INTELLIGENT INFORMATION RETRIEVAL THROUGH SEMANTIC WEB SERVICE DISCOVERY METHODS

Tauqeer Ahmad Usmani\textsuperscript{1} and Durgesh Pant\textsuperscript{2}

\textbf{Abstract}: Conventional Information retrieval system is lack a uniform semantic description for information, so it is hard for users to find more relevant information and to realize the semantic share in information resource. There are a lot of challenging tasks for information retrieve, such as how to make the managed resource have a machine understandable meaning so as to find what users really need and how to realize the semantic searching by means of the domain knowledge. The retrieval system is intelligent, which can not only improve the recall rate and precision rate, but also allows users to search information with natural language, so it provides more suitable and more perfect search service for users. In this paper, we provide a review of current Semantic Web Services (SWS) discovery approaches.

\textbf{Keywords}: Semantic web, Precision, Recall, Intelligent Information retrieval, Ontologies.

1. INTRODUCTION

The conventional information retrieval techniques are based on the keyword matching and full text search technology, mainly by means of catalogue, indexes, keywords and other methods to achieve. The advantage of this technique is simple, fast and easy to implement, but there are several more important issues.

1. Huge information in the query results irrelevant and every hundreds of piece of information is returned at every turn.

2. The query results are displayed chaotically, and the search engine could not differentiate the types of information when sorting the results.

3. Search algorithm uses keywords matching rather than semantic matching. This phenomena of multi-meaning words leads to the recall ratio is difficult to guarantee the search results contains a large number of invalid information which makes the precision rate is difficult to satisfy. Theses problems are inevitable in the keyword matching search algorithm.

The conventional information mechanism is difficult to adapt to the needs of the times for lack of the necessary intelligence and new intelligent information retrieval mechanism is expected on Semantic Web is proposed, which makes it possible to record effectively, helps computers to understand and operate information on the web and facilitate global knowledge exchange. The semantic web layer cake illustration in Fig. 1 demonstrates some key dependencies such as URLs and XML that forms the foundation of the Semantic Web. This figure also shows that future Semantic Web capabilities will deal with trust and providence.

![Semantic Web Layer Cake](image)

Fig. 1: Semantic Web Layer Cake
Information sharing is important and fundamental to the Semantic Web. It’s not only about transferring a stream of bits from one system to another but also about conveying a set of concepts between systems or across knowledge domains. When information is semantically (rather than syntactically) described, it can be shared across domain boundaries by defining the concepts of the foreign domain in terms of concepts we are familiar with. Once these relationships between the concepts of each domain are established, anyone can take advantage of them. Figure 2 illustrates this concept by presenting very simple definitions of the Semantic Web and the World Wide Web and drawing relationships between the concepts of each. On the left side of the figure is very high level description of the WWW. On the right side is the same for the Semantic Web. The circles in the figure represent concepts, while the arrows represent relationships between concepts. The arrows that are dashed connect the concept of WWW with the concept of Semantic Web.

Semantic web is a promising vision that is based on the idea that adding machine understandable semantic information to web resources will facilitate automation of many tasks, including integration of distributed application. OWL is emerging standard representation language for ontologies, and as such has good tool support. Also, the OWL Services (OWL-S) ontology is written in OWL, providing further motivation for using this way of representing ontologies.

2. INTELLIGENT INFORMATION RETRIEVAL FRAMEWORK

Fig. 3: Precision and Recall

\[
\text{Precision} = \frac{|\text{Retrieve } \cap \text{ Relative}|}{|\text{Retrieve}|} \\
\text{Recall} = \frac{|\text{Retrieve } \cap \text{ Relative}|}{|\text{Relative}|}
\]

Introduction of Intelligent Information Retrieval \( P \) (Precision) and \( R \) (Recall) are usually used to measure the performance of the retrieval algorithms, and their relations shows as Fig. 3. Precision \( P \) describes the proportion of the valid information in the search results, which reflects the usefulness for users Recall \( R \) describes the ratio relation of useful information to all information in the search results which meet users' need, and on the other hand reflects the omission of useful information [3]. The objective of information retrieval algorithm is not only to obtain a higher precision rate but also a better recall rate, but the conventional information retrieval algorithm is often difficult to achieve balance between them. In order to ensure a high recall rate in the retrieval process, most of search engines like Google adopt full-text search technology, which will add objects to the final search result set once the search words are contained, but since the retrieval process is for the entire World Wide Web, it will lead to tens of thousands search results. Since conventional information retrieval algorithms return a
large number of unrelated semantic search results, in fact it damages the Information Retrieval precision \( P \). In addition the widely available synonyms in natural language and openness of the World Wide Web leads to the phenomenon that the same resource object may have a variety of expressions, whereas conventional information retrieval performance in the retrieval process depends over-reliance on the query the user raises and turns a blind eye of “shaped synonymous” information, which leads to the recall rate (R) performance lower. Therefore, in order to alleviate the crisis of the “rich data poor information” and make better use of information resources in the World Wide Web, a new intelligent information retrieval technology should be researched. In the information retrieval algorithms, intelligent Information Retrieval is a typical application of Semantic Web in the field of information retrieval, introduces the process of Semantic information for resource objects, improves precision rate, recall rate and performance of information retrieval systems. The biggest difference between semantic retrieval and conventional information retrieval is the process of original information; The indexing information of conventional information retrieval is just a description of the contents of the original document, which can not reflect the role and status of document in the field and through which other related document information can not learned; But in the intelligent information retrieval system the meta-data of documents is organized and arranged on the structure of domain ontology, and it not only reflects the internal information of the document, but also reflects the relationship between the document and other documents.

3. SEMANTIC WEB SERVICE DISCOVERY METHODS

A. OWL-S based

Broadly speaking, OWL-S based approaches use the service profiles and domain ontologies to decide whether there is a match between a requested service and an advertised service. A hybrid approach combining logic-based semantic matching with token-based similarity metrics. The matchmaker is called OWLS-MX where the “\( X \)” stands for five different instances (\( M_0 \) to \( M_4 \)) of the generic hybrid matching scheme depending on which syntactic similarity metric is used. OWL-S based discovery system that employs a two-staged consistency-based matchmaking approach where in the first stage, service profiles that potentially match the request are short listed, and in the second stage, the concrete information about these services is obtained by querying the services directly to check for consistency with the requirements. I-Wanderer [5] performs service discovery in a P2P environment by making use of a clustering technique based on OWL-S functional descriptions at the peer nodes.

The Semantic Web is rapidly becoming a reality through the development of Semantic Web markup languages such as OWL. These markup languages enable the creation of arbitrary domain ontologies that support the unambiguous description of Web content. Web-accessible programs and devices are among the most important resources on the Web. These Web Services have garnered significant interest from industry, and standards are being developed for low-level descriptions of Web Services including WSDL which provides a communication level description of the messages and protocols. To make use of a Web service, a software agent needs a computer-interpretable description of the service and the means to access it. The OWL ontology for Web Services, called OWL-S, could make Web Services computer-interpretable in order to perform the tasks.

An Upper Ontology for Services in OWL, abstract categories of entities are defined in terms of classes and properties. OWL-S uses OWL to define a set of classes and properties specific to the description of services. The class Service is at the top of the OWL-S ontology. The ontology of services provides three essential types of knowledge about a service:

1. The class Service presents a ServiceProfile: “What does the service provide for and require of agents?”
2. The class Service is described By a ServiceModel: “How does it work?”
3. The class Service supports a ServiceGrounding: “How to access the service?” The ServiceProfile provides information about a service that can be used by an agent to determine if the service meets its rough needs, and if it satisfies constraints such as security, locality, and quality requirements. The ServiceModel enables an agent to: (1) perform an in-depth analysis of whether the service meets its needs, (2) compose service description from multiple services to perform a specific task, (3) coordinate the activities of different agents, and (4) monitor the execution of the service. The ServiceGrounding specifies the details of how to access the service such as protocol and message formats, serialization, transport, and addressing. Grounding is a mapping from an abstract to a concrete specification of those service description elements that are required for interacting with the service; basically, the inputs and outputs of atomic processes. Building upon SOAP and WSDL technologies, the OWL-S ontology-based Web services can be dynamically invoked by other services on the Web.
The upper ontology for services requires all three of the following properties to be fully characterized: presents, described by and supports. In general, the ServiceProfile provides the information needed for an agent to discover the service while the ServiceModel and ServiceGrounding objects provide information for an agent to use the service.

We will describe each of these three classes (ServiceProfile, ServiceModel, and ServiceGrounding) in the following sections.

**ServiceProfile**

A service profile provides a high-level description of a service and its provider and is used by discovery registries to request and advertise services. It includes: a human readable description, a specification of functionalities, and functional attributes. A transaction in the Web service marketplace involves three parties: the service requester, the service provider, and the infrastructure components. The service requester (buyer) seeks a service to complete its work and the service provider (seller) offers a service. The infrastructure components facilitate the process, such as registries to match the request with the offers available. Within the OWL-S ontology, the ServiceProfile describes the services. A ServiceProfile describes a service as a function of three basic types of information: what organization provides the service, what functions the service computes, and what features characterize the service. There is a two-way relationship between a service and a profile which is expressed by the properties presents (relates an instance of service and an instance of profile) and presented by (specifies the inverse of presents).

**ServiceModel**

A detailed perspective of a service can be viewed as a process. A subclass of the ServiceModel is defined as the ProcessModel which draws upon Artificial Intelligence, planning and workflow automation to support a wide array of services on the Web. Figure 4 presents the relationships of ServiceModel and includes the Process and Process Control.

**ServiceGrounding**

In OWL-S, both the ServiceProfile and the ServiceModel are thought of as abstract representations; only the ServiceGrounding deals with the concrete level of specification. The grounding of a service specifies the details of how to access the service, including protocol and message formats, serialization, transport, and addressing.

**B. WSMO based**

WSMO-based approaches view service discovery as a task of fulfilling goals which are abstractions of user’s desires. Compared to OWL-S, WSMO is capable of modeling mediation for handling heterogeneities that often arise in open environments. Keller, et al. [6] exploits this unique feature by proposing a generic discovery model that provides efficient pre-filtering of relevant services and accurate contracting of services that fulfill a given requester goal. Two systems in SWS Challenge have used the WSMO approach to solve all the discovery scenarios. The first system SWE-ST [7] uses the Glue discovery engine that is WSMO compliant. The ontologies are modeled using F-logic which is expressive enough to handle relationships among classes and instances, formulas that use logic operators, quantifiers, and rules. The second system from DERI [8] is based upon WSMX – the WSMO execution environment, models services, goals and ontologies directly in WSMT, and expresses rules using WSML. Matchmaking is performed with WSML reasoner and KAON interface. It uses the abstract state machine formalism for allowing dynamic interactions with the providers. Klusch et al. [9] adopt the hybrid approach in OWLSMX to support WSMO by developing a system, namely, WSMO-MX. Similar to OWLS-MX, WSMO-MX uses logic-based matching as well as text retrieval strategies. As WSMO is based on WSML, WSMO-MX is based on WSML-MX. The basic idea behind WSML-MX is to extend the WSML variant WSML-Rule such that users can specify constraints about the manner semantic matching of desired service capability elements shall be performed by a matchmaker. Experimental results showed that WSMO-MX has high precision but is time consuming due to high computation costs associated with both logic-based matching and text retrieval.

**C. SAWSDL and WSDL-S based:**

maker to support syntactic matching, quality of service (QoS) matching, and semantic matching in addition to context and the coverage information of concepts. Overall, the extended approach showed higher precision under similar runtime performance. Moreover, the METEOR-S infrastructure facilitates scalable discovery on web scale. The METEOR-S approach has been extended to SAWSDL as well. Kourtesis, et al. [12] propose a semantic registry approach for discovery by using SAWSDL for annotating service interfaces, OWL for modeling service capabilities and DL reasoning for matchmaking. Klusch, et al. propose SAWSDL-MX by extending their hybrid discovery approach to support SAWSDL [13].

4. CONCLUSION

WSMO-based approaches are currently the most dominant due to better performance characteristics, heterogeneity support and their holistic approach towards service discovery, through W3C recommended service description are fast catching up. An intelligent search engine framework based on Semantic search is proposed, which has the ability of semantic processing and natural language interface. Semantic retrieval depends on natural language understanding, data mining technology and so on. There are some shortcomings to be further researched and improved, such as updation problem followed by new information is stored and so on.

REFERENCES


