Ontology-based Support for Digital Government*

Athman Bouguettaya[1] Ahmed Elmagarmid[2] Brahim Medjahed[1] Mourad Ouzzani[1]

[1]Department of Computer Science
Virginia Tech
{athman,brahim,mourad}@vt.edu

[2]Department of Computer Sciences
Purdue University
ake@cs.purdue.edu

1 Introduction

Digital government can be defined as the civil and political conduct of government using information and communication technologies [4, 6]. The Web has laid, among other emerging technologies, a broad foundation for digital government. A recent study conducted by *The Council for Excellence in Government* indicates that "66% of Americans favor the President's appointing a high-level official who would oversee and encourage government use of the Web and other technologies to make government services and information more readily available to the public" [2].

The development of techniques to efficiently access government databases and services is at the core of our research in digital government. For that purpose, we have teamed with the *Indiana Family and Social Services Administration* (FSSA) to help the needy citizens collect social benefits. FSSA serves families and individuals facing hardships associated with low income, disability, aging, and children at risk for healthy development. To expeditiously respond to the citizens' needs, the FSSA must be able to seamlessly integrate geographically distant, heterogeneous, and autonomously run information systems. In addition, FSSA applications and data need to be accessed through one single interface: *the Web*. In such a framework, case officers and citizens would transparently access data and applications as homogeneous resources.

The FSSA is composed of dozens of autonomous departments located in different cities and counties statewide. Each department's information system consists of a myriad of databases. To access government information, case officers first need to locate the databases of interest. This process is often complex and tedious due to the heterogeneity, distribution, and large number of FSSA databases. To deal with this problem, FSSA databases are organized as *distributed ontologies*. An *ontology* defines a taxonomy based on the semantic proximity of information interest [3]. Each ontology focuses on a single common information type (e.g., disability). It dynamically groups databases into a single collection, generating a conceptual space with a specific content and scope. The use of distributed ontologies elicits the filtering and reduction of the overhead of discovering FSSA databases.

Several rehabilitation programs are provided within FSSA to help disadvantaged citizens. Currently, FSSA case officers must deal with different situations that depend on the particular needs of each citizen (disability, children health, housing, employment, etc.) For each situation, they must typically delve into a potentially large number of applications and determine those that best meet the citizens' needs. Examples of FSSA applications include ICES (*Indiana Client Eligibility System*) and ISETS (*Indiana Support Enforcement Tracking System*). To facilitate the process of collecting benefits, we wrapped each FSSA application with an *e-service*. Simply put, an *e-service* is an application functionality that can be programmatically invoked from the Web [5]. Similarly to FSSA databases, we organized the e-services space into *distributed ontologies* (*vocabularies*). Each vocabulary is composed of a set of attributes that describe the basic properties of FSSA e-services. The use of e-services caters for a dynamic discovery of FSSA applications.

The uniform use of distributed ontologies is the core of our approach. We propose an infrastructure that enables uniform access to a large number of government databases and e-services. A prototype called *WebDG* (*Web Digital Government*) has been implemented. In Section 1, we present our ontological approach to organize government databases. In Section 3, we describe the modeling of government applications as e-services. In Section 4, we provide an overview of the architecture and implementation of *WebDG*. In Section 5, we provide some concluding remarks.

*This research is supported by grant 9983249-EIA from the National Science Foundation's Digital Government Program. The fourth author is partly supported by a USENIX fellowship.

Proceedings of the 27th VLDB Conference, Roma, Italy, 2001

2 Ontological Organization of Government Databases

Part of FSSA case officers' task is to interact with local and state agencies to extract government information. However, the large number of FSSA databases makes it extremely difficult to query the available information space. To tackle this problem, we segmented FSSA databases into *distributed ontologies*.

Ontologies describe coherent slices of the information space. Databases that store information about the same topic are grouped together. For example, all databases that may be of interest to disabled people (e.g., *Medicaid* and *Independent Living*) are members of the ontology *Disability* (Fig. 1). For the purpose of this project, we have identified eight ontologies within FSSA, namely *Family*, *Visually Impaired*, *Disability*, *Low Income*, *At Risk Children*, *Mental Illness and Addiction*, *Health and Human Services*, and *Insurance*. A representative sample of these ontologies is presented in Fig. 1. In this proposed framework, individual databases join and leave the formed ontologies at their own discretion. An overlap of two ontologies depicts the situation where a database stores information that is of interest to both them. For example, the *Medicaid* database belongs to *Family*, *Visually Impaired*, and *Disability* ontologies.

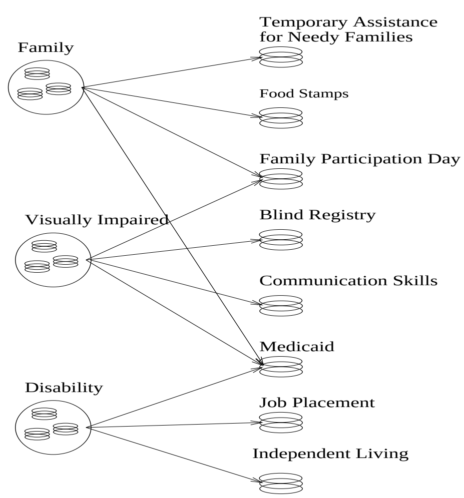

Figure 1: Sample FSSA Ontologies

The FSSA ontologies are not isolated entities. They are related by *inter-ontology relationships*. These relationships are dynamically established based on users' needs. They allow a query to be resolved by member databases of remote ontologies when it cannot be resolved locally. The inter-ontology relationships are initially determined statically by the ontology administrator. They essentially depict a *functional* relationship that would dynamically change over time.

Locating databases that fit users' queries requires detailed information about the content of each database. For that purpose, we surround each FSSA database by a *co–database*. A *co–database* is an object-oriented database that stores information about its associated database, ontologies and inter-ontology relationships. A set of databases exporting a certain type of information (e.g., *disability*) is represented by a class in the co–database schema. This class inherits from a pre-defined class, *OntologyRoot*, that contains generic attributes. Examples of such attributes include *Information-type* (e.g., "Disability" for all instances of the class *Disability*) and *Synonyms* (e.g., "Handicap" is a synonym of "Disability"). In addition to these attributes, every subclass of the *OntologyRoot* class has some specific attributes that describe the domain model of the underlying databases. For example, a subset of the class *Disability* is defined in the following:

Class Disability *ISA* OntologyRoot {
  attribute String County;
  attribute Person Citizens;
  ....
}

3 Discovering Government Services

The process of collecting social benefits within the FSSA is currently time-consuming and performed *ad-hoc*. Indeed, case officers must manually execute different FSSA applications. Since the number of applications is large, locating the ones that best fit the citizens' needs is usually a cumbersome task. Furthermore, the applications are distributed over different FSSA departments which makes this task even more complex. To tackle this problem, we wrapped each FSSA application with an *e-service*. The main benefits of adopting the e-services approach are the following: First, the Web is ideal as a one-stop interface for both case officers and needy citizens. Second, e-services allow the use of pre-existing (legacy) FSSA applications without requiring their modification. Third, they allow the low overhead bridging of the heterogeneity of FSSA applications. Heterogeneity occurs at different levels including the interface, programming language, business logic, and back-end systems. Fourth, they cater for the dynamic discovery of FSSA applications.

To facilitate e-service discovery, we defined *ontologies* (*vocabularies*) for FSSA e-services. Each vocabulary is composed of a set of attributes that describe the basic properties of FSSA e-services. A representative of the vocabularies defined for our project is given in the following:

Location
  Division: String
  Bureau: String
  City: String

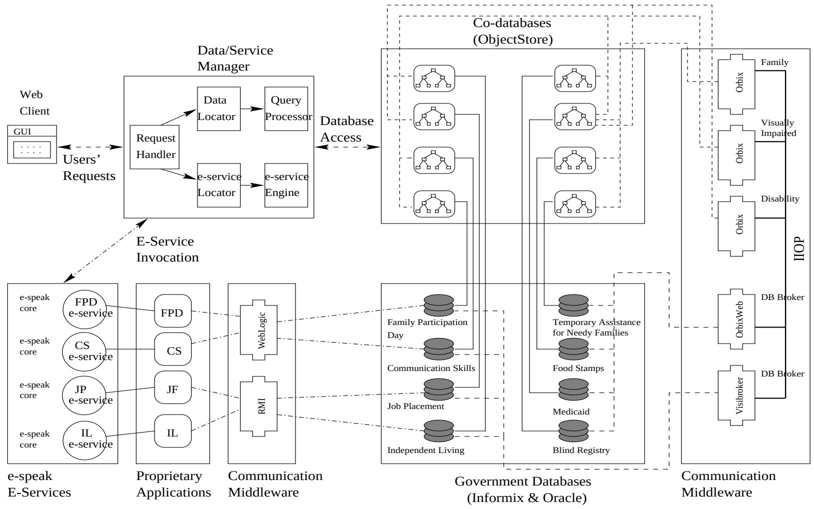

Figure 2: WebDG Architecture

Topic

Addiction: Boolean
Adoption: Boolean
Children: Boolean
Courses: Boolean
Disability: Boolean
Elderly: Boolean
Family: Boolean
Food: Boolean
Hard of Hearing: Boolean
Health: Boolean
Housing: Boolean
Insurance: Boolean
Jobs: Boolean
Pregnancy: Boolean
Visually Impaired: Boolean

Assistance

Cash: Boolean
Supportive: Boolean

FSSA e-services are advertised using XML and according to the defined vocabulary. The use of XML is a natural fit for *WebDG* because of its growing popularity as a standard for encoding and exchanging information on the Web. In the following, we give a subset of an XML document used to advertise FSSA e-services:

```
<?xml version="1.0"?>
   .......
   <attr name="Bureau" required="true">
       <value >Vocational Rehabilitation </value>
   </attr>
   <attr name="Disability" required="true">
       <value >true </value>
   </attr>
```

The aforementioned attributes are of three types: *location-based*, *topic-based*, and *assistance-based*. *Location-based* attributes allow to specify the division (e.g., family and children, mental health) and bureau (e.g., child support, disability determination) the e-service belongs to. Citizens and case officers can also query e-services based on the city the e-services are offered in. For example, a disabled citizen may be interested in registering for an *independent living* course only if this course is offered within her/his city. *Topic-based* attributes allow the specification of the type of e-services users are interested in. For example, a single mother can get the benefits she is entitled to by selecting the *family* and *children* attributes. *Assistance-based* attributes allow the specification of the type of assistance users prefer. This assistance would be either in cash (e.g., TANF– *Temporary Assistance for Needy Families*) or in kind (e.g., WIC– *Women Infant and Children*).

4 Implementation

In Fig. 2, we present the global architecture of the *WebDG* prototype. *WebDG* enables the access to databases and services within three FSSA divisions: *family resources*, *vocational rehabilitation*, and *blind and visually impaired*. Each division's information system uses *Informix* and *Oracle* databases. For the purpose of this project, we considered four rehabilitation programs, namely FPD (*Family Participation Day*), CS (*Communication Skills*), JF (*Job Finder*), and IL (*Independent Living*).

Two types of requests are supported by *WebDG*: querying databases and invoking FSSA applications. All requests are received by the *data/service manager* (Fig. 2). The *request handler* is responsible for routing requests to the *data locator* or the *service loca-*

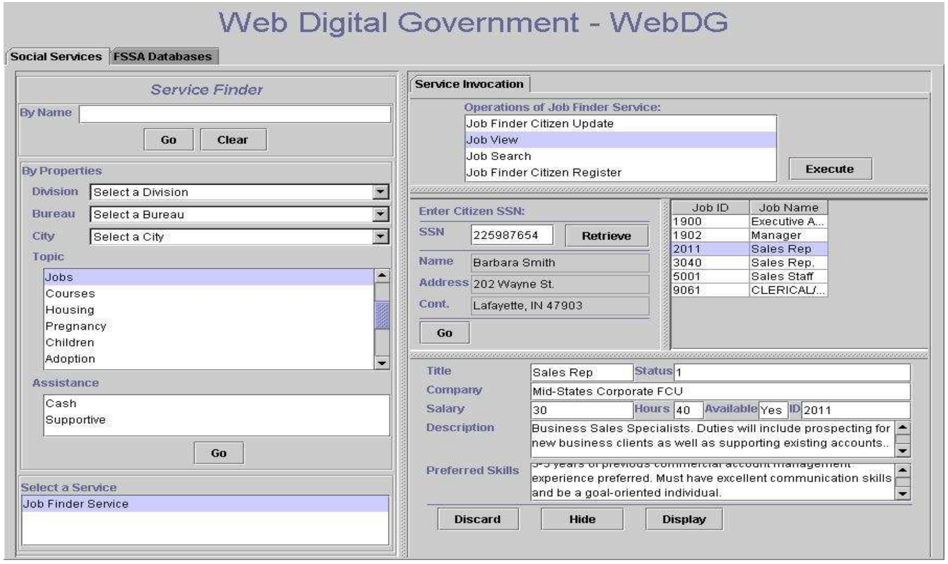

Figure 3: WebDG Interface

tor. Querying requests are forwarded to the *data locator*. Its role is to educate users about the information space and locate relevant databases. All information necessary to locate FSSA databases is stored in *co–databases* (*ObjectStore*). The co–databases are linked to three different Orbix ORBs (one ORB per ontology). Users can learn about the content of each database by displaying its corresponding documentation in HTML/text, audio, or video formats. Once users have located the database of interest, they can then submit SQL queries. The *query processor* handles these queries by accessing the appropriate database via JDBC gateways. Databases are linked to OrbixWeb or VisiBroker ORBs. Since we are using CORBA 2.0 compliant ORBs, all inter-ORBs communications are handled by the IIOP protocol.

The *e-service locator* and *engine* components handle application invocations. The role of the *e-service locator* is to educate users about the application space and locate e-services of interest. For an early deployment of our prototype, we have used HP's *e-speak* [1], an e-service platform to define and locate e-services. We are currently investigating the use of other emerging e-services standards. These include the *Web Services Description Language* (WSDL) and *Universal Description Discovery and Integration* (UDDI).

Locating e-services can be performed in two ways. The first way is by specifying their name (Fig. 3). This is mainly useful for "expert" users of the system. The second way is by specifying the e-service properties as defined in the FSSA vocabulary. Once users have located the e-service of interest, they can directly interact with the e-service by invoking its operations through the *service engine*. Examples of operations for the *Job Finder* e-service include searching jobs (*Job Search*) and displaying selected jobs (*Job View*).

5 Conclusion

We presented in this paper an infrastructure for accessing government data and services. We used distributed ontologies as a basic model for organizing and dynamically discovering a large number of government databases and e-services. A prototype has been implemented and is available on-line (http://www.nvc.cs.vt.edu/˜dgov/demo).

References

[1] *The E-speak Developer Site, Hewlett-Packard (HP).* http://www.e-speak.net.

[2] *E-Government: The Next American Revolution.* http://www.excelgov.org, January 2001.

[3] A. Bouguettaya, editor. *Introduction to the Special Issue on Ontologies and Databases*, Distributed and Parallel Databases Journal, 7(1). Kluwer Publishers, July 1999.

[4] A. Bouguettaya, M. Ouzzani, B. Medjahed, and J. Cameron. Managing Government Databases. *IEEE Computer*, 34(2), February 2001.

[5] F. Casati, U. Dayal, and M. C. Shan. Report on the VLDB Workshop on Technologies for E-Services. *ACM SIGMOD Record*, 29(4), December 2000.

[6] A. Elmagarmid and W. J. McIver. The Ongoing March Toward Digital Government. *IEEE Computer*, 34(2), February 2001.