

연구계획서

연구계획	<p>(1) 연구배경</p> <p>1. 고도화된 개인 맞춤형 기업 및 진로 추천 플랫폼의 필요성</p> <ul style="list-style-type: none"> - 대학원생 맞춤형 취업/진로 정보 제공의 부재 <ul style="list-style-type: none"> • 학부생들에 비하여 대학원생들은 취업 및 진로 관련 정보가 상대적으로 부족하며, 이는 대학원생들의 진로 선택에 어려움을 야기함 • 각 개인의 세분화된 전공과 학기, 논문 등 다양한 조건을 고려한 맞춤형 취업 및 특화된 진로 정보 제공이 필요함 - 개인의 취업/진로 정보 탐색의 비효율성 <ul style="list-style-type: none"> • 직접 채용 정보를 검색하고 기업을 조사해 이력서를 작성, 제출하는 기존 방식은 시간이 많이 소요되고, 종종 개별 연구 분야와 회사의 필요 인제상이 일치하지 않는 경우가 존재함 • 또한, 외부에 잘 알려지지 않은 높은 잠재력을 가진 우수 기업들을 놓칠 위험이 있어, 학생들의 전문 지식 활용과 경력 발전에 제한을 줄 수 있음 - 상용화된 플랫폼의 개인별 맞춤 서비스 부족 <ul style="list-style-type: none"> • 기존의 상용화된 취업/진로 플랫폼은 넓은 범위의 전공에 대해 포괄적인 정보를 제공하는데 초점을 맞춤 • 개인화된 전문 지식 및 전문성을 지닌 대학원생들의 맞춤형 수요를 충족시키지 못함 <p>2. 선수 개발된 진로 추천 시스템의 한계점</p> <ul style="list-style-type: none"> - 기존의 [웹 크롤링 기반의 대학원생 취업/진로 지원 플랫폼]은 학생과 기업의 키워드를 추출한 후 코사인 유사도를 활용하여 매칭시키는 방식을 사용함 - 기존 서비스는 1) 추천 시스템의 정확도 부족; 2) 알고리즘 고도화를 위한 개인 맞춤화를 위한 데이터 부족; 3) 구체적인 채용 정보 제공 미흡; 4) 상용화된 서비스 부재라는 한계점을 지니고 있음 - 추천 시스템의 정확도 부족 <ul style="list-style-type: none"> • 학생 맞춤형 기업 추천을 위하여 네이버 뉴스 기사와 구글 스칼라 데이터에서 키워드를 추출하는 방식을 사용함 • n-gram을 사용하여 키워드 후보군을 생성하는 기존의 방식은 의미가 동일한 키워드를 중복 추출하거나, 고정된 n의 범위에서 벗어난 키워드를 추출하지 못하는 문제가 발생함 • 또한, 뉴스 기사 특성상 기업 및 진로와 관련된 핵심 내용에서 벗어난 무의미한 텍스트가 다수 존재함 • 기존의 방식은 텍스트 정제 능력 부족으로 인하여 '다수 발견'과 같은 부정확한 키워드를 추출함 - 알고리즘 고도화를 위한 개인 맞춤화를 위한 데이터 부족 <ul style="list-style-type: none"> • 저작권 문제와 개인정보 보호로 인하여 학생들의 개인적인 데이터의 이용이 어려움 • 이에 대한 보완책으로 이메일 주소를 사용한 구글 스칼라 키워드 추출 같은 간접적인 데이터 추출 방법을 사용함 • 하지만 출판된 논문이 없거나, 계정 주소가 다른 경우, 동일한 방법을 사용할 수 없다는 점에서 여전히 개인화된 데이터 부족 문제가 발생함
-------------	---

- 구체적인 채용 정보 제공 미흡
 - 선수 개발된 플랫폼은 대기업에 한정된 추천 서비스만 제공하며 경쟁력 있는 중소/중견 기업 및 스타트업을 반영하지 못함
 - 추천된 기업들에 대한 맞춤 채용 공고를 제공하지 못함
 - 근무 강도, 연봉 수준과 같은 평가 조건에 대한 개인별 선호도가 반영되지 못함
- 상용화된 서비스의 부재
 - 추천 시스템의 상용화를 위해서는 충분한 정확도와 다양성을 달성해야 함
 - 하지만 선행 개발된 플랫폼의 경우 앞서 서술한 한계점으로 인하여 상용화 될 정도의 성능을 달성하지 못함

(2) 연구의 목적 및 내용

1. 고도화된 개인 맞춤형 기업 및 진로 추천 플랫폼 개발 목적

- 기업의 규모 및 사업 분야를 고려하고 대학원생 개인의 전문 지식과 기업이 요구하는 취업 조건이 충족되는 맞춤형 기업 추천 및 취업 정보 제공함
 - 자신의 전공 지식 및 연구 경험을 최대한 활용할 기회를 제공하며 효과적으로 진로를 선택할 수 있도록 기여함
 - 개인 맞춤형 정보 제공을 통하여 취업 성공률을 향상시키고 시간과 노력을 줄일 수 있음
- 개인의 선호도와 우선순위를 반영하여, 대학원생들에게 맞춤형 채용 정보를 강화된 형태로 제공함
 - '커리어 향상', '업무와 삶의 균형', '급여 및 복지' 등의 기업 요소를 통합적으로 파악 후 기업 추천 순위에 반영함
 - 개선된 맞춤형 서비스를 통해 장기적으로 직업에 대한 만족도를 유지하고 진로 선택에 있어 만족스러운 결과를 이끌어내도록 지원함

2. 기존 진로 추천 시스템의 한계점 보완

- 추천 시스템 정확도 향상
 - 1) GPT4 API를 활용한 목적 적합한 맞춤형 키워드 추출
 - 기존의 n-gram을 대신하여 GPT4 API를 활용한 키워드 추출이 가능함
 - 추가적인 스크립트 명령어를 사용하여 기존 방식의 한계점인 무의미한 키워드 생성을 방지할 수 있음
 - 노이즈가 제거된 목적 적합한 키워드만을 사용하여 더욱 높은 정확도의 추천 시스템 구축이 가능함
 - 2) Text-Ranking 알고리즘을 사용한 키워드 추출
 - GPT4 API는 사용 시 비용이 청구된다는 한계점이 존재함
 - 이에 대한 대안으로 높은 신뢰도와 안정성을 지닌 Text-Ranking 알고리즘을 사용할 수 있음
 - 잡음 데이터가 다수 존재하는 네이버 뉴스 기사 키워드 추출 시 GPT4 API를, 정제된 데이터로 구성된 구글 스칼라 논문에는 Text-Ranking 알고리즘을 사용하는 방식으로 혼용할 수 있음
- 개인화 데이터 활용의 확대
 - 1) 학생 소속 학과 자체를 키워드로 사용
 - 학과의 특징과 학과명 자체를 키워드로 제공 받음
 - 이러한 키워드를 모델의 입력값으로 활용 시, 해당 학과의 특성을 고려한

	<p>맞춤형 기업 추천이 가능함</p> <p>2) 학생의 구글 스칼라에 등재된 논문들의 초록 활용</p> <ul style="list-style-type: none"> • 논문의 초록에서 키워드를 추출하여 사용함 • 추출된 키워드를 모델의 입력값으로 활용 시, 개인의 연구 분야에 따른 추천이 가능함 <p>3) [논문 추천 시스템]을 통해 축적된 데이터 활용</p> <ul style="list-style-type: none"> • 서비스 중인 [논문 추천 시스템]을 통해 구축된 개인 관심 분야 키워드를 사용함 • 제공 받은 키워드를 활용하여 개인의 세부적인 연구 분야를 파악 가능함 <p>4) 지도교수의 구글 스칼라 프로필 키워드 활용</p> <ul style="list-style-type: none"> • 개인별 지도교수의 구글 스칼라 프로필에서 추출한 키워드를 사용함 • 추출된 키워드를 모델의 입력값으로 활용 시, 소속 연구실의 세부 연구 분야를 파악 가능함 <p>5) 국가 R&D 과제 open API를 사용한 데이터 활용</p> <ul style="list-style-type: none"> • 국가 R&D 과제 open API를 활용하여 지도교수가 참여하고 있는 기업 및 정부 과제와 관련된 한글 키워드 및 연구 내용을 키워드로 사용함 • 추출된 키워드를 모델의 입력값으로 활용 시, 소속 연구실의 세부적인 연구 내용과 방향성을 명확히 파악할 수 있음 <p>- 맞춤형 채용 정보 제공 기능 강화</p> <p>1) 다양한 규모 및 산업 구조의 추천 기업 선정</p> <ul style="list-style-type: none"> • 기존 MK500지수 기반의 100개 대표기업을 활용하면서, 초격차 스타트업 1000+ 프로젝트에 선발된 150개의 유망한 기업들을 추가적으로 활용함 <p>2) 데이터 정제를 통한 채용 공고 정보 추출과 처리</p> <ul style="list-style-type: none"> • 채용 공고는 일반적으로 텍스트 형식보다 pdf 파일의 이미지 형식으로 제공되는 경우가 많기 때문에 데이터를 재구성하는 과정이 필수적으로 수반됨 • 기업마다 제공하는 채용 공고의 구성 형태가 다양하기 때문에 이미지 형태의 채용 공고문의 경우, 광학 문자 인식(OCR) 알고리즘을 통해 텍스트로 변환함 • OCR을 통해 텍스트로 변환된 채용 정보는 GPT4 API를 통해 불필요한 정보를 제거하고 표준화된 양식에 맞게 재구성하여 모든 채용 공고를 의미있게 정리할 수 있음 <p>3) 우선순위 기반 맞춤형 기업 정렬</p> <ul style="list-style-type: none"> • '블라인드' 리뷰 데이터를 크롤링하여 각 기업들의 '커리어 향상', '업무와 삶의 균형', '급여 및 복지', '사내 문화', '경영진'과 같은 기업 평가 항목에 대한 평점을 획득함 • 획득한 평점으로 가중치를 계산해 기업과 학생 간의 유사도를 계산하여 대학원생들이 선택한 우선 순위에 따라 학생과 기업 간의 최적의 매칭을 도출함 <p>- 웹기반의 상용화 가능한 서비스 개발</p> <p>1) 웹 기반의 상용화된 서비스 제공을 위한 클라이언트-서버 시스템 구축</p> <ul style="list-style-type: none"> • 클라이언트는 대학원생의 정보를 입력받아 서버에 전달하고, 서버는 전달 받은 정보를 이용하여 대학원생에게 특화된 기업 및 진로 추천 결과를 도출함 • 결과를 다시 클라이언트에게 보내는 것으로 서비스를 이용하는 대학원생
--	---

	<p>에게 맞춤 추천 서비스를 제공함</p> <p>2) 대학 포털 사이트를 활용한 서비스 접근성 향상</p> <ul style="list-style-type: none"> • 고려대 포털 사이트 또는 KUchive에서 학생 정보와 연동하여 서비스를 제공함 • 이는 대학원생들이 서비스에 접근하기 용이하게 하며, 추가적으로 잠재적 서비스 이용자들을 끌어올 수 있음 <p>3) 이메일을 이용한 정보 알림 서비스 제공</p> <ul style="list-style-type: none"> • 웹사이트를 이용한 서비스 접근 방식만을 제공할 뿐만 아니라, 이메일을 이용한 기업 추천 및 채용 공고에 대한 알림을 추가적으로 제공함 • 추천된 기업들의 다양한 채용 공고들을 간편하게 확인 가능함 <p>(3) 연구방법 및 일정</p> <p>1. 학생 맞춤형 관련 기업 추천</p> <p>1) 데이터 전처리 및 정제</p> <ul style="list-style-type: none"> - 학생의 개인 관심 분야 및 세부 연구 분야를 파악하기 위하여 5가지 경로의 데이터를 정제함 <ul style="list-style-type: none"> • 학과 정보: 학생의 소속 대학 및 학과명을 키워드로 사용함(예: 정보대학, 인공지능학과) • 구글 스칼라: 학생의 이메일 주소를 기반으로 구글 스칼라에 등재된 논문 초록을 데이터로 활용함 • [논문 추천 시스템] 데이터: 해당 서비스를 통해 축적된 학생의 관심 분야를 키워드로 사용함 • 지도교수 키워드: 소속 지도교수의 구글 스칼라에 기재된 관련 분야를 키워드로 사용함 • 연구 과제 키워드: 소속 지도교수명을 국가 R&D 과제 open API에 검색함으로써, 기재된 한글 및 영문 키워드를 사용함 <p>2) 구체적인 학생 키워드 추출 방법</p> <ul style="list-style-type: none"> - 앞서 기술한 5가지 데이터 중 '구글 스칼라'는 키워드를 추출하기 위하여 추가적인 전처리 방법이 요구됨 - GPT4 API를 사용하기 힘든 경우, Text-Ranking 알고리즘을 대안으로 사용하여 키워드 추출 가능 <ul style="list-style-type: none"> • 웹 크롤링: 학생 이메일 주소를 입력값으로, 이에 연동된 개인 구글 스칼라 등재 논문들의 초록 텍스트를 BeautifulSoup 파이썬 라이브러리를 활용하여 크롤링함 • GPT4 API 기반 키워드 추출: GPT4 API를 활용하여 해당 초록이 입력값으로 주어졌을 시, 적절한 n개의 키워드를 추출 가능 • 맞춤형 키워드 추출: "Extract n keywords regarding the author's personal research field"라는 구체적인 스크립트를 추가 기재함으로써, 목적 적합한 키워드만 추출 가능 • Text-Ranking 알고리즘: 각 텍스트를 노드, 텍스트 간의 관계를 엮지화시킨 후 점수에 따라 노드를 나열하여 키워드를 추출하는 방식으로, 높은 신뢰성을 지니고 있음 <p>3) 기업 관련 네이버 뉴스 기사 키워드 추출</p> <ul style="list-style-type: none"> - 기업의 키워드를 추출하기 위하여 기업 관련 뉴스 기사를 크롤링 및 정제 후 GPT4 API를 활용하여 관련 키워드를 추출함 <ul style="list-style-type: none"> • 뉴스 기사 검색: 기업의 다양한 관심 분야를 파악하기 위하여 네이버 뉴
--	---

	<p>스 검색란에 [기업명]과 [수집 키워드]를 함께 검색하여 도출된 최근 2년 뉴스 기사를 사용함(예: 삼성물산 신사업)</p> <ul style="list-style-type: none"> • 웹 크롤링: BeautifulSoup 파이썬 라이브러리를 활용하여 해당 뉴스 기사 제목 및 본문을 크롤링함 • GPT4 API 기반 맞춤형 키워드 추출: 구글 스칼라 키워드 추출과 유사하게, "Extract n keywords regarding the company and research fields in this article"라는 추가적인 입력 스크립트를 활용하여, 각 기업별로 적합한 키워드를 추출 가능함 <p>4) 학생별 맞춤형 기업 추천</p> <ul style="list-style-type: none"> - 앞선 단계에서 추출한 학생 키워드와 기업 키워드를 활용하여 학생별로 적합한 기업을 추천함 - 뛰어난 성능을 지닌 딥러닝 자연어 처리(NLP) 모델인 사전 학습된 BERT를 활용함 <ul style="list-style-type: none"> • 키워드 벡터화: 사전 학습된 BERT에 각 키워드를 입력값으로 제공하면 특정 잠재공간(latent space)에 위치한 벡터 형태로 변형됨 • 코사인 유사도 활용 거리 계산: 학생과 기업 키워드로부터 도출된 벡터들 간의 유사도를 코사인 유사도(cosine similarity)를 사용하여 계산함 • 기업 순위화: 학생별로 도출된 각 기업과의 유사도를 내림차순으로 정렬하여 상위 50개 기업을 추천함 <p>2. 학생 맞춤형 기업 채용 정보 제공</p> <p>1) 채용 정보 홈페이지 선별</p> <ul style="list-style-type: none"> - 추천된 기업들의 채용 공고를 확인함 <ul style="list-style-type: none"> • 일반 채용 정보 포털: 사람인, 잡코리아, 인쿠르트 • IT 및 스타트업 업계 채용 정보 포털: 로켓펀치, 원티드, 프로그래머스 • 공공 기관 및 대학 공고(예: 고려대학교 정보대학 공지사항) <p>2) 채용 공고 정보(상시, 공채, 인턴쉽, 연구장학생 등) 획득 및 정제화</p> <ul style="list-style-type: none"> - 텍스트 정보의 경우, 파이썬에서 사용 가능한 라이브러리 중 하나인 APScheduler 또는 Cron 작업을 이용하여 24H 간격으로 주기적인 크롤링을 진행함 <ul style="list-style-type: none"> • 웹사이트 접속: Requests 라이브러리를 사용하여 웹사이트에 접속해 HTML 소스를 가져옴 • HTML 소스 파싱: BeautifulSoup 라이브러리를 사용하여 HTML 소스를 분석하고 필요한 정보를 추출함 - 크롤링을 통해 얻은 텍스트 데이터를 정제하고 구조화하는 작업을 진행함 <ul style="list-style-type: none"> • HTML 태그 제거: BeautifulSoup 등을 사용, 텍스트에서 HTML 태그를 제거함 • 불필요한 문자 및 공백 제거: 정규 표현식 또는 문자열 관련 함수를 사용해 불필요한 문자와 공백을 제거함 • 구조화: 채용 정보의 명확한 구조화를 위해 JSON이나 CSV 같은 형태로 저장함 - 크롤링 상태 확인 및 저장 <ul style="list-style-type: none"> • 이전 상태 비교: 웹사이트의 채용 공고 목록에서 각 공고의 게시 날짜와 상태(상시, 공채, 인턴쉽, 연구장학생 등)를 확인하고, 실시간으로 변경사항이 있는지 체크함 • 업데이트: 새로 발견된 채용 공고나 업데이트된 정보를 데이터베이스에 저장함
--	---

	<ul style="list-style-type: none"> - 이미지 정보의 경우, Tesseract OCR을 사용해 이미지 파일에서 텍스트를 추출함 <ul style="list-style-type: none"> • 이미지 전처리: 이미지 전처리 과정에서 이미지의 명암 대비를 조절하고, 노이즈 제거, 회전, 텍스트가 있는 영역 표시 등의 작업을 진행해 문자 인식률을 높임 • OCR 작업: Tesseract OCR을 사용하여 이미지 파일에서 텍스트를 인식하고 디지털 텍스트로 변환함 • 텍스트 정제: 텍스트 인식 오류로 인한 오타자를 수정함 <p>3) 통합 텍스트 데이터 선별 작업</p> <ul style="list-style-type: none"> - 크롤링과 OCR을 통해 추출된 통합 텍스트 데이터 처리는 '채용 일시', '모집 분야', '우대사항' 등 유의미한 정보만을 선별함 <ul style="list-style-type: none"> • GPT4 API 기반 맞춤형 가이드: 통합 텍스트 데이터는 "Extract specific information such as hiring date, recruitment field, and number of people from text data"와 같은 추가적인 스크립트와 함께 GPT4 API를 통해 규격화된 양식에 맞게 요약되어 제공됨 <p>4) 학생 맞춤형 채용 정보 제공</p> <ul style="list-style-type: none"> - 선별된 통합 텍스트 데이터와 학생 데이터를 사전학습된 BERT에 입력값으로 넣어 나온 각각의 결과 간의 코사인 유사도를 측정함 - 코사인 유사도는 학생과 해당 기업간의 연관성(우선순위)을 내림차순으로 나타냄 <ul style="list-style-type: none"> • 가중우선순위 생성: 크롤링된 기업 평가 항목('커리어 향상', '업무와 삶의 균형', '급여 및 복지', '사내 문화', '경영진')을 표준화 시켜 만든 가중치를 우선순위에 곱해 평가 항목별 '가중우선순위'를 생성함 - '가중우선순위'를 통해 학생이 선택하는 기업 평가 항목에 맞게 기업 정렬이 가능함 <p>3. 서비스 상용화를 위한 웹 플랫폼 구현</p> <p>1) 웹 기반의 상용화된 서비스 제공을 위한 클라이언트 서버 시스템 구축</p> <ul style="list-style-type: none"> - 장고 프레임워크 혹은 소켓 라이브러리를 이용하여 파이썬 기반의 클라이언트 서버 시스템을 구축 - 클라이언트는 웹에서 사용자의 학번을 입력받아 서버에 전달함 - 서버는 클라이언트로부터 전달받은 학번을 이용하여 대학원생의 정보(학과, 등재 논문, 참여 연구 등)를 얻어냄 - 획득한 데이터들을 기업 및 채용 공고 추천 모델에 인풋으로 넣어 추천 결과를 도출함 - 웹클라이언트는 추천 결과를 전달 받아 웹 사이트의 정해진 위치에 시각화함 <p>2) 대학교 포털 사이트를 활용한 서비스 접근성 향상</p> <ul style="list-style-type: none"> - KUchive 사이트의 경우 AI선배(신입생 프로그램 추천) 및 포트폴리오 관리 서비스를 제공하고 있음 - 해당 서비스들과 유사하게 사이트에 로그인하는 것만으로 개발된 추천 서비스를 이용할 수 있게 함 <p>3) 이메일을 이용한 정보 알림 서비스 제공</p> <ul style="list-style-type: none"> - 고려대학교 대학원의 경우 이메일을 통한 추천 논문 알림 서비스를 제공하고 있으며 동시에 해당 서비스에 대한 설문조사를 진행함 - 이와 유사한 방식으로 이메일을 이용한 추천 기업 및 채용 공고에 대한 알림을 제공함 - 서비스를 이용한 대학원생들의 설문조사 결과를 모델의 성능 및 서비스의 질을 향상시키기 위한 추가적인 평가 데이터로 이용함
--	--

4. 실제 서비스 예시



그림 2 KUchive에 내재된 본 서비스 시연

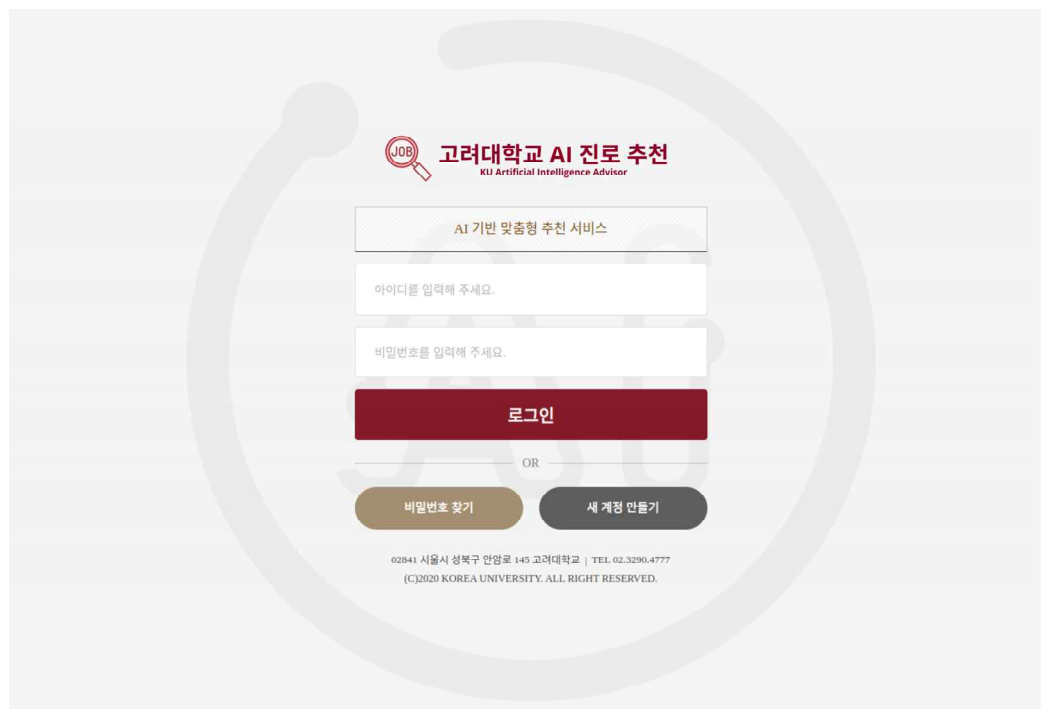


그림 3 본 서비스 로그인 페이지

정렬 기능을 통해 개인 선호에 따라 설정 가능

고려대학교 AI 진로 추천
KU Artificial Intelligence Advisor

반갑습니다! 000님.

관심 직무 키워드는
개발, 프론트엔드 (Front-End),
애플리케이션 (Application) 입니다.

맞춤 채용 공고 이메일 알림 받기

이메일 주소

그림 4 진로 추천 플랫폼 메인 페이지

고려대학교 AI 진로 추천
KU Artificial Intelligence Advisor

맞춤 기업 채용 공고 일정

Recruit Calendar

그림 5 리쿠르트 캘린더 부가 서비스 기능

5. 추진 일정

	추진 일정 (단위: 주)											
	1	2	3	4	5	6	7	8	9	10	11	12
데이터 전처리 및 정제	■	■	■									
학생 및 기업 관련 키워드 추출				■	■							
채용 공고 정보 크롤링 및 정제				■	■	■	■					
유사도 분석						■	■	■				
테스트 서버 구성 및 프로토타입 출시								■	■	■		
정식 서비스 개시 및 진행										■	■	■
연구 결과 보고서 작성											■	■

그림 6 연구 추진 일정표

파급효과

- 고려대학교 대학원생을 대상으로 한 취업/진로 지원 웹 플랫폼 서비스를 제공하여 본교의 취업 시장 경쟁력 제고 및 대학원생들의 재학 만족도 증진
 - 고려대학교 대학원생을 대상으로 하는 취업/진로 지원 웹 플랫폼 서비스
 - 데이터 구축부터 분석을 통한 서비스 제공까지 일괄적으로 진행하는 통합형 솔루션
 - 취업/진로에 대한 간단한 정보뿐 아니라 구체적인 채용 공고를 함께 제공
- 인공지능 기반의 데이터 정제 과정을 활용한 데이터 구축 자동화를 통해 기존 데이터 구축 방법의 문제점 해결
 - 웹 크롤링으로 수집된 비정형 데이터의 정제 및 목적에 부합한 정보를 포함한 대규모 데이터베이스 구축을 위한 GPT4 API 적용
 - GPT4 API를 기반으로 하여 높은 신뢰도와 안정성을 지닌 데이터 구축 및 분석 가능
- 대중화 및 확장 가능한 대규모 데이터셋 구성
 - 구축된 대규모 데이터를 향후 디지털정보처에서 기업 및 취업 진로와 관련된 분석에 활용될 수 있을 것으로 기대
 - 고려대학교 대학원생에 한정된 데이터 구축 방법이 아니기에 다양한 분야로의 확장 가능
- 대학원 취업/진로 추천 서비스 제공을 진행하면서 추가적으로 수집한 대학원생 데이터 및 설문조사 데이터를 활용하여 솔루션 모델의 성능을 향상
- 학생들의 관심 분야 및 연구 키워드를 파악하기 위해 학생 이메일과 구글 스칼라뿐만 아니라 지도교수 및 과제 데이터를 사용하는 GPT4 API 기반의 모델 사용하여, 결과물을 보다 정확하고 안정적으로 추출 가능