Development of Research Topic Map for Analyzing Institute Performed R&D Projects-based on NTIS Data

Kwang-Nam Choi1, Nam-kyu Kang1, Tae-Hyun Kim1, Cheol-JooChae1, Hoe-Kyung Jung2 and Myung-Seok Yang1*

1NTIS Center, Korea Institute of Science and Technology Information, YuSeonggu, Dae-jeon - 34141, Republic of Korea; knchoi@kisti.re.kr, ngkang@kisti.re.kr, heemang@kisti.re.kr, cjchae@kisti.re.kr, msyang@kisti.re.kr

2Department of Computer Engineering, Pai-chai University, Seogu, Dae-jeon - 35345, Republic of Korea; hkjung@pcu.ac.kr

Abstract

Objectives: To figure out how to provide analysis results on changes in research topic trends of the institute which have performed national R&D projects through a knowledge-map after analyzing the data based on national R&D projects provided by National Science & Technology Information Service (NTIS). Methods/Statistical Analysis: In a knowledge map, in general, basic data are created by extracting target data; analysis rules are applied to perform analysis. The analyzed results are stored in the result database used in a knowledge map and shown using a knowledge map tool. To analyze research topics of the research institutes, Network analysis was conducted, and results were stored through the following procedures (Step 1: Create basic data, Step 2: Analysis on the size of R&D expenses, Step 3: Analysis of association relationship, Step 4: Expression of knowledge map). Findings: The Research Topic Map was implemented to help users easily figure out changes in research fields that are they are interested in. The Research Topic Map makes it easy to figure out changes in the research topics of the research institutes performed national R&D projects. It provides primary information such as the number of the research projects involved in the user-interested national R&D programs and size of R&D expenses. In addition, it offers diverse analysis results by analyzing changes in research fields. First, according to the comparison of changes in the size R&D expenses in top R&D fields, changes in key and promising sectors are found. In addition, it is able to estimate changes in total R&D expenses. Therefore, growth factors for research institutes can be examined. Particular research fields exist depending on the characteristics of college, hospital or research institute. Users are able to check changes in research fields considering the latest trends. Since correlations among research fields can be examined, furthermore, it would be possible to figure out research trends in a convergence field. Improvements/Applications: The Research Topic Map visualizes Topic Trends information (ex: changes in research fields, changes in R&D expenses, relationship among research fields, etc.) as well as information on the projects performed by research institutes. It is meaningful in that service users are able to get data insight on research institutes.

Keywords: Knowledge-Map, NTIS, R&D Data Insight, R&D Projects, Topic Trends

1. Introduction

As the amount of data dramatically increases recently, there has been a rising demand for data analysis. In the past, knowledge was produced by just processing and accumulating data. If it matured, it became wisdom. This process is called ‘Data-Information-Knowledge-Wisdom (D-I-K-W). As huge data are accumulated and utilized broadly, the D-I-K-W cycle is being replaced with the Data-Insight-Action (D-I-A) cycle, which means a shift of paradigms to the discovery of the meaning of data through analysis and their execution.
In general, a knowledge map refers to the summarization and supply of key information in the acquired knowledge in an easy-to-understand visualized form by through data processing. The ultimate goal of the knowledge map is to help users have an outlook on the field they are interested in just like the map containing geographical location information. It is an efficient tool which makes users gets insight on data by visualizing analysis results 2.

This study aims to figure out how to provide analysis results on changes in research topics at the research institutes which have performed national R&D projects through a knowledge map after analyzing the data in NTIS 3.

1.1 Science-Map (NISTEP)
Science-Map produced by National Institute of Science and Technology Policy in Japan extracts different research fields from journal-papers in two-dimensional knowledge-map. It is provided information about correlations amongst different research fields and density of core papers in the particular field as follows is shown in Figure 1 4.

1.2 SJR (SCIMAG Journal & Country Rank)
The SJR analyzes co-citation relationships from Scopus DB by categorizing the database into 27 scientific areas and further into 313 subareas depending on the Scopus classification. It is provided to visualize Knowledge-Map about network relationship in specific area by each country. The subarea of scientific area is defined as a node on the knowledge-map, the size of the node is number of papers in each subarea, and the weighted-link represent the numbers of co-citations as follows in Figure 2. It has provided information of distribution in a certain scientific area by each country and observed the network relationship. That is good service at the point of reviewing the national R&D trend analyzing co-citation relationships and theme-specific and country-specific research area distribution but not provided access to another information about specific details regarding the scientific area (http://www.scimagojr.com 5).

1.3 Map of Scientific-Research in Korea
Map of Scientific-Research is project how to represent knowledge-map that analyzed from scientific research project funded by National Research Foundation of Korea (NRF), and represented co-relations between such information of different research projects. The data were provided from 4,614 scientific research projects performed in the area of social science. The main column for data analysis is the project title, project research field, researcher, researcher’ gender, researcher’ age group, and institution extracted from project information Database and it was performed Network-Analysis about the relationships between these pieces of information.

It is like knowledge map represented in general, the node (red color) represent each research organization and the sub-node (blue color) as each scientific field, and the size of the sub-node is the number of scientific-research-

<table>
<thead>
<tr>
<th>Table 1. The number of the research institutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
</tr>
<tr>
<td>No. of Institutes</td>
</tr>
</tbody>
</table>

Number of research institutes (redundancy-excluded)
projects performed. The links represent relationships between different organizations as follows in Figure 3. It is just only used for the study report, not displayed on the WWW, but that is a good research understanding research trends and collaborative relationships by different organization performed R&D Project.

1.4 Implication
Both Science-map and SJR generator represented Knowledge-Map about relationship on research fields using journal data (Scopus-Database). They provide closeness among research fields to use knowledge map in general. In addition, the purpose of the Map of scientific research is to represent co-relationships among research institutes in a certain field in a knowledge map. In this study, Research Topic Map aims to find out key research fields and changes in research scale.

Table 2. Analysis data schema

<table>
<thead>
<tr>
<th>Table</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Briefings</td>
<td>Project number, year, title (Korean), title (English), available expenditure</td>
</tr>
<tr>
<td>Project details</td>
<td>Science and technology classification standard 1, Science and technology classification standard 2, Science and technology classification standard 3,</td>
</tr>
<tr>
<td>Project participants</td>
<td>Researcher ID, name, institute ID, name of the institute</td>
</tr>
<tr>
<td>Classification codes</td>
<td>Standardized code of national science and technology classification, name of the classification</td>
</tr>
</tbody>
</table>

In terms of classification information needed to express Research Topic Map, the National Science and Technology Standard Classification Information in the project information was used. First, annual key research fields were extracted by analyzing project information by the research institute. Then, the Research Topic Map was developed by expanding the analysis of association relationships up to related research fields, using network analysis technique. With the Research Topic Map, users are able to find out research scale by the research field and changes in key research sectors intuitively.

2. Research Topic Map
The number of the research institutes performed national R&D programs has continuously increased since 2010 as seen in Table 1. The Research Topic Map proposed in this study provides a knowledge map by analyzing and visualizing data in order to find out the key research fields of the research institutes participating in national R&D projects and changes in R&D expenses and research fields.

The Research Topic Map can immediately visualize and provide the data of the year users want to know as well as annual analysis results focusing on the research institutes searched by the user in a systematic manner by applying the web-based data analysis platform.

2.1 National Science and Technology Standard Classification System
The National Science and Technology Standard Classification System (hereinafter referred to as the ‘Standard Classification System’) is a national standard framework designed for the efficient management and use of science and technology-related information. It consists of 16 categories (ex: Math, physics, chemistry, etc.), 207 divisions and 1,648 sections in three science and technology fields (nature, life and artificial).

The NTIS utilizes science and technology standard classification codes for the standard classification of the related information such as project information and personnel information. To provide information in research fields, in this study, the Research Topic Map was developed, focusing on divisions among the classification codes included in the project information.

2.2 Analysis Subject
To establish the Research Topic Map, the data on the projects executed by research institutes are essential. For this, the information as seen in Table 2 among the R&D information of the NTIS was collected as the subject of analysis. The project information contains diverse columns. As shown in Table 1, the Standard Classification System in the project information was used to figure out the research institutes’ research fields. In addition, the institute’s ID was used to identify each research institute. Furthermore, the R&D expenses on the projects were used in estimating the size of R&D expenses.

2.3 Development Procedures
In a knowledge map, in general, basic data are created by extracting target data; analysis rules are applied to perform analysis. The analyzed results are stored in the result database used in a knowledge map and shown using a knowledge map tool.
The Research Topic Map proposed in this study follows general knowledge map creation rules for data analysis. To analyze research fields by the year, however, analysis was conducted, and results were stored through the following procedures. The storage structure on the results is as seen in Table 3.

<table>
<thead>
<tr>
<th>Column</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source_Node_ID</td>
<td>Research code</td>
</tr>
<tr>
<td>Source_Node_LABEL</td>
<td>Name of research field</td>
</tr>
<tr>
<td>Target_Node_ID</td>
<td>Related research code</td>
</tr>
<tr>
<td>Target_Node_LABEL</td>
<td>Name of related research field</td>
</tr>
<tr>
<td>Total_cost</td>
<td>R&amp;D expenses</td>
</tr>
<tr>
<td>Organization_ID</td>
<td>Code of Institute</td>
</tr>
<tr>
<td>Org_Name</td>
<td>Name of institute</td>
</tr>
<tr>
<td>Year</td>
<td>Year of the project executed</td>
</tr>
</tbody>
</table>

Step 1: Create basic data: Basic data are created by extracting annual project information using research institute IDs.

Step 2: Analysis on the size of R&D expenses: The projects are searched by the year in sequence in the basic data, and three standard codes in the project information are mapped with the division (hereinafter referred to as the 'Research Code'). Then, the sum of the R&D expenses which belong to the Research Code is stored.

Step 3: Analysis of association relationship: In relationships among the Research Codes, if a person who took part in the Research Code-related project participates in the other research project, it is deemed that the two research fields are related to each other. As a result, their linkage is stored.

Step 4: Expression of knowledge map: The analysis results are stored in the Result Table as stated and visualized as a knowledge map. The Research Codes are expressed in nodes, and the node color represents 'category' in the Science & Technology Standard Classification. The node size refers to the sum of R&D expenses, and a link is created when research fields are related to each other. The link thickness represents the closeness of the relations.

2.4 Implementation

The Research Topic Map was implemented to help users easily figure out changes in research fields they are interested in. If the institute is selected, and its name is entered as follows, in Figure 4, the network analysis results appear. Specifically, quantitative analysis results such as the number of target projects by the year, number of research fields and total R&D expenses are provided.

2.5 Utilization

The Research Topic Map makes it easy to figure out changes in the Research Topic Trends of the research
institutes performed national R&D projects. It provides primary information such as the number of the research projects involved in the user-interested national R&D programs and size of R&D expenses. In addition, it offers diverse analysis results by analyzing changes in research fields.

Figure 5. Research topic map (ex: KISTI)

First, according to the comparison of changes in the size R&D expenses in top R&D fields, changes in key and promising sectors are found. In addition, it is able to estimate changes in total R&D expenses. Therefore, growth factors for research institutes can be examined.

Particular research fields exist depending on the characteristics of college, hospital or research institute. Users are able to check changes in research fields considering the latest trends. Since correlations among research fields can be examined, furthermore, it would be possible to figure out research trends in a convergence field.

3. Acknowledgment

This research was supported by Maximize the Value of National Science and Technology by Strengthen Sharing/Collaboration of National R&D information funded by the Korea Institute of Science and Technology Information (KISTI).

4. Conclusion

The Research Topic Map visualizes Topic Trends information (ex: changes in research fields, changes in R&D expenses, relationship among research fields, etc.) as well as information on the projects performed by research institutes. It is meaningful in that service users are able to get data insight on research institutes.

There should be further studies for the utilization of diverse analyzing results by studying relationships among researchers as well as among research institutes.

References