INTELLIGENT AGENT AIDED USE OF UNSTRUCTURED INFORMATION IN DECISION SUPPORT

Dejan Lavbič, Viljan Mahnič
University of Ljuljana
Faculty of Computer and Information Science
Tržaška 25, 1000 Ljubljana, Slovenia
Tel: +386 1 476 83 67; fax: +386 1 476 87 04
e-mail: Dejan.Lavbic@fri.uni-lj.si

ABSTRACT

There is more than 90% of unstructured information in organizations, but still majority of information engineers mainly deal with structured data. The answer may simply be that we know this best and basically all information systems are based on structured repositories like relational databases. Decision support systems often prove as inefficient due to the lack of intuitive support and active involvement of business user which drives to limited view of the decision problem. Data warehouses are often used for this purpose, but like many other tools for decision support they are also based on structured data. This article deals with integration of data warehouse and unstructured information with the use of ontologies and intelligent agents. The purpose of this synergy is to support associative thinking of business user and facilitate systems support.

1 INTRODUCTION

Majority of business intelligence (BI) systems are based on structured data, ie. data with predefined format (number, character, date etc.) and known location within electronic record. It is still a very demanding task to build systems like this due to increasing emphasis on searching while the quantity of data is rising uncontrollably.

Unstructured data from various sources, like forms, e-mails or documents, contains a lot of information that can be applicable in existing information systems. It is a very important fact that data don't contribute added value if they are not employed in the process of decision making.

The problem of using unstructured data was already addressed by Bill Inmon [1] who argues that accessing unstructured data is only the first step in filtering of usefull information. When unstructured data are read it has to be edited and sorted by priority. Problem in unstructured data is exactly that – it is unstructured. Because structure and format are undefined it is hard to determine important parts of unstructured data.

In data warehouse we work mainly with structured data. They alone are not sufficient for optimal decision making because majority of usefull information is still located in documents and not relational databases. This article therefore presents integration of both types of information into unified system. The purpose of this system is to efficiently aid business user in decision making process.

The article is organized as follows. Section 2 briefly presents background on data warehouses and section 3 on intelligent agents and Multi-Agent Systems. The main part is presented in section 4 with presentation of architecture of solution, conceptual plan and short evaluation. Finally section 5 concludes this contribution.

2 DATA WAREHOUSES

Organisations within their day-to-day operations collect immense quantity of data. These data are usually kept in transactional systems with primary task of maintaining consistency, effective updating, safe simultaneous access, effective execution of numerous small transactions and securing the data in case of failures. Transaction system also provide quick response time especially when using SQL queries [2].

Organisation of transactional oriented databases is not suitable to be directly used in business applications. Business users and analysts require aggregated and summarized information, comparisons by time and space dimensions, synthesis by numerous records, trends in data and other complex analyses that support their tactical and strategical decision support process.

Transactional systems were not designed for decision support, therefore new type of systems to support decision support has evolved. They are usually named analytical systems or BI systems. Data warehouse (DW), optimized for complex analyses and query execution for more than thousand records, plays the main role in these systems [3, 4]. The most widely used BI solutions are OLAP (On-Line Analytical Processing) systems and introduce functionality of interactive exploring of data warehouse.

Data warehouse combines data from heterogeneous data sources to provide united source of information that user can query. Data in DW is organized in a way that enables direct support to management functions.

Relational data model is not appropriate for DW, because all tables are treated symmetrical and users have difficulties comprehending it and nevertheless is also less

effective in execution of analyses that require several conditions. Therefore dimentional data model is used for OLAP analyses with fact table and related dimensional tables. Every record in fact table contains measurable fact according to combination of values from multiple dimensions.

Data warehouses present an important foundation for decision support systems. They form the platform for development of analytical application and are based under assumption that underlying data is at desired quality and cleanse in preceding ETL (Extract-Transform-Load) process.

3 INTELLIGENT AGENTS AND MULTI-AGENT SYSTEMS

Multi-Agent Systems (MAS) are gradually becoming a new paradigm in development distributed business information systems. With agent oriented technology more complex information systems can be developed with the help of natural decomposition of problems, abstraction and flexibility of management of organizational changes [5].

Interest in research of intelligent agents and MAS has been increasing in the last two decades. The mainstream of research is focused mainly on business information systems [6, 7, 8], confirming that MAS is very appropriate for decision support information system development. Resemblance between agents from MAS paradigm and human actor in business organizations can be seen in characteristics and coordination styles. This leads to a modeling approach where intelligent agent in MAS plays a role of actor in business organizations. While the popularity and the quantity of applications with agent technology have gone up in recent years, the recent developments in this area include innovative approaches and architectures for management of integration between different systems.

Many definitions of an agent exists nowadays so it is hard to point one out. They all agree that an agent is a computer system, located in its environment and is capable of autonomous actions in this environment with a purpose of achieving predefined goals. Intelligent agent also share the properties like autonomy, reactivity, proactivness and social behaviour. Agents are also often modeled using abstract concepts like knowledge, belief, desire and intention, while objects on the other side simply encapsulate their inner structure with methods and attributes. Level of autonomy differs from agents to objects – object don't have control over their execution and are initiated by other entity, while agents can, after receiving a request, decide whether to execute requested action or not.

In the following section 4 the use of intelligent agents in the domain of data warehouse will be shown with the emphasis on organizational view of MAS and roles of agents in integration and acquiring of information.

4 THE USE OF INTELLIGENT AGENTS IN DATA WAREHOUSES FOR INTEGRATION WITH UNSTRUCTURED INFORMATION

4.1 Introduction

The most important part of the decision support system is the business user who makes all decision. Business user is usually not a specialist in modern technologies in knowledge management but is rather an expert in a problem domain where he works. This section therefore presents framework for implementation of data warehouse and unstructured information integration. The purpose of this approach is to facilitate decision support process where business analyst has overall view of the data, required for making decisions.

4.2 Architecture

Figure 1 presents organizational structure of intelligent agents supporting data warehouse. There are several roles of agents: **OLAP Agent** (OA), **Information Retrieval Agent** (IRA) and **Knowledge Discovery Agent** (KDA).

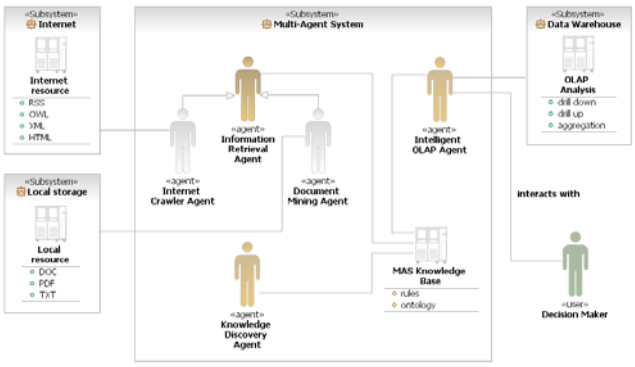

Figure 1: *Use case of MAS for decision support.*

IRA is responsible for acquisition of information from various sources and is furthermore specialized in **Internet Crawler Agent** (ICA) and **Document Mining Agent** (DMA). The role of ICA is to autonomously search for information available on the Internet, related to the OLAP analyses. As depicted in metamodel in figure 2 ICA acquires different information (news, forum opinions, stock reports etc.) about the selected problem domain. Resources being searched are also identified automatically to achieve highest level of autonomy as possible. DMA is on the other hand oriented towards finding usefull data from local document repository, including specifications, contracts etc. in different formats (DOC, PDF, TXT etc.). Role of KDA is employing methods of artificial intelligence and analytical methods for identification of important facts and patterns in existing data. The derived knowledge is therefore directly used to aid business users in their decision making.

The central element of Multi-Agent System is MAS knowledge baze (MAS-KB) with its data vocabulary, used for standardization of used concepts. As depicted in figure 1 all collaborating agents use MAS-KB either for storing

found information or deriving new knowledge. With advent of Semantic Web the availability of tools for implementation of this kind of mechanisms is increasing.

4.3 Integration of unstructured information with DW

Before proceeding with the use of unstructured information in DW let us identify some possible approaches of integration that are available on the market [9, 10]. There are approaches using ETL tools for unstructured information examination and extracting usefull information and transforming to structured form. Another approaches emphasize semantic handling. Hybrid approach that we introduce is based on several concepts from basic taxonomy (hierarchical structure of classification) and ontology (rules and different view on taxonomies) to describing data for enabling context and associative thinking [11].
The idea of integration is depicted in figure 2 where all important elements are presented. Metamodel includes elements of data warehouse and unstructured information from various sources.

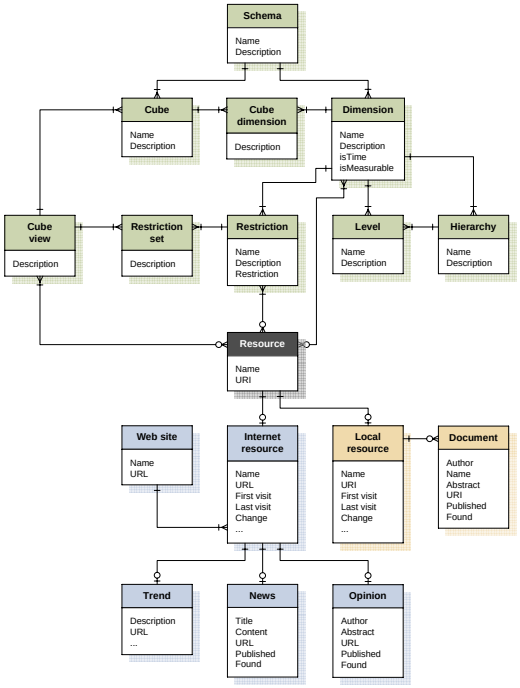

Figure 2: *Metamodel of using unstructured information in OLAP analyses.*

Elements relating to DW are as follows – schema, cube, dimension, cube dimension, cube view, level, hierarchy, restriction and restriction set. Element related to unstructured information is depicted as resource and further specialized into internet resource and local resource.
The most important part of the integration is linking element between unstructured information and process of OLAP analysis. To benefit from overall view on the information space we introduce the use of topic maps [12, 13] that contain metadata needed for navigation between data. Problem with OLAP analyses nowadays lies in preprepared and static reports. These reports have been prepared by users with technical background and will be used by business users who generally don't posses that kind of technical expertise. Let us consider for example a business user from a trading company that makes a decision about a supplier of chocolate bars. Decision maker is primarily interested in movement of sales of several chocolate products in the nearby history, which branch was the most successful, what is the most appropriate stock quantity etc. All these analyses can be accomplished by using OLAP tool on structured data found in DW, but generally this is not sufficient to make an optimal business decision.

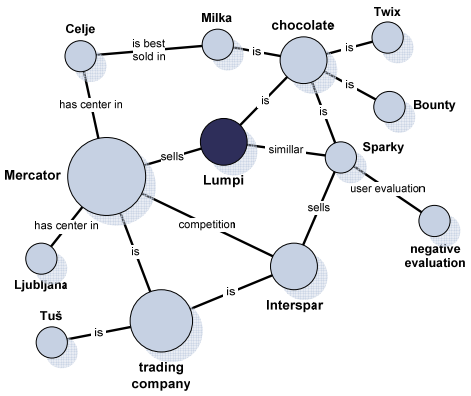

Figure 3: *Ontology in a form of Topic Map for traversing through information space.*

It can be very important how selected supplier is presented in the media or are there any associated scandals, open contracts, what is their financal report and nevertheless what is the general opinion of potential buyers that would trading company include in its sell program. All of these can be achieved with Topic Maps as a mechanism for management of metadata within data warehouse. It is all based on associative thinking paradigm where focus is in selected moment concentrated on specific concept and its context. If someone is interested in Lumpi chocolate bar (see figure 3) then this element becomes the observed concept with all related properties – who currently sells this product, what is the general opinion on forums, which products are in the same scope etc.

4.4 Implementation and evaluation

Implementation of presented approach is progressing in a form of prototype in Java programming language. Tools, used for implementation, are as following:

- **Mondrian** – OLAP server that enables interactive analyses on various datasources.
- **JPivot** – OLAP client used by business analyst used for standard OLAP functions – slice, dice, drill down, drill up etc.
- **Ontopia Omnigator** – supporting tool for building Topic Maps with emphasis on navigational aspects.
- **JADE** – platform for developing agents.

Concept of user interface used by business user is depicted in figure 4. When viewing and dynamically generating

reports business user has ability to use related information from the internet and local document repository and therefore aiding decision making process.

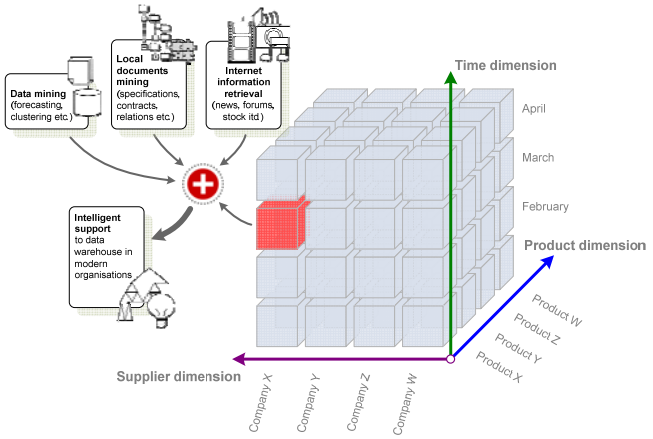

Figure 4: *Dynamic support to decision support process in data warehouses.*

In comparison to existing OLAP tools our approach present several advantages:

- When viewing information from selected problem domain our approach enables connections to other sources within observed context. Existence of navigation on information space where user can freely traverse based on mutual dependencies.
- When viewing report, it is feasible to view all related documents (news on the internet, documents from local repository, financal reports etc.)
- When making decisions all opinions are welcomed (internet forums).

5 CONCLUSION

In decision making information plays the most important role and that remained unchanged. The only thing that changed are types of decisions. In transition from industry oriented to knowledge oriented economy more information is needed to endorse decision support process. Dynamics currently present in business environment demands less information that can be classified into structured information model. It is kind of a paradox that information specialists label themselves as experts for information technology but they still constantly ignore more than 90% of relevant information. This is related to ratio between unstructured and structured information which reaches 9:1, while majority of information specialists mainly deals with minor part of unstructured information. Nevertheless unstructured information are still priceless source for decision support process. Business processes were always dependant on unstructured information and the quantity of unstructured information is only rising not vice versa. A good example is gaining importance of World Wide Web, e-mails, forums etc.

This article presented concept of linking unstructured information with data warehouse as one of the most important systems for decision support. Solution is presented by using agent oriented approach with emphasis on cooperation with business user while searching for information and exploiting navigational support. Problem was addressed with the use of ontologies as a metalevel above data warehouse and contains all required information about associations between data and metadata needed for navigation through information space. Solutions tries to support the process of associative thinking at decision making and enable business user access to required information as entirely as possible.

References

[1] B. Inmon. Information Management: The Noise is Terrific, *DM Review Magazine*. December 2004. http://www.dmreview.com/article_sub.cfm?articleID=1014519.

[2] C. J. Date. *An Introduction to Database Systems, 8th edition*, Addison-Wesley, 2003.

[3] W. H. Inmon. *Building the Data Warehouse, 3rd edition*. Wiley Publishing. 2002.

[4] R. Kimball. *The Data Warehouse toolkit – The Complete guide to dimensional modeling, 2nd edition*. John Wiley & Sons. 2002.

[5] R. Kishore, H. Zhang, R. Ramesh. Enterprise integration using the agent paradigm: foundations of multi-agent-based integrative business information systems. *Decision Support Systems*. Elsevier. 2003.

[6] N. Kang, S. Han. Agent based e-marketplace system for more fair and efficient transaction. *Decision Support Systems, Volume 34*. Elsevier. pp. 157-165. 2003.

[7] G. Tewari, J. Youll, P. Maes. Personalized location-based brokering using an agent-based intermediary architecture. *Decision Support Systems, Volume 34*. Elsevier. pp. 127-137. 2003.

[8] S. T. Yuan. A personalized and integrative comparison – shopping engine and its applications. *Decision Support Systems, Volume 34*. Elsevier. pp. 139-156. 2003.

[9] W. N. McCrosky. The Integration of Unstructured Data into a Business Intelligence System. *DM Review Magazine*. December 2004. http://www.dmreview.com/editorial/dmreview/print_action.cfm?articleId=1015671

[10] J. Ladley. Beyond the Data Warehouse: Primer for Unstructured Data and Semantics. *DM Review Magazine*. October 2005. http://www.dmreview.com/editorial/dmreview/print_action.cfm?articleId=1039846

[11] I. Lajovic. Panoramski pogled na baze podatkov. *Proceedings of Dnevi Slovenske Informatike 2005*. Slovenian society Informatika. Portorož. 2005.

[12] K. Olseviceva. Development of Topic Maps E-Learning Portal. *Electronic Journal of e-Learning, Volume 4, Issue 1*. March 2006.

[13] S. Pepper, G. Moore. XML Topic Maps (XTM) 1.0. *TopicMaps.org Specification*. http://www.topicmaps.org/xtm/1.0. 2001