A Natural Language-Based Multi-Agent System for Legal Research

Vijayan Sugumaran
Oakland University

Kimberly Davis
Mississippi State University

Follow this and additional works at: http://aisel.aisnet.org/amcis2001

Recommended Citation
http://aisel.aisnet.org/amcis2001/83

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2001 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
A NATURAL LANGUAGE-BASED MULTI-AGENT SYSTEM FOR LEGAL RESEARCH

Vijayan Sugumaran
School of Business Administration
Oakland University
sugumara@oakland.edu

Kimberly Davis
Management and Information Systems
College of Business & Industry
Mississippi State University

Abstract

Conducting legal research can be a complex task for seasoned attorneys, and is even more daunting for laypeople. The proliferation of online legal resources requires sorting through the mass of information to obtain the most relevant and meaningful information. Various search techniques can be used on the Internet to narrow and refine one’s search, but the syntax required to search online sources varies. An intelligent agent system could be used to ease the search process by reformulating natural language queries into the appropriate syntax to launch a comprehensive search of relevant online sources to obtain the pertinent information. This paper discusses the benefits to be gained by the use of such an intelligent agent and proposes an architecture for a multi-agent system that implements a natural language based approach for query processing, in conjunction with a domain ontology.

Introduction

Legal research is an art that requires some understanding of the topic being researched, as well as an understanding of the various sources from which the law and legal information may be derived. Legal documents are beginning to be released, exchanged and stored in electronic form both in the public and private sector. For example, many federal government documents are now available electronically. As more and more legal resources become accessible through the Internet (Royles and Bench-Capon 1999), anyone with Internet access can conduct legal research without much legal training or access to special libraries or proprietary databases like Lexis and Westlaw. However, there are several problems that a layperson faces in successfully retrieving documents that are relevant to the legal research that he or she is conducting.

When a user retrieves documents from an online database containing legal resources, he or she creates a query that is executed and the results displayed (Gelbart and Smith 1991). If the query specification is too narrow, the search may not yield enough hits; however, if it is too broad or ill-defined, the search may result in too many hits. Another problem that users face is the lack of knowledge of existing resources and their idiosyncrasies. The legal researcher is faced with identifying appropriate resources and correctly specifying the search query both syntactically and semantically (Moens et al. 2000). In order to minimize this cognitive load and complexity, intelligent agent technology could be used to shield the user from such syntax requirements. Ideally, the user should be able to specify what he or she is looking for in natural language. Intelligent agents could then be deployed to translate search requests into distributed queries, communicate with each of the resources via the Internet and retrieve the necessary documents, synthesize and then present them to the user (Debnath et al. 2000). Thus, the technical complexity of the search could be hidden from the user to allow them to focus on the legal analysis. Therefore, the objectives of this research are to develop:

• an approach for obtaining meaningful results from a natural language search of online legal resources, and
• an architecture for a multi-agent system that implements the natural language based approach for query processing, in conjunction with a domain ontology.
Background

Legal information systems have been an active area of research for almost a decade. Several knowledge-based and retrieval systems have been reported in the literature (Valente 1995; Edwards 1997); however, most of the systems have several limitations and are domain specific. Intelligent agent technology is being utilized to ease the “information overload” problem caused by the exponential growth in the amount of information as well as the number of sources available over the Internet.

In the area of artificial intelligence research, “intelligent agents” are being developed to work on behalf of an individual, as a personal assistant, to obtain information of interest to the user in an intelligent fashion (Helmers 1997). “In contrast to other programs which invariably react to the same keywords or actions in the same manner, the object of these agents is to act in a way geared towards one certain person and, in a manner of speaking, independently and intelligently” (Helmers 1997). Ultimately, the goal is for the intelligent agent to “learn” how to obtain the most relevant information for a particular user based on that user’s preferences and other profiling factors (Helmers 1997). The use of agents can facilitate the efficient retrieval of pertinent information for users, thereby freeing them to pursue other activities (Sapherstein 1997).

Intelligent agents often use ontologies to provide context and resolve conflicts. An ontology refers to a set of concepts or terms that can be used to describe some area of knowledge or build a representation of it (Swarthout 1999). In other words, ontologies “provide definitions for the vocabulary used to represent knowledge” (Fikes 1996) and define the basic terms and rules for combining related concepts (Neches et al. 1991). Terms are usually organized into a taxonomy. An ontology is useful in this research because it provides a means of understanding the terms and concepts in the legal domain (Bench-Capon et al. 1997). It thus provides additional information that could be used to determine what artifacts are related and, therefore, might also be of interest to the user, whether or not the user has specified them initially. In this way, a more “intelligent” response to a user’s query can be generated.

Architecture of the System

This section presents the agent-based architecture for the proposed system with a Web-based natural language front end that will provide the user with intelligent query processing and retrieval capabilities. An overview of the architecture is shown in Figure 1. The architecture can be considered as having two parts, a client and a java enabled server. The client side consists of a basic Web browser. The server side is comprised of the following agents: interface agent, query agent, coordination and execution agent, resource interface agents, and results organization agent. The server side also has repositories that store domain ontologies, past queries, and user profiles. These components are briefly described in the following paragraphs.

The “interface agent” consists of two modules: one that deals with capturing the user’s requirements and one that displays the results to the user in an appropriate format. The “query agent” consists of two modules: initial query and query refinement. The initial query module uses natural language processing techniques to translate the user’s query into structured query language. The query refinement module enhances the initial query by making use of the domain specific information contained in the ontology. The “coordination and execution agent” is responsible for generating the distributed queries, and contains meta-knowledge about the legal resources available and the agents that interface with them. The coordination agent also manages this federation of “resource interface” agents, which enforce the specific syntax (or communication protocol) to be used in interacting with its target resource. The resource interface agents work together (Jennings et al. 1998) in generating the results to sub-queries and the coordinator agent synthesizes these results to generate the overall result for the search query. This result is then processed by the “results organization agent” based on the user preferences.

The second set of components on the server side is the repositories. The “user profiles” repository contains user preferences, authentication information, and the domains that are of most interest to the user. The “domain ontology” repository contains domain-specific knowledge that gives the system its intelligence (Upschold and Tate 1998). It enables the system to identify related terms and concepts that are used in refining the search query (Osborn and Sterling 1999). The “past queries” repository stores previous queries that have been executed. These query-templates are reused during query refinement.

Let us consider a sample scenario to illustrate the inner workings of the system. A user may initiate a search using the following natural language query: “Am I bound by the license agreement that was displayed on the screen when I installed a copy of Microsoft Word on my computer?” This initial query might be processed, on the server side, by an intelligent agent as follows.

The user’s personal profile could be used to first limit the search, based on the user’s geographical location, to the user’s jurisdiction. For example, if the user is located in Massachusetts it would be appropriate to begin with a search of cases in the U.S. Court of Appeals, First Circuit. The domain ontology should also map concepts contained in the initial query to legal terminology. For example, the “license agreement that was displayed on the screen” is known as a “clickthrough” or “clickwrap”
agreement. When an initial search yields no results relating to the enforceability of “clickthrough” or “clickwrap” agreements, the ontology should provide for expansion of the concept to include the analogous concept of “shrinkwrap” agreements in a refined query. The word “bound” might be expanded to include terms such as “enforceable” and “enforceability.” The phrase “Microsoft Word” should be expanded to include the term “software” and perhaps related concepts such as “hardware”, “technology”, and “databases.” Thus, the search might be formulated as follows: ((enforceab!, bind!, bound) /p (clickwrap, clickthrough, shrinkwrap)) and (Massachusetts, Circuit=1). When no cases are revealed in the First Circuit, the agent should broaden the search to include all other Circuits. The agent should also identify the Electronic Signatures in Global and National Commerce Act, enacted in June 2000 by President Clinton, as a related topic that the user may wish to research further to determine the impact, if any, it may have on the enforceability of shrinkwrap and clickthrough agreements. Ultimately, the results should be organized by the results organization agent by weighting the relevance of the retrieved information, and guidance for further inquiries should be given to the user.

**Implementation**

A proof-of-concept prototype implementation of the system is currently underway. The prototype is being implemented using Jess, a java-based expert system shell (Friedman-Hill 2000), JavaBeans, and JSP (Java Server Pages). The query agent and the coordination and execution agents are being implemented using JSP technology. The different modules of the agents utilize several javabeans for: gathering user input; generating initial queries in structured query language; accessing and retrieving artifacts from the target databases; formatting the results; and dynamically creating html documents to be served back to the user. The query refinement module and the ontology are being implemented using Jess. Query refinement is accomplished through rules written in Jess that make use of domain specific knowledge contained in the ontology.

**Summary**

This paper proposes an approach for searching and retrieving legal documents from disparate sources using domain-dependent knowledge contained in ontologies along with natural language parsing techniques, which enables users to execute more intelligent queries. The methodology is currently being implemented in a prototype system using Java Server Pages, Jess, and JavaBeans. In addition to facilitating the execution of intelligent queries, the system will also provide a mechanism for developing user profiles and reusing past queries. Further work is needed to refine and test the prototype and to build up ontologies for different segments within the legal domain.

**References**

Available upon request.