GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

AN ENHANCED VERSION OF ONTOLOGY CLOUD COMPUTING OPEN ARCHITECTURE (OCCOA) WITH QUERY RETRIEVAL AND REFINEMENT MECHANISM

Sachin Gahlawat[*1] and Er. Geetanjali Chawla[2]
[*1,2]CSE department, GGITC Ambala, India

ABSTRACT

We all are living in the world of clouds. User has become adaptive to latest technology trends of information and communication technologies. Cloud computing has made our lives practical and keeps providing us services like consulting, data management and storage in efficient way. Cloud computing is one of emerging areas that is prevailing in industries at grandiose rate. It is combination of Internet and centralized network servers forming a mesh called CLOUD. In this paper, various aspects of cloud computing and its applications are presented. Some differences are provided between network models of cloud computing that are governing organizations and enterprisers. For achieving fault tolerance strategy, there is need to introduce multiple clouds and resource management architecture.

It is believed that existing model must be improved time to time to ensure efficient computing of tasks. The paper presents an improved version of cloud computing detailed architecture. It lists deficiencies between existing cloud information architecture and proposed ontology based architecture. Two additional modules have been introduced in model viz Query Retrieval and Query Refinement. Refinement of queries is done using Rocchio formula that extracts results based on relevance criteria i.e. by distinguishing relevant and non relevant results. They are introduced in order to get efficient indexed results after transforming user query.

Keywords- *Cloud computing, Ontology, Deployment models, Query Retrieval Mechanism and ANEKA .*

1. INTRODUCTION

The concept of cloud computing was originated from telephony networking scheme. Before evolution of cloud computing, there used to be virtual networks that perform work of connecting multiple computers connected through base coaxial cable. This task is performed in 1990's. The use of virtual networks eliminated the hardware circuits between producers and consumers. It is compatible to use with varying network from time to time. But it did not work for long time in technology industry.

With advent of time and modern techniques of technology, it failed to cope with multiple work environments and producing multiple tasks due to decentralized environment. It did not provide services to multiple users at same time and failed to maintain interoperability with servers and clients.

It leads to evolution of cloud computing that works in distributive environments with multiple sources of information. Cloud computing is one of increasing trends in world of technology. It is given as name "dujour" says Gartner's Ben Pring [1]. It is big idea that will revolutionize the change in IT sector services. Although cloud computing effect is ruling minds of technologists, scientists and organizations but it is still occupied by some challenges as listed below:

(i) Data Protection: - Data Security is one of major element that needs to be taken care of. Cloud vendors fear of losing confidential and identity of their consumers. In cloud model, service providers are responsible for maintaining data security and enterprises have to believe them.

(ii) Data Recovery and Availability: - All applications are designed by considering some laws or rules that are called as Service level agreements (SLA's). There are teams designed to support data availability at anytime. These teams perform following tasks:

- Data Replication
- System monitoring
- Maintenance
- Recovery from failure

(iii) Management Capabilities: - Although there are many multiple cloud providers, but management scale is not satisfactory. There is great need to improve on Scalability and balancing features.
The paper is divided into following sections: Section 2 presents concise overview of cloud computing and its network models. Section 3 lists deficiencies in existing cloud computing open architecture and replaces it with proposed model. Section 4 concludes about given paper.

2. ASPECTS OF CLOUD COMPUTING

Cloud computing is combination of various technologies like Grid computing, Virtualization, Autonomic computing, Ubiquitous computing, P2P computing and many more. Time to time updations in existing computer resources at various data centers is one of the factor that led to development of cloud computing.
Use of cloud computing is useful to users because:

- It is inexpensive.
- It is convenient to use.
- Users can access data and use applications with the help of PC and Internet access.
- Software applications need not to be installed on computer; they can be directly accessible through Internet.
- Cloud computing produces applications to market very quickly by using most appropriate resources satisfying user needs.

Cloud computing performs services in ascending order i.e. we can use an acronym "ASC" which stands for Application, Storage and Connectivity.

2.1. Cloud Deployment Models

They are also referred to as cloud computing types- Public Cloud, Private Cloud and Hybrid Cloud. *Public Clouds:* - They provide services to anyone globally with an Internet connection. They are owned by company which has its certain services. E.g. YouTube and yahoo are sites from which user access data in form of emails, attachments, videos from any device that has internet connection.

Private Clouds: - They provide services privately and has private network to limited number of people using that network.

Hybrid Cloud: - It is combination of both public and private cloud models. It has ability to add private cloud with resources of public cloud that helps to manage unexpected workload on cloud.

2.2 Specifications of Hybrid Cloud Model (HCM)

It is developed by Doddle, Morsel and Smith. It has three interfaces- Query Interface of Amazon EC2, SOAP (Simple Object Access Protocol) of EC2 and REST (Representational State Transfer) interface of private cloud.

- Query interface of E2 uses a query string placed in cloud to implement management operations of resource manager. Amazon is cloud provider, provides list of defined parameters and their values to be included in query string. These query strings are sent by resource manager and HTTP GET messages are sent to URL's specified by cloud provider in order to perform management operations. In this way EC2 interface is mapped.
- SOAP is different from query interface in a way that it does not need actual parameters to perform each operation.
- REST assigns uniform resource indicator (URI) to each local resource. Then resource is manipulated via HTTP and mapping is done as usual.

3. DEFINING ARCHITECTURE OF CLOUD

Need of Cloud computing Architecture: -

- It is built in order to cope with large scale complexity data processing.
- Reduces problem of using multiple machines at one time.
- Provides distribution and coordination of complex tasks on various machines.
- User can switch to another machine for accessing resources in case of failure of one machine.

Detailed View of Architecture: - It is described in form of various modules defining their internal functions and components.

(A) **Cloud Value added services**: - It consists of five abstraction layers:

a. *Physical Layer (Hardware as a Service)*:- It is bottom layer consisting of cloud providers, servers, operating systems, devices and switches. Customers of this layer are big industrialists who requires large amount of hardware as service. It performs data processing.
b. *Software Kernel*: - It is second layer and acts as interface between HaaS and Software infrastructure layer. Haas performs data processing whereas S/W infrastructure layer operates the hardware. This layer manages server's hardware resources and performs programs to run in parallel form.

c. *Software Infrastructure*: - It provides network resources to two layers namely: Software environment and Application layer above it. This layer leads to generation of new software environments and applications that will be delivered to end users in form of services.

Services in this layer are as follows:
(i) Computational resources (IaaS) also called Infrastructure as a service. This is available to customers in form of virtual machines (VM). There are many virtualization techniques like para virtualization, hardware assisted virtualization, live migration etc that enables a single server to act as multiple virtual machines.
Examples of clouds that offer IaaS are Amazon elastic compute cloud (EC2), Reservoir architecture

(ii) Data Storage as a Service (DaaS): - It stores data of users on servers located in remote locations. Examples of clouds in DaaS are Amazon S3 (simple storage device), zector zumodrive.

(iii) CaaS (communication as a service):- It provides communication that is reusable, schedulable, configurable and also encrypted. This communication enables CaaS to perform services like network security, real time adjustment to provide better bandwidth and network monitoring.

Examples include VOIP (Voice over internet telephone), E-mail and Video conferencing.

d. *Software environment or PaaS (Platform as a service)*:- Users of this layer include cloud application developers who use applications to implement and distribute their resources via internet. Developers are provided with programming language and set of API's.
Examples of S/W environments are GoogleApp engine and salesforce.com
e. *Application Layer or SaaS (Software as a service)*:- This layer acts as an interface between cloud applications and end users to offer them in demand. In this layer, we don't have to install software on computers as all cloud software is located in providers' data centers.

(B) Cloud IT Infrastructure Management
This module deals with management of hardware and software. Use of Virtual machines reduces the workload pressure on various cloud providers by following virtualization techniques like hardware virtualization, para-virtualization etc.

(C) Cloud Information Model
The aim of this module is to represent cloud resources and information retrieved from large collection of web documents. For performing this task, the existing cloud computing model follows "GrepTheWeb" Architecture.
It allows users to select documents according to their query. Large amount of results and document links are produced. These documents are treated as input and Grep architecture finds documents that match the user's query. It is compared with proposed model in next section of paper.

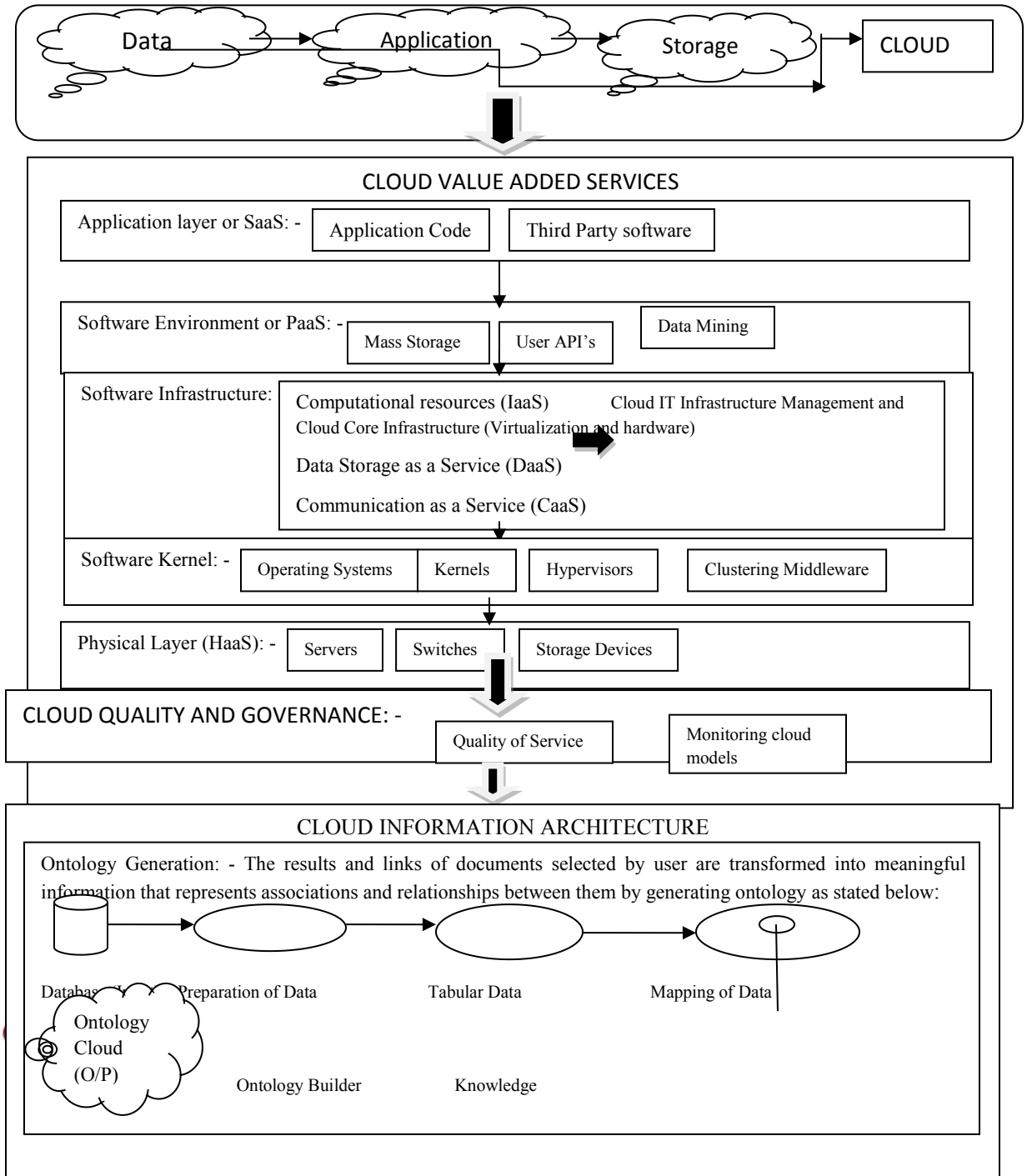
(D) Cloud Quality and Governance
This module deals with factors like Quality of Service (QoS), monitoring cloud models and updating them according to latest technology trends, defines model standards according to Service Level Agreements (SLA's). It deals with authentication and identity management of services accessed through cloud clients (web browsers, applications) to cloud users.

3.1 Multiple Clouds and Resource Management Architecture

The aim of this module is to represent cloud resources and information retrieved from large collection of web documents. For performing this task, the existing cloud computing model follows "GrepTheWeb" Architecture.
It allows users to select documents according to their query. Large amount of results and document links are produced. These documents are treated as input and Grep architecture finds documents that match the user's query. It is compared with proposed model in next section of paper.
For multiple clouds utilization, there is need to manage multiple clouds interfaces to achieve level of fault tolerance.

Need of Resource Management Architecture: - It is needed to monitor, maximize and distribute the computational resources available from each cloud's business. There are several multiple clouds managements including Internet protocols and some cloud federations. Inter-cloud protocols consists of six layers viz Actual physical layer, physical metaphor layer, platform layer, communication layer, management layer and end points layer.

3.2. View of proposed Ontology Cloud Computing Open Architecture (OCCOA)



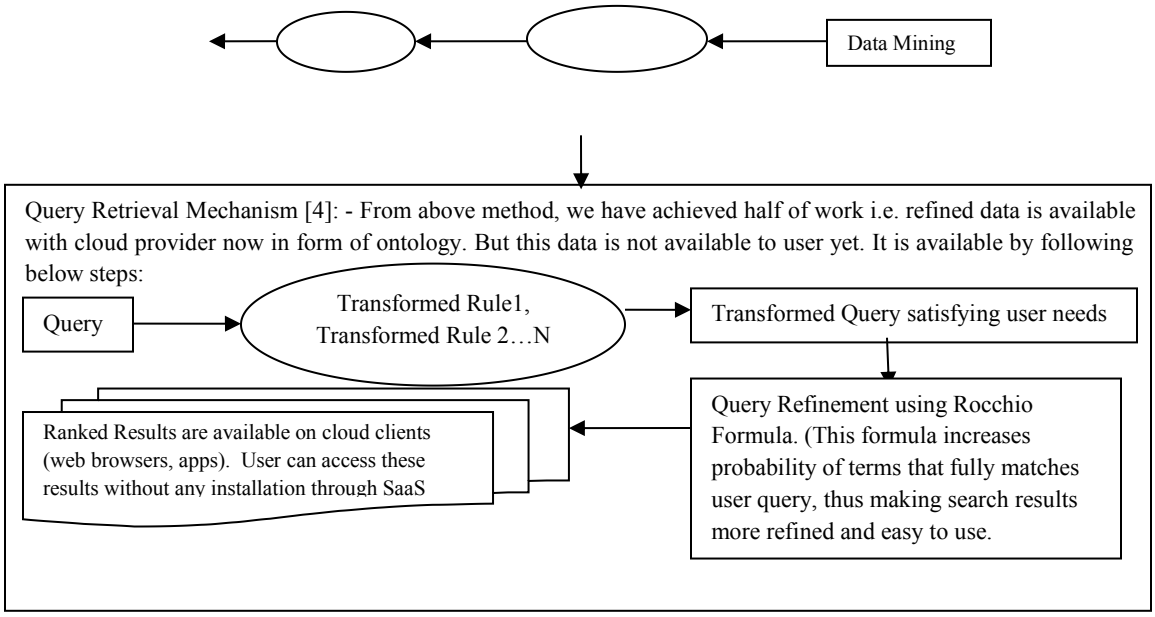


Fig 1: Proposed OCCOA

3.3 Working of Query Retrieval Mechanism

The given query is extracted in following manner as:

- The query entered by user contains terms which may or may not be relevant. It is made relevant by use of Filters.
- Then filtered query process applies various ontology based rules to create a separate query which executes independently using full text search engine.
- After applying rules, we get transformed query corresponding to transformation.
- The query is executed and ranked results are produced on cloud clients (web browsers and applications).

3.4 Query Refinement using Rocchio Formula

Extracting relevant terms and documents from huge collection of search result links is cumbersome task. Although query retrieval process makes this task much easier but our aim is to make model more refined and scalable. It can be done by refinement of transformed query produced after applying generation of ontology and transformation rules.

A tern can have different weights in each relevant document, so there is need to refine query. Query Refinement means calculation of old weights of expanded query terns in order to produce new weights of same query terns. These query terms are transformed into dummy document that is used for Indexing.

Here is formula used that calculates new weights of query terms and produces optimal results by discarding non relevant terms. It is called **Rocchio Formula**.

Aim: - The aim of this formula is to increase weights of terms that occur in relevant documents and decrease the weights of terms occurring in non relevant documents.

Equation: -

$$\mathbf{Q_a\ (new) = x * Q_a\ (old) + y * 1/\ (RD) * \sum wt_{aRD} - z * 1/\ (NRD) * \sum wt_{aNRD}}$$

Where Q_a (new) = New weight of query term a

Q_a (old) = old weights of tern s

RD = Relevant documents judged by user

NRD = Non- Relevant documents judged by user

wt_{aRD} = Weights of terms in relevant documents

wt_{aNRD} = Weights of terms in non relevant documents

$\sum wt_{aRD}$ = All weights of RD are added together

$\sum wt_{aNRD}$ = All weights of NRD are added together

y = It is constant that gives average of weights of terms in RD

If new weight is negative, then it will be discarded automatically.

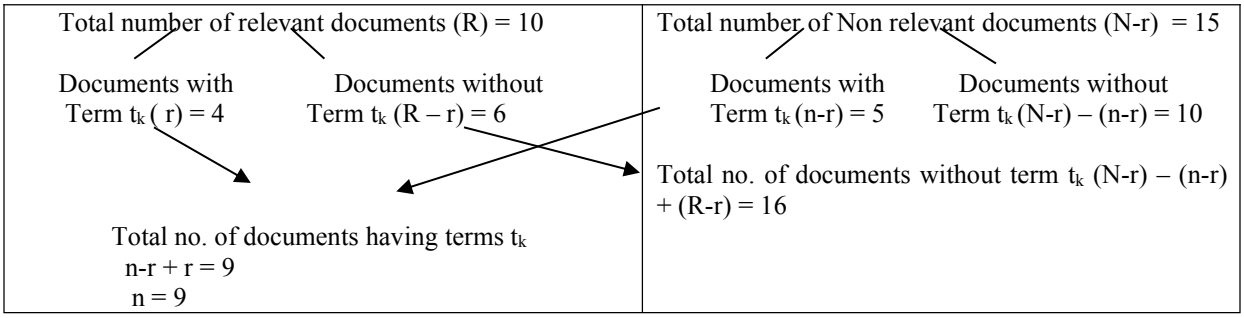
3.5 Discrepancies found in existing cloud computing information architecture model

The existing cloud information architecture uses "GrepTheWeb" architecture for extracting documents from huge collection of online documents. But, it has few shortcomings:

- It requires two inputs before processing starts. One input is list of documents selected by user from Internet and second is Regular Expression. Regular expression is set of terms, phrases that are related to user's query.
- For maintaining two inputs, user has to devote lot of time by analyzing whole documents and it leads to more manual work.
- After user involvement, the application will respond. It retrieves subset of documents that matches given query.
- It does not filter documents selected by user and does not performs indexing on them. Indexing means creation of metadata.
- Grep architecture does not maintain relationships between terms retrieved by it.
- The existing model does not have provision of reusing the extracted terms for further need. Again user has to select documents from web and whole process repeats.

4. ANALYSIS OF DATA USING ROCCHIO FORMULA & IMPLEMENTING IT ON ANEKA PLATFORM

Given data



According to [16],
Total number of documents N= 25
Total number of documents with term t_k (n) = 9
Total number of relevant documents (R) = 10
Total number of relevant documents with term t_k (r) = 4
From above data,
P_k = Probability of term t_k occurring in relevant documents
= 4/10 = 2/5
U_k = Probability of term t_k occurring in non - relevant documents
= 5/15 = 1/3
$X = P_k / (1- P_k)$
= (2/5)/ (3/5) = 2/3
$Y = U_k / (1 – U_k)$
= (1/3) / (2/3) = ½
Odd Ratio or Weighting Function $W_k = X/Y = 4/3$
Ranking function W = log (X/Y) = log (4/3) = 0.20068

On the basis of above graph and probability values, we can find new weight function for terms from old weight function by using Rocchio Formula.

Q_a (new) = x * Q_a (old) + y * 1/ (RD) * $\sum wt_{aRD}$ – z * 1/ (NRD) * $\sum wt_{aNRD}$

Here Q_a (old) = 4/3
Relevant Documents (RD) = 10

Non relevant documents (NRD) = 15
$\sum \mathbf{wt_{aRD}} = 4 + 6 = 10$
$\sum \mathbf{wt_{aNRD}} = 5 + 10 = 15$
X = 1, y = (4+6) / 2 = 5, z = (5+10) / 2 = 15/2 = 7.5
So, Q_a (new) = 1 * (4/3) + 5 * (1/10) * 10 – 7.5 * (1/15) * 15 = 4/3 + 5 – 7.5 = - 3.7
Since new weight function is negative, so it is discarded and old function is considered as relevance function

4.1. Implementing data on ANEKA

ANEKA [17] is one of first platform for developing applications on cloud. It utilizes extra CPU cycles in order to access desktop services from heterogeneous networks**.** It is market oriented platform or framework.
It supports open-ended set of abstractions and features for distributed computing and deployment scenarios. ANEKA cloud is combination of multiple resources connected to each other in a network. These resources can be modified as per user demands by using Virtualization.

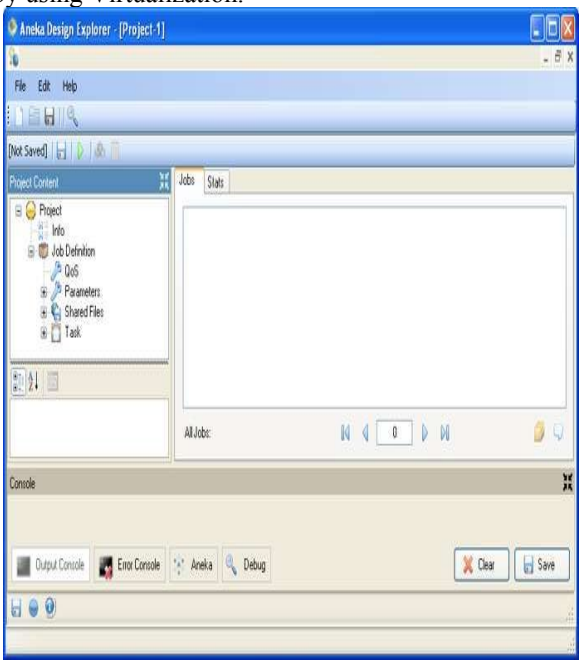


Fig 2: Design explorer of ANEKA

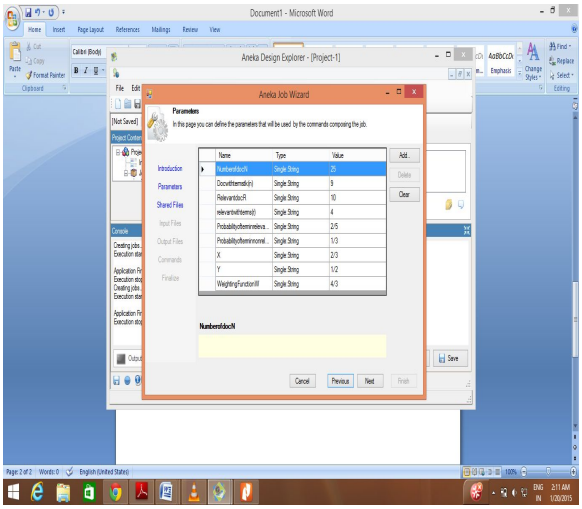


Fig 3: Adding given dataset parameters in ANEKA job wizard

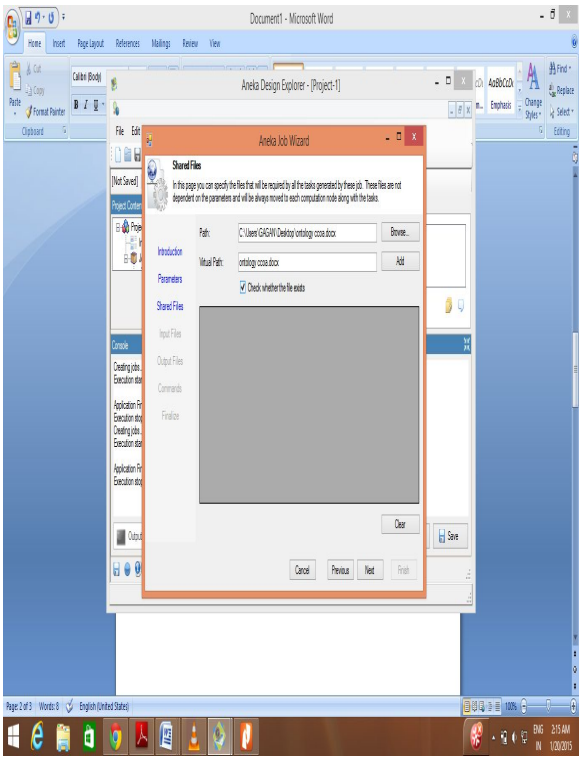


Fig 4: Given file ontology ccoa.docx is being shared by ANEKA platform

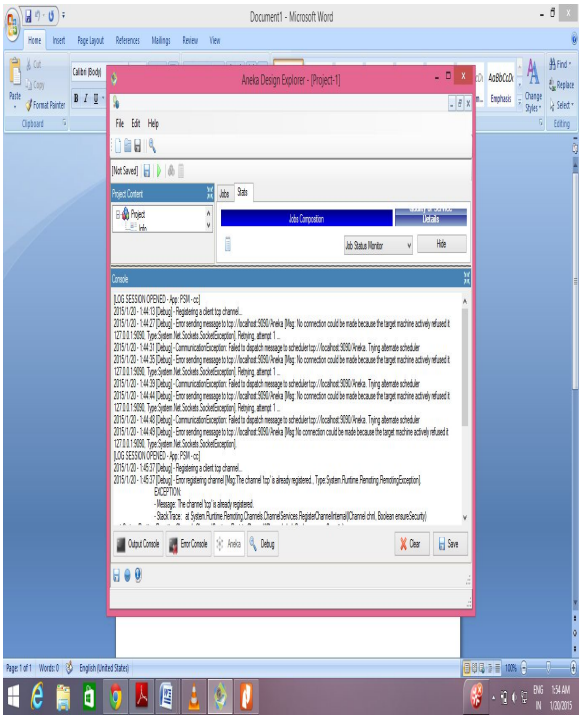


Fig 5: ANEKA TCP client connection on local host

5. CONCLUSION

The paper presents pros and cons of cloud computing in world of information and technology. It describes cloud computing from infancy phase to developing phase. Its roots are still growing and occupying minds of various technologists and researchers. There are several models of cloud computing like Multi Source information architecture, security architecture, open architecture but none of them is able to provide any solution for refining of query terms extracted from web documents.

Its solution lies in concept of Ontology. It means to create relationships and associations between extracted documents so that user can easily identify those documents and use their services from cloud clients. Cloud providers provide these services as PAY PER USE. The proposed cloud information architecture deals with ontology generation module for unified representation of resources and messages stored on cloud clients. After accumulation of data at one place in cloud, it becomes easier for user to access data and services using Query retrieval mechanism. This mechanism filters results and refines them as relevant or non relevant. Unlike Grep architecture, proposed model also provides provision of reusing produces results for future use. It makes our search refined and efficient. The dataset is analyzed and its implementation can be done on ANEKA cloud platform.

REFERENCES

[1]. G. GRUMAN, E. KNORR, What cloud computing really means. InfoWorld, (2009, May). [Online].Available: http://www.infoworld.com/d/cloudcomputing/ what-cloud-computing-really means- 031

[2]. SECURITY GUIDANCE FOR CRITICAL AREAS OF FOCUS IN CLOUD COMPUTING [Online].Available: http://www.cloudsecurityalliance.org/guidance/csaguide.pdf

[3]. L. WANG, J. TAO, M. KUNZE, A. C. CASTELLANOS, D.KRAMER, W.KARL, Scientific Cloud Computing: Early Definition and Experience, In HPCC '08, pp. 825–830.

[4]. Gagandeep Singh Narula, "Information Retrieval through Semantic Web-An Overview", In proceedings of CONFLUENCE-2012 at Amity School of Engg and Technology (ASET).

[5]. V Krishna Reddy, Dr. L.S. Reddy, "Security Architecture of Cloud Computing", International Journal of Engineering Science and Technology (IJEST).

[6]. "Security and high availability in cloud computing environments", IBM Global Technology Services Technical White Paper, IBM, June 2011

[7]. Tim Mather, Subra Kumaraswamy, Shahed Latif "Cloud Security and Privacy", O'Reilly Media, 2009

[8]. V. Krishna Reddy, B. Thirumal Rao, Dr. L.S.S. Reddy, P.Sai Kiran "Research Issues in Cloud Computing " Global Journal of Computer Science and Technology, Volume 11,Issue 11, July 2011

[9]. Gartner's 2008 Data Center Conference Instant Polling Results: Virtualization Summary – March 2, 2009

[10]. Rao Mikkilineni, Vijay Sarathy "Cloud Computing and Lessons from the Past", Proceedings of IEEE WETICE 2009, First International Workshop on Collaboration & Cloud Computing, June 2009.

[11]. Liang-Jie Zhang and Qun Zhou "Cloud Computing Open Architecture" published in IEEE International Conference on Web Services in 2009 at IBM T.J. Watson Research Center, New York, USA

[12]. Kui Ren, Cong Wang, and Qian Wang "Security Challenges for the Public Cloud" IEEE Computer Society 2012

[13]. Wang Zhijun, Zhang Ni, Qiao Zizhi, Lin Min "A Survey on Cloud Computing Security "IEEE OMEE conference 2012.

[14]. Aashish Bhardwaj "Cloud Security Assessment and Identity Management Proceedings of 14th International Conference on Computer and Information Technology (ICCIT 2011) 22-24 December, 2011

[15]. Christopher M.Moyer , "Building Applications in the Cloud Concepts , Patterns and Projects "Pearson ISBN 976-81-3176406000

[16]. Rifat Ozcan, Y. Alp: 'Concept Based Information Access using Ontologies and Latent Semantic Analysis', Technical Report, 2004-08

[17] Rajkumar Buyya, and Ivona Brandic, Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility, Future Generation Computer Systems, 25(6):599-616, Elsevier Science, Amsterdam, The Netherlands, June 2009.