



# World Scientific News

An International Scientific Journal

WSN 41 (2016) 76-82

EISSN 2392-2192

---

---

## Information Retrieval System for Internet of Things: A Survey

**C. Indhu\***, **S. Sivakumari**, **R. Praveena Priyadharsini**

Department of Computer Science Engineering,  
Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India

\*E-mail address: [indhu.jul02@gmail.com](mailto:indhu.jul02@gmail.com)

### ABSTRACT

Internet of things (IoT) enables the communication between the users and the devices with internet. The information that is retrieved from the internet should be based on the users' query by enabling the authenticate access. The retrieved information should be of topic related, semantic, context aware, indexing system and ontology based retrieval system. The most observable applications for Information Retrieval are Web search engine. There are multiple papers based on the information retrieval system. The idea of this paper is to analyze the literature on the retrieval of information based on the query relevant context.

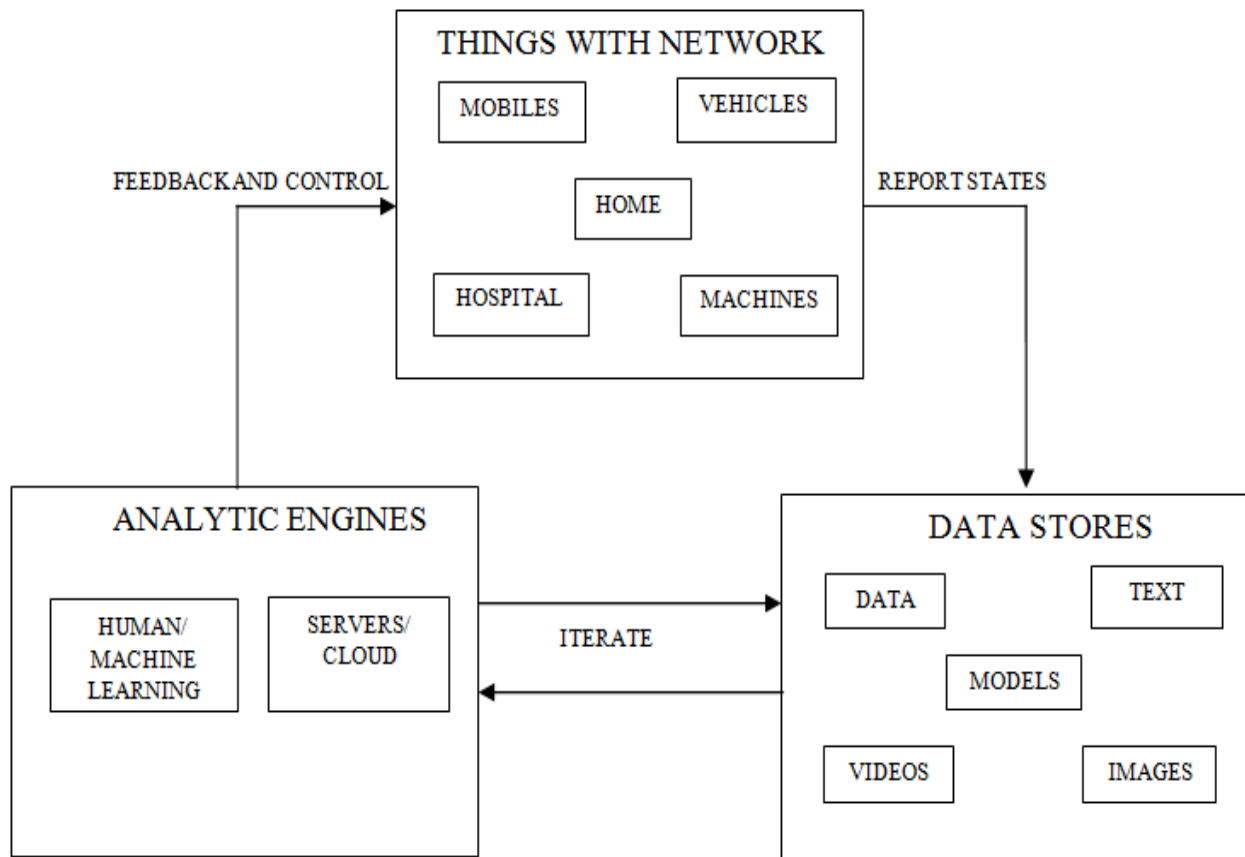
**Keywords:** IoT, communication, Information retrieval

### 1. INTRODUCTION

The Internet of things links smart objects to the Internet. It can allow an exchange of data never available before, and brings users' information in a more secure way. The Internet of Things has been most correlated with machine-to-machine (M2M) communication in manufacturing, power utilities and data transfer. Products built with M2M communication capabilities are often referred as being smart. The IoT extends internet connectivity beyond traditional devices to a diverse range of devices and everyday things that utilize devices to

communicate and interact with the external environment through the Internet. It makes our life more efficient.

There is no limit to the opportunities in the field of IoT until our imagination continues. Like internet, IoT has the potential to change the world. The Internet of Things will disintegrate the sectors and can reduce the cost. It empowers the physical objects to communicate with the internet and be interactive, responsive, and a base of information access and transfer.



**Fig. 1.** Interaction between the components in IoT

Information retrieval is mainly focused on searching methods and structure of the information within files, databases, web documents, etc. Information retrieval is characterized as a process of identifying a set of documents that are relevant to selected topic or query [1].

The steps of information retrieval processes are

- Obtains the information regarding the basic resources, related attributes, etc.
- Integrates the object related information, realizes the intelligent indexing and integration of the information
- Utilizes the computing technologies and process the information related to objects to realize decision and control in the real world.

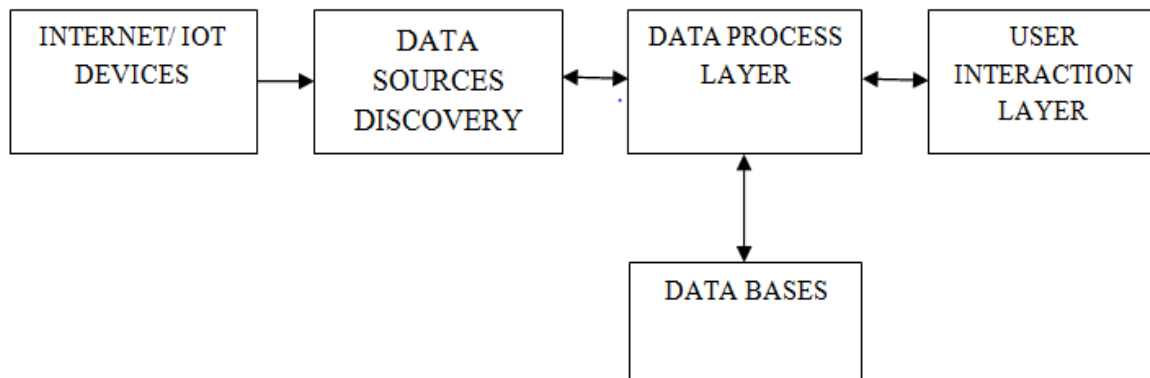


Fig. 2. Information Retrieval model

## 2. RELATED WORKS

### 2. 1. Search engine based Information Retrieval

Chiu C. Tan et al [5] proposed a Micro search in ubiquitous computing environments for embedded devices. Information inside a small device are indexed by user's queries are resolved accurately. The Conventional search engine is designed with limited hardware and algorithms cannot be used. Information retrieval techniques are adopted for query resolution, and propose an efficient top-k query resolution algorithm.

Huajing Li et al [7] discussed the problems of the current CiteSeer architecture and proposed a new architecture for a next generation CiteSeer application. The modular web services and pluggable service components are the main basis. The proposed new data model was basically different from that of the old CiteSeer model. The system has data flexibility, good scalability and query efficiency.

Haodong Wang et al [13] proposed Snoogle, a search engine in which user's queries to find a particular mobile object or a list of objects. Snoogle information is indexed and user queries are processed by using information retrieval, and communication overhead is reduced by bloom filters. To protect sensitive information from the users Security and privacy protections are designed into Snoogle. Communication costs are reduced in the system by making use of compressed Bloom filter of object data and the false positives are maintained at a low rate. A Flexible security method is introduced using public key cryptography that protects user privacy.

Norasykin Mohd Zaid et al [16] proposed an Ontology-based search framework consist elements such as user, query, query interface, search process, and result. The query interface is represented as keywords or sub-keywords structures based on the ontology stratified structure. The relationships between terms to be defined and keywords are enabled by ontology. In Ontology, the information is retrieved by sharing common vocabularies. After the query is sent, the search query is processed in the system, and the search process looks for the matched documents and the user gets their result. The obtained results can be reviewed and if the users are not fulfilled with the results the process can be repeated again.

### 2. 2. Digital library based Information Retrieval

Jeremy Stribling [6] presented OverCite, a design based on a distributed hash table (DHT) for a cooperative and distributed library. Document search and retrieval service is provided to researchers worldwide. The basic evaluation results in servicing more queries per second, and the total storage capacity is increased in a system of  $n$  nodes by  $n/4$  factor. The OverCite is advantageous as it supports the scaling of large data sets and document alerts.

Robert Kahn et al [8] proposed naming, identifying, and invoking digital object methods in a system which provides a well-suited national-level enterprise and great flexibility. It's possible to locate digital objects without any prior knowledge about the object or its locations. The users have a benefit of admitting added conventions that various users may.

### **2. 3. Indexing Information Retrieval**

C. Lee Giles et al [2] presented an autonomous citation indexing system which indexes academic literature in electronic format. CiteSeer understands the different formats of citation parsing, citation identifying in a different format to the same paper and also the contents if citations are identified. The literatures are retrieved by citation links such as journals, authors, the evaluation and ranking of papers, etc. based on the research identifications and the number of citations.

Steve Lawrence et al [3] discussed the creation of digital libraries of scientific literature on the web. The works are efficient location and full-text indexing of articles, autonomous citation indexing, information extraction, displaying the citation context and query sensitive summaries, detection of similar documents and overlapping of documents, correction of distributed errors, user profiling, analyzing the graphs, computation of authorities and hubs. The details of several aspects are proposed by focusing the citation indexing components.

Mathias Lux [11] presented LIRe (Lucene Image Retrieval) which is an open source Java library for retrieving the content based image. Global image features are provided in which indexing and retrieval are offered. It is based on an embedded text search engine which is lightweight. The Integration of applications is done without depending on a database server.

### **2. 4. Context aware information retrieval**

Michael J. Covington et al [4] introduced environment roles, and created a uniform access control framework used to secure context-aware applications. Security architecture is presented to support security policies which make use of environment roles, to control access to resources. Security-relevant context is captured by using the well-developed concepts which access the requests. The access control frameworks are highly versatile and accurate.

Michalis Giannikos et al [15] proposed design and implement architecture for retrieving information in the IoT. In the proposed scheme, information retrieval is limited using the access control policies and information is organized to make easy the development of context aware services. The paper presented lookup architecture for the IoT which tries to achieve a dual goal: access control is provided for the information items, the development of context aware services is facilitated. In contrast to similar approaches, this schema object is associated with much information, but users are allowed to access only a subset: the items that are associated with the same attributes as the users. It associates an item with a user role unauthorized access of the item is protected and its context-specific.

Charith Perera et al [17] present the context aware fundamentals and necessary backgrounds in the IoT model. The evaluation results highlights the past lessons and finds

possible ways for future works. The importance of context awareness is shown clearly. The goal has been achieved by helping us to understand the happenings in the past and the future plans effectively and efficiently.

## **2. 5. Semantic based Information Retrieval**

Xiaomin Ning [9] present SemreX which implements the retrieval of literature on a large scale efficiently and a single access point are browsed. It proposes the concept of Semantic Association to reveal relationships explicitly or implicitly between semantic entities by merging with the ontology-based information visualization technique. It makes easier for retrieving semantically relevant information, and the users' current search intentions are captured by the content relationships while preserving the scientific knowledge.

Feng Zhao et al [18] proposed an Acrost retrieval system for IoT environment based on the topic centric and semantic aware retrieval system. Two topic centric collectors are combined to obtain the initial contents of the information. By accumulating the conditional random fields-based and regular expression based approaches, the metadata is extracted. The relevant contents are ranked and the query is parsed for achieving the semantic aware retrieval. The model improves the efficiency and response time.

## **2. 6. Data mining based Information Retrieval**

Hoefel, Guilherme [10] presents a highly accurate and scalable new two-stage approach to learning a sequence classifier that is easily used in data mining applications. This approach integrates the conditional random fields (CRFs) and the support vector machines (SVMs). The Maximum-margin nature of SVMs gives benefit to the approach and it is highly accurate. The training set is same for every SVM, CRF input has fewer features, and so it's scalable. As it joins the existing published software in an unequivocal way it's easy to use.

Xin Xin et al [12] proposed a unified approach, Constrained Hierarchical Conditional Random Fields, to accomplish the three labeling tasks simultaneously. This approach described the composite structural dependencies. The inference process avoids the logical errors based on the Viterbi algorithm. The prototype model of use- oriented semantic academic conference calendar is developed. The calendar is built automatically for the user by finding, extracting and by updating the semantic information from the web. Other applications like finding experts, finding sponsors are done with semantic conference data.

Milos Radovanović et al [14] investigated a novel facet of the dimensionality curse, referred as hubness and demonstrated by the Inclination of some documents. The origins of hubness are a consequence of high (intrinsic) dimensionality of data which is not based on skewness and sparsity and of term frequency distribution. Hubness process is the behavior of similarity measures in vector spaces with high dimensionality. The deliberation begins with the classical VSM, but more advanced variations like Okapi BM25 are generalized as conclusions.

## **3. CONCLUSION**

This paper presents the broad view on various research articles relating to the various methodologies of information retrieval in IoT. It explains various concepts relating to the query

based information retrieval. We discussed about the various models and techniques and believe that the work will help researchers to obtain the overview of Information Retrieval (IR).

## References

- [1] Ricardo Baeza- Yates, Berthier Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, 1999. ISBN 0-201-39829-X
- [2] C. Lee Giles, Kurt D. Bollacker, Steve Lawrence. CiteSeer: An Automatic Citation Indexing System. *DL '98: Proceedings of the third ACM conference on Digital libraries*. May 1998 Pages 89–98. <https://doi.org/10.1145/276675.276685>
- [3] Steve Lawrence, Kurt Bollacker, C. Lee Giles. Indexing and Retrieval of Scientific Literature. *CIKM '99: Proceedings of the eighth international conference on Information and knowledge management*, November 1999 Pages 139–146. <https://doi.org/10.1145/319950.319970>
- [4] Michael J. Covington, Wende Long, Srividhya Srinivasan, Anind Dey, Mustaque Ahamad, Gregory Abowd, *SACMAT '01: Proceedings of the sixth ACM symposium on Access control models and technologies*, May 2001 Pages 10–20. <https://doi.org/10.1145/373256.373258>
- [5] Chiu C. Tan, Bo Sheng, Haodong Wang, and Qun Li. Microsearch: A Search Engine for Embedded Devices Used in Pervasive Computing. *ACM Transactions on Embedded Computing Systems* Volume 9 Issue 4 March 2010 Article No.: 43 pp 1–29. <https://doi.org/10.1145/1721695.1721709>
- [6] Jeremy Stribling. OverCite: A Cooperative Digital Research Library. In *Proceedings of the 3rd IPTPS*, February 2005.
- [7] Huajing Li, Isaac G. Councill, Levent Bolelli, Ding Zhou, Yang Song, Wang-Chien Lee, Anand Sivasubramaniam, C. Lee Giles. CiteSeerX - A Scalable Autonomous Scientific Digital Library, *InfoScale '06: Proceedings of the 1st international conference on Scalable information systems*, May 2006 Pages 18–es. <https://doi.org/10.1145/1146847.1146865>
- [8] Robert Kahn, Robert Wilensky. A framework for distributed digital object services. *International Journal on Digital Libraries*, Volume 6, Issue 2, April 2006 pp 115–123, <https://doi.org/10.1007/s00799-005-0128-x>
- [9] X. Ning, H. Jin and H. Wu, SemreX: Towards Large-Scale Literature Information Retrieval and Browsing with Semantic Association, *2006 IEEE International Conference on e-Business Engineering (ICEBE'06)*, 2006, pp. 602-609, doi: 10.1109/ICEBE.2006.87
- [10] Hoefel Guilherme, Charles Elkan. Learning a two-stage SVM/CRF sequence classifier. *CIKM '08: Proceedings of the 17th ACM conference on Information and knowledge management*, October 2008 Pages 271–278. <https://doi.org/10.1145/1458082.1458120>
- [11] Mathias Lux, Savvas A. Chatzichristofis. LIRe: Lucene Image Retrieval - An Extensible Java CBIR Library. *MM '08: Proceedings of the 16th ACM international*

- conference on Multimedia, October 2008 Pages 1085–1088.  
<https://doi.org/10.1145/1459359.1459577>
- [12] XinX in, Juanzi Li, Jie Tang, Qiong Luo. Academic Conference Homepage Understanding Using Constrained Hierarchical Conditional Random Fields. CIKM '08: Proceedings of the 17th ACM conference on Information and knowledge management, October 2008 Pages 1301–1310, <https://doi.org/10.1145/1458082.1458254>
- [13] H. Wang, C. C. Tan and Q. Li, Snoogle: A Search Engine for Pervasive Environments, in *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 8, pp. 1188-1202, Aug. 2010, doi: 10.1109/TPDS.2009.145
- [14] Milos Radovanović, Alexandros Nanopoulos, Mirjana Ivanović, On the Existence of Obstinate Results In Vector Space Models. SIGIR '10: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, July 2010 Pages 186–193. <https://doi.org/10.1145/1835449.1835482>
- [15] M. Giannikos, K. Kokoli, N. Fotiou, G. F. Marias and G. C. Polyzos, Towards secure and context-aware information lookup for the Internet of Things, *2013 International Conference on Computing, Networking and Communications (ICNC)*, 2013, pp. 632-636, doi: 10.1109/ICCNC.2013.6504160
- [16] Norasykin Mohd Zaid, Sim Kim Lau, Mohd Nihra Haruzuan Mohamad Said, Ontology-based Search System Using Hierarchical Structure Design. *Procedia - Social and Behavioral Sciences*, Volume 97, 2013, Pages 566-570, <https://doi.org/10.1016/j.sbspro.2013.10.274>
- [17] Perera, Charith, A. Zaslavsky, P. Christen and Dimitrios Georgakopoulos. Context Aware Computing for The Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials* 16 (2014) 414-454
- [18] Feng Zhao, Zheng Sun, Hai Jin, Topic-centric and semantic-aware retrieval system for internet of things. *Information Fusion*, Volume 23, 2015, Pages 33-42, <https://doi.org/10.1016/j.inffus.2014.01.001>