Chapter #

Supporting Data and Services Access in Digital Government Environments

Athman Bouguettaya[1], Mourad Ouzzani[1], Brahim Medjahed[1]
Ahmed Elmagarmid[2]
[1]*Department of Computer Science, Virginia Tech*
[2]*Department of Computer Sciences, Purdue University*

Abstract: We describe a Web-based architecture, named *WebDG*, to support emerging Digital Government applications. We use an *ontological* approach to organize government data and services (applications). The data space is segmented into logically inter-related clusters of databases to accelerate *metadata* and data *discovery*. Equally, services are segmented into *vocabulary*-based taxonomies to provide a fast mechanism for their retrieval and enactment. As a proof of concept, we use government social services, namely, the Indiana Family and Social Services Administration (FSSA), as a case study for the deployment of our novel techniques. The implementation uses several state-of-the-art, Web-based middleware and e-services technologies as building blocks for *WebDG*. These include, CORBA, EJB, and HP's e-speak products.

Key words: Database, E-Service, Ontology, Digital Government

1.INTRODUCTION

The use of databases has taken a new level of importance with the widespread use of the Web. The efficient and effective use of databases has as a result become a central target to government services. The considerable growth in the US economy has largely been possible because of the increasing use of information technology. The use and management of data plays a central role in the created dynamics. Because of the strategic importance of data management in the context of the Web, two recent reports underlined the strategic importance of Internet-based technologies that ought to be supported by government agencies. The first report, a government sponsored report to tackle the issues facing the nation at the dawn of the 3rd millennium, called IT^2 [Gro99], stresses that the efficient access of data on the Web is of utmost importance to both citizens and government agencies. The second report submitted to the President of the United States, called the PITAC report [JK99], and prepared by eminent US scientists makes a very strong argument in investing in strategic IT areas. Specifically, they make a strong statement with regard to investing in data management.

The information revolution has led organizations all over the world to heavily rely on large numbers of databases to conduct their everyday business. These databases are usually deployed in wide-area network based environments. In such highly dynamic environments it has been very difficult to exert any formal control over the changes taking place in the information space (e.g., registration of new information sources), let alone elicit cooperative tasks. The most prominent application requiring this type of environment is Digital Government. To really take advantage of the abundant connectivity generated by the widespread use of the Internet, there is a need to empower both novice and expert users. This requires a sophisticated infrastructure that can support flexible tools to manage the description, location and access to Internet/Web databases.

We propose the use of *distributed ontologies* of information repositories. In a large network of autonomous databases that would potentially span the globe, a meaningful organization and segmentation of databases would have to be based on simple ontologies that describe coherent slices of the information space. This meta-information would represent the domain of interest of the underlying information repositories. In this approach, collections of databases that store information about the same topic are grouped together. These ontologies are used to capture the structure and semantics of the information space. In general, they act as a global conceptual schema against which queries are formulated. We propose to investigate the design and implementation of these distributed ontologies for government databases in the context of the Web. The result is a Web-based infrastructure, called *WebDG* that elicits the transparent access to government databases where each database would be presented to the user as one homogeneous element of a much larger database.

Government welfare and social services within the Indiana *Family and Social Services Administration* (FSSA) are used in this research as a case study. This government agency consists of dozens of autonomous departments providing services to needy and indigent citizens. The current process of serving citizens is inefficient. In addition, it is costly to both the agency and citizen. One of the major challenges facing this agency has been the seamless interoperation of multiple, isolated, heterogeneous, and autonomous information systems. We have teamed up with the Indiana Family and Social Services Administration (FSSA) to help move their database management technology to a level where the FSSA and the citizens would be served effectively and efficiently.

The remaining of this chapter is organized as follows. In Section 2, we discuss issues and challenges in moving government to the Web era. Section 3 describes the FSSA's information systems, services, applications, and goals. In Section 4, we present an ontology-based approach for databases

and applications (services). In Section 5, we present the deployment of WebDG, a Web-based Digital Government System. We describe the implemented architecture and an accompanying scenario.

2.DIGITAL GOVERNMENT

Governments are by far the most complex organizations in a society. They provide the legal, political, and economic infrastructure to support the common daily needs of citizens and businesses. Governments consist generally of a large and complex network of institutions at different levels, i.e., local, tribal, state, and federal. This opens tremendous opportunities for information technology to profoundly improve and transform the way governments function. In this context, Digital Government has emerged as a distinct research area. Digital government or e-government is defined as the civil and political conduct of government using information and communication technologies [EM01]. Some of the most important issues in enabling Digital Government are reported in different reports and workshops [DBK99, EM01, JK99]:

- **Data Integration**: Governments usually store large amounts of information over distributed, autonomous and heterogeneous databases. The objective would be to provide an efficient integrated access to this information. It should be accessible to both lay citizens and experts. Different aspects need to be investigated including ontological integration, middleware support, and query processing.
- **Scalable Information Infrastructure**: A Digital Government infrastructure should be able to scale with respect to the growing numbers of underlying systems and users (potentially all citizens and businesses). Adding a new information system should not require major modification in the infrastructure. In addition, the infrastructure needs to support a large spectrum of heterogeneity and high volumes of information.
- **Access to Government Services**: Governments provide a large number of services to help citizens. Discovering and accessing services that best fit citizens' needs often require expert knowledge. The aim would be to enable a seamless and transparent access to government services through a Web-based infrastructure.
- **Privacy and Security**: These two issues are very critical because of the very sensitive information (e.g., citizens' personal information) and the nature and requirements of some government services (e.g., Internet voting). Failure to provide adequate solutions to privacy and security challenges may hinder the wide deployment of Digital Government.

- **Accessibility**: One of the major goals of Digital Government is to be accessible to all citizens regardless of their physical and social conditions. This means that particular attention should be given to disadvantaged citizens.

Our proposed architecture, *WebDG*, addresses all the above issues save security, privacy, and accessibility.

3.CASE STUDY: GOVERNMENT SOCIAL AND WELFARE SERVICES

While the outcomes of this research are generic enough to be applicable to a wide range of applications, we specifically target the general area of government social services as a case study. The Indiana *Family and Social Services Administration* (FSSA) is our partner in this research. We have worked very closely with the FSSA to address the problem of organizing and accessing the glut of information generated by the variety of departments and other autonomous entities the agency deals with.

The FSSA is moving ahead to help strengthen the ability of families to succeed in their communities. It provides services to families who have issues associated with low income, mental illness, addiction, mental retardation, disability, aging, and children who are at risk for healthy development. The FSSA has several programs to assist citizens for their special needs. These programs interact with their federal counterpart to address issues requiring access to data from other agencies (state and local governments). Federal agencies also need this information for better planning and budgeting. This interaction is also required for reporting and auditing purposes. It is important to note that each program usually maps to a separate information system that in turn maps to several databases. The most important systems used by FSSA are summarized in *Table 1*.

All systems mentioned in *Table 1* interact with the *US Department of Health and Human Services* (HHS) information systems as mandated by law. The purpose has so far been largely for reporting and auditing purpose. However, this is expected to include other goals (e.g., planning and budget allocation). Additionally, some systems interact with other federal agencies such as the *Internal Revenue Service* (IRS) to intercept money owed for child support. Some systems also interact with the Justice Department systems for fraud detection and enforcement. Interfaces between the FSSA systems and other state and federal systems are all dissimilar and no standard interfacing for data transfer and exchange exists. Problems of scalability and extensibility are a direct consequence of *ad hoc* solutions. 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.

Table 1. Overview of FSSA Systems

Agency	*Application*	*Architecture*	*DBMS*	*# of Sites*	*# of Stations*
Family and Children	Indiana Client Eligibility System	Mainframe	IMS	140	3,800
	Indiana Child Welfare Information System	Client-server	Oracle	95	1,200
Child Support	Indiana Support Enforcement Tracking System	Mainframe and distributed AS/400s	DB2 and DB2/400	184	900
	Providers	N/A	N/A	N/A	N/A
	County operations	N/A	N/A	N/A	N/A
Disability, Aging, and Rehabilitative Services	Client Rehabilitative Information system	Client-server	N/A	30	400
	Vocational Rehabilitation Claims	Client-server	MS Access	1	10
	Bureau of Disabilities and Determination Services	Client-server	N/A	9	100
	Bureau of Aging and In-Home Services	Islands of LANs	MS FoxPro	17	100
Mental Health	State operating facilities	Client-server	N/A	9	880
Family Resources	Temporary Assistance for Needy Families	Mainframe batch	IBM DB2	N/A	N/A
Family and Social Services Admin.	Data Warehouse	N/A	N/A	400+	7,500

4.ONTOLOGY-BASED SUPPORT FOR GOVERNMENT DATABASES AND SERVICES

Part of FSSA case officers' task is to interact with local and state agencies to extract government information and provide government services. However, the large number of FSSA databases and applications (services) makes it extremely difficult to query the available information space.

The Indiana FSSA consists of dozens of autonomous departments located in different cities and counties. Each department's information system consists of a myriad of databases and applications. 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 number of FSSA databases.

The FSSA provides several rehabilitation programs 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.

To elicit the filtering and reduction of the overhead of discovering FSSA databases and applications, we propose the use of distributed ontologies. The uniform use of distributed ontologies is the core of the proposed approach. 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. Furthermore, each FSSA application is wrapped by an *e-service*. Simply put, an *e-service* is an application functionality that can be programmatically invoked from the Web [CDS00]. Similarly to FSSA databases, we organize the e-service space into *distributed ontologies*, also called *vocabularies*. Each vocabulary is composed of a set of attributes that describe the basic properties of FSSA e-services.

4.1 Modeling Government Databases into Distributed Ontologies

WebDG adopts an *ontology*-based organization of the diverse databases to filter interactions, accelerate information searches, and allow for the sharing of data in a tractable manner. Key criteria that have guided our approach are: scalability, and design simplicity.

An *ontology* defines a taxonomy based on the semantic proximity of information interest [Bou99]. 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* (*Figure 1*). This topic-based ontology provides the terminology for formulating queries involving a specific area of interest. It aims to reduce the overhead of locating and querying information in large networks of databases. As an information source may contain information related to more than one domain of interest, it may belong to more than one ontology. Note that our definition of an ontology is a little different from those found in other areas of research (e.g., linguistics or AI). In our

approach, a concept is defined locally and its definition may change over time. The participating information systems are responsible for the meaning of a concept.

Each ontology focuses on a single common area of interest. It provides domain specific information and terms for interacting within the ontology and its underlying databases. This generates a *conceptual space* that has a specific content and scope. The formation/modification of an ontology is a semi-automatic process. Privileged users (e.g., the database administrators) are provided with tools to maintain the different ontologies. 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 *Figure 1*.

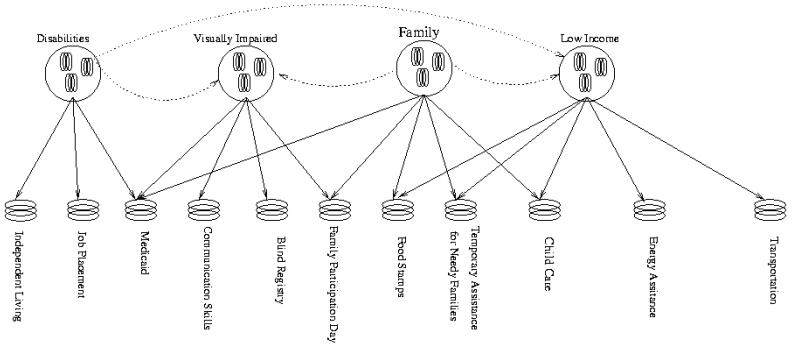

Figure 1. Example of Ontologies within FSSA

The FSSA ontologies are not isolated entities. They are related by *inter-ontology relationships*. These relationships are dynamically established based on user requests' patterns. They allow a query to be resolved by member databases of remote ontologies. The inter-ontology relationships are initially established statically by the ontology administrator. They essentially depict *functional* relationships that dynamically change over time.

Locating databases that fit users' queries requires detailed information about the content of each database. For that purpose, the concept of *co--databases*} are introduced. These are metadata repositories (data about groups of databases) that surround each FSSA database. To avoid the problem of centralized administration of information, metadata repositories are distributed over information networks.

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 as follows:

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

4.2Ontological Organization of Government Services

The current process of collecting social benefits within the FSSA is *ad-hoc* and time-consuming. Typically, the case officers must manually execute different FSSA applications related to the welfare programs. Since the number of applications is large, locating the ones that best fit citizens' needs is usually a cumbersome task. Furthermore, the applications are distributed over different FSSA departments. This makes this task even more complex. To address this issue, we wrapped each FSSA application participating in *WebDG* by an *e-service*. The main benefits of adopting the e-service approach are the following: (1) E-services allow the use of pre-existing (legacy) FSSA applications without requiring their modification. (2) They allow the bridging of the heterogeneity of FSSA applications. Heterogeneity occurs at different levels that include the interface, programming language, business logic, and back-end systems. (3) They cater for the dynamic discovery of FSSA applications that best meet citizens' needs.

To facilitate e-service discovery, we define a vocabulary for FSSA e-services. The vocabulary is composed of a set of attributes that describe the basic properties of FSSA e-services. Each e-service is advertised using XML according to the defined vocabulary (*Figure 2*). The use of XML is a natural fit because of its growing popularity as a standard for encoding and exchanging information over the Web. The following example shows a subset of the XML document used to advertise an FSSA e-service:

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

FSSA Vocabulary
- Location
  - Division: String
  - Bureau: String
  - City: String
- Topic
  - Addiction: Boolean
  - Adoption: Boolean
  - Children: Boolean
  - Disability: Boolean
  - Elderly: Boolean
  - Employment: Boolean
  - Family: Boolean
  - Food: Boolean
  - Hard of hearing: Boolean
  - Health: Boolean
  - Housing: Boolean
  - Insurance: Boolean
  - Pregnancy: Boolean
  - Training: Boolean
  - Visually Impaired: Boolean
- Assistance
  - Cash: Boolean
  - Supportive: Boolean

Figure 2. FSSA Vocabulary

The aforementioned attributes are of three types: *location-based*, *topic-based*, and *assistance-based* (*Figure 2*). *Location-based* attributes 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 in 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 kind of assistance users prefer. The assistance would be either in cash (e.g., TANF- *Temporary Assistance for Needy Families*) or in kind (e.g., WIC- *Women Infant and Children*).

5.DEPLOYMENT OF WEBDG

In this section, we describe the details of *WebDG*, an architecture to support Digital Government applications. We also describe real FSSA application using *WebDG*.

5.1WebDG Architecture

In Figure 3, we present the global architecture of *WebDG* using the FSSA scenario. A representative set of eight FSSA databases (Family Participation Day, Communication Skills, TANF, Food Stamps, Medicaid, Blind Registry, Job Placement, and Independent Living) has been identified. These databases are either under Oracle or Informix DBMSs. We have also deployed four FSSA applications. These are wrapped by e-services.

Users (i.e., citizens and case officers) access *WebDG* through a Web browser. They download a Java applet that opens a socket connection to the rest of the system. Two types of requests can be submitted: querying FSSA databases or invoking FSSA applications. The data/service manager receives all requests. The data/service manager is composed of five components: *request handler*, *data locator*, query *processor*, *service locator*, and *service engine*.

The *request handler* is responsible for routing requests to either the *data locator* or the *service locator*. The data locator's role is to educate users about the information space and discover relevant databases. Three ontologies are currently implemented in the prototype: *family*, *visually impaired*, and *disability*.

Information necessary to locate and access FSSA databases is stored in *distributed metadata repositories* (one repository per database). These repositories contain information such as ontology membership, database name and location, etc. The metadata repositories are stored in object-oriented databases (ObjectStore DBMS). They are registered as CORBA objects to three different Orbix ORBs (one ORB per ontology). CORBA provides a robust distributed object infrastructure for implementing distributed applications.

While browsing the available information space, users can learn about the content of each database by displaying its documentation in the format provided by the database (e.g., HTML/text, audio, or video formats). Once users have located a database of interest, they can submit SQL queries directly to individual databases. The *query processor* handles these queries by accessing the appropriate database via JDBC gateways. Like the metadata repositories, FSSA databases containing the actual data are registered as CORBA objects. *WebDG* uses two ORBs, namely OrbixWeb and VisiBroker.

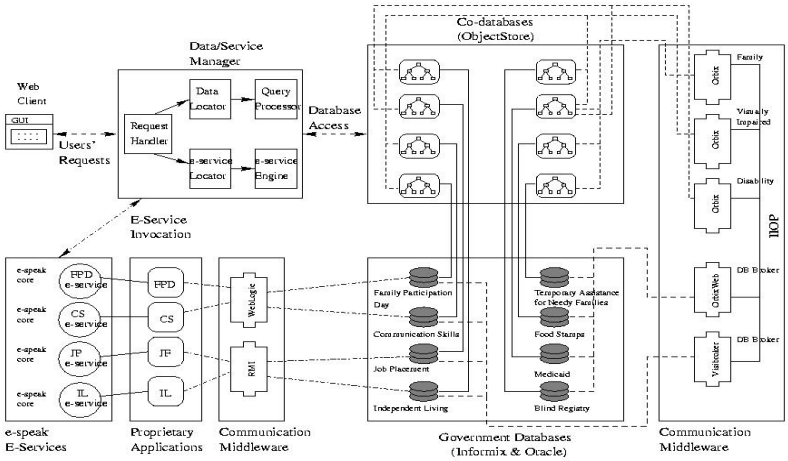

Figure 3. The Global Architecture

The service locator and engine components deal with application requests. Because the number of applications offered within FSSA is large, helping users to find applications of interest is of prime importance. To facilitate this discovery, each application is wrapped by an *e-service*. We have used *HP's e-speak* [HP], an e-service platform to define and invoke e-services.

Users can locate e-services by specifying either the e-service name or properties. Frequent system users usually use the first alternative. Service properties are part of a vocabulary defined for government welfare services. Examples of properties include the *service category* (e.g., *health*, *housing*) and the *bureau* the service is member of (e.g., *VRS*). All e-services are registered in an *e-speak core* by specifying their interfaces and properties using the defined vocabulary. Once users have located the e-service of interest, they can directly interact with it by invoking its operations through

the *service engine*. Examples of operations for the *Job Finder* e-service include searching jobs and setting up interviews. The current prototype includes four e-speak compliant e-services: *FPD* (*Family Participation Day*), *CS* (*Communication Skills*), *JF* (*Job Finder*), and *IL* (*Independent Living*). The e-services wrap proprietary FSSA applications. These applications use two communication middlewares, namely EJB (WebLogic server) and RMI, to implement social and rehabilitation programs.

5.2Application Scenario

Citizens come to the FSSA for specific needs. They may be unemployed, unable to support their families, have children with disability, etc. Based on the ontological approach in organizing data and services, the *WebDG* system helps case managers ensure that applicants receive all benefits they are entitled to. We describe a scenario in which a disabled citizen visits an FSSA officer looking for a job and other rehabilitation programs within FSSA.

In a first step, the case officer checks if the citizen is registered within the Vocational and Rehabilitation Services bureau. The officer browses the existing ontologies and finds that the disabilities ontology is relevant. Three databases, Job Placement (JP), Independent Living (IL), and Medicaid (MDAID) are displayed. The case officer selects the Job Placement (JP) database and submits a query to this database (see *Figure 4*). The query's result shows that the citizen is registered within the JP database.

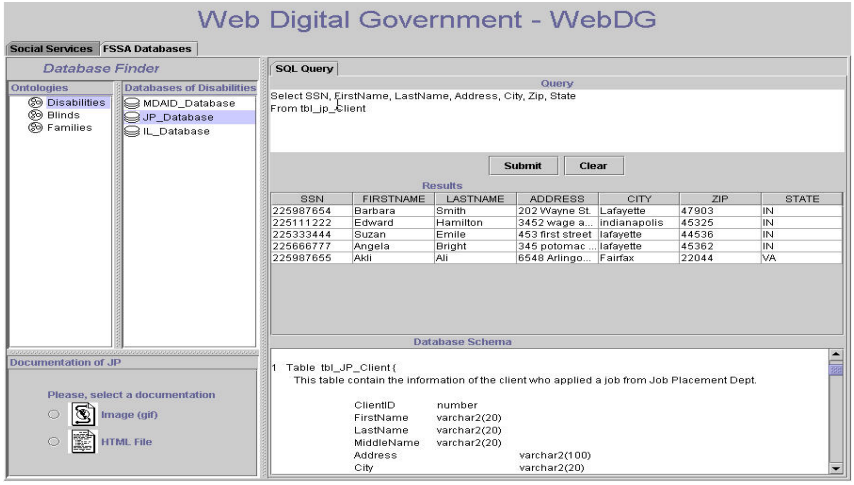

Figure 4. Database Finder

Next, the case officer looks for e-services that provide job openings for disabled people. The disability and job properties of the FSSA vocabulary are used. The system returns the Job Finder e-service. The case officer invokes a Job Finder's operation (Job Search) that looks for job openings. Appropriate search criteria such as skills, minimum hourly wage, and number of hours are specified. The case officer invokes the Job View operation (see *Figure 5*) of the Job Finder e-service to view the different openings relevant to the citizen's expertise. Openings can be discarded, hidden from the citizen, or made available for browsing.

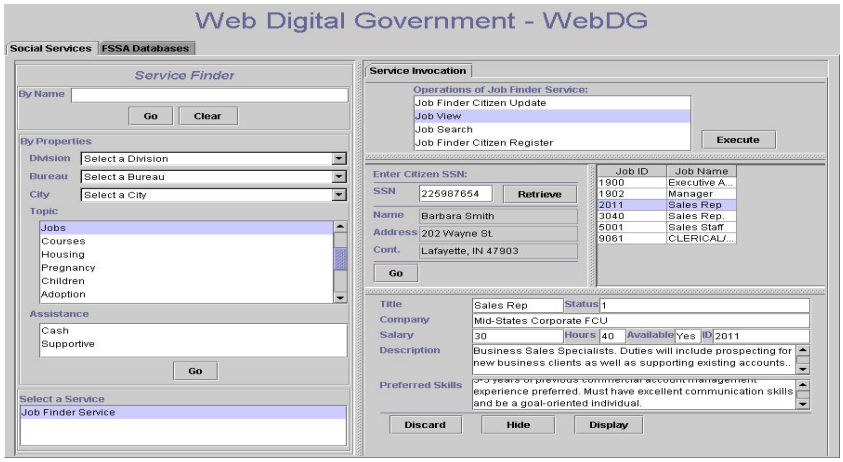

Figure 5. Service Finder

6.CONCLUSION

Digital Government environments are characterized by a large number of interacting databases and applications (e-services). In this chapter, we described *WebDG*, a system that seamlessly integrates databases and e-services based on an ontological approach. We implemented the system to support government welfare and social services within the Indiana FSSA. Case managers use a Web-based one-stop shop to serve needy citizens to have access to the different FSSA programs and services. Citizens have access to the system to check the application status and invoke operations on selected e-services (e.g., set up a job interview). The *WebDG* system has been fully implemented using different technologies have been used

including CORBA, Java, RMI, EJB, and HP e-speak. A demo is available online at http://www.nvc.cs.vt.edu/~dgov.

ACKNOWLEDGMENT

This research is supported by the grant 9983249-EIA from the National Science Foundation's Digital Government program. The second author is partly supported by a USENIX scholarship.

REFERENCE

[Bou99] A. Bouguettaya, editor. Introduction to the Special Issue on Ontologies and Databases, Distributed and Parallel Databases Journal, 7(1). Kluwer Publishers, July 1999.

[CDS00] 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.

[CS01] F. Casati and M.-C. Shan. Models and Languages for Describing and Discovering E-Services. In SIGMOD 2001, May 2001.

[DBK99] S. S. Dawes, P. A. Bloniarz, and K. L. Kelly. Some Assembly Required: Building a Digital Government for the 21st Century. Technical report, Centre for Technology in Government, http://www.ctg.albany.edu/research/workshop/, March 1999.

[EM01] A. Elmagarmid and W. J. McIver. The Ongoing March Toward Digital Government. IEEE Computer, 34(2), February 2001.

[Gro99] IT2 Working Group. Information Technology for the Twenty-First Century: A Bold Investment in America's Future. National Coordination Office for Information Technology Research and Development, http://www.ccic.gov, January 1999.

[HP] HP. E-speak Developer Site. http://www.e-speak.net/.

[JK99] Bill Joy and Ken Kennedy. IT Research: Investing in our Future. President's Information Technology Advisory Committee Report to the President, February 1999.