

텍스트마이닝을 활용한 농업 R&D 키워드 분석

김지훈, 김성섭*
농촌진흥청 농산업경영과

A Study on the Analysis of Agricultural R&D Keywords Using Textmining Method

Ji-Hoon Kim, Seong-Sup Kim*
Farm and Agribusiness Management Division, Rural Development Administration

요약 본 연구는 농업 R&D의 추세를 살펴보고자 텍스트마이닝 기법을 활용하여 농업 R&D에 해당하는 키워드를 분석하였다. 분석자료는 NTIS의 국가연구개발사업 과제정보를 활용하였으며, 2003년부터 2018년까지의 농업 R&D의 주요 키워드를 연도별 및 연구개발단계별로 구분하였다. 텍스트마이닝을 위해 키워드의 TF-IDF를 계산하여 점수별로 순위를 매기었으며, 유사한 키워드별로 그룹화하여 해석하였다. 주요 분석 결과는 다음과 같다. 첫 번째, 신기술의 도입과 외부 환경에 변화에 따른 농업 R&D 트렌드가 변화해가고 있다. 시간이 흐를수록 새로운 키워드가 대두되고 있으며, 기초연구 단계에서는 '기후변화'가, 응용연구 단계에서는 'ICT'와 '스마트팜'이, 개발연구 단계에서는 '수출' 키워드가 주되게 등장하고 있다. 두 번째, 연구개발 단계에서 시차를 가지고 키워드 변화가 나타나고 있다. 기초연구-응용연구-개발연구 순으로 주요 키워드가 변화하고 있으며, 대표적으로 '기후변화'와 '신품종' 키워드가 연구개발단계별로 연계되어 있었다. 세 번째, 농업 R&D의 대표적인 키워드는 '벼' 키워드로 나타났다. 그러나 '녹색 및 기후변화 대응'과 '가공 및 유통기술' 같이 국내의 농업 환경 변화에 따라 연구의 방향성과 목적이 변화하고 있었다.

Abstract This study analyzed keywords for agricultural R&D using the textmining method to examine the trend of agricultural R&D. Data used for the analysis included R&D project information provided by NTIS, and the research and development step by year from 2003 to 2018 were classified and applied. The TF-IDF approach was used as the analysis method, and ranking was derived based on score. Furthermore, we analyzed by grouping for similar keywords. The main analysis results are as follows. First, agricultural R&D trends are changing according to the introduction of new technologies and changes in the external environment. Second, keyword changes appeared with a time lag in the R&D step. The main keywords are changing in the order of basic research - applied research - development research. Third, the main keyword of agricultural R&D was 'rice.' However, the direction and purpose of the research were changing according to changes in the domestic and foreign agricultural environments.

Keywords : Agricultural R&D, Textmining, R&D, Keyword analysis, R&D keyword

본 논문은 농촌진흥청 연구사업(세부과제번호 : PJ01503601)의 지원에 의해 이루어진 것임.

*Corresponding Author : Seong-sup Kim(Rural Development Administration.)

email: kss2486@korea.kr

Received October 21, 2020

Revised November 16, 2020

Accepted February 5, 2021

Published February 28, 2021

1. 서론

4차 산업혁명 시대에 진입하면서 경제, 과학, 기술 분야 등 전 분야에서 빠른 변화가 나타나고 있다. 이러한 흐름에 발맞추어 R&D 투자 전략 역시 혁신적이고 미래를 선도할 수 있는 기술을 중심으로 수립하도록 요구되고 있다. R&D는 단기적인 성과 달성의 목표보다는 장기적인 관점에서 이루어져야 하며, 많은 인력과 높은 비용이 소요되기 때문이다. 따라서 R&D 분야에서 중장기적으로 이뤄지는 연구 동향·추세를 파악하고, 이를 R&D 수립 단계에 반영하는 것이 중요하다. 최근 농업 분야의 R&D는 IoT 활용 스마트팜 기술, 무인 드론, 빅데이터 등 다양한 기술 트렌드와 농업을 접목한 새로운 R&D 전략을 수립하고 있다. 한편, 농업 분야 R&D는 공공성, 불확실성, 비독점성, 지역성 등의 성격으로 타 분야의 R&D에 비해 장기적인 기술 개발 기간이 요구된다[1]. 또한, R&D 연구 성과가 농업 현장에 전달되는 과정이 다른 분야에 없는 지도체계라는 독특한 메커니즘이 존재하고, 시장환경, 자연환경의 변화에 대응해야 한다는 관점에서 새로운 기술 트렌드를 반영하는 속도와 방향성이 다른 분야와 차이를 보이고 있다[2]. 따라서 농업 분야 R&D는 개별적인 연구 동향·추세를 파악할 필요가 있다. R&D 트렌드를 파악하는 방법으로 연구 성과물을 활용하는 방법이 있다. 여기서 연구 성과물에는 크게 논문 성과 혹은 특허 성과로 구분하여 활용되고 있다. 장인실·조한익은 대상관계를 주제로 한 국내 학술지 연구물을 통해 국내 대상관계 관련 연구의 추이를 분석하였으며, 민지은·주은정·시아주칭은 KCI 등재 학술지를 대상으로 언어네트워크 분석을 통해 국내 외식관련 연구의 동향 살펴보았다[3-4]. 또한 서원철·박현석·윤장혁은 특허 내 기술키워드를 바탕으로 단어동시출현분석과 네트워크 분석을 접목하여 국가 R&D 연구동향을 분석하였으며, 김방룡·홍재표·고순주는 빅데이터 기술에 대한 특허분석을 통해 빅데이터 기술개발 동향을 분석하였다[5-6]. 위와 같은 방법은 정량적인 지표를 가지고 계량적으로 분석하기 때문에 과학적으로 트렌드를 분석할 수 있는 장점을 가진다. 그러나 논문은 학문적인 의미의 성과 특징을 가지며, 특허는 산업적인 의미의 성과 특징을 가지고 있다. 이는 R&D 성과를 바탕으로한 트렌드 분석에 유용하지만 R&D 전과정을 살펴보기에는 한계를 가진다. 농업 분야의 R&D는 상기 언급하였듯이 지도체계라는 독특한 메커니즘과 다양한 분야가 복합적으로 적용되고 있으며 새로운 기술 트렌드를 반영하는 속도 및 방향성이 타 분

야와 차이를 가지고 있다. 이 때문에 농업 R&D는 성과 차원에서 트렌드를 분석하기보다 R&D 전과정 관점에서 트렌드 분석이 필요하다. 본 연구에서는 농업 분야 R&D의 트렌드를 파악하기 위해 과거부터 수행되어 온 국가 연구개발사업 과제 정보를 활용하였다. 국가과학기술지식정보서비스(NTIS : National science and Technology Information Service, 이하 NTIS)는 국가 R&D에 대한 정보를 종합적으로 추적하고 있기 때문에 전체적인 관점에서 농업 R&D의 트렌드를 파악할 수가 있다. 한편 트렌드를 분석하기 위하여 빅데이터 분석 기법 중 텍스트마이닝 기법을 활용하였다. 텍스트마이닝은 비정량적인 데이터, 특히 문자와 같은 데이터를 정량화하여 분석할 수 있는 특징을 가지고 있다. 이는 많은 텍스트 정보를 포함하고 있는 데이터를 계량적으로 분석할 수 있는 장점을 가지고 있으므로 다양한 분야에서 트렌드 분석을 가능케 한다. 정철우·김재준은 건설분야의 트렌드를 텍스트마이닝 기법으로 분석한 바 있으며, 최용인·이예원은 국가연구개발사업의 키워드 흐름을 비교 분석하여 과학 기술정책의 트렌드를 분석하였다. 또한 김민정·김철주는 '전통' 용어의 트렌드를 텍스트마이닝과 소셜 네트워크 분석을 수행한 바 있다[7-9]. 그러나 농업분야 R&D에 관한 키워드 분석은 수행된 바가 없다. 본 연구의 목적은 과거부터 현재까지 진행되어온 농업 R&D 키워드를 분석하는 것이다. 본 연구는 키워드를 중심으로 농업 R&D가 추진되어온 경과를 파악하고, 향후 농업 R&D를 위한 시사점을 모색하고자 한다. 본 연구는 빅데이터를 중심으로 활용성이 높아지고 있는 텍스트마이닝 기법을 농업 R&D분야에 처음 적용하였다는 점에서 큰 의의가 있다. 논문의 구성은 다음과 같다. 먼저 2장에서 농업 R&D의 정의와 현황을 살펴보고, 3장에서는 분석 방법을 설명한다. 4장에서는 분석 결과를 해석하고 5장에서는 결론을 제시한다.

2. 농업 R&D 정의 및 현황

2.1 농업 R&D 정의

일반적으로 연구개발이란 자연과학기술에 대한 새로운 지식이나 원리를 탐색하고 해명해서 그 성과를 실용화하는 활동으로 정의한다. 세부적으로 한국산업기술진흥협회는 R&D를 연구단계별로 기초연구, 응용연구, 개발연구로 구분하고 있다. 기초연구란 지식의 진보를 목적으로 행하는 연구로 특정 응용을 노리지 않는 것 혹은 특

정의 사업적인 목적 없이 과학지식의 진보를 목적으로 하는 연구 활동을 의미한다. 응용연구란 지식의 진보를 목적으로 행하는 연구로 실제 응용을 직접 노리는 연구 활동, 또는 제품과 공정에서 특정한 상업적 목적을 가지고 행한 연구 활동을 말한다. 개발연구란 기초연구 및 응용연구 등에 의한 기존 지식을 활용해 새로운 재료, 장치, 제품, 시스템, 공정 등의 도입 또는 개량을 목적으로 한 연구 활동을 의미한다[10]. 한편 OECD는 R&D를 ‘인간·문화·사회를 망라하는 지식의 축적 분을 늘리고 그것을 새롭게 응용함으로써 활용성을 높이기 위해 체계적으로 이뤄지는 창조적인 모든 활동’이라 정의하고 있다[11]. 일반적인 정의와 OECD 정의를 기반으로 농업 R&D를 정의를 한다면 “농업 관련 지식의 총량을 늘리고, 관련 분야의 정책적 목적 달성을 위해 체계적인 틀을 거쳐 이루어진 모든 창의적인 행위들을 포괄하는 일련의 과정”이라 할 수 있다. 농업분야 R&D는 농업 R&D의 특성에 기인하여 정부 차원에서 농업 연구개발 어젠다 및 방향의 기획과 연구 수행이 이루어지고, 연구 성과는 민간 분야로 지도·보급되고 있는 구조를 가지고 있다. 민간 차원에서도 농업 R&D의 기획부터 연구 성과 활용까지 이루어지고는 있지만, 국가 차원에서 주관하는 농업 R&D가 연구개발 규모에 있어서 대부분을 차지하고 있는 상황이다. 이에 본 연구에서는 농업 R&D의 범위를 정부에서 수행하고 있는 기초·원천 연구를 포함한 모든 국가 R&D 사업으로 한정지어 분석하였다.

2.2 농업 R&D 현황

본 연구는 한국과학기술기획평가원(KISTEP : Korea Institute of Science and Technology Evaluation and Planning, 이하 KISTEP)에서 제공하고 있는 NTIS의 국가연구개발사업 과제정보를 활용하였다. NTIS는 R&D 사업, 과제, 인력, 연구시설 및 장비, 성과 등 국가연구개발 사업에 대한 통계와 상세정보를 제공하고 있다. NTIS는 매년 국가연구개발사업 정보를 업데이트하여 제공하기 때문에 과거부터 현재까지 농업 R&D의 연구 키워드 동향을 파악하기에 적합하다. 농업 분야 R&D 정보를 살펴본 결과, R&D를 주관하는 주요 기관은 농림축산식품부, 농촌진흥청, 산림청 등이었다. 최근, 프로젝트 사업으로 산업통상자원부, 과학기술부 등 타 부처와 공동으로 주관하는 R&D 과제도 증가하고 있지만, 과제 비중에 있어서 3개 부처가 가장 많은 부분을 차지하고 있다. 따라서 본 연구에서는 농업 R&D 과제의 범위를 ‘농림축산식품부(Ministry of Agriculture Food and Rural

affairs, 이하 MAFRA), ‘농촌진흥청(Rural Development Administration, 이하 RDA)’, ‘산림청(Korea Forest Service, 이하 KFS)’에서 주관하는 연구개발과제로 한정하였다. Table 1은 농업 R&D 과제 및 예산 현황이다. 농업 R&D 과제 수는 2012년을 기점으로 급증하였다. 2012년을 기점으로 연구개발사업 과제가 세분화되었기 때문이다. 2012년 이후 농업 R&D 과제 수는 증감을 반복하면서 증가하는 추세를 나타내고 있고, 농업 R&D 예산 역시 증가하는 추세이다. 전체 R&D 과제 대비 비중으로 살펴본 결과 농업 R&D 과제 비중과 예산 비중이 낮아지고 있는 것으로 나타났다. 기간별 비중을 살펴보면, 과제 수는 2012년 이후 감소하고 있으며 2018년 10.7% 수준으로 감소하였다. 예산 대비 비중 역시 감소하여 2018년 4.4% 수준으로 감소하였다. 이는 전체 연구개발에서 농업 분야가 차지하는 비중이 감소하고 있음을 의미한다.

Table 1. Task and Budget by NTIS and Agricultural R&D

(Unit : Task, Million won, %)

Year	NTIS R&D		Agricultural R&D	
	Task	Budget	Task	Budget
2003	25,916	4,903,612	1,920 (7.4%)	327,001 (6.7%)
2006	32,114	8,763,909	2,086 (6.5%)	458,874 (5.2%)
2009	39,565	13,534,682	2,397 (6.1%)	727,380 (5.4%)
2012	49,948	18,977,183	7,553 (15.1%)	970,127 (5.1%)
2015	54,433	23,038,353	6,581 (12.1%)	958,115 (4.2%)
2018	63,697	23,099,143	6,838 (10.7%)	1,005,619 (4.4%)

Table 2. Task and Budget by Agricultural R&D supervision institution

(Unit : Task, Million won)

Year	MAFRA		RDA		KFS	
	Task	Budget	Task	Budget	Task	Budget
2003	733	64,832	1,104	223,868	83	38,301
2006	665	80,057	1,398	331,139	23	47,678
2009	816	212,999	1,398	444,361	183	70,020
2012	2,197	339,693	5,057	546,021	299	84,413
2015	1,703	244,661	4,550	604,582	328	108,872
2018	1,915	254,022	4,594	646,698	329	104,899

Table 2는 주관기관별 농업 R&D 현황이다. 농촌진흥청이 과제와 예산 모두 높은 비중을 차지하는 것으로 나타났다. 기간별 비중은 다소 차이가 있으나 큰 변화는 없었다.

Table 3은 농업 R&D의 연구단계별 현황이다. 2018년 기준 응용연구개발 단계, 개발연구개발 단계, 기초연구개발 단계 순으로 비중이 높은 것으로 나타났다. 농업 R&D 분야에서 응용 연구개발단계 차지 비중이 큰 이유는 농생명자원의 품종 개량, 친환경 방제 기법 등 응용을 목적으로 하는 연구가 활발하게 진행되고 있기 때문이다.

Table 3. Task by Agricultural R&D Research Level
(Unit : Task)

Year	Basic Research Level	Apply Research Level	Develop Research Level	Other Research Level
2003	260	1,019	641	0
2006	396	878	773	39
2009	642	729	827	199
2012	2,883	2,127	2,363	180
2015	2,511	1,719	1,795	556
2018	1,261	2,740	2,191	646

3. 분석 방법

3.1 텍스트 마이닝(Text mining)의 개념

텍스트 마이닝은 비정형 데이터인 텍스트 데이터를 대상으로 하여 그들의 암묵적인 정보를 추출하는 과정으로 정의할 수 있다[12]. 텍스트 마이닝은 비정형 및 반정형 데이터에 대해 자연어 처리 기술과 문서 처리 기술을 적용하여 유용한 정보를 추출, 가공하는 목적으로 활용된다. 실생활에서 만들어지는 대부분의 자료는 문서형태를 가지고 있으며 논문, 신문, 잡지 기사, 여론조사, 이메일 등 텍스트를 베이스로 하는 반구조적인 데이터의 형태를 가지고 있다. 반구조적인 데이터는 정형화된 데이터가 아니기에 기존의 통계분석이나 데이터 마이닝 기법 적용에 있어서 부적합하다. 이러한 비정형 데이터를 DB화 하는 방법 가운데 하나가 텍스트 마이닝이며, 다양한 분야에서 활용 및 응용되고 있다. 텍스트 마이닝을 통한 분석과정은 Fig.1처럼 텍스트 자료를 자료처리 과정(data processing)과 자료 분석(data analysis)을 거친 후에 의미 있는 추세와 패턴을 발견 절차로 흐르게 된다. 자료 처리 과정은 정보 검색, 정보추출, 자연어 처리 등의 절차

를 통하여 수집한 데이터를 가공 및 정제하는 과정이다. 자료 분석은 데이터 마이닝, 기계학습(Machine Learning), 통계학 등을 활용하여 분석하는 과정을 말한다[13].

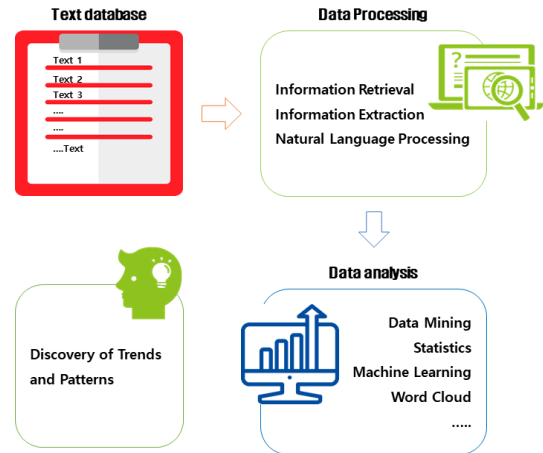


Fig. 1. Text mining analysis processing

자료처리 과정은 '정보 검색 및 추출'과 '자연어 처리'로 나눌 수 있다. 정보 검색은 사용자가 원하는 키워드를 기반으로 원하는 정보가 포함된 텍스트 데이터가 들어 있는 문서를 탐색하는 것이다. 정보 검색은 사용 목적에 따라 웹 검색(web search), 개인 정보 검색(personal information retrieval), 기업이나 기관, 특정 영역 검색(enterprise, institutional, and domain-specific search) 등 세 가지의 형태로 구별된다[14]. 정보추출은 토큰화(tokenization), 문장 분할(sentence segmentation), 품사의 배치(part of speech assignment)와 독립적 개체들로서의 인식(identification of named entities)을 포함하는 일련의 과정을 거치는 것으로, 정보 검색이 사용자가 필요한 정보와 관련된 텍스트가 들어 있는 특정한 문서를 찾는 것인데 반해 정보추출이란 특정한 문서로부터 구체적인 정보를 정제하는 것을 말한다. 텍스트 마이닝의 자연어 처리는 구조와 형태가 복잡한 자연어를 컴퓨터로 분석하기 위해 가공하는 단계이며 한국어나 영어와 같이 문맥에 의존하는 자연어를 대상으로 하기에 자료처리 과정에서 자연어 처리는 필수적이다. 자연어 처리는 크게 '형태소 분석', '동사 분석', '의미 분석', '화용 분석'으로 나뉜다. 자료처리 과정이 원시자료의 가공 단계라고 한다면 자료 분석은 데이터 마이닝, 기계학습, 통계학 등을 활용하여 의미 있는 결과를 도출하는 과정이다[15].

3.2 TF-IDF

텍스트마이닝을 통해 추출된 연도별 농업 R&D의 키워드가 어느 정도로 중요한지를 살펴보기 위하여 TF-IDF(Term Frequency Inverse Document Frequency, 이하 TF-IDF) 방법을 사용하였다. TF-IDF는 정보 검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치이다. 문서의 핵심어를 추출하거나, 검색 엔진에서 검색 결과의 순위를 결정하거나, 문서들 사이의 비슷한 정도를 구하는 등의 용도로 사용할 수 있다. 단어빈도(Term Frequency, 이하 TF)는 특정한 단어(t)가 문서(d) 내에 얼마나 자주 등장하는지를 나타내는 값이다. 단어 빈도가 높을수록 문서에서 중요하다고 생각할 수 있다.

$$tf(t,d) = f_{t,d} \quad (1)$$

한편, 특정 단어 자체가 특정 문서 내에서 자주 사용되는 경우에 특정 단어가 전체 문서군에 흔하게 등장한다고 판단할 수 있다. 이와 같은 오류를 피하고, 특정 단어가 전체 문서군(D)에 얼마나 등장하는지를 도출하기 위하여 역문서 빈도(Inverse Document Frequency, 이하 IDF)를 구하게 된다. IDF는 어떤 문서에 등장한 특정 단어가 다른 문서들에서는 잘 나오지 않은 단어인지 측정하는 척도를 의미한다. IDF는 문서 빈도(Document Frequency, 이하 DF)를 역수 취한 값으로 구할 수 있는데, 역수의 특성상 무한히 커지는 것을 방지하기 위하여 log 값을 씌우게 된다.

$$df(t,D) = \frac{|\{d \in D : t \in d\}|}{|D|} \quad (2)$$

$$idf(t,D) = \log \frac{|D|}{|\{d \in D : t \in d\}|} \quad (3)$$

즉 TF-IDF는 현재 해당 문서에서 특정 단어의 빈도와 이 단어가 다른 문서에도 잘 나타나지 않은 단어인지를 고려하여 각 특정 단어에 대해 중요도를 보는 것이다.

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \quad (4)$$

3.3 분석 절차

본 연구에서는 연구개발단계를 구분하여 키워드 분석

을 수행하였다. 주요 키워드가 연구개발단계별로 각기 다른 특성을 가지기 때문에, 연구개발단계별로 키워드 차이가 있는지를 세부적으로 살펴보기 위함이다. Fig.2는 농업 R&D 키워드 분석을 위한 분석 절차이다. 먼저, 농업 R&D 범주에 해당하는 NTIS 과제정보의 한글 키워드를 데이터프레임으로 설정하였다. 둘째, 자연어 처리를 통해 데이터 정제 및 가공을 한 이후, TF 및 IDF를 도출하여 TF-IDF 가중치를 계산하였다. 여기서 정제 및 가공은 Fig.1의 Data processing 과정을 참고하여 실시하였다. 데이터 정제로는 NTIS 과제정보 상에서 한글 키워드를 연도별, 연구개발단계별로 추출하는 과정이며, 데이터 가공은 추출된 키워드에 문장부호 제거, 조사 제거, 형태소 분리 등을 포함하는 토큰화 과정을 의미한다. 셋째, 계산된 가중치를 바탕으로 연구개발단계별 및 연도별 농업 R&D 키워드의 순위를 매겨 정리하였다. 마지막으로 순위를 기초로 농업 R&D가 어떠한 방향으로 전개되었는지를 살펴보았으며, 주요 키워드에서 의미가 유사한 키워드별로 그룹화하여 해석하였다. 농업 R&D 그룹화를 위한 기준은 농업 R&D 분야의 WBS(Work Breakdown Structure), CPF(Critical Performance Factor)를 바탕으로 작성된 농업 R&D 일람을 활용하였다. R&D일람은 농업 부문의 R&D 활동과 성과 전반에 대한 내용을 설명하고 있으나, R&D 대상에 대한 설명이 모호하다. 예를 들어 같은 신품종 연구개발이라도 그 대상이 품목에 따라 성과는 크게 달라질 수가 있다. 이에 본 연구에서는 대상을 명확하게 설정하기 위하여 농생물자원에 해당하는 식량작물, 특용원예작물, 축산물 분류를 추가하였다. 농생물자원 분류 자원을 R&D 세부 활동 및 성과와 농생물자원에 따라 9개 R&D 활동과 3개 농생물자원으로 정의하면 Table 4와 같다[16]. Table 5는 텍스트마이닝 분석을 통해 도출한 키워드를 Table 4를 기준으로

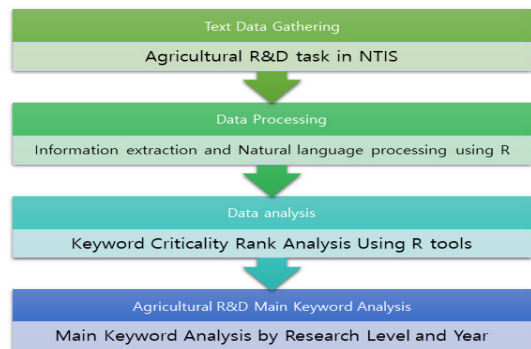


Fig. 2. Process of analysis

그룹화하여 정리한 것이다. 텍스트마이닝을 통해 도출한 키워드는 Table 4의 기준으로 그룹화하는 과정에서 정확성을 높이기 위하여 NTIS 원자료 상 DB와 비교 검토하였다.

여기서 '9. 보급 및 지도'의 경우 주요 키워드에 선정되지 않았기에 제외하였다. 농업 R&D의 주요 키워드를 분석하기 위해 사용한 프로그램은 R 3.6.1이다.

Table 4. Agricultural R&D List

Category	Group Name
R&D Activity	1. Breeding improving
	2. Basic-based research
	3. Manufacturing Technology
	4. Material Development Technology
	5. Green and Climate Change Response Technology
	6. Processing, Distribution Technology
	7. Sales-Consumption Technology
	8. Policy and Plan
	9. Diffusion and Coach
Agricultural Resource	10. Crop Area
	11. Horticultural and Herbal Area
	12. Animal Area

Data source : J. R. Lee et al(2012)[16]

Table 5. Keyword Group by Agricultural R&D List

Group No.	Keyword
1	Line, Collection, New variety, Breed, Seed, Variety
2	Diversity, Protein, Database, Monitoring, Virus, Revelation, Disease, Big data, Ecosystem, Biologically Active, Stress, Plants, Safety, Gene, Genetic Resources, Hereditary, Genom, Crop, Transcriptome, Transformation, Heavy metal, Character Evaluation
3	ICT, High quality, Pesticide, Pest control, Disease and Pest, Agricultural water, Pest, Industrialization, Production, Productivity, Minor crop, Smart farm, Nurture, Cropping System, Cultivation, Natural Enemy, Resistance, Eco-friendly, Quality
4	Functionality, Microbe, Food, Energy, Cosmetics
5	Climate change, Green house gases
6	Fermentation, Management, Post harvest management, Utilization promotion, Storage
7	Administration, Commercialization, Export
8	Economic efficiency, Farmer, Brand
10	Potato, Rice, Food crop, Maize, Feed, Bean
11	Citrus, Pepper, Strawberry, Pear, Cabbage, Chrysanthemum, Peach, Apple, Medicinal crop, Ginseng, Vegetable, Tomato, Grape
12	Insect, Silkworm, Pig, Hanwoo

4. 분석 결과

Table 6은 기초단계 키워드 분석 결과이며, Table 7, 8은 각각 응용단계와 개발단계의 키워드 분석 결과이다. 연구개발단계별 농업 R&D의 키워드는 연도별로 1위부터 10위까지 도출하였다.

Fig.3은 연도별 기초연구단계의 농업 R&D 키워드를 그룹화한 결과이다. 기초연구 단계에서 주된 특징은 생명공학의 '기초기반 연구' 위주의 키워드가 상위에 나타나고 있는 점이다. 대표적으로 '유전자(Gene)', '유전체(Genome)', '형질전환(Transformation)'과 같은 키워드가 상위에 있었다. 이는 농업 R&D의 기초연구는 생명공학의 기초 이론 지식의 축적을 바탕으로 혁신적이고 고부가 가치 창출을 위한 연구가 기초 과학이 활발히 이뤄지고 있음을 의미한다. 다른 특징으로는 '녹색 및 기후 변화 대응기술'의 키워드인 '기후변화(Climate change)'는 '09년부터 대두되기 시작하였다. 이는 과거의 농업 환경에서 이상 기후의 증가, 온난화로 인한 식물생육 한계선의 북상 등과 농업 환경 변화에 대응하기 위한 R&D가 새로운 주요 키워드로 부각되는 것으로 나타난다.

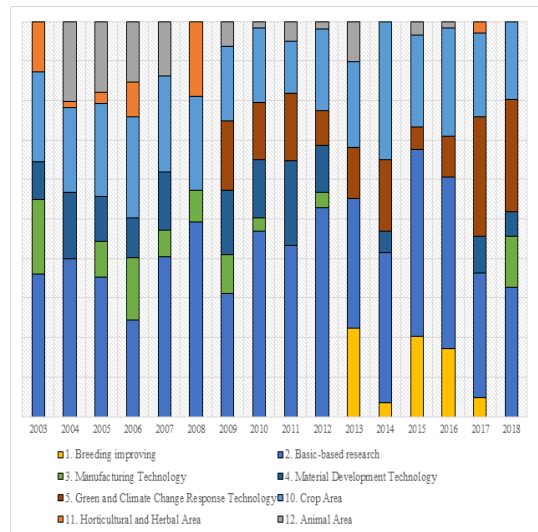


Fig. 3. Result of Agricultural R&D Keyword Textmining in Basis Research level

Table 6. Ranking Agricultural R&D Keyword Textmining by Basis Research level

Year Rank	2003	2004	2005	2006	2007	2008	2009	2010
1	Gene	Genetic resource	Gene	Rice	Gene	Genetic resource	Transformation	Gene
2	Rice	Rice	Rice	Gene	Rice	Gene	Gene	Transformation
3	Genetic resource	Gene	Genetic resource	Genetic resource	Genetic resource	Rice	Rice	Functionality
4	Resistance Transformation	Silkworm	Functionality	Transformation	Transformation	Genom	Climate change	Virus Rice
5	Bean	Functionality Hanwoo Transformation	Virus Bean	Resistance Bean	Genom	Transformation	Genetic resource	Climate change
6	Citrus	Virus	Silkworm	Functionality Pig / Hanwoo	Microbe Resistance	Resistance	Virus	Genetic resource
7	Pesticide	Microbe	Hanwoo Transformation	Pesticide	Functionality Bean / Pig Hanwoo	Virus	Functionality	Pig
8	Pepper Revelation	Biologically active Evaluating properties	Pig / Pear Genom	Pest Control Pear	Virus	DNA / Cabbage Bean	Resistance	Greenhouse gases
9	Functionality Protein Microbe	Grape	Pesticide	Insect Grape	Diversity Stress	Citrus	Pig Microbe	Microbe
10	DNA / Stress Pest control Hereditary	Pesticide	Resistance	Microbe	Hereditary	Collection Grape Monitoring	Plants	Ecosystem
Year Rank	2011	2012	2013	2014	2015	2016	2017	2018
1	Gene	Rice	Gene	Genom	Climate change	Climate change	Genom	Climate change
2	Transformation	Gene	Genetic Resource	Rice	Genom	Genom	Climate change	Rice
3	Rice	Climate change	Rice	Gene	Rice	Rice	Rice	Pest Control Genetic Resource
4	Genom	Genetic Resource	Diversity	Diversity	Gene	Gene	Gene	Bean
5	Functionality	Genom	Genom Variety	Climate change	Functionality	Database	Bean	Agricultural water
6	Climate change	Transformation	Collection	Collection	Bean	Functionality Bean	Database	Fertilizer usage
7	Proein Genetic Resource	Functionality	Climate change	Variety	Heavy metal Variety	Pesticide Pest Control	Pest Control	Soil physical properties
8	Disease and Pest	Variety	Pig	Genetic Resource	Ginseng	Genetic Resource Heavy metal	Food crop	Soil chemical properties
9	Pig	Disease	Bean	Pig	Crop	Crop	Functionality Transformation	Genetic Resource Resistance
10	Diversity	Hanwoo	Transformation	Bean	Monitoring	Transcriptome	Transcriptome	Genom

Table 7. Ranking Agricultural R&D Keyword Textmining by Apply Research level

Year Rank	2003	2004	2005	2006	2007	2008	2009	2010
1	Rice	Rice	Rice	Rice	Rice	Rice	Functionality	Rice
2	Transformation	Transformation	Transformation	Transformation	Bean	Gene Bean	Rice	Functionality
3	Gene	Gene	Gene	Gene	Functionality Variety	Variety	Pest Control Transformation	Natural Enemy
4	Bean	Functionality	Variety	Pest Control	Transformation	Functionality	Gene	Disease and Pest Feed
5	Pig	Bean	Pig Bean	Resistance Variety	Genetic resource	Transformation	Pig	Apple
6	Administration	Genetic resource	Functionality Pear	Functionality Pig Genetic resource Eco-friendly	Breed	Breed	Cropping system	New variety Cultivation
7	Resistance Hanwoo	Resistance	Resistance Eco-friendly	Bean	High Quality	Virus Genetic resource	New variety Feed Eco-friendly Bean	Pest control
8	Economic efficiency	Hanwoo	Genetic resource	Quality	Gene	Disease and Pest	Natural Enemy	Safety Tomato
9	Proein	Variety	Rice	Microbe	Pig	Line Pig Pest control	Export	Export Eco-friendly
10	Variety	Citrus Rice	High quality	Virus Apple	Quality	Potato Ginseng	Management Disease and Pest	Pig / Quality Transformation
Year Rank	2011	2012	2013	2014	2015	2016	2017	2018
1	Rice	Rice	Rice	New variety	New variety	New variety	New variety	Pesticide
2	Apple	New variety	Functionality	Rice	Variety	Rice	Climate change	Genom
3	Pest Control	Pest Control	New variety	Small area	Rice	Variety	Variety	Rice
4	Natural Enemy	Apple	Disease and Pest	Pest Control	Small area	Productivity	Rice	Climate change
5	Functionality Hanwoo	Functionality Eco-friendly	Pest Control	Functionality	Pest Control	Management Pest control	Productivity	Bean
6	Transformation	Disease and Pest	Variety	Hanwoo	Management Cultivation	ICT	Management	Minor Crop
7	Variety Pest	Hanwoo	Pest Transformation	Transformation	Disease and Pest Hanwoo	Apple	Big data	Variety
8	Eco-friendly	Vegetable	Quality	Variety Pest	Bean Transformation	Disease and Pest Bean	Pesticide	Big data
9	Export New variety Cultivation	Variety	Apple	Disease and Pest	Quality	Utilization promotion	ICT	Smart farm
10	Disease and Pest Genom Feed Tomato	Breed	Small area	Cultivation	Apple	Big data	Smart farm	ICT

Table 8. Ranking Agricultural R&D Keyword Textmining by Development Research level

Year Rank	2003	2004	2005	2006	2007	2008	2009	2010
1	Rice	Rice	Functionality	Functionality	Functionality	Rice	Functionality Variety Quality	Variety
2	Variety	Transformation	Transformation	Rice	Rice	Functionality Variety	Breed	Functionality Breed
3	Transformation	Functionality	Rice	New variety	New variety	Apple	Export	Export Quality
4	Breed	Gene	Gene	Variety Transformation	Transformation	Strawberry	High quality Rice	Pig Rice
5	Genetic resource	Genetic resource	Apple	Breed	Apple	High quality Transformation	Energy	Rice / Ginseng Transformation
6	Bean	Microbe Breed	Bean	Gene	Pepper Breed	Pest control New variety Ginseng	Cultivation	Cultivation
7	Apple	Food Bean	Resistance	Genetic resource	Cosmetics Eco-friendly	Peach	Pig / Ginseng Transformation	Eco-friendly
8	Quality	Potato	Genetic resource	Bean	Pear	Production Eco-friendly	Strawberry Microbe Biologically active	Gene
9	Pig / Virus Genetic resource	Resistance Variety	Eco-friendly	Pig	Genetic resource	Cultivation Bean	Brand	High Quality
10	Citrus Functionality	Virus	Pig	Food Quality	High quality Variety Quality Pest	Pear / Quality Hanwoo	Industrialization	Farmer
Year Rank	2011	2012	2013	2014	2015	2016	2017	2018
1	Functionality Variety	Variety	Variety	Variety	Variety	Rice	Variety	Export Variety
2	Quality	New variety	New variety	Rice	New variety	Export	Rice	Breed
3	Export	Breed	Breed	Functionality	Rice	Variety	Export	Rice
4	Pig	Functionality	Functionality	Breed	Breed	Breed	Breed	Nuture
5	Breed	High Quality	Rice	New variety	Disease and Pest	New variety	New variety	New variety
6	Ginseng Cultivation	Production	Post harvest mangement	Disease and Pest	Export	Microbe	Nuture	Maize
7	Transformation	Export	Disease and Pest	Post harvest mangement	Seed	Nuture	Functionality	Functionality
8	New variety Rice Gene	Chrysanthemum	Ginseng Quality	Production	Post harvest mangement	Apple	Rice	Bean
9	Commercialization	Ginseng	High Quality Chrysanthemum	Seed	Medicinal crop	Rice Bean	Medicinal crop	Quality
10	Eco-friendly	Management Rice	Production	Medicinal crop	Storage	Genetic resource	Fermentation	Ginseng

Fig.4는 연도별 응용연구단계의 농업 R&D 키워드를 그룹화한 결과이다. 응용연구 단계에서 주된 특징은 생산 기술과 품종육성 및 개량 기술 그룹의 키워드가 상위에 나타나고 있는 점이다. 특히 생산기술 그룹 내에서 '친환경(Eco-friendly)', '방제(Pest control)', '병해충(Disease and Pest)'과 같은 친환경 안정생산에 해당하는 키워드가 '16년도부터 나타나고 있는 'ICT', '스마트팜(Smart farm)'과 같은 신(新)기술 키워드로 연구의 방향이 전환되고 있다. 이러한 추세는 향후 4차 산업혁명과 연관되어 대두되고 있는 'AI', '무인 자동화'와 같은 새로운 기술 트렌드의 연계 가능성도 전망할 수 있는 대목이다. 품종육성 및 개량 그룹에서는 '신품종(New variety)' 키워드가 상위에 나타나고 있다. 이는 조사료(이탈리안그라스, 오차드그라스 등), 맥류(쌀보리, 겉보리 등), 과수류(복숭아, 사과, 감 등) 등 다양한 작물과 연계된 신품종 개량 연구가 활발하게 나타나고 있기 때문이다. 다른 특징으로는 기초연구 단계에서 '09년부터 대두되고 있는 녹색 및 기후변화 대응기술의 키워드가 응용연구 단계에서는 '17년부터 나타나고 있다. 이는 기초연구에서 응용연구로 이어지기까지 시차(time lag)가 발생하기 때문이다.

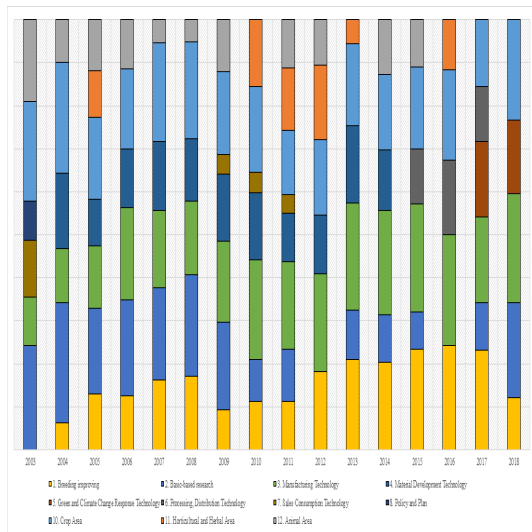


Fig. 4. Result of Agricultural R&D Keyword Textmining in Apply Research level

Fig.5는 연도별 개발연구단계의 농업 R&D 키워드를 그룹화한 결과이다. 개발연구 단계에서 주된 특징은 품종육성 및 개량기술 키워드가 상위에 나타나고 있는 점이다. 세부적으로 해당 키워드의 과제정보를 살펴보면 응용연구 단계에서 주요 키워드로 나타난 조사료, 맥류, 과수

류, 과채류 등의 개발연구가 나타나고 있었다. 이는 위와 마찬가지로 응용연구 단계에서 개발연구 단계로 연계까지 시차가 발생하고 있으며, 향후 실용화 및 상업적 목적을 가지고 품종육성 및 개량 연구가 연계될 것으로 전망된다. 다른 특징으로는 가공 및 유통기술 키워드 가운데 '수출' 관련 키워드는 '09년부터 나타나고 있다. 이러한 추세는 품종육성 및 개량기술과 소재 개발기술 키워드와 연계되어 지속될 것으로 전망된다. 한편, 전체 연구개발 단계에서 식량작물 분야의 '벼(Rice)' 키워드가 상위에 나타나고 있다. 이는 농업 R&D에서 벼와 관련된 연구가 과거부터 지금까지 계속하여 진행되고 있음을 의미한다.

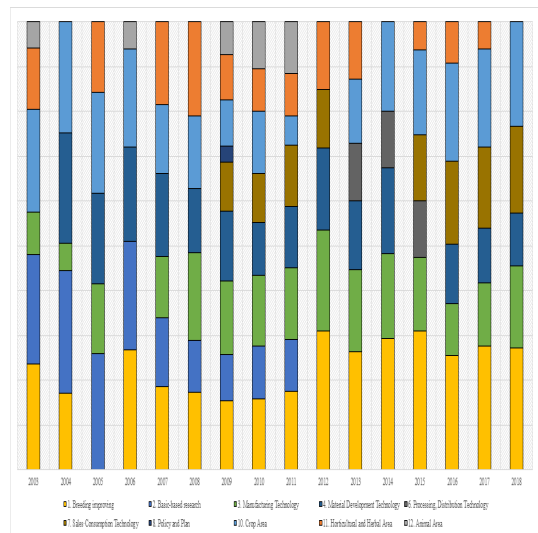


Fig. 5. Result of Agricultural R&D Keyword Textmining in Development Research level

5. 결론

본 연구는 농업 R&D의 트렌드를 살펴보기 위해 텍스트마이닝 기법을 활용하여 농업 R&D의 주요 키워드를 분석하였다. 분석을 위해 KISTEP에서 제공하고 있는 NTIS의 국가연구개발사업 과제정보를 활용하였다. 분석 기간은 '03년~'18년까지이다. 분석은 연구개발단계별로 농업 R&D의 키워드가 차이가 있는지를 세부적으로 살펴보기 위하여 연구개발단계를 구분하여 분석하였다.

주요 분석 결과는 다음과 같다. 첫째, 신기술의 도입과 외부 환경에 변화에 따른 농업 R&D 트렌드의 변화이다. 과거에서부터 최근 연도에 다다를수록 새로운 주요 키워드가 등장하고 있다. 기초연구 단계에서는 '기후변

화('09~)가, 응용연구 단계에서는 'ICT', '스마트 팜'(16~)가, 개발연구 단계에서는 '수출('09~) 키워드가 대두되고 있다. R&D는 중장기적으로 수행되기 때문에 '녹색 및 기후변화 대응기술', '생산기술', '가공 및 유통기술'을 중점으로 R&D가 지속될 것으로 전망된다. 둘째, 연구개발 단계에서 시차를 가지고 키워드 변화가 나타나고 있다. 기초연구-응용연구-개발연구 순으로 주요 키워드가 변화하고 있다. 대표적으로 '기후변화' 키워드가 기초연구 단계에서 응용연구 단계로 이어지고 있으며, '신 품종' 키워드는 응용연구 단계에서 개발연구 단계로 이어지고 있다. 이는 단계적으로 R&D가 수행되는 만큼 초기 R&D 수립에서 연구전략 수립 및 기획이 중요하는 것을 시사한다. 셋째, 농업 R&D의 대표적인 키워드는 '벼' 키워드로 나타났다. 모든 연구개발 단계에서 시간의 흐름에도 불구하고 '벼' 키워드는 상위권에 랭크되고 있다. 그러나 '벼' 연구의 방향성과 목적은 환경 변화에 따라 변화하고 있었다. 2000년대 초반에는 기초기반 연구와 품종육성 및 개량 연구를 중심으로 수행되었다면, 이후 생산기술 녹색 및 기후변화 대응, 가공 및 유통기술을 중심으로 연구가 수행되고 있다. 향후에도 '벼' 연구개발 방향은 국내·외 농업 환경 변화에 대응하면서 이슈에 맞는 연구가 수행될 것으로 예상된다. 특히 개발연구 단계에서 중점적으로 수행되는 가공 및 유통기술 분야로 트렌드 변화가 나타날 것으로 전망된다.

농업부문의 R&D 투자는 비단 농업만이 아니라 우리나라 전체 경제 성장에 기여해 왔다. 녹색혁명을 통해 식량 자급에 기여한 것은 국가적으로 중요한 의미를 지니는 일이었으며, 그 외 신품종이나 비료 및 농약, 그리고 농기계 등의 개발을 통해 생산성 증대에 기여하였다. 특히 농업부문의 생산성 증대를 통해 타 산업 분야에 노동력을 제공함으로써 국가 경제 성장의 밑거름이 되었다. 이러한 관점에서 최근 국가적으로 강조되고 있는 4차 산업혁명 대응과 관련해서 농업 R&D의 중요성은 더욱 부각되고 있다[17]. 그러나 최근 국가 R&D 내에서 농업 R&D가 차지하는 비중은 감소하는 것으로 나타났다. 특히 농업 R&D 예산은 증가하고 있지만, 타 분야에 비해 상대적으로 저평가되고 있다. 위와 같은 상황에서 향후 농업 R&D 방향성에 대한 제고가 필요하다. 본 연구는 국내 농업 R&D의 과거 추세를 살펴보고, 미래지향적인 농업 R&D 전략 방안 수립을 위한 연구의 일환으로써 의미를 가진다. 본 연구는 우리나라의 농업 R&D 키워드 트렌드를 분석하여, 향후 R&D 방향에 대한 예측을 하고자 하였다. 그러나 위 연구를 바탕으로 미래 정책 방향의

수립의 근거 자료로 활용하는데 한계를 가진다. 국내뿐만 아니라 해외 선진국의 농업 R&D 트렌드를 분석하여 글로벌 농업 R&D 트렌드를 살펴보는 것이 필요하며, 우리나라에 적합한 농업 R&D 방향성을 제시하는 것이 필요하다. 우리나라의 NTIS와 같이 주요 선진국들에서도 R&D 과제정보를 제공하고 있다. 미국의 경우 STAR METRICS, 유럽의 경우 CORDIS(Community Research & Development Information Service)에서 R&D 정보를 제공하고 있다. 위 기관에서 제공하는 공공 R&D 데이터를 바탕으로 트렌드 분석이 가능하지만, 이를 위해 데이터 수집에 많은 시간이 소모되며 데이터베이스 구축 작업을 위한 역량이 필요하다. 이 때문에 본 연구에서 국내 농업 R&D 트렌드 동향에 범위를 한정하여 분석한 점은 본 연구의 한계점으로 제시할 수 있으며, 향후에 선진국의 농업 R&D의 트렌드를 추가 분석하는 것이 필요로 한다. 더 나아가 뉴스, 신문기사, SNS, 블로그 등 다양한 데이터를 바탕으로 TF-DI(Term Frequency - Date Index), 언어네트워크와 같은 기법을 활용하여 최신의 농업 주요 이슈를 반영한 트렌드 분석을 실시하려는 노력이 필요하다.

References

- [1] J. R. Lee, "Understanding of Future Agricultural R&D and Our Choice, *Science & Technology Policy*, vol.23, no.1, pp 4-19, 2013
- [2] J. W. Yoon, Y. W. Chae, Y. J. Ahn, S. S. Kim, "A Priority Analysis of Excellent Agricultural R&D Performance Considering Region-wide Diffusion Effect", *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 21, No.4, pp.471-478, 2020 DOI : <https://doi.org/10/5762/KAIS.2020.21.4.471>
- [3] W. C. Seo, H. S. Park, J. H. Yoon, "An Exploratory Study on the Korean National R&D Trends Using Co-Word Analysis", *Journal of Information Technology Applications & Management*, vol.19 no.4 pp.1-18, Nov. 2012. DOI : <https://doi.org/10.21219/IITAM.2012.19.4.001>
- [4] P. R. Kim, J. P. Hong, S. J. Koh, "Big Data Technology R&D Trend through Patent Analysis", *Electronics and Telecommunications Trends*, vol.29, no.2, pp.33-41, April. 2014. DOI : <https://doi.org/10.22648/ETRI.2014.1.290204>
- [5] I. S. Jang, H. I. Jo, "Trend Analysis of the object Relations Theory in Korea : Focused on the KCI journals", *Korean Journal of Social Science*, vol.38, no.3, pp.109-140, Nov. 2019, DOI : <http://doi.org/10.18284/jss.2019.12.38.3.109>

[6] J. E. Min, E. J. Ju, Zhuqing Xia, "The Analysis of Foodservice & Restaurants Research Trend in Korea: Using Semantic Network Analysis", *Culinary Science & Hospitality Research*, vol. 26, no.5, pp.147-159, May, 2020, DOI : <https://doi.org/10.20878/cshr.2020.26.5.015>

[7] C. W. Jeong, J. J. Kim, "Analysis of Trend in Construction Using Textmining Method", *Journal of the Korean Digital Architecture Interior Association* no.12, vol.2, pp.53-60, 2012

[8] Y. I. Choi, Y. W. Lee, "Comparison of Keywords in Korean Science and Technology Policy and the National Research and Development Projects (2003-2017)", *Korea Technology Innovation Society Academic Conference*, Korea Technology Innovation Society, Korea, pp.80-98, 2018.11

[9] M. J. Kim, C. J. Kim, "Exploring Tradition Terminology Trends based on Keyword Analysis (1920~2017)", *Journal of The Korea Contents Association*, vol.18, no.12, pp.421-431, 2018, DOI : <https://doi.org/10.5392/JKCA.2018.18.12.421>

[10] Korea Development Institute(KDI), KDI Economic Information Center Dictionary of Current Term, KDI, c2019, Available From : <https://eiec.kdi.re.kr/material/wordDic.do?styp=all&key=R&D> (accessed Oct. 1, 2020)

[11] OECD, "Frascati Manual 2015(Guidelines for Collecting and Reporting Data on Research and Experimental Development)", OECD Publishing, 2015, pp.46-47, DOI : <http://dx.doi.org/10.1787/9789264239012-en>

[12] T. H. Jo, "Concept and Application of Text Mining", *Journal of Scientific & Technological Knowledge Infrastructure*, No.5, pp.76-85, 2001

[13] Christopher D. Manning, Marie-Cathrine de Marneffe, Anna N. Rafferty, "Finding Contradictions in Text", *Proceedings of ACL-08:HLT, Association for Computational Linguistics*, Columbus Ohio USA, pp.1039-1047, June 2008

[14] Andreas Hotho, Andreas Nürnberger, Gerhard Paaß, "A Brief Survey of Text Mining", *Ldv Forum, Computer Science Bibliography*, Vol.20, No.1, 2005

[15] Jay Jarman, "Combining Natural Language Processing and Statistical Text Mining: A Study of Specialized Versus Common Languages", Ph.D dissertation, University of South Florida, 2011

[16] J. R. Lee, C. S. Lee, Y. J. Jeong, S. K. Kim, M. J. Um, H. J. Yang, H. S. Yoon, S. W. Nam, "Exploratory Study on the Understanding of Public Agricultural R&D and Enhancing the Management of R&D Performance", *Policy Research, Science and Technology Policy Institute*, Korea, page.54, 2012, DOI : <https://doi.org/10.978.896112/2108>

[17] J. H. Heo, S. S. Kim, C. E. Lee, O. S. Kwon, "Economic Effects of Public R&D and Investment Demand in the

Agricultural Sector", *Research Report R826*, Korea Rural Economics Institute, 2017, DOI : <https://doi.org/10.23000/TRKO201800042640>

김 지 훈(Ji-Hoon Kim)

[정회원]



- 2017년 2월 : 단국대학교 환경자원경제학과 유통경제및식품경제 (경제학석사)
- 2016년 9월 ~ 2018년 12월 : 과학기술정책연구원 연구원
- 2020년 3월 ~ 현재 : 농촌진흥청 전문연구원

<관심분야>

농업 R&D, 경제성분석, 투자영향분석

김 성 섭(Seong-Sup Kim)

[정회원]



- 2013년 8월 : 충북대학교 농업경제학과 농업경영전공 (경제학석사)
- 2017년 8월 : 충북대학교 농업경제학과 농업경영전공 (경제학박사)
- 2019년 2월 ~ 현재 : 농촌진흥청 농업연구사

<관심분야>

농업위험관리, 농업투자분석, 농업R&D, 농업회계