

국내 Sci-Hub 이용 현황 분석

An Analysis of Usage of Sci-Hub in Korea

서 지 영 (Jeeyoung Suh)*

목 차

- | | |
|----------------------|---------------------|
| 1. 서 론 | 4. 국내 Sci-Hub 이용 현황 |
| 2. Black Open Access | 5. 결 론 |
| 3. 연구 방법 | |

초 록

본 연구는 저작권을 우회하여 유료로 제공될 학술 논문의 원문을 무료로 제공하고 있는 Sci-Hub에 대해 살펴보고, 2018년 Sci-Hub 측에서 공개한 데이터셋을 바탕으로 국내 Sci-Hub 이용 현황을 분석하여 접근제한에 처한 연구자들의 이용행태를 이해하기 위한 기초적인 바탕을 확립하기 위해 수행되었다. 따라서 Sci-Hub와 관련된 현황을 파악한 후에, 데이터셋을 다운로드된 지역별, 학술 논문의 주제, 출판사, OA 여부, 출간된 연도로 범주를 설정하고 분석하였다. 연구 결과 다운로드가 가장 많이 된 지역은 서울과 수도권이었으며, 자연과학 분야의 논문과 학술지가 가장 많이 다운로드되었고, 대략 20%정도의 논문이 Open Access 상태였다. 2010~2017년에 발간된 논문들이 가장 많이 다운로드 되었으며, IEEE의 논문이 가장 많이 다운로드되어 이를 종합해보면 다운로드된 시점 기준 최근에 발간된 자연과학 계열의 학술논문이 가장 많이 다운로드 되었다는 것을 보여주었다.

ABSTRACT

This study was conducted to examine Sci-Hub, which provides the original text of academic papers to be provided for a fee by bypassing copyright, and to establish a basic basis for understanding the usage behavior of researchers with access restriction in Korea by analyzing the domestic Sci-Hub usage based on the dataset released by Sci-Hub in 2017. Therefore, after grasping the current status of the world related to Sci-Hub, the categories were set and analyzed by region where the dataset was downloaded, subject matter of academic papers, publisher, OA status, and published year. As a result of the study, the most downloaded areas were Seoul and the metropolitan area, and papers and journals in the field of natural science were downloaded the most, and about 20% of papers were in Open Access state. The papers published between 2010 and 2017 were the most downloaded, and IEEE's papers were the most downloaded, showing that recently published academic papers in the natural sciences were the most downloaded by the time they were downloaded.

키워드: 싸이허브, 페이월, 접근 제한, 무단 공유, 학술 출판사

Sci-Hub, Paywall, Access Restriction, Unauthorized Sharing, Academic Publisher

* 성균관대학교 사서교육원 강사(doukong@skku.edu / ISNI 0000 0005 0858 5565)

논문접수일자: 2022년 10월 26일 최초심사일자: 2022년 11월 7일 게재확정일자: 2022년 11월 16일
한국문헌정보학회지, 56(4): 473-490, 2022. <http://dx.doi.org/10.4275/KSLIS.2022.56.4.473>

※ Copyright © 2022 Korean Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

1.1 연구의 필요성 및 목적

1950년대까지 대부분의 학술지는 구독을 기반으로 하는 학회나 기관 같은 비영리 조직에 의해 운영되다가, 2차 세계 대전을 거치며 학문 분야의 세분화와 그에 따른 연구 결과의 폭발적인 증가로 인해 상업 출판사들이 점차 독점적인 자리를 차지하게 된다. 21세기에 학술 출판은 연구 결과의 공유라는 기존의 역할 외에도 매우 수익성 있는 산업이 되었으며, 전 세계 연구자들의 전문적 지위에 큰 영향을 끼친다 (Fyfe, 2018). 그렇기 때문에 연구자들은 학술지의 비용 특성이나 가격에는 관심이 없고 학술지의 평판과 범위에만 관심이 있으며 도서관은 학술지를 구독할 수밖에 없는 상황이지만 대형 상업 출판사들의 독점적인 지위를 이용한 구독료의 인상이 도서관의 재정에 큰 부담이 되는 것은 주지의 사실이다. 출판되는 모든 학술지를 구독할 수 있는 기관은 아무도 없으며 이것이 Open Access(이하 OA) 운동이 중요해진 이유 중의 하나이다(Houle, 2017).

3B라고 불리는, 2001년 12월 OSI(Open Society Initiative)의 결과인 BOAI(Budapest OA Initiative) 선언, 2001년 Bethesda 선언, 2003년 Berlin 선언에 의해서 본격적인 OA 운동이 시작된 지 20여 년이 되었지만, 아직 OA는 처음에 기대했던 수준에 미치지 못하고 있는 것이 사실이다.페이월로 인해 연구자들이 자유롭게 접근할 수 없는 정보가 존재하고 OA 운동이 지지부진한 가운데 페이월 문제를 다른 각도로 돌파하고자 한 움직임이 Black Open

Access이다(Björk, 2017). Black Open Access는 저작권법을 우회하여 유료로 이용될 자료를 무료로 제공하고 있기 때문에 논란의 대상이 되고 있다.

2015년 Elsevier와 미국화학협회(American Chemical Society: ACS)가 저작권 침해와 컴퓨터 사기 및 남용법 위반에 대한 손해배상을 청구하며 본격적으로 논란의 대상이 된 Sci-Hub는 이러한 Black Open Access 중에서 대표적인 사이트이다. 이 배일에 가려진 Sci-Hub를 분석하기 위하여 『Science』의 Bohannon은 Sci-Hub의 창립자로 알려진 알렉산드라 엘바키안(Alexandra Elbakyan)에게 요청하여 6개월(2015년 9월부터 2016년 2월) 동안 있었던 2,800만 건의 다운로드 데이터를 공개하였고 Bohannon이 이를 분석한 기사를 발표한 이후로 Sci-Hub는 그 어떤 리포지터리보다 더 유명하게 되었다(Baich, 2017).

엘바키안은 자신의 트위터 계정과 Sci-Hub 홈페이지를 통하여 최근 한 달 동안 Sci-Hub에서 다운로드 요청이 가장 많았던 국가 20개국을 공개하고 있는데, 2022년 9월 기준 한국에서는 653,789개의 다운로드 요청이 있었다. 또한, 2018년 Sci-Hub 측에서 공개한 데이터셋에 의하면 전국 99개 도시에서 3,063,053의 적지 않은 다운로드 요청이 있었다. 이는 전 세계 국가 순위로는 12위, 비율로는 2%에 해당한다.

학술지와 관련된 논의점은 접근 확대로 인한 이점과 접근 제한으로 인한 해로운 영향으로 압축된다(LaDue, 2018). 접근 확대로 인한 이점에 관한 연구뿐 아니라 접근 제한이 연구자들에게 미치는 영향 또한 연구되어야 할 중요한 문제이다. 페이월의 접근 제한 문제가 야기

시킨 Sci-Hub는 학술 정보 유통 과정에서 무시할 수 없는 존재임에도 불구하고 국내에서는 아직 이에 대한 분석이 이뤄지지 않고 있다. 그러나 연구자들에게 그들이 원하는 학술 정보를 제공해야 하는 대학·연구 도서관의 입장에서 이러한 현상을 정확히 이해할 필요가 있다.

따라서 본 연구는 이와 같은 필요에 근거하여 Sci-Hub와 그를 둘러싼 현황에 대해서 살펴본 후 2018년 Sci-Hub 측에서 공개한 데이터셋을 분석하여 이러한 비합법적 플랫폼을 통해 학술 정보에 접근하고 있는 연구자들의 이용 행태를 분석할 수 있는 기초적인 토대를 마련하고자 한다.

1.2 선행 연구

2016년 Bohannon은 Sci-Hub 측에 로그데이터 공개를 요청하였고 그에 따라 2015년 6월부터 2016년 2월까지의 데이터셋이 공개된다. 이 데이터셋은 다운로드된 자료의 DOI, 다운로드 날짜와 시간, 다운로드의 귀속 위치,¹⁾ 국가로 구성되어 있고 개별 IP 어드레스는 공개하지 않았다. 본격적으로 이 데이터셋이 분석되기 이전에는 대부분의 요청이 개발도상국에서 발생할 것으로 예상하였으나 Si-Hub의 논문 요청 중 1/4은 학술지의 접근성이 가장 좋은 경제협력개발기구(OECD) 34개 회원국으로부터 온 것으로 추정되었고 가장 강력한 사용 중 일부는 미국과 유럽 대학의 캠퍼스에서 일어나고 있는 것으로 보인다(Bohannon, 2016). Elsevier가 소송에서 승리하면서 2015년 11월 sci-hub.org

의 도메인이 압수되었지만²⁾ 수많은 미리링이 미국 관할권 밖에서 일어나면서 한 달 후 새로운 도메인명으로 다시 등장했고(Priego, 2016) 이 때문에 Sci-Hub의 주소는 수시로 바뀌지만 위키피디아, 레딧, 트위터 등에서 새로운 주소를 찾는 일이 어려운 일은 아니며 최근에는 텔레그램봇을 통해 다운로드를 하는 방법도 있다.

Bohannon(2016)의 분석에 의하면 남극을 제외한 모든 대륙에서 다운로드 요청이 있었는데 가장 다운로드 요청이 많았던 도시는 테헤란으로 127만 건의 요청이 있었다. 2017년 한 해 동안 다운로드 요청이 많았던 상위 20개국은 <표 1>과 같다.

다운로드된 학술 논문의 출판사 중 가장 높은 비율을 차지한 곳은 Elsevier였고, 그 뒤를 Springer Nature - Institute of Electrical and Electronics Engineers(IEEE) - American Chemical Society(ACS)- Wiley-Blackwell이 차지했다(<그림 1> 참조).

Greshake(2017)는 Sci-Hub의 말뚝치(Corpus)와 데이터셋을 비교·분석하였는데 다운로드된 학술 논문의 95% 이상이 1982년 이후에 출간된 것이었고, 35%가 다운로드된 시점 기준 2년 이내에 발간된 것이었다. 1900년 이전에 출간된 학술 논문은 0.04% 정도였다. 전체 다운로드의 80%가 9개 출판사를 통해서 발간되는 것으로 확인되어 출판사의 과점성이 더욱 두드러지는 것으로 나타났다. 다운로드된 학술지는 총 177,000종으로 대략 25% 정도가 OA 상태였는데 가장 많이 다운로드된 학술지 20개 중 12개가 화학분야의 학술지임이 드러났다.

1) 이용자의 프라이버시를 보호하기 위해 이용자의 지리적 위치를 가장 가까운 도시의 지리 좌표계로 표시하였다.
2) 이때 Sci-Hub의 서버가 옮겨지면서 공개된 데이터셋에 2015년 11월의 데이터는 포함되어 있지 않다.

〈표 1〉 2017년 다운로드 요청 상위 20개국

번호	국가	다운로드 요청 수	번호	국가	다운로드 요청 수
1	중국	24,938,665	11	스페인	3,089,349
2	인도	13,143,462	12	한국	3,063,053
3	미국	11,991,045	13	독일	2,881,769
4	브라질	7,004,792	14	캐나다	2,808,743
5	이란	6,619,989	15	이탈리아	2,677,048
6	인도네시아	5,382,067	16	말레이시아	2,403,398
7	프랑스	4,208,280	17	네덜란드	2,194,240
8	러시아	4,013,310	18	콜롬비아	2,041,680
9	멕시코	3,271,085	19	튀르키예	2,004,756
10	영국	3,150,705	20	대만	1,999,026



〈그림 1〉 가장 많이 다운로드된 상위 15개 출판사
(출처: Bohannon 2016, p. 509)

Himmelstein et al.(2018)은 2017년 3월 21일 CrossRef에서 다운로드된 DOI에 의해 식별된 81,609,016개를 ‘Scholarly Literature’라고 정의하고 Sci-Hub가 공개한 DOI를 비교한 결과 전체 논문의 68.9%의 ‘Scholarly Literature’가 Sci-Hub에 저장되어 있었다. 나머지 31%의 경우도 Sci-Hub에 저장이 되어 있지 않다가 보다는 Sci-Hub 측에 아직 요청되지 않은 자료라고 Himmelstein은 『Science Insider』와의 인터뷰

뷰에서 주장했다(McKenzie 2017). ‘Scholarly Literature’는 학술지에 실린 학술 논문이 가장 큰 비중이었는데, 그것의 77.8%를 Sci-Hub 측에서 제공하고 있었다. Sci-Hub에서 다운로드된 학술 논문의 대부분이 구독 기반 유료 학술지에서 출간된 것(85.1%)이었다. 주제별로 보면 화학 분야가 93.0%로 가장 높았고 컴퓨터 사이언스 분야가 76.3%로 가장 낮았다. 적용 범위가 75% 미만인 주제 분야는 없었다. 출판사 별로 보

면 Elsevier의 96.9%, Springer Nature의 89.7%, Wiley-Blackwell의 94.7%, Taylor & Francis의 92.6%, Waters Kluwer의 88.3%, Oxford University Press의 88.3%, Sage Publication의 90.9%, ACS의 98.9% 학술 논문을 확보하고 있었으며 341개의 출판사의 경우는 완전히 확보한 것으로 나타났다.

La Due(2018)는 미국 연구자들의 Sci-Hub 이용을 분석하여 미국에서의 Sci-Hub 이용이 필요에 의한 것이 아닌 편리성에 의한 것이라는 주장에 대한 타당성 조사하였다. 그를 위해 Sci-Hub 다운로드의 지리적 특성이 대학의 제도적 특성과 어떤 관련이 있는지, 지리적 위치의 인구통계와 어떤 관련이 있는지를 분석하였는데 이 분석 결과에 따르면 대학원생과 교수진이 많고 연구비 지원 수준이 높은 기관에서 Sci-Hub의 사용 빈도가 높은 것으로 나타나 필요성이 편리성보다 Sci-Hub를 이용하는 원동력에 가깝다는 것을 시사하였다.

González-Solar & Fernández-Marcial(2019)는 도서관이 Sci-Hub가 가지고 있는 즉시성을 제공하지 못하고 있지만 대학에서 연구 지원 서비스의 발전으로 이용자가 정보 출처에 대한 조연과 지침을 구하기 위해 계속해서 도서관을 찾는다는 연구 결과를 언급하며 도서관이 Sci-Hub 같은 불법 시스템에 대한 정보 질의에 직면할 준비를 하고 있어야 한다고 말하고 있다.

이처럼 학술커뮤니케이션에서 무시할 수 없는 흐름이 된 Sci-Hub와 관련된 국내 연구가

아직 이뤄지지 않고 있기 때문에 먼저 Sci-Hub 이용 실태에 대한 분석은 반드시 필요하며 더 나아가 Sci-Hub를 이용하는 연구자들의 이용 요구와 행태에 대한 연구도 뒤따라야 할 것이다.

2. Black Open Access

본 연구에서는 먼저 구독 기반의 접근방법과 기대보다 느리게 진전되고 있는 OA, 기존 저작권에 대한 무지 혹은 무시가 결합하여 새롭게 등장한 Black Open Access와 페이지 뒤에 존재하는 유료 학술 논문을 무료로 다운받을 수 있게 제공하는 해적 사이트인 Sci-Hub에 대해서 살펴보고, 2018년에 Sci-Hub에 의해서 공개된 2017년 한 해 동안의 로그파일 데이터셋을 다양한 관점에서 분석하여 국내에서의 Sci-Hub 이용 실태를 밝히고자 한다.

2.1 그림자 도서관

Björk(2017)가 Black Open Access³⁾라고 부를 것을 제안한 새로운 채널들은 학술 소셜 네트워크(Academic Social Network, 이하 ASN)와 해적판 불법 사이트(Pirate Copy Site)/그림자 도서관(Shadow Library)으로⁴⁾ “저작권 규칙을 적용하는 리포지터리와는 대조적으로, 이 새로운 채널은 대부분 출판된 학술 기사의 디지털 복제본을 불법으로 제공하고 복제한

3) 게릴라 OA(Guerilla Open Access), 해적 OA(Pirate Open Access), 도적 OA(Rogue Open Access), 로빈 후드 OA(Robin Hood Open Access)라고도 불린다.

4) Björk(2017)는 그의 논문에서 해적판 불법 사이트라고 칭했으나 대다수의 연구에서 그림자 도서관이라 칭했기 때문에 본 연구에서도 그림자 도서관이라는 호칭을 사용한다.

다.”고 하였다.

소셜 네트워크의 특성과 연구 결과물을 결합하여 연구자의 필요에 맞게 조정된 연구자 전문 소셜 네트워크인 ASN을 통하여 연구자들은 자신의 학술 논문을 업로드하여 이용자들이 이용할 수 있게 함으로써 이용자가 학술 정보에 접근할 수 있는 또 다른 통로가 된다(Martín-Martín et al., 2018). 연구자들은 학술 논문을 학술지에 게재할 때 저작권을 양도한다는 사실을 모르고 자신의 학술 논문을 공유하기도 한다. 현재 가장 지배적인 두 플랫폼은 Acadimia.edu와 ResearchGate이다. 2021년 현재 Acadimia.edu는 1억 5천만 명, ResearchGate는 2,000만 명의 회원을 보유하고 있다. Borrego(2017)가 2016년 13개 스페인 대학의 기관 리포지터리와 ResearchGate의 연구 결과물을 비교한 결과, 이들 대학의 연구자가 2014년에 발표한 학술 논문의 54.8%가 ResearchGate에 원본으로 업로드되어 있었다. 이들에게 기관 리포지터리에 업로드하지 않은 채로 ResearchGate에만 업로드한 이유를 조사했더니 기관 리포지터리에 대한 존재나 이용 방법에 대한 무지, ResearchGate의 이용 편리성과 ResearchGate가 기관 리포지터리보다 학술 논문의 가시성이 더 높은 점 등 때문에 기관 리포지터리에 업로드를 하지 않은 것으로 나타났다.

그림자 도서관이란 “사회적·정치적 의제를 가진 이용자가 저작권이 있는 자료를 업로드하고 다운로드할 수 있는 디지털 텍스트 리포지터리 및 커뮤니티”를 말한다(Ekström, 2017). 그중 가장 규모가 큰 그림자 도서관인 Library Genesis(LibGen)는 2008년경 러시아어 텍스트 컬렉션을 통합하기 위해 시작한 그림자 도서관이다. LibGen은 소스 코드, 카탈로그, 테라바이트 크

기의 컬렉션을 포함한 자체 라이브러리 인프라를 오픈 소스로 배포하였고 이에 따라 LibGen 주변에 그림자 도서관의 활발한 생태계가 조성되었다.

2.2 Sci-Hub 작동 방식

2009년 카자흐스탄 대학에서 생체검사 신경과학 석사과정에 있던 알렉산드라 엘바키안(Alexandra Elbakyan)은 대학이 구독 기반의 데이터베이스를 제공하지 않았기 때문에 연구를 진행하는 데 비용이 지나치게 많이 들고, 큰 어려움을 느끼는 상황에 대해 문제의식을 느끼고 2011년 이용자의 신원을 익명화하는 툴인 Anonymizer의 소스 코드를 일부 변경하여 페이지를 우회하여 학술 논문을 다운로드받는 스크립트의 형태로 Sci-Hub를 개발하였다. 이는 페이지에 부딪혀 학술 논문에 접근할 수 없는 연구자들에게 접근을 제공하는 수많은 그림자 도서관 중의 하나로 LibGen의 대표적인 미러 사이트이다(Elbakyan, 2017).

처음에는 자체 리포지터리가 없이 임의의 대학 프록시를 이용하여 이용자가 제공한 학술 논문의 URL이나 DOI로 학술 논문을 다운로드받아 제공하고 시간이 지나면 자동으로 서버에서 파일을 제거하는 방식이었다. 그러다 2013년 중국에서 요청수가 폭발적으로 증가하면서 LibGen에 요청받은 파일이 있을 경우에는 LibGen으로 리디렉션을 하고, 파일이 없을 경우에는 Sci-Hub가 취득한 인증정보⁵⁾를 이용하여 대학의 데이터베이스에서 학술 논문을 다운받아서 이용자에게 제공한 후 LibGen에도 저장하는, LibGen을 일종의 외부 리포지터리로 이

용하는 방식으로 바꾸었다. 이러한 방식은 2014년 무렵까지 계속되다가 2015년부터는 자체 리포지토리를 운영하고 있는 상황이다(Otani & Bando, 2018).

Sci-Hub측의 주장에 의하면 2022년 현재 100TB 크기의 자체 데이터베이스에 88,343,822개의 문헌을 보유하고 있으며 이중 80%가 학술지에 실린 학술 논문, 6%가 프리시딩, 5%가 단행본의 장(章)이다. 77%가 1980~2022년에 출간되었으며 주요 학술지의 95%이상을 확보하고 있다. 주제별로 보면 의학 24,577,530 화학 16,460,921 생물학 15,499,507 인문학 12,592,316 물리학 8,658,518건 등이다.

2.3 Elsevier, ACS와의 소송

Elsevier는 2015년 6월에 미국 뉴욕 남부지방법원에 Sci-Hub, LibGen, Elbakyan에 대해 저작권 침해와 컴퓨터 사기 및 남용법 위반에 대한 손해배상을 청구하는 민사소송을 제기했다. 2015년 10월 Elsevier 측의 승리로 Sci-hub.org는

폐쇄되었으나 미국 관할권 밖에 수많은 미러링 사이트가 존재하며 오히려 이로 인해 Sci-Hub라는 존재가 부각되는 부작용을 가져왔다. 이후로도 Sci-Hub는 주소를 계속 옮기면서 운영이 되고 있다.

3. 연구 방법

2018년 Sci-Hub 측에서 공개한, 2017년 한 해 동안의 로그 데이터 셋은 <표 2>와 같이 구성되어 있다.

이 데이터셋을 이용하여 다음과 같은 범주로 데이터를 분석하였다.

- 지역별 다운로드 분포 분석
 - 다운로드된 학술 논문의 출간 연도 분석
 - 다운로드 빈도
 - 다운로드된 논문들의 출판사 분석
- DOI의 prefix를 이용하여 이 실린 학술 논문의 출판사 빈도 분석

<표 2> Sci-Hub 2017 로그 데이터 셋 항목

항목	예
timestamp(yyyy-mm-dd hh:mm:ss)	2017-01-08 11:21:22
DOI	10.1002/ptr.5697
IP identifier	483050
User identifier	735171
Country according to GeolP	South Korea
City according to Geolp	Ansan-si
Latitude	37.3218778
Longitude	126.8308848

5) 인증정보를 획득하는 방법에 대해서는 Sci-Hub와 출판사 측의 주장이 다르다. Sci-Hub는 자신들의 취지에 동조하는 많은 자원봉사자들로부터 자격증을 제공받는 등 밝힐 수 없는 방법으로 수집을 하지만 피싱을 하지는 않는다고 하였고, 출판사 측에서는 일부 정보가 피싱되었을 것이라고 주장하고 있다.

- 다운로드된 학술 논문들의 OA 여부 Unpaywall의 API를 통한 학술 논문들의 OA 여부 분석
- 다운로드된 학술지의 주제 분류 Web of Science와 KCI 학술지 분류를 교차 분석하여 분류

4. 국내 Sci-Hub 이용 현황

4.1 국내 지역별 다운로드 분포 분석

2017년 한 해 동안 국내 Sci-Hub 다운로드 요청 건수는 3,063,053건이다. 이것을 월별로 집계해 보면 <그림 2>와 같다.

이를 보면 계속적으로 2017년 한 해 동안 이용이 점점 증가하는 추세였으며 9월에 다운로드 횟수가 급등한 것이 보인다.

공개된 GeoIp를 기반으로 국내에서 다운로드된 상황을 지역별로 구분해 보면 <표 3>과 같다. 이 중 20,752건은 지리정보가 훼손되어 지

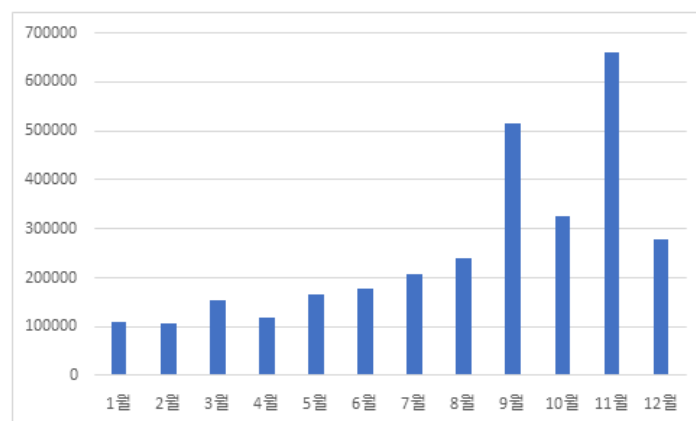
역을 특정할 수 없었다. Sci-Hub 측에서는 구체적인 지리 위치를 밝히지 않기 위하여 다운로드가 발생한 지점에서 가장 가까운 행정구역(예: 서울은 서울시청 37.566535, 126.9779692 등)의 경도와 위도로 위치를 표시하였다.

지역별 요청을 지도에 표시한 것은 <그림 3>과 같다.

대도시를 중심으로 전국 99개의 도시에서 이용요청이 있었음을 알 수 있다.

학술 정보의 수요가 가장 높을 것으로 예상되는 공공연구기관, 대학, 기업체 등 연구수행 주체의 분포와 다운로드 요청 횟수의 상관관계를 살펴보았다. 이를 위해 연구수행주체별 연구개발수행조직, 연구개발비, 연구자 수를 지역별로 정리하면 <표 4>, <표 5>, <표 6>과 같다. 이들의 상관관계를 알기 위해서 R4.2.1로 분석해보았을 때 다운로드횟수와 연구수행주체별 연구개발수행조직, 연구개발비, 연구자 수 간에 양의 상관관계를 갖고 있는 것이 보인다 (<그림 4> 참조).

한국의 경우 연구수행주체들이 서울과 수도

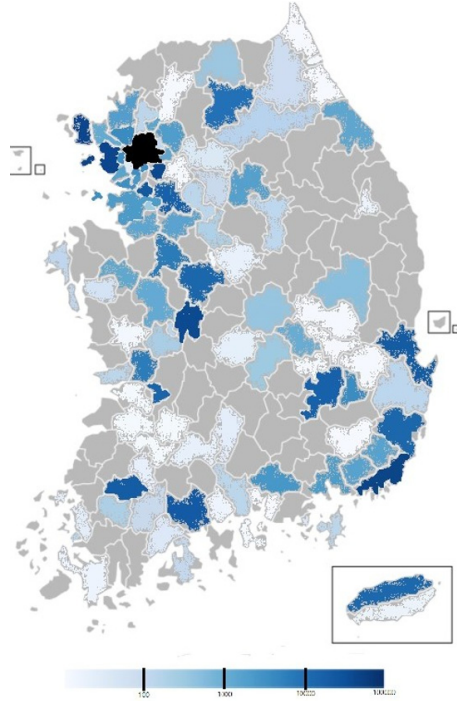


<그림 2> 월별 다운로드 횟수

〈표 3〉 국내 지역별 다운로드 횟수

행정구역	도시	다운로드 횟수	행정구역	도시	다운로드 횟수	행정구역	도시	다운로드 횟수
서울특별시	서울특별시	2,480,527	강원도	강릉	1,625	전라남도	나주	303
부산광역시	부산광역시	53,526		원주	2,669		무안	44
대구광역시	대구광역시	26,932		춘천	9,688		목포	199
인천광역시	인천광역시	54,994		화천	544		순천	27,865
광주광역시	광주광역시	26,182		홍천	223		고흥	106
대전광역시	대전광역시	54,204		고성	27		화순	903
울산광역시	울산광역시	14,677		태백	17		보성	51
경기도	성남	57,417		양양	2		담양	24
	부천	10,026		속초	2		해남	17
	용인	20,177		인제	92		광양	14
	수원	32,269			구미	1,289		
	포천	3			경산	6,972		
	안양	6,829	충청북도	음성	324	김천	474	
	화성	3,483		제천	145	안동	743	
	의정부	2,072		청주	17,891	포항	26,262	
	양평	57		괴산	31	상주	559	
	안성	1,985		증평	3	영천	7	
	평택	1,670		영동	37	군위	4	
	시흥	1,990	충청남도	공주	1,310	칠곡	2	
	고양	6,254		천안	8,188	의성	8	
	군포	3,787		태안	149	김해	2,109	
	오시	475		예산	1,806	하동	188	
	안산	5,065		논산	391	진주	3,977	
	과천	5,611		계룡	201	양산	2,415	
	남양주	2,801		부여	7	거제	172	
파주	2,712	홍성		76	경상남도	남해	7	
하남	1,479	전라북도	군산	156	사천	1		
김포	1,138		익산	7,983	밀양	1		
이천	196		전주	16,730	통영	61		
양주	189		김제	7	함양	32		
구리	100		정읍	4	창원	3,137		
여주	95		임실	3	제주도	제주	15,650	
광명	47		남원	28	서귀포	17		
광주	5		순창	8	N/A		20,753	

계 3,063,053



〈그림 3〉 전국 지역별 다운로드 현황

〈표 4〉 2017년 연구수행주체별 연구개발수행조직(지역)

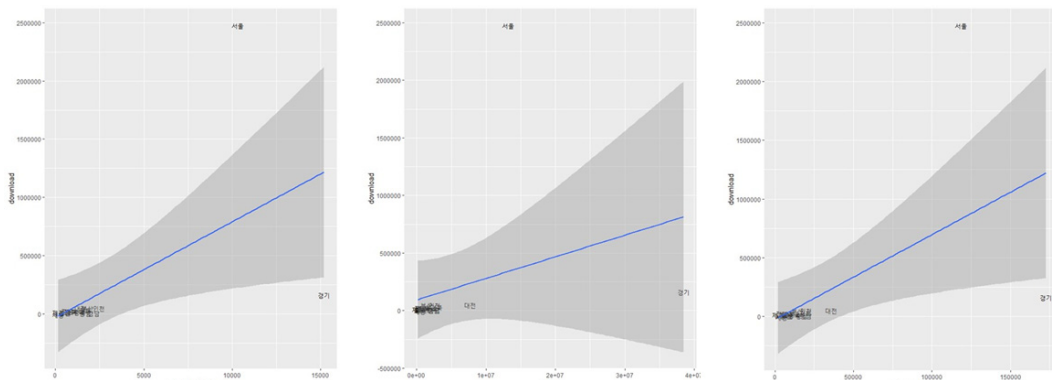
연구수행 주체별(1)	연구수행 주체별(2)	2017																	
		연구개발 수행조직	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
총계	소계	43,510	10,310	1,810	1,651	2,469	807	1,420	521	154	15,180	541	1,389	1,681	889	669	1,683	2,187	149
공공연구기관	소계	569	124	29	27	15	19	35	12	17	64	30	25	20	38	26	38	33	17
	국공립	127	9	6	1	6	3	4	4	-	11	10	8	11	10	9	15	11	9
	정부출연	160	21	11	10	2	9	23	6	14	12	7	6	3	12	4	6	10	4
	지방자치단체출연	45	4	4	1	-	3	1	-	-	2	5	3	2	7	3	6	3	1
	기타비영리	203	72	7	14	7	4	7	2	3	35	6	7	3	8	7	11	7	3
	국공립병원	15	4	1	-	-	-	-	-	-	1	1	1	1	1	3	-	2	-
사립병원	19	14	-	1	-	-	-	-	-	3	1	-	-	-	-	-	-	-	
대학	소계	410	85	23	15	16	12	18	5	4	77	18	20	28	18	17	26	23	5
	국공립	89	12	5	5	3	4	5	1	-	6	7	6	6	5	6	5	11	2
	사립	321	73	18	10	13	8	13	4	4	71	11	14	22	13	11	21	12	3
기업체	소계	42,531	10,101	1,758	1,609	2,438	776	1,367	504	133	15,039	493	1,344	1,633	833	626	1,619	2,131	127
	정부투자기관	49	6	-	3	3	1	13	3	-	6	2	1	1	2	3	2	2	1
	민간기업	42,482	10,095	1,758	1,606	2,435	775	1,354	501	133	15,033	491	1,343	1,632	831	623	1,617	2,129	126

〈표 5〉 2017년 연구수행주체별 연구개발비(지역)

연구수행주체별(1)	연구수행주체별(2)	2017																	
		연구개발비	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
총계	소계	78,789,187	13,169,674	1,403,273	1,237,953	2,548,150	793,269	7,698,499	729,870	483,714	38,462,499	450,289	2,212,864	2,561,356	1,032,155	549,121	2,846,783	2,453,699	156,024
공공연구기관	소계	9,543,217	1,071,572	317,280	272,819	235,553	157,946	4,145,370	79,896	292,336	841,387	99,166	357,904	289,992	378,335	142,171	247,749	548,811	64,930
	국공립	797,894	54,516	112,669	20,819	62,923	1,316	17,158	8,538	-	50,531	39,066	71,968	59,802	132,165	35,050	35,604	44,009	51,760
	정부출연	7,416,927	697,205	134,766	121,628	143,454	68,258	4,104,638	65,594	291,025	530,677	43,398	240,588	168,084	188,122	52,626	68,128	489,376	9,358
	지방자치단체출연	197,143	31,973	11,656	7,700	-	27,964	2,672	-	-	3,632	12,929	10,548	820	22,445	22,444	30,846	9,004	2,510
	기타비영리	993,528	165,558	58,183	122,306	29,176	60,407	20,902	5,764	1,311	243,599	2,948	34,780	61,163	35,601	31,819	113,171	5,539	1,302
	국공립병원	17,752	16,026	6	-	-	-	-	-	-	409	52	20	123	2	231	-	883	-
	사립병원	119,973	106,295	-	366	-	-	-	-	-	12,539	773	-	-	-	-	-	-	-
대학	소계	6,682,523	2,501,918	422,944	306,097	207,069	226,446	506,691	139,657	31,196	927,351	167,073	151,918	203,453	230,466	127,134	332,923	144,190	55,997
	국공립	2,957,916	807,570	279,063	264,317	25,883	218,750	445,761	110,691	-	55,363	98,852	104,688	62,665	156,010	110,988	39,406	123,749	54,159
	사립	3,724,608	1,694,348	143,881	41,780	181,186	7,696	60,930	28,966	31,196	871,988	68,221	47,230	140,788	74,456	16,146	293,517	20,441	1,838
기업체	소계	62,563,447	9,596,183	663,049	659,037	2,105,528	408,873	3,046,438	510,317	160,183	36,693,761	184,050	1,703,042	2,067,911	423,354	279,817	2,266,110	1,760,698	35,097
	정부투자기관	890,135	24,494	-	4,321	41,962	695	581,731	11,472	-	120,488	8,288	6,655	1,240	12,050	24,471	46,343	4,262	1,663
	민간기업	61,673,312	9,571,689	663,049	654,716	2,063,566	408,178	2,464,707	498,845	160,183	36,573,273	175,762	1,696,387	2,066,671	411,304	255,346	2,219,767	1,756,436	33,434

〈표 6〉 연구수행주체별 연구자 수(지역)

연구수행주체별(1)	연구수행주체별(2)	2017																	
		연구원	서울	부산	대구	인천	광주	대전	울산	세종	경기	강원	충북	충남	전북	전남	경북	경남	제주
총계	소계	482,796	118,541	14,371	11,781	19,635	7,722	35,745	7,807	4,109	172,583	6,668	12,324	17,139	9,126	4,493	19,335	19,584	1,833
공공연구기관	소계	36,552	5,413	1,097	1,116	923	556	10,708	652	2,545	4,170	899	1,584	1,023	1,653	720	1,292	1,752	449
	국공립	5,168	201	166	91	360	79	208	159	-	308	511	789	366	780	209	466	252	223
	정부출연	22,905	2,718	442	339	392	229	10,346	313	2,491	2,427	103	460	313	544	201	163	1,367	57
	지방자치단체출연	1,310	261	68	82	-	48	28	-	-	35	145	40	26	181	135	191	49	21
	기타비영리	6,070	1,391	419	596	171	200	126	180	54	1,321	60	291	304	146	140	472	51	148
	국공립병원	350	238	2	-	-	-	-	-	-	11	11	4	14	2	35	-	33	-
	사립병원	749	604	-	8	-	-	-	-	-	68	69	-	-	-	-	-	-	-
대학	소계	102,877	37,214	5,420	3,948	3,638	3,702	8,987	2,196	430	12,164	3,631	3,217	3,423	3,573	1,352	6,013	3,081	888
	국공립	40,772	8,409	2,680	3,134	837	3,152	7,832	1,371	-	1,088	2,112	1,759	800	2,389	1,054	910	2,397	848
	사립	62,105	28,805	2,740	814	2,801	550	1,155	825	430	11,076	1,519	1,458	2,623	1,184	298	5,103	684	40
기업체	소계	343,367	75,914	7,854	6,717	15,074	3,464	16,050	4,959	1,134	156,249	2,138	7,523	12,693	3,900	2,421	12,030	14,751	496
	정부투자기관	2,979	112	-	23	57	10	1,673	28	-	303	72	33	16	65	78	456	33	20
	민간기업	340,388	75,802	7,854	6,694	15,017	3,454	14,377	4,931	1,134	155,946	2,066	7,490	12,677	3,835	2,343	11,574	14,718	476



〈그림 4〉 다운로드 횟수와 연구개발수행조직수, 연구비, 연구자수와의 상관관계

권 지역에 몰려 있기 때문에 이 지역의 데이터가 패턴을 좌우하므로 수도권을 제외한 데이터로도 상관관계를 분석해 보았다(〈그림 5〉 참조). 이러한 분석에서도 다운로드 요청횟수와 연구수행주체별 연구개발수행조직, 연구개발비, 연구자 수 모두 양의 상관관계를 보였다.

이를 보면 연구개발을 수행하는 조직과 연구자 수가 많고, 연구개발비가 많이 수행될수록 Sci-Hub의 다운로드 횟수도 많다는 것을 예측해볼 수 있다.

4.2 다운로드된 학술 논문의 출간 연도 분석

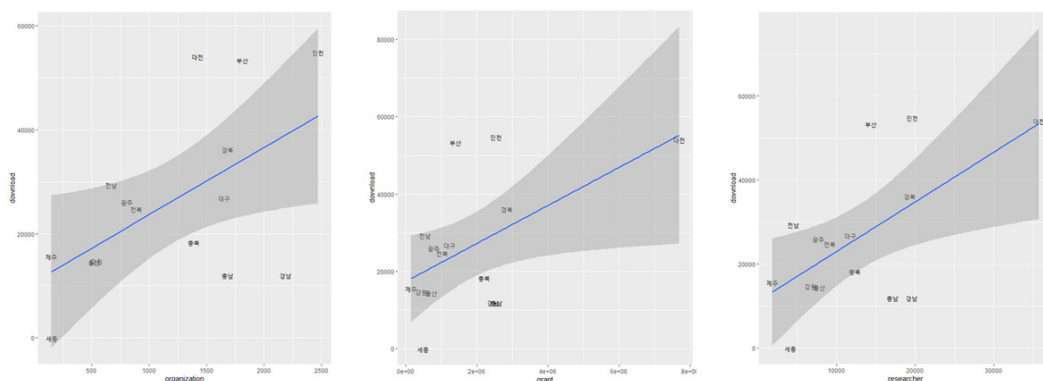
국내에서 다운로드된 학술 논문이 발간된 연도를 그래프로 표시해보면 〈그림 6〉과 같다. 가장 오래된 논문은 1748년 The Royal Society Publishing社에서 발간한 『Philosophical Transactions』에 실린 “VI. Divers means for preserving from corruption dead birds, intended to be sent to remote countries, so that they may arrive there in a good condition. Some of the same means may be employed for preserving quadrupeds,

reptiles, fishes, and insects, by M. de Reaumur, F. R. S. and Memb. Royal. Acad. Sc. Paris, translated from the French by Phil. Hen. Zollman, Esq: F. R. S”으로 비 OA 논문이다.

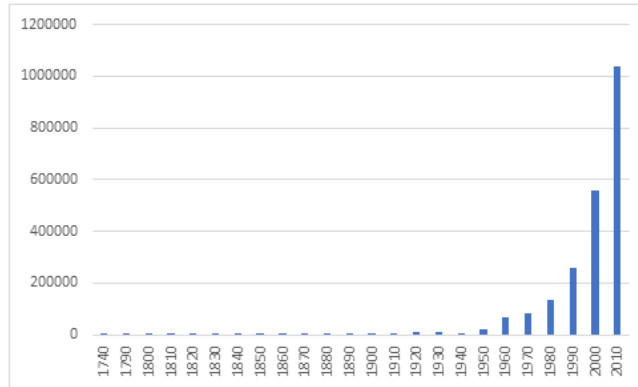
그림에서 보듯이 발간 연도 범위는 1748년~2018년으로 상당히 넓어서, 다양한 요구가 존재한다고 할 수 있다. 그러나 횟수로 보면 1990년대 발간된 논문이 11.6%, 2000년대 발간된 논문이 25.1%, 2010년대 발간된 논문이 46.6%로 83.3%를 차지한다. 2010년대를 좀 더 자세히 보아도 2017년 기준 최근 발간된 논문이 많이 다운로드되었다는 것을 알 수 있다.

4.3 다운로드 빈도

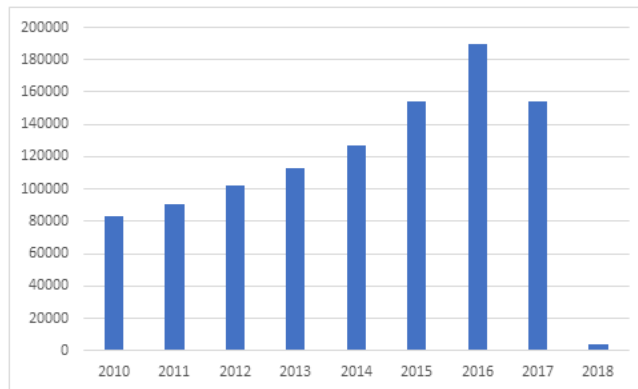
1회만 다운로드된 논문은 총 1,856,749편(81%)이다. 1회만 다운로드된 논문의 발간 연도를 보면 1800년대 이전 14,369편(1%), 1900~1949년 38,178편(2%), 1950~1999년 489,558편(27%), 2000~2009년 456,621편(26%), 2010~2017년 783,884편(44%)으로 2010~2017년 사이에 발간된 논문들이 가장 높은 비율을 차지



〈그림 5〉 수도권 제외 다운로드 횟수와 연구개발수행조직수, 연구비, 연구자수와의 상관관계



〈그림 6〉 다운로드 된 논문의 발간 연도



〈그림 7〉 2010년대 발간 분포

하고 있다.

2회~49회 다운로드된 논문은 총 498,516편이다. 2회~49회 다운로드된 논문의 연도를 보면 1800년대 2,316편(1%), 1900~1949년 6,630편(2%), 1950~1999년 77,782편(19%), 2000~2009년 94,007편(23%), 2010~2017년 233,369편(56%)으로 역시 2010~2017년에 발간된 논문이 가장 높은 비율을 차지하고 있다. 50~99회 다운로드된 논문은 총 98편이다. 발간 연도를 보면 1923년 논문이 1편, 1950~1999년 5편, 2000~2009년 12편, 2010~2017년 79편이다.

2010~2017년도의 논문이 가장 많이 다운로드되었다. 100회 이상 다운로드된 논문은 19개 논문이다. 발간 연도를 보면 1950~1999년 5편, 2000~2009년 2편, 2010~2017년 12편이다.

4.4 다운로드된 논문들의 출판사 분석

DOI의 prefix를 이용하여 다운로드된 논문들을 발간한 출판사 및 학회를 분석하면, 총 916개 사이며, 다운로드 횟수 상위 10개 출판사는 〈표 7〉과 같다. 상위 5개 사(IEEE, Springer

Nature, Wiley-Blackwell, AAAS, World Scientific)가 전체의 62.3%를 차지했고 상위 10개 사로 확대하면 81.2%에 달한다. 국내의 경우 다른 국가와 비교해 특이한 점은 대부분의 국가에서 가장 많이 다운로드된 출판사인 Elsevier가 7위인데 비해 국내에서 가장 많이 다운로드된 출판사는 IEEE라는 점이다. 학술 출판 업계의 빅 파이브인 Elsevier, Springer Nature, Taylor & Francis, Wiley-Blackwell, Sage Publication은 총 1,071,061건으로 32.7%를 차지한다. 이는 학술 출판 시장에서 특정 출판사가 차지하는 비중이 매우 높다는 것, 그들이 출판하는 학술 정보의 정보 접근 제한 비율이 매우 높다는 것을 보여준다. 1회만 다운로드된 출판사 및 학회는 146개 사이다.

국내 학회에서 발간된 논문이 다운로드된 경우도 있었는데, 69 학회에서 1,240회가 다운로드되었다.

4.5 다운로드된 학술 논문들의 OA 여부

학술 논문의 OA 여부의 정보를 제공하는

Unpaywall의 API를 이용해 Sci-Hub에서 다운로드된 학술 논문 2,278,529편의 OA 여부를 조사해본 결과 비(非) OA 논문이 1,777,068(78.0%), OA 논문이 492,241(21.6%)로 비(非) OA 논문이 더 큰 비중을 차지했다.

한국에서 가장 많이 다운로드된 논문은 총 261회 다운로드된 Elsevier社의 『Microchemical Journal(Vol. 14, Issue 2, June 1969)』에 실린 “N-bromosuccinimide as a direct titrant for thiourea and thioacetamide using bordeaux red as an indicator”(10.1016/0026-265x(69)90042-3)으로 비(非) OA 논문이다.

100회 이상 다운로드된 논문 19편의 OA 여부를 보면 2편이 OA, 17편이 비(非) OA 논문으로 OA 논문이 11%를 차지했다. 다운로드된 전체 학술 논문의 비 OA 논문의 비중보다 다운로드가 많이 된 학술 논문 중에서 비(非) OA의 비중이 조금 더 높았다. 이를 볼 때 Sci-Hub의 이용이 편리성 때문에 사용되는 경우도 있지만 접근 제한 정보로의 접근을 위해 Sci-Hub가 주로 사용된다고 할 수 있다.

〈표 7〉 다운로드가 많이 된 상위 10개 출판사

순위	출판사 및 학회	다운로드 횟수	비율(%)
1	Institute of Electrical and Electronics Engineers(IEEE)	677,730	22.1
2	Springer Nature	596,699	19.5
3	Wiley-Blackwell	288,028	9.4
4	American Association for the Advancement of Science(AAAS)	186,213	6.1
5	World Scientific	159,465	5.2
6	American Chemical Society(ACS)	157,028	5.1
7	Elsevier	151,225	4.9
8	Informa UK	125,380	4.1
9	The Royal Society of Chemistry	100,490	3.2
10	Institution of Electrical Engineers	43,557	1.4

4.6 다운로드된 학술지의 주제 분류

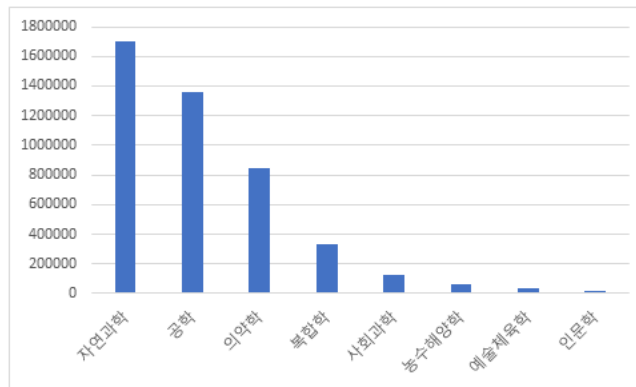
국내의 경우 다운로드된 학술지는 총 11,816 종(種)이다. 이를 Web of Science와 KCI의 학술지 분류를 교차 참고하여 공학, 농수해양학, 복합학, 사회과학, 예술체육학, 의약학, 인문학, 자연과학으로 분류하여 비교해보았다.

다운로드된 학술지의 주제를 보면 자연과학 분야가 1,700,724회로 가장 많이 다운로드되었고 공학 1,355,839회 그 뒤 의약학 850,017회 복합학 328,223회 사회과학 129,211회 농수해양학

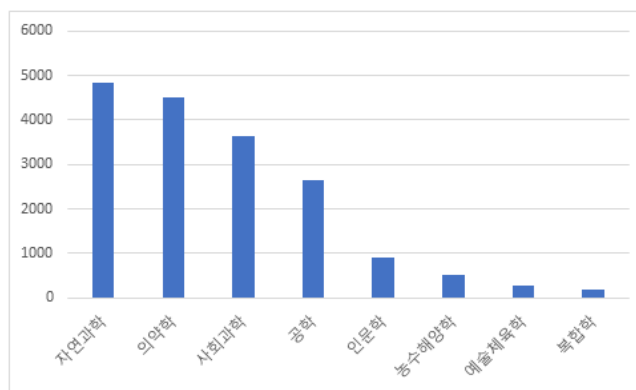
62,094회 예술체육학 33,593회 인문학 19,347회 다운로드되었다(〈그림 8〉 참조).

다운로드된 학술지 종류는 자연과학이 4,825종(種)으로 가장 많았고 그 뒤 의약학 4,515종 사회과학 3,646종 공학 2,643종 인문학 917종 농수해양학 504종 예술체육학 266종 복합학 177종이었다(〈그림 9〉 참조).

가장 많이 다운로드된 학술지는 American Association for the Advancement of Science (AAAS)의 『Science』였다. 가장 많이 다운로드된 학술지 10종은 〈표 8〉과 같다.



〈그림 8〉 다운로드된 학술지의 주제 분류



〈그림 9〉 다운로드된 학술지의 주제별 종류 수 비표

〈표 8〉 가장 많이 다운로드된 학술지 10종

	학술지명	출판사	다운로드 횟수
1	Science	American Association for the Advancement of Science(AAAS)	169,524
2	Nature	Springer Nature	94,744
3	IEEE Transactions on Nuclear Science	IEEE	22,307
4	IEEE Transactions on Magnetics	IEEE	21,794
5	Journal of the American Chemical Society	American Chemical Society(ACS)	19,077
6	IEEE Transactions on Applied Superconductivity	IEEE	15,201
7	Scientific Reports	Springer Nature	15,099
8	Applied Physics Letters	American Institute of Physics	14,675
9	IEEE Journal of Quantum Electronics	IEEE	13,653
10	Scientific American	Springer Nature	13,045

5. 결 론

본 연구는 유료로 이용될 논문을 무료로 다운로드받게 해주는 그림자 도서관인 Sci-Hub의 국내 이용 현황을 분석하여 차후 접근제한에 처한 연구자들이 그들에게 필요한 정보를 입수하지 못했을 때 어떻게 행동하는지를 이해하기 위한 기본적인 토대를 마련하기 위해서 수행됐다. 따라서 2018년에 Sci-Hub 측에서 발표한 데이터셋을 다각도로 분석하였다.

다운로드 횟수를 지역별로 정리해 보면 전국 99개 도시에서 다운로드 요청이 있었었다. 서울과 경기지역에서 가장 다운로드가 많이 발생했는데, 이는 서울과 경기지역에 연구개발수행조직이나 연구자 수가 많기 때문인 것으로 상관관계 분석 결과가 나왔다. 서울과 경기지역을 제외하고 분석하여도 연구개발수행조직, 연구자 수, 연구개발비와 다운로드 횟수는 양의 관계를 보였다.

2017년 한 해 동안 다운로드가 된 추세를 보

면 꾸준한 우상향 상승세를 보였고, 특히 9월에 폭발적으로 이용이 늘어났다.

다운로드된 학술 논문의 출간 연도를 보면 2010~2017년에 출간된 논문이 가장 많이 다운로드되었고, 발간 연도의 범위는 1748년~2017년으로 범위가 상당히 넓은 것으로 보아 다양한 이용 요구가 존재하는 것으로 보인다. 1990년대 발간된 논문이 11.6%, 2000년대 발간된 논문이 25.1%, 2010년대 발간된 논문이 46.6%로 83.3%를 차지하였다.

가장 많이 다운로드된 상위 5개의 출판사는 IEEE, Springer Nature, Wiley-Blackwell, AAAS, World Scientific으로 62.3%를 차지하고 있었고, 빅파이버가 시장을 차지하는 비율은 32.7%였다. 이를 보면 출판사의 과점현상이 Sci-Hub를 이용하는 주된 이유 중 하나라는 것을 추측해 볼 수 있다.

다운로드된 논문 중 OA 논문은 21.6%였다. 100회 이상 다운로드된 논문 중에 OA 논문의 비중은 11%로 OA 논문의 비중이 줄어든다.

Sci-Hub가 이용되는 주된 이유는 편리성보다는 접근 제한성 때문으로 보이지만, 전체적으로 OA 논문이 차지하는 비율이나 국내 학회에서 발간된 학술 논문들도 다운로드된 상황을 보았을 때 편의성에 의한 이용도 존재하는 것으로 보인다.

다운로드된 학술지의 주제를 보면 자연과학-공학-의약학-복합학-사회과학-농수해양학-예술체육학-인문학 순으로 나타나 사회과학이나 인문학보다는 자연과학, 공학, 의약학이 큰 비중을 차지하고 있었다.

Sci-Hub의 이용은 주로 자연과학, 공학, 의약학 분야에서페이월 등의 접근 제한에 의해

입수할 수 없는 논문을 열람하기 위해 전국에 걸쳐 적지 않은 수의 이용자들이 이용하고 있는 것으로 보인다. 그러나 Sci-Hub는 저작권을 위반하고 있는 사이트이다. 이용자들이 이러한 사이트를 이용하고 있는 것은 페이월의 높은벽, Sci-Hub의 편리성과 신속하지 못한 도서관 서비스에 대한 불만 등이 그 이유인 것으로 보인다. 그렇기 때문에 학술 커뮤니케이션에 종사하고 있는 도서관의 입장에서 Sci-Hub와 이용자에 대한 명확한 이해가 필요하며 이 분석을 바탕으로 Sci-Hub와 그 이용자에 대한 연구가 시작되어야 할 것이다.

참 고 문 헌

- Baich, T. (2017). Diminishing the Perceived Need for Black Open Access. Interlending and Document Supply Conference, Paris, France.
- Björk, B. C. (2017). Gold, green, and black open access. *Learned Publishing*, 30(2), 173-175.
- Bohannon, J. (2016). Who's downloading pirated papers? everyone. *Science*, 352(6285), 508-512.
- Borrego, Á. (2017). Institutional repositories versus ResearchGate: the depositing habits of Spanish researchers. *Learned Publishing*, 30(3), 185-192.
- Ekström, B. (2017). The Illicit Information Community: Information - Practical Reflections on the Shadow Library AAARG. BOBCATSSS 2017: Improving Quality of Life Through Information, Tampere.
- Elbakyan, A. (2018). Sci-Hub download log of 2017 [Data set]. Available: <https://zenodo.org/record/1158301#.Y1bEMXZBxPY>
- Elbakyan, A. (2019, Mar 31). Sic-Hub and Alenxandra basic information. Available: <https://engineering.wordpress.com/2019/03/31/sci-hub-and-alexandra-basic-information/>
- Fyfe, A. (2018). Publishing the philosophical transactions: the social, cultural and economic history of a learned journal. *Impact*, 2018(4), 33-35.

- González-Solar, L. & Fernández-Marcial, V. (2019). Sci-Hub, a challenge for academic and research libraries. *El Profesional de la Información*, 28(1), 1-12.
- Greshake, B. (2017). Looking into Pandora's box: the content of Sci-Hub and its usage. *F1000Research*. Available: <https://f1000research.com/articles/6-541>
- Himmelstein, D. S., Greene, C. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R., & Greshake T. B. (2018). Sci-Hub provides access to nearly all scholarly literature. *eLife*. Available: <https://elifesciences.org/articles/32822>
- Houle, L. (2017). Sci-Hub and LibGen What if... Why not. IFLA WLIC 2017 Wroclaw Satellite Meeting, Wroclaw, Poland.
- LaDue, J. O. (2018). Exploring the Convenience Versus Necessity Debate Regarding SCI-HUB Use in the United States. Doctoral dissertation. University of Pittsburgh, United States.
- Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS one*. Available: <https://doi.org/10.1371/journal.pone.0127502>
- Martín-Martín, A., Costas, R., van Leeuwen, T., & López-Cózar, E. D. (2018). Evidence of open access of scientific publications in Google scholar: a large-scale analysis. *Journal of Informetrics*, 12(3), 819-841.
- McKenzie, L. (2017, July 27). Sci-Hub's cache of pirated papers is so big, subscription journals are doomed, data analyst suggests. *Science*. Available: <https://www.sciencemag.org/news/2017/07/sci-hub-s-cache-pirated-papers-so-big-subscription-journals-are-doomed-data-analyst>
- Nazarovets, S. (2018). Black open access in Ukraine: analysis of downloading Sci-Hub publications by Ukrainian internet users. *Nauka innov*, 14(2), 19-26.
- Otani, S. & Bando, K. (2018). A pirate site Sci-Hub: recent trend and usage analytics from Japan. *The Journal of Information Science and Technology Association*, 68(10), 513-519.
- Priego, E. (2016). Signal, not solution: notes on why Sci-Hub is not opening access. *The Winnower*. Available: <https://thewinnower.com/papers/3489-signal-not-solution-notes-on-why-sci-hub-will-not-open-access>
- Tenopir, C. & King, D. W. (1997). Trends in scientific scholarly journal publishing in the United States. *Journal of Scholarly Publishing*, 28(3), 135-170.