PAAM 97

4/22

Proceedings of the Second International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology

Conference Organisation
The Practical Application Company

Conference and Programme Chair
Barry Crabtree

Sponsorship Co-ordinator
Clive Spenser

21st-23rd April 1997
Westminster Central Hall, London , UK

PAAM, P.O. Box 137, Blackpool, Lancashire, FY2 9UN, UK
Tel: +44 (0)1253 358081, Fax: +44 (0)1253 353811
Email: proceedings@pap.com

ISBN 0 9525554 6 8
Published by The Practical Application Company Ltd

Information Brokering in an Agent Architecture

David Martin
Artificial Intelligence Center
SRI International
martin@ai.sri.com

Hiroki Oohama
Laboratory of Information Technology
NTT Data
hama@lit.rd.nttdata.co.jp

Douglas Moran
Artificial Intelligence Center
SRI International
moran@ai.sri.com

Adam Cheyer
Artificial Intelligence Center
SRI International
cheyer@ai.sri.com

Abstract

To date, document identification based on keyword matching strategies has been the basis of most efforts to provide assistance in accessing information resources on the Internet. However, in view of the limitations on human browsing time and the evolution of more capable software agents, we can expect rapid expansion in the use of fully queryable information sources (such as databases and knowledge bases) and *semiqueryable* sources (such as form-based query pages on the World Wide Web, and collections of Web pages that are structured by textual markups).

The ability to obtain information from a wide variety of queryable and semi-queryable sources is a prerequisite for the success of many types of software agents. Providing access to these sources — whether for end users or for software agents acting on behalf of users — poses a number of interesting challenges, many of which can themselves be addressed using agent-based approaches.

This paper describes a working prototype Information Broker system, developed within the Open Agent Architecture framework, that provides transparent access to a variety of information sources, each encapsulated as an independent agent. In this system, a broker *meta-agent* provides flexible mediation services, accepting queries expressed in broker or source schemas, gathering and integrating all available responses from the relevant sources, and allowing for the addition or deletion, at runtime, of participating information sources. Other broker features support the use of conversions, normalizations, and other basic domain knowledge in queries, and *persistent queries* (for which the Broker notifies requestors of changes in information sources that could affect their results). Source agents implement caching and retrieval strategies that alleviate problems related to the long access times and unreliability of Internet sources.

1 Introduction

The rapid growth of the World Wide Web and other forms of Internet and intranet access, and the need for automated assistance in making use of these resources, are now widely appreciated. The most immediate need, and the one receiving the greatest attention to date, is in the area of document retrieval; that is, the identification of textual documents that are relevant to some topic of interest. The topic of interest is generally indicated by an expression containing keywords, and the goal of the retrieval is to locate documents from which a human reader can extract useful information. A number of systems — such as NTT Data's *InterInfo*[1] in the commercial realm and *Amalthaea* [11] in the realm of agent research — are providing increasingly sophisticated approaches to document retrieval.

1.1 Structured and Semistructured Information Sources

An equally important set of challenges, but one that has not yet become as widely recognized, exists in the area of structured and semistructured information sources. By *structured* information sources, we mean sources that can be queried by using well-defined, general query languages, such as relational databases, object-oriented databases, and knowledge bases.

By *semistructured* information sources, we include a variety of sources that contain sufficient structure to be treated as databases, even though they do not provide the full generality and power of a query language. In the context of the World Wide Web, this structure is usually provided in one (or both) of two ways: by an informal form-based query interface or by the presence of HTML (HyperText Markup Language) markups and other textual markers used according to site-specific conventions.

We refer to a semistructured source that provides an informal form-based query interface as a *semiqueryable* source. It should be apparent that, even when such an interface is provided, its "query" capability usually falls far short of the power and generality of a structured information source. When a source provides structure by textual markers, we call that a *semistructured textual* source.

An example of a semistructured source — in which both types of structure are present — begins with a Web page that allows one to ask for hotels by filling in one or more of the following items: a location, a hotel chain, or a class of accommodation (economy, standard, luxury). This page is *semi*queryable, because even though it allows for the construction of certain simple queries, there are many others that cannot be constructed.[2] The result of this query is a list of hypertext links to hotel pages, each of which contains the same basic data in more or less the same format. Each of these pages is a *semistructured textual* source of information, and each can be transformed into a set of data records, using parsing techniques, given that the format is known.

[1] All product names mentioned in this document are the trademarks of their respective holders.

[2] For instance, one cannot ask, in a single query, for luxury hotels in Amsterdam and economy hotels in Paris.

1.2 The Need for Information Brokering

An *information brokering* system is one that provides coordinated access to a heterogeneous collection of structured and semistructured information sources. There are three key reasons why structured and semistructured information sources — and thus, information brokering systems — will be of rapidly increasing importance in the Internet and intranet worlds. First, human browsing time is both limited and, at least in the workplace, extremely expensive. Whereas *document retrieval* returns a body of text from which a human browser can extract some required data, *data retrieval* from structured and semistructured sources returns the required data itself. Thus, in performing tasks where the data requirements are well defined, there is significant economic pressure to make use of structured and semistructured sources.

Second, enterprises have large investments in legacy databases, and naturally want to leverage these in the context of the World Wide Web. This is especially true in light of the decentralization of corporate resources, and other trends in business process engineering. One straightforward way to leverage existing databases is simply to make them accessible to employees via Web interfaces. But the potential for leverage goes much further, when one considers the variety of ways in which legacy databases can be used in combination with other enterprise information resources, and with information from sources on the Internet. One role of information brokering systems is to facilitate the creation and maintenance of applications that rely on such combinations of information sources.

Third, queryable information sources are an essential requirement for the evolution of software agents that provide services within the context of the Internet or an intranet. By *services*, we refer not just to document retrieval, but to the full gamut of services that can be built around networked information. To provide a travel planning service or a stock tracking service, an agent must be able to retrieve data about a specific domain, in a form it can make use of — which requires access to structured and semistructured information. (Although natural language processing technologies have made impressive advances in recent years, there is still a long way to go before a program could reliably obtain this required data from *unstructured* text.) Moreover, many agents will need to access a wide and changing variety of information sources (consider, for example, an agent that finds the best available price for some product). As new sources become available, they will need to quickly acquire access to those sources. Information brokering systems can provide this access in a way that can be reused by numerous application agents.

1.3 Information Brokering and Agent Technology

Software agents will be very active *consumers* of the services provided by information brokering systems. We also argue that agent technology yields a promising approach to constructing the *providers* of these services. That is, the individual information sources as well as the brokering component(s) can very naturally be instantiated as agents, and their efforts coordinated by using the mechanisms of an agent architecture.

One other general observation may be made here regarding the relationship between infor-

mation brokering and agent technology: challenges encountered in coordinating access to information agents may be viewed as special cases of those encountered in coordinating the activities of all types of agents. Thus, many of the techniques developed to coordinate the satisfaction of a query by multiple agents may well be applicable to the more general problem of facilitating the activities of agents in completing various other types of tasks.

To summarize, the role of queryable information sources will expand rapidly in the context of Internet and intranet resources. Their use in these contexts poses several interesting challenges. Addressing these challenges is the focus of the Information Broker project, developed within the framework of the Open Agent Architecture (OAA). Following a brief discussion of these challenges in the next section, the remainder of this paper describes the system that has been developed in the initial work on this project. Section 3 provides an overview of the system, with background information about the OAA. Section 4 describes the system's architecture, and subsequent sections describe its individual components.

2 Challenge Areas

Our focus in this work has been on providing capabilities in two areas: *mediation* allows for the transparent interoperation of heterogeneous information sources; flexible *retrieval strategies* address issues such as long access times and unreliability of Internet resources.

2.1 Mediation

Mediation is a process that permits a requestor to get information from a wide variety of sources, without having to be aware of the identities, locations, schemas, access mechanisms, or contents of those sources. A component that performs mediation presents a single schema to its requestors, accepts queries expressed in that schema, and handles all the details of getting the appropriate data from the relevant information sources — each of which is likely to operate with a different schema.

The capabilities provided by a mediator may be broken down into three subareas: delegation, translation, and optimization. *Delegation* is the process of selecting the appropriate sources from which to satisfy each subquery of a given query. *Translation* of each subquery into the schemas of the selected sources must take place before the subqueries are sent to the sources — and results returned from the sources must be translated back into the schema of the original query. *Optimization* results in a query execution plan that obtains results from the selected sources as efficiently as possible, and exploits parallelism wherever possible.

It should be noted that, in their essence, the problems of mediation are not unique to networked environments. Some of these problems have already been studied for some time, under the heading of *heterogeneous databases*. However, the emergence of the Internet has triggered a renewed interest in these problems.[3] This is not only due to the massive volume

[3]Other projects currently exploring these issues are described in [1], [5], [6], [7], [8], [12], and [13].

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.