Information Retrieval System Based On Ontology

[1] Prof..DeepentiH.Deshmukh, Assistant Professor,Department Of Computer Science &Engineering,PRMITR, Badnera(Mah)
dhdeshmukh@mitra.ac.in
[2] Prof. Shwetambari G. Pundkar ,Assistant Professor, Department Of Computer Science & Engineering ,PRMITR, Badnera(Mah)
sgpundkar@mitra.ac.in
[3] Prof. Poonam B. Lohiya ,Assistant Professor, Department Of Computer Science & Engineering ,PRMITR, Badnera(Mah)
pblohiya@mitra.ac.in
[4]Prof. Snehal H. Kuche ,Assistant Professor, Department Of Computer Science & Engineering ,PRMITR, Badnera(Mah)
shkuche@mitra.ac.in

Abstract:
Over the years, the volume of information available through the world wide web has been increasing continuously, and never has so much information readily available and shared among so many people. Unfortunately, the unstructured nature and huge volume of information accessible over network have made it difficult for users to shift through and find relevant information.

The information retrievals commonly used are based on keywords. These techniques used keyword lists to describe the content of information, but one problem with such list is that they do not say anything about the symantic relationships between keywords, nor do they take into account the meaning of words or phrases.

Keywords:
HozoEditor,XML,DataMining,KnowledgeRepresentation,Clustering, Information Retrieval, Ontology

1. Introduction

Users frequently have problems expressing their information needs and translating those needs into requests. This is because information needs cannot be expressed appropriately in the terms used by the system, and partly because it is not unusual for users to apply search terms that are different from the keywords information system use.

Using conceptual knowledge to help users formulate their requests is a method of introducing conceptual knowledge to information retrieval. Another is to use conceptual knowledge as the intrinsic feature of the system in the process of retrieving the information. The semantic knowledge attatched to information united by means ontologies i.e. the concepts attatched to information are mapped into these ontologies. Machines can then determine these concepts in different piesces of information are actually the same to the positions they hold in ontologies. Thus ontologies is useful method for moving keyword based to concept based information retrieval.

In this paper **(Model-Based Method for Projective Clustering),** a probability model is first proposed to describe projected clusters in high-dimensional data space. Then, author present a model-based algorithm for fuzzy projective clustering that discovers clusters with overlapping boundaries in various projected subspaces. Clustering high-dimensional data is a major challenge due to the curse of dimensionality. To solve this problem, projective clustering has been defined as an extension to traditional clustering that attempts to find projected clusters in subsets of the dimensions of a data space. It proposed an extended Gaussian model which meets the general requirements of projective clustering well. It also derived an objective clustering criterion based on the model, allowing the use of a k-means type paradigm.

In this paper **(Clustering Sentence-Level Text Using a Novel Fuzzy Relational Clustering Algorithm)** This paper presents a novel fuzzy clustering algorithm that operates on relational input data; i.e., data in the form of a square matrix of pair wise similarities between data objects. The algorithm uses a graph representation of the data, and operates in an Expectation-Maximization framework in which the graph centrality of an object in the graph is interpreted as likelihood. Results of applying the algorithm to sentence clustering tasks demonstrate that the algorithm is capable of identifying overlapping clusters of semantically related sentences, and that it is therefore of potential use in a variety of text mining tasks. It also includes results of applying the algorithm to benchmark data sets in several other domains. The FRECCA algorithm was motivatedin fuzzy clustering of sentence-level text, and the need for an algorithm which can accomplish this task based on relational input data. It is capable of identifying softer clusters. The major disadvantage of the algorithm is its time complexity.

In this paper **(An Input–Output Clustering Approach to theSynthesis of ANFIS Networks)**A useful neural network paradigm for the solution of function approximation problems is represented by adaptive neuro-fuzzy inference systems (ANFIS). When dealing with numerical samples of the function to be approximated, clustering procedures represent an attractive approach to the synthesis of ANFIS networks. However, some of these procedures can be affected by serious drawbacks depending on the data space where clustering is applied: input only, output only, or joint input–output. In order to overcome these problems, the hyperplane clustering synthesis (HCS) of ANFIS networks has been proposed in this paper.

In this paper **(A Novel Approach for Mining and Fuzzy Simulation of Subnetworks From Large Biomolecular Networks)** In this paper, we present a novel method to model the regulatory system that executes a cellular function, which can be represented as a biomolecular network. Our method consists of three steps. First, the biomolecular network is derived using data-mining approaches to extend the initial conceptual biomolecular network from the literature search, etc. Secondly, once the whole biomolecular network structure is complete, a novel scale-free network clustering approach is applied to obtain various subnetworks. Itpresent a new method for adaptive modeling of biomolecular networks. computational hypotheses are tested and refined through cycles of fuzzy logic based simulation and laboratory experiments. Author use fuzzy logic methods, previously developed for gene networks, as a robust and general representation for heterogeneous quantitative, qualitative, and linguistic biomolecular data.

In this paper **(A Neuro-Fuzzy System Modeling With Self-Constructing Rule Generation and Hybrid SVD-Based Learning)** Author propose a novel approach for neuro-fuzzy system modeling. A neuro-fuzzy system for a given set of input–output data is obtained in two steps. First, the data set is partitioned automatically into a set of clusters based on input-similarity and output-similarity tests. Membership functions associated with each cluster are defined according to statistical means and variances of the data points included in the cluster. Then, a fuzzy IF–THEN rule is extracted from each cluster to form a fuzzy rule-base. Second, a fuzzy neural network is constructed accordingly and parameters are refined to increase the precision of the fuzzy rule-base.Author has described a novel neuro-fuzzy system modeling approach which incorporates self-constructing rule generation and hybrid SVD-based learning.

- In June 2002 Robert Stevens describes the initial stages of building an ontology of bioinformatics and molecular biology. The conceptualization is encoded using the ontology inference layer (OIL), a knowledge representation language that combines the modeling style of frame-based systems with the expressiveness and reasoning power of description logics (DLs).
- In 2oo3 Sebastian Hübner, Rainer Spittelfocussed on a tourist looking for information regarding water quality, including ontologies for describing vocabularies and catalogues as well as search mechanisms for keywords with respect to their semantics.
- In 2007 Pablo Castells, Miriam Ferna´ndez, and David Vallet researchers proposed a model of ontology based KB scheme for the semi-automatic annotation of documents, and a retrieval system.

2. Proposed Methodology

2.1 Aim And Objectives

The main objective is to provide efficient means to move from keyword based to concept based information retrieval utilizing ontologies as reference for conceptual definitions. The main aspects in this project are the following:

1. Recognition and mapping of information in documents and queries into the ontologies.
2. Improvement of retrieval process by use of similarity measures derived from knowledge about relations between concepts in ontologies.
3. To mend the ideas of such ontological indexing and ontological similarity into a realistic information retrieval scenario.

The first part concerns extraction of semantic knowledge from texts and a word sense disambiguation of this knowledge such that it can be mapped into the ontology in use. The second part concerns the development of scalable similarity measures, where the idea is to compare concepts on behalf of the structure in the ontology. Finally the object of the last part is how to bring first two parts into play in information retrieval.

2.2 Algorithm(Database Clustering)

1. *Select Key Word*
2. *Select Directory*

3. *Load deirectory Files*
4. *Cluster Directory as per content of Keyword.*
5. *Save Clusters*
6. *Stop*

2.3 Algorithm (DataMining& Knowledge representation)

1. *Input Mining string.*
2. *Segment reference string into it's sub combinational string in forward direction*
3. *Start searching cluster from sub reference string according to it'sdecresing order of weight.*
4. *List all script statements from xml files.*
5. *Construct knowledge from script statement.*

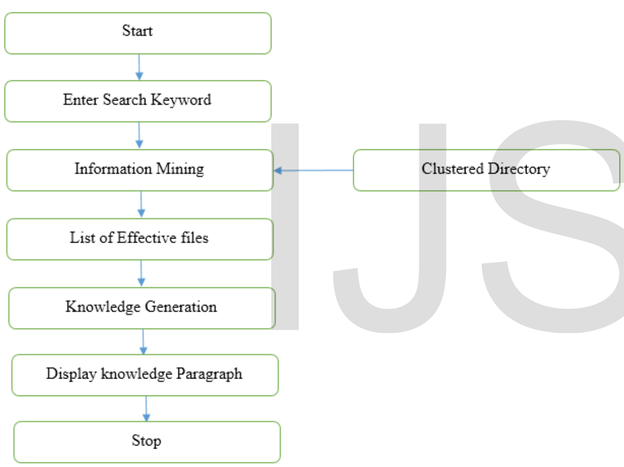

Figure 3.1 System Flow Diagram(Knowledge Representation)

KR is the study of how what we know can at the same time be represented as comprehensibly as possible and reasoned with as effectively as possibly[12]. The simplest analysis shows difference between procedural and declarative knowledge. KR is very important for knowledge based systems. A selected KR scheme should have appropriate inference methods to allow for reasoning. Popular KR schemes are Rules, Semantic Nets, Schemata(Frames and Scripts) and Logic. Balance must be found between effective representations, efficiency and understandability for effectiveness. Effective KR should be used to represent the most important aspects of the real world, such as action, space, time, mental events.Future work is to improve the parameters considered.

Knowledge-representation is the field of artificial intelligence that focuses on designing computer representations that capture information about the world that can be used to solve complex problems. The justification for knowledge representation is that conventional procedural code is not the best formalism to use to solve complex problems such asexpert systems. Knowledge representation makes complex software easier to define and maintain than procedural code. For example, talking to experts in terms of business rules rather than code lessens the semantic gap between users and developers and makes development of complex systems more practical.

2.4 A KR is a Set of Ontological Commitments

If, as we have argued, all representations are imperfect approximations to reality, each approximation attending to some things and ignoring others, then in selecting any representation we are in the very same act unavoidably making a set of decisions about how and what to see in the world. That is, selecting a representation means making a set of ontological commitments. [2] The commitments are in effect a strong pair of glasses that determine what we can see, bringing some part of the world into sharp focus, at the expense of blurring other parts.

These commitments and their focusing/blurring effect are not an incidental side effect of a representation choice; they are of the essence: a KR *is* a set of ontological commitments. It is *unavoidably* so because of the inevitable imperfections of representations. It is *usefully* so because judicious selection of commitments provides the opportunity to focus attention on aspects of the world we believe to be relevant.

The focusing effect is an essential part of what a representation offers, because the complexity of the natural world is overwhelming. We (and our reasoning machines) need guidance in deciding what in the world to attend to and what to ignore. The glasses supplied by a representation can provide that guidance: In telling us what and how to see, they allow us to cope with what would otherwise be untenable complexity and detail. Hence the ontological commitment made by a representation can be one of the most important contributions it offers.

There is a long history of work attempting to build good ontologies for a variety of task domains, including early work on an ontology for liquids [12], the lumped element model widely used in representing electronic circuits (e.g., [3]), as well as ontologies for time, belief, and even programming itself. Each of these offers a way to see some part of the world.

The lumped element model, for instance, suggests that we think of circuits in terms of components with connections between them, with signals flowing instantaneously along the connections. This is a useful view, but not the only possible one. A different ontology arises if we need to attend to the electrodynamics in the device: Here signals propagate at finite speed and an object (like a resistor) that was previously viewed as a single component with an I/O behavior may now have to be thought of as an extended medium through which an electromagnetic wave flows.

Ontologies can of course be written down in a wide variety of languages and notations (e.g., logic, LISP, etc.); the essential information is not the form of that language but the content, i.e., the set of concepts offered as a way of thinking about the world. Simply put, the important part is notions like connections and components, not whether we choose to write them as predicates or LISP constructs.

The commitment we make by selecting one or another ontology can produce a sharply different view of the task at hand. Consider the difference that arises in selecting the lumped element view of a circuit rather than the electrodynamic view of the same device. As a second example, medical diagnosis viewed in terms of rules (e.g., MYCIN) looks substantially different from the same task viewed in terms of frames (e.g., INTERNIST). WhereMYCIN sees the medical world as made up of empirical associations connecting symptom to disease, INTERNIST sees a set of prototypes, in particular prototypical diseases, to be matched against the case at hand.

3. Result Analysis

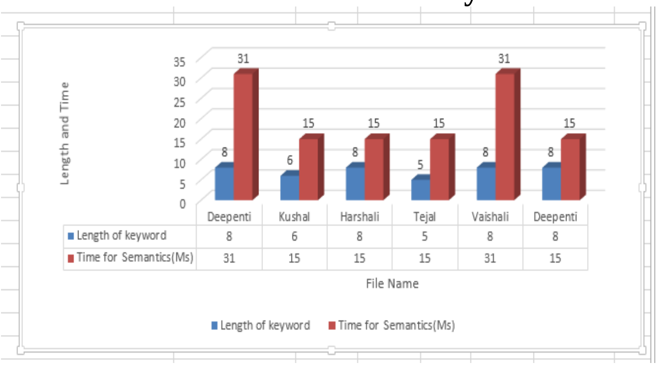

Table 4.1 Length of Keywords Vs Time for Semantics

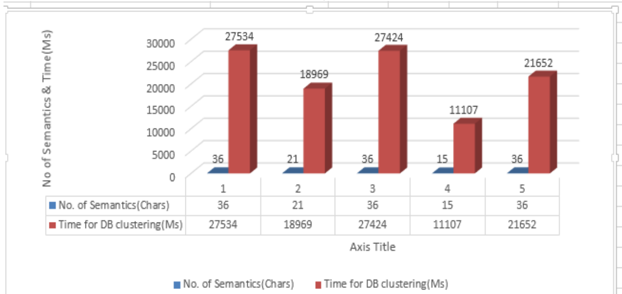

Table 4.2 Length of SemanticsVs Time for DB Clustering

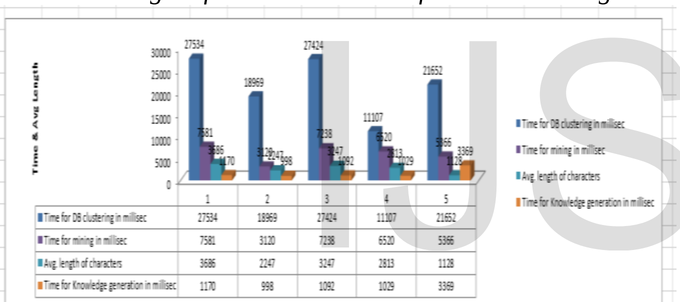

Table 4.3 Length of SemanticsVs Time for DB Clustering Vs Mining Time Vs Knowledge Generation Time

Conclusion

There are various knowledge representation schemes in AI. All KR techniques have their own semantics, structure as well as different control mechanism and power. In this paper, trying to build the intelligent system that can learn itself by the query and have a power full mechanism for representation and inference. The semantic net and script are very powerful techniques in some respects so the aim is to take the advantage of these techniques under one umbrella.

References

[1]Deepenti H. Deshmukh, Prof. S.D.Deshpande December 2013 A Review of Ontology Based Information Retrieval . International Journal of Advance Research in Computer Science and Management Studies Vol.1, Issue 7.

[2]Pablo Castells, Mirium Fernandez, David Vallet Feb 2007 An Adaptation of the Vector Space Model for Ontology Based Information Retrieval . IEEE Transactions on Knowledge and Data Engineering Vol.19, No.2.

[3] Dieter Fensel 2002. Ontology Based Knowledge Management, IEEE.

[4] YajingJhao, Jing Dong, TuPeng, October 2009 Ontology Classification for Semantic Based Software Engineering. IEEE Transactions on Services Computing Vol.2 No.4.

[5] Robert Stevens, Carole Goble, Ian Horrocks, Sean BechhoferJune 2002 Building A BioInformatics Ontology Using OIL. IEEE Transactions on Information Technology in BioMedicine Vol.6, No.2.

[6] A.Th.Schriber, Barbara Dubbeldam, Bob Weilinga, 2001 Ontology Based Photo Annotation, IEEE.

[7] Sebastian Hubner, Rainer Spittel, UbboVisser, 2004, Ontology Based Search for Interactive Digital Maps, IEEE Intelligent Systems.

[8] Thomas R. Gruber, August 1993, Towards

Principles for the Design of Ontology Used for Knowledge Sharing,Internaional Journal of Human Computer Studies 43.

[9] M. Kerber, 2004/05 "Knowledge Representation I", Introduction To Ai 06-08775.

[10] B. Lamiroy1 And L. Najman2,3, "Scan-To-Xml: Using Software Component AlgebraFor Intelligent Document Generation",

[11] Ying Zhang Alex Golovinsky Mark Yim Craig Eldershaw, "An Xml-Based Scripting Language For Chain-Type Modular Robotic Systems" Palo Alto Research Center, 3333 Coyote Hill, Palo Alto, Ca, Usa.

[12]Gaurav Harit1, K. J. Jinesh2, Ritu Garg3 C.V Jawahar2, And Santanu Chaudhury3," Managing MultiligualOcr Project Using Xml".

[13] Alexandre Delteil1, Catherine Faron2, And Rose Dieng1, "Building Concept Lattices By Learning Concepts From Rdf Graphs Annotating Web Documents".

[14] Jean-François Baget 1,2, Olivier Corby1, Rose Dieng-Kuntz1, "Griwes: Generic Model And Preliminary Specifications For A Graph-Based Knowledge Representation Toolkit"

[15] Victoria Uren, Bertrand Sereno, Simon Buckingham Shum &Gangmin Li, September, 2003

[16] H. Cherfi1, O. Corby1, C. Faron Zucker1,2, K. Khelif1 And M.T. Nguyen1," Semantic Annotation Of Texts With Rdf Graph Contexts".

[17] Philippe Fournier-Viger, Evaluating Spatial Representations And Skills In A Simulator-Based Tutoring System, Ieee Transactions On Learning Technologies, Vol. 1, No. 1, January-March 2008.

[18] Li Guoliang, FengJianhua, Zhou Lizhu, Keyword Searches In Data-Centric Xml Documents Using Tree Partitioning*, Tsinghua Science And Technology,Issn 1007-0214 02/ 21 pp7-18Volume 14, Number 1, February 2009.

[19] Li Yingchun ,A Systems-Science-Based Knowledge Explanation Method, Tsinghua Science And Technology Issn1007-0214 11/20 Pp49 – 56Volume 6, Number 1, March 2001.

[20] Randall Davis, Howard Shrobe, And Peter Szolovits, What Is A Knowledge Representation?, Ai Magazine Volume 14 Number 1 (1993) (© Aaai).

[21] DovDori, The Visual Semantic Web: Unifying Human And Machine Semantic Web Representations With Object-Process Methodology,

[22] Philippe Martin, Knowledge Representation In Rdf/Xml, Kif, Frame-Cg And Formalized-English.