Identification of Candidate Services for Optimization of Web Service Composition

Gopinath Ganapathy and Chellammal Surianarayanan

***Abstract*—In automation of web service discovery, there is always a need to consider Quality of Service (QoS) attributes during matching. An analysis of literature concerning the evolution of various web service discovery methods with special emphasis to quality driven service discovery has been carried out in this work. It is found that little efforts have taken into consideration the quality attributes, 'importance' and 'trustworthy' of a web service during automated service discovery. In this work, a practical analysis has been performed to check whether Page Rank algorithm can be employed to find out 'important' services during discovery. Issues in employing the Page Rank algorithm are analyzed. Improved approaches, Trust Rank and Link Variable Trust Rank in finding 'trustworthy' of services during discovery are discussed. An optimization approach is suggested with the above algorithms towards an efficient web service composition by identifying 'important' and 'trustworthy' candidate services in dynamic and interactive web service composition scenarios.**

With this approach, the time involved in the execution of selection and composition of services is reduced by reducing the number of services those require semantic matching. This is done by filtering out those services which do not meet the predefined threshold values for 'importance' and 'trustworthy' of services. Preliminary results with a typical set of functionally similar web services are presented.

***Index Terms* — dynamic web service composition, importance, Link Variable Trust Rank, Page Rank, Trust Rank, trustworthy.**

I. Introduction

When a web service is described using a semantic description standard, it becomes a machine processable entity and automated service discovery becomes feasible. Despite the existence of rapidly growing web services and the disparate ways in which services are being described, the state-of-art Service Oriented Computing (SOC) paradigms demand only highly compatible and accurate services to be discovered and composed. Further, due to the existence of huge number of services, an automated service discovery generally meets with an appreciable number of functionally equivalent services. In this context, Quality of Service (QoS) attributes such as availability, performance, security, reliability, reputation, accessibility, response time, throughput, latency, memory usage, CPU usage etc are being used to distinguish and rank functionally similar services. From literature it is found that dynamic service composition needs a number of optimizations to be applied to reduce the number of candidate services for semantic matching as any semantic reasoner requires an appreciable amount of time in fulfilling a discovery request. In order to produce highly accurate and quality services, QoS attributes are being considered along with functional attributes during matching.

In this work, an analysis has been conducted to check the feasibility of using 'importance' and 'trustworthy' attributes as preprocessing/filtering measures in identifying candidate services those require matching of complete semantics of services during discovery. This work is organized as follows. Section I gives an introduction to the work, Section II describes a survey on various discovery methods and the need for the consideration of QoS attributes during discovery, Section III presents the related work, Section IV discusses the applicability of Page Rank to service discovery, Section V describes the applicability of Trust Rank to service discovery, Section VI proposes the optimization approach to identify reduced set of candidate services for composition and Section VII concludes the paper.

Manuscript received March 18, 2010.
Gopinath Ganapathy is with the School of Computer Science and Engineering, Bharathidasan University, Tiruchirrappalli, TN, 620 023, India (phone: 91-431-2331-663; fax: 91-431-2331-662; e-mail: gganapathy@gmail.com).
Chellammal Surianarayanan, Member IAENG (106383) is with the School of Computer Science and Engineering, Bharathidasan University, Tiruchirrappalli, TN, 620 023, India. (e-mail: chelsrsd@rediffmail.com).

II. Survey

Traditional Universal Description Discovery Integration (UDDI) based web service discovery mechanism is key word based and it retrieves services that contain particular keywords from user's query. Since query keywords might be synonyms, ('zip', and 'pin') and homonyms, ('order' in context of purchase order and 'order' of items in a shelf) simple key word based service discovery leads to low recall and low precision of the retrieved services [1]. Another problem with keyword-based service discovery approaches is that they cannot completely capture the semantics of the user's query because they do not consider the relations between the keywords. Despite difficulties of UDDI mechanism described in [2], it disassociates itself from the service descriptions standard and it does not register any information from service description [3].

As a key to limitations in UDDI key word, Web Service Description Language (WSDL) based discovery such as [4], [5] have been proposed. Though WSDL based method exploits the structure of web service description, it is purely a text based approach of discovery. In contrast to simple WSDL based text matching, WordNet enhanced techniques such as [6] have come up wherein the user's query of similar service is expanded with its synonyms. In [7], authors proposed a suite of service similarity assessment methods namely classical Vector Space Model (VSM), WordNet powered VSM, WSDL structure matching and WordNet powered WSDL matching along with experimental evaluations which prove that WordNet powered methods outperform the classical methods.

Practically, the current distributed computing models such as web service automation and dynamic web service composition cannot be accomplished by mere keyword based or text based matching methods, as service description such as WSDL do not explicitly define the purpose of a web service as intended by the provider [8], [9]. Various standards such as, Semantically Annotated Web Service Description Language (SAWSDL), Web Services Modeling Language (WSML) and Web Ontology Language for Services (OWL-S) are being used to specify and discover services semantically. SAWSDL extends WSDL with pointer to semantics that facilitates web service automation [10]. In WSML a service is described in terms of logical axioms and constraints in Ontologies through which it can be accessed for orchestration and choreography [11]. OWL-S (formerly known as DARPA Agent Markup Language, DAML-S) includes three primary sub-ontologies, namely service profile to describe what the service does, process model to describe how the service is used and grounding to describe how to interact with the service. How service capabilities are described using OWL-S has been presented in [12]. With the above semantic service description, a service becomes a machine process able entity without ambiguity and sufficient semantics. This facilitates the maximal automation and dynamism in service discovery and composition [13]. Further, depending on the way in which service is described, various logic reasoning based methods are being used for discovery. These methods generally perform signature matching and specification matching with Input, Output, Precondition and Effect (IOPE) capabilities.

The Quality attributes of a service need to be considered along with functional attributes during service discovery for two reasons, firstly, to facilitate dynamic web service composition which needs highly compatible and accurate services, secondly as Quality of Service (QoS) can distinguish and rank the functionally similar services. In [14], the existence of various languages and Ontologies/models to describe the Quality of Service (QoS) and the existing interoperability issue due to the disparate ways of description of QoS requirements are discussed in detail. The various ways of classification of QoS attributes are described in [15], [16] and [17].

III. Related Work

A significant number of research works such as [18], [19], and [20] have been proposed for matching QoS requirements along with IOPE matching. It is found that most of the previous works considered attributes such as availability, performance, security, reliability, reputation, response time, throughput, latency, memory usage, and CPU usage during discovery. Link-based Page Rank model for defining importance of services has been proposed in [21].

Hyperlink based approaches such as Selective Hypertext Induced Topic Search (SelHITS) proposed in [22] which calculates hub and authority values of web pages and Trust Rank proposed in [23] which assigns quality score to web pages gain momentum in meeting the needs of identifying authority and quality pages among huge and rapidly evolving Web.

Whereas in this work, Page Rank, Trust Rank and Link Variable Trust Rank algorithms are employed in web services to check their applicability in finding out the values of non-functional attributes, ‘importance’ and ‘trustworthy’ of services. Further, it is proposed to use these attributes in filtering and identifying the quality candidate services to enhance the efficiency of service discovery.

IV. Page Rank Employed To Web Service Discovery

Each web service is exposed through its interface description such as WSDL. Though a WSDL file is full of service semantics and it is more relevant resource to developer community rather than information seekers, from the Internet/Web point of view, the Uniform Resource Locator (URL) of a WSDL is still equivalent to web page. To justify the above treatment a typical WSDL file is tested for its contents and links. It is found that a WSDL generally does not contain any reference/hyperlink going out of it to other web pages/resources except in <documentation> tag. Further, generally, the developers intend to give only the context information in the <documentation> tags. Secondly a WSDL is tested for its in-links with Yahoo Site Explorer and it is found that the in-links of a WSDL files are typically web pages. With the above analysis, the URL of a WSDL can be treated as a simple web page with in-links but not out-links as shown in Fig 1.

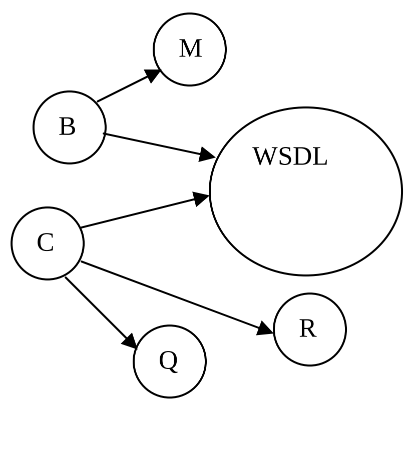

Fig 1 Treatment of WSDL as web page

The WSDL shown in Figure 1 has two in-links ‘B’ and ‘C’ and no out-links. The page ‘B’ propagates its importance to two pages, ‘M’ and ‘WSDL”. The page ‘C’ propagates its importance to three pages, ‘Q’,’R’ and ‘WSDL’. With the above structure it is understood that Page Rank algorithm can be employed to WSDL as Page Rank algorithm involves only the in-links of WSDL and out-links of each in-link pointing to WSDL. The Page Rank computation for a WSDL does not require out-links of WSDL. The Page Rank of a WSDL is equal to the sum of Page Rank of all its in-links.

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

With the above perspective, Google Page Rank algorithm has been employed to a set of functionally similar Short Message Service (SMS) WSDLs, collected from Internet. The Google API assigns Page Rank value to web pages with a non linear scale of value ranges from 0 to 10 based on the number of in-links such as a page with number of in-links between 6 to 25 is assigned with Google Rank of 2 and a page with number of in-links between 25 to 125 is assigned with a Google Rank of 3. The number of in-links of the above URLs has been found with Yahoo Site Explorer. The Page Rank of the URLs of WSDL have been found using Google API. The number of in-links of web site of WSDL, Page Rank of the web site wherein the WSDL is located, the number of in-links of the WSDL URL and Page Rank of the entire WSDL URL are given as a ranked list in Table 1. Ranking of WSDLs is performed in two steps, firstly with the Page Rank of the site and then with Page Rank of the entire WSDL URL

Table 1. List of services ranked with Page Rank

#	WSDL URLs	No of lnlinks of site	Page Rank of Site	No of in-links of the WSDL	Page Rank of WSDL
1	http://ws.strikeiron.com/globalsmspro2_5?WSDL	13	6	25	2
2	http://www.webservicex.com/SendSMS.asmx?WSDL	57	5	2	1
3	http://www.esendex.com/secure/messenger/soap/SendService.asmx?wsdl	627	4	7	2
4	http://www.info-me-sms.it/Ws.php?wsd	66	4	2	0
5	http://www.aswinanand.co/sendsms.php?wsdl	573	3	24	0
6	http://www.ss.mio.it/webserices/sendmessages.asmx?WSDL	142	3	8	0
7	http://ws.cdyne.com/SmsWS/SMS.asmx?wsdl	17	2	12	2
8	http://www.abctext.com/webservices/SMS.asmx?WSDL	9	2	6	1
9	http://ws.acroscommunications.com/SMS.asmx?WSDL	6	0	3	0

From the above table, it is understood that a combination of Page Rank of web site through which the WSDL is exposed and Page Rank of the entire WSDL can very well be applied to find important services during service discovery. Page Rank value of web services can be used in two ways namely to rank the functionally similar services and to optimize interactive and dynamic web service composition by choosing only those services for which the importance/authority is greater than a user defined threshold value. For a typical user defined threshold Page Rank of 3, the ranked and filtered list of candidate services for composition is given in Table 2.

Table 2. Page Rank filtered list of candidate services

#	WSDL URLs	No of lnlinks of site	Page Rank of the Site	No of inlinks of WSDL	Page Rank of WSDL
1	http://ws.strikeiron.com/globalsmspro2_5?WSDL	13	6	25	2
2	http://www.webservicex.com/SendSMS.asmx?WSDL	57	5	2	1
3	http://www.esendex.com/secure/messenger/soap/SendService.asmx?wsdl	621	4	7	2
4	http://www.info-me-sms.it/Ws.php?wsdl	66	4	2	0

V. TRUST RANK EMPLOYED TO WEB SERVICE DISCOVERY

Page Rank algorithm assigns a global importance score to a web page based on the importance of web pages pointing to it. While Page Rank is a good approach to assign importance score to a service, it does not qualify a site whether it is a good one or bad one, i.e. the Page Rank algorithm is vulnerable to web spam and many research efforts such as [24] are proposed for identifying web spam. The Trust Rank algorithm is another link based algorithm used to detect web spam. It assigns trust score to sites based on the quality of the site. It is based on the assumption that good sites seldom point to bad sites and it can be used in detecting the spam links which are pointing to the URL of the WSDL.

The trustworthy score for each service in the test data has been obtained with the Trust Rank tool available from http://www.linkvoodoo.com/site_inlinks.php. This tool internally selects a small set of expert evaluated seed pages and then crawls out from these seed pages to find reliable pages and assigns a quality/trust score to the WSDL. Hence the importance found with Page Rank algorithm can be improved by removing the spam links pointing to WSDL. The test data ranked with trust score is given in Table 3.

Table 3. List of services ranked with trust score

#	WSDL URLs	Trust score (%)
1	http://ws.strikeiron.com/globalsmspro2_5?WSDL	29.92
2	http://www.esendex.com/secure/messenger/soap/SendService.asmx?wsdl	28.05
3	http://ws.cdyne.com/SmsWS/SMS.asmx?wsdl	27.61
4	http://www.abctext.com/webservices/SMS.asmx?WSDL	24.62
5	http://www.aswinanand.com/sendsms.php?wsdl	23.62
6	http://www.webservicex.com/SendSMS.asmx?WSDL	23.16
7	http://ws.acrosscommunications.com/SMS.asmx?WSDL	23.16
8	http://www.sms.mio.it/webservices/sendmessages.asmx?WSDL	22.59
9	http://www.info-me-sms.it/Ws.php?wsdl	15.34

VI. Methodology

From Table 1 and Table 3, it is found that the difference in ranking services based on Page Rank and Trust Rank is significant. The service "http://www.webservicex.com/SendSMS.asmx?WSDL" which got ranked at the second place by the Page Rank, got assigned a trust score of 23.16% and ranked at the 6th place out of 9. Similarly though the Page Rank of "http://www.abctext.com/webservices/SMS.asmx?WSDL" is only 2 and ranked at 8th place by Page Rank; its trustworthy score is 24.62% and got ranked at 4th place by Trust Rank. It is essential that the important services found with Page Rank have to be tested for trust score as the original Page Rank of a service might be altered by a spam which cannot be detected by Page Rank

When an analysis is made with the above ranking procedures, it is understood that in the context of web service discovery, Page Rank algorithm can be employed with an intention to measure the quality attribute 'importance' of a service i.e. it gives a measure of number of important in-links referring to a service and Trust Rank can be employed to find the quality of a WSDL file against spam.

During service composition, the component services that are able to provide the required functionalities are discovered and composed. The number of services providing a given functionality may be large and constantly changing [25]. In a dynamic service composition, the component services are identified during execution time with a set of criteria for selecting services. Selection of service is usually performed with a semantic service matching approach such as OWLS-MX, WSMO-MX and SAWSDL-MX. Though the requirements of dynamic web service composition are met with semantic service matching techniques, semantic service matching task is expensive in terms of time, for an example, time required to fulfill a request composed of single service with ten concepts is of the order of 4 to 5 seconds [26]. Hence a number of optimizations are required for efficient service composition. Generally optimization is performed in two ways; firstly by reducing the number of services to a semantic reasoner so that the execution time in selecting and composing services will be reduced and secondly by improving the performance of a semantic reasoner with its internal architecture. As the second approach varies from reasoner to reasoner depending on its architecture, the first method is a preferable approach.

An optimization approach shown in Fig 2 is suggested to reduce the number of services to a semantic matcher, by identifying relevant, important and trustworthy candidate services for composition.

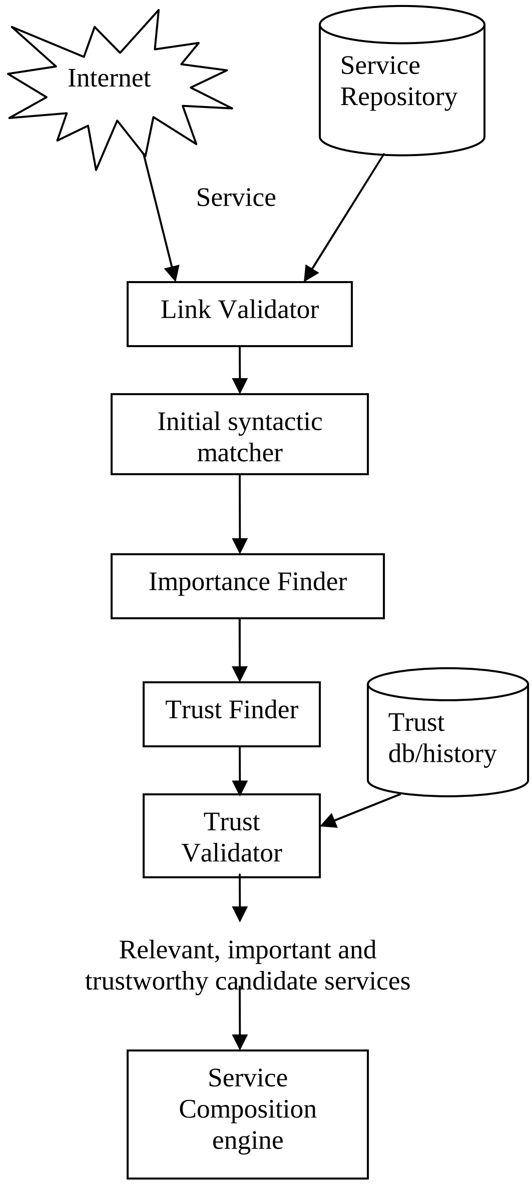

Fig 2. Optimization Approach

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Web services stored in a service repository or services from Internet arrive are initially tested with any link validator such as the Web Link Validator available from REL software. This tool takes the service URL and produces information about broken link, redirected link and status of the link. Any broken link found with this validator is not considered for service matching. Then the services are checked for their relevance with the required/queried service using any syntactic service matcher. A simple signature matching is performed syntactically to produce a set of syntactically relevant web services. The importance of the relevant services is found using the Google Page Rank API. Then only those services whose 'importance' is more than predefined 'importance' are chosen for further matching. The important services are evaluated for their trust score using a Trust Rank finder. Trustful services are found by comparing the trust score of a concerned service with predefined trust score.

Trust Rank algorithm is based on the assumption that good sites are seldom point to bad site, however, good sites may not always point to other good sites as there are possibilities like web spammers may leave many spam comments in good blogs which point spam sites. The algorithmic concept described in [27] is used to identify web spam in-links to a WSDL. The concept is based on the assumption that the web structure of spam change drastically in short period of time.

The WSDLs whose inbound links vary abruptly are found using a Trust Validator. The validator compares the number of in-links of a WSDL to the previously stored in-links information from a Trust db/history. Services with abruptly changing in-links are filtered out. Initial trust db is constructed for the services available from various service portals and directories such as "http://www.xmethods.net", "http://remotemethods.com", "www.wsindex.org" and "http://webservices.seekda.com". If there is not trust history for a service, the current number of in-links is stored in the database and used for future requests. With the above preprocessing steps, the approach results in a reduced set of relevant, important and trustworthy candidate services ready of semantic matching by service composition execution engine.

Experiments with the above optimization approach to a set of syntactically similar functional SMS services, the initially detected number of SMS services, 9 is reduced to 4 services for the predefined value of Page Rank 3 and trust score 24%. It reduced the number of syntactically relevant services from 9 to 4 and thus saves the execution time of composition engine by approximately 5*4 = 20 seconds as any ontology reasoner consumes a minimum discovery time for single service with 10 concepts is 4 seconds.

VII. Conclusion

Experiments have been conducted to study the applicability of Page Rank, Trust Rank and Link Variable Trust Rank algorithms to find 'importance' and 'trustworthy' of services and utilizing those values in identifying important and trustworthy candidate services after a syntactic matching. An optimization approach has been suggested to identify a reduced set of candidate/component services during dynamic composition. Initial results with proposed approach showed that it is an effective mechanism to obtain a reduced set of syntactically relevant services with predefined 'importance' and 'trustworthy'.

References

[1] Qingcong Lv, Jiliang Zhou, Qiying Cao, "Service Matching Mechanisms in Pervasive Computing Environments", International Workshop on Intelligent Systems and Applications, 23-24 May 2009, Wuhan, pp. 1-4

[2] Colin Atkinson, Philipp Bostan, Oliver Hummel and Dietmar Stoll, "A Practical Approach to Web Service Discovery and Retrieval" 2007 IEEE International Conference on Web Services (ICWS 2007), pp.241-248

[3] Jorge Cardoso, John AMiller and Savitha Emani, "Web Services Discovery Utilizing Semantically Annotated WSDL", Reasoning Web LNCS 5224, C.Baroglio et al, Springer Berlin / Heidelberg, Volume 5224/2008, pp. 240-268

[4] Kyong-Ha Lee, Mi-young Lee, Yun-Young Hwang, Kyu-Chul Lee, "A Framework for XML Web Services Retrieval with Ranking" 2007 IEEE International Conference on Multimedia and Ubiquitous Engineering (MUE'07), pp.773-778

[5] Pierluigi Plebani and Barbara Pernici, "URBE: Web Service Retireval based on Similarity Evaluation", IEEE Transactions on Knowledge and Data Engineering, November 2009, Vol. 21, no. 11, pp. 1629-1642

[6] Ricardo Sotolongo, Carlos A. Kobashikawa, Fangyan Dong, Kaoru Hirota, "Algorithm for Web Service Discovery Based on Information Retrieval Using WordNet and Linear Discriminant Functions", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.12, No. 2 pp. 182-189

[7] Yiqiao Wang, Eleni Stroulia, "Structural and semantic matching for assessing Web service Similarity", International Journal of Cooperative Information Systems (IJCIS) 14(4), 2005, pp. 407-438

[8] Kaarthik Sivashanmugam, John A Miller, Amit P. Sheth, Kunnal Verma, "Framework for Semantic Web Process Composition", International Journal of Electronic Commerce, M.E. Sharpe Inc publishers, 2005, Vol 9. Issue 2 (Number 2/Winter 2004-5), pp. 71-106

[9] Tanveer Syeda-Mahmood, Gauri Shah, Rama Akkiraju, Anca-Andrea Ivan, Richard Goodwin, "Searching Service Repositories by Combining Semantic and Ontological Matching", IEEE International Conference on Web Services (ICWS'05), Orlando, Florida, July 2005, pp. 13-20

[10] Jacek Kopecky, Tomas Vitvar, Carine Bournez, Joel Farrell, "SAWSDL: Semantic Annotations for WSDL and XML Schema", IEEE Internet Computing, IEEE Computer Society, 2007, Vol. 11, No. 6, pp. 60-67

[11] Matthias, Klusch, Frank Kaufer, "WSMO-MX: A hybrid Semantic Web service matchmaker" Web Intelligence and Agent Systems, IOS Press publishers, Vol. 7, Issue 1, pp. 23-42

[12] Ruiqiang Guo, Jiajin Le, XiaoLing Xia, "Capability matching of Web services based on OWL-S", Proceedings of 16th International Workshop on Database and Expert Systems Applications", 22-26 August 2005, pp. 653-657

[13] Katia Sycara, Massimo Paolucci, Anupriya Ankolekar, Naveen Srinivasan, "Automated discovery, interaction and composition of Semantic Web Services", Journal of Web Semantics, Elsevier, December 2003, Vol. 1, No.1, pp. 27-46

[14] Vuong Xuan TRAN and Hidekazu TSUJI, "A Survey and Analysis on Semantics in QoS for Web Services", International Conference on Advanced Information Networking and Applications, May 2009, pp. 379-385,

[15] J. O'Sullivan, "What's in a Service? Towards Accurate Description of Non-functional Service Properties", International Journal of Distributed and Parallel Databases, Springer Verlag, 2002, Vol 12, pp. 117-133

[16] S.P. Ran, "A Model for Web Service Discovery with QoS", ACM SIGecom Exchanges, Vol. 4(1), ACM Press, 2003, pp. 1-10

[17] SeokHyun Yoon, DongJoon Kim, SangYong Han, "WS-QDL containing static, dynamic, and statistical factors of Web services quality", Proceedings of the IEEE International Conference on Web Services, IEEE Computer Society, 2004, pp. 808

[18] Marco Conti, Mohan Kumar, Sajal K. Das and Behrooz A Shirazi, "Quality of Service Issues in Internet Web Services" IEEE Transactions on Computers, Vol. 51, No. 6, June 2002, pp. 593-594

[19] Y. Makripoulias, C. Makris, Y. Panagis, E. Sakkopoulos, P. Adamopoulou, A. Tsakalidis, "Web Service discovery based on

Quality of Service," AICCSA, IEEE International Conference on Computer Systems and Applications, 2006., pp.196-199.

[20] Mossab A. Al Hunaity, "Towards an Efficient Quality Based Web Service Discovery Framework", Proceedings of 2008 IEEE Congress on Services – Part I, IEEE Computer Society, 2008, pp. 261-264

[21] Camelia Constantin, Bernd Amann, David Cross Amblard, "A Link-based Ranking Model for Services", Springer-Verlag, Berlin Hiedelberg, R. Meersman, Z. Tari et al. (Eds), OTM 2006, LNCS 4275, 2006, pp. 327-344

[22] Amit C. Awekar, Pabitra Mitra, Jaewoo Kang, "Selective Hypertext Induced Topic Search", Proceedings of the 15th International Conference on World Wide Web, ACM, May22-26 2006, pp. 1023-1024

[23] Zoltan Gyongyu, hector Garcia-Molina, Jan Pedersent, "Combating Web Spam with Trust Rank", Proceedings of the 30th VLDB Conference, Toronto 2004, pp. 576-587.

[24] Andras A. Benczur Karoly Csalogany Tamas Sarlas, "Link Based Similarity Search to Fight Web Spam", AIRWEB'06, August 10, 2006, Seattle, Washington, USA

[25] Liangzhao Zeng, Boualem Benatallah, Marlon Dumas, Jayant Kalagnanam and Quan Z. Sheng, "Quality Driven Web Services Composition", 12th International World Wide Web Conference, Budapest, ACM, 2003, pp. 411-421

[26] Sonia Ben Mokhtar, Anupam Kaul, Nikolaos Georgantas and Valerie Issarny, "Towards Efficient Matching of Semantic Web Service Capabilities", International Workshop on Web Services Modeling and Testing (WS-MaTe 2006)

[27] Qi Chen, Song-Nian Yu, Sisi Cheng, "Link Variable Trust Rank for Fighting Web Spam", Proceedings of the 2008 International Conference on Computer Science and Software Engineering", IEEE Computer Society, 2008, Vol. 4, pp-1004-1007

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)