Smart Cloud environment through Semantic Retrieval

M.Thangaraj, G.Sujatha

Abstract— **The WWW is a huge wide distributed global in- formation centre for many information services. The availability of web resources are enormous that whatever a user needs at a given moment can be found. These resources are not only for data items, but functionality delivered through some sort of service architecture. Cloud computing has emerged as one of the foremost computing models for providing services over the internet. As the range of available cloud services increases, the matter of service discovery and selection arises. To meet out the user's needs, the selection process is integrated with semantic technologies to facilitate the discovery of cloud resources. The architecture proposed here integrates semantic annotation and indexed techniques to cloud services with the ontology evolution which will produce semantically enhanced cloud services.**

Index Terms— Cloud computing, Semantic Web, Semantic Annota- tion, Ontology, Semantic Retrieval

I. INTRODUCTION

The future internet will be based on services, evolving from a mere repository of knowledge to a new platform for information interchange and business transactions. Large organizations are more and more exposing their business process through web services technology for the large scale development of software as well as for the sharing of their services within and outside the organization. New technologies such as Software-as-a-Service (SaaS), Platform- as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) and the cloud computing models promise to create new levels of efficiency through large-scale sharing of functionality and computing resources [3]. SaaS is the model representing the actual application that end users leverage to accomplish this objectives. PaaS is the model representing frameworks and common functions on which application can be built. IaaS is the services that provide processing, storage, networks and other fundamental computing resources. By which consumers can deploy arbitrary software. The term cloud service referring to services in any of three categories. In the current socio-economic era, more and more cloud services become available. The exponential increase in the number of services available makes it harder for users to find those cloud services that meet their requirements. And hence one of the main concerns in the cloud computing community is that of service search and retrieval. [2],[11].

Semantic web (SW) paradigm improves the search processes by providing semantic descriptions for clouds. The semantic web " targets to build an extension of the current web in which information is given well-defined meaning better enabling computers and people to work in co-operation". The semantic web which has been a big buzz word for the past years, promises to bring semantic information in the form of standardized meta data. Ontologies are the backbone technology for SW. An ontology can be defined as a "formal and explicit specification of a shared conceptualization". Ontologies provide knowledge using five kinds of components: Classes, Relations, Attributes, Axioms, and Instances. The use of ontologies and semantic technologies can overcome the limitations of traditional search. The various semantic search models are analysed in the survey [7]. The framework proposed [8] retrieves information efficiently by using a semantic querying and retrieval with an improved ranking algorithm.

In the cloud environment, finding the appropriate service manually is time consuming, the addition of semantics to clouds will improve the search process. Standard information extraction and language processing techniques are used to extract ontology-driven meta data files and annotate clouds. This proposed system will assist the user in discovering the appropriate services with the semantic information available.

This motivates an ontology-based framework which enables the semantic annotation of the services available in the cloud. The annotated cloud services are then exploited for the semantic search. The main aim is to improve the search and retrieval of cloud services relevant to the user needs by a back-end which will create semantic repository and the semantic module comprised of Semantic Annotator,Semantic Indexer,Semantic Query Converter,Semantic searcher and ranker.The main contribution is integrating cloud services and Semantic web by providing a semantic

Manuscript received Aug 22, 2014

M.Thangaraj, Madurai Kamaraj University, Madurai 625 021,India

G.Sujatha, Sri Meenakshi Govt. Arts College For Women(A), Madurai 625 021,India

platform to retrieve most relevant cloud services.Structure of the Paper The rest of the paper is organized as follows. Some related work is analysed in Section II. The components that take part in the framework and their overall architecture are described in section III. In section IV the evolution metric of this system is discussed. Finally conclusion and future work are put forward in section V.

II. RELATED WORK

Cloud computing is an emerging technology that permits to offer computing services over the Internet [12]. In this socio-economic climate, the affordability of cloud computing has gained popularity among current innovations. In these circumstances, more and more cloud services become available. Consequently, it is difficult for service consumers to find and access those cloud services that full-fill their requirements.

Semantic approaches have proven to be very effective in improving search processes [10]. It is time consuming to find the appropriate services manually. The addition of semantics to the cloud landscape can help automate the search process [6]. This approach requires the generation of semantic descriptions for all the available cloud services since the manual process is time consuming [1].

By adding the semantic annotations to the cloud descriptions and assist the user in discovering the most relevant services. Providing suggestions from the ontology is the main concept of this proposed work.Although several systems for ontology- based annotation have been proposed, there is not a standard approach for semantic annotation. [9]. A cloud-focused search tool [4] , [5] that makes use of such semantic descriptions to get accurate results only from the keyword-based searches.

III. PROPOSED SYSTEM

The architecture of the proposed framework is shown in Fig. 1. This approach is based on the two main components namely Back-end and Semantic module. The system works as follows. Initially the description of the services in the cloud are semantically represented and annotated. The second stage is from these annotations a semantic index is created using a modified vector space model. The next stage is semantic searching to retrieve the matching services which is then arranged in the order of the relevancy in the final stage. All these functions are carried out by the Semantic Module.

A. Backend

This will receive both domain ontologies and a natural language description of cloud services as input. The ontology is evaluated using the information available in wikipedia and other domain related concepts. The main idea here is to keep the domain ontologies up-to-date. The terms extracted from clouds is combined with their concepts or synonyms to enrich the ontology knowledge. This back-end supports both the query and the cloud services.

B. Semantic Module

The functionality of the Semantic module includes expansion of Query by providing ontology suggestions, Annotate and Index the cloud service descriptions and providing services to the user arranged in the order of their relevancy.

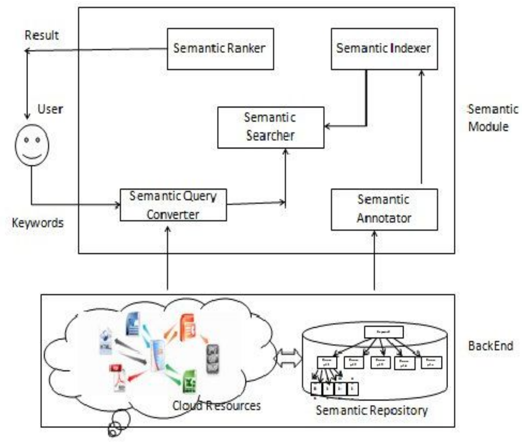

Fig. 1. Architecture of the System

1) Semantic Annotator : It obtains a semantic annotation for the cloud service descriptions in accordance with the domain ontologies and the external thesaurus. The cloud service descriptions are annotated with the classes and individuals of the domain ontologies. The outcome of the semantic annotation module is a list of semantic annotations defined in terms of the ontology. The classes and individuals in the annotations represent terms that have been extracted from the cloud service descriptions.

2) Semantic Indexer : Each annotated document is stored in a database with weight assigned which reflects the importance of the terms and concepts in that particular document. Weights are calculated by using a modified tf-idf algorithm.The term frequency (tf) is the local weighting factor shows the importance of the term within a particular cloud service descriptions and the cloud frequency (cf) is the global factor shows the importance of the term within the entire collection of cloud services.The inverse cloud frequency (icf) relates cloud frequency to the total number of cloud services(N) in the collection.

In this scenario, the cloud services descriptions are the documents to be analysed. For each description, an index is calculated based on the modified vector space model [8]. The outcome of semantic indexing module is a list of semantic concepts and its relations with assigned weights.

3) Semantic Query Converter : The query presented by the user in search of cloud services is merely keywords. This proposed framework provides suggestions to the users with the help of the ontology or thesaurus. By using these, the query is expanded and is converted into semantic query.

4) Semantic Searcher : This module is responsible for finding services in the cloud using the semantic query. The matching cloud services are indexed in the implemented index structure in order to retrieve the interesting cloud services of all the the query words. Then the semantic search component calculates a similarity value between the query q and each services. To find the similarity value the cosine similarity (Eqn.1)is used. The matched cloud services are ranked based on their relevancy value.

SC = cosθ =
q(w + c).s(w,c) |q(w + c)||s(w,c)| (1)

5) Semantic Ranker : The retrieved cloud services is ranked with the relevancy of the user constructed semantic query. During this comparison, the topics of the initial query are taken into account. When searching for relevant services, one should consider what the user is intended to find in the first place. In all the comparisons the weights of topics are taken into account so as to determine the relevance between the pages.
This relevance is measured with the weights calculated by the improved dynamic ranking algorithm. The weight is a function of term frequency and the tag based frequency. The term frequency is the local weighting factor which reflects the importance of the term within a particular document. The global weighting factor considers the importance of a term within the entire collection of cloud services knows as cloud service frequency(cf). The inverse cloud service frequency (icf) which relates the service frequency to the total number of cloud services in the collection (N) is computed. Being a semantic rank algorithm these values are calculated for the keyword as well as for the concept of the keywords.(Eqn.2

wt = tf(w)+log |N| cf(w)
+tag(w)+tf(c)+tag(c)+log |N| cf(c) (2)

Algorithm : Semantic Annotator and Semantic Indexer
Input : Cloud Domain Ontology,Cloud Services
Output : Annotated Cloud Services
Parameters : CS-Cloud Services, w- terms of a cloud service
tf - term frequency, cf - cloud frequency, c-concepts
icf - inverse cloud frequency
CS=$CS_1, CS_2, ..., Cs_n$
for each i cloud service descriptions
for each term w in CS_i
for each concept in Ontology
if $CS_i(t)$ matched with the concept
$CS_i(t) = CS_i(t) U getconcept(c_i)$
count tf(w) , tf(c), tagf(w), tagf(c)
for each cloud service
{
if present(t(w)) then cf(w)=cf(w)+1
if present (t(c)) then cf(c)=cf(c)+1
}
$icf(w) = \log \frac{|N|}{cf(w)}$
$\mathrm{icf(w)} = \log_{cf(w)}^{|N|}$

Algorithm : Semantic Query Converter and Semantic Search
Input : Raw Query Terms Q($qw_1, qw_2, ...qw_n$)
Output : Ranked cloud services matches the user's need
Parameters : CS-Cloud Services, w- terms of a cloud service
Q = Q - Special characters - S
Q = Porterstemmer:Stem(Q)
Q = Q U Ontology Concepts
Q = Q U Web Suggestions
Q = Q U thesauruses : Synonyms(Q)
Q = Q U extractlinks(URL)

for each word in the query
$h(i).add = QW_i$

for each Cloud service descriptions j
{
if qw_i matches with Cloud description
$h(j).add = h(j) U CloudService$
}

Calculate the term frequency of the query words
result = h(1,2)
for each word i in the hashtable
{
$result = result \wedge h(i, 2)$
}
calculate score1 as $tf(words) + tf(concepts) \square icf \square icf$
calculate score2 as tagf(words)+tagf(concepts)
if score1 > threshold and score2 > threshold Present the results in the

IV. PERFORMANCE EVOLUTION

Once the semantic indexes have been created, the experiment starts. This experimental evolution aims at analysing whether the semantic search engine module of the proposed platform is useful. Eleven Topic-based queries are issued. For each query a set of cloud services were manually selected. At the same time, the semantic search system was asked to perform the same task in an automatic way. The system's results were then compared to those produced by the manual selection.
The evaluation has been done by the calculation of the parameters Precision,Recall and F-Measure. The precision score is obtained by dividing the number of services suggested automatically by the system that had also been selected in the manual

process, and the total number of series suggested automatically by the system. The recall score is obtained by dividing the number of services suggested automatically by the system that have been also selected in the manual process, by the total amount of services selected manually. The F-Measure is the weighted harmonic mean of the precision and recall scores. The results of the experiments are plotted in graphs. Fig.2 compares the precision values of the eleven topic based queries for smart cloud and the manual cloud retrieval and in Fig.3 compares the recall values.

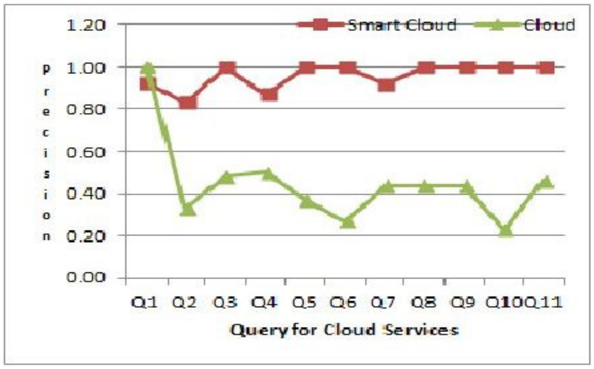

Fig. 2. Result of Query Matching - Precision

CONCLUSION

Cloud computing technology provides computing services over the Internet. It is closely related to distributed computing and grid computing. The affordability of cloud computing has increased its popularity among today's innovations in current socio-economic world. The growth of cloud computing industry led to an exponential rise in the number of services being delivered from the cloud. And hence the time and effort required to find the service manually to meet the needs of an individual or an organization is substantially increased. In order to find the right service(s) with the appropriate functionalities the semantic retrieval of cloud services has

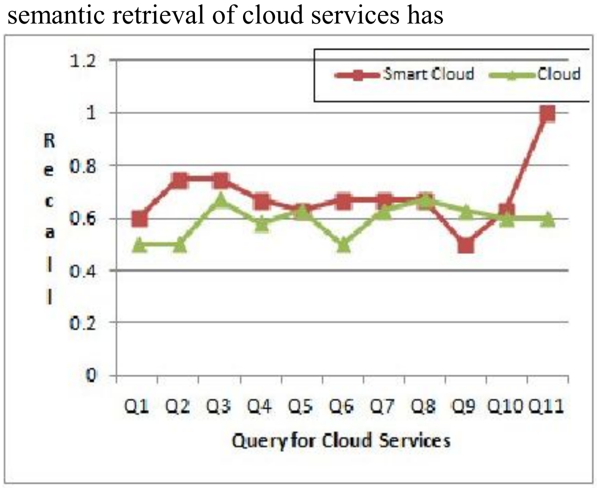

Fig. 3. Result of Query Matching - Recall

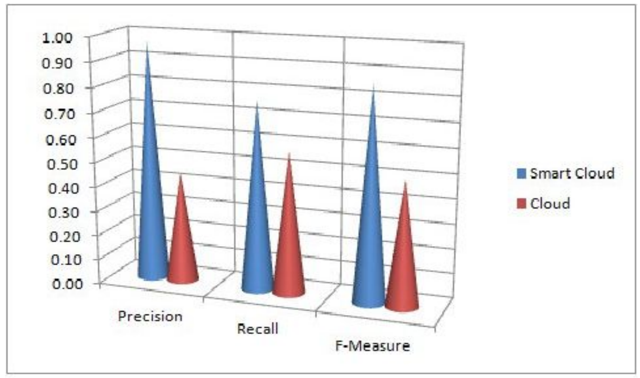

Fig. 4. Precision, Recall, F-Measure

been proposed in this paper. The framework proposed here automatically annotates the description of the cloud services using a domain ontology. This framework has been implemented taking into account a multi ontology environment to support several domains.The ontologies are created dynamically and hence they are up-to- date to changing requirements. The main advantages of this framework over other related work are generation of semantic annotations from unstructured documents, multiple ontologies, dynamic ontology creation, automatic annotation. The results of this experiment when compared with the state-of-art tools with a satisfactory outcome, are very promising.

This system can be extended in future by introducing personalization concept for cloud services retrieval, updated as a recommender cloud system and tried to implement semantic retrieval for multi language support.

REFERENCES

[1]cravegna, Fabio et al User-system cooperation in document anno- tation based on information extraction, In Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management,EKAW02, Springer Verlag.2002

[2]J. Kang, K. Mong Sim, Cloudle: a multi-criteria cloud service search engine, in:Proc. Services Computing Conference, APSCC, 2010, pp. 339346.

[3]P. Mell, T. Grance, The NIST definition of cloud computing, NIST Special Publications (800 Series), SP 800-145, 2011. Available at: http://csrc.nist.gov/ publications/nistpubs/800-145/SP800-145.pdf.

[4]Miguel ngel Rodrguez-Garca, Rafael Valencia-Garca, Francisco Garca- Snchez, J. Javier Samper-Zapater :Semantic annotation and retrieval of services in the Cloud", Springer Verlag Berlin Heidelberg, 2011.

[5]Miguel ngel Rodrguez-Garca, Rafael Valencia-Garca, Francisco Garca- Snchez, J. Javier Samper-Zapater :"Creating semantically-enhanced cloud services environment through ontology evolution", Future Gen- eration Computer Systems 32 (2014) 295-306, Elsevier.

[6]G. Ortegn-Cortzar, J.J. Samper-Zapater, F. Garca-Sanchez, Adding semantics to cloud computing to enhance service discovery and access, in: Proceedings of the 6th Euro American Conference on Telematics and Information Systems, EATIS 2012, Valencia, Spain, 2012.

[7]M.Thangaraj, G.Sujatha, A Study on Searching Mechanisms in Seman- tic Web, in: International Journal of Computer Science & Emerging Technologies (E-ISSN: 2044-6004) 34, Volume 2, Issue 1, February 2011.

[8]M.Thangaraj, G.Sujatha, An architectural design for effective informa- tion retrieval in semantic web, in: Expert Systems with Applications 41(2014) 8225-8233.

[9]V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Mota, F. Ciravegna,Semantic annotation for knowledge management: require- ments and a survey of the state of the art, Journal of Web Semantics 4 (1) (2006) 1428.

[10]Vidoni, Renato et al , An intelligent framework to manage robotic au- tonomous agents, Expert systems with Applications 2011,38(6) :7430- 7439.

[11]M. Wang, J. Zhou, S. Jing, Cloud manufacturing: needs, concept and architecture, in: Proc. Computer Supported Cooperative Work in Design, CSCWD, 2012, pp. 321327.

[12]Zhang, Qi et al ,Cloud Computing : State-of-the-art and research challenges, Journal internet serv Appl 2010,1(1):7-18.

M. Thangaraj received his post-graduate degree in Computer Science from Alagappa University, Karaikudi, M.Tech. degree in Computer Science from Pondicherry University and Ph.D. degree in Computer Science from Madurai Kamaraj University, Madurai, TN, South India in 2006. He is now the ASSOCIATE PROFESSOR of Computer Science Department at Madurai Kamaraj University. He is an active researcher in Web mining, Semantic Web, Information Retrieval and Big-Data Analytics and has published more than 60 papers in Journals and Conference Proceedings. He is a member in the Society of Digital Information and Wireless Communications (SDIWC) and is a senior member of IACSIT. He has served as program chair and program committee member of many international conferences held in India for about one decade.

G.Sujatha received her post-graduate degree in Computer Science from Alagppa Uuniversity, Karaikudi, M.Phil. degree in Computer Science from Mother Teresa University, Kodaikanal. She is now the Associate Professor in Computer Science, Sri Meenakshi Govt. Arts College for Women, Madurai,TN, India. She is currently a Ph.D. scholar in Department of Computer Science, Madurai Kamaraj University, Madurai, TN, India. Her research interests are focused on Semantic Search and the Information Retrieval.