Methodology of Analysis of Users' Information Demand through Search Keyword Analysis

Eunji Yu / KISTI
NTIS Center
Daejeon, Korea
eunjiyu08@kisti.re.kr

Wonkyun Joo / KISTI
NTIS Center
Daejeon, Korea
joo@kisti.re.kr

***Abstract*—The search records in a portal service which provides diverse information services reflect users' demand for information and services. Therefore, the analysis of search records in the portal service makes it possible to figure out and measure the demand for user-wanted information or services. This study suggests a model which investigates users' information demand for national Research and Development (R&D) data after analyzing the National Science & Technology Information Service (NTIS) users' search records. The NTIS has collected national R&D data such as government R&D programs, projects, research outcomes and equipment. It appears that that users' information demand measured through the methodology in this study would be helpful in developing service strategies and establishing services.**

Keywords-Government R&D; Text Analysis; Topic Modeling; Information Demand.

I. INTRODUCTION

A portal is a service designed to provide diverse information services to allow users to search wanted information on a certain site without visiting every website. Recently, the portal service has attempted to provide user-customized services after improving the conventional information search-oriented service. For the effective promotion of this mission, it is critical to figure out what information and services users want. For this, one of the most commonly used methods is to analyze the keywords of the users who use the portal. It appears that users' keywords are the data reflecting their demand for information and services in a realistic manner. Therefore, keywords have been deemed reasonable data needed to figure out users' demand. In fact, there have been a lot of efforts to improve information services based on the analyzed results in diverse fields. This study aimed to analyze the keywords of the NTIS (National Science & Technology Information Service) [1], which has collected and provided national R&D information, such as national R&D programs, projects, research outcomes and equipment. In addition, the study investigates users' demand and its trend with a goal of proposing a model which can analyze how users' information demand is changing.

The rest of the paper is structured as follows. In Section II, conventional studies on the research and keyword analysis methods designed to measure users' information demand are briefly mentioned. In Section III, keywords are analyzed and a scheme to construct a model designed to figure out users' information demand is explained. Lastly, the contribution, limitations of this study and future research directions are stated.

II. RELATED RESEARCH

2.1. NTIS Service

NTIS provides various services for researchers, decision makers, and governmental officers by converting national R&D information such as programs, projects, human resources, research facilities and equipment into information in a 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 outcomes [2].

2.2. Demand Measurement

With the development of information technology, it has become more important to predict users' information demand to properly respond to the fast changing trends. To figure out the exact demand, a lot of resources are needed, such as optimum data, which contains users' demands and domain experts' knowledge. Based on these resources, in fact, a variety of demand prediction techniques have been used. They are divided into quantitative and qualitative techniques. One of the most popular qualitative techniques is a questionnaire survey using experts' opinions. With high accuracy, this method is widely used among businesses. However, with this technique, it is difficult to define a scope of experts or consumers. In terms of quantitative techniques, there are regression analysis, time series analysis and diffusion model-based demand prediction method [3]. When the demand is predicted using a quantitative technique, it is very important to decide what data would be analyzed. The technique is used when optimum data can be secured. During time series analysis, in particular, it is usually adopted in the fields where data can be easily collected. Furthermore, this method is able to model complicated causality, such as diverse variables and time difference, and predict the future by discovering the flow of past data and new trends.

2.3. Keyword Analysis Service

Lately, portal services have provided a service which measures users' interest with the accumulated keywords and analyzes trends. For example, Google has introduced Google Trend [4], which shows keyword query count in an index

ISBN: 978-1-61208-587-6

format based on the keywords of the users from around the world. Naver [5], the largest portal site in the Republic of Korea, has also provided a keyword statistics service called 'Naver Trend'. Since the nation's search engine market share is dominated by Naver (over 70%), it is appropriate to measure the domestic interest on keywords [6]. The predictions by Google or Naver trend analysis have little difference, even compared to reliable statistical data. To examine users' demand through analysis of keywords, therefore, there should be studies on the subjects which meet the conditions further specified. The information to be analyzed should be representative or comprehensive or should have many users, like Google and Naver. Therefore, in this study, the search data of NTIS which has the Korean government R&D information was selected as the analysis target.

III. Analysis Model for Users' Information Demand

A model proposed in this study is shown in Figure 1. The topic analysis methodology used in the model in this study derives a set of the keywords extracted by calculating the importance of the terms using probability based on the TF-IDF (Term Frequency-Inverse Document Frequency) [7] values. The topic analysis differs from conventional models in that it can prevent similar or the same keywords from being classified into a different category when keywords are analyzed based on the TF only. Lastly, the importance of the keywords derived through the topic analysis is visualized, using probability values, as shown in Figure 3.

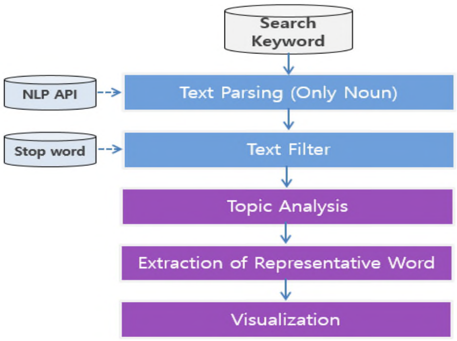
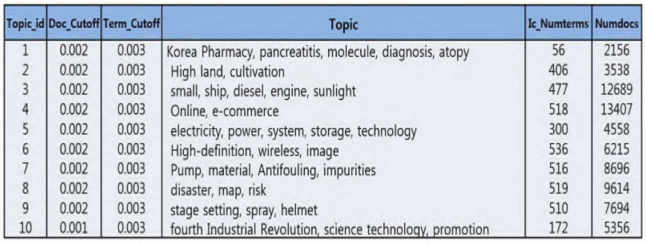


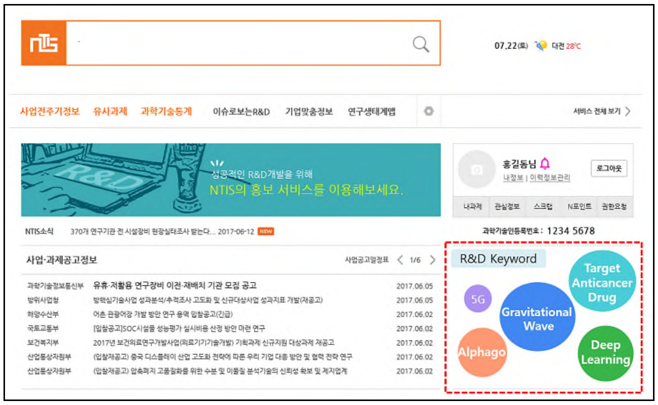
Figure 1. Analysis model for users' information demand

In Figure 1, the process can be explained as follows: In Step 1, pre-treatment on the keywords about national R&D information in the NTIS for the past year is performed. Nouns are only extracted by reflecting users' search query properties, and homonyms are processed using a synonym dictionary. Furthermore, unused terms are filtered with a stopword dictionary. Since the unstructured data 'texts' are analyzed, a pre-treatment process which handles homonyms and disuse is very important in this methodology, even having influence on final analysis results. In Step 3, topics are analyzed. The result of topic analysis is shown in Figure 2.



Topic_id	Doc_Cutoff	Term_Cutoff	Topic	Ic_Numterms	Numdocs
1	0.002	0.003	Korea Pharmacy, pancreatitis, molecule, diagnosis, atopy	56	2156
2	0.002	0.003	High land, cultivation	406	3538
3	0.002	0.003	small, ship, diesel, engine, sunlight	477	12689
4	0.002	0.003	Online, e-commerce	518	13407
5	0.002	0.003	electricity, power, system, storage, technology	300	4558
6	0.002	0.003	High-definition, wireless, image	536	6215
7	0.002	0.003	Pump, material, Antifouling, impurities	516	8696
8	0.002	0.003	disaster, map, risk	519	9614
9	0.002	0.003	stage setting, spray, helmet	510	7694
10	0.001	0.003	fourth Industrial Revolution, science technology, promotion	172	5356

Figure 2. Result of topic analysis



과학기술인등록번호 : 1234 5678

NTIS소식 370개 연구기관 전 시설장비 현장실태조사 받는다... 2017-06-12 NEW

사업·과제공고정보 사업공고일정표 1/6

과학기술정보통신부	유휴·저활용 연구장비 이전·재배치 기관 모집 공고	2017.06.05
방위사업청	방핵심기술사업 성과분석/추적조사 고도화 및 신규대상사업 성과지표 개발(재공고)	2017.06.05
해양수산부	어촌 관광어장 개발 방안 연구 용역 입찰공고(긴급)	2017.06.02
국토교통부	[입찰공고]SOC시설물 성능평가 실시비용 산정 방안 마련 연구	2017.06.02
보건복지부	2017년 보건의료연구개발사업(의료기기기술개발) 기획과제 신규지원 대상과제 재공고	2017.06.02
산업통상자원부	(입찰재공고) 중국 디스플레이 산업 고도화 전략에 따른 우리 기업 대응 방안 및 협력 전략 연구	2017.06.02
산업통상자원부	(입찰재공고) 압축폐지 고품질화를 위한 수분 및 이물질 분석기술의 신뢰성 확보 및 제지업계	2017.06.02

Figure 3. Visualization example of analysis result

As a result, we are able to find out users' demand for national R&D information and demand trends at a glance, as shown in Figure 3.

IV. Conclusions

This study has proposed a model which can measure users' demand by analyzing the keywords in the NTIS and even figure out the demand trends on national R&D information. This research is significant in that it suggested methodology through which users' demand for national R&D information, which is relatively difficult in terms of demand prediction and its changes, can be analyzed as a field in which a huge amount of information is stored, diverse topics are mixed, and new research topics are continuously produced. In future studies, however, it is necessary to verify this methodology by analyzing actual NTIS keywords. In addition, keywords are simpler than conventional studies in which texts are analyzed in terms of pre-treatment. To derive filtered results, however, a high-quality glossary and a stopword dictionary in national R&D information must be developed in advance.

Acknowledgment

This research was supported by Construction of NTIS funded by the Ministry of Science, ICT and Future Planning.

REFERENCES

[1] National Science & Technology Information Service, http://www.ntis.go.kr [accessed July 2017]

[2] MyeongSeok Yang et al., “A study on the NTIS service,” Journal of Korea Technology Innovation Society, 13 2013, pp.294-304.

[3] “Methodology on Short-Term Demand Forecasting: ARIMA model,” 34, (2004), pp.68-80.

[4] Goodle Trends, https://trends.google.com/trends/ [accessed July 2017]

[5] Naver Trends, http://datalab.naver.com/keyword/trendSearch.naver [accessed July 2017]

[6] B. Cogille, J. Wolfers and E. Zitzewitz, “Using Prediction Markets to Trak Information Floes: Evidence from Google,” American Economic Association, 2009, pp.1-44.

[7] J. Han, M. Kamber, Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann Pulishers, San Francisco, 2011.