A Framework Enabling Semantic Search in Health Service Ecosystems

Hai Dong[1], Farookh Khadeer Hussain[2], Elizabeth Chang[3]

Digital Ecosystems and Business Intelligence Institute, Curtin University of Technology
GPO Box U1987, Perth, WA 6845, Australia
[1]hai.dong@cbs.curtin.edu.au
[2]farookh.hussain@cbs.curtin.edu.au
[3]elizabeth.chang@cbs.curtin.edu.au

Abstract— The health service ecosystem is a subdomain of the service ecosystem, which is an ecological view of the health service domain. Due to the importance and the speciality of health services, service consumers have rigorous requirements regarding the efficiency of health service search engines. However, as most health service consumers do not have enough domain knowledge, their service queries are sometimes not able to denote their actual service requests. Moreover, we find that the existing health service search engines lack the ability to help the consumers to disambiguate their service queries, which may impede the efficiency of the search engines. In this paper, by means of semantic search technologies, we design a framework enabling user query disambiguation in the health service ecosystem. The framework embodies a health service ontology for domain knowledge-based user query disambiguation and an ECBR algorithm for accurate service retrieval. In order to evaluate the framework, we build a system prototype and perform a series of experiments on it. Conclusions from the evaluation are drawn in the paper.

I. INTRODUCTION

A service ecosystem is an ecological view of the service domain in our business world [1, 2]. From this perspective, the service domain is cognized as an ecosystem, in which the objects involved in the service activities are species of the ecosystem and the services are the resources on which that species can live. In this service ecosystem, obviously the species is comprised of individuals and organizations, and the services contain all actual services in the service domain. Since the species are providers and consumers of the services, the features of the services reflect the features of the species. Consequently, we can deduce that the services in the service ecosystem are diverse and geographically dispersed owing to the commonly shared features of the species. Moreover, the service ecosystem can be divided into many subdomains. As one of its subdomains, the health service ecosystem inherits similar features from the service ecosystem. Because of the diversity and the geographical distribution of the health services, a species needing a service may not have the knowledge regarding the service request and thus might not find the species who can provide the service in a timely manner. Considering that time is crucial to health services, the health service ecosystem requires a reliable solution to establish an efficient and time-conscious link between the diverse and distributed species, in order to facilitate the species' survival.

Semantic search is an innovative technological field for disambiguating queries and enhancing search performance [3], which is able to assist the species who do not have the knowledge about a service search to more accurately denote their search intentions. Hence, there is an urgent need for a semantic service search engine within the health ecosystem. Nevertheless, having surveyed the current commercial health service search engines, we find that most of them lack semantic supports, resulting in an inability to disambiguate user queries. Hence, the current commercial health service search engines cannot fulfil the above requirement. In this paper, we present a framework that facilitates semantic search within the health service ecosystem. This framework makes use of semantic search technologies for the retrieval of information from the health service ecosystem.

The rest of the paper is organized as follows: in Section 2, we conduct a general survey of the current semantic search research field and find the common issues in this field; in Section 3, we present the framework of the proposed solution for solving the search issue in the health service ecosystem; in Section 4, we introduce a health service metadata schema for standardizing health service information over the Internet and a health service ontology for modelling the domain knowledge of the health service ecosystem; in Section 5, we provide a mathematical model for the semantic search; in Section 6, we implement a series of experiments in order to validate the proposed framework; conclusions are drawn in the final section.

II. RELATED WORKS

In this section, we briefly introduce Semantic Web and ontology as well as review semantic search surveys in the existing literature.

A. Semantic Web and Ontology

Semantic web is an ongoing project, the aim of which is to provide meaning to the data in the web. By means of this project, the web is able to understand people's or machines' requests to use web contents [4]. To realize this objective, many semantic web language standards and semantic web

development tools have been developed, such as OWL (Web Ontology Language), RDF (Resource Description Framework), RDFS (Resource Description Framework Schema), SPARQL (SPARQL Protocol and RDF Query Language), Protégé and so forth. Ontology is a form of domain knowledge representation [5], which comprises concepts, properties of concepts and relations between concepts. Ontology is the backbone of the semantic web, being an efficient tool for representing knowledge and for knowledge sharing between people-people, people-machine and machine-machine.

B. Semantic Search

In 2003, Guha et al. [3] provided a new terminology – "*semantic search*" for the field of web search, which is distinct from the traditional information retrieval methods. The purpose of the semantic search is to assist users to denote their search intentions and assist search engines to understand the meaning of users' queries in terms of semantic web technologies. Since then, much research has been undertaken in the area of semantic search. Moreover, the literature indicates that three groups of researchers have conducted comprehensive surveys of the existing semantic search technologies.

In 2004, Hildebrand et al. [6] conducted a survey of 32 semantic search applications. They divide these semantic search applications into the phases of query construction, search algorithm and result presentation, and analyse the three phases from the perspective of functionality and interfaces. From the survey, they concluded that there are three drawbacks to the current semantic search applications: lack of evaluation of semantic search algorithms, lack of user evaluation of the interfaces and lack of APIs and middleware support.

In 2007, Mangold [7] conducted a survey on existing semantic search technologies within a limited scope. The semantic search approaches: 1) working on XML documents, 2) requiring user query formulation, 3) working on particular network environments, are excluded from this survey. Twenty-two semantic search approaches were selected by means of criteria which were established from the perspectives of architecture, coupling, transparency, user context, query modification, ontology structure and ontology technology. The author divided the 22 semantic search approaches in terms of their technological features, which are Simple HTML Ontology Extension (SHOE) [8-11], Inquirus2 [12], TAP [3], hybrid spreading activation [13, 14], Intelligent Semantic Web Retrieval Agent (ISRA) [15], librarian agent [16], Semantic Content Organization and Retrieval Engine (SCORE) [17], TRUST [18], audio data retrieval [19, 20], and Ontogator [21, 22]. The conclusions educed from the survey include: ignorance of analysis of query modification techniques, ignorance of incorporating semantic search engines, ignorance of analysis of user acceptance, adaptability of ontologies in semantic search engines, ignorance of ontology-based document ranking, ignorance of integration with Document or Content Management System (DMS/CMS), and lack of performance or scalability evaluation.

In 2008, Dong et al. [23] conducted a survey on current semantic search technologies, which embody 27 research achievements. We classified the research achievements by means of product types, technological features and retrieved objects. Six categories of semantic search technologies were identified, which include semantic search engines, semantic search technologies, hybrid semantic (keyword-based enhanced by semantic) search engines, XML (files) search engines, ontology (files) search engines and semantic multimedia (files) search engines. Four primary drawbacks were inferred from the survey: ignorance of ontological differences between designers' and users' perceptions, ignorance of evolving knowledge structure, low precision and high recall, and lack of evaluations.

These conclusions deduced from the above surveys indicate that the common limitation of the existing semantic search technologies is the lack of evaluations. In this research, we attempt to resolve this issue, namely to evaluate the proposed semantic search algorithm in detail.

III. PROPOSED SOLUTION

In this section, we propose a conceptual framework for a Health Service Semantic Search Engine (HSSSE), in order to disambiguate user queries and enhance search performance in the health service ecosystem.

The HSSSE framework is extended from our previous work in the transport service domain [24, 25]. The difference between the HSSSE framework and the previous work is that we modify some parts of the previous framework, including the knowledge base and functions of the interface, so that it can adapt to the health service domain.

The system architecture of the HSSSE is displayed in Fig. 1. The HSSSE primarily comprises three parts, which are described as follows:

- **Semantic search engine.** It is responsible for interacting with users for query disambiguation and service information retrieval by a series of human computer interactions (HCIs). In addition, a WordNet API is utilized for query filtering and expansion.
- **Health service knowledge base.** It is designed with the purpose of providing specific health service domain knowledge for query disambiguation. The service knowledge is stored in the form of a health service ontology, which is a hierarchy of health service concepts. The health service ontology will be discussed in Section 4.
- **Health service metadata base.** It stores the information regarding health service advertisements and health service provider profiles. The service information is abstracted from web documents over the Internet, and standardized to service metadata according to a unified health service metadata schema. A service metadata is annotated by concepts of the ontology

stored in the health service knowledge base, by associating it with one or more relevant concepts. The health service metadata schema will be discussed in Section 4.
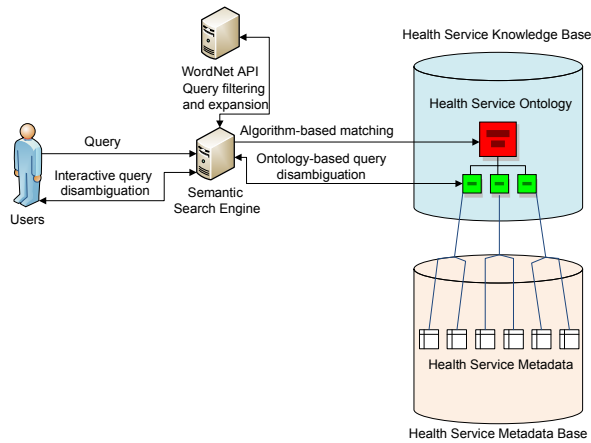

Fig. 1. System architecture of the Health Service Semantic Search Engine

The workflow of the HSSSE system is depicted as follows:

1. Once a user sends a query to the semantic search engine, the query words are sent to the WordNet API for query filtering and query expansion. As a result, the keywords in the query are preserved when they can be found from the WordNet, and the synonyms of the keywords are retrieved from the API.
2. The semantic search engine computes the similarity values between the expanded query and each concept in the health service ontology by means of a mathematical model. The matched concepts are then displayed to the user and ranked according to their similarity values. The mathematical model will be described in Section 5.
3. The search module employs the following HCI approach for query disambiguation:

   a. The user can choose the preferred health service concept from the retrieved concepts by means of a view-based approach.
   b. Once the user clicks a concept, if the concept is a parent class in the health service ontology, all its children classes are retrieved from the health service knowledge base and displayed to the user as well as ranked according to their similarity values against the query.
   c. The user can then choose the preferred health service concept from the retrieved view of concepts. Step 3a to Step 3c is a recursive process unless the user chooses a bottom-level concept from the health service ontology.
   d. Once the user chooses a bottom-level concept from the health service ontology, the health service metadata annotated by the health service concept are retrieved from the health service metadata base and ranked based upon their domain-specific quality values based on an extended CCCI metrics methodology. Further details regarding the CCCI metrics can be referenced from [26].

The major feature of the semantic search approach is that it provides a restricted view-based approach for users to retrieve the appropriate health service concepts to denote their queries. There are two main advantages of this approach, which are described as follows:

1. This approach does not purely match concepts with a query based on their similarity values with the query. On the contrary, it displays all the children classes of a matched concept to the user, and does not omit concepts that could be relevant to the query. Moreover, this approach allows users' subjective perceptions towards their queries to work in the query disambiguation process, which potentially enhances user satisfaction towards the concept selection, since users play the most decisive roles in this process.
2. The concepts that comprise a view are selected based on the similarity values between the concepts and a query, which ignores concepts irrelevant to the query, reduces the scope of user viewing, and facilitates the view-based search. This eliminates the limitation of the view-based search whereby views with large-scale options could bring users obvious inconvenience, when considering the time cost factor.

In the next section, we will introduce the unified health service metadata schema employed to standardize health service information and the health service ontology that specially represents health service domain knowledge.

IV. HEALTH SERVICE METADATA SCHEMA AND HEALTH SERVICE ONTOLOGY

A. Health Service Metadata Schema

As introduced in Section 3, before a health service advertisement or a health service provider profile can be retrieved from the Internet, it must be standardized in terms of a unified health service metadata schema, which allows both users and computers to easily understand the meaning of the service information. Consequently, we designed a semantic-focused crawler to standardize health service information. Here, we introduce only the unified health service metadata schema; further details regarding the semantic focused crawler can be referenced from [27, 28].

There are two types of schemas involved in health service metadata schema, which are Health Service Description Entity (HSDE) schema and Health Service Provider (HSP) schema [27].

The HSDE schema is used to build a health service metadata that describes a health service provided by a service provider, which is defined by the following properties:

- **healthServiceName.** This property provides the name of a health service provided by a service provider, e.g. massage therapy.
- **serviceDescription**. The *serviceDescription* stores the detailed description of a health service. This property can be an arbitrary amount, which depends on the number of information snippets describing a health service entity.
- **linkedConcepts.** This property is used to store URIs of relevant bottom-level health service ontology concepts in order to realize the annotation process.
- **provider.** This property is used to reference the relevant HSP metadata by storing their URIs.

The HSDE schema in RDFS is shown below:

```
<rdfs:Class rdf:about="& kb;Health_Service_Description_Entity"
      rdfs:comment="Health_Service_Description_Entity"
      rdfs:label="Health_Service_Description_Entity">
      <rdfs:subClassOf rdf:resource="& rdfs;Resource"/>
         <rdf:Property rdf:about="& kb;healthServiceName"
               rdfs:label="healthServiceName">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Description_Entity"/>
               <rdfs:range rdf:resource="& rdfs;Literal"/>
         </rdf:Property>
         <rdf:Property rdf:about="& kb;serviceDescription_1"
               rdfs:label="serviceDescription">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Description_Entity"/>
               <rdfs:range rdf:resource="& rdfs;Literal"/>
         </rdf:Property>
         …
         <rdf:Property rdf:about="& kb;serviceDescription_#"
               rdfs:label="serviceDescription">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Description_Entity"/>
               <rdfs:range rdf:resource="& rdfs;Literal"/>
         </rdf:Property>
         <rdf:Property rdf:about="& kb;linkedConcepts"
               rdfs:label="linkedConcepts">
               <rdfs:range rdf:resource="&
         kb;Health_Service"/>
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Description_Entity"/>
         </rdf:Property>
         <rdf:Property rdf:about="& kb;provider"
               rdfs:label="provider">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Description_Entity"/>
               <rdfs:range rdf:resource="&
         kb;Health_Service_Provider"/>
         </rdf:Property>
</rdfs:Class>
```

The HSP schema is used to build an HSP metadata that describes a health service provider, which consists of the following properties:

- **providerName.** This property is used to store the name of a health service provider, e.g. the name of a hospital.
- **providerProfile.** This property is used to store the descriptive information about the profile of a health service provider.
- **address.** This property is used to store the address information of a health service provider.
- **contactDetails.** This property is used to store the contact information of a health service provider, including phone number, fax number, URL of website, email etc.
- **services.** *services* is the inverse property of the *provider* property of the HSDE schema, which is used to store the URI(s) of the HSDE metadata relevant to a HSP metadata.

The HSP schema in RDFS is displayed below:

```
<rdfs:Class rdf:about="& kb;Health_Service_Provider"
      rdfs:comment="Health_Service_Provider"
      rdfs:label="Health_Service_Provider">
      <rdfs:subClassOf rdf:resource="& rdfs;Resource"/>
         <rdf:Property rdf:about="& kb;providerName"
               rdfs:label="providerName">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Provider"/>
               <rdfs:range rdf:resource="& rdfs;Literal"/>
         </rdf:Property>
         <rdf:Property rdf:about="& kb;providerProfile"
               rdfs:label="providerProfile">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Provider"/>
               <rdfs:range rdf:resource="& rdfs;Literal"/>
         </rdf:Property>
         <rdf:Property rdf:about="& kb;address"
               rdfs:label="address">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Provider"/>
               <rdfs:range rdf:resource="& rdfs;Literal"/>
         </rdf:Property>
         <rdf:Property rdf:about="& kb;contactDetails"
               rdfs:label="contactDetails">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Provider"/>
               <rdfs:range rdf:resource="& rdfs;Literal"/>
         </rdf:Property>
         <rdf:Property rdf:about="& kb;services"
               rdfs:label="services">
               <rdfs:domain rdf:resource="&
         kb;Health_Service_Provider"/>
               <rdfs:range rdf:resource="&
               kb;Health_Service_Description_Entity"/>
         </rdf:Property>
</rdfs:Class>
```

B. Health Service Ontology

The health ontology is a four-tier hierarchical structure with 218 health service concepts, in which the lower level concepts are the specification of the upper level concepts. The root concept defines the boundary of the general health service domain and the properties of a generic health service. The second tier concepts define the six health service subdomains, which are allied health service, dentistry service, medicine service, nursing service, pathological and clinical laboratory service, and hospital and clinic service. The third and fourth tier concepts are the further specification of the subdomain concepts, which inherit the properties from its superclass and also have own domain-specific properties. It is noted that the health service ontology is designed by referring to the relevant knowledge from Wikipedia (http://en.wikipedia.org/wiki/), Open directory Project (ODP) (http://www.dmoz.org), and hundreds of international health service websites. Further details regarding the health service ontology can be referenced from [27].

V. QUERY-CONCEPT MATCHING ALGORITHM

In order to realize the function of computing similarity values between ontological concepts and queries, we design an Extended Case-based Reasoning (ECBR) algorithm. The principle of the ECBR algorithm is to compare a group of query terms and their synonyms with the *serviceDescription* property values of a service concept. For a *serviceDescription* property, if a query term is contained in it, a value of 1 will be awarded; if a synonym of a query term is contained in it, a value of 0.5 will be awarded. After the comparison process, the sum of the values for the comparison of all query terms will be normalized by the length of the terms in the *serviceDescription* property; thus, its value should be between 0 and 1. Since one concept may have more than one *serviceDescription* property, the maximum value among them is the similarity value between the concept and the group of query terms. The ECBR algorithm for computing the similarity value between a concept C and a query Q is mathematically shown below:

$$sim(C,Q) = \max_{SD_i \in C} \left(\sum_{t_{ih} \in SD_i} \frac{match(t_{ih},Q) + match_s(t_{ih},S)}{l_{SD_i}} \right) \quad (1)$$

with

$$match(t_{ih},Q) = \begin{cases} 1 & if\ \exists q_t\ |(t_{ih} = q_t) \wedge (q_t \in Q) \\ 0 & otherwise \end{cases} \quad (2)$$

$$match_s(t_{ih},S) = \begin{cases} 0.5 & if\ \exists s_k\ |\ (t_{ih} = s_k) \wedge (s_k \in S) \\ 0 & otherwise \end{cases} \quad (3)$$

where SD_i is a *serviceDescription* property of the concept C, t_{ih} is a term that occurs within SD_i, and l_{SDi} is the frequency of all terms appearing in SD_i; q_t is a term that occurs within the query Q, and S are synonyms of Q, which consists of a group of terms s_k.

The ECBR model is very simple to implement, and it does not need to generate index terms before matching, which saves pre-processing time. It can also adapt to the flexibility of concepts that often needs to regenerate index terms in most of the index term-based algorithms. Since the algorithm is independent of index terms, it does not have the issue of index term dependency. More in-depth information can be referenced from [29, 30].

After obtaining the similarity value between a query and a concept, a threshold value needs to be configured in order to prevent the phenomenon of query flooding, and to filter irrelevant concepts for a query, as well as to select a limited number of concepts to compose a view for users. Therefore, in the next section, we will implement a series of experiments for the optimal threshold value selection.

VI. SYSTEM EVALUATION

In this section, we empirically evaluate the ECBR model employed in the HSSSE system. The goals of this experiment are: (1) to find the proper threshold value for the ECBR model; and (2) to evaluate the feasibility of the ECBR model in the HSSSE.

A. Performance Indicators

In order to find the proper threshold value and evaluate our ECBR model, three performance indicators from the field of information retrieval are employed, which are precision, recall, and harmonic mean.

Precision in the information retrieval is used to measure the preciseness of a retrieval system. In this experiment, precision is the proportion of retrieved and relevant concepts in all retrieved concepts for a query, which are represented below as:

$$Precision = \frac{number\ of\ retrieved\ and\ relevant\ concepts}{number\ of\ retrieved\ concepts} \quad (4)$$

Recall in the information retrieval refers to the measure of effectiveness of a query system. In this experiment, recall is the proportion of retrieved and relevant concepts in all relevant concepts for a query, which can be represented as:

$$Recall = \frac{number\ of\ retrieved\ and\ relevant\ concepts}{number\ of\ relevant\ concepts} \quad (5)$$

It is important to note that the number of relevant concepts can be determined only by a peer-reviewed method, as the estimation of relevance between concepts and queries requires detailed knowledge of all concepts in the collection, which can only be manually implemented in the current situation.

Harmonic mean (F-measure) in the information retrieval is used as an aggregated performance scale for the search engine. In this experiment, harmonic mean is the mean of precision and recall, which can be represented as:

$$\text{Harmonic Mean} = \cfrac{1}{\cfrac{1}{\text{Precision}} + \cfrac{1}{\text{Recall}}} \qquad (6)$$

When the harmonic mean value reaches the highest, it means the aggregated value between precision and recall reach to the highest simultaneously.

B. Experiments

We configure an experiment to evaluate the performance of our ECBR model in the HSSSE system from the perspective of information retrieval, In order to obtain the most precise statistical data, we instantiate 100 queries which cover almost every domain in the health service field. Therefore, all the indicators' results will be averaged by 100. We will use the peer-reviewed method to test the performance of our ECBR model on the three indicators, by the level of different threshold values that range from 0 to 1.0 with an increment of 0.05. By means of the harmonic mean that leverages the two indicators – precision and recall, the optimal threshold value for the ECBR model can be determined. This task can realize the first and second goal of our experiment.

Fig. 2 displays the performance of the ECBR model on precision for different threshold values. It is found that the overall trend of the ECBR on precision is ascending along with the augmentation of the threshold value. This is because the higher threshold values can filter more irrelevant concepts to a query. It can be obviously observed that Threshold Value 0.5 is a watershed in this curve, due to a more than 70% raise of the precision value at this point. The precision value reaches the peak and remains stable after the threshold value reaches 0.75 or higher.
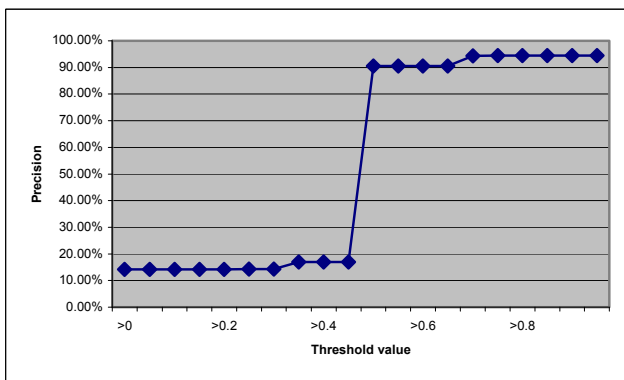

Fig. 2. Precision of the ECBR model @ different threshold values in the HSSSE system

Fig. 3 shows the performance of the ECBR model on recall for different threshold values. It is observed that the overall trend of the ECBR model experiences a gradual fall with the increase of the threshold value. The reason for this is that higher threshold values can reduce the number of retrieved and relevant concepts. Similar to its trend on precision, Threshold Value 0.5 is a divide in the curve of the ECBR on

recall, where there is more than a 4% fall at this point compared with the previous point. The descending trend stops when the threshold value is higher than 0.7.
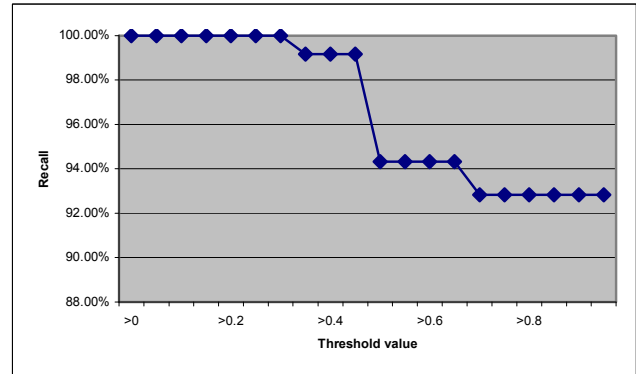

Fig. 3. Recall of the ECBR model @ different threshold values in the HSSSE system

Fig. 4 depicts the performance of the ECBR model on harmonic mean for different threshold values. As the aggregation of precision and recall, the trend of the ECBR on harmonic mean is jointly affected by the trends on precision and recall. It can be seen that the overall trend of the ECBR on harmonic mean is analogous to its trend on precision, as a result of that the overall variation interval of the ECBR on recall is relatively smaller (7.17%) compared with its overall variation interval on precision (80.36%). Therefore, the impact of the precision trend on the harmonic mean trend appears more obvious than the recall trend. Therefore, analogously, Threshold Value 0.5 is a milestone for the trend of the ECBR model on harmonic mean, and the harmonic mean values reaches the top when the threshold value is higher than 0.75.
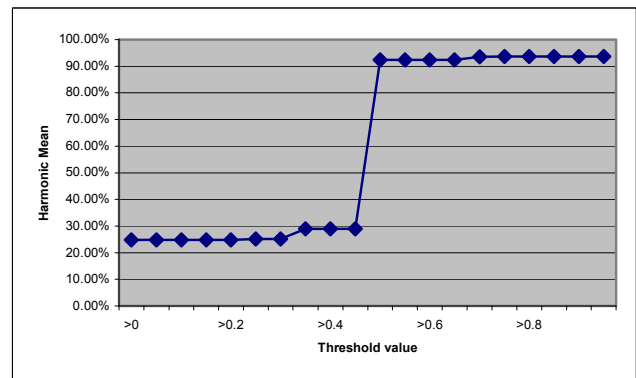

Fig. 4. Harmonic mean of the ECBR model @ different threshold values in the HSSSE system

We draw two conclusions to this experiment, which can be described as follows:

- With regard to the first goal involved in this experiment, we employ the aggregated metrics – harmonic mean as the benchmark to determine the optimal threshold value for the ECBR model. It is

observed that the harmonic mean value reaches the top and remains stable when the threshold value is higher than 0.75. Consequently, 0.75 is an optimal threshold value for the ECBR model in this experiment.

- With regard to the second goal, we judge the validity of the ECBR model based on its performance at its optimal threshold value. At the 0.75 point, its precision value is 94.50%, its recall value is 92.83% and its harmonic mean value is 93.66%, which can be regarded as an outstanding performance for the ECBR model from the perspective of information retrieval. Thus, we can conclude that the feasibility of the ECBR model is proven by this experiment.

VII. CONCLUSION AND FUTURE WORK

In this paper, in order to address the troublesome situation of the service search engines, and help users to disambiguate their service queries and enhance searching precision in the health service ecosystem, we proposed a HSSSE framework by integrating semantic search technologies into the health service search process. We designed a restricted view-based search module, by selecting relevant ontological concepts to compose a limited view from which users can choose in order to denote their service queries. An ECBR algorithm is employed for determining relevant concepts for constructing a view for a query. Through our experiment, we primarily prove the feasibility of this framework in the health service ecosystem.

Our future work will focus on the following areas:

- We will adopt the index term strategy to enhance the efficiency of our ECBR model.
- We will compare our model with other typical information retrieval models to further prove its feasibility.
- We will compare our framework with other similar products in this field to further prove its feasibility.

REFERENCES

[1] R. Veryard, *Plug and Play: Towards the Component-Based Business*. London: Springer, 2000.

[2] H. Zhuge and X. Shi, "Toward the Eco-grid: A harmoniously evolved interconnection environment," *Communications of the ACM*, vol. 47, pp. 78-83, 2004.

[3] R. Guha, R. McCool, and E. Miller, "Semantic search," in *WWW2003*, Budapest, Hungary, 2003.

[4] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," in *Scientific American Magazine*. NY: Scientific American, Inc., 2001.

[5] T. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, pp. 199-220, 1995.

[6] M. Hildebrand, J. v. Ossenbruggen, and L. Hardman, "An analysis of search-based user interaction on the Semantic Web," S*tichting Centrum voor Wiskunde en Informatica*, Amsterdam 2007.

[7] C. Mangold, "A survey and classification of semantic search approaches," *International Journal of Metadata, Semantics and Ontologies*, vol. 2, pp. 23-34, 2007.

[8] J. Davies and R. Weeks, "QuizRDF: search technology for the semantic web," in *the 37th Hawaii International Conference on System Sciences (HICSS 2004)*, Big Island, HI, USA, 2004, p. 40112.

[9] E. García and M.-Á. Sicilia, "Designing ontology-based interactive information retrieval interfaces," in *Workshop on Human Computer Interface for Semantic Web and Web Applications (HCI-SWWA)*, New York, USA, 2003, pp. 152-165.

[10] J. Heflin and J. Hendler, "Searching the Web with SHOE," in *2000 AAAI Workshop*, Menlo Park, CA, USA, 2000, pp. 35–40.

[11] S. Luke, L. Spector, and D. Rager, "Ontology-Based Knowledge Discovery on the World-Wide Web," in *AAAI 96 workshop*, Portland, Oregon, USA, 1996, pp. 96-102.

[12] E. J. Glover, S. Lawrence, M. D. Gordon, W. P. Birmingham, and C. L. Giles, "Web search - your way," *Communications of the ACM*, vol. 44, pp. 97-102, 2001.

[13] F. Crestani and P. L. Lee, "WebSCSA: web search by constrained spreading activation," in *IEEE ADL 99 - Advances in Digital Libraries Conference*, Baltimore, Maryland, USA, 1999, pp. 163-170.

[14] C. Rocha, D. Schwabe, and M. P. Aragao, "A hybrid approach for searching in the semantic web," in *WWW '04*, New York, USA, 2004, pp. 374-383.

[15] A. Burton-Jones, V. C. Storey, V. Sugumaran, and S. Purao, "A heuristic-based methodology for semantic augmentation of user queries on the web," in *the 22nd International Conference on Conceptual Modeling - ER 2003*, Chicago, IL, USA, 2003, pp. 476-489.

[16] N. Stojanovic, "On analysing query ambiguity for query refinement: the librarian agent approach," in *22nd International Conference on Conceptual Modeling - ER 2003*, Chicago, IL, USA, 2003, pp. 490-505.

[17] A. Sheth, C. Bertram, D. Avant, B. Hammond, K. Kochut, and Y. Warke, "Managing semantic content for the web," *IEEE Internet Computing*, vol. 6, pp. 80-87, 2002.

[18] C. Amaral, D. Laurent, A. Martins, A. Mendes, and C. Pinto, "Design and implementation of a semantic search engine for Portuguese," in *the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal, 2004, pp. 247-250.

[19] L. Khan, D. McLeod, and E. H. Hovy, "Retrieval effectiveness of an ontology-based model for information selection," *The VLDB Journal - The International Journal on Very Large Data Bases*, vol. 13, pp. 71-85, 2004.

[20] Y. Labrou and T. Finin, "Yahoo! as an ontology: using Yahoo! categories to describe documents," in *the Eighth International Conference on Information and Knowledge Management (CIKM' 99)*, New York, NY, USA, 1999, pp. 180-187.

[21] E. Hyvönen, S. Saarela, and K. Viljanen, "Ontogator: Combining view- and ontology-based search with semantic browsin," in *XML 2003*, Kuopio, Finland, 2003.

[22] E. Mäkelä, E. Hyvönen, and S. Saarela1, "Ontogator — A semantic view-based search engine service for Web applications," in *The Semantic Web - ISWC 2006*, Athens, GA, USA, 2006, pp. 847-860.

[23] H. Dong, F. K. Hussain, and E. Chang, "A survey in semantic search technologies," in *the 2008 2nd IEEE International Conference on Digital Ecosystem and Technologies (DEST 2008)*, Phitsanulok, Thailand, 2008, pp. 403-408.

[24] H. Dong, F. K. Hussain, and E. Chang, "A service search engine for the industrial digital ecosystems," *IEEE Transactions on Industrial Electronics*, Accepted. DOI: 10.1109/TIE.2009.2031186

[25] H. Dong, F. K. Hussain, and E. Chang, "A human-centered semantic service platform for the digital ecosystems environment," *World Wide Web*, vol. 13, pp. 75-103, 2010.

[26] H. Dong, F. K. Hussain, and E. Chang, "A QoS-based service retrieval methodology for digital ecosystems," *International Journal of Web and Grid Services*, vol. 5, pp. 261-283, 2009.

[27] H. Dong, F. K. Hussain, and E. Chang, "A framework for discovering and classifying ubiquitous services in digital health ecosystems," *Journal of Computer and System Sciences*, accepted. DOI: 10.1016/j.jcss.2010.02.009

[28] H. Dong, F. K. Hussain, and E. Chang, "Focused crawling for automatic service discovery, annotation and classification in industrial digital ecosystems," *IEEE Transactions on Industrial Electronics*, accepted. DOI: 10.1109/TIE.2010.2050754

[29] H. Dong, F. K. Hussain, and E. Chang, "A transport service ontology-based focused crawler," in T*he 4th international conference on semantics, knowledge and grid (SKG '08)*, Beijing, 2008, pp. 49-56.

[30] H. Dong, F. K. Hussain, and E. Chang, "An ontology-based webpage classification approach for the knowledge grid environment," in *The Fifth International Conference on Semantics, Knowledge and Grid (SKG '09)*, Zhuhai, China, 2009, pp. 120-127.