



A Middle Layer Open Architecture to Search Query based Web Services over the Public Web

Ritu¹
Student, M.Tech. (CSE)
ITM University, Gurgaon

Neha Singh²
Asstt. Professor, M.Tech. (CSE)
ITM University, Gurgaon

Abstract: Web Services are the server side procedures provided to the web clients and web programmers to improve the application environment. When the service search is performed respective to particular keyword, there exists number of possible web service with almost same functionality. In such case, the comparative analysis is required in terms of availability, response time and security features of a web service so that effective selection of web service will be done. In this present work a web service search environment is defined that will first search the webservice based on keyword analysis and later on will index them under the availability and response time parameters. The work includes the indexing of web services under different parameters where each parameter will be assigned by a specific weightage. In this paper, the algorithmic work is defined for the web service identification and selection over the web..

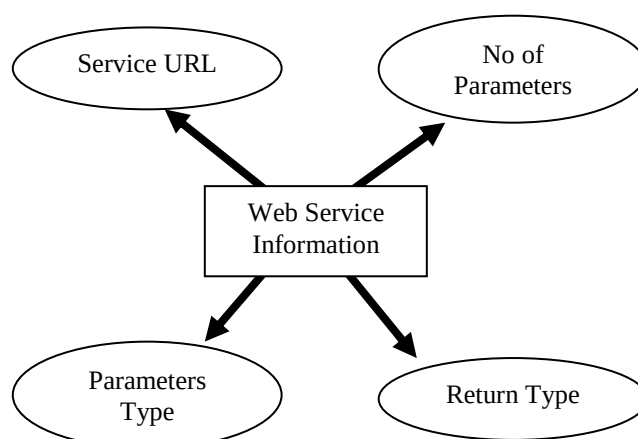
Keywords: Middle Layer Architecture, Web Service Selection, Availability, Response Time

I. Introduction

A web service is actually a web method that is available over the web publicly or with authentication. Web service helps a programmer or the client to consume the web service. There are number of service providers that provide the web contents in the form of services. To identify the service over the web, web service crawler program is required. This kind of program search and scan all the methods present on the services pages and provide the relative information to the user. The information presented by the crawler includes, the parametric information and the return type. The parametric information includes the number of parameters in the web service and the type of service parameters. This all information is incorporated in the WSDL (Web Service Description Language) incorporated within the service itself. Most of the available such crawler can provide the detail after specification of the service page, but they does not compare the available services under different parameters such as response time and the availability[1][2][4].

A web site crawler is the software, program or the bot that search throughout the website for all its pages respective to the defined keyword. It check the pages and identify the keyword relevant pages over the web. It build the list of web pages and index them based on the relevancy and present it to the user. Another aspect of

Figure 1 : Web Service Parameters Information



crawler is the level search. The level search is a about to search the internal links on a web page. The number of levels means the number of links within the web pages. The number of possible crawled web URLs are been generated by the server side software. One of the major challenges of these kind of web crawlers are to avoid the duplicate content search over the web. The web content crawling is performed effectively and presents it to the web client with the relative URL and the file format. In some crawler the relevancy vector as well as the matched content ratio is also available. The information collected for a web service is shown in figure 1.

The crawling process is used by the search engines to avail the web pages and web content for the downloading purpose. These kinds of search engines having the heavy traffic over the web server so that the main aim is to reduce the server efforts so that the effective search over the web will be drawn. To perform this, most the available servers maintain the index table that contains the contents as well as the service URL[5][8][10].

When a web search is completed, the information provided the search engine for the client end are listed as under

1. Provide the Downloadable Link or the Browsing Link for the page
2. Provide the major tag list to represents the author
3. Represents the server name and the meta data information related to web page.

The web service crawling is the improved form of web page crawling that extracts the service pages from the repository and identifies the list of available web methods and the attributes described in figure 1. The server starts by parsing the server and identifies the hyper links and notifies them as the server list. Each time a new link found, the index list is updated..

II. Literature Review

There is lot of work already done in the area of web service crawling and cloud search over the web. The information retrieval is one of the major interest area for the researchers. Such kind of work to identify the available sources over the cloud and grid servers was presented by Ioannis Katakis in year 2012. In this paper, author has defined a tagging approach to improve the capabilities of search system. The presented tool is an effective powerful tool to perform the software search that help the end users to perform the prediction based on the tag search. The multi label classification problem was focused and resolved by the author [1]. Another work on the software information retrieval on cloud architecture was performed by Marios D. Dikaiakos. Author defined a keyword support system to search the contents over the web and to define a software discovery analysis so that effective node search can be performed over the web. The work performed by the user was independent to the infrastructure, size of server etc. Author also deals with the unstructured nature of software and unavailability. Author also defines the work on search based on meta contents in the web information [2]. In year 2012, Bernaro Ferreira has defined a private data search on the storage cloud and performs the management of the web contents over the web. Author defined the prototype description of the available information present on the web under the internal structure analysis. This work includes the evaluation of the contents over the private web information system. The author simulated the work for the amazon.com and performs the information retrieval on the private contents of the website [3]. Maria-Elena Hernandez has also performed a search based on the tag analysis on the semi structural cloud architecture. The information search was focused on a specific kind of web servers called Clinical web servers. Author defined an interactive and light weighted search with synchronized support based on visual analysis and filtration of query results. Authors also perform the analysis based on the web interfaces and they provide the key term analysis over the web. Author taken a semi structured information analysis to derive the support based query results over the web [4].

Another work for the web content analysis was performed by Aameek Singh in year 2009. Author defined a outsource search on the outsourced storage. Author defined a file system based search approach that integrates the access control and index the search mechanism so that uniform framework was provided to the system The presented system was uniform in terms of access control and trust aware. Author defined a primitive search to encapsulate the access control with the index definition over the web [5]. Venkateshprasanna H.M. defined an enterprises search along with automatic synthesis analysis to identify the relevant information over the web. In this paper, author defined a new approach so that based on the descriptive analysis. The presented search mechanism. Author defined the new approach to perform an integrated search based on the knowledge optimization. The optimized results are presented by the author to high degree of productivity [6]. Another work on query log optimization to achieve the secure content retrieval was performed by Yuko Arai in year 2010. Author defined a secure query execution model in the key feature in the real time environment. Author presented a tuple extraction scheme on the plain text table with bloom filter based index and performs the search and replace on the keyword analysis. Author also perform the analysis on security based on the attack model and

analyze the query log. The work also includes the web content protection for the fake queries [7].

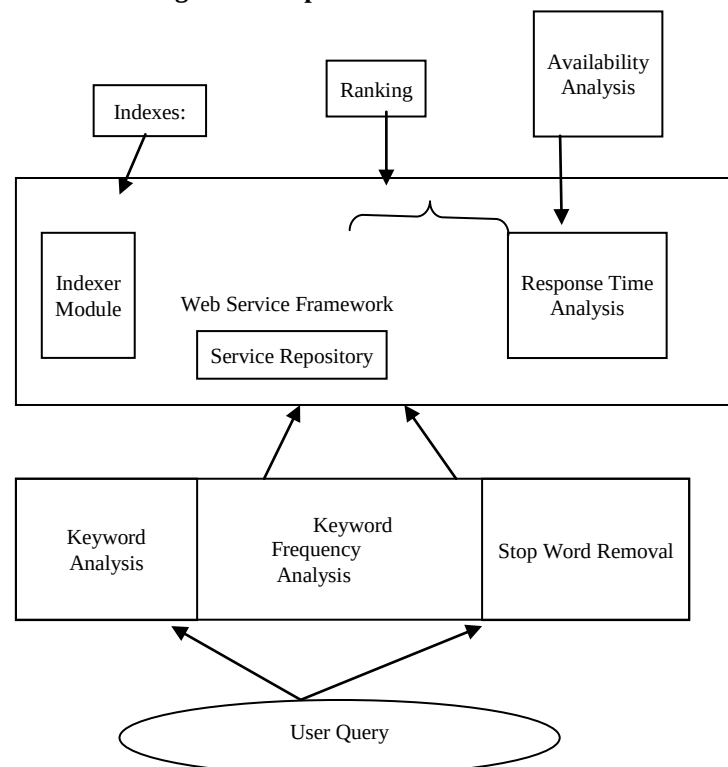
Another work on the personalization of services in an architecture was proposed by Hang Guo in year 2009. Author defined work on the personalized services. The work was implemented for the public cloud environment for the study of services over the cloud. Author also deal with the composition of services in two parts called the personalized data search over the web and to prove and present the optimized results [8]. In this work a multimedia search is been defined with the cloud based client architecture. Author defined a key trend in computing and also discusses the opportunities and challenges associated with large scale multimedia analysis and search. Author defined a research on the kind of search engine and improves the knowledge extraction for the clients. Author also deals with the decision making activities available online [9]. Kiran kumar has defined a work on the provenance on the data classification and provide a local storage system to perform the search based on the requirement anlaysis. Author defined a feasible and attractive approach to perform the search. The work tested on public web architecture [10]. Another work on the web, grid and cloud leveraging was performing under the service search by S. Heinzl in year 2009. Author defined an assisted application to achieve the high computing resources and to define combinational search in web application environment. Author presented an index results under the rank based approach. The application analysis was also performed by the author[11]. Another work based on video based search was performed on Zhi Ye in year 2010. Author defined a location search service by allowing the multimedia information retrieval over the web. Author defined a summarization policy to reterive the web contents and based on the content analysis the indexing of the URL was done. Author defined a probabilistic mechanism to perform the relationship analysis and identify the co-occurrence over the information retrieval [12].

III. Proposed Work

In this presented work a middle layer architecture is been defined over the web. This middle layer exists between the web servers and the web clients. The presented work will be incorporated in the middle layer of the proposed architecture. This middle layer will work as the processing layer between the web clients and the web. The client will pass it the information in the form of keyword search or the topic search. This input query will be analyzed by the processing layer to perform the filtration. The filtration process will remove the unrequired words from the query and will retrieve the actual keywords from the query. The query will also be analyzed respective to the keyword frequency based on which the priorities will be assigned to the keywords. Once the keywords will be retrieved the next work is to use this filtered query as the input query to the system. The basic architecture used in this proposed work is shown in figure 2.

After the derivation of client side query, the complete control shifted to the server side. Here the processing

Figure 2: Proposed Model



layer will perform a search over the web to identify the available web services that are relevant to the user requirement. The relevancy is here been performed under three parameters. First parameter is the existence of the keyword in the service url itself that give the most relevant vector and second is the query part exist in the service url that give second level of relevancy. The third parameter is the meta data information available regarding the web service is having the query contents.

After the relevancy check obtained from the work, the next work is to perform the analysis of the analysis of service under different parameters. The parameters considered here are the availability analysis, response time analysis and the security aspect. The availability is the periodic analysis that will check whether the service is in start mode or in stop mode. The response time check will be performed on the services available in start mode. The third aspect is the security that will check for the secure content retrieval is possible or not. Once these parameters are defined the next work is to perform the indexing on these web services.

To perform the indexing, the ranking is assigned to each web service. The ranking is here defined parametric with weighted values. The weightage is here defined under these defined parameters.

$$\text{Rank} = w1 * \text{relevancyVector} + w2 * \text{Availability} + \text{Vector} + w3 * \text{ResponseTime} + w4 * \text{SecurityAspect}$$

Once the ranking is assigned to these web services. The results in the rank based index order will be shown to the user. To present the result a user friendly interface will be taken. The results will include the service url and the information about the parametric constraints. The constraints will be shown in ratio form. Based on this output, user can select the most effective webservice for the service retrieval.

IV. Conclusion

In this paper, an open web architecture is defined under the parametric analysis to search query based web services over the web. In this model, a weighted model is suggested under multiple parameters to index the web services. The parameters considered here are based on the efficiency, reliability and security vector so that most effective service search can be performed.

V. References

- [1]. Ioannis Katakis, "Automated Tagging for the Retrieval of Software Resources in Grid and Cloud Infrastructures", 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing 978-0-7695-4691-9/12© 2012 IEEE
- [2]. MARIOS D. DIKAIKOS, "Minersoft: Software Retrieval in Grid and Cloud Computing Infrastructures", ACM Transactions on Internet Technology, Vol. 12, No. 1, Article 2, Publication date: June 2012 ACM 1533-5399/2012/06
- [3]. Bernardo Ferreira, "Management and Search of Private Data on Storage Clouds", SDMM'12, December 3-4, 2012, Montreal, Quebec, Canada. ACM 978-1-4503-1615-6/12/12
- [4]. Maria-Elena Hernandez, "Synchronized Tag Clouds for Exploring Semi-Structured Clinical Trial Data".
- [5]. AAMEEK SINGH, "Search-as-a-Service: Outsourced Search over Outsourced Storage", ACM Transactions on the Web, Vol. 3, No. 4, Article 13, Publication date: September 2009. ACM 1559-1131/2009/09
- [6]. Venkateshprasanna H.M., "Enterprise Search through Automatic Synthesis of Tag Clouds", COMPUTE '11, March 25-26, Bangalore, India ACM 978-1-4503-0750-5/11/03
- [7]. Yuko Arai, "Query Log Perturbation Method for Privacy Preserving Query", ICUIMC'10, January 14-15, 2010, SKKU, Suwon, Korea. ACM 978-1-60558-893-3
- [8]. Hang Guo, "Personalization as a Service: the Architecture and a Case Study", CloudDB'09, November 2, 2009, Hong Kong, China. ACM 978-1-60558-802-5/09/11
- [9]. Wei-Ying Ma, "Rethinking Multimedia Search in the "Clients + Cloud" Era", LS-MMRM'09, October 23, 2009, Beijing, China. ACM 978-1-60558-756-1/09/10
- [10]. Kiran-Kumar Muniswamy-Reddy, "Provenance as First Class Cloud Data".
- [11]. S. Heinzl, "MIRO: A Mashup Editor Leveraging Web, Grid and Cloud Services", iiWAS 2009, December 14-16, 2009, Kuala Lumpur, Malaysia. ACM 978-1-60558-660-1/09/0012
- [12]. Zhi Ye, "Video Based Mobile Location Search with Large Set of SIFT Points in Cloud", MCMC'10, October 29, 2010, Firenze, Italy. ACM 978-1-4503-0168-8/10/10