Visualization and Structure Analysis of Legislative Acts: A Case Study on the Law of Obligations

Innar Liiv, Anton Vedeshin, Ermo Täks
Department of Informatics, Tallinn University of Technology
15 Raja Street, 12618 Tallinn, Estonia
innar.liiv,anton.vedeshin,ermo.taks@ttu.ee

ABSTRACT

This paper introduces visualization methods for an initial approximation of the structure in any legislative act, procedures for relevance calculation and metrics for calculating the distance between the sections. To evaluate the methodology in real world scenarios, experimental results for the law of obligations are included and analyzed. The results of the study show that it is possible to get insightful and interpretable results when analyzing only the inner-structure of the legislative act constructed by references between the sections. Methods and metrics presented in this paper can be directly applied to current keyword-based searching systems of legal texts for relevance ranking enhancement.

1. INTRODUCTION

Individual sections in legislative acts are not of equal importance. Evaluating the influence of an individual section typically requires expert judgement and understanding of the whole system and its underlying structure. Our goal is to develop a methodology that could present approximate estimations of importance based on the structure of references in the legal text. Such approach requires no extra information and minimal or no changes from the current systems available. Possible applications include: advanced expert systems for citizens, enhanced search engines, legal drafting tools. Evaluation of influence could also help officials to estimate the impact of changes to be made in the legal text. A related research has been reported by Palmirani et al. [1] concerning the extraction of references from unstructured legal text. The initial parsing phase of the research is rather similar - as references are expressed in a natural language, the research group developed a prototype to identify references using regular expressions. However, the goal of that study was not to use the extracted information for further analysis, but to standardize and rewrite (tag) the references to allow unique and undubious identification. In this article, we take that step forward and present two kinds of approaches for such an analysis - human-interpretable visualization techniques that enable the perception of the inner structure of any legal text and computer-implementable methods for ranking and measuring the sections.

ICAIL '07 June 4-8, Palo Alto, CA USA

2. PROPOSED METHODOLOGY

§ 2325. Sales price[U.S.Code:Title 42,Chapter 24,Subchapter II]
(b) Other properties
In all other cases the sales price to priority purchasers shall be the appraised value less the deductions provided by **section 2326** of this title, except that sales made under **sections 2343** **(b)** and **2343** **(c)** of this title shall be made at the prices set forth therein.
(c) Appraised value of interest in commercial property
The appraised value of the Government's interest in commercial property shall, in the cases where renegotiation of the lease is requested by the lessee under the provisions of **section 2201** **(e)** of this title be based upon the renegotiated lease if any is agreed on.
§ 2326. Deductions from sales price
(d) Additional deduction to persons purchasing property without benefit of indemnity provisions
Persons purchasing property pursuant to the provisions of section **2342** of this title, who do not desire to avail themselves of the indemnity provisions contained in **sections 2363** **to** **2366** of this title, shall be entitled to an additional deduction of 10 per centum of the appraised value of the property in addition to any other deduction set forth in **this section**.

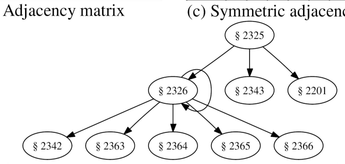
(a) Input: unstructured legal text with references in natural language

	§2201	§2325	§2326	§2342	§2343	§2363	§2364	§2365	§2366
§2201	0	0	0	0	0	0	0	0	0
§2325	1	0	1	0	1	0	0	0	0
§2326	0	0	1	1	0	1	1	1	1
§2342	0	0	0	0	0	0	0	0	0
§2343	0	0	0	0	0	0	0	0	0
§2363	0	0	0	0	0	0	0	0	0
§2364	0	0	0	0	0	0	0	0	0
§2365	0	0	0	0	0	0	0	0	0
§2366	0	0	0	0	0	0	0	0	0

(b) Adjacency matrix

	§2201	§2325	§2326	§2342	§2343	§2363	§2364	§2365	§2366
§2201	0	1	0	0	0	0	0	0	0
§2325	1	0	1	0	1	0	0	0	0
§2326	0	1	1	1	0	1	1	1	1
§2342	0	0	1	0	0	0	0	0	0
§2343	0	1	0	0	0	0	0	0	0
§2363	0	0	1	0	0	0	0	0	0
§2364	0	0	1	0	0	0	0	0	0
§2365	0	0	1	0	0	0	0	0	0
§2366	0	0	1	0	0	0	0	0	0

(c) Symmetric adjacency matrix



(d) Visualization: Legal text as a map of relationships

	§2326	§2325	§2201	§2343	§2366	§2365	§2364	§2363	§2342	Weight
§2326	1	1	0	0	1	1	1	1	1	31
§2325	1	0	1	1	0	0	0	0	0	55
§2201	0	1	0	0	0	0	0	0	0	61
§2343	0	1	0	0	0	0	0	0	0	61
§2366	1	0	0	0	0	0	0	0	0	69
§2365	1	0	0	0	0	0	0	0	0	69
§2364	1	0	0	0	0	0	0	0	0	69
§2363	1	0	0	0	0	0	0	0	0	69
§2342	1	0	0	0	0	0	0	0	0	69

Numerical example of the conformity weight calculation for §2326:
COUNT(§2326=1)+
COUNT(§2325=1)+
COUNT(§2201=0)+
COUNT(§2343=0)+
COUNT(§2366=1)+
COUNT(§2365=1)+
COUNT(§2364=1)+
COUNT(§2363=1)+
COUNT(§2342=1)=
=7+3+8+8+1+1+1+1+1=31

(e) Matrix *(c)* after seriation, sections ordered by their conformity

Figure 1: A walkthrough of the proposed methodology

We use legal text from U.S. code as a simple example (Fig.1) to establish common understanding of the methodology. The first step is to extract references from the initial unstructured legal text using regular expressions. This parsing and conversion process is not trivial due to variety of possible phrases for referencing, lack of textual referencing standards, references to multiple sections, ranges and comma separated nonconsecutive sections. References identified in the example (Fig.1.a) have been highlighted to illustrate the construction of adjacency matrices (Fig.1.b and Fig.1.c). Self-references were allowed, but references to subsections were merged with references to parent sections (underlined in the example). From adjacency matrix, we can draw a graph (Fig.1.d) to get the first insight of the complexity in the legislative act's reference system. Extra constraints should be set for minimizing the crossings of the edges - this allows the natural emergence of the inner structure and clusters. Such visualization method is informative and intuitive for manual inspection, but unsuitable for automatic numerical relevance calculations. A classical centrality measure, *betweenness*, from the area of social network analysis [2], should be used in this case to get a numerical importance estimations for sections comparable to manual judgements inferred from Fig.2. The final step is to apply matrix seriation method called conformity analysis [3] (see Fig.1.e for a numerical example of the calculation), which enables us to align the sections according to the typicality of their connections. For bigger datasets, in order to open the inner structure of references more efficiently, we will pursue the following additional procedure [3]. After calculation of conformity weights, we will find the least typical element, eliminate it and recalculate the weights. Eventually, after repeating the procedure until no elements are left, we will get the list of eliminations, which gives us better rank order (e.g. new order for rows and columns) than pure scale of conformity. It should be noted, that in our case, the scale is inverted, as we are dealing with a sparse structure, where the most typical section is a section with no references.

3. EXPERIMENTS AND CONCLUSION

The aim of the experiments is to investigate the applicability of proposed methods in real world scenarios. The Estonian Law of Obligations (*Võlaõigusseadus*) was selected for this research because of the large amount of sections (1075) and references between them (480). Due to the dimensions of the adjacency matrix (1075x1075), we plotted matrices as bitmap images (Fig.3), black dots represent references (positive values in the matrix). In the sense of a graph problem, we are dealing with a sparse matrix and a very specific structure. From the initial matrix, (Fig.3,left) we are able to identify a majority of connections near the diagonal, which can be explained as referencing mainly neighbouring sections and seldom to other regions. From practical point of view, such structure to be analyzed is very similar to knowledge discovery from retail transaction history, where assortment changes extremely frequently and purchases are not linked with loyalty cards. With both of those cases and structures, we are dealing with indirect relations and patterns, where classical data mining methods like clustering and association rules fail to give reasonable results. Seriation result is a chain of relations with underlying typicality scale and direction. It can also be used to reorder matrix elements (Fig.3,right), which results a great compression of the reference area, allowing the emergence of chains and clusters. Both of the visualization methods complement each other: if graph visualization (Fig.2) describes only the attraction of the elements, then seriation illustrates both - structure and the lack of structure, including the balance between elements attracting and repelling in the system. Such methods enable us to present the legislative act as a map of relationships with clusters, chains and hubs. Finally, as to the implementation status of our methodology, we have a running prototype. Implementation (PHP scripting language), along with the initial legislative act, transformed dataset and results are available electronically upon request for benchmarking and research purposes.

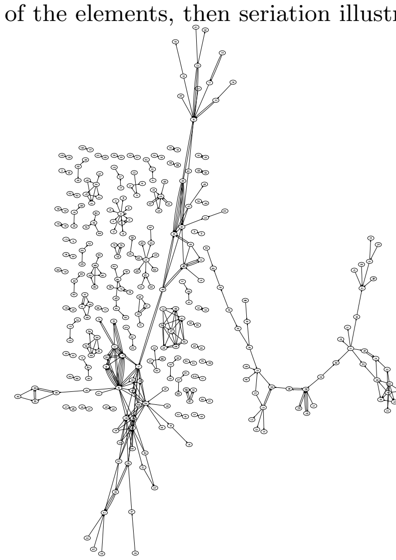


Figure 2: Chains of references and clusters

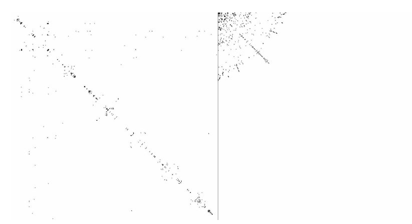


Figure 3: Initial and reordered matrix of references

4. ACKNOWLEDGMENTS

This work was supported in part by the Estonian Information Technology Foundation under Grant 06-03-99-08.

5. REFERENCES

[1] Palmirani, M., Brighi, R., and Massini, M. Automated extraction of normative references in legal texts. In *Proc. of ICAIL'03*, pp. 105-106, 2003.

[2] Freeman, L. C., Centrality in social networks: Conceptual clarification. *Social Networks*, Vol 1, pp. 215-239, 1979.

[3] Võhandu, L. Some Methods to Order Objects and Variables in Data Systems. *Trans. of Tallinn Technical University*, No. 482, pp. 43-50, 1980.