

# Technology-Based Decision-Making Support System

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

KISTI, 52-11 Eueon-dong, Yuseong-gu  
Daejeon, KOREA 305-806  
jhm@kisti.re.kr

**Abstract.** This paper describes a decision-making support system focused on technologies, R&D agents, and R&D results. To deal with heterogeneous literatures and metadata, we introduce text mining and Semantic Web-based service platforms. InSciTe, a decision-making support system developed by us, provides a through process including analysis as well as ETL, verifies search and analysis results, connects its information with Semantic Web open sources in the level of RDF, and generates automatic summary reports. This system is significant in the sense that it has been implemented about a year earlier than similar projects such as CUBIST and FUSE.

**Keywords:** Decision-Making Support, Text Mining, Semantic Web, Technology Intelligence Service.

## 1 Introduction

Thanks to the dramatic development of the web, the range of the use of information in information services based on technological information likewise is changing gradually. Information is usually represented as metadata which can be defined data about data. There are two kinds of metadata; explicit metadata as formal documentation of resources and implicit metadata as automatic creation of description of resources. While, in the past, explicit metadata were loaded on to information analysis tools such as VantagePoint and Thomson Data Analyzer [1] [2], and human analysts gave insights to support decision-making processes under a clear goal, recent projects such as CUBIST, CTI, and FUSE have tended toward the additional use of implicit metadata hidden in documents by using automatic language processing techniques such as text mining, information extraction, and semantic annotation [3] [4] [5]. Such a tendency stems partly from the dramatic increase in the Web documents, but neither explicit metadata nor implicit metadata is sufficient in supporting decision-making processes.

Though we know implicit metadata is a very useful resource to get insights, it is difficult to discover meaningful data, to transfer to structured information, and to further merge with explicit metadata in a uniform way. A case study on ACRC<sup>1</sup> shows this difficulty clearly. Only 23% of civil complaints are currently analyzed by human analysts, and no more to be interpreted only by humans to be of any use. Thus

---

<sup>1</sup> The Anti-Corruption & Civil Rights Commission.

the results produced by the analysts are far from a comprehensive and systematic analysis. It also often leads mistrust on the results due to a large variation in its quality which is mainly caused by difference in the experience levels of analysts.

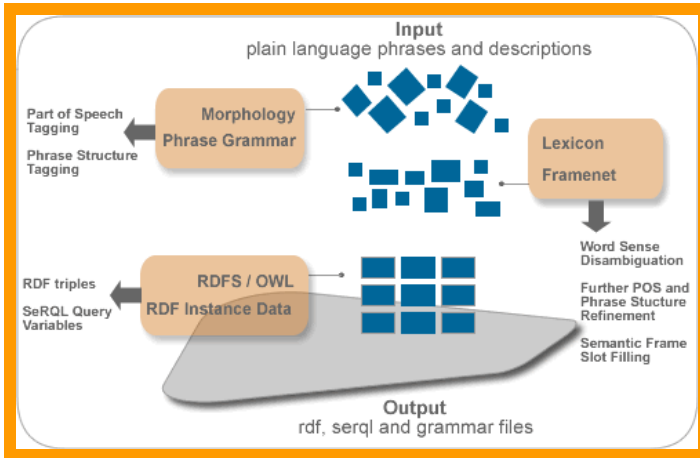


Fig. 1. Process for Transferring Implicit Metadata into Structured Information<sup>2</sup>

We would like to describe how explicit and implicit metadata are merged in an information infrastructure to implement decision-making support system. Semantic Web technologies such as ontology model and reasoning were introduced to provide such infrastructure, and text mining technology to acquire structured information from textual documents. The system includes a lot of service components which interact with search and reasoning engines.

## 2 System Description

The growth and availability of data and, therefore, our need to consider it in decision-making and planning is growing exponentially [6]. Thus, the purpose of this study lies in supporting decision-making, especially for establishing R&D strategies, by adopting ontologies, implicit metadata, and explicit metadata based on technology-oriented information acquired from patents and academic papers, and using a Semantic Web-based service platform to implement them as an integrated system called InSciTe<sup>3</sup> [7]. To achieve this goal, the following preconditions are necessary:

- Bring relevant information and analytics together and present them so as to help human decision-makers improve the quality of their decisions.
- Require ETL (extract, transform, and load) on different kinds of literatures such as patents and academic papers, and further resolve semantic differences as well as syntactic differences.

<sup>2</sup> <http://www.monrai.com/products/cypher/img/ad-framework.gif>

<sup>3</sup> Intelligent Science and Technology Service.

- Use semantic annotations and Semantic Web-based service platform to perform the ETL process effectively.

In the decision-making process for the establishment of R&D strategies, the key elements are technologies, R&D agents (researchers, institutes, and nations), and R&D results (patents and academic papers). In particular, as a medium that links R&D agents and R&D results, technologies serve to determine the direction of the establishment of R&D strategies. This study focuses on effectively supporting decision-making by extracting, from academic papers and patents, the relationships among technologies and, with them at the center, analyzing the activities of R&D agents from the perspectives of cooperation and competition.

InSciTe, a technology intelligence<sup>4</sup> service developed by us, considerably differs from VantagePoint, which is representative information, especially for patents and academic papers, analyzer, in data size, analysis cost, user expertise, and focus strength. A detailed comparison of the two is shown in Table 1.

As shown in fig. 2, InSciTe lies on 'Decision Support' level. The level requires a text mining platform for extracting meaningful information from scientific literatures as one of preceding tasks and a service platform which can deal with both explicit and implicit metadata. We use SINDI<sup>5</sup> (see Fig. 3) as a tool to acquire technical terms and relations among them.

The major functions of InSciTe can be summarized as follow:

- Providing a through process including analysis as well as ETL<sup>6</sup> by combining text mining and Semantic Web technologies
- Verifying search and analysis results by applying reasoning verification to show how the results were induced
- Connecting with Semantic Web open sources in the level of RDF for enhanced information accessibility (e. g., Linked Data, Open Calais)
- Blending heterogeneous sources for multi-faceted explanation (e.g. academic view vs. business view)
- Generating summary reports through automatic processing with hierarchical condition check (e.g. technology report and institution report)

The data set used to implement InSciTe encompasses about 336,000 academic papers, over 400,000 patents, and about 70,000 technological terms in the field of green technology. They are loaded into the following ontology model. It includes a Topic class representing technical domain, Person, Institution, and Nation classes

---

<sup>4</sup> Technology intelligence can be defined as an activity that enables companies to identify the technological opportunities and threats that could affect the future growth and survival of their business ([http://en.wikipedia.org/wiki/Technology\\_intelligence](http://en.wikipedia.org/wiki/Technology_intelligence)). It slightly differs from competitive intelligence which widely covers the action of defining, gathering, analyzing, and distributing intelligence about products, customers, competitors and any aspect of the environment needed to support executives and managers in making strategic decisions for an organization ([http://en.wikipedia.org/wiki/Competitive\\_intelligence](http://en.wikipedia.org/wiki/Competitive_intelligence)) in that technology intelligence focuses on technology-oriented analysis rather than overall analysis.

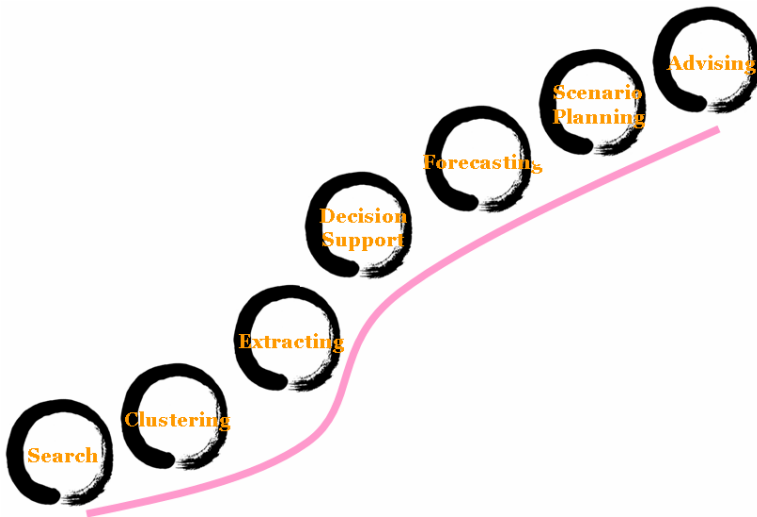
<sup>5</sup> Scientific **IN**formation **DI**scovery Platform.

<sup>6</sup> **E**xtract, **T**ransform, and **L**oad.

denoting research agents, and Patent and Article classes representing research results, and consists of 17 classes, 57 datatype properties, and 37 object properties.

**Table 1.** Detailed Comparison of InSciTe and VantagePoint

|                              | <b>InSciTe</b>                            | <b>VantagePoint</b>  |
|------------------------------|---|--|
| <b>Data size</b>             | - tens of millions of records             | - 20,000 records   |
| <b>Target users</b>          | Planners, experts, chief officers, et al. | Analysts, consultants                                      |
| <b>DB</b>                    | Metadata, full-text (DB2OWL)              | Bibliographic database (import filter)                     |
| <b>Dimension of analysis</b> | Multi-dimensional                         | 2-dimensional (co-occurrence matrices, maps, and networks) |
| <b>Text mining level</b>     | Entity/relation extraction                | Keyword extraction   |
| <b>Service type/method</b>   | Canned services<br>Pull and push services | DIY, scripting<br>Pull services                            |
| <b>Type</b>                  | Web-based                                 | Stand-alone  |
| <b>Others</b>                | Ontology model                            | Expectancy value using the Bernoulli process               |



**Fig. 2.** Value Pyramid with the Difficulty Level of Implementation (Modified from [8])

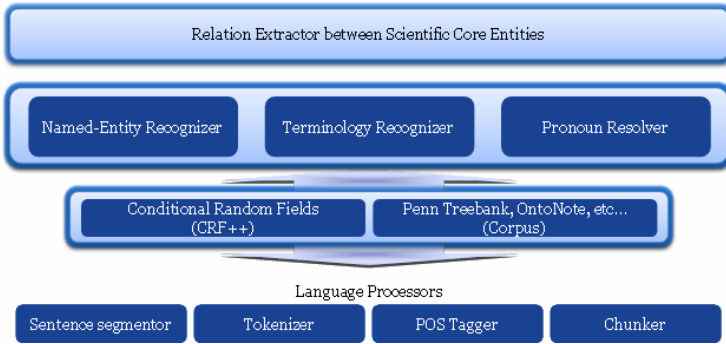


Fig. 3. SINDI Platform

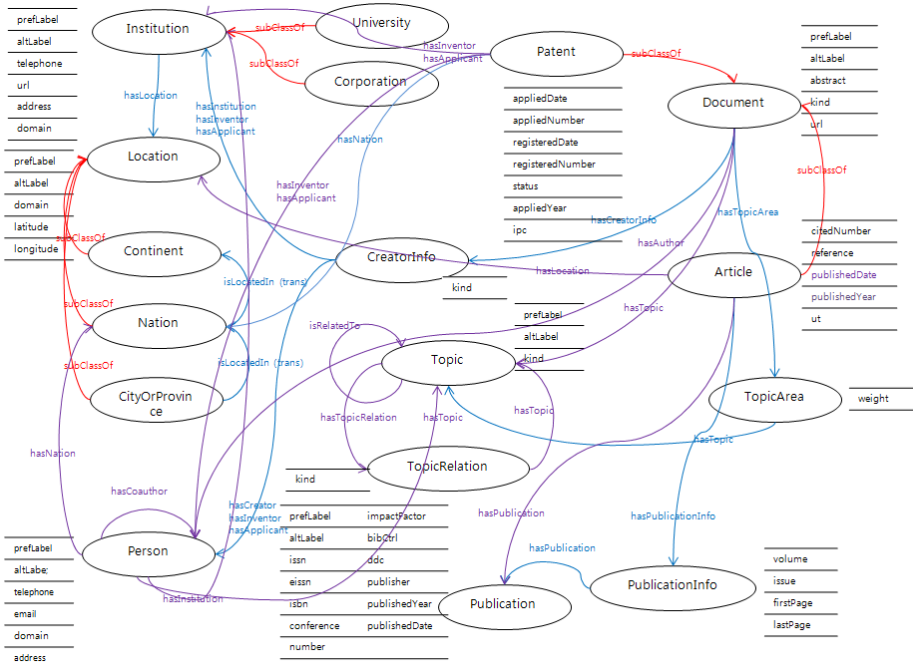


Fig. 4. Ontology Model for InSciTe

As for representative services that support decision-making, there are: Agent/Technology Map, which shows the correlation between R&D agents and related technologies with respect to a given technology from the perspectives of cooperation and competition (see Fig. 5); Technology Trends Map, which shows R&D direction of representative agents for a given technology; Agent Network, which shows R&D agents' cooperation behaviors on both abstract and concrete

levels; and Trends/Institute Summary Report, which automatically summarizes technological trends and research institute's information on R&D.

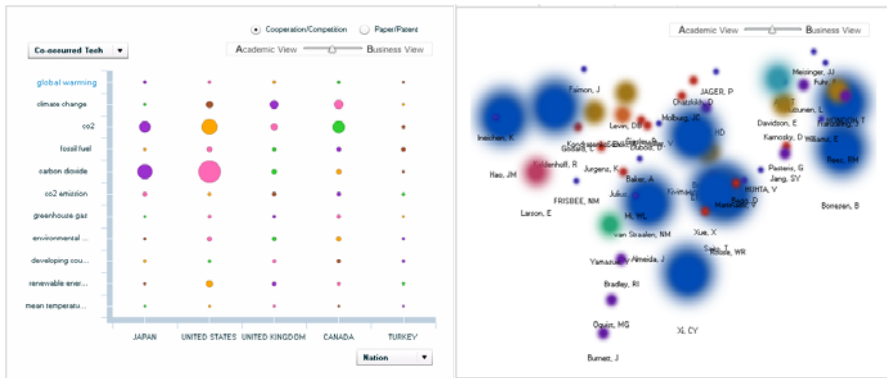


Fig. 5. InSciTe Service Screen (Left: Agent/Technology Map, right: Agent Network)

### 3 Conclusion

This paper showed the trends of information service and emphasized the importance of the use of both explicit and implicit metadata for supporting decision-making. Our decision-making support system consists of two platforms to achieve the goal; text mining platform and Semantic Web-based service platform.

We have several future works such as constructing and getting test suites for sharable test environment and evaluating usefulness of this system targeting human decision makers and research scientists.

### References

1. <http://www.thevantagepoint.com/>
2. [http://thomsonreuters.com/products\\_services/legal/legal\\_products/intellectual\\_property/Thomson\\_Data\\_Analyzer](http://thomsonreuters.com/products_services/legal/legal_products/intellectual_property/Thomson_Data_Analyzer)
3. <http://kn.theiet.org/news/sep10/cubist.cfm>
4. <http://cisti-icist.nrc-cnrc.ca/eng/ibp/cisti/newsletters/cisti-news/2009april.html>
5. [http://www.iarpa.gov/solicitations\\_fuse.html](http://www.iarpa.gov/solicitations_fuse.html)
6. McComb, D.: The CIO's Guide to Semantics v3 (2009), <http://semanticarts.com/wordpress/wp-content/uploads/2011/01/The%20CIO's%20Guide%20to%20Semantics%20v3.pdf>
7. Lee, S., Lee, M., Kim, P., Jung, H., Sung, W.-K.: OntoFrame S3: Semantic Web-Based Academic Research Information Portal Service Empowered by STAR-WIN. In: Aroyo, L., Antoniou, G., Hyvönen, E., ten Teije, A., Stuckenschmidt, H., Cabral, L., Tudorache, T., et al. (eds.) ESWC 2010. LNCS, vol. 6089, pp. 401–405. Springer, Heidelberg (2010)
8. Bousfield, D., Fooladi, P.: Scientific, Technical & Medical Information: 2009 Final Market Size and Share Report, Outsell Inc. (2010)