**공통프로젝트 기획안(국방과제 해결을 위한 웹크롤링 자동화프로젝트)**

작성일자 : 2025년 4월 30일

| 조 명 | 안되면 되게 해조 |
| --- | --- |
| 조 원 | 조장: 이인철  조원: 김태영, 손태인, 이수현, 하은지 |
| 프로젝트 주제 및 개요 | 0. 주제 : 국내외 방산분야 상위 기업 특허 분석을 통한 기술 편중 및 수준 비교  1. 배경:   1. 글로벌 동향  * AI·사이버 기술의 전략적 가치: 러-우 전쟁의 우크라이나는 AI 기반 무인기(자율항법), 전자전(EW), 예측물류 등으로 전장 우위 확보 * 특허·R&D 경쟁: 2023년 글로벌 방산 AI 시장 R&D 부문 320억 달러 투자, 예상 연평균 성장률 14.2%(’25~’30년)  1. 국내 현주소:   2-1) 투자 부문   * 방위사업특례제도, 미래국방가교기술개발 사업 신설 * 민간분야 벤처기업 R&D 지원, 국방기술로의 연계 시도   2-2) 기술격차 부문   * AI·사이버 분야 기술격차 4.1년(과학기술정보통신부, ’23) * ‘지능형 전장 인식과 판단’ 분야에서는 12개 정부 주도 사업 중 4개 사업만 연구에 착수, 나머지 대다수 사업은 3~4년뒤에 연구가 진행될 예정 * 국내 방산 선도기업 한화에어로스페이스, LIG넥스원 등의 일부 기술 및 소재 개발 해외 의존도 심화       2. 프로젝트 목표:  - 방산 시장에서 시장 영향력이 크며 방산기술의 전반적 흐름과 방향을 파악하는데 효과적인 매출이 높은 기업 선정.  - 매출 상위권 기업과 방위사업청의 방산기업 현황 데이터를 통해, 현 한국 방산 기업의 대부분 기술 분야를 확인 가능한 11개의 기업을 선정.  - 방위산업 분야의 연구개발 및 특허 출원은 단순한 민간 기업의 자발적 활동을 넘어, 국가 안보와 전략적 자립을 위한 국가 주도의 정책적 방향과 밀접하게 연관.   * 국내외 방산업체 매출 상위 11개 기업 특허기술 수집 * 국가과학기술표준분류-IPC 코드에 기반한 수집 데이터 기술 분야 분류 * 국내외 방산기업의 주요 기술 분야별 특허 출원 시계열 분포 시각화 * 국내외 기술 분야 비율 시각화 * 국내외 특허 기술 발전 동향 및 기술편중 확인 * 국방과학 기술 수준이 높은 해외 방산업체와의 비교   3. 타겟:   1. 국가, 유관 기관  * 방위사업청 * 국방부 및 산업통상자원부 등 국가 관련 기관  1. 민간기업  * 방산중소벤처기업협회 * 방위산업 관련 산학협력단 * 방산기술혁신펀드[[1]](#footnote-0)   - 벤처기업 R&D 부서  4. 기대효과:   1. 국가, 유관 기관  * 방위사업청의 미래국방가교기술개발사업, 방산기업 원스톱 지원센터의 서비스와 같은 기술 자립, 미래전 대비 기술 확보를 목표로 한 정부의 전략적 지원체계 보조 * 국내외 방산기업 투자 수준과 방산 기술의 발전 속도와 트렌드를 파악하고 동향 분석 * 민간 기업과 협업을 통한 기술사업화 전략 수립 보조 * 효율적인 R&D 관련 투자 예산 배분  1. 민간기업  * 벤처기업 육성을 위한 특허현황 조회 서비스 제공 * 대학 TLO(기술이전센터) 및 산학협력단이 특허 현황을 실시간 공유하여 기술이전, 라이선싱, 공동연구 활용 * 성장 잠재력이 높은 분야를 선제적으로 발굴, 민간 투자 유치로 산업 경쟁력을 강화 |
| 프로젝트 수행 방향 및 내용 | ● 프로젝트 내용  1. 데이터 수집  - 선정 기준  - 매출:  매출 규모는 방위부문에서 글로벌 시장의 영향력이 크며, 글로벌 네트워크 및 수출역량을 나타내는 지표로 기술의 트렌드의 전반적인 흐름과 방향을 파악할 수 있음.  - 특허 기술:  방산 기술은 폐쇄적이고 민감한 분야이기 때문에, 외부 정보 부족을 보완하는 민간 공개 정보원인 특허 기술을 참고   1. 데이터 수집 대상   (1) 방산 기업의 특허 데이터  1-1) 국내 데이터  주요 분석 대상 국내 기업 (24년 4분기 기준 매출액 Top 11):  공공데이터포털의 ‘방위사업청\_방산업체 지정현황’ 을 참고하여 선정  목록 :  STX엔진, SK오션플랜트, KAI, HD현대인프라코어, 현대트랜시스, 현대위아, 현대로템, 한화에어로스페이스, 한화시스템, 풍산, LIG넥스원  12,024개의 특허 기술을 KIPRIS 지식 재산 정보 검색 서비스(  https://www.kipris.or.kr/khome/main.do)에서 수집  1-2) 해외 데이터  주요 분석 대상 해외 기업(23년도 기준 매출액 Top 11):  스톡홀름국제평화연구소의 세계 방산매출 순위 top100 을 참고하여 선정 목록 :  Lockheed Martin Corporation(미), RTX corporation(미), Northrop Grumman Corporation(미), The Boeing Company(미), General Dynamics Corporation(미), BAE Systems(영), Rostec(러), AVIC(중), NORINCO(중), CETC(중), L3 Harris Technologies(중)  100,806개의 특허 기술을 WIPO-patentscope search 에서 수집  https://patentscope.wipo.int/search/en/search.jsf  (2) 특허 분류 기준   * 특허청의 국가과학기술 - 특허 연계표의 IPC 분류 코드로 특허 기술의 카테고리를 분류 * IPC코드 중심 분류: 예) A01B(로봇/자동화기계), B01D(신재생에너지) 등 * 출처 : 국가과학기술표준 - IPC, CPC   https://www.kipo.go.kr/ko/kpoContentView.do?menuCd=SCD0200273   1. 데이터 수집 경로   2-1) 특허기술  국내 - 특허정보넷 키프리스(KIPRIS): <https://www.kipris.or.kr>  해외 - WIPO (국제특허기구):<https://www.wipo.int>  2-2) 기업 목록  국내 - 공공데이터포털 : <https://www.data.go.kr/>  해외 - 스톡홀름국제평화 연구소의 세계 방산 매출 순위 top100   1. 데이터 수집 방법   3-1) 파일 기반 수집  상위 매출 11개 기업을 KIPRIS에 검색하여 직접 수집  3-2) 웹 크롤링  Python을 사용한 크롤링 진행  Selenium의 webdriver로 WIPO url에 연결  기술명, 기술요약, IPC 분류 코드, 기업명, 출원일자 수집  4) 수집 항목 예시  4-1) KIPRIS 특허 기술 수집 예시      4-2) WIPO 특허기술 수집 예시      5) 수집 범위  5-1) KIPRIS 검색 범위    5-2) WIPO 검색범위    Int.Class (IPC 분류 코드), 특허 기술명(중앙 상단), Applicant(지원자), 출원 날짜(우측 상단), 특허 기술 요약(중앙 글귀)  2. 전처리   1. 특허 기술   1-1) KIPRIS  수집 기준으로 수집한 원본 데이터 (순번 | 발명의명칭 | 발명의명칭(영문) | IPC분류 | 출원일자 | 출원인 | 등록일자 | 법적상태 | 원출원일자 | 요약 | 발명자 | 최종권리자 | 청구항) 에서 일부 추출  1. 발명의 명칭  2. IPC분류 : 기술 분야 분류를 위해 앞의 4글자만 추출 (Ex: G01R 31 -> G01R)  3. 출원일자 : 날짜 타입으로 바꾸기 (Ex: 2024-12-31)  4. 출원인 : 검색한 기업명으로 통합  (Ex: 아즈텍 주식회사(전라남도 영암군...)|한화시스템 주식회사(경북 구미시 … -> 한화시스템)  5. 요약 : 기술 요약  6. 기술분류 : 국가과학기술표준분류에서 IPC분류 코드가 속한 중분류로 새로운 컬럼 추가 (Ex: G06F -> 소프트웨어)  11,616개의 데이터로 전처리  1-2) WIPO  크롤링으로 수집한 데이터(IPC | date | tech\_name | tech\_desc | company) 전처리  1. tech\_name : 발명의 명칭  2. IPC : 기술 분야 분류를 위해 앞의 4글자만 추출, IPC 분류가 없는 데이터는 삭제 ( Ex: B60N 2/02 -> B60N)  3. date : 출원일자, 날짜 타입으로 바꾸기 (Ex: 2024-12-31)  4. company : 출원인, 대소문자, 언어 구분없이 기업명이 들어간 행만 추출 (Ex: AVIC HARBIN AIRCRAFT INDUSTRY GROUP CO., LTD -> AVIC)  5. tech\_desc : 기술 요약  6. 기술분류 : 국가과학기술표준분류에서 IPC분류 코드가 속한 중분류로 새로운 컬럼 추가 (Ex: G06F -> 소프트웨어)  91,831개의 데이터로 전처리  3. 분석(EDA)  - 연도 별 ICT 분류 코드, 기술 분류 빈도수 측정  - 기술 분야 별 특허 기술 요약 토큰화, 단어 빈도수 측정  - 국내/국외 특허 기술의 연도 별 ICT 분류 코드 비중 비율 시각화  - 국내/국외 특허 기술 출원 시기 시각화  - 기술 분야 별 특허 기술 요약 단어 시각화  4. 시각화  - 국내외 기술 분야 별 비중 비율 막대 그래프  - 국내외 기술 분야 별 파이차트  - 국내, 국외 기업 별 특허 기술 분야 시계열 분포 (산점도)  - 국내, 국외 기업 기술의 연도별 출원, 누적 출원 수 선 그래프  - 최신 특허 기술(2019년 이후) 상위 10개 분야 누적 막대 그래프  - 국내외 출원 수 시계열 그래프  - 국내, 국외 기술 분야 별 특허 기술 요약의 워드 클라우드  5. 대시보드 화면  예시:    ● 데이터 소개  - 수집 출처  - 방산업체 목록:   방위사업청\_방산업체 지정현황(수정일 : 2024-04-11)  스톡홀름 국제 평화 연구소- 세계 100대 방산기업 매출 순위  - 기술 분류 :  국가과학기술-특허(IPC,CPC) 연계표  (https://www.kipo.go.kr/ko/kpoContentView.do?menuCd=SCD0200273)  - 방산업체 매출 확인:  전자공시 시스템 다트(<https://dart.fss.or.kr/>)  기업 정보 확인(https://comp.fnguide.com)  - 특허 기술:  KIPRIS(<https://www.kipris.or.kr/khome/main.do>)  WIPO(<https://patentscope.wipo.int/search/en/search.jsf>)  - 데이터 설명  1) 국내 방산기업 특허 기술:  - 대상 : 방산업체 매출 상위 11개 기업  - 형식 : xlsx  - 수집 방법 : 파일 기반 수집  - 내용 :  발명의 명칭, 명칭 영문, IPC분류, CPC분류,  출원일자, 출원인, 등록번호, 등록일자, 공고일자,  법적상태, 원출원일자, 요약, 발명자, 최종권리자, 청구항  - 분량 : 약 12,000개의 특허 기술 정보  2) 해외 방산기업 특허 기술:  - 대상 : 방산업체 매출 상위 11개 기업  - 형식 : csv  - 수집 방법 : 웹 크롤링  - 내용 :  발명의 명칭, Applicant(기업명), date(출원일자),  Int.Class(IPC 분류 코드), 기술 설명 요약  - 분량 : 약 100,000개의 특허 기술 정보  3) IPC 분류  - 대상 : 국가과학기술표준분류 - 특허분류(IPC,CPC)  - 형식 : xlsx  - 수집 : 파일 기반 수집  - 내용 : IPC 분류 코드와 분류 코드가 의미하는 대분류, 중분류, 소분류  - 데이터 모델링  DB 예시:   | IPC | year | name | tech\_name | tech\_des | | --- | --- | --- | --- | --- | | F41A | 2017 | 다산기공 주식회사 | 총기용 소염기 | 기술 내용 요약1 | | D06F | 2009 | 다산기공 주식회사 | 더블 셔츠 프레스 머신 | 기술 내용 요약2 | | H04B | 2023 | 단암시스템즈 주식회사 | 무선 신호 송신을 위한 자동 이득 제어 방법 및 그를 위한 장치 | 기술 내용 요약3 |   ● 수행 도구  Python : 전처리, 웹 크롤링, 시각화에 사용  Excel : 파일 기반 수집 데이터 확인에 사용xcel : 파일 기반 수집 데이터 확인에 사용 |
| 프로젝트 조직  (구성원 및 역할) | ● 역할분담  조장: 이인철 - 문서 작업(기획서, 매뉴얼, 데이터 명세서, PT자료), 웹 페이지 크롤링, PT 발표, 대시보드 설계 및 구현  조원1: 손태인 - 문서 작업(기획서, 매뉴얼, 데이터 명세서), 데이터 수집, 웹 페이지 크롤링, 데이터 저장, 대시보드 구현  조원2: 김태영 - 문서 작업(회의록, PT자료), 데이터 수집, 웹페이지 크롤링, 대시보드 설계, 발표 PT 제작  조원3: 이수현 - 데이터 및 텍스트 전처리, 정제, 변환 작업, 시각화, 대시보드 구현  조원4: 하은지 - 데이터 및 텍스트 전처리, 정제, 변환 작업, 시각화 , 대시보드 설계 |
| 프로젝트  추진 일정 | ● 일정  4/28(월) : 주제 선정 및 일정 수립  4/28 ~ 4/30 : 크롤링 대상 선정, 분류 기준 선정, 사전조사(코딩 결과물 예시 프로토타입 생성 및 크롤링 대상 홈페이지 보안 확인)  5/1 ~ 5/4: KIPRIS에서 데이터 수집 및 전처리, WIPO 크롤링 자동화 코드 작성  5/5 ~ 5/6: WIPO에서 수집한 데이터 전처리, SQL DB에 저장  5/7 ~ 5/8: 전처리된 데이터 시각화 작업  5/9 ~ 5/10: 대시보드 구현  5/11 ~ 5/12: 발표 PPT 제작, 대본 작성 |

<참고 문헌>

# 1.정보통신기획평가원. (2022). 『글로벌 패권 경쟁 관점의 국방기술 혁신 동향』. 정보통신기획평가원. 2.한국개발연구원(KDI). (2022). 『한국 방위 산업, 어느 수준에 와 있을까?』. KDI 경제정보센터. 3.삼일PwC경영연구원. (2024). 『키워드로 보는 방위산업의 현재와 미래』. 삼일회계법인.[PwC+3PwC+3한국병무정책학회+3](https://www.pwc.com/kr/ko/insights/industry-focus/defense-industry.html?utm_source=chatgpt.com) 4.손창호. (2018). 『데이터마이닝을 통한 방위산업기술 분석 연구: 특허분석을 중심으로』. 한국산학기술학회논문지, 19(10), 101-107.[DBpia+2KCI+2ScienceON+2](https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002396599&utm_source=chatgpt.com) 5.국회도서관. (2023). 『국내 방위산업의 지식재산권 경쟁력 강화 방안』. 국가전략정보포털.[국가전략정보포털+1국가전략정보포털+1](https://nsp.nanet.go.kr/plan/subject/detail.do?nationalPlanControlNo=PLAN0000049066&utm_source=chatgpt.com) 6.국방기술진흥연구소. (2024). 『2024 국가별 국방과학기술 수준 조사서』. 국방기술진흥연구소.[코시스](https://www.kocis.go.kr/koreanet/view.do?seq=1050548&utm_source=chatgpt.com) 7.방위사업청. (2023). 『지역 방산 중소기업의 현장 목소리에 방사청과 지자체가 함께 응답하다』. 방위사업청 보도자료. 8.Precedence Research. (2024). 『AI in Aerospace and Defense Market Size, Share and Trends 2024 to 2034』. Precedence Research.[Precedence Research](https://www.precedenceresearch.com/ai-in-aerospace-and-defense-market?utm_source=chatgpt.com)

# 9.스톡홀름국제평화연구소(SIPRI). (2023). 『세계 100대 방산기업』. SIPRI.[뉴스스페이스+1조선일보+1](https://www.newsspace.kr/news/article.html?no=4697&utm_source=chatgpt.com)

# 10.김성진. (2023). 『절충교역 관점에서 본 한국의 방위산업 및 무기체계 국제거래 현황과 과제』. 한국국방연구원.

<참고 기사>

1.김형민. (2025년 2월 17일). 미래戰 핵심기술 방산AI 외국 의존 심각… 주요국과 4.1년 격차. 『동아일보』.

2.김선아. (2025년 1월 22일). 한국 국방과학기술 세계 8위···미국 대비 82% 수준. 『코리아넷뉴스』. [코시스](https://www.kocis.go.kr/koreanet/view.do?seq=1050548&utm_source=chatgpt.com)

3.김정영. (2024년 8월 7일). [단독] 현대로템, 세계 방산 기업 ‘톱100’ 신규 진입…KAI '제외'. 『더구루』.[더구루+1뉴스스페이스+1](https://www.theguru.co.kr/news/article.html?no=75473&utm_source=chatgpt.com)5.김정영. (2025년 2월 19일). 세계 방산 톱 100에 한국 기업 4곳 포진. 『조선일보』

1. 한국성장금융투자운용,주요 자펀드 운용사, 국방기술진흥연구소 등 [↑](#footnote-ref-0)