

Semantic Search : A Survey

Jinsoo Park

College of Business Administration &
Graduate School of Business,
Seoul National University
(*jinsoo@snu.ac.kr*)

Namwon Kim

College of Business Administration,
Seoul National University
(*telew@snu.ac.kr*)

Minjung Choi

College of Business Administration,
Seoul National University
(*snumj04@snu.ac.kr*)

Zhe Jin

College of Business Administration,
Seoul National University
(*kimcher@snu.ac.kr*)

Youngseok Choi

College of Business Administration,
Seoul National University
(*aquinas9@snu.ac.kr*)

.....

Since the ambitious declaration of the vision of the Semantic Web, a growing number of studies on semantic search have recently been made. However, we recognize that our community has not so much accomplished despite those efforts. We analyze two underlying problems : a lack of a shared notion of semantic search that guides current research, and a lack of a comprehensive view that envisions future work. Based on this diagnosis, we start by defining semantic search as the process of retrieving desired information in response to user's input using semantic technologies such as ontologies. Then, we propose a classification framework in order for the community to obtain the better understanding of semantic search. The proposed classification framework consists of input processing, target source, search methodology, results ranking, and output data type. Last, we apply our proposed framework to prior studies and suggest future research directions.

.....

Received : December 5, 2011 Accepted : December 20, 2011

Type of Submission : English Fast-track Corresponding author : Jinsoo Park

1. Introduction

Since the advent of the World Wide Web (WWW), the needs for effective search (methodologies and technologies) have been fueled by the unprecedented avalanche of information. In

light of this, the world-wide vision of the Semantic Web (Berners-Lee et al., 2001) was declared by the WWW Consortium (W3C) a decade ago. The term "Semantic Web" is mainly used to describe the model and technologies proposed by the W3C. These technologies in-

clude the Resource Description Framework (RDF), a variety of data interchange formats, and notations such as RDF Schema (RDF/S) and the Web Ontology Language (OWL). These are to provide a formal description of terms, concepts, and relationships within a given domain to help retrieve data.

Although many of the technologies proposed by the W3C already exist and are used for various purposes in practice, the Semantic Web as a global vision remains largely unrealized and its feasibility have been questioned. Highly interesting is that this is in stark contrast to the academia where a plethora of research is being produced with the fancy titles tagged as “semantic.”

Thus, we maintain that this alleged failure is not with the lack of academic interest or efforts but with the lack of a central theme that guides current research and envisions future. In the middle of this lies semantic search which is believed to solve the limitations of the current Web and accomplish the vision of so-called “Web 3.0.”

Information Systems (IS) society suffers from two intertwined problems regarding the investigation on this topic. First, although there have been innumerable attempts solely or partially devoted to semantic search, they have never been well-categorized, leading to research outcomes unclear and hard to understand. Because many authors and their papers differ in scope, perspectives, depth, and width, we realize that our community necessitates a comprehensive view

not only to better understand the concept but also to advance our understanding of it. Second, contributions made by current research are not well appreciated by the practitioners as well as academia because there is no consensus on what “true” semantic search is. Without the shared definition, many of the efforts accumulated so far do not seem to comprise a big picture. Thus, we argue for a well-defined definition of semantic search to strengthen our knowledge base in a more effective way.

2. Semantic Search in Literature

To the best of our knowledge, there is no widely accepted definition of semantic search although a handful of suggestions have been made in literature (Ding et al., 2004; Guha et al., 2003; Wu et al., 2006). This is partially due to the fact that the notion of the Semantic Web, from which the concept of semantic search originated, has not yet been formally defined despite our frequent reference to it. Some studies have failed to address this and are problematic in that they have too broad a concept and void of an appropriate definition.

Guha et al. (2003) suggests that semantic search is to augment and improve traditional search results (based on Information Retrieval technology) by using data from the Semantic Web. In Ding et al. (2004), semantic search engines are required to find appropriate ontologies and instance data, and characterize the Semantic Web. Recently, Wu et al. (2006) defined seman-

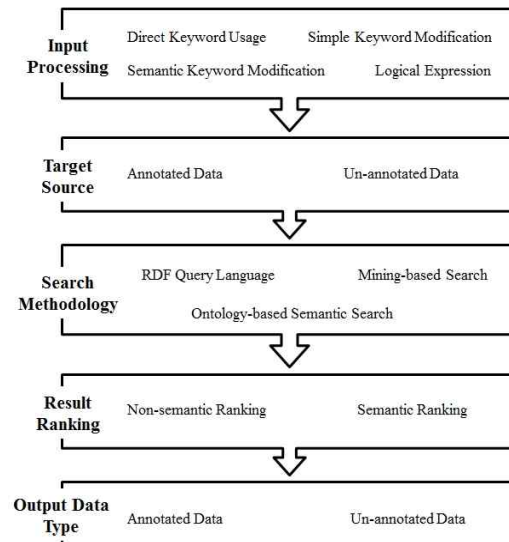
tic search as finding the semantic associations between objects with providing a graph-based user interface.

We define *semantic search* as the process of retrieving desired information (output data) in response to the keyword request (input) via the use of ontologies or other mechanisms alike (target sources and search methods) to understand users' implicit as well as explicit intent. Note that this is in alignment with our classification framework that will follow in the next section.

3. Classification

Based on our definition above, we propose a classification framework as shown in <Figure 1>. It is logically comprised of five distinct components : input processing, target source, search methodology, results ranking, and output data type. First, *input processing* deals with the keywords created by users per se or their modification. Second, *target source* classifies the data type a search engine surfs for retrieving information, such as traditional text or semantic data. Next, in the *search methodology* dimension, semantic search engines are categorized into the development of query languages, use of ontologies, and application of data mining with regards to the technology they adopt. Fourth, *methods for result ranking* are concerned with whether semantic relations are considered or not in displaying search results. Last, the final *output data type* is classified into semantically annotated data versus un-annotated data. Note that this

framework follows the consecutive steps the user may go through when using search systems.



<Figure 1> Framework for Semantic Search

3.1 Input Processing

The first step in conducting semantic search is processing the keyword from users. In the context of semantic search, it is difficult for normal search systems to return semantically satisfactory results to the end user if only based on the keyword the user has entered. This is largely because the viewpoints of search systems and users do not always match due to the semantic ambiguity of input keywords, leading to unsatisfactory answers. Appropriate query processing is, thus, of prime importance for efficiently searching Web resources.

Although some studies do not adopt any keyword modification techniques, most of the

search systems maximize their performance by clarifying the meaning of the keyword input from the user using various modification techniques. Keyword modification techniques may have two possible approaches : simple keyword modification and semantic keyword modification. One of the alternative ways of keyword modification is to present query as a logical expression. This can be a very effective method to provide semantics of user intention to machine because the logical expression allows the user to specify the condition and other constraints for the description of user intention. Thus, in our study, we categorize input processing methods into four sub-categories, i.e., (1) direct keyword usage, (2) simple keyword modification, (3) semantic keyword modification, and (4) logical expression.

3.1.1 Direct Keyword Usage

Most of the search systems in this category find the document which contains the exact word that matches the user's input. Some semantic search engines do not use any query modification. Swoogle (Ding et al., 2004), one of the most largely cited metadata search engines, finds the metadata which includes the exact same tag as the user input. XSearch (Cohen et al., 2003) is another example. It returns XML (eXtensible Markup Language) documents without any query extension.

3.1.2 Simple Keyword Modification

In semantic knowledge retrieval systems, an ontology can provide the primitives needed

to formulate keyword descriptions. Especially, simple ontologies, e.g., keyword hierarchies, taxonomy, or thesaurus, might also benefit information retrieval techniques. Using these kinds of simple external knowledge resources, the range of user input could be extended to its synonym set. OntoSeek (Guarino et al., 1999) improves its search performance employing WordNet, a large linguistic resource, to clarify the sense of a word in the input. Simple NLP (Natural Language Processing) techniques are an alternative approach for keyword modification. A question answering system for Semantic Web, AquaLog (Lopez et al., 2005) combines several powerful techniques, including both GATE NLP platform and WordNet to make sense of natural language keywords. However, in this kind of simple modification, the effect of modification is limited to resolve the lexical ambiguity based on external resource. For this reason, semantic keyword modification can be a good alternative for keyword extension strategy.

3.1.3 Semantic Keyword Modification

For more accurate descriptions of the user's input, semantically richer ontology-based keyword annotations are frequently used. Such annotations are not atomic keywords but can be more detailed structured descriptions that are linked with other resources. With the help of ontologies, the user can also express search keywords more precisely and unambiguously, which leads to better precision rates (Hyvönen et al., 2003). Furthermore, through ontological class de-

definitions and inference mechanisms, implicit semantic relationships between the concepts in user input can be clearly defined. Ontogator, which uses controlled vocabularies or hierarchical thesauri for finding keyword-based annotated resource, is one good example of this approach (Hyvönen et al., 2003). OntoBroker (Stefan et al., 1998) has a user query interface, in which the technical ontology concepts of object, class, attribute, and value are used to form logical query expressions.

3.1.4 Logical Expression

Using the formal logical expression as an input can be a good alternative for keyword modification. Especially, logical query which is based on a specified sophisticated query language is more expressive in presenting semantic intent of the users, although it requires training and thus might pose a challenge to non-expert users. SMART (Battista et al., 2007) utilizes the Manchester OWL Syntax to define description logics (DL) queries (i.e., queries defined by a class description) to facilitate semantic keyword formulation. An important feature of this language, which makes it more readable and easier to understand than traditional DL syntax, is that it uses an infix notation rather than a prefix notation for keywords.

3.2 Target Source

In this section, we describe the target source of a search engine. Target source can be classified into either un-annotated or annotated

data. Basically, semantic search is an extension of the current Web which targets text, image, and sound (Mayfield and Finin, 2004). Therefore, the first category of target sources is, by nature, un-annotated data. Second, industry develops machine-readable metadata languages such as RDF/S and OWL, XML (Garcia and Sicilia, 2003). These markup languages linked terms up with Web pages to provide a machine-readable definition and their relationship with other terms. These standards are reflected in <Table 1> which summarizes the methodologies of semantic search in order to capture patterns of semantic search according to its target source or methodologies.

3.2.1 Semantically Annotated Data

Semantic annotation refers to a specific metadata generation and usage schema targeted to enable new information access methods and extend existing ones (Kiryakov et al., 2004). Moreover, semantic annotation is based on the understanding that the named entities mentioned in the documents constitute important part of their semantics. The annotation mark is attached to text documents, images, audio, video, or other kinds of information. There are a number of annotation methods commonly used for browsing and searching, and one of them is tagging that mostly accounts for semantic annotation. Unlike tagging, we can encode information in a machine-readable form and find information that share the semantic meaning by using markup languages (Mayfield and Finin 2004). RDF/S and

OWL are representative markup languages that define the terms with formal semantics and relationship among other terms. This aims at assigning pre-defined semantic descriptions or entities to the data objects for their representation. For example, SemRank suggests a heuristic method to determine the rank of semantic associations among those entities (Anyanwu et al., 2005).

3.2.2 Un-Annotated Data

In general, when traditional data is used as sources within semantic search engines, it requires additional annotation process unlike annotated data. Because the traditional data could not manifest themselves as descriptive data stored in Web-accessible databases (Chebotko and Lu, 2009), a large number of semantic search engines attach the semantic meaning to the traditional data via annotation step (Heflin and Hendler, 2000; Cohen et al., 2003; Khan et al., 2004; Cheng and Qu, 2009). For example, SHOE (Heflin and Hendler, 2000) retrieves Web pages as concepts in ontology. It annotates URLs of Web pages through application of SGML (Standard Generalized Markup Language) and XML, and then retrieves the Web pages. After this annotating step, traditional data could be used by semantic search engines.

3.3 Search Methodology

3.3.1 Query Language for RDF Source Data

In the Semantic Web, many search technologies retrieve information from tagged data

such as RDF sources, which require substantial formatting and markup to achieve machine understandability. Nowadays, the voluminous RDF descriptions, sufficiently expressive declarative languages, are already appearing. The quantity of RDF data has also been growing steadily in recent years. Better knowledge about their meaning, usage, accessibility, or quality will considerably facilitate an automated processing of Web resources (Gregory et al., 2002). The method to take advantage of RDF sources data, and therefore promote retrieval performance, is to combine information retrieval approaches with inference searching based on RDF query language. SPARQL is emerging as the de-facto RDF query language, and is also a W3C Recommendation. There are some other RDF query languages such as DQL (Doctrine Query Language), N3QL, R-DEVICE, RDFQ, RDQ, RDQL, RQL/RVL, SeRQL and so on. Some search technology examples in this category include XSearch (Cohen et al., 2003), TAP (Guha and McCool, 2003), Falcon (Cheng and Qu, 2009) and SEMPLORE model (Wang et al., 2009).

3.3.2 Ontology-based Semantic Search

Ontology is believed to be a tool for realizing the Semantic Web, which focuses on technological reusability and interoperability. It provides a general knowledge-sharing space which can allow the various systems or organizations to have consistent and unified understanding on semantic information. Ontology as an explicitly

formal specification of knowledge can transform information into semantic metadata which is used to describe the information entities. Using ontology to build the semantic-oriented metadata can support the concept modeling, information searching and exchanging. Ontology provides search engines with the function of a semantic match. It is different from traditional search engines that can only complete the search directly and exactly; ontology-based semantic search engines rely on certain ontology structures which can extend the search to the semantic level.

By using information resource annotation technology, ontology-based information search promotes the retrieval from text-matching to concept-matching. Therefore, applying ontology to the process of retrieval has become a hotspot in the current study. SHOE uses the way of natural navigation to guide users to point out the object of their inquiries (Heflin and Hendler, 2000). SPIRIT developed tools and techniques to support spatial search on the Internet based on ontology (Jones et al., 2002). TAP implements semantic expansion of keyword-based search and returns ontology instances in line with the query words (Guha and McCool, 2003). An ontology-based spatial query expansion method is developed that supports retrieval of information relevant to space by trying to derive its geographical query footprint (Fu et al., 2005). SMART (Battista et al., 2007) provides a Web-based application that adopts an ontology-centric model to perform semantic query answering over heterogeneous yeast biological knowledge.

3.3.3 Mining-based Search

Even under the banner of the Semantic Web era, semantic data in the Web are still not rich enough to take the advantages of semantic search engine. Thus, the effectiveness of the semantic search is largely dependent on its capability of analyzing the traditional Web data, which requires capturing the relationship between the objects or concepts described within. In this regard, data mining in terms of semantic search plays an important role in transforming human understandable content to machine understandable semantics.

Cheng et al. (2008)'s study is a compelling example adopting vector space model (VSM) that takes up one part of the mining technique for information retrieval. Another stream in semantic search literature includes the study of social networks which can discover personal or organizational information. Ontocopi (Alani, 2002), a tool for finding communities of practice in social network, extracts and identifies patterns of certain connections between entities relating to a community. Flink (Mika, 2005) is a personal information search system in the research community. Its mining component employs a co-occurrence analysis technique applied to social network extraction. Another mining component also performs the additional task of finding topic interests, i.e., associating researchers with certain areas of research.

3.4 Results Ranking

Result ranking is the next phase when

search engine obtained some results. Nowadays, there is a great quantity of data on the WWW and search engines generally provide overabundant retrieval results when searching on the Web. Most of the answers are usually irrelevant or insufficient because traditional Web technology establishes the resources in the Web, which are only human-understandable but not for search engines. In reality, however, the user often ignores most of the results and tends to consider just the first few that are returned by traditional search engines. It is clear that the effective techniques for ranking the result data are required. Based on our study of the recent research, we classify the major ranking approaches into non-semantic ranking and semantic ranking.

3.4.1 Non-Semantic Ranking

Some of the semantic search engines employ existing ranking methods, which does not take into deep consideration rich semantics of relations of the classes in the ranked result. Among the non-semantic ranking methods is using the frequency of the terms. It mainly relies on the number of query keywords presented in the target documents. Cohen et al. (2003), for example, referred to term frequency in the document by employing extended vector space model to rank the search result. Another example is a rule-based approach, whose focus is confined to the concepts or objects in the result themselves, not the relations between them. In FalconS (Cheng et al., 2008), for example, the SW ob-

jects (what they call Semantic Web object in their developed artifact) in the results are ranked by a combination of their relevance to the query and their popularity. The popularity score is calculated based on the number of RDF documents that the SW object is used in.

3.4.2 Semantic Ranking

The task of semantic ranking is analyzing across different sources of data and ranking retrieval results by uncovering previously unknown and potentially semantically related with user query. Semantic ranking technology is a core subject for Semantic Web applications, such as semantic search and semantic service.

In recent years, some researchers have presented several kinds of information retrieval middleware which allows users to utilize some semantic discovery protocols in semantic and ubiquitous environments. Some of the methods derive ranking results from the identification of the semantic relatedness by calculating the conceptual distance between the query and the context (Rada et al., 1989). Another route is using ontology. Karanastasi and Christodoulakis (2007) present an ontology-driven semantic ranking methodology based on domain specific ontologies where a semantic relatedness measure was used for disambiguation of a natural language query. There are also some other methods. Choi et al. (2005) introduce three service-matching algorithms based on the marriage-matching algorithm to ranking that takes into account hier-

archical structure of the domain.

3.5 Output Data Type

The last component for semantic search is output data type of search results. Output display should be done to provide users with organized information for decision making about whether information is relevant to the queries.

Similar to the target source, the output form is classified into semantically annotated data and un-annotated data. Semantically annotated output is generated from either originally annotated target source or those un-annotated which are later annotated by the search engines. As for the un-annotated data, such type of output is provided when search engines do not process un-annotated target source.

4. Discussion

Drawing from our analysis presented above, we introduce a classification scheme for semantic search research, adopting basic search processes on the Web as each component of our framework. At the core of semantic search underlie different disciplines and various techniques (i.e., information retrieval, machine learning, natural language processing, ranking algorithm, computational/natural linguistics, ontology, and so forth). Many studies included in our survey are mainly focused on how they utilize and apply these various disciplines to improve the search results. It is thus a comprehensive if

not perfect, and at the same time unambiguous, easily understood classification framework that our academic society at large are currently in need.

A first finding from our exploration is that the semantic search framework has not yet been formalized although many semantic search systems have been devised and implemented. Although we selected and thoroughly studied more than twenty search systems (engines), we have encountered the difficulty in applying a unified perspective to view semantic search. This is the reason why we have proposed a novel yet comprehensive framework with regards to the intuitive steps of the semantic search process. We posit that our proposed framework might not be complete or exhaustive enough to cover all extant research. Granted, we do believe that our study can be a steppingstone that helps and motivates prospective researchers to participate in this intellectual journey towards realizing true semantic search.

Second, we found a particular tendency in the methodologies used in each search process. In the input processing phase, for instance, semantic modification takes up a dominant portion of the techniques for processing user keywords. We contend that logical expression may be another good option for input processing methods in that users are able to specify conditions or constraints for their query by using logical expression. On the contrary, however, very few studies published heretofore adopt this type of method in processing user input. This can large-

ly be explained by its complexity for non-experts who are not familiar with logical query language. Likewise, we found most of the semantic search systems use semantically annotated data as target source. This is because it is very useful in finding a semantic relationship between concepts through ontology.

Our last finding revolves around the problem that semantic search systems have in com-

mon with improving the quality in comparison to that of conventional Web search. Even though many kinds of semantic search engines use semantically annotated data format as target source, in the Web scenario, only a small fraction of knowledge is semantically annotated. In case of unstructured knowledge, one unavoidable problem is how to choose an appropriate metadata like ontology or RDF for annotation which en-

<Table 1> Studies on Semantic Search

	Input Processing	Target Source	Search Methodology	Result Ranking	Output Data Type
Stefan (1998)	Semantic Modification	Annotated	Ontology	Semantic	Annotated
Guarino et al. (1999)	Simple Modification	Annotated	Query Language, Ontology	No Ranking	Annotated
Heflin and Hendler (2000)	Semantic Modification	Un Annotated	Ontology	No Ranking	Annotated
Alani et al. (2002)	Semantic Modification	Annotated	Ontology, Mining	Semantic	Annotated
Jones et al. (2002)	Semantic Modification	Both	Ontology	Semantic	Annotated
Karvounarakis et al. (2002)	Logical Expression	Annotated	Query Language	No Ranking	Annotated
Shah et al. (2002)	Simple Modification	Both	Ontology, Mining	Hybrid	Both
Sheth et al. (2002)	Simple Modification	Both	Ontology	N/A	Un-Annotated
Cohen et al. (2003)	Direct Keyword Usage	Un Annotated	Query Language	Non-Semantic	Un-Annotated
Ehrig et al. (2003)	Semantic Modification	Both	Ontology	N/A	Both
Guha and McCool (2003)	Semantic Modification	Annotated	Query Language	No Ranking	Annotated
Hyvönen et al. (2003)	Semantic Modification	Annotated	Ontology	Semantic	Annotated
Stojanovic (2003)	Semantic Modification	Annotated	Ontology	Semantic	Annotated
Calvanese et al. (2004)	Semantic Modification	Annotated	Query Language, Ontology	No Ranking	Annotated
Ding et al. (2004)	Direct Keyword Usage	Annotated	Ontology, Mining	Semantic	Annotated
Khan et al. (2004)	Simple Modification	Un Annotated	Ontology	Semantic	Annotated
Rocha et al. (2004)	Semantic Modification	Annotated	Ontology	Semantic	Annotated
Anyanwu et al. (2005)	Semantic Modification	Annotated	Ontology	Semantic	Annotated
Fu et al. (2005)	Semantic Modification	Annotated	Ontology	Semantic	Annotated
Lopez et al. (2005)	Simple Modification	Annotated	Ontology	Semantic	Annotated
Mika (2005)	Direct Keyword Usage	Un Annotated	Mining	No Ranking	Un Annotated
Celino et al. (2006)	Semantic Modification	Annotated	Ontology	No Ranking	Annotated
Wu et al. (2006)	Semantic Modification	Annotated	Ontology	Semantic	Annotated
Battista et al. (2007)	Logical Expression	Annotated	Ontology	No Ranking	Annotated
Cheng et al. (2008)	Semantic Modification	Un Annotated	Ontology, Mining	Non-Semantic	Un-Annotated
Wang (2009)	Semantic Modification	Both	Query Language	Semantic	Both

ables clarifying the semantic relationship between classes. Consequently, the semantic power of search systems which use unstructured data source is very limited. This means the quality of a search result happens to be strongly dependent upon whether the source knowledge is structured or not.

<Table 1> summarizes a selection of papers that applied our classification framework. Note that we only included those papers that satisfied our definition of semantic search provided earlier although our original collection was a lot more.

Our analysis of extant research so far lets us come close to what we believe as semantic search. Ideally, semantically annotated data help the search engine to process target source as they are mapped to the concepts in ontology for sharing the semantics. Considering the current Web environment where such type of data are scarce, however, semantic search systems are to transform un-annotated target source into annotated ones to accomplish the semantic search. This will not only allow more sophisticated search mechanism but also bring better search results.

5. Future Research

Given our findings above, our contributions to the research community are two-fold : providing a semantic search framework and projecting future research. As for the framework, we expect researchers to adopt this unified

framework from this moment on. Only on the shared theoretical foundation can we accumulate our knowledge; and only on this sound accumulation can each piece of work be meaningful to the greater vision. In this regard, we are pleased that we have paved a way to positioning our research in a rightful location and shed a light on the expected research.

Thus, we set forth promising research directions on semantic search. To enhance the quality of a search result, one should try to cover not only structured data but also unstructured data by improving methodology for annotating unstructured knowledge. In addition to this attempt, the input processing phase also has more room for improvement. Existing studies mainly use a semantic query modification or logical expression to extract the user's intention exactly. However, these kinds of query processing approach only focus on extending user input semantically rather than finding more semantic information from users or user inputs. In this regard, utilizing user context including user information or query history may be a good strategy to maximize the relevancy of a search result. Indeed, some works already reflect user context as a search resource (Anyanwu et al., 2005; Lopez et al., 2005; Stojanovic, 2003), but more sophisticated methodologies for formalizing user context as semantic information are needed. Besides, personal information of a user can be an important source because the position or situation of the users can affect the expectation level toward a search result. Park et al.

(2011) serves as a good example which reflects the user's personal information in the input processing phase. Last but not least, with the endeavor to improve the quality of result knowledge, more comprehensive and rigorous evaluation metrics are to be developed for better semantic search.

References

- Anyanwu, K., A. Maduko, and A. Sheth, "Sem Rank : Ranking Complex Relationship Search Results on the Semantic Web", in *Proceedings of the 14th International World Wide Web Conference*, New York, (2005), 117~127.
- Alani, H., K. O'Hara, and H. Shadbolt, "Ontocopi : Methods and Tools for Identifying Communities of Practice", *Intelligent Information Processing*, Vol.221(2002), 225~236.
- Bamba, B. and S. Mukherjea, "Utilizing Resource Importance for Ranking Semantic Web Query Results", in *Proceedings of the Second International Workshop on Semantic Web and Databases (SWDB)*, Toronto, Canada, 2004.
- Battista, A. D. L., N. Villanueva-Rosales, M. Palenychka, and M. Dumontier, "SMART : A Web-based, Ontology-driven, Semantic Web Query Answering Application", in *Proceedings of the International Semantic Web Conference*, Busan, South Korea, 2007.
- Berners-Lee, T., J. Hendler, and O. Lassila, The Semantic Web, *Scientific American Magazine*, 2001.
- Calvanese, D., D. Lembo, M. Lenzerini, and R. Rosati, "What to Ask to a Peer : Ontology-based Query Reformulation", in *Proceedings of the 9th International Conference on the Principles of Knowledge Representation and Reasoning (KR)*, 2004.
- Celino, I., E. D. Valle, D. Cerizza, and A. Turtat, "Squiggle : a Semantic Search Engine for indexing and retrieval of multimedia content", in *Proceedings of the 1st International Workshop on Semantic-Enhanced Multimedia Presentation Systems (SEMPs)*, Athens, Greece, (2006), 20~34.
- Chebotko, A. and S. Lu, *Querying the Semantic Web : An Efficient Approach Using Relational Databases*, LAP Lamber Academic Publishing, 2009.
- Cheng, G. and Y. Qu, "Searching Linked Objects with Falcons : Approach, Implementation and Evaluation", *International Journal on Semantic Web Information System*, Vol.5, No.3(2009), 49~70.
- Cheng, G., W. Ge, H. Wu, and Y. Qu, "Searching Semantic Web Objects Based on Class Hierarchies", in *Proceedings of the WWW Workshop on Linked Data on the Web*, 2008.
- Cheng, G., W. Ge, and Y. Qu, FALCONS : Searching and Browsing Entities on the Semantic Web, in *Proceedings of the World Wide Web Conference*, 2008.
- Chenxi, L., X. G. Rong, Z. H. Jun, Y. Yong, "Using Probabilistic Latent Semantic Analysis for Personalized Web Search", in *Proceedings of the 7th Asia-Pacific Web Conference (AP-Web)*, Shanghai, China, March 29~April 1, 2005, Reprinted in *Lecture Notes in Computer Science*, 3399, New York : SpringerLink, (2005), 707~717.
- Choi, K. H., K. M. Lee, H. J. Shin, and D. R. Shin, "Efficient Algorithm for Service Mat-

- chmaking in Ubiquitous Environments”, in *Proceedings of the 6th International Conference on Web Technologies (EC-Web)*, Copenhagen, Denmark, Reprinted in *Lecture Notes in Computer Science*, 3590, New York : SpringerLink, (2005), 258~266.
- Cohen, S., J. Mamou, Y. Kanza, and Y. Sagiv, “XSearch : A Semantic Search Engine for XML”, in *Proceedings of the 29th International Conference on Very Large Databases*, Berlin, Vol.29(2003), 45~56.
- Ding, L., T. Finin, A. Joshi, R. Pan, R. S. Cost, and Y. Peng et al., “Swoogle : A Search and Metadata Engine for the Semantic Web”, in *Proceedings of the 13th ACM Conference on Information and Knowledge Management*, Washington D. C., (2004), 652~659.
- Ehrig, M., C. Tempich, J. Broekstra, van F. Harmelen, M. Sabou, and R. Siebes et al., “SWAP-ontology-based knowledge management with peer-to-peer technology”, in *Proceedings of the 1st National Workshop Ontologie-basierendes Wissensmanagemen*, 2003.
- Fu, F., C. B. Jones, and A. I. Abdelmoty, “Building a Geographical Ontology for Intelligent Spatial Search on the Web”, in *Proceedings of IASTED International Conference on Databases and Applications (DBA)*, 2005.
- Garcia, E. and M. A. Sicilia, “Designing Ontology-Based Interactive Information Retrieval Interfaces, on the Move to Meaningful Internet Systems, R. Meersman and Z. Tari (Eds.)”, *OTM Workshop*, Vol.2889(2003), 152~165.
- Gregory, K., A. Sofia, C. Vassilis, and P. Dimitris, “RQL : A Declarative Query Language for RDF”, in *Proceedings of the 11th international conference on World Wide Web*, 2002.
- Giuglea, A. M. and A. Moschitti, “Knowledge Discovery Using FrameNet, VerbNet and PropBank”, in *Proceedings of the Workshop on Ontology and Knowledge Discovering (ECML)*.
- Grimes, S., “Breakthrough Analysis : Two+Nine Types of Semantic Search”, *Information Week*, 2010.
- Guarino, N., C. Masolo, and G. Vetere, “Onto-Seek : Content-based Access to the Web”, *IEEE Intelligent Systems*, Vol.14, No.3(1999), 70~80.
- Guha, R. and R. McCool, “TAP : a Semantic Web Platform”, *Computer Networks*, Vol.42, No.5(2003), 557~577.
- Guha R., R. McCool, and E. Miller, “Semantic Search”, in *Proceedings of the 12th International Conference on World Wide Web (WWW)*, Budapest, 2003.
- Heflin, J. and J. Hendler, “Searching the Web with SHOE, in Artificial Intelligence for Web Search”, Papers from the *AAAI Workshop*, WS-00-01, AAAI Press, Menlo Park, CA, (2000), 35~40.
- Hildebrand, M., van J. R. Ossenbruggen, and L. Hardman, “An Analysis of Search-based User Interaction on the Semantic Web”, *Technical Report, Centrum voor Wiskunde en Informatica (NL)*, 2007.
- Hyvönen, E., S. Saarela, and K. Viljanen, “Ontogator : Combining View-and Ontology-based Search with Semantic Browsing”, in *Proceedings of the International SEPIA Conference*, Helsinki, (2003), 18~20.
- Jones, C. B., R. Purves, A. Ruas, M. Sanderson, M. Sester, M. Kreveld van et al., “Spatial Information Retrieval and Geographical On-

- tologies : An Overview of the SPIRIT project”, In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR), Tampere, Finland, (2002), 387~388.
- Karvounarakis, G., S. Alexaki, V. Christophides, D. Plexousakis, and M. Scholl, “RQL : A Declarative Query Language for RDF”, in *Proceedings of the 11th International Conference on World Wide Web* (WWW), (2002), 592~603.
- Khan, L., D. McLeod, and E. Hovy, “Retrieval Effectiveness of an Ontology-based Model for Information Selection”, *The VLDB Journal*, Vol.13(2004), 71~85.
- Karanastasi, A. and S. Christodoulakis, “Ontology-driven Semantic Ranking for Natural Language Disambiguation in the Ontonl Framework”, in *Proceedings of the 4th European Semantic Web Conference* (ESWC), Innsbruck, Austria, 2007.
- Kiryakov, A., B. Popov, I. Terziev, D. Manov, and D. Ognyanoff, “Semantic Annotation, Indexing, and Retrieval”, *Web Semantics : Science, Services and Agents on the World Wide Web*, Vol.2, No.1(2004), 49~79.
- Mayfield, J. and T. Finin, “Information Retrieval on the Semantic Web : Integrating Inference and Retrieval”, in *Proceedings of SIGIR Workshop on the Semantic Web*, Toronto, Canada, 2004.
- Latifur, K., D. McLeod, and E. Hovy, “Retrieval Effectiveness of an Ontology-based Model for Information Selection”, *The VLDB Journal*, Vol.13, No.1(2004), 71~85.
- Lee, J. H., M. H. Kim, and Y. J. Lee, “Information Retrieval Based on a Conceptual Distance in IS-a Hierarchy”, *Journal of Documentation*, Vol.49, No.2(1993), 188~207.
- Li, L., Y. Yang, and B. Wu, Ontology-Based Matchmaking in e-Marketplace with Web Services, in *Proceedings of the 7th Web Technologies Research and Development* (AP-Web), Reprinted in *Lecture Notes in Computer Science*, 3399, New York : SpringerLink, (2005), 620~631.
- Lin, C., G. R. Xue, H. J. Zeng, and Y. Yu, Using Probabilistic Latent Semantic Analysis for Personalized Web Search, in *Proceedings of the 7th Asia-Pacific Web Conference* (AP-Web), Shanghai, China, March 29~April 1, 2005, Reprinted in *Lecture Notes in Computer Science*, 3399, New York : Springer Link, (2005), 707~717.
- Lin, D., “An Information-Theoretic Definition of Similarity”, in *Proceedings of the 15th International Conference on Machine Learning*, (1998), 296~304.
- Lopez, V. and E. Motta, “Aqualog : An Ontology-portable Question Answering System for the Semantic Web”, in *Proceedings of the 2nd European Semantic Web Conference*, Heraklion, Greece, 2005.
- Mika, P., “Flink : Semantic Web Technology for the Extraction and Analysis of Social Networks”, *Journal of Web Semantics*, Vol.3, No.2(2005).
- Park, H., S. Rho, and J. Park, “A Link-based Ranking Algorithm for Semantic Web Resources : A Class-oriented Approach Independent of Link Direction”, *Journal of Database Management*, Vol.22, No.1(2011), 1~25.
- Rada, R., H. Mili, E. Bicknell, and M. Blettner, “Development and Application of a Metric on Semantic Nets”, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.19, No.1

- (1989), 17~30.
- Resnik, P., "Using Information Content to Evaluate Semantic Similarity in a Taxonomy", in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Quebec, Canada, Vol.1(1995), 448~453.
- Rocha, C., D. Schwabe, and M. P. Aragao, A Hybrid Approach for Searching in the Semantic Web, in *Proceedings of the 13th International Conference on World Wide Web (WWW)*, (2004), 374~383.
- Shah, U., T. Finin, A. Joshi, R. S. Cost, and J. Mayfield, "Information Retrieval On The Semantic Web", in *Proceedings of the 10th International Conference on Information and Knowledge Management*, 2002.
- Sheth, A., C. Bertram, D. Avant, B. Hammond, K. Kochut, and Y. Warke, "Managing Semantic Content for the Web", *IEEE Internet Computing*, Vol.6, No.4(2002), 80~87.
- Stefan, D. F., S. Decker, M. Erdmann, and R. Studer, "Ontobroker : The Very High Idea", in *Proceedings of the 11th International Flairs Conference (FLAIRS)*, Sanibal Island, USA, 1998.
- Stojanovic, N., "On Analysing Query Ambiguity for Query Refinement : The Librarian Agent Approach", in *Proceedings of the International Conference on Conceptual Modeling (ER)*, (2003), 490~505.
- Wang, H., Q. Liu, T. Penin, L. Fu, L. Zhang, T. Tran, Y. Yu, and Y. Pan, "Semplere : A Scalable IR Approach to Search the Web of Data", *Journal of Web Semantics*, Vol.7, No.3(2009), 177~188.
- Wei, W., P. M. Barnaghi, and A. Bargiela, "Search with Meanings : An Overview of Semantic Search Systems", *International Journal Communications of SIWN*, Vol.3(2008), 76~82.
- Wu, H., G. Cheng, and Y. Qu, "Falcon-S : An Ontology-Based Approach to Searching Objects and Images in the Soccer Domain", in *Proceedings of the International Semantic Web Conference (ISWC)*, 2006.
- Zhou, Q., W. Chong, X. Miao, H. Wang, and Y. Yu, "Spark Adapting Keyword Query to Semantic Search", in *Proceedings of the 6th International Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference (ISWC/ASWC)*, 2007.

Abstract

시맨틱 검색 : 서베이

박진수* · 김남원** · 최민정** · 김 철*** · 최영석***

시맨틱 웹(Semantic Web)의 비전에 대한 공표가 이루어진 이래로 이와 관련한 많은 연구가 진행되어 왔다. 그러나 지금까지의 연구가 성공적이었다는 판단을 하기에는 아직 이르다. 본 논문은 시맨틱 관련 연구분야의 두 가지 문제점을 진단한다. 첫째는 ‘시맨틱 검색’이라는 개념의 합의된 정의가 없다는 것이고, 둘째는 장래의 유관 연구를 바라볼 수 있는 종합적이고 체계적인 시각이 부족하다는 것이다. 이러한 진단 아래, 본 논문은 시맨틱 검색의 개념을 ‘사용자의 입력에 따라 온톨로지와 같은 시맨틱 기술을 이용하여 원하는 정보를 얻는 행위’로 정의한다. 또한 시맨틱 검색에 대한 이해를 돕기 위해 시맨틱 검색 엔진 분류 프레임워크를 제안하였다. 본 연구에서 제안하는 프레임워크는 (쿼리) 입력문의 처리, 타겟 소스, 검색 방법론, 검색결과와 서열화, 출력 결과물의 데이터 종류, 이렇게 다섯 가지 부분으로 나뉜다. 마지막으로 본 논문은 제시한 프레임워크를 응용하여 기존의 연구결과물을 분석하고 앞으로의 연구 방향을 논하는 것으로 끝을 맺는다.

Keywords : 시맨틱 웹, 시맨틱 검색, (쿼리)입력문의 수정

* 서울대학교 경영대학/경영전문대학원 부교수

** 서울대학교 경영대학 석사 졸업

*** 서울대학교 경영대학 박사과정

저 자 소 개



박진수

The University of Arizona에서 경영정보시스템을 전공하여 경영학 박사를 취득했으며, University of Minnesota의 Carlson School of Management에서 조교수, 고려대학교 경영대학에서 조교수를 역임했다. 현재 서울대학교 경영전문 대학원/경영대학에 부교수로 재직 중이다. 현재 국제저널인 Journal of Database Management와 International Journal of Principles and Applications in Information Science and Technology의 편집위원으로 활동하고 있으며 MIS Quarterly, IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Computer, ACM Transactions on Information Systems (TOIS), Information Systems Frontiers, Communications of the AIS, Journal of Global Information Technology Management (JGITM), International Journal of Electronic Business, 경영정보학연구 등 국내외 우수 전문학술지에 다수의 논문을 게재하였다. 주요 관심분야는 온톨로지, 정보 시스템 통합, 지식 공유, 에이전트, 시맨틱 모델링, 웹 정보시스템 등이 있다.



김남원

한국 과학기술원(KAIST) 산업공학과를 졸업하였으며, 서울대 경영대학에서 경영정보 전공으로 석사학위를 취득하였다. 주요 관심분야는 정보 시스템 보안, 시맨틱 웹, Social Network에서의 개인 정보 보호 등이 있다.



최민정

서울대학교 경영학과를 졸업하고 동 대학원 경영정보시스템 전공으로 석사학위를 취득하였다. 주요 관심분야는 데이터 모델링, 온톨로지, 지식관리시스템 등이다.



김 철

현재 서울대학교 경영대학에서 경영정보시스템 전공 박사과정에 재학 중이다. 주요 관심분야로는 시맨틱 웹, Ontology Engineering, 정보시스템에서의 정보 보안 등이 있다.



최영석

서울대학교 전기공학부를 졸업하였으며, 현재 서울대학교 경영대학에서 경영정보시스템 전공 박사과정에 재학 중이다. 주요 관심분야는 정보 시스템 통합, 온톨로지, 시맨틱 웹, Social Network 등이 있다.