



보도 일시	2023. 2. 12.(일) 낮12시	배포 일시	2023. 2. 10.(금) 14:30
담당 부서	정보고객지원국 정보관리과	책임자	과 장 양기성 (042-481-5134)
		담당자	사무관 송상용 (042-481-5359)

의약품 실험데이터 및 인공지능(AI) 학습데이터 45만 건 무료 개방
- 특허청 '특허정보활용서비스'에서 2.13(월)부터 이용 가능 -

- 특허청(청장 이인실)은 2월 13일(월)부터 특허정보활용서비스(KIPRIS^{Plus}, plus.kipris.or.kr)를 통하여 특허공보에 포함된 의약품 실험데이터 및 실험데이터 추출을 위한 인공지능(AI) 학습데이터 총 45만 건을 무료로 개방한다고 밝혔다.
- 특허정보활용서비스(KIPRIS^{Plus})는 특허청이 운영하는 데이터 개방 유통망(플랫폼)으로서 국내외 13개국 주요 산업재산권(특허·상표·디자인) 공보와 특허행정 정보 등의 데이터 상품을 제공하고 있으며,
 - 최근에는 다국어 번역, 그림(image) 검색 등을 위한 인공지능(AI) 학습데이터까지 개방하여 총 115종의 데이터 상품을 파일 또는 공개에이피아이(OpenAPI*) 형태로 개방하고 있다.
- * Open Application Programming Interface : 표준화하여 구축된 데이터를 실시간으로 호출하여 누구나 쉽게 활용할 수 있는 공개 인터페이스
- 이번에 개방하는 의약품 실험데이터는 의약품 분야의 특허공보에 포함된 그림(image) 형태의 표를 가공·분석하여 의약품 성분명, 실험방법, 실험값 등을 기초자료(데이터베이스)로 구축한 정보이고,

- 인공지능(AI) 학습데이터는 ① 표·그래프·화학식 등 다양한 그림 (image) 형태의 실험데이터에서 표 형태만을 추출하기 위한 그림 (image) 분류 정보, ② 표 데이터를 정확하게 파악하기 위한 표 구조 (행×열) 정보 및 ③ 실험에 사용된 성분명, 실험값 등을 자동분류하기 위한 실험데이터 속성 분류 정보를 포함하고 있다.
- 이러한 의약품 실험데이터 및 인공지능(AI) 학습데이터를 활용하면, 특허·논문 등에 포함되어 있는 그림(image) 형태의 ‘표 실험데이터’를 문자(text) 형식의 분석 가능한 데이터로 변환 추출할 수 있다.
- 이를 기반으로 지식재산 서비스 업체들은 특허공보의 실험데이터 추출·활용 서비스를 개발하고, 관련 기업 및 연구기관들은 특허 공보에 포함된 실험 실시예, 비교예 등을 자유롭게 분석 활용하여 백신·신약 등 연구개발에 활용할 수 있을 것으로 기대된다.
- 특허청 김기범 정보고객지원국장은 “기술 정보의 보고인 특허 데이터는 실험데이터와 같이 유용한 정보가 많아 공공 및 민간에서 적극 활용한다면, 미래를 이끌어 갈 첨단 기술 분야에서 우리나라의 기술개발 역량이 제고되고 세계 경쟁력은 더욱 강화될 것”이라며
- “특허청은 적극행정의 일환으로 국민의 수요에 맞춘 다양한 지식 재산 데이터를 지속적으로 개방해 나갈 것”이라고 밝혔다.

※ 붙임: 의약품 실험데이터 및 인공지능(AI) 학습데이터 개요

의약품 실험데이터 및 인공지능(AI) 학습데이터

1. 개요

의약품 분야 실험데이터 및 특허 그림(이미지) 분류 학습데이터를 대용량(Bulk)으로 제공

2. 세부정보

특허에 포함된 의약품 성분 및 실험정보를 추출하여 기초자료(데이터 베이스)로 구축한 의약품 실험데이터 및

의약품 분야 특허 명세서에 포함된 이미지(표) 실험데이터를 인식 시키기 위한 인공지능 학습데이터

< 인공지능(AI) 학습데이터 제공 정보 >

- ① (그림(이미지) 분류) 국내 특허공보에 첨부되어 있는 그림(이미지) 유형 (표, 그래프, 화학식, 흐름도 등) 분류 정보
- ② (표 구조) 표 데이터를 정확하게 파악하여 문자(텍스트) 정보를 추출하기 위한 표의 구조(행×열) 정보
- ③ (실험데이터 속성 분류) 실험에 사용된 성분, 실험명칭·값 등을 자동 분류 하기 위한 학습데이터

설명(메타) 정보	분류체계	국내 IP데이터 > 부가공통 > 국내공통
	제공기관	특허청
	업데이트주기	미정
	상품개시일	2023년 2월
	태그	인공지능(AI)
세부 정보	서비스유형	대용량(벌크)데이터
	데이터 건수	44.7만건
	데이터 제공방식	대용량(DB Bulk)
	데이터세부사항	이미지(표), 표 구조정보, 속성정보 포함
	데이터수수료	무료