



관심 분야 유망 회사 검색을 위한

(AI)특허·논문검색

WEB APPLICATION

PPAS

| 'PAPER' | 'PATENT' | 'AI' | 'SEARCH' |

4팀_최환욱(팀장), 박윤경, 서보선, 김하준

>>>> CONTENTS

01.



분석배경 및
주제 선정 이유

02.



웹사이트
제작 **FLOW**
CHART
진행과정

03.



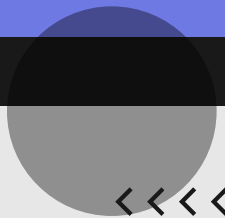
PPAS
웹사이트 시연

04.



주요 가치
및 보완점

ARTIFICIAL INTELLIGENCE (AI)



01.

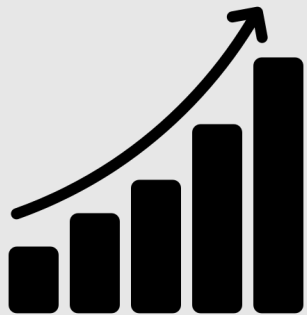
분석 배경 및 주제 선정 이유



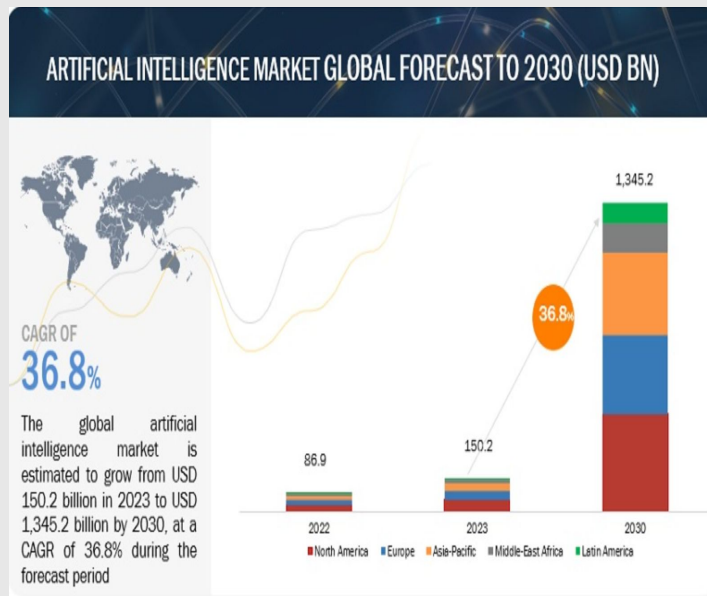
Reason why made it



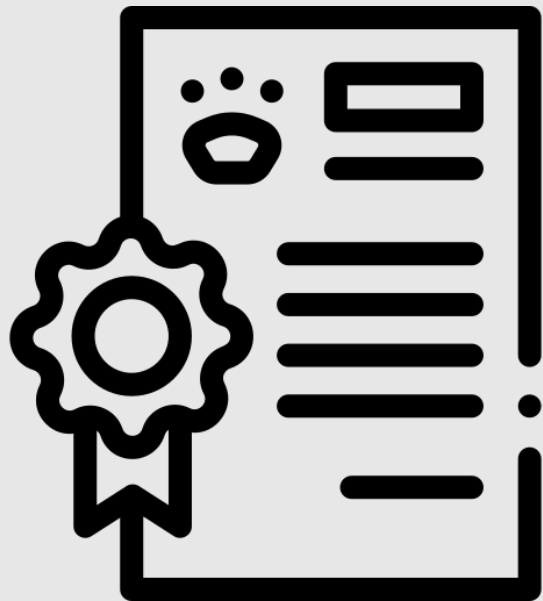
01. | 분석배경



• AI 시장의 성장성:
전 세계 인공지능(AI) 시장
규모가 2023년 1502억달러
(약 200조원)에서
2030년에는 1조3452억달러



01. | 분석배경



특허 정보의 중요성

- 기술발전 트렌드 분석:
특허 검색을 통해 AI기술 발전 추이 파악
(특정 기술이 언제, 어디서, 어떻게 발전 했는지 이해)
- 투자자, 구직자:
성장할 기업, 비전 있는 기업 탐색 가능
- 연구 및 개발(R&D) 방향 설정:
기업의 경우 경쟁사의 기술전략 개발 방향을
분석하여 자신들의 기술과 현재 AI기술의
연구 갭을 파악하여
새로운 연구 및 개발 방향 설정
- 협력 및 라이선싱 매칭:
필요한 기술 특허를 가진 기업 및 연구 기관을
탐색해 새로운 기술 개발 가능

01. | 분석배경

한국의 AI 특허 활용 사례

"특허 기반 투자·대출 사상 첫 6조 돌파"

이해성 기자 ☆

입력 2022.02.22 17:17 수정 2022.02.23 02:08 지면 A17

가

"특허 보유 스타트업 생존율, 성장성 높아"

국내 스타트업 특허창출 성과 높지만 등록·유지에 부담

이정아 기자 leeah@cmn.co.kr

[기사입력 : 2019-05-27 19:27:47]

글로벌 특허출원은 딥테크기업 생존 위한 필수 전략

※ 노원옥 마이셀럽스 변리사 © 입력 2022.11.09 17:23 © 수정 2022.11.11 23:09 댓글 0
♡ 좋아요 0



“특허, 스타트업 생존·성장위해 필수”

입력 2019-06-16 12:00:01 수정 2019.06.16 12:00:01 양종근 기자

하지만 스타트업이 소홀히 한 특허의 효과는 높다. 2017년 전미경제연 연구에 따르면 스타트업은 최초 특허 등록 후 5년간 평균 고용이 54% 늘었고 매출은 79% 뛰었다.

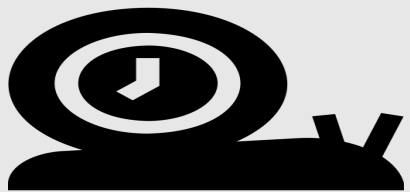
•한국에서는 AI 특허 판매에 대한 구체적인 사례가 많이 알려져 있지는 않지만, AI 관련 특허를 활용한 다양한 사업 모델이 존재합니다.

예를 들어, 중소기업들은 AI 기술 특허를 통해 새로운 비즈니스 기회를 창출하고 있습니다.

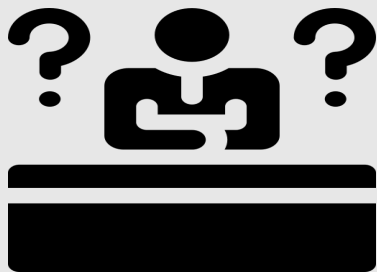
이런 기업들은 AI 특허를 통해 기술력을 인정받고, 이를 바탕으로 파트너십을 맺거나 기술 라이선싱을 통해 수익을 창출합니다.

한국에서는 AI 기술을 보유한 스타트업들이 특허를 기반으로 투자 유치에 성공하는 사례도 많습니다.

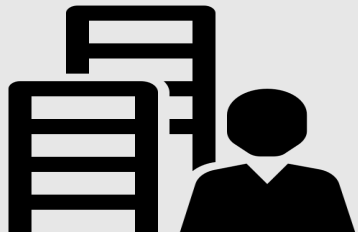
01. | 주제 선정 이유



•특허 출원 공개 시간: 특허는 출원 등록부터 공개까지 평균적으로 약 1년 6개월의 시간이 소요되어 최신 기술 동향 파악이 어려움



•특허 데이터 검색 난이도:처음 특허 데이터 정보 검색을 하는 이들에게 쿼리를 작성하는 방법부터 원하는 정보를 추출하기까지에는 어려움이 있음.



•대기업 위주의 정보:현재 AI특허 데이터를 검색해보면 대부분 대기업 위주의 정보가 주를 이루고 스타트업의 특허 정보 특정이 어려움.

01. | 주제 선정 이유

Solution-특허 출원 등록 시간

한국특허정보원 KIPI	
 한국특허정보원 Korea Institute of Patent Information	
KIPI RIS 특허정보검색서비스	
arXiv	
	
운영	코넬 대학교
종류	학술 데이터베이스

•문제)특허 출원 공개 시간: 특허는 출원 등록부터 공개까지 평균적으로 약 1년 6개월의 시간이 소요되어 최신 기술 동향 파악이 어려움

•해결)최신 기술 동향 파악: 논문 데이터를 추가하여 특허의 출원 등록 시간이 소요되는 문제를 보완하고 최신 AI 기술 동향을 신속하게 파악

•수집사이트 선택 기준:

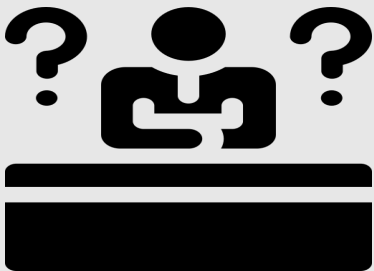
한국특허정보원에서 운영하는 키프리스에서 특허 데이터 수집

논문 데이터를 확인 가능한 대부분의 학술 데이터베이스는 유료이기에 무료로 논문 데이터를 확인 가능한 arXiv를 통해 논문 데이터 수집

01. | 주제 선정 이유

Solution-특허 데이터 검색 난이도

Query: order: -announced_date_first; size: 200; date_range: from 2013-06-24 to 2024-06-24; include_cross_list: True; terms: AND title="artificial intelligence"; OR title=AI; OR title="deep learning"; OR abstract="artificial intelligence"; OR abstract=AI; OR abstract="deep learning"



•문제)복잡한 검색 과정:

특허 데이터는 방대한 양과 복잡한 구조로 인해 검색이 어려움.

비전문가가 필요한 정보를 찾기 위해 많은 시간과 노력이 필요.

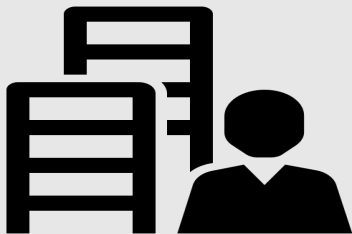
•해결)간편한 검색과 필터링 기능을 제공하여

누구나 쉽게 사용할 수 있도록 설계. 형태소 분석 및 키워드 추출을 통한

데이터 전처리를 통해 일관성을 유지하여 분석 및 활용이 용이.

01. | 주제 선정 이유

Solution-대기업 위주의 정보



•문제)대기업의 특허 정보 위주:

대기업들은 방대한 자원으로 많은 특허를 보유.
중소기업이나 스타트업은 대기업의 특허 정보를
분석하기 어려움.

•해결)상위 출원인 외에 스타트업 중 상위
출원인 섹션을 만들어 특정 분야에서 유망한
스타트업의 정보를 확인할 수 있도록 보완

02.

WEB 제작과정 FLOW CHART

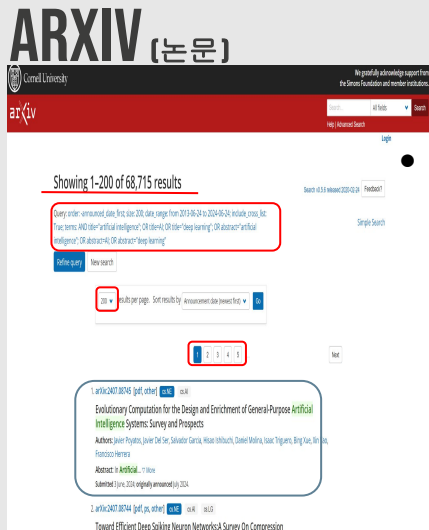


The process of to made it

프로젝트 FLOW CHART

- 데이터 수집 및 전처리: 데이터 소스, 데이터 수집 과정, 전처리 과정
- 웹 서버: Flask를 이용한 웹 서버 구축
- 백엔드 개발: Flask 서버 설정, 라우팅 설정, 검색 및 필터링 기능 구현
- 프론트엔드 개발: UI 설계 및 구현 (사용자 인터페이스 제공)

데이터 수집



데이터 소스:

KIPRIS: 한국특허정보원
특허 데이터베이스

총 364,259 건의
특허 자료 데이터
크롤링 : 페이지당
90건 (출원번호,
출원일자, 출원인,
요약문)에 대해서
4,048페이지로부터
데이터 수집

arXiv: 학술 논문
데이터베이스

총 68,715 건의 논문
자료 데이터 크롤링 :
페이지당 200건
(특허명, 출원일,
요약문, 저자, 링크
정보)에 대해서
345페이지로부터
데이터 수집

> > > 데이터 전처리

데이터 전처리

- ✓ 특허 출원 요건 자체가 상당히 tight하여 누락 데이터 많지 않음
 - 특허 출원의 필요 요건이 아닌 우선권 주장, 국제 출원, 최종권리자의 경우 null 값이 상당수 있어 공백 처리
- ✓ 총 364,259 건의 특허 자료 중 인공지능/AI와 관련 없는 특허가 상당수 포함
 - 이러한 관련 없는 특허를 2차로 거르기 위해 (1) 1차로 추출된 36만여건의 특허의 특허명과 요약문을 KoNLPy 형태소 분석기를 이용하여 한글 명사만을 키워드로 뽑아 리뷰하고 stopwords setting 후 인공지능/AI 관련 키워드만으로 재 분류 진행

```
# KoNLPy 형태소 분석기 초기화
okt = Okt()

# 전체 텍스트 읽기
df['title'] = df['title'].astype(str); df['summary'] = df['summary'].astype(str)

# 불용어(stopwords) 리스트 정의
stopwords = ['상기', '이하', '따라', '본문', '특히', '항구', '발명', '포함', '이상',
             '해당', '또한', '오도', '다수', '결함', '부가', '연관', '가속', '통한',
             '부차', '완료', '제공', '대항', '연결', '구성', '구성', '사용', '상태',
             '미리', '명식', '대하', '변화', '개구', '관용', '제출', '성능', '생산',
             '증거', '과제', '최소', '특성', '따라서', '연속', '일정', '통과', '상',
             '영역서', '별위', '종류', '가능']

# 텍스트 정리 및 불용어 함수 정의
def preprocess_and_tokenize(text):
    text = re.sub(r'[\^~\|가-힣|a-z]', '', text) # 한글만 남기고 모든 문자 제거
    nouns = okt.nouns(text) # 형태소 분석 후 명사 추출
    nouns = [word for word in nouns if len(word) > 1 and word not in stopwords]
    return nouns

# 각 행별로 키워드 추출 및 누적
all_keywords = []
for index, row in df.iterrows():
    text = row['title'] + ' ' + row['summary']
    tokens = preprocess_and_tokenize(text)
    all_keywords.extend(tokens)

# 전체 텍스트에서 빈도 계산 및 상위 빈도 키워드 추출
word_freq = Counter(all_keywords)
```

형태소 분석& 키워드 추출

364,259 → 38,767 건

```
# 필터링할 키워드 설정
keywords = ['인공 지능', '인공지능', '뉴럴 네트워크', '양자 컴퓨팅', '양자 알고리즘', '뉴로모픽', 'ILSVRC', 'chat']

# 'title' 컬럼 또는 'summary' 컬럼에 키워드가 포함된 경우만 남기기
filtered_df = df[df['title'].str.contains('|'.join(keywords), case=False, na=False) |
                 df['summary'].str.contains('|'.join(keywords), case=False, na=False)]

# 필터링된 행의 인덱스를 가져오기
filtered_index = filtered_df.index

# 원본 데이터프레임에서 필터링된 인덱스에 해당하는 행만 남기기
fdf = df.loc[filtered_index]

print(df.shape, fdf.shape)

[100] ✓ 53.8s
(364259, 3) (38767, 3)
```

재 분류 작업

>>> 웹서버 구축



프로젝트 폴더 구조

```
M4_WEB_PROJ/  
|  
| └─ static/  
|   | └─ css/  
|   |   | └─ style.css  
|   | └─ gen/  
|   |   | └─ packed.css  
|   |   | └─ packed.js  
|   | └─ js/  
|   |   | └─ scripts.js  
|   |   └─ plot_01.png  
|  
| └─ templates/  
|   | └─ plot.html  
|   └─ Web.html  
|  
| └─ config.py  
| └─ Google Cloud Translate API Key  
| └─ NanumBarunGothic.ttf  
| └─ Papers_Arxiv.xlsx  
| └─ requirements.txt  
| └─ result_mod_30June2020.xlsx  
| └─ result_mod_sample_30June2020.xlsx  
| └─ Temp_plot.ipynb  
└─ WebApp.py
```

필요한 라이브러리

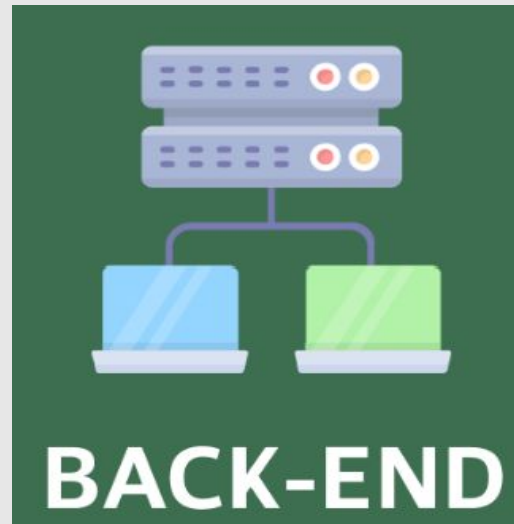
```
from flask import Flask, render_template, request, send_file, make_response, redirect, url_for  
from wtforms import StringField, SelectField, DateField, SubmitField  
import pandas as pd  
import matplotlib.pyplot as plt  
from markupsafe import Markup  
from flask_assets import Environment, Bundle  
import matplotlib.font_manager as fm  
import os  
from datetime import date  
from flask_wtf.csrf import CSRFProtect  
from flask_wtf import FlaskForm # FlaskForm 추가  
import zipfile  
from io import BytesIO  
import uuid  
from googletrans import Translator  
import deepl  
import matplotlib.gridspec as gridspec  
import seaborn as sns  
import re  
import requests
```

> > > 백엔드 개발



백엔드 주요 기능 요약

1. 데이터 수집 및 전처리:
 - Excel 파일 로드 및 날짜 형식 변환
2. 번역 기능:
 - Google, DeepL, Naver PAPAGO 번역 API 사용
3. 품 처리 및 검색:
 - 사용자의 검색 요청을 처리하고, 데이터를 필터링
4. 데이터 시각화:
 - Matplotlib, Seaborn을 활용한 시각화
5. 파일 다운로드:
 - 필터링된 데이터를 CSV 파일로 변환 후 다운로드 제공
6. 웹 페이지 렌더링:
 - 템플릿을 사용하여 검색 결과 및 시각화된 데이터를 웹 페이지에 표시



> > > 프론트엔드 개발

프론트엔드 주요 UI 구성 요소 요약

1. 검색 기능 및 다운로드

- 검색 바: 사용자가 검색어를 입력하고 검색할 수 있는 인터페이스 제공.
- 필터 옵션: 사용자가 검색 범위를 날짜, 카테고리 등으로 필터링할 수 있는 기능 제공.
- CSV 파일 다운로드: 검색된 특허와 논문 데이터를 CSV 파일로 다운로드할 수 있는 기능 제공.

2. 결과 표시

- 검색 결과 페이지: 검색된 특허와 논문의 목록을 사용자에게 표시.

3. 공개 공보 보기

- 특허 및 논문 상세 페이지: 선택된 특허나 논문의 공개 공보를 제공.
- 하이퍼링크: 출원인에 따라 특허의 경우 원티드랩의 기업 및 채용 정보, 논문의 경우 논문 원본 자료로 연결되는 링크 제공.

4. 데이터 시각화

- 인공지능 관련 기술 동향과 트렌드를 시각화하여 사용자에게 그래프 이미지 제공



03.

PPAS

WEB SITE 시연



What is possible with it



[AI]특허·논문검색 WEBSITE

| 'PAPER' | 'PATENT' | 'AI' | 'SEARCH' |

/[AI]/[AI]/



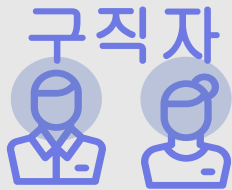
04.

주요 가치 및 보완점

You can enter a subtitle here if you need it



TARGET AND VALUE



구직자

최신 기술 동향 파악:
최신 AI 기술 관련 특허와 논문을
통해 기술 트렌드를 파악하고,
직무에 필요한 역량을 향상.
취업 준비:
AI 분야의 최신 연구 및 기술 동향을
학습하여 면접 대비.

투자자

기술 평가:
투자할 스타트업의 기술력과
혁신성을 특허와 논문을 통해 평가.
시장 분석:
AI 기술의 발전 추세와 시장 동향을
파악하여 투자 전략 수립.
리스크 관리:
특허 분석을 통해 기술적 리스크를
최소화하고, 투자 결정을 더 신뢰성
있게 수행.

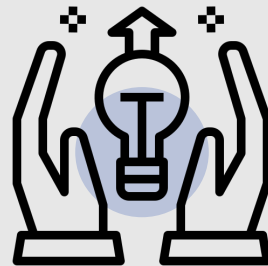
기업

혁신 촉진:
최신 특허와 논문을 통해 연구
개발 방향을 설정하고, 기술
혁신을 촉진.
경쟁력 강화:
시장에서 경쟁사 대비 기술적
우위를 확보.
IP 전략 수립: 특허 정보를
기반으로 효과적인 지적 재산권
(IP) 전략 수립.



SEARCH

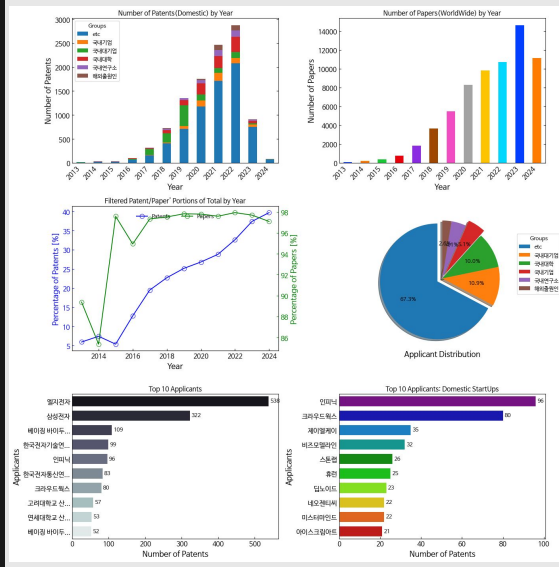
다양한 특허 및 논문
검색 기능



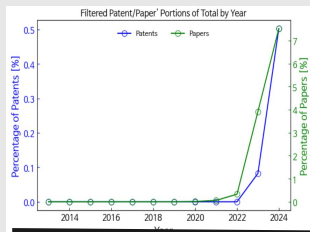
VENUS

데이터 분석으로
가치있는 인사이트
plot제공

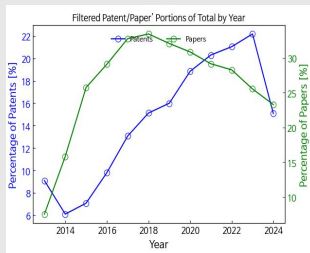
제공 plot 예시



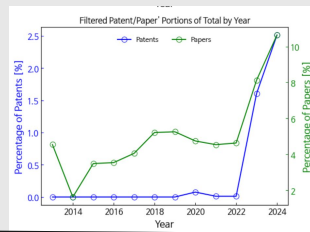
인사이트 예시 거대언어모델



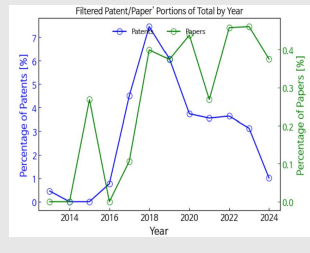
이미지



생성형



블록체인





보완점

데이터베이스: 현재 시스템에서는 쿼리나 실시간 데이터 업데이트가 필요하지 않다고 판단되어, 데이터베이스를 통합하지 않았습니다.
다만, 이는 데이터의 일관성과 효율적인 관리에 제한을 초래할 수 있습니다.

로그인 통한 사용자 관리: 사용자 관리를 위해 로그인 기능을 구현하지 못했습니다. 이는 사용자 맞춤형 서비스 제공에 제한이 있을 수 있습니다.

부트스트랩 적용: 사용자 인터페이스 개선을 위해 부트스트랩을 적용하지 못했습니다. 이로 인해 웹사이트의 반응형 디자인과 사용자 경험이 제한될 수 있습니다.
