

A Knowledge Extraction and Management Component to Support Spontaneous Participation

Lukasz Porwol^(✉), Islam Hassan, Adegbojega Ojo, and John Breslin

Insight Centre for Data Analytics, National University of Ireland,
Galway, Ireland

{lukasz.porwol, islam.hassan, adegboyega.ojo,
john.breslin}@insight-centre.org

Abstract. Harnessing spontaneous contributions of citizens on Social Media and networking sites is a major feature of the next generation citizen-led e-Participation paradigm. However, extracting information of interest from Social Media streams is a challenging task and requires support from domain specific language resources such as lexica. This work describes our efforts at developing a Knowledge Extraction and Management component which employs a lexicon for extracting information related to public services in Social Media contents or streams as part of a holistic technology infrastructure for citizen-led e-Participation. Our approach consists of three basic steps – (1) acquisition and refinement of public service catalogues, (2) organization of the public service names into a lexicon based on different semantic similarity measures and (3) development of a dictionary-based Named Entity Recognizer (NER) or “spotter” based on the lexicon. We evaluate the performance of the NER solution supported by contextual information generated by two well-known general-purpose information NER tools (DBpedia Spotlight and Alchemy) on a dataset of tweets. Results show that our strategy to domain specific information extraction from Social Media is effective. We conclude with a scenario on how our approach could be scaled-up to extract other types of information from citizen discussions on Social Media.

Keywords: e-Participation · Citizen-led e-Participation · Information extraction (IE) · Natural Language Processing (NLP) · Public services · e-Government

1 Introduction

e-Participation involves technology-mediated interaction between citizens and the politics sphere [1]. By leveraging information and telecommunication technology (ICT), in particular contemporary social software technologies, e-Participation facilitates ubiquitous public participation and instant feedback capabilities [2]. Certain contemporary e-Participation solutions attempted to augment the discussion platforms with Social Media as additional communication channel though with very limited success. This is due to the fact that the traditional conceptualizations of e-Participation as a consultative, democratic process with involvement of citizens in policymaking do

not sufficiently address the common spontaneous citizen participation on informal channels such as Social Media [3]. Our experience so far is that state-of-the-art e-Participation solutions integrated with Social Media directly or in particular, leveraging generic, of-the-shelf analytical solutions to harness the vast information on Social Media mainly through general-purpose NER tools, largely fail to achieve the desired performance level. This we argue is due to lack of specific mechanisms to deal with information overload and lack of domain-specific natural language preprocessing tools to complement the popular general-purpose ones which offer relatively narrow range of general concepts like name of a person, organization, location, brand, product, a numeric expression including time, date, money and percent [4]. In particular, state-of-the-art solutions fall short in addressing e-Participation specific concepts such as public service names or policies. Therefore, progress in this area is contingent on developing domain-specific tools for processing political discussion by citizens and e-participation contents in general. Building on our previous work [5], we show how the Knowledge Extraction & Management component of a holistic infrastructure for e-Participation could be implemented using domain-specific lexical resource. We demonstrate and discuss the use of the component through a use case scenario. Thus, by our solution we exemplify the creation of explicit technological bridge between the citizen-political-discussion-sphere (Social Media) and public services sphere. The paper is structured as follows: In Sect. 2 we elaborate on related work in the context of use of social software, in particular Social Media for e-Participation; Sect. 3 presents important concepts essential to understand the of Knowledge Extraction and Management Component development process; In Sect. 3 we discuss the approach; Sect. 4 elaborates on the Knowledge Extraction and Management component creation and presents an example use case scenario. In Sect. 5 we discuss our contribution to the advancement of e-Participation domain with final conclusions presented in Sect. 6.

2 Related Work

2.1 Web 2.0 and Social Media in e-Participation

The last decade witnessed many examples of the use of social software as an infrastructure for realizing certain aspects of e-Participation. Social software is usually referred to as: Web 2.0 Software (or platform) that enables social networking by offering capabilities for people to contact and interact with each other [6]. The main principle of Web 2.0 is collective intelligence, collaborative content creation and composition by the user (here citizen) who contributes towards common knowledge [7]. Many e-Participation projects including HUWY,¹ U@MARENOSTRUM², VID³,

¹ <http://www.huwy.eu/vi>.

² <http://www.umatmarenostrum.eu/>.

³ <http://www.vidi-project.eu/>.

WAVE⁴, VOICES⁵, WEGOV⁶, Puzzled by Policy⁷, PADGETS⁸, SPACES⁹ employed Web 2.0 tools such as digital forums, blogs, wiki's and live-chat to provide dedicated e-Participation environment where citizens can express and discuss their needs, concerns and ideas. Those highly structured platforms, though supposedly well tuned to specific e-Participation needs, in principle suffer from abysmally low participation of citizens. In contrast, very specific, incredibly popular sub-group of social software tools: Social Media are widely used by citizens for spontaneous political discussions though without direct link to the formal e-Participation. This phenomena is referred to in the literature as Duality of e-Participation [8]. Therefore, in response to challenges faced by the dedicated e-Participation platforms some of the solutions indeed, introduced explicit support for the popular Social Media platforms with particular feed integration (in rare cases both ways content exchange is available) [9, 10]. Some more advanced solutions such as presented by PADGETS [11] with injection of special widgets into Social Media enable direct back-loop feed to the dedicated platform. However the big challenge remains unsolved as the prominent e-Participation solutions integrating Social Media largely do not address the issue of volume nor quality (lack of relevant selectivity) of the content produced [12], therefore do not ensure sufficient innovation to enable the dual e-Participation observed by Macintosh [8]. We are aware of certain original attempts to leverage the potential of spontaneous discussions on Social Media, such as the innovative approach presented in WEGOV project [13]. Nevertheless the methodology applied in the project appears to relay on relatively generic Social Media analytics tools (for topic detection, topic popularity, sentiment analysis and seed user detection) without explicit, direct links to the government sphere including for instance: references to governmental services, policy documents or newsletters. Moreover the methodology does not seem to give much of explicit thought to the essential synergy between current government-led solutions and processes, and citizen-led participation. Therefore the solution offered by the project, though advanced, appears to repeat the principles of the already available of-the-shelf, popular Social Media analytics solutions for businesses.

Considering Social Media and politics it is important to recall past miscellaneous attempts to harness Social Media for e-Participation, beyond e-Participation research projects. In particular, as it has been shown that successful Social Media campaign can influence political popularity (hence can have a significant impact on results of elections), many decision makers and government offices employed Social Media as a direct campaign communication channel [14, 15]. Another important e-Participation Social-Media use context has been: improved, Social-Media-supported Disaster and Crisis Management and Policy Development derived from Social-Media-facilitated citizen reporting capabilities [16, 17]. In particular Social Media have been playing

⁴ <http://www.wave-project.eu/>.

⁵ <http://www.give-your-voice.eu/>.

⁶ <http://www.wegov-project.eu/>.

⁷ <http://join.puzzledbypolicy.eu/>.

⁸ <http://www.padgets.eu/>.

⁹ <http://www.positivespaces.eu/>.

increasing role as rapid crowdsourcing and rapid response tools, especially in the events of crisis (including political crisis) [18] and natural disaster [19]. However, the Social-Media-applications for e-Participation, in the cases mentioned, focus rather on the use of popular Social Media platforms directly or use common, of-the-shelf (not a domain specific) analytical methodologies and solutions to harness the spontaneous political discussions what results in moderate performance. Therefore, a solution that would try to comprehensively address specific analytical needs of the e-Participation context, such as: effective methods for identifying political content on Social Media or contextual information clustering and linking is yet to be developed.

2.2 Semantic Web

The Semantic Web (Web 3.0) provides a framework that allows data to be shared and reused across applications, enterprises, and community boundaries [20] advancing the website-based Web 2.0 to the Web of Data. Semantic Web leverages ontologies for information modelling and knowledge representation. Ontologies provide a controlled vocabulary of terms that can collectively provide an abstract view of the domain [21]. Semantic Web technologies and ontologies are being used to address data discovery, data interoperability, knowledge sharing and collaboration problems. Ontologies can be described in RDF (Resource Description Framework) [22] which provides a flexible graph based model, used to describe and relate resources. The application of Semantic Web technologies to e-Government gained significant momentum with applications to several major areas including the use of ontologies to formally model different aspects of e-government; Structuring e-Participation research [23], Enabling personalized service delivery [24]; Enabling interoperability and integration of government resources and services [25].

In our work we leverage Semantic Web for constructing artefacts essential for Natural Language Processing and for storing all the data in a graph, in order to explicitly support better data discovery and interoperability for next generation e-Participation.

2.3 Natural Language Processing

2.3.1 Information Extraction

The goal of Information Extraction (IE) is to derive information structures directly from text with emphasis on the following aspects: identifying relations from textual content [26], automatic instantiation of ontologies and building knowledge bases tools [27]. Common methods on IE have focused on the use of supervised learning – SL techniques [28], self-supervised methods [29] and rule learning [30]. These techniques learn a language model or a set of rules from a set of manually tagged training documents and then apply the model or rules to new texts. The challenge for the SL approaches is the high cost of creating the labelled resources. In contrast, the unsupervised learning (UL) methods (also referred to as Open Information Extraction) attempt to fetch information automatically from the texts themselves [31].

2.3.2 Named Entity Recognition

A named entity can be defined as an entity classified accordingly to predefined set of categories for instance: person, organization, location, brand, product, time, date etc. [4]. The Named Entity Recognition applies multiple “classic” information extraction techniques listed before: SL, SSL and UL. However certain contemporary NER solutions apply lexical resources (e.g. WordNet¹⁰), lexical patterns and statistics computed on large annotated corpus [32]. The common processing pipeline for NER includes detecting named entities, assigning a type weighted by a numeric confidence score and by providing a list of URIs for disambiguation. The lexical resources and terminological databases are essential part of modern NLP systems consisting of large amount of highly detailed and curated entries [33]. A Lexicon can be developed to be domain independent or to support a specific domain. Thus, in our work we focus on the lexicon-based domain-specific approach to NER.

3 Approach

Design science creates and evaluates artefacts that define ideas, practices, technical capabilities and products through which the analysis, design, implementation and use of information systems can be effectively accomplished. Given that the goal of this work is to construct a technical artefact, our research follows the Design Science Research guidelines and process elaborated in [34, 35]. In particular, our objective is to develop a Knowledge Extraction and Management Component (KEMC) as part of a comprehensive 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 and integrating generic NER solutions through dedicated APIs. 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) [36] 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. In line with the DSR process model described in [35].

4 Knowledge Extraction and Management Component (KEMC)

In this section we elaborate on Knowledge Extraction and Management Component implementation encapsulating two core building blocks: the public service lexicon and the NER solution leveraging the generated language resource. First we present the comprehensive infrastructure for e-Participation design to provide the context for KEMC development. Then we present the domain-specific lexicon creation process

¹⁰ <https://wordnet.princeton.edu/>.

algorithm followed by application of the language resource to dedicated NER solutions combining the output of the generic NER solutions.

4.1 e-Participation Infrastructure Design

The infrastructure for e-Participation [5] presented in Fig. 1, is a comprehensive design resulting from extensive works on identifying and consolidating the duality of e-Participation requirements. The black and white components represent the tool-containers while arrows represent the interfaces. The central component – Knowledge Extraction and Management (marked with the red dashed line) is primarily responsible for all Social Media participation data retrieval and processing. The component extracts and analyses the discussion data, i.e., posts, user profiles, discussion topics, threads and performs continuous data quality improvement by filtering and linking related concepts as well as linking data from external sources such as other e-Participation systems, governmental portals or any other places holding valuable e-Participation information. The work presented in this paper attempts to provide an approach and implementation for realising this core e-Participation infrastructure component.

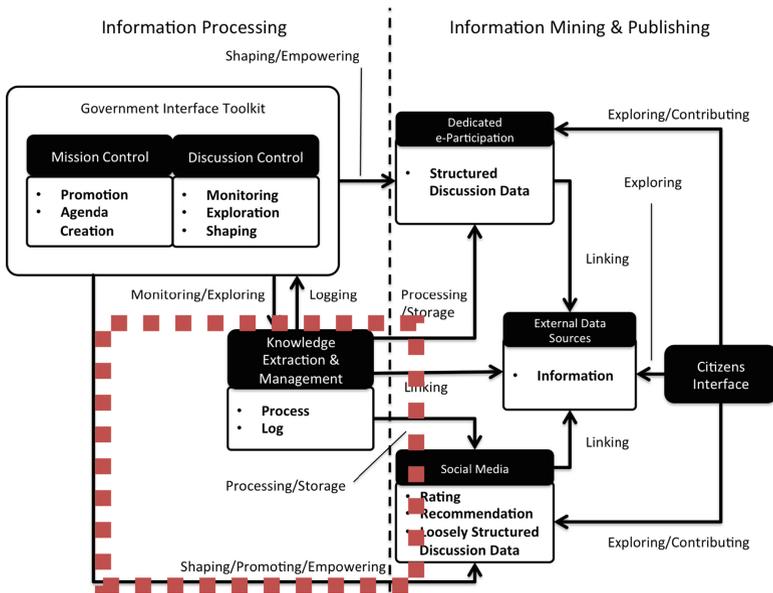


Fig. 1. Infrastructure for e-Participation design

4.2 Lexicon

In order to construct the lexicon we drew from the existing catalogue of services and the Core Public Service Vocabulary¹¹ – a simplified reusable and extensible data model

¹¹ https://joinup.ec.europa.eu/asset/core_public_service/description.

that captures the fundamental characteristics of a service offered by public administrations. We leveraged the Core Public Service Ontology as the basis for creating the initial language resource. Two major datasets were employed as input resource to the development of the lexical resource – United Kingdom and Irish Government Public Service Catalogues. Given that the recall potential of the resource is directly linked to the diversity of its entries, continued update to the lexical resource based on other government public service catalogues is important. Using this language resource in the information extraction task is aimed at enabling the automatic population of the public service ontology using the information extracted from governmental public service documents (different from public service catalogues) available in the form of standard PDF documents or as web documents. We describe the lexicon construction algorithm in Fig. 2. In summary, the construction starts with acquisition of the first list of the Public Services Names and unique identifiers in a form of URLs from the first catalogue. Then the following public service catalogues are aligned to the existing index based on similarity measures. The resulting structure is a simple tree in which parent nodes are the automatically generated generic service name and child or leave nodes are semantically related concrete service names obtained from the different public service catalogues. The alignment procedure is repeated for consequent service catalogues. The generated abstract service names comprise keywords derived from names of the child nodes.

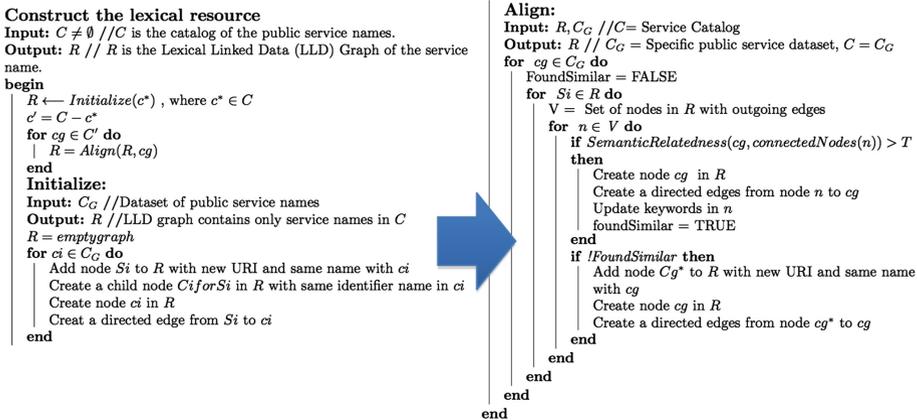


Fig. 2. Lexicon construction algorithm

4.3 Named Entity Recogniser

In this section we briefly explain our NER solution. The JAVA-based basic tool developed incorporates the implementation of the “Public Services Spotter” for Social Media with the domain-specific lexicon (represented in RDF) for Public Services Names as the input dictionary. As shown in the results from initial validation (or evaluation) of the tool in “spotting” public services related information (see example Table 1), the dedicated solution outperforms the generic NER tools (Dbpedia

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_ISSUI NG Passport: https://egov.deri.ie/PublicServices/Service/CA_goc_passport	Passport cards: http://dbpedia.org/resource/Passport_card EU: 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

Spotlight¹² and Alchemy API¹³) which hardly identified any public services. However, off-the-shelf solutions supply rich, contextual information for instance: location, organisation, persons (and more). The tools are also capable of providing statistics-based data, like “hot” topics and keywords. We argue that the final output of KEMC as a combination of Public Services names linked with miscellaneous information in a single semantic graph can deliver a very powerful tool for political Social Media data analytics. In particular, this enables creation of rich data stories with detailed information on where particular Public Services are being discussed and which certain political figures or organisations are usually mentioned around the topics in specific locations. Moreover certain main topics and keywords can be supplied together with sentiment analysis and popularity statistics. To visualise better the range of possibilities we consider a use case scenario in next section.

4.4 Scenario

The goal of this section is to demonstrate the use of KEMC solution through a use case scenario: John Smith (hypothetical character) is an Irish politician promoting

¹² <https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki>.

¹³ <http://www.alchemyapi.com/>.

legislation introducing restrictions on medical card applicants' eligibility. John opens the e-Participation Analyser supported by KEMC. Based on his specific request, the interface generates a dynamic report of places in Ireland from which it appears that citizens express negative sentiment towards the healthcare services. From the information mined from Twitter (public service tweets detected by our NER solution) it is apparent that Galway City (location detected) has the highest rate of negative opinions (sentiment analysis) oscillating around the institution of University Hospital – UH and Merlin Park Hospital – MPH (organisation entities detected). Moreover common topics found are (through topic analysis): prenatal care, physiotherapy and medical card. John tries to identify the key arguments against his policy project therefore he explores the posts and discussions of highest popularity rank with negative sentiment associated with the medical card and public healthcare. After following selected discussions (represented in semantic web format - every posts and discussion is distinguished by unique URL) he realises that the negative opinions come mainly from UH and MPH not accepting the medical card for particular services (prenatal care and physiotherapy) therefore he engages into discussion with citizens on Twitter and explains that the issues mentioned by citizens are of local character (but will be addressed) and ensures citizens that the upcoming legislation will not bring any harm but rather improve the current set of services covered. Moreover now, since John knows that the “hot” topics detected around Public Healthcare Services in Ireland are closely related medical card (similarity measures and graph distance), he suggests relevant common strategy that should be developed in order to facilitate a solution for these problems. The use case scenario presented will be leveraged for real-world experimentation for KEMC deployment. We believe that the direct implication of the use of KEMC will be to enable government to interact with citizen-spaces on Social Media in a more selective, topic-relevant, efficient way and long-term, can contribute significantly to enhancing the delivery of the public services as a result of better understanding of citizen's needs and concerns; hence directly supporting the duality of e-Participation.

5 Discussion

In this paper we have briefly introduced the state-of-the-art use of social software (in particular Social Media) for e-Participation, focusing on the main trends. The last decade has seen a shift in e-Participation from simple Web 2.0 forums to more advanced platforms integrating Social Media like Facebook or Twitter. Social Media are by many folds more widely used by citizens than any e-Participation solution. Moreover, many people incorporated them into everyday activities as they are very easy to use [37] and indeed they became a spontaneous place for, every-day political discussions. Nevertheless, we argue that to date, the current e-Participation solutions do not unleash the full potential of Social Media analytics in the context of e-Government, essential to deal with significant channel-specific obstacles like information overload and varying content quality. The main challenge, which is the lack of dedicated analytical tools for e-Participation, renders most of the contemporary e-Participation solutions' performance insufficient to fully support the duality of e-Participation. Results from our work, in creating Knowledge Extraction and Management component

as part of a holistic infrastructure for e-Participation, in a form a domain-specific NER solution powered by a Public Service Lexicon, provides first significant step towards building explicit connection between the sphere of government and citizen spontaneous discussions on Social Media, thus delivering the base to support duality of e-Participation. We show that dedicated solution outperforms the generic, of-the-shelf analytical solutions; therefore further development of custom solutions (that can follow our universal methodology) is a viable option for the advancement of e-Participation domain. We claim better universality and scalability of the automatically generated lexical resources (based on similarity measures) in comparison to Supervised-Learning-based solutions, which demand significant manual efforts on creating relevant resources. Moreover we claim better alignment of the resource delivered to specific e-Participation need. Nevertheless, we emphasise on the limitation of the resource, created explicitly for Irish and UK context. We demonstrate, through a specific use case scenario that a combination of the dedicated NER, supplied with rich contextual information provided by generic NER solutions can be a very powerful tool for e-Participation information analytics. The example use of KEMC for identifying Social Media discussions and posts related to concrete public services opens up a possibility for completely new set of capabilities for public services' citizen-perception evaluation, hence supporting future improvement and public service integration. Apart from the extensive works carried under WeGov project [13], we are not aware of any other significant attempt at applying advanced Social Media analytics to e-Participation. We acknowledge innovative work by Hagen et al. [38] on leveraging NLP technologies to analyse e-Petitioning content. However, we haven't found any approach that would try to combine and apply dedicated, scalable NER, supported by rich, automatically annotated domain-specific lexical resources, Semantic Web and Natural Language Processing technologies to address the duality of e-Participation.

6 Conclusion

Motivated by the need to provide the necessary step towards supporting the duality of e-Participation, we have presented a technical component – KEMC for extracting, consolidating and enriching (by linking) information from spontaneous discussions on Social Media as part of the comprehensive infrastructure for e-Participation design implementation, advancing the existing e-Participation methods and tools. Results from our work show immediate opportunities for developing and consolidating the domain-specific lexical resources, Semantic Web and NLP technologies into an analytical infrastructure for application to the context of e-Participation. While we have demonstrated theoretically the usefulness of the KEMC component, on the example of public healthcare service information matching supported by miscellaneous contextual information, more detailed and formal evaluations in different contexts are yet to be conducted. Next steps for the research include the implementation of special, information-visualisation-rich dashboards building upon developed analytical component to explain better the capabilities of the comprehensive infrastructure for e-Participation. This will be followed by a set of interviews with politicians and citizens in order to determine the usefulness of presented solution in reaching out the

information about public services and potentially supporting the process of public services improvement. Future steps should also bring series of design and development iterations of the solution with applications of the component to other contexts of e-Participation to advance the domain of Social Media analytics for e-Government.

Acknowledgement. This work has been funded by EU Commission - Grant 645860 (ROUTE-TO-PA).

References

1. Sæbø, O., Rose, J., Skiftenesflak, L.: The shape of eParticipation: characterizing an emerging research area. *Gov. Inf. Q.* **25**, 400–428 (2008)
2. Chadwick, A.: Bringing e-democracy back in: why it matters for future research on e-governance. *Soc. Sci. Comput. Rev.* **21**, 443–455 (2003)
3. Porwol, L., Ojo, A., Breslin, J.: On the duality of e-participation – towards a foundation for citizen-led participation. In: Kö, A., Leitner, C., Leitold, H., Prosser, A. (eds.) EDEM 2013 and EGOVIS 2013. LNCS, vol. 8061, pp. 211–225. Springer, Heidelberg (2013)
4. Grishman, R.: Message understanding conference-6: a brief history. In: Proceedings of COLING 1996 (1996)
5. Porwol, L., Ojo, A., Breslin, J.: Harnessing the duality of e-participation. In: Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance - ICEGOV 2013, pp. 289–298 (2013)
6. Reuter, C., Marx, A.: Social software as an infrastructure for crisis management - a case study about current practice and potential usage. In: Proceedings of 8th International ISCRAM Conference, pp. 1–10 (2011)
7. O’reilly, T.: What is web 2.0: design patterns and business models for the next generation of software. *Commun. Strateg.* **65**, 17–37 (2007)
8. Macintosh, A., Coleman, S., Schneeberger, A.: eParticipation: the research gaps. In: Macintosh, A., Tambouris, E. (eds.) ePart 2009. LNCS, vol. 5694, pp. 1–11. Springer, Heidelberg (2009)
9. Chang, A.: Leveraging Web 2.0 in Government. E-Government/Technology Series Leveraging Web 2.0 in Government
10. Panopoulou, E., Tambouris, E., Tarabanis, K.: eParticipation initiatives in Europe: learning from practitioners. In: Tambouris, E., Macintosh, A., Glassey, O. (eds.) ePart 2010. LNCS, vol. 6229, pp. 54–65. Springer, Heidelberg (2010)
11. Charalabidis, Y., Loukis, E.: Transforming Government Agencies’ Approach to eParticipation through Efficient Exploitation of Social Media. In: ECIS 2011 (2011)
12. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. In: Proceedings of the 2008 International Conference on Web Search and Data Mining - WSDM 2008, p. 183 (2008)
13. Claes, A., Sizov, S., Angeletou, S., Taylor, S., Wandhoefer, T.: WeGOV project: where eGovernment meets the eSociety. In: Initial WeGov toolbox, pp. 1–65 (2010)
14. Effing, R., Van Hillegersberg, J., Huibers, T.: Social media and political participation: are Facebook, Twitter and YouTube democratizing our political systems? *Electron. Participation* **6847**, 25–35 (2011)
15. Moreira, A.M., Ladner, A.: E-Society and e-democracy. *Media* (2009)

16. Kuzma, J.: Asian Government Usage of Web 2.0 Social Media. *Sites J. 20th Century Contemp. French Stud.* 1–13 (2010)
17. Ashley, H., Corbett, J., Jones, D., Garside, B., Rambaldi, G.: Change at hand: web 2.0 for development. In: *Participatory Learning and Action*, vol. 59, pp. 8–20. IIED (2009)
18. Makinen, M., Kuirra, M.W.: Social media and postelection crisis in Kenya. *Int. J. Press/Polit.* **13**, 328–335 (2008)
19. Gao, H., Barbier, G., Goolsby, R.: Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intell. Syst.* **26**, 10–14 (2011)
20. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Sci. Am.* **5**, 28–37 (2001)
21. Schreiber, R., Swick, G.: Semantic web best practices and deployment working group (2006)
22. Frank, M., Eric, M.: RDF primer
23. Wimmer, M.A.: Ontology for an e-participation virtual resource centre. In: *Proceedings of the 1st international conference on Theory and practice of electronic governance - ICEGOV 2007*, p. 89 (2007)
24. Loutas, N., Lee, D., Maali, F., Peristeras, V., Tarabanis, K.: The semantic public service portal (S-PSP). In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) *ESWC 2011, Part II. LNCS*, vol. 6644, pp. 227–242. Springer, Heidelberg (2011)
25. Ojo, A., Estevez, E., Janowski, T.: Semantic interoperability architecture for Governance 2.0. *Inf. Polity* **15**, 105–123 (2010)
26. Embley, D.W., Campbell, D.M., Smith, R.D., Liddle, S.W.: Ontology-based extraction and structuring of information from data-rich unstructured documents. In: *Proceedings of the Seventh International Conference on Information and Knowledge Management - CIKM 1998*, pp. 52–59. ACM Press, New York (1998)
27. Alani, H., Kim, S., Millard, D.E., Weal, M.J., Hall, W., Lewis, P.H., Shadbolt, N.: Web based Knowledge Extraction and Consolidation for Automatic Ontology Instantiation. In: *proceedings of the Workshop on Knowledge Markup and Semantic Annotation at the second International Conference on Knowledge Capture (K-Cap 2003)* (2003)
28. Bikel, D.M., Miller, S., Schwartz, R., Weischedel, R.: Nymble. In: *Proceedings of the Fifth Conference on Applied Natural Language Processing*, pp. 194–201. Association for Computational Linguistics, Morristown (1997)
29. Etzioni, O., Cafarella, M., Downey, D., Popescu, A.-M., Shaked, T., Soderland, S., Weld, D.S., Yates, A.: Unsupervised named-entity extraction from the web: an experimental study. *Artif. Intell.* **165**, 91–134 (2005)
30. Soderland, S.: Learning information extraction rules for semi-structured and free text. *Mach. Learn.* **34**, 233–272 (1999)
31. Dalvi, B., Cohen, W., Callan, J.: Websets: Extracting sets of entities from the web using unsupervised information extraction. In: *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*. ACM (2012)
32. Alfonseca, E., Manandhar, S.: An unsupervised method for general named entity recognition and automated concept discovery. In: *Proceedings of the 1st International Conference on General WordNet, Mysore* (2002)
33. McCrae, J., Aguado-de-Cea, G., Buitelaar, P., Cimiano, P., Asunci, D., Gracia, J., Hollink, L., Tobias, M.D.S.: Interchanging lexical resources on the semantic web. *Lang. Resour. Eval.* **46**, 701–719 (2012)
34. Hevner, A., Chatterjee, S.: *Design Research in Information Systems*. Integrated Series in Information Systems, pp. 9–23. Springer, Boston (2010)
35. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. *J. Manage. Inf. Syst.* **24**, 45–77 (2007)

36. Gabrilovich, E., Markovitch, S.: Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: IJCAI International Joint Conference on Artificial Intelligence, pp. 1606–1611 (2007)
37. Lane, M., Coleman, P.: Technology ease of use through social networking media. *Technology* 1–12 (2012)
38. Hagen, L., Harrison, T.M., Uzuner, Ö., Fake, T., Lamanna, D.: Introducing textual analysis tools for policy informatics: a case study of e-petitions. In: Proceedings of the 16th Annual International Conference on Digital Government Research, pp. 10–19 (2015)