UvA-DARE (Digital Academic Repository)

Measuring Interestingness of Political Documents

Azarbonyad, H.

DOI

[10.1145/2911451.2911485](https://doi.org/10.1145/2911451.2911485)

Publication date

2016

Document Version

Final published version

Published in

SIGIR'16

[Link to publication](#)

Citation for published version (APA):

Azarbonyad, H. (2016). Measuring Interestingness of Political Documents. In *SIGIR'16: the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval: Pisa, Italy, July 17-21, 2016* (pp. 1175). Association for Computing Machinery. <https://doi.org/10.1145/2911451.2911485>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Measuring Interestingness of Political Documents

Hosein Azarbonyad
h.azarbonyad@uva.nl

University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

Political texts are pervasive on the Web covering laws and policies in national and supranational jurisdictions. Access to this data is crucial for government transparency and accountability to the population. The main aim of our research is developing a ranking method for political documents which captures the interesting content within political documents. Text interestingness is a measure of assessing the quality of documents from users' perspective which shows their willingness to read a document. Different approaches are proposed for measuring the interestingness of texts. In this research we focus on measuring political texts' interestingness. As political data sources, we use publicly available parliamentary proceedings.

Keywords

Text Interestingness; Topical Diversity; Political Documents

1. RESEARCH PROPOSAL

Political texts such as parliamentary proceedings are valuable information sources for historians, politicians, and also the public. Access to this data is crucial for government transparency and accountability to the population [2]. Therefore, there is a need to develop techniques to access and analyze the content of political data sources. In this research we focus on measuring the interestingness of political documents. Interestingness of documents could be used for ranking these documents and help the users to not only find the documents that are related to their information need but also focus on more interesting documents. The availability of user-generated text-based reviews stimulated research in automatically computing the interestingness of texts [3, 4]. Our main research question is: **RQ1:** *What aspects of interestingness are covered by current approaches? How can we incorporate the aspects related to documents content in text interestingness measures?* Our main goal is to design a measure for estimating the interestingness of debates. In [3] it is shown that text interestingness is highly correlated with topical diversity on e-books and e-commerce products description datasets. The current measures use either the topical diversity of documents to measure the interestingness of political

debates [1] or the structural information of debates [6]. Both of these approaches do not use the main content of the debates for assessing their interestingness. We are aiming to propose text-based measures of interestingness for political debates and also leverage useful structural information of debates as well as the topical diversity of them.

We break **RQ1.1** down into three research questions. We first study whether interestingness of political debates is correlated with their topical diversity and answer the following question: **RQ1.2:** *Are topically diverse political documents also interesting?* Usually topical diversity of documents is measured by means of topic models. The main intuition is that if a document covers many dissimilar topics it is a diverse document. We propose a parsimonization technique [5] to extract salient information from topic models which makes them more suitable for the task of estimating diversity of documents compared to current topic models. The next research question is: **RQ1.3:** *What are the main deficiencies of current topic models in estimating topical diversity of documents? And how effective is the proposed parsimonious topic model in addressing these deficiencies?* We use the proposed parsimonious topic models to measure the topical diversity of documents. Then using the estimated diversity values for the debates, we analyze the impact of different sources of diversity on making the documents topically diverse. Our next research question is: **RQ1.4:** *What are the main sources of diversity of debates? And how they affect the topical diversity of debates?* As the sources of diversity, we study the effect of diversity of people participated in debates and the diversity of main topics of debates. We incorporate these sources of diversity in our proposed interestingness and diversity measures.

Acknowledgements This research was supported by the Netherlands Organization for Scientific Research (ExPoSe project, NWO CI # 314.99.108; DiLiPaD project, NWO Digging into Data # 600.006.014) and by the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement ENVRI, number 283465.

REFERENCES

- [1] H. Azarbonyad, F. Saan, M. Dehghani, M. Marx, and J. Kamps. Are topically diverse documents also interesting? CLEF'15, 2015.
- [2] M. Dehghani, H. Azarbonyad, M. Marx, and J. Kamps. Sources of evidence for automatic indexing of political texts. ECIR'15, 2015.
- [3] M. Derzinski and K. Rohanimanesh. An information theoretic approach to quantifying text interestingness. In NIPS MLNLP, 2014.
- [4] D. Ganguly, J. Leveling, and G. J. F. Jones. Automatic prediction of aesthetics and interestingness of text passages. COLING'14, 2014.
- [5] D. Hiemstra, S. Robertson, and H. Zaragoza. Parsimonious language models for information retrieval. SIGIR'04, pages 178–185, 2004.
- [6] A. Hogenboom, M. Jongmans, and F. Frasinca. Structuring political documents for importance ranking. NLDB'12, 2012.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGIR '16 July 17-21, 2016, Pisa, Italy

© 2016 Copyright held by the owner/author(s).

ACM ISBN 123-4567-24-567/08/06...\$15.00

DOI: <http://dx.doi.org/10.1145/2911451.2911485>