Using Agents for Information Retrieval

Abstract: In this work the use of information agents in an information retrieval environment is introduced. Initially, a brief revision of some basic theoretical aspects is presented. The contributions in this aspect are: a classification for non-cooperative information agents, based on tasks that the agent has to perform, and the identification of main functionalities which must be provided in cooperative information agents systems. In addition, from a practical point of view, MASIR, a cooperative agents prototype, which provides uniform access through Internet to heterogeneous and dispersed documental databases is explained.

1. Introduction

Nowadays, the advances in Information Technology and Communications and particularly the Internet revolution, allows accessing to a huge amount of information. As the Web size increases, finding out the suitable information becomes a more difficult task. Other problems related are the changes in data (pages appear and/or disappear or are frequently modified), redundancy (duplicated contents) or quality (incorrect or erroneous information). In addition, the problem complexity grows due to the distributed and heterogeneous nature of information. Data sources differ in interface, format, structure, and language, and are distributed among a great number of connected computers.

Information retrieval (IR) techniques try to solve some of the previously introduced problems. These methods provide a way of locating relevant documents based on user requirements (Baeza-Yates, 1999). Besides, advanced systems which help in efficient and unified access to information are needed. This has led to the appearance of new flexible and powerful tools and sophisticated algorithms (currently implemented in search engines) which help users in the process of retrieving relevant information.

In an environment the described before, intelligent agents constitute a key technology for the future of Internet. Agents may assist the user in finding useful, relevant information, managing and overcoming the difficulties associated with “information overload” (Klusch, 1999). They may also inform the user that new relevant data have been published, negotiate a sell or buy of products, participate in electronic auctions, etc. Agents may carry out these tasks independently or working in a coordinated way with other agents.

In this work the system MASIR, which is an example of cooperating agents is introduced. The paper is organised as follows. Section 2 introduce the theoretical aspects of intelligent information agents; while in part 3 concepts about non-cooperative agents are addressed. In section 4, cooperative agents, which is the key technology in the system is exposed in depth. Section 5 focuses in the description of MASIR prototype. Finally, in point 6 conclusions and future work are presented.
2. Intelligent Information Agents

Agents represent an important advance in abstraction level in the process of software development. They may be used for the design and implementation of complex (distributed) software systems (Jennings, 1999). Actually there is no universal definition of Agent concept, but a widely used one states that an intelligent agent is a computer system capable of flexible, autonomous action in some environment; in this definition, flexible means that agent is reactive; proactive and social. (Wooldridge, 1995).

Intelligent Information Agents can be defined as computational software systems which access to multiple heterogeneous and geographically distributed information sources in order to help users in the process of searching relevant information (Klusch, 1999). These agents provide transparent access to many different information sources, simplifying the problem of managing a large amount of data. This implies that agents must interpret the user request, analyse and translate it in the correct way for each source, retrieve the desired information, integrate the results and give them back to the user. All this process must be done preferably in a just-in-time fashion.

Information Agents may be classified according to different criteria. Depending on their ability to co-operate with each other in the execution of tasks, agents may be non cooperative or cooperative. This aspect distinguishes two kinds of systems, the ones which are formed by a single agent and systems composed by a society of interacting agents. Any of the previous types may be classified as rational, which are utilitarian in an economic sense, adaptive which are able to adapt themselves to changes in the environment or mobile which travel autonomously through the Internet (Klusch, 1999).

3. Non Cooperative Information Agents

Klush classification is based in the characteristics of the agent internal behaviour. Nevertheless, we thought it is possible to introduce a different classification taking into account the tasks that the agent has to accomplish. This classification is closer to final user point of view, and it is similar to the ones used in some Internet agents repositories (Botspot, Agentland, ...). According to this criterion, several kinds of non-cooperative information agents can be distinguished:

- **Search agents**, help the user to retrieve information from heterogeneous and distributed sources. They provide a fast and simple way to obtain relevant information. Examples of search agents are Bullseye (http://www.intelliseek.com), CiteSeer (http://www.researchindex.com) and Copernic (http://www.copernic.com/).

- **Monitor agents** control the changes in different information sources (for instance, changes in web structures, updated news in a newspaper...) or the appearance of new sources related with a particular subject in order to warn the user. Usually, the notification of changes is sent to user by e-mail. Examples of monitor agents are Mind-It (http://minder.netmind.com/) and Informant (http://informant.dartmouth.edu/)

- **Filter agents** reduce the incoming information keeping only the relevant data from the user point of view. The preferences of the user are stored in his personal profile. Examples of filter agents are InfoScan
Browser agents help the user in navigation through the Web. It is usual to have browser agents integrated into the navigator. They highlight dead links, emphasised keywords in the web page, anticipate relevant links, etc. Interquick (http://interquick.deerfield.com/) and Letizia (Lieberman, 1995) are examples of browser agents.

Agents for electronic commerce offer commercial services, in order to save time and money. They are able to recommend a certain product, to compare prices among different electronic shops, etc. Some examples are MySimon (http://www.mysimon.com/) and Pricerunner (http://www.pricerunner.com/).

4. Cooperative Information Agents

Many of current investigation related to intelligent agents has focused on individual agent capabilities and structure. Nevertheless, in order to solve complex problems these agents must cooperate with other agents in a heterogeneous environment. A multiagent system (MAS) is a set of agents which communicate among them and cooperate in task execution (Sycara, 1998). This interaction allows solving problems that are beyond the individual capabilities of each individual agent. Interaction in MAS differs from other computing paradigms since it takes place at knowledge level using a high-level agent communication language like KQML, FIPA-ACL (Labrou, 1998). Basically, the advantages of an approach based on cooperative agents are: simplicity, flexibility, robustness, scalability and the integration of existing legacy systems.

In this cooperative agents environment two main functionalities must be provided: a mechanism for linking the different agents and a way of solving the heterogeneity of managed information.

4.1. Intermediation

Cooperation among information agents can be achieved using different interaction techniques. Taking into account the role played, three different types of agents can be distinguished (Klusch, 2001).

1. Provider agents offer their capabilities to users and other agents.
2. Requester agents use information and services offer by provider agents.
3. Middle agents mediate for a correct communication among providers and requesters.

The process of mediation done by middle agents is based on the following steps: (1) Provider agents advertise their capabilities to one or more middle agents, describing the service they provide. (2) Middle agents are responsible for storing all these advertisement. (3) A requester agent asks for locating and connecting to provider agents which offer a desired service. (4) The middle agent using the stored advertisements returns the result.

Depending on the kind of result returned, two types of middle agents are distinguished:

1. Matchmaker agent, the result is an ordered list of provider agents which offer the requested service. Once this result is received by requester agents, it is the responsible for contacting the provider agent, negotiating and performing the transaction.
2. **Broker agent**, in contrast with matchmaker, performs the complete transaction. This means that there is no direct communication between provider and requester agent, because all the operations go through the broker agent. The main tasks of these agents are contacting appropriate provider agents, negotiating, performing and controlling the transaction and return the results of service to the requester agent.

Given the fact that different types of middle agents provide different performance trade-offs what types of middle-agents are appropriate depends on the application (Klusch, 2001).

### 4.2. Ontologies

One of the main problems for cooperation in an agents society is the semantic heterogeneity of information that the agents must manage. Semantic heterogeneity considers the content of an information item and its "intended" meaning. In order to achieve this semantic heterogeneity, the meaning of the interchanged information has to be understood across the agents society. An ontology can be defined as an explicit specification of a conceptualisation (Gruber, 1994); that is, a representation (with a set of concepts and the relationships among them) of an abstract and simplified view of the world.

In agents societies, ontologies can be used to describe the semantics of the requests and service descriptions and to make explicit the content of the different information sources. They also reduce conceptual and terminological conflicts providing a unified framework.

The middle-agent overcomes the semantic heterogeneity by means of a knowledge-based process which relays on using ontologies. That is, the use of ontologies enable shared understanding among different agents with different aims and different viewpoints of the global system.

### 4.3. Examples of Cooperative Multiagent Systems

**RETSINA** is an open multi-agent system, which performs an information retrieval and integration directed by goals, in support of a variety of decision-making tasks (Sycara, 2001). The process of mediation in RETSINA basically relies on service matchmaking. The specification of capability and service is based on an Agent Capability Description Language (ACDL) called LARKS (Language for Advertisement and Request for Knowledge Sharing). Application domain knowledge in agent advertisements and requests can be currently specified as local ontologies written in a specific concept language ITL or by using WordNet.

**InfoSleuth** is an agent-based system that can be configured to perform many different information management activities in a distributed environment (Bayardo, 1997). It is composed of seven kinds of software agents which, all together, provide a number of complex query services which involve solving ontology-based queries over the dynamically changing resources. InfoSleuth architecture uses a process of matchmaking, that is, the broker agent maintain a knowledge base with information about all the agents in the system and their capabilities and uses this knowledge to match required services with the provider agent.

### 5. MASIR

Multiagent System for Information Retrieval (MASIR) is a prototype based in cooperative information agents. The system allows uniform access trough Internet to several heterogeneous documental databases. The prototype allows, by
means of an unique interface, to pose a query, send it to several databases, get the answers from the databases and finally integrate the results and present them to user. The scalability of the system permits the integration of new sources, with independence of their data model, structure, query language, DBMS or location.

The global architecture of the system is presented in Figure 1. The system is integrated by four kinds of agents: interface, wrapper, broker, and ontology.

- Each user interacts with the system through its own interface agent. It is responsible for capturing the query, translate it to the suitable FIPA-ACL sentence (Request) in order to send it to broker agent and facilitate the user navigation through the obtained results. The interface agent, making use of the information supplied by ontology agent, generates dynamically, in order to make the query and to present the results, a user domain specific interface.

- The ontology agent contains an abstraction of the databases schemata, linking each concept to the databases where it is relevant. The ontology agent also provides to interface agents and broker the domain terms for the query and the semantic relationships. A more detailed description can be found in (Cuesta, 2001).

- The broker agent is the system core and acts as an intermediary among user and the various documental databases. It takes the request received from interface agent, composes subrequests, and distributes them among wrapper agents (using FIPA-ACL Contract-Net). In order to achieve this objective, it asks the ontology agent what the databases implied in the request are. It also synchronises answers to query, manages detected errors, integrates the answers received from several wrapper agents and sends them back to interface agent.

![Figure 1: MASIR Architecture](image)

- The wrapper agent knows the specific features of its related source. It translates the subrequest received from broker agent in order to access its...
associated documental database and obtain the required data. Each request may imply, for example, the execution of SQL or OQL sentences or the utilisation of text retrieval techniques. In addition, the wrapper agent contains all the details of the semantic knowledge for that source. When a new source is added, a wrapper agent will be created associated with it; meanwhile the ontology agent will update its knowledge base.

The comparison of MASIR with the systems presented in point 4.3, can be made taking into account the intermediation model and the use of ontologies. From the first point of view, both RETSINA and InfoSleuth are based on a matchmaking process. This implies a direct interaction between Requester agents and selected Provider agents, which is performed independently from the matchmaker. This avoids, for example, data transmission bottlenecks or single point of failure at the matchmaker but increases communication overhead among agents. MASIR, in contrast uses the brokering model. The main reason is avoiding communication overhead due to the necessity of solving source heterogeneity through the use of ontology agent.

From an ontologically point of view MASIR has a higher level of description than RETSINA or InfoSleuth because MASIR maintains a global agent ontology. This agent provides an abstraction of the information sources schemas, linking each concept to the specific source where it is relevant. RETSINA use only local ontologies to specify agent advertisements and requests instead of using them to integrate the information sources. On the other hand, InfoSleuth has a set of specific (local) domain ontologies and does not provide way to integrate the local ontologies.

6. Conclusions and future work

In this work, the advantages of using agents for information retrieval has been presented. Besides, a new classification for non-cooperative information agents has been proposed; this taxonomy is based on tasks that the agent has to perform. In the construction of cooperative agents systems, two important aspects must be taken into account: the model for intermediation among the different agents and how ontologies are used to solve the heterogeneity of information.

In contrast with other systems, MASIR uses a middle agent with brokering intermediation and a global ontology agent for solving the semantic conflicts. The system is able to grow; when a new source is added, an associated wrapper agent is created and the ontology is actualised incorporating the source new semantics.

As future work, many aspects will be studied. One of the key challenges is related with the communicational aspect of MASIR; at the moment, the interaction protocols are being refined. At ontological level, the development of mechanism for facilitating ontology creation and modification is being studied. In addition, the adaptation of MASIR to “FIPAOntology-Server” is currently under consideration, in order to allow the integration of external ontologies. Another current work area is the user of adaptative interface agents with capability of learning from the user behaviour and adapting its knowledge model to facilitate customisation and suitable task delegation.
References


