1.

라스팜 페이턴트 서치 시스템입니다. 저희는 더 쉽고 다양한 방식으로 특허 검색을 하기 위해 특허들 사이의 관계와 유사성, 데이터를 통해 더 정확하고 커스터마이징된 서칭 시스템을 만들었습니다.

2.

기존의 특허 검색은 일반적인 검색식을 통해 키워드가 포함된 관련특허들을 보여주는 것에 그쳤는데요.

3.

저희는 시각적 정보와 알고리즘을 통한 정확성을 제공해 사용자에게 최대한 많은 도움을 주려 시도했습니다. 사용자들은 완벽한 검색어를 입력하지 않아도 됩니다. 유사한 특허를 찾을 수 있으며 특허이미지 검색이 가능합니다. 시각화 서비스 또한 제공받을 수 있습니다.

4.

서비스 기능을 말씀드리겠습니다. 두가지 검색기능 중 텍스트 검색에서는 검색 키워드와 관련된 단어들을 워드클라우드 형태로 시각적 정보를 제공해 검색하고 싶은 정확한 단어를 찾게 돕거나 키워드를 추가할 수 있습니다. 또한 검색한 특허들을 군집화한 결과를 시각화하고 군집을 각각 데이터표로 보여주어 같은 키워드 안에서도 세부적으로 특허들 간의 유사성을 쉽게 볼수 있게 만들었습니다. 이미지 검색에서는 사용한 이미지와 유사성이 높은 특허들과 비슷한 이미지를 보여줍니다. 키워드 보다 이미지로 검색하는게 더 효율적인 사용자들에게 유용할 것입니다.

5.

서비스 기능과 관련해 이 사이트를 유용하게 쓸 사용자는 첫번째로 관련특허와 유사한 특허들을 추가로 검색하고 싶은 분. 두번째, 한 키워드를 통해 검색한 카테고리안에 있는 특허들간의 유사성, 관계에 대한 정보를 얻고 싶은 분, 세번째, 이미지를 통해 특허ID를 찾고싶거나 가지고 있는 정보가 이미지로 한정되어 있는 분. 이외에도 특허 정보를 얻고 싶은 사용자라면 많은 도움이 될거라 확신합니다.

6.

임베딩 기반의 텍스트 검색과 랜드스캐핑 서비스입니다.

7.

사용한 데이터의 정보입니다. 텍스트 검색에서는 미국 특허상표청의 페이턴츠뷰 API 에서 utility 타입의 특허들을 2001년도부터 2019년도까지로 사용했습니다. Utility, 즉 발명특허는 데이터의 대다수를 차지할 뿐만 아니라 보편적으로 알려진 정의의 특허기 때문에 선택했고 속도와 효율성을 위해 사람들이 가장 많이 찾을 2001년부터 2019년까지의 최근 20년데이터로 한정했습니다. 데이터 테이블은 title과 abstract가 포함된 patent, assignee ID 가 포함된 patent assignee, cpc 정보가 들어있는 cpc current, classification 라벨값이 있는 wipo,wipofield입니다.

8.

Patent 테이블입니다. 전체 오백구십만개의 로우 중 사백이십만개를 사용했습니다.

9.

Patent Assignee와 cpc current 테이블입니다.

10.

Wipo, wipo field 테이블입니다.

11.

전처리 과정입니다. Html tag제거 후 문장을 sentencize해서 tokenize하고 special token들을 추가해 id값을 넣었습니다. 그리고 bert모델에 넣어 각각의 평균을 구했습니다.

12.

메타데이터 처리입니다. 각 피쳐가 얼마나 많은 Cpc와 assignee, wipo 를 포함하는지를 구하는 과정입니다. wipo에선 sector title 피쳐가 중복되어 가장 많은 sector title 을 1로 정했습니다.

13.

Heterogeneous graph는 서로 다른 node와 edge를 가지는 그래프로 assignee와 cpc를 노드와 엣지로 한 그래프를 트레인셋으로 HAN에 넣어서 예측한 wipo label 값에 대한 정확도 train accuracy는 0.8963, micro f1은 0.8963, macro f1 은 0.8313입니다. 임베딩에는 웨이트값인 Z를 사용합니다.

14.

이미지 벡터라이징 기반 이미지검색 서비스입니다

15.

이미지 검색에 사용된 데이터는 patentsview api에서 patent\_id 가 포함된 patent, cpc code가 있는 cpc current 테이블과 patent\_id로 GooglePatent에서 크롤링한 썸네일 이미지 데이터입니다. cpc code 는 특허청에서 분류한 범용코드라고 생각하시면 됩니다.

16.

이미지 검색 알고리즘은 Convolutional – Deconvolutional Autoencoder 모델을 사용헀습니다.

크롤링한 모든 구글 페이턴트 썸네일 이미지들을 64x64로 리사이즈 했습니다. convolution – maxpool 로 인코딩 deconvolution – upsample 로 디코딩 한 후 임베딩 값에 cpc code 정보를 추가하기 위해 cpc code 앞 3자리로 127개의 cpc code 임베딩값을 구합니다.

17.

서칭 과정입니다. Input image를 인코딩해 기존 db 의 특허 이미지 임베딩값과 코사인유사도를 구해 상위 5개 출력합니다. 임배딩된 127개의 Cpc code classification precision은 0.116 이고 상위5개 이미지 중 같은 특허의 이미지가 존재할 확률은 0.99입니다.

18.

사이트 소개입니다.

19.

메인화면과 서비스 기능 안내, 이미지와 텍스트 검색 페이지입니다.

20.

텍스트 검색바와 랜드스캐핑, 워드클라우드 입니다.

21.

이미지 검색바와 검색결과입니다. 이미지와 특허id, 커서를 올리면 특허 제목과 설명, 구글 페이턴트로의 연결링크가 출력됩니다.

실제 사이트 검색 시연

메인화면입니다. 서비스 안내 탭입니다. 프로젝트 팀 멤버입니다. 기타 문의나 연락은 여기로 부탁드립니다.

텍스트 검색

텍스트검색에 키워드를 넣고 워드클라우드 버튼을 누르면 워드클라우드가 생성됩니다. 워드 클라우드 안에 있는 텍스트를 눌러 키워드를 추가, 수정이 가능합니다. 키워드가 수정됨과 동시에 워드클라우드도 현재의 키워드와 관련있는 단어들로 바뀌어 제공됩니다. 현재 키워드들로 검색하기를 누르면 군집화된 특허들의 랜드스캐핑 결과와 특허데이터들 출력되고 구글페이턴트로의 연결링크가 나타납니다. 군집들을 활성화/비활성화 할 수 있으며 군집별로 따로 데이터표를 출력해 볼 수 있습니다. 각 특허들이 점으로 나타나는데 그 점들 사이의 거리로 유사성을 판단합니다. 가까울수록 유사한 특허인거죠. 특허데이터표에는 특허ID, 제목, 출원국가와 날짜 등이 표시되며 column 을 기준으로 정렬이 가능할 뿐만아니라 검색 결과 내에서도 각 column의 입력내용이 포함된 특허의 검색이 가능합니다. 또한 특허들을 원하는 개수만큼 불러올 수 있습니다.

이미지 검색

이미지 검색화면에서 파일을 업로드 한 후 현재 이미지로 검색하기를 누르면 유사한 특허이미지와 특허ID를 출력합니다. 이미지 위에 커서를 올리면 특허 이름과 설명, 구글페이턴트로 연결링크가 출력됩니다.

이상으로 라스팜 페이턴트 서치 시스템의 소개를 마치겠습니다 시청해주셔서 감사합니다.