Mining Collaboration Opportunities to Support Joined-Up Government

Rilwan Basanya[1,2], Adegboyega Ojo[1], Tomasz Janowski[1] and Franco Turini[2]

[1]UNU-IIST Center for Electronic Governance, P. O. Box 3058, Macao
[2]Department of Computer Science, University of Pisa, Italy
{rilwan, ao, tj}@iist.unu.edu, turini@di.unipi.it

Abstract. Governments strive to achieve improvements in delivering public services, developing and implementing public policies, responding to crisis situations, and optimizing the use of public resources, among others. Achieving such goals requires collaboration across different levels and functions of government, and across public and private sectors in a Joined-Up Government. Establishing such collaboration requires information on prospective participants including their goals, resources, processes and services. Such information is rarely available in structured forms e.g. in databases, but instead scattered over government portals, publications and other textual sources. This paper proposes the use of semantic text mining for extracting collaboration-related information (focusing on government collaboration) from unstructured data sources. The proposed solution applies natural language processing techniques supported by the relevant domain and process ontologies. It consists of three steps: 1) extracting process-related information from textual sources, 2) creating process ontology instances from extracted information and 3) mining shared and integrated processes based on process instances and the service goal hierarchy in the domain ontology. The paper describes the rationale of and approach adopted in this research, the progress achieved in implementing step 1, the challenges encountered and how we intend to address them in pursuing subsequent steps.

Keywords: Joined-Up Government, Collaborative Networks, Goal Hierarchy, Data Mining, Text Mining, Process Mining

1 Introduction

Governments face enormous pressures in coping with their regular responsibilities of delivering the required services to their stakeholders – citizens, businesses and governments, and developing effective policy interventions to societal issues in various national and international contexts characterized by financial and economic pressures, increasing terrorism-related threats to national security, and increasing incidents of health pandemics and natural disasters. A common element in various responses by governments to these challenges is the adoption of a holistic, whole-of-government approach. This approach aims at creating a single virtual organizational view [1] from various government and non-government entities, based on

collaboration across different levels and functions of government as well as collaboration with non-state actors from the private and voluntary sectors [2].

Developing collaboration structures and relationships requires coordination and integration of services, processes and information-related assets like databases, applications, devices, hardware and networks, offered by collaborating parties. Specifically, establishing any collaboration requires information on the shared goals, prospective participants, their interests and resources [1][3]. However, such information is rarely available as structured data (e.g. in databases) or from a single source. The information is generally unstructured and found in disparate sources like agency portals and online publications related to the entities of interest. This makes the discovery, analysis and integration of information on prospective collaboration parties difficult. Even when such information is available, integrating this information to achieve the objectives of a joined-up government remains difficult due to various organization issues and in the absence of explicit domain ontologies.

Text mining, a data mining technique for unstructured or textual data, which also draws from various disciplines including information extraction, semantic web and computational linguistics, offers significant promise for automatic extraction of important terms, concepts and patterns. Presently, there are a number of tools for text mining which enable annotation of text with user information. When considering text mining in a specific context, like government-wide collaboration, specific domain structures (e.g. goals, services and processes) would be of interest. In this regard, process mining [4] offers concrete technique for mining process-related structures. However, workflow logs used in process mining are structured.

In this work, we are concerned with textual descriptions of such processes and the services they implement. We aim at investigating how text mining, and particularly semantic text mining (text mining supported by relevant ontologies) could be used in discovering various collaboration structures from their descriptions in various textual sources, provided by government agencies, private sector organizations and other entities within the joined-up government context. Our final goal is to enable semi-automatic discovery of shared and integrated processes to realize collaborations and joined-up government. We adopt ontology-based tagging of terms in collaboration descriptions, using service-oriented process ontology obtained from our analysis and augmentation of existing e-government service models. The tokenized terms are subsequently combined, guided by their meanings expressed in the ontology, into collaboration structures. This paper contributes to the final goal through the synthetic analysis and augmentation of existing e-government service description models.

The rest of this paper is structured as follows. Section 2 presents the background in semantic text mining, information extraction, collaboration networks and joined-up governments. Section 3 explains the problem to be addressed. Section 4 discusses the approach adopted to solve this problem, followed by solution steps in Section 5. Finally, Sections 6 and 7 present discussion and conclusions respectively.

2 Background

A number of foundational concepts are essential in explaining our motivation and research approach. These concepts, drawn from the domains of Text Mining, Collaborative Frameworks and Joint-Up Government are briefly introduced below.

Text Mining: Data mining [5] refers to the extraction of knowledge from large amounts of data. A key step in knowledge discovery, it is an iterative process aimed at identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Text mining intersects with Natural Language Processing (NLP), machine learning, data mining, and information retrieval. Knowledge discovery in text, however, involves intelligent text data analysis aimed at uncovering previously unknown patterns in existing unstructured textual resources [6]. By appropriately integrating techniques from each of these disciplines, useful new methods of discovering knowledge from large textual sources can be developed. For our purpose, we explore such techniques in extracting useful patterns in describing e-government services provided by prospective collaborators. Information Extraction (IE) is a key component in text mining, as it enables the application of data mining to unstructured text corpora to discover predictive rules useful for improving the performance of text mining tasks [7]. Handling the representation of background knowledge in IE is a very important challenge [8] and ontologies are used to supply such knowledge and annotations, as they effectively enable the representation of domain schemas containing technical terms, domain concepts and the relationships between them.

Collaborative Frameworks: Collaboration involves exchanging information, altering activities, sharing resources, and enhancing the capacity of another for mutual benefit and a common purpose. Each participating organization aims to help others become the best that they can be at what they do [9]. Studies in network theory and its effect on inter-agency collaboration [10] have also shown enhanced cooperation among government agencies. For our purpose, we adopt the definition of collaboration as the act or process of "shared creation" or discovery [11] involving the creation of new value by doing something new or differently.

Joined-Up Government: Joined-Up government denotes public service agencies working across portfolio boundaries to achieve shared goals and an integrated government response to pressing issues [12]. The rationale for joined-up government is two-fold [13]: 1) avoiding duplication of efforts through improved information flow resulting in increased situational awareness and enhanced capacity for strategic planning and intelligent decision-making; and 2) harmonization of efforts among all stakeholders recognizing the co-dependent nature of their shared and individual goals, and the required compromise and understanding among collaborators in jointly achieving the goals towards attaining a more agile government enterprise. Achieving such goals must be based upon collaboration across different levels and functions of government, and across public and private sectors. However, the complexities of configuring all possible collaborations among agencies are enormous and require a systematic approach dealing with the various scenarios. We seek an environment where collaboration structures for all agencies – processes, capabilities and competencies – can be extracted and modeled along with the collaboration objective to be achieved, leading to the configuration of a collaborative initiative.

3   Problem

Our major goal is to investigate how data mining can be used to support the discovery of information to build collaborative government services towards joined-up government. Since government services are usually described in textual forms on government portals and in other sources, we employ semantic text mining to identify different parts of service descriptions and on this basis discover opportunities for developing shared and integrated services. Discovering such services will optimize the use of financial, human, data, organizational and technical resources. It will also enable improvements in the delivery of government services.

There are two inherent challenges to this problem. The first is addressing the contextual semantic issues posed by textual sources containing service descriptions. For example, a step such as "Issue license" extracted from two different service descriptions e.g. driving license or construction license, could be very different in terms of implementation complexity. The second is mining processes representing shared and integrated services from process ontology instances, as opposed to using structured event logs which is the standard practice in traditional process mining.

4   Methodology

Our approach is to first model Joined-up Government as a Collaborative Networked Organization (CNO) whose members are agencies involved in shared or integrated service arrangements. We then develop the required ontologies to guide the extraction of information from textual descriptions of services and to represent the extracted information as process models. Two ontologies are required for this purpose: 1) a domain ontology to define and relate e-governance concepts and provide a hierarchy of service delivery goals, and 2) a process ontology to represent extracted process models. Since our process models are service-oriented, the process ontology is developed by extending existing e-government service-related ontologies by additional elements discovered through our review of service descriptions available on six national government portals – South Africa, United States, United Kingdom, Canada, Korea, and Germany. Fig. 1 below shows three major steps of our approach to the problem, with steps 1 and 2 executed concurrently.
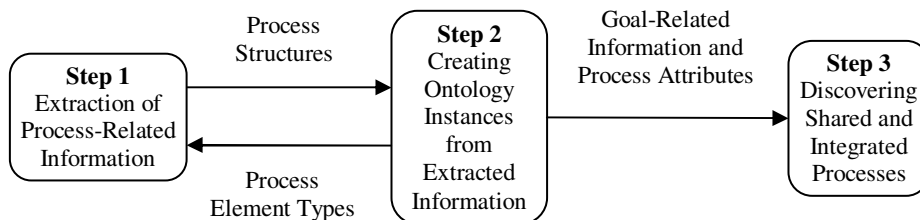

Fig. 1: Three-Step Research Methodology

Using these ontologies, models and recognizers will be developed to identify and tag different elements (e.g. goals, subjects or eligibility criteria) of a service from its descriptions. The extracted information will be stored in the process ontology as process instances. With such instances captured in the process ontology, the final task will be to: 1) discover common sub-processes across process models as a basis for building shared services and 2) suggest possible super-processes as a basis for seamless or integrated services using the service delivery goal hierarchy.

5   Solution

This section describes in some detail the three steps in Figure 1.

Step 1 - Extraction of Process-related Information: This step involves pre-processing process-related descriptions to conform to some standardized and structured process syntax. We gather service descriptions from various government websites and carry out rigorous analysis to extend existing public service description models. This task involves selection of sentences describing key service components and processes. It also involves decomposing the sentences into tokens for the recognizer to identify named entities and co-references. Next, the tokens will be combined into structures based on their meanings in the ontology. Figures 2 and 3 show the examples of two service descriptions on the South African Government's (www.gov.za) and Alberta Government's (www.alberta.ca) portals.

The resulting service-oriented process ontology consists of the following elements: 1) Service – a clear reference to the nature of the service, 2) Type – the type of service based on the authorization, certification, control, and production types [14], 3) Source – the government offering the service, 4) Subject – the actors involved in different roles in delivering the service including service requestor, provider and their collaborators, 5) Eligibility – the rules determining which requestor can receive the services, 6) Evidence – the confirmation of receipt of request to provide a service, 7) Location – the place where interactions in the service delivery process take place, 8) Duration – how long will it take to deliver the service, 9) Legal – the legal provision e.g. government decrees empowering the provider to provide the service and specifying which requestors can receive the service, 10) Payment – fees to be paid for receiving the service, 11) Outcome – the result of the service, 12) Review – what can be done if the requestor is not satisfied with the outcome, 13) Suspension – if a positive outcome can be suspended and for how long, 14) Validity – when a positive outcomes will expire, 15) Contact – who can be reached for enquiries on the service, and 16) Appeal – if and how to challenge the negative outcome of a service request.

Following the development of the process ontology, the next task is to develop recognizers for each element and tag extracted text with the appropriate element. In summary, this step involves: 1) preprocessing, 2) development of service description-based recognizers, and 3) extracting service elements from their textual descriptions.

Fig. 2: Sample Service Description from www.gov.za

Fig. 3: Sample Service Description from www.alberta.ca

Step 2 - Creating Process-Ontology Instances: This step involves developing a formal process ontology based on the service description elements and the domain ontology. The extracted text is created as individuals or instances in the developed process ontology, where each instance represents a complete service description. This is followed by creating relationships between the instances based on the service goal hierarchy defined in the domain ontology. The goal hierarchy specifies how a higher level or conjugate goal is obtained by fulfilling simpler goals in addition to goal prioritization. For instance, the service goal related to Obtaining Licenses can be fulfilled by goals Successfully Filing an Application and Obtaining Decision on the

Application. Therefore, a process that implements the submission of applications and another process implementing decision making on applications could be combined into a single process with the goal of license service delivery.

Step 3 - Discovering Shared and Integrated Services: Our target is to extract two main types of process information structures. The first is processes common to several agencies. For example, most agencies require authorization services to evaluate or decide on applications based on some criteria. The second is integrated processes that could be constructed from the processes of various entities towards fulfilling a higher-level goal. For instance, responding to a health pandemic may require that screening services (checking travelers) at all ports of entry are linked with ambulatory services (for transfer) and with emergency services (treatment) at hospitals.

6 Discussion

So far, our research has focused on the Step 1 – Extraction of Process-Related Information. To this end, we attempted to extract process-related information from the online sources provided by the national governments of United States, United Kingdom, Canada, Germany, Korea and South Africa. However, we were confronted with a number of challenges that further highlight that meaningful extraction, parsing and construction of processes and other higher-level structures from these descriptions require semantic support. Here we describe three specific challenges. The first challenge is non-standard form of service descriptions. We observed that there is no standard form, nor a controlled language, for describing available government services even within the same government. In general, the descriptions present across various government portals use free narrative forms. We also discovered that service descriptions and presentations vary according to the government's level of e-government maturity. For example, some portals like the US federal government's www.usa.gov tailored and personalized their service descriptions to individuals based on their interactions with the portal. The second challenge concerns the existence of huge variations in the details of service descriptions. Some descriptions, for instance, may not contain process-related information. The third challenge is inconsistent naming of services, with the same or related services referenced differently across different government portals. For example, Child Support Grant, Child Care and Development Fund, Child Care Subsidy, and Child Care Benefit all refers to essentially the same services. We are currently working to address these challenges through the techniques employed in resolving semantic conflicts in general.

7 Conclusions

The goal of this paper is to present the research approach and initial implementation of this approach in automating the identification of shared or integrated services to underpin Joined-up Government – one of the most important forms of Collaborative Network Organizations in the public sector domain targeting complex, cross-sector challenges. Unlike existing work on service development and composition that relies

on structured repositories and service descriptions, we seek to identify opportunities for shared and integrated services from non-structured textual descriptions commonly found on government portals and in government publications using semantic text mining. Despite various semantic challenges encountered in the course of this work, we consider the discovery task feasible. In our view, once the implementation of step 1 in our research methodology (extraction of process-related information) is complete, creating ontology instances from extracted information (step 2) and discovering shared and integrated processes (step 3) through ontology-based reasoning will be more tractable. This is the direction of our future work.

References

1. Camarinha-Matos, L., Afsarmanesh, H., On Reference Models for Collaborative Networked Organizations. International Journal of Production Research. 46, 2453-2469 (2008).
2. Stanton, T.H., Improving Collaboration by Federal Agencies: An Essential Priority for the Next Administration. Management (2008).
3. Reillly, Thom, Collaboration in Action - An Uncertain Process. The Haworth Press. (2001).
4. de Medeiros, A.K.A., van der Aalst, W., and Pedrinaci, C., Semantic Process Mining Tools: Core Building Blocks. 16th European Conference on Information Systems, p. 1953–1964. Citeseer (2008).
5. Han, J., Kamber, M., Data Mining: Concepts and Techniques, CA: Morgan Kaufmann, San Francisco (2006).
6. Novak, P.K., Vavpetic, A., Trajkovski, I., Lavrac, N., Towards Semantic Data Mining with g-SEGS. Business, pp. 3-6. (1990).
7. Nahm, U.Y., Mooney, R.J., Text Mining with Information Extraction. Proceedings of the 4th International MIDP Colloquium, Bloemfontein, pp. 141-160. South Africa (2005)
8. Stavrianou, A., Andritsos, P., Nicoloyannis, N., Overview and Semantic Issues of Text Mining. ACM SIGMOD Record. 36, 23 (2007).
9. Himmelman, A.T., Collaboration for a Change: Definitions, Decision-Making Models, Roles, and Collaboration Process Guide. Available through Partnering Intelligence at http://www. partneringintelligence. com/documents/5.02 _Collaboration% 20for% 20a% 20Change. doc.
10. Bardach, E., Can Network Theory Illuminate Interagency Collaboration. Workshop on Network Analysis and Innovations in Public Programs, LaFollette Institute of Public Affairs, University of Wisconsin-Madison, (1994).
11. Thomson, A.M., Perry, J.L., Collaboration Processes: Inside the Black Box. Public Administration Review (2006).
12. Christensen, T., Lægreid, P., The Whole-of-Government Approach to Public Sector Reform. Public Administration Review. 67, 1059–1066 (2007).
13. Baker, J., Quick Impact Projects: Towards a Whole of Government Approach. Paterson Review. 8, 1–21. (2007).
14. Ojo, A., Janowski, T., Estevez, E., Domain Models and Enterprise Application Framework for Developing Electronic Public Services. Proceedings of the 6th International EGOV Conference, Regensburg, p. 157–164. Germany (2007).