InSciTe Advanced: Service for Technology Opportunity Discovery

Mikyoung Lee[1], Seungwoo Lee[1], Jinhyung Kim[1],
Dongmin Seo[1], Pyung Kim[1], Hanmin Jung[1], Jinhee Lee[1], Taehong Kim[1],
Hee Kwan Koo[1], Won-Kyung Sung[1]

[1] Korea Institute of Science and Technology Information (KISTI),
245 Daehangno, Yuseong-gu, Daejeon, KOREA 305-806
{jerryis, swlee, jinhyung, dmseo, pyung, jhm, jhlee29,
kimtaehong, hkkoo, wksung}@kisti.re.kr

Abstract. This paper describes *InSciTe Advanced*, which is a technology opportunity discovery (TOD) service to support decision-making on R&D planning. It models TOD based on the life cycle of technologies, analyzes and visualizes associate relations between technologies and agents. Concretely it provides following five major services: (a) *trends and predictions*, (b) *technology levels*, (c) *relationship paths*, (d) *roadmaps*, and (e) *competitors and collaborators*. These services were implemented by combining Semantic Web technologies with text mining technologies and applying to technological literature.

Keywords: Semantic Web, Technology Opportunity Discovery, Technology Trends and Prediction, Life Cycle of Technology, Relationship Path

1 Introduction

Recently, global competition among companies gets intensified gradually and the life cycle of technologies and products gets shorter. In this circumstance, it is essential for companies to discover and identify new technologies and businesses which may give global competitiveness to them after 5 to 10 years. Especially small and medium-sized enterprises need to make use of trends on competing and emerging technologies in near future and their involved industries for their business planning. However, most small and medium-sized enterprises are busy coping with current competitions and could rarely make long-lasting efforts or have systematic methodologies for discovering new technologies and businesses due to high costs. This is a reason why a national system which can timely provide such information for small businesses is asked urgently. This paper introduces such a system, named *InSciTe Advanced*[1], which discovers new technologies or businesses based on technology opportunity discovery (TOD) model and also uncovers complex relationship paths among technologies and research agents. This is very advanced from the previous *InSciTe*, which had focused on only direct relationships between technologies and agents.

[1] http://inscite.kisti.re.kr/2011/

Technology opportunity means a chance of technical innovation [1] and technology opportunity discovery indicates activities that search where such a chance exists. Recent global competition among technologies is summarized as a process of a discovery and preoccupancy of technology opportunities. Therefore, it can be said that core competitiveness of a country or a company on research and development lies in their ability to analyze and apply technology opportunities.

To discover technology opportunity, *InSciTe Advanced* adopts Semantic Web technologies as a framework for representing and managing semantic data and also employs text mining technologies as a tool for automated and intelligent acquisition of semantic data. It analyzes web resources as well as digital contents on science and technology including academic papers and industrial patents, detects technological issues and discovers emerging technologies.

2 System Architecture

InSciTe Advanced system consists of several components, which are roughly grouped into five parts: (1) extracting technical terms and their relations from literature, (2) linking to external semantic data, (3) managing semantic data based on Semantic Web technologies, (4) measuring life cycles of technical terms, and (5) integrating and visualizing information. The overall architecture is shown in Fig. 1 in detail.

SINDI-CORE/LINK is a text mining engine that recognizes core and significant entities from technical literature like papers and patents and generates higher-dimensional, technical knowledge by extracting and processing correlations between the recognized entities [2]. The extracted data as well as metadata are fed into OntoURI, which converts the data into RDF triples and manages URIs of semantic objects. OntoURI is a component of OntoFrame, which is a Semantic Web platform [3]. The URIs are resolved and connected to external semantic data like Linked Open Data (LOD) by OntoURIResolver. OntoURIResolver is a service component for URI resolution and management of various types of objects. Using data from LOD, it resolves the identity of ambiguous objects, recommends a typical URI and name of a

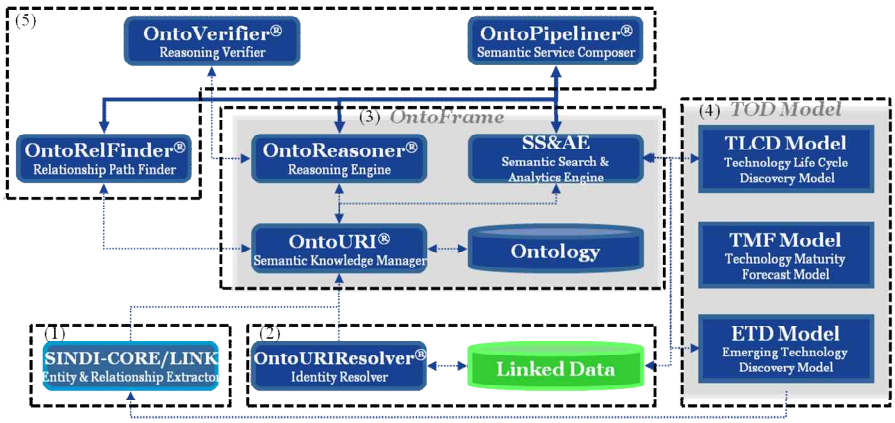

Fig. 1. System architecture of *InSciTe Advanced*

given object. TOD model is a logical model for discovery of emerging technologies [4] and prediction of phase and speed on a technology life-cycle (see Section 3). All data calculated and predicted by TOD model are also loaded into OntoFrame.

Most *InSciTe Advanced* services are provided through OntoPipeliner [5]. It asks many single-facet queries to OntoFrame and OntoRelFinder and integrates results of the queries into several multi-facet services with appropriate visualization. OntoRelFinder is an inter-instance path-finding engine that fast traverses and visualizes all possible paths among more than two objects of large-scale ontology database (see Section 4) and OntoVerifier [6] provides verification on inference-related analyzed results. This engine traces the inferred results back to get the ground facts and applied inference rules and shows how the results were induced in order to get users' confidence in the analysis results.

3 Technology Opportunity Discovery

Technology opportunity discovery consists of 3 sub-models: technology life cycle discovery (TLCD) model, technology maturity forecast (TMF) model, and emerging technology discovery (ETD) model. The TLCD model decides emerging phase of a specific technology through feature selection and analysis – i.e., decision tree learning – extracted from papers and patents information. We adopt emerging phases composed of 5 steps as in the Gartner's hype cycle: irruption, frenzy, turning point, synergy, and maturity. The TMF model decides technology development speed of a specific technology, i.e., the time required for reaching to the 5th step using the exponential moving average (EMA) method [14]. EMA, a variant of moving average, gives more weight to the latest data. The ETD model decides emergence of technologies based on the following 5 types of emerging technology (ET):

- Trigger ET is a technology newly appeared in recent 2 years and has bigger

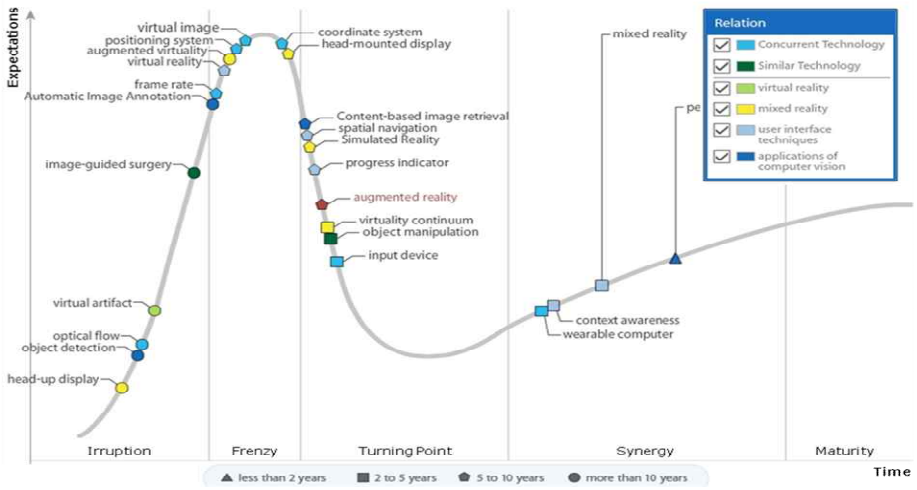

Fig. 2. Example of discovered life cycle of an emerging technology '*augmented reality*' and its related technologies

growth rate than average.

- Associated ET is a technology associated or similar to emerging technologies in recent 2 years.
- Pre-stable ET is a technology which will arrive at the final step in following 2 years.
- Reference ET is a technology commonly referred by other emerging technologies.
- Derived ET is a technology newly appeared and derived by other emerging technologies.

Fig. 2 shows an example of discovered life cycle of an emerging technology '*augmented reality*' and its related technologies.

4 Finding Relationship Paths

InSciTe Advanced finds indirect relationships – i.e., relationship paths – among technologies and research agents as well as direct relationships. Finding relationship paths from a large-scale ontology is a very time-consuming task since it requires exhaustive search like breadth-first search or depth-first search as known in graph theory [7]. To solve this issue and achieve reasonable response time, we have designed an efficient path-finding algorithm guided by ontology schema – in the concrete, *rdfs:domain* and *rdfs:range*, which restrict domain and range classes of a property, respectively. All possible class-level paths are computed in advance – when an ontology schema is given – and then all possible individual-level paths between two given individuals are searched within the class-level paths between the classes of the individuals. This approach reduces search space considerably by excluding unnecessary searches in advance.

Fig. 3 shows an example of relationship paths found between '*cloud computing*' and '*virtual reality*'. The path includes research agents and technologies which are indirectly as well as directly related to the given two technologies. The indirect relationship enriches information required for technology opportunity discovery.

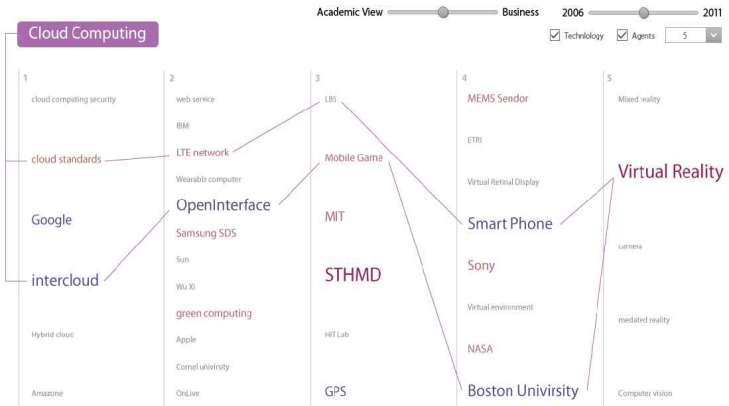

Fig. 3. Relationship paths between '*cloud computing*' and '*virtual reality*'

5 Other Advanced Services

InSciTe Advanced provides five major services, among which two services were already explained in Section 3 and 4. The remaining three are *technology level, roadmap* and *competitor/collaborator*. *Technology level* displays technology entry points, gaps and collaborations among research agents (see Fig. 4). *Roadmap* recommends promising technology candidates for each research agent considering current technologies in its possession (see Fig. 5). *Competitor/collaborator* analyzes and compares research agents in the views of technological similarity, competition and collaboration (see Fig. 6).

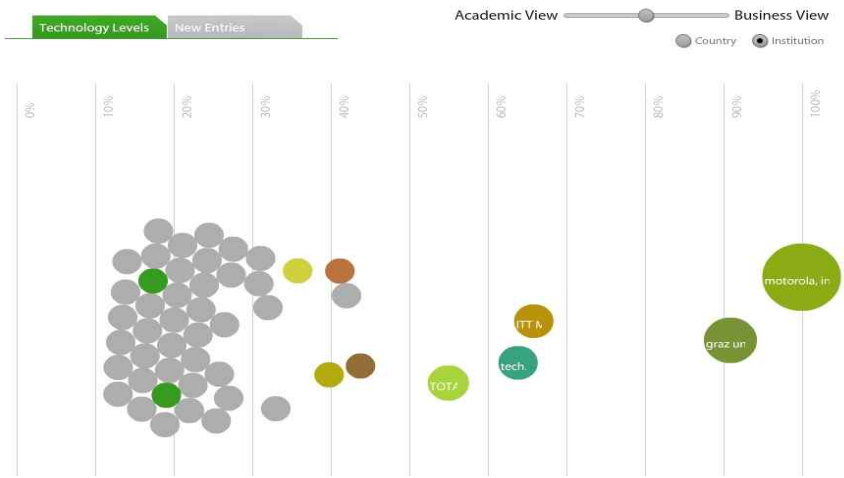

Fig. 4. Technology gaps among institutions for '*augmented reality*'

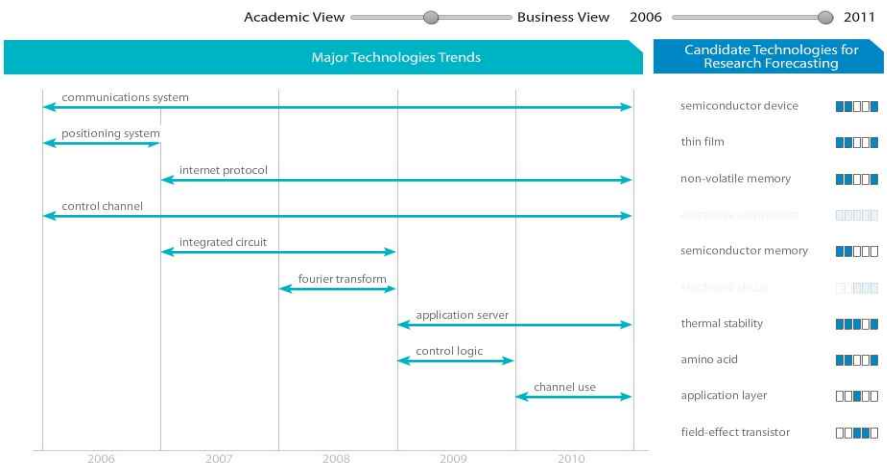

Fig. 5. Recommendation of promising technologies for '*Motorola, inc.*' considering its current technologies

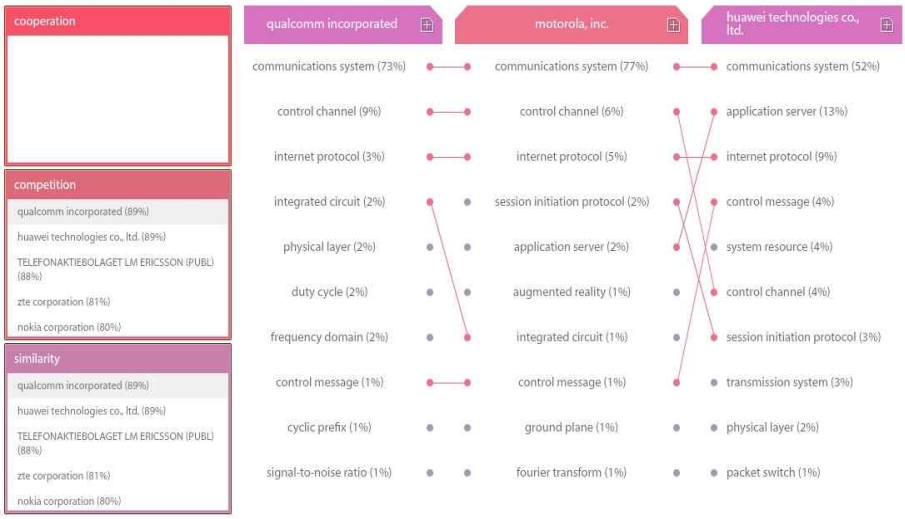

Fig. 6. Comparison of research agents in technologically similar, competitive or collaborative relationship with '*Motorola, inc.*'

5 Conclusion

This paper describes *InSciTe Advanced*, a service for technology opportunity discovery. It discovers emerging technologies based on technology life-cycle model and analyzes R&D status of research agents in competition or collaboration. Text mining and Semantic Web technologies allow *InSciTe Advanced* to analyze technological literature automatically and provide such intelligent services.

In the future, we plan to develop customized services for strategy planning of research institutes and firms.

References

1. Porter, A.L., Detampel, M.J.: Technology Opportunities Analysis. Technological Forecasting and Social Change. vol. 49, no. 3, 237--255 (1995)
2. Jeong, C., Choi, S., Choi, Y.: Fundamental Study on Extracting Relations between Technical Terms using Verb-based Patterns. In: 2009 International Conference of Maritime Information and Communication Sciences. (2009)
3. 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, L. Aroyo, et al. (eds.), The Semantic Web: Research and Applications, LNCS vol. 6089. (2010)
4. Lee, J., Kim, J., Jung, H., Sung, W.-K.: Toward Discovering Emerging Technologies Based on Decision Tree. In: IEEE CPSCom2011. (2011)

5. Lee, S., Jeong, S., Kim, H., Jung, H., Lee, M., Song, S., You, B.: OntoPipeliner: An Ontology-based Automatic Semantic Service Pipeline Generator. Expert Systems with Applications, vol. 38, no. 8. (2011)
6. Lee, S., Seo, D., Kim, P., Lee, M., Jung, H., Sung, W.-K.: Indexing Triple Dependencies for Inference Verification. In: ICCC2010. (2010)
7. Migliore, M., Martorana, V., Sciortino, F.: An algorithm to find all paths between two nodes in a graph. Journal of Computational Physics. vol. 87, no. 1, 231--236 (1990)

Appendix

A. Minimal requirements

- *InSciTe Advanced* is an application that mainly targets research planners who have responsibility for establishing R&D strategy rather than general researchers surveying R&D information. It is very useful when research planners want to find new business areas or emerging technologies dedicated to them.
- The main data source is technical literature such as papers and patents. It currently covers about 6,700,000 papers from IEEE proceedings and journals since 2006 and other proceedings and journals since 2009 and 3,100,000 patents of US, Europe and Japan. We also used PubMed and Wikipedia data to augment the relations between technologies extracted from the papers.
- All data used in *InSciTe Advanced* are represented in RDF format. All entities such as technology terms, research agents and research outcomes are represented in URIs and their relationships are represented in RDF triples. Such entities and their relationships are further processed to derive useful implicit knowledge. The applied Semantic Web technologies give us sufficient flexibility in composing useful technology opportunity discovery services, which is the main advantage of *InSciTe Advanced* compared to existing analytic tools.

B. Additional features

- *InSciTe Advanced* was designed to be a simple and easy Web application considering unskilled users. The analyzed results are visualized appropriately to help users' understanding.
- *InSciTe Advanced* detects emerging technologies and predicts their emerging phases and maturing speeds to help research planners to get insight for technology opportunity discovery.
- *InSciTe Advanced* finds indirect relationships as well as direct relationships among technologies and research agents. The indirect relationship enriches information required for technology opportunity discovery.
- *InSciTe Advanced* compares research agents having a specific technology in the views of technological entry point, level, similarity, competition and collaboration. It also recommends promising technologies dedicated to each research agent considering its current technologies.