

ANALYSING CSTP REPORTS UNDER THE OECD SEMANTIC ANALYSIS FRAMEWORK

CSTP-TIP Workshop: Semantic Analysis for Innovation Policy 12 March 2018

DKI Team:
Frédéric Abrazian
Mary-Ann Grosset
Jan-Anno Schuur
Thierry Vebr

STI Team:
Andrés Barreneche
Alina Deniau
Michael Keenan
Blandine Serve



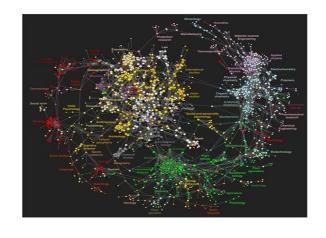


THE OECD SEMANTIC ANALYSIS FRAMEWORK

Taxonomies/Ontologies

Breaking the barriers

Of LANGUAGE and WORDS



Create concepts and link them so that machines can use them as a cloud of knowledge (Semantic layer)

Capture and formalise Subject-Matter-Experts' knowledge

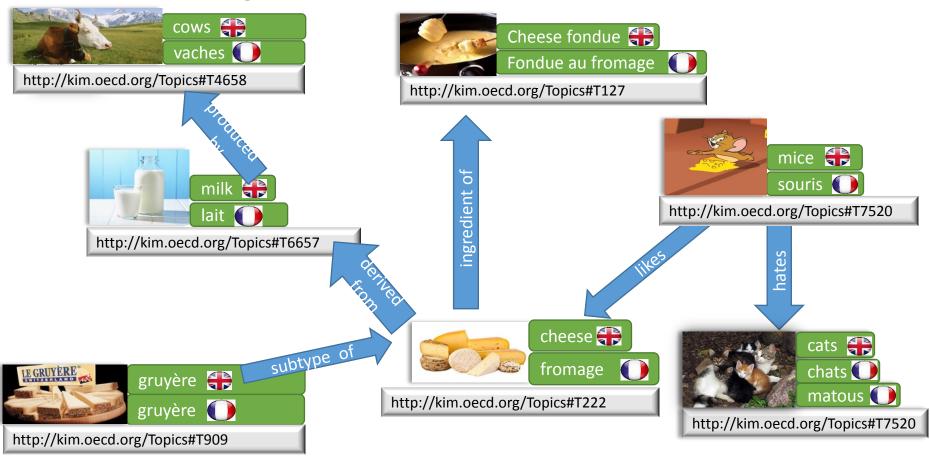
Identify synonyms, related terms in official languages

Use URIs not labels

Use in:

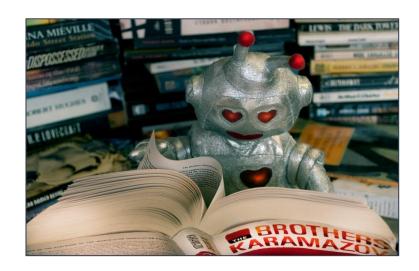
Semantic analysis tools
Search and discover tools

Breaking the barriers



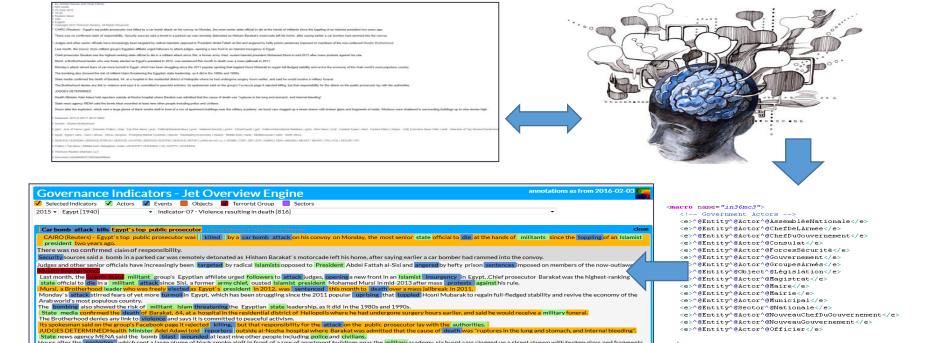
How do we create the semantic robots?

- Use lexicon (taxonomies)
- Use text patterns (part of speech)
- Identify relationships (ontologies)
- Test the results on a set of documents
- Debug Disambiguate
- Test on the complete corpus
- Put in production using Web Services



36 different robots in production

Semantic enrichment



Hours after the explosion, which sent a large plume of black smoke aloft in front of a row of apartment buildings near the military academy, six burnt cars clogged up a street strewn with broken glass and fragments of metal.
Windows were shattered in surrounding buildings up to nine stories high."It felt like the apartment was lifted up and dropped right back on the ground.
When I felt it I said to my family, this is it, death has come for us," said 17-year-old Khaled Youssef.
Sis! soffice mourned Barakat's death and canceled celebrations prepared for Tuesday to mark the anniversary of mass unrest that led to Mursi's fall, and said the perpetrators would draw "the most severe

The government, however, declared June 30 a national holiday in commemoration of the 2013 protests, state media said. "These kinds of vicious attacks will not deter the state from continuing its path of development, the adoption of rights, and realizing the hopes and aspirations of the Egyptian people," the presidency's office said in a statement according to MENA. The Egyptian judiciary says it is independent of the

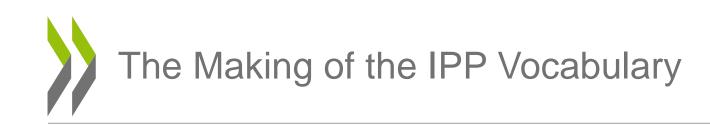
nment and military. But some of Egypt's judges have drawn accusations of blatant bias by handing down lengthy jail terms and mass death sentences against Islamists.

Returns articles covering protests, demonstrations, manifestations, etc...

Search for protests in Africa



ANALYSING TEXTS TIP, CSTP and Research Policy



IPP Vocabulary has +1000 unique concepts

relevant within the field of STI policy

MAIN SOURCES USED

- Consultations with STI colleagues
- Oxford Handbook of Innovation
- Semantic analysis of OECD / World Bank flagship publications
 - Ranking of the **10 000** most frequently used terms

Features

- +300 synonyms
- Generic concepts (e.g. firms)
 are not used for tagging

Limitations

- Flat structure, not taxonomy has been built on top
- Needs updating (e.g. digital economy and inclusive innovation)

Source	Number of items	Text analysed	Time-period
CSTP <u>substantive</u> reports	782	Full text	1993-2017
Subset: TIP <u>substantive</u> reports	100	Full text	1994-2016
Research policy articles	2 527	Titles + abstracts	1988-2016

Output from Luxid system:

- Maximum of 20 IPP topics identified per item.
- For CSTP+TIP documents: A total of 12 390 topics identified.



Treatment of semantic analysis results

- Creation of word clouds for topics (50 most recurrent).
- Calculation of yearly occurrences of IPP topics identified as keywords.
- **Normalisation** of yearly occurrences by number of items analysed in the respective year (to make occurrences comparable over time).
- Data aggregation in time-periods of 8 years: 1993-2000, 2001-2008 and 2009-2017.
- **Ranking** topics in terms of **variance** across the 3 timeperiods to identify key trends.