Integrating Agent Based Information Outsourcing Techniques on Data Warehousing Systems

Orlando Belo Alcino Cunha

Departamento de Informática
Escola de Engenharia
Universidade do Minho
Largo do Paço, 4709 Braga Codex, PORTUGAL
Voice: +351(53)604470 Fax: +351(53)604471 Email: obelo,alcino@di.uminho.pt

ABSTRACT

In the last few years, information outsourcing has been a current activity on large companies. Information as become a regular trading commodity. It is a well known fact that direct mailing companies acquire databases, or other kind of information sources, from other companies with names and addresses of potential clients. Also, enterprise managers are frequently concerned with the current status and welfare of their clients and suppliers. Commonly, they use to appeal to specialized external information providers which may, under certain conditions, provide them specific profiles about such potential commercial partners. On companies with effective means of information processing, it is very probable that such needs of information may be directly satisfied with data stored and managed in the companies' data warehouses.In this paper we propose a protocol based on economic principles that enables the automatic negotiation of information transference between data warehouse systems of different companies. The protocol was designed to be used by a community of intelligent agents responsible to ensure and support all the operational tasks related to information outsourcing among companies.

1. INTRODUCTION

A particular market can be viewed as a macro chess game where enterprises play for the conquer of a bigger market parcel. Enterprises with the best resources and strategies of decision-making make the difference. Their managers try to get in time the best information to support effectively decision-making tasks. They must deal with large volumes of information, frequently obtained from heterogeneous sources of information. The application of data warehousing techniques on decision-making processes is not a new approach. Enterprises have been developing and installing specialized repositories of data, oriented by subjects and competence areas, with the goal to support managers day-by-day activities of decision making. Today, any enterprise to be competitive and able to react, in an appropriated manner, to market requisites and mutations, needs to have an effective support to decision-making, which means effective and flexible *Data Warehousing Systems* (DWS) [15] [8] [9].

In the last few years information outsourcing has been a current activity on large corporations, as a form to complement their own local services of decision-making. Several reasons may justify such practice. Enterprise managers, namely the ones involved with commercial tasks, are frequently concerned with the current status and welfare of their clients and suppliers. In order to develop and establish safe and profitable commercial activities with them, managers need to know about their current credibility and positioning in the market. Thus, they use to appeal to specialized external information providers which may, under certain conditions, provide them some profiles about such potential commercial partners. Latter, the information acquired will be confronted with the one stored on local enterprise's *Data Warehouses* (DWs) and, after combined with the managers' expertise, may be an excellent support to effective decision making.

In order to optimize their own services of information outsourcing, and reduce operational costs, it would be very convenient to change their own data integration mechanisms, being able to manage automatically information transference between different companies. Additionally, it is also necessary to ensure effective processes for external gathering of information, be able to react to system failures, and develop sophisticated forms of information exchange among distinct DWSs. Through the combination of DWSs techniques with the ones related to agent based computing it is

0-7803-5731-0/99$10.00 ©1999 IEEE

possible to reach such goals and operational requisites.

2. DATA WAREHOUSING SYSTEMS SUPPORTED BY AGENTS

The scientific area of *Agent Based Computing* (ABC) [5] [13] has being growing significantly in the last few years. Agent based applications has demonstrated their real viability on a diversified range of areas [4] [6]. The modular approach provided by ABC, combined with the possibility to design and construct agents that emulate human expertise and knowledge on specific application domains, makes this technological "arena" very attractive to the conception, development, and implementation of DWSs. The technological field of agent based DWSs [7] involves the analysis, planning, and implementation of computational scenarios with the ability to integrate, in their own space, several communities of intelligent agents. These entities are able to do, autonomously, all the activities related to the regular tasks of a DWS, according with their own expertise and based on an agenda previously defined by their supervisors. When applied to a DWS, ABC techniques allow to improve its performance, making possible to transform conventional search and integration mechanisms in more flexible systems with self-adaptation abilities to day-by-day needs of the enterprise's decision-makers. By allocating regular DWS activities to an agent, or to an agent community, it is possible to: reduce management, monitoring and maintenance costs of a DWS; optimize selection, extraction, and integration processes; improve the quality of information through the application of intelligent techniques on consistency and redundancy control activities; ensure higher levels of confidence on decision-making processes; and develop adaptative interfaces according with previous user interaction processes.

3. AN ECONOMIC BASED APPROACH

Usually, information outsourcing involves costs. The information gathered on external specialized information providers is normally paid. Thus, in order to implement a real DWS with information outsourcing abilities, we decided to apply a set of economic principles to such processes, and integrate them on the knowledge bases of the agents that we intend to put in control of a DW information contracting system. Wellman [14] argues that it is possible to apply economic principles to solve almost any problem where the following conditions hold: the fundamental problem to be solved is one of resource allocation; all the involved agents act rationally in order to achieve their most preferred outcomes; the decision making is inherently decentralized.

In this scenario all these conditions hold. The most important resource of a company is its own money. In this case, each one has to decide how it should use it in order to get the information it needs: for example, should it acquire the information from company A or B, or should it spend it in building a new database from scratch? Since that negotiation is between different companies, whose main objective is to maximize profit, the second conditions follows immediately: each company should act rationally in order to acquire the information it needs at the minimum price. The last condition also holds trivially since we are talking about different companies, and it is clear that the result of the negotiation process can not be controlled by just one of them, being, on the contrary, an interactive process that tries to conciliate all different expectations. The protocol that will support the negotiation process between the companies is the *Contract Net Protocol* (CNP) [12, 3]. This protocol reproduces in a precise manner the interactions that occur in real markets when an entity wants to determine the best partner to execute a particular task. Two roles coexist in this protocol: a) the *managers*, that announce the tasks that must be allocated and select the best candidates for their execution; and b) the *contractors*, that answer to the announces with a bid that reflects their suitability to the execution of a task in the hope that eventually they will be selected to execute it.

The economic principles are applied in the decisions that must be made in the *announcing*, *bidding* and *awarding* phases of the CNP. The approach followed in modeling this decisions is a restriction of our previous work in the area of resource allocation in multi-enterprise environments [2, 1], which in turn was based of Sandholm work in the area of automated contracting using extensions of the CNP [10, 11].

4. THE INFORMATION CONTRACTING MODEL

The first problem to be solved when trying to implement a protocol that involves different companies is that of finding a common language that is understandable by everyone. In this particular setting, this problem is harder than usual, since the object of interest is not restricted and can be information of any kind. A similar problem occurs, for example in frameworks for Electronic Data Interchange, and the first solution that occurred to us was similar to the one

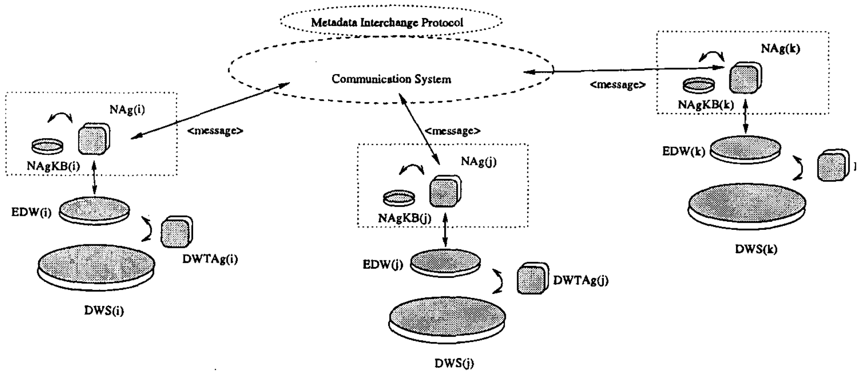

Figure 1. Agent-Based Data Warehousing Model for Information Contracting.

used there: all information requests are codified in *Structured Query Language* (SQL) and all companies that are interested in joining the automated contracting community must agree on a common conceptual scheme for data interchange among their DWs, which is integrated on the overall model's metadata interchange protocol layer (see Figure 1.).

This first attempt of solution revealed itself a very restrictive one, since it forced the re-conversion of all the DWs involved on the information contracting network, even if the information that a company is willing to sell was only a small part of it. Also, it does not allow for a company to participate in more than one automated information contracting network with different agreed schemas. To overcome these problems we propose a slightly different solution: a) there exists a common scheme agreed by all parties; b) however, instead of reconverting the DWs to the agreed scheme, every involved company maintains a separate DW - the *Exportable Data Warehouse* (EDW) - that contains the subset of the company main DW that can be negotiated and that is organized according to the agreed common scheme; c) there exists a *Data Warehouse Translation Agent* (DWTAg) for each of the companies, whose task is to maintain the EDW actualized according to the defined refresh rate of the main DW; d) all information requests are codified in SQL according to the agreed scheme. In each of the companies, besides the EDW and the DWTAg, there exists a Negotiation Agent (NAg), whose main task is to implement the negotiation protocol by executing the *announcing*, *bidding* and *awarding* phases of the CNP. The NAg may assume simultaneously both roles of the CNP: *manager* when trying to acquire information to its company and *contractor* when answering to announces from other NAgs. A pictorial view of the framework's model can be seen in Figure 1.

5. THE NEGOTIATION PROTOCOL

A typical run of the CNP in this framework proceeds as follows:

1. One of the NAg announces that it is willing to acquire a particular piece of information. This is done by sending to all the remaining NAgs of the community an announcing message.
2. After receiving the announce, each contractor NAg determines if its company is suitable to provide the required information and, in case of an affirmative answer, evaluates the information price and sends to the manager NAg its bid.
3. After receiving all the bids, or when a stipulated deadline arrives, the manager NAg selects the best one, sends to the wining contractor an award message and regret messages to all the others.
4. The awarded contractor evaluates if it still can provide the information (for example, meanwhile it could have sold it to an NAg that did not want it to be resold). If it can not, it sends a commitment breaking message to the manager and pays the stipulated penalty.
5. If the manager receives a commitment breaking message it restarts the negotiation process (back to step 1).

Since the awarding phase (steps 3 to 5) is very simple, we will only present in this paper the detailed description of the announcing phase (step 1) and the bidding phase (step 2).

The Announcing Phase

Each announce contains the following fields:

- **Announce identification.** Each announce should be uniquely identified within each company.

- **Source identification.** Identification of the NAg that is issuing the announce. The identification includes the IP address and port number where the NAg is expecting the bids.

- **Issuing time.** Time stamp that indicates the issuing time of the announce.

- **Requested information.** This is the main field of the announce. It is a SQL query that indicates which information the company is trying to acquire. The query should be stated according to the previously agreed schema.

- **Information time.** Instant of time when the requested information should be delivered.

- **Bid information.** Sometimes, in order to decide which bid is the best, it is not sufficient to compare the prices. For example, one may want to know how many rows the answer will have. This information should be sent in the bid together with the proposed price. Like the requested information, this request can also be stated as a SQL query on the agreed schema.

- **Bid deadline.** Deadline for sending bids to this announce. The bids received after the specified time will not be considered. The NAg determines this instant automatically by subtracting to the information time the usual delay necessary to select the wining bid and send the awards.

- **Eligibility condition.** Boolean expression whose terms may be SQL queries. Its purpose is to prevent the reception of bids without "quality". This notion of "quality" is announce dependent and can be used to state that, for example, one does not want to receive a bid unless the information it concerns has a specific minimum number of rows.

- **Maximum price.** This field is also an eligibility condition that prevents the reception of bids with an exaggerated price.

- **Disclosure time.** Sometimes, a company wants to guarantee that the information it is going to acquire will not be sold to another company at least for a certain period of time (of course, that this exclusivity will be reflected on the price). This field states the time when the contractor will be able to sell the information again. If its value equals the information time then the manager does not want exclusive access to the information.

- **Commitment breaking penalty.** This is the price that a contractor has to pay if it sends a bid to this announce and later, if it is awarded, refuses to provide the information. The value of this field is defined by the system manager and should compensate the company for having to delay the information acquisition.

The NAg makes the announces according to the *Table of Announces* (TA) that is updated by the system manager (a kind of *to do list* stored in the negotiation agent's NAgKB - *Negotiation Agent's Knowledge Base*). Besides containing the most important fields of an announce, the TA also has a mechanism to schedule repeated announces (for example, one may want to know the evolution of stock exchange every day and it would be very uninteresting for the system manager to have to schedule the same announce every day).

The Bidding Phase

The following steps must be carried out by a contractor NAg when an announce is received:

1. Determine if the requested information will be available at the specified information time.
2. Determine if the bid information is also available.
3. Check if the eligibility condition holds.
4. Check if the same information was already promised to somebody else during the period between the information time and the disclosure time.
5. Calculate the price of the information according to the specified disclosure time.
6. Check if the calculated price is less or equal then the maximum price.
7. Analyze the query requested in the bid information in order to determine if it contains valuable information.
8. Check if the bid can arrive to the manager NAg before the bid deadline.
9. Compose the bid and send it to the manager NAg.

Whenever possible, this steps were ordered according to its difficulty: when one of step fails it is not necessary to carry out the remaining ones and so, this ordering minimizes the time spent in analyzing the bids. The remaining of this section will be devoted to analyze the fifth and seventh steps, i.e., how to determine the price of the information and how to determine if the bid information contains valuable information. The remaining steps are straightforward. The task of determining the price of the requested information will be divided into two smaller tasks: first, we will present a sketch of how to calculate the base price, i.e., the price when the disclosure time equals the information time; after calculating the base price, we will show how to bias that price according to required interval of exclusivity.

The first step to calculate the base price belongs to the system manager and consists of valuing the raw information contained in the EDW. More specifically, for each table of the EDW it is necessary to establish the price of each row (given a table t this value will be denoted by $ppr(t)$), and to decide which fields are more valuable by giving to each a factor between 0 and 1 (given a table t and a field f this factor it will be denoted $value(t, f)$). Assuming that $tables(EDW)$ is a set with all the table names of the EDW and that $fields(t)$ is a set containing all the fields of table t, there is an obvious invariant about the values of the fields:

$$\forall t \in tables(EDW) \cdot \sum_{f \in fields(t)} value(t, f) = 1$$

After establishing the price of the information contained in the EDW, the required SQL query is analyzed in order to determine precisely the fields and number of rows from each table that are necessary to evaluate its result. For example, if the specified query requests the mean value of a particular field in a table, the number of necessary rows to answer it is the number of rows of the table. Suppose that for each announce a and table t the number of rows necessary to answer the query specified in a is denoted by $nrows(a, t)$ and that the set of necessary fields is denoted by $nfields(a, t)$. The base price of the announce a (denoted by $bprice(a)$) is determined as:

$$bprice(a) = \sum_{t \in tables(EDW)} nrows(a, t) \times ppr(t) \times \beta$$

being β defined as

$$\beta = \sum_{f \in nfields(a,t)} value(t, f)$$

In order to bias the base price we will begin by defining when two announces are incompatible. Given an announce a, $it(a)$ denotes the information time of a, and $dt(a)$ denotes the disclosure time of a. The fact that two announces a and b are incompatible is denoted by $a\#b$. The predicate $\#$ is defined as follows:

$$a\#b = (it(a) \geq it(b) \wedge it(a) < dt(b))$$

or

$$a\#b = (it(b) \geq it(a) \wedge it(b) < dt(a))$$

Let A be the set of all the pending announces, i.e., the announces to which the contractor NAg answered with a bid but did not receive an award or a regret. Let $price(a)$ denote the final price that was sent in the bid for an announcement a and $cbp(a)$ denote the commitment breaking penalty stipulated by the manager of a. Let $A\#a = \{b \in A | a\#b\}$ be the set of all announces in A that are incompatible with a (notice that $\#$ is overloaded). The maximum amount of money that the contractor NAg can potentially loose if it answers to a new announce a is

$$\Delta_a^+ = \sum_{b \in A\#a} price(b) + cbp(b)$$

This value assumes that all the conflicting bids sent by the agent will be accepted by the respective managers, and corresponds to a very optimistic behavior of the NAg. If the NAg has a pessimistic behavior it would expect no awards and hence $\Delta_a^- = 0$. Ideally the character of a company's NAg should reflect the character of its manager and so we decided to include a parameter α $(0 \leq \alpha \leq 1)$ that allows this configuration. If α is near 1 the NAg will behave very optimistically and for values near 0 it will reproduce a pessimistic behavior. Given this parameter α, the amount of money that the contractor NAg can potentially loose if it answers to a new announce a will be determined by:

$$\Delta_a = \alpha \times \Delta_a^+$$

The final price that will be sent in the bid for an announce a is determined as:

$$price(a) = bprice(a) + \Delta_a$$

This equation bias the base price according to the amount one may potentially loose by accepting a. This bias already takes into account the money one would not receive by having to reject awards incompatible with a and the respective penalties.

After presenting the method to determine the base price of the requested information it is straightforward to determine if the bid information contains valuable information. The contractor NAg applies the same method to the query of the bid information field and determines the price of that information. If that price is zero or bellow a specified threshold the NAg submits the bid.

6. CONCLUSIONS

In these paper we presented a generic framework for information contracting among a network of DWSs.

The use of intelligent negotiation techniques enables to: improve information interchange processes among heterogeneous DWs; optimize information contracting; improve local decision-making mechanisms; and increase DWS robustness, since it reduces direct human intervention. However, even reducing human dependency, enterprise managers are directly responsible for agents behavior. Their knowledge bases, that integrate representations and mechanisms to support their own decision making abilities, are fulfilled with managers' expertise and knowledge related to specialized areas of decision-making - like the estimation of prices for the (raw) data stored in the DWs. Additionally, each manager can also parameterize the negotiation agents of its own company according to its level of optimism, and the information available on the DWs under its administration.

Although a particular mechanism to determine the price of information outsourcing was presented hear, the framework is flexible enough in order to accommodate different estimation prices mechanisms, augmenting the possibilities for system configuration. In fact, we are currently working in a formal approach to this topic in order to categorize precisely a large range of approaches to this specific problem. We also intend to improve the mechanism to determine if the information that is necessary to submit in a bid is valuable or not to the current status of an enterprise's decision-making process. We also intend to specify the negotiation protocol in a formal language and use temporal logic to specify some desirable properties that it should satisfy.

7. REFERENCES

[1] Alcino Cunha and Orlando Belo. An electronic commerce framework for resource allocation among multi-agent enterprises. In *10th International FLAIRS Conference (FLAIRS'97)*, Daytona Beach, Florida, May 1997.

[2] Alcino Cunha and Orlando Belo. A multi-agent based approach for load distribution in multi-enterprise environments. In *IASTED International Conference on Applied Informatics (AI'97)*, Innsbruck, Austria, February 1997.

[3] R. Davis and R.G. Smith. Negotiation as a metaphor for distributed problem solving. In A. Bond and L. Gasser, editors, *Readings in distributed problem solving*, pages 333–356. Morgan Kaufmann, 1988.

[4] F.Cheong. *Internet Agents - Spiders, Wanderers, Brokers, and Bots.* New Riders Publishing, 1996.

[5] M.Wooldridge and N.Jennings. Intelligent agents: theory and practice. *Knowledge Engineering Review 10*, 1995.

[6] O.Belo. A Hardmetal Tools and Wear Parts Production System Simulation. In *Proceedings of the 8th European Simulation Symposium (ESS'96)*, Genoa, Italy, October 1996.

[7] O.Belo. Gathering the Right Information at the Right Time - An Agent Based Approach to Data Warehouses Loading Processes. In *Proceedings of the First International Conference on Enterprise Information Systems (ICEIS99)*, Setubal, Portugal, March 1999.

[8] P.Gray and H.Watson. *Decision Support in the Data Warehouse.* The Data Warehousing Institute Series from Prentice Hall Ptr, 1998.

[9] R.Barquin and H.Edelstein, editors. *Building, Using, and Managing the Data Warehouse.* The Data Warehousing Institute Series from Prentice Hall Ptr, 1997.

[10] Tuomas W. Sandholm. An implementation of the contract net protocol based on marginal cost calculations. In *Eleventh National Conference on Artificial Intelligence (AAAI-93)*, pages 256–262, Washington D.C., 1993.

[11] Tuomas W. Sandholm and Victor R. Lesser. Issues in automated negotiation and electronic commerce: Extending the contract net framework. In *First International Conference on Multiagent Systems (ICMAS - 95)*, pages 328–335, San Francisco, 1995.

[12] Reid G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, C-29(12):1104–1113, December 1980.

[13] S.Russel and P.Norvig. *Artificial Intelligence - A Modern Approach.* Prentice Hall, 1995.

[14] Michael P. Wellman. The economic approach to artificial intelligence. *ACM Computing Surveys*, 27(3):340–342, September 1995.

[15] W.Inmon. *Building the Data Warehouse.* John Wiley & Sons, 1996.