

인적 사항



김전도 KIM JeonDo
🎂 92년생(30살)
📞 010-3538-4325
📍 서울 관악구 신림동
✉ zzang20044@naver.com

인사말

안녕하세요. 저는 학부에서 컴퓨터공학을, 대학원 석사과정에서는 통계학 및 머신러닝, 딥러닝을 공부했습니다. 이후 기업에서의 1년 10개월은 짧은 시간이었지만 학생 땀 겪어보지 못한 즐거운 실무 경험이었습니다. 앞으로도 많은 경험을 쌓으며 더욱 성장하길 원합니다. 감사합니다.

학력

연세대학교 (서울) 정보대학원 정보보호전공 석사 졸업
2017.03 - 2019.02
상명대학교 (서울) 컴퓨터과학과 학사 졸업
2012.03 - 2017.02
대성고등학교 (서울) 자연계 졸업
2008.03 - 2011.02

주요 보유 기술

Programming Language:

Python

Machine Learning, Deep Learning:

scikit-learn, keras 등의 ML 라이브러리 및 프레임워크
Gradient Descent, CNN 등의 딥러닝 알고리즘 구현

Database:

Mysql

Environment:

Linux, Windows

프로젝트

프로젝트명: [인텔리콘연구소] 계약서 조항 파서 개발(Python)

기간: 2021.01 - 2021.05

내용: 계약서 조항 파서는, 이전 단계에서 문서 변환 파서가 HWP, DOCX 등의 확장자를 가지는 계약서 문서를 HTML source로 변환해주면, 이를 입력으로 받아서 계약서의 조항호목('제1조', '(1)' 등의 머리 번호)을 기준으로 계층적으로 구분된 자료 구조 객체를 반환하는 파서입니다. 텍스트, 테이블, 이미지, 메모에 해당하는 자료구조 클래스와 이를 재귀적으로 담을 수 있는 그릇 형태의 자료구조 클래스를 만들어서 이용했고, 최종 결과 객체는 이후 테스트에서 두 개 계약서 간 유사 조문 매칭 등의 작업을 위해 사용됩니다.

프로젝트명: [인텔리콘연구소] 유렉스 유사 문단 추천기(Python)

기간: 2020.01 - 2020.12

내용: 유렉스란 법률 도메인에서의 검색 시스템입니다. 유사 문단 추천기는 사용자가 쿼리를 입력하고 법, 판례 등의 범주에서 문서를 특정 지으면 그 문서의 문단들과 쿼리를 비교하여 특정 문단을 추천해주는 시스템입니다.

Semantic 방법으로는 TF-IDF Weighted Cosine Similarity, Term Matching 방법으로는 BM25와 피인용 조항의 BM25 값을 이용하는 간접추천을 사용하였고, Term Matching 성능을 향상 시키기 위하여 한자어-한글 변환, 법률어-일상어 변환, 상위 계층 문단 상속 등의 처리를 하였습니다.

최종적으로 문단 추천 개수 선정을 위하여 값의 범위가 정해져있지 않은 BM25 및 간접추천 점수를 정규화해주고, 추가적으로 Cosine Similarity 값까지 총 세 개의 척도를 이용하여 상, 중, 하로 구분하여 추천 결과를 반환하는 Flask API Endpoint로 개발하였습니다.

프로젝트명: [인텔리콘연구소] 한글 판례 PDF Parser(Python)

기간: 2019.05 - 2019.12

내용: 법원에서 제공하는 판례 PDF를 이후 ML, DL 등의 방법론으로 활용하기 위해 정형화된 데이터로 변환시키는 데이터 엔지니어링 작업이었습니다. 텍스트, 이미지, 테이블 파싱이 요구되었고 텍스트 및 이미지 파싱은 pymupdf(fitz), 테이블 파싱은 camelot 패키지를 이용하였고, 오류 데이터, 문단 분리 등의 이슈가 있었습니다.

오류 데이터 이슈로 일부 텍스트가 이미지로 간주되는 경우, char 단위로 복제되는 경우 등이 있었으며 OCR, BBOX(페이지 상의 위치 정보) 등을 활용하여 문제를 해결했습니다.

문단 분리 이슈는 실제 하나의 문단이 한 줄씩 다수의 줄로 분리되는 문제였으며 문제 해결을 위해 각 줄의 BBOX, 문단 머리 번호(ex. '가.', '㉔', '1. '), 형태소의 품사 정보 등을 이용하였습니다.

프로젝트명: [인텔리콘연구소] OCR Parser(Python)

기간: 2019.09 - 2019.11

내용: 계약서 이미지를 문단 Level의 데이터로 변환시키는 데이터 엔지니어링 작업이었습니다. Parser로 Google Vision API를 이용하였으며, 주요 처리 작업은 정렬, 오류 데이터 제거, 병합, 인덱싱, 회전이었습니다.

프로젝트명: [대학원] 기사 댓글의 감성에 영향을 주는 요인분석(Python)

기간: 2018.09 - 2018.12

내용: 연애 기사의 Feature가 기사의 댓글 감성에 어떤 영향을 주는지를 살펴본 연구입니다. 연구의 Phrase 1에서는 댓글 2000개를 라벨링하고 scikit-learn의 TfidfVectorizer와 Logistic Regression을 이용하여 감성 분류기를 만들고 Phrase 2에서는 기사별 댓글 감성 분류를 통해서 종속 변수로 댓글 감성 점수가 할당된 네이버 기사 1100여개에 대하여 기사의 Feature(기사 소재의 성별, 직업 등)를 이용한 회귀분석을 진행하였습니다.

프로젝트명: [대학원] 사이버불링 억제를 위한 신고시스템 성공 요인
(Experiment research)

기간: 2018.08 – 2018.12

내용: 본 연구는 억제이론 및 IS(Information Systems)성공모델을 활용하여 사이버불링 억제를 위한 신고시스템 성공 요인을 규명한 연구입니다. 억제의 성격을 가지는 신고시스템을 성공적으로 디자인할 수 있는 연구모델을 만들었으며, 시스템을 어떻게 하면 사용 활성화할 수 있는지에 초점을 뒀고 포커스 그룹 인터뷰 및 시나리오 기반의 실험 결과 사이버불링 억제, 신고시스템과 관련된 실무 관리자들에게 신고시스템 디자인과 관련된 실무적 시사점을 제공하였습니다.

활동

활동명: 사내 파이썬 심화 프로그래밍 온라인 강의

기간: 2020.12 – 2021.02

내용: 사내에서 주당 3시간씩 파이썬 프로그래밍 강의를 수강했습니다. 주요 내용으로 Decorator, Generator, Concurrent 등을 다뤘으며 알고 있던 본 개념들을 다시 익히는 시간이 되었습니다.

활동명: 데이터분석 전문가(ADP) 스터디

기간: 2020.10 –

내용: 데이터분석 전문가란 데이터 분석과 관련된 국가공인자격(증)입니다. 필기 스터디 및 시험 응시를 통해 20.12에 필기 시험을 합격하고 실기 스터디를 진행하고 있습니다.

활동명: 사내 딥러닝 온라인 강의

기간: 2020.06 – 2020.10

내용: 회사에서 정기적으로 온라인 딥러닝 강의를 수강할 수 있는 시간을 제공해주었습니다. 이를 통해 딥러닝의 수치 미분, 역전파를 이용한 경사하강법과 CNN, RNN 등을 딥러닝 프레임워크 없이 numpy 패키지로 구현하며 딥러닝을 복습하였습니다.

활동명: Kaggle Platform 이용

기간: 2019.01 – 2019.05

내용: Kaggle Competitions에 참여했습니다. 라이브러리 및 프레임워크로 sklearn, keras를 주로 이용하였으며 Batch Normalization, Dropout 등 개념적으로만 알고 있던 것들을 사용해보는 기회가 됐고 이론으로만 공부했을 때는 필요성을 잘 알지 못했던 Feature Engineering의 중요함을 알 수 있었습니다.

활동명: 연세대 바른ICT연구소 활동

기간: 2017.03 – 2018.12

내용: 바른ICT연구소는 건전한 ICT 문화 구축에 기여하는 것을 목적으로 하는 교내 연구소이며, 연구조원으로서 개인정보보호에 관한 세미나 및

프로젝트에 참여하는 것이 주된 활동이었습니다. 본 연구소에서 주관하는 'Asia Privacy Bridge Forum', 'Barun ICT Research & ISACA Korea Conference' 등의 참여하였고 한국인터넷진흥원, ISACA 등의 기관의 개인정보에 관련된 프로젝트에 참여하였습니다.

활동명: 토픽 모델링 기반 청원 데이터 이슈 분석

기간: 2018.08 – 2018.11

내용: 청와대 청원 웹에서 청원 데이터를 스크래핑하고 전처리한 후 토픽 모델링,

네트워크 분석을 통해 이슈를 분석하였습니다. 이후 데이터 수집부터 인사이트 도출까지의 전과정에 흥미를 느끼고 남북 문제, 영화 평가, 암호 화폐 등에 대한 유사한 분석을 진행하였습니다.

활동명: 상명대 음성인식 기술을 이용한 인터랙티브 기기 개발

기간: 2016.05 – 2016.09

내용: 실생활에서 리모컨, 열쇠 같이 사소하지만 당장 필요한 물건의 위치를 기억하지 못해, 이를 찾기 위해 번거로운 수고를 겪어야 하는 일이 흔히 발생하는데 이러한 불편을 아두이노를 활용하여 특정 소리의 패턴에 반응하여 각 패턴에 따라 상이한 LED와 사운드를 출력하는 기기를 개발하였습니다.

활동명: 트레인스토리 영업 업무

기간: 2015.04 – 2015.12

내용: 트레인스토리는 모형 기차를 독일(주로 marklin)에서 수입해서 그것을 기반으로 디오라마를 제작, 판매하는 회사입니다. 친구가 운영하는 곳 이었고 당사에서 영업 업무를 담당했습니다.

활동명: 상명대 42대 총학생회 '사이좋은' 선거운동 및 임원 활동

기간: 2014.10 – 2015.04

내용: 총학생회 선거운동에 참여하여 운동본부를 홍보하고 선거를 독려하는 등의 홍보 활동을 진행하였고 당선 이후, 시설복지팀 임원직을 맡아서 역할을 수행하다 상기 언급한 트레인스토리에서 일하기 위해 중도 사퇴하였습니다.

활동명: 상명대 암호학 연구원 활동

기간: 2014.10 – 2015.04

내용: 암호학 국가 과제에 연구원으로 참여하였습니다. C 언어를 기반으로 DES, AES, LEA 등의 암호 알고리즘을 기존 방법대로 구현하고, 복잡도를 개선하는 작업을 진행하였습니다.

경력

기업명: 인텔리콘연구소 (전문연구요원으로 재직중)

직급: 연구원

기간: 2019.05.15 -

자격사항

자격명: 정보처리기사