

Intelligent Retrieval Model : Object Oriented Search Methodology on Web

Dr. Pushpa R. Suri¹ and Harmunish Taneja²

¹ Maharishi Markandeshwar University, Mullana, Ambala, Haryana, India

Received: 21 June 2011 Accepted: 18 July 2011 Published: 28 July 2011

Abstract

Information handling through improved search mechanisms is and will remain the driving force behind rich patterns of access to web contents. Keyword search offers only an imperfect solution to information discovery over web as it misses many relevant documents of today's users. By its fundamental nature, keyword search is sharply focused to find the exact terms specified in the query. The proposed Intelligent Retrieval Model exploits the object oriented concepts for information representation. This kind of conceptual modeling of web based inheritance boosts data structuring flexibility and caters to the density of diverse web information needs. The guiding principle of proposed model for intelligent retrieval over World Wide Web is the adequate representation of relevant complex relationships between the various entities in the real world. The result approves better hierarchical relationship recognition among web objects leading to less complex computing.

Index terms— Intelligent Retrieval Model (IRM), World Wide Web (WWW), Inheritance, Information Computing.

1 INTRODUCTION

WW is expanding into the dynamic repository for information on virtually every topic including people and their relations to one another as well as to contents. The dynamism is changing web in truly fundamental ways [11], mainly due to companies such as Facebook, Google, and Yahoo! Ideally information retrieval over web needs methodologies that offer high performance and intelligent retrieval [8].

The proposed Intelligent Retrieval Model (IRM) resorts to object oriented methods for being intelligent over the conventional search techniques. As the complexity of the web shoots, the efficiency of existing search algorithms drops [7]. Therefore there is a need to strike a balance between complexity of web information representation and efficiency of information computing. The most conventional information retrieval technique works on the basic principle of query based pattern matching of the search terms entered by the user and displaying as output the document with exact combination of the keywords that may include Boolean operators. The Boolean text systems reflects the inability of the relational model to express a wide range of relationships between diverse objects on web.

2 IRM

accommodates more complex relationships than the Boolean ones due to the object oriented background. The relational data model [1] [2] [3] has been the most popular method for information handling. Although the relational model is easy to implement and is simple, it is inefficient for handling multidimensional data in present scenario of emerging web. Relational model suffers severe drawbacks [4] The rest of the paper is organized as follows. Section 2 presents the background of object based system in contrast to relational data model. Section 3 presents a conceptual model embedding the object oriented concepts in web search methodology followed by the

architecture of proposed intelligent retrieval model, further illustrated with the examples of various web domains. Results are discussed in the section 4 and finally paper is concluded in the section 5.

3 II.

Traditional information handling over web has been very successful, but is limited to certain types of applications [7]. As web experiences an explosion in the diversity of its users, as business users expand to newer types of applications, and grow older ones, their attempts to use WWW encounter the "Relational barrier," where RDBMS technology suffer pitfalls to provide the performance and functionality desired. This barrier is encountered when extending information models to support relationships, new data types, and direct W 1. Limitations on data structure flexibility particularly in handling hierarchically structured data. To represent the specific information about a sub part of information, individual tables for each sub parts with many more tables to establish hierarchical links are required. 2. Relationships lack the explicit specification. The information by combination of different relations among entities is hard to extract especially by a naïve user. 3. Query language knowledge is mandatory by all users. Also in languages such as SQL [5], user needs to specify the query in specified format. 4. In contrast to relational model, proposed model provides a more elegant paradigm for information handling. It stems on object oriented approach that allows automatic hierarchical information representation by storing data explicitly at various levels of abstraction.

Global Journal of Computer Science and Technology Volume XI Issue XIV Version I 2011

4 August

5 RELATED WORK

with complex operations. Attempts to scale the barrier with relational technology lead to an explosion of tables, many joins, poor performance, poor scalability, and loss of integrity [9]. Object oriented information computing offer a path beyond this barrier. Object oriented information representation enjoys enormous advantages over a relational model. The core of object oriented system is an object that is a behavioural unit constituted of fields and methods to manipulate those fields. An object's state at a certain moment is the values of its fields. Object is the combination of procedures and data which performs all the operation with the help of message passing between various objects. A class / property inheritance system consists of a set of concepts with a strictly transitive inheritance relation and a set of inheritance properties associated with each concept [10]. Object oriented systems support data abstraction which means various procedures specific to a class or a particular category of information can be clubbed together thereby increasing the modularity of information computing. In contrast to this relational data model do not support any one to many correspondences without creating new tables thereby making the search tedious [4].

Modern information computing is expected to provide the desirable object-oriented features through which the complex relations within the applications can be expressed in an enhanced way for simplifying web information handling. The study on the existing information computing approaches disclose the lack of standard data models for handling web based information management applications. The existing hypertext data models [13][14] reflect the whole database in terms of n-tuples but fail to meet the terms with the object-oriented approach. Object-oriented approach adopted in [15] [16] lacks the mapping to standard logical model. A hybrid data model that combines the advantages of the relational model with the positive features of semantic data models represents database as a graph and hierarchy to some extent through grouping constructs [17]. But the approach lacks to provide object orientation and hypertext semi-structured data handling. Conventional information search over web is ruled by the keyword based technique that fails to capture relationships between the words entered in query leading to irrelevant documents as output.

6 III.

WWW has billions of related documents that may prove to be appropriate for a users query. But relationship between two concepts over WWW needs to be recognized. The proposed system identifies object oriented concepts embedded in the web search to describe the various queried topics in which the user might be interested and the relationships while preserving a degree of generality.

7 a) Conceptual Model of web based Inheritance

The class inheritance system consists of set of concepts with a strictly transitive inheritance relations and set of inherited properties associated with each concept [6] [12]. Hierarchy describes the ranking or ordering in general of abstraction. The class structure of an object oriented information system consist of an "isa" hierarchy often to describe single or multiple inheritance and the object hierarchy describes a "partof" or "whole / part" hierarchy to denote aggregation [18]. Inheritance describes the relation between the classes that allow us for the definition and implementation of one class to be based on that of other existing classes. A new class inherits the properties from its parents and may introduce new property that extend, modify or defeat its inherited property [21] unless an exception is specified. Fig. 1 elaborates the proposed conceptual model embedding the object oriented concepts for enhanced web search. Inheritance is described as the ability of sharing the attributes and operations between

objects and their class. Typically the object classes are objects themselves, so they also inherit the attributes from other classes. Inheritance tree of Fig. 1 exhibits an object inheritance network as a directed graph.

8 INTELLIGENT RETRIEVAL MODEL (IRM)

Inheritance is both a form of expansion where the behavior and data associated with the child class or sub class is always an extension of the parent class, and a contraction because a child is always a more specialized or restricted form of parent class [19].

DigitalCamera is a subclass of Camera and Price and therefore inherits traits from both super classes. Any path in the graph that links two nodes (classes in this case) represents the relation between two reflected concepts. A path with only one link is simplest possible relation as shown in Fig. ??; the relation is a one link. Fig. ?? : One Link Inheritance hierarchy of object oriented system can be described with different types of associations [20]. Complex relations may have two or more links. Aggregation facilitates collection of elements together to form a case of abstraction. It also hides the details of its parts. This information hiding results in aggregate that may be considered as a collection of parts that together form something new that is more than just the sum of its parts [19]. Inheritance often can be regarded as the feature that distinguishes the object oriented paradigm from other programming languages [20]. The proposed model for enhance web search methodology generalizes the relatedness concept on the following levels:

9 ?

IRM does not limit itself to one domain. It has been designed as a general tool which may contain any conceptual knowledge.

10 ?

The concepts are organized in a hierarchical way, with classes able to inherit some characteristics of their superclasses. This produces even more complex relations, but also gives the ability to generalize them in order to determine conceptually simple relations.

IRM help the user in finding not only the more general or more specific keywords but also the relationship that may exist between the objects obtained through keywords. ? Query handler : User enters the keyword based search into the interface which is derived from the normal dialogue box from the conventional search engine. The request is forwarded to the query handler which generates the web object well matched with the queried keywords. The Meta data of tags describing the object attributes are also stored.

? Knowledge space : It implements the trustworthy deep web knowledge categorization into the multiple domains of the desired search area. This component help to organize dispersive information on the surface web in a way that facilitates creation of object lexicons corresponding to each domain based on vertical search.

? Object Mapper : The surrounding text of the web object on the web page provides the object description. The mapper converts the same in the structured form and filters the noisy information. ? Object Indexer : User information is far beyond the keyword based search. To meet such a requirement, IRM offers object oriented object indexing. The output from relationship graph extract the structure attributes of the desired web objects by discovering the best matched structured information from the web object mapper. The hierarchical structure of the domain knowledge is exploited and the entry in the pseudo page corresponding to the object information is reranked by cross verification with the super class and the sub class from the inheritance graph. The hierarchical links between the search objects stored by IRM are shown in Fig. ??. Consider a search object song made up of three categories Western, Ghazals and Folk. Let each of these categories is made up of sub categories as shown in the Fig. ??. A song is represented as an individual Class (C1). Proposed model abstract the search information by creating a set of classes Ghazals (C2) and Gz_Hindi (C3) at different levels. An object representation can naturally capture the inherent structure of search over WWW. Objects can be created at various levels of web search and IRM supports storage of hierarchical links between objects along with the objects themselves. IRM facilitate creation of new subclass with specific group of objects having common features and this new class object will represent the common abstract property e.g. in Fig. 9 the objects Fk_Punjabi and Gz_Punjabi and Gz_Hindi share a common property singer. Hence a new subclass having these three objects as its instances can be created which will then be a subclass of the generic class C2 and C3, as shown in Fig. ??(a) and 8(b) respectively.

11 RESULTS

IRM implementation shows the efficient representation of relational information between the web search results that leads to enhanced search. Fig. 11 provides the illustration as screen shots for the query "Panasonic Lumix DMC 10K" that is a digital camera. All DMC 10K" Similar illustration for searching another electronic gadget "Nikon Coolpix P500" is given in Fig. 12. The object instances are grouped similar to the exemplification as shown in Fig. 12. The basic relations between the concepts in the IRM are more general. Beside products and their findings, IRM also considers other types of knowledge that user commonly desire for example: the components, price etc. and its associated elementary relation for the products at various levels.

12 CONCLUSION

The quality of information computing as in keyword based Boolean expression query formulation is unable to express relationships. Entity representation as objects provides natural ease of viewing WWW at different levels of abstraction. The proposed model IRM reduces the number of components as per user needs that leads to reduced computational complexity. It facilitates the hierarchical representation of web objects that ensures efficient capture of all desired information with less latency. The results confirm identification of not only one to many relationships resulting in tree structures but also many to many relationships resulting in lattice by using IRM. The main advantage of the proposed work is the freedom of regrouping and the ability to visualize the same object from different aspects for efficient web search. From the point of view of user, information computing over web seems to have embedded relations among the objects and the drawbacks of keyword based searching over web are effectively reduced by the object oriented approach. In addition to possible applications in product search, this work could improve methods for information computing as in indexing and querying. The indexer will obtain suggestions on related concepts and relationships on which a search result may be indexed. These suggestions are not restricted to hierarchical organization, but also include inter hierarchical information.



Figure 1: Fig. 1 :

¹© 2011 Global Journals Inc. (US)

²August

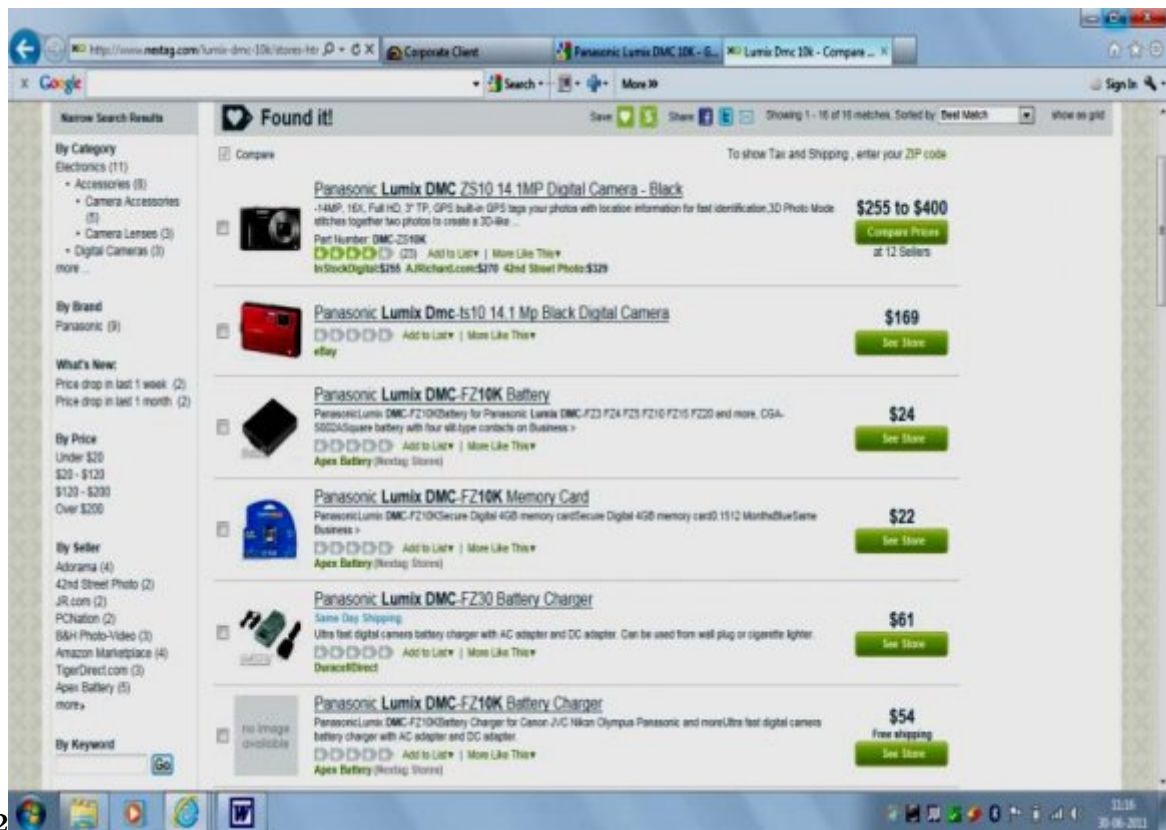


Figure 2: Fig. 2 :

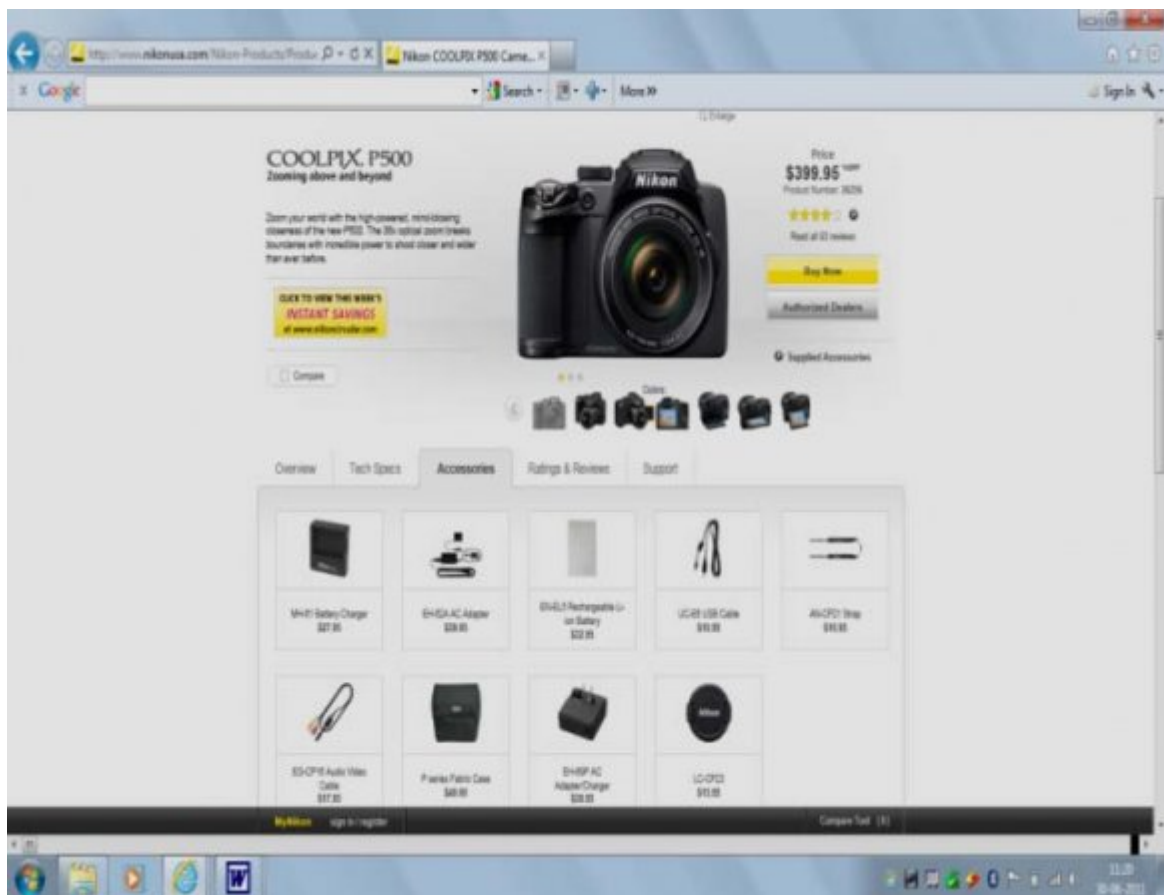


Figure 3:

-
- [Wm and Tompa ()] ‘A data model for flexible hypertext database Systems’. Frank Wm , Tompa . *ACM Transactions on Information Systems* 1989. 7 (1) p. .
- [Codd ()] ‘A relational model of data for large shared data banks’. E F Codd . 10.1145/362384.362685. *Communications of the ACM* 1970. 13 (6) p. .
- [Abadi and Cardelli ()] *A Theory of Objects*, M Abadi , L Cardelli . 1996. NY; New York: Springer-Verlag.
- [Budd ()] *An introduction to object-oriented programming*, T Budd . 2001. Addison-Wesley. (3 rd Edition)
- [Nicholas Linda and Rosenberg ()] *Canto: a hypertext data model*, Charles K Nicholas & Linda , H Rosenberg . 1993. John Wiley and Sons limited, Electronic Publishing. 6 p. .
- [Cho et al. ()] ‘Estimating nested selectivity in object-oriented and objectrelational databases’. W Cho , K Hong , W Loh . *Information and Software Technology* 2007. 49 (7) p. .
- [He and Jerome ()] ‘Evaluating the Dynamic Behavior of Database Applications’. Z He , D Jerome . *Journal of Database Management* 2005. 16 (2) p. .
- [Marino et al. ()] ‘Expressing Structural Hypertext Queries in Graph log’. P Marino , O Consens & Alberto , Mendelzon . *Hypertext ’ 89 Proceedings*, 1989.
- [Codd ()] ‘Extending the database relational model to capture more meaning’. E F Codd . 10.1145/320107.320109. *ACM Transactions on Database Systems (TODS)* 1979. 4 (4) p. .
- [Choudhury et al. ()] ‘GDM: A New Graph Based Data Model Using Functional Abstractionx’. Sankhayan Choudhury , Nabendu Chaki & Swapan , Bhattacharya . *Journal of Computer Science and Technology* 2006. 21 (3) p. .
- [Larson ()] ‘Introduction to Information Retrieval’. R Larson . *Journal of the American Society for Information Science and Technology* 2010. (4) p. .
- [Copeland and Maier ()] ‘Making smalltalk a database system’. George Copeland , & David Maier . 10.1145/971697.602300. *ACM SIGMOD Record* 1984. 14 (2) .
- [Blaha and Rumbaugh ()] *Object -Oriented Modeling and Design with UML, 2 nd Edition*, Michael Blaha , & James Rumbaugh . 2007. Pearson Education.
- [Booch Booch ()] *Object-oriented analysis and design with applications*, G Booch & Booch . 2007. Addison-Wesley. (3 rd Edition)
- [Taivalsaari ()] ‘On the notion of inheritance in’. A Taivalsaari . *ACM Computing Surveys* 1996. 28 (3) p. .
- [Uiiman ()] *Principles of Database Systems*, J D Uiiman . 1983. Rockville, MD: Computer Science Press.
- [Donald et al. (1974)] ‘SEQUEL: A structured English query language’. D Donald , F Chamberlin & Raymond , Boyce . 10.1145/800296.811515. *Proceedings of the ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control*, (the ACM SIGFIDET (now SIGMOD) workshop on Data description, access and controlAnn Arbor, Michigan) 1974. May 01-03, 1974. p. .
- [Wang ()] ‘Solving Relational Database Problems with ORDBMS in an Advanced Database Course’. Ming Wang . *Proceedings of Information Systems Educators Conference ISECON*, (Information Systems Educators Conference ISECONNashville Tennessee, USA) 2010. 1376. p. .
- [Belkin ()] ‘Some (What) Grand Challenges for Information Retrieval’. Nicholas J Belkin . *Advances in Information Retrieval*, 2008. Springer. p. .
- [Abiteboul et al. ()] *The Lorel Query Language for Semi Structured Data*, Serge Abiteboul , Dallan Quars , Jason Mchugh , Jennifer Widom , Janet L Wiener . 1997. AQM.
- [Touretzky ()] *The Mathematics of Inheritance sytems*, D Touretzky . 1986. Pitman, London; Los Altos, CA: Morgan Kaufmann Publishers, Inc.