INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

Indexed & Listed at:

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.,

Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],

Index Copernicus Publishers Panel, Poland with IC Value of 5.09 & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 2022 Cities in 153 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Management Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic and Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email address: infoijrcm@gmail.com.

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____

THE EDITOR
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF .**

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '_____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:

a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mention the following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE:** The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS:** The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT:** Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT:** Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS:** All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS:** All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT:** The main text should follow the following sequence:

INTRODUCTION

REVIEW OF LITERATURE

NEED/IMPORTANCE OF THE STUDY

STATEMENT OF THE PROBLEM

OBJECTIVES

HYPOTHESES

RESEARCH METHODOLOGY

RESULTS & DISCUSSION

FINDINGS

RECOMMENDATIONS/SUGGESTIONS

CONCLUSIONS

SCOPE FOR FURTHER RESEARCH

ACKNOWLEDGMENTS

REFERENCES

APPENDIX/ANNEXURE

It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES & TABLES:** These should be simple, crystal clear, centered, separately numbered &self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS:**These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES:** The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, KurukshetraUniversity, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

IMPLEMENTATION OF MULTI AGENT SYSTEMS WITH ONTOLOGY IN DATA MINING

VISHAL JAIN
RESEARCH SCHOLAR
COMPUTER SCIENCE & ENGINEERING DEPARTMENT
LINGAYA'S UNIVERSITY
FARIDABAD

GAGANDEEP SINGH
STUDENT
GURU TEGH BAHADUR INSTITUTE OF TECHNOLOGY
GGS INDRAPRASTHA UNIVERSITY
DELHI

DR. MAYANK SINGH
ASSOCIATE PROFESSOR
COMPUTER SCIENCE & ENGINEERING DEPARTMENT
LINGAYA'S UNIVERSITY
FARIDABAD

ABSTRACT

In past years, there was no concept of Ontology and Semantic Web. At that time, researchers uses the concept of Knowledge Management Solutions (KM) for identifying patterns and trends from large quantities of data. With advent of decisional data processes in 1990's, concept of Databases and Data Warehouse came into existence. Keeping concept of Data Mining and Distributed Databases in mind, this paper emphasis on Evolution of Ontology, role of Knowledge Discovery (KD) processes, formation of new concepts and approaches with ontology in Multi Agent Systems (MAS). It also solves the problem of classifying and analyzing web documents by making use of ONTOLOGY WEB CONTENT MINING methodology and implementing them by using WORDnet.

KEYWORDS

Data Mining, Knowledge Discovery (KD), Peer-to-Peer Architecture, Multi Agent Systems (MAS), Ontology, Distributed Data Mining (DDM), Agent Based Distributed Data Mining (ADDM), WORDnet.

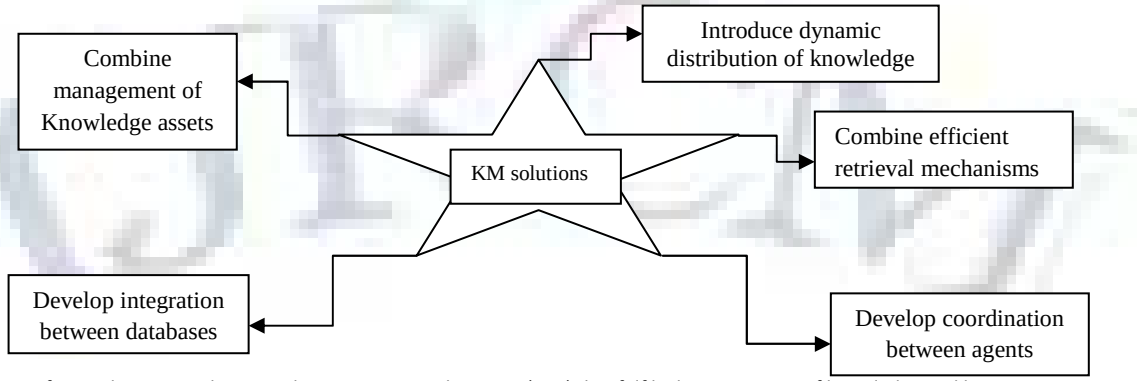
INTRODUCTION

In recent years, there was a great demand of Knowledge Management (KM) solutions and are used in organization as tools for performing many tasks. These tasks include:

- ✓ Document Management and Workflow Management
- ✓ Web Conferencing
- ✓ Data Warehouse and Decision Support Systems

But these conventional KM solutions are unable to become part of organization and committee because these solutions are based on centralized architecture which is storehouse of central knowledge only that is accessed through standard ontology. These KM solutions are not able to access decentralized information located at different networks. *After these uncertainties with KM solutions, the concept of Ontology and Semantic Web was introduced* [1]. Besides this, the architecture given by conventional KM solutions is not suitable for integration of knowledge and business requirements. It is evident from requirements of KM solutions as shown below:

FIGURE - 1: REQUIREMENTS OF KM SOLUTIONS



According to requirements of KM solutions, we have used *Peer-to-Peer architecture* (P2P) that fulfils the integration of knowledge and business processes.

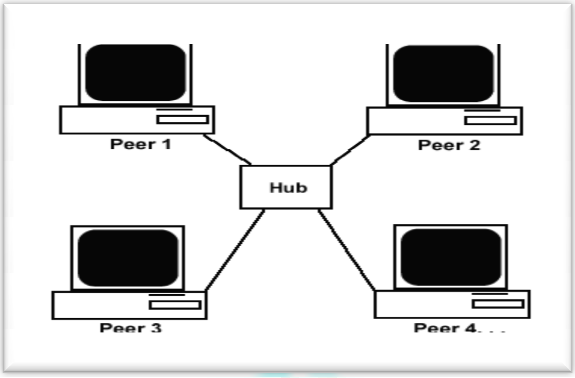
Peer-to-Peer (P2P):- It is technology that gives way of enabling distributed control of knowledge and diversity of contexts. It describes large collection of concepts related to distributed knowledge and it facilitates all requirements of KM solutions. This P2P architecture represents organization as distributed processes that manage the KM of individual users and maintain coordination among them

"*Although P2P is powerful paradigm when it is used to satisfy user requirements of Knowledge level KM solutions. But it is not suitable to complex systems* [2]."

The answer to this statement lies below:

- o It doesn't give information about the social skills of peers. Peers are the users in P2P terminology.
- o Current implementation of P2P based systems trust peers as entities with limited social skills like Search, Locate and Send. But in KM solutions, we need complex social skills like collaboration, coordination etc. These skills are not available with P2P systems.

FIGURE - 2: P2P ARCHITECTURE



The remaining sections of paper are as follows: Section 2 gives information about *Evolution of Ontology in Multi Agent Systems*. It also relates to concept of agents. This section presents approach involved in formation of concepts and improving knowledge sharing among agents .Section 3 describes concepts of *Data Mining and its classification*. It is categorized into *Distributed Data Mining* (DDM), its architecture, technology and protocols used for systems. It also defines brief introduction about *Multi Agent Based Distributed Data Mining* (MADDM) and its various agents. Section 4 deals with methodology for classification of web documents. It is done through experimentation with *WORDnet*.

2. EVOLUTION OF ONTOLOGY IN MULTI AGENT SYSTEMS

With advent of concept of Ontology and Semantic Web, we are able to use KM solutions in different environments where agents can define their own ontology. An agent is able to perform following tasks with use of Ontology:

- ✓ Agents can distribute collection of data and enable communication of data in different environments and applications.
- ✓ Interface Description Languages and services are provided for different environments where Interface Language refers to defining of data objects and their location.
- ✓ Agents are able to access some distributed models to enable interaction between processes like CORBA (Common Object Request Broker Architecture), RMI (Remote Method Invocation).

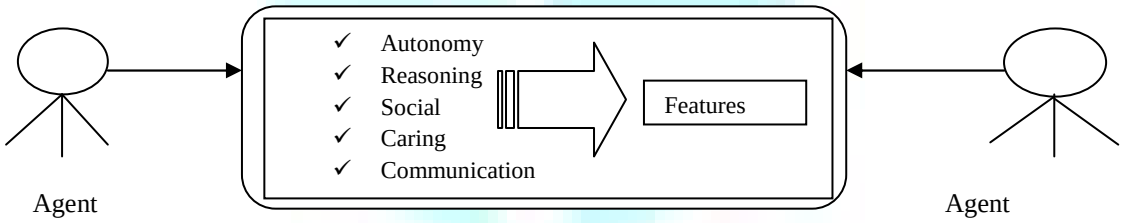
2.1 MULTI AGENT SYSTEMS (MAS)

Problem: - While using KM solutions for integration of knowledge processes, we have used P2P architecture for enabling distributed control of knowledge resources at centralized location via peers. Peers are directly connected to Hub. But this architecture is not suitable for accessing knowledge from complex systems. Then what is way to retrieve knowledge from systems in complex KM environment.

Solution: - Multi Agent Systems (MAS)

Analysis: - Systems where individual self authorized agents derive new facts with the help of other agents are called *MULTI AGENT SYSTEMS*. In these systems, individual agents create their different models and prototypes instead of following standard ontology. *An agent is defined as an entity with/without body*. Agents use different kinds of knowledge sources and resolve differences among themselves to provide best answers of queries in complex KM environment. We can also define MAS in terms of characteristics of agents as follows:

FIGURE 3: CHARACTERISTICS OF AGENTS



Ontology is evolved by learning concepts and relationships by taking guidance from other agents. We need to make aware of concepts among agents in order to improve knowledge sharing which is efficient for communication.

Approach involved: - There are following assumptions to be kept in mind which are as follows:

- Assume given organization has n agents *Ag1…..Agn* where each agent manages knowledge for its organization.
- Each agent knows concept which is denoted by *Ck.*
- Each agent has positive and negative thoughts (*ti*) with respect to that concept.
- Each agent has learning algorithm (*Li*) or classification mechanism.
- Each agent has its unique Ontology denoted as *Oi.*
- Each agent has set of features for representing concepts. It is denoted as *fi*.

It is represented as: **Agn = {Li, ti, fi, Oi}**

(a) **Learning Algorithm (Li):** - An agent learns a concept under supervision of teachers. It is also possible that teachers are not well expert about given queries, in that case they can use learning algorithm to give example regarding queries.

(b) **Set of thoughts/examples (ti)**: - In this, positive and negative examples are taught to agents by teacher. These examples are related to given problem domain. Using this classification capability, learner agent is able to decide whether example is positive or negative.

(c) **Set of features (fi)**: - They are most important factor to represent concepts. Features are selectively used by each agent to represent different concepts. In Multi Agent concept learning, we have to collect most related features from different sources of knowledge in order to develop a comprehensive concept.

(d) **Ontology (Oi)**: - In terms of concepts and knowledge sharing, it is defined as mixture of Meta concepts and fine grained structure. Meta concepts are concepts which are divided into sub concepts until fine grained concept level is reached. It is preferred to use own ontology rather than using standard ontology. We can achieve ontology- based semantic integration [3, 4] which means that we can resolve differences that arises during run time interaction of system and agents. It is done in following ways:

- ✓ Using a single centralized ontology for all agents and application domain.
- ✓ Merge source ontology into a common ontology to prevent overlapping of concepts by various agents.
- ✓ Search set of mappings or matches between two ontologies when it is difficult to merge them.

2.2 OVERVIEW OF CONCEPTS

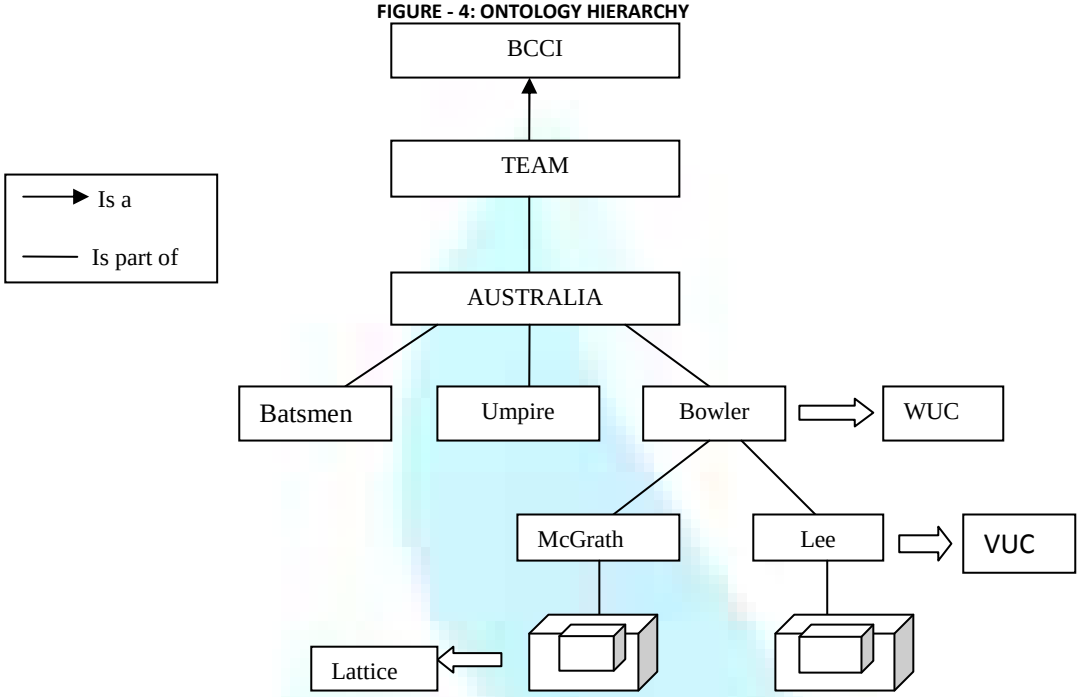
Before forming concepts, we categorize concepts in two levels:

(a) **Well Understood Concepts (WUC):** - Concepts which are known and clearly understood are called WUC. They have nodes in their ontology to show concepts. Agent searches its ontology to find concept. If features are matched, agent is expert and it has positive and negative examples which are assigned to WUC.

(b) **Vague Understood Concepts (VUC):** - Concepts that are not clearly mentioned in ontology are called VUC. They are implicit i.e. they are understood by indirectly means. VUC are described through lattice structure that shows relations and concepts in form of hierarchy. If features are not matched related to ontology, then there is no WUC and agent will draw lattice to find VUC.

Below we have shown features of Cricket Association in hierarchical manner. It is shown below:

FIGURE - 4: ONTOLOGY HIERARCHY

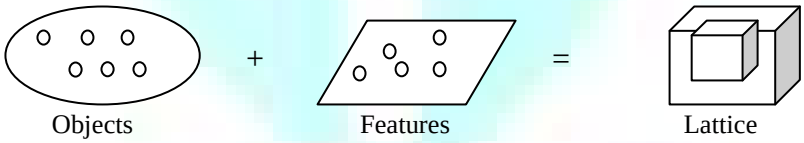


2.3 FORMATION OF CONCEPTS

In formation of concepts, *object refers to documents.* Agents have some objects and features corresponding to particular concept or set of concepts. It has following processes:

- Agent builds formal document and its related lattice. Lattice is structure that is improved with addition of objects and features.
- Concepts in lattice are generated automatically and under guidance of supervisor.
- Concepts are clearly labeled by the supervisor.

FIGURE - 5: FORMATION OF CONCEPTS

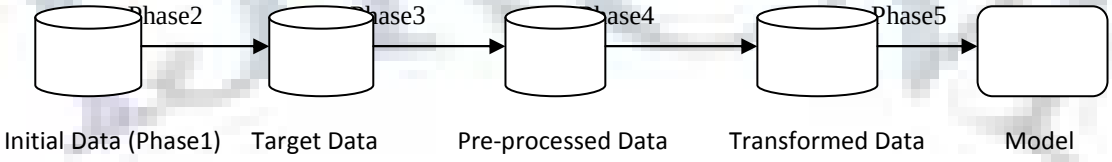


3. INTRODUCTION TO DATA MINING

Data mining is a technology that is used for identifying patterns and ways from large quantities of data or other information repositories. This technology works in a way that it adopts data integration method to generate Data Warehouse. With help of algorithm, it extracts agent information.

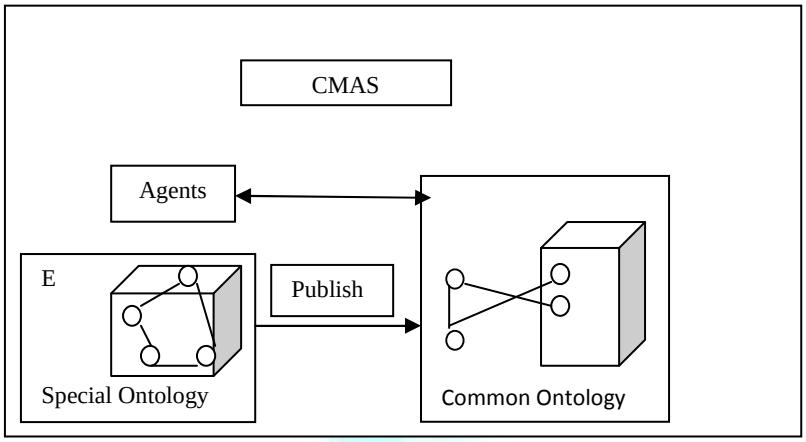
Data Mining involves use of several agents and data sources in which agents are configures to work on basis of mining algorithms to solve multiple tasks. Data Mining is originated from *Knowledge Discovery from Databases* (KDD) [5].

FIGURE - 6: KDD PROCESSES



Knowledge Discovery (KD): - It is defined as discovery of interesting, implicit and unknown knowledge from large databases. Conventional Data mining methods like Classification, Association have been introduced but they are unable to extract information from complex systems. To facilitate knowledge discovery in complex systems, we establish ontology to describe basis of system knowledge. *Special Ontology is written by experts in some specific format. Common Ontology is used to describe knowledge of system. It is written in formal language that is understood by all agents.*

FIGURE - 7: ONTOLOGY IN COMPLEX MULTI AGENT SYSTEMS [6]



Concept of Data Mining suffered from many challenges as shown below:

- There is huge amount of data located at different sites. To identify patterns from them can easily exceed limit.
- Since Data Mining involves large datasets therefore it is preferred to distribute data into fragments in order to achieve parallel processing.
- In many cases, data which is to be mined is produced at either higher rate or it comes in small packets. It may lead to wastage of data and less accuracy.
- Data extracted after mining process is not secure to transfer on all networks.

3.1 DISTRIBUTED DATA MINING (DDM)

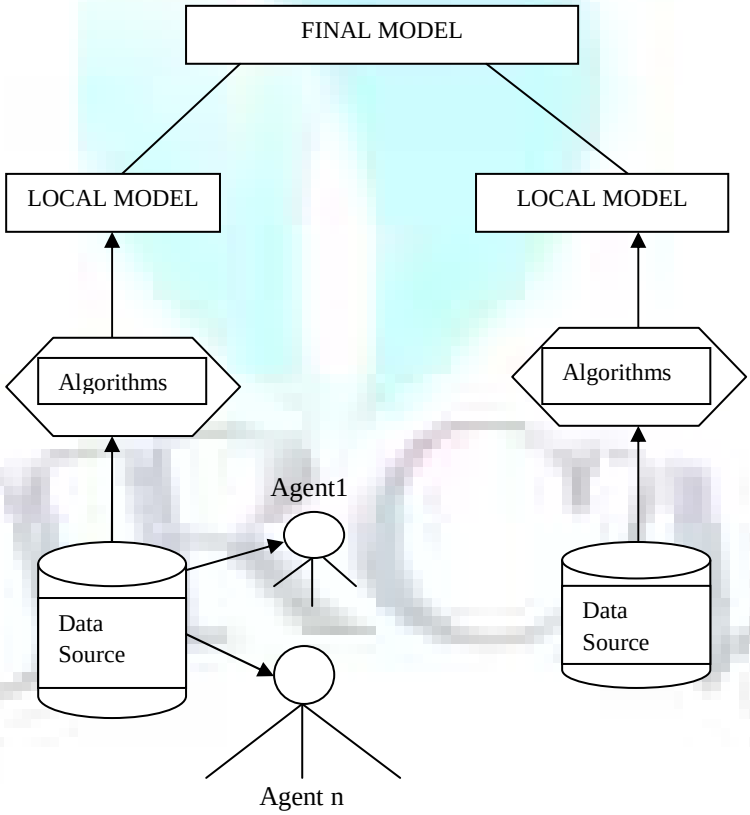
Problem: - Since huge amount of data is located at different sites and data is not transferred to other sites due to privacy issues. Single Data Mining techniques deals with centralized structure and to achieve faster processing, we have to distribute data sources into fragments.

Solution: - DDM

Analysis: - Objective of DDM is to perform data mining operations based on type and availability of data sources. It has algorithms to perform mining operations at centralized location and leads to distributed data.

Concept of DDM arises from its decentralized architecture that reaches every network. It increases security of data when it is transferred to other networks. DDM is complex system that focuses on the distribution of resources over wide network as data mining processes. It extracts useful patterns from distributed heterogeneous databases. *It has framework designed as an integrated GUI based environment that constructs the design process of multi agent system* [7]. It makes use of agents for updating these systems from time to time. It is shown below:

FIGURE - 8: DDM FRAMEWORK



3.2 ARCHITECTURE

Multi agent systems architecture are used for both *Distributed Data Mining (DDM*) and *Distributed Classification (DC)* systems. This architecture is organized as two level procedures as follows:

- First level deals with producing classifications on basis of particular data.
- Second level deals with source based classification.

DDM*: - It comprises components that handle source based data and tasks. It operates only in same host as that of data sources.* Distributed Data Mining Multi agent System (DDMMAS) *is categorized into two parts as shown in table 1.*

***DC**: - It is composed of components that handle metadata. It operates in any host*. Distributed Classification Multi Agent System (DCMAS) *is categorized into two parts as shown in table 2.*

Table - 1: DDM MAS

SOURCE BASED PART	META LEVEL PART
It has following components: • *Data Source Management Function:* - It participates in distributed design of ontology and collaborates with Meta level agents in testing and training method. • *KDD Agent of data source:* - *It* trains and tests source based classification agents. • *Training and Testing Services:* - It is not an agent. This component consists of software classes or tools to implement KDD methods.	It has following components: • *KDD Master/Meta Learning Agent:* - It performs training and testing of Metadata. It manages design of Meta model used in decision making. • *KDD Agent of Meta level:* - It trains and tests Meta level classification agent

FIGURE - 9: ARCHITECTURE OF DDM MAS

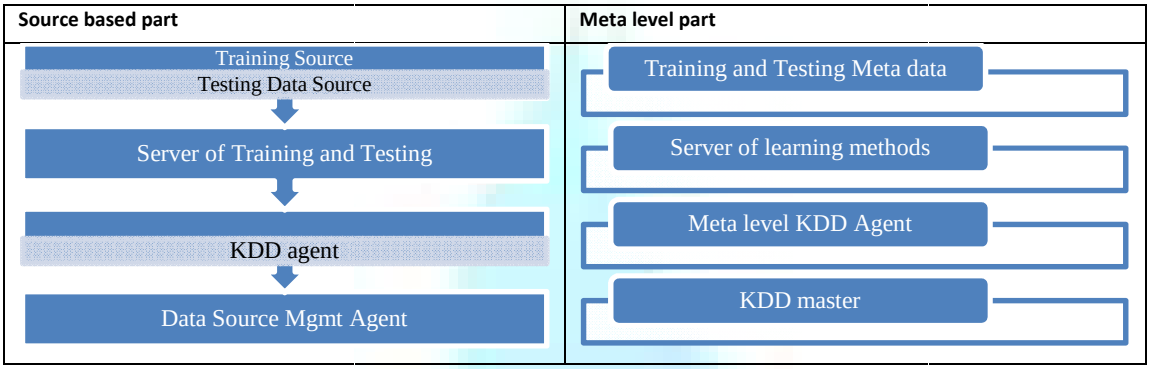
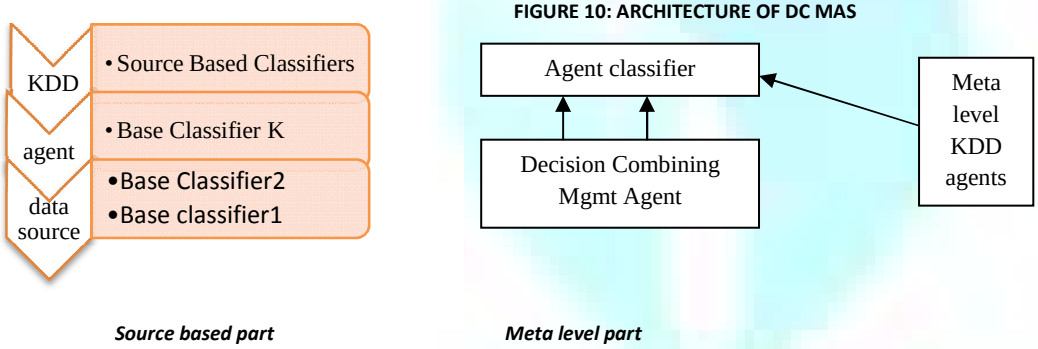


TABLE - 2: DC MAS

SOURCE BASED PART	META LEVEL PART
It has following components: • *Source Based Classification Agents:* - It consists of Base Classifiers. These classifiers are tested by KDD agent of data source.	It has following components: • *Agent Classifier:* - It is trained and tested by Meta level KDD agent. • *Decision Combining Management Agent:* - It coordinates operations of Agent classifier of Meta level and Meta level KDD agent

FIGURE 10: ARCHITECTURE OF DC MAS

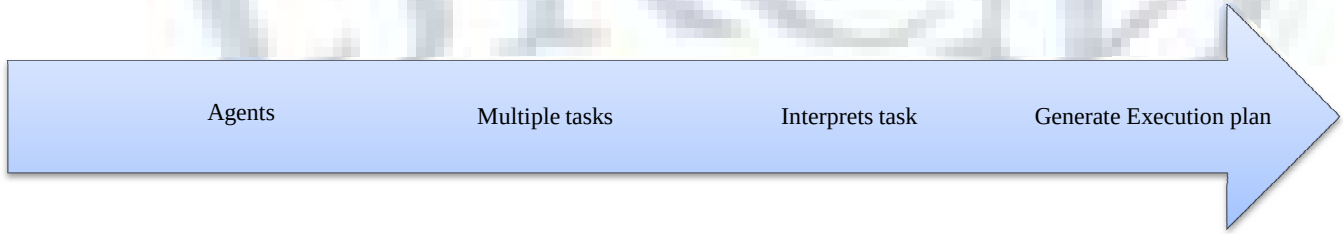


3.3 PROCESS FOR MINING DECENTRALIZED DATA BY AGENTS

Since decentralization of data is the main requirement for DDM systems. Data is distributed at each data site. Autonomous agents perform multiple tasks based on configuration of systems. Then agent interprets the configuration and generates execution plan to complete multiple tasks.

Agents can be transferred from one data site to other data sites. *It leads to dynamic organization of DDM system*. Each agent can view limited system. This limitation leads to better security as it prevents unwanted tasks from different hosts.

FIGURE 11: DECENTRALIZATION PROCESS



3.4 MULTI AGENT BASED DISTRIBUTED DATA MINING SYSTEM (MADM)

Problem: - In DDM systems, we can't obtain exact result from various processes as knowledge obtained from one model is different from another model irrespective of data is same. This makes DDM systems incompatible. Then what is way to obtain accurate knowledge among distributed systems.

Solution: - MADM systems

Analysis: - MADM system involves use of various agents to complete its task. It uses KQML as standard language that facilitates interaction among agents [8]. It has various agents which are described in table3 along with their functions.

TABLE 3: VARIOUS MADM AGENTS AND THEIR FUNCTIONS

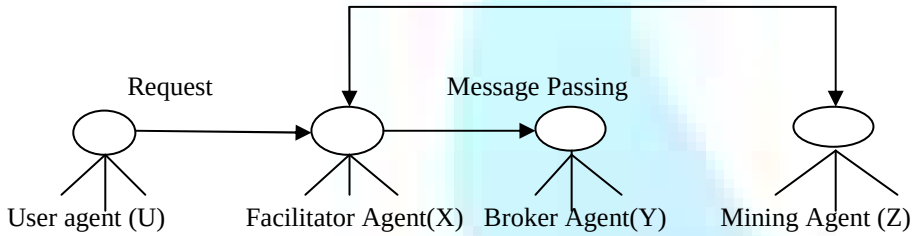
AGENTS	FUNCTIONS
(a) Interface Agent (User Agent)	It interacts with user to provide requirements and displays results. It has interface module that contains method for inter agent communication.
(b) Facilitator Agent (Management Agent)	It activates different agents. It receives questions from interface agent and may take the help of group of agents to solve those questions.
(c) Resource Agent (Data Agent)	It maintains Meta data information about data sources. It generates queries based on user request and sends their results to user agent.
(d) Mining Agent	It implements Data mining techniques and algorithms.
(e) Result Agent	It observes mining agents and other results from them. After obtaining results, these agents show results to user agent by integrating with manager agent.
(f) Broker Agent	It is advisor agent that can send reply to query of an agent with name and ontology of respective agent.
(g) Ontology Agent	It maintains and provides knowledge of ontology to solve queries related to ontology

3.5 FLOW OF SYSTEM OPERATIONS

Pre decision made in order to mine given data sources is called Data Mining Task Planning. Data mining task planning requires compensation between *Facilitator Agents* and *Mining Agents* through message passing.

Consider *User Agent* is denoted U. *Facilitator Agent* is denoted by X. *Broker Agent* is denoted by Y. *Mining Agent* is denoted by Z. If U sends request to X to ask for Data mining with other agents in system. Then X tries to compensate with Y to determine which agents are suitable for performing task. Mining Agent Z is responsible for completion of task while X is used for planning. When Z is completed, it shows results to X and X passes them to U.

FIGURE - 12: FLOW OF SYSTEM OPERATIONS



4. CLASSIFYING AND ANALYSING WEB DOCUMENTS

Many information extraction methods and techniques were used but they all are in vain. So we need more intelligent system to gather useful information from huge amount of data.

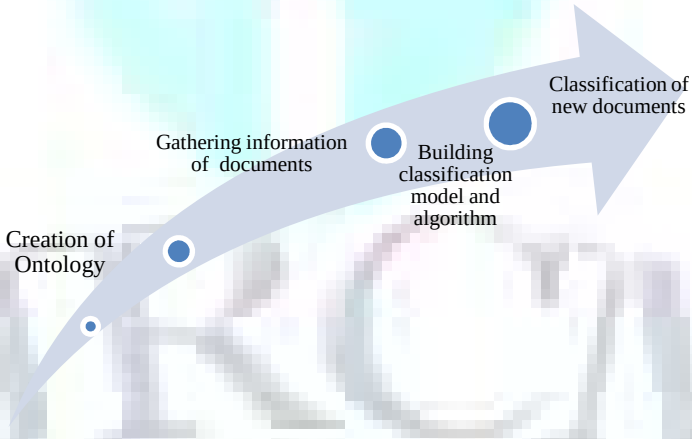
Problem: - To find meaningful and informative documents with help of Data Mining algorithms and then interpreting mining results in expressive way.

Solution: - Ontology Based Web Content Mining Methodology

Approach involved: - The proposed methodology uses concept of *Domain Ontology* [9]. Domain Ontology organizes concepts, relations and instances into given domain. This approach is used because it resolves synonyms and reducing confusion among agents.

Analysis: - ***Ontology Based Web Content Mining*** *represents conceptual information about given domain. It shows document representation, extraction of relevant information from text documents and creates classification models. This methodology is followed that uses the ideas and principles of Data Mining to analyze web data.*

FIGURE - 13: STAGES OF ONTOLOGY WEB BASED CONTENT MINING



In this section, we have evaluated proposed methodology in two specific domains- *weather domain* (web pages containing information about weather forecasting and analysis) and *GoogleTM* collection (web pages containing news).

4.1 ABOUT WORDnet

WORDnet was invented by A. Miller [10]. It is a large lexical database of English. Nouns, Verbs, Adverbs, Adjectives combine into sets of synonyms. Each of synonyms represents concept and solves queries through search and lexical results. In this paper, we have used WORDnet version 3.1. It contains 155287 synonyms. WORDnet is one of best example of ontology used in experiments.

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: ontology Search WordNet

Display Options: (Select option to change) Change

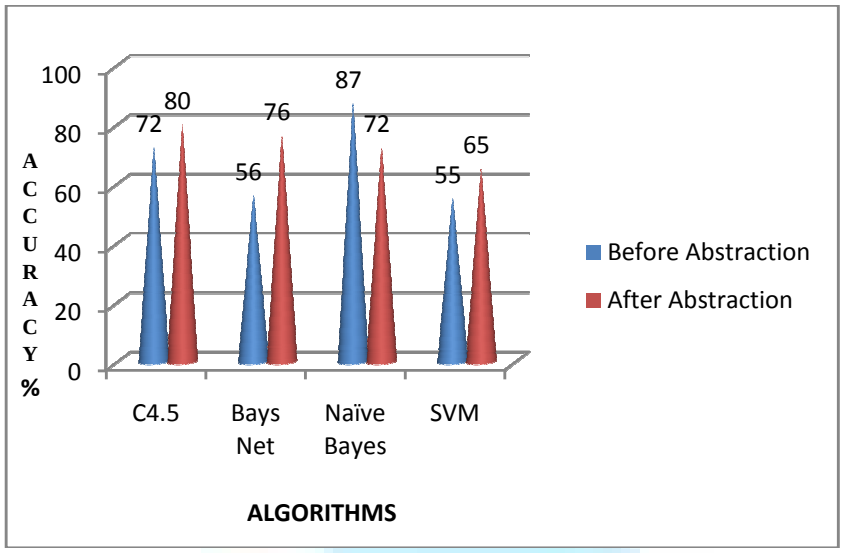
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **ontology** ((computer science) a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations)
- S: (n) **ontology** (the metaphysical study of the nature of being and existence).

4.2 EXPERIMENT

This experiment is done to improve accuracy in classification of web documents using Domain Ontology. We tested system in two domains- weather forecasting and GoogleTM. There is specific abstraction level (k) related to each domain. Our system is tested for both domains before and after abstraction. We have used WORDnet for experimental evaluations.



RESULTS

The above experiment shows that classification Data Mining algorithms like C4.5, Bayes Net and SVM are improved by using WORDnet. We can classify simple words and expressions of different datasets. Naïve Bayes is algorithm that shows accurate result before abstraction and shows less accurate result after abstraction.

5. CONCLUSION

From this paper, we conclude that we have described role of Knowledge Management (KM) solutions in extracting information from simple systems. But these solutions are not proved to be useful for extracting information from complex systems. This uncertainty has evolved a concept of Ontology. With help of Ontology, we are able to use KM solutions in different environments which require use of self authorized agents. Agents can define their own ontology. We have presented an approach for increasing communication between agents that make use of Ontology as part of knowledge representation. This approach is accomplished by making use of algorithms. This paper also shows Multi Agent technology for DDM and DC systems along with their architecture. It also involves use of various agents involved in MADM systems. We presented a new Ontology- based methodology for classification and analysis of web documents by various agents. We have conducted an experiment using WORDnet. The main credit of this work goes to domain ontology in representing documents. Use of WORDnet leads to improve in classifying web documents with the help of synonyms as it has large collection of similar words related to particular search.

6. REFERENCES

1. A.Susaria, D.Liu and A.Winson, "Peer-to-Peer Knowledge Management", "Handbook of Knowledge Management, vol 2, ch 39", Nov 2002.
2. Accessible from L.Steels, "The Origin of Ontologies and Communication Conventions in Multi Agent Systems", "J. Autonomous Agents and Multi- Agent systems, vol 1", March 2004, pages 165-194.
A. Doan, J. Madhavan, A. Halevy, "Ontology Matching: Machine Learning Approach", "Handbook on Ontologies in Information Systems, S. Staab and R. Studer", May 2004, pages 397-416.
3. Fu. L, "Knowledge Discovery based on complex multi agent systems", "Communications of ACM Vol.42 No.11", Jan 2004, pages 47-50.
A. Giannella, R. Bhargava and H. Kargupta, "Multi-Agent Systems and Distributed Data Mining", "Lecture Notes in Computer Science", 2005, pages 1-15.
4. Litvak, M. Last, Kisilevich, "Improving Classification of Multi-Lingual Web Documents using Domain Ontologies", "ECML/PKDD-2005", October 2005.
5. Miller, G.A., Beckwith, Gross, "Wordnet: An Online Lexical Database", "International Journal of Lexicography", 2004, pages 235-244.
6. N.F. Noy, M.A. Musen, "Algorithm and Tool for Automated Ontology Merging and Alignment", "Proceedings 17th National Conference on Artificial Intelligence (AAAI-2000)", Nov 2000.
7. R. Wille, "Concept lattices and conceptual knowledge systems", "Computers & Mathematics with Applications", September 2004, pages 493-515.
8. T. Gruber, "The Role of Common Ontology in Achieving Sharable, Reusable Knowledge Bases" Principles of Knowledge Representation and Reasoning: "Proceedings of the Second International Conference, Cambridge, MA", 1991, pages 601-610.
9. U. Fayyad, R. Uthuruswamy, "Data Mining and Knowledge discovery in databases", "Communications of the ACM, 39(11)", 1996, pages 1-15.
10. Vladimir Gorodetsky, Oleg, Karsev, "Multi Agent technology for distributed data mining and classification", "In IAT, IEEE Computer Society of India", 2003, pages 438-441.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail **infoijrcm@gmail.com** for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals