

6조 프로젝트 매뉴얼

목차

1. 데이터 수집

1-1. KIPRIS

1-2. WIPO

1-3. IPC 분류

2. 데이터 전처리

2-1. 국내 데이터 전처리

2-1-1. 추출

2-1-2. 정제 및 변환

2-1-3. 추가

2-2. 해외 데이터 전처리

2-2-1. 컬럼명 변경

2-2-2. 정제 및 변환

2-2-3. 기술분류 컬럼 추가

3. 시각화

3-1 막대그래프 상위 20개 기술분류별 비율 국내, 해외 비교

3-1-1. 데이터 처리

3-1-2. 그래프 생성 과정

3-1-3. barplot(막대그래프) 축 정보

3-1-4. 그래프 분석, 전략적 인사이트

3-2 막대 국내외 연도별 상위 10개 기술 누적 특허 출원 수(2020~)

3-2-1. 데이터 처리

3-2-2. 그래프 생성 과정

3-2-3. barplot(막대그래프) 축 정보

3-2-4. 그래프 분석, 전략적 인사이트

3-3) 산점도 - 국내외 기업별 특허 출원일자와 특허기술(2004~)

3-3-1. 데이터 처리

3-3-2. 그래프 생성 과정

3-3-3. scatter plot(산점도) 축 정보

3-3-4. 그래프 분석, 전략적 인사이트

3-4) 선그래프 - 국내외 기술별 누적 특허 출원 수

3-4-1. 데이터 처리

3-4-2. 그래프 생성 과정

3-4-3. line plot(선 그래프) 축 정보

3-4-4. 그래프 분석, 전략적 인사이트

3-5) 선그래프 - 국내외 월별 특허 출원 수(2004~)

3-5-1. 데이터 처리

3-5-2. 그래프 생성 과정

3-5-3. line plot(선 그래프) 축 정보

3-5-4. 그래프 분석, 전략적 인사이트

3-6) 파이차트 - 국내외 상위 10개 기술 분포

3-6-1. 데이터 처리

3-6-2. 그래프 생성 과정

3-6-3. 그래프 분석, 전략적 인사이트

3-7) 워드 클라우드 - 국내외 기술 분야별 특허 기술 요약

3-7-1. 데이터 처리

3-7-2. 그래프 생성 과정

3-7-3. 그래프 분석, 전략적 인사이트

<1. 데이터 수집 매뉴얼>

프로젝트 진행을 위해 필요한 데이터 수집 내용과 방식을 작성한 매뉴얼이다.

수집 방법과 수집 예시를 설명한다.

- 파일 기반 데이터 수집 (KIPRIS, IPC 분류 코드 데이터 수집에 활용한 기법) :

기관이나 시스템이 제공하는 데이터를 파일 형태(XML, CSV, XLSX 등)로 내려 받아 로컬이나 서버에 저장한 후, 가공/분석하는 방식의 데이터 수집 방법

사이트 선택 -> 회원가입 및 신청 -> 데이터 수신(제공 데이터 다운로드) -> 데이터 파싱(XML 또는 CSV로 변환 및 처리)의 작업을 거쳐 수집한다.

- 웹 크롤링 (WIPO에서 데이터 수집에 활용한 기법) :

웹 페이지의 HTML 페이지를 자동으로 방문하고, 그 안의 정보를 추출해 수집하는 기술

웹 상의 여러 페이지를 자동으로 순회하며, 정보 수집, 데이터베이스 구축을 위해 활용한다.

1-1) KIPRIS (파일 기반 수집)

KIPRIS는 대한민국 특허 관련 정보를 검색하고 열람할 수 있는 공식 온라인 특허 정보 포털이다. 지식재산권(특허, 실용신안, 디자인, 상표 등) 정보를 무료로 제공받을 수 있다.

이 포털에서 한국 방산기업 매출 상위 11개 기업을 선정하여 파일 기반 데이터 수집 방법으로 특허 기술 데이터를 수집한다.

선정한 기업은 '현대트랜시스, 한화에어로스페이스, 현대위아, 현대로템, 한국항공우주산업, 풍산, LIG넥스원, 한화시스템, HD현대인프라코어, STX엔진, SK오션플랜트' 이다.

(방산 기업 목록은 공공 데이터 포털의 '방위사업청_방산업체 지정현황' 데이터에서 추출.)

(매출은 각 기업의 2024년 4분기 사업보고서를 통해 방산분야 매출을 확인하였다.)



KIPRIS 메인 화면

지식재산정보 검색 서비스에서 기업명을 검색하여 법적상태가 공개, 등록된 특허 기술만 필터링하고 하단의 검색 범위를 기준으로 수집한다.

발명의명칭 | IPC 분류 | CPC분류 | 출원일자 | 출원인

등록번호 | 등록일자 | 공고일자 | 법적상태(법적상태가 등록과 공개인 값만)

원출원일자 | 요약 | 발명자 | 최종권리자 | 청구항

저장 항목 검색 범위

- 수집 데이터 예시

컬럼명	내용
발명의명칭	주파수 보상 방법 및 이를 이용한 저궤도 위성용 에스-밴드 수신기
발명의명칭(영문)	method for frequency compensation and a s-band receiver for low earth orbit satellite using the method
IPC분류	H04B 1/12(2006.01) H04B 1/16(2006.01)
CPC분류	H04B 1/12(2013.01) H04B 1/16(2013.01)

출원일자	2010.08.13
출원인	(주)씨트렉아이(대전광역시 유성구...)
등록번호	1011313850000
등록일자	2012.03.22
공고일자	2012.04.03
법적상태	등록
원출원일자	
요약	본 발명은 주파수 보상 방법 및 이를 이용한 저궤도 위성용 에스-밴드 수신기에 관한 것으로, 에스-밴드(S-band) 신호를 수신하는 신호 수신부; 인가되는 전압에 따라 국부 주파수를 발진하는 국부 발진부; ... (요약)
발명자	박인용(대전광역시 서구...)
최종권리자	(주)씨트렉아이(대전광역시 유성구...)
청구항	[청구항5]제1항 또는 제3항 또는 제4항의 상기 에스-밴드 수신기를 포함한 저궤도 위성.

1-2) WIPO (웹 크롤링)

WIPO는 국제 지식 재산권 보호와 협력을 위한 유엔(UN) 산하의 국제기구이다.

특허, 상표, 디자인 등 지식재산권을 국제적으로 보호하고 조율하기 위해 설립된 UN 산하 국제 기구이다.

해외 방산기업 매출 상위 11개 기업의 특허 기술 정보를 수집한다.

(방산 기업 목록은 스톡홀름국제평화 연구소의 '세계 방산 매출 순위 TOP 100'에서 발췌.)

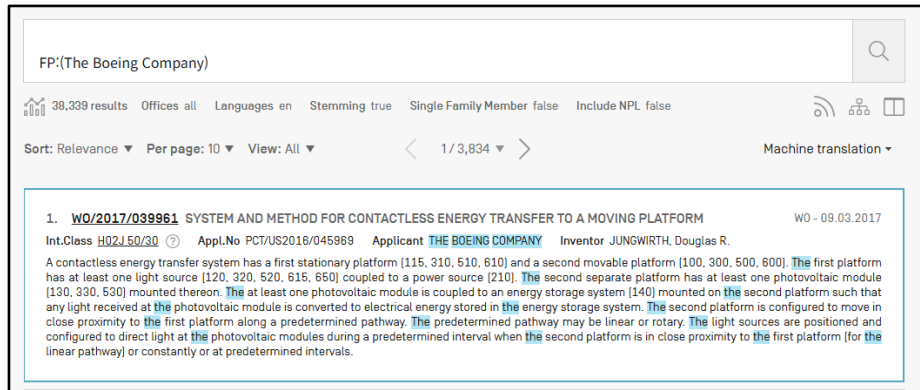
The screenshot displays the WIPO PATENTSCOPE Simple Search page. At the top, there is a header with the WIPO logo and navigation links. Below this, a search bar is visible. The main content area is titled 'PATENTSCOPE Simple Search' and contains a brief description of the database, including the number of patent documents and the availability of PCT publications. A search input field is located at the bottom of the page, with a dropdown menu for 'Field' and a 'Search terms...' input box. The page also includes links for 'Feedback', 'Search', 'Browse', 'Tools', and 'Settings'.

WIPO 검색 화면

- 크롤링 코드

WIPO 특허 정보 자동 수집을 설계한 프로그램 코드이다.

Int.Class(IPC 분류 코드), 특허 기술 명(영문), 출원 일자, Applicant(기업), 특허 기술 요약 내용을 수집한다.



검색 내용 예시

내용 요약:

4) : crawl_wipo(회사명, 페이지 수) -> 사용자가 지정한 회사 이름으로 정한 페이지 수 만큼 특허 기술 검색

5) ~ 8): 크롬(Chrome 실행)

13 ~ 14): WIPO 검색 사이트 접속

17) ~ 25): 기업명을 검색하고 특허 기술(Patent) 페이지로 이동

35) ~ 41): 기본값인 한 페이지에 10개씩 조회에서 200개씩 조회로 변경

44) ~ 54): 현재 페이지 html 정보 수집 -> 특허 기술 이름, 설명, 출원 일자, IPC, 기업명 이 담긴 정보 수집

56) ~ 86): 수집한 정보(특허 기술 명, 특허 기술 요약, 출원일자, IPC 분류 코드, 기업명)를 기록

90) ~ 99): 사전에 정한 페이지 수만큼 반복

101) ~ 112): '(기업명)_크롤링.csv' 파일로 저장, 수집한 데이터 양 출력, 종료

```
2- # WIPO 페이지 특허 기술 크롤링 코드
3- def crawl_wipo(company_name, max_pages=3):
4-     service = Service(ChromeDriverManager().install())
5-     options = webdriver.ChromeOptions()
6-     driver = webdriver.Chrome(service=service, options=options)
7-     wait = WebDriverWait(driver, 10)
8-     results = []
9-
10-     try:
11-         # url 입력
12-         driver.get('https://www.wipo.int/search/en')
13-         wait.until(EC.presence_of_element_located((By.ID, 'query')))
```

```

14-
15-     # 기업명 검색
16-     search_element = driver.find_element(By.ID, 'query')
17-     search_element.clear()
18-     search_element.send_keys(company_name)
19-     wait.until(EC.element_to_be_clickable((By.XPATH, '//*[@id="button-addon2"]'))).click()
20-
21-     time.sleep(10)
22-
23-     # 특허 기술로 이동
24-     wait.until(EC.element_to_be_clickable((By.XPATH, '//*[@id="patentscopeEntry"]/a'))).click()
25-
26-     # 새 탭으로 전환
27-     original_window = driver.current_window_handle
28-     for window in driver.window_handles:
29-         if window != original_window:
30-             driver.switch_to.window(window)
31-             break
32-
33-     page = 1
34-     # per page 클릭
35-     driver.find_element(By.XPATH, '//*[@id="resultListCommandsForm:perPage:input"]').click()
36-     time.sleep(3)
37-
38-     # 페이지당 200개로 전환
39-     driver.find_element(By.XPATH, '//*[@id="resultListCommandsForm:perPage:input"]/option[4]').click()
40-     time.sleep(3)
41-
42-     # html 태그로 크롤링
43-     while page <= max_pages:
44-         print(f"[{company_name}] 크롤링 중... {page} 페이지")
45-
46-         wait.until(EC.presence_of_element_located((By.ID, 'resultListForm:resultTable_data')))
47-         html = driver.page_source
48-         soup = BeautifulSoup(html, 'html.parser')
49-         tech_list = soup.find('tbody', {'id': 'resultListForm:resultTable_data'})
50-
51-         if not tech_list:
52-             print(f"[{company_name}] 데이터 없음 (페이지 {page})")
53-             break
54-
55-         adj_tech_list = tech_list.find_all('tr')
56-         for idx, row in enumerate(adj_tech_list):
57-             try:
58-                 # 기술명
59-                 tech_name_tag = row.find('span', {'class': 'trans-section needTranslation-title'})
60-                 # 기술 요약
61-                 des_tag = row.find('span', {'class': 'trans-section needTranslation-biblio'})
62-                 # 출원 일자
63-                 date_tag = row.find('span', {'id': f'resultListForm:resultTable:{idx}:resultListTableColumnPubDate'})
64-                 # IPC 분류 코드
65-                 code_tag = row.find('a', {'target': '_blank'})
66-                 # 기업명
67-                 company_tag = row.find('span', {'class': 'ps-field--value ps-patent-result--applicant notranslate'})
68-
69-                 # 기술명
70-                 tech_name = tech_name_tag.text.strip() if tech_name_tag else ""
71-                 # 기술 요약
72-                 tech_des = des_tag.text.strip() if des_tag else ""
73-                 # 출원 일자
74-                 tech_date = date_tag.text.strip() if date_tag else ""
75-                 # IPC 분류 코드
76-                 code = code_tag.text.strip() if code_tag else ""

```

```

77-         # 기업명
78-         company = company_tag.text.strip() if company_tag else ""
79-
80-         # 결과 묶기
81-         results.append([code, tech_date, tech_name, tech_desc, company])
82-
83-     except Exception as e:
84-         print(f"[{company_name}] 파싱 실패 (페이지 {page}, 항목 {idx}): {e}")
85-         continue
86-
87-     # 다음 페이지 버튼 클릭
88-     try:
89-         next_btn = wait.until(EC.element_to_be_clickable((By.ID, 'resultListCommandsForm:j_idt1991')))
90-         next_btn.click()
91-         page += 1
92-         time.sleep(4)
93-     except Exception:
94-         print(f"[{company_name}] 다음 페이지 없음 또는 클릭 실패 (페이지 {page})")
95-         break
96-     # 다음 페이지 버튼 못누르면 멈춤
97-     if soup.find('a', {'id': 'resultListCommandsForm:j_idt1991'}) == None:
98-         break
99-
100-    # 저장
101-    df = pd.DataFrame(results, columns=['IPC', 'date', 'tech_name', 'tech_desc', 'company'])
102-    df.to_csv(f"[{company_name}] 크롤링.csv", index=False, encoding='utf-8-sig')
103-    print(f"[{company_name}] 크롤링 완료: 총 {len(df)}건 저장됨.")
104-    return results
105-
106-    except Exception as e:
107-        print(f"[{company_name}] 전체 에러 발생: {e}")
108-        return results
109-
110-    finally:
111-        driver.quit()

```

- 수집 데이터 예시

컬럼명	내용
IPC	E03B 9/08
date	14.07.2011
tech_name	VALVE BOX HAVING MEANS FOR BLOCKING THE COLLAR THEREOF RELATIVE TO THE BASE OF THE COLLAR, COLLAR AND BASE, AND METHOD FOR INSTALLING THE VALVE BOX
tech_desc	The present invention relates to a valve box having means for blocking the collar thereof relative to the base of the collar. According to the invention, the collar (3) of the valve box (1) includes at least one flat section (8a), the opening (16) of portion (6) of the base (2) includes at least one planar portion forming a flat section (16a), and the O-ring (17) is housed in a groove (18), ... (요약)
company	NORINCO

1-3) IPC 분류 (파일 기반 수집)

특허청에서 제공하는 국가과학기술-특허(IPC,CPC) 연계표를 내려 받는다.

과학기술 관련 정보/인력/연구개발사업 등을 효율적으로 관리할 수 있도록 구축되어 있는 국가과학기술표준분류와 특허통계의 집계기준인 국제특허분류 (IPC,CPC)간 연계표를 통해 과학기술과 특허정보의 상호 연계분석이 가능하도록한다.

수집한 특허 기술 데이터의 기술 분류를 위해 사용한다.

- 수집 데이터 예시

[illegible]

<2. 데이터 전처리 매뉴얼>

수집한 데이터 전처리(추출, 정제, 변환, (변수) 추가 등) 과정을 작성한 매뉴얼이다.

KIPRIS에서 수집한 국내 상위 매출 11위 방산기업의 특허 기술 데이터와

WIPO에서 수집한 해외 상위 매출 11위 방산기업의 특허 기술 데이터 전처리 내용을 설명한다.

KIPRIS 데이터는 11,616개로 전처리 되었고 WIPO 데이터는 91,831개로 전처리 되었다.


2-1) 국내 데이터 전처리

2-1-1. 추출

수집한 KIPRIS 특허 기술 데이터에서 “발명의 명칭, IPC분류, 출원일자, 출원인, 요약” 컬럼
(column, 표의 세로축, 열을 의미 이후에도 컬럼이라고 표현)만 추출

5개의 컬럼 중 발명의 명칭, 요약 컬럼은 원본 데이터의 핵심 내용을 담고 있기 때문에 원본을 유지한다.

- 예시

항목	내용	
발명의 명칭	항공기 추진 로터의 블레이드를 작동시키기 위한 구동 장치	
IPC 분류	B64D 27/24, H02P 5/00, H02K 11/20, B64D 45/00	
출원인	(주)경인테크, 한화에어로스페이스(주)	
등록일자	2024.12.26	
요약	하우징, 롤러스크류, 2개의 직렬 연결된 모터, 모터 모니터링부, 모터 제어부로 구성된 블레이드 작동 구동 장치. LVDT 검출 센서로 위치를 정밀 측정하여 정밀 제어 수행. 센서부를 통해 출력전류, 온도, 습도, 진동주파수 측정. 이상 시 비상 구동 요청 및 제어 가능.	

2-1-2. 정제 및 변환

데이터에 오류가 있는 부분을 보완하는 정제, 데이터의 형태를 바꿔 분석/모델에 맞도록 하는 변환 작업의 요약 내용이다.

IPC분류, 출원일자, 출원인 컬럼을 대상으로 실행하였다.

컬럼명	정제 및 변환	예시
IPC분류	앞 4글자만 추출 후 카테고리 타입으로 변환	G01R 31 -> G01R
출원일자	yyyy-mm-dd 형식으로 데이터 타입 변환	2024.01.31 -> 2024-01-31
출원인	기업명으로 변환	에스케이오션플랜트(주)(경상남도 고성군...) -> SK오션플랜트

2-1-3. 추가

‘국가과학기술표준분류-특허분류(IPC_CPC)_연계표’(1-3의 수집 데이터)에서 기술분류(중분류)와 IPC 분류 코드만 추출하여 새로운 데이터프레임인 ‘IPC코드-기술분류 연계표’를 생성한다.

- ‘IPC코드-기술분류 연계표 예시

IPC 분류	기술 분류
A01B	로봇/자동화기계
A01C	농업기계/설비
A01D	농업기계/설비
A01F	농업기계/설비
A01G	식량작물과학

KIPRIS 특허 데이터프레임의 추출한 컬럼(2-1의 컬럼)으로 된 원본 데이터프레임에 ‘IPC코드-기술분류 연계표’를 ‘IPC분류’ 컬럼을 기준으로 병합하여 새로운 컬럼 ‘기술분류’ 생성

컬럼명	정제 및 변환	예시
기술분류	결측치는 ‘기타’로 대체, 카테고리 타입으로 변환	항공시스템, 로봇/자동화기계, 무기화학

- 전처리 완료한 국내 데이터프레임

	발명의명칭	IPC분류	출원일자	출원인	법적상태	요약	기술분류
0	소를 회수하는 방법 및 장치	H01L	2014-10-15	풍산	등록	리 시스템에 다시 제공된다.	반도체소자/시스템
1	헤드 노출 미세 조정 장치	B41J	2023-10-17	풍산	공개	방향으로 제어하여 조정한다.	소프트웨어
2	헤드 노출 미세 조정 방법	B41J	2023-10-17	풍산	공개	조정하는 단계를 포함한다.	소프트웨어
3	기관 처리용 챔버 개폐장치	H01L	2014-01-20	풍산	등록	고 진밀하게 체결할 수 있다.	반도체소자/시스템
4	가능한 날개의 수납 전개장치	F42B	2020-12-22	풍산	등록	수납 전개장치를 제공한다.	국방플랫폼
5	카드를 이용한 영커 시공방법	F16B	2021-07-19	풍산	등록	완성할 수 있는 효과가 있다.	자동차/철도차량
6	수산 부산물 자원화 시스템	B09B	2021-06-29	풍산	등록	어되는 건조부;를 포함한다.	폐기물관리/자원순환
7	론 투하탄용 투하장치 조립체	B64D	2023-10-06	풍산	공개	장치 조립체(10)를 제공한다.	항공시스템
8	동축 로터 기반 복합형 드론	B64U	2023-10-06	풍산	공개	복합형 드론가 제공된다.	기타

2-2) 해외 데이터 전처리

2-2-1. 컬럼명 변경

분석을 하기 전 데이터를 이해하기 쉽고 일관되게 만들기 위해 KIPRIS 특허 기술 전처리 데이터와 컬럼명을 동일하게 변경한다.

변경 전 컬럼명		변경 후 컬럼명
IPC	->	IPC분류
date	->	출원일자
tech_name	->	발명의명칭
tech_desc	->	요약
company	->	출원인

2-2-2. 정제 및 변환

5개의 컬럼 중 발명의 명칭, 요약 컬럼은 원본 데이터의 핵심 내용을 담고 있기 때문에 원본을 유지한다.

IPC분류, 출원일자, 출원인 컬럼을 대상으로 실행 하였다.

컬럼명	정제 및 변환	예시
IPC분류	앞 4글자만 추출 후 카테고리 타입으로 변환	E03B 9/08 -> E03B
출원일자	yyyy-mm-dd 형식으로 데이터 타입 변환	14.07.2011-> 2011-07-14
출원인	대소문자, 언어 구분 없이 기업명이 들어간 행만 추출하여 기업명으로 변환	NORINCO INTERNATIONAL COOPERATION LTD. -> NORINCO

2-2-3. 기술분류 컬럼 추가

WIPO 특허 데이터프레임의 전처리 데이터(2-1~2-2 내용) 에

'IPC코드-기술분류 연계표'를 'IPC분류' 컬럼을 기준으로 병합하여 새로운 컬럼 '기술분류' 생성

컬럼명	정제 및 변환	예시
기술분류	결측치는 '기타'로 대체, 카테고리 타입으로 변환	항공시스템, 로봇/자동화기계, 무기화학

- 전처리 완료한 국내 데이터프레임

	발명의명칭	IPC분류	출원일자	출원인	요약	기술분류
0	und Anzeigeeinrichtung	G06F	2015-06-25	AVIC	ie Abtastelektrode dient.	정보이론
1	NG SYSTEM OF AIRCRAFT	F15B	2024-05-30	AVIC	ents of hydraulic pumps.	국방플랫폼
2	R DISTRIBUTION SYSTEM	H02J	2024-06-13	AVIC	ected to a storage battery	전력IT
3	ETHOD FOR USING SAME	B64C	2024-06-13	AVIC	e retractable steel cable.	항공시스템
4	PPARATUS AND METHOD	B64D	2024-06-13	AVIC	caused by fuel overflow.	항공시스템
5	paration method thereof	C04B	2020-04-10	AVIC	ation degree is prepared.	건설환경설비기술
6	OR CREEP-AGE FORMING	B21D	2016-08-25	AVIC) in the second direction.	생산기반기술
7	ment monitoring device	G01C	2024-04-05	AVIC	in monitoring precision.	천문학

<3. 시각화 매뉴얼>

3-1) 막대그래프 - 상위 20개 기술분류별 비율 국내, 해외 비교

국내/해외 방산 기업의 특허 기술 분야의 분포, 편중 비교를 위해 상위 20개 기술분류를 도출하고 기술분류 별 비율을 시각화 한다.

3-1-1. 데이터 처리

국내 기업 특허 데이터(2-1 국내 특허 기술 데이터 전처리)들을 불러오고, 데이터 프레임을 연결하는 함수인 concat()을 사용하여 컬럼을 기준으로 세로로 병합

해외 기업 특허 데이터(2-2 해외 특허 기술 데이터 전처리)들을 불러오고, 데이터 프레임을 연결하는 함수인 concat()을 사용하여 컬럼을 기준으로 세로로 병합

막대 그래프 제작할 때, 국내와 해외 데이터의 비교를 위해 병합한 데이터프레임에 각각 '기업분류' 컬럼 추가

컬럼명	추가한 값
기업분류	국내

컬럼명	추가한 값
기업분류	해외

3-1-2. 그래프 생성 과정

- 기술분류 컬럼 추가 후 국내 기업 특허 기술 데이터프레임과 해외 기업 특허 기술 데이터프레임 병합
- '기술분야의 빈도 수 기준 상위 20개 추출
- 기술분류와 기업분류 컬럼을 기준으로 데이터를 그룹화하고, 각 조합의 데이터 개수를 계산하는 'count' 컬럼 생성 (Ex: '기업분류 - 국내' 별 '기술분류 - 항공시스템'의 개수)
- 국내/해외 기업에서 각 기술분류가 차지하는 비율을 확인하기 위해 국내/해외 'count' 컬럼의

- 값을 국내/해외 전체 행 개수로 각각 나눔
5. <국내/해외 기술 분류 상위 20개 비율 비교 막대그래프> 생성

3-1-3. barplot(막대그래프) 축 정보

X축	Y축
기술 분야 별 출원된 특허 수의 비율	기술분류 컬럼에서 상위 20개

3-1-4. (1)그래프 분석

국내의 경우, 자동차/철도차량 부문에서 압도적으로 높은 특허를 출원
 해외는 항공시스템, 나노기술, 고분자화학 등의 영역에서 강한 기술주도권

3-1-4. (2) 전략적 인사이트

-정책 방향

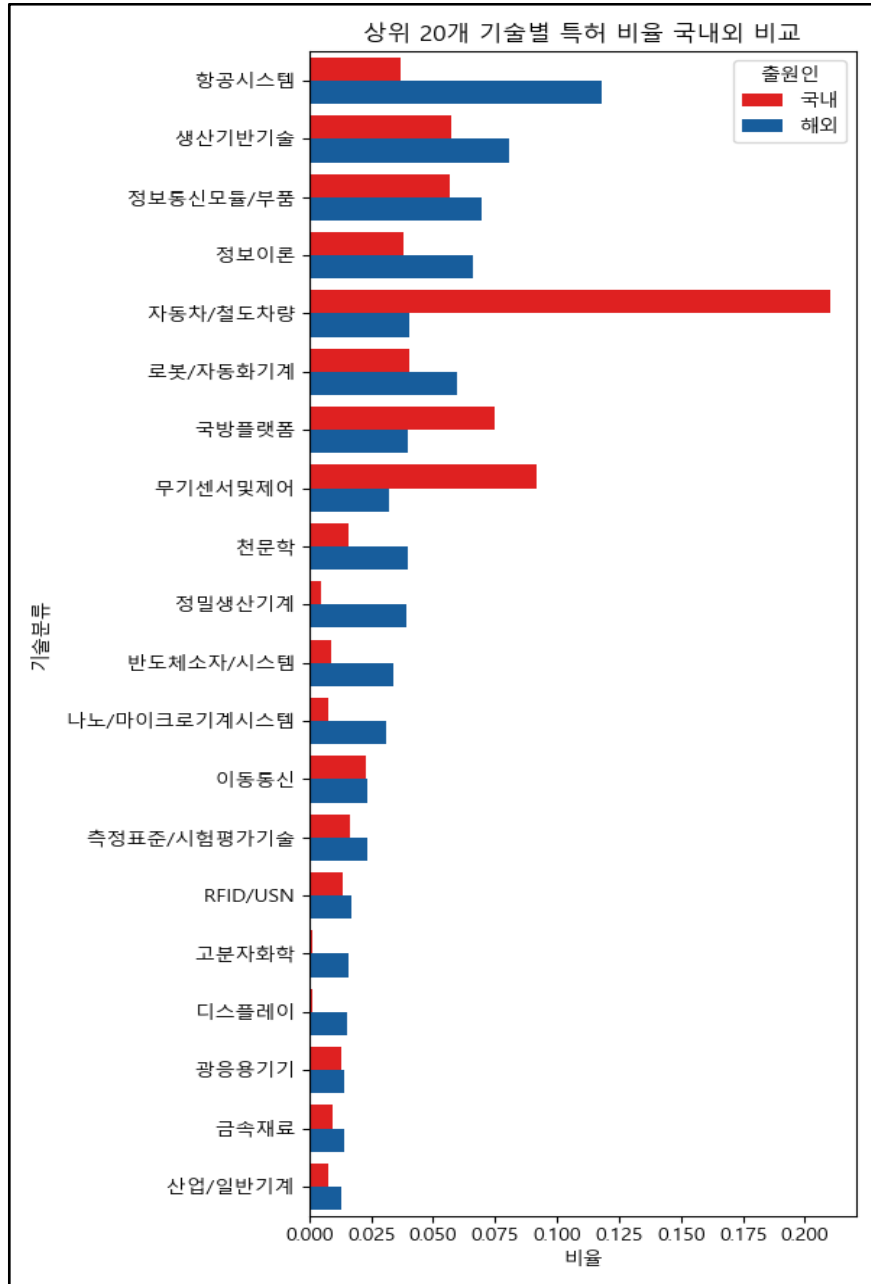
국내의 상대적 열위 분야(항공, 나노소재 등)에 대한 산학연 연계 및 R&D 투자 확대 필요성을 시사한다

-투자 전략

국내의 상대적 강세 분야인 정보통신/센서 기술의 경우 민군 활용도가 높아, 중소기업/스타트업 기반 기술사업화를 시도할 수 있다

-기술 보완

첨단소재, 고분자화학 등의 기술 열위 분야는 우선적으로 글로벌 협력을 활용하되, 장기적으로는 기술 국산화 과제를 수행하여 해외 기술 의존도를 낮춰야 한다



<국내/해외 기술 분류 상위 20개 비율 비교 막대그래프>

3-2) 막대그래프 - 국내외 연도별 상위 10개 기술 누적 특허 출원 수(2020~)

국내/해외 특허 기술(2-1 국내 특허 기술 전처리 데이터, 2-2 해외 특허 기술 전처리 데이터)의 연도별 출원된 특허 기술의 기술 분야를 확인하기 위해 시각화

예시로 2020년~2025년의 빈도수 상위 10개 기술 분야에 대한 누적 특허 출원 수를 시각화 한다.

3-2-1. 데이터 처리

국내/해외 특허 기술 데이터(2-1 국내 데이터 전처리, 2-2 해외 데이터 전처리)의 '출원일자' 컬럼에서 연도 부분만 추출해 '연도' 컬럼을 생성

컬럼명	추출 내용
출원일자	2020.01.01 -> 2020

컬럼명	컬럼 내용
연도	해당 특허 기술의 출원 연도

3-2-2. 그래프 생성 과정

1. 기술분류와 연도를 그룹화 후 연도별 개수를 집계
2. 빈도수 상위 10개 기술분야 추출
3. 국내/해외 연도별 특허 수 누적 막대그래프 생성

3-2-3. barplot(막대그래프) 축 정보

X축	Y축
상위 10개 기술분류	연도별 특허 수의 합

3-2-4. (1) 그래프 분석

국내 방산 기업에서 가장 많이 출원된 기술 분야는 '자동차/철도차량'이며 2022년에 약 230개의 특허 기술이 출원되었다.

특허가 공개되는데 오랜 기간이 필요하다는 특징에도 불구하고, 2025년 출원된 기술 분야는 자동차/철도차량, 국방플랫폼, 건설시공/재료가 있었다.

정보통신모듈, 정보이론 등 신기술 영역이 중상위권에 진입하여 첨단 IT 기반 융합기술의 중요도가 상승했음을 알 수 있다.

한편 해외 상위 11개 기업은 '항공시스템', '정보기술' 중심의 첨단 기술에서 우위를 선도하고 있으며 다양한 산업 분야에 비교적 고르게 특허 기술이 출원된 모습을 보였습니다.

3-2-4. (2) 전략적 인사이트

-정책 방향

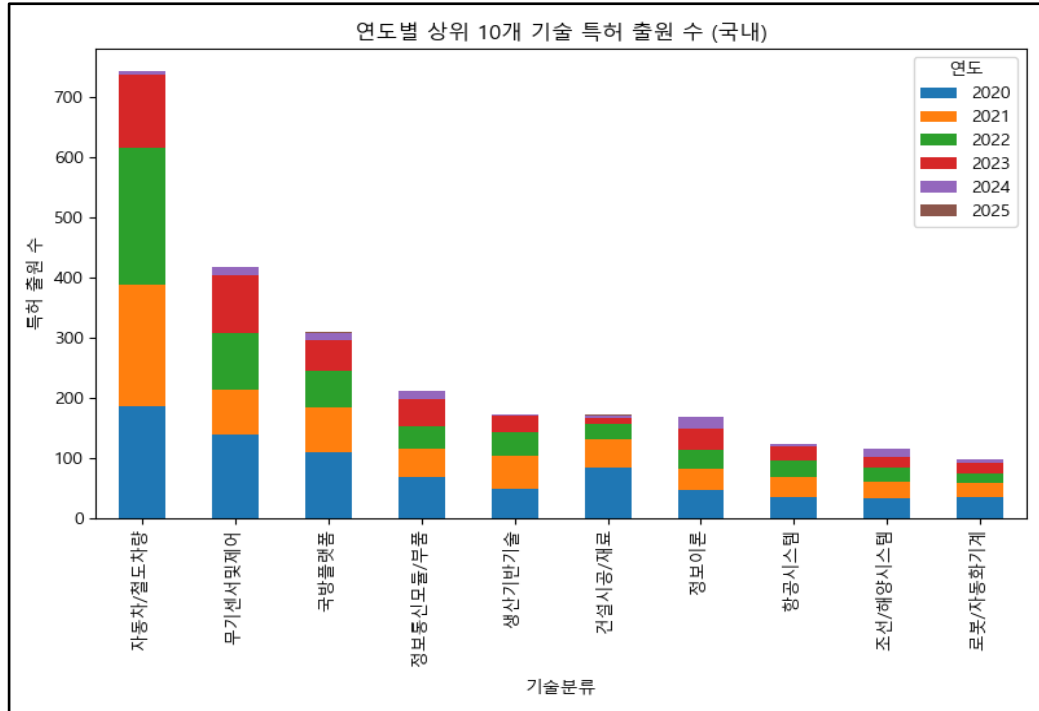
국가는 항공, 통신 분야 등 첨단 기술을 포함한 다양한 산업군에 기술 다변화를 유도해야 한다

-투자 전략

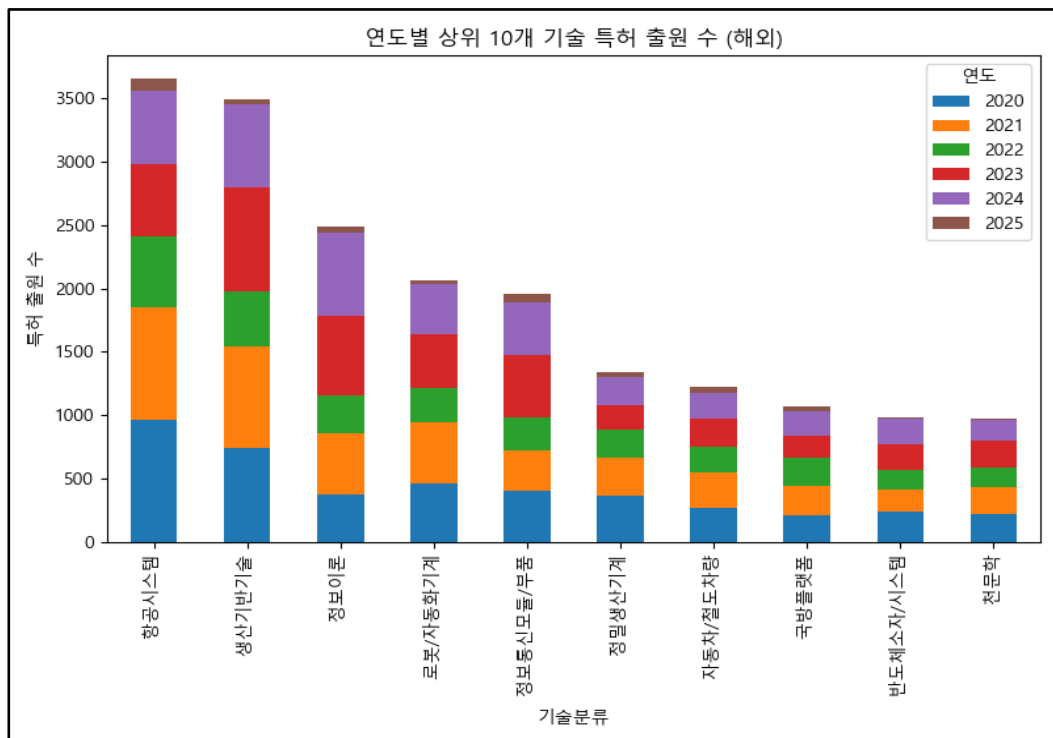
무기센서, 정밀생산기기 등 민군의 융합이 가능한 기술군을 타겟으로 투자해야 한다

-기술 보완

상위 11개의 대기업 주도가 약한 '정보이론','로봇' 등에서 중소기업/스타트업의 진입 기회가 있다



<2020 ~ 국내 연도별 상위 10개 특허 기술 출원 수 누적 막대그래프>



<2020 ~ 해외 연도별 상위 10개 특허 기술 출원 수 누적 막대그래프>

3-3) 산점도 - 국내외 기업별 특허 출원일자와 특허기술(2004~)

방산 기업이 어느 시기에 어느 기술의 특허를 출원했는지를 확인하기 위한 시각화

국내 방산 기술의 시간에 따른 진화 양상 및 트렌드 추적을 통해 향후 국방 전략기술 10대 분야 ('인공지능', '유·무인자율', '양자', '우주', '에너지', '첨단소재', '사이버·네트워크', '센서·전자전 추진' 'WMD(대량살상무기) 대응') 으로의 전환 여부를 분석 할 수 있음.

국내 방산 기업인 한화에어로스페이스, KAI, LIG 넥스원을 대상으로 예시 그래프 생성

3-3-1. 데이터 처리

국내/해외 특허 기술 데이터(2-1 국내 특허 기술 전처리 데이터, 2-2 해외 특허 기술 전처리 데이터) 에서 '출원일자', 기술분류, IPC분류' 컬럼만 추출하여 사용

'출원일자' 컬럼 연도에 해당하는 부분을 추출하여 '연도' 컬럼 생성

- 국내 데이터

컬럼명	변환한 내용
출원일자	'출원일자' 컬럼을 날짜형으로 바꾼 후, 2004년부터의 데이터만 추출

- 해외 데이터

컬럼명	변환한 내용
출원일자	'출원일자' 컬럼을 날짜형으로 바꾼 후, 국내 특허 기술 데이터와 비교를 위해, 2004년부터의 데이터만 추출

3-3-2. 그래프 생성 과정

1. 기술 분류를 그룹화하여 개수 집계 후 정렬하는 코드 작성

예시:

- groupby('기술분류', observed=False).size() -> 국내/해외 각각 기술분류별 총 출원 건수를 집계하기 위해, 데이터를 '기술분류' 컬럼을 기준으로 그 개수를 그룹화함

sort_values(ascending=False) -> 개수가 많은 순서대로 정렬

2. 추출한 데이터에서 빈도수 상위 20개의 기술분류 추출하고, 기술분류 이름 기준으로 IPC분류 정렬 순서 설정
3. y라벨 만들기
 - 3-1. 중복 IPC에 대해 첫 기술분류 기준으로, 중복 제거 후 IPC분류별 대표 기술분류 매핑 (딕셔너리 형태)

3-2. 기술분류 이름 리스트 생성 (카테고리 순서대로)

3-3. 중복 기술분류 이름을 그룹으로 묶고 가운데 인덱스만 표시

4. (2)에서 추출한 데이터 프레임으로 scartter plot을 그리고, (3)에서 만든 y축 라벨을 적용

5. 국내와 해외 각각 scartter plot 시각화

3-3-3. scatter plot(산점도) 축 정보

X축	Y축
2004년부터의 출원일자	상위 20개 기술분류랑 매핑된 IPC분류

3-3-4. (1) 그래프 분석

한화에어로스페이스는 2020년~2024년에 무기화생방/화력탄약 기술 연구, 개발이 집중

한국항공우주산업(KAI)는 대부분 기간에 항공 시스템 기술의 연구, 개발에 집중

LIG 넥스원은 무기체계기술, 항법, RFID 등 전반적인 ICT 기반 무기 시스템에 고르게 기술을 보유함과 동시에 출원시기의 분포가 넓어 지속적이고 활발한 투자 추세를 보임

이 3개의 기업은 2010년 이후 출원이 증가, 2020년 이후에도 여전히 많은 출원이 이루어지고 있음.

이 3개의 기업은 모두 '정보통신/보드/부품', '정밀이론', '조사/항법시스템' 등의 특정 분야에 집중적으로 출원하는 경우가 존재

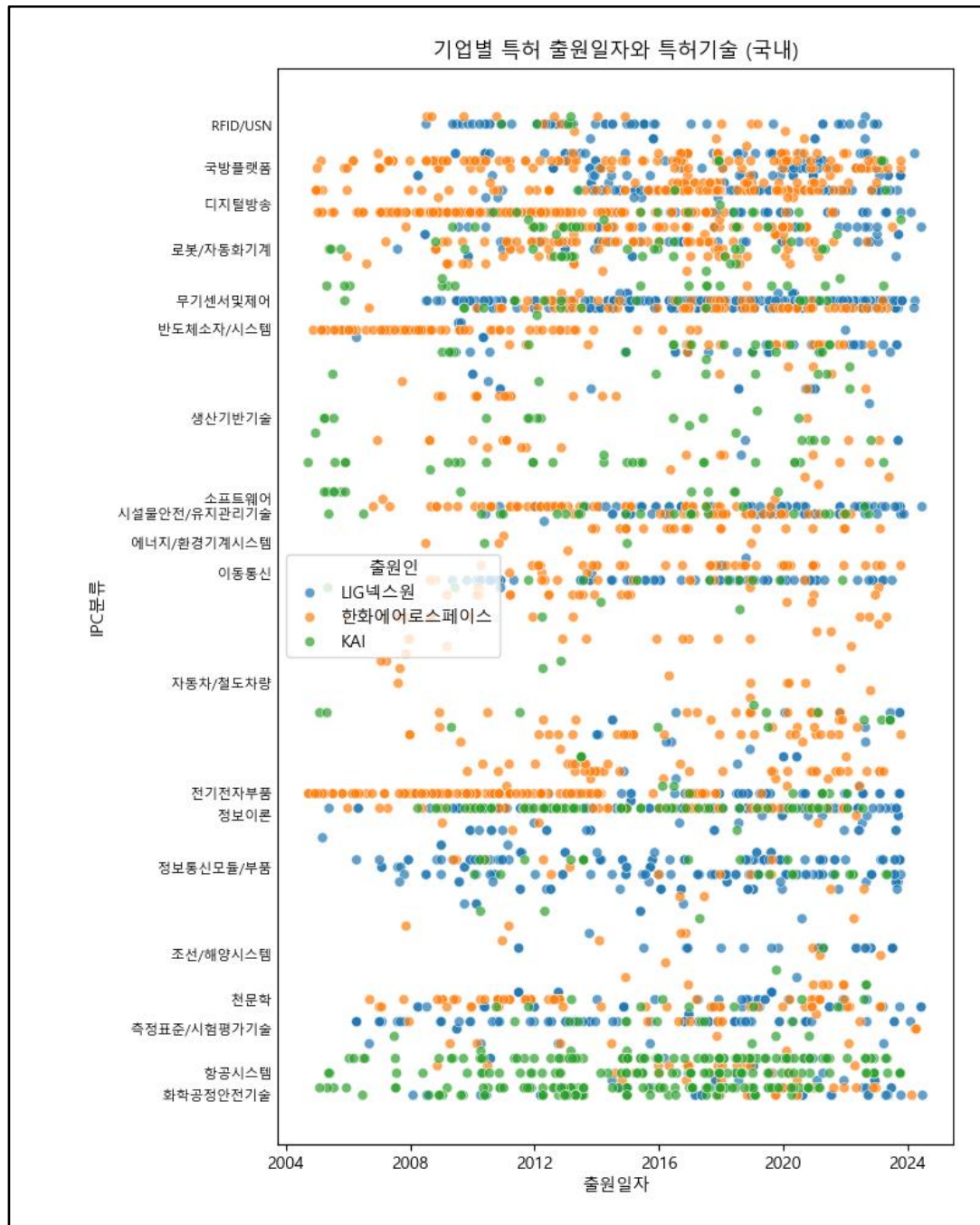
3-3-4. (2) 그래프 분석

-투자 전략

산업 내 전략기술 및 중복투자 영역에서 산학연 협력 또는 기술 차별화 전략 수립을 고려할 수 있다

-정책 방향

정부나 방위사업청에서 특허 포지셔닝 기반 R&D 중복 방지와 효율적 예산 배분을 할 수 있다



<국내 기업별 특허 기술의 출원 일자를 표현한 산점도 그래프>

3-4) 선그래프 - 국내외 기술별 누적 특허 출원 수

3-2) 막대그래프와 비슷한 내용을 담고 있으나, 기술 분야 별로 모든 출원일자에 대한 특허 출원 수를 확인할 수 있다.

2-1) 국내 전처리 데이터의 '기술 분류' 중 '항공시스템', '생산기반기술', '정보통신모듈/부품', '정보이론', '로봇/자동화기계'를 예시로 시각화

3-4-1. 데이터 처리

'출원일자' 컬럼 연도에 해당하는 부분을 추출하여 '연도' 컬럼 생성

컬럼명	변환한 내용
출원일자	출원일자 >= 2004-01-01, 2004년부터의 데이터 추출 후 연도 컬럼 생성

데이터프레임에서 출원일자와 기술분류별 누적 출원 건수를 계산하는 코드 작성

예시 :

- groupby(['출원일자', '기술분류'], observed=True).size() -> 출원일자와 기술분류 컬럼을 그룹화하여 출원 수 계산

- unstack(fill_value=0) -> 데이터 프레임을 재구조화함. 출원일자는 인덱스가 되고, 기술분류가 컬럼으로 올라감. 없는 것은 0으로 채움

- sort_index().cumsum() -> 재구조화한 데이터 프레임을출원일자(행) 기준으로 정렬한 후 누적합을 계산

- 위와 동일한 과정으로 그룹화하여 변형한 데이터 프레임 예시

기술분류	RFID/USN	가스 에너지	가정용기기/전자용기기	건설 시공/재료	건설환경설비 기술	고분자재료	고분자화학	광대역통신	광응용기기	국방플랫폼	...	철도교통기술	측정표준/시험평가기술	치료/진단기기	폐기물관리/자원순환	항공시스템	화학공정	화학공정안전기술	화학제품	환경예측/감시/평가	기타
출원일자																					
2012-01-30	53	1	0	86	1	2	3	3	25	124	...	53	47	3	2	62	0	19	0	0	8
2012-01-31	54	1	0	86	1	2	3	3	25	124	...	53	47	3	2	62	0	19	0	0	8
2012-02-01	54	1	0	86	1	2	3	3	25	124	...	53	47	3	2	62	0	19	0	0	8
2012-02-02	54	1	0	86	1	2	3	3	25	124	...	53	47	3	2	62	0	19	0	0	8
2012-02-03	54	1	0	86	1	2	3	3	25	124	...	53	47	3	2	62	0	19	0	0	8
2012-02-06	56	1	0	86	1	2	3	3	25	124	...	53	47	3	2	62	0	19	0	0	8

<재구조화 후 기술분류별 출원 수를 누적합 한 데이터프레임>

3-4-2. 그래프 생성 과정

1. 출원일자를 월 단위로 변환해 2004년부터 시작하는 기술분류 데이터들만 추출
2. 국내와 해외 각각 누적 선그래프로 시각화

3-4-3. line plot(선 그래프) 축 정보

X축	Y축
월별 출원일자	누적 특허 출원 수

3-4-4. (1)그래프 분석

그래프로 기술기가 가파른 기간에 해당 분야 기술 개발, 연구가 활발했음을 확인하여 동향을 파악할 수 있다

기술의 성장 속도 부문에서 국내는 기술별 격차가 적고 완만하게 상승하는 반면, 해외는 특정 분야가 빠른 속도로 독주하는 추세를 보인다.

국내 특허 출원은 방산 내재화를 목적으로 제조/센서 기반 기술 중심으로 이뤄지는 반면, 해외의 경우 진입장벽이 높은 항공/정보통신 등의 첨단 핵심 기술을 중심으로 이뤄진다

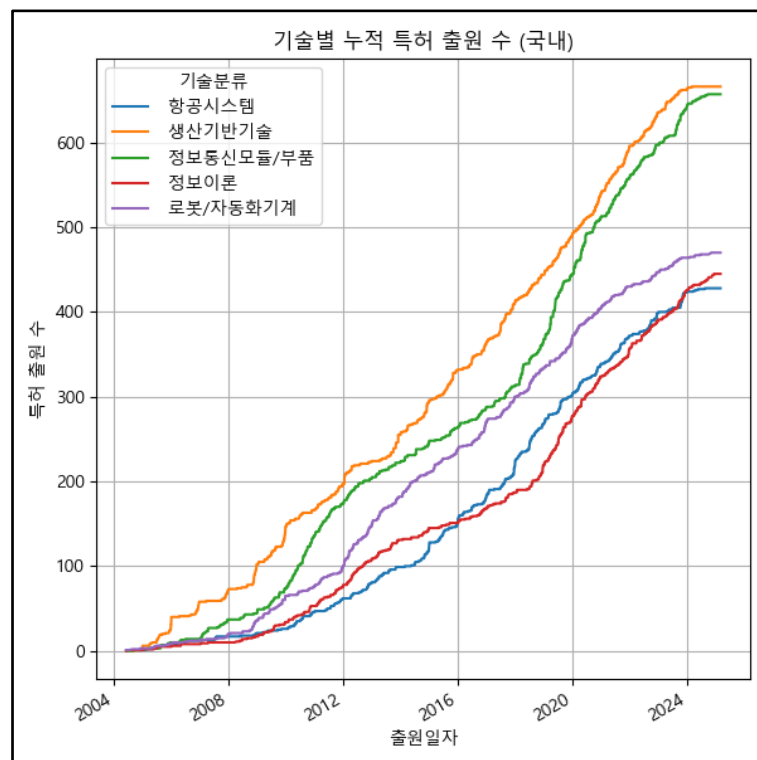
3-4-4. (2)전략적 인사이트

-정책 방향

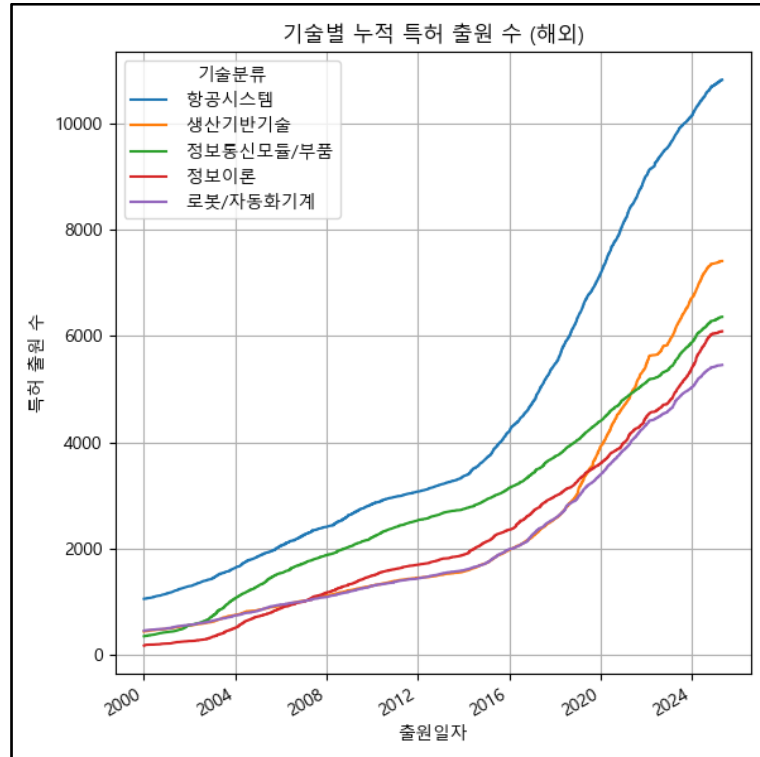
정부의 국가 핵심기술 육성사업 및 전략 R&D 자금 배분 우선순위 결정의 지표로 활용할 수 있다

-투자 전략

기술별 성장 곡선 분석을 통해 R&D 투자 시점을 판단할 수 있다



<특허 출원 동향 파악을 위해 국내 기술별 특허 출원 수를 누적하여 그리는 선 그래프>



<특허 출원 동향 파악을 위해 해외 기술별 특허 출원 수를 누적하여 그리는 선 그래프>

3-5) 선그래프 - 국내외 월별 특허 출원 수(2004~)

3-4) 선그래프와 유사한 내용을 내포하지만, 더 좁은 범위에서 기업의 월별 특허 출원 수 추이를 시계열로 분석한 시각화 자료

3-5-1. 데이터 처리

국내/해외 특허 기술 데이터(2-1 국내 특허 기술 전처리 데이터, 2-2 해외 특허 기술 전처리 데이터)에서 'IPC분류' 컬럼의 결측치를 제거

'출원일자' 컬럼을 날짜형으로 변환

국내 방산기업 특허 기술과 비교를 위해 2004년 이후의 데이터만 추출

컬럼명	변환한 내용
출원일자	날짜 타입으로 변환 후, 2004년부터의 출원일자를 월 단위로 변환

3-5-2. 그래프 생성 과정

1. 월별 출원 건수를 추출하여 저장하고 datetime 형식의 타임스탬프로 변환
2. 국내와 해외 각각 누적 선그래프로 시각화

3-5-3. line plot(선그래프) 축 정보

X축	Y축
월별 출원일자	월별 특허 출원 수

3-5-4. (1)그래프 분석

시계열 추세를 확인하여 국내 방산 기업의 기술 연구, 개발은 특정 시기에 짧은 기간 급증했다가 바로 급감하는 파형(peak)임을 확인 할 수 있다

또한 국내는 특정 시점에 기술 집중이 이뤄진 뒤 이후 급격히 출원 수가 감소하는 산발적 집중 형태는 프로젝트 중심의 단기 기술 집중 출원 전략을 시사한다

해외 방산 기업의 기술연구, 개발은 월별 변동은 있으나 전반적으로 균형 잡힌 계단식 상승 구조를 유지하고 있다

장기적이고 점진적인 증가 형태는 기술 축적형의 기업 내부 R&D 성장을 시사한다

3-5-4. (2)전략적 인사이트

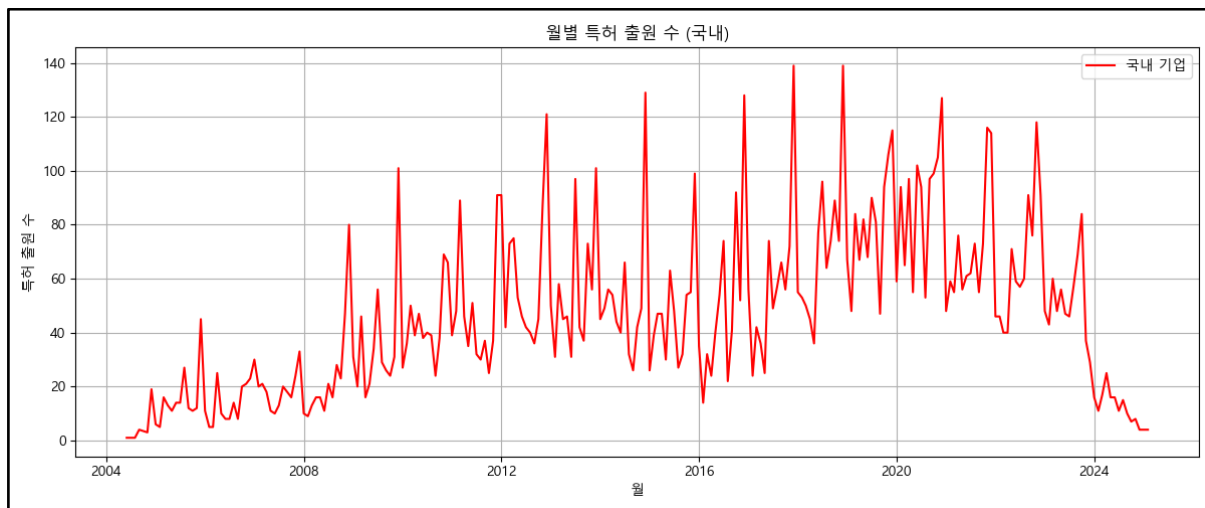
-정책 방향

국내 방산 기술 개발은 지속성 확보가 관건이므로, 연속형 R&D 투자와 함께 중장기 과제를 기획하여야 한다

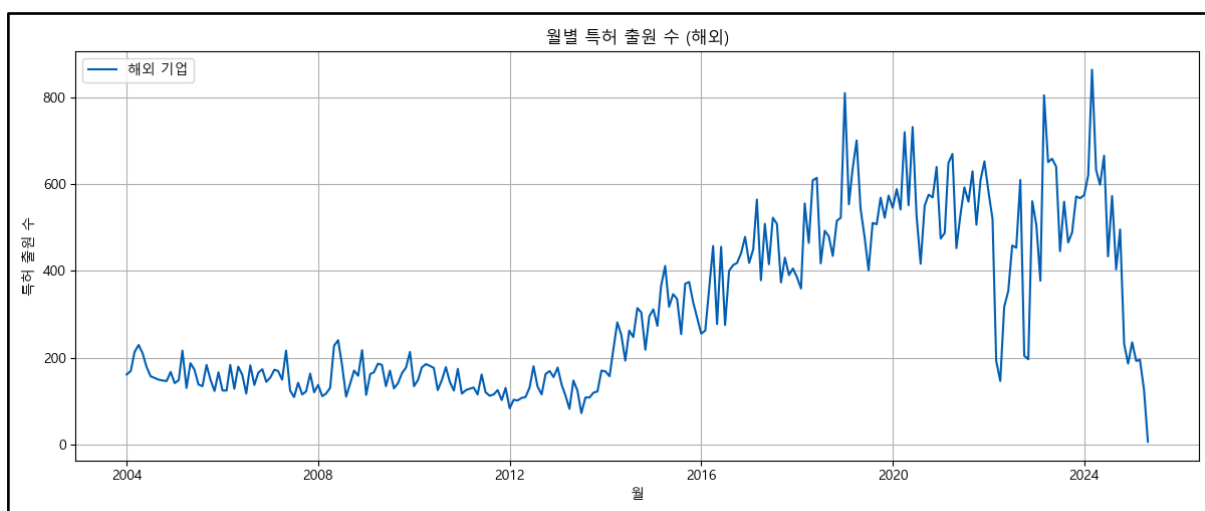
-기업 전략

국내 기업은 개발과 출원의 시점을 분산화하여 기술 개발의 흐름을 일관적으로 유지하는 전략을 고려할 수 있다

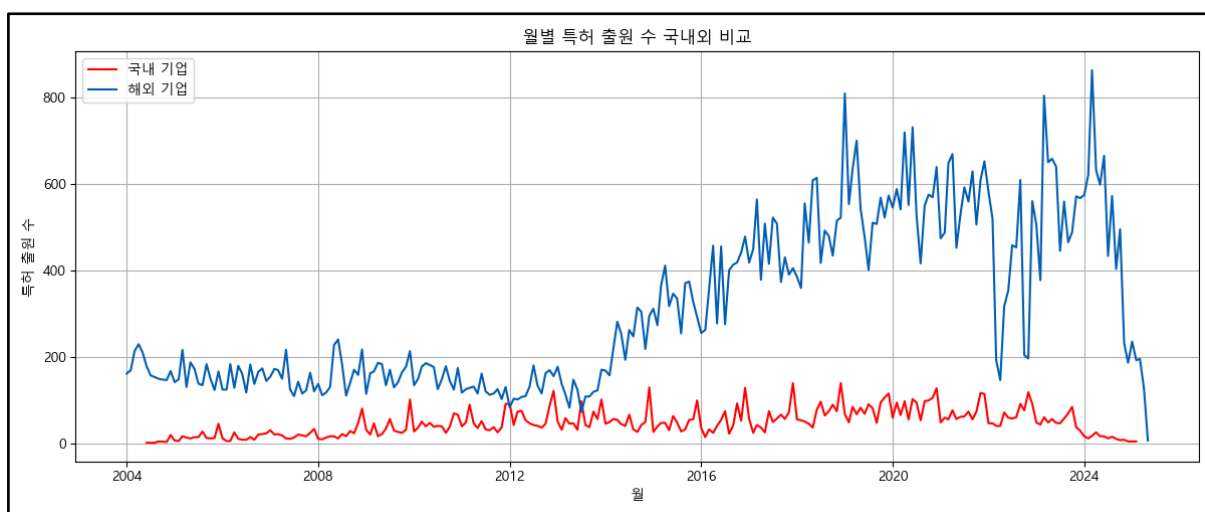
또한 장기적 내부 과제를 수행하는 해외기업을 벤치마킹하여 장기 특허 운영 전략을 수립할 필요성이 있다



<국내 월별 특허 출원 수 선 그래프(2004~)>



<해외 월별 특허 출원 수 선 그래프(2004~)>



<국내, 해외 월별 특허 출원 수 비교 선 그래프(2004~)>

3-6) 파이차트 - 국내외 상위 10개 기술 분포

국내/해외의 기술 분야의 편중을 확인하기 위한 시각화

국내/해외 특허 기술 데이터(2-1 국내 특허 기술 전처리 데이터, 2-2 해외 특허 기술 전처리 데이터)의 '기술분류' 컬럼이 갖는 빈도수에 대해 비율을 시각화

3-6-1. 데이터 처리

국내/해외 특허 기술 데이터의 '기술분류' 컬럼에서 상위 10개 기술분류 추출

국내/해외의 상위 10개 기술분류의 빈도수 리스트 생성

예시:

기술분류	자동차/철도차량	무기센서및제어
국내 count	2442	1067
해외 count	3727	NaN

3-6-2. 그래프 생성 과정

1. 최대 10개의 색상을 갖는 파이차트 색상 팔레트 설정
2. subplot()를 사용함 - 하나의 Figure(그래프 창) 안에 여러 개의 그래프(axes)를 배치할 수 있게 해주는 함수
3. 하나의 figure안에 두 개의 subplot을 생성하고, 그래프들이 겹치지 않도록 함
4. 반복문을 통해 상위 10개 기술분류의 빈도수 리스트를 가져와 두 개의 데이터에 대해 반복
5. 파이차트 시각화

lables	autopct
기술분류명	'%.1f%%' 형식(소수점 첫째 자리까지 표시하고 뒤에 '%'를 붙임) 기술분류의 특허 건 수를 비율 표

3-6-3. (1)그래프 분석

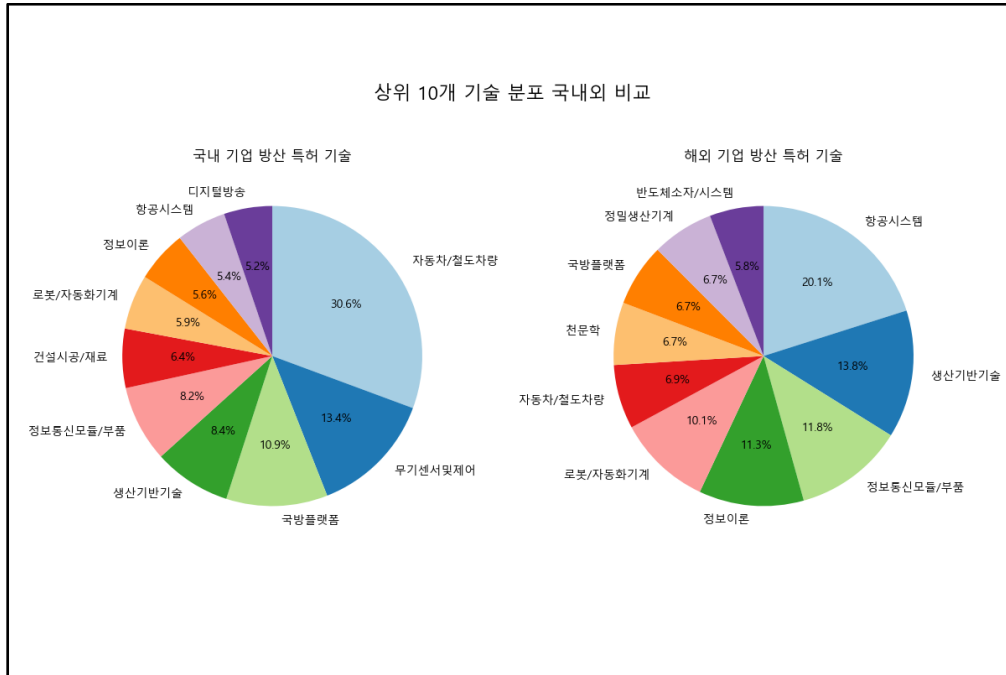
국내 방산 기업의 특허 기술은 자동차/철도차량 분야(31.6%)에 과도하게 집중되어 기술 쏠림 현상이 나타나고 있다

반면 해외의 경우 항공시스템(23.1%), 생산기반(17.4%), 정보이론(13.4%), 로봇(12.3%) 등 다양한 분야에서 균형된 기술 개발이 이뤄지고 있으며 상위 기술 간의 편차가 크지 않다

3-6-3. (2)전략적 인사이트

-정책 방향

국내 기술 포트폴리오의 다변화와 전략적 분산이 필요하다



<국내외 상위 10개 기술 분포를 확인할 수 있는 파이차트>

3-7) 워드 클라우드 - 국내외 기술 분야별 특허 기술 요약

워드 클라우드(Word Cloud)란, 문서나 비정형 데이터에서 자주 등장하는 단어들을 출현 빈도에 따라 시각적으로 달리 표현한 그림이다.

핵심 키워드를 도출하거나, 빅데이터를 요약하는 역할을 한다.

국내/해외의 출원된 특허 기술에 대해 분야별로 기술 흐름과 핵심 키워드를 한눈에 시각적으로 파악하기 위해 워드 클라우드를 제작한다.

3-7-1. 데이터 처리

병합한 국내와 해외 데이터(2-1 국내 특허 기술 전처리 데이터와 2-2 해외 특허 기술 전처리 데이터) 불러오기

국내/해외 특허 기술 데이터의 '기술분류' 컬럼에서 상위 10개만 추출

-> 국내: 자동차/철도차량, 무기센서및제어, 국방플랫폼, 정보통신모듈/부품, 생산기반기술, 로봇/자동화기계, 정보이론, 항공시스템, 디지털방송, 건설시공/재료,

-> 해외: 항공시스템, 생산기반기술, 정보통신모듈/부품, 정보이론, 로봇/자동화기계, 자동차/철도차량, 전문학, 국방플랫폼, 정밀생산기계, 반도체소자/시스템

1) 국내

- 상위 10개 기술분류 컬럼을 한 개씩 선택하고, 요약 컬럼의 모든 값을 문자열 타입으로 변환, 결측 값을 빈 문자열로 대체

- '요약' 컬럼을 추출하여 시리즈로 변환 후 리스트 형식으로 변환
- 형태소 분석기 Okt() 생성 후 tokens_n이라는 빈 리스트를 생성
- 반복문을 통해 요약 컬럼에서 명사만 추출하여 리스트로 반환
- 추출된 명사 리스트를 tokens_n 리스트에 확장
- 추출된 명사 중 '상기'와 '발명'은 각 기술마다 워드 클라우드에 나타낼 필요가 없으므로 제거
- 리스트 컴프리헨션을 이용해 tokens_n 리스트의 각 토큰을 순회하면서 길이가 2이상인 명사들만 tokens_c 리스트에 저장
- tokens_c 리스트에 있는 명사 토큰들의 빈도수를 계산

2) 해외

- 요약 컬럼이 영어 문자 이므로, 영어 불용어 리스트를 가져옴
- 상위 10개 기술분류 컬럼을 한 개씩 선택하고, 요약 컬럼의 모든 값을 문자열 타입으로 변환하고, 결측값을 빈 문자열로 대체
- 정규표현식($r'\Ww+':$ 단어 하나를 찾는다. 즉, 1개의 단어를 단위로 문장을 분할)으로 단어 단위로 추출하고 명사 태그 리스트를 설정
($'NN'$: 단수 일반 명사, $'NNS'$: 복수 일반 명사,
 $'NNP'$: 단수 고유 명사, $'NNPS'$: 복수 고유 명사)
- 요약 컬럼을 소문자로 변환하여 대소문자를 통일하고, 단어를 추출
- 추출된 단어 중 불용어를 제거하고, 품사 태깅(각 단어에 해당하는 품사를 자동으로 붙이는 작업)을 함
- 모든 텍스트에서 추출된 명사들을 tokens 리스트에 합침
- tokens 리스트에 있는 명사 토큰들의 빈도수를 계산

3-7-2. 그래프 생성 과정

- WordCloud() 워드 클라우드 객체를 생성
- 워드 클라우드를 시각화 함

font_path	max_font	max_words	width	height	background_color
C:/Windows/Fonts/malgun.ttf	50	100	400	200	white

<해외 - 항공시스템 워드 클라우드>