J. Internet Comput. Serv.
ISSN 1598-0170 (Print) / ISSN 2287-1136 (Online)
http://www.jics.or.kr
Copyright ⓒ 2020 KSII

KISTI 과학기술정보서비스의 콘텐츠 활용 분석☆

A Study on the Content Utilization of KISTI Science and Technology Information Service

강 남 규[1] 황 미 녕[2*]
Nam-Gyu Kang Mi-Nyeong Hwang

요 약

한국과학기술정보연구원(KISTI)가 제공하는 과학기술정보서비스는 일반적인 정보서비스와 유사하게 구축된 콘텐츠를 이용자가 쉽고 편리하게 검색하고 조회할 수 있도록 만들어진 서비스이다. NDSL은 KISTI의 핵심 과학기술정보서비스로서 약 1억 38백만건의 콘텐츠를 제공하고 있으며 2019년 1년간 약 93백만회의 페이지 뷰를 보이고 있다. 본 논문에서는 NDSL이 제공하고 있는 학술논문, 보고서, 특허 등의 과학기술정보가 웹 서비스(https://www.ndsl.kr)를 통해서 어떻게 검색되고 활용되는지와 입력된 검색 질의어 등을 분석을 통해 다양한 인사이트를 도출하였다. 콘텐츠 구축 현황, 콘텐츠 유형별 활용 현황과 활용 방식 등의 일반적인 통계 이외에도 월별/요일별/시간대별 콘텐츠의 이용 행태, 조회수 구간별 활용 콘텐츠의 유형, 콘텐츠 유형별 1회 검색 대비 조회 비율, 학술논문의 발행년도별 이용 현황 비교, 국내 학술논문의 활용과 KCI 영향력 지수와의 관계, 학술논문과 특허 활용에 대한 특징 등 콘텐츠 유형별 활용성에 대한 분석을 진행하였다. 또한 콘텐츠 유형별 검색 질의어의 구성 방식, 질의어의 단어수 분석, 검색 질의어와 시기성의 관계 등의 검색 질의어에 대하여 분석하였으며 NDSL에서는 한글 성명 검색이 많다는 특징도 발견할 수 있었다. 이러한 분석 결과를 토대로 정보시스템 구축 관점에서의 NDSL 서비스 개선 방안을 제안하고자 한다. 본 논문에서는 콘텐츠 활용에 대한 행태 분석 결과를 이용자의 검색 결과에 동적으로 반영, 비로그인 이용자에 대한 식별을 통해 타겟팅 서비스가 가능하도록 시스템을 구축, 질의어 확장하는 등의 것을 NDSL 개선 사항으로 제안하고자 한다.

☞ 주제어 : 과학기술정보, 과학기술정보서비스, 콘텐츠 활용 분석, 질의어 확장, 이용자 프로파일 정보

ABSTRACT

The Science and Technology Information Service provided by the Korea Institute of Science and Technology Information (KISTI) is a service designed to allow users to easily and conveniently search and view content that is built similar to the general information service. NDSL is KISTI's core science, technology and information service, providing about 138 million content and having about 93 million page views in a year of 2019. In this paper, various insights were derived through the analysis of how science and technology information such as academic papers, reports and patents provided by NDSL is searched and utilized through web services (https://www.ndsl.kr) and search query words. In addition to general statistics such as the status of content construction, utilization status and utilization methods by type of content, monthly/weekly/time-of-day content usage, content view rate per one-time search by content type, the comparison of the use status of academic papers by year, the relationship between the utilization of domestic academic papers and the KCI index we analyzed the usability of each content type, such as academic papers and patents. We analyzed query words such as the language form of query words, the number of words of query words, and the relationship between query words and timeliness by content type. Based on the results of these analyses, we would like to propose ways to improve the service. We suggest that NDSL improvements include ways to dynamically reflect the results of content utilization behavior in the search results rankings, to extend query and to establish profile information through non-login user identification for targeted services.

☞ keyword : science and technology information, S&T information service, content utilization analysis, extended query, user profile information

1. 서 론

한국과학기술정보연구원(KISTI)에서 제공하고 있는 NDSL(National Digital Science Library)은 산·학·연 연구자

1 Convergence Service Center, Korea Institute of Science and Technology Information, Daejeon, 34141, Korea
2 Div. of National S&T Data, Korea Institute of Science and Technology Information, Daejeon, 34141, Korea
* Corresponding author : (mnhwang@kisti.re.kr
[Received 30 March 2020, Reviewed 14 April 2020(R2 23 June 2020), Accepted 3 July 2020]

☆ This research was supported by KISTI Program (Development of S&T Knowledge Infrastructure Convergence Service / No. K-20-L01-C07-S01)

http://dx.doi.org/10.7472/jksii.2020.21.4.87

들에게 고품질의 정보를 제공하여 국가과학기술혁신에 기여하고자 개발·운영된 과학기술정보 서비스 플랫폼으로 논문, 특허, 보고서, 동향, 연구자, 연구기관 등 약 1억 38백만건 이상의 콘텐츠에 대한 검색과 콘텐츠 유형별 전문검색서비스를 제공한다[1, 2]. 연구자는 KISTI가 제공하는 웹(www.ndsl.kr)과 모바일(m.ndsl.kr) 서비스를 통하거나 NDLS 오픈 서비스인 NOS를 이용한 외부서비스를 통해 NDSL 콘텐츠를 이용할 수 있으며, 이러한 방식들로 2019년 1년 동안 NDSL에서는 약 93백만회의 페이지 뷰를 나타내었다[3].

본 논문에서는 NDSL을 이용하는 3가지 방식 중 웹(http://www.ndsl.kr) 서비스를 대상으로 2019년 1년간의 이용 현황을 분석하고자 한다. 웹 서비스는 이용자가 뚜렷한 목적과 의지를 갖고 NDSL에 직접 접속하여 활용하는 방식이기 때문에 정확한 콘텐츠의 이용 현황을 조사할 수 있을 것으로 판단하였다[4, 5, 10]. 검토 대상으로는 페이지 뷰 현황, 검색과 조회 현황, 학술논문, 보고서, 특허 등의 콘텐츠 유형별 이용 현황, 검색 유형, 검색 질의어 등이며, 분석을 통해 다양한 인사이트를 도출하고 정보시스템 개발 관점에서 향후 NDSL 웹 서비스가 개선해야 할 내용을 제안하고자 한다[6]. 도출된 개선 사항으로는 콘텐츠 활용 행태 분석 결과를 반영하여 검색 결과의 출력 순서를 동적으로 구성하는 방식과 비로그인 이용자에 대한 프로파일 구축을 통해 이용자 타겟팅 서비스[9]와 시기성이 반영된 연관어 제시 등을 제안한다.

2. NDSL 페이지 뷰 현황

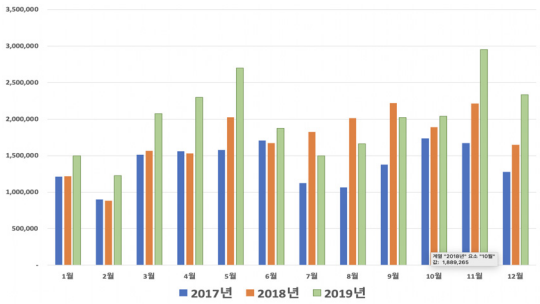
이용자는 NDSL이 제공하는 웹, 모바일, NOS 서비스를 통해 약 1억 38백만건의 콘텐츠를 검색하고 조회할 수 있다. 표 1은 2019년도 NDSL 페이지 뷰를 서비스 유형별로 구분한 것으로 웹 서비스는 NDSL 전체 이용의 25.8%를 차지하고 있으며, NOS의 이용이 73.2%로 가장 많은 페이지 뷰를 보이고 있다.

(표 1) 2019년 NDSL 유형별 페이지 뷰 (단위 : 천건)
(Table 1) 2019 NDSL page view by type (unit 1,000)

유형	웹	모바일	NOS	합계
건수	24,189	899	68,624	93,713

NDSL 이용의 약 73.2%를 차지하고 있는 NOS는 대학교, 연구소, 공공기관 등 350여개의 기관과 서비스에서 활용하고 있으며, NOS의 이용은 오픈API를 통해 NDSL의 서비스를 직접 활용하거나, 데이터의 유통 표준 프로토콜을 사용하여 NDSL 데이터를 직접 내려받아 활용할 수 있다. NOS가 NDSL 이용의 약 3/4를 차지하고 있으나, NOS를 활용하고 있는 기관이나 서비스의 성격에 따라 NDSL 콘텐츠를 이용하는 목적이 달라질 수 있으므로 본 논문에서의 분석 범위에는 포함하지 않았다.

NDSL 주요 기능인 과학기술정보의 검색은 로그인을 이용하지 않아도 누구나 사용할 수 있다. 2019년 로그인과 비로그인 상태에서의 페이지 뷰를 비교하면 로그인 2%, 비로그인 98%로 대부분의 NDSL 서비스 이용이 비로그인 상태로 이루어짐을 알 수 있다.



(그림 1) 2017~2019년 NDSL 페이지 뷰 월별 현황
(Figure 1) 2017~2019 NDSL page view monthly status

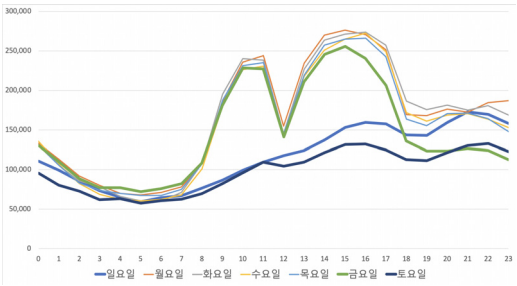
2019년 월 평균 페이지 뷰는 2,015천건으로 2017년 대비 약 621천건이 증가하였다. 월별 페이지 뷰를 살펴보면 그림 1과 같이 1~2월과 7~8월에 적은 이용을 보이며 11~12월에 많은 이용을 보인다. 이것은 2018년과 2017년에도 비슷한 양상을 보이나, 2018년 7월부터 9월까지 일시적으로 페이지 뷰가 늘어난 것에 대해서는 다른 요인이 있는지 세밀한 분석이 필요할 것이다.

표 2처럼 요일별 페이지 뷰를 살펴보면 주중과 주말이 확연하게 구분됨을 알 수 있었으며, 주중에서는 월요일과 화요일이 주말에서는 토요일보다 일요일이 더 많은 페이지 뷰가 관찰되었다.

(표 2) 2019년도 요일별 페이지 뷰 (단위 : 천건)
(Table 2) page view by day of 2019 (unit 1,000)

일	월	화	수	목	금	토
2,793	3,996	3,947	3,774	3,802	3,506	2,368

그림 2는 일요일부터 토요일까지의 시간대별 페이지 뷰를 그래프로 나타낸 것으로서, 요일에 따른 시간대별 페이지 뷰 현황을 살펴보면 주중의 이용 패턴과 주말의 이용 패턴이 상이함을 알 수 있다. 월요일부터 금요일까지 페이지 뷰의 패턴은 매우 유사하며, 10~11시와 14~16시에 가장 많은 페이지 뷰를 보인다. 그러나 금요일은 페이지 뷰가 감소하는 시점이 다른 요일들보다 조금 더 빠르다는 것을 알 수 있는데, 월요일 등의 요일들에서는 16~17시 사이에 페이지 뷰가 감소하기 시작하지만, 금요일은 15시를 기점으로 페이지 뷰가 감소함을 알 수 있다. 또한, 토요일과 일요일에서는 18~19시를 제외하면 10시부터 22시까지 꾸준히 증가하는 것이 매우 특징적이다.

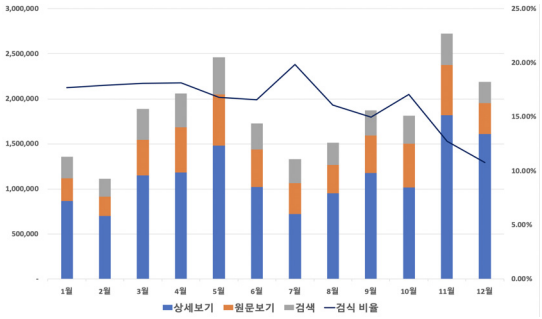


(그림 2) 요일별 페이지 뷰 현황
(Figure 2) page view status by day

3. NDSL 검색과 콘텐츠 조회 현황

NDSL에서는 통합검색, 빠른검색, 상세검색, 고급검색, 결과내 재검색 등을 다양한 검색 기능을 제공하고 있으며, 콘텐츠의 조회 기능으로 상세보기와 원문보기를 제공하고 있다. 그림 3은 2019년도 월별 검색과 콘텐츠 조회에 대한 횟수를 그래프로 나타낸 것이며, 1년간 검색과 콘텐츠 조회수는 각각 3,533천건과 18,512천건으로 검색 비율은 약 16.39%, 1회 검색을 통해 약 5.1회의 콘텐츠 조회했음을 알 수 있다. 그러나 11월과 12월의 검색 비율은, 다른 월들과 비교해볼 때, 각각 12.73%와 10.74%로서 1회 검색을 통해 많은 수의 콘텐츠를 조회했다는 것을 알 수 있다.

NDSL 콘텐츠 조회의 경우, 상세보기 13,709천건, 원문보기 4,802천건으로 상세보기가 원문보기보다 약 3배 더 많았다. 원문보기인 경우 NDSL에서 직접 제공하는 PDF 또는 XML 파일의 직접 다운로드가 77%를 차지했으며 나머지 23%의 경우는 외국 출판사 등으로의 외부 연결을 통한 원문 조회를 하였다.



(그림 3) 2019년 NDSL 검색과 콘텐츠 조회 현황
(Figure 3) 2019 NDSL search & view status

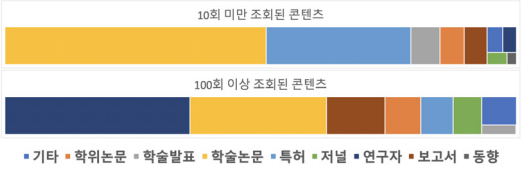
표 3은 NDSL이 제공하는 콘텐츠에 대하여, 콘텐츠 유형별로 2019년도에 어떤 콘텐츠를 많이 그리고 빈번히 조회했는지를 알아보기 위한 표이다. 2019년도에 1회 이상 조회된 콘텐츠는 5,110천개이며, 이 콘텐츠들이 총 18,512천회 웹을 통해 조회되었다. 그 중 많이 조회된 콘텐츠로는 학술논문, 보고서와 특허임을 알 수 있고, 동향정보와 보고서는 조회된 콘텐츠 1건당 약 19.23회와 12.41회의 평균 조회수를 나타내고 있어 빈번하게 조회됨을 알 수 있었다. 또한, 콘텐츠 구축 대비 활용은 보고서와 동향정보가 가장 높다는 것을 알 수 있다.

(표 3) 콘텐츠 유형별 조회 현황 (단위 : 천건)
(Table 3) content view by content type (unit 1,000)

유형	전체 콘텐츠 구축 총량	2019년 조회 콘텐츠 수	총 조회수	1건당 조회 빈도
학술논문	85,068	2,558	7,217	2.82
보고서	330	306	3,799	12.41
특허	38,870	1,392	3,292	2.36
학위논문	1,109	243	1,097	4.51
동향	54	47	906	19.23
연구자	782	109	668	6.11
학술발표	11,622	282	625	2.22
저널	390	66	418	6.33
기타	250	106	490	4.61
합계	138,475	5,111	18,512	

2019년도에 1회 이상 조회된 콘텐츠 5,110천개에 대한 이용 현황을 분석하였는데, 1회만 조회한 콘텐트는 약 2,938천개이며, 총 10회 이하로 조회된 콘텐트는 총 4,815천개로 조회된 전체 콘텐츠의 94.23%를 차지하고 있다. 총 100회 이상 열람된 콘텐츠는 약 9.5천개로 0.18%로 확

인되었다. 그림 4는 10회 이하와 100회 이상 열람된 콘텐츠를 대상으로 콘텐츠 유형을 분석한 결과이다. 100회 이상 자주 열람되는 콘텐츠는 보고서와 학술논문, 그리고 동향정보로 확인되었다.



(그림 4) 10회 미만, 100회 이상 조회된 콘텐츠 현황 비율
(Figure 4) percentage of content view 〈 10 times, and ≥ 100 times

4. NDSL 콘텐츠 유형별 조회 현황

4.1 학술논문 활용 현황

NDSL에서 가장 많은 건수를 제공하고 있는 학술논문은 중국, 일본, 해외, 국내, KCI 등의 데이터 입수처로 구분되며, 표 4는 2019년도에서의 학술논문 유형별 조회 현황과 2019년도에 조회된 학술논문 중 최근(출판년도가 2010년 이후와 2015년 이후)에 출판된 학술논문의 조회 현황을 나타내고 있다.

학술논문 중 국내 학술논문과 해외 학술논문의 이용이 전체 학술논문 조회수의 92%를 차지하고 있으며, KCI와 국내 학술논문의 경우 최근에 10년 내에 출판된 학술정보가 다른 학술논문들보다 많이 이용된다는 것을 알 수 있다.

(표 4) 학술논문 입수처 유형별 조회 현황 (단위 : 천건)
(Table 4) paper view by academic paper source type (unit 1,000)

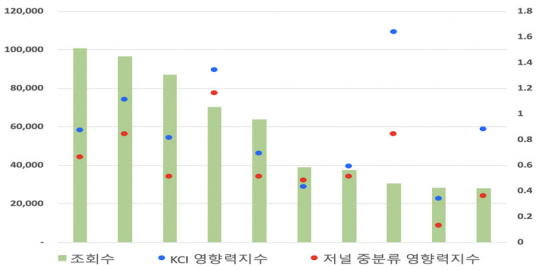
유형	2019년도 조회 현황		2015~2019년 발행된 논문		2010~2019년 발행된 논문	
	조회 논문수	총 조회수	조회 논문 비율	총 조회 비율	조회 논문 비율	총 조회 비율
중국	60	80	17.13	22.73	38.64	42.85
해외	1,686	2,490	15.18	20.86	36.88	43.52
일본	48	67	21.36	29.17	29.33	35.00
국내	610	4,155	19.58	32.32	43.64	59.22
KCI	154	425	53.40	50.56	78.22	80.15
합계	2,558	7,217				

학술논문 중 저널정보가 없는 59천여개의 학술논문을 제외하고, 학술논문의 입수처 유형별로 많이 활용된 저널의 현황을 확인하였다. 학술논문에 포함된 ISSN번호를 이용하여 저널을 구분하였으며, 표 5는 2019년 1년간 조회된 학술논문의 저널 수를 입수처 유형별로 표기한 것으로 100회 이상 조회된 학술논문의 저널과 그에 해당하는 학술논문의 비율도 함께 표현하였다. 중국, 해외, 일본에서 입수된 학술논문의 경우 100회 이상 조회된 학술논문의 저널이 전체 조회된 학술논문의 저널과 비교할 때 약 10% 정도임을 알 수 있다. 반면에 국내와 KCI에서 입수된 학술논문의 저널 50%에서 대부분의 학술논문 조회수가 발생함을 알 수 있다.

(표 5) 학술논문 입수처 유형별 저널 현황 (단위 : 건)
(Table 5) journal view by academic paper source type (unit 1,000)

유형	조회된 전체 논문의 저널수	100회 이상 조회된 논문의 저널수(비율)-①	①에서의 논문 조회 비율
중국	1,838	211 (11.48%)	48.34%
해외	38,244	4,955 (12.96%)	71.82%
일본	1,664	158 (9.50%)	57.92%
국내	2,339	1,167 (49.89%)	99.13%
KCI	2,582	1,185 (45.89%)	85.18%

국내 학술논문 중 가장 많이 조회된 저널 10개에 대하여 KCI 등재여부와 KCI 영향력지수를 조사하였는데, 영향력지수는 한국연구재단의 기준년도 2018년 KCI IF(2년) 값을 사용하였다. TOP-10 저널 모두 KCI 등재지로 확인되었으며 KCI 영향력지수 평균은 0.87, 최대값 1.64, 최소값 0.34를 나타내고 있었다.

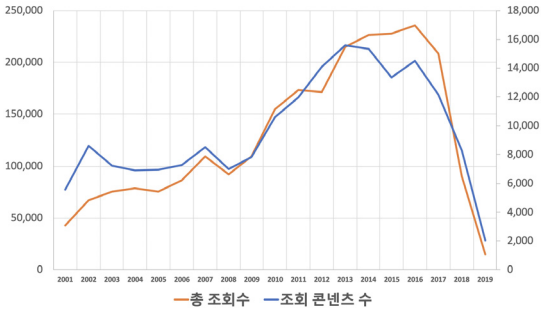


(그림 5) 국내 학술논문 중 많이 조회된 저널 TOP-10의 조회수 및 KCI 영향력 지수
(Figure 5) paper view & KCI Impact factor in domestic academic paper

그림 5와 같이 조사된 결과로부터 국내 학술논문의 이용과 KCI 영향력지수의 관계성을 이끌어내기는 어려울 것으로 예측하였으나, 저널의 영향력지수가 그 저널이 속한 분야(중분류)에서의 KCI 영향력지수보다는 상회하고 있어 TOP-10 저널이 해당 분야 내에서 많이 인용되고 있음을 간접적으로 알 수 있었다.

4.2 보고서 활용 현황

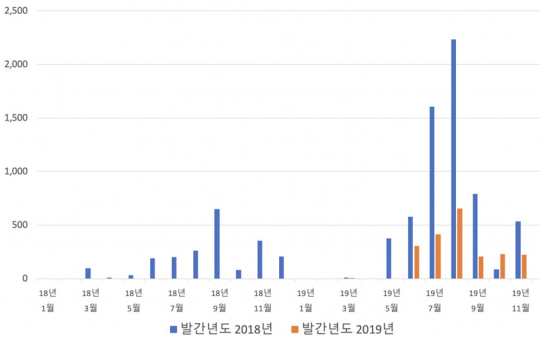
학술논문 다음으로 NDSL에서 많이 이용되는 콘텐츠는 보고서이다. 보고서 중에서도 연구보고서가 많이 이용되고 있으며, 그림 6은 2019년 1년간 조회된 연구보고서를 그 연구보고서의 발간 년도 별로 나타낸 것으로, 특징적인 것은 가장 최신의 정보인 2018년도 또는 2019년도의 보고서가 아닌 2016년도에 발간된 연구보고서가 가장 많이 이용되었다는 것이다.



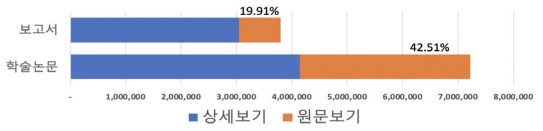
(그림 6) 2019년도 연구보고서 활용 현황
(Figure 6) 2019 research report view

이러한 상황은 연구보고서가 연구 종료 시점에 발행되고 이것을 연구 수행기관으로부터 수집한 후 DB 구축 과정을 거쳐 서비스하는 등의 시차가 반영된 것으로 판단한다. 그림 7에서와 같이 2018년과 2019년도에 발간된 연구보고서가 NDSL에 서비스되는 시점을 그래프로 표현한 것이다. 이 그래프에서 볼 수 있듯이 2018년에 발간된 보고서가 2019년 하반기에 주로 서비스가 된다는 것을 알 수 있다.

3장의 표 3에서와 같이, NDSL에서 보고서는 학술논문에 이어 가장 많은 이용된 콘텐츠이며, 1건당 조회빈도도 12.41회로 다른 콘텐츠에 비하여 1건당 조회빈도가 매우 높은 콘텐츠이다. 그림 8은 보고서와 학술논문과 비교하여 상세보기와 원문보기의 비율을 나타낸 것이다. NDSL에서의 상세보기와 원문보기 비율이 1:3인 것과 비교할 때 보고서와 학술논문의 원문보기 비율이 많이 다름을 알 수 있다.



(그림 7) 발간년도 2018~2019년 보고서의 서비스 시점
(Figure 7) service time for 2018 ~ 2019 research report



(그림 8) 보고서와 학술논문의 상세보기와 원문보기 비교
(Figure 8) metadata view vs pdf view of academic paper and research report

4.3 특허 활용 현황

NDSL에서 제공하고 있는 특허는 한국, 미국, 유럽, 일본 특허이며 표 6과 같이 미국 특허가 가장 많이 활용되었으며, 한국특허는 1건당 조회빈도가 가장 많은 것으로 확인되었다. 그리고 미국과 한국 특허가 전체 특허 조회수의 85.01%의 비중을 차지하고 있다.

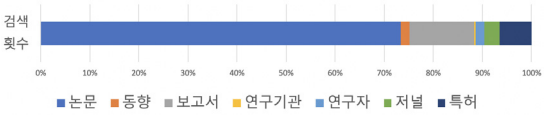
조회수 기준으로 미국 특허의 87.53%가 등록 특허이며, 나머지는 공개 특허가 조회되었다. 한국 특허에서는 52.41%가 등록특허, 38.78%가 공개특허, 나머지는 실용신안이 조회되었다.

(표 6) 특허 유형별 조회 현황 (단위 : 천건)
(Table 6) content view by patent type (unit 1,000)

유형	2019년 조회 콘텐츠 수 (비율)	총 조회수 (비율)
한국	250 (17.96%)	1,208 (36.70%)
미국	823 (59.09%)	1,591 (48.32%)
유럽	170 (12.20%)	272 (8.24%)
일본	149 (10.75%)	221 (6.27%)

5. NDSL 검색 질의어 분석

1년간 NDSL 웹 서비스에서의 검색은 약 3,533천건 실행되었으며, 검색 대상별 검색 실행은 그림 9와 같다. 그림 9의 논문에는 학술논문과 학위논문, 학술발표가 함께 검색된 결과이다. 논문 검색이 전체 검색의 73.35%를 차지하고 있으며 그 뒤를 보고서가 13.22%를 나타내고 있다. 논문 검색이 많은 이유는 NDSL 웹 서비스에서 통합검색을 실행하면 논문을 대상으로 우선 검색하기 때문으로 판단한다.



(그림 9) 검색 대상별 검색 현황
(Figure 9) status by search target

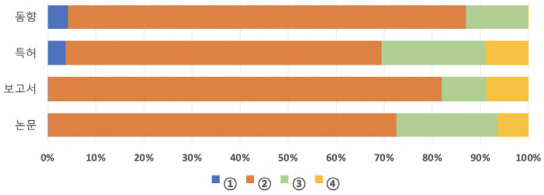
검색 유형으로는 통합검색, 상세검색, 고급검색 등으로 나눌 수 있는데, 통합검색 비율이 97.82%로 거의 대부분의 검색이 통합검색으로 이루어짐을 알 수 있다. 그러나 검색 유형을 검색 대상별로 구분할 경우, 특허 검색에서는 상세검색의 비율이 6.14%에 달하고 있어 다른 유형과는 다른 양상을 보이고 있다.

검색 질의어의 언어 형태로는 한글, 영어, 특수문자, 한자, 일본어 등이 사용되었으며, 이것을 형태별로 정리하면 표 7과 같다. 전체 검색의 71.71%가 한글(숫자 포함)로만 검색되고 있으며 20.29%가 영어(숫자 포함)로만 검색되었다. 한글과 영어가 혼합된 형태는 약 6.64%에 그치고 있다.

(표 7) 검색 질의어 유형 (단위 : 천건)
(Table 7) type of search query

질의어 형태	질의어 예시	건수
한글·영어 제외한 문자 또는 숫자로만 구성-①	魯桑	48
한글과 숫자로만 구성-②	4대강 오염	2,534
영어와 숫자로만 구성-③	NO RAW 264.7	717
한글·영문 혼합-④	드론 ESC	234

그림 10은 표 7의 ①부터 ④까지의 검색 질의어 유형을 검색이 많이 실행된 논문, 보고서, 특허, 동향 대상으로 세분화한 결과를 보여주고 있다.



(그림 10) 검색 대상별 검색 질의어 유형
(Figure 10) search query type by search target

위 그래프로부터 NDSL의 검색은 한글 중심의 검색이 이루어지고 있음을 알 수 있으며, 각 콘텐츠별로 사용하고 있는 언어는 조금씩 차이를 보인다. 특허와 논문에서는 다른 콘텐츠보다 영어로 된 검색 질의어가 많이 입력되고 있음을 알 수 있는데, 전체 학술논문의 4.5%, 특허의 14.3%가 한글 콘텐츠임을 고려할 때 한글 검색질의어의 입력이 많은 부분을 차지하고 있는 것이 특징적이다.

검색 질의어의 공백을 기준으로 단어수를 조사하였다. 한글과 영문 등의 구분 없이 1개의 단어로 이루어진 검색은 전체의 44.81%이며 2개 단어로의 검색은 32.15%로 확인되었고 5개 단어 이하의 검색은 전체 검색의 91.63%로 조사되었다. 질의어의 언어까지 고려할 경우, 1개 단어로 이루어진 검색에서의 한글 질의어는 81.93%, 영어 질의어는 14.07%를 나타냈다.

공백이 포함된 단어의 처리를 위하여 검색 질의어에서 공백을 제거한 후 분석을 진행하였는데, 가장 많이 검색된 질의어로는 미세먼지, 인공지능, 블록체인, 드론, 미세플라스틱 등으로 최근 이슈가 되거나 활발하게 연구가 이루어지는 분야의 키워드들임을 확인할 수 있었다. 또한 1개 단어 질의어 중 성명으로 추정되는 3음절의 단어가 많이 발견되었는데, 이 단어를 NDSL이 제공하고 있는 연구자 정보와 매핑했을 때 약 201천개의 질의어가 매핑되었다. 즉 1개 단어로 된 질의어 중 성명으로 검색한 것이 12.72%임을 알 수 있다.

6. NDSL 정보서비스 개선 방향

5장까지 설명한 것과 같이, NDSL 정보서비스에서는 두드러진 이용 패턴을 보인다. NDSL 구축 콘텐츠 총량에 대비하여 이용자가 조회하고 있는 콘텐츠의 양은 매우 적으며, 지난 1년간 웹 서비스를 통해 조회된 콘텐츠는 전체 구축 콘텐츠의 3.7%에 지나지 않고 있다. 또한, 비로그인 이용자가 서비스 이용의 대부분을 차지한다는 것과

논문, 보고서, 특허 등의 콘텐츠가 많이 이용되고 있으며 콘텐츠 유형별로 활용에 대한 특징이 뚜렷하다는 것이다. 그리고 검색 질의어의 패턴도 매우 특징적인데, 검색 질의어로 한글 단어가 주로 많이 사용되고 있으며 사용된 단어의 수도 매우 적은 것을 알 수 있다. 따라서 NDSL 정보서비스가 개선되어야 할 방향은 이용자의 콘텐츠 활용과 검색 행태를 잘 반영해야 할 것이다.

NDSL에서는 조회가 되는 콘텐츠와 그렇지않은 콘텐츠를 구분할 필요가 있을 것이다. 조회가 되지 않았다는 것은 검색결과로 노출되지 않았다는 것을 의미할 수 있다. 이것은 이용자의 검색 질의어에서 사용되는 언어와 상당한 관련이 있을 수 있는데, 해외 학술논문의 양이 절대적으로 많은 NDSL에서 한글 검색이 주로 이루어지고 있기 때문이다. NDSL에서는 조회수가 적은 해외 학술논문의 검색결과 노출을 확대할 것인지 아니면 검색 속도 향상을 기대할 것인지에 대한 의사 결정이 필요하다. 한글로 입력된 검색 질의어에 대해서도 해외 학술논문 검색 결과를 보여주기 위해서는 질의어를 영어로 변환하는 등의 전처리 작업과 대역어 사전이 필요할 것이다. 다른 대안으로서, 조회가 빈번한 콘텐츠와 조회가 빈번하지 않은 콘텐츠를 분리하여 색인하고 검색 결과를 선택적으로 제공함하여 NDSL에서의 빠른 검색 속도를 이용자에게 제공하는 것이다.

정해진 순서에 따라 검색된 결과를 제공하는 랭킹도 NDSL 콘텐츠 활용 행태에 따라서 동적으로 제공할 필요가 있을 것이다. 일반적인 정보검색 서비스가 높은 유사도에 최신의 정보 중에서도 많이 이용된 콘텐츠 중심으로 검색 결과를 제공한다면, NDSL에서는 콘텐츠 유형이나 검색 질의어에 따라서 출력의 순서를 변경할 수도 있어야 할 것이다. 동일한 학술논문이라도 입수처가 국내와 KCI인 것을 우선하여 출력해주거나 KCI 중에서도 저널 중분류 영향력 지수가 높은 논문을 먼저 제공하는 것도 방법이 될 수 있다. 입력된 검색 질의어에 따라서도 출력 순서가 변할 수도 있는데, 예를 들어, 질의어가 성명으로 판단된다면 연구자 정보를 제일 먼저 출력해주는 것이 바람직할 것이다.

NDSL을 이용하는 98%의 비로그인 이용자에 대한 별도의 프로세스가 필요할 것이다. 현재의 NDSL에서는 로그인해야만 그 이용자의 식별이 가능하며, 식별된 경우에는 이용자의 프로파일 정보를 이용하여 '코로나'로 검색을 했을 때 지구과학, 천문우주 분야의 정보를 제공하거나 생명과학, 바이오 분야의 정보를 선별적으로 제공하는 등의 타겟팅 서비스가 가능하다. 비로그인 상태에서도 로그에 저장되는 HTTP 헤더정보를 이용하면 동일한 이용자로 추정되는 로그들의 집합을 묶을 수 있고, 이것들을 대상으로 질의어의 입력 행태와 콘텐츠 조회 현황 등의 서비스 이용 행태를 파악한다면 비로그인 이용자에 대한 프로파일 정보를 만들 수 있다. 이러한 방식으로 만들어진 프로파일 정보는 비로그인 이용자에 대한 타겟팅 서비스를 제공할 때 활용할 수 있다.

NDSL에서는 2개 단어 이하로 입력되는 경우가 전체 검색의 76.96%에 달하고 있다는 점에서 검색 질의어에 대한 확장을 고려해야 할 것이다. 또한 질의어로 사용되는 단어에는 시기성이 포함되어 있다. 2019년 11월의 질의어 '코로나'와 2020년 2월의 '코로나'는 다른 의도를 갖고 이용자가 입력한 검색어일 것이다. 정보서비스가 이용자의 숨겨진 의도를 파악하여 검색 결과를 제공한다는 것은 어려울 것이다. 그러나 이용자의 선택을 통해 숨겨진 의도를 파악하는 것은 비교적 쉬울 수 있다. 즉, 적은 수의 단어가 검색 질의어로 입력이 되었을 경우 시기성이 반영된 연관어들이 자동으로 제공되고, 이용자가 그것을 선택함으로써 검색 결과의 만족도와 신뢰도를 높여주는 것이다. 이를 위해서는 엑소브레인(exobrain)의 토픽 트랜드 분석 API 등과 같은 시기성이 반영된 연관어 정보를 제공하는 기능 등의 연계가 필요할 것이다[7, 8].

7. 결　론

NDSL은 KISTI가 제공하는 핵심 과학기술정보서비스이다. 매일 수많은 콘텐츠가 구축되고 있으며, 불특정 다수의 과학기술 종사자에 의해 정보검색과 콘텐츠 조회가 이루어지고 있다. 또한, 현재의 NDSL 서비스 정책이 변하지 않는 한 비로그인 상태에서도 NDSL의 주요 기능들을 사용할 수가 있을 것이다.

본 논문은 2019년 1년간의 NDSL 웹 서비스 로그를 분석하여 NDSL이 제공하는 콘텐츠의 활용과 검색의 유형 등을 살펴보았다. NDSL 웹 서비스의 98%는 비로그인 이용자가 사용하고 있으며, 학술논문, 보고서, 특허가 다른 콘텐츠에 비교하여 많이 이용되고 있다. 또한 콘텐트의 상세보기가 원문보기보다 3배 많았으며, 1년간 10회 이하로 조회된 콘텐츠가 1년 전체 조회된 콘텐츠의 94.23%가 됨을 알 수 있다. 학술논문에서는 국내 학술정보가 가장 많이 이용되고 있으며, 이용이 빈번한 국내 학술정보의 저널은 동일 분야의 저널보다도 영향력지수가 높다는 것을 알 수 있었다. 연구보고서의 이용 활성을 위해서는 보

고서 발간과 동시에 서비스가 될 수 있는 체계가 필요하다는 것도 알 수 있었다. NDSL 검색의 97.82%는 통합검색으로 수행되고 있으며 전체 검색 질의어의 71.71%가 한글로 검색한 것이었다. 또한 2개 단어 이하의 검색 질의어가 전체 검색의 76.95%을 차지하고 있었다.

위와 같은 특징들을 NDSL 웹 서비스 개선에 반영하기 위하여 본 논문에서는 몇 가지 방안을 제안하였다. 콘텐츠의 활용성을 분석하여 조회되는 콘텐츠와 그렇지 않은 콘텐츠를 구분하여 콘텐츠의 활용을 증대하는 방안을 적용하거나 검색 결과를 선택적으로 제공하는 방식을 도입할 필요가 있을 것이다. 또한, 검색 질의어의 유형이나 콘텐츠 활용 행태에 따라서 검색 결과를 동적으로 구성하여 제공하는 방식도 적용할 필요가 있다. 그리고 비로그인 이용자를 위한 타켓팅 서비스를 위해 로그 정보를 활용한 프로파일 구축도 필요할 것이다. 마지막으로 이용자의 의도를 정확히 파악하기 위해 검색 질의어의 확장도 고려해야 할 것이다.

2019년 1년간 NDSL 웹 서비스를 분석하는 것이 NDSL 전체를 대표할 수는 없을 것이다. 그러나 일부일지라도 분석을 통해 시사점을 도출하고 이를 서비스 개선에 반영한다면 전체 NDSL 서비스가 좋아질 것으로 예상한다. 향후 연구과제로는 NDSL 개선을 통한 효용성 분석과 NDSL 로그의 73%를 차지하는 NOS의 이용 현황을 분석하고 NOS와 웹 서비스의 상관관계도 함께 분석하고자 한다.

참고문헌(Reference)

[1] SK Kim, WJ Kim, TS Lee, SY Bae, “An Improvement study in Keyword-centralized academic information service – Based on Recommendation and Classification in NDSL -”, Journal of Korean Library and Information Science Society, Vol. 49, No. 4, pp. 265-294, 2018. https://doi.org/10.16981/kliss.49.4.201812.265

[2] NDSL, http://ndsl.kr

[3] MH Hun, HJ Lee, HS Kim, “Effect of NDSL Open Service(NOS) on Sharing S&T Information”, ICCC 2014, pp. 297-298, 2014. https://www.koreascience.or.kr/article/CFKO201431749164056.j

[4] SK Kim, “A Study on the Information Needs and Using Behavior of Science-Technology Information Users”, Proceedings of the Korean Society of Computer Information Conference, Vol. 24, No. 2, 2016, pp. 65-67, 2016. https://www.koreascience.or.kr/article/CFKO201623070249493.j

[5] TY Kim, JY Baek, HJ Oh, “An Analysis of Library User and Circulation Status based on Bigdata Logs A Case Study of National Library of Korea, Sejong”, Journal of Korean Library and Information Science Society, Vol. 49, No. 2, pp. 357-388, 2018. https://doi.org/10.16981/kliss.49.201806.357

[6] NG Kang, MH Cho, OS Kwon, “A Relation Analysis between NDSL User Queries and Technical Terms”, Journal of Information management, Vol. 39, No. 3, pp. 163-177, 2008 https://doi.org/10.1633/JIM.2008.39.3.163

[7] Exbrain, http://exobrain.kr

[8] Adams.ai, http://adams.ai

[9] KS Lee, JW Yoon, “Rapid Hybrid Recommender System with Web Log for Outbound Leisure Products”, KIISE transactions on computing practices, Vol. 22, No. 12, pp. 646-653, 2016. https://doi.org/10.5626/KTCP.2016.22.12.646

[10] J Srivastava, R Cooley, M Deshpande, PN Tan, “Web usage mining: discovery and applications of usage patterns from Web data”, ACM SIGKDD Explorations Newsletter, Vol. 1, pp. 12-23, 2000. https://doi.org/10.1145/846183.846188

◐ 저 자 소 개 ◑

강 남 규(Nam-Gyu Kang)
1997년 충남대학교 컴퓨터공학과(공학사)
1999년 충남대학교 대학원 컴퓨터공학과(공학석사)
2019년 충남대학교 대학원 컴퓨터공학과(공학박사)
1999년~현재 한국과학기술정보연구원 책임연구원
관심분야 : 정보검색, 온톨로지, 데이터베이스, 블록체인 etc.
E-mail : ngkang@kisti.re.kr

황 미 녕(Mi-Nyeong Hwang)
2000년 부산대학교 전자계산학과(이학사)
2002년 부산대학교 대학원 전자계산학과(이학석사)
2002년~현재 한국과학기술정보연구원 선임연구원
관심분야 : 인공지능, 데이터 모델링, 시맨틱 웹, 빅데이터 etc.
E-mail : mnhwang@kisti.re.kr