PAPER AND DEMO

Title: E-Government: Human-Centered Systems for Business Services

Authors:[1]

Nabil R. Adam	CIMIC-Rutgers University	adam@adam.rutgers.edu
Francisco Artigas	Rutgers University	artigas@cimic.rutgers.edu
Vijayalakshmi Atluri	Rutgers University	atluri@cimic.rutgers.edu
Soon Ae Chun	Rutgers University	soon@cimic.rutgers.edu
Sue Colbert	Office of IT, State of NJ	aebied@oit.state.nj.us
Melania Degeratu	Columbia University	melania@cs.columbia.edu
Adel Ebeid	Office of IT, State of NJ	aebied@oit.state.nj.us
Vasileios Hatzivassiloglou	Columbia University	vh@cs.columbia.edu
Richard Holowczak	City University of NY	richard_holowczak@baruch.cuny.edu
Odysseus Marcopolus	Office of IT, State of NJ	omarcopolus@oit.state.nj.us
Pietro Mazzoleni	University of Milan	mazzolen@cimic.rutgers.edu
Wendy Rayner	CIO, State of NJ	wwr@capitol.statehouse.state.nj.us

Topic Areas:

Cooperation among State Government agencies
Delivery and tracking of public services

Contact Author:

Vijay Atluri, MSIS Department, Rutgers University, 180 University Avenue, Newark, NJ, 07102.
email: atluri@cimic.rutgers.edu
Phone: 973-353-1642

Web Page:

http://cimic.rutgers.edu/dgov

[1] Soon Ae Chun, Melania Degeratu and Pietro Mazzoleni are student authors

1 Introduction and Overview

Advances in information technologies have positioned electronic commerce (EC) at the forefront of business and commerce [1]. EC has also come to the forefront at federal and state agencies, many of whom have been mandated to adopt EC technologies to help streamline and transform their core business processes. In response to these mandates, the NJ State government has recently established the E-Government Services (eGov) division in the Office of Information Technology (OIT). The goal of this initiative is to redefine the way in which government, citizens, and businesses exchange information, goods, and services by shifting the NJ government services paradigm from a labor-intensive, agency-focused delivery system to the one with immediacy and cohesiveness with no barriers of time or distance. An integral portion of this initiative is to develop an effective and efficient *human-centered e-government portal* to support small and medium sized enterprises in New Jersey that fosters the establishment of new businesses and facilitates effective long-term interactive relationships with them. Under NSF Digital Government Initiative funding, Rutgers-CIMIC has been collaborating with the State of New Jersey in designing and building a prototype for a business Portal. Presently, establishing a business in New Jersey (or in any other state) is an arduous task fraught with pitfalls due to the wide range of agencies that a prospective business must interact with. Thus, our system is geared towards addressing the following pressing questions posed by any enterprenuer.

1. What information is available for my business?: Many state agencies (over 60 in NJ alone) have established web sites that offer information about the services provided. The information offered comes in diverse forms: data gathered from forms, local maps, census data, unemployment data, traffic patterns, environmental data, real-estate, the co-location of potential customers and the co-location of competing or complimentary businesses, the availability of the labor pool, appropriate transportation corridors, the cost of renting, buying or building office space, a showroom, a storefront, etc., the state and local tax rates, zoning restrictions, government incentives, and many others, depending on the nature of business. Data gathered from forms submitted to these agencies are seldom exchanged. In our portal, the scattered information and services are aggregated and only a relevant subset are identified and tailored for each user's personal and business requirements.

2. Which agencies are relevant for my business registration process and in which order should I visit them?: Registering a new business requires interacting with at least the following different state agencies: The Division of Taxation, Div. of Unemployment and Disability Insurance, Div. of Worker's Compensation, The Compensation Rating and Inspection Bureau, Div. of Commercial Recording, and the local County Clerk's office (for trademark and other local regulations). Some tasks (e.g. obtaining various permits) must often be obtained in a specific sequence, and knowledge of the appropriate sequence is frequently inexplicit and not shared among agencies.

3. Why do I have to submit redundant information?: Opening a new business requires submitting many forms and data associated with them to different agencies. Once agencies and their respective forms have been identified, they must be filled out and faxed or mailed in. In most cases, such forms overlap considerably in terms of the basic level of data requested. Once this data is acquired by an agency, it is unlikley to be shared with other agencies.

A system that addresses these concerns must be capable of guiding entrepreneurs through the various processes involved in a specific order (*workflow*), and must bring together the wide range of state agencies that an entrepreneur must interact with (*interoperability*). In addition, it should facilitate a secure transaction environment and provide a means to authenticate entrepreneurs.

2 The NJ Business Portal System

In this section, we present the business portal system - its architecture (Figure 1), each of its components, and the status of our research prototype. While describing each component, we identify the research problems and describe our approaches to addressing them.

2.1 Decentralized Dynamic Workflow Management System

Registering a new business in the State requires the collaboration and coordination of various state agencies as they execute their relevant tasks. Although the core basic tasks of business registration remain the same, many tasks may diverge from this basic set, depending on the user's personal and business profiles (e.g. business types and varieties and its location may require different set of permits and licenses).

To guide entrepreneurs through the process of establishing their business, we employ a workflow system that is capable of (1) generating a *customized workflow* that is *dynamic* in nature, based on user parameters, requirements and constraints; (2) automatically executing the different processes involved at appropriate agencies in the appropirate sequence by authorized individuals, adhering to the business policies of those agencies; and (3) visually reporting the progression of the workflow to the entrepreneurs.

The *Customized workflow Generation module* is responsible for the automatic generation of a customized workflow for each user, by taking into account the *services and tasks* and their *business rules* publicized by each State agency, and the *user profiles*. The user profiles are captured via a Web based user interface using rule-based interview sessions.

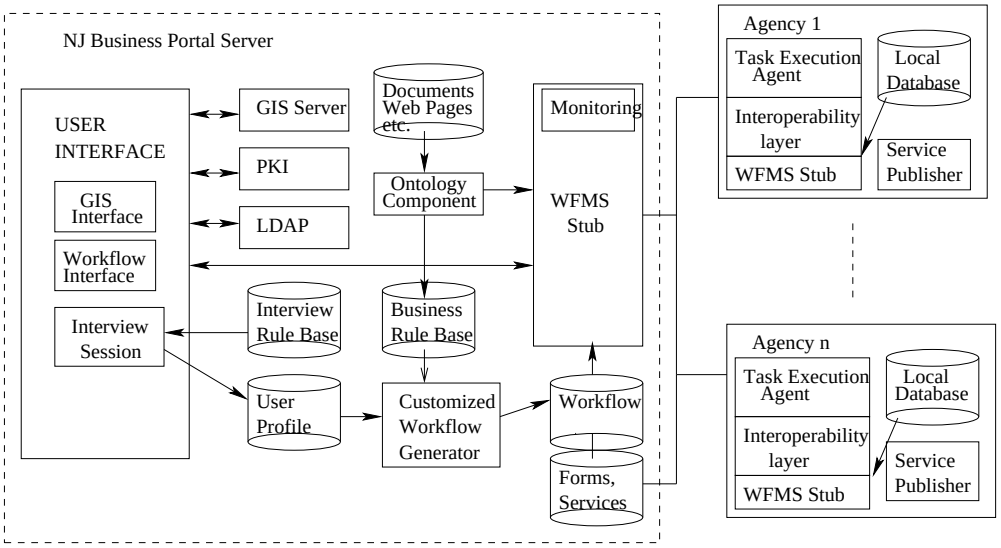

Figure 1: NJ Business Portal Architecture

In this project, we are developing a *decentralized dynamic workflow management system* (DDWFMS) that eliminates the need for a centralized WFMS which has known problems such as limited scalability and performance bottlenecks. Our DDWFMS relies on (1) on-the-fly construction of *self-describing workflows* and (2) their execution with the help of *WFMS stub* located at each agency. A self-describing workflow is a partition of the workflow that is associated with the tasks' definitions, the agencies responsible for their execution, and the control flow information. When an agency receives a self-describing workflow, the WFMS stub extracts the tasks to be executed at that location, evaluates the control flow conditions upon their completion, constructs the new self-describing workflow(s), and forwards them to the appropriate agencies. The WFMS stub is also responsible for forwarding the dynamic changes that may occur during the execution of the workflow. In our prototype, we have used interactive Adobe Acrobat Portable Document Format (PDF) forms to capture the user profiles, from which File Data Format (FDF) data is extracted. The FDF is then used to generate the customized workflow, which is shown graphically to the users with its execution status information (c.f. demo). To adopt the decentralized workflow execution, we have provided a range of alternatives to the agencies, based on their existing infrastructure.

2.2 Data interoperability layer

In order to support the distributed workflow, we require: (1) the capability to perform data interoperability between different state agency information systems application programs and (2) the ability to automatically mod-

ify existing programs in case of changes to related schema and necessary application programs upon addition of new information systems. Our approach to providing light-weight data interoperability employs eXtensible Markup Language (XML). Since we have not implemented the interoperability services at this time, we do not provide further details.

2.3 Ontology: Common Language across Agencies

Effective communication between agencies and workflows handling data from multiple State agencies requires a common language, or at least a mechanism for representing semantic differences. Further, knowing important relationships between concepts and terms in the domain (e.g., the various subtypes of a business entity) is crucial for designing workflows that customize processing according to the appropriate user circumstances and validate the data entered by the user. Rather than manually creating the appropriate workflows and entering constraints on data fields by hand, we are relying on a domain ontology that encapsulates the concepts in the small business initiation domain, the terms that are used to represent these concepts, and the relationships between them.

Ontologies have been used as repositories for concepts and relationships in several applications, and even large scale ontologies such as SENSUS [3] with more than 90,000 entries have been constructed. Usually the construction and expansion of the ontology entails significant manual labor; knowledge engineers together with domain experts collect the concepts and populate the ontology with definitions and relationships. The resulting ontology, while still useful for applications, is expensive to develop, maintain, or expand to other domains. We take a different approach, focusing instead on the automatic acquisition of ontological information from raw, unnanotated text. We apply text mining techniques to discover important terms, organize the terms into concepts, and identify several types of relationships between concepts, generalizing recent work such as [4].

Our early efforts in this area have focused on identifying recurring text templates that connect terms in a consistent fashion. We start from a small pilot collection of HTML documents (24,583 words) provided by the New Jersey Business Services Gateway (`http://www.state.nj.us/njbgs/`). We parse the HTML content in these documents, automatically assign part of speech information, and detect terms with a statistical tool [2]. We subsequently examine the text for patterns involving two terms and a linking verb, as well as patterns involving one or more term and specialized constructs such as apposition and parenthetical expressions. manually constructed. We intend to automate the acquisition of those patterns in the future.) Instantiated patterns are collected across texts and validated using a frequency-based test. Patterns that are consisted and frequent enough are classified (using manually derived, at this point, clues) into one of several types (*is-a*, *has-part*, *has-attribute*) that our system recognizes and added to the ontology database.

Even from our small pilot collection of documents, we have extracted several hundreds of interesting patterns and relationships. These include, for example, the ontology fragment, which captures that a *limited liability partnership*, which is a particular type of *business entity*, must have a *business location*, which in turn is a subtype of the general concept *address*; the latter, in turn, has subfields such as *street*, *number*, and *zip code* that *business location* inherits. Such relationships will be automatically used to guide workflow creation and adaptation to personalized user needs; for example, by automatically detecting the valid choices for business subtypes and presenting to the user only the appropriate data fields for the business type she is trying to create.

Our work is moving to a large-scale replication of our pilot pattern collection experiments, using the full New Jersey Complete Registration Package collection. We plan to focus on automating the components of the system (such as the relationship classifier) which currently use a pre-constructed model, replacing such models with models that expand and adapt to new data. As multiple patterns suggest potentially conflicting information, we are working on methods for automatically detecting erroneous relationships extracted during the text mining phase and separating those from genuine differences between sources or agencies. We are also looking at issues coming into play when data from broader, uncontrolled sources and of uncertain quality is processed by our system.

2.4 Large Scale Public Key Infrastructure

To foster secure communications between members of large service organizations through the web, the NJ Office of Information Technology has implemented a Directory Service based upon the Lightweight Directory

Access Protocol (LDAP) that is coupled with a Public Key Infrastructure (PKI) that includes digital certificates. This infrastructure allows members to create and maintain a detailed profile about themselves, to search the Directory listings, to develop groups for communication and document sharing, and to publish their public key for use in secure communications among members. A registrant has to maintain a certificate key pair. The public key is stored within the member's listing in the LDAP Directory, and the private key is securely stored in the member's browser/mail client. The State serves as a Certification Authority that issues and verifies the identification, creates and issues a recognized and trusted certificate, and maintains a revocation list of certificates. An entrepreneur has the ability to create his or her profile using the LDAP directory Service and is issued a certificate and public/private key pair by the State. This allows the new business to interact with State agencies and other organizations in a secure and trustworthy manner by digitally signing forms and documents submitted to the various agencies. The current PKI and LDAP infrastructure has the capacity to scale up to millions of users.

2.5 GIS services

Location information by means of a functional Geographic Information System (GIS) provides important assistance to business entrepreneurs in at least two fronts. First, an entrepreneur may be better prepared to make a siting decision for his or her business if he/she has the appropriate geographic information for the area being considered. Information such as the location of currently established businesses, traffic movement statistics, boundaries of economic incentive zones, census statistics on household income, education, average size and age distribution of family units and the ethnic makeup of neighborhoods are all factors that may want to be considered. We have developed an interactive mapping technology that provides the ability to effectively deliver this information to the user via the WWW.

Second, the state and local governments have strict zoning and permitting regulations to prevent pollution, health hazards and destruction of wildlife. These regulations based on geographic coordinates impose restrictions in the development of certain sensitive areas. In all cases, one must have a permit to excavate, dredge, fill or place any structure on any of the regulated areas as described in geo-registered state and local maps. Geographic coordinates on the surface of the earth where there are restrictions that require special permits are obtained from state and local maps and entered and managed by a GIS system. When a workflow for permit is requested by a user, the geographic coordinates of the area being proposed (identified by clicking on an interactive map) are examined to determine if there are any restrictions associated with those particular coordinates. If there are restrictions (e.g. need a permit to excavate because the coordinates correspond to a location classified as a wetland in the state wetlands map), the system then generates a workflow that takes these restrictions into consideration and adds the appropriate tasks to the workflow. Currently we are working on this module.

References

[1] N. Adam, O. Dogramaci, A. Gangopadhyay, and Y. Yesha. *Electronic Commerce: Technical, Business, and Legal Issues*. Prentice Hall, 1999.

[2] Steve Finch and Andrei Mikheev. A workbench for finding structure in texts. Washington, D.C., April 1997.

[3] Eduard H. Hovy. Combining and standardizing large-scale, practical ontologies for machine translation and other uses. In *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC)*, Granada, Spain, 1998.

[4] Mark Sanderson and Bruce Croft. Deriving concept hierarchies from text. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 1999.