The role of knowledgebase in semantic searching of images in big data particularly with respect to social networking sites

P. Bavithra Matharasi
Lecturer, Dept. of MCA,
Mount Carmel College,
Bangalore.
bavithramatharasi@gmail.com

N. Vijayalaksmi
Lecturer, Dept. of MCA,
Mount Carmel College,
Bangalore.
vijinitya@yahoo.co

Dr. A.Senthilrajan
Director
Computer Centre
Karaikudi
agni_senthil@yahoo.co

***Abstract*– The technological improvement has bought people closer virtually. Exchanging of knowledge has become easier through Social Networking Sites (SNS). There are numerous sites available today in the market. They are not used for mere fun and entertainment but even for serious business, research works, education and various medical reasons. The amount of data involved here is voluminous, variety and comes with more velocity, hence SNS are dealing with Big Data. There are both opportunities and challenges involved while handling big data. To analyze and make quick decisions, to find the behavioural pattern of users, to do research work or to study about disease, big data is best suited. Searching is the major activity done in this big data. Here we have made an attempt to develop a working model for semantic searching of images. In this suggested model, semantic searching of images is possible by integrating various social networking sites. Core of this model is a powerful knowledgebase. This paper highlights this working model.**

Keywords: Social Networking sites, Big Data, Knowledgebase, semantic image searching

I. INTRODUCTION

Due to the advancement in technology, people are socially able to stay in with their friends, relatives or anybody on earth. The advent of Social Networking Sites (SNS) has bought people in different demographic very closer. Distance is no matter for building relations. The purpose may be different like business, finding friends, education, government initiative, entertainment or medical reasons but linked via SNS. When the user exchange information, the amount of data involved is surplus. Users share text, images, audio and video, so a variety of data is involved. Further the data accumulation with respect to time is increasing day by day. Thus the data to be handled in this SNS are Big Data. When the data becomes Voluminous, Varity and comes with more Velocity, managing becomes more complex. There are both challenges and opportunities in big data. Bringing out the insight about the data to get more knowledge is feasible. But an important challenge put front in big data is mining or searching. When the traditional searching is diminishing in the course of time, semantic search is the need of the time. Thus leads to researches in NLP, ontology, Machine Learning and etc. are ongoing in text based dataset. Searching for images with relevant to context and intent of searching is semantic searching which is more challenging. Here we have made an attempt to develop a model that is used for semantic search of images in big data particularly with respect to social networking sites. There are different reasons for designing this model: (1) SNS is the weapon to connect people in different demographic (2) there are numerous advantages in this connection (3) the data involved here is big data (4) semantic search of images offers more advantages.

II. ADVANTAGES OF SNS

Social Media Networking is a buzzword and has become almost part and parcel of an individual. Social network are no more for mere fun and entertainment. More domains like business and financial sector, educational field[1], Science [2] and Researching communities[3], Disaster Management[4][5][6], Government[7] and Law enforcement[8], Medical[9] and Health care[10] are to name few are growing serious with social networking. Thus there are numerous advantages in using social media. The major technical operation performed here is searching.

III. BIG DATA

Big Data has become a buzzword that is the everywhere around us. Big data is characterized by 3Vs, namely Volume, Variety and Velocity. Big Data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process the data within a tolerable elapsed time [11]. Every day, we create 2.5 quintillion bytes of data, so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data[12].There is anexponential growth of both structured and unstructured data with respect to time. Structured data could be defined as the data that is organized in a highly format that is manageable. Structured data is ready for seamless integration into a database or well structured file format such as XML. Unstructured data is raw and unorganized data. Mining

through unstructured data can be cumbersome and costly. Examples of unstructured data include books, documents, medical records, and social media posts. There are both challenges and opportunities involved in big data. The main activity here is analyze and yield valuable new insights[13]. Another major activity involved in big data is searching.

IV. SEMANTIC SEARCHING

There basically two types of searches: Traditional Navigation searching and Semantic searching [14]. In the navigational searching the search engine retrieves all the web pages by comparing with the keyword. Even the unnecessary documents are retrieved in this type of searching. Semantic search is not applicable to navigational searches. In the semantic searching, a search is made with respect to the intent and context of searching. The user provides the search engine with a phrase which is intended to denote an object about which the user is trying to gather/research information. There is no particular document which the user knows about and is trying to get to. Rather, the user is trying to locate a number of documents which together will provide the desired information. Let us see Semantic Search with a real world example [15]. Let us say we are working on a computer and someone asks "Do you have windows there", it means whether you have Microsoft Windows operating system on your computer. But if we are approaching a Realtor looking for office space and ask him the same question, it takes a completely different meaning. In that context, the question means whether the office space you are discussing has any windows and ventilation in the room.

A. Factors to be considered

While performing a semantic searching, we have to consider the fowling factors

1. **Context of search**: The context is the circumstance or the environment in which the user is performing the search. As mentioned in the above example, the context of the user in the first circumstance in computer, whereas the later one is realtors looking for office.
2. **Location of search**: It is another factor that should be known for a search is location. Where does the user search from? For example, a user may wish to do online shopping for some beverages or food of preference in Bangalore. The search should specify the availability of the product depending on the landscape and delivery. There is no use in listing or displaying the entire product in all geographic location – unless mentioned by the user.
3. **Relevance of searching**: When the user is searching for a particular item or brand, more suggestions that are relevant to search could be listed
4. **Pre-knowledge**: Before the user starts a search, the system should have some knowledge about that user; context of searching item. So that the search will give a better result. From the above mentioned example, if a user is asking about windows, the system should know about both windows operating system and windows in a house or room. Also some knowledge should be pre-defined to use the context of searching in the right place.

B. Pre-requisition

To perform a semantic search a great deal of knowledge is needed. The above mentioned factors are to be considered. Following tools can be used along with the knowledge base.

1. Knowledge base: A powerful database along with inference logic
2. Dictionary: A lookup dictionary will help in to look at the meaning of words in the phrase
3. Basic knowledge or predefined knowledge: This knowledge could be rule based, object based, metadata format or any form of representation
4. 'alt' text: The alternate text for an image could be very useful in semantic searching
5. Tag: This is also a useful mechanism for searching an image and inferring about the user
6. Clustering: It is very important to perform a semantic search. The images are classified under different headings like business, entertainment, education, research etc. and images may be interlinked among these headings too. Clustering algorithm should be efficient which will help in this searching. Knowledge about clustering is also stored in the knowledge base

V. WORKING OF THIS MODEL

The model is as shown in the Fig 1. There are various components like an interface, knowledge base, tools for searching, result collecting and presenting the results.

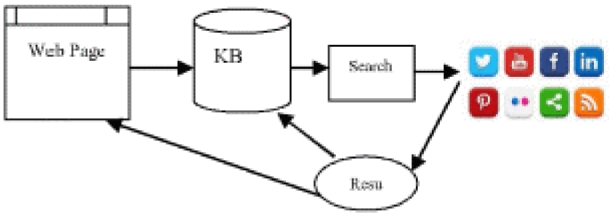

Fig. 1 Proposed Model

A. Interface

It is a web page through which the users interact with the system. They can search for images in two different modes: (1) images with text (2) images using images. It is shown in Fig. 2. User places the query through this interface and the results are displayed.

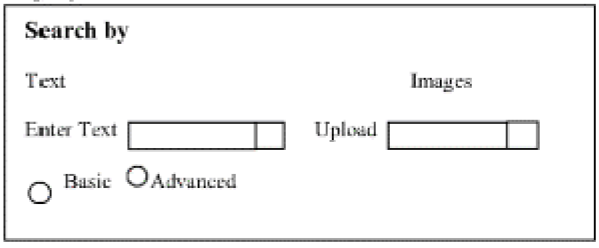

Fig. 2. Search Screen

B. Role of KB

In this model, knowledge base is the core, where data is not only acquired and stored but also represented and manipulated. As the information grows, it becomes more and more difficult to keep track of the relationships between the items. The most recent advancement of knowledge-based systems has been to adopt the technologies for the development of systems that use the Internet. The Internet often has to deal with complex, unstructured data that can't be relied on to fit a specific data model. The technology of knowledge-based systems and especially the ability to classify objects on demand is ideal for such systems. The model for these kinds of knowledge-based Internet systems is known as the Semantic Web [16]

As knowledge-based systems became more complex the techniques used to represent the knowledge base became more sophisticated. Rather than representing facts as assertions about data, the knowledge-base became more structured, representing information using similar techniques to object-oriented programming such as hierarchies of classes and subclasses, relations between classes, and behavior of objects. As the knowledge base became more structured reasoning could occur both by independent rules and by interactions within the knowledge base itself. Here we are using both queries and object oriented programming to store the data. For example, procedures stored as demons on objects could fire and could replicate the chaining behavior of rules[17]. Another important component is the use of a clustering algorithm that is used to classify images under predefined heading. This classified heading will be the indexing for searching. In this way a classifier also can play the role of an inference engine[18].

C. *Knowledge Representation*

Representation of knowledge plays an important role in designing knowledge base systems and expert systems. Nowadays there are many various knowledge models which have already been suggested and applied. In the books [19], [20], [21] and [22] we can find popular methods for knowledge representation in designing knowledge base systems. They include predicate logic, semantic nets, frames, deductive rules. Many new methods and techniques were presented in [23], [24], [25], and [26]. Among these methods neural networks and fuzzy logic can be used for computational intelligence. Some methods are suitable for representing and processing semantics such as conceptual graphs[21],[22]and [27]

D. *Characteristic of KB in this model*

Since the whole model is based on a knowledge base. It is the vital entity. It should be very efficient and powerful. It should have the following characteristics:

1. The Knowledge base should be able to hold voluminous data. The storage capacity of the Knowledge base should be enormously high and support for networking.
2. It should support for variety of unstructured data. Since the dataset to be used in the Social Networking Sites are of multimedia and unstructured, the knowledge base should be able to support and access.
3. Knowledge base is well build to support for various representation of knowledge. The knowledge is represented in the form of simple queries, metadata(to describe an object or entity), objects based, rule based or frame based.
4. When a new inference rule or knowledge arrives, the KB should be able to modify itself
5. An indexing mechanism is included as a result of clustering. Since this model is intended to integrate various SNS for a better search, clustering algorithm is used for classifying the images under different heading. So the searching will start from this index. Once the result is found from the sites, it is returned to the knowledge base for updation and sent to the interface too.

VI. CONCLUSION

Searching in big data is a big challenge as the unstructured data is growing exponentially with respect to time. This traditional search has its own disadvantages and searching becomes obsolete. Semantic search gives some meaning in searching. Social networking paves wave for the people to be connected and changes the life itself. There are numerous advantages of various social networks. So this domain has been chosen for our research work. A new theoretical model is suggested that tries to integrate various social networking sites. Knowledge base is the core of this model.

REFERENCE

[1] Gee, James Paul (2004). Situated Language and Learning: A Critique of Traditional Schooling. London: Routledge

[2] http://www.mysciencework.com/news/6400/social-networks-for-scientists

[3] Ferri F., Grifoni P., Guzzo T. (2012). "New forms of social and professional digital relationships: the case of Facebook". Social Network Analysis and Mining Journal, Vol. 2, pp. 121-137

[4] Appleby, L. (2013). Connecting the Last Mile: The Role of Communications in the Great East Japan Earthquake. London: Internews Europe

[5] Montgomery, D. (2013, July 2). CalgaryHerald. Retrieved August 29, 2013, from Calgary Herald: http://blogs.calgaryherald.com/2013/07/02/calgarians-fight-disaster-with-social-media/

[6] Williamson, R. A., & Antoniou, N. (2012). Data Policies in Support of Climate Change and Disaster Management Application. 63rd International Astronautical Congress. Naples: International Astronautical Federation

[7] Atari, S., Stewart-Weeks, M. & McCalla, J. (2011). Web 2.0 in Next-Generation Government and Governance: A Middle East Point of View. Cisco Internet Business Solutions Group (IBSG)

[8] http://source.southuniversity.edu/social-networking-and-law-enforcement-trends-0208.aspx#sthash.v231dnBG.dpuf

[9] Comprehensive listing of medical applications using social networking via Dose of Digital – http://www.doseofdigital.com/healthcare-pharma- ocial-media-wiki/

[10] Li J. Improving chronic diseases self-management through social networks. PopulHealth Manag 2013

[11] Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big Data': Big gaps of knowledge in the field of Internet.International Journal of InternetScience, 7, 1-5.http://www.ijis.net/ijis7_1/ijis7_1_editorial.html

[12] http://www-01.ibm.com/software/in/data/bigdata/

[13] Katina Michael, University of Wollongong Keith W. Miller, University of Missouri–St. Louis, June 2013 (Vol. 46, No. 6) pp. 22-240018-9162/13/$31.00 2013 IEEE Published by the

IEEE Computer Society Big Data: New Opportunities and New Challenges

[14] Guha, R.; McCool, Rob; Miller, Eric (May 24, 2003)."Semantic Search". WWW2003. Retrieved July 13, 2012

[15] John, Tony (March 15, 2012). "What is Semantic Search?". Techulator. Retrieved July 13, 2012

[16] Berners-Lee, Tim; James Hendler and Ora Lassila (May 17, 2001). "The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". Scientific American

[17] Mettrey, William (1987). "An Assessment of Tools for Building Large Knowledge- BasedSystems". AI Magazine. Volume 8 Number 4

[18] MacGregor, Robert (June 1991). "Using a description classifier to enhance knowledge representation".IEEE Expert 6 (3). Retrieved 10 November 2013

[19] John F. Sowa, Knowledge Representation: Logical, Philosophical and Computational Foundations, Brooks/Cole, 2000

[20] John F. Sowa, Knowledge Representation: Logical, Philosophical and Computational Foundations, Brooks/Cole, 2000

[21] Michel Chein & Marie-Laure Mugnier, Graph-based Knowledge representation: Computational foundations of Conceptual Graphs, Springer-Verlag London Limited 2009

[22] Frank van Harmelem, Vladimir, and Bruce, Handbook of Knowledge Representation, Elsevier, 2008

[23] Amit Konar, Computational Intelligence : Principles, Techniques and Applications, Springer-Verlag Berlin Heidelberg, 2005

[24] Leszek Rutkowski, Computational Intelligence: Methods and Techniques, Springer-Verlag Berlin Heidelberg, 2008

[25] ToshinoriMunakata, Fundamentals of the New Artificial Intelligence: Neural, Evolutionary, Fuzzy and More, Springer-Verlag London Limited, 2008

[26] M. Tim Jones, Artificial Intelligence : A System Approach, Infinity Science Press LLC, 2008

[27] F. Lehmann, Semantic Networks in Artificial Intelligence, Elsevier Science Ltd, 2008