Cross-Context Linking Concepts Discovery in E-Government Literature

Bojan Cestnik[1,2](✉) and Alenka Kern[3]

[1] Temida d.o.o., Ljubljana, Slovenia
bojan.cestnik@temida.si
[2] Jozef Stefan Institute, Ljubljana, Slovenia
[3] Housing Fund of the Republic of Slovenia, Public Fund, Ljubljana, Slovenia
alenka.kern@ssrs.si

Abstract. To conduct their business, organizations are nowadays challenged to handle huge amount of information from heterogeneous sources. Novel technologies can help them dealing with this delicate assignment. In this paper we describe an approach to document clustering and outlier detection that is regularly used to organize and summarize knowledge stored in huge amounts of documents in a government organization. The motivation for our preliminary study has been three-fold: first, to obtain an overview of the topics addressed in the recently published e-government papers, with the emphasis on identifying the shift of focus through the years; second, to form a collection of papers related to a preselected terms of interest in order to explore the characteristic keywords that discriminate this collection with respect to the rest of the documents; and third, to compare the papers that address a similar topic from two document sources and to show characteristic similarities and differences between the two origins, with a particular aim to identify outlier papers in each document source that are potentially worth for further exploration. As a document source for our study we used E-Government Reference Library of articles and PubMed. The presented case study results suggest that the document exploration supported by a document clustering tool can be more focused, efficient and effective.

Keywords: Document clustering · Linking concepts discovery · E-government · Public housing · Social media

1 Introduction

Every modern organization in both government and private sector needs to process, organize and store information that is required to conduct its business. In this task, ontologies typically play a key role in providing a common understanding by describing concepts, classes and instances of a given domain. They are frequently built manually by extracting common-sense knowledge from various sources in some sort of representation. Many computer programs that support manual ontology construction have been developed and successfully used in the past, such as Protégé [1].

Since manual ontology construction can be a complex and demanding process, there is a strong need to provide at least partially automated support for the task. With the emergence of new text and literature mining technologies, large corpora of

H.J. Scholl et al. (Eds.): EGOV 2016, LNCS 9820, pp. 19–30, 2016.
DOI: 10.1007/978-3-319-44421-5_2

documents can be processed to semi-automatically construct structured document clusters [2]. Resulting document clusters can be viewed as concepts (classes, topic descriptions) that can be used to describe domain properties in the form of topic ontologies. In recent years, various tools that help constructing document clusters from texts in a given problem domain were developed and successfully implemented in practice [2]. One example of such tool that enables interactive construction of clusters of text documents in a selected domain is OntoGen [3]. It can be used to extract concepts from input documents and organize them into high-level topics. By using modern data and text processing techniques OntoGen supports individual phases of ontology construction by suggesting concepts and their names and defining relations between them [4].

Literature mining is a process of applying data mining techniques to sets of documents from published literature. Essentially, literature mining is a technique used to tame the complexity of high dimensional data and extract new knowledge from the available literature. It can be used in many ways and for various purposes, also, for example, when dealing with problems spawning from economic crisis that the society is facing in our time. For instance, in [5] the authors analyze and compare innovation in public and private sectors. They identify three factors for improved interest for innovation in public sector. First, the requirements and expectations of the public sector services have grown considerably. Second, the number of complex problems that the public sector has to face in the areas like public safety, poverty reduction, and climate mitigation has also grown. And third, innovative capabilities of governments and localities play an important role in the competitive globalization game [5].

Documents that are of interest for an organization might come from various sources. They can be stored in the organization's Intranet storage, or can reside in a more or less organized form and format on the Internet. Among many publicly accessible potential sources we can identify semi-structured Semantic Web entities and Linked Data sources, as well as more organized public libraries such as Medline and PubMed [6], E-Government Reference Library [7], and Google Scholar [8]. A general text processing management and ontology learning process from text consists of several steps [e.g. 2]. First, the documents (natural language texts) and other resources (e.g. semi-structured domain dictionaries) are obtained from designated sources. Then, they are preprocessed and stored on text processing server. In the next step, domain ontology is built with ontology learning and ontology pruning algorithms. In the last step, the constructed ontology is visualized, evaluated and stored on a repository for further use and exploration.

The main motivation for our case study was to demonstrate how the text processing can be used for public documents and government data. We wanted to present the utility and evaluation of the approach from the interested parties' (i.e. public bodies) viewpoint. In particular, our aim was to offer some interesting insights, such as how the document clustering technology can be used to identify mutual subsets of papers from one context (document source) that were more close to the subset of papers from the other context. Such a cross-context approach to linking term discovery has been introduced in medical field [e.g. 9–11] and has been used to identify hidden relations between domains of interest with a great success.

In the case study described in this paper we used E-Government Reference Library of articles [7] and PubMed [6] as a document source. In the first experiment we obtained an overview of the topics addressed in the recently published e-government papers. In particular, we were interested in the shift of focus of the papers through the years; the keywords describing document clusters gave us clues about which topics are trending in certain time periods. In the second experiment we formed a cluster of papers related to a preselected term (in our case we used two arbitrarily selected terms: "social media" and "housing") in order to explore the characteristic keywords that discriminate this cluster with respect to the rest of the documents. The underlying assumption was that while it is often easy to automatically collect data, it requires considerable effort to link and transform them into practical information that can be used in concrete situations. In the third experiment we combined the papers addressing the similar topic from two document sources, e-Government Reference Library and PubMed. Then, we identified characteristic similarities and differences between the two origins, with a particular aim to identify outlier papers that are worthy of further exploration for finding potential cross-context concept links. Here, the underlying assumption was that while the majority of papers in a given domain describe matters related to a common understanding of the domain, the exploration of outliers may lead to the detection of interesting associating concepts among the sets of papers from two disjoint document sources. In addition, focusing on a potentially interesting subset of outlier papers might considerably reduce the size of article corpora under investigation. The presented case study results suggest that the document exploration aided by OntoGen can, in comparison to the traditional manual one, be more focused, efficient and effective.

This paper is organized as follows. In the Sect. 2 we describe the construction of the input sets of documents. In the Sect. 3 we describe the methods used in the study and present three cases in which OntoGen was used to generate and visualize clusters of documents with similar properties. In Sect. 4, we assess and discuss the main lessons learned from the case study. The paper is concluded in Sect. 4.

2 Document Sources

Documents and papers that are of interest for an organization can be obtained from many publicly accessible sources on the Internet. There are several semi-structured Semantic Web entities and Linked Data sources, as well as more organized public libraries such as Medline and PubMed [6], E-Government Reference Library [7], and Google Scholar [8]. Majority of the contemporary published papers can be, depending on the copyright issues, obtained in an electronic form from the Internet. It is particularly useful when a set of documents from a selected domain is available in some sort of standard format.

One such example is E-Government Reference Library – EGRL – [7] that in the current version 11.5 contains 9.690 references of peer-reviewed articles predominantly in English language. It is available in XML format for public download and use. Another example of a resource of papers on the Internet is PubMed [6], which contains papers largely from the medical field.

The first step in the process of text mining and document clustering is retrieval and preprocessing of text documents. For our study we took 7.810 documents from the EGRL library in XML format as an input for further processing. Text mining and document clustering methods were shown to produce useful results on scientific papers when used on titles and abstracts [12]. Therefore, in the preprocessing phase we excluded the papers that contain only title in the XML file and included only those library papers that have also their abstracts available. There are 5.223 such papers in the library. Each relevant paper was described with the year of publication, the title and the abstract. Short statistics of the included papers according to the year of publication is shown in Table 1. The first input document collection was used in the experiments described in Subsects. 3.1 and 3.2.

To process the papers that address a similar topic from two document sources we prepared the second input document collection from the PubMed papers responding to the search string “social media” and “government”. The criteria for the search were arbitrarily selected with the aim to focus on the papers related to “government” topic and narrow the number of retrieved papers. Note that any other specific topic of interest can be used instead of “social media”. The concrete search query was “government AND social AND (media OR network)”. As a result, we obtained 9.690 papers, from which 5.327 papers had abstracts and were published after the year 2004. The second input document collection was used together with the first document collection in the experiment, described in Subsect. 3.3.

Table 1. Number of papers from E-Government Reference Library [7] by the year of publication. In the last two columns the papers with included abstract are given.

	All papers		With abstracts	
Publication year	Number	%	Number	%
2002 and before	502	6.4	283	5.4
2003	288	3.7	211	4.0
2004	404	5.2	270	5.2
2005	465	6.0	243	4.7
2006	353	4.5	93	1.8
2007	592	7.6	210	4.0
2008	353	4.5	297	5.7
2009	687	8.8	449	8.6
2010	650	8.3	428	8.2
2011	702	9.0	431	8.3
2012	793	10.2	469	9.0
2013	763	9.8	682	13.1
2014	698	8.9	606	11.6
2015	560	7.2	551	10.5
Total	7.810	100.0	5.223	100.0

3 Document Clustering with OntoGen

The process of forming clusters of documents from a set of documents and naming them by keywords can be considered as creating topic ontology in a domain under study. Ontologies include descriptions of objects, concepts, attributes and relations between objects. They conceptualize and integrate the domain terminologies that can be identified in text. Therefore, ontologies reflect the content and the structure of the knowledge as it can be recognized through the use of terms in the inspected collection of texts. Note that the documents that are used in the construction of topic ontologies must be carefully selected before they are processed and considered for analyses.

Ontologies for a given domain can be constructed manually using some sort of language or representation. In manual extraction, an expert seeks common sense concepts and organizes them in hierarchical form. Since manual ontology construction is a complex and demanding process, several computerized programs have been created that support semi-automatic construction of ontologies from a set of documents [e.g. 2]. Based on text mining techniques that have already been proved successful for the task, OntoGen [4] is a tool that enables the interactive construction of ontologies from text in a selected domain. Note that OntoGen is one representative of the tools that help constructing ontologies from texts. With the use of machine learning techniques, OntoGen supports individual phases of ontology construction by suggesting concepts and their names, by defining relations between them and by the automatic assignment of text to the concepts. The most descriptive words of each concept are obtained by the SVM [13] from the documents grouped in each cluster.

The input for OntoGen is a collection of text documents. Documents are represented as vectors; such representation is often referred to as Bag of Words (BoW) representation [14]. In the BoW vector space model, each word from the document vocabulary stands for one dimension of the multidimensional space of text documents. This way, the BoW approach can be employed for extracting words with similar meaning. Therefore, it is commonly used in information retrieval and text mining for representing collections of words from text documents disregarding grammar and word order, which enables to determine the semantic closeness documents. BoW vector representation can also be used to calculate average similarity between the documents of a cluster. The similarity is also called cosine similarity, since the similarity between two documents is computed as cosine of the angle between the two representative vectors.

3.1 Topic Focus Shift Through Time

In the first experiment we set a goal to acquire an overview of the topics (keywords) prevailingly addressed in the recently published e-government papers. In particular, we were interested in the shift of topic focus of the papers through the years. The characteristic keywords describing document clusters, which were generated automatically with OntoGen, gave us clues about which topics are trending in certain time periods. By using OntoGen users can construct a complex ontology more efficiently and in shorter time period than manually. They can create concepts, organize them into topics and also assign documents to concepts. Simultaneously, they have full control over whole process (therefore semi-automatic) by choosing or revising the suggestions provided by the system [3].

We constructed a topic ontology with OntoGen from the abstracts of 5.223 papers from EGRL [7], shown in Table 1 and Fig. 1. The topics represent temporal divisions (clusters) of documents according to the year of publication and are labeled with the most descriptive words. The topic ontology from Fig. 1 can be regarded as a structure of folders for the input set of papers. In such way it can enrich our prior knowledge about the domain, motivating creative thinking and additional explanations of the constructed concepts. Moreover, the descriptions of clusters (keywords) in Fig. 1 can be used to analyze trends in the published topics. For example, keyword “media” (or “social media”) appeared in the descriptions only after year 2011 and gained more importance after 2013. Keyword “citizens” is spotted from 2005 on, while “cities” gained importance in 2015 with the smart cities initiative. Many other interesting relations can be observed directly from Fig. 1. Note that average similarity measure for each cluster is also shown in Fig. 1.

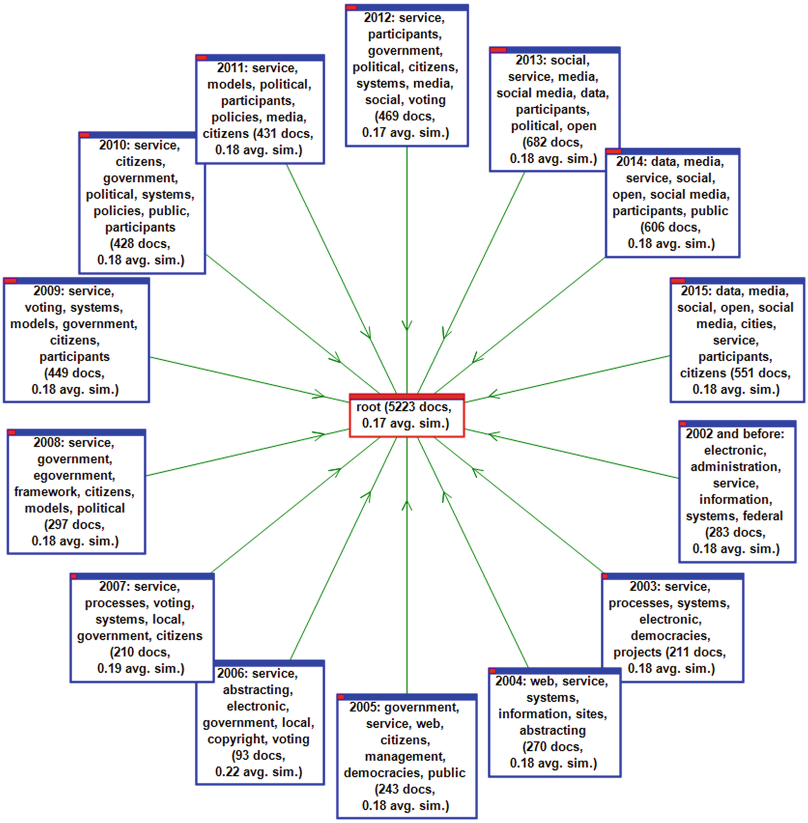
2012: service, participants, government, political, citizens, systems, media, social, voting (469 docs, 0.17 avg. sim.)
2011: service, models, political, participants, policies, media, citizens (431 docs, 0.18 avg. sim.)
2013: social, service, media, social media, data, participants, political, open (682 docs, 0.18 avg. sim.)
2010: service, citizens, government, political, systems, policies, public, participants (428 docs, 0.18 avg. sim.)
2014: data, media, service, social, open, social media, participants, public (606 docs, 0.18 avg. sim.)
2009: service, voting, systems, models, government, citizens, participants (449 docs, 0.18 avg. sim.)
2015: data, media, social, open, social media, cities, service, participants, citizens (551 docs, 0.18 avg. sim.)
root (5223 docs, 0.17 avg. sim.)
2008: service, government, egovernment, framework, citizens, models, political (297 docs, 0.18 avg. sim.)
2002 and before: electronic, administration, service, information, systems, federal (283 docs, 0.18 avg. sim.)
2007: service, processes, voting, systems, local, government, citizens (210 docs, 0.19 avg. sim.)
2003: service, processes, systems, electronic, democracies, projects (211 docs, 0.18 avg. sim.)
2006: service, abstracting, electronic, government, local, copyright, voting (93 docs, 0.22 avg. sim.)
2005: government, service, web, citizens, management, democracies, public (243 docs, 0.18 avg. sim.)
2004: web, service, systems, information, sites, abstracting (270 docs, 0.18 avg. sim.)

Fig. 1. 5.223 papers from EGRL library clustered according to the year of publication. Each cluster is described with SVM [13] keywords that characterize the contained papers.

3.2 Grouping Papers by Selected Characteristic Keywords

In the second experiment we generated a special cluster of papers related to a preselected term (in our case we used two arbitrarily selected terms: “social media” and “housing”) in order to explore the characteristic keywords that discriminate this cluster with respect to the rest of the documents. The underlying assumption was that while it is often easy to automatically collect data, it requires considerable effort to link and transform them into practical information that can be used to help decision makers in concrete situations. As input we took the abstracts of 5.223 papers from EGRL and manually (overriding OntoGen’s document similarity feature) constructed four clusters. In the first cluster we included documents containing term “social media” (503 papers); the remaining 4.720 documents were included in the second cluster. In the third cluster we included documents containing term “housing” (21 papers); the remaining 5.202 papers were included in the fourth cluster. Then, we generated SVM keyword descriptions for each cluster that distinguish it from its counterpart cluster (the first from the second, and the third from the fourth cluster). The goal was to explore the characteristic keywords that discriminate the documents in one cluster with respect to the rest of the documents. In our case, we wanted to identify common concepts (keywords) between the two clusters, since “social media” and “housing”are both topic of high interest for our organization, and pinpoint the most relevant papers describing the two topics.

The four clusters and descriptions are shown in Fig. 2. The cluster for “social media” is described with the following keywords: “social, media, social media, networks, political, social networks, community, twitter, participants, citizens”, while the remaining cluster is described by “service, systems, government, models, data, citizens, public, information, participants, processes”. The cluster for “housing” is described with the following keywords: “housing, community, service, digital, divide, digital divide, social, citizens, website, government website”, while its counterpart cluster is described by “service, government, systems, citizens, models, public, data, participants, political, social”. The descriptions of two distinguished clusters share two common keywords: “social”and “citizens”. The central document for “social media” ncluster is the document with id 1998 [15], while the central document for “housing” cluster is the document with id 6588 [16]. The two documents were used for more detailed preliminary study of the two topics and for finding new, potentially uncovered ideas for social media applications in housing.

3.3 Combining Papers from Two Document Sources

In the third experiment we combined the papers addressing the similar topic from two document sources, e-Government Reference Library and PubMed. Our aim was to identify characteristic similarities and differences between the papers from the two origins. In particular, we were interested in outlier papers that are worthy of further exploration for finding potential cross-context concept links [e.g. 11]. Here, our assumption was that while the majority of papers in a given domain describe the matters related to a common understanding of the domain, the exploration of outliers

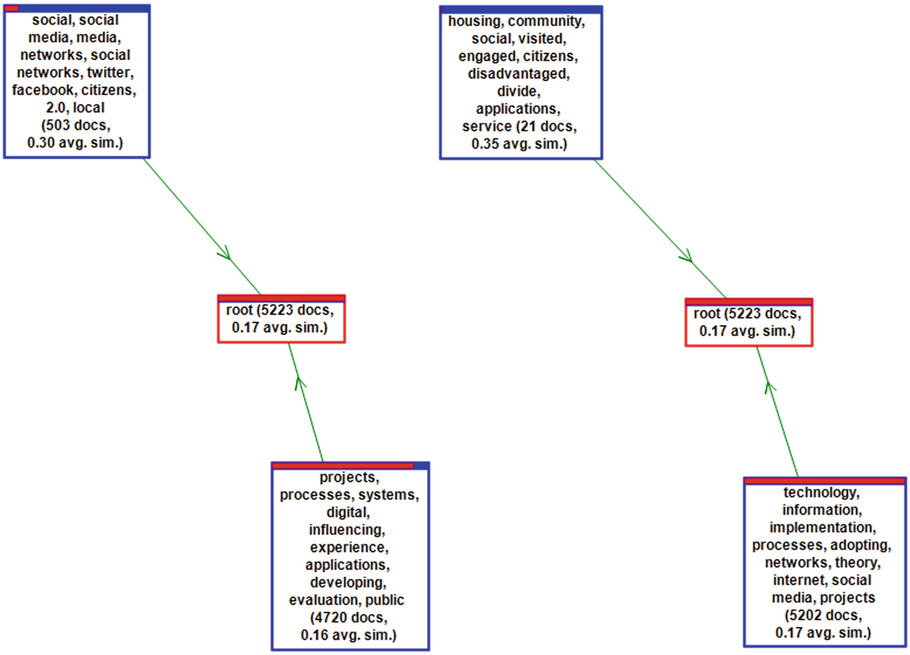
social, social media, media, networks, social networks, twitter, facebook, citizens, 2.0, local (503 docs, 0.30 avg. sim.)
root (5223 docs, 0.17 avg. sim.)
projects, processes, systems, digital, influencing, experience, applications, developing, evaluation, public (4720 docs, 0.16 avg. sim.)
housing, community, social, visited, engaged, citizens, disadvantaged, divide, applications, service (21 docs, 0.35 avg. sim.)
root (5223 docs, 0.17 avg. sim.)
technology, information, implementation, processes, adopting, networks, theory, internet, social media, projects (5202 docs, 0.17 avg. sim.)

Fig. 2. Two document clusters for preselected terms "social media" (left) and "housing" (right). The characteristic keywords that discriminate the two clusters with respect to the rest of the documents are shown in the rectangles.

may lead to the detection of interesting associating concepts among the sets of papers from two disjoint document sources. In addition, focusing on a potentially interesting subset of outlier papers might considerably reduce the size of article corpora under investigation, which might also help decision-makers narrowing down the mere quantity of papers to read for further study.

For practical purposes, we have joined the first and the second input document collections to obtain 10.550 papers with abstracts. Then, we have constructed with OntoGen two clusters of documents based on their similarity. I the papers from the two sources were completely different, the two clusters would most probably contain the documents from one document source, respectively. However, the situation depicted in Fig. 3 shows that this assumption is only partially correct. The two top level clusters are labeled "health, careful, patients" and "service, citizens, government". The first cluster (lets denote in with P) contains 8.416 documents, while the second one (denoted with E) contains 5.734 documents. Second level clusters reveal that in cluster P there is a majority of papers (4.749) from PubMed and only a minority (67 papers in cluster denoted P-E) of papers from eGov field. The situation is reversed in cluster E: here, the majority is from eGov (5.156 papers) and slightly bigger minority from PubMed (578 papers in cluster denoted E-P).

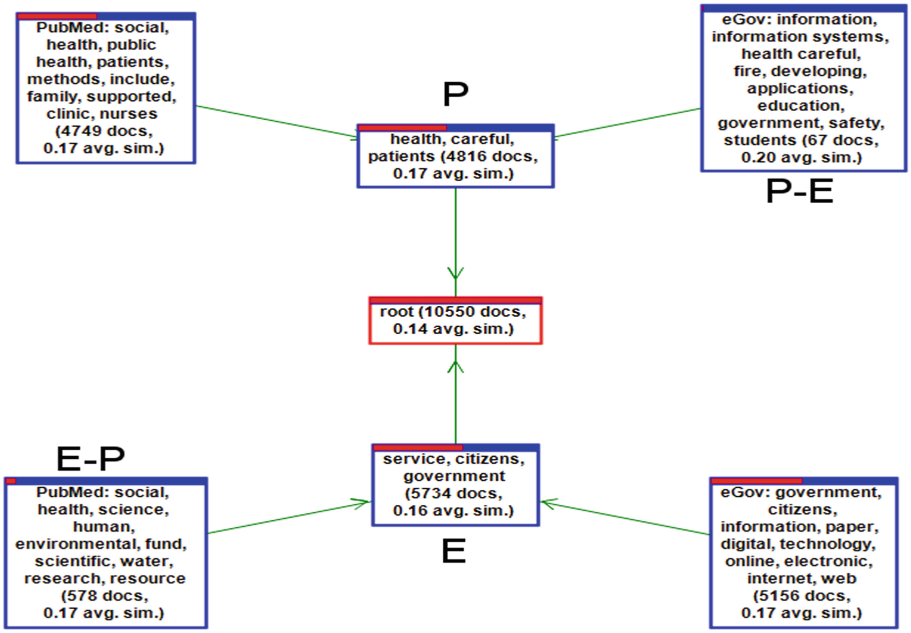
PubMed: social, health, public health, patients, methods, include, family, supported, clinic, nurses (4749 docs, 0.17 avg. sim.)
P
health, careful, patients (4816 docs, 0.17 avg. sim.)
eGov: information, information systems, health careful, fire, developing, applications, education, government, safety, students (67 docs, 0.20 avg. sim.)
P-E
root (10550 docs, 0.14 avg. sim.)
E-P
PubMed: social, health, science, human, environmental, fund, scientific, water, research, resource (578 docs, 0.17 avg. sim.)
service, citizens, government (5734 docs, 0.16 avg. sim.)
E
eGov: government, citizens, information, paper, digital, technology, online, electronic, internet, web (5156 docs, 0.17 avg. sim.)

Fig. 3. Combining papers from two sources: e-Government Reference Library and PubMed. Clusters containing outlier documents are shown in bottom-left and top-right rectangle.

In cluster P-E there are 67 documents from EGRL library that are described with keywords “information, information systems, health careful, fire, developing, applications, education, government, safety, students”. They are “outliers” from eGov (EGRL) library because they are more similar to PubMed documents. Clearly, they prevailingly deal with the health-related issues. On the other hand, in cluster E-P there are 578 documents from PubMed that are more similar to EGRL library documents. The can be described with the following keywords: “social, health, science, human, environmental, fund, scientific, water, research, resource”.

In our preliminary study we took into account the outlier papers from both P-E and E-P clusters and formed combined blended input document collection for further analysis. Our aim was to investigate the potential of outlier clusters for uncovering linking concepts between the two fields in our further work. In order to reduce the search space, the white list of interesting potential linking concepts for further consideration (shown in Table 2) that was prepared with OntoGen and further refined and validated by the domain expert.

All the listed terms appear to be interesting to the domain expert that was included in the process. The identified outlier papers for each term seem worth for further exploration. For example, the single outlier paper from EGRL that contains the term “family” states how job clarity, effective communications with management, a participatory management approach, organizational support of career development, opportunities for advancement, and **family-friendly policies** are all significant variables affecting the job satisfaction of IT employees [17]. The two papers from EGRL

Table 2. The list of potential linking terms between outliers E-P (PubMed) and P-E (eGov library of documents). Number of outlier papers containing each term is shown.

Term	Number of papers	
	E-P (PubMed)	P-E (eGov)
Safety	14	5
Media	96	2
Privacy	12	1
Family	13	1
Education	32	7
Disability	6	2
Disadvantage	3	1
Economy	13	1
Low income	2	2
Financial incentive	3	1
Electronic health	1	3
Public fund	9	1
Big data	1	1

that include term “disability” deal with health status impact to information consumers [18] and regional disparities in occurrences of diseases due to unsafe water resources in China [19]. We have observed that the last paper is indexed also in the PubMed library. When considering “disadvantage” as a linking term, the outlier document indexed in EGRL that deals with poverty and health in the good society [20] was identified. It is actually a book published by Palgrave Macmillan and is definitely worth reading and referencing in further studies. Last but not least, we found two outlier documents containing term “big data”. The first document indexed in PubMed deals with big data analysis framework for healthcare and social sectors in Korea [21], while the second document is indexed in EGRL and deals with incentivizing health information exchange [22].

4 Conclusion

In this paper we describe three experiments in using text processing and clustering methods to model and visualize existing but often overlooked knowledge that is hidden in documents and papers. The issue addressed is the information integration in e-Government domain ontologies and their visualization through the similarity maps. The ontologies were constructed semi-automatically with the computational support of OntoGen [3] using scientific papers from EGRL [7] and PubMed as input. The use of OntoGen has enabled a quick insight into a given domain by semi-automatically generating the main ontology concepts from the domain’s documents.

Our observations show that ontologies help gaining understanding in a given subject area. Therefore, using tools for semi-automatic ontology construction from textual data can significantly speed up the process of becoming acquainted with the domain of interest. We can first generate top-level domain ontology concepts and thus obtain a general overview and understanding of the domain, and only then concentrate on reading an extra load of information. In such a way, semi-automatically constructed ontologies actually helped us to review and understand the variety of topics of interest prior to further investigation.

Encouraged by the growing demands for public innovation, one of the aims of this article was also to explore technological possibilities for supporting creative processes in public sector. In order to exploit existing but often overlooked knowledge that is hidden in public information we investigated the potential of text processing and document clustering. In the third experiment we focused on identifying outlier documents from two document sources (PubMed and EGRL libraries), since the exploration of outliers may lead to the detection of interesting associating concepts among the two sets of documents. We have demonstrated that focusing on a potentially interesting subset of outlier papers considerably reduces the size of document corpora under investigation. Our observations show that using tools for semi-automatic ontology construction from text can significantly speed up the process of becoming acquainted with the domain of interest, thus making the process more focused and effective.

Acknowledgement. The presented work was partially carried out for the Housing Fund of the Republic of Slovenia, Public Fund. The authors wish to thank the Fund's management for their continuous cooperation and support.

References

1. Gennari, J., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubezy, M., Eriksson, H., Noy, N.F., Tu, S.W.: The evolution of protégé: an environment for knowledge-based systems development. Int. J. Hum.-Comput. Stud. **58**(1), 89–123 (2003)
2. Kietz, J.U., Mädche, A., Mädche, E., Volz, R.: A method for semi-automatic ontology acquisition from a corporate intranet. In: EKAW-2000 Workshop "Ontologies and Text", Juan-Les-Pins (2000)
3. Fortuna, B., Grobelnik, M., Mladenić, D.: System for semi-automatic ontology construction. In: Demo at ESWC 2006, Budva, Montenegro (2006)
4. Fortuna, B.: OntoGen: Description. http://ontogen.ijs.si/index.html. Accessed 15 Dec 2015
5. Sørensen, E., Torfing, J.: Enhancing collaborative innovation in the public sector. Adm. Soc. **43**(8), 842–868 (2011)
6. PubMed, 15 December 2015. http://www.ncbi.nlm.nih.gov/pubmed
7. Scholl, H.J.: E-Government Reference Library (EGRL) version 11.5 (2015). https://catalyst.uw.edu/webq/survey/jscholl/22768. Accessed 3 Jan 2015
8. Google Scholar (2016). https://scholar.google.com/. Accessed 15 Dec 2015
9. Swanson, D.R., Smalheiser, N.R., Torvik, V.I.: Ranking indirect connections in literature-based discovery: the role of medical subject headings (MeSH). J. Am. Soc. Inf. Sci. Technol. **57**(11), 1427–1439 (2006)
10. Weeber, M., Vos, R., Klein, H., de Jong-van den Berg, L.T.W.: Using concepts in literature-based discovery simulating Swanson's Raynaud–fish oil and migraine–magnesium discoveries. J. Am. Soc. Inf. Sci. Technol. **52**(7), 548–557 (2001)

11. Petrič, I., Cestnik, B., Lavrač, N., Urbančič, T.: Outlier detection in cross-context link discovery for creative literature mining. Comput. J. **55**(1), 47–61 (2012)
12. Cestnik, B., Urbančič, T., Petrič, I.: Ontological representations for supporting learning in business communities. In: Smrikarov, A. (ed.) e-Learning 2011: proceedigs of the International Conference on e-Learning and the Knowledge Society, pp. 260–265. ASE Publishing House, Bucharest (2011)
13. Ayed, Y.B., Fohr, D., Haton, J.-P., Chollet, G.: Keyword spotting using support vector machines. In: Sojka, P., Kopeček, I., Pala, K. (eds.) TSD 2002. LNCS (LNAI), vol. 2448, pp. 285–295. Springer, Heidelberg (2002)
14. Sebastiani, F.: Machine learning in automated text categorization. ACM Comput. Surv. **34** (1), 1–47 (2002)
15. Effing, R., van Hillegersberg, J., Huibers, T.W.: Social media participation and local politics: a case study of the Enschede council in the Netherlands. In: Wimmer, M.A., Tambouris, E., Macintosh, A. (eds.) ePart 2013. LNCS, vol. 8075, pp. 57–68. Springer, Heidelberg (2013)
16. Sipiror, J., Ward, B.: Bridging the digital divide for e-government inclusion: a United States case study. Electron. J. e-Gov. **3**(3), 137–146 (2005)
17. Kim, S.: IT employee job satisfaction in the public sector. Int. J. Public Adm. **32**(12) (2009). Special Issue: Reforms of Welfare Administration and Policy
18. Goldner, M.: How health status impacts the types of information consumers seek online. Inf. Commun. Soc. **9**(6), 693–713 (2006)
19. Carlton, E.J., Liang, S., McDowell, J.Z., Li, H., Luo, W., Remais, J.V.: Regional disparities in the burden of disease attributable to unsafe water and poor sanitation in China. Bull. World Health Organ. **90**(8), 578–587 (2012)
20. Cattell, V.: Poverty, Community, and Health: Co-operation and the Good Society. Palgrave Macmillan, New York (2011)
21. Song, T.M., Ryu, S.: Big data analysis framework for healthcare and social sectors in Korea. Healthcare Inf. Res. **21**(1), 3–9 (2015)
22. Jarman, H.: Incentivizing health information exchange: collaborative governance, market failure, and the public interest. In: Proceedings of the 15th Annual International Conference on Digital Government Research, pp. 227–235. ACM, New York (2014)