

비정형 문서에서 정형데이터의 변환



팀 이름 : REMODY
담당 교수 : 안희철 교수님
조장 : 이정훈, 조원 : 김효일, 오태현, 이경민

1. 작품 개요 및 필요성

설문조사

PDF 문서를 사용하는 경우가 많다.
PDF 글이 많아 필요한 데이터를 얻기 힘들다.
PDF를 재활용하여 관련 데이터를 이용하고 싶다.

Q. PDF 문서 변환 사이트를 이용하는 이유?

관련 데이터를 얻기 위해 40%	영어 해석 30%	자료 수정 30%
-------------------	-----------	-----------

Q. 기존 PDF 변환 사이트에서 어려움을 겪는 이유?

이상한 해석 60%	깨지는 어휘 20%	변경되지 않은 사진 12%	기타 8%
------------	------------	----------------	-------

많은 사용자들이 PDF 문서 변환사이트를 이용하고
있지만 관련 서비스가 잘 되지 않는다는 점을 파악

2. 개발 환경 및 사용 기술

1. 클라이언트 & 서버

- ① React : 손쉬운 SPA구성과 성능을 위해 사용 (웹 프론트)
- ② GraphQL-yoga : 서버 제공 및 테스트와 API DOC를 제공
- ③ Prisma : GraphQL을 이용한 ORM(Object CRUD 기능)
- ④ ElasticSearch : 해쉬맵으로 하여 빠른 검색 제공

2. Module

- ① Konlpy : 자연어 및 의미기반 언어 처리 프로세스
- ② word2Vec : 입력한 "key word"에서 유사한 데이터 추출 모듈

3. Cloud Service

- ① EC2 : 컴퓨팅파워를 클라우드에서 제공하는 웹 서비스
- ② RDS : 안전성을 확보한 데이터베이스 서버



그 외에도 유저 저장(MySQL), 프록시 서버(NGINX), 유저정보 암호화(JWT) 사용

3. 개발 내용 및 과정

서비스 목표 수립

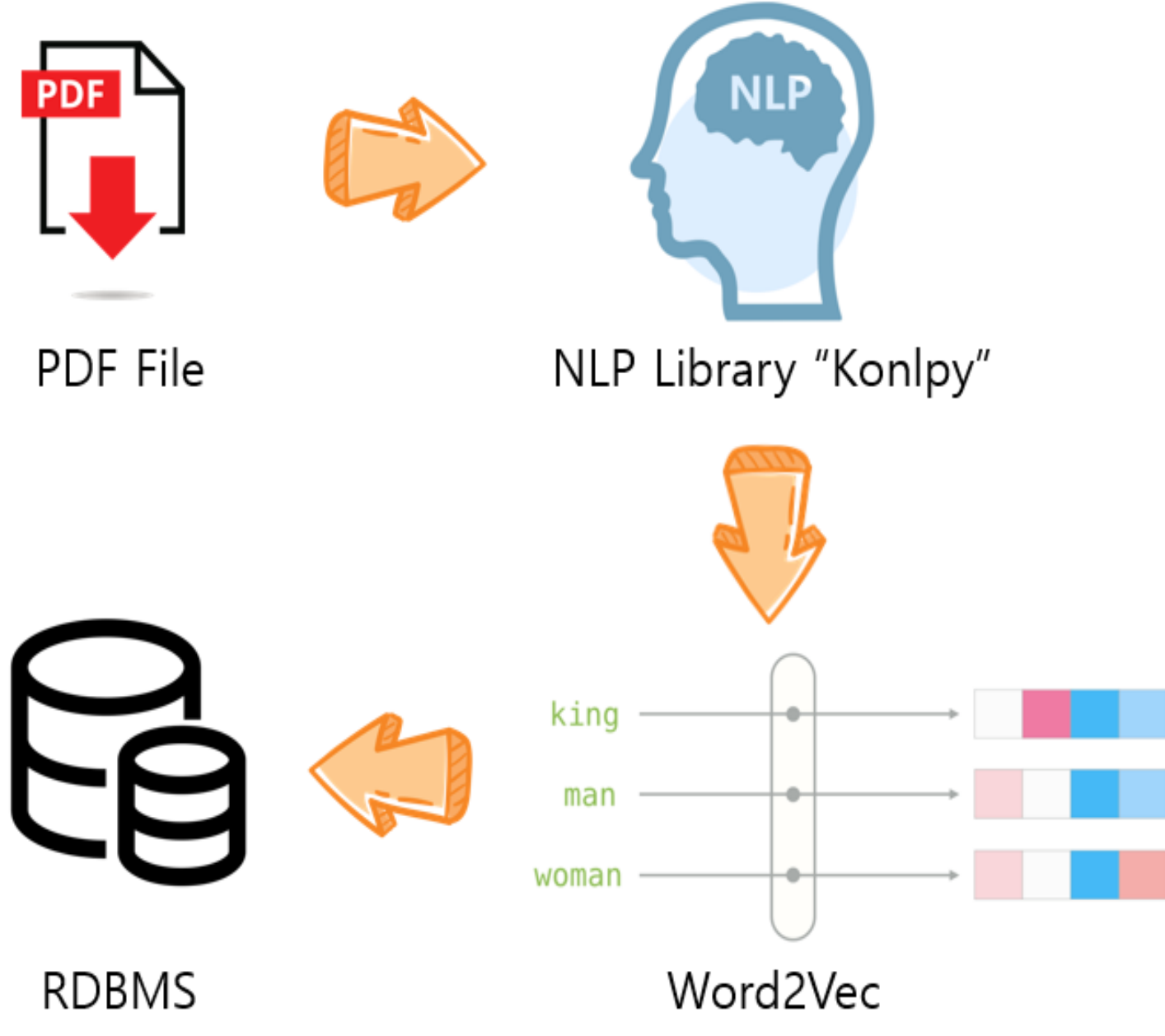
✓ 정확한 데이터를 얻을 수 있는 PDF 변환 사이트를 만들자

✓ 사용자가 편리하게 이용할 수 있는 기능을 추가하자

IDEA



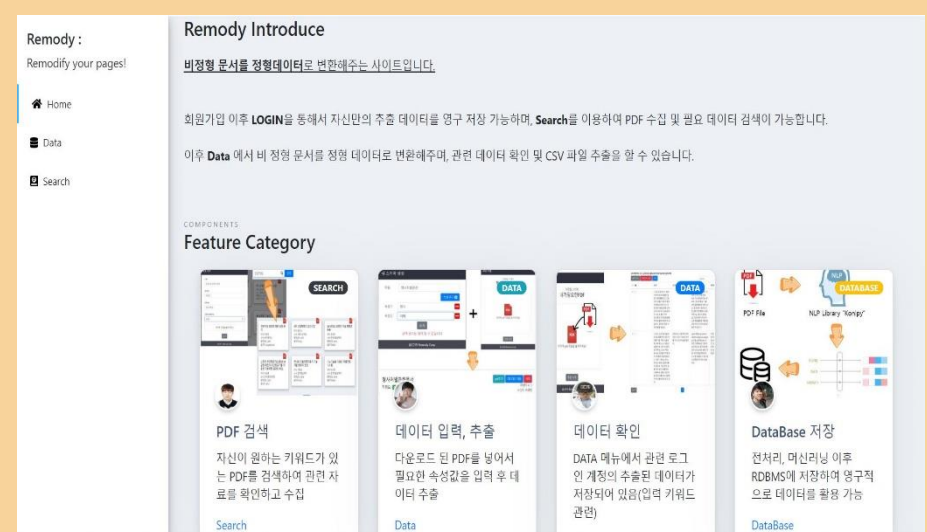
사용자 UI



모듈 처리 과정

4. 개발 결과

1. 사용하기 쉬운 UI



I 제작 의도

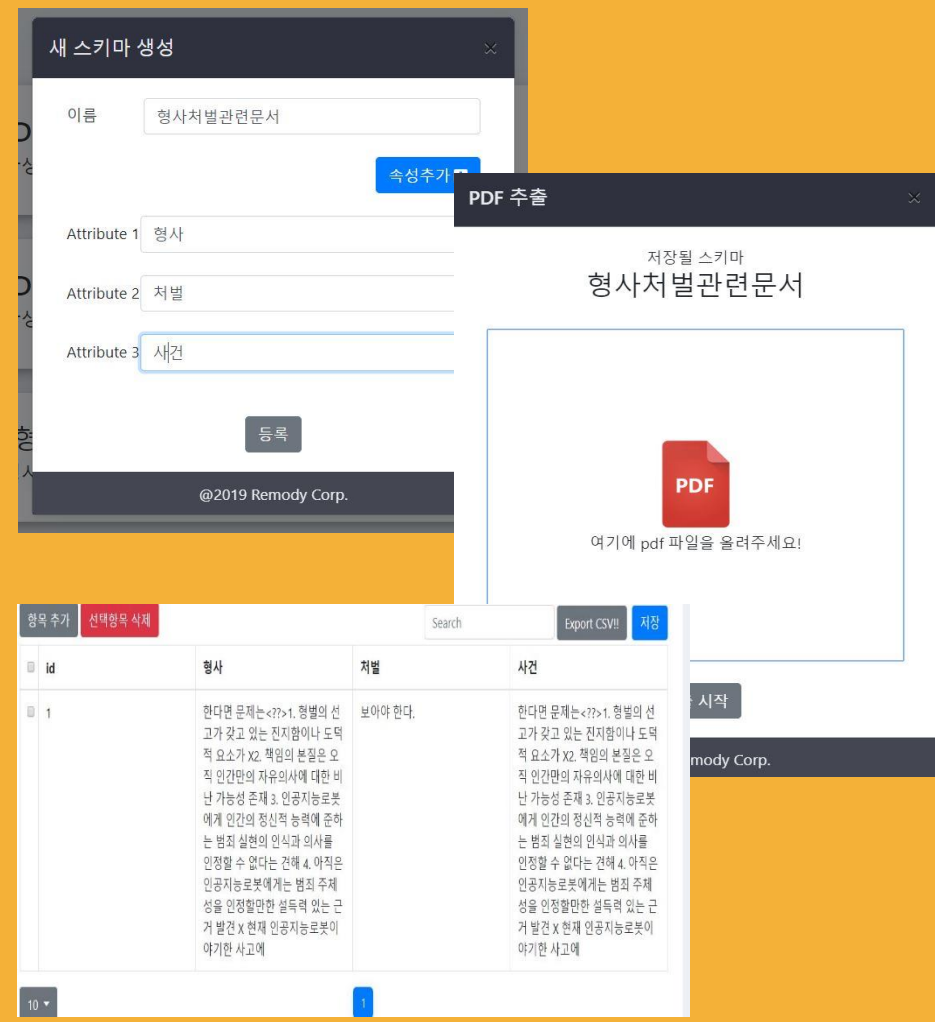
기존 사이트 들은 소개 및 메뉴
구성이 제대로 되어 있지 않았음.

메인 페이지에 설명을 적어 돕으로
써 사용자의 접근성을 더 높일 수
있었다.

2. 정확한 데이터 추출

I 제작 의도

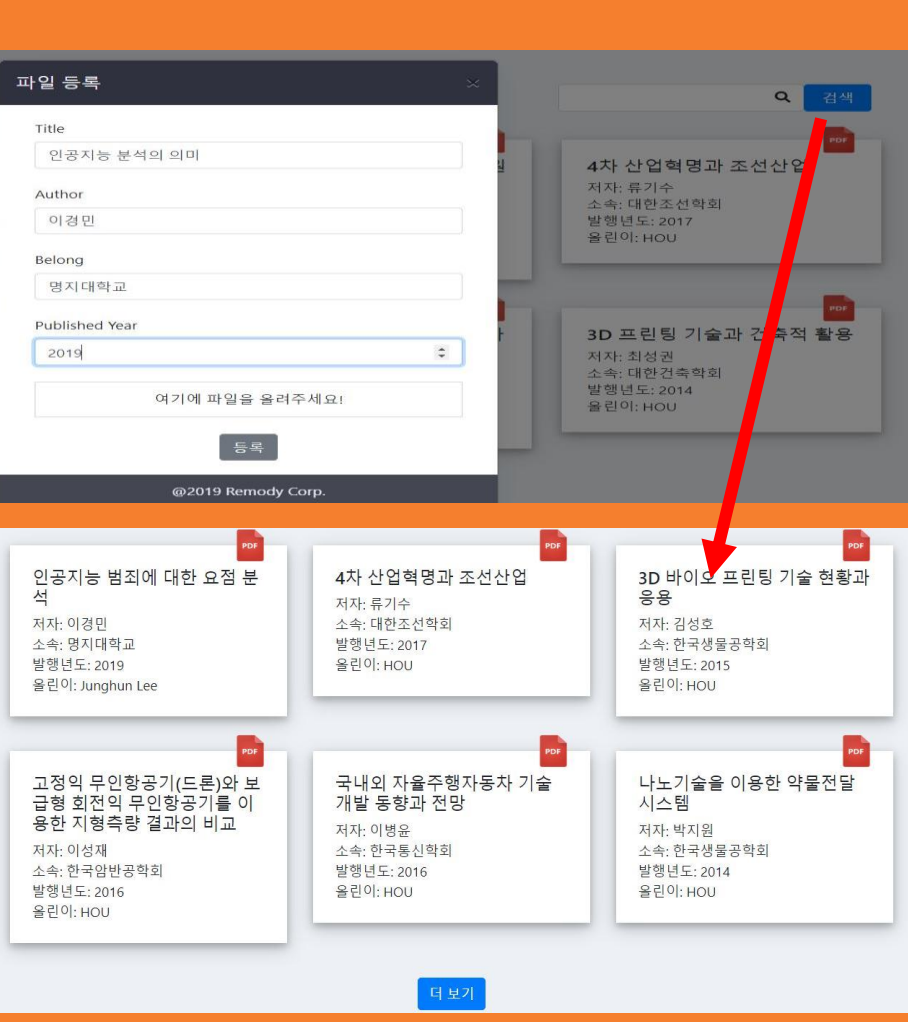
가지고 있는 PDF 에서 원하는
"KEYWORD" 를 입력하여 관련
데이터를 확인 및 수집한다.



3. PDF 검색

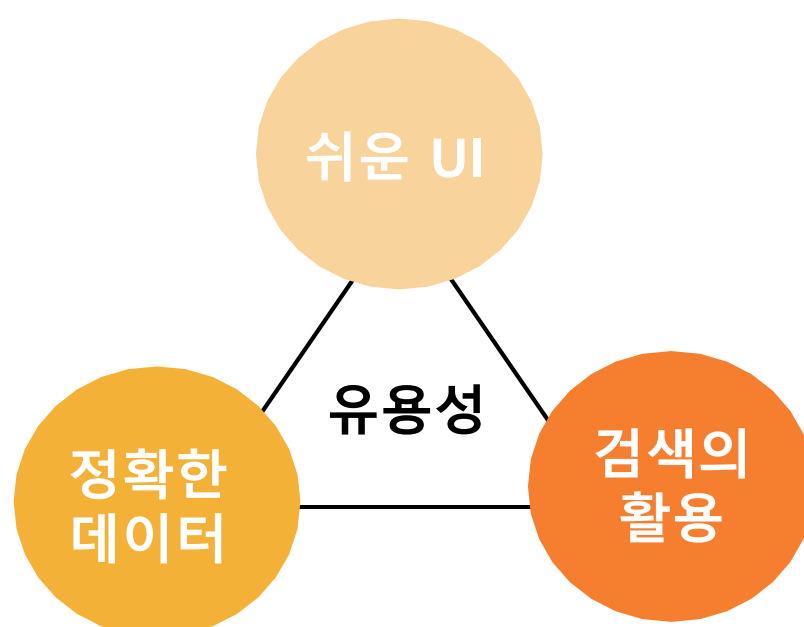
I 제작 의도

자신이 원하는 키워드가 있는
PDF를 검색하여 관련 자료를
확인하고 수집한다.



5. 결론 및 기대효과

- 현재 서비스를 사용하기 쉽게 유용성 제공
- 관련 문서의 데이터를 수집하여 활용 함으로 써 유용성이 높음
- 찾으려는 데이터가 포함된 문서를 검색 하여 다른 문서와 함께 유용하게 활용



활용 분야

- 1. 논문 제작에 도움
- 2. 정형 데이터로 변환하여 영구보관
- 3. 정형 데이터 가공 및 공유