Original Research Paper

An Ontological Crawling Approach for Improving Information Aggregation over eGovernment Websites

Heru Agus Santoso, Junta Zeniarja, Ardytha Luthfiarta and Bima Jati Wijaya

Department of Informatics, Intelligent System Lab CISU, Dian Nuswantoro University, Semarang 50131- Indonesia

Corresponding Author:
Heru Agus Santoso
Department of Informatics,
Intelligent System Lab CISU,
Dian Nuswantoro University,
Semarang 50131- Indonesia
Email: heru.agus.santoso@dsn.dinus.ac.id

Abstract: E-Government applications in developing countries are still lagging behind e-Governments in advanced countries. For example, the use of information integration for Web portal content is still very limited. This paper proposes an automated approach for information aggregation over e-Government portals using ontological approach. The study uses data obtained from 10 local government Websites in the Central Java province-Indonesia. The data in the form of HTML Web document text, meta-data, hyperlinks and other rich-contents are effectively crawled. This paper focuses on the development of a crawler, which consists of two main modules, i.e., multi-tread downloader and scheduler. The use of ontology in the focused crawler producesamore effective result as compared to the Breadth First Search (BFS) approach as it reaches 37% of effectiveness in terms of the number of relevant documents downloaded.

Keywords: E-Government, Information Aggregation, Focused Crawler, Semantic Web, Ontology

Introduction

The implementations of e-Government in developing countries, including in Indonesia have not been providing an optimum benefit for government itself, citizens and businesses. For example e-government in Indonesia, United Nations ranks Indonesia at 106th in the world (United_Nation, 2014). The number of unmaintained e-Government portals or Web sites indicate the low score of e-Government Development Index (EDGI). As such, local governments website need further utilization (Hermana and Silfianti, 2011). Nurdin *et al.* (2012) reported that from 489 local e-Government implementation, 55% is still in the emerging stage. In other words, most of e-Government implementations in Indonesia are still at the level of creation and dissemination and some of them are in the level of enhancement.

This paper proposes an approach for e-Government information aggregation in an automatic manner. The approach produces a local e-Government database which can be utilized toimprove information aggregation of e-Governments. It can be further applied, such as for news or content provider using Really Simple Syndication (RSS), for benchmarking e-Govermnet website and so on. To achieve this goal, the technique namely crawling is employed. Carwling is the process of exploring and aggregating Web contents to ease searching and extraction process of information (Zheng *et al.*, 2013). Instead of using general crawling approach commonly used by commercial search engines, our approach uses focused crawling method. The focused crawling method can fulfill personal or organizational needs of specific information topics, to maintain Web portal content, aggregation of information locally, even to provide complex information (Batsakis *et al.*, 2009). The main difference between focused crawling and other methods lies in determining the relevance of aggregated information. Focused crawling uses a set of rules based on content, link structure Uniform Resaource Locator (URL) and anchor text to keep focused in the aggregation processes. However, one drawback of focused crawling is, it highly depends on specific topics. Thus, it cannot aggregate pages from relevant domains (Kozanidis, 2008; Pal *et al.*, 2009; Tan and Mitra, 2010; More and Govilkar, 2013). This paper proposes a method to overcome such limitation by using an ontological approach.

Ontology has a graph structure and weights that allows number of relevant topics to be analyzed with the combination of document content and link structure on the Website. Ontology helps crawler to understand the meaning of information, e.g., understanding equivalence of terms and distinguishing the difference of two objects. For instance, the system

can distinguish "Apple" in the context of fruits from the one in the context of electronic devices brand. This kind of feature is crucial for information aggregation. The use of ontology will support semantic description of resources (Davies *et al.*, 2006). Hence, it can help user to query specific objects in a semantic basis.

Ontology-based focused crawling proposed in this paper uses the Semantic Web technology. The Semantic Web in more general term refers to a set of technologies recommended by the World Wide Web Consortium (W3C) used to share information over the Web. W3C has recommended RDF to represent data over the Web and Resource Description Framework (RDFS)/Web Ontology Language (OWL) as the formal languages for authoring ontologies. OWL, which uses Description Logic (DL) as the formal logical foundation, describes knowledge of a specific domain using classes, properties and their relationships. The ontology is used to determine information relevancy which will be aggregated. The crawling approach is expected to be more efficient than the common techniques, because the information which will be collected should be more relevant, minimizing cost of time, bandwidth and storage.

Literature Review

This section presents literatures review that consists of two main sections, i.e., related work of web crawling, including the implementation of crawling method in eGoverment domain and its related theories.

Related Work

The Studies in the field of Web-crawling have been attracting many researchers, especially for those who interested in the field of Web engineering. A number of publications can be summarized as follows.

Kumar and Vig (2012) proposed the use of learning process to determine selected information to be aggregated. Term Frequency-Inverse Document Frequency (TF-IDF) is taken into account as the weight of relevant document. TF-IDF is used to determine how important a word in documents collection. The resulting score will assist the priorities of crawler to download Web pages. Crawling process is then performed sequentially and continuously. The weakness of this model is the use of words as input. As it is known that word is different from concept. For example, the word "apple" and the concept of "apple", if they are used in the crawling process will produce different output, e.g., apple as gadget different from apple as fruit. Hence, the drawback of the approach can be overcome with the use of ontology to determine the contextually and relationship among existing concepts.

Bedi *et al.* (2013) used Social Web and Semantic technology to develop a focused crawling. They named it as Social Semantic Focused Crawler to get data and relevant information from the Web. Search pattern follows general pattern when the objects of the search were limited only to the bookmarking sites Social Bookmarking Site (SBS). SBS is a collection of sites that provide services for users to add, delete and share bookmarks. Crawler uses the concept of ontology to obtain relevant bookmarks. The output of this approach is relevant bookmarks.

Batsakis *et al.* (2009) propose the design and implementation of focused crawling which its main module consists of input, downloads Web pages, processing the content, assignment of priorities and expansion modules. The model is an enhancement of classical focused crawling, namely the use of keywords in the form of certain topics. The model is also equipped with the ability to estimate relevant pages based onset of topics provided. Estimation is performed using Hidden Markov Model (HMM) to obtain the path that leads to relevant Web pages. As explained earlier, the model has weakness because it only focused on topics that are very specific therefore; it is not able to aggregate relevant topics.

Ontologies have been integrated for the enhancement of focused web crawler as reported in (Kozanidis, 2008; Bedi *et al.*, 2013) and so forth. However, none of them have been implemented in e-Government domain. On the other hand, the uses of crawling techniques in e-Government have been significantly useful in several works. The study conducted by Sabucedo and Anido-Rifón (2010) showed that government services can be determined automatically using the crawling technique. Service search process is done by the following steps: Firstly determining the location of services the public wants. Initial input consists of a Web address or URL that leads to a Web services provider. Secondly, when the service is found, the information is then collected and analyzed. Furthermore, the crawler runs in two operations, namely: (a) Web exploration using information that exists in micro format, which uses HTML tag to store metadata that is processed automatically. (b) Perform a scanning process of HTML to get links that lead to related Web pages. In general, this model provides a mechanism to store and collect information about the services provided by the government.

McNutt and Pal (2011) reported that crawling techniques can be used to analyze and map the structure of links in e-Government Website. For example, Issue Crawler will follow specific links to observe the structure of communities and aggregate their relevant information. Once the crawling process is done, it is followed by the analysis of links that are found. It can be used to determine the level of influence on community structure in a certain region.
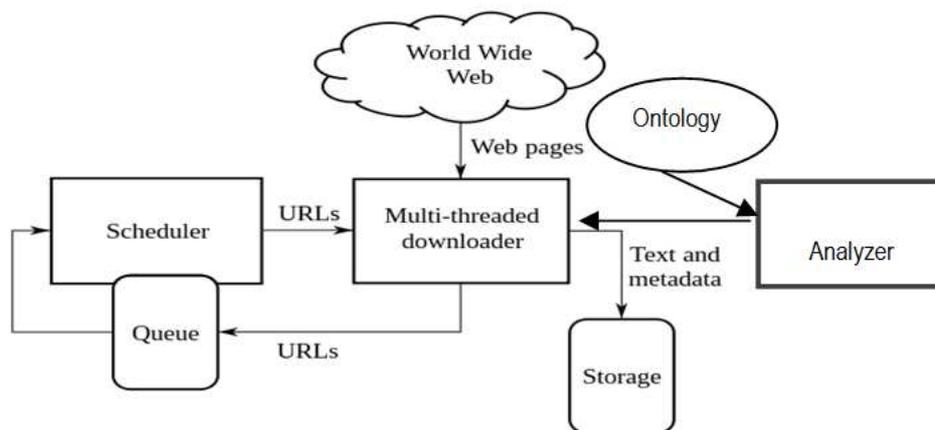
Fig. 1. The architecture of ontology-based focused crawler

Study conducted (Margetts and Dunleavy, 2007) reported that Web crawling can be used to identify the characteristics of e-Government. These characteristics of e-Government can be: (1) the size and content of Website, (2) the ease of finding information services and (3) its visibility on search engines. The approach is used to collect the contents and structures of the 26 sites in the UK government. It is implemented by visiting each site for further size and structure identification and for pattern relationships analysis between Web pages. Yahoo Site Explorer is integrated to get the pages that are connected to the site.

Based on the above related works, we conclude that the use of Web crawling in domain of e-government needs further study. Furthermore, the works particularly in the use of crawling techniques to maintain connectivity between e-Government portal content have not been appeared. Thus, the study of crawling applications by using ontological approach for information aggregation over e-Government website is very relevant.

Theoretical Background

Crawler is the software to crawl and collect Web contents locally (Kozanidis, 2008). The results of collecting such contents are used to facilitate searching and extracting information on the Web. A focused crawler, as one of crawler techniques, is commonly used to aggregate specific topics for organization or individual needs, for instance to maintain Web portal contents, to collect the documents locally or meet the needs of complex information (Batsakis *et al.*, 2009).

Generally, the architecture of ontology-based focused crawler is depicted in Fig. 1. Ontology provides certain topics or criterion that has been prepared. Given a set of Web pages as input, the crawler will extract existing hyperlinks on Web pages; specify URLs that must be visited based on certain criterion and priorities. Web

page address represented by URL will then be placed in a certain order (queue). The relevant downloaded pages will be stored in a local storage. Next, they can be analyzed and further processed. As described in Fig. 1, the architecture of the crawler consists of three main components, namely multi-threaded downloader, scheduler and analyzer. Multi-threaded downloader allows the downloading process carried out simultaneously by using a schedule set by the scheduler. While the analyzer module is used to analyze Website contentsthat have been downloaded.

In computer science, ontology constitutes a formal description of concepts and their relationships, which provides common terms and vocabularies for a given domain. In the Semantic Web community, the most cited definition of ontology is given by Gruber (Antoniou and Van Harmelen, 2004). He defined ontology as an "explicit specification of a conceptualization" which means that through conceptualization, ontology makes implicit knowledge of a particular domain become an explicit knowledge. As such ontological entities used in the conceptualization should be explicitly defined by means of specific formal languages such as RDFS and OWL. Gomez-Perez *et al.* (2006) stated that the aim of ontologies is to represent specific knowledge of domain of interest in a generic way. Hence, ontology can be used as common vocabulary that is shareable among different applications. This feature enables arbitrary applications to directly infer specific information rather than just displaying.

The Semantic Web in more general term refers to a set of technologies recommended by the W3C used to share information over the Web. W3C has recommended RDF to represent data over the Web and RDFS/OWL as the formal languages for authoring ontologies. RDF is the statement about Web resource, practically known as RDF triples. RDFS is a

more expressive language as compared to RDF. RDFS is used to structure the RDF resource. On the other hand, OWL is the most expressive language as compared to RDF and RDFS. OWL which uses DL as the logical foundation formally describes knowledge of a specific domain using classes, properties and relationships.

If ontologies are used to describe and share resources in a machine-processable manner, ontologies need to be formally specified using a specific ontology language. Ontology language such as OWL is a formal language by which ontology is created. It must be expressed in a concrete notation. Based on its formal semantics, ontology language consists of two major groups, namely the Frame-based language and the DL-based language (Antoniou and Van Harmelen, 2004). DL-based languages have been developed in the area of KR research. DL-based language describes the representation of knowledge as a set of concepts and roles and their relationships. Concepts and roles are comparable to frames and slots in the frame-based language. The important aspect of DL-based language is its well-understood theoretical properties (Antoniou and Van Harmelen, 2004), which provides automatic reasoning using machine-readable concept descriptions as well as automatically generates class taxonomy. OWL and its extension, OWL 2 are the examples of DL-based language. In February 2004, W3C has recommended OWL as the standard ontology language. Although OWL 2 has been introduced as the extension and revision of OWL in October 2009, however, OWL remains compatible with OWL 2.

In recent years, every citizen all over the world hunger for improvement of public services, i.e., being more responsive to the dynamics of politic, economic, social and technology policies. A good example for this is the reality that public administration service should be closer with the society in daily life and more proactive rather than just being reactive (Batsakis *et al.*, 2009). E-Government is a transformation of public-sector internal and external relationships through the use of Information and Communication Technology (ICT) to promote greater accountability, efficiency, cost-effectiveness and create greater constituency participation (Bedi *et al.*, 2013). E-Government is an effort by government to provide people with more convenient access to government information and services, improve the quality of services and provide greater opportunities to participate in the democratic institution and processes (Calvanese *et al.*, 2007). In e-Government, the support in exchange of data, information and knowledge exchange is becoming the key issue (Ahmad and Hasibuan 2012).

Proposed Method

Preliminary Study

Our preliminary studies have been conducted related to ontological approaches and e-Government, mostly about extraction and management of ontology. The result of studies have been published in journal such as (Santoso *et al.*, 2010; 2011a; 2011b) and international conferences (Santoso *et al.*, 2009; 2011c).

The method proposes in the paper focuses on crawler development, especially on the development of multi-threaded downloader and scheduler by using stages as follows:

a. Exploring the contents of 10 local government websites in Central Java, including metadata and other rich-contents. This stage is conducted in order to describe the profile of the current local government's Website

b. Identification of the structure, behavior and requirement analysis for a crawler that are needed for modeling of multi-threaded downloader and scheduler

c. Ontology development

d. At this stage, meth ontology method is carried out (Gomez-Perez *et al.*, 2006). The method consists of the following steps: (a) Specification: Determine scope of egov ontology based on terms and characteristics; (b) conceptualization: Organize and structure the knowledge; (c) formalization: Transform the knowledge using a formal language, i.e., OWL; (d) implementation: Use the ontology coded in OWL as background knowledgeof crawler. Jena is employed in this phase; and (e) maintenance: Supporting activities for ontology implementation, which consists of acquisition of new knowledge, integration, evaluation, ontology documentation and configuration management

e. Design and Implementation of crawler model

f. Software engineering method used in this stage is a Model-Driven Engineering (MDE) (Ga *et al.*, 2009) in which the software design is developed based on the transformation of the model that has been generated in the previous stage, e.g., database design and ontology. Furthermore, the developed design is then implemented into the program code to produce the crawler

g. Simulation with real data is carried out to ensure that the crawler has been free from disadvantages and anomalies. Next, it is followed by the validation and evaluation

Results

As an initial model of Web crawler development, the crawler is intended to be able to aggregate Web contents from 10 local governments Website in Central Java-Indonesia. It crawled over Websites and stored their data corresponding to respective file types by implementing BFS algorithm. Then, analyze the results in details, i.e., (1) the number of files obtained from the respective website, (2) any files obtained from crawling, (3) the size of each website, (4) relevant contents obtained (5) The link structure affects the amount of contents that can be crawled (6) The amount of contents indicate of the lack of security in a Website.

The crawler deals with dynamic contents by crawling over e-government Websites, then analyzes all type of documents containing in the Web including static and dynamic HTML (produced by on-the-fly manner). As the crawler cannot access data stored in the restricted database, it browses history logs and crawls seed page and URL information such as site maps.

BFS-Based Crawling Result

BFS-based web crawling program is used to get the content from 10 government websitesin Central Java.The result is then stored into a MySQL database inlocal hard drive. After that, the result is analyzed for comparison in the development of crawling method with better algorithms. The issue is how to crawl relevant contents. Figure 2 shows crawling flowchart of BFS-based crawl algorithm. As the most important part of crawling, it generates information from 10 local government websites.

Table 1 shows the result of BFS-based crawler which its flowchart depicted in Fig. 2. Information obtained from Table 1 shows that the total yields based on BFS crawling achieve overall size is 1.025.808KB or 1GB of file sizes. The percentage ratio of HTML and PDF documents are 97.27 and 2.73% respectively. The largest average of file size is owned by the City of Semarang website or http://semarangkota.go.id/ with the size 87.30 KB and the smallest one is owned by Cilacap, i.e., 25,38KB. The most data are obtained from http://pekalongankab.go.id/ with the total number of files are 10,995; consisting of 245 HTML files and 10,995 PDF files with the size reach 493.400KB.

By using BFS approach, all documents contained in the Website with particular depth will be downloaded. Hence, BFS approach is considered ineffective to aggregate data or information from Websites. As it downloads all files instead of relevant data or information, the issue is the main motivation of the study.
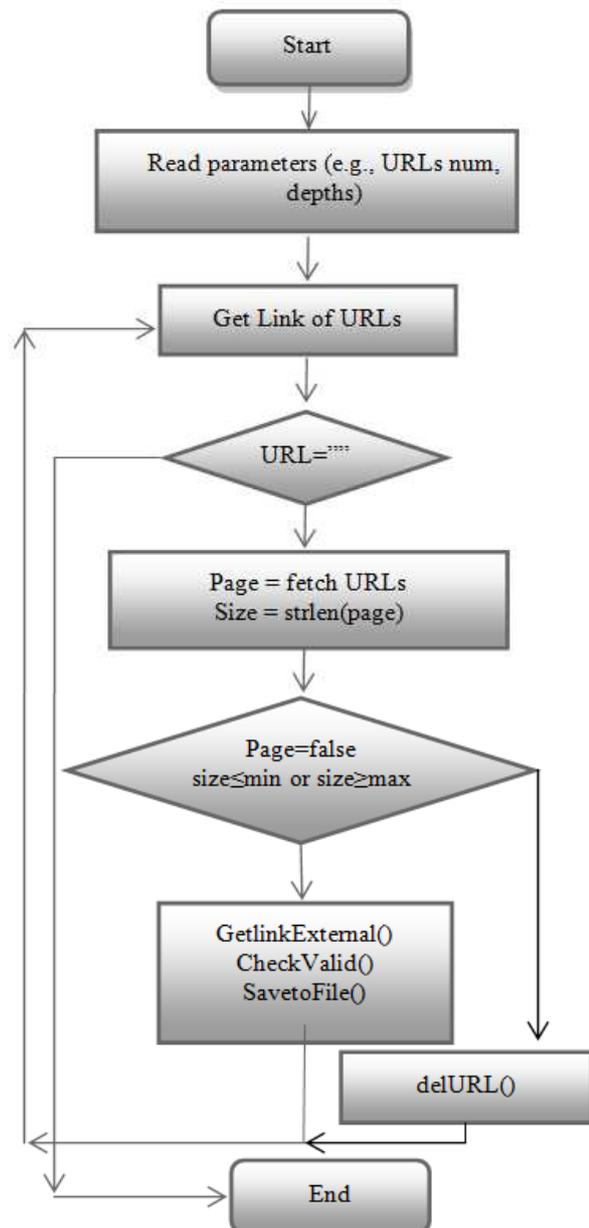

Fig. 2. Flowchart of BFS-based crawler

Ontology Development

Ontology used in this paper is developed based on Indonesian Government regulation. The eleven main classes created for this study are Capital Investment, Housing, Development Planning, Environment, Agriculture and Food Security, Spatial Planning, Youth and Sports, Health, Public Works, Education, Cooperatives and Small and Medium Enterprises. Since we use the ontology as the background knowledge in the crawling process, technically, the approach can be applied to all instances of government in Indonesia.

Fig. 3. The excerp of e-Gov ontology used in the crawler

Table 1. The result of BFS Crawler

Local Gov. Website	File type		Doc Num	Size (KB)	Avg (KB)
	HTML	PDF			
pekalongankab.go.id	10,750	245	10,995	493,400	44.87
kebumenkab.go.id	3,190	166	3,356	230,000	68.53
semarangkab.go.id	2,780	60	2,840	108,510	38.21
banyumaskab.go.id	531	0	531	15,216	28.66
magelangkab.go.id	241	42	283	20,900	73.85
patikab.go.id	122	0	122	7,382	60.51
semarangkota.go.id	39	1	40	3,492	87.3
rembangkab.go.id	22	0	22	570	25.91
tegalkab.go.id	19	0	19	1,196	62.95
cilacapkab.go.id	8	0	8	203	25.38
Total	19.725	554	20.279	1.025.809	682

In this study, ontology O is defined as 4-tuples:

$$O = (C, R, I, Ch) \qquad (1)$$

Where:
C = The set of concepts or classes
R = The set of roles
I = The set of instances or individuals
Ch = Concept hierarchy

The excerpt of e-Gov ontology used in this paper is depicted in Fig. 3. It explains the relationship of the 11 classes. Set of instances of each classes are 1750 Web document terms, while set of roles used are mostly Sub Class Of.

Preprocessing and Weighting

The purpose of preprocessing stage is to get the whole terms and indexing (in ascending manner) which is then stored into the variable terms All Document. The variable is used as a reference for each document vector space model. The preprocessing and weighting is carried out described as follows:

a. docN is a variable represented documentswhich has been through in preprocessing stage, i.e.,

tokenization, stop words removal and stemming. Terms are stored interm Doc

b. Read each term contained interm All Document then do a search on term Doc as well as counting the number of frequency of occurrence is stored into variable NewT. After completion of the check frequency NewT if more than 0 do additions to the document frequency

c. Add NewT into the res variable as a store results

The calculation of inverse document frequency and weighting is then performed. Implementation of the IDF calculation of the number of documents obtained divided by the frequency of occurrence of the term in each document. Term Frequency (TF) is the number of important terms occurrence. DF is the number of documents containing important terms. IDF is the invers of DF:

$$IDF(t) = \log\left(\frac{D}{DF}\right) \qquad (2)$$

$$Weight(t) = TF.IDF = TF.\log\left(\frac{D}{DF}\right) \qquad (3)$$

Similarity computation is carried out to search relevant documents. For similarity computation between keyword and web documents, this study uses cosine similarity as follows:

$$\frac{\sum x_i \cdot y_i}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum y_i^2}} \qquad (4)$$

Where:

x,y = The weight of terms in query and document respectively.

The Implementation of Ontology-based Focused Crawler

The Implementation of focused crawling using ontological approach is carried out over 10 e-Government website sites in Central Java Indonesia. All parameters, e.g., Website URLs, depth and value level of relevance are stored in a text file. The crawler then reads the file to g*et al*l parameters.Crawling process is performed with the predetermined relevant threshold value, i.e., 27%. Total storage reaches 258 MB which consists of 12,772 documents.

Figure 4 and 5 show the result of ontology-based focused crawling over 10 local government Websites. In terms of number of documents downloaded, BFS approach produced 20,279 documents, while ontology-based focused crawler produced 12,772 documents. Hence, our approach reached 37% of effectiveness. By using ontology as the background knowledge in the crawling process, it produces more relevant results. Classes and their contained instances in the ontology determine the relevancy of Web documents which were aggregated. This proves that ontology-based focused crawling is more efficient than BFStechnique; because the information collected are more relevant, minimizing cost of time, bandwidth and storage.
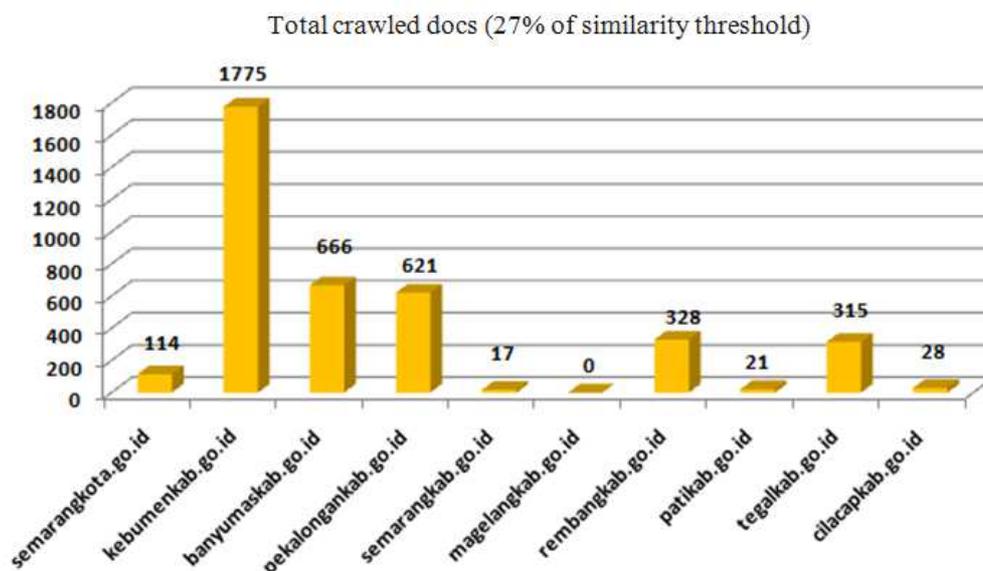

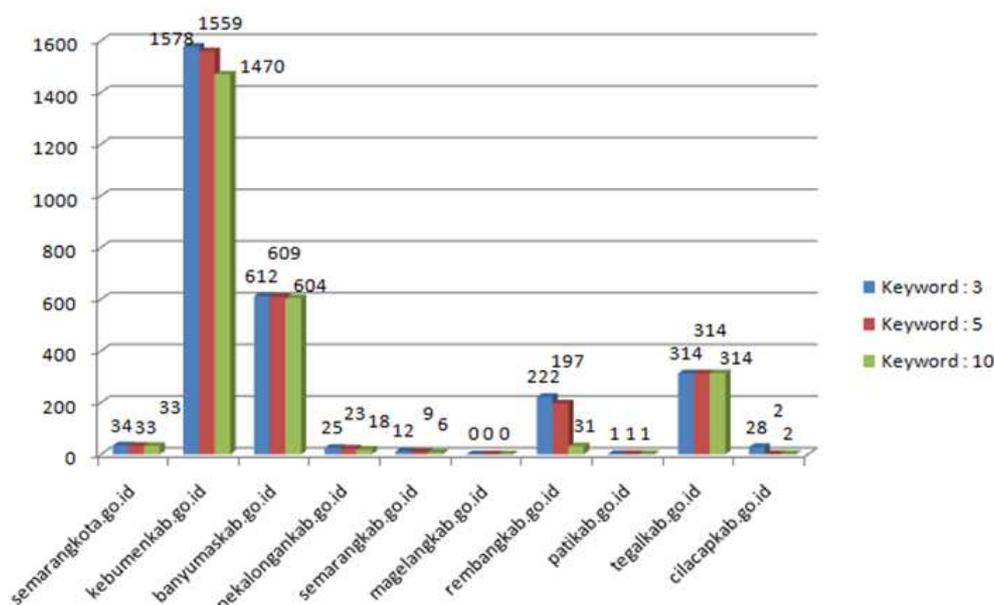Fig. 4. The result of ontology-based focused crawling based on similarity threshold

Fig. 5. The result of ontology-based focused crawling based on number of keywords

Conclusion

The use ontology in the focused crawler provides more effective results as compared with the result from BFS approach. The approach crawled only relevant Website contents instead of the whole Web contents. Hence, it is easer for further processing application such as for news or contents provider, feeder forReally Simple Syndication (RSS) or benchmarking e-Government Websites and so on.

Acknowledgement

Funding Information

Author's Contributions

Heru Agus Santoso: Conception, design, implementation and final approval/submission.

Junta Zeniarja: Data acquisition and interpretation.

Ardytha Luthfiarta: Writing and reviewing the manuscript.

Bima Jati Wijaya: Drafting the article.

Ethics

This article is original and contains unpublished material. The author approved the manuscript and confirms that no ethical issues arise.

References

Ahmad, M. and Z. A. Hasibuan, 2012. E-Government based on cloud environment in Indonesia. Seminar Nasional Aplikasi Teknologi Informasi (SNATI).

Antoniou, G. and F. Van Harmelen, 2008. A Semantic Web Primer. 2nd Edn., MIT Press, Cambridge, ISBN-10: 0262012421, pp: 264.

Batsakis, S., E. Petrakis and A. Milios, 2009. Improving the performance of focused web crawlers. Data Know. Eng., 68: 1001-1013. DOI: 10.1016/j.datak.2009.04.002

Bedi, P., A. Thukral and H. Banati, 2013. Focused crawling of tagged web resources using ontology. Comput. Electrical Eng., 39: 613-628. DOI: 10.1016/j.compeleceng.2012.09.009

Calvanese, D., G. Giacomo, D. Lembo, M. Lenzerini and R. Rosati, 2007. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. J. Automated Reason., 39: 385-429. DOI: 10.1007/s10817-007-9078-x

Davies, J., R. Studer and P. Warren, 2006. Semantic Web Technologies: Trends and Research in Ontology-Based Systems. 1st Edn., John Wiley and Sons, Hoboken, N.J., ISBN-10: 0470030348, pp: 326.

Gasevic, D., D. Djuric and V. Devedzic, 2009. Model Driven Engineering and Ontology Development. 2nd Edn., Springer Science and Business Media, Dordrecht, ISBN-10: 364200282X, pp: 378.

Gomez-Perez, A., M. Fernandez-Lopez and O. Corcho, 2004. Ontological Engineering: With examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. 1st Edn., Springer, London, ISBN-10: 1852335513, pp: 420.

Hermana, B. and W. Silfianti, 2011. Evaluating e-government implementation by local government: Digital divide in internet based public services in Indonesia. Int. J. Bus. Social Sci., 23: 156-163.

Kozanidis, L., 2008. An ontology-based focused crawler. Proceedings of the 13th international conference on Natural Language and Information Systems: Applications of Natural Language to Information Systems, (LIS' 08), Springer-Verlag Berlin, pp: 376-379. DOI: 10.1007/978-3-540-69858-6_48

Kumar, M. and R. Vig, 2012. Learnable focused meta crawling through web. Procedia Technol., 6: 606-611. DOI: 10.1016/j.protcy.2012.10.073

Margetts, H. and P. Dunleavy, 2007. Government on the Internet: Progress in Delivering Information and Services Online. 1st Edn., The Stationery Office, London, ISBN-10: 0102947244, pp: 45.

McNutt, K. and L. Pal, 2011. Modernizing government": Mapping global public policy networks. Governance, 24: 439-467.
DOI: 10.1111/j.1468-0491.2011.01532.x

More, M.R. and S. Govilkar, 2013. Profile based document specific crawling. Int. J. Eng. Res. Technol.

Nurdin, N., R. Stockdale and H. Scheepers, 2012. Benchmarking indonesian local e-government. AIS Electronic Library.

Pal, A., D.S. Tomarfea and S.C. Shrivastava, 2009. Effective focused crawling based on content and link structure analysis. Int. J. Comput. Sci. Inform. Security, 2: 1-5.

Sabucedo, L.A. and L.A. Rifon, 2010. Locating and crawling e government services a light-weight semantic approach. J. Universal Comput. Sci., 16: 1117-1137. DOI: 10.3217/jucs-016-08-1117

Santoso, H.A., Z.T. Abdul-Mehdi and S.C. Haw, 2009. Semantic enhancement framework for e government using ontology versioning approach. Proceeding of The 6th Conference on Information Technology and Application, (ITA' 09), Hanoi, Vietnam, pp: 1-12.

Santoso, H.A., S.C. Haw and C.S. Lee, 2010. Change detection in ontology versioning: A bottom-up approach by incorporating ontology metadata vocabulary. Proceedings of the International Conferences on Database Theory and Application, Bio-Science and Bio-Technology, Dec. 13-15, Jeju Island, Korea, pp: 37-46.
DOI: 10.1007/978-3-642-17622-7_5

Santoso, H.A., Z.T. Abdul-Mehdi and S.C. Haw, 2011a. Ontology extraction from relational database: Concept hierarchy as background knowledge. Knowledge-based Syst., 24: 457-464.
DOI: 10.1016/j.knosys.2010.11.003

Santoso, H.A., S.C. Haw and C.S. Lee, 2011b. On detecting and representing bottom-up change in ontology versioning. Int. J. Database Theory Applic., 4: 79-98.

Santoso, H.A., S.C. Haw and C.S. Lee, 2011c. Software reuse: MDA-based ontology development to support data access over legacy applications. Proceedings of the 2nd International Conference on Software Engineering and Computer Systems, Jun. 27-29, Kuantan, Pahang, Malaysia, pp: 130-142.
DOI: 10.1007/978-3-642-22203-0_12

Tan, Q. and P. Mitra, 2010. Clustering-based incremental web crawling. ACM Trans. Inform. Syst. DOI: 10.1145/1852102.1852103

United_Nation, 2014. United Nation e-Government Survey 2014. Retrieved 01 Oct, 2016.

Zheng, Q., Z. Wu, X. Cheng, L. Jiang and J. Liu, 2013. Learning to crawl deep web. Inform. Syst., 38: 801-819. DOI: 10.1016/j.is.2013.02.001