

사용자의 국가 R&D 정보 이용 접근성 향상을 위한 큐레이션 서비스 : 이슈로 보는 R&D 사례를 중심으로

Curation Service to Improve User's Access to National R & D Information : Focusing on Issues R&D Service

유은지, 최광남, 황유나
한국과학기술정보연구원

Eun-ji Yu(eunjuu@kisti.re.kr), Kwang-Nam Choi(knchoi@kisti.re.kr),
Youna Hwang(ynhwang@kisti.re.kr)

요약

국가 R&D 데이터는 기초과학 연구부터 산업화 부분까지 전 분야에 대한 정보를 포괄하고 있지만 전문적인 용어로 표현되며 이로 인해 대중의 이용에는 어려움이 있다. 이에 NTIS 는 국가 R&D 데이터를 이용한 데이터 큐레이션 서비스를 개발하여 국가적인 현안과 사회적 이슈에 대해 국가 R&D 정보를 선별하여 제공하고 있다. 이에 본 연구에서는 NTIS 의 데이터 큐레이션 서비스인 이슈로 보는 R&D 서비스 분석을 통해 국가 R&D 정보를 이용한 데이터 큐레이션 서비스 구축 방안을 제안하고자 하며 이 서비스가 국가 R&D 정보에 대한 사용자의 접근성 향상에 미친 영향도 분석하였다. 이슈로 보는 R&D 서비스는 뉴스기사에서 추출한 이슈와 관련된 국가 R&D 과제, 성과, 주요 연구기관 등을 매핑하여 정리, 제공한다. 패키징에 이용된 데이터는 모두 오픈되어 있고 관련 자료는 보고서 형식으로 정리되어 PDF 파일로 제공된다. 또한 해당 프로세스를 자동화 하여 관리자 뿐 만 아니라 NTIS 이용자라면 누구나 개인적인 이슈패키징을 가능하게 하였다. 이 밖에 Special Issue 코너를 개설하며 주요 현안에 대한 사용자의 접근,이용 편의성을 높였고 코너 개설 이후 접속자의 페이지뷰가 증가한 것으로 나타났다.

■ 중심어 : | 이슈로 보는 R&D | 데이터 큐레이션 | 과학기술표준분류 | 국가과학기술지식정보서비스(NTIS) |

Abstract

National R & D data covers information in all fields from basic science research to industrialization, but it is expressed in technical terms, which make it difficult for the public to use. Accordingly, NTIS developed and launched the data curation service 'R&D issue service', which selects national R&D information on national and social issues and provides them to the public. Therefore, this study aims to analyze the effect of a data curation service on NTIS users' access to R&D data and suggest how to develop the curation service. The R&D issue service extracts issue from the news article and provide related national R&D projects, achievements and major research institute. All raw data used for the service are open to the public, organized in a report format and provided as PDF files. In addition, automative process is developed for all NTIS users to make individual issue packaging like administrator. The results show that 'R&D issue service' launching increases users' access and convenience to R&D data related to major issues, and the number of page views of users increased after the service was opened.

■ keyword : | Issue R&D | Data Curation | National S&T Standard Classification System | NTIS |

* 본 연구는 2020년도 한국과학기술정보연구원(KISTI) 주요사업 과제로 수행한 것입니다.

접수일자 : 2020년 05월 28일

심사완료일 : 2020년 09월 08일

수정일자 : 2020년 09월 08일

교신저자 : 황유나, e-mail : ynhwang@kisti.re.kr

I. 서론

정보 기술의 발전과 함께 하루 수천, 수백억건의 데이터가 발생하고 공유, 유통되고 있다. 기하급수적으로 늘어나는 데이터로 인해 수용자가 접근 할 수 있는 정보의 물리적 양은 늘어났지만 그 만큼 무의미한 정보도 함께 증가하며 필요한 정보를 선별하는 과정의 중요성도 높아지고 있다.

1970년대 미디어의 발전과 함께 유통되는 정보가 증가하기 시작하며 계층 간 정보를 통한 지식 습득에 차이가 있으며 이는 계층 간 격차를 더 유발한다는 '지식 격차가설(Knowledge Gap Hypothesis)' 이 등장했다 [1]. 이는 정보화 시대를 거쳐 오며 새로운 정보매체에 대한 접근성을 가진 사람과 그렇지 못한 사람 사이에 차이가 유발된다는 정보격차이론 (Information Gap Theory) 으로 변화한다[2]. 하지만 정보 기술 발전과 함께 정보 격차 유발요인은 정보에 대한 '접근'의 문제 보다는 정보에 대한 '이해와 분별력'의 문제로 옮겨지게 된다[3].

한 연구자의 연구 데이터에 접근했을 때 사업가는 연구자의 기술사업화 실적, 대학은 과제와 논문연구실적 그리고 학생이라면 연구자의 강의평가 데이터에 보다 집중할 것이다. 또한 '코로나바이러스(Corona Virus)'의 연구경향을 알고 싶은 사람이라면 코로나 바이러스의 세계 피해상황에 대한 뉴스 정보 보다는 각 국가의 연구실적, 백신 연구, 특허 등의 정보가 필요할 것이다. 이처럼 빅데이터 시대에 지식정보 습득에 있어 중요한 것은 내가 알고자 하는 분야에 대해 얼마나 많은(how much)정보를 이용하는가 보다는 얼마나 정확한(how accurate) 정보를 적절하게(appropriate)이용하는가이다. 이러한 이유로 생산된 데이터 중 이용자에게 필요한 정보를 찾아내고 배열하는 중간과정, 즉 '데이터 큐레이션(data curation)'의 중요성이 높아지고 있다.

이처럼 지식정보시대에 정보를 제대로 이용하기 위해서는 정보를 선별, 정리하는 큐레이션 과정이 중요하다. 그리고 이용자에게 꼭 필요한(High Quality)큐레이션이 이루어지기 위해서는 해당 정보를 얼마나 잘 이해하고 있는지가 가장 중요하다. 정보에 대한 이해가 바탕이 되어야 이를 통해 필요한 데이터를 선택, 분류

할 수 있기 때문이다. 국가 R&D정보는 국가과학기술정보라는 데이터의 특수성과 수집과정의 어려움으로 인해 일반 이용자가 접근하기 어려운 정보였다. 2013년 공공데이터 개방 정책이 시행된 이후 정부에서 국가과학기술지식정보서비스(National Science & Technology Information Service, 이하 NTIS)를 통해 국가가 수집하고 있는 연구개발 정보(437개의 표준항목) 중 80%가 넘는 항목을 공개, 개방하고 있다[4]. 정보의 공유, 이용에 대한 관심이 높아지면서 국가 R&D 정보 접근성도 높이기 위해 다양한 서비스를 시도하고 있지만 2019년 11월 기준 일반이용자 비율은 38%로 연구자, 기관 이용자 (61%)에 비해 낮다. 특히 국가 R&D 데이터의 경우 접근이 용이하다 해도 전문적인 내용과 수집 항목의 용어 등이 낯선 만큼 이용자의 니즈(needs)에 따른 정보 검색에 어려움이 있다.

이에 NTIS에서는 국가 R&D 정보에 대한 접근과 이용편의성을 높이기 위해 데이터 큐레이션 서비스인 '이슈로 보는 R&D'를 개발하였다. 이슈로 보는 R&D는 국가적인 현안과 사회적 이슈에 대하여 과제, 논문, 특허, 연구자 및 연구기관 등 국가 R&D 관련 정보를 선별하여 발 빠르게 제공하는 서비스다.

이에 본 연구에서는 특수한 성격을 띠고 있는 국가 R&D 정보를 이용한 데이터 큐레이션 서비스 구축 방안을 제안하고자 한다. 이를 위해 NTIS의 이슈로 보는 R&D 서비스를 분석 하고 이와 함께 데이터 큐레이션 서비스가 국가 R&D 정보에 대한 사용자의 접근성 향상에 미치는 영향에 대해서도 분석하고자 한다.

이를 위해 우선 '이슈로 보는 R&D'의 서비스 구조와 구축 절차 등을 설명하고 해당 서비스에 접근한 이용자의 페이지뷰 로그를 통해 사용자 접근성에 대해 분석하고자 한다.

II. 관련연구

1. 데이터 큐레이션(Data Curation)

큐레이션(curation)은 과거 전시품, 작품 등을 수집하여 소개하는 것을 일컫는 말로 주로 박물관, 미술관과 같은 곳에서 사용되며 자료 전시, 홍보 활동 등을 하

는 사람을 일컫는 의미로 널리 쓰였다. 하지만 빅데이터 시대로 오며 예술품이 아닌 '데이터(data)'를 수집, 전시, 홍보하는 과정을 데이터 큐레이션(data curation)으로 표현한다. 대표적인 큐레이션 서비스로는 데이터 네이버의 오픈캐스트와 아마존의 콘텐츠 추천 서비스 그리고 유튜브의 동영상 추천 서비스가 있다.

데이터 큐레이션은 데이터의 중요성이 커지면서 함께 강조되고 있다. 데이터 큐레이션에 대한 연구는 주로 도서관학에서 수행되었다[5]. 도서관에서 소장하고 있는 수많은 정보를 이용자 맞춤형으로 제공하고자 주로 연구되었으며 그 외에 교수학습 지원과 연구자원을 위해서도 데이터 큐레이션 서비스가 연구되었다[6][7]. 최근에는 이 외에도 디지털 데이터, 방송콘텐츠, 소셜 데이터 그리고 과학데이터등 다양한 분야에서 큐레이션 서비스에 대한 연구가 진행되고 있다[8-11].

기존연구에서 데이터 큐레이션에 대한 정의는 크게 두 가지로 이용의 차원으로 분류할 수 있다.

데이터의 관리 차원에서 데이터 큐레이션은 데이터의 생애 주기에 맞춰 저장, 관리하며 보존, 재사용을 도모하는 활동으로 정의되며 데이터의 이용 차원에서는 데이터를 정리, 관리, 그리고 분석과 재평가를 통해 활용, 공유 가능하게 함으로써 이용자에게 서비스하거나 재사용이 가능한 메타데이터를 생성하여 부가가치를 제고하는 활동에 집중하는 것으로 정의된다[5][6][12][13].

이 밖에 데이터 큐레이션 과정에서 정보를 필터링하는 주체에 따라 생산자가 주체일 경우에는 데이터 큐레이션으로 그리고 사용자가 주체일 경우에는 소셜 큐레이션으로 정의되기도 한다[9].

이에 본 연구에서는 데이터 큐레이션에 대해 데이터의 수집 관리 보다는 이용자의 활용에 보다 집중하여 '방대한 데이터 속에서 사용자가 필요로 하거나 가치 있는 정보를 선별하고 나열하여 사용자에게 보다 편하게 접할 수 있는 방법으로 제공하는 일련의 과정을 '데이터 큐레이션'으로 정의하고자 한다.

이처럼 다양한 데이터를 이용한 큐레이션 서비스에 대한 분석연구가 시행되고 있지만 R&D 데이터를 이용한 큐레이션 서비스 연구는 매우 드문 실정이다. 이에 본 연구에서는 국가 R&D 정보데이터를 바탕으로 큐레

이션 서비스를 제공하고 있는 '이슈로 보는 R&D' 서비스를 분석하고자 한다.

2. 국가과학기술정보서비스(NTIS)

국가과학기술정보서비스(NTIS: National Science and Technology Information Service)는 2008년 서비스 오픈 이후 현재까지 국가연구개발사업과 관련된 정보를 통합 제공하는 국가 R&D 지식포털로서 운영되고 있다. 2016년까지는 연구자를 대상으로 정보를 제공하던 것에서 나아가 2017년부터는 대상을 확대하여 누구나 정보를 확인할 수 있다. NTIS에서는 국가 R&D 정보를 총괄 수집하고 관리하는 것 외에도 수집한 R&D 정보를 이용하여 국가차원의 공동 활용을 지원하고 사용자의 데이터 접근이 용이하도록 다양한 서비스를 제공 하는 등 수집된 데이터를 이용해 다양한 서비스를 지원한다. NTIS 를 이용한 연구도 활발히 진행 중이며 관련 연구는 NTIS 의 서비스에 대한 직접적인 연구과 NTIS 에서 제공한 데이터를 이용한 간접적인 연구로 구분 할 수 있다.

NTIS 의 서비스에 대한 직접적 연구로는 NTIS 에서 제공하는 서비스에 대한 시스템 구축 연구, 서비스의 경제적 파급 효과 연구, 시스템 개선 연구 등이 있다 [14-17].

NTIS의 데이터를 이용한 연구로는 NTIS 와 타 서비스의 연계에 대한 분석, NTIS 에서 제공하고 있는 과제 정보, 성과정보 를 이용한 국가 R&D 에 대한 분석연구 등이 있다[18-20].

III. 연구내용 및 방법

1. R&D정보 큐레이션 서비스 구축 절차

1.1 현안 및 이슈 추출

국가 R&D정보의 경우 전문적인 용어로 표현되고 전문 기술 키워드가 쓰이기 때문에 해당 용어에 익숙하지 않은 일반 이용자의 경우 국가 R&D 정보에 대해 어렵다고 느끼게 된다. 그러나 실제 뉴스에서 언급되는 많은 사회 이슈들은 국가 R&D 연구 주제와 밀접하게 관련되어 있다. 한 예로 '미세먼지' 의 경우 여러 미디어

매체에서 보도하고 있는 주요 이슈로 이에 대한 뉴스는 쉽게 접할 수 있다. 이 같은 이슈는 사회적 논의 뿐 아니라 국가 R&D 분야에서도 미세먼지 저감, 미세먼지 집진기, 미세먼지 측정 등을 주제로 활발하게 연구되고 있다.

이 점을 활용하여 이슈로 보는 R&D 에서는 일반 사용자가 일상에서 쉽게 접하고 있는 뉴스와 정책 동향에서 큐레이션의 주제가 되는 현안 및 이슈를 추출하고 이를 기반으로 관련된 국가 R&D 정보를 큐레이션 함으로써 일반 이용자가 국가 R&D 정보를 쉽게 접하고 이해할 수 있도록 하고자 하였다. 이를 위해 과학 관련 뉴스, 이슈와 같은 콘텐츠를 제공하는 과학뉴스 매체, 정책 동향 원문을 수집하고 있다. 원문 수집 대상은 [표 1]과 같다.

표 1. 이슈로 보는 R&D 원문 수집 대상

매체		설명
정책 동향	정책브리핑	정부부처의 정책 동향 및 정책 브리핑
	ITFIND	ICT 뉴스, 동향정보
뉴스	다음(DAUM)	사회, 경제, IT 분야
	한겨레	의료-건강, IT, 미래, 과학, 기술, 환경 분야
	중앙일보	과학, 산업미래, 환경 분야
	전자신문	통신방송, SW개암성장기업, 소재부품, 전자자동차유통, 경제금융, 산업과학정책 분야
	사이언스타임즈	기초응용과학, ICT-로봇, 자연환경에너지, 보건의학, 항공우주, 신소재신기술 분야

[표 1]과 같이 다양한 매체로부터 국가 R&D 연구 주제를 추출하기 위해 매일 기사 원문을 자동 수집한다. 수집된 뉴스, 정책 동향은 텍스트 형태로 되어 있으며, 텍스트 원문으로부터 형태소 분석을 통해 주요 키워드를 추출한다. 여기서 주요 키워드란 수집한 원문 기사 내에서 빈도수가 높은 키워드들로 코로나바이러스, 미세먼지와 같은 키워드의 형태이다. 이때 추출된 주요 키워드는 이슈 주제의 후보가 되고, 다음 단계인 클러스터링 분석을 통해서 이슈 주제 후보로 선정된다.

또한, 이 과정에서는 정확하고, 활용 가능한 키워드를 추출하기 위해 국가 R&D 과제 정보를 기반으로 구축된 NTIS 국가 R&D 용어사전을 활용한다. 용어 사전은 국가 R&D 용어, 영문대역어, 용어의 과학기술표준분

류, 불용어로 구성되어 있어 다음 단계인 클러스터링과 큐레이션을 위한 정보 패키징을 효과적으로 수행 할 수 있도록 필터링 하는 과정을 거치는 것이라고 볼 수 있다. 국가 R&D 용어는 국가 R&D 과제에서 언급되는 연구 주제, 기술 키워드 등이 추출되도록 하고, 불용어는 분석에 들어가지 않도록 제외시키는 리스트로 한 글자 문자, 국가 R&D 연구와 관련이 낮은 키워드, 출현 빈도가 낮은 키워드 등으로 구성되어 있다.

용어의 과학기술표준분류는 해당 용어가 국가 R&D 연구 과정에서 활용 될 때 어떤 분야의 연구였는가를 나타내는 분류 정보이다. 예를 들어 최근 몇 년간 광범위한 분야에 활용되고 있는 '인공지능'이라는 키워드는 과학기술표준분류 상의 '통합 재난관리' 분류를 갖고 재난 분야에서 활용 될 수도 있고, '보건/의료' 분류를 갖고 의료분야에서 활용될 수 있다. 이렇게 하나의 용어는 여러 개의 과학기술표준분류 정보를 가질 수 있으며, 패키징 과정에 활용 할 때는 큐레이션 관리자가 큐레이션 하는 이슈 주제에 따라 선택한다. 선택된 분류는 관련 국가 R&D 과제를 큐레이션 할 때 중요한 역할을 한다.

이렇게 필터링 과정을 거친 키워드를 기반으로 큐레이션 주제를 추출하기 위한 원문별 키워드셋(set)에 대해 클러스터링을 수행한다. NTIS 에서는 분석도구인 R 을 기반으로 K-Means 계산법을 적용하여 분석한다. 이 과정에서 분석 대상이 되는 원문의 추출 기간, 클러스터링 개수 등 분석 결과에 영향을 미칠 수 있는 조건들을 큐레이션을 수행하는 관리자가 조정 할 수 있도록 설계하였다.

설정된 기준에 따라 클러스터링 분석이 완료되면 몇 개의 주제 후보가 자동으로 추출된다. 큐레이션 관리자는 NTIS 과제 DB를 참조하여 주제 후보 키워드에 대한 관련 R&D 과제가 대략 몇 건 정도 있는지 확인하고 최종 이슈 주제를 결정하게 된다. 어떤 주제에 대해 국가 R&D 연구 현황을 파악하기 위해서는 최근 5년 간 수행된 과제 건수가 최소 30 ~ 50 건 이상이어야 되기 때문에 관리자가 서비스 전에 이 부분을 확인해야 한다.

1.2 큐레이션 주제 관련 R&D정보 패키징

앞서 설명한 것처럼 키워드 추출, 키워드 필터링, 클러스터링, 관리자 검토 과정을 거쳐 큐레이션 주제가 결정되면 해당 주제에 대한 국가 R&D 연구 현황을 제공하기 위한 패키징 과정이 진행된다. 전체적인 패키징 과정을 도식화 한 것은 [그림 1]과 같다.

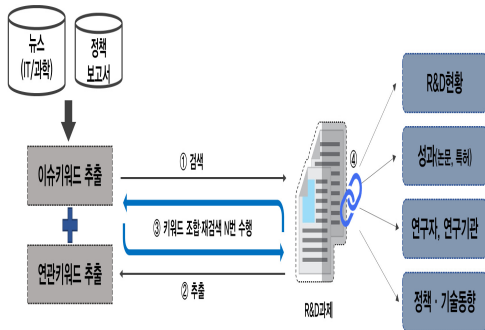


그림 1. R&D 과제 패키징 프로세스

첫 번째 단계는 큐레이션 주제 키워드를 기반으로 국가 R&D 과제 정보 추출하는 단계이다. 추출된 과제 이후 함께 제공되는 논문, 특허, 정책 동향, 주요 연구자 및 연구기관을 추출하는 기준이 되기 때문에 매우 중요하다.

예를 들어 '코로나 바이러스'가 큐레이션 주제가 될 경우 용어사전에서 '코로나 바이러스'의 분류정보를 조회하고 원문의 주제와 관련 있는 '보건/의료' 분류를 선택한다. 최근 5년 동안의 국가 R&D 과제정보에서 과학기술표준분류가 보건/의료이면서 코로나 바이러스 키워드가 포함된 과제를 후보 과제로 1차 추출한다. 키워드에 따라 적게는 몇 십 건에서 몇 천 건의 과제가 추출된다. 최종 서비스 대상 과제는 사용자가 원하는 정보만 빠르게 취사선택하여 확인할 수 있도록 50~200 건 정도의 과제가 패키징에 포함되도록 한다. 후보 과제가 이보다 더 많이 추출된 경우에는 주제키워드와 연관키워드를 조합하여 추가로 AND 검색을 수행한다. 이를 통해 서비스에 적합한 과제데이터를 추가적으로 필터링하는 과정을 거치게 된다. 예를 들어, 미세먼지를 주제 키워드로 하여 서비스하고자 하면 서비스 건수보다 많은 약 5,000건의 과제가 추출된다. 기사 원문에서 추출한 연관 키워드를 추가로 검색하면 기사에서 다루고 있는 이슈에 따라 미세먼지와 관련된 질환을 주제로

이야기하는지, 미세먼지 집진기와 같은 미세먼지 저감을 위한 기술에 대해 이야기하는지에 따라 제공될 이슈에 적합한 과제가 필터링된다. 여기서 연관 키워드란 원문에서 추출된 키워드와 1차로 추출된 과제에서 추출된 고빈도 키워드를 말한다. 이렇게 주제 키워드와 연관 키워드의 조합으로 적정 과제 건수로 필터링 될 때까지 여러 번의 AND 검색을 수행하게 된다.

두 번째 단계는 국가 R&D 성과 정보 추출 단계이다. 앞선 단계에서 적정 과제 건수에 도달하면 관련 국가 R&D성과 정보를 추출한다. 성과 정보를 추출하는 방식을 2가지가 있다. 첫 번째는 패키징 된 과제에서 유발된 논문, 특허를 추출하는 것이다. 과제에서 유발된 성과를 측정하는 것은 투입 예산 대비 성과를 측정하는 것으로도 볼 수 있기 때문에 유의미한 정보라고 볼 수 있다. 그러나 과제에서 유발된 성과가 반드시 코로나 바이러스의 키워드를 포함하거나 해당 키워드를 주제로 논문이나 특허가 작성되지 않았을 수 있다는 우려가 있다. 이는 대형 과제일수록 다양한 연구 주제를 포함하고 있고 성과는 다양한 주제 중 한 가지에 초점을 맞춰 도출될 수 있기 때문이다. 따라서 두 번째 방법에서는 과제 추출에 사용된 키워드를 검색 로직을 통해 성과 추출에 사용하여 실제로 관련 키워드들이 사용된 논문과 특허를 추출한다. 어떤 방법을 사용할 지는 성과 정보를 활용하기 위한 조건에 따라 적절하게 수행한다.

세 번째 단계는 해당 분야의 연구를 수행하는 연구자와 연구기관 정보이다. 추출된 과제를 수행한 연구책임자와 연구 수행기관을 추출하여 과제의 연구비를 기준으로 상위 연구자와 연구기관을 각각 20 건을 추출하여 제공한다. 추가적으로 NTIS 참여인력 정보 API를 통해 과제를 수행한 주요 연구자의 기본 정보, 논문, 지식재산권, R&D참여과제, 저역서 등 상세 정보를 확인할 수도 있다. 연구자와 연구기관의 과제 상세내역까지 확인할 수 있어 수행 이력을 참고하여 협업 연구자 또는 연구기관을 탐색 하는데도 활용 될 수 있다. 마지막으로 큐레이션 주제키워드가 포함된 정책 동향과 기술동향을 패키징에 포함시킨다. 활용되는 정보는 정책브리핑에서 연계된 것과 NTIS 정책, 기술동향 DB에서 추출한 것으로 최근 1년을 기준으로 추출하여 이전 과정에서 추출된 과제, 논문, 특허, 정책 동향, 주요 연구

자 및 연구기관 등을 함께 묶어 패키징 형태로 제공한다.

앞서 설명한 현안 및 이슈 추출, 큐레이션 주제에 대한 관련 R&D 추출 과정을 수행하기 위한 기능들을 포함한 이슈로 보는 R&D 서비스 설계도는 [그림 2]와 같다. 원문, 국가 R&D 정보를 기반으로 크게 원문관리, 원문분석, 용어관리, 패키지관리와 같은 기능들로 구성되어 있다.

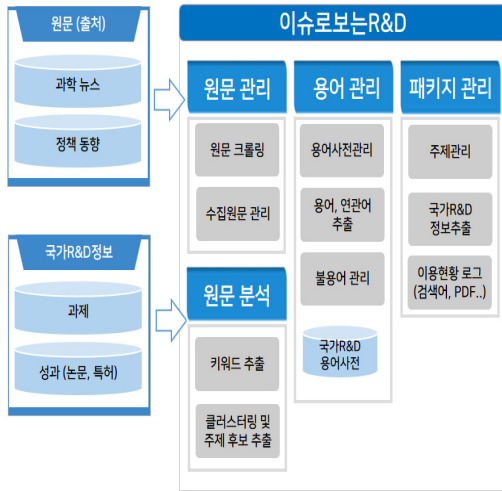
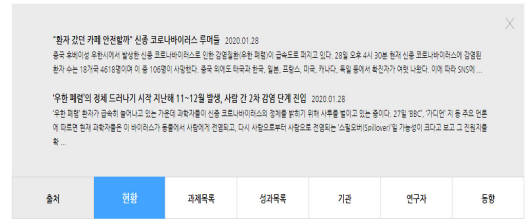
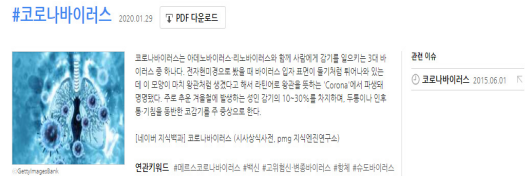


그림 2. 이슈로 보는 R&D 설계도

원문 관리 기능은 큐레이션 주제를 추출하기 위해 과학뉴스, 정책 동향 등 크롤링의 대상이 되는 외부 사이트나 원문 파일 등에 대한 출처 정보를 관리하는 기능이다. 해당 기능에서 수집된 원문 정보를 기반으로 키워드 추출, 클러스터링 및 주제후보를 도출하는 과정은 원문 분석 기능을 통해 수행한다. 키워드를 기준으로 분석이 수행됨에 따라 어떤 키워드를 뽑아 분석에 활용하고 필터링 하느냐가 매우 중요하다. 따라서 용어 추출, 관련어 추출, 불용어 관리, 신규용어 추출 등 용어 관리 기능을 구축하여 이를 활용해 원문 추출부터 패키징까지 폭 넓게 활용하고 있다. 마지막으로 패키지 관리 기능을 통해 큐레이션 주제를 관리하고 관련된 국가 R&D 정보를 추출하고 정제하기 위한 세부 기능들이 구축되어 있다. 최종적으로 이용자에게 제공된 큐레이션 정보에 대한 페이지뷰 로그, PDF 다운로드 제공 현황 등을 관리하는 기능까지 구축되어 있다.

2. 서비스 제공 현황

2.1 서비스 현황



패키징 단계를 거쳐 제공되는 서비스 화면은 [그림 3]과 같다. 서비스 이슈에 대한 관련 이미지, 네이버백과 또는 위키백과를 연계한 이슈에 대한 정의, 연관키워드, 관련이슈, 관련기사, 현황그래프, 관련 과제, 성과, 정책 동향, 주요 연구자 및 연구기관 정보가 제공된다. 관련 현황으로는 최근 5년 간 연도별 관련 과제, 연구비, 참여인력, 성과 건수, 연도별 과제구분정보별 과제 건수 등이 제공된다. 키워드 검색을 통해 현재까지 제공된 이슈 주제를 검색하는 것도 가능하다.



그림 3. 이슈로 보는 R&D 예시 (코로나바이러스)

또한 사용자가 패키징 결과를 좀 더 수월하게 이용할 수 있도록 패키징 결과를 보고서 형태인 PDF 파일로 제공하고 있다. 패키징에 포함된 과제, 논문, 특허 상세 정보 등 모든 자료들도 엑셀 형태로 다운받을 수 있게 함으로써 사용자의 이용 편의성을 높이기 위해 노력하였다.

2.2. Special Issue

기존에 이슈로 보는 R&D 서비스는 서비스 화면에 게시된 날짜순으로 NTIS 홈페이지에 노출되었다. 이 경우 게시일이 오래되면 해당 이슈가 오래 지속되고 있는 사회의 주요 현안일 경우에도(예: 코로나바이러스) 관련된 패키징 자료가 첫 화면에 나타나지 않아 사용자의 접근이 어려웠다. 이에 2020년 3월부터는 [그림 4]에서 볼 수 있듯이 주요 이슈와 관련된 이슈패키지들을 그룹핑하여 패키징의 게시일과 관계없이 지속적으로 메인에 노출시키는 ‘Special Issue’ 코너를 NTIS 홈페이지 상단에 추가하여 게시하였다.

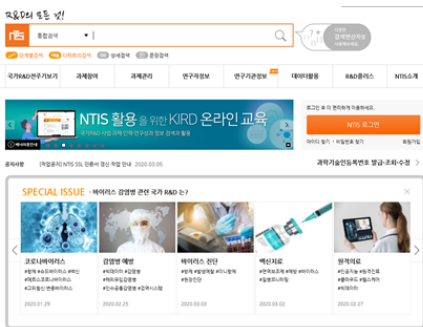


그림 4. NTIS 홈페이지의 Special Issue 코너

‘코로나 바이러스’ 이슈의 경우 2020년 1월 29일 처음 게시된 이후 Special Issue 코너가 개설되기 전인 3월 2일까지의 페이지뷰는 694회로 나타났다. 하지만 3월 3일 서비스 개설 후 3월 24일까지 약 3주간의 페이지뷰는 1,820회로 사용자의 페이지 뷰가 약 2.5배 증가하였다.

또한 [표 2]에서 볼 수 있듯이 비슷한 시기에 게시된 주제 중 Special Issue 에 그룹핑 된 주제가 그렇지 않은 주제보다 평균 페이지뷰가 약 3배 더 많은 것으로

나타났다. 일반 이슈의 평균 페이지뷰가 Special Issue 코너 개설 전 보다는 증가한 것으로 나타나며 해당 코너의 개설이 큐레이션 되어 제공되는 국가 R&D 정보에 대한 사용자의 접근성과 향상에 긍정적인 영향을 미쳤음을 보여준다.

표 2. Special Issue 와 일반이슈의 페이지뷰 비교

	주제	게시일	페이지뷰	평균 페이지뷰
Special Issue	바이러스 진단	2020-03-03	567	303.25회
	역학조사	2020-03-04	302	
	진단키트	2020-03-17	173	
	유전자 증폭기술	2020-03-18	171	
일반 이슈	당뇨병 검사	2020-03-06	110	96.5회
	필서	2020-03-09	62	
	박테리오파지	2020-03-10	117	
	사이클로덱스트린	2020-03-13	97	

2.2 연구 분야 분류에 따른 서비스 현황

서비스를 개시한 2014년부터 현재까지 패키징된 개수는 약 1,200 건으로 환경, 로봇, 인공지능, 바이오 등 매우 다양한 분야의 이슈가 서비스 되었다. 이렇게 서비스된 이슈의 연구 분야를 이슈 패키징에 포함된 과제의 과학기술표준분류 정보를 기준으로 분석하였으며 상위 20개 연구 분야이며 과학기술표준분류 중 중분류 데이터를 기준으로 하였다.

[그림 5]에서 볼 수 있듯이 ‘이슈로 보는 R&D’에서 가장 많이 서비스된 주제는 보건의료 분야이다. 의약품/의약품개발, 의생명과학, 임상의학, 치료/진단기기 등 보건의료와 관련된 주제가 제공된 것을 확인 할 수 있었다. 두 번째 분야는 정보/통신 분야의 주제로 소프트웨어, 정보이론, 자동화기계, 컴퓨팅, 정보보호 등의 주제가 눈에 띄게 나타난 것을 볼 수 있었다. 실제 데이터를 살펴보면 로봇, 인공지능 관련 기술이 많이 언급된 것을 확인할 수 있었으며 이는 우리 사회에서 보건의료 그리고 정보통신 분야의 이슈가 주요사회 이슈로 떠오른 적이 많았다는 것을 보여준다.



그림 5. 과기표준분류 내 과제건수 비율

3. 이용자가 만드는 이슈 R&D

기사 원문에서 이슈를 추출하여 관련 R&D 정보를 제공하는 이슈로 보는 R&D 서비스가 운영되면서 이용자 개인 관심분야에 대한 패키징 정보를 제공받고자 하는 수요가 지속적으로 발생했다. 이에 따라 NTIS 에서는 2019년 12월부터 이용자가 직접 본인의 관심 분야에 대한 관련 R&D 정보를 패키징 할 수 있도록 신규 기능을 개발하여 서비스 하고 있다.

그림 6. 이슈로 보는 R&D 예시 (친환경 먹거리)

기존 서비스와 패키징 절차는 동일하지만 기사 원문에서 추출했던 이슈 키워드를 이용자가 직접 본인의 관

심 분야, 전문 분야에 기반 하여 주제를 선택할 수 있는 점이 가장 크게 다르다고 볼 수 있다. 사용자 이슈 패키징 서비스에서는 이용자가 주제 키워드를 입력하면 NTIS 과제 DB와 용어사전을 참조하여 이용자가 입력한 키워드와 관련 있는 키워드를 [그림 6] 과 같이 추천한다. 이는 논문, 특허 등 어떠한 정보를 검색 할 때, 해당 분야에 대해 정확히 모를 경우 검색에 도움을 주기 위해 연관키워드를 제공 하는 것과 유사한 역할이다. 이용자가 입력한 키워드와 추천 키워드를 이용해 검색식을 완성하면 NTIS 과제DB에서 관련 과제를 자동으로 추출한다. 이후의 논문, 특허, 정책 동향 추출 과정은 기존 서비스와 동일하다. 2019년 12월 말 서비스 오픈 이후 2020년 4월까지 약 100여 개의 주제가 해당 서비스를 통해 패키징되어 게시되었다.

IV. 결론 및 향후 연구

본 연구에서는 NTIS 의 이슈로 보는 R&D 서비스 분석을 통해 국가 R&D 정보를 이용한 데이터 큐레이션 서비스 구축의 관점에서 상세 기술되었으며 구축된 서비스가 이용자의 접근성 향상에 미치는 영향에 대해서도 기술하였다.

이슈로 보는 R&D 서비스는 이용자가 일상에서 쉽게 접하는 뉴스 기사에서 사회적 이슈와 현안을 추출하여 이와 관련 있는 국가 R&D과제, 논문, 성과, 정책 동향, 주요 연구자 및 연구기관 등 국가 R&D정보를 매핑 하여 관련 현황을 요약된 형태로 제공하여 이용자가 R&D 분야에 대해 접근하는 장벽을 낮추고, 관련 현황을 신속하게 파악할 수 있도록 하였다. 또한, 이용자의 서비스 활용성을 고려하여 패키징된 정보의 웹 서비스 뿐만 아니라 PDF 형태로 보고서를 다운 받거나 패키징에 포함된 과제, 성과 정보를 엑셀 형태로 다운 받을 수 있도록 구현함에 따라 사용자가 다양한 방법으로 활용할 수 있도록 하였으며 Special Issue 코너를 개설하여 홈페이지 상단에 게시함으로써 주요 현안이슈에 대한 이용자의 접근 편의성을 높일 수 있도록 하였다.

추가적으로 R&D 정보에 대한 이용자 친화적 서비스 제공을 위해 이용자가 직접 만드는 기능을 추가하여 이용자 관심 분야에 따라 관련 R&D 현황을 분석할 수 있

도록 구현 하였다.

본 서비스는 각 기관에서 수행하고 있는 국가 R&D 정보를 사회 주요 이슈에 맞춰 큐레이션 하여 제공함으로써 이 전에는 접근과 이용이 어려웠던 국가 R&D 정보를 사용자 친화적으로 제공하였는데 그 의의가 있다. 또한 이슈에 알맞은(fit) 데이터 큐레이션을 위한 이슈 선택, 과제·성과 추출 등의 전 과정에 있어 프로세스 과정을 자동화 함으로써 누구나 해당 서비스를 이용하여 자기만의 이슈 패키지를 구성할 수 있다는 것에서 서비스 개방성, 사용자 접근성, 활용성을 모두 포함하고 있다는 장점이 있다. 이처럼 국가 R&D 데이터를 사용자에게 보다 쉽게, 그리고 정확하게 이용할 수 있는 서비스를 구축함으로써 이를 통해 R&D 정보, 과학데이터 등 특수한 데이터에 대한 접근 편의성과 이용편의성을 높일 수 있는 프로세스와 그 효과를 제시하였다는 것에 의의가 있다.

하지만 현재까지 구축한 본 서비스는 다음과 같은 한계점이 있다. 첫째, 큐레이션 구성 과정이 자동화되었기 때문에 누구나 접근, 이용가능하다는 장점이 있지만 그 때문에 유사한 이슈일지라도 사용자에게 따라 검색에 이용하는 키워드가 다를 경우 패키징된 결과물이 다를 수 있다는 한계가 있다. 또한 이로 인해 해당 서비스를 통해 제공되고 있는 큐레이션 데이터의 정확성과 통계적 유의성 면에서 한계가 있다는 단점이 있다. 이는 서비스를 구축할 때 고려해야하는 목적이 있어서 서비스 접근성, 개방성과 데이터의 정확성, 폐쇄성이 상충하는 과정에서 유발되는 문제로서 향후 해당 서비스를 보다 정교화 하는 과정에서 보완해야 할 점이라고 생각되며 향후 이를 보완한 시스템 구축 방안에 대한 연구를 제안한다. 또한 본 연구의 경우 서비스의 구축과 접근 편의성 향상에 대한 분석이 이루어졌지만 실제 서비스를 이용한 사용자 분석은 이루어지지 못했다는 한계점이 있다. 이에 향후 추가 분석이 이루어진다면 서비스 접근, 사용 기록 외에 이용자의 심층인터뷰를 추가하여 보다 다각적인 분석을 시도해 보기를 제안한다.

- [1] P. J. Tichenor, G. A. Donohue, and C. N. Olien, "Mass media flow and differential growth in knowledge," *Public opinion quarterly*, Vol.34, No.2, pp.159-170, 1970.
- [2] B. M. Compaine, "Information gaps: myth or reality?," *The digital divide: Facing a crisis or creating a myth*, pp.105-118, 2001.
- [3] 김문조, 김종길, "정보격차 (Digital Divide) 의 이론적·정책적 재고," *한국사회학*, 제36권, 제4호, pp.123-155, 2002.
- [4] 국가과학기술지식정보서비스, <http://www.nts.go.kr>
- [5] 박우정, 노영희, "특화도서관의 큐레이션 서비스 개발에 관한 연구," *한국비블리아학회지*, 제30권, 제1호, pp.53-75, 2019.
- [6] 이정미, "교수학습활동 지원 개선을 위한 대학도서관의 데이터 큐레이션 연구," *한국문헌정보학회지*, 제54권, 제1호, pp.175-195, 2020.
- [7] 김진희, 최서연, 임철일, 함윤희, "연구지원을 위한 데이터 큐레이션 사서교육 프로그램 개발," *교육문화연구*, 제25권, 제6호, pp.757-779, 2019.
- [8] 최동훈, 박재원, 김병규, 신진섭, "커뮤니티 주도적 과학 데이터 큐레이션 협업 환경의 개발," *한국콘텐츠학회논문지*, 제17권, 제9호, pp.1-11, 2017.
- [9] 안영희, 박옥화, "디지털 큐레이션 정책을 위한 프레임워크 개발," *한국도서관 정보학회지*, 제41권, 제1호, pp.167-186, 2010.
- [10] 이해성, 권준희, "빅 데이터 시대의 개인화 데이터 큐레이션에 대한 연구," In *Proceedings of KIIT Conference*, pp.124-127, 2013.
- [11] 이유자, 유홍식, "EBS의 주제별 방송콘텐츠 큐레이션 서비스가 VOD이용량에 미치는 영향 및 성공적 큐레이션의 특징에 관한 연구," *정보사회와 미디어*, 제18권, 3호, pp.291-317, 2017.
- [12] 이유경, 정은경, "데이터 큐레이터의 핵심 직무 요건 고찰에 관한 연구," *한국비블리아학회지*, 제26권, 3호, pp.129-150, 2015.
- [13] 진보라, 정영미, "데이터 큐레이션 정책 개발을 위한 가이드라인에 관한 연구," *한국도서관정보학회 하계 학술발표회*, pp.91-103, 2016.
- [14] 이병희, 손강렬, "사용성 향상을 위한 국가 R&D 정보 내비게이션 서비스의 사용자 니즈 분석과 정보 시각화," *한국콘텐츠학회논문지*, 제13권, 제2호, pp.512-521, 2013.

참 고 문 헌

- [15] 김태현, 양명석, 최광남, “국가 R&D 정보활용을 위한 전문용어사전 구축,” 한국콘텐츠학회논문지, 제19권, 제10호, pp.217-225, 2019.
- [16] 박성욱, “국가 지식정보시스템 개발의 경제적 효과분석-한국과학기술정보원 (KISTI) 의 연구활동을 중심으로,” 정보관리연구, 제39권, 제3호, pp.73-94, 2008.
- [17] 남연화, 강주연, 김태영, 오효정, “NTIS 일반 이용자 그룹별 정보 요구 분석에 기반한 정보시각화 개선방안 연구,” 한국도서관·정보학회지, 제47권, 제1호, pp.361-382, 2016.
- [18] 유진환, 김승인, 김희연, 이소희, 최용수, 이병희, “NTIS-NDSL 연계 서비스'R&D Navigator'사용성 평가,” 디지털디자인학연구, 제13권, 제1호, pp.35-46, 2013.
- [19] 서상혁, 이선영, 이병희, “국가 R&D정보 이용자의 고객가치 및 고객만족도 영향요인 분석,” 기술혁신학회지, 제15권, 제4호, pp.837-861, 2012.
- [20] 황유나, 이승환, 최광남, “정부의 R&D 투자가 혁신 성과와 재정성가에 미치는 영향분석: 연구개발 단계를 중심으로,” 기술혁신학회지, 제22권, 제4호, pp.503-525, 2019.

최 광 남(Kwang-Nam Choi)

정회원



- 1994년 2월 : 충남대학교 컴퓨터공학과(공학석사)
- 2017년 2월 : 배재대학교 컴퓨터공학과(공학박사)
- 1994년 7월 ~ 현재 : 한국과학기술정보연구원 책임연구원 / NTIS 센터장

〈관심분야〉 : 정보검색, 정보분석, 빅데이터

황 유 나(You-Na Hwang)

정회원



- 2011년 8월 : 고려대학교 과학기술학(이학석사)
- 2017년 2월 : 고려대학교 과학기술학(과학기술학박사)
- 2017년 4월 ~ 2018년 4월 : 고려대학교 반도체기술연구소 연구교수
- 2018년 5월 ~ 2020년 4월 : 한국과학기술정보연구원 박사후 연구원

〈관심분야〉 : 국가 R&D 데이터 분석, 연구분야 분석, 네트워크 분석, 데이터 큐레이션

저 자 소 개

유 은 지(Eun-Ji Yu)

정회원



- 2014년 2월 : 국민대학교 비즈니스 IT전문대학원(경영정보학석사)
- 2014년 11월 ~ 현재 : 한국과학기술정보연구원 연구원

〈관심분야〉 : 데이터마이닝, 텍스트분석, 소셜네트워크분석, 딥러닝