

Image Search Semantic Search Using Approximate Methodology

V.V.Sunilkumar, Nunna Sai Sindhura

Associate Professor, Dept. of CSE, PBRVITS, Kavali, India

M.Tech Student 2nd year, Dept. of CSE, PBR VITS, Kavali, India

ABSTRACT: The difficulties of taking care of the unstable development in information volume and intricacy cause the expanding requirements for semantic questions. The semantic inquiries can be deciphered as the connection mindful recovery, while containing inexact outcomes. Existing distributed storage frameworks for the most part neglect to offer a sufficient capacity for the semantic inquiries. Since the genuine esteem or worth of information intensely relies on upon how productively semantic pursuit can be done on the information in (close) continuous, vast divisions of information wind up with their qualities being lost or altogether lessened because of the information staleness. To address this issue, we propose a close ongoing and practical semantic questions based strategy, called FAST. The thought behind FAST is to investigate and misuse the semantic relationship inside and among datasets by means of connection mindful hashing and reasonable level organized tending to altogether decrease the handling inactivity, while bringing about acceptably little loss of information hunt precision. The close constant property of FAST empowers quick recognizable proof of connected records and the huge narrowing of the extent of information to be prepared. Quick backings a few sorts of information investigation, which can be executed in existing accessible stockpiling frameworks. We lead a true utilize case in which youngsters revealed missing in a greatly swarmed condition (e.g., an exceptionally prominent picturesque spot on a pinnacle vacationer day) are distinguished in a convenient manner by breaking down 60 million pictures utilizing FAST. Quick is additionally enhanced by utilizing semantic-mindful namespace to give dynamic and versatile namespace administration for ultra-substantial capacity frameworks. Broad exploratory outcomes show the proficiency and viability of FAST in the execution enhancements.

I. INTRODUCTION

Storage systems are facing great challenges in handling the files from many data intensive application such as business transactions, scientific computing, and social network webs, mobile applications, information visualization, and cloud computing. Approximately 800 Exabyte of data were created in 2009 alone. According to a recent survey, 1,780 data center in 26 countries. The hierarchical directory tree based metadata management scheme used in almost all file systems today. The most important functions of namespace management are file identification and lookup File system namespace as an information-organizing infrastructure is fundamental to system's quality of service such as performance, scalability, and ease of use.

II. PROBLEM DOMAIN

Network Security relies on layer of protection and consists of multiple components including network monitor and security software addition to hardware and appliance. This all increase security of computer network. Network security covers variety of computer network like both public and private.

III. PROBLEM DEFINITION

Petabyte or Exabyte-scale data sets and Gigabit data streams are the frontiers of today's file systems. Storage systems are facing great challenges in handling the deluge of data stemming from many data-intensive applications such as business transactions, scientific computing, social network webs, mobile applications, information visualization, and cloud computing. Approximately 800 Exabyte of data were created in 2009 alone. This reflects a reality in which we are generating and storing much more data than ever and this trend continues at an accelerated pace.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

This data volume explosion has imposed great challenges to storage systems, particularly to the metadata management of file systems. For example, many systems are required to perform hundreds of thousands of metadata operations per second and the performance is severely restricted by the hierarchical directory-tree based metadata management scheme used in almost all file systems today.

IV. OBJECTIVE AND SCOPE OF THE PAPER

The most important functions of namespace management are file identification and lookup. File system namespace as an information-organizing infrastructure is fundamental to system's quality of service such as performance, scalability, and ease of use. Almost all current file systems, unfortunately, are based on hierarchical directory trees. This namespace design has not been changed since it was invented more than 40 years ago. As the data volume and complexity keep increasing rapidly, conventional namespace schemes based on hierarchical directory trees have exposed the many weaknesses.

V. SEMANTIC SEARCH SYSTEMS

Conventional search techniques are developed on the basis of words computation model and enhanced by the link analysis. On one hand, semantic search extends the scope of traditional information retrieval paradigm from mere document retrieval to entity and knowledge retrieval; on the other hand, it improves the conventional IR methods by looking at a different perspective: the meaning of words, which can be formalised and represented in machine processible format using ontology languages such as RDF¹ and OWL². For example, an arbitrary resource or entity can be described as an instance of a class in an ontology; having attribute values and relations with other entities. With the logical representation of resources, a semantic search system is able to retrieve meaningful results by drawing inference on the query and knowledge base. As a simple example, meaning of the query for "people in School of Computer Science" will be interpreted by a semantic search system as individuals (e.g., professors and lecturers) who have relations (e.g., work for or affiliated with) with the school. On the contrary, conventional IR systems interpret the query based on its lexical form. Web pages in which the worlds "people" and "computer science" co-occur, are probably retrieved. The cost is that users have to extract useful information from a number of pages, possibly query the search engine several times. As we will see shortly, other inference mechanisms based on logical rules and inductive approaches have also been evaluated to enable a system to interpret and understand ad-hoc queries.

To facilitate our study, we classify the research of semantic search into six categories in accordance with their objectives, methodologies, and functionalities. Categorisation and discussion of semantic search systems based on alternative criteria such as user-system interaction can be found in [31].

5.1 Document-oriented Search

Document-oriented search can be thought of as an extension of the conventional IR approaches where the primary goal is to retrieve documents, such as web pages and text documents [24, 35, 28], document fragments [10], scientific publications [44, 18] and ontologies [19]. In search systems of this category, documents are annotated using logic-based knowledge representation languages with respect to annotation and domain ontologies to provide approximate representations for the topics or contents of original documents. The retrieval process is carried out by matching user queries with the resulting semantic annotations.

The early work in SHOE search system facilitates users constructing constrained logical queries by specifying attribute values of ontology classes to retrieve precise results [24]. The main limitations are the manual annotation process and lack of inference support. There has been considerable work on speculating the significance of logical reasoning for semantic search systems. OWLIR [35] adopts an integrated approach that combines logical inference and traditional information retrieval techniques. A document is represented using original text and semantic markup. The semantic markup and the domain ontology are exploited by logical reasoners to provide enhanced search results. Conventional IR system based on the original text is integrated into the semantic search system to provide

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

complementary results in case no result is returned by the semantic search. In a similar work, a semantic search framework for annotation, indexing and retrieval of documents called KIM is introduced by [28]. The automated annotation framework is based on information extraction technologies concerning named entities. With the semantic annotation of the content, constrained queries with regard to entity type, name, attribute and relation are formulated to obtain precise results. In addition, a query pre-processing step with logical inference enables the system to interpret user queries in meaningful ways [28]. Besides retrieval of full documents, the idea of text retrieval based on document fragments is implemented in the DOSE system [10]. The intention is to provide users short text fragments which are likely to contain relevant information rather than the whole document. One of the common limitations of the above discussed work is that the context of development is based on a “closed world” assumption that does not take heterogeneous nature of the semantic web into consideration (e.g., sources may publish different ontologies in similar domains).

Several semantic search systems have also been developed to provide alternative ways for researchers to explore and browse large repositories of scholarly articles [44, 17]. The IRIS [44] provides an inference engine to perform reasoning with rules over a computer science literature domain ontology to elicit implicit knowledge. During user interaction with the system, semantically related, broader and narrower concepts (the IRIS domain ontology is built on SKOS³) are recommended to facilitate user browsing and refining queries. FacetedDBLP [18] is a faceted browser that helps users to browse scientific publications based a number of facets. The concept facet is built using the GrowBag algorithm which exploits keywords co-occurrences to construct concept hierarchies [17].

A semantic web document is different from a standard document (e.g., web page, email and scholarly article) because it is authored primarily for automatic machine processing, though it is human readable. Further, unlike hyperlinks between Web documents, links between semantic web documents are labeled with meanings. To cope with the problems, an ontology ranking algorithm which is based on a “rational surfer model” rather than “random surfer model”, OntoRank, is introduced in [20]. An search engine called Swoogle [19] is implemented to search ontology documents on the Web.

5.2 Entity and Knowledge-oriented Search

Entity and knowledge-oriented search method expands the scope of conventional IR systems which solely retrieve documents. Systems based on this method often model ontologies as directed graphs and exploit links between entities to provide exploration and navigation. Attribute values and relations of the retrieved entities are also shown to provide additional knowledge and rich user experiences [22, 23, 38, 21].

TAP is one of the first empirical studies of large scale semantic search on the Web. It improves the traditional text search by understanding the denotation of the query terms, and augmenting search results using a broad-coverage knowledge base with an inference mechanism based on graph traversal [22]. Recently a number of systems such as CS-Aktive [38], RKB Explorer [21], and SWSE [23] have been implemented based on the principles of “open world” and “linked data⁴”, which coincide with spirit of the semantic web of being “distributed”. CS-Aktive and RKB Explorer provide unified views of information (using graphical interfaces) collected from a significant number of heterogeneous data sources. To resolve the problem that heterogeneous sources may publish different information about same set of entities, CS-Aktive uses its reference ontology as the mediator [38], while RKB Explorer implements set of consistent reference services, which are essentially knowledge bases of URI equivalence generated using heuristics [21]. A notable feature of the SWSE system is a hybrid data integration solution for the collected data, such as existing RDF datasets, XML database dumps, various static and live data sources. Data consolidation is realised using a simple mechanism through analysis of inverse functional properties of entities [23]. Retrieved results are ranked by tuning PageRank algorithm using context information (i.e., provenance of data) [27].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

5.3 Multimedia Information Search

In ontology-based image annotation and search systems, re-sources are annotated by domain experts or using text around images in documents. Early works adopt manual approach which suppresses scalability [37]. The work in [43] performs query expansion to retrieve semantically related images by drawing logical inference on its domain ontology (e.g., using subsumption relation between concepts). Falcon-S [45] anno-tates images using metadata by crawling and parsing pages on the Web. Disambiguation of resources with same labels is re-solved using context information derived from user queries [45]. Squiggle [13] is a semantic framework to help build-ing domain-specific semantic search applications for indexing and retrieving multimedia items. Its knowledge representation model is based on the SKOS vocabulary which enables the sys-tem to suggest meanings of queries by a simple inference pro-cess, e.g., suggest alternative labels or synonyms for an image. As a result, images annotated with one label can be retrieved using the image's alternative labels.

5.4 Relation-centered Search

Relation-centered semantic search approach pays special at-tention to relations between query terms implicitly expressed by users. It usually performs an additional query pre-processing step through the use of external language lexicons, or an infer-ence process using a knowledge base [33, 31, 28, 32].

AquaLog is a question-answering system. It implements similarity services for relations and classes to identify syn-onyms of verbs and nouns appearing in a query using Word-Net⁵. The obtained synonyms are used to match property and entity names in the knowledge base for answering queries [33]. SemSearch supports natural language queries and translates them into formal queries for reasoning [31]. An entity referred by a keyword is matched against a subject, predicate or object in the knowledge base using combinations.

The above two systems process entity relations using lan-guage lexicon or word combinations. While in KIM [28] and OntoLook [32], relations between query terms are inferred from knowledge bases to aid the retrieval process, for exam-ple, in KIM a query “company, Redwood Shores” could be understood by the system that the intention of the query is to retrieve documents mentioning the town and companies with geographical constrains (i.e., companies located in the town), but not the word “company” [28]. OntoLook assembles query terms to concept pairs and send these pairs to the knowledge base to retrieve all relations which have been asserted in the knowledge base.

5.5 Semantic Analytics

Semantic analytics, also known as semantic association analy-sis, is introduced by Sheth et al. to discover new insights and actionable knowledge from large amounts of heterogeneous content [39]. It is essentially a graph-theoretic based approach that represents, discovers and interprets complex relationships between resources. By modeling RDF database as directed graph, relationships between entities in knowledge base are represented as graph paths consisting of a sequence of links.

Search algorithms for semantic associations such as breadth first and heuristics-based search are discussed in [40, 5]. Effec-tive ranking algorithms are indispensable because the number of relationships between entities in a knowledge base might be much larger than the number of entities. In [2] the ranking eval-uates a number of parameters: context, subsumption, trust, rar-ity, popularity, and association length. In another work called SemRank, a blend of semantic and information-theoretic tech-niques are used to analyse and rank the semantic associations [4]. Semantic association analysis has been deployed in several real world applications, such as national security application [40] and conflict of interest detection [3]. A relation robustness evaluation algorithm for semantic associations has been imple-mented in the MANA multimedia presentation generation sys-tem to automatically select multimedia objects for presentation generation [8].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

5.6 Mining-based Search

Effectiveness of the semantic search depends largely on the quality and coverage of the underlying knowledge base. The search methodologies discussed so far either utilise explicit knowledge, which is asserted in the knowledge base, or implicit knowledge, which is derived using logical inference with rules. Another kind of knowledge, which we refer to as “hidden knowledge”, cannot not be easily observed using techniques such as information extraction, natural language processing, logical inference, and semantic analytics. For example, “who are the experts in semantic web research community?”, “Which institutions ranks highly in the machine learning research area?”. Such knowledge can only be derived from large amount of data by using some sort of sophisticated data analysis techniques. We refer to approaches that utilise techniques to infer hidden knowledge as mining-based semantic search.

Flink is a semantic web application for extraction, aggregation and visualization of an online social network that consists of professional work and social connectivity of semantic web researchers [36]. It allows users to identify prominent researchers in the semantic web community based on popular measures in social network analysis (i.e., degree, betweenness, and closeness centrality measures [42]). Ontocopi uses a method that combines breadth first and spreading activation search to identify communities of practice from its knowledge base [1]. Arnetminer [41] is another semantic-enhanced application which is featured by a number of mining services, e.g., expert finding. Using researchers’ publication, the expert mining service utilises the PLSA model to extract and rank experts with respect to user queries.

VI.CONCLUSION

This paper proposes a near real-time scheme, called FAST, to support efficient and cost-effective searchable data analytics in the cloud. FAST is designed to exploit the correlation property of data by using correlation-aware hashing and manageable flat-structured addressing. This enables FAST to significantly reduce processing latency of correlated file detection with acceptably small loss of accuracy. We discuss how the FAST methodology can be related to and used to enhance some storage systems, including Spy-glass and SmartStore, as well as a use case. FAST is demonstrated to be a useful tool in supporting near real-time processing of real-world data analytics applications.

REFERENCES

- [1] H Alani, K O’Hara and N Shadbolt, Ontocopi: Methods and tools for identifying communities of practice. *Intelligent Information Processing 2002*, Vol. 221, pp. 225–236.
- pp. B Aleman-Meza, C Halaschek-Wiener, I B Arpinar, C Ramakrishnan and A P Sheth, Ranking complex relationships on the semantic web. *IEEE Internet Computing*, Vol. 9, No. 3, 2005, 37–44.
- [2] B Aleman-Meza, M Nagarajan, C Ramakrishnan, L Ding, P Kolari, A P Sheth, I B Arpinar, A Joshi and T Finin, Semantic analytics on social networks: experiences in addressing the problem of conflict of interest detection. *WWW 2006*, ACM, 407–416.
- [3] K Anyanwu, A Maduko and A P Sheth, Semrank: ranking complex relationship search results on the semantic web. *WWW 2005*, ACM, pp. 117–127.
- [4] K Anyanwu and A P Sheth, -Queries: Enabling Querying for Semantic Associations on the Semantic Web. *WWW 2003*, pp. 690–699.
- [5] D Artz and Y Gil, A survey of trust in computer science and the semantic web. *J. Web Sem.*, Vol. 5, No. 2, 2007, pp. 58–71.
- [6] R A Baeza-Yates and B A Ribeiro-Neto, *Modern Information Retrieval*, ACM Press / Addison-Wesley, 1999.
- [7] P M Barnaghi and S A Kareem, Relation robustness evaluation for the semantic associations. *EJKM*, Vol. 5, 2007, pp. 265–272.
- [8] A D L Battista, N Villanueva-Rosales, M Palenychka and M Dumontier, Smart: A web-based, ontology-driven, semantic web query answering application. *Proceedings of ISWC2007 2007*.
- [10] D Bonino, F Corno and L Farinetti, Dose: A distributed open semantic elaboration platform. *ICTAI 2003*, pp. 580–588.
- [11] S Brin and L Page, The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, Vol. 30, No. 1–7, 1998, pp. 107–117.
- [12] T Bumers-Lee, J Hendler and O Lassila, The semantic web. *Scientific American*, Vol. 284, No. 5, 2001.
- [13] I Celino, E D Valle, D Cerzza and A Turati, Squiggle: a semantic search engine for indexing and retrieval of multimedia content. *Proceedings of SAMT 2006 2006*, pp. 20–34.
- [14] H Chen, Machine learning for information retrieval: Neural networks, symbolic learning, and genetic algorithms. *JASIS*, Vol. 46, No. 3, 1995, pp. 194–216.
- [15] P Cimiano, *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [16] S C Deerwester, S T Dumais, T K Landauer, G W Furnas and R A Harshman, Indexing by latent semantic analysis. *JASIS*, Vol. 41, No. 6, 1990, pp. 391–407.
- [17] J Diederich and W T Balke, The semantic growbag algorithm: Automatically deriving categorization systems. *ECDL 2007*, Vol. 4675, pp. 1–13.
- [18] J Diederich, W T Balke and U Thaden, Demonstrating the semantic growbag: automatically creating topic facets for facetddblp. *JCDL 2007*, ACM, p. 505.
- [19] L Ding, T Finin, A Joshi, R Pan, R S Cost, Y Peng, P Reddivari, V Doshi and J Sachs, Swoogle: a search and metadata engine for the semantic web. *CIKM ’04 New York, NY, USA, 2004*, pp. 652–659.
- [20] L Ding, T W Finin, A Joshi, Y Peng, R Pan and P Reddivari, Search on the semantic web. *IEEE Computer*, Vol. 38, No. 10, 2005, pp. 62–69.
- [21] H Glaser and I C Millard, Rkb explorer: Application and infrastructure. *Proceedings of Semantic Web Challenge 2007*.
- [22] R V Guha, R McCool and E Miller, Semantic search. *WWW 2003*, pp. 700–709.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

- [23] A Harth, A Hogan, R Delbru, J Umbrich, S Oriain and S Decker, Swse: Answers before links!. Proceedings of Semantic Web Challenge 2007.
- [24] J Heflin and J Hender, Searching the web with SHOE. Artificial Intelligence for Web Search Menlo Park, CA, 2000, pp. 35–40.
- [25] M Hildebrand, J van Ossenbruggen and L Hardman, An Analysis of Search-based User Interaction on the Semantic Web, Tech. rep., Centrum voorWiskunde en Informatica (NL), 2007.
pp. T Hofmann, Probabilistic latent semantic analysis. UAI 1999, 289–296.
- [26] A Hogan, A Harth and S Decker, Reconrank: A scalable ranking method for semantic web with context. Proceedings of SSWS2006 2006.
- [27] A Kiryakov, B Popov, I Terziev, D Manov and D Ognyanoff, Semantic annotation, indexing, and retrieval. J. Web Semantics., Vol. 2, No. 1, 2004, pp. 49–79.
- [28] J M Kleinberg, Authoritative sources in a hyperlinked environ-ment. SODA 1998, pp. 668–677.
- [29] S Lawrence, C L Giles and K Bollacker, Digital libraries and Autonomous Citation Indexing. IEEE Computer, Vol. 32, No. 6, 1999, pp. 67–71.
- [30] Y Lei, V S Uren and E Motta, Semsearch: A search engine for the semantic web. EKAW 2006, pp. 238–245.
- [31] Y Li, Y Wang and X Huang, A relation-based search engine in semantic web.. IEEE Trans. Knowl. Data Eng., Vol. 19, No. 2, 2007, pp. 273–282.
- [32] V Lopez, M Pasin and E Motta, Aqualog: An ontology-portable question answering system for the semantic web. ESWC 2005,
pp. 546–562.
- [33] E Makel`a, Survey of Semantic Search Research, Tech. rep., Semantic Computing Research Group,Helsinki Institute for Information Technology, 2007.
- [34] J Mayfield and T Finin, Information retrieval on the semantic web: Integrating inference and retrieval. Proceedings of Workshop Semantic Web at SIGIR 2003 2003.
- [35] P Mika, Flink: Semantic web technology for the extraction and analysis of social networks. J. Web Semantics, Vol. 3, No. 2-3, 2005, pp. 211–223.
- [36] A T Schreiber, B Dubbeldam, J Wielemaker and B J Wielinga, Ontology-based photo annotation. IEEE Intelligent Systems, Vol. 16, No. 3, 2001, pp. 66–74.
pp. N Shadbolt, N Gibbins, H Glaser, S Harris and M M C Schraefel, Cs active space, or how we learned to stop worrying and love the semantic web. IEEE Intelligent Systems, Vol. 19, No. 3, 2004, 41–47.
- [37] A Sheth, I B Arpinar and V Kashyap, Relationships at the Heart of Semantic Web: Modeling, Discovering, and Exploiting Complex Semantic Relationships, Springer-Verlag, 2002.
- [38] A P Sheth, B Aleman-Meza, I B Arpinar, C Bertram, Y Warke, C Ramakrishanan, C Halaschek, K Anyanwu, D Avant, F S Arpinar and K Kochut, Semantic association identification and knowledge discovery for national security applications. J. Database Manag., Vol. 16, No. 1, 2005, pp. 33–53.
- [39] J Tang, J Zhang, D Zhang, L Yao, C Zhu and J Li, Arnetminer: An expertise oriented search system for web community. Proceedings of ISWC2007 2007.
- [40] S Wasserman and K Faust, Social network analysis: methods and applications, 1st ed., Cambridge Univ. Press, Cambridge, 1997.
- [41] W Wei and P M Barnaghi, Semantic support for medical image search and retrieval. BIEN '07 Anaheim, CA, USA, 2007,
pp. 315–319.
- [42] W Wei, P M Barnaghi and A Bargiela, The Anatomy and Design of A Semantic Search Engine, Tech. rep., School of Computer Science, University of Nottingham Malaysia Campus, 2007.
- [43] H H Wu, G Cheng and Y Z Qu, Falcon-s: A ontology-based approach to searching objects and images in the soccer domain. ISWC 2006.
- [44] J Zhang, J Tang, L Liu and J Li, A mixture model for expert finding. To Appear in Proceedings of 2008 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2008) 2008.

BIOGRAPHY



Mr.Vemula.V. SUNIL KUMAR has received his B.Tech in Electrical Communication Engineering and M.Tech degree in Computer science from JNTU, Hyderabad 2008. He is dedicated to teaching field from the last 12 years. He has guided 12 P.G Students and 25U.G students. His research included CN- MANETs, Neural Networks, Image processing, Embedded systems. At present he is working as Assoc Professor on PBR VITS,kavali.



Mrs.Nunna Sai Sindhurareceived B.Tech in Computer Science and Engineering from PBR Visvodaya Institute of Technology And Scienceaffiliated to the Jawaharlal Nehru technological universityAnathapur in 2014, and pursuing M. Tech in Computer Science and Engineering from PBRVisvodaya Institute of Technology And Science affiliated to the Jawaharlal Nehru technological university Anantapur in 2017, respectively.