

Harnessing Spontaneous Participation on Social Media – Implementing the Knowledge Extraction Component

Lukasz Porwol

Islam Hassan

Adegboyega Ojo

John Breslin

Insight Centre for Data Analytics
National University of Ireland, Galway
name.surname@insight-centre.org

ABSTRACT

Recent efforts to mainstream social media-based and citizen-led political deliberations to complement traditional government-led e-Participation, must among others address a number of technical challenges. One of these challenges is how to automatically filter and extract meaningful information or knowledge from streams of social media data contributed by citizens to inform agenda setting or serve as feedback for specific policy issues. A central resource in developing an effective information extraction system is the Lexicon of terms and relationship among terms in the domain of interest. This poster summarizes our work in developing a lexicon of public services names as the kernel of the knowledge extraction component of our Social Software Infrastructure for spontaneous participation. Our lexicon-based knowledge extraction process is scalable and could be easily extended to capture other information of interest from social media or related platforms.

Categories and Subject Descriptors

H.4.2 [Types of Systems]: e-Participation System

General Terms

Management, Measurement, Design, Standardization, Theory

Keywords

e-Participation, Information Extraction, Social Media

1. INTRODUCTION

e-Participation employs technology-mediated dialogue between citizens and the politics sphere and between citizens and administration [10] for effective, public participation and feedback [1], while also introducing new ways of political participation [2]. A common perception about e-Participation initiatives is that they have been less than successful [11], in part due to the abysmally low participation of citizens. Macintosh et al. in [5] has argued for the need to mainstream social-media-based citizen-led political deliberations to complement traditional e-Participation. We believe that harnessing relevant contents on the social web is contingent on providing information extraction solutions to capture citizen discussions related to public services or policies of interest from social media streams. While there are general purpose information extraction solutions such as Named Entity Recognisers (NER) for names of persons, locations or organizations [6] such solutions for domain specific entities (e.g.

policy domain or public service names are less available. NER solutions are often based on language resources such as lexicons (e.g. WordNet), which define entities of certain types that are of interest and possibly relationships among these entities. This work describes our efforts at developing a technology solution for identifying contents (or discussions) that are related to public services on social media platforms. This solution is part of a holistic Social Software Infrastructure (SSI) for e-Participation [8]

2. MOTIVATION

General purpose entity extractors such as AlchemyAPI, DBpedia Spotlight, Extractiv, OpenCalais and Zemanta [9] are insufficient for identifying domain-specific entities or names from contents on social media. Identifying different types of entities that are contained in contents generated by citizens on social media require the development of specific names entity recognition and classification solutions. Thus, identifying major issues of interest around public services from social media discussion requires the development of a dedicated NER component for Public Services. We describe below our approach to constructing the public service name recognizer.

3. APPROACH

Our objective is to develop a component for extracting information from social media as part of a Social Software Infrastructure for e-Participation. To achieve this goal, we construct two technical artefacts – 1) a Lexicon of public service names and 2) a Named Entity Recogniser based on the lexicon, following the Design Science Research tradition [4] and [7]. Design science creates and evaluates artifacts that define ideas, practices, technical capabilities and products through which the analysis, design, implementation and use of information systems can be effectively accomplished. The development of the lexicon is based on the national public service catalogues. The two datasets were employed as input into a process, which automatically related public service names, based on a set of semantic similarity and relatedness measures, including Explicit Semantics Analysis (ESA)[3] and WordNet-based measures. The resulting graph of Public Service Names is subsequently employed to develop a NER or spotter using an open source dictionary based spotter framework.

4. INFRASTRUCTURE

The process for constructing the lexicon is shown in Fig.1. The lexicon is the core element of the Knowledge Extraction & Management SSI (Fig.2) component. Two major datasets were employed as input resource to the development of the lexical resource – United Kingdom and Irish Government Public Service Catalogues. Additional catalogues are planned to be added to the two initial catalogues over time to improve diversity and consequently the recall properties of the lexicon. The language resource created enables population of the public service ontology

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.

dg.o 2015, May 27-30, 2015, Phoenix, AZ, USA
ACM 978-1-4503-3600-0/15/05.
<http://dx.doi.org/10.1145/2757401.2757460>

using the information extracted from governmental public service documents.

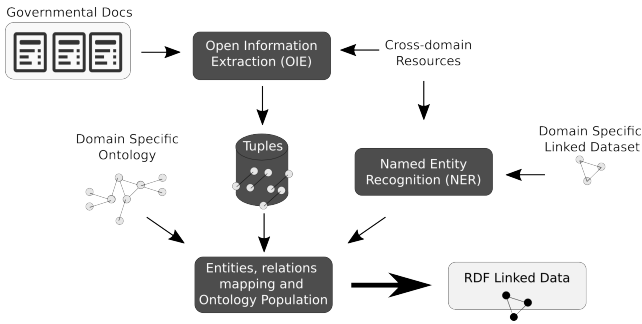


Fig.1: Lexicon construction

The populated Public Service Ontology, represented in RDF/XML - Linked Data publishable format, (Semantic Web knowledge graph representation standard) informs the recognition task to be performed by the NER solution.

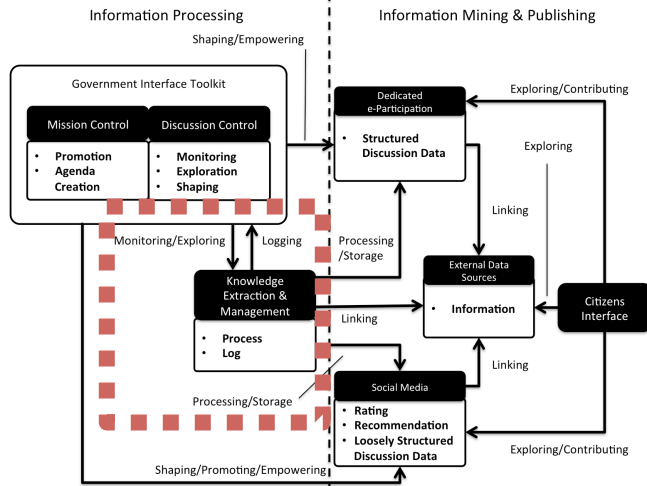


Fig. 2: SSI

5. RESULTS

Results show that the NER solution constructed based on the lexicon is effective in recognizing public service related contents in Social Media streams. As shown in the Table 1, while our tool successfully matched public services names, generic and mainstream NER tools like DBpedia Spotlight and Alchemy API identified more general concepts like locations and names of organizations. However, we exploit the complementarity of general purpose NER tools to provide better context and entity linking opportunities for public service names identified.

6. CONCLUSIONS

Motivated by the need to support the duality of e-Participation, we have presented an infrastructure for extracting information from spontaneous discussions on Social Media as part of the comprehensive SSI design implementation, advancing the existing e-Participation methodologies. Results show immediate opportunities for developing and consolidating dedicated lexical resources supported by Semantic Web and NLP technologies into a Knowledge Extraction & Management component for application to the context of e-Participation. While we have demonstrated the usefulness of the software infrastructure,

detailed formal evaluations of the NER system are still on-going in terms of recall and precision accuracy over large datasets of contributions on Social Media platforms such as Twitter and Facebook.

Table 1: NER performance comparison

Tweet	LEX NER	DBpedia	Alchemy
Ireland issuing passport cards for EU travel, like US/Canada. Can take photo from smartphone	Passport issuing: https://egov.deri.ie/PublicServices/Service/UK_PASSPORT_ISSUING Passport: https://egov.deri.ie/PublicServices/Service/CA_goc_passport	passport cards: http://dbpedia.org/resource/Passport_card EU: http://dbpedia.org/resource/European_Union Canda: http://dbpedia.org/resource/Canada smartphone: http://dbpedia.org/resource/Smartphone	US/Canada : Country EU: Organization Ireland: Country
Why Rwanda Plans to Issue Biometric Passport - Rwanda will start issuing biometric passports to her citizens which	Passport: https://egov.deri.ie/PublicServices/Service/CA_goc_passport	Rwanda: http://dbpedia.org/resource/Rwanda biometric passports: http://dbpedia.org/resource/Biometric_passport	Rwanda: Country
@SenatorLesniak No hunting licence for ANYONE that hasn't completed a gun safety course.	https://twitter.com/christwords199/status/553672403046240256	hunting licence: https://egov.deri.ie/PublicServices/Service/UK_HUNTING_LICENCE	gun safety: http://dbpedia.org/resource/Gun_safety

Nevertheless, current results are promising and provide good arguments for the feasibility of constructing effective Social infrastructure for Social Media based e-Participation.

7. ACKNOWLEDGEMENT

This work has been funded by EU Commission - Grant 645860

8. REFERENCES

- [1] Chadwick, A. 2003. Bringing E-Democracy Back In: Why it Matters for Future Research on E-Governance. *Social Science Computer Review*. 21, 4 (Nov. 2003), 443–455.
- [2] Dijk, J.A.G.M. Van 2000. Models of Democracy and Concepts of Communication. (2000).
- [3] Gabrilovich, E. and Markovitch, S. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. *IJCAI International Joint Conference on Artificial Intelligence*. (2007), 1606–1611.
- [4] Hevner, A. and Chatterjee, S. 2010. Design Research in Information Systems. *Integrated Series in Information Systems*. Springer US. 9–23.
- [5] Macintosh, A. et al. 2009. eParticipation : The Research Gaps. (2009), 1–11.
- [6] Nadeau, D. and Sekine, S. 2007. A survey of named entity recognition and classification. *Linguisticae Investigationes*. 30, 1 (2007), 3–26.
- [7] Peffers, K. et al. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*. 24, 3 (Dec. 2007), 45–77.
- [8] Porwol, L. et al. 2013. Harnessing the duality of e-participation. *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance - ICEGOV '13*. (2013), 289–298.
- [9] Rizzo, G. and Troncy, R. 2011. Nerd: Evaluating named entity recognition tools in the web of data. *Workshop on Web Scale Knowledge Extraction (WEKEX'11)*. (2011).
- [10] Sabo, O. et al. 2008. The shape of eParticipation: Characterizing an emerging research area. *Government Information Quarterly*. 25, 3 (Jul. 2008), 400–428.
- [11] Sæbø, Ø. et al. 2011. Understanding the dynamics in e-Participation initiatives: Looking through the genre and stakeholder lenses. *Government Information Quarterly*. 28, 3 (Jul. 2011), 416–425.