Discovering an Encyclopaedic Novel

a case study in automatically analysing Harry Mulisch's *The Discovery of Heaven* (1992)

https://github.com/LvanWissen/Ontdekking/

Encyclopaedic novel

- Edward Mendelson
  - Technologies and sciences
  - Broad range of subjects
- Examples:
  - *Ulysses* (1922)*, Faust* (1808)
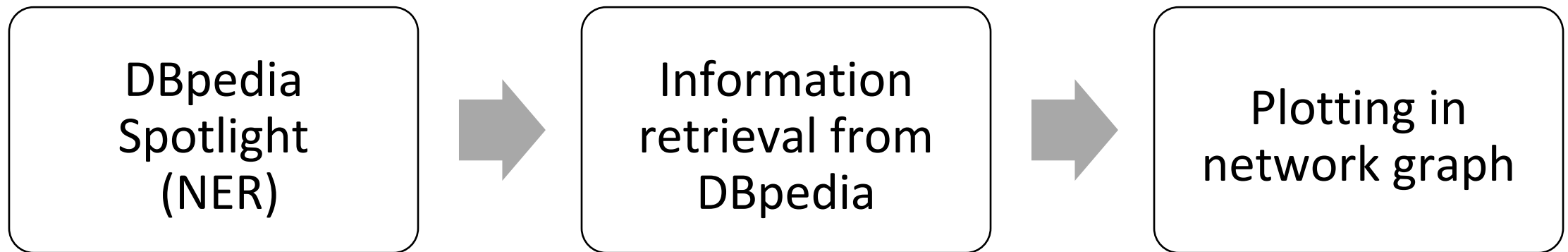 - *The Discovery of Heaven* (1992)

Encyclopaedic novel: Relevancy

- Large (publicly) available datasets
  - Wikipedia

- Intertextuality and hypertext

- Network of knowledge
  - Facts and references

- *Discovery of Heaven*:
  - Example: Opposition Alpha - Beta

Method: Step by step

Case study: *The Discovery of Heaven* (1992) by Harry Mulisch

DBpedia Spotlight (NER)		Information retrieval from DBpedia		Plotting in network graph

https://github.com/LvanWissen/Ontdekking/

Spotlight

- Named Entity Recognition
  - DBpedia URI

- In *The Discovery of Heaven*:
  - 18.110 entities
  - 4.775 unique entities

Ik ben makelaar in koffie, en woon op de Lauriergracht No 37.

http://nl.dbpedia.org/resource/Koffie

http://nl.dbpedia.org/resource/Makelaar   http://nl.dbpedia.org/resource/Lauriergracht

[I am a coffee-broker, and live at No. 37 Laurier Canal Amsterdam.]

(*Max Havelaar*, 1860)

DBpedia

- 'Structured' Wikipedia data
  - RDF triples
  - SPARQL

- Using DBpedia as ontology
  - 'Hierarchy'
  - Simple Knowledge Organisation System

DBpedia: Method

URI_NL   →   URI_EN   →   **Subjects (categories)**   **Broader categories**
                                                                                                                                         Vertebrates

http://nl.dbpedia.org/resource/Heikikker   http://dbpedia.org/resource/Moor_Frog   **Broader categories**

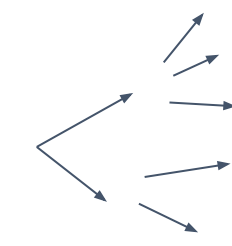   Fauna

                                                                                              Amphibians   Animals

                                                                                              Arctic land   Amphibious
                                                                                              animals   organisms

                                                                                              ...

About: Moor frog

An Entity of Type : species, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

The moor frog (Rana arvalis) is a slim, reddish-brown, semiaquatic amphibian native to Europe and Asia. It is a member of the family Ranidae, or true frogs.

dct:subject
- dbc:Fauna_of_Finland
- dbc:Amphibians_of_Europe
- dbc:Animals_described_in_1842
- dbc:Arctic_land_animals
- dbc:Fauna_of_Siberia
- dbc:Rana_(genus)
- dbc:Amphibians_of_Asia

skos:broader
- dbc:Vertebrates_of_Europe
- dbc:Amphibians_by_continent

skos:broader
- dbc:Vertebrates_by_continent
- dbc:Fauna_of_Europe

...

skos:broader
- dbc:Subfields_by_academic_discipline
- dbc:Biology

https://github.com/LvanWissen/Ontdekkin

DBpedia: Method

- Continue until reached:
  - http://dbpedia.org/resource/Category:Disciplines
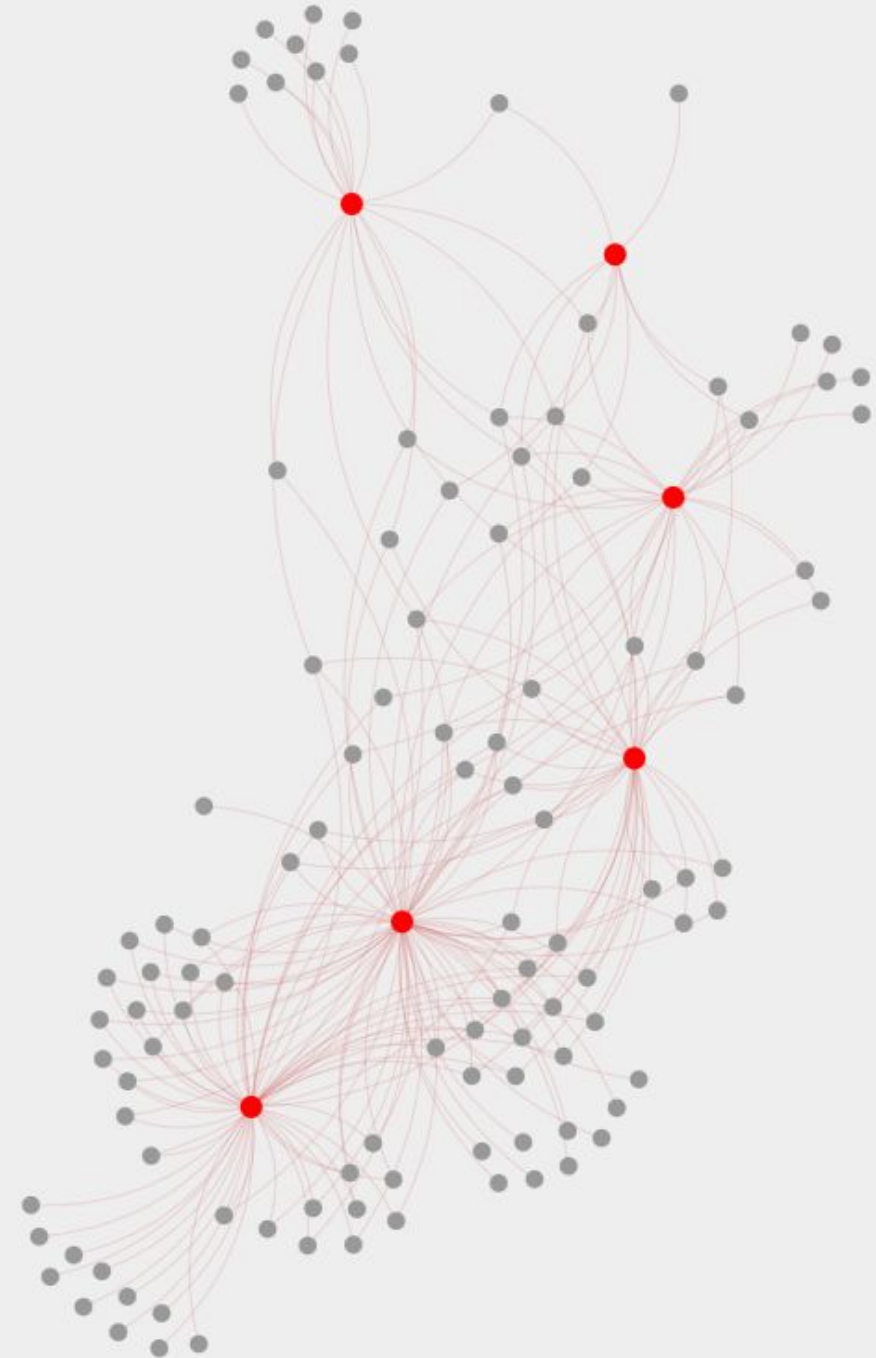 - Fixed number of tries (preventing infinite loops)

Result:

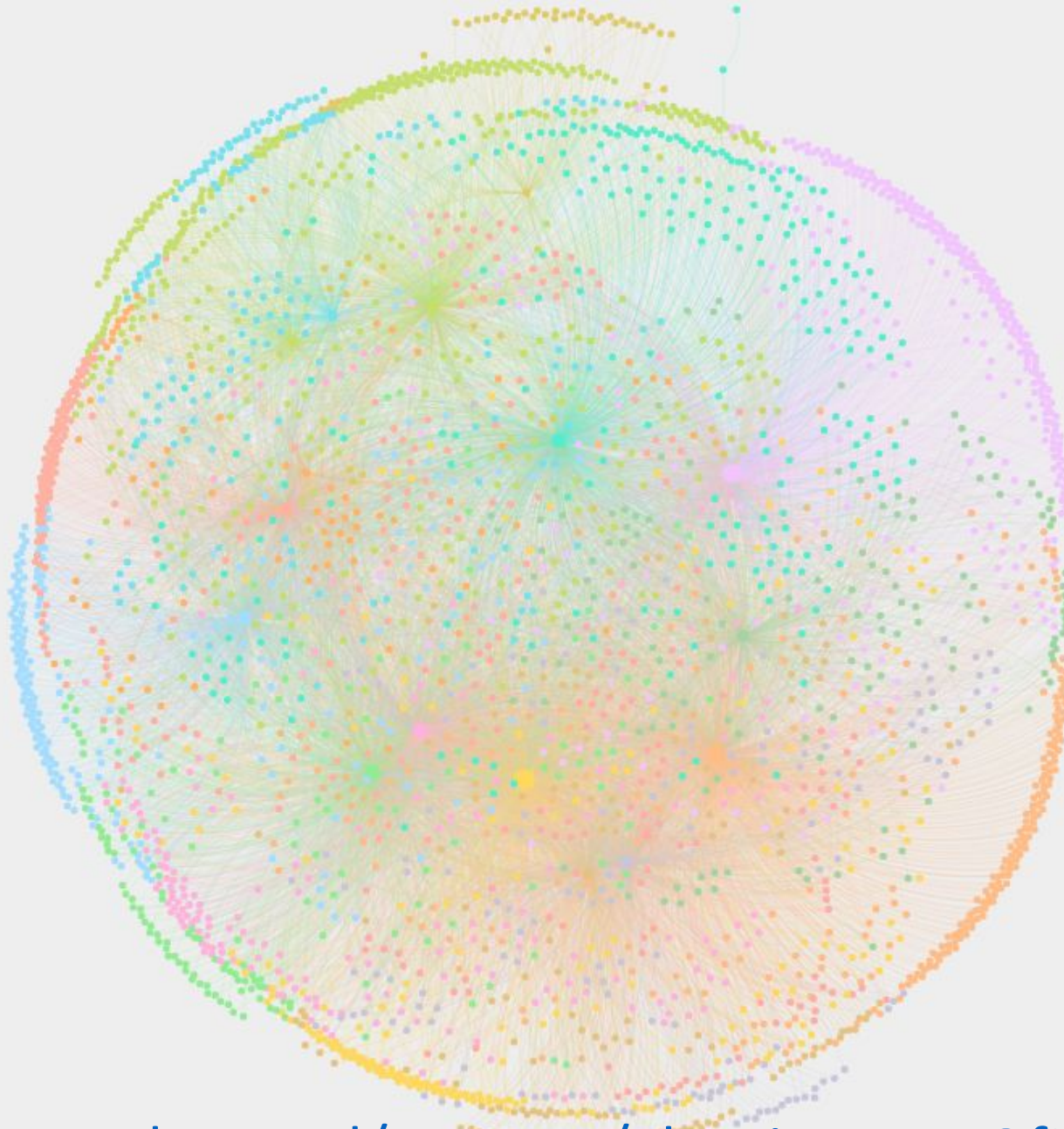- Route to top category for every URI

Disciplines

- Anthropology
- Archaeology
- Biology
- Chemistry
- Culinary arts
- Economics
- Geography
- Health
- History

- Law
- Mathematics
- Musicology
- Philology
- Philosophy
- Physics
- Psychology
- Religion
- Sports

Network graph

- Route list with categories (hierarchy)

- Nodes
  - Entities + Categories
- Edge is drawn if:
  - a discipline occurs in route list

- Visual representation:
  - Gephi

https://github.com/LvanWissen/Ontdekking/

Output

http://kyoto.let.vu.nl/~vanerp/TheDiscoveryOfHeaven/

Evaluation & feedback

- Named Entity Recognition: Is Spotlight the best tool for this?

- Available information in DBpedia?
  - Better than nothing?
  - More datasets?

- Would this be useful for humanities?

https://github.com/LvanWissen/Ontdekking/