OntoFrame S3: Semantic Web-Based Academic Research Information Portal Service Empowered by STAR-WIN

Seungwoo Lee, Mikyoung Lee, Pyung Kim, Hanmin Jung, and Won-Kyung Sung

Korea Institute of Science and Technology Information (KISTI),
335 Gwahangno, Yuseong-gu, Daejeon, Korea 305-806
{swlee,jerryis,pyung,jhm,wksung}@kisti.re.kr

Abstract. We have developed a prototype of a practical knowledge-driven semantic portal, OntoFrame S3, which provides various reasoning-based analysis services on academic research information. To realize this semantic portal, we developed and applied several Semantic Web and linguistic technologies. Through this demonstration, we will show how Semantic Web technologies can be utilized for information connection and fusion in the academic research information service sector and empowered by linguistic knowledge.

Keywords: Academic research information portal service, semantic search, ontology, reasoning, semantic word network.

1 Introduction

According to the US National Science Foundation (NSF)[1], researchers are spending more than half of their total research and development hours for hunting information. Thus, to allow researchers to have more time to spend on research and development itself, it is crucial to reduce time spent on gathering information. Currently, leading search engines only provide a keyword-based matched list as the result of a search query, which is limited in terms of accuracy and efficiency of information comprehension. A new type of information service is required that can find the information desired by the researcher, and then connect, combine, and analyze it to provide as much value to the user as possible. To address this need, we have developed a prototype of a knowledge-driven semantic portal, OntoFrame S3[2], which provides various analysis services on academic research information combining OntoFrame and STAR-WIN, which are implemented by Semantic Web and linguistic technologies.

In this demonstration, we will show you various semantic analysis services focused on topics and researchers. Topic-focused services include trends, leading researchers and researcher networks on a topic and researcher-focused services include research areas, research outputs, co-working researchers, and similar researchers of a researcher. More detail description about these services is given in the next section.

[1] http://www.nsf.gov/
[2] S3 means STAR-WIN-enabled Semantic Service.

L. Aroyo et al. (Eds.): ESWC 2010, Part II, LNCS 6089, pp. 401–405, 2010.

2 Related Work

There are some research information service systems similar to OntoFrame S3. BiomedExperts (www.biomedexperts.com) and Authoratory (authoratory.com) provide useful services such as domain experts, researcher and article details, statistics, researcher network and researcher map in biomedical domain mainly dependent on PubMed[3] database. ResearchGATE (researchgate.com) also provides similar network community services like domain experts, researcher details, and researcher network. It requires users to provide the profiles of their own and colleagues. These service systems have very similar features to ours in the viewpoint of service functions. However, they are typical database applications based on data mining technologies whereas OntoFrame S3 is developed with Semantic Web technologies such as ontology and reasoning, which enable to achieve a flexible and precise service in both connecting knowledge and planning services.

3 Technologies Used

OntoFrame S3 shows a service example that combines Semantic Web technologies, implemented on OntoFrame, and linguistic technologies, implemented in STAR-WIN.

OntoFrame is an information service platform that uses Semantic Web technologies. It includes OntoURI – a semantic knowledge management tool that creates ontology instances by referring ontology schemata and resolves co-references between ontology individuals; and OntoReasoner – a reasoning engine that stores and infers ontology-based RDF triples and answers SPARQL queries. It also involves Mariner – a commercial search engine provided by DiQuest (www.diquest.com) – to provide search functionalities. STAR-WIN (Science & Technology Assister – Word Intelligent Network) is a semantic word network of technical terms that represents semantic and conceptual relationships among the terms in science and technology domain. In the following subsections, we explain each component in detail.

3.1 Ontology Engineering

The goal of OntoFrame S3 is to provide connection, fusion, and analysis services on academic research information to enable researchers to effectively obtain information needed for their work. In order to achieve the goal, we have developed the KISTI Reference and Academic Ontologies[4]. They model research agents such as persons and institutions, their accomplishments such as articles, reports and patents, publications which indicate specific journal issues or proceedings, locations and topics. In contrast with other research-related ontologies, our ontology connects researchers to their affiliations of which they were members at the time they had their accomplishments. It also connects institutions to their locations such as countries, states and cities. The ontology schemata are composed of 16 classes and 89 properties, and designed using RDF, RDFS and OWL (strictly, OWL-Lite) vocabularies.

[3] http://www.ncbi.nlm.nih.gov/pubmed/

[4] http://isrl.kisti.re.kr/ontologies/ReferenceOntology1_0.owl,
http://isrl.kisti.re.kr/ontologies/AcademicOntology1_0.owl

Currently, OntoFrame S3 is in service as a practical prototype (http://www.ontoframe.kr/S3/english). It contains about 4 million articles in English, which have been written by 12.7 million researchers, on 400,000 topics, who work for 90,000 institutions, spanning 410,000 locations. In total, the system has a total of about 826 million RDF triples. These data were populated partly from CiteSeer OAI (Open Archive Initiative) metadata[5] and partly from NDSL (National Digital Science Links) metadata[6]. We applied an elaborate process to identify and disambiguate same-name authors and populate them into ontology instances [1].

3.2 Storing and Reasoning

The populated ontology instances are stored and inferred using OntoReasoner [2]. That is, OntoReasoner consists of roughly two parts: Triple store and Reasoner. The triple store gets ontology instances and stores them to back-end DBMS such as MS-SQL Server according to predefined table schema. It also can answer to a query represented in SPARQL, which is converted into appropriate SQL and executed. The reasoner performs rule-based reasoning based on RDF Semantics[7] and OWL Semantics[8] in ways of forward-chaining. It currently supports entailments of full RDFS and some of OWL vocabularies such as *owl:inverseOf*, *owl:sameAs*, and *owl:TransitiveProperty*.

3.3 Semantic Word Network

OntoFrame S3 is empowered by STAR-WIN to get linguistic flexibility. STAR-WIN is a semantic work network for Korean language developed by KISTI and represents semantic and conceptual relationships between words [3]. It covers technical terms[9] of science and technology domain. It also includes Korean-English translation information for technical terms. Thus, it is highly applicable to linguistic information applications such as information retrieval and machine translation.

Linguistic flexibility of OntoFrame S3, provided by STAR-WIN, includes following three points: auto-complete of search keywords, semantic word network for keyword extension, and Korean-English translation of search keyword. Auto-complete of keyword also functions as keyword recommendation. Semantic word network shows various relations such as broad and narrow terms, sibling terms, similar terms and other related terms. Keyword translation makes it possible to query in Korean although the ontology instances only have English literals.

4 Services

OntoFrame S3 is designed to be topic and researcher-centric (Fig. 1). This is because authors (i.e., researchers) and research topics are useful starting points for finding academic research information. When a user performs a search, a keyword-based search engine is used to find results, but URIs for the terms are also retrieved and

[5] http://citeseer.ist.psu.edu/oai.html
[6] http://www.ndsl.kr
[7] http://www.w3.org/TR/rdf-mt/
[8] http://www.w3.org/TR/owl-semantics/
[9] Currently, it contains about 350,000 technical terms.

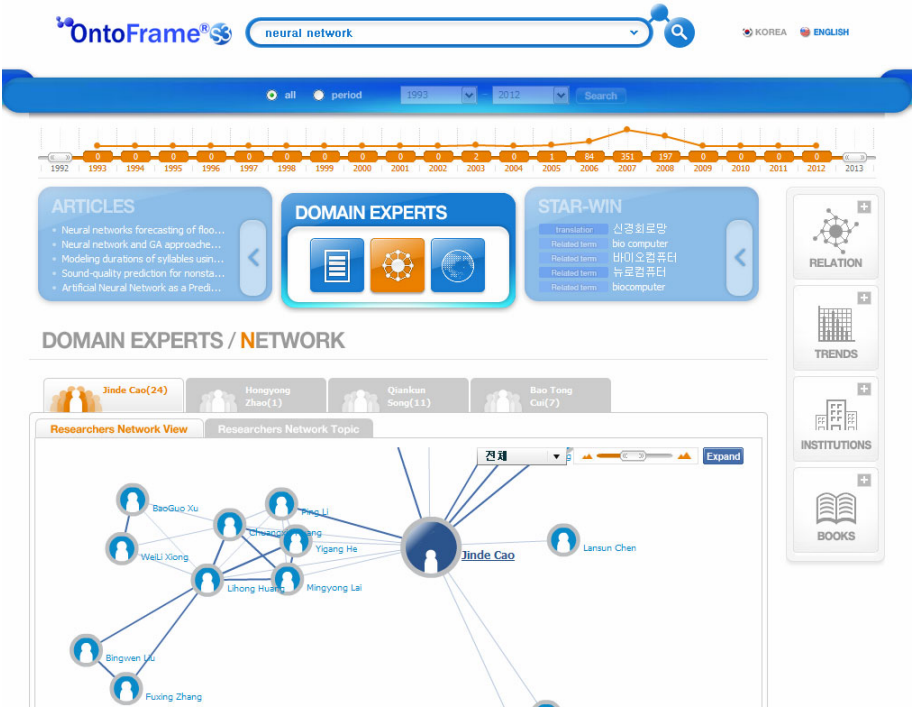

Fig. 1. OntoFrame S3 service snapshot

used to search the ontology instances for more information relating to the term. The results are presented to the user by appropriate visualization.

The semantic portal currently implements and provides many service components such as topic trends, identification of leading researchers or research institutes for a topic, main research topics of a researcher, researcher publication network, research trends of a publication network, recommendation of associated researchers, similar researchers, similar papers, and geographical distribution of researchers. These services provide knowledge that has been analyzed by connecting and fusing fractional information along with proper visualization.

The service prototype has been designed and developed considering user-centric service layout and usability. We selected three main services considering user preference and distinguished these services from add-on services considering service priority. Through this, each service component was properly placed at main area or right area of service user interface.

5 Conclusion

We have developed a prototype of a knowledge-driven semantic portal, OntoFrame S3, which provides various analysis services on academic research information combining OntoFrame and STAR-WIN. This semantic portal service shows how Semantic Web

technologies can be utilized for information connection and fusion in the academic research information service sector and empowered by linguistic knowledge.

We will continue to evaluate the services and develop new useful services to open real services to the public. Thus, we plan to measure the usefulness of service attributes by comparatively evaluating OntoFrame S3 and other similar service systems mentioned in the Related Work.

References

1. Kang, I.-S., Na, S.-H., Lee, S., Jung, H., Kim, P., Sung, W.-K., Lee, J.-H.: On co-authorship for author disambiguation. Information Processing and Management 45, 8–97 (2009)
2. Lee, S., You, B.-J.: Memory and DBMS-based Hybrid Reasoning. In: Korea Computer Congress 2009, Korea, pp. 122–123 (2009) (in Korean)
3. Lee, S.-H., Choe, H.-S., Yoon, H.-M., You, B.-J.: KISTI-STAR: Semantic Technology based Science and Technology Knowledge Information Service. In: Service Science Society 2009 Spring Conference, Korea, pp. 1–8 (2009) (in Korean)