Creating Context Networks in Dutch Legislation

Radboud Winkels
Alexander Boer
Ivan Plantevin

Problems with Dutch Legislative Portal

- Where in the law are we?
- Incoming references?
- Earlier or later versions?
- Other relevant sources of law?

Research Question

Given a particular document (article) in *focus*, can we determine other relevant documents purely on the basis of 'objective' meta-information?

- No interpretation of the content of the documents;
- No use of metadata added by other sources than the official owners and publishers of the documents.

The Web of Law

- Sources of Law form a network through references:
  - Implicit and explicit
  - Within and between documents
- The network grows over time
  - New documents and references
  - Different versions of same documents
- Analyse the network
  - Structure of the domain
  - Importance of documents
  - Trends over time

Related Research for Legislation

- Boulet e.a. for French law (2009 a, b)
  - Entire network at level of laws
  - One specific law
- Winkels & Boer (2013)
  - Graph representation of weighted local network
  - Just two seed nodes

Now:

- A unweighted context network
- Text representation closer to official portal from *any* focus node

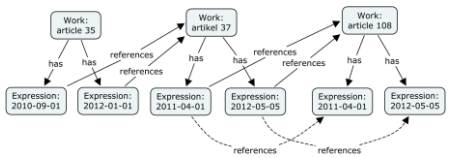
Additional Information

- **Internal - General:** A list of the most important texts in the current law.
- **Internal - Focus:** Texts in current law that are citing the text in focus or are cited by it, ordered by importance.
- **External - Focus:** Texts from other laws that are citing the text in focus or are cited by it, ordered by importance.
- **Versions - Focus:** A list of different versions of the text in focus.

Creating a Context Network

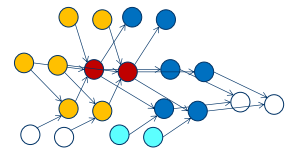
MetaLex Document Server contains all data from official portal as:

- CEN MetaLex XML
- RDF linked data (> 290 million triples)
- Work – Expressions (FRBR levels)



Creating a Context Network -2

- SPARQL queries for **incoming** and **outgoing** references of **focus** node
- SPARQL queries for versions of focus node



Prototype Network

- Generated off-line
- Six laws tax domain
  (cf. Winkels & Hoekstra, JURIX 2012)
- Citations within six laws are resolved and retrieved
- Others included but not further analysed
- 7,992 nodes and 13,496 edges

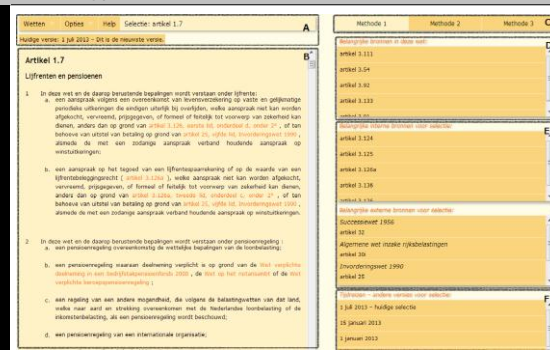
Analysing the Network for Importance

1. *In degree*
   number incoming citations node
2. *Degree centrality*
   number of nodes a node is connected to
3. *Betweenness centrality*
   number of shortest paths that pass through node
   (most 'expensive')

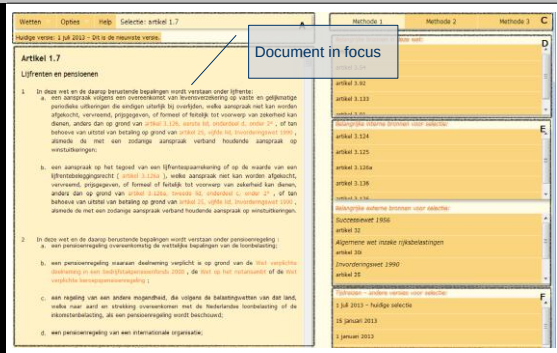
The Prototype

- Creating, analysing network and collecting versions in Python modules
- Django server application
- Client in HTML5 and jQuery Javascript
- All recent browsers except IE
- Runs on most devices
- http://justinian.leibnizcenter.org/wetten

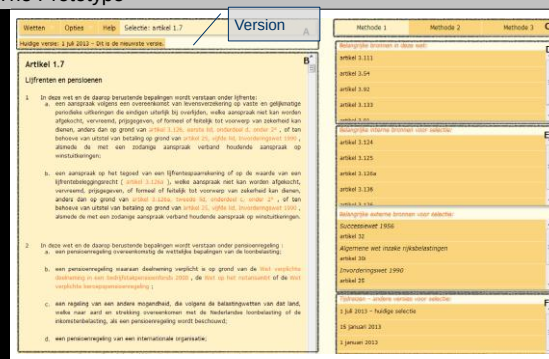
The Prototype



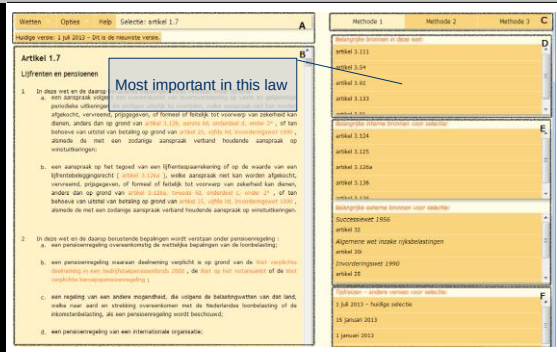
The Prototype



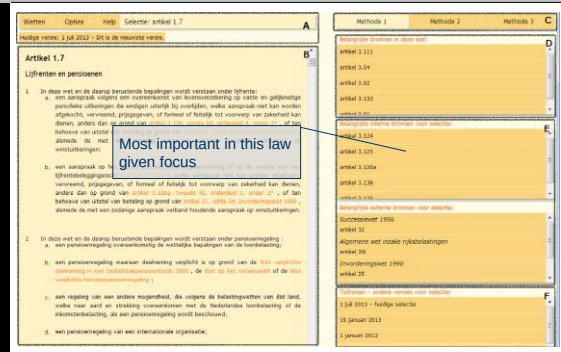
The Prototype



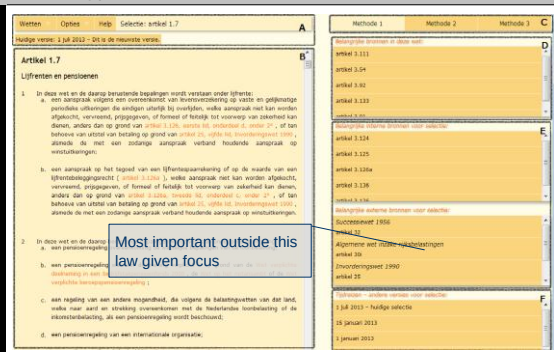
The Prototype



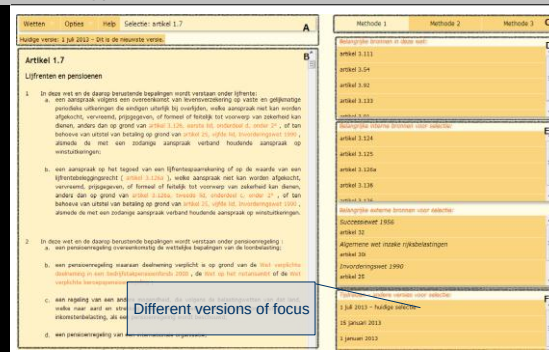
The Prototype



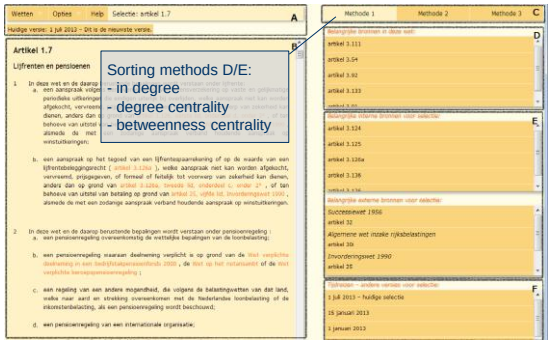
The Prototype



The Prototype



The Prototype



Formative Evaluation

Three Experts Tax Administration

- List important documents useful but depends on task
- Some other relevant documents missing, but not in BWB format and MDS
- Important documents given focus very useful but bug when time travelling
- No preference for sorting method (results quite similar)
- Time travelling very useful but not all future versions in database

Similarity Sorting Methods

Overlap and identical ranking for 7 lists:

1. Top 5 of most important texts in the Income Tax Law 2001 (Frame D)
2. Top 5 *internal* important texts for the three texts listed as most important in that law: articles 3.111, 10.1 and 2.5 (Frame E)
3. Top 5 *external* important texts for the three texts listed as most important in that law: articles 3.111, 10.1 and 2.5 (Frame E)

Overlap = $\frac{overlap\ results}{total\ results}$ Identical positions= $\frac{results\ same\ position}{total\ results}$

Means for all Lists

Comparison between:	Overlap	Identical positions	
In degree – degree centrality	0.60 (s = 0.23)	0.23 (s = 0.18)	Expected
In degree – betweenness centrality	0.40 (s = 0.26)	0.14 (s = 0.15)	
Degree centrality – betweenness centrality	0.66 (s = 0.25)	0.34 (s = 0.32)	?

The network has been built for only six laws that are *closely* related: group-like structure. Much cited and citing texts often lie on many shortest paths.

Conclusions and Future Work

Can we determine relevant context given a particular document in *focus*, purely on the basis of 'objective' meta-information?

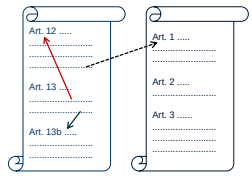
- **Yes**, but room for improvement and further research
- Larger, less related network
- Fix time-travel bug
- Include other portal functionality (table of content, search facility)

Conclusions and Future Work

- Include other sources of law
  - Case law, doctrine
  - New project: ***OpenLaws.eu***
- More (also summative) evaluation!
  - More users, different tasks
  - Text based visualisation or graph?
  - Importance network vs other sources (e.g. text books)
  - Weighted or unweighted network?

Weighing the Context Network

- Superiority
  - Laws vs decrees; outward > inward?
- External or Internal
- Anaphoric or Cataphoric
  - Anaphoric to *definitions* > cataphoric



Weighing the Context Network

- Superiority
  - Laws vs decrees; outward > inward?
- External or Internal
- Anaphoric or Cataphoric
  - Anaphoric to *definitions* > cataphoric
- Dynamics
  - Recency of change
  - Number of expressions of a work
- Centrality
  - In-degree, out-degree or combination

Questions or Suggestions

?!

winkels@uva.nl
www.LeibnizCenter.org