Abstract

Search engines today are based on decades old technology patched with new solutions. When specifying a search, users enter a small number of terms in the query. Yet the query describes the information need and is commonly based on the words that people expect to occur in the types of document they seek. This gives rise to a fundamental problem, in that not all documents will use the same words to refer to the same concept. Therefore, not all the documents that discuss the concept will be retrieved by a simple keyword-based search. Furthermore, query terms may of course have multiple meanings (query term polysemy). As conventional search engines cannot interpret the sense of the user’s search, the ambiguity of the query leads to the retrieval of irrelevant information.

Converse to the problem of polysemy, is the fact that conventional search engines that match query terms against a keyword based index will fail to match relevant information when the keywords used in the query are different from those used in the index, despite having the same meaning (index term synonymy). Although this problem can be overcome to some extent through thesaurus-based expansion of the query, the resultant increased level of document recall may result in the search engine returning too many results for the user to be able to process realistically. In addition to an
inability to handle synonymy and polysemy, conventional search engines are unaware of any other semantic links between concepts.

Many search engines fail to take into consideration aspects of the user’s context to help disambiguate their queries. User context would include information such as a person’s role, department, experience, interests, project work etc.

The results returned from a conventional search engine are usually presented to the user as a simple ranked list. The sheer number of results returned from a basic keyword search means that results navigation can be difficult and time consuming. Generally, the user has to make a decision on whether to view the target page based upon information contained in a brief result fragment.

Most of them use the inverted index method and its statistics/popularity flavored derivatives. The search problem is directly attributable to the limitations of the inverted index method as the underlying platform. Any kind of semantic enrichment requires handling and organizing semantically rich data, and a very short processing time. This requirement exceeds what is expected from an inverted index regardless of hardware capacity.

The search community has provided various solutions to such a problem by means of query expansion and semantic annotation. Query Expansion is addressed using linguistic knowledge in the form
of WordNet/thesauri. Query expansion is also addressed using the domain specific knowledge captured in taxonomic model of Ontologies. By the very complex nature of the knowledge representation of a domain, very few relationships are addressed while solving this problem in QE. State of the art query expansion uses synonyms and hierarchical relationships only, expanding either in the directions of parent or children. Semantic annotation requires the annotation of the resources by using the terms from a knowledge base. Such a process of manual creation of the domain knowledge and also annotating the documents with the knowledge is quite expensive. As a result, annotations often are incomplete or erroneous, resulting in decreased search performance.

The use of ontologies and associated metadata can allow the user to more precisely express their queries, thus avoiding the problems above. Users can choose ontological concepts to define their query or select from a set of returned concepts following a search in order to refine their query. This can improve the accuracy of a search. Having submitted a query, the user is then presented with a ranked list of documents relevance to the query. Therefore it is suggested here, the future of the search engines lies in supporting more of the information and management process, as opposed to incremental and modest improvements to relevance ranking of documents. In this approach, software supports the process of actually reading and analyzing relevant documents, rather than merely listing them and leaving the
rest of the information analysis task to the user. Corporate knowledge workers need information defined by its meaning, not by text strings. They also need information relevant to their interest and to their current context. They need to find not just documents, but sections and information entities and even digests of information created from multiple documents. The generation of ontologies and the creation of metadata attributing information to them is the key to access knowledge domain specific entities.

Metadata, therefore, captures the wide range of intrinsic or extrinsic information about a variety of objects. These intrinsic or extrinsic characteristics and features are described in the individually structured data elements that facilitate object use, identification and discovery.

The way that current service-oriented infrastructure handles and manages services metadata is not adequate and effective for metadata to help services discovery and knowledge sharing. There are no problems for humans to understand XML-based metadata because we know the meaning of these English words, the question is: “can machines understand and consume them?”, so that they can perform automatic processing with regards to the use of Web/Grid services. Clearly without further assumptions, the answer will be no. The Semantic Web / Grid are extensions of the current Web/Grid in which information and services are given well-defined meaning, better enabling computers and people to work in cooperation. We believe
that the first step towards the Semantic Web/Grid is to make the Web/Grid full of rich SMD, in other words, metadata with semantics.

Our approach targets towards providing the relevant results using the domain specific knowledge. To meet the expectations, the solutions for the problem discussed above, we have used domain knowledge represented in the form of OWL Ontology. Our approach uses query expansion method to expand the query and the traditional basic keyword search as the search mechanism. Our system fits the query terms in the ontology graph in an appropriate way and exploits the surrounding knowledge to retrieve the relevant results. We find that we can fix the context of the query and also bring in semantically related terms into the picture using this approach. The resulting enhanced query is given to the underlying basic keyword search system. As a result, we find that we can achieve substantial improvement in both precision and recall compared to the basic keyword search system.

We also proposed our approach towards the next generation service-oriented computing infrastructure with rich metadata and semantic support and also presented an integrated framework for Semantic Meta Data management for Web/Grid services.

If a program wants to compare conceptual information across two knowledge bases on the Web, it has to know when any two given terms are being used to mean the same thing. It intends to capture all
metadata of Web/Grid services and the concepts related to domain in which these services operate. It further models these metadata, concepts and their relations in a structure using commonly agreed terms. The purpose is to abstract the ontological entities of metadata and put them in context, thus giving them meaning. Our framework uses ontologies to perform metadata and context modeling in which entities such as services will be conceptualized as ontological concepts and an entity’s metadata will be conceptualized as its properties. Context modeling will conceptualize all other entities related to the concerned entity and establish relations among them via concepts properties. Overall context modeling will create a self-contained ontology in which metadata can be interpreted unambiguously by both humans and machines. Ontology-based metadata and context modeling provides a common communication language for Web/Grid service providers and consumers.

The key features of the approach are as follows: Firstly, ontologies are used for metadata and context modeling, thus help towards interoperability and machine understandability. Secondly, knowledge acquisition, i.e., service metadata collection and semantics tagging, is carried out semi-automatically through a formal knowledge binding process also known as semantic annotation. Thirdly, Web ontology languages are used for SMD knowledge representation, thus enabling knowledge sharing and effective reuse.
Finally a case study, describes the applicability of social network analysis to the semantic web, particularly discussing the multi-dimensional networks that evolve from ontological trust specifications.

In many computing systems, information is produced and processed by many people. Knowing how much a user trusts a source can be very useful for aggregating, filtering, and ordering of information. Furthermore, if trust is used to support decision making, it is important to have an accurate estimate of trust when it is not directly available, as well as a measure of confidence in that estimate. Much research has focused on authentication of resources, including work on digital signatures and public keys. Confidence in the source or author of a document is important, but trust, in this sense, ignores many important points. Just because a person can confirm the source of documents does not have any explicit implication about trusting the content of those documents. The web itself has shown the patterns of a small world network, in clustering and diameter. Viewing the current web as a graph, where each page represents a node and the hyperlinks translate to directed edges between nodes, has produced some interesting results.

The key idea is that by introducing ontology, an algorithm for finding trust from the resulting network and different methods for accessing the network through a web service or applications, this is a first step for showing how non-security based efforts can become part
of the foundation of the web of trust. A representation of social relationships needs to be fine grained enough so that we can capture all the detail from the individual sources of information in a way that these can be later recombined and taken as an evidence of a certain relationship. Network analysis itself can be a help, as it has a rich vocabulary of characterizing social relationships.