# 기업 R&D 진단 및 벤치마킹 시스템 기술

KiSTi 한국과학기술정보연구원 Korea Institute of Science and Technology Information www.kisti.re.kr

**특 허 명** 데이터분석 장치 및 그 동작 방법

**Keyword** 임베딩 벡터, 데이터분석장치, 기계학습

**발 명 자** 이재민/하태현

## 기 술 성

### 기술 개요

- 본 특허는 분석대상 데이터(ex: 특허, 상표 및 논문 데이터)에 특화된 엠베딩(Embedding) 벡터를 추출하여 이를 데이터 분석에 활용하기 위한 기술임

### 기존 기술 문제점

- 연구 및 기술개발, 그리고 보유기술과 비즈니스 영역(제품, 서비스)의 추출 등을 위한 데이터 분석은 다양한 분야에 걸쳐서 그 중요도가 날로 커지고 있음. 특히 특허 및 논문 데이터 분석의 경우 키워드 기반의 검색엔진을 활용하는 것이 일반적임.
- 이에, 특허 데이터 분석에 키워드 기반의 검색 엔진을 활용하는 경우, 사용자가 원하는 기술 또는 특징이 컨텍스트로 포함된 특허 데이터뿐만 아니라, 전혀 무관한 특허 데이터인 노이즈까지 검색결과로 나올 수 있기 때문에 이에 기반한 데이터 분석의 실효성이 높지 않음

### 기술의 특징 및 우수성

#### ▶ 기술의 특징

- 특정기업이 보유한 분석대상 데이터에 대해서 추출되는 임베딩(Embedding) 벡터의 평균값을 특정기업에 대한 대표 임베딩 벡터로 산출하며, 특정기업의 대표 임베딩 벡터를 특정기업과는 다른 타 기업의 대표 임베팅 벡터와 비교하여 기업간 유사도를 판단 할 수 있음.
- 기본 언어모델(BERT)을 가져와 분석대상 데이터의 분류코드로 각각 학습하고 파인 튜닝(fine-tuning)하여 해당 문헌에 최적화된 임베딩 벡터를 추출하고 이를 활용함
- 분석대상 데이터에 등장하는 단어, 표현 등 해당 데이터에 특화된 임베딩 벡터를 기계학습을 기반으로 추출하며 이를 활용한 다양한 데이터 분석 서비스를 제공하는 것임

#### ▶ 기술의 우수성

- 분석대상 데이터에 등장하는 단어, 표현 등 해당 데이터에 특화된 임베딩 벡터를 기계학습을 기반으로 추출하여 이를 활용한 데이터 분석 서비스를 제공함으로써, 데이터 분석에 있어서 그 실효성이 높음
- 특정기업 보유한 분석대상 데이터 사이의 임베딩 값을 비교하여 기술과 상관도가 낮은 저장 데이터를 걸러낼 수 있음

# 분석대상 데이터의 임베딩 벡터와 데이터 분석 기술

## 상세설명

- 본 기술의 데이터 분석 장치는 전용언어모델을 생성하는 생성부, 임베딩 벡터를 추출하는 추출부, 및 분석 서비스를 제공하는 제공부를 포함
- 생성부는 분석대상 데이터의 텍스트를 대상으로 사전훈련언어모델을 파인 튜닝(fine-tuning)하여 사전훈련언어모델로부터 분석대상 데이터를 위한 전용언어모델을 생성
- 생성부는 분석대상 데이터가 가지는 특정 데이터 필드의 텍스트가 입력 값이 되고, 분석대상 데이터를 분류하는 분류코드가 출력값이 되는 학습 데이터 셋을 구성하여, 사전훈련언어모델을 통해 이를 학습하는 방식으로 전용언어모델을 생성될 수 있도록 함
- 나아가, 생성부는 분석대상 데이터에 대한 학습 데이터 셋의 구성이 완료되면, 분석대상 데이터에 부여될 수 있는 분류코드를 포함한 완전연결계층을 사전훈련언어모델에 연결시켜, 완전연결계층을 기반으로 사전훈련언어모델을 지도 학습하여 분석대상 데이터만을 위한 전용언어모델을 생성함

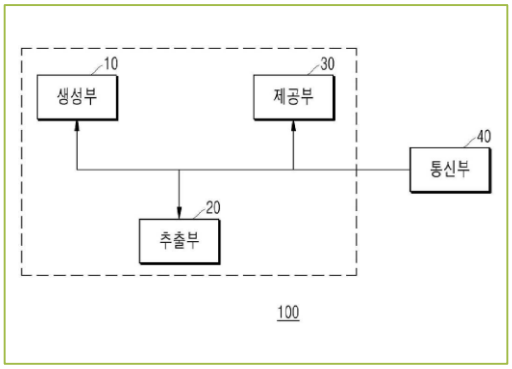


그림1 데이터 분석 장치의 구성도

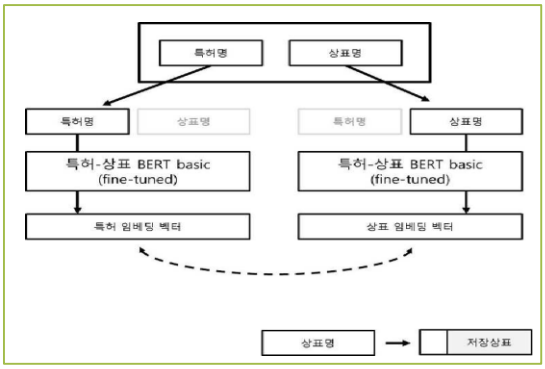


그림2 데이터 분석 장치의 동작방법

## 기술완성도 (TRL)

기술완성도 : TRL6 (Full Scale 시제품 개발)

| TRL1 | TRL2 | TRL3 | TRL4 | TRL5 | TRL6 | TRL7 | TRL8 | TRL9 |
|---|---|---|---|---|---|---|---|---|
| 기술원리 발표 | 기술컨셉 설정 | 기술컨셉 증명 | Lab Scale 시제품개발 | **구현환경 적용실험** | **Full Scale 시제품개발** | 유사 상용품 개발 | 상용품 완성 | 상용품 실시 |

## 활 용 분 야

## 활용분야 및 적용제품

### 활용분야

- ◆ 기업 분석 연구
- ◆ 특허/상표 출원시 유사 기술/상표 검색
- ◆ 특허 분쟁 시 요소 기술 비교 분석
- ◆ 논문 분석 연구

### 적용제품

- ◆ 특허/상표를 기반으로 한 기업 분석 서비스
- ◆ 등록을 원하는 특허/상표 내용에 근거한 유사 특허/상표 검색 및 추천 서비스
- ◆ 논문을 기반으로 한 기업 분석 서비스

# 인공지능 기반 스마트 영상학습 데이터 제작 기술

## 산업동향(기술 동향 및 트랜드 등 )

- 클라우드 컴퓨팅은 규모의 경제, 빅데이터 구축, 디바이스의 연산용량 한계로 주목을 받아 왔음. 인터넷 연결의 폭발적 증가와 데이터의 초대용량화, 실시간 처리 필요성 증가로 엣지 컴퓨팅이 함께 부각되고 있음. 엣지 컴퓨팅은 사물이나 기기의 엣지 단에서 데이터 분석, 처리를 분담함으로써 데이터 과다 트래픽 발생을 막고, 안정적으로 실시간 처리를 하는 것으로 이미 통신기업이나 서버기업에서 시스템을 개발해 보급 시도 중임
- 공공데이터포털과 빅데이터 플랫폼을 중심으로 데이터의 원활한 연계 및 활용을 위한 표준화 및 품질관리 노력 중임

## 시장전망(목표시장 규모 및 전망)

- 디지털 뉴딜의 핵심 산업인 '데이터' 산업의 전 체 시장규모는 19조2,736억원('20)으로 전년 대비 14.3% 성장하였으며,공공데이터 개방 건수는 55,561건('21.3)으로 전년 대비 63.4% 증가하여 데이터경제 활성화를 위한 기반을 제공하고 있음
- 서비스의 특성상 여러 사업과 융합될 수 있지만, 4차 산업혁명과 관련된 서비스는 대부분 IT 기술과 기존 산업의 융합으로 발생함. 이와 같은 추세에 따라 IT 서비스 시장은 지속적 성장으로 긍정적인 전망
- 세계 IT 서비스 지상은 연간 3.2% 이상의 성장률을 보이고 있으며, 2020년 이후 3.9%의 성장률이 예상됨. 각 부분별 성장률을 살펴보면 컨설팅/SI은 연 5.0%, 아웃소싱은 3.4%, 서포트/트레이닝은 2.8%로 전망
(출처 : IDC(International Data Corporation))

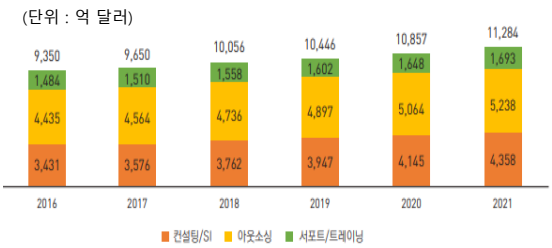


그림3 세계 IT 서비스 시장 전망

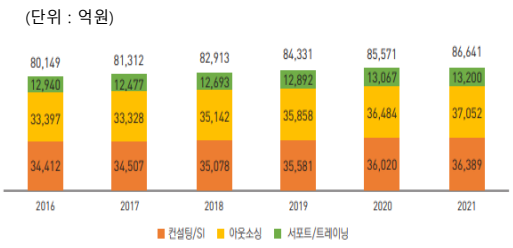


그림4 국내 IT 서비스 시장 전망

## 지재권현황

| 권리현황 | 특허출원번호 | 발명의 명칭 |
|---|---|---|
| 출원 | 10-2021-0013234 | 데이터분석장치 및 그 동작방법 |

## 문 의 처


**기술이전**

한국과학기술정보연구원 Korea Institute of Science and Technology Information

담 당 자 ooo
연 락 처 042-869-0915
이 메 일 kwsim@kisti.re.kr

**기술문의**

한국과학기술정보연구원 Korea Institute of Science and Technology Information

담 당 자 이재민 책임연구원
연 락 처 042-
이 메 일 @kisti.re.kr