Structuring Technological Information for Technology Roadmapping: Data Mining Approach

BYUNGUN YOON[1*], ROBERT PHAAL[2], DAVID PROBERT[2]

[1]Department of Industrial & System Engineering, Dongguk University,
3-26, Pil-dong 3ga, Chung-gu, Seoul, 100-715, KOREA
[2] Centre for Technology Management, Institute of Manufacturing,
Department of Engineering, University of Cambridge
Mill Lane, Cambridge CB2 1RX, UK

Most of research in technology roadmapping has focused on practical applications of technology roadmap, and methods to enhance its operational process. Thus, despite a demand for supports through well-organized information, little attentions have been paid to a study on how/which information can be utilized in a technology roadmapping. Therefore, this paper aims to propose a methodology to structure technological information in order to support the process. To this end, four methods are suggested to provide useful and substantial information: summary, information extraction, search and clustering. This research identifies the characteristics of data which can be potentially used in roadmapping and presents an approach to yielding important information from such raw data through various data mining techniques including text mining, multi-dimensional scaling and K-means clustering. Moreover, this paper explains how this approach can be applied in each step of roadmapping (preliminary, mapping and managing steps).

Key-Words: *Technology roadmapping, data mining, technological information, navigation*

1 Introduction

Technology roadmap has been regarded as one of the most influential, frequently used tools for strategic management of technology. It has great potential for supporting the development and implementation of integrated strategic business, product and technology plans, providing the information, process and tools to produce it (Phaal et al., 2004). Thus, both governments and firms usually generate such technology roadmaps to achieve their own specific objectives. At the national/industry level, forecasting the future of technology is definitely important to obtain the competitiveness of a nation and assign budgets to promising technology/industry areas. At the firm level, technology roadmapping can provide various graphical forms to explore the relationship among markets, product and technologies over time (McCarthy, 2001). In short, it is defined as a visual aid which crystallizes the links between research programs, development programs, capability targets and requirements (Kostoff and Schaller, 2001) or a map of the unfolding evolution of technologies and products (Rinne, 2004).

Various objectives of technology roadmapping enable to introduce diverse types of technology roadmaps. Although many researchers proposed the taxonomy of technology roadmap, the most popular classification of it is a work of Kappel which divided a large number of roadmaps into four groups: science-technology roadmaps, product-technology roadmaps, industry roadmaps and product roadmaps (Kappel, 2001). Such roadmaps can exhaustively cover the different aims of various organizations, meeting their needs. However, in a practical situation, most of analysts are struggling to draw a roadmap because traditional methods of developing a roadmap depend on their individual experiences and knowledge. Thus, although templates for generating technology roadmaps have been proposed and practically used, information which has been gathered from a large number of sources cannot be fully utilized. It is attributable to a lack of methods and processes to structure technological information. Many practitioners and researchers visualize or summarize such information to accomplish a variety of benefits. First, structuring information allows analysts to view large quantities of information, which human being cannot perceive easily otherwise. Second, such techniques can help retrieve appropriate information

from large datasets by recognizing patterns and trends. Finally, they can also help elicit meaningful implications by reducing the dimensions of original data, eliminating noise that normally exists in multivariate data (Engelsman and van Raan, 1992).

This paper aims at addressing how analysts can use information from technological documents in roadmapping and proposing the methods that structure the contents of documents and retrieves relevant documents. For this, we analyze how useful information can be extracted from documents to enhance the performance of roadmapping. In addition, this paper suggests how to manage documents systemically to improve their reachableness and applicability.

2 Background

2.1 Technology Roadmap
Technology roadmap has proven to be an effective technique to achieve successful technology development and product innovation. Robert Galvin, former CEO at Motorola defined roadmapping as an extended look at the future of a chosen field of enquiry composed from the collective knowledge and imagination of the brightest drivers of change in that field (Galvin, 1998). This advantage of roadmapping enables it to be widely adopted in industry with anecdotal evidence suggesting that the ratio of medium-to-large manufacturing firms in the UK deploying roadmapping has now increased to more than 50%.

Technology roadmaps are basically developed to plan the evolution of a product by linking a business strategy to product/technology (Albright and Kappel, 2003). In addition, they help to forecast science and technology development by reducing complexity through refined and synthesized representation of information (Saritas and Oner, 2004). Technology roadmaps are also utilized to enable successful knowledge management and intellectual property management for supporting knowledge creation and new patent development (Yasunaga and Yoon, 2004).

2.2 Text Mining
Text mining aims at performing knowledge discovery from unstructured databases. Since documents in a textual format have no data field, it is very difficult to analyze such textual information. Thus, the starting point of text mining is to define key features of documents in order to transform unstructured textual information into structured information. This technique is applied to cluster similar documents (Dhillion and Modha, 1999) or classify newly generated documents into a category (Sebastiani, 2002). More intelligent application of text mining covers identifying the subjects of documents and summarizing their contents (Clifton, 2003).

There are so many different techniques and methods for text mining in order to find new structures and patterns. Since the objective of text mining is to provide a more efficient way to browse and search large collections of documents (Atkinson, 2000), the major techniques can be divided into several classifications according to the tasks in the discovery process. The general criterion of classification is the kinds of extracted information such as keywords, words and terms. In case of keywords, the assumption is that keywords can represent the whole contents of documents and knowledge discovery can be performed through the analysis of the keywords (Manning and Schutze, 1999). A method to use words assumes that a document is labeled by each of the words which occur within it. Finally, terms which mean a sequence of words have meaning in the domain and mining is performed on the basis of the extracted terms labeling each document. The advantage of this approach is that terms tend to represent more important information than keywords and words because it can generate the meaning of label (Feldman et al., 1998).

3 Research Framework

3.1 Techniques for Structuring Technological Information
This paper presents four techniques to structure technological information. Such techniques include summarization, information extraction, clustering and navigation that traditional text mining techniques have handled for a long time. Since current methods to draw a technology roadmap are totally dependent on the intuition of experts, even applying existing text mining techniques to roadmapping has valuable meaning, facilitating the process with well-organized information. However, in addition, the proposed processes to apply those techniques will be customized to be suitable to a roadmapping process.

3.1.1 Summarization
The objective of this technique is to compress full contents by separating significant sentences from relatively less important sentences. Since analysts

cannot skim all pages of all documents due to the size of enormous information, this technique can help researchers collect appropriate information by reducing the quantity of documents. Thus, selecting important sentences is critical in summarizing the whole contents without a loss of meaningful ideas. In addition, this paper borrows text mining techniques to analyze such sentences because manual works require intensive efforts and a long time to complete the process. Therefore, this paper proposes to investigate significance of each sentence on the basis of critical keywords. The first step is to examine the importance of keywords, considering technological features. The second step is to derive critical keywords from documents. In the final step, important sentences of which the importance is over a cut-off value are selected by investigating how much critical keywords are included in each sentence.

3.1.2 Information Extraction

The most important issue on managing technological information is how meaningful information can be extracted from documents which analysts are unable to fully examine with their intellectual capability. Thus, the computerized approach will play a critical role when it can provide information which the manual work cannot derive or it is impossible for analysts to conduct due to a cost and time issue. In this paper, two functions are addressed to achieve the objective, including *ontology generation* and *trend analysis*.

(1) Ontology generation

When we intend to conduct technology analysis, understanding the characteristics of the technology is a short-cut for successful analysis. The core information to understand technology is the ontology of technology because it can provide the hierarchical structure of sub-technologies, presenting parallel and vertical relationships among them. For this, patent documents are a suitable data source to generate the ontology of a specific technology. They contain technological features in detail, which means that researchers can easily investigate the relationships between patents. On the basis of patent documents, multi-layer K-means clustering is applied to create a hierarchical structure of technology. The proposed brand-new clustering iterates K-means clustering until it can get leaf-nodes, meaning sub-technologies in the lowest level.

In the first clustering, all patent documents are divided into several clusters, examining an optimal number of clusters. When deciding the number of

clusters, analysts should check how far the central points of obtained clusters are from those of the other clusters and how closely patents in a cluster gathered around the central point of the cluster. It can be regarded as the criterion of clear clustering because an inter-distance between clusters is far and an intra-distance in a cluster is close. The first-layer of ontology is derived through this process, giving each clusters appropriate names respectively. In the second clustering, the derived clusters are subsequently decomposed into several sub-clusters through the same process. These clusters are located in the sub-level of the clusters which are obtained in the first clustering, generating a hierarchical structure of clusters.

(2) Trend Analysis

The typical characteristic of technology roadmap is to show the historical change of technology development. Therefore, trend analysis gives researchers ample information for technology roadmapping. Among various possible trend analyses, this paper chooses three types of trend analysis. First, the trend in the number of patents related to a technology helps grasp the position of technology life cycle, indicating whether a technology is in an embryonic, growing or declining phase. This sort of graph presents the change of the number of granted/applied patents over time. Thus, while the abscissa of a point is a year when a technology is granted, the ordinate of the point is the number of granted patents. Second, the trend of keywords provides a list of emerging keywords and declining keywords, enabling to anticipate the future of technology. Since a design process is usually started from devising influential keywords, such technological keywords can support to draw the future part of roadmaps. Third, the trend of morphology of technology can be suggested for an intensive analysis. As a matter of fact, keywords per se cannot offer direct, concrete information to come up with new technology. Instead, the morphology of technology assists researchers to understand the real forms and characteristics of technology and forecast the detailed configurations of promising technology.

3.1.3 Clustering

In general, clustering techniques enables users to examine the relationships among data by grouping associated data into several clusters. Technological documents are also grouped through K-means clustering in this paper. The first step is to extract keywords from all documents and generate keyword vectors of documents, consisting of the frequency of

important words that frequently occur in documents. The second step is to analyze the relationships among documents by calculating the similarity between a pair of documents. The third step is to group associated documents through K-means clustering. In the final step, each cluster should be named by considering the characteristics of the clusters. Although the process to derive the ontology and clusters of technology is similar, the remarkable difference of two techniques is their structures. While the ontology of technology has a hierarchical structure of sub-technologies, all clusters of technology have a parallel topology. Therefore, clustering can be flexibly conducted because it doesn't need to consider the issue that in the ontology of technology, sub-technologies in a same level have to have a same topology.

3.1.4 Navigation

When analysts generate a roadmap, they used to require a bunch of relevant documents instead of voluminous pile of documents. Therefore, a navigation function to permit searching for appropriate technological documents of interest is definitely helpful for efficient roadmapping. The process to generate a navigation map is as follows. The first step is to extract keywords from all documents and create keywords vectors of documents in the same method as clustering. The second step is to project the documents into 2-dimensional space based on relationship of keywords, applying multi-dimensional scaling (MDS) in SPSS, computer software for statistical analysis. The function of this software produces a map which locates

each document in a coordinate of the 2-dimensional space. The final step is to link the navigation map with real documents in order to allow users to refer to contents of the documents. Therefore, a schematic navigation map needs to link real data sources by clicking a node in the navigation map. Various types of technological documents are located in the navigation map, providing the relationship between the different types of documents as well.

3.2 Application of Proposed Techniques in Roadmapping

The main objective of this paper is to apply various techniques to technology roadmapping. Thus, it is extremely critical to present how the proposed four techniques can be utilized in roadmapping. This paper aims at suggesting the method to use such techniques in three major steps (a preliminary step, a mapping step and a managing step) which this research defines. First of all, in the preliminary step, a list of keywords related to products can be used to identify product feature concepts from reports. In addition, ontology of technology can help analysts define a technology area. In the mapping step, a set of emerging/declining keywords and morphology over time assists in identifying the characteristics and titles of technology in each year. Moreover, a navigation map to search for appropriate documents and a summary of selected documents from the navigation map can be used to get in-depth information of a specific technology in a specific period. Clusters of documents are useful to refer to documents of analogous technology because
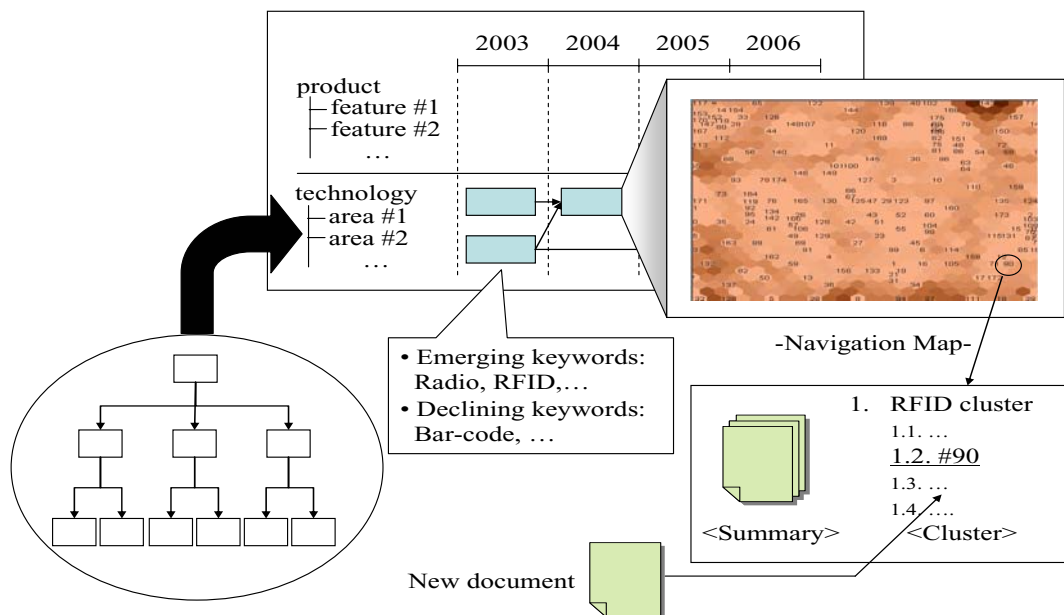

Fig. 1. Schematic Application of the Proposed Techniques

such clusters include a number of documents which have similar technological characteristics. In the managing (ex post) step, newly collected documents can be classified into a predefined category through a rule. Furthermore, after generating a navigation map and ontology of technology, they need to be updated for the continuous monitoring of technology change. Since the most important feature of technology roadmapping is the chain of technological evolution, the periodic renewal of a roadmap is quite essential for a quick response of technological breakthrough. The schematic application of the proposed techniques is depicted in Fig. 1.

4 Illustration

In order to illustrate the proposed methodology, this paper collects 271 patent documents, 47 roadmaps and 85 academic papers which are related to RFID (Radio Frequency Identification), an emerging technology that has been increasingly used in logistics and supply chain management (SCM) in recent years (Ngai et al., 2007). Basically, it is made up of two components: the transponder, which is located on the object to be identified, and the reader, which, depending upon the design and the technology used, may be a read or write/read device (Finkenzeller, 2003). Since the RFID technology has been considered as a promising technology that has potential to increase the performance of economic systems, many researchers have paid attentions on theoretical modeling and more practitioners have been developing new technology and RFID systems. The technology has sufficient reasons to become a subject to develop roadmaps and apply the proposed methodology.

In the first place, this research extracts meaningful information on the ontology of RFID technology. As a result of K-means clustering, a hierarchical structure of RFID technology is derived, consisting of four levels. Fig. 2 presents the ontology of RFID technology. This technology can be broken down into five sub-technologies in the first level. Security, middleware, hardware, software and application are the main categories of RFID technology. Subsequently, those categories are decomposed into subordinate level, providing more detailed sub-technologies. For instance, 'security' technology consists of 'encryption' and 'authentication' related technology. The hardware category is decomposed into the lowest level of technology because the technology has various types of sub-technology and includes a large amount of sub-technology.
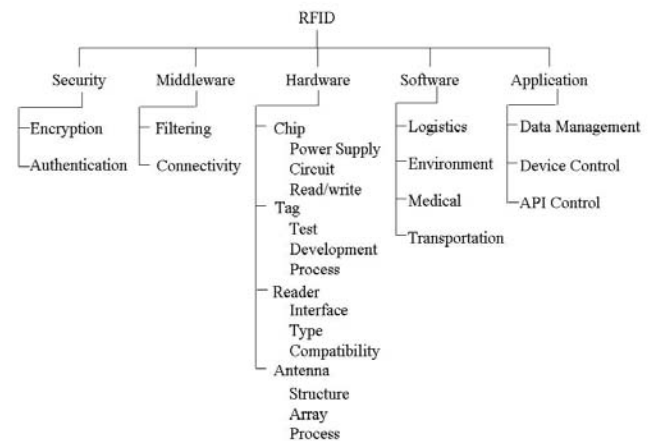

Fig. 2. The Ontology of RFID-related Technology

A navigation map for RFID-related patents is generated through MDS of keyword vectors of documents. The method to create a navigation map might be two-folds. The first method is to include all types of documents (patents, reports and academic papers) which are collected. On the contrary, the second method makes a navigation map with single type of documents respectively. However, the objective of this paper is to pursue an efficient management of all documents. Therefore, a navigation map had better include the whole data set for a complete investigation of a broad spectrum of information. Fig. 3 portrays a navigation map of RFID-related documents and illustrates a linked document. This navigation map also indicates that while some associated documents are gathered despite the different types of documents, most of documents have a tendency to be grouped together with similar documents.
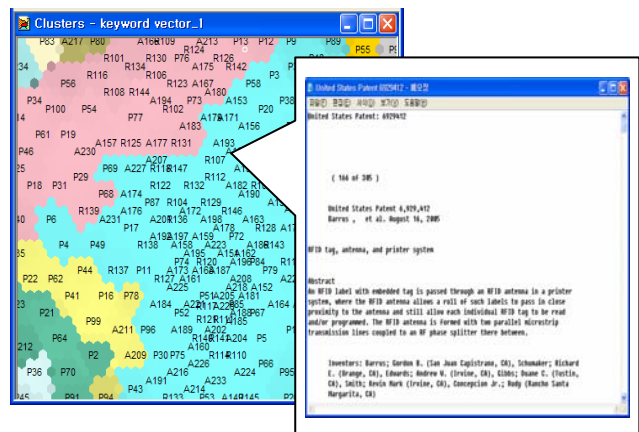

Fig. 3. A Navigation Map of RFID documents

4 Conclusion

This paper proposed four techniques to structure technological information in order to support developing a technology roadmap. Information extraction, clustering, summarization and navigation can help analysts draw technology roadmapping quickly and correctly by providing well-organized information. In addition, this paper present a real case to apply the suggested methodology to RFID-related documents for a deep understanding of the process.

Although the proposed approach can offer a method to manage technological information systemically, this paper has some limitations. First, while the size of keywords of patents is suitable to managing and analyzing, the range of keywords from reports and roadmaps are too broad to analyze intensively. It is attributable to typical characteristics of such documents because they contain a variety of contents. Second, the criterion of collecting documents from different data sources is not clear. While academic papers and patent documents are collected by searching keywords in the abstract of each document, reports are gathered by conducting a full-text search of keywords. Even though it aims at analyzing the application of RFID by involving all reports associated with the technology, it might bring about imprecise analysis of technology. Moreover, this approach is hard to be applied to an emerging product because sufficient documents for a precise analysis cannot be collected due to a short period of technology development. In this sense, future research should be conducted to enhance the performance of the proposed methodology to structure technological information by tackling the presented issues.

References:

Albright, R.E. and Kappel, T.A. (2003). Roadmapping in the Corporation, Research-Technology Management, 46(2): 31-40.

Atkinson, J. (2000). Text Mining: Principles and Applications. Revista Facultad de. Ingenieria, 7: 57-62.

Clifton, C. (2004). TopCat: Data Mining for Topic Identification in a Text Corpus. IEEE Transactions on Knowledge and Data Engineering 16(8): 949-964.

Dhillon, I.S. and Modha, D.S. (2001). Concept Decompositions for Large Sparse Text Data Using Clustering. Machine Learning, 42(1): 143-175.

Englesman, E.C. and van Raan A.F.J. (1994). A Patent-based Cartography of Technology. Research Policy, 23: 1-26.

Feldman, R., Fresko, M., Hirsh, H., Aumann, Y., Liphstat, O., Schler, Y. and Rajman, M. (1998). Knowledge Management: A Text Mining Approach. Proceedings of the 2nd PAKM, Basel, Switzerland.

Finkenzeller, K. (2003). RFID Handbook Radio-Frequency Identification Fundamentals and Applications, John Wiley & Sons, Ltd., England.

Galvin, R. (1998). Science Roadmaps. Science, 280: 803.

Kappel, T.A. (2001). Perspectives on Roadmaps: How Organizations Talk about the Future, Journal of Product Innovation Management, 18 (1): 39–50.

Kostoff, R.N. and Schaller, R.R. (2001). Science and Technology Roadmaps. IEEE Transactions on Engineering Management, 8(2): 132–143.

Manning C. and Schutze H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.

McCarthy, J.J., Haley, D.J. and Dixon, B.W. (2001). Science and Technology Roadmapping to Support Project Planning, Proceedings of the PICMET'01, Portland.

Ngai, E.W.T., Cheng, T.C.E., Au, S. and Lai, Kee-hung. (2007). Mobile Commerce Integrated with RFID Technology in a Container Depot. Decision Support Systems, 43: 62-76.

Phaal, R., Farrukh, C., and Probert, D. (2004). Technology Roadmapping – A Planning Framework for Evolution and Revolution. Technological Forecasting & Social Change, 71: 5-26.

Rinne, M. (2004). Technology Roadmaps: Infrastructure for Innovation, Technological Forecasting & Social Change, 71(1–2): 67–80.

Saritas, O. and Oner, M.A. (2004). Systemic Analysis of UK Foresight Results: Joint Application of Integrated Management Model and Roadmapping. Technological Forecasting & Social Change, 71(1): 27-65.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. ACM Computing Surveys, 34(1): 1-47.

Yasunaga, Y. and Yoon, T. (2004). Technology Roadmapping with Structuring Knowledge and Its Advantages in R&D Management. Proceedings of IEMC, Singapore.