International Journal of Management and Humanities (IJMH)
ISSN: 2394-0913, Volume-3 Issue-5, August 2018

Information Retrieval for Ontology-Based Biological Data

Khin Myo Sett

Abstract: ***This paper provides a brief overview of the information retrieval process and also describes the main classic information retrieval models such as keyword-based information retrieval and ontology-based information retrieval. A brief overview of image retrieval system is discussed in this chapter. Some categories of image retrieval such as keywords-based image retrieval, content-based image retrieval, and ontology-based image retrieval are surveyed in this chapter. Evaluation methods of ontology-based biological information system are also presented. Graph database and other types of graph database are illustrated and Neo4j graph database in this research area are described. This paper describes the architecture of ontology-based biological information system in detail. Taxonomy and ontology concept are discussed in the ontology construction model. Biology graph database structure is discussed over the state of the art which aims to motivate and introduced the strength of neo4j database internal design. This work discusses the implementation of the proposed system. Java implementation is also provided in this chapter. Evaluation framework is discussed for the ontology based biological information system. Unified Modeling Language (UML) diagrams are described in this chapter. This work describes the query results and discussed the analysis of these experimental results.***

Index Terms: ***Keyword-Based Information Retrieval, Ontology-Based Information Retrieval, Biological Information System, Graph Database.***

I. INTRODUCTION

Currently, information retrieval is an essential system for everywhere and the processing time very fast from the data storage is required. Many researchers have discussed the numerous data storage formats and data representation techniques. Information Extraction (IE) is the task of locating specific pieces of data within a natural language document. The advent of the internet has given IE a particular commercial relevance. IE is a process which takes unseen texts as input and produces fixed format, unambiguous data as output. The most popular storage format is ontology which is suitable for semantic information retrieval from the specific domains.

The meaning of ontology is sharing and reusing the knowledge in computational form. The term ‘Ontology’ has various definitions in various texts, domains and applications. In Philosophy and Linguistics, ontology is “The study of existence”, “A theory of what there is in the world”, or “A taxonomy of the world Concepts”. The most popular definition of ontology in information technology and AI community in theoretical view is “A formal explicit specification of a shared conceptualization” or “An abstract view of the world we are modeling, describing the concepts and their relationships” [38]. Biological oriented ontologies are viewed as essential for integrating a huge range of biological data. Ontology-based information services are applied in learning aids and research areas. In this thesis, biological taxonomies are constructed as ontologies by using neo4j graph database server.

The motivation for this thesis is to support the need of the information extraction for students and teachers who are studying and teaching bioinformatics. All the publications of biological information are written in documents, books form, and currently, most of web search engines are based on purely statistical techniques. While they are not able to figure out the meaning of a query, they can provide only answers by returning the statistically most appropriate answer to a user’s query – based on some measures for computing similarity. The search engines such as Google, Yahoo, NCBI, Bioportal, Animal.com etc. are in fact providing a baseline quite difficult to outperform. Due to the nature and the maturity of the underlying statistical techniques, they are more robust and to the scale to the size of Web, as opposed to semantic technologies. So, the problem is that students or teachers or researchers cannot search easily for relevant biological data or images for their needs and they can search by using link pages. Therefore, they take a lot of time for searching. This research is intended to extract easily for relevant and related information or images. Moreover, in this research, the role of ontology and the use of ontology in bioinformatics are investigated.

This thesis emphasizes the three portions of analysis such as different information retrieval query processing, ontology based biological data construction and graph data storage. The use of ontologies for the description of biological knowledge has increased rapidly as the community has recognized the value of this approach. Annotating biological data with ontological terms provides an explicit description of some of the data’s features. Developing and maintaining the ontologies in biology requires manual creation, deletion and correction of concepts and their definitions within the ontology, as well as annotating biological data to concepts of the ontology.

The motivations of this research are focused with three factors. The first one is needed in the life sciences for alternatives to keyword-based approaches based in the traditional information retrieval framework and also image retrieval framework. The second one is extensive work in text mining which is being done to identify animal-animal interactions and their breeding habits, feeding habits, habitats,

Revised Version Manuscript Received on 12 July, 2018.

Dr. Khin Myo Sett, Department of Computer Studies, Mandalay University, Mandalay, Myanmar, E-mail: khinmyosett.mu@gamil.com

Locations, locomotion, bodies cover, etc. Finally, ontologies based relationships are constructed in neo4j graph database and have been discussed in biological dataset.

II. RELATED WORKS

Information retrieval is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be based on metadata or on full-text (or other content-based) indexing.

Automated information retrieval systems are used to reduce what has been called "information overload". Many universities and public libraries use IR systems to provide access to books, journals and other documents. Web search engines are the most visible IR applications. An information retrieval process begins when a user enters a query into the system. Queries are formal statements of information needs, for example search strings in web search engines. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of relevancy.

An object is an entity that is represented by information in a database. User queries are matched against the database information. Depending on the application the data objects may be, for example, text documents, images, audio, mind maps or videos. Often the documents themselves are not kept or stored directly in the IR system, but are instead represented in the system by document surrogates or metadata.

Most IR systems compute a numeric score on how well each object in the database matches the query, and rank the objects according to this value. The top ranking objects are then shown to the user. The process may then be iterated if the user wishes to refine the query.

Quang Hieu Vu et al. [48] have given explanation that database management systems offer a comprehensive solution to data storage, while they require deep knowledge of the schema, as well as the data manipulation language, in order to perform effective retrieval. Since these requirements pose a problem to lie or occasional users, several methods incorporate keyword search (KS) into relational databases. However, most of the existing techniques focus on querying a single DBMS. On the other hand, the proliferation of distributed databases in several conventional and emerging applications necessitates the support for keyword-based data sharing and querying over multiple DMBSs. In order to avoid the high cost of searching in numerous, potentially irrelevant, databases in such systems, G-KS is proposed as a novel method for selecting the top-K candidates based on their potential to contain results for a given query. G-KS summarizes each database by a keyword relationship graph, where nodes represent terms and edges describe relationships between them. Keyword relationship graphs are utilized for computing the similarity between each database and a KS query, so that, during query processing, only the most promising databases are searched. An extensive experimental evaluation demonstrates G-KS that achieve better results than the current state-of-the-art technique on all aspects, including precision, recall, efficiency, space overhead and flexibility of accommodating different semantics. David W. Embley et al. [12] have presented a new approach to extracting information from unstructured documents based on an application ontology that describes a domain of interest. Starting with such ontology, the system has been formulated rules to extract constants and context keywords from unstructured documents. For each unstructured document of interest, which is extracted its constants and keywords and apply a recognizer to organize extracted constants as attribute values of tuples in a generated database schema. Generally, researchers have fixed all the processes and change only the ontological description for a different application domain. In experiments their system have been conducted on two different types of unstructured documents taken from the Web, their approach attained recall ratios in the 80% and 90% range and precision ratios near 98%. And the researchers have proposed a framework for an ontology-based system that extracts and structures information found in data-rich unstructured documents. Except for ontology creation, the processes in their framework are automatic and do not require human intervention. A prototype system has also been built based on this framework and has been applied two application areas—car advertisements and a computer jobs listing. As raw data for these applications, researchers have used documents placed on the Web by the Salt Lake Tribune and the Los Angeles Times. So, near 99% recall and precision on tuning data and roughly 90% recall and 98% precision on test data are conducted and obtained as result. They have observed that most of the errors in recall and precision were due to incomplete lexicons and incomplete ontologies. Without changing the framework, better lexicons and richer ontologies will overcome both of these short comings. Improvements in heuristics, front-end processing, and back-end processing are also possible.

Martin Labsky [24] has presented that the information extraction from websites using extracting ontologies and described that the advent of the Semantic Web initiatives around 2000, modeling parts of reality with domain ontologies became increasingly popular and a number of ontology authoring tools appeared. IE techniques became the natural choice to populate these ontologies with instances from text (semi-)automatically. The CREAM approach defines a methodology and tooling support for visually annotating web pages and for storing annotations as instances of ontology classes. The original approach was manual but it was soon coupled with a trainable IE system Amilcare that utilized the LP2 named entity extraction algorithm and a GATE annotator for preprocessing. The combined system was named Semi-automatic CREAM (SCREAM).

Siegfried Handschuh, Steffen Staab, and Fabio Ciravegna [16] have defined that CREAM is a comprehensive framework for creating annotations, relational metadata in particular, the foundation of the future Semantic Web. The new version of S-CREAM (semi-automatic creation of metadata) have been presented here supports metadata creation with the help of information extraction in addition to all the other nice features of CREAM, like comprises inference services, crawler, document management system, Ontology guidance/fact browser, document editors/viewers, and a meta ontology. Fabio Ciravegna [9] has presented (LP)2, an algorithm for learning to extract implicit events from documents of different types, has described the algorithm and explained experiments where the algorithm reaches excellent results. How the different features of the algorithm are contributed to such results. In particularly, the researcher has focused on the contribution of linguistic information. (LP) 2 has become the basis for two adaptive IE systems.

Borislav Popov et al. [32] have described that the semantic annotation platforme (KIM) which provides a novel Knowledge and Information Management infrastructure and services for automatic semantic annotation, indexing, and retrieval of documents. It provides mature infrastructure for scalable and customizable information extraction (IE) as well as annotation and document management, based on GATE. In order to provide basic level of performance and allow easy bootstrapping of applications, KIM is equipped with upper-level ontology and a knowledge base providing extensive coverage of entities of general importance. The ontologies and knowledge bases involved are handled using cutting edge Semantic Web technology and standards, including RDF(S) repositories, ontology middleware and reasoning. From technical point of view, the platform allows KIM-based applications to use it for automatic semantic annotation, content retrieval based on semantic restrictions, and querying and modifying the underlying ontologies and knowledge bases. While the previous approaches used ontologies mainly for enumerating the extractable concepts, defining their inheritance hierarchy and for storing their extracted instances, another approach emerged that attempted to augment the ontology itself with extraction knowledge

III. RESEARCH METHODOLOGY

For the proposed system, firstly bio data are created as ontological commitments. For example, in this system Class Animalia that may be created based on ontology for teaching aid. Firstly, bio data was created in spreadsheet as CSV format file which supports that simply represents relationships. The CSV format can be obtained from any row data, databases or Excel export. Each line must contain at least two elements. Each row of the file is a node or an edge. Working with spreadsheets are comfortable can be manipulated data in this environment. That is why; this is often a very suitable way of conceptual data based on ontology.

A. Creation of Ontology Based Bio Data

First, create a node table which has a unique identifier (first column: node), a name (second column), and a type/label (third column) and as needed (e.g. Scientific Name, feeding habit, breeding habit, habit, location, locomotion, body cover). Among these columns, third column to the last column are properties or attributes. These data are imported in a meaningful way into a graph database management system like Neo4j, and it has some additional information about the relationships between these data. Data are interconnected between their nodes vice versa.

The two CSV format files: nodes.csv and rels.csv are created with all nodes and relationships. Actually, biological ontology is constructed based on the biological data focuses the Animalia domain. Animals are multicellular, eukaryotic organisms of the kingdom Animalia which is taxonomic kingdom comprising all living or extinct animals. Generally, Animalia consists of six core classes: Chondrichthyes, Osthychthyes, Amphibia, Reptilia, Aves and Mammalia.

Chondrichthyes or cartilaginous fishes are jawed fish with paired fins, paired nares, scales, a heart with its chambers in series, and skeletons made of cartilage rather than bone. Osthychthyes or bony fish are a taxonomic group of fish that have bone, as opposed to cartilaginous, skeletons. A class Amphibia or amphibian is a cold-blooded vertebrate animal that is born in water and breathes with gills. As the larva grows into its adult form, the animal's lungs develop the ability to breathe air, and the animal can live on land. A reptile is one of those mostly scaly and slimy animals that are cold-blooded and have backbones. Snakes, salamanders, frogs, turtles, and crocodiles are all reptiles. A class mammalia or mammal is an endothermic amniotes distinguished from reptiles and birds by the possession of hair, three middle ear bones, and mammary glands.

The four relationship types are created in this biological domain such as has, is_a, is_a_ClassName, lives_in. The Animalia Class is the super class (root node in graph database) of this systems.

B. Biological Ontology Graph Database

The key benefits of using a DBMS is that data are safely and accurately shared to restricted users or even to the public. To achieve such goals, the basic concepts of a DBMS, including the ACID (Atomicity, Consistency, Isolation and Durability) properties, are to be considered in the database design. To control the validity of data changes occurring when the user performs updates to the database, the atomicity concept is applied.

Graph is a collection of nodes connected by edges. To apply graphs into the data model, a data entity is described as a node with node attributes, which is the same as in Object Oriented Database Management (OODM). Nodes are connected by edges to illustrate their binary relationship. Edges are typed and always stored pointers to start and end nodes. This data model is suitable for storing less descriptive data entities with complex relationships. With some limitations of these database models, a data model was designed by using a combination of the OODM and graph data model in programmatic data layer (conceptual data layer) storing tangible biological entities as objects and their relationships in binary relationship of the graph model to with the high complexity data. In this thesis, classes were designed by adapting the class description in the Biological ontology. Classes were strictly designed based on object-oriented programming concepts are as follows.

1. Data abstraction and encapsulation, properties and data structure of a class have to be protected from procedures and users. Data in the classes can only be accessed through the class interface called operator.

This design concept is used to control data consistency inside the object.

2. Inheritance is a mechanism of code reusability, which is a powerful object-oriented approach. Subclasses were generalized from its ancestor class called superclass and inherited the basis of class definition; such as attributes and methods.

Biological core classes such as Chondrichthyes, Osthychthyes, Amphibia, Reptilia, Aves and Mammalia and relationship types and their relations, properties types and their properties are defined as physical entity class mimicking node in the graph data model. Relationships among the classes are represented separately as relation class mimicking edge in the graph data model.

C. Implementation of Search Engine

Search engine is created as Search Class which has Group Search () method. Mynodes Class is created for graph node referred to as each animal. For the detailed search, any animal can be queried by common name and scientific name. For the group search, the complex query can be retrieved with various features such as feeding habit, breeding habit, etc. The complete class diagram can be represented in Figure 4.5, Class Diagram of Search, Admin Search and Database Operations

User input class name such as Chondrichthyes, the search engine finds the nodes which relationships with Common name Chondrichthyes and relationship type is “is_a_ClassName”. Aslo finds concurrently the nodes which relationships with Scientific name Chondrichthyes and relationship type is “is_a_ClassName”. The nodes with same class “Chrondrichthyes” node are extracted by search engine. The user input with desired common name, “Eagle-Ray” is retrieved from these nodes and displays the detailed information of class name, body cover type, breeding habit type, feeding habit type, locomotion type and scientific name.

The user input with desired scientific name, “Myliobatis” is retrieved from these nodes and displays the detailed information of class name, body cover type, breeding habit type, feeding habit type, locomotion type and scientific name. The output description for detailed search is same for searching common name and scientific name.

Group search provides five options types and also mix any option types. Group Search method has two arguments which are one ArrayList<Relationship> and one ArrayList<mynodes>. User input option types ArrayList<Relationship> are passed to GroupSearch and user input option values ArrayList<mynodes) are carried also. GroupSearch method finds the nodes which are agreement of parameter ArrayList<Relationship> and ArrayList<mynodes) and add to the ArrayList<mynodes> type collection. And then return that list. Public ArrayList<mynodes> GroupSearch(ArrayList<Relationship> rlst,

ArrayList<mynodes> nodelist)

IV. EXPERIMENTAL RESULTS AND EVALUATION

Implementation of ontology-based biological information system is presented in this chapter. The first implementation with Neo4j Community Server installation, configuration, and biology based ontology graph nodes are created in this database server. Biological ontology model is conceptually implemented in biological domain and class orientation of superclass and subclasses hierarchy structure is discussed. Java implementation of Mynode class, query engine class and search and admin search for any simple query and more complex query are also presented. Secondly, in the java implementation, detail search, group search and admin user for database operations are provided. The third implementation of graph engine component with neoclipse configuration and illustrated graph nodes are discussed in detail. Biological ontology plays an important role in representing the unstructured/semi-structured data of this thesis. Java based application is developed to collect different resumes through web search and to convert into common structured format. The most common using standard format can be obtained by the use of ontology. Graph Database Query Syntax and Results are presented.

A. Analysis of Bio Data

In this system, there are 149 elements in animalia to find out how many of them had both Carnivore and Viviparous. The number of elements in Carnivore and Viviparous can be found by calculating and drawing Venn diagram (Figure 5.3). The query results of bio data are given below.

n (C) = number of elements in Carnivore is 94

n (V) = number of elements in Viviparous is 67

$n (C \cap V)$ = number of elements in Carnivore and Viviparous is 30.

$n (C \cap V) = n (C) + n (V) - n (C U V) = 94 + 67 - 131 = 30.$

Be sure that the symbols in your equation have been d The results are generally consistent with the results of the analyses. The figures are fully consistent with the query results.

B. Evaluation

Ontology is a fairly complex structure and it is often more practical to focus on the evaluation of different levels of the ontology separately rather than trying to directly evaluate the ontology as a whole. In this thesis, biological ontology is evaluated of Hierarchy or taxonomy level. An ontology typically includes a hierarchical is-a relation between concepts. Although various other relations between concepts may be also defined, the “is-a” relationship is particularly important and may be the focus of specific evaluation efforts.

The biological ontology contains other such as is_a_ClassName, has, lives_in. The ontology to be evaluated may also be mapped to an upper ontology that defines constructs that are not in the KR language. For example, an upper ontology may define class, relation, property, attribute, facet, quality, or trope.

Task-based evaluations offer a useful framework for measuring practical aspects of ontology deployment, such as the human ability to formulate queries using the query language provided by the ontology, the accuracy of responses provided by the system’s inferential component, The degree of explanation capability offered by the system, the coverage of the ontology in terms of the degree of reuse across domains, The scalability of the knowledge base, and the ease of use of the query component [29].

Biological ontology identifies and defines a concise set of eight ontology quality criteria and presents as follows:

- Accuracy
- Adaptability
- Clarity
- Completeness
- Computational efficiency
- Conciseness
- Consistency
- Organizational fitness

C. Class Evaluation

Biological ontology has seven classes and 292 subclasses. Animalia is a super class (root node) and it has six subclasses. 292 subclasses are inherited from the corresponding super classes as shown in Table 1. Mammalia class has 108 subclasses which is the largest subclasses in this biological ontology.

Table 1 Biological Ontology Class Name and Number of Sub Classes

No	Class Name	Number of Sub Classes
1	Animalia	6
2	Chondrichthyes	20
3	Osthychthyes	42
4	Amphibia	22
5	Reptilia	64
6	Aves	30
7	Mammalia	108
Total		**292**

D. Relationship Evaluation

Biological ontology has four relationship types such as is_a_ClassName, is_a, has, lives_in, etc. Relationship types and number of relationships are described in Table 2.

Table 2. Relationship Types and Relationships

No	Relationship Type	Relationships
1	is_a_ClassName	286
2	is_a	286
3	Has	242
4	lives_in	143
Total		**957**

E. Precision and Recall

In pattern recognition and information retrieval with binary classification, precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also known as sensitivity) is fraction of relevant instances that are retrieved.

Precision = True class A / (True class A + False class A)

Recall = True class A / (True class A + False class B)

Where,

True class A (TA) - correctly classified into class A
False class A (FA) - incorrectly classified into class A
True class B (TB) - correctly classified into class B
False class B (FB) - incorrectly classified into class B

The result data analysis on feeding habit and breeding habit of this biological ontology are described in Table 3 and graph shows in Figure 1. According to the biological data, it can be seen that two complex search group queries for feeding habit (Carnivore) and breeding habit (Viviparous) is 30 and feeding habit (Carnivore) and breeding habit (Oviparous) is 63.

Table 3 Analysis of Feeding Habit and Breeding Habit

Feeding Habit	Breeding Habit	Bio Data	TA	FA	FB	Precision	Recall
Carnivore	Viviparous	30	30	0	0	100%	100%
Carnivore	Oviparous	63	63	0	0	100%	100%

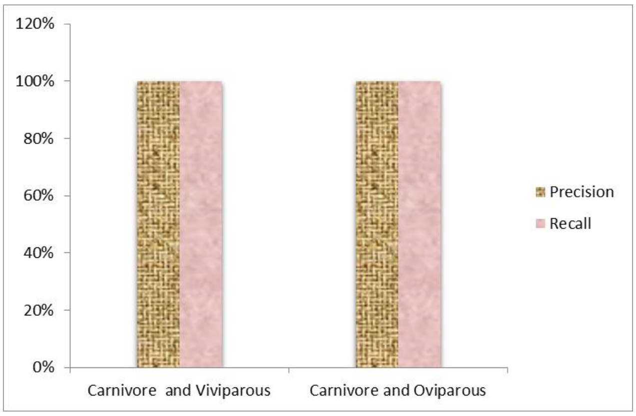

Figure 1 Comparing Result of the Two Complex Search Group Query

The result data analysis on feeding habit, breeding habit and habitat of this biological ontology are described in Table 4 and graph shows in Figure 2. According to the bio data, it can be seen that three complex search group queries for feeding habit (Carnivore), breeding habit (Viviparous), habitat (Marine Water) has 12, feeding habit (Carnivore), breeding habit (Viviparous), habitat (Terrestrial) has 13, feeding habitat (Carnivore), breeding habit (Viviparous), habitat (Arboreal) has 1 and feeding habit (Carnivore), breeding habit (Viviparous), habitat (Tundra) has 4 according to our biological data.

Table 4. Analysis of Feeding Habit, Breeding Habit and Habitat

Feeding Habit	Breeding Habit	Habitat	Bio Data	TA	FA	FB	Precision	Recall
Carnivore	Viviparous	Marine Water	12	10	1	2	91%	83%
Carnivore	Viviparous	Terrestrial	13	12	0	1	100%	92%
Carnivore	Viviparous	Arboreal	1	1	0	0	100%	100%
Carnivore	Viviparous	Tundra	4	4	0	0	100%	100%

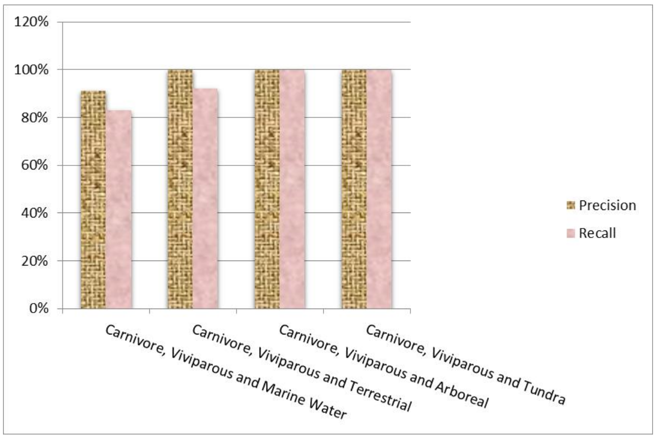

Figure 2 Comparing Result of the Three Complex Search Group Query

V. CONCLUSION

Ontology-based biological information system is proposed in this paper which provides the efficient and semantically related information retrieval features. Queries can be processed single search for detailed description, group search, and graph view with cypher query with colorful graph nodes. Ontologies based relationships are constructed in neo4j graph database and it can be used in the life sciences for alternatives to keyword-based approaches based in the traditional information retrieval framework and also image retrieval framework. Because of the cypher query language, it is also provided to retrieve fast access query. It can also retrieve relationship between living things.

This research supports ontology-base information retrieval system which provides the relevant research information requested by users. These semantic relationship-based queries system is implemented by applying biological domain. The semantic search by ontology can also provide the information of bio data in the specific search area at different types. Animalia and its respective properties, common name, scientific name, feeding habit, breeding habit, habitat, locomotion, and body cover can be retrieved by ontology search. Then, this proposed system can be matched between these types of bio data, e.g. query for two or three or more complex queries.

REFERENCES

1. Aitken, S. and S. Reid, Evaluation of an Ontology-Based Information Retrieval Tool, AIAI, University of Edinburgh, December 2014, http://citeseerx.ist.psu.edu
2. Antezana, E,. M. Kuiper and and V. Mironov, Biological Knowledge Management: The Emerging Role of the Semantic Web Technologies, Briefings in Bioinformatics, Advance Access publication, Oxford University Press, Vol. 10, No 4, pp 392-407, May 2009
3. Batra, S. and C. Tyagi, Comparative Analysis of Relational And Graph Databases, International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Vol. 2, Issue 2, May 2012
4. Bordoloi, S. and B. Kalita, Designing Graph Database Models from Existing Relational Databases, International Journal of Computer Applications (0975-8887), Vol. 74, No. 1, July 2013
5. Bouramoul, A., M. K. Kholladi and B. L. Doan, An Ontology-based approach for semantic ranking of the web search engines results, http://www.semantic-web-journal.net/sites/default/files/swj248.pdf
6. Brank, J., M. Grobelnik and D. Mladenić, A Survey Of Ontology Evaluation Techniques, Jozef Stefan Institute, Slovenia, http://eprints.pascal-network.org
7. Calado, P. et al 9, Searching Web Databases by Structuring Keyword-Based Queries, ACM 1-58113-492-4/02/0011, CIKM'02, Mclean, Virginia, USA, November 2002
8. Chen, Y. et al, Keyword Search on Structured and Semi-Structured Data, SIGMOD'09, Providence, Rhode Island, USA, ACM 978-1-60558-551-2/09/06, July 2009
9. Ciravegna, F., (LP)2: Rule Induction for Information Extraction Using Linguistic Constraints, Technical Report, CS-03-07, Computer Science, University of Sheffield, September 2003
10. Coto, A. L., The Use of Ontology for Improving Image Retrieval and Annotation, Technical Report, kmi-08-08, KMi Institute, September 2008
11. El-Qawasmeh, E. et al, A Framework For Processing Keyword-Based Queries In Relational Databases, Journal of Theoretical and Applied Information Technology, JATIT, Jordan, 2010, http://www.jatit.org
12. Embley, D. W. et al, Ontology-Based Extraction and Structuring of Information from Data-Rich Unstructured Documents, Brigham Young University, Provo, Utah 84602, USA, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.785…
13. Franz, M. N. and D. Thau, Biological Taxonomy and Ontoloy Development: Scope and Limitations, Biodiversity Informatics, Vol. 7, 2010, pp. 45-66
14. Ghige, P. M and Ruhi R. kabra, Relational Keyword Search System, International Journal of Engineering Research and General Science, Vol. 2, Issue 6, ISSN 2091-2730, November 2014
15. Goyal, N. and N. Sigh, A Review on Different Content Based Image Retrieval Techniques Using High Level Semantic Features, International Journal of Innovative Research in Computer and Communication Engineering, (An ISO 3297:2007 Certified Organization), Vol. 2, Issue 7, ISSN(Online): 2320-9801, ISSN(Print): 2320-9798, July 2014, http://www.ijircce.com
16. Handschuh, S., S. Staab and F. Ciravegna, S-CREAM-Semi-automaticCREAtion of Metadata, http://files.ifi.uzh.ch/ddis/iswc_archive/iswc/pps/...handchuh.pdf
17. He, H. and A. K. Singh, Graphs-at-a-time: Query Language and Access Methods for Graph Databases, SIGMOD'08, Vancouver, BC, Canada, ACM 978-1-60558-102-6/08/06, June 2008
18. Holzschuher, F. and R. Peinl, Performance of Graph Query Languages, EDBT/ICDT '13, Genoa, Italy, ACM 978-1-4503-1599-9/13/03, March 18 2013
19. Jakonienė, V. and P. Lambrix, Ontology-based Integration for Bioinformatics, Proceedings of the 31st VLDB Conference, Trondheim, Norway, 2005, https://www.ida.liu.se/~patla/publications/05odbis.pdf
20. Joshi, K. Raj, Graph Visualization Using the NoSQL Database,May 2013, http://hdl.handle.net/10365/22972
21. Joshi, M. D. et al, Image Retrieval and Re-Ranking Techniques-A Survey, Signal and Image Processing: An International Journal (SIPIJ), Vol. 5, No. 2, April 2014
22. Jouili, S. and S. Tabbone, Hypergraph-Based Image Retrieval for Graph-Based Representation, Elsevier, 2012, http://dx.doi.org/10.1016/j.patcog. 2012.04.016, 2012
23. Krishnan, A., Knowledge bases, Ontologies and Key-Value Stores, www.cbrg.ethz.ch /education/SDB/L4.pdf
24. Labsky, M., Information Extraction from Websites using Extraction Ontologies, Dissertation, Doctor of Philosophy in Computer Science, 2008
25. Magesh, N. and P. Thangaraj, Semantic Image Retrieval Based on Ontology and SPARQL Query, International Conference on Advanced Computer technology (ICACT), Proceeding published by International Journal of Computer Applications® (IJCA), 2011
26. Malafsky, G.P. and B. D. Newman, Organizing Knowledge with Ontologies and Taxonomies, TECHi2, Fairfax, VA, http://www.techi2.com
27. McColl, R. et al, A Performance Evaluation of Open Source Graph Databases, PPAA '14, Orlando, Florida, USA, ACM 978-1-4503-2654-4/14/02, February 2014, http://dx.doi.org/10.1145/2567634.2567638
28. Miller, Justin J., Graph Database Applications and Concepts with Neo4j, Proceeding of the Southern Association for Information Systems Conference, Atlanta, GA, USA, March 2013
29. Obrst, L. et al, The Evaluation Of Ontologies: Toward Improved Semantic Interoperability, July 2006, www.mel.nist.gov

30. Ortego, G. M., A Semantic Middleware to enhance current Multimedia Retrieval System with Content-based functionalities, © Servicio Editorial de la Universidad del País Vasco Euskal Herriko Unibertsitateko Argitalpen Zerbitzua, ISBN: 978-84-694-6824-1, 2011
31. Partner, J., A. Vukotic and N. Watt, Neo4j in Action, Neo4j in Action MEAP, Version 13, MEAP Edition, Manning Publications, 2014
32. Popov, B., et al, KIM-Semantic Annotation Platform, http://www.researchgate.net/pro -file/Atanas_Kiryakov/publication/221466086_KIM
33. Raguenaud, C., J. Kennedy and P. J. Barclay, A Database System for Supporting Taxonomic Work, www.soc.npier.ac.uk/~prometheus/prometheus_1/.../pro00.ps
34. Ramirez, J. A. V., Medical Image Retrieval Using Multimodal Semantic Indexing, Universidad Nacional de Colombia, Engineering School, 2013
35. Revuri, S ., S. R. Upadhayaya and P. S. Kuma, Using Domain for Efficient Information retrieval, COMAD 2006, Delhi, India, December 2006, © Computer Society of India, 2006
36. Schriml, L. M. et al, Disease Ontology: A Backbone for Disease Semantic Integration, Nucleic Acid Res. Jan 2012; 40 (D1): D940-D946. Published Online Nov 12, 2011, doi:10.1093/nar/gkr972 PMCID:PMC32 45088, Nov 2011
37. Schulz, S., H. Stenzhorn and M. Boeker, The Ontology of Biological Taxa, Bioinformatics, Vol. 24, ISMB, 2008, pp 1313-1321, doi:10.1094/ bioinformatics/btn 158, 2008
38. Shamsfard, M. and A. A. Barforoush, The State of Art in Ontology Learning: A Framework for Comparison (Accessed Date 2014, December) citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.5679…pdf
39. Silvescu, A., D. Caragea and A. Atramentov et al, Graph Databases (Accessed Date 2014, March) www.cse.iitk.ac.in/users/smitr/.../Graph%20Databases.pdf
40. Singaravelan, S. and D. Musugan, Survey on Content Based Image & Information
41. Retrieval, International Journal of Computational Engineering Research, ISSN:2250-3005, IJCER, Jan-Feb 2012, Vol. 2, Issue No.1, pp 168-171, 2012
42. Sklyar, N., Survey of Existing Bio-Ontologies, September, 2001, citeseerx.ist.psu.edu/viewdoc/download?/doi=10.1.1.226.2803…
43. Neo4j Team, The Neo4j Manual V I.9. M04, available at http://neo4j.org/, January 2013
44. Torres, R. S and A. X. Falcao, Content-Based Image Retrieval:theory and Application, RITA. Vol. XIII. Numero 2.2006, 2006
45. Vicknair, C., et al, A Comparison of a Graph Database and a Relational Database, ACMSE ’10, April 15-17, 2010, Oxford, MS, USA Copyright 2010 ACM 978-1-4503-0064-3/10/04
46. Vijayarajan, V., et al, A Review: From Keyword Based Image Retrieval to Ontology Based Image Retrieval, International Journal of Reviews in Computing, 31th December 2012, Vol. 12© 2009 - 2012 IJRIC & LLS, ISSN: 2076-3328, http://www.ijric.org, E-ISSN: 2076-3336, Dec 2012
47. Vogt, L., The Future Role of Bio-Ontologies for Developing a General Data Standard in Biology: Chance and Challenge for Zoo-Morphology, © Springer-Verlag, doi: 10.1007/s00435-008-008-5, October 2008
48. Vrandečić, D., Ontology Evaluation, June 2010, www.aifb.kit.edu/images/b5/OntologyEvaluation.pdf
49. Vu , Q. H. et al, A Graph Method for Keyword-based Selection of the top-K Databases, ACM SIGMOD’08, Vancouver, BC, Canada, 2008
50. Yasmin, M., M. Sharif, and S. Mohsin, Use of Low Level Features for Content
51. Based Image Retrieval: Survey, Research Journal of Recent Sciences, Vol. 2 (11), pp 65-75, ISSN 2277 – 2502, Res. J. Recent Sci, November 2013
52. Yu, B. et al, Effective Keyword-based Selection of Relational Databases, SIGMOD'07, Beijing, China, ACM 978-1-59593-686-8/07/0006, June 2007