한국경제포럼 제 12 권 제 4 호 125-181

한국 특허 데이터 프로젝트: 내용과 방법*

이 지 홍** · 임 현 경*** · 김 상 동**** · 송 근 상***** · 정 재 유******

논문초록 본 논문은 "한국 특허 데이터 프로젝트(Korea Patent Data Project, KoPDP)"의 내용과 방법에 대해 소개한다. KoPDP는 1948-2016년 동안 한국 특허청에 출원된 모든 (실용)특허와 1976-2016년 사이 한국에 위치한 출원인 및 발명자가 미국 특허청에 출원·등록한 모든 특허에 대한 자료를 수집하고, 출원인-기업명 매칭(matching)을 통해 각종 기업 관련 정보를 제공하는 DataGuide 5.0과 연결하였다. 그 결과 14,803개의 상장 및 비상장 한국 기업의 재무 및 금융 정보를 비롯하여 해당 기업이 한국과 미국에서 출원·등록한 특허 정보에 대한 포괄적인 데이터베이스를 구축하였다. 한국 특허의 경우 전체 표본 특허의 45% 이상이, 미국 특허의 경우 제1 출원인이 한국인 표본 특허의 약 87%가 한국 기업 정보와 연결되었다. 본 논문은 출원인-기업명 매칭 과정에 대한 자세한 설명과 함께 한국과 미국 특허 간에 일관된 산업 분류체계를 제공한다.

핵심 주제어: 특허 자료, 기업 자료, 혁신
경제학문헌목록 주제분류: C80, O30

투고 일자: 2019. 6. 15. 심사 및 수정 일자: 2019. 9. 26. 게재 확정 일자: 2019. 12. 20.
* 이 연구는 서울대학교 박양숙-정영호 기초학문 후원기금 사업으로 지원되는 연구비에 의하여 수행되었음.
** 교신저자, 서울대학교 경제학부 & 경제연구소 교수, e-mail: jihonglee@snu.ac.kr
*** 제1저자, 박사과정, Department of Economics, University of Wisconsin-Madison, e-mail: hyunkyeong.lim@wisc.edu
**** 공동저자, 서울대학교 경제학부 석사과정, e-mail: tippingpts@snu.ac.kr
***** 공동저자, 서울대학교 경제학부 석사과정, e-mail: sks1107@snu.ac.kr
****** 공동저자, 박사과정, Department of Economics, Michigan State University, e-mail: jungja10@msu.edu

Ⅰ. 서 론

특허 및 특허 인용에 관한 정보는 혁신 활동을 분석하는 데에 매우 유용한 도구로 경제, 경영, 재무 등 다양한 학문 분야에서 광범위하게 사용되고 있다. 특허 자료를 활용한 혁신 관련 연구는 최근 들어 Hall, Jaffe, and Trajtenberg (2001, 2005)의 선구적인 연구 결과의 산물인 NBER Patent Data Project (PDP)로 인해 비약적인 발전을 이루었다. NBER PDP는 Hall, Jaffe, and Trajtenberg (2001)의 방법을 토대로 1963-2006년 동안 미국 특허청 (US Patent and Trademark Office, USPTO)에 출원·등록된 특허에 대한 공공 데이터베이스를 구축하고, 이를 Compustat이 제공하는 기업 정보와 "매칭 (matching)"하였다.[1] 그 이후로 Thoma et al. (2010)은 유럽 특허에 대한 유사한 데이터베이스를 구축하였으며 Compustat-USPTO 매칭은 Lee and Lim (2019)에 의해 2017년도까지 연장되었다.[2]

Hall, Jaffe, and Trajtenberg (2005)는 Compustat-USPTO 매칭을 통해 관측이 가능해진 기업 간 혁신 활동의 편차를 이용하여 기업의 시장 가치 (토빈의 Q)와 특허 및 특허 피인용 스톡 (stock) 사이에 유의미한 양의 상관관계가 존재한다는 것을 밝혔다. 이러한 결과는 특허 등록 수와 피인용 횟수가 기업이 축적한 무형의 '지적 자산 (knowledge capital)'을 측정하기 위한 도구로 쓰일 수 있다는 점을 확인해 주었다. R&D 지출 또한 기업 가치에 영향을 미치는 중요한 요소이며 특허 출원과도 높은 상관관계에 있지만, 그 자체로는 기업의 혁신 활동의 성공 여부와 그 질적인 측면에 대해 어떠한 정보도 주지 못한다는 단점을 가지고 있다. 특허 자료는 혁신 활동의 양적·질적 결과에 관한 정보 이외에도 기술분류체계 (technology classification system), 출원인과 발명자 정보, 특허 인용 관계 등을 통해 다양한 측면에서 혁신 활동의 본질을 파악할 수 있는 중요한 실마리를 제공한다.

본 논문은 "한국 특허 데이터 프로젝트 (Korea Patent Data Project, KoPDP)"의 내용 및 방법에 대해 서술한다. 본 프로젝트는 1948-2016년에 걸쳐 한국 특허청 (Korea Intellectual Property Office, KIPO)에 출원된 모든 특허와 1976-2017년 동안 한국에 위치한 출원인 및 발명자가 미국 특허청에 출원·등록한 모든 실용 특허에 대한 정보를 수집하고, 출원인-기업명 매칭 (matching)을 통해 한국에서 각종 기업

1) https://www.nber.org/patents/.

2) 이외에도 Kogan et al. (2017)과 Autor et al. (2019) 등에서 NBER PDP의 Compustat-USPTO 매칭 결과를 연장하였다.

관련 정보를 제공하는 DataGuide 5.0과 연결하였다. 그 결과 14,803개의 상장 및 비상장 한국 기업의 재무 및 금융 정보를 비롯하여 해당 기업이 출원·등록한 한국 및 미국 특허 정보에 대한 포괄적인 데이터베이스를 구축하였다. 출원인과 기업 정보가 매치된 등록 특허 수는 한국의 경우 전체 표본 특허(총 1,551,653개)의 45% 이상이며, 미국의 경우 제1 출원인이 한국인 표본 특허(총 191,102개)의 약 87%에 달한다.

본 논문에서는 특허의 인용 정보를 비롯하여 표본 특허에 담겨 있는 주요 정보와 이를 토대로 산출된 기초 통계에 대해 설명한다. 본 프로젝트의 매칭 알고리즘(matching algorithm)과 사용 방법 또한 자세히 기술하였다. 기업명 표준화(name standardization) 및 문자열 알고리즘(string matching algorithm)은 NBER PDP, Thoma et al. (2010), Lee and Lim (2019)의 알고리즘을 참고하였으며, 주요 알고리즘은 모두 Python으로 구현하였다. 또한 KIPO와 USPTO에 포괄적으로 적용할 수 있는 특허의 산업 분류체계를 구축하였다. 특허의 산업 분류는 국제특허분류(International Patent Classification, IPC)를 바탕으로 하여 Schmoch et al. (2003)이 제시한 분류체계를 확장하였다.

KoPDP의 목적은 본 프로젝트가 구축한 데이터베이스와 방법이 관련 연구를 수행하고자 하는 연구자들에게 널리 사용되도록 함에 있다. Lee (2019)는 매치된 기업 및 개별(특허·기업) 데이터베이스에서 해당 기업을 식별할 수 있는 기업 고유의 코드에 대한 목록과 함께, KoPDP의 결과물을 반복 구축(replicate)하고 개선 및 연장하는 데에 필요한 관련 알고리즘과 공개 가능한 데이터를 모두 무료로 제공한다.[3] 이와 같은 데이터베이스의 구축 및 공유가 한국 경제의 혁신 활동을 평가하고 혁신에 영향을 미치는 기업 내외부적 요소들을 이해하는 데 도움이 될 것으로 기대한다. 최근 특허 자료를 사용하여 한국 경제의 혁신 역량을 진단하는 몇몇 연구가 이루어졌으나(Kwon, Lee and Lee, 2017; 이지홍·임현경·정대영, 2018) 기업 자료와 한국과 미국을 망라한 특허 자료를 함께 사용하여 혁신 활동을 분석한 연구는 아직 없는 것으로 보인다.

1990년대부터 한국 기업 및 개인의 특허 활동은 대내외적으로 급격히 증가하였다. 물론 모든 혁신 활동이 특허로 출원·등록되는 것은 아니다. 어떤 경우에는 혁신 활

3) Lee, J. (2019): "Korea Patent Data Project (KoPDP)," https://doi.org/10.7910/DVN/AUYERV, Harvard Dataverse. 데이터베이스의 사용 방법은 "Overview.pdf" 파일을 참조하기 바란다.

동의 결과가 영업 비밀(trade secret)로 유지되기도 하며, 경영 혁신과 같이 특허화(patentable)할 수 없는 혁신도 존재한다. 특허 제도 또한 완벽하다고 할 수 없다. 그러나 한국 경제가 점차 세계적으로 선도적인 위치에 도달하면서 새로운 지식과 기술을 대변하는 특허가 제공하는 정보의 중요성이 높아지고 있다. 특히 해외에서 출원·등록되는 특허는 한국 경제의 혁신 활동과 그 역량의 국제적 위치를 파악하는 데 유용한 정보를 제공한다.

본 프로젝트는 혁신 성장에 대한 관심이 그 어느 때보다 높은 현시점에 혁신에 관한 다양한 연구를 가능하게 할 것이다. 해외에서는 이미 NBER PDP가 구축한 자료를 사용하여 많은 영향력 있는 연구가 수행되었다. 거시 및 국제경제학(Jones, 2009; Akcigit, Celik, and Greenwood, 2016), 산업조직론(Bloom, Schankerman, and Van Reenen, 2013), 노동경제학(Kerr and Lincoln, 2010; Bell et al., 2018), 금융경제학(Aghion, Van Reenen, and Zingales, 2013; Tian and Wang, 2014), 도시경제학(Ellison, Glaeser, and Kerr, 2010), 불평등(Aghion et al., 2018) 등 거의 경제학 전 분야에 걸쳐 NBER PDP의 데이터베이스가 중요한 역할을 하고 있다. 이제는 한국 경제학계에서도 이러한 흐름에 주목할 때이다.

본 논문의 구성은 다음과 같다. 제Ⅱ절에서는 한국 특허청 특허 자료의 구성 및 데이터베이스 구축 방법에 대해 설명한다. 또한 데이터베이스에 구축된 변수들을 자세히 살펴봄으로써 각 변수가 어떻게 사용될 수 있는지 소개한다. 제Ⅲ절에서는 미국 특허청의 한국 특허 자료를 살펴본다. 데이터 수집 방식 및 구성은 한국 특허청 자료와 비슷하기 때문에 한국 특허청 특허 자료와의 차이점을 위주로 설명한다. 제Ⅳ절은 KoPDP에서 이용한 기업 정보 자료 DataGuide 5.0에 대해 소개한다. 특히 각 기업을 식별해주는 고유의 식별코드와 매칭 작업에 사용된 기업명 및 주소에 대해 자세히 살펴본다. 제Ⅴ절에서는 매칭 작업을 단계별로 구체적으로 살펴보고 그 결과를 소개한다. 제Ⅵ절은 특허의 산업을 어떻게 분류할지에 대해 논의한다. 본 논문에서는 각 특허의 국제산업분류(International Patent Classification, IPC)를 토대로 기존 NBER의 산업 분류에 따라 특허의 산업을 분류하는 방식을 제안하고 있다. 마지막으로 제Ⅶ절은 본 프로젝트에서 구축한 데이터베이스의 한계점 및 사용 시 유의 사항에 대해 다룬다.

Ⅱ. 한국 특허청 특허 자료

1. 자료의 출처·가공 및 구조

본 프로젝트는 한국특허정보원의 KIPRIS Plus Open API를 구매하여 한국 특허청(Korea Intellectual Property Office, KIPO)에 출원·등록된 특허의 서지 상세 정보, 인용 정보, 특허 패밀리 정보, 그리고 국가 R&D 사업자 정보를 수집 및 가공하였으며, 국제특허분류(International Patent Classification, IPC)를 토대로 각 특허의 산업을 분류하였다. 한국특허정보원은 KIPRIS Open API를 통해 특허 데이터에 접근할 수 있는 데이터 서버(request url)와 서버에 데이터를 호출하는 규칙을 설정할 수 있도록 한다.[4] 설정된 규칙에 따라 원하는 특허를 출원번호로 특정하여 각종 특허 정보를 호출할 수 있다. 본 프로젝트에서는 Python의 Request 패키지를 활용하여 연도별로 특허의 출원번호를 반복적으로 호출하여 데이터를 수집하였다.

그 결과 1948-2016년에 걸쳐 출원된 2,670,914개의 특허에 대한 데이터를 구축하였다. 출원된 특허가 공개되기까지는 약 2-3년의 시차가 존재하기 때문에, 2017-2018년에 출원된 특허 중 일부는 아직 공개되지 않아 본 프로젝트에서 구축한 데이터베이스에 포함하지 않았다. 그러나 등록 일자를 기준으로 할 경우 2017-2018년에 등록된 특허들은 데이터베이스에 등장한다. 출원된 특허 중 1,551,653개는 등록 특허에 해당하며, 등록 특허는 (1) 출원된 특허 중 등록번호가 존재하고 (2) 법적 상태가 '등록 결정(심사 전치 후)', '등록 결정(일반)', '등록 결정(재심사후)', '등록 결정(취소 환송 후)', '심판원에 의한 등록 결정', '이의신청 후 등록 유지'에 해당하는 것을 기준으로 추려내었다. 각 특허의 xml 파일로부터 추려낸 정보는 주제별로 'basicinfo.dta', 'assignee.dta', 'citation.dta', 'family.dta', 'invt_loc.dta', 'RND.dta', 'ipc.dta', 'industry.dta'로 정리하였다. 각 데이터 파일에 구축되어 있는 변수 목록은 부록에 정리하였다.[5]

4) 여기서 규칙이란 다양한 특허 정보 중 구체적으로 어떤 특허 정보를 호출할지를 결정하는 조건을 의미한다.

5) 저작권 관계로 KIPRIS에서 수집한 데이터는 Lee(2019)에서 제외되었다.

2. 특허의 날짜 정보 및 '출원-등록-공개'

문서화된 각 특허에는 발명자의 특허출원서가 접수된 날짜(출원 일자, application date)와 특허가 등록된 날짜(등록 일자, registered or grant date)가 부여되어 있다. 본 데이터는 1948년부터 출원된 모든 특허의 출원일 및 등록일 정보를 'basicinfo. dta'에 정리하였으며 각각을 STATA 날짜로 변환하여 정리하였다. 2, 670, 914개의 모든 출원 특허가 등록되는 것은 아니며 'basicinfo. dta'의 '등록 여부(regi)' 변수를 통해 해당 특허가 최종적으로 등록된 특허인지 식별할 수 있다.

실제 발명 시점은 특허의 등록 시점보다는 출원 시점에 가까울 것이다. 발명자들은 기술의 발명 직후 최대한 빠른 시일 내에 관련 기술에 대한 특허를 출원할 유인이 있기 때문이다. 반면 특허의 등록 시점은 특허청에서 검토하는 데에 걸리는 시간에 의존하게 된다. 평균적으로 출원된 특허의 등록 여부를 결정하는 데까지는 약 2. 75년이 소요되고 있다. 또한 특허의 출원이 공개되는 데까지도 시차가 존재한다. 한국 특허청에 출원된 특허는 출원 후 1년 6개월 혹은 등록 결정 시점 중 더 이른 쪽의 시점에 공개되고 있다.

위와 같은 이유로 특허 기술의 발명 시점을 지정할 때에는 출원 일자를 사용하는 것이 바람직할 수 있다(Griliches, Hall, and Pakes, 1991). 그러나 데이터의 단절(truncation) 문제로 인해 특허의 출원 일자를 사용하는 경우 관측치가 크게 줄어들 수 있다는 문제가 있다. 다시 말해, 최근에 가까울수록 출원-등록-공개까지 걸리는 시차로 인해, 출원되었으나 공개되지 않아 데이터베이스에 정보가 없는 경우가 많을 수 있다. 따라서 특허 기술의 발명 시점을 부여하는 데에 있어서 실제 발명 시점 및 데이터 단절 문제 등 다양한 이슈를 고려할 필요가 있다.

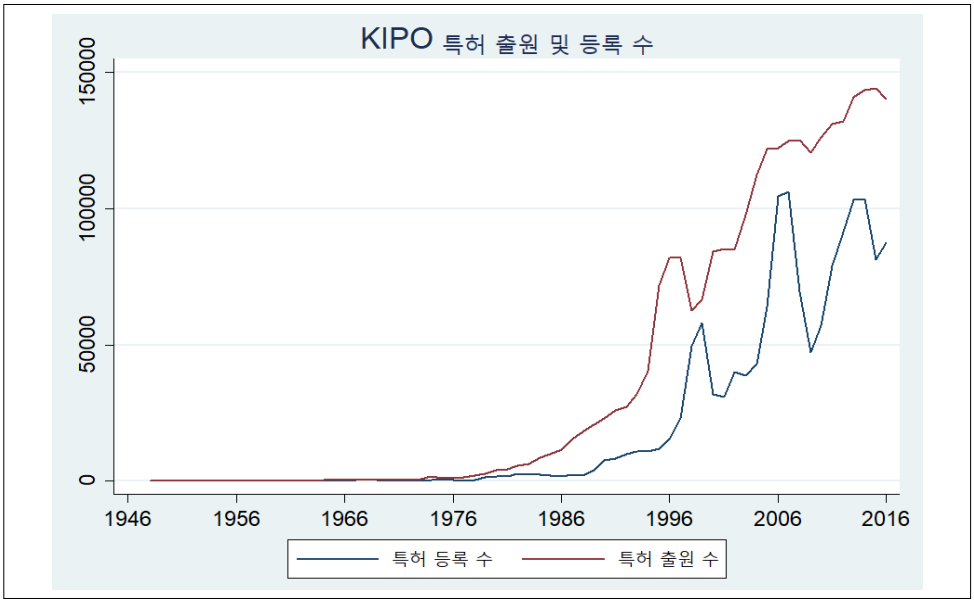
3. 특허 출원 및 등록 수

〈그림 1〉은 한국 특허청에 출원 및 등록된 특허 수를 연도별로 나타내고 있다. 데이터 단절 문제를 고려하여 2016년까지 출원된 특허만을 고려하였다. 앞서 언급했듯이 출원된 특허가 공개되기까지 최대 1년 6개월이 소요되며 등록되기까지는 평균적으로 2-3년이 소요되기 때문이다. 그래프에서 특허 등록 수의 그래프가 특허 출원 수의 그래프보다 오른쪽에 있다는 점을 통해 특허의 출원 시점부터 등록 시점까지 시차가 존재한다는 것을 확인할 수 있다. 또한 모든 특허가 등록되는 것이 아니므로 시차를

고려하더라도 등록 수가 출원 수보다 적은 것을 확인할 수 있다.

한국은 특히 1990년대 중후반에 접어들면서 급격히 특허의 출원 및 등록 숫자가 증가하고 있다. 특허 정보가 존재하는 시작 시점인 1948년부터 90년대 중후반까지의 출원 및 등록 수는 이후에 비하면 매우 미미한 수준이다. 90년대 중후반 이후에는 연도별로 다소 차이가 있으나 전반적인 추세는 꾸준히 증가하고 있다. 1995년에는 특허 출원 및 등록 수가 각각 71,862개, 11,821개였으나 2015년에는 각각 144,125개, 81,355개로 특허 출원 수는 6배 이상 증가하였다.

〈그림 1〉



4. 출원인 및 발명자 정보

한국 특허청에서는 각 출원인(assignee)별로 고유의 식별코드인 '특허 고객 번호(구 출원인코드)'를 부여하고 있다. 모든 출원인은 특허 출원 시 출원서에 특허 고객 번호를 기재해야 하며, 한 번이라도 한국 특허청에 출원한 기록이 있는 사람은 특허 고객 번호를 가지고 있다. 특허 고객 번호는 총 12자리이며 '출원인 종류(1자리)-등록 연도(4자리)-일련번호(6자리)-체크 Digit(1자리)'로 구성된다.[6] 각 출원인에 대

6) 특허청예규 제 93호 (시행 2016.10.4) 참조.

해서는 특허 고객 번호 외에도 이름과 국적, 주소에 대한 정보가 공개되며, 하나의 특허에는 여러 명의 출원인이 존재할 수 있다. 본 데이터베이스의 'assignee.dta'에는 특허의 출원번호별로 모든 출원인의 특허 고객 번호, 순번, 한글 이름, 영문 이름, 국적 그리고 주소에 대한 변수가 구축되어 있다. 이때 특허 고객 번호는 하나의 출원인에 하나의 번호가 부여되는 것이 원칙이나 사명 변경 등의 오류로 인해 하나의 출원인에 여러 개의 번호가 부여되는 경우도 존재하는 것으로 확인됐다. 이와 같은 문제는 매칭(matching) 과정(Ⅴ장 참조)에서 발견된 경우 하나의 특허 고객 번호로 통합했으며, 그 결과로 새 변수 'kiprisidH'를 구축하였다.

〈출원인 종류[7]〉

1: 국내 법인(국내 상법상 법인 - 2007. 7. 이전)	-54.44 %
2: 국가기관(국가기관 및 기타 법인 - 2007. 7. 이전)	6.88 %
3: 삭제됨(각급 시험 연구 기관 - 2007. 7. 이전)	5.40 %
4: 국내 자연인	17.13 %
5: 외국 법인, 외국 국가기관	15.66 %
6: 외국 자연인	0.47%
7: 법인이 아닌 사단·재단(2007. 7.부터 추가)	0%

전체 등록 특허의 99.98%는 출원인이 존재하며 출원인이 존재하지 않는 경우는 아직 기업, 대학, 정부 혹은 개인에게 특허권이 법적으로 귀속되지 않은 상태에 있는 경우에 해당한다. 이 특허들은 해당 특허를 발명한 발명자에게 소유권이 있는 상태이며 추후에 소유권을 다른 출원인에게 귀속시킬 수도 그렇지 않을 수도 있다. 본 데이터의 대부분의 특허는 국내 법인 기업에서 소유하고 있으며(54.44%), 15.66%는 외국 법인 기업 혹은 외국 국가기관에서 소유하고 있다. 17.13%는 국내 자연인에게 귀속되어 있는데 여기서 자연인은 법인으로 등록되지 않은 개인을 의미한다. 2007년 7월부터 법인이 아닌 사단 혹은 재단이 출원인 종류 중 하나로 추가되었으나 본 데이터에서 출원인 종류가 7번으로 분류되는 등록 특허는 존재하지 않았다.

각 특허의 법적 권리를 소유하는 출원인 외에 특허 기술 발명의 창작 행위에 직접 가담한 발명자에 관한 정보도 존재한다. 한국 특허청은 발명자에 대해서는 고유 식별

7) 각 %는 등록 특허 기준 각 종류의 출원인이 소지한 특허의 비중을 뜻한다.

번호를 부여하지 않는다. 다만 발명자의 이름과 국적 그리고 주소에 대한 정보는 공개된다. 출원인과 마찬가지로 하나의 특허에 여러 명의 발명자가 존재할 수 있다. 본 데이터베이스의 'invt_loc.dta'에는 각 특허의 출원번호별로 발명자의 순번과 주소가 정리되어 있으며, 주소에서 광역시·도 단위의 주소를 추출하여 'dist'라는 변수로 따로 정리해두었다.

한편, 출원인과 발명자는 일치하지 않을 수 있으며, 따라서 출원인과 발명자의 주소 역시 일치하지 않을 수 있다. 일례로 〈그림 2〉의 특허 출원번호 1020060040692 ("이동통신 단말기에서 전자사전을 이용한 단어 검색 방법 및 장치")의 경우 출원인은 '삼성전자주식회사'에 해당하지만 발명자의 경우 총 4명이 있으며 출원인과 4명의 발명자의 지리 정보는 모두 일치하지 않는다. 한국 특허청은 발명자에 대해서는 고유의 식별 번호를 부여하지 않는다.

〈그림 2〉

이동통신 단말기에서 전자사전을 이용한 단어검색 방법 및장치
Apparatus and Method For Searching Words WithElectronic Dictionary In The Mobile Station

상세정보 | 공개전문 | 공고전문 | 기타공고 | 등록사항 | 통합행정정보

서지정보 | 인명정보 | 행정처리 | 청구항 | 지정국 | 인용/피인용 | 패밀리정보 | 국가R&D연구정보

(51) Int. Cl. H04B 1/40(2015.01.01)
(52) CPC
(21) 출원번호/일자 1020060040692 (2006.05.04)
(71) 출원인 삼성전자주식회사
(11) 등록번호/일자 1008089910000 (2008.02.25)

다운로드 | 크게보기

▸ (71) 출원인

번호	이름	국적	주소
1	삼성전자주식회사 SAMSUNG ELECTRONICS CO., LTD. (119981042713)	대한민국	경기도 수원시 영통구...

▸ (72) 발명자

번호	이름	국적	주소
1	이석곤	대한민국	경북 구미시 ...
2	손재곤	대한민국	대구 북구...
3	김기태	대한민국	경북 구미시 ...
4	한용희	대한민국	대구 북구...

5. 인용 정보

특허의 인용 정보는 발명된 기술들 사이의 관계를 보여주는 중요한 정보를 제공한다. 예컨대 특허 B가 특허 A를 인용했다는 것은, 특허 A가 특허 B를 발명하는 데에 필수적인 선행 기술이라는 것을 의미한다. 동시에 특허 B가 특허 A에 관한 기술에는 독점적 권리를 행사할 수 없음을 의미한다. 출원인은 특허를 출원하는 시점에 선행 기술에 대한 모든 지식을 명시할 법적 의무가 있다. 그러나 어떠한 기술을 관련 선행 기술로 볼 것인지는 해당 기술 분야의 전문가로서 해당 특허와 실질적으로 관련이 있는 선행 기술 특허를 추가 또는 첨가할 능력이 있는 심사관의 심사를 거쳐 결정된다.

따라서 특허의 인용은 다음과 같은 맥락에서 특허 기술들 사이의 연결 고리를 파악하는 정보를 제공한다고 할 수 있다. 첫째로, 특허의 인용 정보는 기술 및 지식의 전파 양상을 파악하는 데에 중요한 자료를 제공한다. 즉, 특허 B가 특허 A를 인용했다는 것을 특허 A와 관련된 지식이 특허 B로 전파되었다고 보는 것이다(Jaffe, Trajtenberg, and Henderson, 1993; Caballero and Jaffe, 1993). 둘째로, 특허의 피인용 횟수는 해당 특허가 해당 기술 분야에서 얼마나 중요한 위치에 있는지 혹은 얼마나 가치 있는 기술인지를 보여주는 척도로 사용될 수 있다(Trajtenberg, 1990; Lanjouw and Schankerman, 2004; Hall, Jaffe, and Trajtenberg, 2005; Kwon, Lee and Lee, 2017). 이 밖에도 Hall, Jaffe, and Trajtenberg (2001) 에서는 특허의 인용 정보를 토대로 각 특허의 고유성(originality), 일반성(generality) 등을 측정하는 지수를 고안하였다.

한국 특허청에서는 특허의 인용 정보를 매우 세분화하여 제공하고 있다. 특허의 인용이 어떤 단계에서 발생한 것인지 혹은 인용 및 피인용된 특허가 등록된 특허인지 등의 여부를 다음과 같은 인용 코드와 표준 인용 식별코드로 구분하고 있다.

예를 들어, 인용 정보 중 심사관이 인용한 등록 특허만을 보고 싶은 경우, 피인용 문헌 인용 구분 코드(cited_typecode)가 'E0802'이고 표준 인용 식별코드(cited_code)가 'B1'인 경우를 보면 된다. 심사관 인용은 1992년부터, 출원인 인용은 2010년부터 법적으로 제도화되었으며, 이 때문에 법제화 이전에는 관련이 있는 특허의 인용이 누락되어 있을 가능성이 존재한다. 심사관 인용은 특히 특허 기술들의 선행 관계 및 관련성에 있어 정확도가 높다고 할 수 있다. 한편, 출원인 인용은 특허 기술의 발명 과정에서 출원인이 어떠한 선행 기술들로부터 영향을 받았는지에 대한 정보를 제공한다. 따라서 특허의 질을 측정하는 대리 변수로는 심사관 인용을 사용하

고 지식의 전파 측면을 살펴볼 때에는 출원인 인용을 사용하는 것이 보다 적절할 것이다.

〈표 1〉 인용구분

인용 구분	코드	설명
발송문서	E0801	심사관이 발송 문서(의견제출통지서, 거절 결정서) 작성 시 인용한 건
선행기술조사문헌(=심사관인용)	E0802	심사관이 출원서, 선행기술조사보고서, 의견제출 통지서 등의 인용 문헌 정보를 바탕으로 심사 보고서에 최종적으로 선택한 건
선행기술조사보고서	E0805	선행기술조사 전문 기관(특허정보진흥센터, 윕스 등)에서 작성한 선행기술조사보고서에서 인용된 경우
출원서인용문헌이력정보(=출원인인용)	E0806	출원인이 출원서에 기재한 선행 기술 문헌 정보(출원 시점에 작성)

〈표 2〉 표준 인용 식별코드

표준 인용 식별코드	구분	피인용 문헌 번호
A	특허 공개	공개 번호
B1	특허등록	등록번호(마지막 네 자리가 '0000'인 경우) 또는 공고 번호
U	실용 공개	공개 번호
Y1	실용 등록	등록번호 또는 공고 번호

(1) 평균 인용·피인용 수

연도별 평균 인용(피인용) 수는 2002-2016년에 등록된 특허를 대상으로 당해 연도에 출원된 각 등록 특허가 인용한(피인용된) 특허의 수를 연도별로 평균 낸 것이다.

C_{it}: t년도에 출원된 등록특허 i가 인용한/피인용된 특허 수

N_t: t년도에 출원된 등록특허 수

$$AvgCit_t = \sum_i C_{it}/N_t$$

산업별 평균 인용·피인용 수는 마찬가지로 2002-2016년에 등록된 특허를 대상으로 2002-2016년 동안 각 산업에 속하는 특허들의 평균적인 인용 혹은 피인용 수를 의

미한다.[8]

C_{ij}: 산업 j에 속하는 등록특허 i가 인용한/피인용된 특허 수
N_t: 산업 j에 2002－2016년 기간 동안 출원된 등록특허 수
$AvgCit_j = \sum_i C_{ij}/N_j$

〈그림 3〉과 〈그림 4〉는 각각 KIPO의 연도별 평균 인용・피인용 수와 산업별 평균 인용・피인용 수를 보여준다. 연도별 '평균 피인용(전체)' 수는 점차 감소하고 있으며 연도별 '평균 인용(전체)' 수는 점차 증가하고 있다.[9] 산업별 평균 인용・피인용 수에서는 화학과 기타 산업의 평균 인용 수 및 화학과 정보통신 산업의 피인용 수가 다소 높은 것으로 나타났다.

연도별 평균 인용・피인용 수는 데이터 단절의 문제로 인해 왜곡이 될 수 있다는 점에 주의해야 한다. 〈그림 3〉에서 '평균 피인용(전체)'은 각 특허가 출원된 시점부터 2016년까지 인용을 받은 총 횟수를 가지고 연도별 평균을 구한 것이다. 이때 평균 피인용 수는 시간이 흐름에 따라 점차 줄어들고 있는 양상을 보이며 특히 최근에 가까울수록 급격히 감소하고 있다. 하지만 과거에 출원된 특허일수록 인용 받을 수 있는 기간이 길다는 점과 최근의 특허 중에는 누락된 등록 특허가 존재할 수 있다는 점을 고려하면 이러한 감소 추세가 왜곡되었을 가능성을 추론할 수 있다.

이와 같은 문제를 해결하기 위해 '평균 피인용(2년)'은 각 특허가 출원 후 2년 동안 받은 인용 횟수를 기준으로 연도별 평균을 구해보았다. 즉 위의 C_{it}가 t년도에 출원된 등록 특허 i가 2년 동안 받은 인용 숫자가 되는 것이다. 〈그림 3〉에서 '평균 피인용(2년)'은 가장 최근에는 오히려 증가하는 양상을 보이고 있다. 다만 2016년에는 다시 감소하는데, 이는 누락된 등록 특허 때문인 것으로 추정된다.

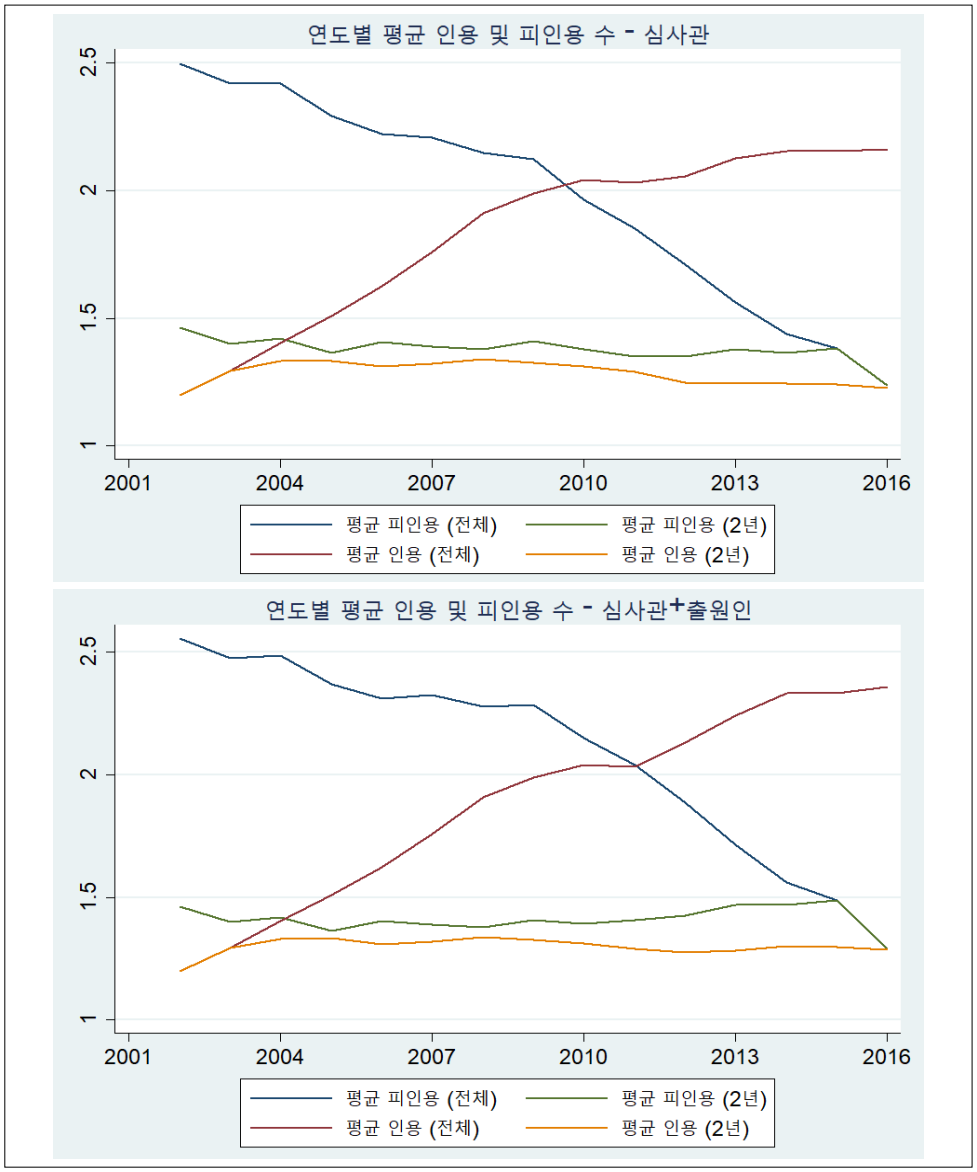
마찬가지로 연도별 평균 인용 수의 경우에도 인용 가능 기간을 각 특허가 출원된 시점으로부터 2년 전에 출원된 특허로 제한하면 연도별로 증가하는 추세가 사라진다. 인용 기간을 제한하기 전에 연도별로 증가하는 추세가 나타나는 것은 시간이 흐를수

8) 산업 분류에 대해서는 VI절에서 기술한다.

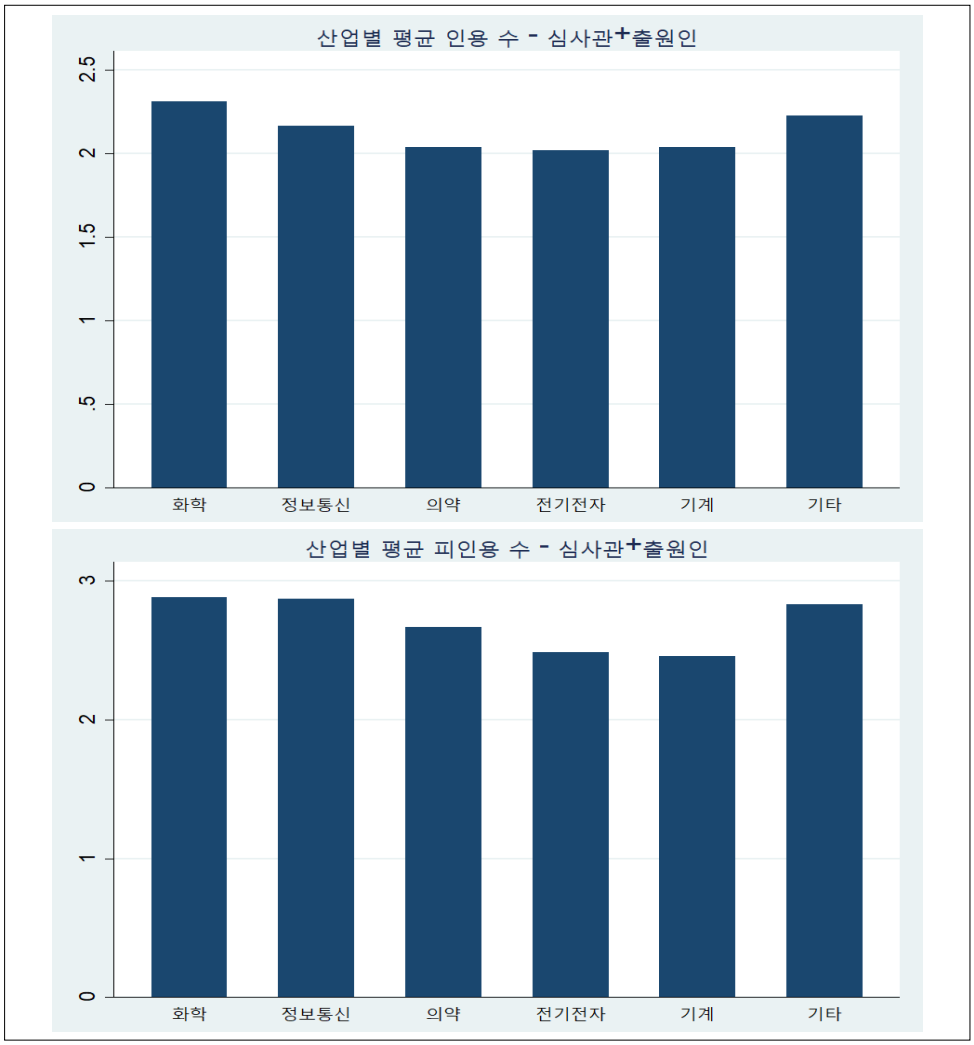
9) 피인용 수의 분포는 높은 비대칭도(skewness)를 나타낸다. 5년 단위로 계산한 결과, 2002-2006년에 등록된 특허들의 피인용 횟수 평균은 2, 중간값은 2.96, 2007-2011년에는 평균 2, 중간값 2.60, 그리고 2012-2016년에는 평균 1, 중간값 1.97이었다.

록 많은 특허가 출원되고 또 축적된 특허의 수가 증가하기 때문에 인용할 수 있는 특허의 수가 증가하면서 나타나는 자연스러운 현상이다. 이처럼 인용 가능 기간을 제한하는 방법 외에도 피인용 횟수의 왜곡을 해결하기 위한 다양한 방법이 존재하는데, 아래 Ⅷ절 1항을 참조하기 바란다.

〈그림 3〉



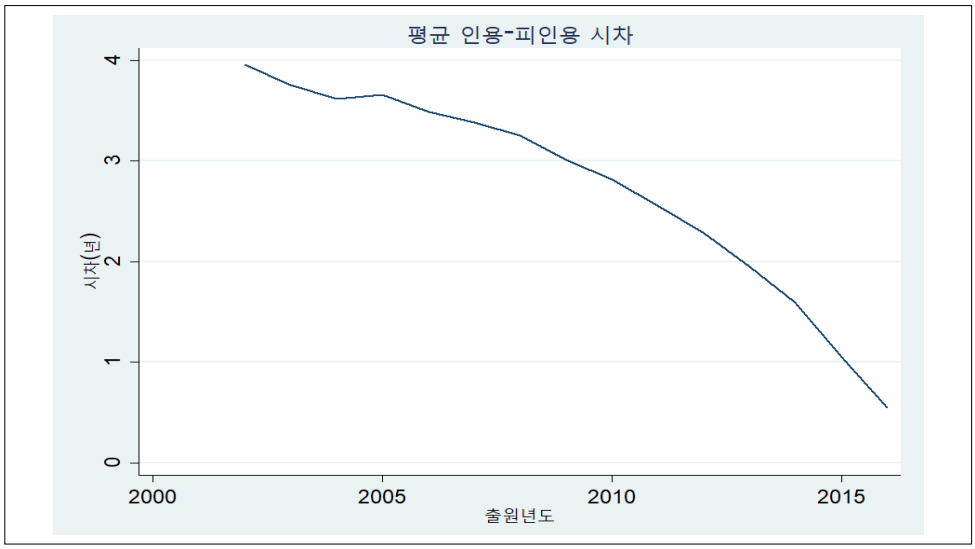
〈그림 4〉



(2) 평균 인용-피인용 시차

〈그림 5〉는 평균 인용-피인용 시차를 피인용된 특허의 출원 연도에 따라 보여주고 있다. 여기서 인용-피인용 시차는 2002년부터 2016년까지 등록된 특허 중 피인용된 특허가 처음으로 인용되기까지 소요되는 시간을 출원 연도를 기준으로 계산한 것이다. 평균 인용-피인용 시차는 각 특허의 시차를 출원 연도별로 평균 낸 것이다. 평균 인용-피인용을 계산하는 데에는 심사관과 출원인 인용을 모두 대상으로 하였다. 연도별 평균 인용-피인용 시차는 감소하고 있다.

〈그림 5〉



(3) 자기 인용

특허의 자기 인용은 통상적으로 출원인 본인이 자신의 특허를 직접 인용하는 경우를 뜻한 다.[10] 특허의 자기 인용은 지식의 전파보다는 해당 기술을 보다 더 깊이 있게 습득했다는 것(appropriation)을 의미할 가능성이 높다. 이러한 이유로 지식의 전파를 연구할 때에는 보통 자기 인용을 제외하고 다른 출원인의 특허 인용만을 사용하고 있다.

KIPO 특허의 경우 출원인 각자에게 고유하게 부여되는 특허 고객 번호로 자기 인용을 식별할 수 있다.[11] 즉 인용-피인용 관계에 있는 두 특허의 출원인 특허 고객 번호가 일치한다면 이는 자기 인용이 발생한 것으로 식별하는 것이다.

〈그림 6〉은 KIPO의 연도별 평균 자기 인용·피인용 비율과 산업별 자기 인용 비율을 보여주고 있다. 연도별 평균 자기 인용·피인용 비율은 해당 연도에 출원된 특허를 대상으로 각 출원인이 인용하거나 피인용된 특허의 수 중 자기 인용이 차지하는 비율을 구하여 연도별로 평균을 구한 것이다. 시간이 흐를수록 피인용 특허 중에서는

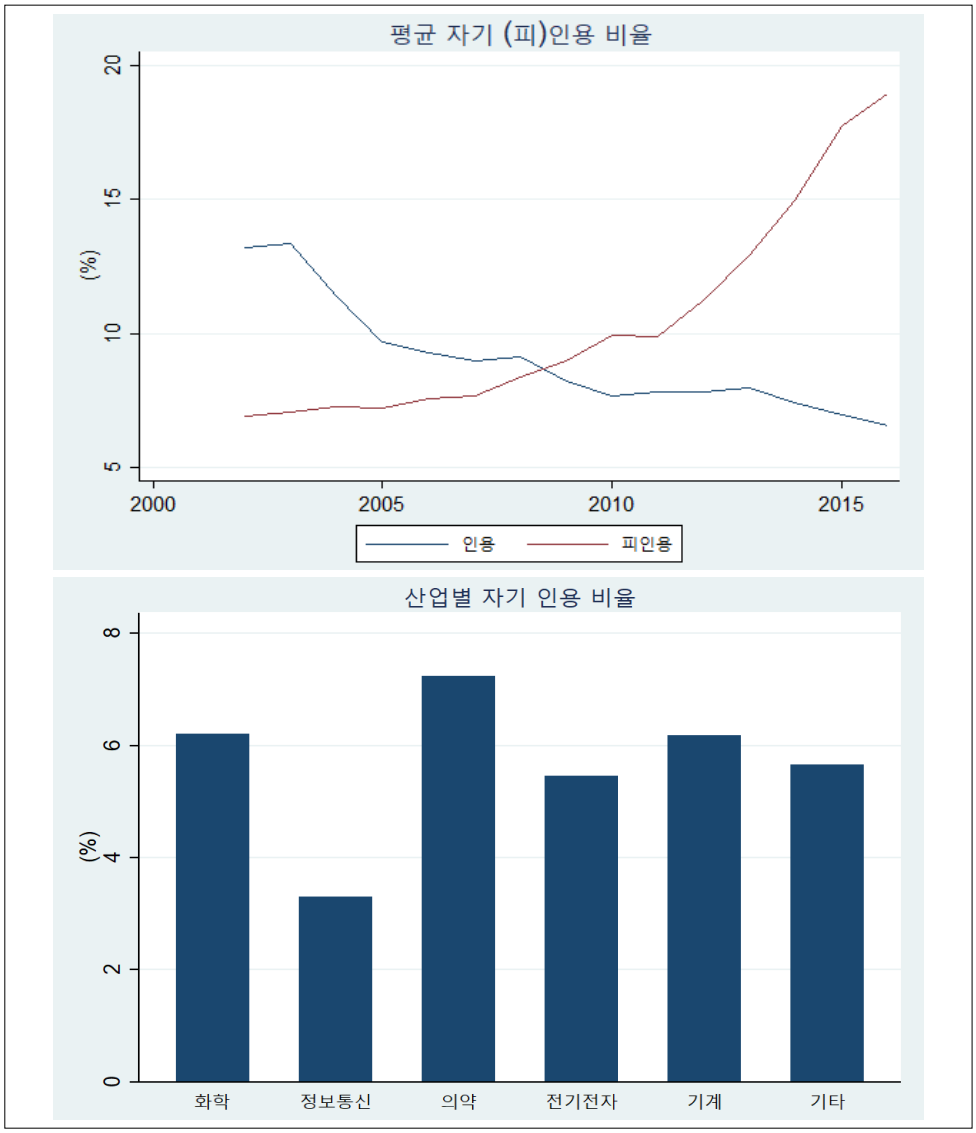
10) 이때, 특허의 출원인과 발명자는 서로 다를 수 있기 때문에 출원인을 기준으로 자기 인용을 정의하는 것과 발명자를 기준으로 자기 인용을 정의하는 것에는 차이가 발생할 수 있다.

11) 그러나 특허 고객 번호가 다르다고 해서 반드시 다른 기업 혹은 다른 재무제표를 보고하는 기업으로 보는 데에는 어려움이 있다(Ⅶ절 3항 참조).

자기 인용의 비율이 증가하고 있으며 인용 특허 중에서는 자기 인용의 비율이 감소하고 있다.

한편, 산업별 자기 인용 비율은 해당 산업에 속하는 특허들이 2002-2016년 동안 인용하거나 피인용된 특허의 수 중 자기 인용에 해당하는 특허의 비율을 구한 것이다. 흥미롭게도 의약 분야의 자기 인용 비율이 다른 산업 분야에 비해 다소 높은 것을 확인할 수 있다.

〈그림 6〉



6. 특허 분류

한국 특허청에 출원된 특허는 국제특허분류(International Patent Classification, IPC)에 따라 분류되어 있다. IPC는 1968년에 처음 도입된 특허 분류체계이며, 한국에서는 1981년 6월부터 사용하기 시작하였다. 국제특허분류는 "섹션(Section)-클래스(Class)-서브 클래스(Subclass)-메인 그룹(Maingroup)-서브 그룹(Subgroup)"으로 이루어져 있으며 가장 대분류에 해당하는 섹션은 총 8개로 이루어져 있다. 예를 들어, Figure 2의 특허 출원 번호 1020060040692의 IPC는 H04B 1/40이며 H는 섹션, 04는 클래스, B는 서브 클래스, 1은 메인 그룹, 40은 서브 그룹을 의미한다. 하나의 특허에는 여러 개의 IPC가 부여될 수 있으며 한국 특허청에서는 이 중 특별히 중심이 되는 IPC를 구분하고 있지 않다.

국제특허분류는 급변하는 기술 분야의 체계를 반영하기 위해 매년 새로운 분류가 생기거나 기존의 분류가 없어지면서 업데이트가 되고 있다. 1968년에 처음 국제특허분류가 등장했을 때에는 8개의 섹션, 103개의 클래스, 594개의 서브 클래스가 있었으나 2006년에 8판이 등장했을 무렵에는 8개의 섹션 129개의 클래스, 639개의 서브 클래스로 업데이트되었다. 본 데이터베이스에는 출원 시점의 IPC에 대한 정보를 구축하였다.

- 섹션: 총 8개의 섹션으로 대문자 A-H 중 1개로 표시
  1. A - 생활필수품
  2. B - 처리 조작; 운수
  3. C - 화학; 야금
  4. D - 섬유, 지류
  5. E - 고정 구조물
  6. F - 기계공학; 조명; 가열; 무기; 폭파
  7. G - 물리학
  8. H - 전기
- 클래스: 섹션을 세분화한 것으로 2자리의 숫자로 표시
- 서브 클래스: 각 클래스는 1개 이상의 서브 클래스를 포함하며 1자리의 대문자로 표시
- 메인 그룹: 서브 클래스를 세분화한 것으로 1-3자리의 숫자로 구성되며 사선(/)

과 숫자 '00'을 뒤에 붙임

- 서브 그룹: 메인 그룹을 세분화한 것으로 메인 그룹 및 사선(/) 이후에 '00' 이외의 다른 숫자를 1-3자리 붙임

IPC는 한국 특허에 대해서만 부여되는 것이 아니라 전 세계적으로 공용되고 있다는 점에서 다른 국가의 특허 데이터와 쉽게 연계할 수 있다는 장점이 있다. 본 프로젝트에서도 미국 특허청에 출원된 특허들의 IPC 정보를 구축하였으며, 따라서 한국 특허청의 특허와 미국 특허청의 특허 모두 통일성 있게 IPC를 토대로 분류할 수 있다.

한편 NBER Patent Data Project에서는 미국특허분류(United States Patent Classification, USPC)에 따라 특허를 분류하고 있으며 400개 이상의 메인 USPC를 다시 6개의 카테고리(category)와 44개의 서브 카테고리(sub-category)로 묶어 특허의 산업을 분류하고 있다. 미국 특허청 특허 데이터를 기반으로 진행된 기존의 많은 연구들이 USPC를 토대로 한 특허의 산업분류를 많이 이용하고 있기 때문에, 본 프로젝트에서도 서브 클래스 수준의 IPC를 NBER의 44개 서브 카테고리에 대응시켜 기존 문헌과의 연계성을 높이고자 하였다. 특허의 산업 분류와 관련된 자세한 내용은 VI절을 참조하기 바란다.

7. 특허 패밀리

1국 1특허의 원칙에 따라 해외에서 특허권을 획득하기 위해서는 별도로 해외 출원이 필요하며, 해외 출원을 하는 방법에는 전통적인 출원 방법과 Patent Cooperation Treaty(PCT) 국제출원 방법이 있다. PCT 출원은 국적국 또는 거주국의 특허청에 하나의 출원서를 제출함으로써 그로부터 정해진 기간 이내에 특허 획득을 원하는 국가로부터 PCT 국제 출원의 출원일을 인정받을 수 있도록 하는 제도이다. 다만 기존에 출원된 특허에 대한 우선권을 주장하는 목적으로 출원하는 경우, 기존 출원 특허의 출원일로부터 12개월 이내에 PCT 국제 출원을 하여야 우선권 주장을 인정받을 수 있다.[12]

패밀리 특허는 공동 우선권을 기초로 하는, 개별 국가 특허청 혹은 PCT 국제 출원

12) 출처: http://www.kipo.go.kr/kpo/user.tdf?a=user.pct.info.BoardApp&c=1001&catmenu=m08_ 01_01.

되고, 공개 또는 등록된, 발명의 내용이 일치하거나 기본적으로 일치하는, 한 무리의 특허를 말한다. 한국 특허청은 한국 특허청에서 정리하는 'Family' 특허 패밀리와 유럽 특허청에서 정리하는 'DOCDB' 특허 패밀리 정보를 모두 제공한다. 본 데이터베이스에도 한국 특허청의 패밀리 특허 정보와 유럽 특허청의 DOCDB 패밀리 정보가 모두 구축되어 있다. 각 특허별 특허 패밀리 문헌의 종류는 '문헌 코드(litkind)' 변수를 통해 확인할 수 있으며, 주로 "B, B1, B2"가 등록된 특허에 해당한다.[13]

각 특허별 패밀리의 크기는 해당 특허의 질을 측정하는 도구로 사용되기도 한다. 특허 패밀리의 크기가 크다는 것은 그만큼 여러 국가에 특허를 출원 및 등록할 유인이 있는 특허라는 것을 의미하기 때문이다(Lanjouw and Schankerman, 2004). 본 데이터베이스의 경우 전체 등록 특허 중 약 33.26% 가 특허 패밀리를 보유하고 있으며 평균 약 3.36개의 특허 패밀리를 보유하고 있다.[14] 주요 국가별로는 미국 특허 평균 1.82개, 유럽 특허 평균 1.53개, 일본 특허 평균 1.43개를 보유하고 있다.[15]

Ⅲ. 미국 특허청 한국 특허 자료

1. 자료의 출처·가공 및 구조

미국 특허청(USPTO)에서는 특허 정보가 담긴 Bulk Data를 무료로 제공한다.[16] 본 프로젝트에서는 XML 형태로 된 Bulk Data를 다운로드한 후 Python의 'dom.minidom' 패키지를 이용하여 XML 파일을 파싱(parsing)함으로써 1976-2017년 동안 미국 특허청에 등록된 모든 실용특허 데이터에 대한 정보를 구축하였다.

이 중에서 '첫 번째 출원인'의 국가 정보가 "KR"(한국)에 해당하는 특허를 추려내어 특허 출원번호, 출원일 및 등록일, 그리고 출원인 정보를 정리하였다

13) 다음의 링크를 통해 국가별 문헌 코드를 확인할 수 있다: http://abpat.kipris.or.kr/abpat/biblio/po/POCD1010.jsp?cntry=us.

14) 한국 특허청 특허 패밀리를 단 한 개라도 보유하는 경우를 대상으로 하며, 문헌 코드가 "B" 또는 "B1" 또는 "B2"인 경우에 대해 계산하였다. 미국의 경우 2001년 이전까지는 등록된 패밀리 특허에 대해 "A"라는 문헌 코드를 부여하였기 때문에 미국에 한하여 "A"까지 고려하여 계산하였다.

15) 미국 특허는 US Patent and Trademark Office(USPTO), 유럽 특허는 European Patent Office(EPO), 그리고 일본 특허는 Japan Patent Office(JPO)에서 우선권을 인정하는 패밀리 특허를 의미한다.

16) https://www.uspto.gov/learning-and-resources/bulk-data-products.

(basicinfo_uspto_kr, dta, assignee_uspto_kr. dta). 1976-2017년 동안 총 191, 102개의 특허가 한국 출원인에 의해 미국 특허청에 등록되었다. 한편, 2002-2017년 동안 미국 특허청에 등록된 특허 중 '첫 번째 발명자'의 국가 정보가 "KR"(한국)인 경우를 추려내어 이들의 특허 출원번호, 발명자 이름 및 주소 등을 정리하였다 (invt_loc_uspto_kr. dta).[17] 2002-2017년 동안 총 171, 810개의 특허가 한국 발명자에 의해 미국 특허청에 등록되었다.

미국 특허 데이터의 경우에는 등록된 특허의 데이터만 존재한다. 다시 말해 출원되었으나 등록되지 못한 특허의 데이터는 본 데이터베이스에 잡히지 않는다. 미국 특허청의 전체 실용특허에 대한 정보는 'basicinfo_uspto. dta', 'assignee_uspto. dta', 'citation_uspto. dta', 'ipc_uspto. dta'로 정리하였으며 미국 특허청에 등록된 한국 실용특허(제1 출원인 혹은 제1 발명자 국가 정보 기준)의 정보는 'basicinfo_uspto_kr. dta', 'assignee_uspto_kr. dta', 'invt_loc_uspto_kr. dta'에 따로 추려내어 정리하였다.

2. 특허의 날짜 정보 및 출원-등록 시차

한국 특허청 특허 자료와 마찬가지로, 미국 특허청에 등록된 특허도 특허마다 출원일자와 등록 일자가 부여되어 있다. 1976년부터 미국 특허청에 등록된 모든 실용특허의 출원일 및 등록일 정보를 'basicinfo_uspto. dta'에 정리하였으며, 마찬가지로 각각을 STATA 날짜로 변환하였다. 이 중 첫 번째 출원인의 국가 정보가 한국인 경우에는 'basicinfo_uspto_kr. dta'에 따로 정리되어 있다. 한국 특허청 특허 자료와의 차이점은 등록된 특허에 대한 자료만 존재한다는 점이다. 미국 특허청의 경우에도 출원 후 등록까지 약 2년 정도의 심사 과정을 거쳐야 하기 때문에 출원과 등록 사이에 시차가 발생하게 된다. 따라서 한국 특허청 특허 데이터와 동일한 이유로 데이터 단절 문제가 발생할 수 있다.

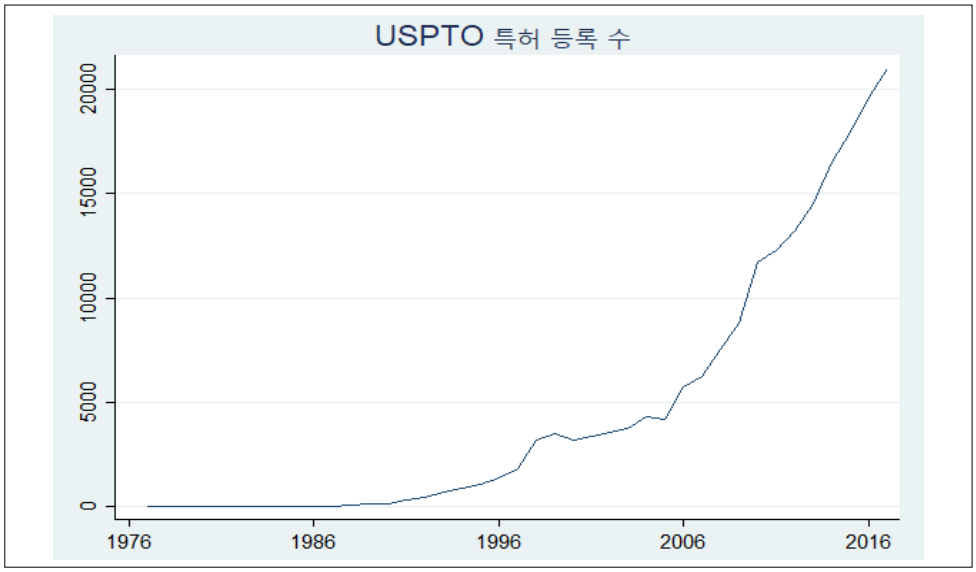
3. 특허 등록 수

〈그림 7〉은 미국 특허청에 등록된 실용특허 중 첫 번째 발명자의 국가 정보가 한국인 특허의 수를 등록일을 기준으로 하여 연도별로 나타내고 있다. 미국에 등록되

17) 2002년 이전에는 발명자의 지역 정보에 오류가 많아 2002년부터 첫 번째 발명자의 국가가 "KR"인 특허를 추려내고 이들의 주소를 광역자치단체 수준에서 정리하였다.

고 있는 한국 출원인의 특허 수는 1976년부터 현재까지 지속적으로 증가하고 있다. 1995년에는 등록 특허가 1,085개이었는데 2015년에는 17,929개로 15배 이상 증가하였다.

〈그림 7〉



4. 출원인 및 발명자 정보

미국 특허청에서는 출원인별 고유의 식별코드를 부여하고 있지 않으며, 출원인이 직접 기입한 이름만을 제공하고 있다. 그러나 출원인의 이름 정보에 오타와 변형이 많이 존재하기 때문에 출원인의 이름만을 토대로 출원인을 식별하는 데에는 어려움이 있다. Lee and Lim (2019)에서는 미국 특허청의 모든 실용특허 출원인을 대상으로 이름을 정리한 후, 이를 토대로 출원인 고유의 식별코드를 부여하였다. 특히 미국 기업에 대해서는 미국 기업 데이터와의 매칭(matching) 작업을 통해 추가적으로 식별코드를 정리하였다.

한편, 본 프로젝트에서는 첫 번째 출원인의 국가 정보가 "KR"인 191,102개의 특허를 추려서 출원인의 이름을 정리한 후 이를 토대로 출원인 고유의 식별코드를 부여하였다.[18] Lee and Lim (2019)의 출원인 식별코드는 4-8자리의 숫자로 이루어져 있으

18) 자세한 내용은 V절 특허 자료와 기업 정보의 매칭 참조.

며 본 프로젝트에서 부여한 출원인 식별코드는 총 4자리의 숫자로 이루어져 있다.[19] 출원인의 종류에 대해서는 미국 특허청에서 아래와 같이 코드를 부여하고 있다.

- 1: 출원인 정보 없음
- 2: 미국 비정부 기관(대부분 기업)
- 3: 미국 외 비정부 기관(대부분 기업)
- 4: 미국 자연인
- 5: 미국 외 자연인
- 6: 미국 연방 정부
- 7: 미국 외 정부 기관

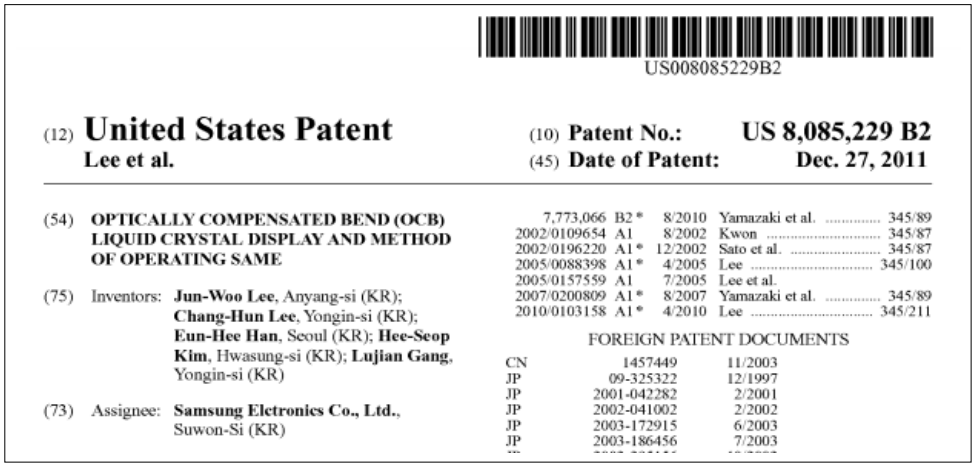
발명자에 대해서는 한국 특허청과 마찬가지로 고유의 식별 번호가 존재하지 않으며, 미국 특허청 특허의 경우에도 발명자와 출원인의 주소는 다를 수 있다. 본 프로젝트에서는 2002-2017년 동안 첫 번째 발명자의 국가 정보가 "KR"인 171,180개의 특허를 추려 광역자치단체 수준에서 주소 정보를 정리하였다. 첫 번째 발명자의 국가 정보가 "KR"에 해당하는 특허 중에서는 164,101개의 특허의 첫 번째 출원인 국가 정보가 "KR"에 해당한다. 〈그림 8〉의 예시에서 출원번호 8085229 특허의 출원인은 'Samsung Electronics Co. Ltd'로 주소는 'Suwon-Si'로 되어있지만 5명의 발명자의 주소는 국적은 동일하나 광역자치단체 수준에서 모두 다르다.

한국 특허청 특허 자료와는 달리, 미국 특허청 특허 자료의 경우 심사관 인용과 출원인 인용의 구분이 어려운 상태이다. 따라서 본 데이터의 USPTO 특허 인용 정보는 출원인과 심사관 인용을 구분 없이 모두 포함하고 있다. 또한 USPTO 특허 인용 정보는 모든 등록 특허에 대한 인용 및 피인용 관계만을 파악할 수 있다. 'citation_uspto.dta'에는 1976-2017년 동안 미국특허청에 등록된 모든 실용특허의 출원번호별로 해당 특허가 인용한 모든 USPTO 실용특허의 출원번호가 정리되어 있으며 이 중 첫 번째 출원인 혹은 발명자가 "KR"에 해당하는 경우를 식별함으로써 한국 특허를 추려낼 수 있다.

---

19) Lee and Lim(2019)의 출원인 식별코드에 대해서는 Lee and Lim(2019) 부록(A.5) 참조. Lee and Lim(2019)에서 첫 번째 출원인 국가 정보가 한국인 기업에 대한 식별코드는 본 프로젝트의 식별코드와 일치하지 않으며, 한국 기업에 대해서는 추가적인 정리 작업이 수행되지 않았기 때문에 본 프로젝트의 식별코드에 비해 부정확하다.

〈그림 8〉



US008085229B2

(12) **United States Patent**
Lee et al.

(10) **Patent No.: US 8,085,229 B2**
(45) **Date of Patent: Dec. 27, 2011**

(54) **OPTICALLY COMPENSATED BEND (OCB) LIQUID CRYSTAL DISPLAY AND METHOD OF OPERATING SAME**

(75) Inventors: **Jun-Woo Lee**, Anyang-si (KR); **Chang-Hun Lee**, Yongin-si (KR); **Eun-Hee Han**, Seoul (KR); **Hee-Seop Kim**, Hwasung-si (KR); **Lujian Gang**, Yongin-si (KR)

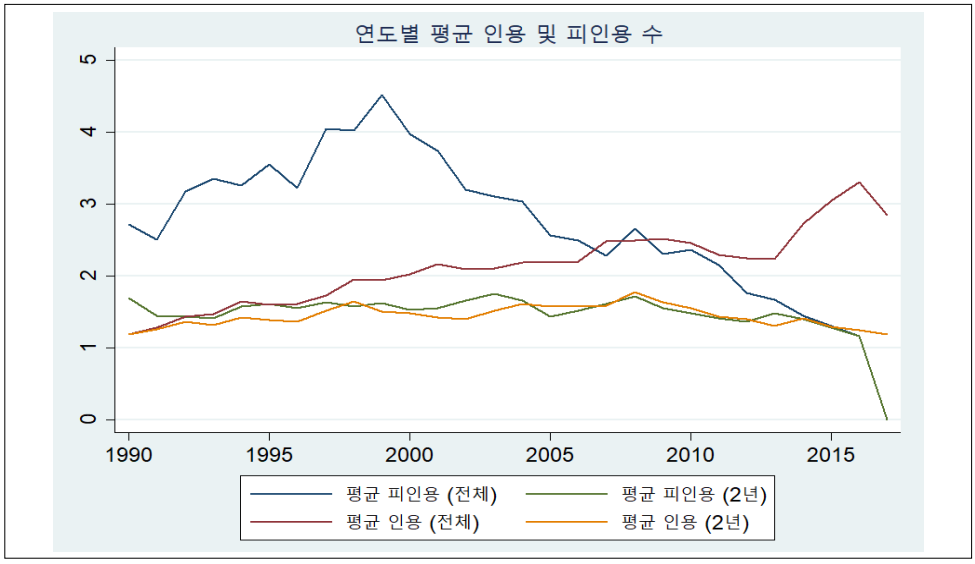
(73) Assignee: **Samsung Electronics Co., Ltd.**, Suwon-Si (KR)

7,773,066	B2 *	8/2010	Yamazaki et al.	345/89
2002/0109654	A1	8/2002	Kwon	345/87
2002/0196220	A1 *	12/2002	Sato et al.	345/87
2005/0088398	A1 *	4/2005	Lee	345/100
2005/0157559	A1	7/2005	Lee et al.	
2007/0200809	A1 *	8/2007	Yamazaki et al.	345/89
2010/0103158	A1 *	4/2010	Lee	345/211

FOREIGN PATENT DOCUMENTS

CN	1457449	11/2003
JP	09-325322	12/1997
JP	2001-042282	2/2001
JP	2002-041002	2/2002
JP	2003-172915	6/2003
JP	2003-186456	7/2003

〈그림 9〉



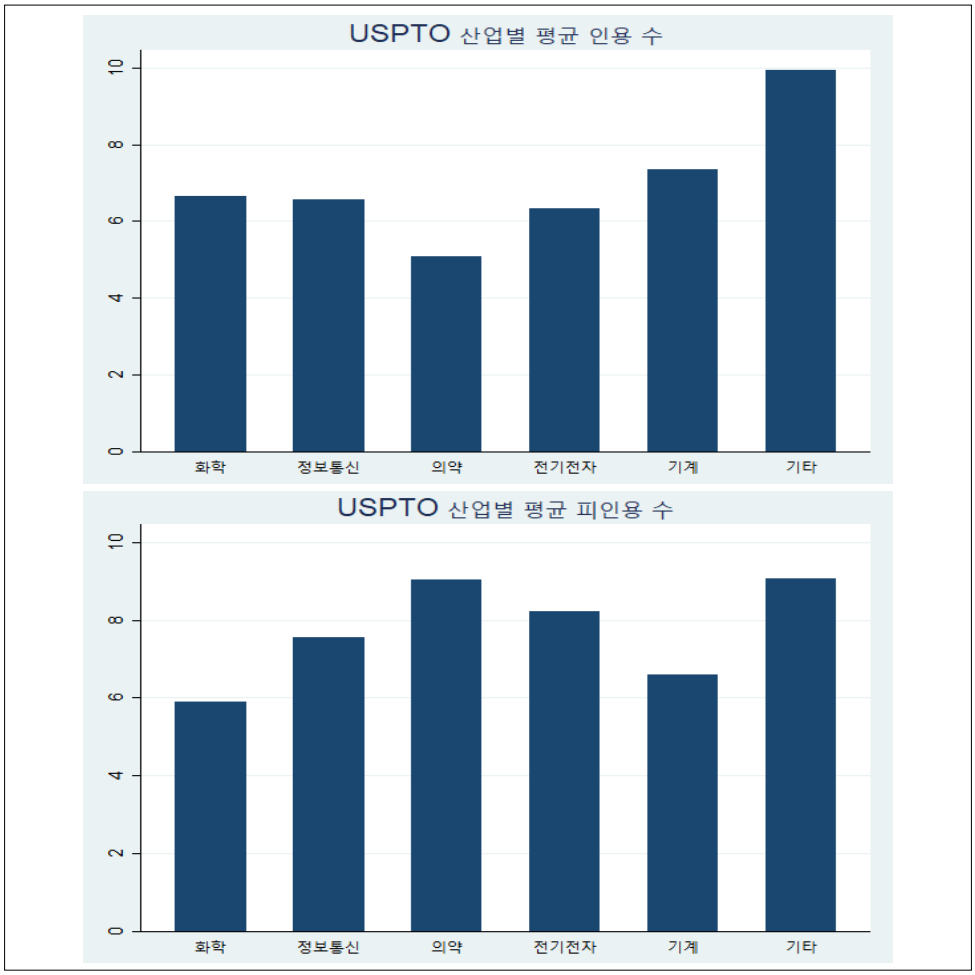
5. 인용 정보

〈그림 9〉는 출원인 국가 정보로 식별한 USPTO 한국 특허의 연도별 평균 인용 및 피인용 수를 보여주고 있다. 계산 방식은 KIPO의 경우와 동일하며, 마찬가지로 인용·피인용 횟수의 왜곡을 조정하기 위해 인용 기간을 2년으로 제한해보았다.[20] KIPO의 결과와 유사하게 인용 기간을 제한하지 않은 평균 피인용 수는 1980년대 초

반에 급격히 증가하나 2000년대 이후로는 점차 감소하고 있으며, 평균 인용 수는 다소 증가하는 양상을 보인다. 그러나 인용 기간을 제한하면 뚜렷한 증가 혹은 감소의 양상이 사라지고 대체로 일정하게 유지되고 있다. 다만 인용 기간을 제한한 평균 피인용 횟수의 경우 데이터 단절 문제가 있는 최근 시기에는 감소하는 모습을 보인다.

〈그림 10〉은 출원인 국가 정보로 식별한 USPTO 한국 특허의 산업별 평균 인용 및 피인용 수를 보여준다. 산업별 평균 인용 및 피인용 수도 KIPO와 동일한 방식으로 계산하였다. 의약 산업의 평균 피인용 수가 다소 낮은 반면 기타 산업의 평균 피인용 수가 높고, 의약과 기타 산업의 평균 피인용 수가 다소 높은 것을 확인할 수 있다.

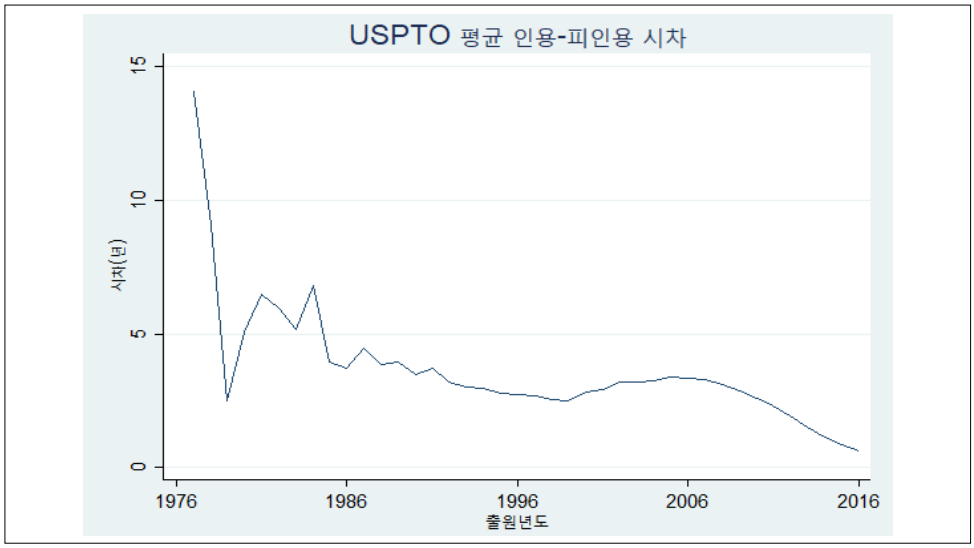
〈그림 10〉



20) 1976-2017년에 걸쳐 등록된 USPTO 특허를 대상으로 하였다.

〈그림 11〉은 출원인 국가 정보로 식별한 USPTO 한국 특허의 인용-피인용 시차를 보여준다. 즉, USPTO 한국 특허가 다른 USPTO 등록 특허에 의해 첫 번째로 인용을 받기까지 소요되는 시간의 연도별 평균이다. 계산 방식은 KIPO의 경우와 동일하게 적용하였다. 연도별 평균 인용-피인용 시차는 KIPO의 경우와 마찬가지로 1970년대 후반에 급격히 감소한 이래로 지속적으로 감소하는 양상을 보인다.

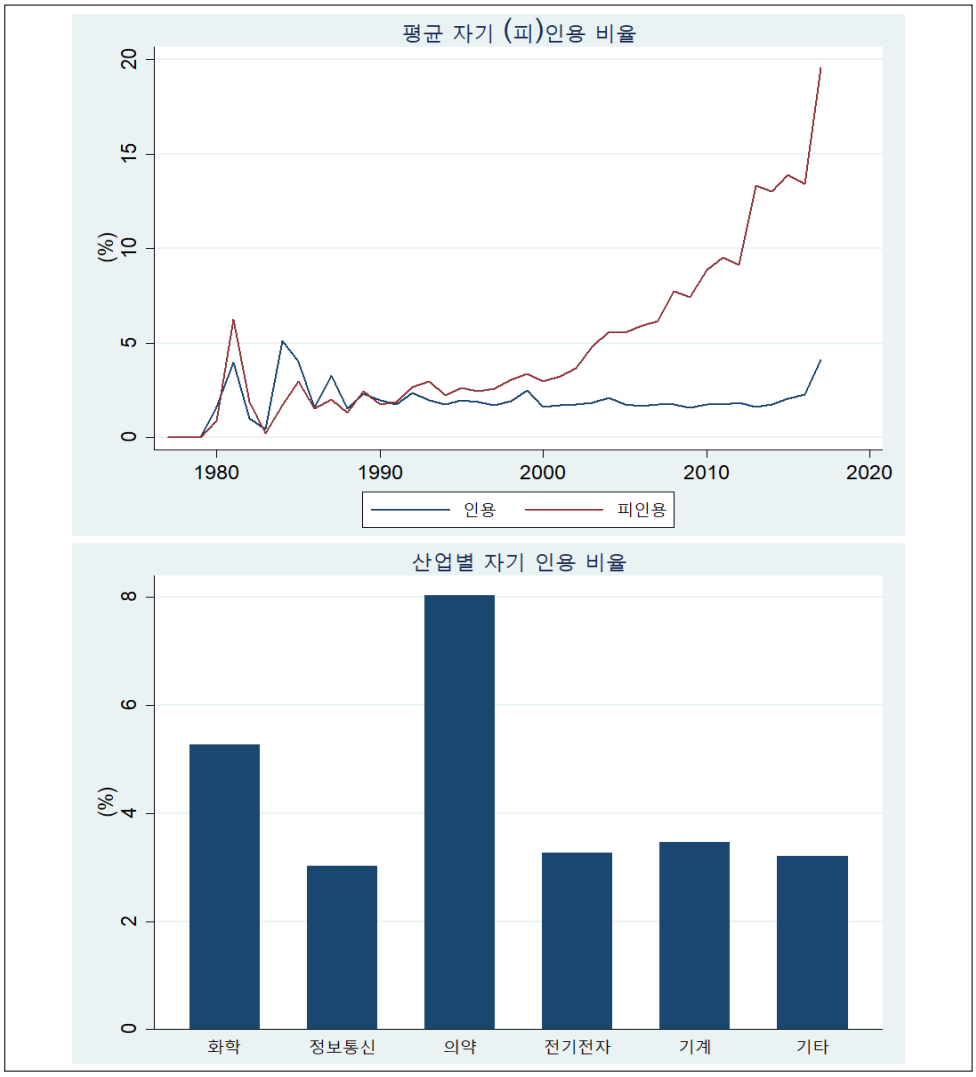
〈그림 11〉



한편, USPTO의 경우 고유하게 부여되는 출원인 번호가 존재하지 않기 때문에 자기 인용을 식별하는 것이 다소 까다롭다. 본 프로젝트에서는 출원인의 이름을 토대로 고유의 ID를 부여하였으며, 새롭게 부여된 ID를 토대로 자기 인용을 식별하였다.[21] 〈그림 12〉는 연도별 및 산업별로 출원인 국가 정보로 식별한 USPTO 한국 특허의 자기 인용 비율을 보여주고 있다. 연도별 평균 자기 (피)인용 비율 및 산업별 자기 인용 비율은 KIPO와 동일한 방식으로 계산하였으며, 산업별 자기 인용 비율의 경우 1976-2017년 동안 출원된 등록 특허의 누적 데이터를 사용하였다. KIPO의 경우에 비해 연도별 평균 자기 인용 비율은 비교적 안정적이나, 평균 자기 피인용 비율은 1990년대 후반부터 급격히 증가하는 모습이 특징적이다. 한편, 산업별 자기 인용 비율은 KIPO의 경우와 마찬가지로 의약 분야에서 두각을 보인다.

21) 보다 자세한 내용은 V절을 참조.

〈그림 12〉



6. 특허 분류

미국 특허청 특허는 2015년 5월까지 미국특허분류(US Patent Classification, USPC)를 사용하였으나 2015년 5월 이후로 협력적특허분류(Cooperative Patent Classification, CPC)를 사용하고 있다. 기존 NBER PDP에서 약 400개 이상의 USPC를 6개의 카테고리와 36개의 서브 카테고리로 재구성하여 산업을 분류하였으며, 이와 같은 특허의 산업 분류는 산업별 이질성(heterogeneity)을 분석하는 데에

많이 사용되고 있다. 그러나 미국 특허의 분류체계 변경으로 인해 2015년 5월 이후에 등록된 특허에 대해서는 더 이상 NBER PDP의 산업 분류를 사용할 수 없게 되었다.

그러나 국제특허분류(IPC)의 경우 USPC 혹은 CPC와 더불어 1976년부터 현재까지 지속적으로 부여가 되고 있다. 따라서 본 프로젝트에서는 미국 특허청 특허에 대해서도 마찬가지로 앞서 언급한 바와 같이 IPC에 따라 특허를 분류하였으며, IPC의 서브 클래스를 NBER 서브 카테고리에 대응시킨 매칭을 토대로 기존 NBER의 산업 분류를 그대로 사용할 수 있도록 하였다. 이 매칭은 단순히 미국 특허청에 등록된 한국 특허뿐만 아니라 미국 특허청 특허 전체에 적용할 수 있다. 미국 특허청 특허는 한국 특허청 특허와 마찬가지로 하나의 특허에 여러 IPC가 부여될 수 있으나, 2005년까지는 미국 특허청에서 'Main IPC'를 지정하고 있으며 2006년부터는 'Main IPC'를 따로 지정하고 있지 않지만 첫 번째 IPC를 'Main IPC'에 상응하는 것으로 보고 있다.

Ⅳ. 기업 정보; F&Guide DataGuide 5.0

특허 데이터로부터 얻을 수 있는 정보들을 최대로 활용하기 위해서는 특허 데이터를 다양한 다른 데이터와 결합하여 살펴볼 필요가 있다. 특허 데이터에만 한정하여 분석을 진행하게 되면 분석의 내용 및 범위가 국한될 가능성이 크기 때문이다. 예를 들어 발명자의 위치 정보를 해당 지역과 관련된 데이터와 결합하여 기술 혁신의 지역적 차이를 살펴볼 수 있을 것이며, 그 외에도 특허의 산업 분류, 특허 출원 및 등록 시점, 출원인에 관한 정보 등을 다양한 데이터와 결합하여 보다 확장된 연구를 진행할 수 있을 것이다.

본 프로젝트에서는 특허의 출원인을 기준으로 출원인이 기업에 해당하는 경우 기업 정보를 매치하여 특허 데이터의 확장성을 높이고자 하였다. 다양한 한국 기업 데이터베이스 중에서도 5만 개 이상의 상장·비상장 기업에 대한 각종 재무 및 금융 데이터를 보유하고 있는 FnGuide의 DataGuide 5.0을 사용하였다.[22]

22) DataGuide 외에 국내 기업 연구에 많이 이용되고 있는 Kisvalue와 비교하였을 때, DataGuide에는 비상장 외감 기업까지 포함하여 총 57,217개의 기업이 특허 데이터와 매치 대상이 되는 반면 Kisvalue에서는 비슷한 수준에서 31,920개의 기업이 매치 대상이 된다. 무엇보다 Kisvalue에는 DataGuide와 달리 기업의 영문명이 존재하지 않아서 USPTO에 특허를 출원한 국내 기업 출원인과 기업 데이터를 매치하는 데에 어려움이 있다. 저작권 관계로 DataGuide의 자료는 Lee(2019)에 공개되어 있지 않다. DataGuide 사용과 관련해서는 http://www.dataguide.

DataGuide 5.0은 한국상장회사협의회에서 분류한 상장 기업 및 외부감사 대상 법인에 대한 누적된 자료를 제공한다. 한 기업이 외부감사 대상 법인으로 지정되어 B 코드를 부여받으면, 기업의 해산이나 합병, 외부감사 대상 법인 지정 해제 이후에도 데이터가 남아 검색 시기와 관계없이 존재했던 기업에 대한 패널 자료를 확보할 수 있다. 기업이 상장되어 A 코드를 부여받았다가 상장이 폐지된 경우에도, 그 기록이 누적되어 상장폐지 이후에도 과거 상장 기록을 조회할 수 있다. 이는 KISVALUE의 경우 자료를 수집하는 시기에 외감 법인 혹은 상장 법인으로 남아있는 기업만 조회가 가능하여 표본 선택 문제가 발생할 수 있다는 점과는 상반된다.

DataGuide 5.0의 구성, 변수 목록 및 기업 범위에 대해 〈표 3〉, 〈표 4〉, 〈표 5〉에 정리하였다.

〈표 3〉 DataGuide 5.0 구성

종류	기간	출처	내용
금융통계	1999-	금융감독원	금융회사 및 금융업종 경영정보 데이터
산업	2000-	국토해양부, 지식경제부, 대한석유협회 등	국내 주요 산업별 협회 등 3천여 항목의 산업 DB
해외	2000-	이데일리, NRI	주요 글로벌 인덱스 및 주가, 도쿄 증권거래소 상장 기업 Data
주식	1980-	KOSCOM	주가, 수정주가, 수익률, 베타 등
이벤트	1980-	금융감독원	자본금 변동, 자기주식, 배당, 잠재적 보통주식, 기타 공시 내역 등
지분	2003-	금융감독원	상장된 기업의 주요주주 변동 내역 및 내부자 변동 사항
재무	1983-	한국상장회사협의회	상장 기업, 외부감사 기업, 일반 기업 재무제표, 재무비율, FCF
경제	1970-	한국은행, 통계청, 이데일리 등	한국은행, 통계청 등 21,000여 항목의 경제 데이터
컨센서스	2000-	국내 전 증권사 및 연구소	기업별 분기 및 연간 추정 재무 데이터, 투자 의견 등에 대한 Consensus 및 상세 추정 내역 DB

출처: FnGuide DataGuide 상품 소개서.

co.kr/DG5Web/index.asp를 참조하기 바란다.

〈표 4〉 DataGuide 5.0 주요 변수 목록

재무	재무제표	재무상태표 (총자본, 총부채, 연구개발비, 급여), 포괄 손익계산서, 현금흐름표, 자본 변동표, 이익잉여금 처분
	재무비율	안정성, 성장성, 수익성, 활동성, 생산성 (부가가치, 자본/노동 생산성)
	상대가치표	주당 지표, 주가 배수
	절대가치표	투하자본, 현금흐름, 자본비용
	배당	배당금, DPS, 배당률, 배당수익률, 배당성향
주식	가격, 수익률, 거래량, 주식 수, 시가총액 (보통/우선), 대차거래, 공매도, 신용거래, 투자자별 매매, 베타, 주가 배수 지표	
컨센서스	샤프 에스티메이트, 주주별 지분	
일반사항	기업 정보(영문 기업명, 본사 주소, 서울 주소, 산업 분류 코드, 상장/폐지 일자, 신용등급), 투자 참고, 임직원 정보(직원수)	
해외종목	가격, 수익률, 거래량	
경제	국민소득, 경기/산업, 고용/임금, 물가, 금융, 무역/국제수지/외환, 해외 통계	

출처: FnGuide DataGuide 상품 소개서.

〈표 5〉 DataGuide 5.0 기업 커버리지(Coverage)

기준	갯수	비고
MKF2000	1932	유가증권시장 및 코스닥시장에 상장된 전체 기업을 구성 종목으로 하는 주가지수
MKF500	503	한국증권선물거래소의 유가증권시장과 KOSDAQ에 상장된 전체 기업 중 시가총액 상위 500개 종목을 대상으로 개발한 지수
FnGuide Industry Group 27	503	기존 FICS(Fnguide Industry Classification Standard)의 중분류인 Industry Group(총 25개)에서 규모가 작은 의료 장비 및 서비스와 부동산을 각각 의료와 기타 금융에 합치고, 소분류 중 규모가 큰 화학, 금속 및 광물, 건설, 조선 등 4개를 별도의 업종으로 구분하여 사용자들이 가장 많이 비교하는 27개의 업종으로 주가지수를 산출하고 재무와 추정치를 합산한 데이터
FnGuide Universe	284	최근 3개월 동안 3개 이상의 증권사에서 투자의견을 제시한 종목으로 FnGuide Universe를 구성하여, 이들 종목만을 대상으로 한 과거 실적 및 추정치 재무 데이터와 시장 데이터에 대한 합산과 유동주식 비율을 감안한 시가총액 가중 방식의 지수를 산출하여 제공하는 서비스
KSE + KOSDAQ	2048	

KSE	768	
KOSDAQ	1280	
KONEX	152	창업 초기의 중소, 벤처기업들이 자본시장을 통해 필요한 자금을 원활하게 조달할 수 있도록 개설된 중소기업 전용 주식시장
K-OTC	120	코스피나 코스닥에 상장되지 못한 장외 기업들의 자금조달을 돕기 위한 주식시장
외감 기업(상장협 분류)	56614	
벤처 투자 대상 기업 (KOSDAQ)	572	
외감 기업(표준산업분류)	56247	

출처: FnGuide DataGuide 상품 소개서.

1. 기업 식별코드(Symbol)

DataGuide에서는 각 기업을 식별하기 위해 고유의 Symbol을 부여하고 있다. DataGuide Symbol에는 크게 A-type과 B-type의 두 가지 종류가 있으며 첫 번째 자리가 A 또는 B 그리고 뒤로 6자리의 숫자가 이어지는 총 7자리의 코드이다. A-type은 상장회사에 부여되는 Symbol이며 B-type은 각종 비상장 회사에 부여되는 Symbol이다. 특히 A-type의 경우 'A' 뒤에 이어지는 6자리의 숫자는 주식 종목 코드에 해당한다.

하나의 기업에 여러 개의 Symbol이 부여되는 경우가 존재하는데, 이는 주로 A-type과 B-type이 하나씩 부여된 경우에 해당한다. 대부분 비상장 기업이 상장되며 기존의 B-type Symbol에 더해 A-type Symbol을 새로 부여받았거나, 상장 기업이 상장폐지되면서 B-type의 Symbol을 추가로 부여받게 된 경우이다. 본 프로젝트에서는 이 같은 경우에 대해 '사업자등록번호', '법인등록번호', '설립일'이 모두 같은 경우에는 동일 기업으로 보고 A-type Symbol만을 부여하고 B-type Symbol을 제거하였다. 상장 혹은 상장폐지 등의 변화를 반영하지 않은 이유는 일관성을 위한 것으로, 매칭(matching) 단계(제Ⅴ절 참조)에서 기업의 사명 변경 혹은 M&A 등 시점에 따른 기업의 변화 역시 반영하지 않았기 때문이다. 많은 경우 비상장 기업보다는 상장 기업을 대상으로 분석하기 때문에 A-type Symbol을 남겨두었다. 또한 기업의 한글명이 동일하고 본사 주소가 '광역자치단체-기초자치단체' 수준에서 일치하여 동일 기업으로 간주되나 복수의 Symbol을 부여받은 경우에는 설립일이 늦은 것은 지사라고 판단하여 해당 기업에 부여된 Symbol을 제거하였다.

2. 기업명 및 주소

기업의 재무 정보 외에도 기업의 이름 및 주소는 기업 데이터를 특허 데이터와 결합하는 데에 중요한 정보를 제공한다. 특히 DataGuide는 국내 기업의 국문명과 영문명을 함께 제공하고 있어 지금까지 개발된 영문 기반의 다양한 문자열 알고리즘을 이용하여 기업과 특허 출원인을 매치할 수 있다. 또한 기업의 주소를 시-군-구 단위까지 상세하게 제공하고 있다. 국내 기업의 경우 특히 같은 이름을 가진 기업들이 다수 존재하는데, 이런 경우 매치 가능성이 높은 기업의 위치 정보를 비교하여 복수의 매치된 기업-출원인 쌍 중 어느 것이 제대로 된 매치인지 분별할 수 있는 기준을 제공한다. 그러나 기업명의 경우 특히 영문명은 철자에 오류가 많으며, 주소의 경우에도 다양한 변형이 존재한다(서울, 서울시, 서울특별시). 따라서 이러한 부분에 대한 추가적인 정리 작업이 필요하다.[23]

V. 특허 자료와 기업 정보의 매칭(Matching)

1. 매칭 대상

본 프로젝트는 아래의 총 3가지 데이터베이스를 토대로 특허 출원인과 기업 재무정보를 매치하였다. 매칭의 최종 목표는 KIPO의 특허 고객 번호와 USPTO 출원인에 부여한 ID를 DataGuide의 Symbol에 매치하여 각종 금융 및 재무 정보를 이용할 수 있도록 하는 것이다.

가. KIPO 특허 데이터에는 총 392,435개의 특허 고객 번호(출원인)가 존재하며 이들은 1948-2016년에 걸쳐 총 2,669,409개의 특허를 출원하였으며 이 중 1,551,354개가 최종 등록되었다.[24]

나. USPTO 특허 데이터는 출원인 이름에 대한 표준화(standardization과 harmonization)를 거친 후 총 7,502개의 출원인이 식별되었으며 이들은 191,102개의 첫 번째 출원인이 한국인인 등록 특허를 보유하고 있다.[25]

---

23) 보다 자세한 내용은 V절 참조.

24) 이후 매칭 과정에서 추가적인 정리를 거친 후 총 392,264개의 출원인이 식별되었다. V절 2항 및 3항 3목 참조.

25) 이후 매칭 과정에서 추가적인 정리를 거친 후 총 6,556개의 출원인이 식별되었다. V절 2항 및

다. DataGuide에는 총 55,079개의 DataGuide Symbol이 존재하며 이 중 3,345개는 상장 기업, 51,734개는 비상장 기업에 해당한다.

2. 출원인, 발명자 및 주소 표준화(Name Standardization)

KIPO와 DataGuide에는 출원인 이름과 기업명이 국문과 영문의 두 가지 형태로 주어지며, USPTO에는 출원인 이름이 영문으로 주어진다. 그러나 출원인과 기업 이름의 형태가 정형화되어있지 않고, 다수의 오류를 내포하고 있어 이름을 이용하여 매칭할 시 문제가 발생할 수 있다. 특히 USPTO는 출원인 고유의 식별코드가 없어 각 출원인을 식별하는 데에 어려움이 있다.

이를 해결하기 위해 NBER PDP에서 개발한 알고리즘을 이용하여 출원인과 기업의 영문명을 정리하였으며, 국문명의 경우에는 NBER PDP 알고리즘의 부호 정리를 그대로 따르고 추가로 이름에서 "주식회사", "(주)", "(유)", "(재)", "(합)"에 해당하는 부분을 제거하였다. 또한 KIPO와 DataGuide에는 출원인 및 기업의 주소가 국문으로 주어지는데, 이 역시 통일된 형태로 기재되어 있지 않아서 STATA 코드를 이용해 추가로 정리하였다. 예컨대 '서울, 서울시, 서울특별시'를 모두 '서울특별시'로 통일하였다.

(1) NBER PDP Name Standardization 알고리즘

① STATA '.do 파일'로 되어있으며 영문명을 입력하면 'standard name'과 'stem name'을 출력한다.

② 'standard name'

i) 기업명을 모두 대문자로 변경한다. (Samsung Electronics Co., Ltd. → SAMSUNG ELECTRONICS CO., LTD.)

ii) 기업명에서 빈번히 등장하는 약어(abbreviation)들을 정리하고, 기업명에서 'Co.', 'Ltd.', 'Inc.'와 같은 접미사가 다양한 형태로 등장하는 것을 통일해주며, 빈번히 발생하는 철자 오류 및 부호 등을 정리해준다. (SAMSUNG ELECTRONICS CO., LTD. → SAMSUNG ELECTRONICS CO LTD)

4항 3목 참조.

③ ‘stem name’

iii) ‘standard name’에서 접미사에 해당하는 부분을 제거한 형태를 돌려준다. (SAMSUNG ELECTRONICS CO LTD → SAMSUNG ELECTRONICS)

한편, 출원인 고유의 식별코드가 존재하지 않는 USPTO의 경우에는 ‘standard name’을 기준으로 ‘Assignee ID’를 부여하였으며(즉, ‘standard name’이 다르면 다른 ‘Assignee ID’를 가진다), 이후 매칭 결과를 토대로 ‘Assignee ID’를 보완하였다.

3. KIPO-DataGuide 매칭

출원인 및 기업의 이름과 주소를 표준화한 후 각 KIPO의 특허 고객 번호와 DataGuide의 Symbol에 대해서 아래와 같은 절차를 거쳐 매칭을 진행하였다. 일차적으로 출원인과 기업의 법인등록번호를 토대로 매칭을 진행한 후, 이 과정에서 매치되지 않은 출원인과 기업의 경우에는 문자열 알고리즘을 적용하여 추가로 매치하였다. 출원인 및 기업의 이름에 오류가 많이 존재하고 동명의 기업이 많아 이름을 토대로 매치하는 것에 비해 법인등록번호를 토대로 매치하는 것이 훨씬 정확하다고 판단하여 법인등록번호를 이용한 매치를 우선시하였다. 본 프로젝트에서 개발한 문자열(string matching) 알고리즘은 NBER PDP의 알고리즘을 토대로 하고 있으며 한국 기업명의 특수성을 고려하여 국문명과 지리 정보의 사용을 추가적으로 고려하였다.

(1) Step 1: 법인등록번호 이용

① 55,079개의 DataGuide 기업 중에서 54,750개의 기업에 대해 법인 번호가 존재하여 이들을 대상으로 매칭을 진행하였다.

② DataGuide에서는 기업별로 쉽게 법인등록번호를 얻을 수 있으며 KIPO 출원인에 대해서는 특허청에서 제공하는 “법인등록번호-특허 고객 번호”의 연결을 Python을 이용하여 크롤링(crawling) 하였다.[26]

③ 결과: 16,264쌍의 ‘DataGuide Symbol-특허 고객 번호’ 매칭이 완성되었다.[27]

---

26) 특허청 홈페이지에서 법인등록번호를 검색하면 특허 고객 번호를 얻을 수 있다. (https://www.patent.go.kr/jsp/kiponet/mp/apagtinfo/ReadApAgtInfoInput.jsp 참조.)

27) 여기서는 ‘DataGuide Symbol-특허 고객 번호’ 쌍 중 하나의 특허 고객 번호에 여러 개의

(2) Step 2: 문자열 알고리즘 적용

① Step 1에서 매치되지 못한 DataGuide Symbol과 특허 고객 번호를 문자열 알고리즘을 적용하여 매칭하였다.

② 이때 매칭은 NBER PDP의 문자열 알고리즘을 아래와 같이 변형하여 사용하였다.

i) 각 특허 고객 번호에 대해 출원인의 국문명이 기업의 국문명과 정확하게 일치하는 DataGuide Symbol을 찾는다(perfect match).

ii) 만약 위의 경우에 해당하는 매치가 없으면, 영문명의 'stem name'을 이용한 매치 점수가 임곗값 이상이고 광역자치단체 수준의 주소가 일치하는 경우를 매치한다(score-based match).

A. 매치 점수는 이름을 이루는 단어가 데이터상에서 나타나는 빈도를 토대로 계산한다.

B. 매치 점수 계산 방식

ⓐ X: KIPO 출원인 'stem name'의 토큰(token) 집합[28]

ⓑ Y: DataGuide 기업 'stem name'의 토큰(token) 집합

ⓒ X_k와 Y_k를 각각 X, Y 집합의 원소라고 하고, $f(X_k)$와 $g(Y_k)$를 각각 X_k와 Y_k가 KIPO와 DataGuide의 출원인 'stem name'에 등장하는 총 횟수라고 하자.

ⓓ 이때, KIPO 출원인 X를 DataGuide 기업 Y에 매치할 때의 매치 점수는 다음과 같이 계산한다:

$$w_{X_k} := \frac{100}{f(X_k)} \text{ and } w_{Y_x} := \frac{100}{g(Y_k)}$$

$$SC(X,Y) := \sum_{X \cap Y} w_{X_k}$$

DataGuide Symbol이 매치되는 중복을 허용한다. 이후 V절 3항 3목의 중복 처리 방법을 적용한다.

28) '광신하이테크'의 'stem name'은 'GWANG SIN HI TECH'이며, 이 경우 $X = \{GWANG, SIN, HI, TECH\}$가 된다.

$$RSC(X,Y) := \frac{\sum_{X \cap Y} w_{X_k}}{\sum_{X} w_{X_k}}$$

ⓐ DataGuide 기업 Y를 KIPO 출원인 X에 매치할 때의 매치 점수 $SC(Y,X)$와 $RSC(Y,X)$도 비슷하게 정의가 가능하다.

ⓑ NBER PDP의 기준을 적용하여 $SC(X,Y) > 110$이거나 $SC(X,Y) > 100$ & $RSC(X,Y) > 45$인 경우에만 매치한다.

iii) 하나의 DataGuide Symbol에 6개 이상의 특허 고객 번호가 매치된 경우에는 NBER PDP를 따라 매치가 부정확하다고 판단하여 제외하였다.[29]

iv) 매치 방향에 따라 결과가 달라질 수 있으므로 매치 점수를 토대로 한 결과는 양방향의 교집합만을 고려하였다.[30]

③ 결과

i) 출원인이 자연인인 경우와 등록 특허가 하나도 없는 출원인을 제외하고 총 864쌍이 국문명의 완벽한 일치로 매치되었다.

ii) 출원인이 자연인인 경우와 등록 특허가 하나도 없는 출원인을 제외하고 총 18쌍이 매치 점수 및 주소 정보로 매치되었다.

(3) 중복 처리

① 170쌍이 하나의 특허 고객 번호에 여러 개의 DataGuide Symbol이 매치되었으며 이 경우 아래의 기준에 따라 처리하였다.

i) DataGuide Symbol이 A-타입(상장회사)인 경우를 남겼다.

ii) 그럼에도 중복이 있는 경우 시-군-구 단위의 주소가 일치하는 경우를 남겼다.

iii) 그럼에도 중복이 있는 경우 'standard name'을 비교하여 정확하게 일치하는 것을 남겼다.

② 447쌍이 하나의 DataGuide Symbol에 여러 개의 특허 고객 번호가 매치되었다. 특히 등록 특허를 보유하고 있는 출원인과 관련하여, 이와 같은 오류의 주

29) 하나의 특허 고객 번호에 6개 이상의 DataGuide Symbol이 매치된 경우는 없었다.

30) KIPO를 DataGuide에 매치하거나 DataGuide를 KIPO에 매치할 수 있다.

된 원인은 사명 변경 혹은 회계 정보를 함께 공시하는 지사가 있는 경우에 해당한다. 따라서 우선 ①의 ii) - iii)을 거친 후, 여전히 하나의 DataGuide Symbol에 여러 개의 특허 고객 번호가 매치된 경우는 특허 고객 번호의 오류로 보고 특허 고객 번호를 하나로 합쳐주었다(harmonization).

③ 결과: 등록 특허가 한 개도 없는 출원인을 제외하고 총 14,803쌍의 'DataGuide Symbol-특허 고객 번호' 매칭이 완성되었다.

4. USPTO-DataGuide 매칭

USPTO 출원인의 경우에는 법인등록번호를 따로 제공하지 않기 때문에 특허 패밀리 정보를 이용하여 우선적으로 매치하였다. 그리고 이 과정에서 매치되지 않은 출원인과 기업의 경우에는 문자열 알고리즘을 적용하여 추가로 매치하였다. 문자열 알고리즘은 NBER PDP의 알고리즘을 이용하였으며, USPTO 출원인의 경우 출원인의 한글 이름을 제공하지 않기 때문에 출원인과 기업의 영문명만을 이용하였다. USPTO 출원인의 경우 위치 정보가 존재하지만 지나치게 오류와 변형이 많아 이 정보는 이용하지 않았다.

(1) Step 1: 특허 패밀리 정보 이용

① Thoma et al. (2010)의 방법을 차용하여 패밀리 정보로 연결된 USPTO와 KIPO의 출원인을 연결한 후, 앞 단계의 KIPO-DataGuide 매칭을 이용하여 최종적으로 USPTO 출원인에 DataGuide Symbol을 매치하였다.[31] 구체적인 방법은 아래와 같다.

i) 각 'Assignee ID'가 출원한 USPTO 특허를 모으고, 패밀리 정보를 이용하여 관련 KIPO 특허들을 연결한다.

ii) 연결된 KIPO 특허들의 출원인을 파악하고, 하나의 특허에 여러 명의 출원인이 존재하는 경우에는 1/n의 가중치를 부여한다.

iii) 'Assignee ID'당 연결된 KIPO 출원인의 명단을 중복을 허용하여 만든 후,

31) Thoma et al. (2010)은 본 연구와 유사하게 유럽 기업이 EPO뿐 아니라 미국에 등록한 특허도 함께 매치하였다.

가중치를 고려한 비중이 가장 높은 KIPO 출원인을 매치해준다.

iv) 매치된 'Assignee ID-특허 고객 번호' 쌍에 기존에 특허 고객 번호에 매치된 DataGuide Symbol을 연결한다.

② 상위 두 개의 출원인 비중이 동일한 경우 매치하지 않았다.[32]

③ 결과: 2,577쌍의 'Assignee ID-DataGuide Symbol' 매칭이 완성되었다.

(2) Step 2: 문자열 알고리즘 적용

① Step 1에서 매치되지 못한 'Assignee ID'와 DataGuide Symbol 전체에 문자열 알고리즘을 적용하여 매치하였다.[33]

② Lee and Lim(2019)에서 USPTO와 Compustat을 매치하기 위해 NBER PDP의 알고리즘을 Python 언어로 옮긴 코드를 그대로 사용하였다. 구체적인 방법은 아래와 같다.

i) 각 'Assignee ID'에 대해 USPTO 출원인과 'standard name'이 정확하게 일치하는 DataGuide Symbol을 찾는다(perfect match).

ii) 만약 위의 경우에 해당하는 매치가 없으면, 'stem name'을 이용한 매치 점수가 임곗값 이상인 경우를 매치한다(score-based match).

A. 매치 점수는 KIPO-DataGuide 매치에서 사용한 것과 동일하게 계산된다.

B. USPTO 출원인의 주소 정보는 매우 오류가 많아 주소 정보는 사용하지 않았다.

iii) 하나의 'Assignee ID'에 6개 이상의 DataGuide Symbol이 매치된 경우에는 KIPO-DataGuide 매치와 같은 이유로 매치하지 않았다.[34]

---

32) 이 경우 둘 중에 어느 KIPO 출원인이 USPTO 'Assignee ID'에 해당하는지 판단하기 어려우며 대부분의 경우 두 개의 출원인이 각각 0.5의 비중을 가지고 있는 경우에 해당하기 때문에 위의 방법을 이용한 식별이 어렵다고 판단하였다. 추후에 문자열 알고리즘으로 추가로 매치될 수 있는 가능성도 고려하여 이러한 경우는 제외하였다.

33) KIPO는 출원인별로 부여된 특허고객번호가 각 출원인을 완벽하게 식별하지만, USPTO의 경우 출원인 이름에 오류가 상당히 많다. 이와 같은 이유로 인해 Step 1에서 DataGuide Symbol에 매치되지 않은 경우를 포괄하기 위해 DataGuide Symbol은 Step 1에서 매치되었더라도 Step 2에서 추가로 'Assignee ID'와 매치될 가능성을 열어놓았다.

34) 'DataGuide Symbol-특허 고객 번호' 쌍 중 하나의 DataGuide Symbol에 6개 이상의 DataGuide Symbol이 매치되 경우는 없었다.

iv) 매치 방향에 따라 결과가 달라질 수 있어서 매치 점수를 토대로 한 결과는 양방향의 교집합만을 고려하였다.

③ 결과

i) 총 633쌍이 국문명의 완벽한 일치로 매치되었다.

ii) 총 78쌍이 매치 점수가 임곗값을 넘어 매치되었다.

(3) 중복 처리

① 573쌍이 하나의 'Assignee ID'에 여러 개의 DataGuide Symbol이 매치되었으며 이 경우 아래의 기준에 따라 처리하였다.

i) DataGuide Symbol이 A-타입(상장회사)인 경우를 남겼다.

ii) 그럼에도 중복이 있는 경우는 특허 패밀리 정보를 이용하여 매치된 경우를 남겼다.[35]

iii) 여전히 중복이 있는 경우는 부정확한 매치라고 판단하여 제외하였다(69쌍).

② 1,412쌍이 하나의 DataGuide Symbol에 여러 개의 'Assignee ID'가 매치되었다. 이는 USPTO의 출원인 이름이 오류가 많아서, 'standard name'을 기준으로 'Assignee ID'를 부여했음에도 불구하고 여전히 식별에 어려움이 있음을 보여준다.

③ 1,412쌍의 중복 매치 중 1371쌍이 특허 패밀리 정보를 이용한 결과에 해당한다. 이러한 경우는 모두 'Assignee ID'의 식별 문제로 보고 각 DataGuide Symbol에 매치된 여러 개의 'Assignee ID'를 하나로 합쳐주었다(harmonization).

④ 결과: 총 2,002쌍의 DataGuide Symbol과 'Assignee ID'가 1-1로 매치되었다.

35) 문자열 알고리즘보다는 패밀리 정보를 이용한 매치 결과가 자의성이 적고 더 정확할 것으로 판단하였다.

5. 매칭 결과

(1) 출원인 기준

① 총 14,803쌍의 'DataGuide Symbol-KIPRIS ID-Assignee ID'가 매치되었다.
② KIPO Only: 12,801쌍(상장회사 1,583쌍)
③ USPTO Only: 88쌍(상장회사 10쌍)
④ Both: 1,914쌍(상장회사 733쌍)

(2) 특허 수 기준

① 각 특허의 첫 번째 출원인이 DataGuide에 매치되어 있을 때 해당 특허가 기업 정보와 매치된 것으로 간주하였다.
② KIPRIS: 전체 등록 특허 중 45.96%(상장회사 36.27%)
⑤ USPTO: 첫 번째 출원인 국가 정보가 한국인 전체 등록 특허 중 87.85%(상장회사 79.28%)

(3) 특허 등록 수 상위 출원인

〈표 6〉 특허 등록 수 상위 10위 출원인

KIPO			USPTO		
등수	출원인 국문명	특허등록수	등수	출원인 영문명	특허등록수
1	삼성전자 주식회사	101437	1	SAMSUNG ELECTRONICS	66959
2	엘지전자 주식회사	53462	2	LG ELECTRONICS	21967
3	현대자동차 주식회사	46495	3	HYNIX SEMICONDUCTOR	12560
4	에스케이하이닉스 주식회사	29089	4	SAMUSNG DISPLAY	9103
5	주식회사 포스코	23495	5	LG DISPLAY	8078
6	주식회사 엘지이아이	17625	6	SAMSUNG SDI	6868
7	삼성에스디아이 주식회사	17131	7	HYUNDAI MOTOR	6161
8	엘지디스플레이 주식회사	15859	8	SAMSUNG ELECTRO MECHANICS	4168
9	삼성전기 주식회사	12604	9	LG CHEM	3278
10	주식회사 엘지화학	12092	10	LG INNOTEK	2523
Total		713055	Total		167885

Ⅵ. 산업 분류

본 프로젝트에서는 기존 NBER의 특허 산업 분류와의 연계성을 위해 Schmoch et al. (2003)의 분류에 따라 623개의 IPC 서브 클래스를 44개의 NBER 서브 카테고리에 매치하였다.[36] 2003년 이후 새로 생긴 IPC 서브 클래스들 중 'indexing scheme'을 제외한 17개의 IPC에 대해서는 각각의 IPC 정의와 Schmoch et al. (2003)에서 서브 카테고리에 대응시킨 국제산업분류(International Standard Industrial Classification, ISIC) 코드의 정의를 대조하여 기존의 44개 중 하나의 서브 카테고리를 부여하였다(〈표 7〉 참조).[37] 〈표 8〉에서 Schmoch et al. (2003)의 결과 외에 추가로 산업을 분류한 IPC에 대한 목록을 제공하고 있다. 이후 NBER의 미국특허분류(USPC)를 기반으로 한 산업 분류의 정의를 참고하여 44개의 서브 카테고리를 6개의 카테고리 수준의 산업 분류로 묶어주었다.

Schmoch et al. (2003)의 산업 분류는 Mancusi(2008), Belderbos et al. (2014) 등 특허 자료를 이용한 지식 전파 및 특허의 시장 가치를 분석하는 다수의 논문에서 사용되고 있다. Schmoch et al. (2003)의 산업 분류는 IPC를 기반으로 하고 있기 때문에 한국 특허청과 미국 특허청의 모든 특허를 망라하는 산업 분류를 가능하게 한다는 장점이 있다. 특히, 미국 특허청의 특허 분류가 2015년을 기점으로 USPC에서 CPC로 바뀌었다는 점에서 미국 특허에 대해 더 이상 NBER의 산업 분류를 사용할 수 없게 되었으며, 이 때문에 IPC를 기반으로 한 통일된 산업 분류가 필요하다. 또한 44개의 서브 카테고리를 NBER의 6개 카테고리로 묶어줌으로써 기존의 NBER 카테고리의 산업별 분류를 토대로 한 분석과 비교가 가능하도록 하였다.

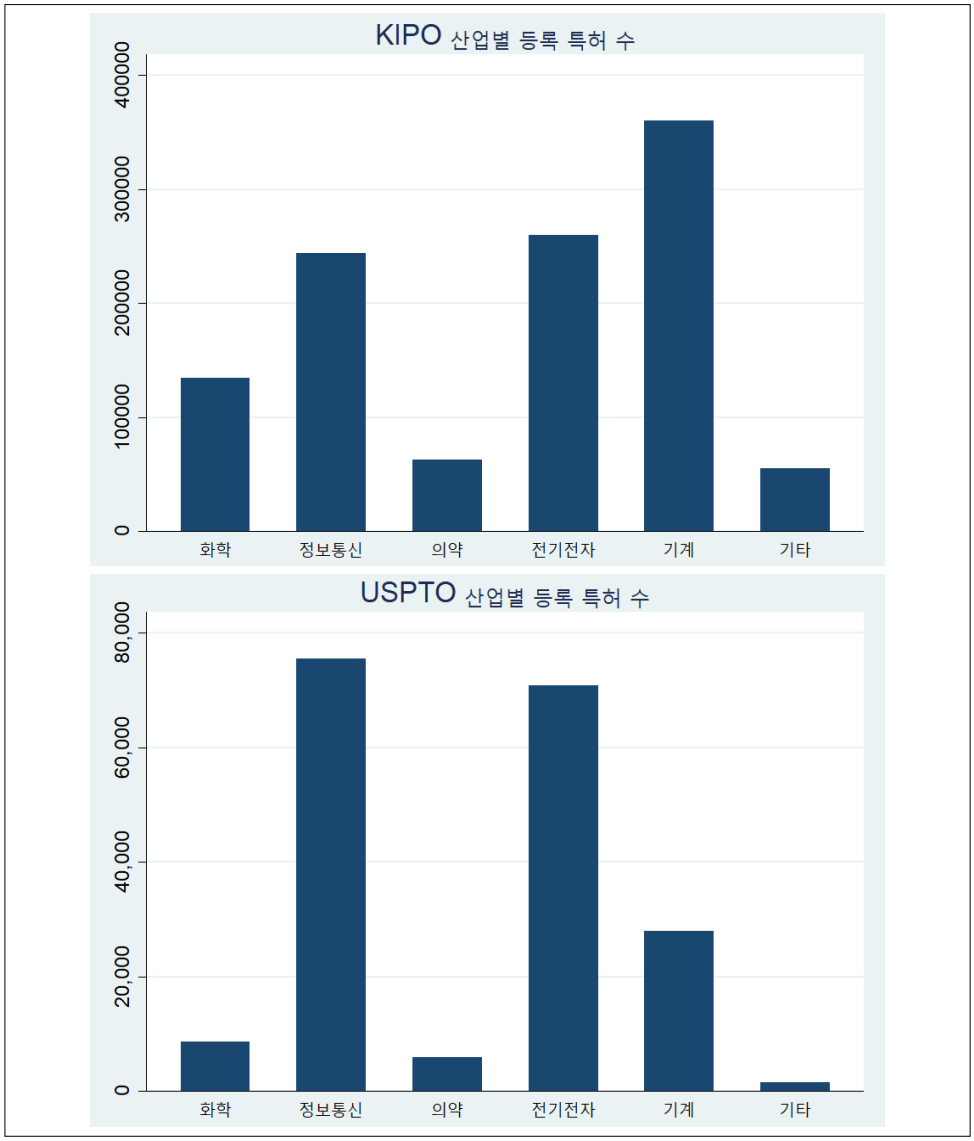
〈그림 13〉은 KIPO와 USPTO에 등록된 산업별 특허 수를 보여주고 있다. 산업별 특허 수는 KIPO와 USPTO 각각 1948-2016년, 1976-2017년 동안 산업별로 등록된 누적 특허 수를 계산하였다. KIPO에서는 기계산업에 해당하는 특허가 비교적 많이 등록되어 있다는 것이 특징이다. 반면 USPTO에는 정보통신 및 전기 전자산업에 해당하는 특허가 많이 등록되어 있고, 상대적으로 기계산업에 해당하는 등록 특허 수는

36) Schmoch et al. (2003)에서는 총 628개의 특허 중 'indexing scheme'이 아닌 620개의 IPC를 서브 클래스 수준에서 44개의 sub-category 산업으로 분류하였으며, 3개의 새롭게 등장한 IPC를 추가적으로 분류에 포함했다.

37) B82Y는 나노테크놀로지에 해당되는 분류인데, 해당 분류는 다른 분류들에 부가적으로 부여되는 특허 분류이기 때문에 따로 sub-category를 부여하지 않았다.

적은 편이다.

〈그림 13〉


KIPO 산업별 등록 특허 수
0
100000
200000
300000
400000
화학
정보통신
의약
전기전자
기계
기타
USPTO 산업별 등록 특허 수
0
20,000
40,000
60,000
80,000
화학
정보통신
의약
전기전자
기계
기타

〈표 7〉 'IPC-산업 분류' 결과

Category	Category Def.	Subcategory	Subcategory Def.
1	chemical	9	petroleum products, nuclear fuel
		10	basic chemical
		11	pesticides, agro-chemical products
		12	paints, varnishes
		14	soaps, detergents, toilet preparations
		15	other chemicals
		16	man-made fibers
		17	rubber and plastics product
		18	non-metallic mineral products
2	computers & communications	28	office machinery and computers
		35	signal transmission, telecommunications
3	drugs & medical	13	pharmaceuticals
		37	medical equipment
4	electrical and electronics	29	electric motors, generators, transformers
		30	electric distribution, control, wire, cable
		31	accumulators, battery
		32	lightening equipment
		33	other electrical equipment
		34	electronic components
		36	television and radio receivers, audiovisual electronics
		38	measuring instruments
		39	industrial process control equipment
5	mechanical	19	basic metals
		20	fabricated metal products
		21	energy machinery
		22	non-specific purpose machinery
		23	agricultural and forestry machinery
		24	machine tools
		25	special purpose machinery
		26	weapons and ammunition
		27	domestic appliances
		40	optical instruments
		41	watches, clocks
		42	motor vehicles
		43	other transport equipment
6	others	1	food, beverages
		2	tobacco products
		3	textiles
		4	wearing apparel
		5	leather articles
		6	wood products
		7	paper
		8	publishing, printing
		44	furniture, consumer goods

〈표 8〉 새롭게 추가된 IPC 목록 및 산업 분류 결과

IPC Subclass	IPC Subclass Definition	Number of Patents(KIPO)	Subcategory	Category
A01P	BIOCIDAL, PEST REPELLANT, PEST ATTRACTANT OR PLANTGROWTH REGULATORY ACTIVITY OF CHEMICAL COMPOUNDS OR PREPARATIONS	293	11	1
A61Q	SPECIFIC USE OF COSMETICS OR SIMILAR TOILET PREPARATIONS	26125	14	1
B33Y	ADDITIVE MANUFACTURING, i.e. MANUFACTURING OF THREE-DIMENSIONAL (3D) OBJECTS BY ADDITIVE DEPOSITION, ADDITIVE AGGLOMERATION OR ADDITIVE LAYERING, e.g. BY 3D PRINTING, STEREOLITHOGRAPHY OR SELECTIVE LASER SINTERING	1986	25	2
B60W	CONJOINT CONTROL OF VEHICLE SUB-UNITS OF DIFFERENT TYPE OR DIFFERENT FUNCTION; CONTROL SYSTEMS SPECIALLY ADAPTED FOR HYBRID VEHICLES; ROAD VEHICLE DRIVE CONTROL SYSTEMS FOR PURPOSES NOT RELATED TO THE CONTROL OF A PARTICULAR SUB-UNIT	24906	42	4
C13B	PRODUCTION OF SUCROSE; APPARATUS SPECIALLY ADAPTED THEREFOR	75	1	1
C40B	COMBINATORIAL CHEMISTRY; LIBRARIES, e.g. CHEMICAL LIBRARIES, IN SILICO LIBRARIES	201	15	1
F24S	SOLAR HEAT COLLECTORS; SOLAR HEAT SYSTEMS	2610	22	2
F24T	GEOTHERMAL COLLECTORS; GEOTHERMAL SYSTEMS	689	22	2
F24V	COLLECTION, PRODUCTION OR USE OF HEAT NOT OTHERWISE PROVIDED FOR	266	22	2
F99Z	SUBJECT MATTER NOT OTHERWISE PROVIDED FOR IN THIS SECTION	1	22	2
G01Q	SCANNING-PROBE TECHNIQUES OR APPARATUS; APPLICATIONS OF SCANNING-PROBE TECHNIQUES, e.g. SCANNING-PROBE MICROSCOPY (SPM)	651	40	4
G04R	RADIO-CONTROLLED TIME-PIECES	313	41	4
G06Q	DATA PROCESSING SYSTEMS OR METHODS, SPECIALLY ADAPTED FOR ADMINISTRATIVE, COMMERCIAL, FINANCIAL, MANAGERIAL, SUPERVISORY OR FORECASTING PURPOSES; SYSTEMS OR METHODS SPECIALLY ADAPTED FOR ADMINISTRATIVE, COMMERCIAL, FINANCIAL, MANAGERIAL, SUPERVISORY OR FORECASTING PURPOSES, NOT OTHERWISE PROVIDED FOR	209768	28	2
G16H	HEALTHCARE INFORMATICS, i.e. INFORMATION AND COMMUNICATION TECHNOLOGY ICT SPECIALLY ADAPTED FOR THE HANDLING OR PROCESSING OF MEDICAL OR HEALTHCARE DATA	791	35	3
G99Z	SUBJECT MATTER NOT OTHERWISE PROVIDED FOR IN THIS SECTION	4	22	2
H02S	GENERATION OF ELECTRIC POWER BY CONVERSION OF INFRA-RED RADIATION, VISIBLE LIGHT OR ULTRAVIOLET LIGHT, e.g. USING PHOTOVOLTAIC (PV) MODULES	8610	29	2
H04W	WIRELESS COMMUNICATION NETWORKS	134907	35	3

Ⅶ. 유의 사항

1. 데이터 단절 문제

데이터 단절과 관련된 이슈는 출원-등록-공개 사이의 시차로부터 비롯된다. KIPO 데이터의 경우 최근에 가까워질수록 출원된 특허가 아직 등록에 필요한 절차를 거치고 있어 등록 특허가 될 가치가 충분함에도 데이터에 등록 특허로 잡히지 않을 수 있음에 유의해야 한다. 따라서 비록 출원일이 실제 발명 시점에 더 가깝지만, 출원일을 사용하여 등록 특허를 분석할 때에는 등록 특허의 누락 가능성을 고려해야 한다. 또한 출원되었으나 등록되지 않은 특허들도 공개되기까지 시간이 소요되기 때문에 이로 인한 출원 특허의 누락 가능성도 고려해야 한다.

USPTO 데이터의 경우에도 마찬가지로 데이터 단절 문제가 있으며 등록된 특허에 대한 데이터만 존재하기 때문에 데이터 단절을 더욱 신중히 고려해야 한다. KIPO와 마찬가지로 최근에 가까워질수록 출원된 특허가 아직 등록에 필요한 절차를 거치고 있어 누락된 등록 특허가 존재할 것이다. 따라서 출원일을 특허의 발생 시점으로 간주할 때에는 등록 특허의 누락 여부를 고려해야 한다.

데이터의 단절 문제는 단순히 해당 연도에 출원 혹은 등록된 특허의 숫자를 헤아릴 때뿐 아니라 특허의 인용 혹은 피인용 횟수를 계산할 때에도 주의를 기울여야 한다. 다시 말해 출원되었으나 아직 공개 혹은 등록되지 않은 특허 혹은 앞으로 출원될 특허들을 고려할 필요가 있다. 먼저 등록 특허만을 대상으로 인용 횟수를 계산하는 경우, 출원되었으나 아직 등록되지 않은 특허를 인용했다면 이는 출원-등록의 시차로 인해 인용 횟수에 포함되지 않게 된다. 더 중요한 문제는 각 특허의 피인용 횟수를 계산할 때 과거에 출원 혹은 등록된 특허에 비해 최근에 출원 혹은 등록된 특허일수록 인용을 받을 수 있는 기간이 짧다는 구조적인 이유로 피인용 횟수가 낮아질 수 있다는 점이다. 특허의 피인용 횟수는 주로 특허의 질을 대리하는 지표로 사용되기 때문에 이와 같은 문제는 최근에 출원 혹은 등록된 특허의 질을 과거에 출원 혹은 등록된 특허의 질보다 낮게 측정할 가능성이 존재한다.

피인용 횟수를 계산할 때에는 시간의 흐름에 따라 인용의 패턴 자체가 바뀔 수 있다는 점도 고려해야 한다. 예를 들어, 1990년의 특허와 2010년의 특허가 동일한 기술적인 파급력을 지니고 있더라도, 2000년대 이후 정보 기술이 발달하여 심사관들이 기존 특허를 더욱 쉽게 찾아낼 수 있다면 2010년의 특허가 더 많은 인용을 받을 수 있

다. 이러한 데이터 단절 문제와 인용 패턴의 변화 문제를 해결하고 인용 횟수로부터 의미 있는 정보를 추출하기 위해 Hall, Jaffe and Trajtenberg (2001) 과 Lerner and Seru (2017) 등에서 세 가지 전처리 방법을 제시한다.

첫 번째 방법은 각 특허가 인용을 받을 수 있는 기간을 설정하여 해당 기간 중 받은 인용 횟수만을 고려하는 것이다. 대표적으로 Lee and Lim (2019) 에서는 10년의 기간을, Bloom and van Reenen (2002) 에서는 5년의 기간을 사용한다. 이를 통해 데이터의 단절 문제를 해결할 수 있지만, 이 방법은 인용 패턴이 시간의 흐름에 따라 크게 변하지 않는다는 가정이 필요하다.

두 번째 방법은 고정 효과 접근법 (fixed effects approach) 으로 모든 종류의 인용 패턴 변화를 소음 (noise) 으로 보고 처리하는 것이다. 이 경우, 개별 특허의 피인용 횟수를 당해의 모든 특허의 평균적인 피인용 횟수로 나누어 줌으로써 해당 특허의 기술적인 파급력의 지표로 사용한다. 이 방법은 단순하고 별도의 구조적인 가정이 필요하지 않다는 장점이 있으나, 실질적인 기술력 변화에 따른 인용 패턴의 변화까지 제거한다는 단점이 있다.

마지막 방법은 준구조적 접근법 (quasi-structural approach) 으로, 인용 시차의 분포에 대해 구조적인 가정을 추가한 후 계량경제학적 추정을 통해 인용 데이터의 소음을 식별하는 것이다. 구체적인 방법에 대한 상세한 설명은 Hall, Jaffe and Trajtenberg (2001) 에 소개되어 있다. 이를 통해 데이터의 소음들을 식별하고 나면 사용자의 선택에 따라 데이터를 보정할 수 있다. 이 방식은 구조적인 가정에 의존한다는 단점이 있으나, 이러한 가정이 현실에서도 충족된다고 믿을 만한 근거가 있는 경우, 위의 두 방법보다 정교하게 의미 있는 정보를 추출할 수 있다는 장점이 있다.

2. 특허 분류와 산업 분류

본 데이터는 국제특허분류 (IPC) 에 따라 특허를 분류하고 있으며 IPC를 기준으로 특허의 산업을 분류하고 있다. 이와 같은 'IPC-산업 분류' 매핑 (mapping) 은 각 특허를 산업에 매핑하는 것이 아니라 IPC를 산업에 매핑하기 때문에 하나의 특허에 여러 개의 IPC가 붙고 각 IPC에 매핑된 산업이 다를 경우 문제가 발생한다. 이를 해결하기 위한 방법으로 KIPO의 경우에는 여러 개의 IPC 중 하나를 무작위로 택하여 해당 IPC에 매칭된 산업 분류를 적용할 수 있다. 한편, USPTO는 2005년까지 여러 개의 IPC 중 중심이 되는 IPC를 지정하고 있으므로 KIPO보다 수월하게 문제를 해결할 수

있다. 즉, 2005년까지는 USPTO에서 지정하는 메인 IPC를 사용하여 산업 분류를 적용하고 2006년부터는 여러 개의 IPC 중 첫 번째 IPC를 사용하여 산업 분류를 적용하는 것이다. 본 데이터베이스의 'ipc_uspto.dta'에는 이와 같은 방식으로 하나의 특허에 하나의 IPC가 부여되어 있다.[38]

특허의 산업 분류와 기업의 산업 분류에는 차이가 있다는 점에 주의해야 한다. 기업의 산업을 분류할 때에는 대체로 한국표준산업분류(Korean Standard Industrial Classification, KSIC) 혹은 FnGuide에서 제공하는 FnGuide Industry Classification Standard(FICS)와 같이 각 기업에 부여되는 산업 분류를 이용한다. KSIC 혹은 FICS와 NBER의 특허 산업 분류는 1:1로 매핑이 되지 않으며 해당 기업이 특정 분야의 산업에 속한다고 해서 법적 권리를 소유하고 있는 특허가 모두 같은 산업에 속하는 것은 아닐 수 있다.

3. 매칭 및 기업 정보 이용

(1) 기업의 시계열적 변화

본 데이터는 법인등록번호 및 기업명을 기준으로 KIPO, USPTO, 그리고 DataGuide를 매치하였다. 이 때문에 기업명의 변화 혹은 M&A 등으로 인한 기업 소유권 변경이 발생하는 경우, 기존의 특허에 대한 소유권 역시 새로운 기업으로 귀속해야 하나 본 데이터에는 이러한 부분이 반영되지 않았다.[39] 따라서 Hall, Jaffe, and Trajtenberg(2005)에서처럼 '계속기록법(perpetual inventory)' 방식을 이용하여 기업별 스톡 변수를 계산하는 경우 다소 과소 추정될 가능성이 존재한다. 이러한 문제는 최근 데이터베이스에 처음 등장하는 기업일수록 높다.

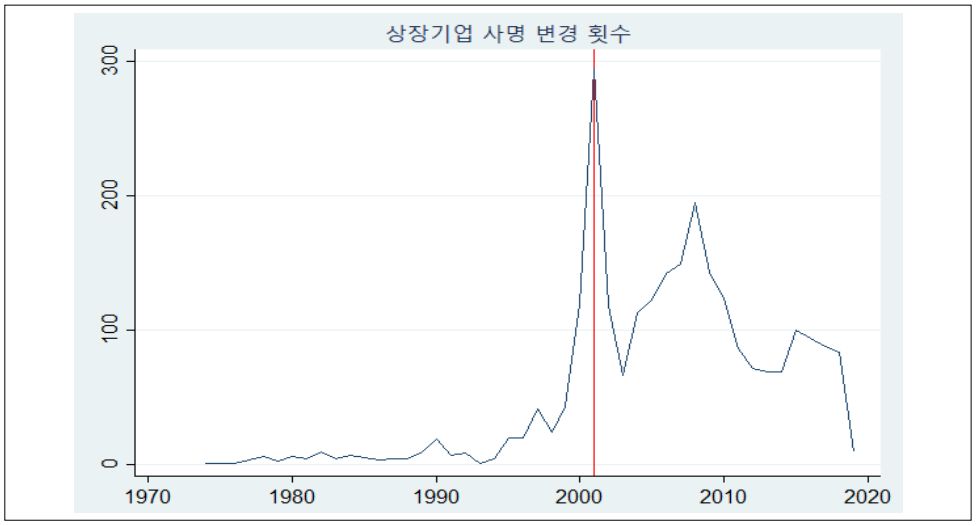
예컨대 〈그림 14〉는 DataGuide에서 1,305개의 A-type Symbol을 가진 기업들의 사명 변경 이력을 1974-2019년 기간에 대해 조회한 결과이다. 조회 결과 사명 변경은 특히 2000년을 기점으로 급증하기 시작했으며, 전체 조회 기간 중 2001년에 295건으

38) Cooperative Patent Classification(CPC)의 경우 메인 특허 분류와 유사하게 'first CPC'를 지정하고 있으며 'first CPC'가 해당 기술의 특성을 가장 잘 반영하는 특허 분류에 해당한다(https://www.uspto.gov/web/offices/pac/mpep/s905.html 참조). IPC와 CPC의 경우 서브클래스 수준까지 매우 유사하고, IPC에 대해서도 'first IPC'가 지정된 점을 이용하였다.

39) 새로운 기업이란 새로운 기업명 혹은 해당 기업을 소유하게 된 기업을 의미한다.

로 최고점을 기록하였다. 2000-2018년 동안 발생한 사명 변경의 평균 건수는 매년 약 118건 수준이다. 이와 같은 사명 변경의 결과가 본 데이터베이스의 매칭 결과에는 반영되지 않았다.

〈그림 14〉



인수합병 이력의 경우, 톰슨-로이터에서 제공하는 Securities Data Company (SDC) Platinum 데이터베이스를 참고하여 연구자의 필요에 따라 보완할 수 있다. SDC Platinum의 M&A 데이터베이스는 1985년 이후 이루어진 147,000건 이상의 세계 M&A에 대한 상세 정보를 뉴스 기사, SEC 공시 정보 등을 통해 수집하여 제공하며, 그중에는 한국 기업의 인수합병 정보 역시 포함되어 있다. 1991년까지는 100만 달러 이상 규모의 거래만 기록되었으나 1992년부터는 그보다 작은 규모의 M&A 정보 역시 포함되어 있다.[40]

(2) 모회사-자회사 관계와 개별 vs. 연결 재무제표의 사용

특허 데이터에 기업 정보를 매치하여 분석할 경우, 각 특허를 어떤 기업의 지적 자산으로 귀속할 것인지가 중요하다. 특히 모회사-자회사 관계에 있는 회사의 경우 다

40) 보다 자세한 내용은 SDC Platinum의 공식 홈페이지(https://www.refinitiv.com/en/products/sdc-platinum-financial-securities)를 참조하기 바란다.

음의 두 가지 방식이 가능하다.

① 직접 특허를 출원한 자회사에 귀속하고 개별 재무제표를 사용.

② 자회사가 재무제표 공시 의무가 없는 경우를 고려하여 특허를 모회사에 귀속하고 연결 재무제표를 사용.

전자의 경우, 재무제표를 공시하지 않는 기업을 분석 대상에서 제외해야 한다는 단점이 있으나, 특허권이 사실상 해당 특허와 전혀 무관한 기업의 자산으로 귀속될 여지가 적다는 장점이 있다. 후자의 경우, 연결 재무제표를 사용하기 때문에 하나의 모회사에 여러 자회사가 존재하고 그중 하나라도 재무제표를 공시하지 않을 경우 모든 자회사들의 특허권을 모회사에 귀속하고 연결 재무제표를 사용해야 한다. 따라서 특허권과 재무제표와의 실질적 연관성이 약화될 수 있다. 다만 후자의 경우 더 많은 특허를 분석 대상에 포함할 수 있다는 장점이 있다. 본 프로젝트에서는 전자의 방식을 사용하였다. 따라서 DataGuide Symbol로 재무 데이터를 불러올 때 반드시 개별 재무제표(non-consolidated level) 데이터를 사용해야 한다.

Ⅷ. 결 론

많은 경제학 연구는 주제에 적합하고 신뢰할 수 있는 데이터를 구축하는 데에서부터 시작된다. 그러나 '지식(knowledge)'이라는 것은 매우 추상적인 개념이기 때문에 이를 어떻게 수치화하여 측정할 것인가에 대한 어려움이 있다. 이러한 맥락에서 기업과 특허 자료를 연결하는 데이터베이스를 구축한 KoPDP가 한국 경제의 혁신 역량에 대한 보다 엄밀하고 과학적인 진단과 처방을 도출하는 데 일조하기를 기대한다.

최근 이지홍・김상동・송근상(2019)은 KoPDP의 자료를 활용하여 지식 자본과 기업 생산성의 상관관계를 분석하였다. 구체적으로 2003-2013년 동안 한국 또는 미국 특허청에 특허를 등록한 이력이 있는 504개의 국내 상장 기업을 살펴본 결과, 특허 스톡이 기업의 유형자산 및 노동 투입 규모에 의해 설명되지 않는 매출 중 일부를 통계적으로 유의하게 설명할 수 있음을 보였다. 특허의 단순 등록 수뿐만 아니라 피인용 횟수 또한 기업 생산성에 관한 유의미한 정보를 내포하고 있는 것을 확인하였다.

위와 같은 결과는 KoPDP가 구축한 자료가 한국 내 혁신 활동의 분석에 활용될 수 있음을 의미하는 근거를 제시한다. 특허 스톡을 혁신 활동의 대리변수로 사용하여 산업・금융・노동 등 다양한 분야의 주요 정책 효과를 분석하는 시도가 가능할 것이다. 거시 및 국제경제 분석에서도 특허 정보가 기존에 사용되어온 혁신 관련 변수들을 보

완하는 역할을 할 수 있을 것이다. 한편 본 연구에서 시도한 한국과 미국 특허의 연계는 국가 간 혁신 활동의 비교연구 등 여러 새로운 연구에 유용하게 활용될 것으로 기대된다.

앞서 언급되었듯이 본 프로젝트의 데이터베이스에는 확장성과 개선의 여지가 많이 남아있다. 예컨대 특허의 인용 관계로부터 더 많은 정보를 추출하여 지식의 전파를 추적한다거나(Jaffe, Trajtenberg, and Henderson, 1993; Kwon et al., 2018), 특허의 '고유성(originality)' 혹은 '일반성(generality)'을 측정하는 지표를 만드는 등(Trajtenberg, Jaffe, and Henderson, 1997) 다양한 목적으로 확장 적용될 수 있을 것이다.

나아가 이 데이터베이스는 본 프로젝트에서 다루지 않은 다양한 데이터베이스를 기업 단위에서 매치함으로써 더욱 그 범위를 넓혀나갈 수 있을 것이다. 예를 들어, EPO나 WIPO와 같은 해외 특허청 데이터를 고려해볼 수 있다. 실제로 NBER PDP의 데이터로부터 시작한 깊이 있고 흥미로운 다수의 연구들이 이처럼 여러 데이터베이스를 결합함으로써 탄생하였다.

■ 참 고 문 헌

1. 이지홍 · 김상동 · 송근상, "지식자본과 기업 생산성: 특허 자료를 중심으로," 『경제논집』, 제58집 2호, 2019, pp. 43-68.
2. 이지홍 · 임현경 · 정대영, "4차 산업혁명과 한국의 혁신 역량: 특허 자료를 이용한 국가 · 기술별 비교 분석," 『경제분석』, 제24집 3호, 2018, pp. 37-82.
3. Aghion, P., U. Akcigit, A. Bergeaud, R. Blundell, and D. Hémous, "Innovation and Top Income Inequality," *The Review of Economic Studies*, Vol. 86, 2018, pp. 1-45.
4. Aghion, P., J. Van Reenen, and L. Zingales, "Innovation and Institutional Ownership," *American Economic Review*, Vol. 103, 2013, pp. 277-304.
5. Akcigit, U., M. A. Celik, and J. Greenwood, "Buy, Keep, or Sell: Economic Growth and the Market for Ideas," *Econometrica*, Vol. 84, 2016, pp. 943-984.
6. Autor, D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu, "Foreign Competition and Domestic Innovation: Evidence from US Patents," forthcoming in *American Economic Review*, 2019.
7. Belderbos, R. A., B. Cassiman, D. Faems, B. Leten, and B. V. Looy, "Co-Ownership of Intellectual Property: Exploring the Value-Appropriation and Value-Creation Implications of Co-Patenting with Different Partners," *Research Policy*, Vol. 43, 2014, pp. 841-852.
8. Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," *Quarterly Journal of Economics*, Vol. 134, 2018, pp. 647-713.

9. Bloom, N. and J. Van Reenen, "Patents, Real Options and Firm Performance," *Economic Journal*, Vol. 112, 2002, pp. C97-C116.
10. Bloom, N., M. Schankerman, and J. Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, Vol. 81, 2013, pp. 1347-1393.
11. Caballero, R. J. and A. B. Jaffe, "How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," *NBER Macroeconomics Annual*, Vol. 8, 1993, pp. 15-74.
12. Ellison, G., E. L. Glaeser, and W. R. Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review*, Vol. 100, 2010, pp. 1195-1213.
13. Griliches, Z., B. H. Hall, and A. Pakes, "R&D, Patents, and Market Value Revisited: Is There a Second (Technological Opportunity) Factor?," *Economics of Innovation and New Technology*, Vol. 1, 1991, pp. 183-201.
14. Hall, B. H., A. B. Jaffe, and M. Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," Working Paper No. 8498, National Bureau of Economic Research, 2001.
15. _____, "Market Value and Patent Citations: A First Look," *RAND Journal of Economics*, 2005, pp. 16-38.
16. Jaffe, A. B., M. Trajtenberg, and R. Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, Vol. 108, 1993, pp. 577-598.
17. Jones, B. F., "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *Review of Economic Studies*, Vol. 76, 2009, pp. 283-317.
18. Kerr, W. R. and W. F. Lincoln, "The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention," *Journal of Labor Economics*, Vol. 28, 2010, pp. 473-508.
19. Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman, "Technological Innovation, Resource Allocation, and Growth," *Quarterly Journal of Economics*, Vol. 132, 2017, pp. 665-712.
20. Kwon, S., J. Lee, and S. Lee, "International Trends in Technological Progress: Evidence from Patent Citations, 1980-2011," *Economic Journal*, Vol. 127, 2017, pp. F50-F70.
21. Kwon, H. S., J. Lee, S. Lee, and R. Oh, "Knowledge Spillovers and Patent Citations: Trends in Geographic Localization, 1976-2015," forthcoming in *Economics of Innovation and New Technology*, 2019.
22. Lanjouw, J. O. and M. Schankerman, "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators," *Economic Journal*, Vol. 114, 2004, pp. 441-465.
23. Lee, J., "Korea Patent Data Project (KoPDP)," https://doi.org/10.7910/DVN/AUYERV, *Harvard Dataverse*, 2019.
24. Lee, J., and H. Lim, "Market Value of Patents: Evidence from the US, 1976-2017," Working Paper, Seoul National University, 2019.
25. Lerner, J., and A. Seru, "The Use and Misuse of Patent Data: Issues for Corporate Finance and Beyond," Working Paper No. 24053, National Bureau of Economic Research, 2017.
26. Mancusi, M. L., "International Spillovers and Absorptive Capacity: A Cross-country

Cross-sector Analysis Based on Patents and Citations," *Journal of International Economics*, Vol. 76, 2008, pp. 155-165.

27. Schmoch, U., F. Laville, P. Patel, and R. Frietsch, "Linking Technology Areas to Industrial Sectors," *Final Report to the European Commission, DG Research*, 2003.
28. Thoma, G., S. Torrisi, A. Gambardella, D. Guellec, B. H. Hall, and D. Harhoff, "Harmonizing and Combining Large Datasets: An Application to Firm-level Patent and Accounting Data," Working Paper No. 15851, National Bureau of Economic Research, 2010.
29. Tian, X., and T. Y. Wang, "Tolerance for Failure and Corporate Innovation," *Review of Financial Studies*, Vol. 27, 2014, pp. 211-255.
30. Trajtenberg, M., "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND Journal of Economics*, 1990, pp. 172-187.
31. Trajtenberg, M., Henderson, R., and A. Jaffe, "University Versus Corporate Patents: A Window on the Basicness of Invention," *Economics of Innovation and New Technology*, Vol. 5, 1997, pp. 19-50.

〈부록: 데이터 파일 및 변수 목록〉

1. 한국 특허청 특허 자료

(1) basicinfo.dta: KIPO 출원 및 등록 특허의 기본 정보

① 출원번호(appnum)

② 출원 일자(appdate)

③ 법적 상태(finaldisposal)

④ 등록번호(registernum)

⑤ 등록 일자(registerdate)

⑥ 심사청구 여부(exmrequest)

⑦ 심사청구항 수(nclaim)

⑧ 출원일의 STATA 날짜 변수(app_date) - constructed

⑨ 등록일의 STATA 날짜 변수(reg_date) - constructed

⑩ 등록 여부(regi)

- 최종 등록된 특허에 대해 1의 값을 갖는 변수, constructed

(2) assignee.dta: KIPO 특허 출원인 정보

① 출원번호(appnum)

② 출원인 순서(order)

③ 출원인 한글명(name)

④ 출원인 영문명(engname)

⑤ 특허 고객 번호(kiprisid)

⑥ 특허 고객 번호 Harmonized(kiprisidH)

- 매칭 과정에서 추가로 정리된 특허 고객 번호, constructed

⑦ 출원인 국가(country)

⑧ 출원인 주소(address)

⑨ 등록 여부(regi)

- 최종 등록된 특허에 대해 1의 값을 갖는 변수, constructed

(3) citation.dta: KIPO 특허 인용 정보

- XML 파싱 결과(citation.csv)는 출원번호 기준으로 정리되어 있지 않고 중복이 많아서 2002-2016년 동안 등록된 특허를 대상으로 정리하였음

① 인용 특허 출원번호(citing)

② 인용 순서(order)

③ 피인용 특허 출원번호(cited_num)

④ 표준 인용 식별코드(cited_code)

⑤ 피인용 특허 인용 구분 코드(cited_typecode)

(4) family.dta: KIPO 패밀리 특허 정보

⑩ 출원번호(appnum)

⑪ 패밀리 특허 순서(order)

⑫ 패밀리 특허 국가(country)

⑬ 패밀리 특허 국가코드(ccode)

⑭ 패밀리 종류(familykind)

⑮ 패밀리 번호(familynum)

⑯ 패밀리 특허 종류(litkind)

⑰ 패밀리 특허 문헌번호(litnum)

⑱ 패밀리 특허 공개번호(opennum)

(5) invt_loc.dta: KIPO 특허 발명자 정보

① 출원번호(appnum)

② 발명자 순서(order)

③ 발명자 한글명(name)

④ 발명자 영문명(engname)

⑤ 발명자 주소(address)

⑥ 발명자 광역자치단체(dist)

⑲ 등록 여부(regi)

- 최종 등록된 특허에 대해 1의 값을 갖는 변수, constructed

(6) RND.dta: KIPO 특허 RND 수여 여부 정보

① 출원번호(appnum)

② 연구과제 순서(order)

③ 연구과제 번호(tasknum)

④ 연구부처(ministry)

⑤ 연구사업명(project)

⑥ 연구과제명(task)

⑦ 주관기관명(inst1)

⑧ 연구 관리 전문 기관명(inst2)

⑨ 과제 기간(term)

⑩ 연구과제 기여율(constribution)

(7) ipc.dta: KIPO 특허 IPC 정보

① 출원번호(appnum)

② IPC(ipc)

③ IPC 부여 날짜(ipcdate)

2. 미국특허청 한국 특허 자료

(1) basicinfo_uspto_kr.dta: 첫 번째 출원인이 한국인 USPTO 특허 기본 정보

① 출원번호(wku)

② 출원 일자(apd)

③ 등록 일자(isd)

(2) assignee_uspto_kr.dta: 첫 번째 출원인이 한국인 USPTO 특허 출원인 정보

① 출원번호(wku)

② 출원인 ID(assgid)

- 출원인 영문명 정리 후 일시적으로 부여된 출원인 ID, constructed

③ 출원인 ID Harmonized(assgidH)

- 매칭 과정에서 추가로 정리된 출원인 ID, constructed

④ 출원인 영문명(engname)

⑤ 출원인 종류(asscode)

⑥ 출원인 국가(cnt)

⑦ 출원인 주소(city)

(3) invt_loc_uspto_kr.dta: 첫 번째 발명자가 한국인 USPTO 특허 발명자 정보

- 발명자 세부 주소에 오류 및 변형이 많아 2002-2017년 동안 등록된 특허를 추려 수작업으로 정리하였음

① 출원번호(wku)

② 발명자 이름(nam_invt)

③ 발명자 국가(cnt)

④ 발명자 세부 주소(cty)

⑤ 발명자 광역자치단체 코드(dist)

⑥ 발명자 광역자치단체 이름(name)

3. 미국 특허청 전체 특허 자료

(1) basicinfo_uspto_all.dta: USPTO 전체 특허 기본 정보

① 출원번호(wku)

② 출원 일자(apd)

③ 등록 일자(isd)

(2) assignee_uspto_all.dta: USPTO 전체 특허 출원인 정보

① 출원번호(wku)

② 출원인 ID(assg_id)

- 자세한 설명은 Lee and Lim(2019) 참조, constructed

① 출원인 영문명1(engname1)

- 출처: XML bulk data 파싱(parsing) 결과 (codes 〉 2_USPTO 〉 basic.py 이용)

② 출원인 영문명2(engname2)

- 출처: USPTO 'ASG NAMES UPRD 69 15NUMSORT.txt' 파일

③ 출원인 종류(asscode)

④ 출원인 국가(cnt)

⑤ 출원인 주(sta)

⑥ 출원인 주소(city)

(3) citation_uspto_all.dta: USPTO 전체 특허 인용 정보

① 인용 특허 출원번호(citing)

② 피인용 특허 출원번호(cited)

(4) invt_loc_uspto_all.dta: USPTO 전체 특허 발명자 정보

① 출원번호(wku)

② 발명자 이름(nam_invt)

③ 발명자 순서(order)

④ 발명자 국가(cnt)

⑤ 발명자 주(sta)

⑥ 발명자 세부 주소(cty)

(5) ipc_uspto_all.dta: USPTO 전체 특허 IPC 정보

① 출원번호(wku)
② IPC(ipc)

4. 매칭 결과
(1) matching_table.dta: 'symbol - kiprisidH'와 'symbol-assgidH' 매칭 결과
① 특허 고객 번호 Harmonized(kiprisidH)
② 출원인 ID Harmonized(assgidH)
③ DataGuide 기업 식별코드(symbol)
④ 매칭 타입(type)
- 해당 DataGuide Symbol을 부여받은 기업이 KIPO에만 특허를 출원했는지, USPTO에만 특허를 출원했는지, 두 특허청에 모두 출원했는지를 식별
⑤ KIPO 출원인과 매칭 단계(phase_kip)
- 법인등록번호 매칭, 문자열 알고리즘 매칭(1단계, 2단계) 중 어느 단계에서 매치된 것인지 식별
⑥ USPTO 출원인과 매칭 단계(phase_uspto)
- 패밀리 특허 이용 매칭, 문자열 알고리즘 매칭(1단계, 2단계) 중 어느 단계에서 매치된 것인지 식별

5. 산업 분류
(1) industry.dta
① IPC(subclass)
② Sub-category(subcat)
③ Category(cat)
④ 추가 여부(added)

Korea Patent Data Project (KoPDP): Contents and Methods*

Jihong Lee** · Hyunkyeong Lim*** · Sangdong Kim**** · Keunsang Song***** · Jae Yu Jung******

Abstract

In this paper, we describe the contents and methods of "Korea Patent Data Project (KoPDP)". The project collects all utility patents granted from the Korea Intellectual Property Office (KIPO) for the period 1948-2016 and the US Patent and Trademark Office (USPTO) for the period 1976-2017. The project also matches their assignees to firms in DataGuide 5.0, a Korean financial database. The resulting dataset includes total 14,803 listed and non-listed Korean firms matched with their Korean and US patents, in addition to a host of accounting and financial information. Over 45% of all sample KIPO patents and 87% of US patents assigned to Korean assignees are matched. We explain the detail of our matching procedures and also provide a coherent industry classification system for both sets of patents.

Key Words: patent data, firm data, innovation
JEL Classification: C80, O30

Received: June 15, 2019. Revised: Sept. 26, 2019. Accepted: Dec. 20, 2019.

* This research was supported by the Park Yang Sook-Chung Yung Ho fund at Seoul National University.

** Corresponding Author, Professor, Department of Economics and Institute of Economic Research, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea, Phone: +82-2-880-6365, e-mail: jihonglee@snu.ac.kr

*** First Author, Graduate Student, Department of Economics, University of Wisconsin-Madison, Phone: +1-608-695-7214, e-mail: hyunkyeong.lim@wisc.edu

**** Co-Author, Graduate Student, Department of Economics, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea, e-mail: tippingpts@snu.ac.kr

***** Co-Author, Graduate Student, Department of Economics, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea, e-mail: sks1107@snu.ac.kr

****** Co-Author, Graduate Student, Department of Economics, Michigan State University, Phone: +1-517-775-6292, e-mail: jungja10@msu.edu