Topic Modelling with Wikidata

Scientific Greetings – October 31, 2021

Houcemeddine Turki
Data Engineering and
Semantics Research Unit,
University of Sfax, Tunisia

Mohamed Ali Hadj Taieb
Data Engineering and Semantics
Research Unit, University of Sfax,
Tunisia

Mohamed Ben Aouicha
Data Engineering and
Semantics Research Unit,
University of Sfax, Tunisia

WIKI DATA CON

2021 - a sustainable
future for Wikidata

Data Engineering and Semantics Research Unit

Affiliated at the University of Sfax, Tunisia
Officially recognized in 2021

- Specialized in data-driven intelligent systems

- Develops real-life applications of the Wikimedia Projects

- Works on the automatic enrichment and validation of Wikidata
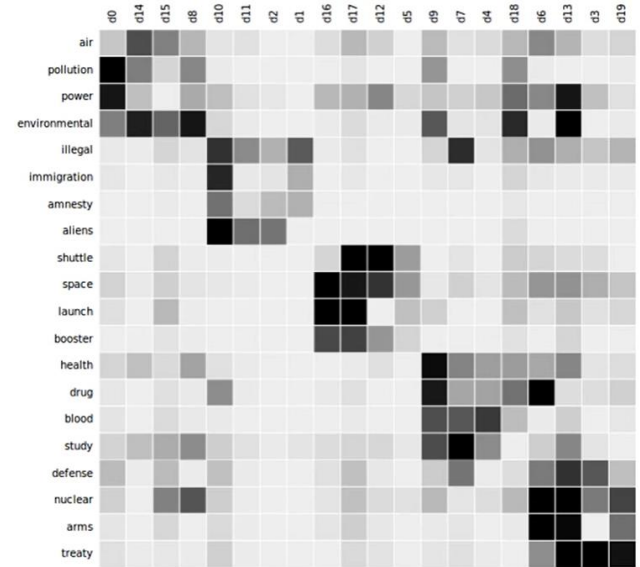
- Collaborates with Wikimedia Tunisia

Data Engineering and Semantics
هندسة البيانات و دلالاتها

WIKI DATA CON
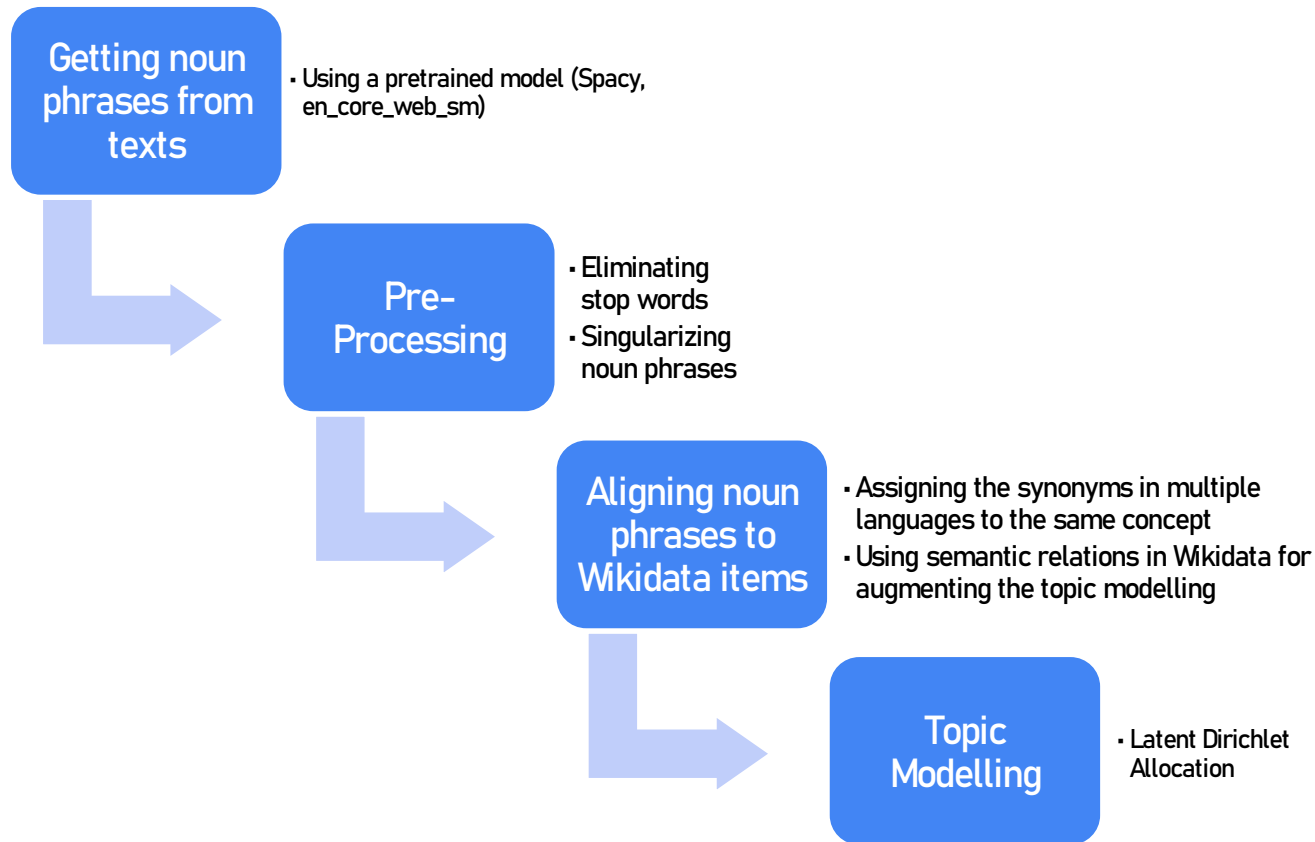2021 - a sustainable future for Wikidata

Topic Modelling

Probabilistic Model allowing the representation of the topics in a given collection of texts

- Latent Semantic Analysis

- Latent Dirichlet Allocation

- Non-negative matrix factorization

Topic Modelling with Wikidata: Our approach

Getting noun phrases from texts

- Using a pretrained model (Spacy, en_core_web_sm)

Pre-Processing

- Eliminating stop words
- Singularizing noun phrases

Aligning noun phrases to Wikidata items

- Assigning the synonyms in multiple languages to the same concept
- Using semantic relations in Wikidata for augmenting the topic modelling

Topic Modelling

- Latent Dirichlet Allocation

WIKI DATA CON

2021 - a sustainable future for Wikidata

Topic Modelling with Wikidata: Useful applications

- Multilingual Classification of Entities: Assigning main subject statements to Wikidata items about scholarly publications

- Social Network Analysis: Topic Modelling of posts about a given area for sentiment analysis and text summarization

- Recognition and validation of semantic relations in Wikidata and other knowledge graphs

- Enrichment of Wikidata and other knowledge graphs

Participation to WikidataCon 2021

- Wikifunctions and Abstract Wikipedia: Community Perspectives
  October 30, 2021 - 13:20–13:40 UTC, Room 2

- Systematic Review Automation driven by Wikidata
  October 31, 2021 - 12:40–13:00 UTC, Room 2

- Wikidata & Education: A Global Panel
  October 31, 2021 - 15:00–15:55 UTC, Room 2

Thanks for your attention!

Get in touch with us:

Houcemeddine Turki
turkiabdelwaheb@hotmail.fr

Mohamed Ali Hadj Taieb
mohamedali.hajtaieb@fss.usf.tn

Mohamed Ben Aouicha
mohamed.benaouicha@fss.usf.tn

WIKI DATA CON
2021 - a sustainable future for Wikidata