Semantic Web based E-Government System

Syed I. Jami* and Zubair A. Shaikh

Mohammed Ali Jinnah University, Pakistan; imran.jami@jinnah.edu, zubair.shaikh@jinnah.edu

Abstract

Objective: This paper focuses on integration of Semantic Web with E-Governance by developing a blog based portal for feedback to implement machine readable G2C in an automated way which is one of the targets of Government 3.0. Methods: We propose to change the representation system of blog based portals for linking data in accordance with a general schema to make it semantically meaningful for machines to perform queries. Citizen Feedback system is employed as a test case to extract raw data and associated metadata in an automated way without parsing any structure. Ontology is developed to help in identification of entities and interpret user comments. Findings: The proposed architecture will make it possible for machines to automatically extract and report the general feedback from comments given by a large community of users. Current Web 2.0 based portals show that for estimation of feedbacks, each blog needs to be evaluated either manually or by employing parsing techniques using Bag of Words approach. To implement our proposed approach, we have developed a social semantic-based networking system through which citizens give comments on the policies/decisions implemented or decided to be implemented by the Government. The system interprets machine-readable comments and generates a summary to report the overall public opinion on a certain government policy, hence reducing the gap between citizens and government - one of the primary objectives of eGovernment as Government 3.0. The use of W3C compliant framework in our implementation will lead to universal adaptability of system. It has been found that the proposed approach resulted in a more accurate and efficient analysis of user comments as compared to manual and NLP/Text Mining techniques within a restricted domain. Application: The current implementation can also be extended to cover other domains with a little modification in the ontology schema, such as education, finance, and healthcare.

Keywords: Auto Summarization, E-Government, Government 3.0, Ontology, Social Semantic Web

1. Introduction

eGovernment is focused at increasing interaction between the citizens and government (G2C) through digital interaction, thus actively engaging the general public in governmental issues. Some examples of e-government portals include that implemented by Singapore, Kingdom of Bahrain and USA. Typical e-services being offered to citizens include job search services, accreditation and validation of certificates, car registration, building permits, access to public libraries and such. Efforts are being made to equip services so as to allow e-government to collaborate with others to enhance interaction by engaging in a shared activity, usually from remote locations.

*Author for correspondence

Using the World Wide Web as a means of medium for communication, government portals interact with their users/citizens to exchange services thorough web-pages developed using a range of technologies/languages.

One of the most important issues for making eGovernment effective is to enable citizens participates in the decision-making process. In our proposed approach we present a platform to ensure that citizen opinions and comments are properly received by public bodies. Opinions matter a great deal in politics and laws can change under the (negative) pressure of public opinion. Politicians almost always try to understand what voters are thinking about pending policy or government regulation proposals. In other words, government's success relies on how effectively it communicates messages to citizens increasing the trust and relations with them through their participation in the decision-making process.

Keeping in view the significance of user feedback on governmental decisions, motivation for our project comes from the fact that up-to-date governmental social web sites are not consistently evaluated in the governmental decision-making process and as such citizens' voices are most of the times heard in a limited audience. As such, we wish to build a platform using which the Government and citizens can come together and collectively collaborate in working for the betterment of their country.

Semantic Web based solution is developed to give the machine readability to the response of general public. An XML based form is developed for the user to give the tagged representation to user opinions. For Semantic Interoperability Ontology is developed using OWL framework. All tools used in this work are W3C compliant for universal adaptability. The Semantic Web based solutions are also proposed in many B2B and B2C based applications through automated agent-based search for information in which Ontologies were developed to map the relationships between entities much similar to how a human brain deduces a relationship between entities. Ontology development tool such as, Protégé have been very helpful to accomplish this task.

Our proposed system is based on such approach and is particularly targeted towards its use by Government organizations. With e-Government application as G2C, Governmental agencies need to follow citizens to assess consumer satisfaction with services they receive. To get feedback from millions of users and then evaluate each of them manually is time consuming and practically impossible. As such, our system will provide the luxury of observing responses in an automated way from a large number of users, presenting valuable information for busy executives and business heads to review their decisions and monitor progress. It will aid the Government to learn about the problems faced by people and foresee how they will be affected if a certain policy is implemented; as such important measures and actions can be taken as appropriate.

2. Background

The main obstacle to providing better support to web users is that, at present, the meaning of web content is not machine-accessible. Through traditional approach, there are tools that can retrieve texts, tokenize (split) them into parts, check the spelling, and count their words. Examples include, Microsoft's NLPW in which uses syntactic techniques based on sentence structure¹. In² provide another tool for text-extraction service using a combination of statistical and grammatical analysis. However, when it comes to interpreting sentences and extracting useful information for users, capabilities of current software are still very limited.

The current work done in this domain is mainly based on Natural Language Processing (NLP) techniques^{3, 4} using sentence splitting and POS tagging to help identify features relevant to software review domain. The relevant features and the related user comments are then classified through machine learning approach using Support-Vector Machines (SVM) as positive, negative or neutral on per feature basis. The detailed survey⁵ provides detailed insight into Sentiment Analysis and Opinion Mining using NLP based algorithms. In another work⁴ sentiment analysis is performed to tag blogs as positive/ negative based on assigning weights to individual sentences and words to quantify the polarity and summing them up. The final score is then classified as positive or negative. Overall summary is produced by merging the individual summaries of positive and negative relevant sentences. In⁶ a system is proposed for opinion mining in the eGovernment domain to automatically find out public's opinion about governmental decisions as shown in blogs. They applied POS tagging to identify nouns and related adjectives on the assumption that nouns represent features of governmental decisions and adjectives their corresponding opinions by public, thus tagging each adjective-noun pair as an opinion phrase. After performing sentiment analysis⁶ developed a decision-tree classifier to cluster the opinion phrases as "supportive" or "non-supportive".

Recently, research has been done to combine semantic web with NLP techniques for the opinion mining task. Besides NLP techniques, the work in² uses Sentic computing⁸ concept to mine patient opinions from free text to *structured topic-tagged polarity*. Similarly, author⁹ utilized ontology to facilitate the sentiment classification task at the level of features specific to movie reviews domain. They used an ontology based approach to extract features from opinions where they identified sentences containing the ontology terminologies.

Several recent Semantic Web based work has been proposed for E-Government including¹⁰ in which an

ontology is modeled for life events for e-government service integration, and¹¹ in which it proposed an infrastructure for Semantic Web Services based e-government system that integrates Business Process Modelling (BPM) and Semantic Annotation with existing Semantic Web frameworks for the modeling and engineering of non-automated government processes. The deliverables of these works can help our framework presented in next section by strengthening our ontology for efficient semantic processing.

3. Proposed Approach

Modular approach is proposed for this architecture as shown in Figure 1. There are three major modules in our work. This includes Input, Semantic Processing and Output. Input module in our system provides X Forms interface for users to provide comments on some policy which is transformed into tagged representation as XML. The responses through X Forms are transformed into Semantic Processing module using RDF/ XML format using the ontology structure through which. User comments gets instantiated into an RDF file in accordance and conforming to the ontology structure defined in an OWL file. This RDF file besides user feedback also stores information about the content related to policies such as: which "policy" is associated with which "department"; proposed by which "official"; which "user" has commented on which "policy" etc. Output module of this system deals with intelligent queries that are executed over this file for semantic processing and results generated.

The ontology used helps in resolving issues of semantic inter-operability and schema heterogeneity by incorporating domain knowledge via abstract data/ domain models taking into account concepts related to diversified government departments and their inter-connectedness.



Figure 1. High-level block diagram of system architecture.

A portal was developed for two different kinds of users, System Administrator (Government Official) and the Citizens. The Citizens would be required to register themselves prior to posting their comments about a certain policy. Once registered, the users will be able to view all the Policies proposed by the Government, until the duration specified by the Administrator. Presented with a policy, user will be posed with a questionnaire, consisting of a range of questions, the answers to which will be used to generate a summary or a blog, viewable to the citizen. Citizens can only comment once on a given policy, they will be presented with their summary on a policy they have previously commented on.

The Administrator has the task of updating the portal with the policies as they are proposed, and to report the citizen response on a policy, to the concerned department (s), after pre-defined period of time. The citizen response on a policy proposed by Defense Ministry (for example) reflects the performance of that Ministry, the performance of the Prime Minister and in turn the overall performance of the President as the Head of the State. The citizen responses can be narrowed down from outright Opposition to Favoring based on the province the citizen represents. This would help the government to better isolate the problems and solve them accordingly, not upsetting others. Graphs and charts have been used to help better in identifying and evaluating the performance.

4. Result and Discussion

The system is evaluated on the basis of the performance of our proposed approach compared with other alternative approaches being used to achieve a similar objective. In this context, we tested our system in terms of efficiency and storage by measuring the query response time and summary generation time. The query response time refers to the time it takes for the computer to analyze all the user comments and present it in a summarized form. This response times were measured against the following four selected queries as shown in Figure 2. The Province includes Sindh, Punjab, Baluchistan, and Khyber Pakhtunkhwa and ratings include Oppose, Lightly Oppose, Neutral, Lightly Favor, and Favor. Summary Generation Time refers to the time it takes to generate summary out of the user comments for one user on some given policy.

Finally the results are compared with Keyword-based Text Mining Approach. To achieve this, a small prototype



Figure 2. Comparison of query response time.

is developed using text-mining approach of keywordbased searching to find out the demographic location of user from comments. The response times of all these approaches were compared to validate our work and to highlight any pros and cons of our system.

The comment density was gradually increased and the corresponding response time for each query was noted. Results are depicted in bar graph in Figure 2 against all the queries and as a line graph in Figures 3 and 4 against the first two queries.

According to the results, the query time varied depending only on the number of ratings being queried. As evident from Figures 3 and 4, the processing time for "all ratings" is approximately 5 times that of "one rating" when the "provinces" variable is kept constant. This is also apparent from Figure 2 as the bars corresponding to "All Provinces-All Ratings" is equal in length with that of "One Province-All Ratings" and same applies for "All Provinces-One Rating" and "One Province-One Rating"; with significant variation between the two categories.

The system generates automated summary of user feedback per user per policy. Comments entered by user in the form of feedback are transformed into a summarized passage. Figure 5 shows the time taken for summary generation as the comment density in the form of number of users was gradually increased.

It is worth noting that processing times for both summary generation and querying is not proportionally related to number of comments as evident from non-uniform line graphs (which are not straight line).

Alternative approaches to automate the process of opinion mining exist in the form of information retrieval and text/data mining techniques. For comparison purposes, we tried to find out the demographic location of user from comments using a simple approach of finding



Figure 3. Response time for all provinces all ratings query.



Figure 4. Response time for all provinces one rating query.



Figure 5. Summary generation time.

specific relevant terms such as Sindh, Punjab etc. within user comments. Using the inverted index approach, location of user comments were decided according to frequency of such keywords above a specified threshold. On a text file of equal size as that of our 240 comments, it took 0.374 seconds to build the index and another 0.032 seconds to search over that for a specific keyword. Although the query time is less as compared to our system's performance but the query needs to be repeated over a large number of query terms as a training dataset- not only for the four provinces but also for the cities within provinces, for example. Moreover, to decide the positive/ negative orientation of user comments, further processing needs to be done in the form of sentiment analysis to classify the feedback. Taking all these costs into account, the overall retrieval time sums up to a larger number than that given by the proposed system.

5. Conclusion

Comparative analysis reveals that semantic web approach for the task of comment evaluation has definite advantages. Firstly, this approach based on directed graph (ontology) traversal is far more accurate than other mining techniques. Further, it allows for precise meaningful queries to satisfy end user needs. Secondly, it is efficient in terms of processing needed to generate query results as discussed.

The results shown in the form of graphs are generated from queries performed on a single file containing both user comments and government specific data. Since policy-specific queries are performed more frequently on user comments only, this can be separated from other data. This other data includes governmental structure metadata, history of various office bearers and details of policies uploaded. Additionally, indexing and distributed computing techniques can be utilized. Taking these into account, it is presumed that the processing time obtained through our evaluation and summarized above would decrease drastically. However, the fast evaluation/retrieval of results comes at a cost of increased storage size. Moreover, comments can be taken in the form of answers to specific questions as compared to free text in the case of NLP techniques. This is because comments need to be classified into conceptual classes already defined in ontology. Identifying classes from free text is a very complicated task and an area under study. As such, the user feedback - main input of our system - is limited to answers to pre-defined questions given by the citizen community. Many interesting avenues can be further explored in this work. This includes security of the feedback to protect the malicious use of data. Querying results can further be improved by introducing efficient mechanism of indexing.

Semantic web can be combined with text mining techniques to evaluate the feedback from free text. An approach called "Sentic Computing" described in as text analysis based on common sense reasoning tools and domain specific ontologies. Research needs to be done about the applicability of such an approach in our context of eGovernment.

6. References

- 1. Leskovec J, Milic-Frayling N, Grobelnik M. Extracting Summary Sentences Based on the Document Semantic Graph. Microsoft Research; 2005. p. 1–9.
- Hirschberg J, Manning CD. Advances in Natural Language Processing. Science. 2015; 349(6245):261–6. https://doi. org/10.1126/science.aaa8685. PMid:26185244
- Shein KPP, Nyunt TTS. Sentiment classification based on ontology and SVM classifier. Proceedings of Second International Conference on Communication Software and Networks (ICCSN), IEEE; 2010. p. 169–72. https://doi. org/10.1109/ICCSN.2010.35
- Balahur A, Lloret E, Boldrini E, Montoyo A, Palomar M, Martínez-Barco P. Summarizing threads in blogs using opinion polarity. Proceedings of the Workshop on Events in Emerging Text Types (eETTS), Bulgaria; 2009. p. 23–31.
- 5. Vinodhini G, Chandrasekaran RM. Sentiment analysis and opinion mining: A survey. International Journal of Advanced Research in Computer Science and Software Engineering. 2012; 2(6):1–11.
- Stylios G, Christodoulakis D, Besharat J, Vonitsanou M, Kotrotsos I, Koumpouri A, Stamou S. Public opinion mining for governmental decisions. Electronic Journal of e-Government; 2010. p. 203–14.

- Cambria E, Hussain A, Durrani T, Havasi C, Eckl C, Munro J. Sentic computing for patient centered applications. Proceedings of the 2010 IEEE 10th International Conference on Signal Processing (ICSP); 2010. p. 1279–82. https://doi.org/10.1109/ ICOSP.2010.5657072
- Cambria E, Hussain A. Sentic computing. Cognitive Computation. 2015; 7(2):183–5. https://doi.org/10.1007/ s12559-015-9325-0
- Zhao L, Chunping L. Ontology based opinion mining for movie reviews. Proceedings of the International Conference on Knowledge Science, Engineering and Management (KSEM), Springer; 2009. p. 204–14. https:// doi.org/10.1007/978-3-642-10488-6_22
- Alqahtani A, Lu H, Lu J. Knowledge-based life event model for e-government service integration with illustrative examples. Intelligent Decision Technologies. 2014; 8(3):189-205 https://doi.org/10.3233/IDT-140188
- Fonou-Dombeu JV, Huisman M. Engineering semantic web services for government business processes automation. Electronic Government and the Information Systems Perspective. Kő A, Francesconi E, editors, Springer International Publications: Switzerland; 2015. p. 40–54. https://doi.org/10.1007/978-3-319-22389-6_4