

대한민국 ICT 인재를 키우는 힘, 한이음 ICT 멘토링 프로젝트 참여기 - 1

FindMap : 빅데이터, AI 기반 맞춤형 검색 및 아카이빙 공유 플랫폼 구축

글. 강자원 컴퓨터시스템응용기술사 · KBS MNC(Media Network Center)팀
아주대학교 우다현, 오승민, 승현수, 박지윤

한이음 ICT 멘토링은 2004년부터 과학기술정보통신부가 지원하고 정보통신기획평가원(IITP)이 주관하는 인력 양성사업의 일환으로 시작되었다. 대학생(멘티)이 ICT 기업전문가(멘토)와 팀을 이루어 프로젝트를 수행함으로써 ICT 실무 역량을 향상하는 대한민국 ICT 분야 대표 멘토링 프로그램이다.

방송과기술 지면을 통해 필자가 2021년 4월부터 11월까지 아주대 학생들과 함께했던 인공지능 관련 프로젝트에 대해 소개하는 시간을 가져보려 한다. 수행 프로젝트에 대한 내용은 학생들이 직접 작성한 결과 보고서를 바탕으로 작성하였다.

한이음 엑스포

한이음 공모전을 통해 수상한 프로젝트들을 중심으로 작품 전시를 진행하는 한이음 엑스포가 2021년 12월 3일 개최되었다. 코로나19 확산 방지를 위해 일반 관람객 없이 최소한의 인원이 참여한 오프라인 행사와 온라인 생방송으로 동시에 진행했으며, 화상회의 앱을 활용해서 시상식을 비롯해 유명 유튜버와 함께하는 성과물 설명회와 특강 등은 실시간 영상으로 방송되고 참여하는 체험형 비대면 행사로 진행되었다. 이번 엑스포에서는 한이음 공모전에 총 499팀이 지원했고, 1차 서류평가와 2차 심층평가, 최종 종합평가 등을 거쳐 최종 195개 팀을 선발했다. 특별히 우리 멘티들은 상위 23개 팀 안에 들어 동상을 받았다. 그리고 추가로 다음과 같은 결실을 맺었다.



그림 1. 수상 사진, 박지윤, 우다현, 오승민, 승현수



한이음 논문학술대회 - 우수상



한국정보처리학회 논문 개재

- 검색환경 개선을 위한 자연어 처리 기반
맞춤형 추천 검색시스템



한이음 공모전 - 동상



이번 한이음 ICT 멘토링 사업에는 약 3,000여 명의 멘티가 2대 1의 경쟁률을 뚫고 선발됐다. 필자는 빅데이터 기반의 AI 관련 프로젝트를 아주대학교 학생들과 수행하였고, 열성적이었던 4명의 멘티 덕분에, 한국정보산업연합회로부터 프로젝트 지도에 기여한 공로가 크다며 공로패도 받게 되었다. 개인적으로 지금까지 학생들을 지도해 오던 중 가장 의미 있는 프로젝트였다. 이런 보람 덕분에 한이음 멘토링을 지속해오는 것이 아닐까 싶다.

프로젝트명 : FindMap (빅데이터, AI 기반 맞춤형 검색 및 아카이빙 공유 플랫폼 구축)

팀 명		아주NICE					
팀 원		이 름	소 속	부서/학과	역할	담당업무	직위/학년
멘 토		강자원	KBS	MNC	멘토링	프로젝트 지도	감독
멘 티	팀장	우다현	아주대학교	소프트웨어학과	PL, 개발	API 서버 개발, DB 설계 및 구축	3학년
	팀원1	오승민	아주대학교	소프트웨어학과	PM, 개발	프로젝트 매니저, Mobile Client 개발	4학년
	팀원2	승현수	아주대학교	전자공학과	개발	Recommender system 개발, NLP 서버 개발	4학년
	팀원3	박지윤	아주대학교	소프트웨어학과	개발	NLP Embedding 학습 데이터 구축, Category Classification 모델 개발, Mobile Client 개발	3학년

프로젝트 소개

‘FindMap’은 사용자 맞춤형 검색 및 아카이빙 공유 플랫폼이다. 사용자에게 맞춤형 검색 결과를 제시하고 사용자의 취향을 예측한 추천 검색어를 제시한다. 또한, 검색 중 저장하고 싶은 게시글이 있을 때 나만의 아카이브에 스크랩하고 관리할 수 있다. 마지막으로 SNS 기능을 추가해서 다른 사람들과 게시글을 공유하며 소통할 수 있다.

프로젝트 사용자 시나리오

아주대학교에 재학하고 있는 김구직 학생은 곧 졸업을 앞둔 취준생이다. 자소서를 작성하는데 자료검색만 평균 3시간이 걸린다. 김구직 씨는 취준생들에게 필요한 정보만 검색할 수 있는 맞춤형 검색이 절실히 필요했다. 그래서 ‘FindMap’ 앱을 다운로드받았다. FindMap을 통해 검색하면 내가 검색한 것의 선호도에 따라 맞춤형 검색 결과를 제공해준다. 또, 검색 결과를 연관검색어와 함께 마인드 맵 형태로 보여준다. 김구직 학생에게 연관검색어를 별도로 입력할 필요 없이 그 단어를 탭하면 선택한 단어를 중심으로 또다시 연관검색어를 제시해주는 맵은 매우 유용한 기능을 제공했다. 그리고 검색 결과를 내 아카이브에 저장할 수도 있다. 또한 김구직 학생이 검색한 결과는 FindMap을 사용하는 다른 사용자들과 공유할 수도 있고 피드에 게시할 수도 있다. 검색 결과를 공유하는 신개념의 SNS다. 취준생들에게 유용한 검색 결과를 아카이빙할 수 있고 함께 공유하며 취업 준비를 위해 소요되는 시간을 절약할 수 있다.

최종 구현 화면



그림 2. 최종 구현된 모바일 앱 및 기능 소개

역할 및 수행내용 설명

● FindMap 앱 Front-end, Client 개발 부문

FindMap의 Mobile Client 개발 담당자 프로필



이름 : 오승민

소속 : 아주대학교 소프트웨어학과 4학년

역할 : PM(Project Manager)

Front-end, client 개발

기능 : Scrum과 XP를 결합한 Agile 방법론 적용

Docker 기반 CI/CD 환경 구축

Flutter를 활용한 iOS, Android 호환 앱 개발

수행 내용

가. Project Manager

개발방법론 : Agile 방법론 중 Scrum + XP

애자일 프로세스를 진행하기 위해 프로덕트 백로그와 스프린트 백로그를 작성하였다. JIRA 툴을 이용해 이슈 단위로 백로그를 만들고 칸반 차트로 진행 상황을 공유하며 프로젝트를 관리하였다. 매주 BDC(Build, Development, Commit)를 이용해 개발 진도에 차질이 없는지 확인하고 문제 상황들을 공유했다. 프로젝트가 원활하게 진행할 수 있도록 회의를 하며 개발을 진행했다. 스프린트가 끝난 후에는 회고를 통해 스토리 포인트를 조절하며 팀원들이 무리하게 개발하지 않도록 신경 썼다.

나. Docker, Docker-compose, CI/CD 환경 구축

개발 서버, 테스트 서버, 호스팅 서버를 따로 두고 프로젝트를 진행했다. 또한, 하나의 서버 안에서 다시 Proxy 서버, API 서버, ML 서버, DB 서버를 따로 만들어 구현하였다. 각 서버를 분리하여 구현한 이유는 의존성 충돌 문제를 해결하고, 개발 환경을 분리함으로써 개발 및 유지보수를 편리하게 할 수 있기 때문이다. 우리는 형상 관리에 gitlab을 사용했기 때문에 gitlab-runner를 이용해 CI/CD를 만들었다.

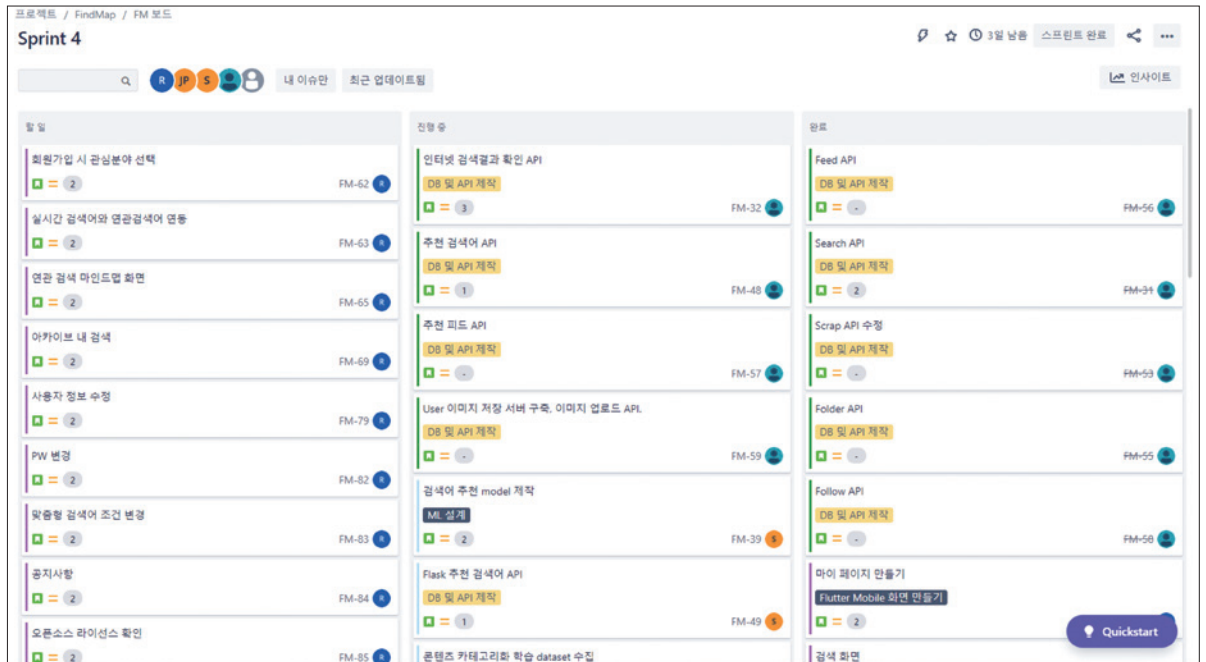


그림 3. 스프린트를 백로그 단위로 관리

● FindMap 앱 인프라 설계 - Back-end 개발

담당자 프로필



이름 : 우다현

소 속 : 아주대학교 소프트웨어학과 3학년

역 할 : PL(Project Leader)

Back-end 개발, API, DB 서버 설계 및 구축

기 능 : AWS를 활용한 인프라 설계 및 구축

API 설계, DB 설계

수행 내용

가. 인프라 설계

AWS RDS(Amazon Relational Database Service)를 구성했다. API Server에서 Nginx 웹 서버를 이용해 트래픽을 관리하고 Node.js 환경에서 Express 프레임워크를 통해 API를 제공하는 일련의 시스템을 구성했다. AWS RDS 서비스를 이용해서 API Server와 DB Server를 분리해서 데이터를 안전하게 관리했다.

나. DB 설계

개념적 모델링, 논리적 모델링 과정을 거쳐 DB 모델을 설계했다. 정규화 규칙을 따르려고 노력했으며, 어떻게 하면 데이터베이스의 낭비 없이 효율적으로 데이터를 관리할 수 있을가에 대해 고민했다. 데이터베이스 모델이 처음부터 완벽하지는 않았지만, API를 만들며 지속적인 수정을 하는 과정에서 더 나은 모델을 만들 수 있었다.

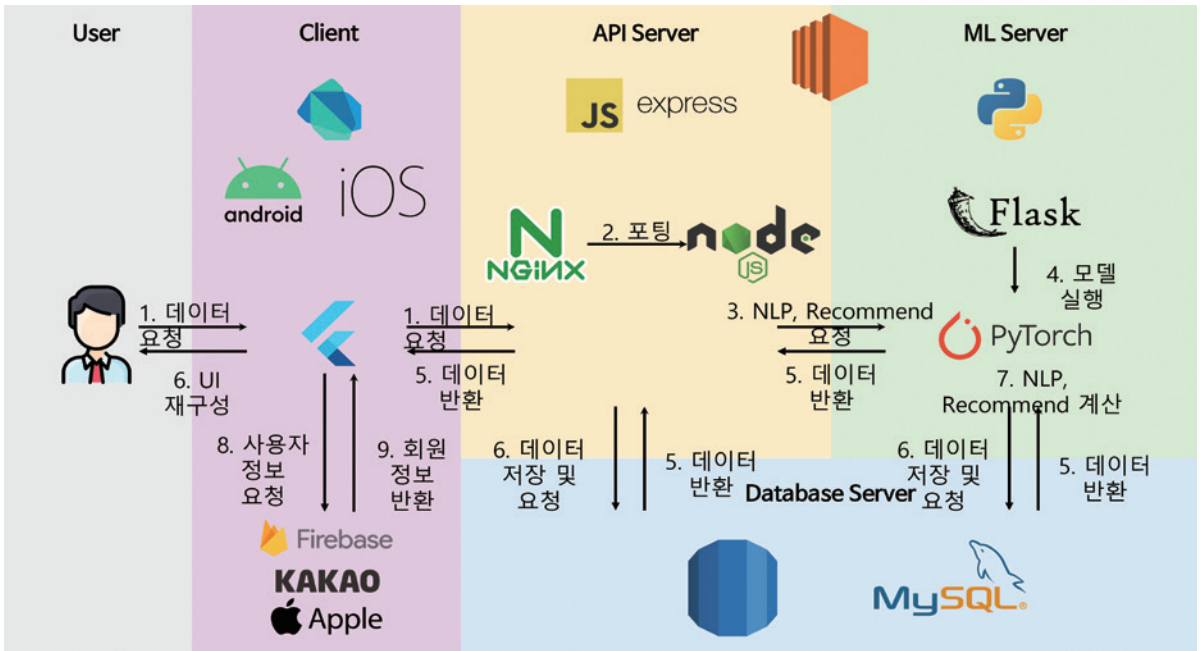


그림 4. 인프라 설계 기반 서비스 흐름도

다. API 설계

서비스에 필요한 API를 client에 제공했다. Restful API 규칙을 최대한 지켜 FindMap에서 제공하는 기능별 분류에 따른 API를 설계했다.

● FindMap 앱 알고리즘 #1 개발 - 추천시스템 및 NLP 서버 개발

담당자 프로필



이름 : 승현수
소속 : 아주대학교 소프트웨어학과 4학년
역할 : Recommender system 개발
 NLP(Natural Language Processing) 서버 개발
기능 : item based collaborative filtering 알고리즘 적용
 Flask를 활용한 머신러닝전용 API 개발

수행 내용

가. Recommend System 개발, Flask API 구축 및 연동
 사용자 검색어 입력 시 선호도 순위 기반 추천알고리즘을 구현하였다. 추천 알고리즘은 검색어의 카테고리를 기준으로 하는 item based collaborative filtering을 사용했다. 이는 검색어 간의 cosine 유사도를 통해서 지저행렬을 만들고 사용자가 어떠한 카테고리를 좋아할지를 예측값으로 나타내주어 그 값이 높은 것(사용자가 좋아할 만한 카테고리의 수치)을 추천해주는 구조이다.

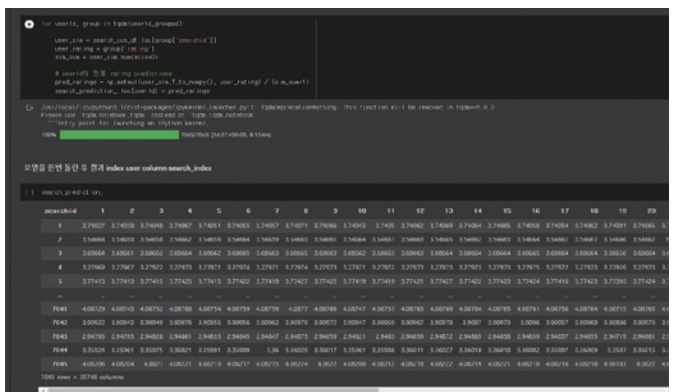


그림 5. 추천알고리즘 구현

나. NLP(National Language Processing) 서버 개발

추천시스템의 Cold start 문제(새로운 또는 어떤 사용자에게 대한 충분한 정보가 수집된 상태가 아니어서 해당 사용자에게 적절한 제품을 추천해주지 못하는 문제)를 해결하기 위해서 NLP를 사용했다. word embedding을 통한 단어 거리를 통해 추천하도록 시스템을 보완하였다. 크롤링을 통한 데이터를 사용하여 KoNLP, Fasttext, LSTM, Pororo 등 다양한 라이브러리를 통해서 성능을 테스트하였고 그 결과에 따라 Fasttext로 word embedding을 활용했다. 또한, 학습 전 데이터 전처리를 통해 데이터의 정확도를 높이려는 작업을 진행했다. 이를 통해서 최종적으로 30개의 유사도가 높은 단어들에서 조사를 제외하고 결과를 출력하도록 구성했다.

● FindMap앱 알고리즘 #2 개발

- NLP Embedding 학습 데이터 구축, Category Classification 모델 개발

담당자 프로필



이름 : 박지윤

소속 : 아주대학교 소프트웨어학과 3학년

역할 : NLP Embedding 학습 데이터 구축

머신러닝을 위한 학습 데이터셋 구축

아카이빙 기능 알고리즘 구현 및 개발

기능 : FastText를 사용한 Text Classification Model 적용

Category Classification 모델 적용

수행 내용

가. Text Classification Model 제작

추천시스템 모델을 개발하기에 앞서 사용자가 검색한 검색어와 콘텐츠로 취향을 유추할 수 있도록 text classification 모델을 도입하기로 했다, 그 중 KoBert, Pororo, FastText 등 여러 모델을 비교하는 테스트를 수행했다. 비교 결과 정확도 및 모델의 무게, 학습 시간 등을 고려하여 FastText를 사용하여 모델을 개발했다.

나. 데이터셋 구축

앱에 적용할 수 있는 형태로 모델을 개발하기 위해서 우리의 모델에 맞는 데이터셋이 필요했다, 그래서 블로그나 커뮤니티 글 등을 크롤링하여 학습시킬 데이터셋을 구축했다. 데이터셋을 구축할 때, 데이터 특성상 html 문서에서 가져온 text이기 때문에 html 문서 내 사용되는 특수문자가 섞여 비정상적인 데이터가 섞이는 것을 방지하기 위해 데이터 전처리 작업을 수행했다.

다. 사용자 취향 계산

Ranking System을 개발하여 OpenAPI를 통해 받아온 데이터를 사용자 맞춤형으로 결과를 제시하는 모델을 제작하였으나, 사용자의 취향을 계산하기 위해 API가 호출될 때마다 계산을 해서 비용이 많이 들고, table join을 다수 수행해서 비효율적이었다. 따

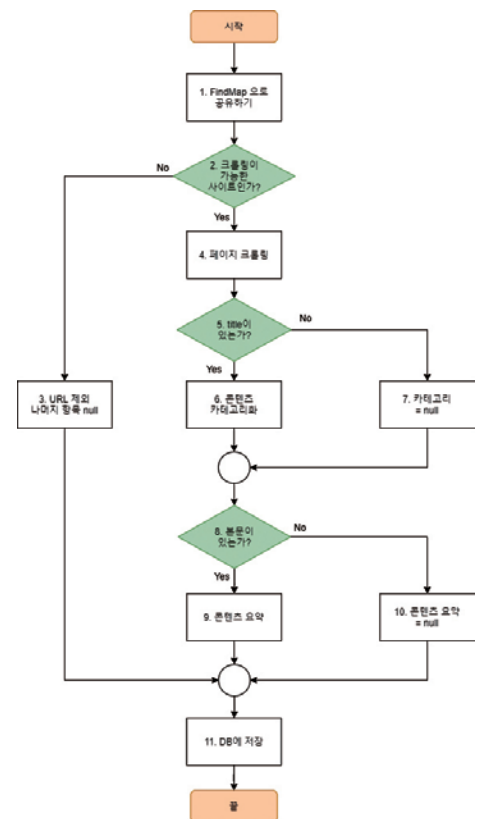


그림 6. 콘텐츠를 FindMap으로 공유 시, 동작하는 내부 알고리즘

라서 성능을 개선하기 위해 DB에서 필요한 칼럼만 가져온 후, python의 dataframe을 사용하여 사용자 취향을 빠르게 계산하여 성능을 개선했다.

라. 콘텐츠 저장을 위한 아카이빙 기능 구현

사용자가 검색한 콘텐츠를 아카이빙하기 위한 알고리즘을 설계하고 기능을 구현했다.

프로젝트 수행 후기

강자원 멘토 : 한이음 프로젝트를 수행하면서 이번처럼 열정적이고 실력 있는 멘티들을 만난 건 처음이었습니다. 매달 프로젝트 수행을 위한 과제를 내주면 언제나 그 이상을 해오는 멘티들은 프로젝트에 정말 진심이었어요. 공모전, 학술대회 등 모든 부분에서 수상하며 멘토에게도 큰 보람을 가져다준 멘티들에게 감사의 마음을 전하고 싶어요. 덕분에 행복한 2021년이었습니다.

오승민 : 이번 프로젝트에서 맡았던 직무는 Front-end 개발과 Project Management였어요. 두 직무 모두 처음 해보는 분야여서 조금 부담이 있었습니다. 개발은 혼자 다양한 시행착오를 해보면 좋은 코드를 만들 수 있으니 큰 부담이 덜 했지만, PM은 제 역할을 못 하면 팀에 피해가 가기 때문에 특별히 많은 시간을 쏟았습니다. 처음 해보는 애자일 프로세스를 제대로 적용하지 못하면 어쩌나 걱정이 많았는데, 팀원들이 모두 흥미를 느끼고 잘 도와주었기 때문에 성공적으로 마무리할 수 있었던 것 같아요. 마지막으로, 부끄러워서 제대로 말하지 못했지만, 이번 기회를 통해 모두에게 감사하고 정말 좋은 팀원이었음을 꼭 말해주고 싶어요.

우다현 : 제가 맡았던 직무는 API 서버 개발이었습니다. 서버 환경을 구성하고, DB를 설계하고, API를 만들었어요. 일상생활에서 앱을 사용하면서 서비스들의 작동 원리에 대해서 궁금했었는데 이번 기회를 통해 확실하게 이해하게 되었어요. 서비스를 만드는 과정에서 많은 흥미를 느꼈고, 앞으로도 계속해서 서비스를 만들어보고 싶다는 생각이 들어요. 이렇게 좋은 경험과 기억을 남길 수 있었던 것은 팀원 합이 좋았기 때문이에요. 다들 정말 적극적으로 참여해주어서 결국에 좋은 성과를 낼 수 있었던 것 같아요. 팀원들과 멘토님 모두에게 정말 감사합니다.

승현수 : 늦은 나이에 개발자가 되고 싶다고 생각을 했습니다. 지금까지 이렇게 팀프로젝트를 체계적으로 해본 경험이 없었습니다. 이번 기회를 통해서 협업의 중요성을 깨닫고 개발자로서 한 걸음 더 내디딜 수 있어서 좋았습니다. 또한 저희팀의 훌륭한 개발자님, 책임감 있는 PM님과 팀장님을 만나서 더 많이 배웠고, 제 부족한 능력에도 불구하고 항상 긍정적인 답을 해주셔서 감사합니다.

박지윤 : 처음 참여해본 대외활동이라 팀원들의 발목을 붙잡지 않을까 걱정을 많이 했어요. 그래서 맡은 분야는 완벽하게 수행하려고 공부를 참 많이 했던 것 같아요. 프로젝트에 참여하면서 여러 가지 분야를 경험해보았는데, 저는 특히 데이터를 위주로 개발을 하다 보니 데이터를 효과적으로 다루는데 관심이 많이 커졌어요. 앞으로 데이터를 어떻게 하면 더 잘 다룰 수 있을지 공부해보려고 해요. 저에게는 좋은 멘토님과 팀원들을 만나 많은 것들을 배우고, 경험한 아주 소중한 기회였어요. 🍀



그림 7. 한이음 엑스포 수상식에 참여 후 뒤줄이 (좌로부터 강자원 멘토, 승현수, 오승민, 박지윤, 우다현)