

[DOI: 10.20472/IAC.2016.023.101](https://doi.org/10.20472/IAC.2016.023.101)

**EUNJI YU**

Korea Institute of Science and Technology Information, Korea

**YOUNGEUN KWON**

Korea Institute of Science and Technology Information, Korea

## **BUILDING NATIONAL R&D TREND KNOWLEDGE MAP USING TOPIC ANALYSIS.**

### **Abstract:**

Recently, big data collection, storage, analysis and visualization technologies have emerged in IT industry. As a result, there have been attempts to discover issues in diverse domains such as politics, society, economy and culture and utilize the results in making a decision. For example, a trend knowledge map and issue tracking designed to extract issues and glimpse at issue change patterns have drawn an attention. This study has proposed a method to construct a visualized national R&D trend knowledge map after analyzing user query data on national R&D information in the National Science & Technology Information Service (NTIS) which has collected and provided the national R&D information. It is expected that the proposed national R&D trend knowledge map would help national R&D researchers plan and implement national R&D projects and programs.

### **Keywords:**

Text Mining, Topic Analysis, Issue Tracking, Visualization, Social Network Analysis

## 1. Introduction

Recently, big data collection, storage, analysis and visualization technologies have emerged in IT industry. As a result, there have been attempts to discover issues in diverse domains such as politics, society, economy and culture and utilize the results in making a decision. Therefore, text-mining technology designed to analyzed unstructured text data has been widely used. Text mining is an analysis technique which comprehensively utilizes diverse technologies such as data mining, natural language processing (NLP), information retrieval, computational linguistics and topic analysis (Mooney and Bunescu, 2006).

Lastly, there have been many studies on topic analysis whose goal is to figure out issue distribution at a certain point, issue tracking (Aggarwal et al., 2014) designed to analyze and trace issue changes at several points of time and a trend knowledge map (Bangrae Lee et al., 2013) targeted to figure out current trend in a visualized form through simple frequency and data processing. Specifically, the goal of topic analysis is to extract issues from multiple document sets while issue tracking or trend knowledge map is aimed to figure out the distribution of major issues by specific period after extracting issues from the posted or connected documents (Chen Liu, Namgyu Kim, 2014). The studies on topic analysis and trend analysis have generated visual performances by analyzing SNS, news and VOCs. However, these cases are mostly analyzed by some businesses which have their own data with a goal of figuring out the latest trend more quickly and developing a marketing strategy.

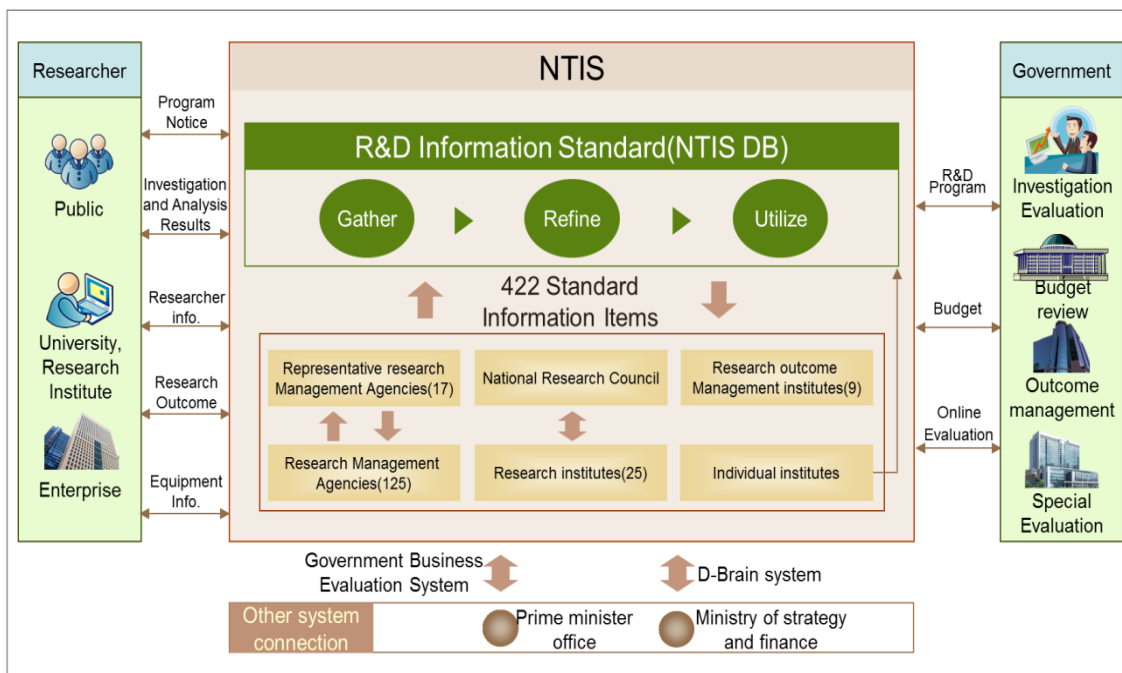
As the importance of data is emphasized, Korean government and public agencies have made some moves to make good use of the data which have been accumulated for a long period of time. One typical example is a public data portal (DATA.go.kr). It offers diverse public information in various fields such as weather, traffic and culture, using open data and API. Under this system, both citizens and businesses are able to get access to the information in an easy and convenient manner. The R&D information knowledge portal and NTIS which have supported R&D activities by collecting national R&D data from all bureaus in an integrated fashion also collect the data relating to national R&D programs, projects, personnel, R&D equipment and research outcome.

So far, more than 4.85 million sets of data have been collected, and diverse services have been provided. However, there has been a rising demand for actual data-based analysis services. Hence, this study attempted to propose a national R&D trend knowledge map construction method, which visualizes personal issue trends after extracting issues from users' search keywords associated with national R&D information in the NTIS, using topic analysis and classifying them by a certain period. This paper is structured as follows: In chapter 2, a proposed national R&D trend knowledge map construction method is stated. Lastly, the contribution and limitations of this study and a direction of future studies are discussed.

## 2. National Science & Technology Information Service (NTIS)

NTIS provides various service for researchers, decision makers, and governmental officers by converting national R&D information such as programs, projects, human resource, and research facilities and equipments and outcome into database by the interconnection with 17 governmental departments conducting national R&D projects for improving the effectiveness of research and development through the research life cycle from the R&D planning and the utilization of the outcome (MyeongSeok Yang et al, 2013).

**Figure 1: NTIS Conceptual Diagram**



The main service of NTIS consists of the national R&D program management service, the national R&D participating human resource information service, the national R&D outcome information service, the national R&D research facility and equipment management service, and the additional information service related to the national R&D.

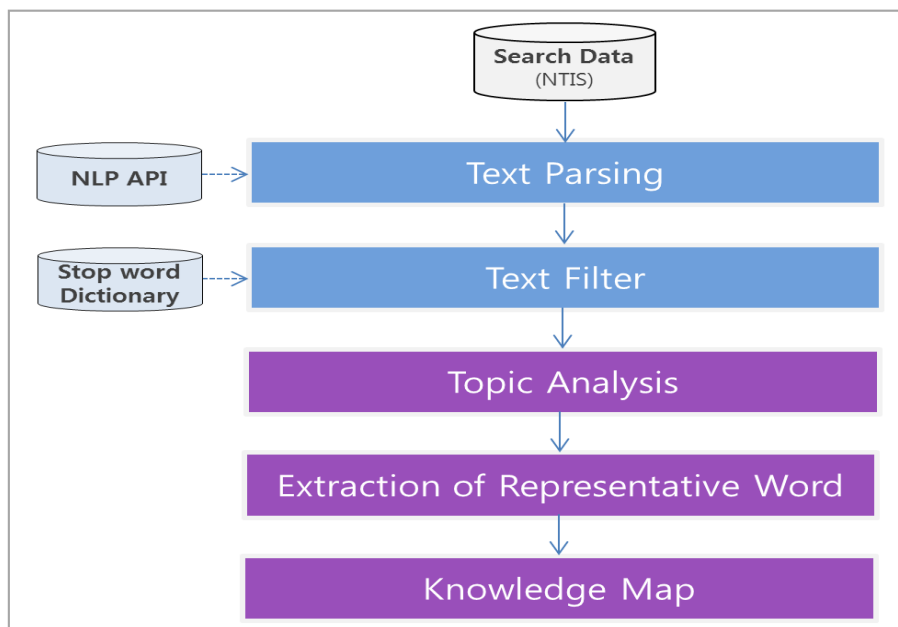
The national R&D program management service provides the current information on national R&D projects, and the survey and evaluation information to be able to utilize by industry, university, research institutes and government. The R&D participating human resource information service has the national R&D participating human resource information service providing the detailed information participating national R&D projects, the recommendation of the evaluation committee service, and the researcher registration service. The national R&D outcome information service supports the utilization of the research outcome generated by national R&D programs from the national level.

The national R&D research facility and equipment management service manages the research facilities and equipments comprehensively configured for the national R&D program, and supports the systematic operation and management of the research facility and equipment of universities and research institutes.

### 3. National R&D Trend Knowledge Map

The national R&D trend knowledge map construction method proposed in this study is shown in <Figure 1> below. The steps stated in <Figure 2> above are as follows: In the 1st step, text parsing targeted to extract target keywords from 320,000 data sets in the NTIS is carried out. It is designed to extract nouns only, by reflecting the characteristics of a user's search queries. In the 2nd step, the extracted terms are filtered after applying a stop list to parsed keywords. A stop list is a list of the terms for filtering to avoid being included in a list of the analysis subjects. In the 3rd step, topic analysis is performed. Topic analysis is a set of the keywords extracted based on the TF-IDF of the terms. In terms of a topic, either monolingual or multilingual topic can be selected and configured. The results of topic analysis are illustrated in Figure 3 below:

**Figure 2: Research Overview**

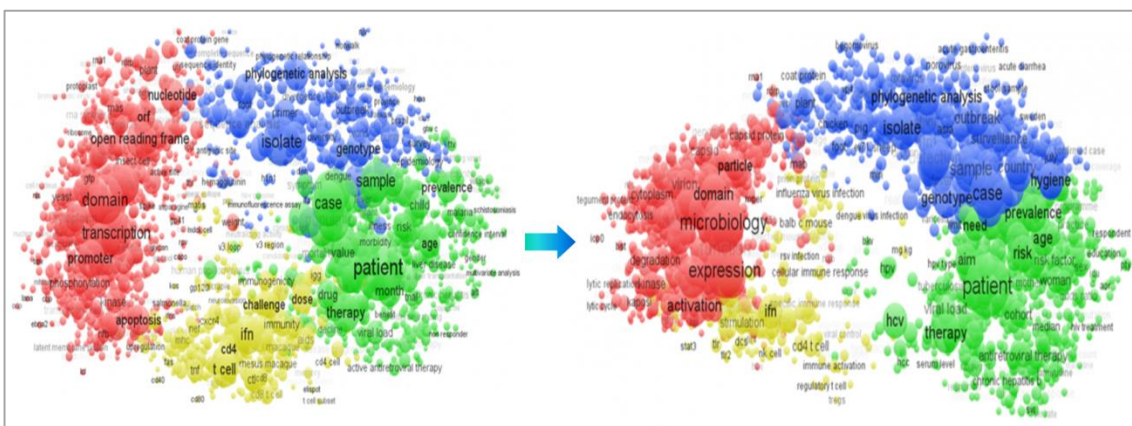


**Figure 3: Example of Topic Modeling**

Topic_id	Doc_Cutoff	Term_Cutoff	Topic	Ic_Numterms	Numdocs
1	0.002	0.003	대한약전,pancreatitis, + 조고속#다중#분자#진단용,아토피성피부염개선,+지진기#술#개발	56	2156
2	0.002	0.003	+빔포밍,mbc,서브틸리스.filtered	406	3538
3	0.002	0.003	+창조기업#과제,+기반#기술#확산,문광섭,+분석#장치,+제어#기술#기반#지능	477	12689
4	0.002	0.003	우미양,+s#분석기,+제기#술,+oled#기술#경영,gmg	518	13407
5	0.002	0.003	온보드,unigraphics,격자,태진기술,+공정#개선	300	4558
6	0.002	0.003	카디,조석규,+수성#접착제,+경제협력#방안,+포르	536	6215
7	0.002	0.003	+온라인#전자상거래,+경북#지역#현장#적용#시험,+저항성#자원#개발	516	8696
8	0.002	0.003	+엘엠티,+강화#복합,지하역,+확산#저항성#동위원소분리#기술#개발	519	9614
9	0.002	0.003	그린건설,+신청하다,주력,음형	510	7694
10	0.001	0.003	생식질,전체론,아토피성피부염개선,+에너지#저장#기능,대한약전	172	5356
11	0.002	0.003	+고화질#무선#영상,+물리#과정,난수,+지역#특산물#자미,우미양	907	34343
12	0.001	0.003	+재해#위험#지도,213002043sb910,변우혁,+패적성#인증,음형	337	12013
13	0.002	0.003	+world#가상,+분석#입력,+통신#이용자#보호#정책,+ibs	465	9102
14	0.002	0.003	+지역#특산물#자미,유태경,+유산균#발효,김이경,농작물재해보험	587	9754
15	0.002	0.003	가능검사,임채신,한국코르크리트,+안정기#술#개발,+기반#전문우	551	20147
16	0.002	0.003	무대장치,스프레이건,+헬멧#시현	429	6780
17	0.002	0.003	침만,항장품,교배,+미세#표면#형상	421	10770
18	0.003	0.003	화산재,iron,+gis#기반#건설,+naocl#생산,+확산#저항성#동위원소분리#기술#개발	520	4562
19	0.002	0.003	+안정기#술#개발,+기반#전문우,에프에스알엔티,+상대#전송#장치#개발,신형식	428	5472
20	0.003	0.003	+ic#연구,+mouse,+oled#기술#경영,음형,+에너지#슈퍼#커패시터#개발	546	6607
21	0.003	0.003	작은와포자충,주파수,+메쉬#투명,+주기#술#개발,하남	367	7169
22	0.002	0.003	+고랭지#재배기#술,무대장치,+급속#발열체,메,+온라인#전자상거래	506	7709
23	0.005	0.003	+중소형#선박#디젤#엔진,+태양광#발전#이용#시스템#개발,국립병원,냉감,진단기	13	12639
24	0.002	0.003	+전력#저장#시스템#기술#실증,e8,지하역,내츄럴엔도텍	491	9667
25	0.001	0.003	+방우#소재#개발,+정량#펄프,+관련#프로그램,+4#차#지방#과학기술#진흥,+불순물#제어	407	22559

There are two ways to express the keywords extracted by topic analysis in a visualized form such as a knowledge map. First, if a trend knowledge map is prepared by period based on co-occurrence counts, it is able to find out changes in national R&D research trend. If the co-occurrence count is high, the keywords are displayed in bold. In the second method, associated keyword rules are extracted through association analysis and visualized, using social network analysis (SNA). In this study, a trend knowledge map will be extracted, using the first method. The trend knowledge map built according to the said processes is shown in <Figure 4> below:

**Figure 4: Example of National R&D Trend Knowledge Map**



Source: *Journal term co-occurrence map for the field of Virology, using a set of 14,158 papers published from 2000 to 2002. Colors used to distinguish clusters of related terms. Data source: [Scopus](http://www.scopus.com)*

The biggest advantage of a trend visualization method is that it allows users to catch a trend change pattern at a glimpse. It appears that a national R&D trend knowledge map construction model proposed in this study would be useful for national R&D researchers in planning and implementing R&D projects.

#### 4. Conclusion

This study proposed a national R&D trend knowledge map construction method which extracts issues through topic analysis on user search results in the NTIS which provides national R&D information and provides the extracted issues in a visualized form by period. The biggest contribution of this study is that it has proposed an analysis model which can analyze national R&D trends after analyzing search keywords on national R&D information in the NTIS and visualizing national R&D related issues. In future studies, it is planned to apply actual data to the proposed model and analyze the results. To provide customized data depending on NTIS' service direction, in addition, it is needed to provide personalized data through the construction of a personal trend knowledge map which can extract personal issues and track changes personal trends and perform studies relating to recommendation services.

#### References

- Mooney, R. J. and R. Bunescu. (2006) Mining Knowledge from Text using Information Extraction. 2006, Vol. 7, No. 1, s. 3-10.
- Aggarwa, A. G.; Waghmare, and A. Sureka. (2014) Mining Issue Tracking Systems Using Topic Models for Trend Analysis, Corpus Exploration, and Understanding Evolution, Proceedings of the 3rd International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering. 2014, s. 52-58.
- Bangrae Lee et al. (2013) Morphological Classification of Knowledge Map for Science and Technology and Development of Knowledge Map Examples in the View of Information Analysis, The Journal of the Korea Contents Association. 2013, s. 461-476.
- Chen Liu, Namgyu Kim. (2014) Individual Interests Tracking: Beyond Macro-Level Issue Tracking. 2014, s. 275-287.
- DATA. Go. Kr
- MyeongSeok Yang et al., (2013) Study on the National Science & Technology Information Service. 2013, s294-304
- National Science & Technology Information Service, <http://www.ntis.go.kr>