Putting Things in Context: Dynamic eGovernment Re-Engineering using Ontologies and Context

Avigdor Gal, Aviv Segev
Technion – Israel Institute of Technology
{avigal@ie, asegev@tx}.technion.ac.il

ABSTRACT

The proliferation of the Internet allows citizens to communicate with public administration using various channels. The diversity of tools allows citizens to become more involved in the daily decision making of public administration, especially when it comes to local governments. However, these new methods of communication require efficient techniques for handling an ever growing flow of information to civil servants. This position paper promotes a technical solution to allow efficient communication between citizens and civil servants, allowing a civil servant a flexible mechanism to automatically sort out any incoming data, analyzing it on-the-fly, directing it to the person who may find it most usable, and using it in evaluating services offered by the government. Such a solution paves the way for interoperability, while at the same time supporting re-engineering using local modifications and lending itself well to the needs of multi-lingual and multi-cultural environments such as the European Union.

1. INTRODUCTION

The tremendous growth of the Internet enables citizens to communicate with public administration using a variety of channels, including email, Web publishing, Internet chat, Web forums, *etc.* The diversity of tools allows citizens to become more involved in the daily decision making of public administration, especially when it comes to local governments. However, these new methods of communication require efficient techniques for handling an ever growing flow of information to civil servants. In particular, manual compilation and analysis of incoming information require, on the one hand, extensive and tedious manual effort, and on the other hand, IT expertise, which goes far beyond the end user level, and thus demands direct involvement of the IT personnel, *e.g.*, in setting up discussion forums.

This position paper promotes a technical solution to allow efficient communication between citizens and civil servants. At the heart of our solution there is an ontology, currently the de-facto standard for representing semantic information. We also use a module we dub "knowledge extractor," which is based on statistical techniques to support text summarization. Together, they give a civil servant a flexible mechanism to automatically sort out any incoming data, analyzing it on-the-fly, directing it to the person who may find it most usable, and using it in evaluating services offered by the government.

The solution proposed in this work is aimed at supporting evolving organizations, dynamically adapting to the needs of the citizens. While current technology (and in particular ontology management systems) are static in nature, their envisioned usage goes well beyond a "design once, use forever" approach. Therefore, we propose a solution that allows minor modifications by using a "context," a lightweight, text-based mechanism that defines the essence of an ontological concept. A context can be modified to adapt to changes over time in the focus of a government following external changes, or citizen requests. It can also serve in defining minor changes in perspectives between local governments. Once changes are accumulated, an ontological change may be needed, representing a major shift in policy, to be captured by redesigning the ontology itself.

Such a solution paves the way for interoperability via a global ontology, while at the same time supporting re-engineering using local modifications and lending itself well to the needs of multi-lingual and multi-cultural environments such as the European Union.

The rest of the paper is organized as follows. We start by specifying the problem in eGovernment terms. Then, we dwell on the technological solution. We conclude with a case-study, to illustrate the benefit of the proposed solution to the eGovernment domain.

2. PROBLEM STATEMENT

Citizen participation in the democratic process has been decreasing for years in European countries. Therefore, there is a need for the introduction of methods and instruments that will contribute to the re-establishment of the public action legitimacy in citizens' mind. According to political science research, such legitimacy stems from good governance and involves trust in the public action, transparency of policy lifecycle (namely, formulation, decision making, implementation, and evaluation) and civic participation. Therefore, any mechanism to allow the success of good governance

should involve a) communication with citizens about service quality, and b) the introduction of new services to improve visibility and transparency of the public action. The extent to which such a mechanism would work depends directly on the involvement of all stakeholders in the debate. In particular, *there is a need to foster debates between citizens, politicians, and civil servants about policy evaluation in order to discuss possible re-orientations.* Important aspects to consider are:

- The decline of citizen participation in general elections throughout the last decade is often associated with a lack of traceability between policies formulated by political parties and politicians and their impact on everyday life. It also probably relates to the fact that many citizens believe their voice is not really heard and that politicians are not really accountable for their promises. In this context, strengthening the legitimacy of political actions requires involving citizens in the debate for policy formulation rather than limiting their interventions at the late step of decision-making through elections. Special care should also be taken that eDemocracy not become an additional barrier to democratic processes rather than boosting it.

- Although the involvement of politicians is obviously critical to the birth of any eDemocracy process, their responsiveness to citizen questions will depend on both the relevance and usefulness of the IT-enhanced processes.

3. PROPOSED SOLUTION

Our proposed solution is combined of two complementary elements, namely *ontologies* and *contexts*. Ontologies were defined and used in various research areas, including philosophy (where it was coined), and most recently, eCommerce applications. In his seminal work, Bunge defines Ontology as a world of systems and provides a basic formalism for ontologies [2]. We follow an engineering approach to ontology management, as will be discussed soon. We consider an *ontology* to be a directed graph, with nodes representing concepts (*things* in Bunge's terminology [1, 2]) and edges representing relationships.

Ontologies are typically considered to be globally accepted (if not universal), serving many applications from the same domain. Research has shown, however, that building global ontologies is a hard task. An alternative that was adopted in practice assumes no global ontology is in place. Rather, local ontologies are developed or adopted. Whenever applications need to exchange information, a process of ontology alignment is performed, which results in a mapping between ontologies. Such a process was shown to carry with it a level of uncertainty [3], which may not suit eGovernment applications. A hybrid approach was also presented in the literature [4], in which each ontology has two parts: a static part (which is the global ontology) and a dynamic part, which evolves either by exporting ontologies or by discovery.

Contexts were defined to be first class objects formulated by McCarthy [5] using a relation $ist(\mathcal{C}, P)$, asserting that a proposition P is true in a context $\mathcal{C}$. For the purpose of this work, we define a *context* to be a set of words defining a document, possibly associated with weights that represent the relevance of a word to a document. Such a definition coincides with certain models in classical Information Retrieval, where a context may represent the number of repetitions of a word in a document. An alternative method for defining contexts is given in Section 3.1.

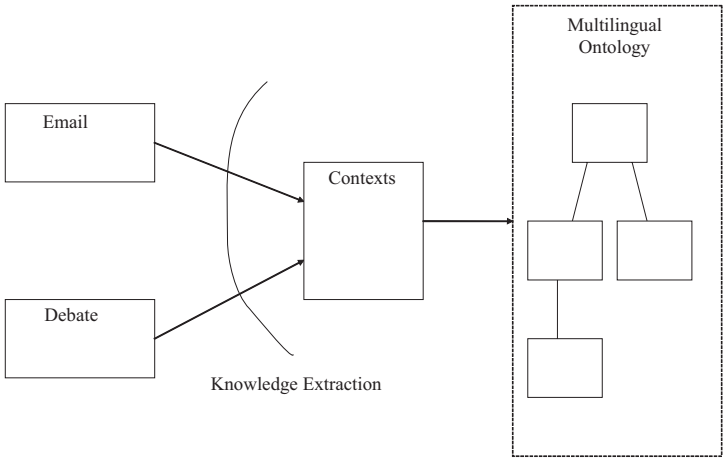

Figure 1: Putting things in context

Ontologies and contexts are joined together, as illustrated in Figure 1 and described herein. Each concept in the ontology is represented by a name and a context. Having such a representation in place has three main benefits as follows:

Flexibility with respect to a global ontology: eGovernment applications come, most likely, with a generic global ontology for eGovernment processes. Such an ontology is the outcome of a well-designed set of concepts and relationships modeled by experts. Once defined, evolving it becomes a difficult task. Here, contexts can serve as the local interpretation of the global ontology, which can be maintained by a civil servant without the involvement of the IT personnel.

Flexibility with respect to language: The multi-lingual culture of Europe may result in a need to adapt the ontology to each language separately. Avoiding such multiple effort is desirable, both for the initial specification of the ontology and for the ontology evolution. Here, the context can serve as the "translation" mechanism, in which ontological concepts are interpreted in the local language. To illustrate this point, consider the English term of a "state," representing an intermediate government level. While Germany uses a translation of the English term "state" for this concept, Poland officially uses a synonym term of a "province." The latter seems to represent an entity with less autonomy (similar to the difference between the federal systems of the US and Canada) than the former. When translating the English word "state" to German, one gets ten possible words, while with Polish, there are 14 possible meanings. The use of a context and the mechanism suggested below for generating the context of a concept (such as "state"), compensates for any under-

specification that may result from the universality of the ontology.

Flexibility with respect to culture: Even without the language barrier, different governments (say, local governments), may have different emphases on eGovernment tasks, representing cultural differences. For example, border cities (*e.g.*, Saarbrucken at the German-French border) may put more emphasis on recognizing language differences than cities in the heart of countries. Also, capital cities may have more sensitivity to minority culture than cities in the periphery. Context can serve as a compensating element in ontologies, adding topics of interest to the global ontology.

3.1 Context Extraction

Several methods have been proposed in the literature for extracting context from text. A set of algorithms was proposed in the IR community, based on the principle of counting the number of appearances of each word in the text, assuming that words with the highest number of appearances serve as the context. Variations on this simple mechanism involve methods for identifying the relevance of words to a domain, using methods such as stop-lists and inverse document frequency. For illustration purposes, we next provide a brief description of a context recognition algorithm and discuss its benefits. This method is based on [6].

The algorithm we present uses the Internet as a knowledge base to extract multiple contexts of a given situation, described in a document. The input to the algorithm is a document. The context recognition algorithm output is a context that attempts to describe the current scenario most accurately. Recall that a context is a list of words or phrases, each describing an aspect of the scenario, presented in the document. The context recognition algorithm consists of the following major phases:

Document Analysis: The document is analyzed using standard IR techniques and decomposed into linguistic meaningful parts (*e.g.*, sentences or paragraphs).

Support Expansion: Each document part is fed to a search engine, yielding a set of documents that best match the document part. This step serves in extending the basis of support for the next step of keyword ranking.

Keyword Ranking: The retrieved documents are analyzed and keywords are ranked according to their number of appearances in the retrieved documents, as well as in the original document.

Context Recognition: Using clustering techniques, the top ranked keywords are selected and output as the algorithm output.

There are a few subtleties that were not captured in this description and the interested reader is referred to [6] for more details. Given the model presented above, there is a benefit in automating context extraction in an eGovernment setting. Recall that we have separated the global ontology from its local variation. The former was designed by an ontology engineer, together with organization experts. The latter serves in specifying local flavors to concepts in the global ontology. An automatic context extraction provides a low-cost specification method. A civil servant provides the context extraction algorithm with a set of documents that belong to a specific ontology concept. These documents are processed and a tentative context is generated. The civil servant then modifies the context to her liking and associates it with the ontology concept.

This simple method capitalizes on the benefits mentioned above. First, the documents are provided in the local language, thus overcoming the language barrier. Second, new documents can be added as they arrive, revising the context as needed without going into the tedious and expensive ontology modification process. Finally, the most likely case is that local governments, even within the same country, will provide the context extraction with different documents, yielding different contexts that reflect the emphases of these governments, while keeping the global ontology intact, for the purpose of inter-communication.

4. QUALEG: A CASE IN POINT

QUALEG (QUALity of service and legitimacy in eGovernment) is a European 6^{th} framework project, aimed at providing an overall organizational and managerial solution to public involvement in public administration activities, combining various Internet technologies into an integrated system that serves as a portal for interacting with citizens. QUALEG supports public administration in organizing and managing public involvement in processes and activities. QUALEG provides an innovative solution to the problem of collecting, analyzing and using public feedback, based on a high level of automation to avoid a continuous involvement of IT personnel.

To illustrate the benefits of combining contexts and ontologies, we present next two of QUALEG services, namely email routing and public agenda. The email routing service routes emails of citizens, sent to some generic government mail box, to relevant civil servants following automatic analysis. Each email is analyzed using the knowledge extractor, yielding a context. This context is compared, in turn, to the contexts associated with the various ontology concepts. A match (even a partial one) indicates relevance, although some tuning may be needed to avoid false positives. An email will be routed to any ontology concept to which relevance was determined and with which an email address is associated. Whenever no routing is possible (either no relevance was identified or no associated email address was found) the email is directed to a manual operator for further processing.

QUALEG interface enables citizens to interact via various means, including emails, chat rooms, questionnaires, *etc.* During this interaction, some concepts repeat more often than others. The QUALEG system collects statistics about the recurrence of concepts in citizen input and determines which of these concepts are more popular on the public agenda. This process is accomplished, once more, by analyzing citizen input via the knowledge extractor and identifying relevant concepts via context comparison.

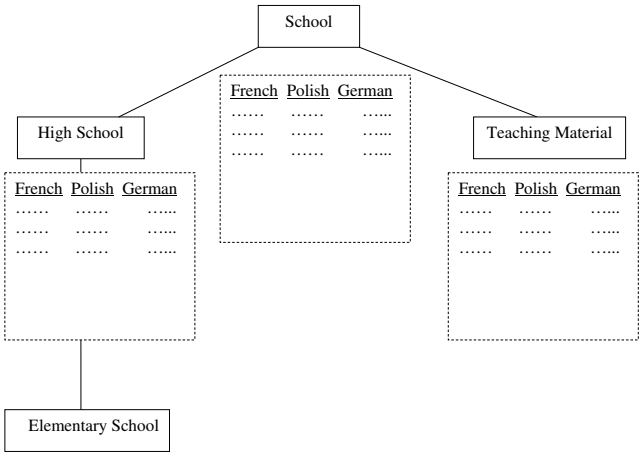

Figure 2: Multilingual Ontology

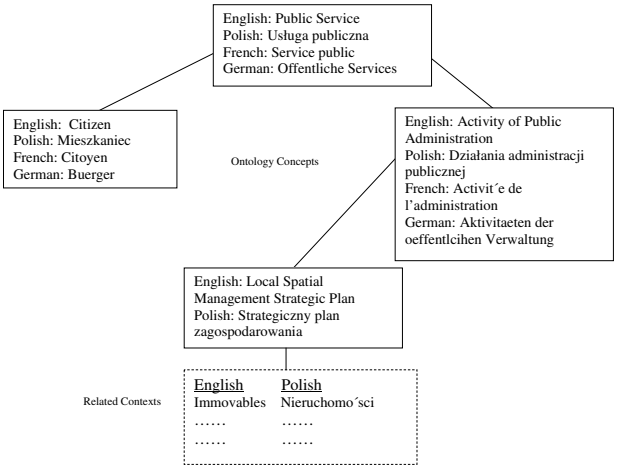

Figure 3: Multilingual Ontology Example

4.1 Examples

The decision on the mapping of contexts to ontology concepts can be made according to the task at hand. We now present three examples to illustrate the benefits of the proposed model.

The first example involves the routing of input, such as an incoming email, to the appropriate place in the organization. Given a distance threshold, t_1, any ontology concept whose context matches an automatically generated context from an email and its distance is lower than the threshold ($d(A, B) < t_1$) will be considered relevant. If such a context has an associated email address, the email will be routed to it.

The second example involves opinion analysis. A relevant set of ontology concepts is identified, as in the case of email routing. The ontology also contains contexts defining various opinions. Such contexts may be globally defined (for the whole ontology) or specific to some concepts. Opinion contexts can be defined in multiple languages. The relative distances of the different opinions of a matching concept are evaluated. If the difference in distance is too close to call (given an additional threshold t_2), the system refrains from providing an opinion. Otherwise, the email is marked with the opinion with minimal distance.

The determination of public agenda is a third task the system can support. If all ontology concepts (of the n relevant concepts) do not exceed the threshold $d(A, B) \geq t_1$, then the email is considered to be part of a new topic on the public agenda and is added to other emails under this concept. Periodically, such emails are clustered and provided to decision makers to determine the addition of new ontology concepts.

Figure 2 presents an example of a multilingual ontology. Each concept is represented by a node with multiple contexts for each language. Figure 3 represents a more specific, detailed example of an ontology. It contains three ontology concepts, namely Citizen, Public Service, and Activity of Public Service. Each ontology concept in English is translated to Polish, French, and German. Next, there are ontology concepts that are relevant only to the local government of Tarnow and therefore they appear only in English and in Polish. Each local government can extend the ontology concepts to include ones that interest it alone and can decide to use existing ontology concepts simply by adding the translation to the local language. Figure 3 represents a graphical example of an ontology created in XML.

Consider the following example of an email in Polish:

Subject: podatek od nieruchomości
Szanowni Państwo,
Zwracam się z prośbą o przesłanie wysokości stawek opłat za podatek od nieruchomości dla osób prawnych w Państwa mieście obowiązujące w latach 2000-2004?
Z poważaniem
Edyta

The context as extracted by the context recognition algorithm include: Podatek, nieruchomości, Polska, and Strona. The first two, translated to value added tax (v.a.t.) and immovables, are very relevant to the topic of the email. The other two contexts, Polska, which is Poland, and Strona, which has multiple meanings such as a party or side, have less relevance to the scenario described in the text.

When mapping the contexts to the ontology, the context of nieruchomości can be identified in the list. Therefore, this document can be mapped to the topic of Local Spatial Management Strategic Plan, which can now be accessed by both English or Polish queries.

The following is an example of a German email:
Strassentheater und das kostenlos und auf hohem iveau. Ein Aushängeschild für Saarbrücken und ein leuchtendes Beispiel für junges wildes Theater, abseits des Mainstreams. Toll.

The context recognition identified the following contexts, described in detail below with the actual values assigned by the algorithm: Art - Perspectives du Theatre (RAHMENPROGRAMM, ORGANISATION, SPIELPLAN, VERANSTALTER): 10 - Long Day School: 1
Abseits des Mainstreams - Perspectives du Theatre: 0 - Long Day School: 0
Oscar Wildes - Perspectives du Theatre: 0 - Long Day School: 0
Jenseits des Mainstreams - Perspectives du Theatre: 0 - Long Day School: 0
Movie - Perspectives du Theatre: 0 - Long Day School: 0
Firma - Perspectives du Theatre (BESUCHER): 1 - Long Day School: 0
Saar - Perspectives du Theatre: 2 - Long Day School: 0
Download - Perspectives du Theatre: 0 - Long Day School: 0
Programm - Perspectives du Theatre (RAHMENPROGRAMM, INFORMATIONEN): 3 - Long Day School:0

The results are represented by the context on the left side. Note that the context is extracted not only in German but also in French. Perspective du Theatre and Long Day School both represent possible ontology concepts. The highest ranking was received by the Perspective du Theatre concept. Note also that some of the contexts are not mapped to any concept. If we look at the concepts that relate to Perspective du Theatre, which include RAHMENPROGRAMM, ORGANISATION, SPIELPLAN, VERANSTALTER, BESUCHER, and INFORMATIONEN, each context can belong to one or more concepts. Therefore, when we examine which of the subtopics related to Perspective du Theatre are relevant to the text, we can classify the document as belonging to the RAHMENPROGRAMM (master program) in Perspective du Theatre which received the total highest score.

5. CONCLUSIONS

In this position paper we proposed a unique combination of a global ontology with a dynamic, easy-to-use context, as a dynamic tool for adopting eGovernment IT tools to a multi-lingual and multi-cultural setting, such as the European Union. We provided examples of the usage of such a solution in the QUALEG project, to enhance the citizen involvement in the public life.

Acknowledgments

The work was partially supported by QUALEG, a European Commission 6^{th} Framework IST project. We also thank the QUALEG partners for their support, collaboration, and contribution.

6. REFERENCES

[1] M. Bunge. *Treatise on Basic Philosophy: Vol. 3: Ontology I: The Furniture of the World*. D. Reidel Publishing Co., Inc., New York, NY, 1977.

[2] M. Bunge. *Treatise on Basic Philosophy: Vol. 4: Ontology II: A World of Systems*. D. Reidel Publishing Co., Inc., New York, NY, 1979.

[3] A. Gal, A. Anaby-Tavor, A. Trombetta, and D. Montesi. A framework for modeling and evaluating automatic semantic reconciliation. *VLDB Journal*, 14(1):50–67, 2005.

[4] J. Kahng and D. McLeod. Dynamic classification ontologies for discovery in cooperative federated databases. In *Proceedings of the First IFCIS International Conference on Cooperative Information Systems (CoopIS'96)*, pages 26–35, Brussels, Belgium, June 1996.

[5] J. McCarthy. Notes on formalizing context. *In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, 1993.

[6] A. Segev. Identifying the multiple contexts of a situation. In *Proceedings of IJCAI-Workshop Modeling and Retrieval of Context (MRC2005)*, 2005.