A Survey on Context based Information Retrieval in Cloud

Vasudeva R[1], Dr. S. N Chandrashekara[2]

[1, 2] CSE, C. Byregowda Institute of Technology

Kolar, Karnataka, India

Abstract:- **Cloud computing is a promising information technique (IT) that can organize a large amount of IT resources in an efficient and flexible manner. Driven by the revolution of information technology in recent years and as large amounts of data are outsourced to cloud storage servers. In today's era information retrieval becomes more complex as the mediums are available to store huge amounts of information as on computer's storages. Effective information retrieval has been the need for the day, with the ever increasing demand for faster and more accurate information based on a dynamic user's context, when using applications in these settings; a user has much to gain by the effective use of implicitly sensed context. In this survey we emphasize on the importance and implications of context in information Retrieval and how it can affect the retrieval system to operate more effectively.**

Keywords: Cloud computing, information retrieval (IR), User Context, cloud search.

1. INTRODUCTION

Cloud computing is witnessing rapid innovations in the recent years. It has two main tasks storing and accessing data and programs by means of Internet rather than usage of a computer's hard drive. The entity cloud presents an extensive range of services. It reduces the complexity of the networks, makes provision for customization, scalability, efficiency etc. Besides, the information stored on cloud is generally not easily lost. Because of its on-demand nature, could typically buy cloud computing the same way you would buy electricity, telephone services, or Internet access from a utility company. It is so easy with the cloud because one can add extra services (or take them away) at a moment's notice as the business needs change [3].

Humans are quite successful at conveying ideas to each other and reacting appropriately.

This is due to many factors: the richness of the language they share, the common understanding of how the world works, and an implicit understanding of everyday situations. When humans talk with humans, they are able to use implicit situational information, or context, to increase the conversational bandwidth. Unfortunately, this ability to convey ideas does not transfer well to humans interacting with computers. In order to use context effectively [1][7], we must understand both what context is and how it can be used. An understanding of context will enable application designers to choose what context to use in their applications.

The definition [1] [5] "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves".

Context-aware applications running on smart devices need to be aware of the abilities and capabilities of themselves and their nearby devices to perform in more intelligent manner. The browsing behaviour [8][7] of web users can change from one day to another and from one time period to another depending on several parameters grouped under the notion of context which may include the time, the location, the current task, the event, etc. In [1][8] specify that "The user context can be assimilated to all factors that can describe his intentions and perceptions of his surroundings".

Information retrieval (IR) [10][3] is the activity of obtaining information system resources that are relevant to an information need from a collection of those resources.

2. RELATED WORK.

The use of context is important in interactive applications. It is particularly important for applications where the user's context is changing rapidly, such as in both handheld and ubiquitous computing. In order to better understand how can use context and facilitate the building of context-aware applications [1]. In handheld and ubiquitous computing, a user's context is very dynamic. When using applications in these settings, a user has much to gain by the effective use of implicitly sensed context.

This paper [14] proposes the use of local context as a way to semantic information retrieval. In model, rather than trying to formalize the contents of the documents among which the search is done, try to automatically build a representation of the context in which the search is done. Consider that search is always done as part of an activity, and that the search context is determined by the activity that is carried out at a particular moment. Many activities that a person is engaged in are carried out with the help of a computer, and leave a digital trace in the files that are created in connection to it; take these files, and the structural relations between the directories in which they are stored, as a starting point for the representation of context[2].

Involving context in any text retrieval process requires the addition of semantic knowledge into the application. Knowledge incorporated applications tend to be process intensive. Further, this process needs to be applied on huge

data hence our application uses Hadoop for data storage and Spark to perform in memory processing to provide faster analysis.

Context-awareness enables applications to provide end-users with a richer experience by enhancing their interactions with contextual information. Several frameworks have already been proposed to simplify the development of context-aware applications. These frameworks are focused on provisioning context data and on providing common semantics, definitions and representations of these context data. Assume that applications share the same semantic, which limits the range of use cases where a framework can be used, as that assumption induces a strong coupling between context management and application logic.

Context Awareness Mobile devices have more features than their computer counter parts, including location information, built-in cameras, and certain social networks.[4] can use information from these additional features to create new research areas that aren't just incremental enhancements of existing studies. IR and data mining of the new information will require novel technologies specifically developed to process information such as time, location, and semantics.
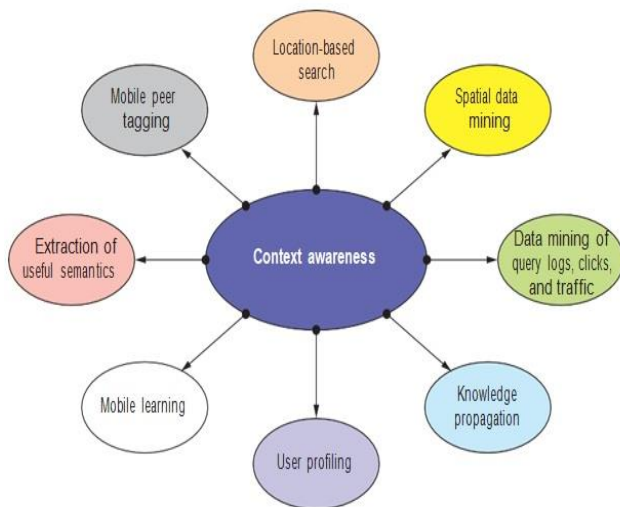

Fig 1: Context awareness [5].

Figure 1 shows some research topics in this area. Specifically, in context-aware information processing, the additional information of mobile users' locations creates vast opportunities for corporations, opening potential channels of communication, sales, and marketing, and creating potential revenue drivers. Mining spatial data generated [5].

In Figure 2, shows context, Mobile devices have smaller screens and less processing power than most computers. Thus, to effectively and efficiently progress in this area, search the research archives to analyze technologies optimized for small screens and low-power and low memory devices. However, also move forward to personalize and adapt these technologies to mobile information. Efficient ways to process large amounts of

data and personalization of small screens need both efficient and scalable algorithms that can perform well in power-limited settings.
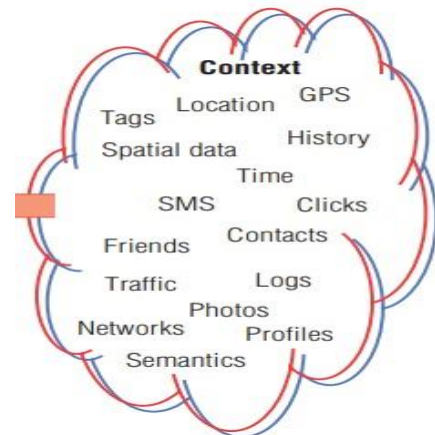

Figure 2: Context [16].

Mobile devices become widely accepted computing paradigms; but the mobile services need to be aware of the dynamical user environment and adapt accordingly to the context. With the increasing amount of multimedia, ontology can add value to the new semantic multimedia services, by considering the contextual information [16].

2.1: Contextual Information Retrieval Systems.
Most classic information retrieval systems "IRS" use statistical methods based on the frequency of occurrences to retrieve relevant documents or either use linguistic or semantic methods. Consequently, it is necessary for IR'S to face challenges associated with the individual differences that exist between users, their tasks and goals. This fact involves a crucial problem nowadays as the information is authored and exploited by millions of different user's profiles, each one with different knowledge, backgrounds, and preferences leading to different search contexts. Thus, to overcome these challenges, the retrieval of information should depend on many factors like: history of interaction, task in hand, time, place and a variety of other factors or variables that are not explicit but implicit in the interaction and surrounding environment, namely the context. These factors can be graded as dimensions, taxonomies or categories of context.

CONTEXT CATEGORIES OR TAXONOMIES.

In Fig. 3, [15] introduce the context of the information access device in addition to the document context, the Spatio-temporal context, the user context and the task context. This taxonomy [8] covers all the categories in a more organized and detailed way. Context modelling has been long acknowledged as a key aspect in information retrieval. The aim of this task is to address two main problems: how to model the user's context, and how to exploit it in the retrieval process in order to provide the user with information that is of most value in his context. Effectively, in contextual IRS, user's behaviour can be predictable grounded on his past interactions with the system
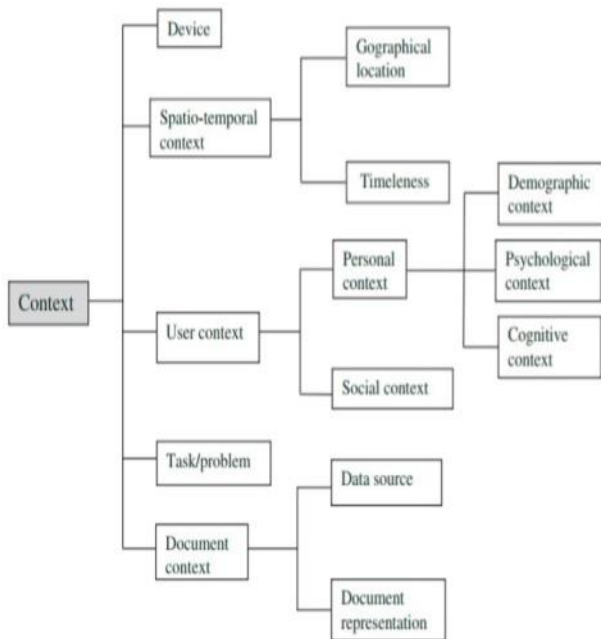
Fig 3: Concept The multi-dimensional concept of context in an IR system [15].

The abundant development of hardware and software, offers opportunities to explore new contextual dimensions, to this fact and from the taxonomies that learned. The author Djalila Boughareb[1] [8] here proposed the taxonomy of context, in which introduce all dimensions introduced. Each context meaning clearly defined and distinguished based on different situation for effective information retrieval.

Different Retrieval techniques
Cloud computing is a collection of computing resources used for storing or accessing data from any distant place [3]. Fig 4 shows search schemes, Keyword search procedures are extensively used and the user is allowed to retrieve chosen data from the storage space.
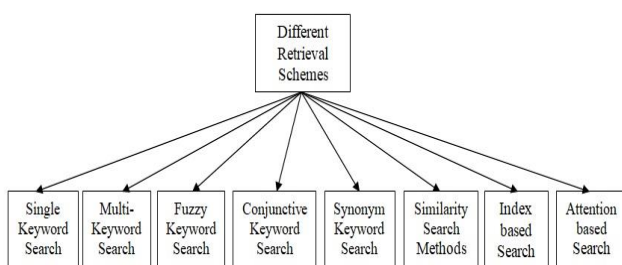

Fig 4: Different Retrieval Schemes [3].

In [7] keywords are indexed along with their context which is extracted from the documents. Here Binary Search Tree is used to store the keywords for improving the performance of the retrieval system in terms of accuracy and efficiency for retrieving more, relevant documents as per the requirements of the user as context of the keywords are also provided. Thus using index for searching purposes provides fast access to document context structure along with an optional searching.

In [8] the contextual information retrieval field have proven that the state when the user conducts a search has a perceptible effect on the user's search behaviour. The search context may include several dimensions such as time, location, history of interaction, current task, etc. In the information retrieval field, it has taken a very important part of research aiming to improve the relevance of the search results.

In [9] Involving context in any text retrieval process requires the addition of semantic knowledge into the application Knowledge incorporated applications tend to be process intensive. Further, this process needs to be applied on huge data hence our application uses Hadoop for data storage and Spark to perform in memory processing to provide faster analysis. The process of context based information retrieval is carried out in three major phases; TF-IDF based word ranking, context based feature vector creation and classifier training. Context Based Information Retrieval in Big Data [9] is a Sequence of process to be carried out in performing effective information retrieval process. The advantage of this method is that it can analyze the entire data corpus in fast and effective manner to perform efficient retrieval of information.

In [11] the semantic-social information plays a central role in discovery, dynamic selection, dynamic composition and substitution of services. The experiment shows that our approach outperforms adaptation ratio and time response.

It is well accepted that a good IR system [10][2] should retrieve as many relevant documents as possible (i.e., have a high recall), and it should retrieve very few non-relevant documents (i.e., have high precision). Unfortunately, these two goals have proven to be quite contradictory over the years. Cluster hypothesis states that documents that cluster together (are very similar to each other) will have a similar relevance profile for a given query. Natural Language Processing (NLP) has also been proposed as a tool to enhance retrieval effectiveness, but has had very limited success.

The Alexandra Dumitrescu, Simone Santini [13] has presented the outline of a model of meaning not based on annotations, one in which the reader's context plays a preponderant role. on the one hand, to build a context based data access client (configured as a plug-in to some word processing or presentation program, if possible) to make context based retrieval on general websites and repositories and, on the other hand, to build a context-based access server. The server that we consider here will be integrated with the user's computer from which it will derive the current context, and with which it will cooperate to support interaction, the relations between geometry and fuzzy logic.

2.2: Ontology based information retrieval approaches and techniques.

The Kamran Munir, M. Sheraz Anjum, al et, [12] reviews the state of the art in ontology-based database information retrieval. Here, a historical overview of information retrieval

approaches is first presented, followed by a detailed analysis of existing ontology-based query systems and data search strategies in relation to three different key aspects that guided the review of such work. These three aspects are: (1) ontology assisted visual or interactive query formulation; (2) ontology based information linking approaches (also known as keyword search); and (3) ontology based query refinement (including query enrichment).

- *Information retrieval from a historical perspective.*

Database information retrieval is the search for information in databases. The need for effective methods to automate Information Retrieval has grown in importance because of the significant increase in the amount of both structured and unstructured information embodied in information sources. An information retrieval (IR) system locates information that is relevant to a user's query. This brief review of past work focuses on the algorithms that take a user's query and retrieve a set of relevant documents [12].

- *Ontology-based query formulation approaches.*

Ontology-based Visual or Interactive query formulation systems are query systems for databases that use visual representations to express related data requests. These systems adapt ontologies for database query generation in order to improve the effectiveness of the human-computer communication.

- *Ontology-based information linking approaches.*

In this approach, data access is enabled by defining links between ontology concepts and relational data. This ontology-to-database mapping mechanism enables a designer to link a data source to an OWL-Lite ontology. Queries are formulated by consulting ontology-to-database mapping rules, but this rule derivation process is carried out manually by ontology and database experts.

- *Ontology-based query formulation approaches.*

This method makes query refinement process aim at enabling end users to make improved formulated query and it attempts to improve information retrieval by replacing or adding extra terms into the initial query. Most of the methods will use query refinement process includes both query rewriting and expansion operations. This approach has ambiguity in relation with both structural and semantic ambiguities.

3. DISCUSSIONS AND SUGGESTIONS.

According to the concepts, contextual information retrieval systems are based on concept of user's context and users modelling approaches, to provide the most useful information need to the users along with considering their context. To achieve this many authors proposed schemes [15][3][7][6]. A major challenge in these techniques is the modularization of user's context, where it difficult to identify and use the knowledge about user context. The

context is extremely dynamic in nature and can change constantly for the same user or during the same search process. Since contextual information retrieval systems based on user related personal information, another issue is the user's context authentication and privacy should be preserved, but most of the users will not prefer to share the private information.

Thus, authors [15] suggest some context model's requirements that address some of discussed issues and others. In general,
- first, in order to improve retrieval results, using the long-term user interest profiles can be efficient, to cope with this, a context-modelling framework has to identify all relevant and useful contextual categories;
- The knowledge about the actual context of a user should be used for 2 aims: first to enhance any information created, modified, or published by the user, and second to offer information created, modified, or published in contexts similar to the current user's context.
- After the system allow users to explicitly provide context information, at the same time a context should basically be recognised automatically;
- The context-modelling framework should rank dynamically any important contextual dimensions used to well performing the similarity match;
- The notion of time is very important on term of recognising the current context and on retrieving information relevant to this context;
- Dynamic behaviour of the users for selecting context to retrieve the information.
- To address the problem of model of context availability and more generally, the availability of more structured context models is to define user-dependent aggregation schemes defined as linear combinations where weights of relevance context dimensions are automatically computed based on the user-specified priority order over the dimensions.

For the issue of privacy , the client-side applications can be useful where some architecture are used to enforces user privacy on social networking sites by shielding a user's personal information from the site and from other users that were not explicitly authorized by the user[14].

4. CONCLUSIONS.

Context becomes a common notion in information retrieval and also it is most challenge in the field. This paper presented the concept of context, the Varity of its categories, context awareness model and different information retrieval techniques that can be used for different context, and then showed some of the ontology based information retrieval schemes. Finally explained some of the challenges and issues really need to observe for context based information retrieval.

5. REFERENCES

[1] Anind K. Dey and Gregory D. Abowd, "Towards a Better Understanding of Context and Context-Awareness", Graphics, Visualization and Usability Center and College of Computing.

[2] Jochen L. Leidner, "Information Retrieval in the Cloud", .

[3] Raghavendra S, Chitra S Reddy, Geeta C M, Rajkumar Buyya, Venugopal K R, S S Iyengar, L M Patnaik "Survey on Data Storage and Retrieval Techniques over Encrypted Cloud Data", International Journal of Computer Science and Information Security (IJCSIS), Vol. 14, No. 9, September 2016.

[4] Kamal Benzekkia,∗, Abdeslam El Fergouguia, Abdelbaki ElBelrhiti ElAlaouia, "A Context-Aware Authentication System for Mobile Cloud Computing", The First International Conference On Intelligent Computing in Data Sciences, Elsevier, 127 (2018) 379–387,

[5] Manouchehr Zadahmada,*, Parisa YousefzadehFarda, Morteza Abbaszadeha," A data model to support context-aware mobile cloud collaboration scenarios", World Conference on Technology, Innovation and Entrepreneurship, Elsevier , 195 (2015) 1602 – 1608.

[6] Ranya Alawadhia, Tahani Hussainb1, "A Method Toward Privacy Protection in Context-Aware Environment", The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40) April 29 – May 2, 2019

[7] Deepika Sharma, Kirti Choudhary, Sandeep Kumar Poonia, "Design And Implementation Of Context Based Information Retrieval System", International Journal of Engineering, Management & Sciences (IJEMS) ISSN-2348 –3733, Volume-1, Issue-6, June 2014.

[8] Djalila Boughareb, Nadir Farah, "Context in Information Retrieval", 978-1-4799-6773-5/2014©2014 IEEE.

[9] N. Karthikeyan, Dr. R. Dhanapal, "Fast and Enhanced Context Based Information Retrieval in Big Data using Hadoop and Spark", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 19 (2015) pp 40699-40702.

[10] Amit Singhal , "Modern Information Retrieval: A Brief Overview", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2001.

[11] Alti Adela*, Sébastien Laborieb, Philippe Rooseb, "Semantic Context-aware Adaptation Platform Architecture", The 4th International Symposium on Frontiers in Ambient and Mobile Systems (FAMS-2014),Elsevier, (2014) 959 – 964.

[12] Introduction to InformationRetrieval" IEEE Computer Society, 1541-1672, 2010.

[13] Alexandra Dumitrescu Simone Santini," Think locally, search globally; context based information retrieval", Conference Paper · September 2009, Reasearhgate.

[14] MalikaYaicia, AnisOussayahb, Mohamed Ahmed Takerrabetb, "Trust-based Context-aware Authentication System for Ubiquitous Systems", 15th international conference on mobile computing and pervasive computing, Elsevier, (2018) 35–42.

[15] Zakariae ALAMI MERROUNI, Bouchra FRIK1, Brahim OUHBI, "Toward Contextual information retrieval: A Review and Trends", Second international conference on intelligent computing in Data Sciences , Elsevier, (2019) 191–200.

[16] Dejan Kovachev, Ralf Klamma," Context-aware Mobile Multimedia Services in the Cloud", Proceedings of the 10th International Workshop, 2009.