Transforming Legal Documents for Visualization and Analysis

Nuno Ramos Carvalho
United Nations University (UNU-EGOV)
Rua de Vila Flor 166,
4810-445 Guimarães,
Portugal
ramos.de.carvalho@unu.edu

Luís Soares Barbosa
University of Minho and
United Nations University (UNU-EGOV)
Rua de Vila Flor 166, 4810-445 Guimarães,
Portugal
barbosa@unu.edu

ABSTRACT[1]

Regulations, laws, norms, and other documents of legal nature are a relevant part of any governmental organisation. During digitisation and transformation stages towards a digital government model, information and communication technologies are explored to improve internal processes and working practices of government infrastructures. This paper introduces preliminary results on a research line devoted to developing visualisation techniques for enhancing the readability and comprehension of legal texts. The content of documents is conveyed to a well-defined model, which is enriched with semantic information extracted automatically. Then, a set of digital views are created for document exploration from both a structural and semantic point of view. Effective and easier to use digital interfaces can enable and promote citizens engagement in decision-making processes, provide information for the public, and also enhance the study and analysis of legal texts by lawmakers, legal practitioners, and assorted scholars.

CCS CONCEPTS

• Human-centered computing~Information visualization • Human-centered computing~Visualization toolkits

KEYWORDS

Document Visualization, Document Exploration, Legal Documents, Law, Digital Transformation, Digital Government

ACM Reference format:

N. Ramos Carvalho, L. Soares Barbosa. 2018. Transforming Legal Documents for Visualization and Analysis. In *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance, Galway, Ireland, April 2018 (ICEGOV'18)*, 4 pages.
DOI: 10.1145/3209415.3209424

1 INTRODUCTION

The rapid development of what is commonly called e-*participation*, or *e-democracy*, witnesses a massive potential of digital technologies, and their progressive integration with social dynamics and innovative policy initiatives, to promote sustainable community development and more inclusive societies. A crucial aspect is, of course, access to the relevant information, and, first of all, to normative information which establishes and rules the quotidian exercise of citizenship. Indeed, access to such information is fundamental to include citizens in deliberation and decision processes (e.g. through participatory surveys or on-line mass deliberation systems [10, 22]). Meeting increasing demand for transparency and open governments, norms and regulations should be easy to read and accessible to everyone, independently of educational background or accessibility constraints [23]. Moreover, with the increase of digitally available data in recent years, the challenge for lawmakers, civil servants, researchers, and the public to use it effectively has increased [14].

The overarching goal of this work is to build visual interfaces to explore legal documents in an efficient way. The generic approach for building the visual artefacts is divided into three major stages: (1) build a model of the document, including its content; (2) enrich the model using a set of natural language processing techniques; (3) build a visually rich representation of the data in the model that captures or emphasizes specific details[2]. During the implementation of the heterogeneous visualisations, Shneidermans' mantra for developing visualisation systems was followed: (i) overview first; (ii) zoom and filter second; and finally, (iii) details on demand [18]. While the two underlying goals of every artefact built are to be: (i) informative, i.e. provide the user with information to help him or her gain knowledge; and, (ii) efficient, i.e. provide the intended perspective or detail of interest as straightforward as possible without adding unnecessary complexity [20].

Related work includes: encouraging debate on public issues using computer-supported visualization of argument maps [17]; using argument visualization to enhance participation in the

ICEGOV '18, April 4–6, 2018, Galway, Ireland

ACM ISBN 978-1-4503-5421-9/18/04…$15.00
https://doi.org/10.1145/3209415.3209424

[2] This paper focuses only on the third stage.

legislation process [15]; United States Code representation and analysis using several hierarchical networks [3]; using diagrams to help describe laws to improve public access [2]; improve business contracts clarity and usability [8, 9]; using parallel word clouds to analyse trends in US court decisions [4]; more examples and approaches available in [5, 6].

The remainder of this paper is organised as follows: after the introduction, Section 2 illustrates and discusses some visualisation approaches. Section 3 concludes the paper, discusses the broader context for this research, and future plans for validating the illustrated approach.

2 VISUALISATION APPROACHES

This section describes some of the approaches currently under development for visualizing and exploring legal documents. Before building the visual artefacts, the document content is conveyed to a generic, well defined model, that captures the document structure and text. After the model is ready, combinations of Natural Language Processing (NLP) techniques (e.g., tokenization parsing, lemmatization, part-of-speech tagging [13]) are used to enrich the model with semantic information, adding, for example, lists of relevant concepts, term frequencies, links between concepts and document elements, etc. These resources are then used to create views that enable structural and semantic browsing. Some examples are described below. This workflow - building the views from the model - also contributes to having an approach as generic as possible, i.e. that works across a heterogeneous set of documents.

2.1 Structural Navigation

Legal documents can have a significant number of pages (e.g., the Portuguese Constitution has a total of 91 pages in PDF format) and different levels of hierarchy (e.g., parts, sections, sub-sections, chapters). Skimming through its content may be a challenge. Tree-maps are a visualisation approach for describing hierarchical structures while allowing for other dimensions to be conveyed using size or colour [12]. Using this approach, large documents can be illustrated in a single page, where each column is used to describe a level of the documents' hierarchy. Figure 1 illustrates a view of the Portuguese Constitution.

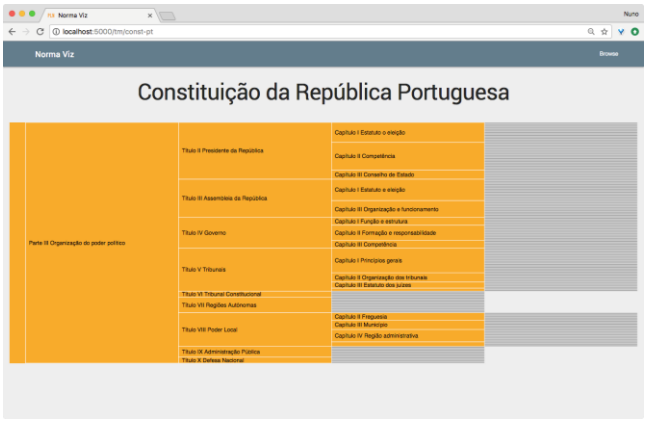

Figure 1: Screenshot of the Portuguese Constitution using a tree-map[3].

The user can quickly view the entire document, divided into sections, chapters, etc., and choose which part he or she wants to see in more detail. This approach is *zoomable*, i.e. you can click on any block to zoom in and make that section the root of the view, given that the articles of the document are always the leaves of the tree. Figure 2 illustrates the same document after selecting an arbitrary chapter.

The structural navigation allows starting with a broader view of the document and zooming in, i.e. get more detailed information, on some area of interest, based on the document structure. This view is particularly useful when the user is not able to come up with keywords, or search terms, to describe the information need.

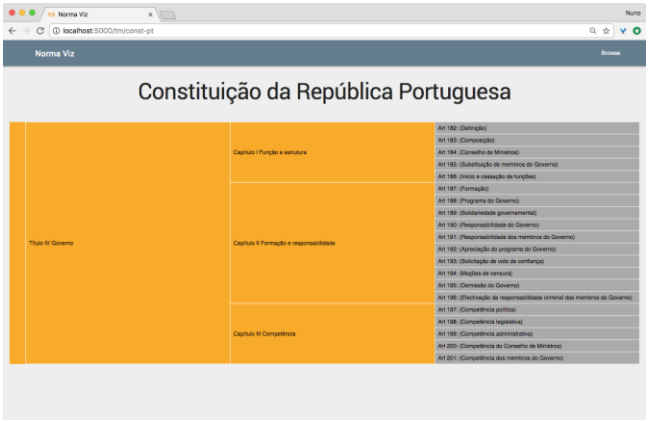

Figure 2: Chapter zoom view of the Portuguese Constitution using a tree-map.

2.2 Semantic Browsing

Complementary to exploring the document using its structural hierarchy, another useful approach is to browse the document

[3] View available from: http://norma-viz.egov.uminho.pt/tm/const-pt (last accessed 30-08-2018, site under development).

based on its semantic information. The semantic information is gathered using natural language processing techniques (mainly statistical ones) and can be materialised using a graphical representation of nodes and arcs – a semantic network (or graph) [7, 19].

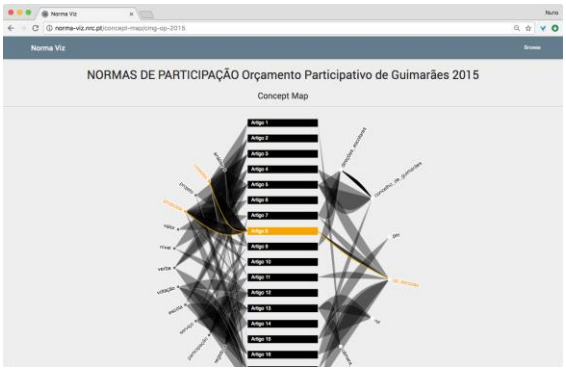

Figure 3: Screenshot of the concept map explorer for a participatory survey regulation[4].

Figure 3 illustrates a view of a document semantic network graph of a local city hall participatory survey regulation. The box nodes in the middle column represent concrete articles in the document, the non-box nodes on the left part of the visualization represent concepts that were found relevant in the text (e.g., citizens, proposals), and the non-box nodes on the right represent entities found in the document (e.g. city hall, school boards). The black and grey arcs between nodes represent a connection, usually because the corresponding concept or entity is addressed in the corresponding article. When a node is selected, as illustrated in Figure 3, the related concepts entities and articles connected to that node are highlighted. This gives an immediate insight on which concepts and entities are related with a specific article. A concept, or entity, can be selected instead, giving information about which articles are related to it.

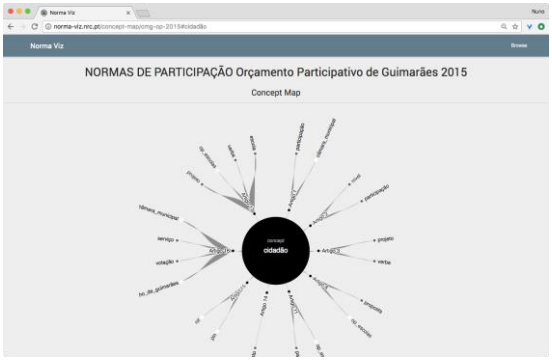

Figure 4: Screenshot of the concept map explorer for a participatory survey regulation centred on a concept.

Furthermore, each node is a link. Thus, when a link is followed, it is possible to see the selected concept, article, or entity as the centre of the network, as well as all the connected nodes that are related to the selected node, and other nodes connected by transitivity relations. This is illustrated in Figure 4. This view allows exploring the document semantically, i.e. navigating based on concepts, and then viewing the document parts related to the concepts of interest. This approach is interesting when searching for specific parts of the document that address some specific concept, as opposed to searching the entire document. The user can easily have details on which parts of the document may be more relevant to fulfil the information need.

2.3. Word Clouds

Word (or tag) clouds provide a visual representation of the vocabulary used in a given context. The most common approach is to vary the font size of each word according to the word frequency in the text (i.e., the more frequent a word is, the bigger the size of that word), colour and positioning can also be used to convey other dimensions of the data [1]. These are useful, for example, to quickly get an overview of the most addressed topics in a document, or collection of documents. The frequency of terms based on Term Frequency-Inverse Term Frequency (TF-IDF), and other related metrics, can also be explored to illustrate words and terms relevant in a given context or more prominent in a specific document belonging to a collection of documents [16]. This is useful to gain a quick insight into which topics are addressed in a document, and also which ones are more prominent.

Figure 5 illustrates a term frequency Word Cloud for the Portuguese Constitution. This immediately shows some relevant concepts in this document (e.g., citizens, president of the republic, law).

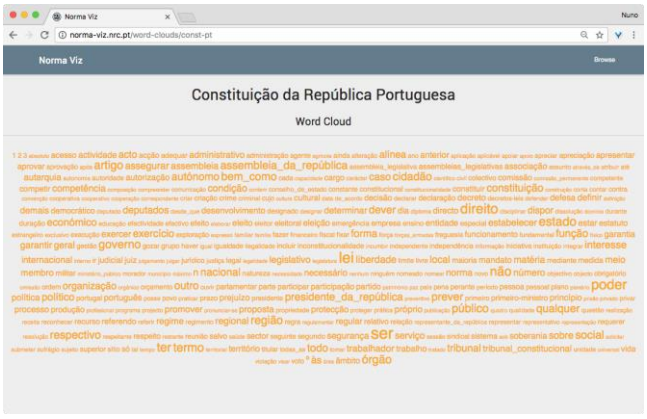

Figure 5: Screenshot of a term frequency word cloud for the Portuguese Constitution[5].

3 CONCLUSION AND FUTURE WORK

[4] View available from http://norma-viz.egov.uminho.pt/concept-map/cmg-op-2015 (last accessed 30-08-2017, site under development).

[5] View available from: http://norma-viz.egov.uminho.pt/word-clouds/const-pt (last accessed 30-08-2018, site under development).

This paper provides an overview of the ongoing research on the automatic derivation of visual interfaces from documents of legal nature. As a direct follow-up, future work will develop into three major areas: (1) improve current visualizations; (2) design new visual artefacts; (3) undertake qualitative studies to validate and measure the effectiveness of different approaches. Future work also includes validating the described approach to measure the benefits of its use. The plan is to build exercises that require specific information from legal documents, and measure metrics (e.g. time to complete, number of steps) while different groups of people attempt to solve them, some groups using the visualization approaches, and others using the more traditionally available document formats and search interfaces. The validation also includes having people answering a survey to have a clear idea of the benefits of this family of approaches.

This work, however, is part of a broader research agenda on techniques, methods and tools for modelling, analysing and validating legal text. Indeed, the recent, but extensive, body of knowledge in formal modelling and validation of software can be harnessed in such a direction, aiming at the systematic improvement of the quality of law. For example, going beyond the visualization dimension, as discussed here, natural language processing techniques [13] can be used to identify and retrieve legal patterns from the relevant textual sources. On the other hand, relational modelling tools, and in particular model-finders, such as Alloy [11], provide a platform for reasoning about (models of) legislation suitably extracted from textual representations, and prototyping alternative formulations.

Interestingly enough, the proper, pragmatic context for this work may be found in the efforts to reduce administrative burden [21], which is a major concern for governments in many countries. Typically, this encompasses initiatives around the re-engineering of administrative processes, provision of multi-service shops for citizens, risk-based approaches to regulation, and enforcement. The review of legislative stocks has been high on the agenda for a number of years now as well. For example, the European Commission established the Regulatory Fitness and Performance Programme (REFIT) under which 300 legislative proposals were withdrawn since 2006 and 53 in 2014 alone, and 350 assessments were carried out before proposing new legislation.

Our research programme, of which the structural visualization component discussed here is the first outcome, will contribute to reducing potential administrative burden acting from the outset on one of its main sources: the design of the legal and normative framework itself.

ACKNOWLEDGMENTS

This work is a result of the project "SmartEGOV: Harnessing EGOV for Smart Governance (Foundations, methods, Tools) / NORTE-01-0145-FEDER-000037", supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (EFDR).

We would like to thank the reviewers for their valuable insight and detailed comments, which aided in improving the paper.

REFERENCES

[1] Bateman, S. et al. 2008. Seeing things in the clouds: the effect of visual features on tag cloud selections. *ACM conference on Hypertext and hypermedia*. 4250, (2008), 193–202. DOI:https://doi.org/10.1145/1379092.1379130.

[2] Berman, D. 2000. *Toward a new format for Canadian legislation - Using graphic design principles and methods to improve public access to the law.*

[3] Bommarito, M.J. and Katz, D.M. 2010. A mathematical approach to the study of the United States Code. *Physica A: Statistical Mechanics and its Applications*. 389, 19 (2010), 4195–4200. DOI:https://doi.org/10.1016/j.physa.2010.05.057.

[4] Collins, C. et al. 2009. Parallel tag clouds to explore and analyze faceted text corpora. *VAST 09 - IEEE Symposium on Visual Analytics Science and Technology, Proceedings* (2009), 91–98.

[5] Curtotti, M. et al. 2013. Software tools for the visualization of definition networks in legal contracts. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law - ICAIL '13*. (2013), 192–196. DOI:https://doi.org/10.1145/2514601.2514625.

[6] Curtotti, M. and McCreath, E. 2012. Enhancing the Visualization of Law. *SSRN Electronic Journal*. (2012), 1–27. DOI:https://doi.org/10.2139/ssrn.2160614.

[7] Feldman, R. and Sanger, J. 2007. The text mining handbook: advanced approaches in analyzing unstructured data. *Imagine*. 34, (2007), 410. DOI:https://doi.org/10.1179/1465312512Z.00000000017.

[8] Haapio, H. 2013. Contract clarity and usability through visualization. *Knowledge Visualization Currents: From Text to Art to Culture*. 63–84.

[9] Haapio, H. 2011. Contract clarity through visualization - Preliminary observations and experiments. *Proceedings of the International Conference on Information Visualisation* (2011), 337–342.

[10] Iandoli, L. et al. 2009. Enabling On-Line Deliberation and Collective Decision-Making through Large-Scale Argumentation: A New Approach to the Design of an Internet-Based Mass Collaboration Platform. *International Journal of Decision Support System Technology*. 1, 1 (2009), 69–92. DOI:https://doi.org/10.4018/jdsst.2009010105.

[11] Jackson, D. 2012. *Software Abstractions: logic, language, and analysis*. MIT press.

[12] Johnson, B. and Shneiderman, B. 1991. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. *Proceeding Visualization '91*. (1991), 284–291. DOI:https://doi.org/10.1109/VISUAL.1991.175815.

[13] Jurafsky, D. and Martin, J.H. 2009. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. *Speech and Language Processing An Introduction to Natural Language Processing Computational Linguistics and Speech Recognition*. 21, (2009), 0–934. DOI:https://doi.org/10.1162/089120100750105975.

[14] Keim Daniel, K.J. and Mansmann, G. rey E. and F. 2010. *Mastering the Information Age Solving Problems with Visual Analytics*.

[15] Loukis, E. et al. 2009. Using argument visualization to enhance e-participation in the legislation formation process. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2009), 125–138.

[16] Manning, C.D. and Raghavan, P. 2009. An Introduction to Information Retrieval. *Online*. 1, (2009), 1. DOI:https://doi.org/10.1109/LPT.2009.2020494.

[17] Renton, A. and Macintosh, A. 2007. Computer-Supported Argument Maps as a Policy Memory. *Information Society*. 23, 2 (2007), 125–133. DOI:https://doi.org/10.1080/01972240701209300.

[18] Shneiderman, B. 1996. The eyes have it: a task by data type taxonomy for informatio nvisualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*. (1996), 336–343. DOI:https://doi.org/10.1109/VL.1996.545307.

[19] Sowa, J.F. 2006. Semantic networks. *Encyclopedia of Cognitive Science*. (2006), 1–32. DOI:https://doi.org/10.1017/S0140525X00028521.

[20] Steel, J. and Iliinsky, N. 2010. *Beautiful Visualization*.

[21] Veiga, L. et al. 2016. Digital Government and Administrative Burden Reduction. *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance*. (2016), 323–326. DOI:https://doi.org/10.1145/2910019.2910107.

[22] Velikanov, C. 2017. Can Deliberative Governance Become Inclusive? *Proceedings of the 18th Annual International Conference on Digital Government Research* (New York, NY, USA, 2017), 531–540.

[23] Zittel, T. et al. 2006. *Participatory democracy and political participation: can participatory engineering bring citizens back in?*. Routledge.