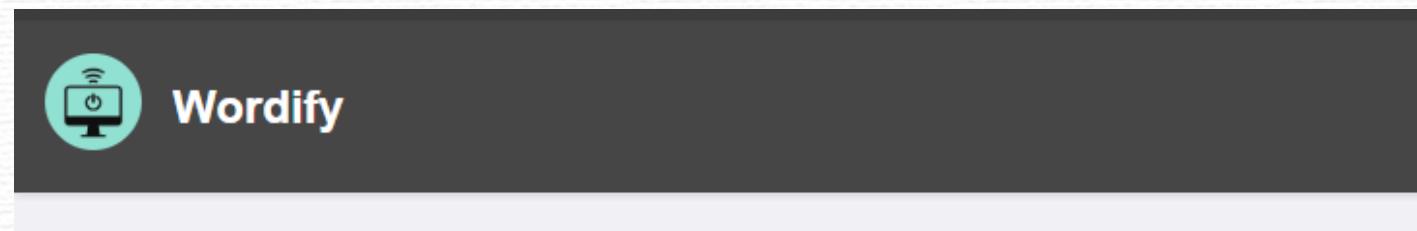
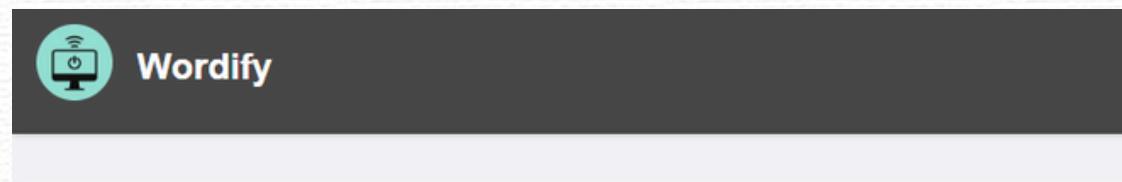
Limitations

- Hard to find definitions in specific languages.



Limitations

- Cross data between DBpedia and Wikidata.



crow

Meaning

heraldic animal

Source Filter

DBpedia Wikidata

Top 5 Languages: (ja: 12.0%, it: 8.0%, fr: 8.0%, en: 8.0%, nl: 8.0%)

Ambiguities

[figura araldica](#) vs [un genere di uccelli della famiglia Corvidae](#) (it)

[animal heraldique](#) vs [oiseau](#) (fr)

[heraldic animal](#) vs [bird of the genus Corvus, or more broadly a synonym for all of Corvus](#) (en)

[wapendier](#) vs [taxon](#) (nl)

[Wen Zhang Dong Wu](#) vs [suzumeMu karasuKe noNiao](#) (ja)

[heraldicna zival](#) vs [ptic iz rodu Corvus ali sirse sinonim za vse vrste Corvus](#) (sl)

[Ri Ben nopyroge-ma-](#) vs [Wen Zhang Dong Wu](#) (ja)

crow

Meaning

A crow is a bird of the genus Corvus, or more broadly a synonym for all of Corvus. Crows are gene

Source Filter

DBpedia Wikidata

Top 5 Languages: (cs: 16.67%, en: 16.67%, fr: 16.67%, ko: 16.67%, ja: 16.67%)

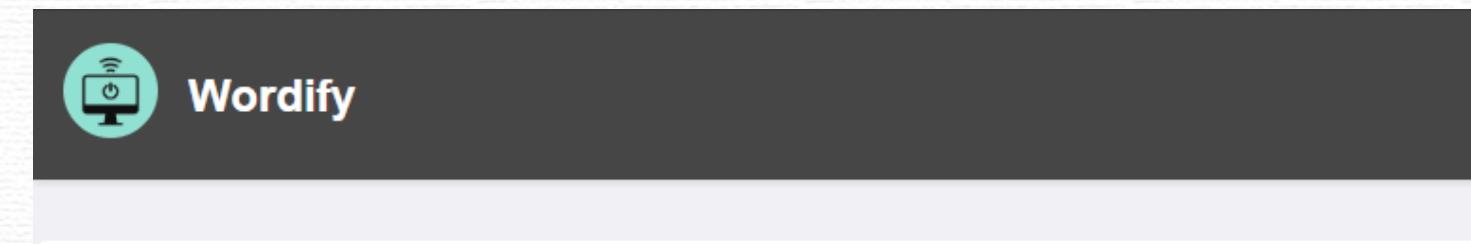
Ambiguities

Limitations

- Impossible to make regionalisms (due to location and languages).



The screenshot shows the Wordify interface for the word "fino". At the top, there's a dark header bar with the "Wordify" logo (a computer monitor icon) and the word "fino". Below the header is a light gray navigation bar with the "Source Filter" button and two radio buttons for "DBpedia" and "Wikidata". The main content area displays the word "fino" in large, bold, black letters. Underneath, there's a "Meaning" section with the text "None". A "Source Filter" section follows, with the "Wikidata" radio button selected. At the bottom, it says "Top 5 Languages: (eo: 100.0%)". A callout box labeled "Ambiguities" is visible at the bottom left.



The screenshot shows the Wordify interface for the word "imperial". The layout is identical to the "fino" interface. It features a dark header bar with the "Wordify" logo and the word "imperial". Below is a light gray navigation bar with the "Source Filter" button and radio buttons for "DBpedia" and "Wikidata". The main content area shows "imperial" in large, bold, black letters. Underneath, there's a "Meaning" section with the text "None". A "Source Filter" section follows, with the "Wikidata" radio button selected. At the bottom, it says "Top 5 Languages: (eo: 100.0%)". A callout box labeled "Ambiguities" is visible at the bottom right.

Limitations

- Limited to single words.

 Wordify

bico de pato

Meaning

None

Source Filter

DBpedia Wikidata

Ambiguities

 Wordify

bico-de-pato

Meaning

O bico-de-pato (*Machaerium nyctitans* (Vell. Conc.) Benth.) é uma árvore da família Fabaceae. Sinonímia botânica: *Machaerium nictitans* (Vell. Conc.) Benth.,

Source Filter

DBpedia Wikidata

Top 5 Languages: (en: 33.33%, pt: 33.33%, sv: 33.33%)

Ambiguities

Thank you!