

# SEMANTIC WEB-BASED SERVICES FOR SUPPORTING VOLUNTARY COLLABORATION AMONG RESEARCHERS USING AN INFORMATION DISSEMINATION PLATFORM

Hanmin Jung<sup>1\*</sup>, Mikyoung Lee<sup>1</sup>, Won-Kyung Sung<sup>1</sup>, and Dong-In Park<sup>1</sup>

<sup>1</sup>Information System Research Lab., KISTI, Korea  
Email: {\*jhm, jerryis, wksung, dipark}@kisti.re.kr

## ABSTRACT

Information dissemination platforms for supporting voluntary collaboration among researchers should assure that controllable and verified information is being disseminated. However, previous related studies on this field narrowed their research scopes into information type and information specification. This paper focuses on the verification and the tracing of information using an information dissemination platform and other Semantic Web-based services. Services on our platform include information dissemination services to support reliable information exchange among researchers and knowledge service to provide unrevealed information. The latter is also divided into the two: knowledgization using ontology and inference using a Semantic Web-based inference engine. This paper discusses how this platform supports instant knowledge addition and inference. We demonstrate our approach by constructing an ontology for national R&D reference information using 37,656 RDF triples from about 2,300 KISTI (Korea Institute of Science and Technology Information) outcomes. Three knowledge services including 'Communities of Practice', 'Researcher Tracing,' and 'Research Map' were implemented on our platform using a Jena framework. Our study shows that information dissemination platforms will make a meaningful contribution to the possibility of realizing a practical Semantic Web-based information dissemination platform.

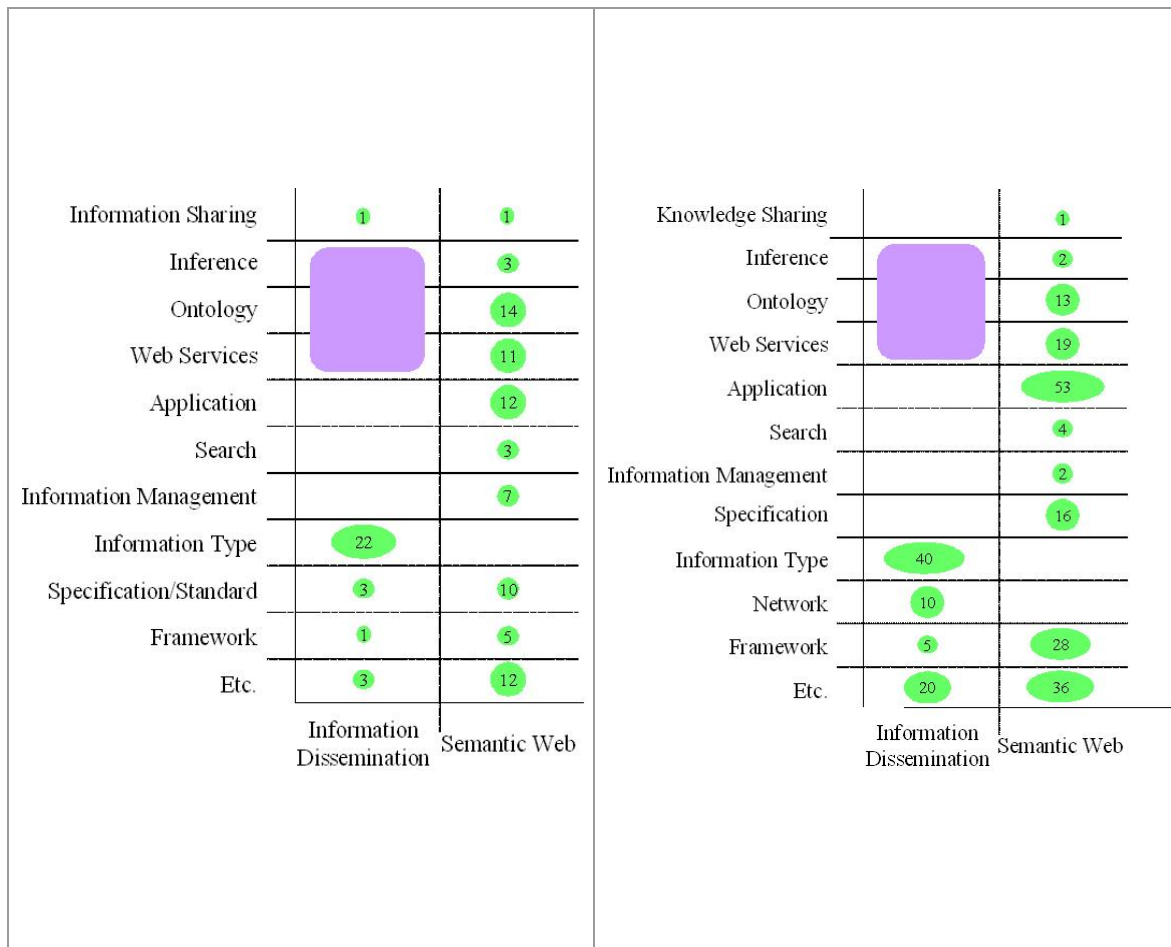
## 1 INTRODUCTION

The importance of systematized information dissemination based on knowledge management systems or content management systems grows more and more in an environment in which information is intemperately created in great quantities. However, studies on knowledgization and information verification, in particular, by Semantic Web-related communities, were independently performed regardless of information dissemination workflow. The separation between the two research fields brought on Semantic Web-based services without consideration of dynamic information life cycles including creation, storage, management, search, and dissemination.

In this paper, we introduce Semantic Web-based technologies that provide knowledge services including knowledgization and inference on an information dissemination platform. The platform also verifies and traces information using a document container. This combination of information dissemination with Semantic Web eventually enables dynamic service for knowledgized information. We apply our platform in support of voluntary collaboration among researchers.

## 2 PREVIOUS STUDIES

Many studies on information dissemination including P2P (Peer-To-Peer) systems have low level information verification mechanisms such as document identifier checking. Clarke et al. (2000) and Lee et al. (2005) assign document IDs to media content files using header information, and thus, identical files would have the same document ID. However, in cases where users intentionally modify header information but it is still necessary to trace document versions, this approach can not provide an appropriate solution. It is also difficult to support collaboration among users because additional metadata can not be inserted into documents or containers. The verification and tracing of registered documents are crucial in disseminating information with reliability. Choi et al. (2004), Lee et al. (2004), and DVB (2003) also studied information verification but did not consider the logical connection of a current document with its ancestor documents or update possibility by other users.

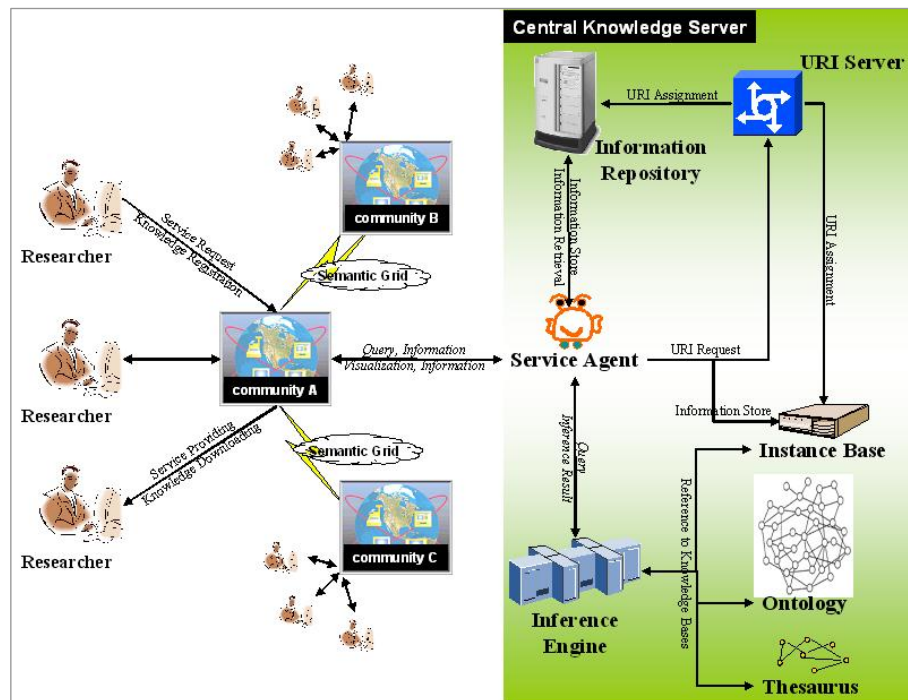


**Figure 1.** Paper Map on Information Dissemination and Semantic Web (Left: Korean Papers, Right: Overseas Papers)

To resolve some of the above problems, we introduce the concept of a “document container” that includes the document itself, metadata, and management information, to control access privileges and to trace versions. It is also possible to provide ontology-based inference by allowing metadata registering. Cho et al. (2004) show a verification method for RDF (Resource Description Framework) documents by using RDF parsing. To eliminate this load of RDF validity check, we designed a URI-based user interface. It automatically assures that a user’s query is valid by restricting input as metadata with URIs (Universal Resource Identifiers).

There are only a few successful Semantic Web applications such as CAS (CS AKTive Space) of AKT (Advanced Knowledge Technologies) project and SEAL (Semantic portAL). Even in Korea, no case can be yet found to manage real data. Previous Semantic Web-based services were designed and implemented to provide static knowledge service, that is, they have no real-time knowledgization process to reflect up-to-date information (Jung et al., 2005b) (Seon et al., 2004) (CAS) (SEAL). On the other hand, our platform adds *DB-to-OWL Conversion* process to update new information instantly. The results of the process, i.e. individuals, are also inserted into the main ontology for keeping knowledge consistency. Figure 1 shows a map comparing papers on both information dissemination and the Semantic Web. The results are from the analysis of 23 information dissemination-related Korean papers and 52 Semantic Web-related from 2000 to 2004. For overseas results, 51 information dissemination-related and 154 Semantic Web-related papers are used (source: <http://ndsl.or.kr/eng/newindex.html>). It is absolutely insufficient for Semantic Web-based services to be on an information dissemination platform. Most of the previous studies focused on the type and the specification of information.

### 3 ONTOFRAME-K: INFORMATION DISSEMINATION PLATFORM



**Figure 2.** Workflow on Information Dissemination Platform (*Community A* is a virtual research community, and it can be connected with other communities such as *Community B & C*.)

OntoFrame-K is an information dissemination platform for supporting voluntary collaboration among researchers. It has a radial shape including a central knowledge server to allow only reliable information exchange and researcher clients to use national R&D reference information and other researchers' information. We also introduce Semantic Web-based architecture for knowledgezation and inference using ontology. OntoFrame-K offers information dissemination services for obtaining intuitive information such as information retrieval and version tracing. It also provides three knowledge services using inference; '*COP (Communities Of Practice)*,' '*Researcher Tracing*,' and '*Research Map*.' An OntoFrame-K can be viewed as a virtual research community for a specific research topic or a study field, and it can be further combined with other OntoFrame-Ks in the manner of connecting between central knowledge servers. A multi-community model is not implemented yet; however, we have a plan for building it with Semantic Web Services and Semantic grid after 2007.

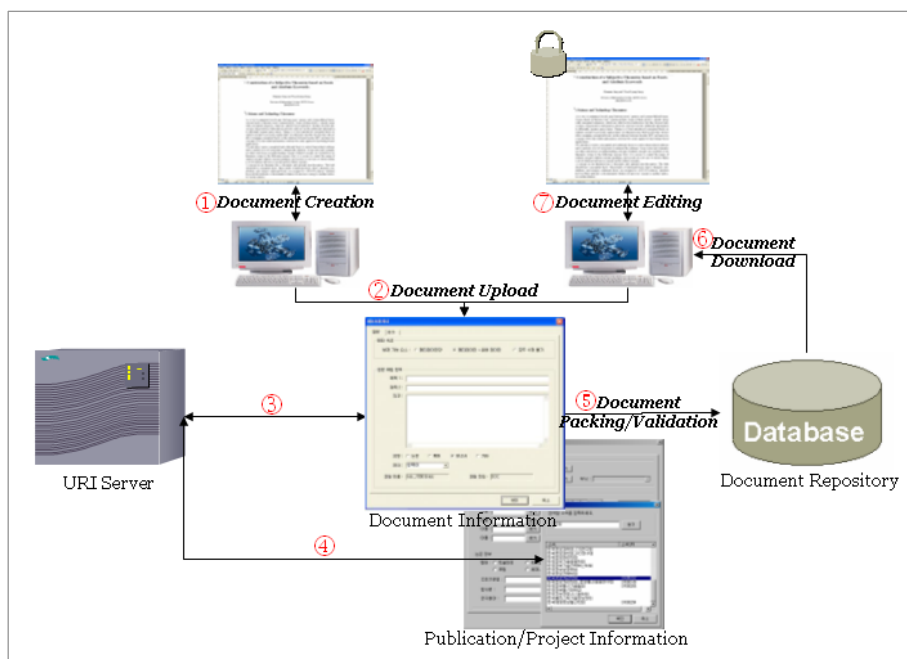
#### 4.1 SERVICES FOR SUPPORTING COLLABORATION AMONG RESEARCHERS

Services for supporting voluntary collaboration among Researchers consist of two components: information dissemination service (*see Section 4.1*) and knowledge service (*see Section 4.2 & 4.3*). The former guarantees voluntary and immediate information exchange on our information dissemination platform. Thus, information dissemination service includes keyword-based document search and related document search with version tracing. On the other hand, the latter, which includes knowledgezation and inference, provides inference-based services beyond simple information retrieval and question-answering.

##### Information Dissemination Service

Information dissemination service is based on a client-server model. To organize and to provide collaboration, a central knowledge server should be constructed. Verified clients request and download their necessary information to it. After being created or modified, information would be uploaded to the server. A successive workflow of the following events establishes the voluntary information dissemination service.

- ① *Document Creation*: Researchers create documents as outcomes such as publications, reports, and intellectual properties. Document editors, e.g. MS Office, would be used.
- ② *Document Upload*: Created documents and document containers (see ⑤ of Section 4.1) are registered into a central knowledge server. Researchers can fill in simple document information such as creation date, file name, file type, and creator, or can add pre-defined metadata for outcomes. In the case of the latter, the metadata are automatically knowledgized for inference (see Section 4.2).
- ③ *URI Assignment for Document*: URI (Universal Resource Identifier) server assigns a URI to each document object as document ID. IDs are automatically connected with document metadata for further providing document contents.
- ④ *URI Assignment for Metadata*: For the fields defined as class type in an ontology, URIs are assigned during publication/project information registration. Each URI should have a prefix for a specific class, e.g. “OR” for organization class. The URI server generates an appropriate URI for a requested field. This process is the same as ① of section 4.2.



**Figure 3.** Information Dissemination Service Process

- ⑤ *Document Packing/Validation*: Previous P2P (Peer-To-Peer) systems have a problem in blocking illegal information dissemination because of the weakness of their own document verification mechanisms. They have few mechanisms to control transformation and reproduction of original data. To provide smooth collaboration among researchers, the information dissemination service should protect unverified documents and manage document history for version tracing. We introduce the document container as a method of satisfying these requirements. All of the registered documents after ② are verified and packed. A document container as the result of document packing additionally contains document metadata such as simple document information (creation date, creator, and so on) and three document IDs for originator, parent, and itself. Users can access document containers corresponding to their authorities. Three access types are possible: Read-Only, Read-Write (metadata + text), and Read-Write (only metadata). On finishing editing in a document container, the document is automatically re-packed. The knowledge server verifies both documents and document containers. In the case of the container, metadata in the uploaded container are compared with related information in document repository.
- ⑥ *Document Download*: Researchers can download information using information retrieval and inference. Selected information is provided in the form of a document container.
- ⑦ *Document Editing*: Document container activates menus within user’s authority when it is opened.

## 4.2 Knowledgization

Knowledgization is defined as the process to systematize national R&D reference information (metadata) using ontology. Inference uses RDF (Resource Description Framework) triples which are the results of knowledgization.

- ① *URI Assignment for Metadata*: This is the process that calls the URI server to acquire URIs for metadata (see ④ of Section 4.1).
- ② *Metadata Storing*: Registered metadata (literals and URIs) are stored into a RDBMS-based metadata registry.

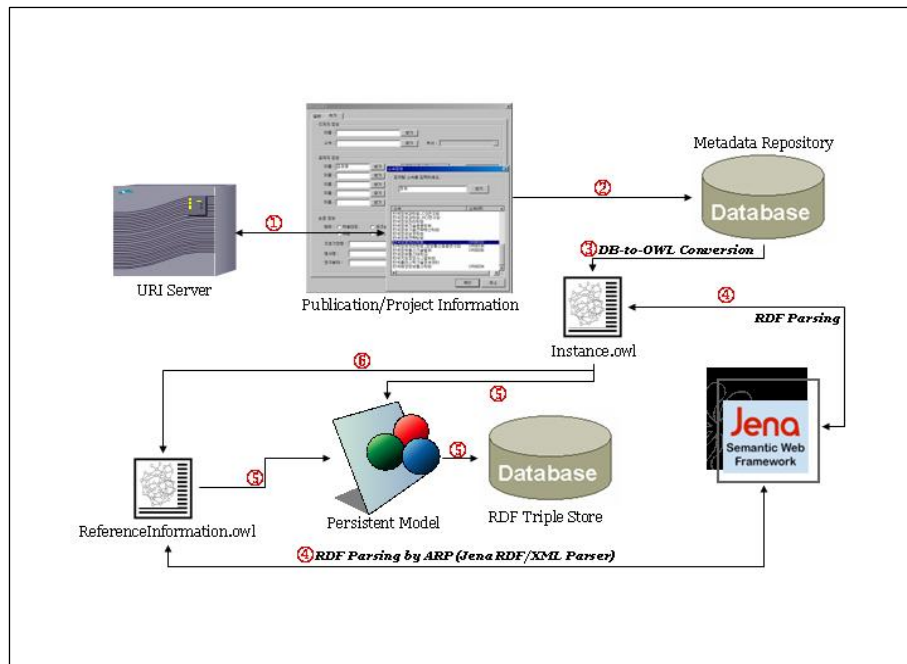


Figure 4. Knowledgezation Process

③ *DB-to-OWL Conversion*: This process creates an OWL (Web Ontology Language) document for instance. It is dependent on service scenarios thus is influenced by ontology modification. As another method, we can do direct DB-to-Triple Conversion without OWL document for instances. However, this alternative is more dependent on service scenarios and ontology. It is also necessary for the design of another module to update an OWL document for the reference information. Our DB-to-OWL Conversion process is indispensable to real-time metadata knowledgezation and inference with up-to-date information. Ontology loading time is also reduced because we only have to reflect an instance OWL document to RDF triple store rather than loading a whole ontology for reference information whenever new information is inserted. Finally, our system establishes a practical service system by instantly using real-time information.

④ *RDF Parsing*: RDF/XML parser (ARF; Another RDF Parser) of Jena (<http://jena.sourceforge.net>) creates a persistent model by parsing OWL document for instances (Instance.owl), or an OWL document for the reference information (ReferenceInformation.owl).

⑤ *Persistent Model Creation*: OWL-based ontology is transferred to RDF triples (see Table 1) and is stored into RDBMS, which is the repository of Jena.

Table 1. Example of RDF Triples for National R&D Reference Information Acquired from KISTI outcomes

Subject	Predicate	Object
nlp:신기술아이디어 사업화타당성 평가사업_2003 <sup>1</sup>	nlp:hasProjectMember	nlp:서성호
nlp:Electrospinning_Nanofiber_양산기술평가_및_시장분석	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	nlp:SuperintendProject
nlp:Electrospinning_Nanofiber_양산기술평가_및_시장분석	nlp:hasBeginningDate	"2003-06-10"^^http://www.w3.org/2001/XMLSchema#string
genid:ARP1344001 <sup>2</sup>	http://www.w3.org/1999/02/22-rdf-syntax-ns#first	nlp:Workshop

<sup>1</sup> "nlp:" is the prefix of "http://nlplab.ulsan.ac.kr/kisti/ontology/BaseInformation.owl#."

<sup>2</sup> Bnode (blank node) generated by Jena.

⑥ *Instance Appending to OWL-based Ontology*: Individuals acquired from DB-to-OWL conversion are appended into an OWL document for national R&D reference information. The OWL document is used for batch processing and ontology optimization. In the case that RDF triple store is initialized, RDF triples are re-created from this OWL document. ④ and ⑤ connected with OWL document for the reference information are for the batch processing.

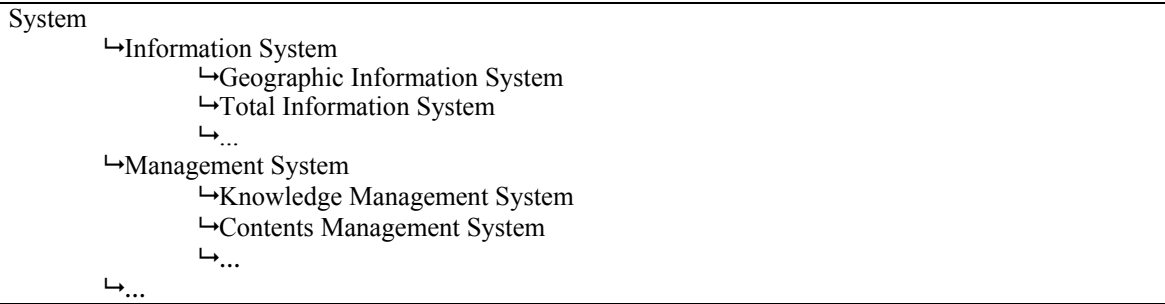
### 4.3 Inference

Information dissemination service provides intuitive and obvious information. On the other hand, inference acquires unrevealed information using facts, rules, and even post-processing to manipulate retrieved information. We introduce expert ranking module for *Communities of Practice*.

① *RDQL Creation*: We design pre-defined RDQL (RDF Data Query Language) templates for three knowledge services, i.e. *Communities of Practice*, *Researcher Tracing*, and *Research Map*. When users select and input arguments (literal for research topic, URI for researcher, and so on), the arguments are inserted into the pre-defined RDQL templates. The subsequent box shows an example of RDQL for *Communities of Practice*. The research topic selected by the user is “system.” It requires project names (?project), cooperating researchers (?member), and project funds (?fund).

```
SELECT    ?project, ?member, ?fund
WHERE    (?project rdf:type nlp:Project),
         (nlp:system nlp:isResearchTopicOf ?project),
         (?project nlp:hasProjectMember ?member), (?project nlp:hasFund ?fund)
```

② *Thesaurus Loading*: To resolve lexical disagreement between system-embedded knowledge and a user’s query (e.g. research topic), the knowledge server adopts a thesaurus (Jung et al., 2006). In the case that the user’s research topic is matched with a concept node of thesaurus, the system automatically expands the user’s topic with all of the concepts in the sub-tree of the matched concept node. Inference uses this expansion to enlarge retrieval coverage. The sub-tree of “system” on thesaurus is as follows.



③ *Inference*: Jena as our inference engine uses RDF triples and inference rules and produces instances or triples. We design five rules for rdfs:subClassOf property and five for root classes (Class:IntellectualProperty, Class:Project, and Class:ResearchTopic). The subsequent rule is for a research topic.

```
(?x nlp:subTopicOf ?y) (?y nlp:subTopicOf ?z) → ( ?x nlp:subTopicOf ?z)
```

④ *Inference Result Generation*: This process accomplishes service-dependent post-processing to make input data for visualization from inference results. It also generates statistical information such as the number of outcomes by year.

⑤ *Inference Result Visualization*: We provide knowledge service including inference with Flash-based visualization to provide information easily to researchers (see Appendix B).



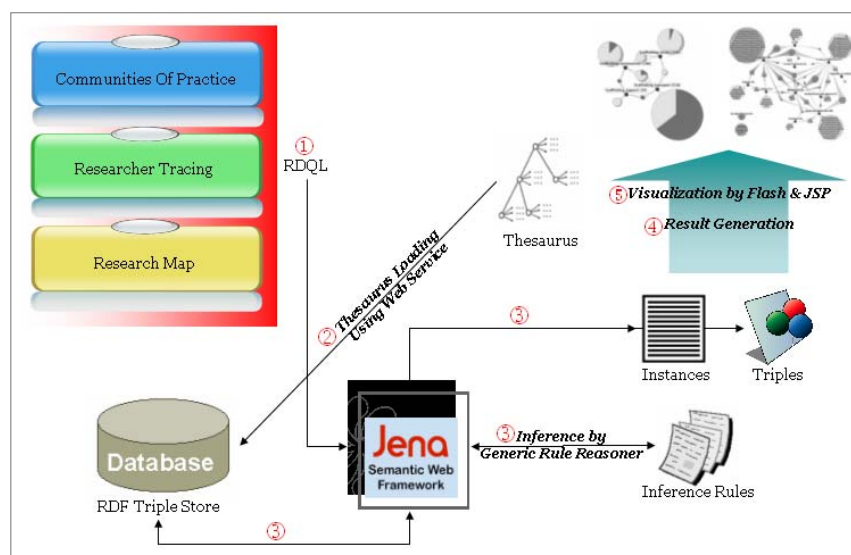


Figure 5. Inference Process

## 5 EXPERIMENTAL RESULTS

Our ontology for national R&D reference information consists of seven top classes: *Event*, *Group*, *Intellectual Property*, *Person*, *Project*, *Publication*, and *Research Topic*. The whole number of classes is 74. We also defined 75 properties and several restrictions including someValuesFrom, allValuesFrom, and minCardinality. As reference information, about 2,300 documents with metadata were collected from KISTI (Korea Institute of Science and Technology Information) outcomes. Jena finally generated 37,656 RDF triples from them. The thesaurus for query expansion includes about 15,000 concepts which were manually selected out of about 50,000 terms extracted by means of term life cycle methodology (Jung et al., 2005a). We currently provide the information dissemination service (see Appendix A) and knowledge service (see Appendix B) on the Web.

## 6 CONCLUSIONS

We described an information dissemination platform for supporting voluntary collaboration among researchers. The platform is based on the Semantic Web to knowledgeize information instantly. It is also able to verify and to trace registered information using a document container. It is possible to say that the platform showed a new paradigm to disseminate information owing to ontology-based inference on the Semantic Web. For future work, we have a plan to achieve a knowledge-based fusion platform by interlocking the services on OntoFrame-K.

## 7 REFERENCES

- Cho, S., Song, B., Cho, H., & Choi, E. (2004) RDF Document Validator and N-Triple Generator for Creation and Maintenance of Metadata. *Journal KIPS 11-B(5)*.
- Choi, S., Lee, S., Lee, S. Cho, S., & Suk, J. (2004) A Study on the Digital Content Dissemination Model Based on Identification and Public Key. *Proc. the 31<sup>st</sup> KISS Fall Conference*.
- Clarke, I., Sandberg, O., Wiley, B., & Hong, T. (2000) Freenet: A Distributed Anonymous Information Storage and Retrieval System. *Proc. the International Workshop on Designing Privacy Enhancing Technologies*.
- Jung, H., Koo, H., Lee, B., & Sung, W. (2005a) Toward Managing the Life Cycle of Terms Using Term Dominance Trend. *Proc. the Pacific Association for Computational Linguistics*.
- Jung, H., Seon, C., Kim, J., Son, J., Sung, W., & Park, D. (2005b) Information Extraction for User's Utterance Processing on Ubiquitous Robot Companion. *Journal LNCS 0302-9743*.
- Jung, H., Sung, W., & Park, D. (2006) Korean Science & Technology Thesaurus Construction Based on Concept/Relation Facets. *Proc. the 3<sup>rd</sup> Global WordNet Conference*.

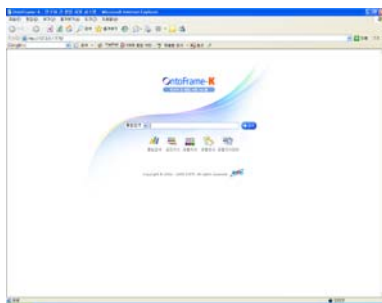
Lee, H., Choi, B., Hong, J., & Seo, J. (2004) Design and Implementation of a Protection and Distribution System for Digital Broadcasting Contents. *Journal KIPS 11-C* (6).

Lee, W., Kim, S., & Jeon, C. (2005) An Efficient False Data Reduction Algorithm in P2P Environment. *Proc. KCC2005*.

Seon, C., Jung, H., Kim, J., & Son, C (2004) Formal Description Generation from Users Query on Web-based Service Platform. *Proc. the 1<sup>st</sup> International Workshop on Intelligent and Ubiquitous Robotics*.  
CAS: <http://triplestore.aktors.org/demo/AKTiveSpace/>

DVB CPCM Functional Model Definitions and Applications. DVB-CPT (2003) SEAL. Retrieved from the WWW April 15, 2007: [http://www.aifb.uni-karlsruhe.de/Projekte/viewProjektenglish?id\\_db=58](http://www.aifb.uni-karlsruhe.de/Projekte/viewProjektenglish?id_db=58)

## 8 Appendix A: Screen Shots on Information Dissemination Service



(a) OntoFrame-K main page



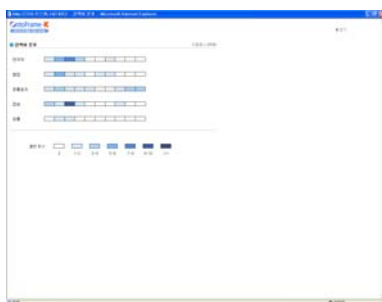
(b) Document Search



(c) Version Tracing



(d) Metadata



(e) Keyword Distribution



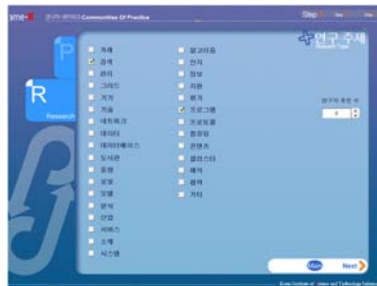
(f) Metadata Registering Interface



## 9 Appendix B: Screen Shots on Knowledge Service



(a) Knowledge Service Main Page



(b) COP: Research Topic Selection



(c) COP: Researcher List



(d) COP: Reference Information



(e) Researcher Tracing



(f) Research Map (for KISTI)