Scrutinizing Web Mining Technology and Multi Agent Systems into Intelligent Information Retrieval System (IIRS)

Vijay Karan Singh Anand
Senior Software Engineer
Prodigi Pvt. Ltd, Hogarth WW
India

Jagjit Singh
Associate Technology Level 2
Sapient Consulting Pvt. Ltd

Harmeet Singh Saimbhi
Business Analyst ERP
Implementation
ProbitySoft Pvt.Ltd

ABSTRACT

Users get muddled while searching for specific documents on web. Retrieving relevant information from reservoir of documents is a cumbersome task. It is like finding a pixel of image from big picture gallery. Web mining is one of technologies that uses concept of data mining to explore useful prefatory patterns and relevant information on a given domain.

The paper describes the use of web based content mining technology that in incorporates role of multi agent systems to perform an intelligent search. In addition to this, an information retrieval (IR) system has been proposed that helps in deducing various inferences by interacting with agents and recording their behavior. This multi agent system (MAS) design can be implemented by using JADE (Java Agent Development Environment)

Keywords
Web mining, Data mining, Multi Agent Systems (MAS), Deductive databases and JADE.

1. INTRODUCTION
Web mining means to retrieve information from web. But this information is useful only if it is relevant or specific to particular domain thus satisfying user's selection. The web is growing at raped rate that is beyond human scope as well as traditional search engines. It evolves the need of data mining technologies that helps in identification of useful patterns and information from large collection of web documents or repositories [1].

The following paper also describes the use of agents in designing of proposed intelligent system. A system is called intelligent because it incorporates the use of multiple agents for distributing collection of data and enables communication of data in different environments and applications [2]. Each agent is associated with its specific behavior and posses several characteristics which helps in making retrieval process as intelligent and affective.

The paper is categorized as follows: section 2 gives brief overview of MAS. It also presents   hierarchical relationship of web mining technologies that is represented as ontology. Section 3 describes web based context mining methodology that is being involved in designing of an IR system. Section 4 presents layout of proposed MAS. Section 5 concludes the following paper.

2. MULTI- AGENT- SYSTEMS (MAS)
The systems in which individual agents derive new inferences and deductions by interacting with other agent in given environment are called as MAS. These systems are best way to retrieve knowledge from complex environment. MAS are superior to knowledge management (KM) solutions that only discover knowledge from centralized systems.

In MAS, an agent can be said as an entity with/without body. Each agent is supplied with following characteristics [3].

Fig 1: Features of Agents

The above features of agents make them intelligent that helps in deriving new inferences by maintaining interoperability among other agents in given system. Agent accesses web database to provide optimal solutions of given problem.
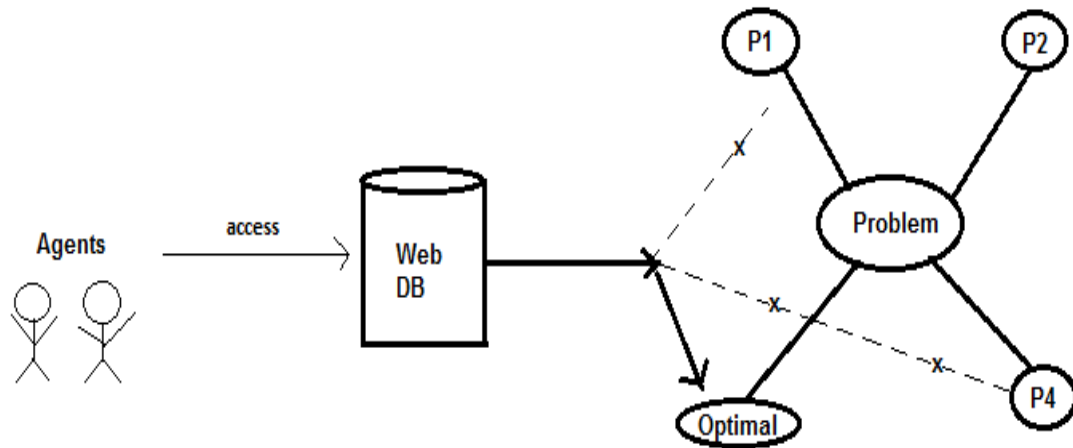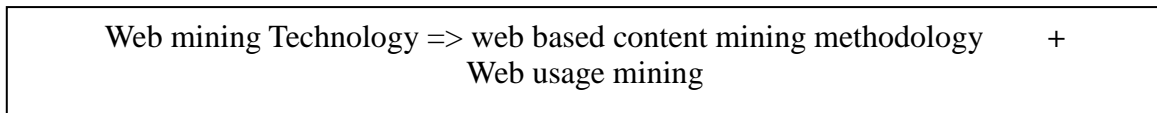
Fig 2: How agents solve complex problems

2.1. Taxonomy of web mining technologies

Web mining technology is categorized mainly into two sub technologies which are further composed of sub parts.

Web mining Technology => web based content mining methodology   +
Web usage mining

Web based content mining is being described in next section. Web usage mining is defined as technology that extracts information and patterns on basis of percentage of user accessing the web. A web toolbar maintains record of recent searches and displays information knowledge based on those searches to users [4] [5]. Web based content mining composed of attributes like multiple agents for performing different tasks (history agent, user agent and facilitator agent) and uses logic rules to derive relationship/ inferences among different users.
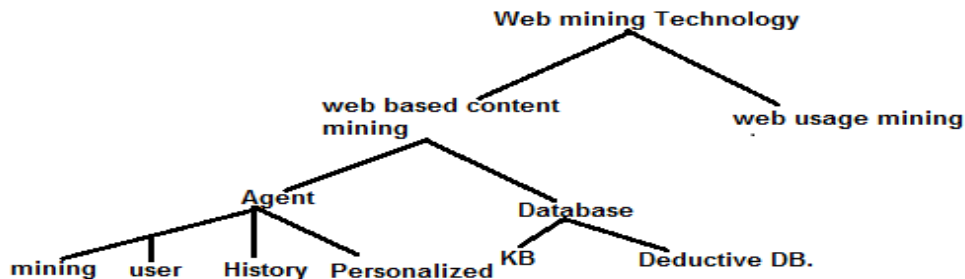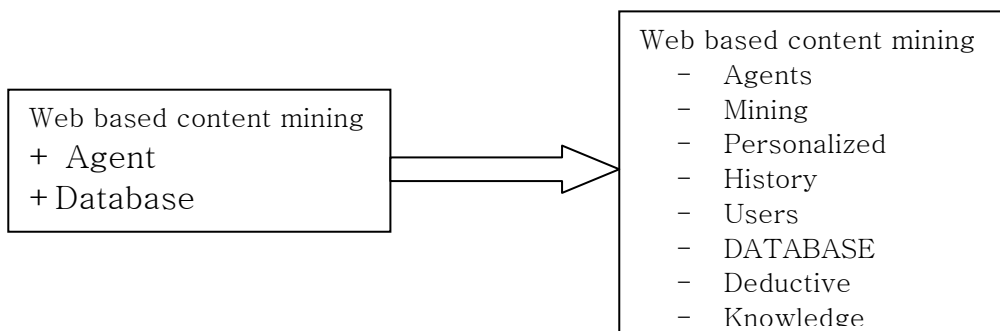

Fig 3: Ontology and web Technologies

Fig 4: Defined classes & Attributes

3. WEB BASED CONTENT MINING METHODOLOGY

This technology extracts contents of web pages and potential patterns by using concept of agents and data mining. It uses multiple agents to perform specific tasks like user agent is one who provide requirements/ searches for a given data. Similarity, there are other agent too. The data that has been sent from web DB gets transformed and stored into knowledge base (KB). In short, we can say that this methodology involve two databases (KB and deductive), multiple agent and concept of data mining.
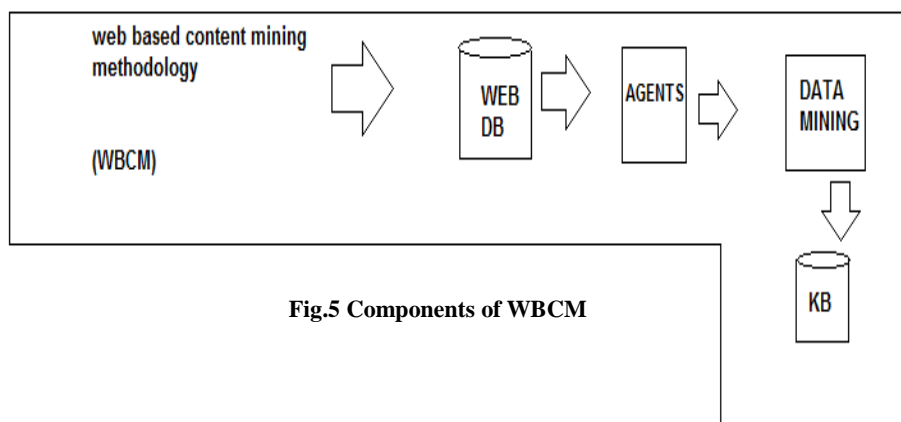

Fig.5 Components of WBCM

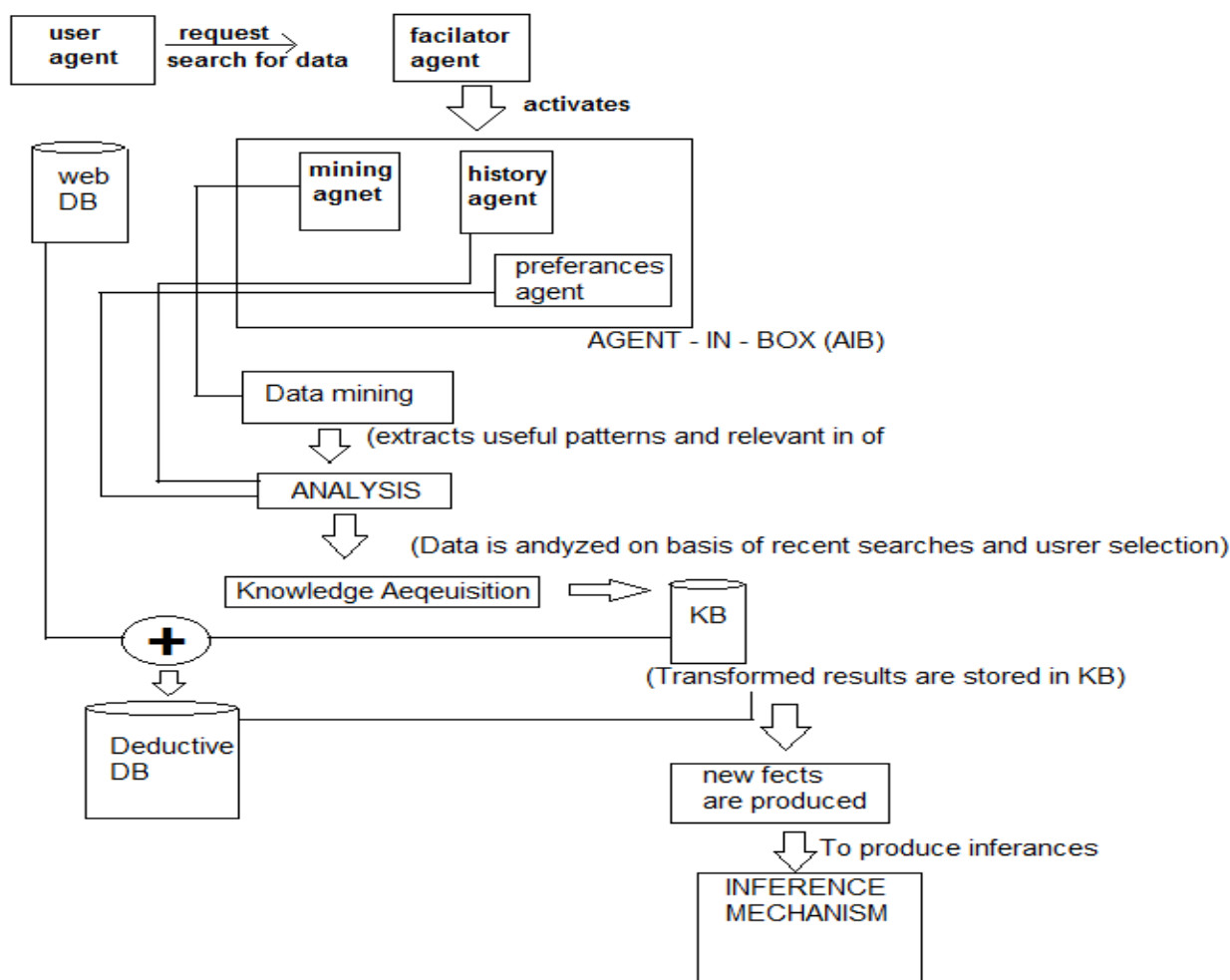## 4. PROPOSED IR SYSTEM

A layout has been show below:-

Fig.6 Layout

4.1. Working

The proposed system works as follows:

(a) The amount of information available on web is stored in web databases.

(b) There is an interface called as agent – in – box (AIB) that consists of various agents to perform specific tasks. AIB acts as interface between web DB and KB.

(c) User agent searches for data by sending request to facilitator agent.

(d) Facilitator agent starts processing request. It activates different agents to extract and analyze data

(e) After processing of data, it is mined to extract potential patterns, trends and information.
This responsibility is taken by mining Agent.

(f) Now, data is being analyzed on basis of user recent searches and preferences. It is made possible by history agent and preferences agent respectively.

(g) It leads to produce some relevant information that in called as knowledge acquisition or knowledge gathering.

(h) The information may be in form of facts. These facts are stored in KB.

(i) Now, apply some rules an facts stored in KB. It will lead to concept of deductive databases

$$\text{Deductive DB} = \text{Web DB} + \text{KB}$$

(j) Rules are defined as follows:-

Superior $(x,y) \leftarrow$ supervise (x,y) **(1)**
Superior $(x, y) \leftarrow$ supervise (x, z) superior (z,y) **(2)**
Subordinary$(x, y) \leftarrow$ Superior (y, x) **(3)**

(k) It leads to generation of new fact to deduce relationship/ inference among different users.

Consider an example →

Let x and Y be two different users who are surfing online shopping website. X has selected tshirt of brand z . Other t-shirt of same brand z is being selected by Y.

So, by applying rule 2, we can say that X and Y are related to each other and share common property "brand"

X→Z supervise (X, Z)
Z→Y Superior (Z, Y)

$$X \rightarrow Y$$ Superior (X, Y)

5. CONCLUSION AND FUTURE WORKS

The paper employs use of web content mining methodology and multi agent systems to propose an intelligent information retrieval system (IIRS). Each agent has its features and associated behaviors that lead to react intelligently and retrieve accurate results. Web mining is one of technologies that uses concept of data mining to explore useful prefatory patterns and relevant information on a given domain. WBCM

involves two databases-knowledge base and web database that is collectively equivalent to deductive database.

As a future work, the implementation of above multi agent inference system can be done using JADE (java agent Development environment). JADE is treated as agent middleware that is used to implement agent development framework [6] [7].

6. REFERENCES

[1] Gagandeep Singh Narula et.al. "Implementation of Data Mining in online shopping system using TANAGRA tool", "International Journal of Computer Science and Engineering (IJCSE), Vol.2 Issue 1, ISSN 2278-9960", Feb 2013, pp 47-58

[2] J. Mayfield, "Ontologies and text retrieval", "Knowledge Engineering Review", 2007.

[3] Vishal Jain, Gagandeep Singh, Dr. Mayank Singh, "Implementation of Multi Agent Systems with ontology in Data Mining", "International Journal Of Research In Computer Application & Management (IJRCM), Vol. No. 3, Issue No.1 ISSN 2231-1009", January 2013, pp 111-117

[4] Xinlin Zhang, Xiangdong Yin, ' Design of an Information Intelligent System based on Web Data Mining' ,International Conference on Computer Science and Information Technology 2008.

[5] WangBin, LiuZhijing, , ' Web Mining Research' , International Conference On Computational Intelligence and Multimedia Applications, IEEE (2003).

[6] Gagandeep Singh et.al., "An Approach For Information Extraction using Jade: A Case Study", 'Journal of Global Research In Computer Science (JGRCS)', Vol.4 No. 4, April 2013, page no. 186-191, ISSN No. 2229-371X

[7] James Huamonte, Kevin Smith, 'The Use of Roles to Model Agent Behaviors for Model Driven Architecture', 2005 IEEE.

[8] Dhanashree S. Deshpande, "A Survey on Web Data Mining Applications", "International Journal of Computer Applications (IJCA), ETCSIT- Number 3", 2012.

[9] M. Preethi, Dr. J. Akilandeswari, "Combining Retrieval with Ontology Browsing", "International Journal of Internet Computing (IJIC), Vol.1, Issue-1", 2011.

[10] Kobkul, Gareth Clayton and Phayung Meesad, "A Hybrid System based on Multi Agent System in the Data Preprocessing Stage", International Journal of Computer Science and Information Security (IJCSIS), Vol.7, Issue 2, pp 199-202, 2010

[11] O. Wrobel, A. Hui, J.M. Joller, "Data Mining for Ontology Building: Semantic Web Overview", "Diploma Thesis- Department of Computer Science WS2002/2003, Nanyang Technological University

[12] Amir Hussain et.al., "Agent based computing from multi agent systems to agent based models: a visual survey", Springer Scientometrics, 89(2), 479-499, 2011.

[13] Jin Oh et.al., "Distributed Intelligent Microgrid control using multi agent systems", (http://www.SciRP.org/journal/eng), 2013