Evaluating the Performance of Semantic Service Search Engine based-on SVM

V.P.HARA GOPAL, M.Tech,
Asst. Professor, Department of IT,
RGMCET, Nandyal, Andhra Pradesh

K.F.BHARATI M.Tech,(Ph.D.)
Asst. Professor, Department of CSE,
JNTUA, Anantapur, Andhra Pradesh.

ABSTRACT
Semantic service search engine is designed for DE, to provide semantic search support. Digital ecosystem (DE) is comprised of heterogeneous and distributed species which can play the dual role of service provider and service requester. DE needs to provide a reliable and trustworthy link between service providers and service requesters. So, semantic service search engine is a conceptual framework of service-ontology-based system which delivers semantic search support to DE. This paper is intended to measure the performance of the semantic service search engine based-on Support Vector Machine (SVM). The performance is evaluated through six performance indicators of information retrieval (IR). They are Precision, Recall, Mean average precision, Harmonic mean, Fallout rate and Mean Reciprocal Rank. The conclusion to this evaluation and future works are provided in the final section.

General Terms
Quality of service (QoS) evaluation, semantic service search engine, support vector machine, information retrieval, performance indicators.

Keywords
Quality of service (QoS) evaluation, semantic service search engine, information retrieval, performance indicators.

1. INTRODUCTION
Semantic Service Search Engine has designed to provide semantic search support for Digital Ecosystem. Digital ecosystem (DE) is comprised of heterogeneous and distributed species which can play the dual role of service provider and service requester [1]. But, DE needs to support in semantic search, which means that it has to provide a reliable and trustworthy link between service providers and service requesters. To solve this issue, a conceptual framework of a service-ontology-based semantic service search engine is developed. In this model, besides semantic search functionality, it provides the Quality of Service based Evaluation and Ranking methodology. Currently available services are less semantic in network [1]. Thus it is not possible for a service requester to know about all other service providers for a service requested. Also, there isn't quality of service information available for any of the services.

This paper illustrates about semantic search interface indulging the ontology structure to the search engine to build a service-oriented architecture. The purpose of this is that the service requestor has to know about all the service providers of a service in an semantic search environment. For this, a Pizza ontology is build to measure the performance of the system based-on Support Vector Machine (SVM) model.

2. RELATED WORKS
Service Search Engine is a framework of semantic search engine, ontology concept, quality-of-service evaluation and ranking methodology. A brief review regarding the research about these is given in this section.

Gabor Nagypal [5] worked out about background knowledge stored in ontologies and semantic metadata which can be optimally exploited for the task of information retrieval. A special emphasis is placed on the issue of imperfect ontologies and metadata which is the reality on the present Web. Kevin and Andrew [10] described some of the problems that are addressed in data mining and knowledge discovery from web, and elements of the WebFountain infrastructure that they have built for addressing them.

Jingyu et al. [2] presented ExpertRec, a collaborative/social Web search engine. With ExpertRec, users share experts' search histories (search expertise) through a Web browser toolbar or a proxy browser. A CBR(case-based reasoning)-based recommendation engine was designed to build recommendations and its core is a scalable method to identify search expertise based on a hierarchical user profile in order to improve users' search quality. Primary evaluation shows that ExpertRec provides some functions which users like.

Larry et al. [7] proposed the role of case-based reasoning in semantic search, and in particular, as it applies to Knowledge Sifter, an agent-based ontology-driven search system based on Web services. The Knowledge Sifter architecture is extended to include a case-based cliché for collaborative semantic search, including case creation, indexing and retrieval services. A collaborative filtering methodology is presented that uses stored cases as a way to improve user query specification, refinement and processing.

Huaqiao Lv et al. [13] proposed that the individual search engine can't settle the problem of quickly obtaining information in information service industry any more. There need a more efficient way to get information. The personalization search engine with the assistant of computer and internet technology establish a personalized information service environment. It actively offers special information and oriented services "taking customer as its center" by the functions of the users ordering, the system recommending and delivering.

Indu Chawla [14] performed research in the area of personalization in web search. Various personalization strategies had proposed and personalized search systems had developed, but found that they all are not optimal. It represented a significant improvement over generic Web search for some queries, while it had little effect and even harms query performance under some situations.

Sadegh et al. [15] showed using data mining it can be improved the re-crawling policy and could improve performance of crawlers in WWW. Many large Web crawlers start from seed pages, fetch every links from them, and continually repeat this process without any policies that help them to better crawling and improving performance of those. Based on proposed coefficients, their proposed crawler had better performance than general purpose web crawlers from freshness point of view.

Minxue et al. [11] build a new effective e-commerce service quality system of Web2.0. First, a new service quality system was built based on TAM model and then the relation between the new service quality system and consumer behavior (satisfaction and loyalty) was proved by the survey data. It extracted 4 key factors namely functional service, interface design, responsiveness and security/privacy. Results showed that, compact and lively interface design, real-time interaction will satisfy the e-service users.

Milly et al. [12] introduced some quality evaluation and assessment methods to assess the quality of web pages. The proposed quality evaluation mechanisms were based on a set of quality criteria which were extracted from a targeted user survey. A weighted algorithmic interpretation of the most significant user quoted quality criteria is proposed. The set of quality criteria allowed implementing a web search engine with quality ranking schemes, leading to web crawlers which can crawl directly quality web pages.

John Gekas [8] presented the concept of web service ranking: a service rank is a quantitative metric that in some way shows the "importance" of a service within a web service network. Elizabeth Chang et al. [9] proposed CCCI (Correlation, Commitment, Clarity and Influence) methodology is an advanced trustworthiness measurement methodology that provides four metrics, and defines the maximum possible correlation value for a business service interaction.

Dong et al. [1] worked to provide a trustworthy and reliable technology for linking service providers and service requesters in the DE environment. Their case study reveals that, currently neither generic search engines nor local search engines can satisfy the advanced requirements of DE. Thus, they design the conceptual framework of a semantic service search engine. They build a focused crawler based-on the transport service ontology [3] derived from ECBR model and then in [1], the performance of it is compared with VSM, LSI and PM models. The proposed function of this search engine is to enable service requesters to retrieve and evaluate services published by service providers.

3. SERVICE ONTOLOGY

3.1 Ontology

According to Berners-Lee, "Ontologies will play an important role in achieving the vision of the Semantic Web". Ontology is a modeling tool for conceptual model, which describes the information system in terms of semantics and knowledge [17]. According to Guarino [19], an ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e., its ontological commitment to a particular conceptualization of the world. Contemporary ontologies share many structural similarities, regardless of the language in which they are expressed and most ontologies describe individuals (instances), classes (concepts), attributes, and relations.

Ontologies can be developed in many ways [18]. Ontologies can be directly converted into taxonomies and formal approach can be followed. A method for the automated transformation of product and service categories into an ontology is discussed by Hepp. However, Dogac et al., defined ontologies based on taxonomies have limited usefulness, as they do not capture the intricate interrelationships among the defined concepts.

A formal approach to ontology design and evaluation has been proposed by Uschold and Grüninger. This well-established ontology engineering method prescribes that a motivating application scenario should be first identified, in which the proposed ontology is expected to be applied. Based on this scenario, a comprehensive list of so called informal competency questions should be defined. These questions must be answered by the ontology and are used to establish the terminology.

Formal languages are necessary whenever an ontology is defined [18]. The Web Ontology Language (OWL) is a standard XML-based language for representing ontologies that has been widely accepted and supported by the research community. OWL builds on the Resource Description Framework (RDF) and RDF Schema Language. OWL comes with a larger vocabulary and stronger syntax than RDF and it is stronger language with greater machine interpretability. OWL is the most popular ontology description language, and is based on Description Logics, which is a family of logic-based knowledge representation formalisms.

OWL provides the following three increasingly expressive sub-languages designed for use by specific communities of implementers and users according to W3C:

- OWL Lite - supports users who primarily need a classification hierarchy and simple constraints.
- OWL DL - supports users who want the maximum expressiveness while retaining computational completeness and decidability. OWL DL includes all OWL language constructs, but they can be used only under certain restrictions.
- OWL Full - is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees.

3.2 Service Ontology

According to Dong et al [1], "The Service Ontology is defined as the conceptualization of the Service, which is identified by a Service Name, defined by a Service Description". They presented the Service Ontology as the combination of the ontology name and a tuple where the elements of the tuple can be complex elements which describe about service, and it is defined as follows.

Service [Service Name, Service Description]

where Service Name denotes the name that can be used to uniquely identify a service. Service Description denotes the meaningful descriptions of a service. The normal form of a service description is a set of words (noun, adjective, or adverb) and a service concept may have many service descriptions.

4. SYSTEM DESIGN

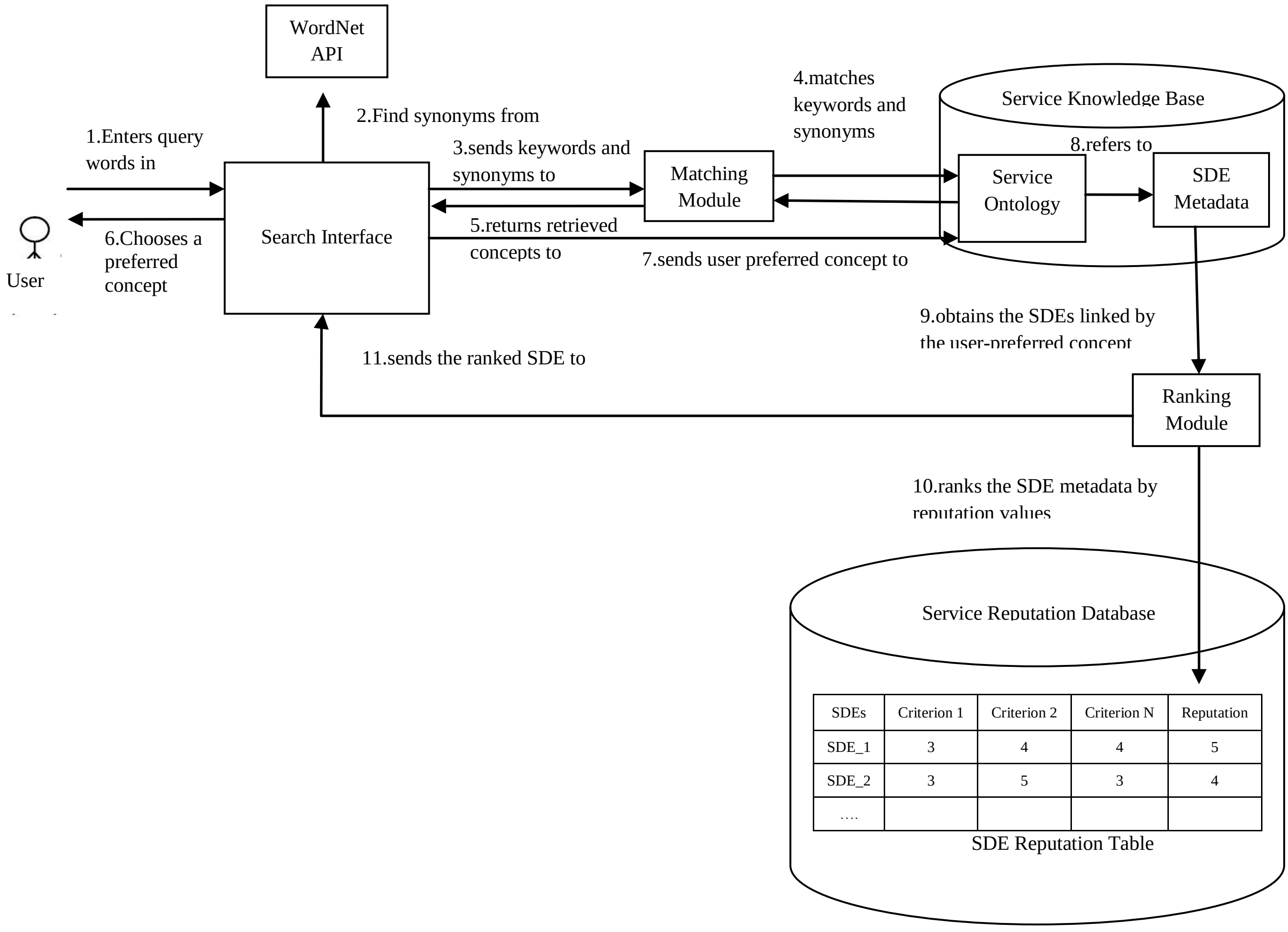
Semantic Service Search Engine in this paper is based on the work proposed in [1]. Hence the design is taken from that work. The system consists of four different modules: Service Knowledge base, service search module, service evaluation module and service reputation database. From those, service search module is chosen to perform evaluation using Support Vector Machine model.

From Fig 1, it can be noticed that a service requestor first enters a set of query terms into the search interface. The search interface sends each term to the Wordnet API and it returns synonyms for each and every term; otherwise, the query term will be filtered. Then it sends the query terms and their synonyms to the matching module. Matching module runs based on the algorithm i.e., SVM to compute the similarity values between the service ontology concepts stored in the service knowledge base and the query terms, its synonyms[16]. The concepts with higher similarity concepts will be returned to the search interface and then ranked accordingly to their similarity values. Then the user can choose choice of preference from the obtained concepts.

Finally quality of service evaluation of a service after the service provider provides the service requested can be done by service requestor by an authentication process. The metrics used for evaluation and ranking methodology is CCCI metrics [9]. Ranking module will store and retrieve the values from the service reputation database, [1] will provide further information about system design of semantic service search engine.

5. PROTOTYPE IMPLEMENTATION

The prototype of semantic service search engine is built with Java. Various API's are incorporated like WordNet API, JAWS (Java API for WordNet Searching), Protégé-OWL API, and MySQL API. As mentioned above, WordNet API will generate synonyms to a word given, and then JAWS is a tool which is useful to connect to the WordNet API on the Java platform. For the service knowledge base, Protégé-OWL API is used where a pizza ontology is built by taking the pizza service providers of Hyderabad into account. To build service reputation database and to store service provider's reputation values, MySQL API is used.



SDEs	Criterion 1	Criterion 2	Criterion N	Reputation
SDE_1	3	4	4	5
SDE_2	3	5	3	4
….				

Fig 1: Semantic Service Search Module

6. EXPERIMENTS AND RESULTS

6.1 Performance Indicators

To evaluate the semantic service search engine, six performance indicators from the field of IR (Information Retrieval) are used, which are Precision, Recall, Mean Average Precision, Harmonic Mean, Fallout Rate and Mean Reciprocal Rank. In [1], first five, i.e., precision, recall, mean average precision, harmonic mean and fallout rate have used and description also given, refer [6] to know in detail.

Mean Reciprocal Rank is a statistic for evaluating any process that produces a list of possible responses to a query, ordered by probability of correctness. The reciprocal rank of a query response is the multiplicative inverse of the rank of the first correct answer. The mean reciprocal rank is the average of the reciprocal ranks of results for a sample of queries Q [4].

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i} \qquad (2)$$

The reciprocal value of the mean reciprocal rank corresponds to the harmonic mean of the ranks.

6.2 Experiment and Results

The experiment is to evaluate the performance of semantic search engine on SVM in the information retrieval prospective using six performance indicators as mentioned above. The results are taken using peer-reviewed method to test the performance. The level threshold values ranges from 0 to 0.9 with an increment of 0.1.

The testing result of SVM model is shown in Table 1. The representations THR, PREC, MAP, REC, HM, FAL, and MRR refer to Threshold, Precision, Mean Average Precision, Recall, Harmonic Mean, Fallout Rate and Mean Reciprocal Rank respectively. To obtain average values, 15 queries are instantiated under pizza service field.

The two columns on the last page should be as close to equal length as possible.

Table 1: Testing Result

THR	PREC	MAP	REC	HM	FAL	MRR
>0	0.206	0.683	0.675	0.316	0.061	0.416
>0.1	0.211	0.691	0.669	0.321	0.057	0.421
>0.2	0.269	0.724	0.593	0.370	0.038	0.448
>0.3	0.368	0.752	0.527	0.433	0.025	0.492
>0.4	0.472	0.791	0.435	0.453	0.019	0.549
>0.5	0.599	0.833	0.371	0.458	0.012	0.635
>0.6	0.685	0.873	0.302	0.419	0.009	0.699
>0.7	0.761	0.892	0.268	0.396	0.005	0.731
>0.8	0.783	0.897	0.241	0.369	0.002	0.752
>0.9	0.801	0.904	0.215	0.339	0.002	0.762

To get optimum threshold value, the Harmonic Mean can be used, as it is dependent on two primary performance indicators: precision and recall [1]. Here, the highest Harmonic Mean value is 0.458 for at a threshold value of 0.5. Hence, it can be said that the optimum threshold value for this system can be assessed as 50% for the Harmonic Mean of 45.8% where Precision and Recall are 59.9% and 37.1% respectively.

Figure 2 gives a significant graph representation for the testing results obtained in Table 1. Further, the testing process of this model can be evaluated by comparing this with other techniques. Thus it is added to the future work.

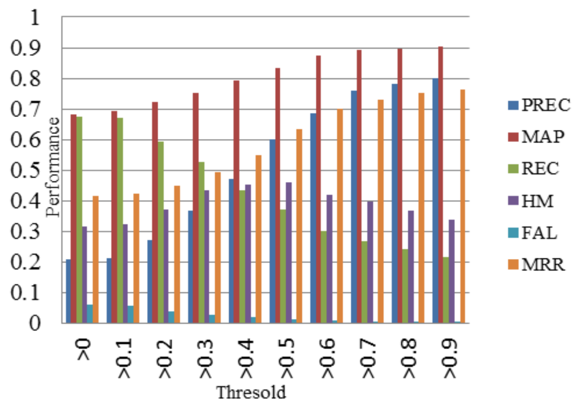


Figure 2 : Testing Result by Graph

7. CONCLUSION AND FUTURE WORK

Semantic service search engine [6] is a conceptual and experimental framework to overcome the disadvantage of traditional search environment. There is much need to rely on the semantic service-oriented search environment, where quality of service (QoS) can be provided. The system provides an interface where service requestor can get information about service provider and its QoS, further service requester can evaluate the performance of the service provider and it will be evaluated by CCCI metrics.

This paper is implemented to perform evaluation on search environment of semantic service search engine using six performance indicators of information retrieval. They are Precision, Recall, Mean average precision, Harmonic mean, Fallout rate and Mean Reciprocal Rank. Thus the testing is done using SVM model and results are evaluated under peer-reviewed method. Further, its effectiveness can be obtained by comparing these results by applying other data mining techniques on semantic service search module.

Future work includes implementing few more data mining techniques on search environment to compare present system and optimizing the framework to show better performance.

8. REFERENCES

[1] Hai Dong, Farooq Khadeer Hussain, Elizabeth Chang, “A service search engine for the industrial digital ecosystem”, IEEE Trans. Ind. Electron., vol.58, no.6, Jun 2011.

[2] Jingyu Sun, Xueli Yu, Ning Zhong, “Anatomy of a collaborative search engine”, 2nd International workshop on social reccomender systems in conjuction with CSCW 2011”, March 19-23, 2011.

[3] Hai Dong, Farooq Khadeer Hussain, Elizabeth Chang, “A transport service ontology-based focused crawler”, Fourth International Conference on Semantics, Knowledge and Grid, IEEE, 2008.

[4] E.M. Voorhees, "Proceedings of the 8th Text Retrieval Conference", pp. 77–82, 1999.

[5] Gabor Nagypal, "Improving information retrieval effectiveness by using domain knowledge stored in ontologies".

[6] R. Baeza-Yates and B. Ribeiro-Neto, "Modern Information Retrieval", New York: Addison-Wesley, 1999.

[7] Larry Kerschberg, Hanjo Jeong, Yong Uk Song, and Wooju Kim, "A Case-Based Framework for Collaborative Semantic Search in Knowledge Sifter," ICCBR 2007, LNAI 4626, pp. 16–30, Springer-Verlag Berlin Heidelberg 2007.

[8] John Gekas, "Web service ranking in service networks", 3rd European Semantic Web Conference,11-14th Jun, 2006.

[9] Elizabeth Chang, Farookh Khadeer Hussain, Tharam Dillon, "CCCI metrics for the measurement of quality of e-service", 2005 IEEE/WIC/ACM International Conference in Intelligent Agent Technology(IAT'05), 2005.

[10] Kevin S.Mc Curley and Andrew Tomkins, "Mining and knowledge discovery from the web", 7th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'04), IEEE, 2004.

[11] Minxue Huang, Xiaoling Li, Xun Zeng, "Service qulaity in Web 2.0 electronic commerce: An integrative perspective from typical service and technique adoption", IEEE, 2005.

[12] Milly Kc, Markus Hagenbuchner, Ah Chung Tsoi, "Quality information retrieval for the world wide web", 2008 IEEE/WIC/ACM International Conference on Web Intelligent Agent Technology, IEEE, 2008.

[13] Huaqiao Lv, Peng Li, Lugang Guo, "Research on the applicationof personalization search engines in information service industry", 2010 International Conference on Educational and Information Technology(ICEIT 2010), IEEE, 2010.

[14] Indu Chawla, "An overview of personalization in web search", IEEE, 2011.

[15] Sadegh Kharazmi, Ali Farahmand Nejad, Hassan Abolhassani, "Freshness of web search engines: Improving performance of web search engines using Data Mining techniques", IEEE, 2009.

[16] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: a Library for Support Vector Machines".

[17] Qizhi Qiu and Qianxing Xiong, "An Ontology for Semantic Web Services," Springer-Verlag Berlin Heidelberg 2007.

[18] Vikram Sorathia, Luís Ferreira Pires, Marten van Sinderen, "An Analysis of Service Ontologies," Pacific Asia Journal of the Association for Information Systems, 2010.

[19] Guarino, N, "Formal Ontology and Information Systems," In: Proceedings of FOIS 1998.