Euroscepticism and Big-Data

Federico Nanni, Simone Paolo Ponzetto and Laura Dietz
Data and Web Science Group
University of Mannheim

Background

Bachelor and Master in **Contemporary European History**

Final year PhD student in **Digital Humanities** at Centre for the History of Universities and Science, University of Bologna

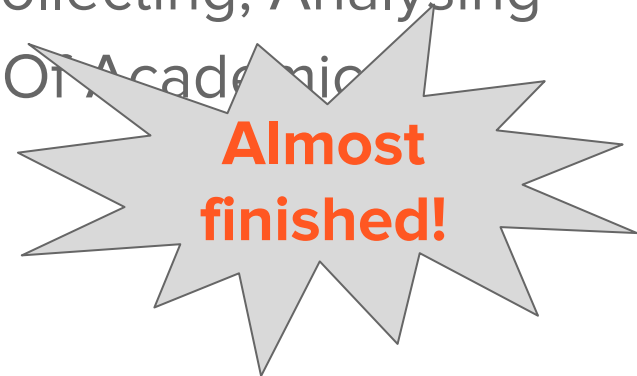
Thesis: “The Web as a Historical Corpus: Collecting, Analysing and Selecting Sources On the Recent Past Of Academic Institutions”

Background

Bachelor and Master in **Contemporary European History**

Final year PhD student in **Digital Humanities** at Centre for the History of Universities and Science, University of Bologna

Thesis: “The Web as a Historical Corpus: Collecting, Analysing and Selecting Sources On the Recent Past Of Academic Institutions”



Working at the University of Mannheim

Researcher at the **Data and Web Science Group** and at the **Political Science Department.**

Project: building **topic-specific collections** for supporting research on the international relations of the United States.

New research proposal

Combining new sources and new methods for advancing our comprehension of **Euroscepticism**.

Today's talk

1. Quick overview on the topic and traditional approaches
2. Present new sources - and how to collect them
3. Describe the work we've been doing on new methods
4. Recap on how everything converges
5. Small surprise

Euroscepticism

Euroscepticism

Euroscepticism

Euroscepticism

Euroscepticism

Literally, it means **criticism** of the European Union (EU).

Hard euroscepticism is a principled opposition to the EU and European integration.

Soft Euroscepticism is support for the existence of, and membership of, a form of European Union, but with opposition to specific EU policies.

Euroscepticism

It does not follow common left-right divisions of the political spectrum, but it could be represented as a **new one-dimensional scaling** (Proksch and Slapin, 2010):

Pro Eu Integration

Against EU Integration

How it is traditionally studied

Traditional approaches:

- **Polls** and **surveys**
- Establishing the position of political parties towards the EU using **party-manifestos** (Conti, 2013)
- analysing the **opinion** of **candidates** (McAllister and Studlar, 2000)
- hermeneutic **close reading** of political news (Vallaste, 2013).

How it is traditionally studied

Limits:

- Strong **top-down** view
- Works that analyse content are often based on **close reading**
- Does **not** easily **capture interactions** (media - elites - people)

Solution: Big Data!

Solution: Big Data!

Define research questions first!

1) Is it true that **Euroscepticism** has been **growing** or only its political expression has suddenly increased?

Define research questions first!

1) Is it true that **Euroscepticism** has been **growing** or only its political expression has suddenly increased?
2) Political **campaign** Anti-EU vs **establishment** Anti-EU?

Define research questions first!

1) Is it true that **Euroscepticism** has been **growing** or only its political expression has suddenly increased?
2) Political **campaign** Anti-EU vs **establishment** Anti-EU?
3) What **role** has the **web** played in it?
    a) What do we mean by “role”?
    b) What do we mean by “web”?

Define research questions first!

1) Is it true that **Euroscepticism** has been **growing** or only its political expression has suddenly increased?
2) Political **campaign** Anti-EU vs **establishment** Anti-EU?
3) What **role** has the **web** played in it?
    a) What do we mean by “role”?
    b) What do we mean by “web”?
4) Different countries **different Euroscepticism**?

Challenges

1. Sources are **difficult to acquire**:
   -> not all countries have a National web archive
2. Sources are **complex to process**
   -> not only big data - but different languages
3. **Methods** are **not** as **reliable** as we think
   -> did someone say topic models? :D

Case Studies

I am currently working on studying the growth of Euroscepticism **during the last 17 years** (1999 - 2016), in three different countries: UK, Germany and Italy.

Case Studies

I am currently working on studying the growth of Euroscepticism **during the last 17 years** (1999 - 2016), in three different countries: UK, Germany and Italy.

Challenges:

- Three different languages
- No National web archive for Italy
- No access to the web archive for Germany (waiting for L3S)
- No text mining on UK web archive

Born Digital Sources Available

Political Manifestos:

- Manifesto Project website offers over 1000 parties from 1945 until today in over 50 countries on five continents.

Parliament Speeches:

- EuroParl covers all plenary debates held in the EP since July 1999, manually translated in 23 languages.
- Transcript of National Parliament debates (e.g. parliament.uk)

Government / Party / Campaign Websites:

- From the Internet Archive

Born Digital Sources Available

Media:

- Internet Archive provides snapshots of media websites
- TREC KBA web archive News Collection (2011-2013)
- Archive-Team News Grabber (still not available)
- Common Crawl News Dataset (2016 - ?)

People:

- Large datasets, e.g. Reddit Corpus (other collections you would suggest?)
- Topic-specific Twitter collections (#brexit, etc)

Methods we are developing

General goal: topic-based analyses

Task: Given a document, identify **central topics**.

Underlying goal: linking topics to the **semantic web.**

Final goal: map how **topical-opinions** spread:

- Country-based
- Across countries

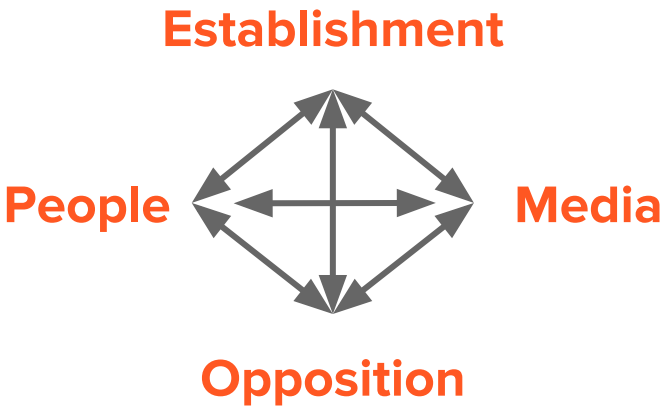
General goal: topic-based analyses

Task: Given a document, identify **central topics**.

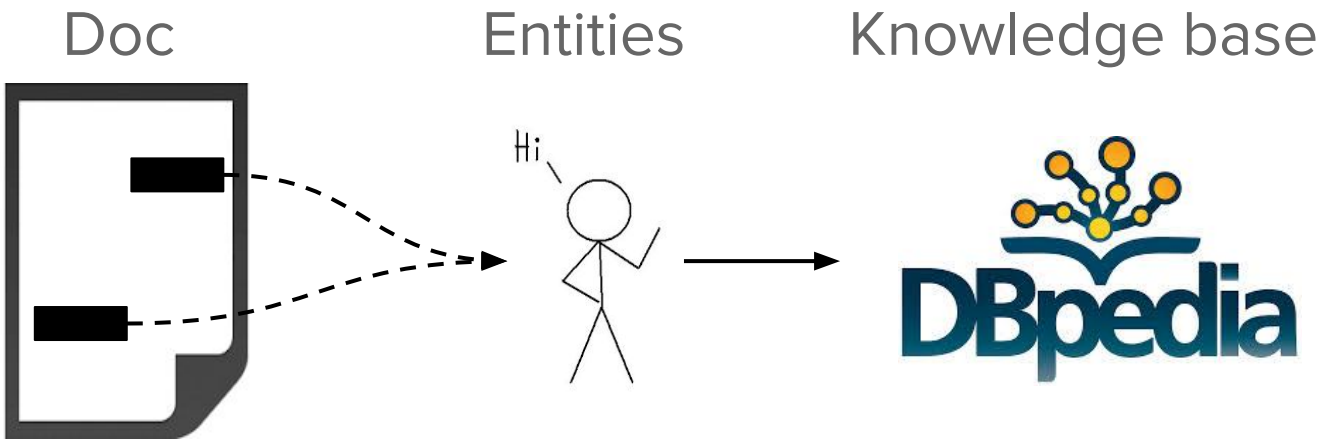
Underlying goal: linking topics to the **semantic web.**

Final goal: map how **topical-opinions** spread:

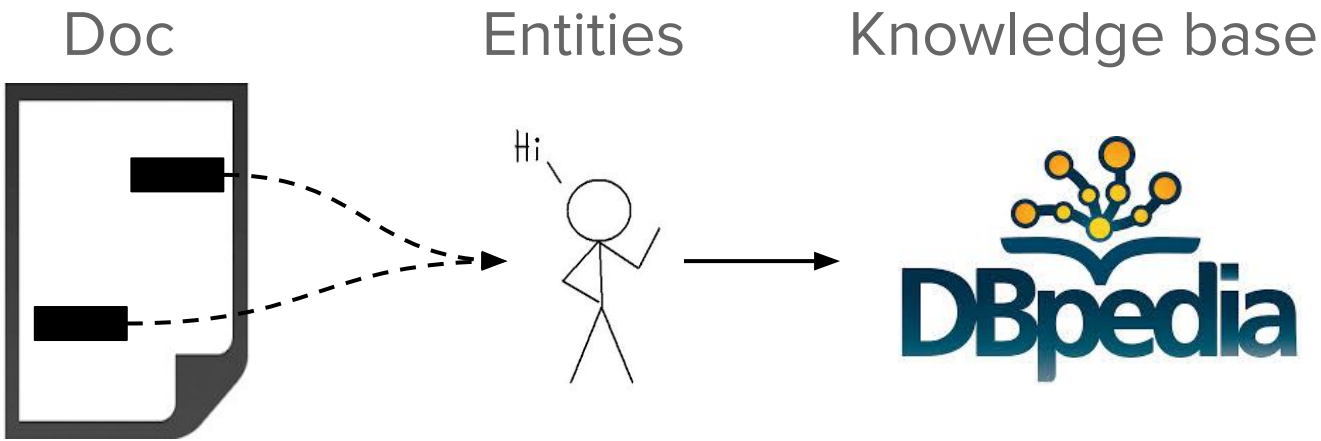
- Country-based
- Across countries



Entities as Topic Labels (with Anne Lauscher)



Entities as Topic Labels (with Anne Lauscher)



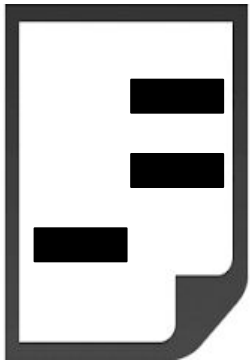
fy Babelfy "Word sense disambiguation and entity linking together!"

TAG·ME

DBpedia Spotlight

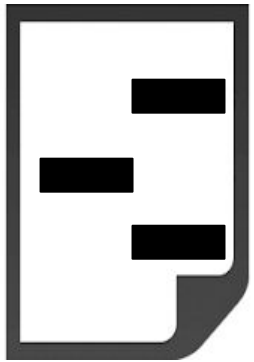
Entities as Topic Labels (with Anne Lauscher)

Doc1



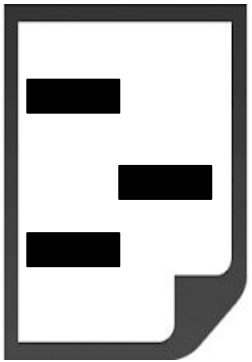
Entity1
Entity2
Entity3

Doc2



Entity1
Entity2
Entity3

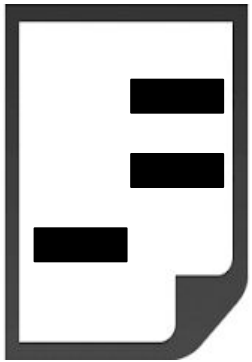
Doc3



Entity1
Entity2
Entity3

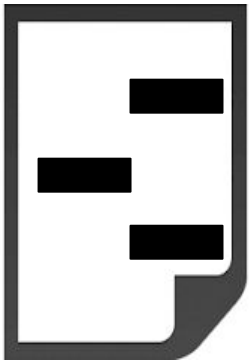
Entities as Topic Labels (with Anne Lauscher)

Doc1



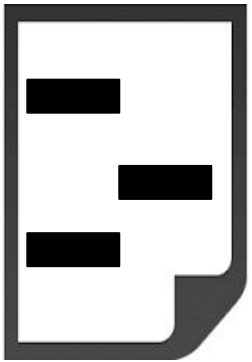
Entity1
Entity2
Entity3

Doc2



Entity1
Entity2
Entity3

Doc3



Entity1
Entity2
Entity3

Labeled LDA!

Entities as Topic Labels (with Anne Lauscher)

Examined most relevant topics addressed by each party in the European Parliament's fifth term (1999-2004).

Les Verts (France)

	Label: Consumer (47%)	**Label: GMO** (34%)
Topic words	product directive consumer safety law market	human health agreement food measure sustainable

Topic-based political scaling (with G. Glavas, C. Zirn)

Measuring lexical and semantic textual similarities between politicians.

Topic-based political scaling (with G. Glavas, C. Zirn)

Measuring lexical and semantic textual similarities between politicians.

Refugee Crisis

Pro Eu Integration

Against EU Integration

Topic-based political scaling (with G. Glavas, C. Zirn)

Measuring lexical and semantic textual similarities between politicians.

Topic-based political scaling (with G. Glavas, C. Zirn)

Measuring lexical and semantic textual similarities between politicians.

Cross-lingual political scaling (with G. Glavas)

Given a series of texts about the same topic in two different languages we scale them from pro-EU to against-EU.

Economic Crisis

Cross-lingual political scaling (with G. Glavas)

Given a series of texts about the same topic in two different languages we scale them from pro-EU to against-EU.

Economic Crisis

Pro Eu Integration

Against EU Integration

We are also working on

Measuring **disputed topics** in UK political speeches (with Mahmoud Hussein).

Detecting fine-grained **controversial aspects** in political topics (with Aya Mobayad).

Identifying topic-based **agreement and disagreement** in political discussions (with Stefano Menini and Sara Tonelli).

Where are we, so far

Task: Given a document, identify **central topics**.

Underlying goal: linking topics to the **semantic web.**

Where are we, so far

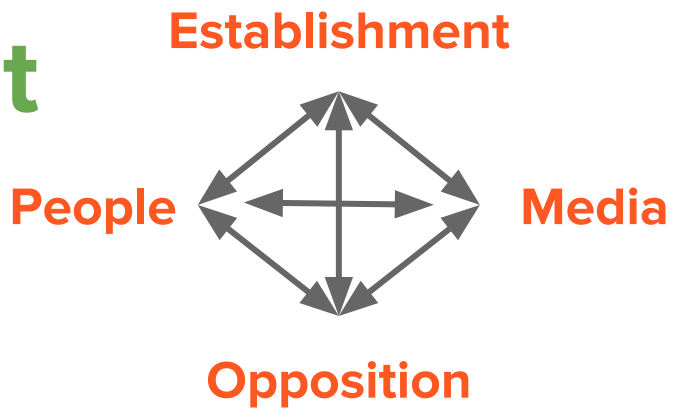
Task: Given a document, identify **central topics**.

Underlying goal: linking topics to the **semantic web.**

Final goal: map how **topical-discussions** spread:

- Country-based
- Across countries

#brexit



Karlsruher Institut für Technologie

UNIVERSITY OF MANNHEIM

Take away

New sources and new methods could guide us in **better understanding** Euroscepticism.

National web archives are **essential** resources to expand this kind of study (esp. for understanding the **role of the web**).

Interdisciplinary competencies and **source + tool criticism** strongly needed, to not get caught by big-data rhetoric.

Hackathon on Euroscepticism and Big Data Summer 2017!

Three days at the beautiful **Lake Garda**, Italy.

20 scholarships available for PhD Students and Post-Docs, from humanities, social sciences and computer science.

Call for application out soon!

UNITELMA SAPIENZA

UNIVERSITY OF MANNHEIM

FONDAZIONE BRUNO KESSLER

TECHNISCHE UNIVERSITÄT DARMSTADT

Questions?

UNIVERSITÄT MANNHEIM

Federico Nanni

Data and Web Science Group

University of Mannheim

federico@informatik.uni-mannheim.de