Personalized Extended Government for Local Public Administrations

Claudio Biancalana[1,2], Alessandro Micarelli[1], and Giuseppe Sansonetti[1]

[1] Roma Tre University, Department of Engineering
Via della Vasca Navale 79, 00146 Rome, Italy
[2] Lazio Innovazione Tecnologica, LAit S.p.A.
Via Adelaide Bono Cairoli 68, 00144 Rome, Italy
claudio.biancalana@laitspa.it,{micarel,gsansone}@dia.uniroma3.it

Abstract. This paper discusses the enterprise organization environment and reports our experience and lessons learned in developing an extension of the traditional virtual enterprise model, we named *personalized extended government (PEG)* model. The aim of such model is to simplify and enhance the effectiveness of e-Government services, by realizing Administration to Administration (A2A) and Administration to Citizen (A2C) processes in a personalized perspective. The features of the proposed model make it suitable for use in local public administrations. As a proof of this, it has been successfully deployed to realize the Italian Open Government Data Portal of Regione Lazio, which allows every citizen to be informed about the employment of public resources on regional territory.

Keywords: personalization, e-Government, virtual enterprise model

1 Introduction

In the late 1990s, the concept of virtual enterprise (VE) model has been introduced [2], where every business organization unit is connected to each other through a data transmission network, in order to explore market opportunities and cooperate, on a temporary basis, to better respond to business opportunities. Hence, a VE can be seen as a heterogeneous network for both enterprises and individuals with integrated cooperation, exploiting ICT technologies and protocols for a specific business process. Over the years, a second model has been developed, substantially similar to the VE model, but based on more stable and long-term agreements: the extended enterprise (EE) model [7]. Recently, an organization model similar to EE has been implemented at a government level and can be recognized in initiatives such as the Italian Open Government Data Portal of Regione Lazio [3]. Such a portal provides a web interface that centralizes access to all open datasets for anyone, in particular for data journalists,

[3] https://dati.lazio.it

public administrations, scientists, and business people. The project has been designed to realize Administration to Administration (A2A) and Administration to Citizen (A2C) processes, and relies on stable and deeply defined agreements. Unlike the EE model, however, the focus is not on business opportunities, but on making e-Government services simpler and more effective. In addition, the need to manage and discover unstructured information has resulted in the gradual awareness of the need to adopt knowledge management systems (KMSs) based on semantic and user-modeling functions. This high degree of similarity allows us to introduce a new definition to refer to the concepts described so far, namely, *personalized extended government (PEG)*, as a personalized and context-oriented extension of EE type. This paper: (1) defines PEG; and (2) provides notes on the design of a KMS that supports PEG.

2 Personalized Extended Government

Hereafter, we refer to personalized extended government (PEG) as an integrated unit of organizations, agreements, protocols and ICT resources able to support public administrations to deploy a context-oriented model to build Administration to Administration (A2A) and Administration to Citizen (A2C) scenarios, in order to simplify and to improve the effectiveness of e-Government services. By analogy with the EE model, we can define the following major PEG model features:

- e-Government service-driven cooperation: A2A and A2C processes are always aimed at providing electronic government services to citizens and businesses, with the goal to simplify and make them more efficient and effective;
- Complementarity: administrations exchange with each other only correct and complete data;
- Process integration and resource sharing: more specifically, data, information and knowledge;
- Interdependence: process integration and resource sharing are carried out according to well defined cooperation agreements.

In order to deploy the PEG organization model, it is necessary to define the following aspects:

- Guidelines for every single participant IT assets integration. This problem is due to different technologies used by every administration and the need to preserve both investments and administration autonomy. For these reasons, it is necessary to define a technological infrastructure that guarantees interoperability regardless of the organization structures and single participant legacy systems;
- Maturity model, which is a structured collection of elements that describe certain aspects of maturity in an organization, for example, to provide a way to define what improvement means for an organization;
- Common governance model through the administrations of all participants and citizens.

Regione Lazio's experience raises an issue: the realization of A2A and A2C isolated processes leads to fragmented knowledge and to a loss of fundamental information used to integrate management relationship between administrations and citizens or enterprises. For this reason, LAit S.p.A. and Regione Lazio have planned a knowledge management system (KMS) design with basic concepts inspired to both EE model and PEG model, in order to develop research ideas in the EE field. The main principles of KMSs should be the following:

- Affordable setup: no more heavy bulked social networks held by central public administrations. As a normal web user can now start a forum or a blog using third party (often free) software, he should also be able to use a web host or a hosting service;
- Accessibility through (semantic?) search engines: in our vision, this is surely something related to the open nature of KMSs, but it would gain some commitment from search engines, which will be able to improve quality of searches through proper indexing of published semantic annotations;
- Scalable open architecture: a given service may explicitly be built upon a KMS, committing to its ontologies and content organization. Vice versa, in an even more open view, independent services may be linked by a given KMS. This would allow users to tag the content of these services according to the OASIS reference ontologies, thus easily putting traditional (non semantic-driven) services immediately into practice. The same process would be applied to standard web pages. People could write web pages directly connected to a KMS making explicit reference to its vocabulary, as embedded RDFa, or they could semantically bookmark an external web page (or annotate part of its content) against that same vocabulary.

2.1 Knowledge Indexing

Before addressing the problem of knowledge retrieval, it is essential to analyze how the system indexes the available information, that is, which representation has been chosen to guarantee an efficient and effective retrieval phase. The requirements are twofold: it is essential that on the one hand knowledge is quickly retrieved by users, on the other hand this knowledge accurately satisfies users information needs in terms of high precision. An indexing system for business companies must also be able to deal with different kinds of information representations, from unstructured documents based on natural language to ontology-based knowledge and relational databases. Moreover, it should provide a comprehensive and homogenous human-computer interface for knowledge retrieval. In order to provide the aforementioned prerequisites, it is necessary to consider different types of information and the degrees of information "richness". Information based on ontological standards, for instance, expresses relationships between typically non-structured information, such as natural language text and metadata. Such metadata usually describe features or classes related to given pieces of information. A typical example is the association between a document and one particular category in a predefined taxonomy. As for information stored in databases, we have an underlying relational model that clearly states the semantic meaning of each piece of information unit, such as price, address, and location, and therefore enables the interpretation/recovery process. In order to define a unique representation that deals with all the different types of available information (i.e., natural language, ontology-based, and databases) we must define a subset of shared features that is possible to generalize, and automatic or semi-automatic methods and techniques for translating information from one of these representations to the internal one. This sort of intermediate representation consists of traditional non-structured information with associated meta-information related to concepts of a taxonomy of the business domain for the given public administration unit (PAU). Briefly, each information unit is classified in a subset of categories from a simplified ontology. Such meta-information can be exploited both in the retrieval phase, to reduce possible ambiguities in the processed information, and to re-organize the knowledge in more efficient ways for further user search activities, such as online hierarchical clustering. Information based on ontological standards does not pose relevant issues. In this case, the source is based on a rich language while our internal representation simplifies some features, such as the kinds of relations between concepts. In our representation, we have relations $\langle u, C\rangle$ where u is the information unit and C is a set of categories in the given taxonomy related to the concept u. We only have IS-A relationships, so it is not hard to extract them from the initial ontology. The selection of the most important concepts from the initial ontology is the only task that knowledge experts have to perform before populating the internal taxonomy. Considering the current amount of unstructured information available within companies, the problem of making such information accessible to users is likely to be the most important issue to solve in our knowledge system. Specifically, having chosen a particular representation for PAU domains, it is necessary to find a technique that allows us to autonomously process the unstructured information and populate the internal knowledge base. In input we have information objects, typically text documents, reports, hypertext pages, etc. These objects are processed for named entity extraction, text segmentation, and text categorization. Given an information object, we initially locate and classify atomic elements in text into predefined categories, such as names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. To this aim, we used named entity recognition (NER) systems based on linguistic grammar-based techniques, statistical models and dictionaries (or gazetteers). NER is a well-known research field, subtask of information extraction, which does not focus on semantic interpretation of languages but on more practical and easier goals, so obtaining excellent results in terms of precision of results. The proper nouns of companies, persons, etc. in output of the NER module are used to increase the weight of these entities during the indexing/retrieval steps. A further step we follow to process the input information is text segmentation. In short, a given document is divided into sequences of words or other similar meaningful units that are separately stored in the knowledge base. This step is useful whenever we have long documents, such as reports or e-books, which cover several different topics (i.e., categories in the internal taxonomy). In order to increase the retrieval precision, it is better to split them into meaningful coherent regions. Our segmentation algorithm is based on the Choi's work [3]. It performs three steps: (1) tokenization, (2) lexical score determination, and (3) boundary identification. Basically, after breaking the document up into a sequence of tokens, we use a similarity measure to analyze the semantic coherence among contiguous text regions. Finally, we determine the boundaries whenever we have relevant variations of the semantic coherence measure. The last process we perform on unstructured information is text categorization. After recognizing semantically coherent information units and - for each unit - its relevant entities, we assign a subset of taxonomic categories to it through text categorization techniques [8]. As for the training phase, we use a subset of documents already categorized by the knowledge expert, performing an *ad-hoc* feature selection that also exploits the aforementioned NER module to assign more weight to terms that correspond to relevant semantic classes, such as proper names and locations. The categorization output is a tuple of couples $\langle c_i, \alpha_i \rangle$, where for each taxonomic category c_i we have a value α_i between $[0, 1]$ that represents the degree of relatedness of the input document to the class. This information is stored along with the document in the knowledge base and it is used during the retrieval and personalization phases, as described in the following section.

2.2 Semantic Querying and Personalization

The most popular paradigm for querying a textual database is to submit short queries. Users express their information needs through a small set of keywords that must be present in the retrieved documents. The retrieval system returns an ordered list of references, based on matching algorithms that assign a relevance weight to each indexed document. In our knowledge base, along with each document, we have a list of assigned categories referenced in the internal taxonomy. In order to exploit this information, the query should include one or more categories that users are interested in. We named these enhanced user needs *semantic queries* [1]. One of the most important problems that occurs while querying a corpus of textual documents is the choice of the right keywords for retrieval. Synonymy (i.e., two words that express the same meaning) and polysemy (i.e., different meanings expressed by one word) of natural language may decrease the recall and precision of the retrieval process [4]. For that reason, we have included a user modeling component to represent the users needs. This component is involved during the querying in order to help disambiguate the meaning of the query terms. Some user modeling strategies have been proposed in the e-Government services field (see, for instance, [6]). The proposed user modeling is based on a concept network paradigm [5] instantiated on the taxonomy of the PAU domain. Concept networks are usually employed as a form of knowledge representation. They consist of graphs of concepts and edges that represent semantic relations between these concepts. We use concept networks to weight which concepts users are more interested in, that is, concepts related to the user needs. In our first prototype, the relations between concepts are not considered.

3 Conclusion

In this paper we have introduced a natural extension of the virtual enterprise model, we called *personalized extended government (PEG)* model, whose aim is to encourage and facilitate the exchange of public domain and community information in a personalized perspective, respecting the public administrations and citizens information needs. The implementation of the proposed model at a government level has represented a strategic roadmap for Regione Lazio ICT Government.

Obviously, the further development of the PEG model first of all involves planning and performing an in-depth experimentation, in order to assess the actual satisfaction of stakeholders. Furthermore, several future developments of all aspects of this work are possible. More specifically, we would like to enhance the effectiveness of the proposed model by exploiting information extracted from social media. In literature, indeed, some works (e.g., see [9]) show how government services can be improved based on user-generated contents found in publicly available social media.

References

1. Biancalana, C., Gasparetti, F., Micarelli, A., Sansonetti, G.: Social semantic query expansion. ACM Trans. Intell. Syst. Technol. 4(4), 60:1–60:43 (Oct 2013)
2. Camarinha-Matos, L.M., Afsarmanesh, H.: The virtual enterprise concept. In: Proc. of the IFIP TC5 WG5.3 / PRODNET Working Conference on Infrastructures for Virtual Enterprises. pp. 3–14. Kluwer, B.V., Deventer, The Netherlands (1999)
3. Choi, F.Y.Y.: Advances in domain independent linear text segmentation. In: Proc. of the 1st North American Chapter of the Association for Computational Linguistics Conference. pp. 26–33. Stroudsburg, PA, USA (2000)
4. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. CACM 30(11), 964–971 (1987)
5. Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User profiles for personalized information access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web, pp. 54–89. Springer-Verlag (2007)
6. Iaquinta, L., Torsello, M.A., Comerio, M., Fanelli, A.M., Semeraro, G.: User segmentation in e-government services. In: Late-Breaking Results, Project Papers and Workshop Proceedings of the 21st Conference on User Modeling, Adaptation, and Personalization, Rome, Italy (June 2013)
7. Oliveira, M., Crowcroft, J., Slater, M.: An innovative design approach to build virtual environment systems. In: Proc. of the Workshop on Virtual Environments. pp. 143–151. ACM, New York, NY, USA (2003)
8. Sebastiani, F.: Machine learning in automated text categorization. ACM Comput. Surv. 34(1), 1–47 (2002)
9. Wan, S., Paris, C.: Improving government services with social media feedback. In: Proceedings of the 19th International Conference on Intelligent User Interfaces. pp. 27–36. ACM, New York, NY, USA (2014)