

# 2조

## Final Presentation

강의데이터를 활용한 자연어 기반  
진로 나침반



# 주제

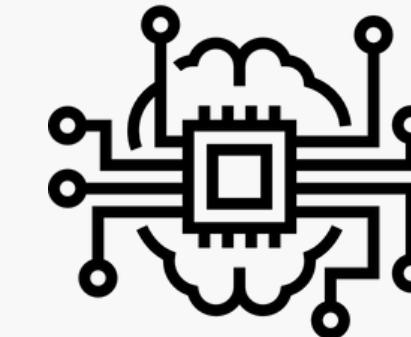
## 강의데이터를 활용한 진로 나침반

### 사용자 입력



“컴퓨터로 사람을 재밌게 해주고 싶어”

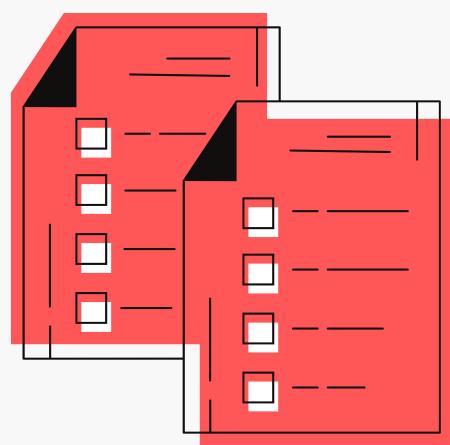
### 예상 답변



당신과 가장 가까운 직업은 게임 개발자이며  
멋진 게임 개발자가 되기 위해  
다음 강좌들을 수강하시는 것을 추천드립니다:)

대학수학, 선형대수학, 자료구조, 알고리즘  
, 컴퓨터그래픽스, 물리시뮬레이션  
유니티 실습, gpu 프로그래밍

# 필요성



[꿈꾸는 경기교육] 진로를 고민하는 모든 청소년에게

## 구체적 로드맵 제시

1 00:00

많은 학생들이 막연하게 하고 싶은 일은 있지만 어떤 공부를 어떻게 해야 하는지 구체적인 로드맵을 알기 어려워합니다.

10급시 내려놓고 다양한 분야 도전  
실패 하더라도 미래에 좋은 양분 될 것  
온라인 활용 진로적성검사 꼭 받아보길

## 진로 기반 과목

현재의 번거로운 설문기반 진로탐색은 직업군을 추천하기는 하나, “내가 하고 싶은 일이 어떤 직업인지”, “그 직업을 위해 어떤 과목을 수강해야 하는지” 정확히 알지 못해 진로 중심 탐색에는 한계가 있습니다. 지금까지는 이를 도울 수 있는 자동화 시스템이 거의 없습니다.

“자연어 기반 진로 목표를 입력하면  
필요한 학문 체계 및 수강과목을 안내해주는 시스템이 필요”

# 수집 데이터



+



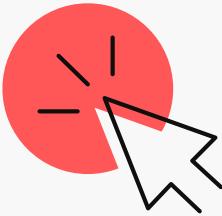
대학생활

- 수강관리
- 수업시간표
- 출석관리(KW출첵)
- 강의계획서 조회
- 수강신청 프로그램
- KLAS 앱 다운로드
- 수강포기 신청
- 교류 학점 신청
- CDP 출석내역
- CDP 출석여부 설문조사



KLAS : 2024년 1&2학기 강의계획서 PDF 학정번호 OCR

# 수집 데이터



고용24 진로 가이드

사용자 입력 → 직업 매칭을 위한 데이터

홈페이지 게시 PDF → 텍스트 추출

사용자에게서 받은 자연어에 SBERT 적용 후,  
고용24 진로, 전공가이드에서 Top-5 추출

## 직업정보 찾기

연구직 및 공학 기술직 > 소프트웨어 개발자 > 게임프로그래머

게임프로그래머

요약하기      하는 일      교육/자격/훈련      임금/만족도/전망      능력/지식/환경      성격/흥미/가치관      옵션

### ▪ 직무개요

게임프로그램의 구조를 설계하고 게임프로그래밍을 수행한다.

### ▪ 수행직무

- 게임기획자, 게임그래픽디자이너, 게임음악가 등으로부터 넘겨받을 자료를 기반으로 게임프로그램의 구조를 설계한다.
- 게임을 제작하기 위한 툴인 게임엔진을 개발한다.
- 영상을 컴퓨터 모니터에 출력하는데 필요한 제반 함수들의 집단인 그래픽 라이브러리를 제작한다.
- 그래픽 특수효과를 제작한다.
- 키보드, 마우스 등의 입력장치 제어루틴을 제작한다.
- 그래픽 파일이나 사운드 파일이 정상적으로 게임 속에서 작동될 수 있도록 해 주는 프로그램을 작성한다.
- 게임을 테스트하여 에러를 수정하고 버그를 찾아낸다.

# 수집 데이터



KMOOC

전공가이드에서 추출한 Top-5 직업에 대한  
직업 설명 → 강좌 매칭을 위한 데이터

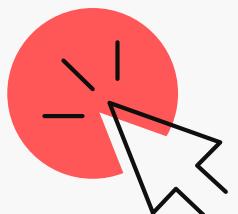
KLAS : Selenium 크롤링 → 텍스트 추출

Kmooc : 공공데이터포털 API 이용 → CSV 저장  
(교과목 개요 & 학습목표 및 학습방법)

직업 설명 2-gram 적용 후 SBERT 적용,  
임베딩된 강의계획서, K-MOOC 강좌정보와  
유사도 매칭, Top-20 추출 후 10개로 압축

교과목명	텍스트마이닝-Text Mining		
학정번호	I040-3-4139-01		
장애학생 지원	장애인 수강 시, 학습 지원(장애인지원실에 등록된 학생의 요청 시 지정 족석 제공 등)		
외국어 강의	<input checked="" type="radio"/> 영어강의	<input type="radio"/> 중국어 강의	<input type="radio"/> 일본어 강의
외국어 비율	50 %		
교과목 개요	텍스트 마이닝은 자연어 처리라는 이름 하에 기업과 사회에서도 점점 더 중요성을 인식하기 시작하였으며, 일반인들도 이에 대한 관심을 갖고 관련 기술을 습득하고 있음. 이와 관련하여, 본 교과목은 텍스트 마이닝과 관련된 수학, 언어학 등의 기본 지식과 파이프라인 방식의 자연어 처리를 위한 핵심 원리, 자연어 처리 기술 기반 응용 시스템, 딥러닝 기반의 자연어 처리 기술 등의 다양한 주제를 포괄적으로 다룰 것임. 이와 더불어, 토픽모델링, 감성 분석 등의 최신 주요 이슈도 포함할 계획임.		
교과목 대표역량	전문지식능력		
학습목표 및 학습방법	- 텍스트 마이닝 및 자연어 처리를 위한 수학, 언어학 등의 기초 지식에 대한 이해 - 자연어 처리 기술 기반 응용 시스템 학습 및 구현 능력 함양		

# 수집 데이터



KMOOC

전공가이드에서 추출한 Top-5 직업에 대한  
직업 설명 → 강좌 매칭을 위한 데이터

KLAS : Selenium 크롤링 → 텍스트 추출

Kmooc : 공공데이터포털 API 이용 → CSV 저장  
(교과목 개요 & 학습목표 및 학습방법)

직업 설명 2-gram 적용 후 SBERT 적용,  
임베딩된 강의계획서, K-MOOC 강좌정보와  
유사도 매칭, Top-20 추출 후 10개로 압축

## 강좌소개

### ■ 수업 개요

이 수업은 텍스트 데이터를 처리하고 활용하기 위한 R 프로그래밍 언어와 활용 기술을 소개한다. 특히, 기존의 통계 지식이나 프로그래밍 경험이 전무 한 인문계열의 학생들도 쉽고 친숙한 방식으로 텍스트 마이닝에 대해 알아가고 활용해보는 기회를 제공한다.

### ■ 수업 대상 및 목표

#### 수업대상

- 디지털 텍스트 분석에 관심은 있지만 코딩 경험이 없는 인문사회계열 학생

#### 수업목표

- 텍스트 마이닝의 기본 개념과 활용을 소개할 수 있다.  
- 기초적인 수준에서 텍스트를 분석할 수 있는 방법을 설명할 수 있다.  
- 개별 프로젝트를 수행할 수 있다.

#### 교과목 개요

텍스트 마이닝은 자연어 처리라는 이름 하에 기업과 사회에서도 점점 더 중요성을 인식  
음. 이와 관련하여, 본 교과목은 텍스트 마이닝과 관련된 수학, 언어학 등의 기본 지식과  
시스템, 딥러닝 기반의 자연어 처리 기술 등의 다양한 주제를 포괄적으로 다룰 것임. 이

#### 교과목 대표역량

전문지식능력

#### 학습목표 및 학습방법

- 텍스트 마이닝 및 자연어 처리를 위한 수학, 언어학 등의 기초 지식에 대한 이해  
- 자연어 처리 기술 기반 응용 시스템 학습 및 구현 능력 함양

## JSON 국가평생교육진흥원\_K-MOOC\_강좌정보API

한국형 온라인 공개강좌(K-MOOC)의 강좌 목록 API 및 강좌 세부 정보 API



4



0



관심

# 직업 설명 데이터

## 대학교수

요약하기	하는 일	교육/자격/훈련	임금/만족도/전망	능력/지식/환경	성격/흥미/가치관	업무활동	전직가능직업
------	------	----------	-----------	----------	-----------	------	--------

하는 일	대학에서 교육, 예술 및 인문학, 사회과학·경영 및 법학, 자연과학·수학 및 통계학, 정보통신기술, 공학·제조 및 건설, 농림수산업 및 수의학, 보건 및 복지 등 분야별 전공과목의 강의를 준비하고, 자신의 전공분야에 대한 연구 및 논문을 발표한다.		
	관련학과	관련자격	훈련정보
교육/자격/훈련	정확히 일치되는 학과정보가 없습니다.	자료가 존재하지 않습니다.	대학 교수

## 직업 통상 업무 설명

# 직업 설명 데이터

## ▪ 직무개요

대학에서 교육, 예술 및 인문학, 사회과학·경영 및 법학, 자연과학·수학 및 통계학, 정보통신기술, 공학·제조 및 건설, 농림수산업 및 수의학, 보건 및 복지 등 분야별 전공과목의 강의를 준비하고, 자신의 전공분야에 대한 연구 및 논문을 발표한다.

## 직업 통상 업무 설명

## ▪ 수행직무

- 과정표와 교과과정을 계획·조정한다.
- 강의를 준비하고, 학생들에게 강의를 하며 세미나 및 실험을 수행한다.
- 시험을 출제하고 그 결과를 평가한다.
- 졸업생의 논문 및 연구계획을 지도하고 조언한다.
- 자신의 전공분야에 대한 연구를 수행한다.

# 직업 설명 데이터

## 필요 기술 및 지식

대학교수가 되기 위해서는 전공과목의 박사학위를 보유하는 것이 기본적으로 요구된다. 일부 전문대학의 문화·예술 관련 분야의 경우 박사학위가 없어도 특별한 업적이나 해당 분야의 오랜 실무경험을 갖춘 사람을 채용하는 경우도 있다. 대학교수가 되기 위해서는 무엇보다 전공 분야에 수준 높은 연구 실적이 중요하며, 이를 위해 국내외 학술지를 통해 연구 실적을 발표하고, 관련 학회 활동에 활발히 참여하는 것이 필요하다.

## 직업 요구 능력 설명

# 직업 설명 데이터 임베딩

## 하는 일

대학에서 교육, 예술 및 인문학, 사회과학·경영 및 법학, 자연과학·수학 및 통계학, 정보통신기술, 공학·제조 및 건설, 농림수산업 및 수의학, 보건 및 복지 등 분야별 전공과목의 강의를 준비하고, 자신의 전공분야에 대한 연구 및 논문을 발표한다.

## 직무개요

대학에서 교육, 예술 및 인문학, 사회과학·경영 및 법학, 자연과학·수학 및 통계학, 정보통신기술, 공학·제조 및 건설, 농림수산업 및 수의학, 보건 및 복지 등 분야별 전공과목의 강의를 준비하고, 자신의 전공분야에 대한 연구 및 논문을 발표한다.

## 수행직무

- 과정표와 교과과정을 계획·조정한다.
- 강의를 준비하고, 학생들에게 강의를 하며 세미나 및 실험을 수행한다.
- 시험을 출제하고 그 결과를 평가한다.
- 졸업생의 논문 및 연구계획을 지도하고 조언한다.
- 자신의 전공분야에 대한 연구를 수행한다.

## 필요 기술 및 지식

대학교수가 되기 위해서는 전공과목의 박사학위를 보유하는 것이 기본적으로 요구된다. 일부 전문대학의 문화·예술 관련 분야의 경우 박사학위가 없어도 특별한 업적이나 해당 분야의 오랜 실무경험을 갖춘 사람을 채용하는 경우도 있다. 대학교수가 되기 위해서는 무엇보다 전공 분야에 수준 높은 연구 실력이 중요하며, 이를 위해 국내외 학술지를 통해 연구 실력을 발표하고, 관련 학회 활동에 활발히 참여하는 것이 필요하다.

## Sbert 기반 임베딩

전처리 하여 합치고,  
직업 설명 데이터로 사용

# 전처리

- 특수 문자 제거
- url 문자열 삭제
- 동어반복 삭제(예 : ‘인공지능(AI)’ → ‘인공지능’)
- 불용어 삭제(예 : ‘이 강좌에서는’, ‘상기 url을 참고하시기 바랍니다’ 등)

- Analyzing Design & Judging Good Design:

- 디자인 분석을 통해 좋은 디자인을 분별할 수 있는 능력 강화

- Advanced Visual Communication Design & Computer Graphic Design:

- 시각디자인 이론·방법론과 컴퓨터그래픽디자인 실습의 심화학습 과정

- UX·UI·APP Design Trend & Vision:

- UX·UI·APP 디자인 트렌드 조사·분석 및 Vision 도출

- UX·UI·APP Design Theory & Methodology:

- UX·UI·APP 디자인 이론과 방법론을 학습하고 디자인 프로젝트에 적용하는 병행학습

- Graphic·APP Design Technique & Tool:

- 그래픽 디자인과 APP 디자인의 테크닉과 관련 소프트웨어 활용능력 개발

- Software:

- Adobe Illustrator, Photoshop, APP Prototyping Tool(Figma), Microsoft PowerPoint



Analyzing Design Judging Good Design 디자인 분석을 통해 좋은 디자인을 분별할 수 있는 능력 강화 Advanced Visual Communication Design Computer Graphic Design 시각디자인 이론 방법론과 컴퓨터그래픽디자인 실습의 심화학습 과정 UX UI APP Design Trend Vision UX UI APP 디자인 트렌드 조사 분석 및 Vision 도출 UX UI APP Design Theory Methodology UX UI APP 디자인 이론과 방법론을 학습하고 디자인 프로젝트에 적용하는 병행학습 Graphic APP Design Technique Tool 그래픽 디자인과 APP 디자인의 테크닉과 관련 소프트웨어 활용능력 개발 Software Adobe Illustrator, Photoshop, APP Prototyping Tool Figma , Microsoft PowerPoint

# 전처리(빈출 중복 단어 제거 + 불용어 제거)

**본 강의는** K콘텐츠의 과거와 현재 동향 및 미래 전망에 대한 이해를 제공하고, 학생들이 한국 콘텐츠의 다양한 측면을 분석하고 평가할 수 있는 능력을 키우는 것을 **목표로 합니다.**

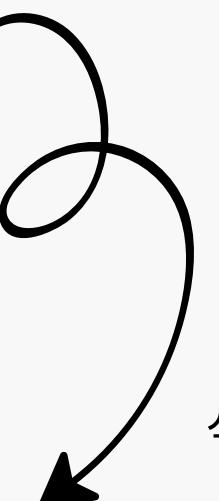
**이 과목은** 학생들의 영어 듣기 능력을 중급 이상으로 향상 시키는 것을 **목표로 하고 있습니다.**

**본 교과목의 수강생은** 건설관리 분야에서 다루는 여러 주제들에 대한 기초 지식과 각 지식영역 간의 관련성을 익힘으로써, 향후 건설산업에서 관리자로서의 역할을 할 수 있는 기본 소양을 배양하는 것이 **본 강의의 목적입니다.**

이해:	1136
수:	895
대한:	813
학습:	752
능력:	564
강의:	487
이론:	475
듣:	452
수업:	439
통해:	428
것:	419
기초:	414
기본:	374
분석:	361
방법:	350
함:	341
학생:	340
기술:	326
관련:	316
지식:	312
과목:	309
사회:	309
개념:	306
위:	302
설계:	298
내용:	280
과정:	274
활용:	264
습득:	261
문화:	256
분야:	240

**학습 목표** 의료 데이터를 탐색하여 분석 과제를 발굴할 수 있다. 의료 데이터를 직접 수집 정리하고 결과를 예측할 수 있다. 의료 데이터 기반의 기획서를 작성할 수 있다. **수강신청 및 학습방법 1. 수강신청 방법 안내** K MOOC <https://new.kmooc.kr> 회원가입 회원가입시 아이디는 kw 학번으로 설정 예시 kw2022091800 사이트 상단 매치업 탭을 통해 교육기관 구분에서 광운대학교 교육 바로가기 클릭 심화로 구성되어 있는 의료빅데이터 분석 및 활용 실무 수강신청 클릭 2. 강의 수강 방법 안내 K MOOC사이트 <https://new.kmooc.kr> 로그인 사이트 우측 상단 본인 아이디 내 강의실 클릭 수강등록 강좌 선택 강의계획서에 따라 온라인 주차별 강좌 수강 및 학습평가 기말고사, 프로젝트 등 참여 강의계획서에 따라 오프라인 기말시험 등 참여 수업운영 방식 본 강좌는 하이플렉스러닝 방법으로 운영되는 강좌입니다. 하이플렉스러닝은 대면 비대면으로 오프라인 활동을 지원하며, 반드시 오프라인 강의실에 참여하지 않아도 출석 및 활동이 가능합니다.

학습목표란에 학습목표와 관련없는 내용이 기재되어있는 경우  
→ 빈출 중복 단어 제거로 일정 수준의 전처리는 가능했지만 완벽한 배제는 어렵다는 한계점

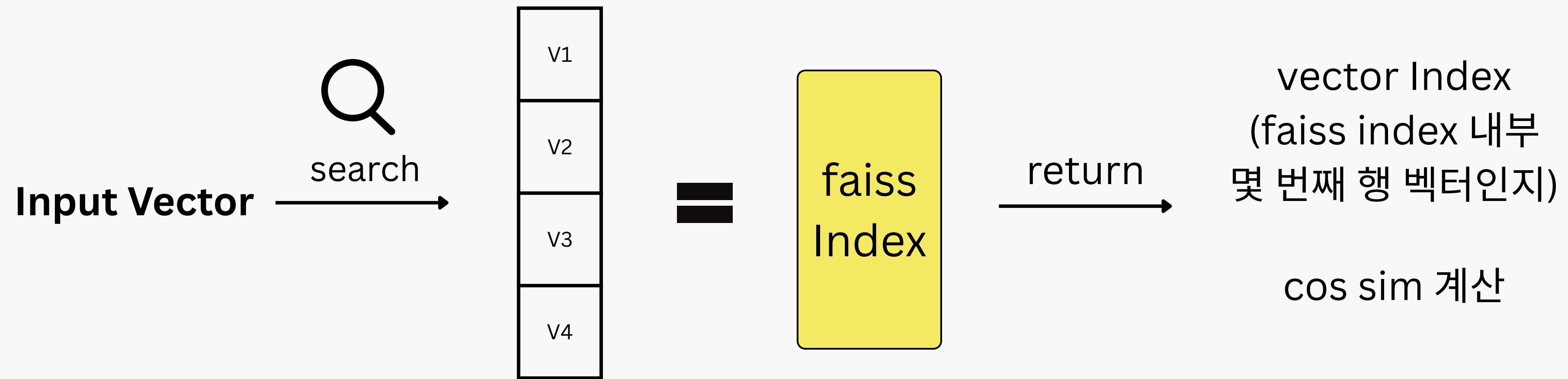


삭제할 단어를 직접 선정

# 의미 약한 단어

```
words_to_delete = ['대면', '비대면', '퀴즈', '학습', '주차', '강좌', '내용', '강의', '소개', '수업', '계획', '운영', '목표', '본', '이해', '교과목']
```

# FAISS 라이브러리



# 문서임베딩

Pre-trained SBERT 사용

“paraphrase-multilingual-mpnet-base-v2”

SBERT → 트랜스포머 구조  
∴ 입력 토큰 개수가 고정

다루고자 하는 문서(직업, 강좌 설명)는 매우 깊  
→ 분할 입력

직업 이름(W1)	직업에 대한 설명 문서 =(문장1, 문장2, ..., 문장 n)
-----------	--

강좌 이름(L1)	강좌에 대한 설명 문서 =(문장1, 문장2, ..., 문장 n)
-----------	--

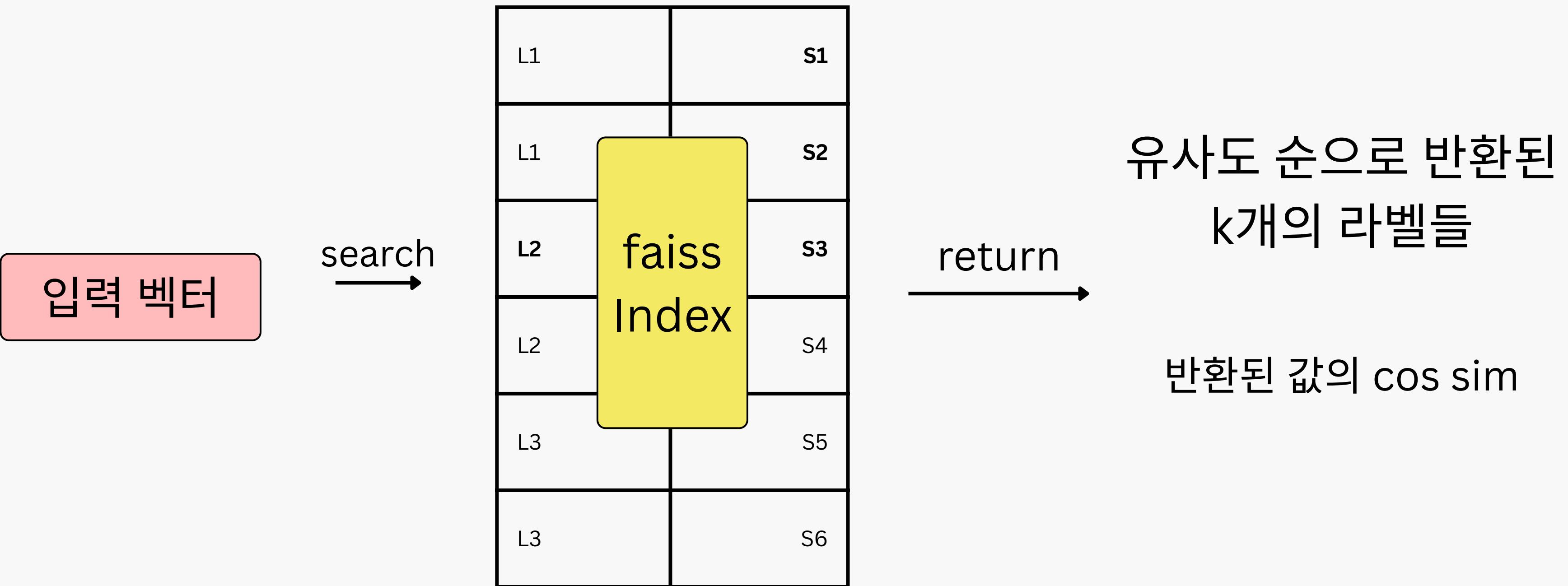
W1	S1
W1	S2
W1	S3
W1	...



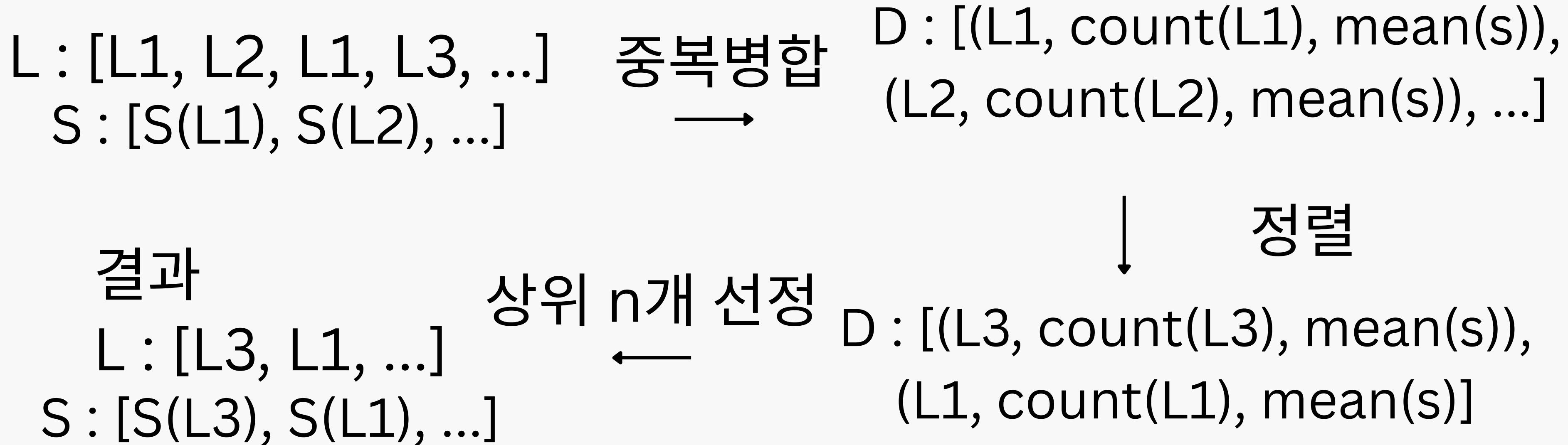
L1	S1
L1	S2
L1	S3
L1	...



# FAISS를 이용한 선정 과정



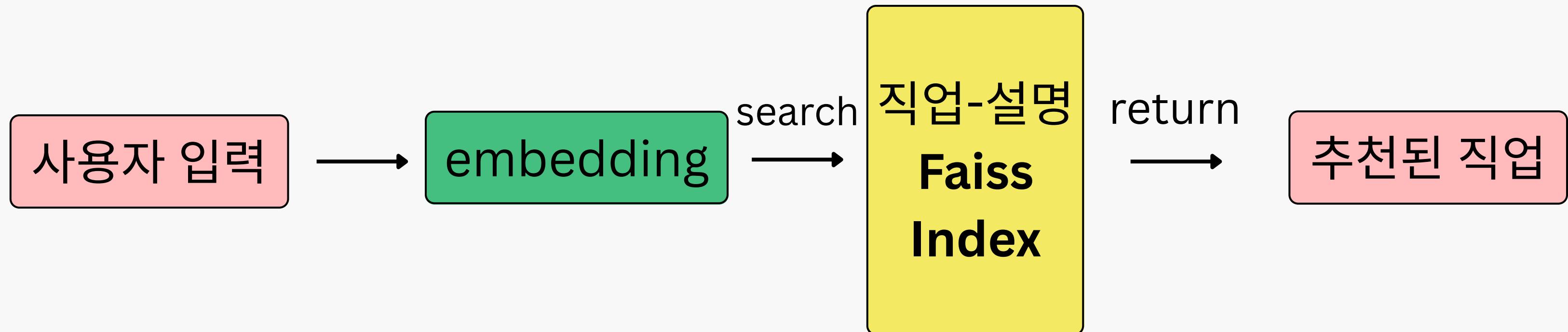
# FAISS를 이용한 선정 과정



# 설명기반 선정



# 설명기반 선정



# 지표기반 직업 추천

## 데이터분석가(빅데이터분석가)

### • 업무수행능력 중요도

#### 직업 내 비교

500여개 직업 종사자들의 자신의 직업에 대해 평가한 업무수행능력 관련 항목별 중요도를 직업 내에 비교하여 본 직업에서 중요성(평가점수)이 높게 나타난 항목을 순서대로 제시함.

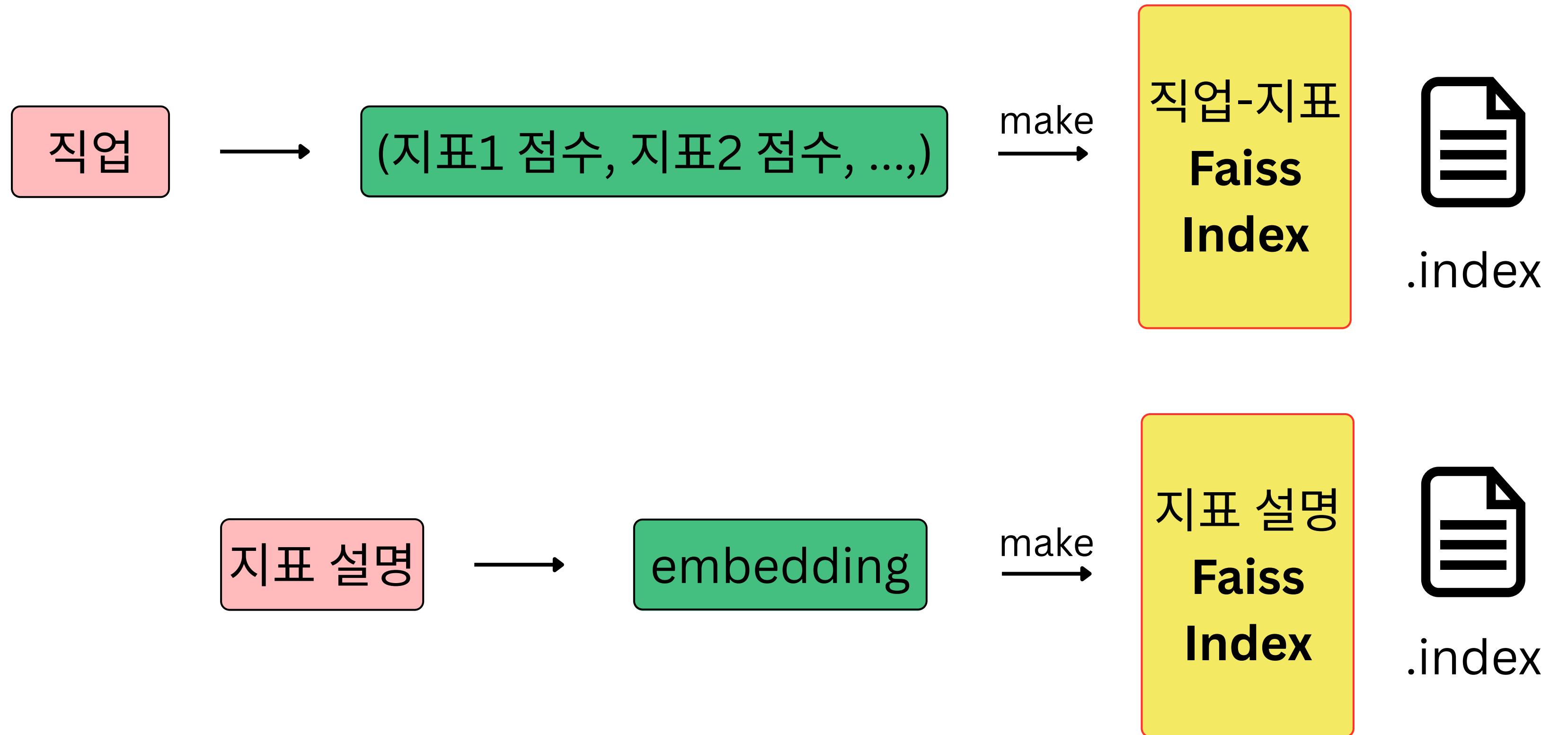
### • 성격 중요도

#### 직업 내 비교

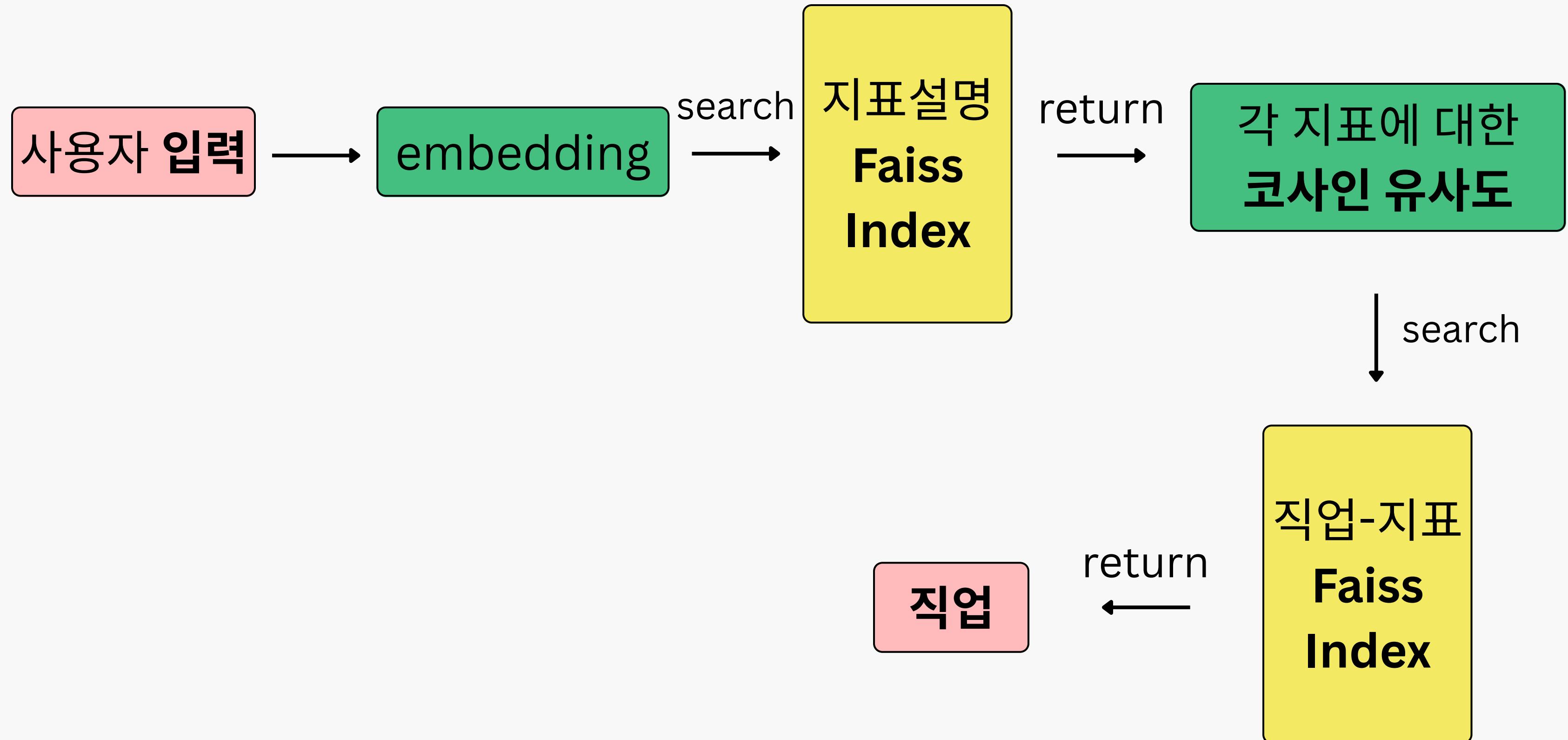
500여개 직업 종사자들의 자신의 직업에 대해 평가한 성격 관련 항목별 중요도를 직업 내에 비교하여 본 직업에서 중요성(평가점수)이 높게 나타난 항목을 순서대로 제시함

평균 (5점 만점)	항목	설명	평균 (5점 만점)	항목	설명
4.0	논리적 분석	문제를 해결하기 위해(혹은 의사결정을 하기 위해) 체계적으로 이치에 맞는 생각을 해낸다	4.5	꼼꼼함	사소한 부분까지도 주의 깊고 업무를 철저히 완수한다
3.9	수리력	어떤 문제를 해결하기 위해 수학을 사용한다	4.4	분석적 사고	문제에 대한 답을 구하기 위해 정보를 분석하거나 논리를 사용한다
3.9	읽고 이해하기	업무와 관련된 문서를 읽고 이해한다	4.3	책임과 진취성	책임을 기꺼이 받아들이고 도전하려 한다
3.8	전산	다양한 목적을 위해 소프트웨어나 인터넷을 활용하거나 프로그램을 작성한다	4.2	신뢰성	믿을 수 있고, 자신이 맡은 책무를 완수한다
3.8	듣고 이해하기	다른 사람들이 말하는 것을 집중해서 듣고 상대방이 말하려는 요점을 이해하거나 적절한 질문을 한다	4.1	성취/노력	도전적인 목표를 설정한 후에 이를 달성하기 위해 노력한다

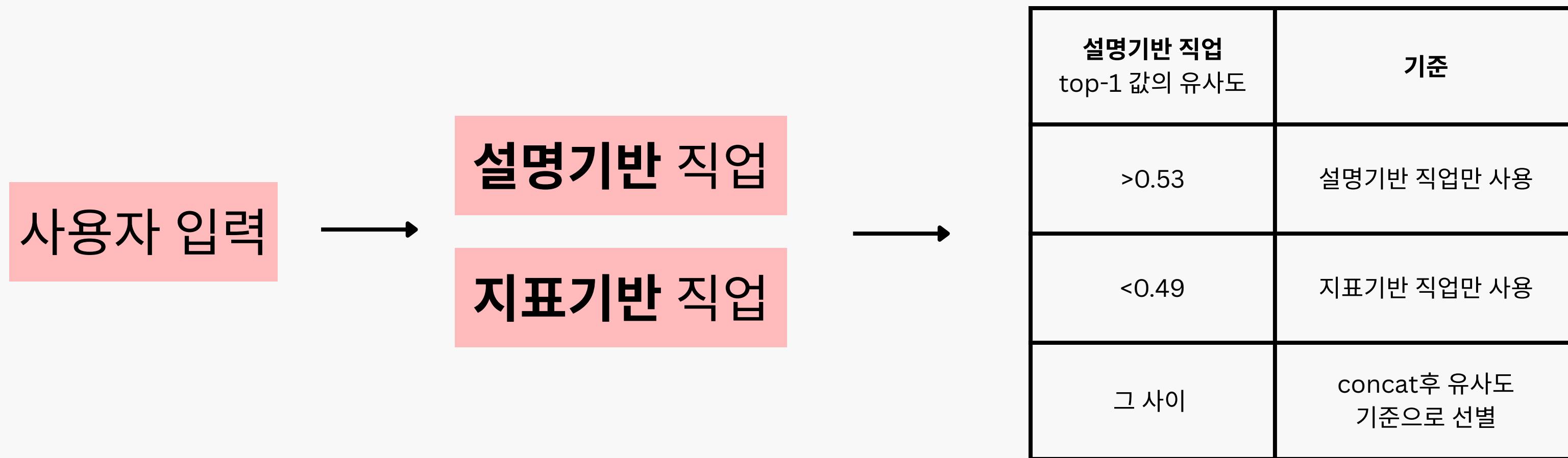
# 지표기반 선정



# 지표기반 선정



# 직업 선정 기준



# 직업 기반 강좌 추천

추천받은 직업명



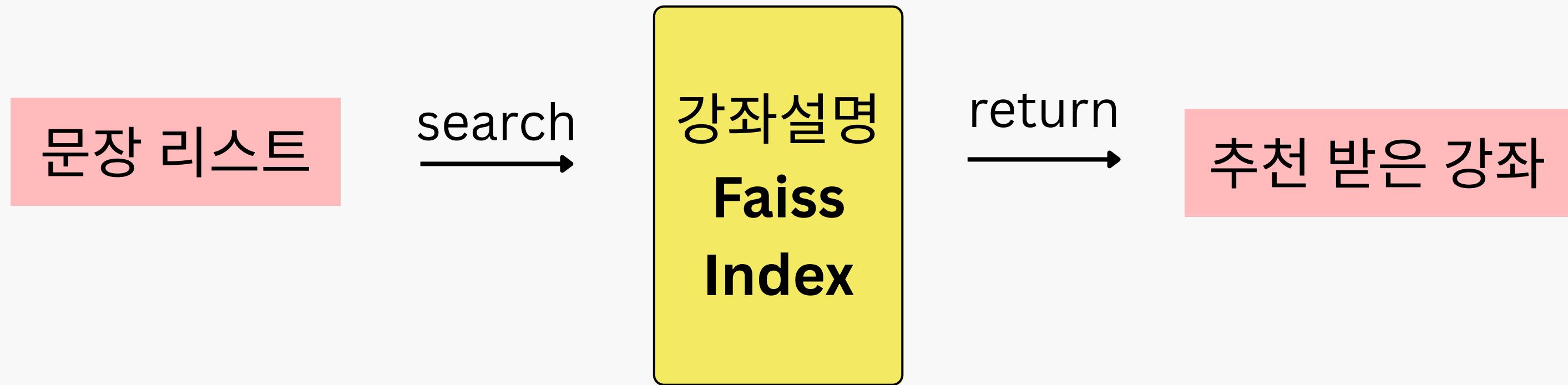
직업_이름	문장
3D프린팅모델러	3D프린터를 조작 운영하여 고 3D프린팅은 다른 분야와 3D프 최근 경향은 미술 관련 조형물 따라서 제품 제작에 있어서 제 그래픽디자인 3D모델링 등의 제품의 형상을 이미지로 디자인 작업이 제대로 진행되고 있는 사상 작업 도색 및 코팅 등 후 출력물의 조건에 맞는 최적의 설계된 디자인대로 프린터가
IT기술지원전문가	IT기술지원전문가는 시스템 관리자 IT기술지원전문가가 되기 위해 이들은 높은 수준의 이론적 지식 또한 IT기술지원전문가가 되기 그리고 수학이나 물리 화학 등 또 복잡한 문제를 해결하기 위 설치지시서에 따라 컴퓨터에 새로운 시스템을 구성하거나 매뉴얼을 읽고 사용자와 협의하 사용자가 상품을 사용하는 데 소프트웨어나 하드웨어를 평가 일상적인 데이터 전송기록과 하드웨어나 소프트웨어에 관한 문제해결을 수행하는 작업자 사용자에게 전달할 서류를 준비
	소프트웨어 등 IT제품의 출시 IT 관련 소프트웨어 및 프로그램 IT 개발자로 일한 경험이 있으 품질관리와 관련해서는 프로젝 트 관리에 대한 대수는 IT 분야

추천받은 직업에 대한 설명  
(긴 문장)

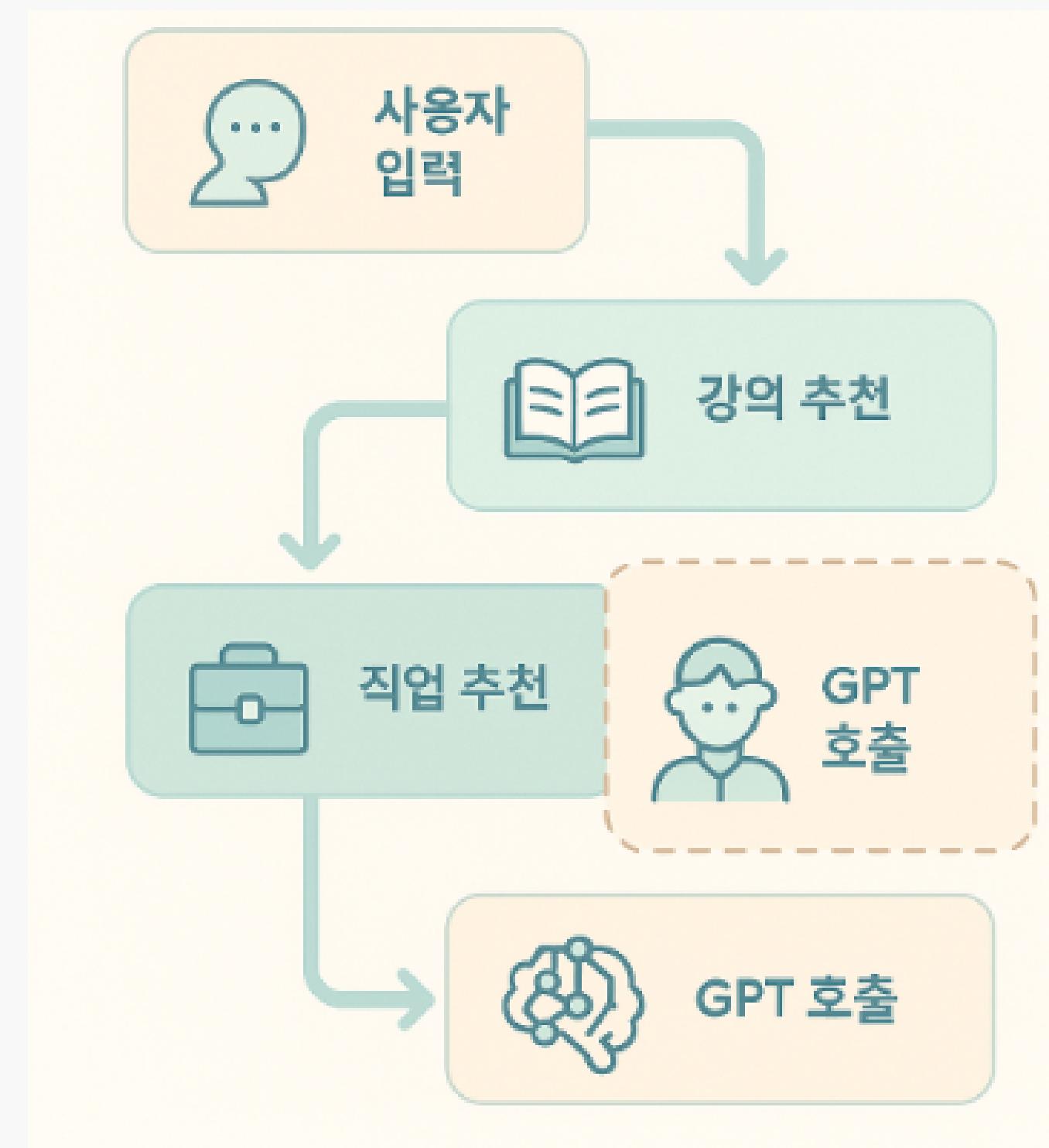


두 문장 단위로 분할  
(2-gram)

# 직업 기반 강좌 추천



# GPT 프롬프트로의 통합



입력받은 사용자의 텍스트

+

FAISS를 이용한 강의 추천 결과 가져오기

+

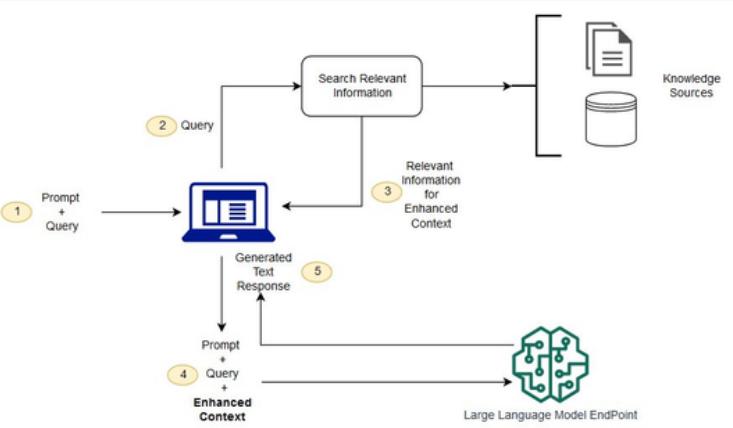
FAISS를 이용한 직업 추천 결과 가져오기

↓

하나의 GPT 프롬프트로 통합

# GPT 프롬프트로의 통합

## 프롬프트 설정

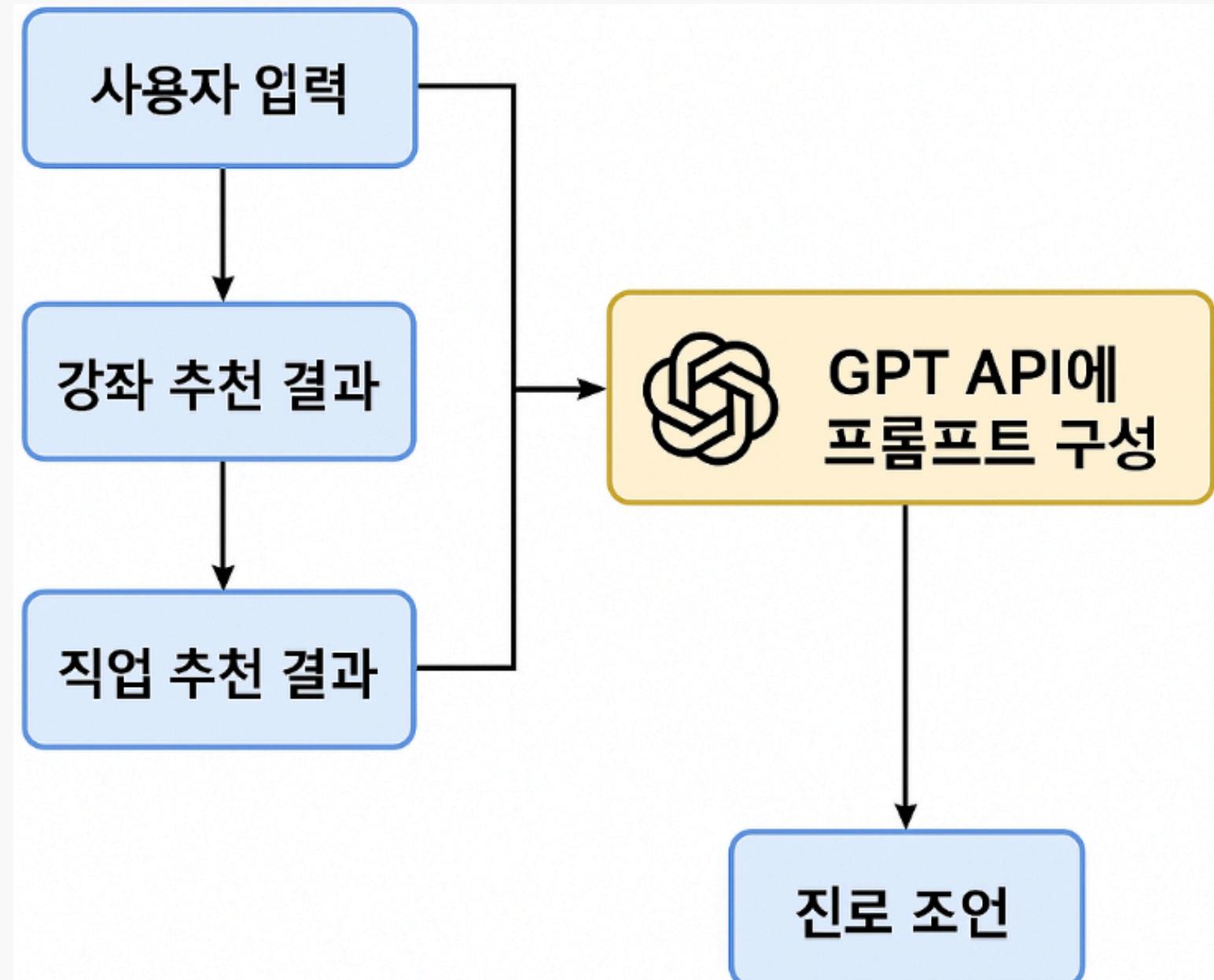


```
prompt = (
    '[사용자 입력]\n'
    f'{user_input}\n'
    '[사용자 입력으로 선정한 직업 목록]\n'
    f'{work_name_out}\n'
    '[사용자 입력과 직업의 코사인 유사도]\n'
    f'{work_sim}\n'
    '[직업을 가지기 위해 도움되는 강의 목록]\n'
    f'{lecture_name_out}\n'
    '[직업과 강의의 코사인 유사도]\n'
    f'{lecture_sim}\n'
    '위 내용은 사용자 입력과 사용자 입력으로 직업, 강의 DB에 RAG를 한 결과\n'
    '위 결과를 참조해 사용자의 진로에 도움이 될 만한 요약 및 조언을 제공하라.\n'
    '친절하고 구체적인 언어로 답변하라.\n'
```

## 모델 설정

```
model='gpt-4o-mini',
messages=[
    {"role": "system", "content": "당신은 진로 추천 전문가입니다."},
    {"role": "user", "content": prompt}
]
```

# GPT API를 활용한 진로 추천 통합



“사용자 입력 + 강좌 추천 결과 + 직업 추천 결과”를 조합  
-> 하나의 GPT 프롬프트로 만들고  
-> GPT API로 호출해서  
→ 최종 진로 추천 응답을 생성

# 시연 예

언젠가는 의료 기술과 데이터를 융합해서 사람들의 건강을 예측하고 돋는 시스템을 만들어보고 싶어요. 막연하지만 그런 기술이 세상을 바꿀 수 있다고 믿고 있어요.

## 출력

안녕하세요! 당신의 목표인 의료 기술과 데이터의 융합을 통해 사람들의 건강을 예측하고 돋는 시스템을 만들고 싶다는 꿈은 정말 멋진 비전입니다. 기술이 발전함에 따라 데이터 분석과 학습 모델은 더욱 혁신적이고 정교해질 것입니다.

## ### 직업 선택

당신이 선택한 직업 목록에서 가장 유망한 직업은 다음과 같습니다:

- \*\*보건의료정보관리사\*\*: 의료 데이터를 체계적으로 관리하고 분석하여 의사 결정 과정을 지원하는 역할입니다. 데이터 관리와 의료 시스템의 이해가 필수적입니다.
- \*\*데이터분석가(빅데이터분석가)\*\*: 대량의 데이터를 수집, 분석하여 유의미한 통찰력을 제공하는 데 중점을 둡니다. 이는 건강 예측 시스템에 중요한 역할을 할 것입니다.
- \*\*재활공학기사\*\*: 의료 기술과 데이터를 이용하여 환자의 재활과 관련된 기기를 설계하고 개발합니다. 이는 기술이 사람들의 삶의 질을 향상시키는 데 기여할 수 있습니다.

이 외에도 방사선사와 통신망운영기술자 같은 직업도 유용할 수 있지만, 상기 직업들이 당신의 목표에 더 근접한 것 같습니다.

## ### 강의 추천

해당 직업에 대비하기 위해 필요한 기술과 지식을 극대화할 수 있는 강의 목록은 다음과 같습니다:

- \*\*빅데이터의 세계 원리와 응용\*\*
- \*\*파이썬과 데이터마이닝\*\*
- \*\*비즈니스 애널리틱스의 기초\*\*
- \*\*SQL 정형 데이터 분석\*\*
- \*\*인공지능과 미래기술 AIFuture Technology\*\*

이 강의들은 데이터 분석, 머신러닝 및 인공지능 관련 기본 지식과 기술을 배울 수 있는 좋은 기회를 제공합니다. 특히 파이썬은 데이터 분석과 머신러닝에서 매우 유용한 프로그래밍 언어입니다.

# 시연 예

저는 애니메이션 속 장면 하나하나가 어떻게 만들어지는지 너무 궁금하고,  
언젠가는 그런 장면을 직접 구성하고 움직이는 일을 해보고 싶어요.

당신이 애니메이션 속 장면을 직접 구성하고 움직이는 일을 하고 싶다는 열망을 가지고 있다는 것은 정말 멋진 일입니다! 관련된 직업으로는 만화영화작가(애니메이터)가

## ### 추천 직업: 만화영화작가(애니메이터)

애니메이터는 캐릭터, 배경, 그리고 전체 스토리를 삶을 불어넣는 역할을 합니다. 당신이 가지고 있는 호기심과 창의력이 이 직업에 잘 맞습니다. 애니메이션을 통해 이야-

## ### 관련 직업

다음으로는 게임그래픽디자이너와 가상현실전문가입니다. 게임그래픽디자이너는 비슷한 시각적 기술을 필요로 하며, 가상현실전문가는 새로운 경험을 설계하고 구현하는 데

## ### 추천 강의

당신의 진로에 도움이 될 만한 강의는 다음과 같습니다:

- \*\*가상현실과 미디어 기술\*\*: 애니메이션의 기술적 기초를 다질 수 있습니다.
- \*\*디지털 이미징설계\*\*: 애니메이션의 시각적 요소를 표현하는 데 필수적인 스킬을 배울 수 있습니다.
- \*\*생성형 AI 활용 실무\*\*: 현재 애니메이션과 게임 디자인에 AI 기술을 접목시키는 방법을 익히는 데 유용할 것입니다.

이들 강의를 수강함으로써, 애니메이션과 관련된 기술과 이론을 보다 깊이 이해할 수 있으며, 직업 세계에 진입하는 데 필요한 경쟁력을 확보할 수 있습니다.

## ### 최종 조언

애니메이션 분야는 끊임없는 학습과 실험이 중요한 작업입니다. 따라서 다양한 프로젝트에 참여하거나 인턴십을 통해 실무 경험을 쌓고, 당신만의 포트폴리오를 만드세요.

# 시연 예

머릿속에 막연히 떠오른 생각이나 이야기를 어떻게든 표현해서 누군가에게 전달할 수 있을 때, 비로소 나를 이해받는 느낌이 들어서 그게 참 좋더라고요.

## 출력

당신의 생각과 이야기를 다른 사람에게 전달함으로써 이해받는 느낌을 중요하게 생각하신다는 것은 의사소통과 창의적인 표현에 대한 강한 흥미와 열정을 가지고

## ### 직업 목록

- \*\*예술치료사\*\*: 사람들의 정서적 어려움을 치료하기 위해 예술을 활용하는 직업입니다. 당신의 표현과 이해받고자 하는 욕구가 더욱 의미 있게 실현될 수 있습니다.
- \*\*출판물기획자\*\*: 글과 콘텐츠를 기획하고 출판하는 작업으로, 창의력과 커뮤니케이션 능력이 필수적입니다. 독자에게 전달할 메시지를 신중하게 구성하는 능력이 필요합니다.
- \*\*조사전문가\*\*: 데이터와 정보를 분석하여 통찰을 제공하는 역할입니다. 자신의 생각을 명확하게 정리하고 표현하는 능력이 큰 도움이 될 것입니다.
- \*\*작곡가\*\*: 음악을 통해 감정을 전달하고 이야기할 수 있는 기회입니다. 당신의 내면을 음악적 언어로 표현하는 매력적인 직업이 될 수 있습니다.
- \*\*문리학원강사\*\*: 다양한 주제를 가르치며 학습자들과의 소통을 통해 만족감을 느낄 수 있는 직업입니다. 교육을 통해 지식을 나누는 것 또한 큰 보람을 제공합니다.

## ### 도움이 될 강의 목록

주목할 만한 강의는 다음과 같습니다:

- \*\*성인학습자를 위한 대학 글쓰기\*\*: 글쓰기 능력을 향상시키는 데 큰 도움이 될 것입니다.
- \*\*교육방법 및 교육공학\*\*: 교육적 접근 방식을 배우고, 효과적으로 지식을 전달하는 데 필요한 기술을 익힐 수 있습니다.
- \*\*내일 내 일은 산업강사\*\*: 교육자로서의 경로를 탐색할 수 있는 강의로, 진로 설정에 큰 도움이 될 수 있습니다.

## ### 조언

이제 다양한 직업과 관련된 강의들을 고려하면서, 자신이 어떤 방식으로 사람들과 소통하고 싶은지를 깊이 생각해 보세요. 각 직업이 요구하는 자질과 당신의 관심사, 책 읽기나 글쓰기 연습, 음악적 기술을 계속 발전시키는 것도 좋습니다. 상상력을 자극하고 자신을 표현할 수 있는 다양한 방법을 적극적으로 찾아보세요.

# 보조 평가지표

직업/강좌 설명으로 만든 임베딩

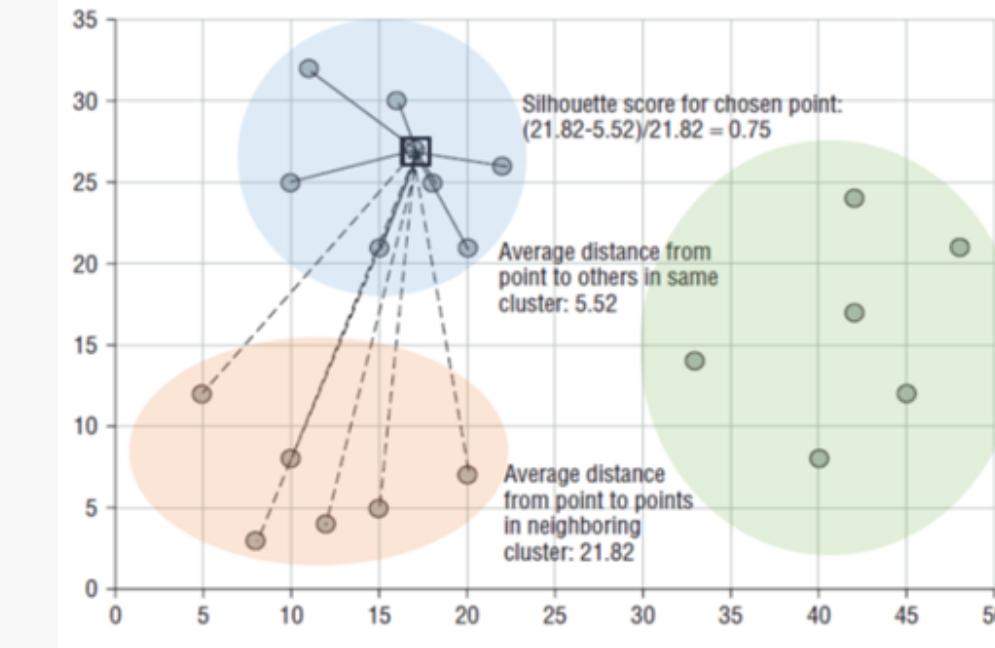
라벨을 가지는 벡터

클러스터링



실루엣 스코어

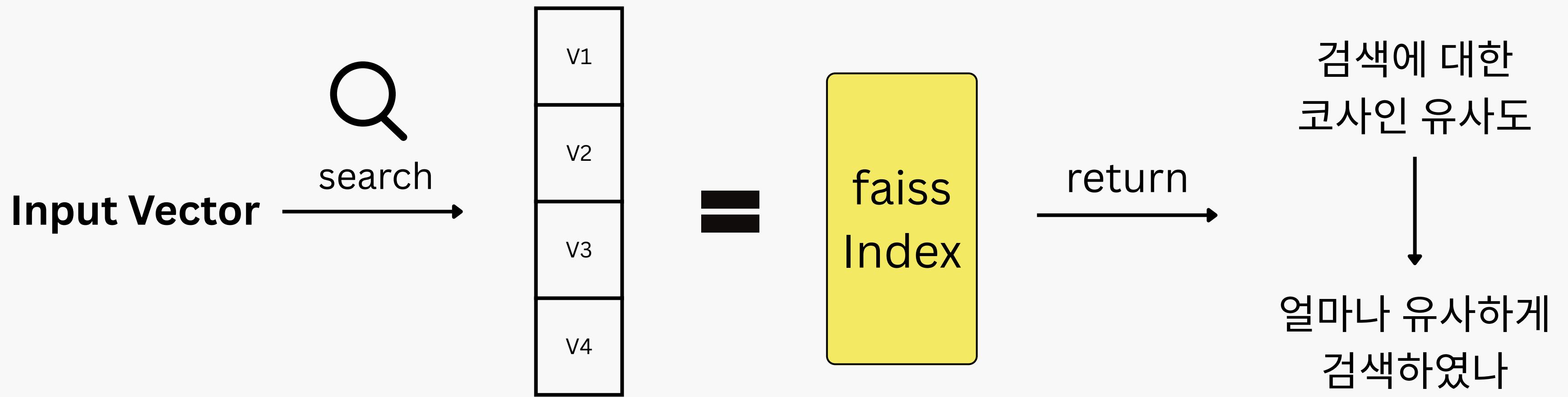
$$s(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))}$$



데이비스-볼딘 지수

CH 점수

# 보조 평가지표

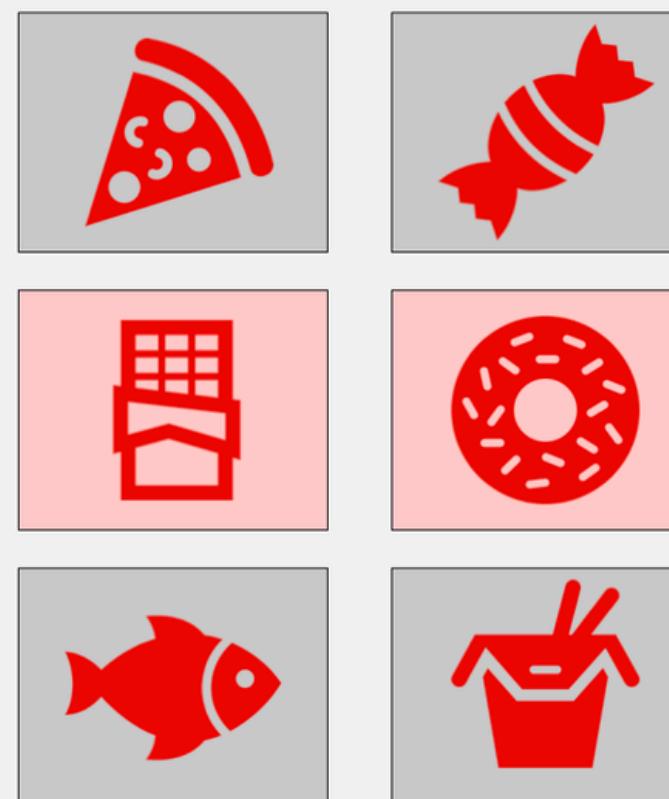


# 주 평가지표

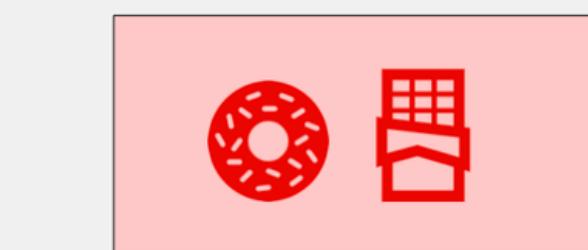
$$\text{Precision at } K = \frac{\text{Number of relevant items in } K}{\text{Total number of items in } K}$$

Precision@K

Interaction results



$$\text{Precision @6} = \frac{\text{Number of relevant items in } K}{\text{Total number of items in } K} = 0.33$$



Not relevant

Relevant



# 주 평가지표

$$\text{Recall at } K = \frac{\text{Number of relevant items in } K}{\text{Total number of relevant items}}$$

Recall@K



# 주 평가지표

F1@K

$$F_{\beta} = \frac{(1 + \beta^2) \times \text{Precision at K} \times \text{Recall at K}}{(\beta^2 \times \text{Precision at K}) + \text{Recall at K}}$$



# 평가방법

사용자입력 → 직업 추천

직업 → 강의 추천

각 작업을 독립적으로 평가(보조, 주 평가지표 이용)

## 테스트셋

사람이 생성한 (사용자 입력, 기대 직업 출력), (직업 이름, 기대 강의 출력) 쌍을 테스트셋으로 사용

input	work	직업이름	추천강좌
나는 환자들의 질병을 정확히 진단해서 적절한 치료 방법을 안내하고 싶어	임상병리사,한의사,가정의학과의사,보건의료정보관리사,간호사,외과의사,마취병리과의사,일...	3D프린팅모델러	3DPrinting제품모델링,3D프린터아이디어를출력하라,실감3D캐릭터모델링,3D모션...
나는 혁신적인 기술을 연구하고 설계해서 사람들의 삶을 편리하게 만드는 발명품을 만들...	제품디자이너,로봇공학기술자,3D프린팅모델러,전자제품및부품개발기술자,고무·플라스틱제품...	IT기술지원전문가	심화클라우드보안엔지니어양성과정,클라우드보안엔지니어양성과정,웨어러블로봇하드웨어의특징과...
나는 재미있고 교육적인 게임을 기획하고 코딩해서 사람들이 즐겁게 플레이할 수 있는 ...	게임프로그래머,게임그래픽디자이너,웹디자이너,웹개발자,모바일앱개발자,만화영화작가(애니...	IT테스터 및 IT QA전문가	디지털서비스프로젝트관리실무,리눅스활용실습-PracticeinLinux,오픈소스를이용...
나는 어려움에 처한 사람들의 권리를 변호하고 사회 정의를 실현하고 싶어	사회단체활동가,사회복지사,작업치료사,법무사및집행관,생활지도원,변호사,일반행정공무원,...	UX-UI디자이너	고객니즈분석을통한융합제품개발전략,신제품설계론,창업가를위한브랜드네이밍기법,머신러닝기반...
나는 학생들의 잠재력을 이끌어내고 흥미를 유발하는 수업을 통해 지식을 전달하고 싶어	중·고등학교교사,문리학원강사,예능강사,안무가,기술·기능계강사,보조교사,초등학교교사,...	가구제조·수리원	집콕테스트마이닝실전및분석,집콕딥러닝시대에도필요한고급기계학습,신제품설계론,디자인프로세...
...	...	...	...
여러 사람과 동시다발로 대화하는 채팅 업무는 집중력이 산만해졌어요. 누군가 메시지를...	속기사,텔레마케터,모바일앱개발자,취업알선원,이동전화기수리원,통신장비설치·수리원,고객...	환경미화원 및 재활용품수거원	폐기물관리및처리,하수관로조사및정보구축,하수관로조사및정보구축,분리선별공학광물및순환자원...
새로운 광고 카피를 기한 내에 완성해야 하는 건 막상 시작하면 손가락이 멈춰 버렸어...	속기사,데이터분석가(빅데이터분석가),광고·홍보·마케팅전문가,시각디자이너,간판제작·설...	환경시험원	진공공학및설계,제조시스템심화교육과정CNCHMI마신세팅및측정,산학프로젝트실습실감형기기...
팀원들의 사적인 고민을 상담해 주는 일은 의외로 큰 감정 소모가 되었어요. 한 사람...	속기사,예술치료사,취업알선원,조사전문가,고객상담원,전산자료입력원및사무보조원,심리상담...	환경영향평가원	환경영향평가-EnvironmentalImpactAssessment,환경법및정책-En...
잡다한 임시 요청을 수시로 처리해야 하는 업무는 내 플로우를 자주 끊어 놓더라고요....	하역·적재종사원,취업알선원,전산자료입력원및사무보조원,고객상담원,패스트푸드준비원,선박...	회계사	사회복지회계행정의이해,회계정보의이해와활용,회계정보의이해,원가및관리회계,회계원리-Pr...
발표 자료를 실시간으로 공유하며 설명하는 건 제겐 너무 긴장되는 순간이었어요. 화면...	속기사,영상·녹화및편집기사,IT테스터및ITQA전문가,미디어콘텐츠창작자(크리에이터),...	회계사무원	원가및관리회계,회계정보의이해,재무관리,회계정보의이해와활용,사회복지회계행정의이해,대학...

# 모델 개선 과정 (초기 설정)

임베딩 모델 : paraphrase-multilingual-MiniLM-L12-v2 (384차원)

문서 분할 : 한 문장으로 분할

임베딩 보조지표

```
work embedding metric
sil : -0.07130 dbi : 4.77893 chi : 6.40043
```

```
lecture embedding metric
sil : 0.28362 dbi : 40621.07584 chi : 5.72854
```

# 모델의 유사도 결과값

## 직업 유사도 결과

	0	1	2	3	4
count	280.000000	280.000000	280.000000	280.000000	280.000000
mean	0.594785	0.589815	0.586971	0.588776	0.592456
std	0.072330	0.066855	0.067275	0.069220	0.069240
min	0.375915	0.409637	0.397551	0.370026	0.398659
25%	0.544058	0.542489	0.537893	0.540182	0.539889
50%	0.598415	0.586990	0.585038	0.588168	0.593761
75%	0.643136	0.636398	0.637695	0.634219	0.641529
max	0.790372	0.761444	0.776188	0.781949	0.763691

## 강좌 유사도 결과

	0	1	2	3	4	5	6	7	8	9
count	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000
mean	0.607716	0.606714	0.606332	0.605130	0.606721	0.606773	0.605388	0.602455	0.600589	0.597072
std	0.070071	0.074080	0.068807	0.072837	0.073278	0.074514	0.072334	0.074635	0.071593	0.072762
min	0.396502	0.379704	0.356296	0.399894	0.325845	0.356743	0.389237	0.366386	0.393304	0.391962
25%	0.561448	0.560714	0.563856	0.559431	0.559842	0.552671	0.552538	0.549415	0.549901	0.545617
50%	0.616078	0.611055	0.609163	0.608196	0.609046	0.610096	0.608290	0.605786	0.607539	0.603024
75%	0.653900	0.657873	0.657034	0.659942	0.658471	0.662105	0.658500	0.658473	0.652621	0.649994
max	0.799572	0.776033	0.819265	0.844283	0.844283	0.834955	0.846662	0.817422	0.770928	0.824187

# 주 지표

직업 추천지표 결과

	<b>precision</b>	<b>recall</b>	<b>f1</b>
count	280.000000	280.000000	280.000000
mean	0.467857	0.233929	0.311905
std	0.281514	0.140757	0.187676
min	0.000000	0.000000	0.000000
25%	0.200000	0.100000	0.133333
50%	0.400000	0.200000	0.266667
75%	0.600000	0.300000	0.400000
max	1.000000	0.500000	0.666667

강좌 추천지표 결과

	<b>precision</b>	<b>recall</b>	<b>f1</b>
count	537.000000	537.000000	537.000000
mean	0.032402	0.009140	0.013979
std	0.059791	0.018807	0.027756
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.100000	0.014925	0.025974
max	0.300000	0.142857	0.193548

# 모델 개선 과정

<임베딩 모델 변경 >

paraphrase-multilingual-MiniLM-L12-v2

→ **paraphrase-multilingual-mpnet-base-v2**

<임베딩 보조지표>

```
work embedding metric  
sil : -0.05499 dbi : 4.58730 chi : 7.22057
```

```
lecture embedding metric  
sil : 0.29573 dbi : 54063.63149 chi : 6.00787
```

# 모델의 유사도 결과값

## 직업 유사도 결과

	0	1	2	3	4
count	280.000000	280.000000	280.000000	280.000000	280.000000
mean	0.609039	0.604918	0.606380	0.605602	0.606472
std	0.062869	0.066654	0.065684	0.067497	0.064969
min	0.453076	0.414593	0.432301	0.436646	0.441177
25%	0.563090	0.554797	0.558230	0.560273	0.563145
50%	0.608898	0.599900	0.608283	0.604237	0.602528
75%	0.652266	0.650644	0.651476	0.649750	0.653744
max	0.767325	0.808695	0.781613	0.828718	0.777224

## 강좌 유사도 결과

	0	1	2	3	4	5	6	7	8	9
count	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000
mean	0.613083	0.608169	0.607354	0.612646	0.608987	0.607792	0.605124	0.604279	0.602705	0.602712
std	0.063055	0.068072	0.069354	0.068969	0.067703	0.066607	0.071095	0.073281	0.066921	0.069291
min	0.408866	0.400352	0.389298	0.407029	0.339081	0.428387	0.370034	0.345692	0.379860	0.391689
25%	0.572479	0.564112	0.563759	0.570236	0.572837	0.561859	0.555790	0.557638	0.557758	0.551496
50%	0.616572	0.611137	0.613959	0.615085	0.615282	0.607738	0.605840	0.611037	0.605798	0.603699
75%	0.656713	0.657396	0.656214	0.656775	0.652457	0.655331	0.655936	0.654379	0.647610	0.653044
max	0.763517	0.774201	0.809799	0.810940	0.806358	0.781862	0.805450	0.810521	0.794687	0.786446

# 주 지표

직업 추천지표 결과

	<b>precision</b>	<b>recall</b>	<b>f1</b>
count	280.000000	280.000000	280.000000
mean	0.517143	0.258571	0.344762
std	0.284345	0.142172	0.189563
min	0.000000	0.000000	0.000000
25%	0.200000	0.100000	0.133333
50%	0.600000	0.300000	0.400000
75%	0.800000	0.400000	0.533333
max	1.000000	0.500000	0.666667

강좌 추천지표 결과

	<b>precision</b>	<b>recall</b>	<b>f1</b>
count	537.000000	537.000000	537.000000
mean	0.030912	0.008514	0.013111
std	0.057414	0.017514	0.026116
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.100000	0.013514	0.023810
max	0.300000	0.107143	0.157895

# 모델 개선 과정

<문서 분할 방법 변경>

한 문장이 아닌 **두 문장 단위**로 분할 (2-gram)

<임베딩 보조지표>

```
work embedding metric  
sil : -0.00928 dbi : 3.41456 chi : 10.51341
```

```
lecture embedding metric  
sil : 0.40087 dbi : 29855.90890 chi : 9.83588
```

# 모델의 유사도 결과값

## 직업 유사도 결과

	0	1	2	3	4
count	280.000000	280.000000	280.000000	280.000000	277.000000
mean	0.591573	0.583986	0.579497	0.574159	0.568294
std	0.068273	0.067280	0.066755	0.066969	0.066673
min	0.414889	0.384809	0.369986	0.397181	0.403016
25%	0.541994	0.544175	0.537316	0.531146	0.523383
50%	0.592765	0.585668	0.578076	0.571277	0.566949
75%	0.642410	0.622895	0.619819	0.616918	0.613622
max	0.754574	0.785294	0.780048	0.801467	0.782600

## 강좌 유사도 결과

	0	1	2	3	4	5	6	7	8	9
count	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000
mean	0.602023	0.598871	0.596174	0.600917	0.598713	0.596602	0.596189	0.595541	0.593501	0.592313
std	0.066316	0.067529	0.070847	0.067862	0.069404	0.071637	0.070308	0.073038	0.074668	0.075324
min	0.401559	0.407029	0.369484	0.389298	0.372408	0.339081	0.312004	0.274558	0.325949	0.310074
25%	0.560550	0.554478	0.550277	0.554803	0.560002	0.551612	0.545638	0.550546	0.541242	0.544577
50%	0.603742	0.600695	0.601392	0.603220	0.602353	0.602384	0.602506	0.602248	0.598418	0.596331
75%	0.647790	0.647077	0.650322	0.649032	0.645120	0.646890	0.641898	0.646930	0.648736	0.641485
max	0.763517	0.774201	0.805451	0.805156	0.778780	0.809799	0.810940	0.810521	0.781862	0.806586

# 주 지표

직업 추천지표 결과

	precision	recall	f1
count	280.000000	280.000000	280.000000
mean	0.512857	0.256429	0.341905
std	0.281533	0.140766	0.187688
min	0.000000	0.000000	0.000000
25%	0.200000	0.100000	0.133333
50%	0.600000	0.300000	0.400000
75%	0.800000	0.400000	0.533333
max	1.000000	0.500000	0.666667

강좌 추천지표 결과

	precision	recall	f1
count	537.000000	537.000000	537.000000
mean	0.034264	0.009304	0.014375
std	0.060618	0.018036	0.027006
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.100000	0.017241	0.029412
max	0.300000	0.107143	0.157895

precision min value	precision min input
0.0	심야에 별빛을 보며 사진을 찍었더니 장노출 사진 속에 은하수 띠가 또렷이 떠올랐어....
0.0	잡다한 임시 요청을 수시로 처리해야 하는 업무는 내 플로우를 자주 끊어 놓더라고요....
0.0	바닷가 조개 껌데기를 주워보니 나선형 구조가 수학 공식처럼 느껴졌어. 그 껌데기를 ...
0.0	카메라로 도심 속 나무들의 변화를 찍어봤는데 계절마다 잎 색이 다채롭게 변하더라고....
0.0	공원 연못에 비친 하늘을 사진으로 찍고 나니 물결에 따라 하늘이 일렁이는 모습이 아...
precision max value	precision max input
1.0	나는 저널리즘 윤리를 준수하며 공정한 기사 작성으로 사회 문제를 알리고 싶어
1.0	나는 관광 가이드를 하면서 한국의 문화와 역사를 해외 관광객에게 알리고 싶어
1.0	나는 사람들의 마음을 울리는 깊이 있는 소설을 집필해서 독자들에게 감동을 선사하고 싶어
1.0	나는 신문사 기자로 일하며 사회적 이슈를 심층 취재하고 싶어
1.0	나는 자동차 엔진 엔지니어링을 연구하여 친환경 하이브리드 차량의 성능을 개선하고 싶어
recall min value	recall min input
0.0	도시 공원의 분수 물줄기가 바람에 따라 리듬을 타는 듯 보여서 고속 카메라로 찍어보...
0.0	등산길에 핀 야생화를 보고 있으면 작은 꽃 하나에도 생명력이 가득하다는 걸 느껴....
0.0	유적지 돌담을 손으로 만져보면 수백 년의 흔적이 고스란히 느껴지는데 그 질감을 스캔...
0.0	수시로 바뀌는 프로젝트 범위는 내 작업 흐름을 완전히 무너뜨리더라고요. 어제까지 확...
0.0	바닷가 조개 껌데기를 주워보니 나선형 구조가 수학 공식처럼 느껴졌어. 그 껌데기를 ...
recall max value	recall max input
0.5	나는 자동차 엔진 엔지니어링을 연구하여 친환경 하이브리드 차량의 성능을 개선하고 싶어
0.5	나는 애니메이션 캐릭터를 기획하고 3D 모델링하여 생동감 있는 애니메이션을 제작하고 싶어
0.5	나는 관광 가이드를 하면서 한국의 문화와 역사를 해외 관광객에게 알리고 싶어
0.5	나는 금융 시장을 분석하여 적절한 투자 전략을 수립하고 수익을 극대화하고 싶어
0.5	나는 청소년 심리 상담 센터를 열어 정신적으로 어려움을 겪는 친구들을 지원하고 싶어
f1 min value	f1 min input
0.0	하나의 프로젝트가 끝나기도 전에 다른 긴급 과제가 시작되는 상황은 버거웠어요. 끝나...
0.0	잡다한 임시 요청을 수시로 처리해야 하는 업무는 내 플로우를 자주 끊어 놓더라고요....
0.0	나는 서양화 기법을 배우고 개인전을 열어 나만의 화풍을 선보이고 싶어
0.0	나는 기상학을 공부해 기후 변화 패턴을 분석하고 예측 모델을 개발하고 싶어
0.0	수시로 바뀌는 프로젝트 범위는 내 작업 흐름을 완전히 무너뜨리더라고요. 어제까지 확...
f1 max value	f1 max input
0.666667	나는 금융 시장을 분석하여 적절한 투자 전략을 수립하고 수익을 극대화하고 싶어
0.666667	나는 저널리즘 윤리를 준수하며 공정한 기사 작성으로 사회 문제를 알리고 싶어
0.666667	나는 가상 자산(암호화폐) 연구를 통해 블록체인 기반 금융 서비스를 개발하고 싶어
0.666667	나는 피아노 연주 실력을 갈고닦아 콘서트홀에서 독주회를 열고 싶어
0.666667	나는 음향 엔지니어링을 배우고 콘서트 현장에서 고품질 사운드를 제공하고 싶어
0 min value	0 min input
0	0.471185 비 오는 창가에 앉아 빗방울이 떨어지는 소리를 듣는데 리듬이 마치 자연의 박자처럼 ...
0 max value	0 max input
0	0.754574 나는 기업 재무 분석을 통해 재무 건전성을 평가하고 투자 전략을 제시하고 싶어

**지표의 극단값의 입력도 파악  
→ 어느 입력에 대해 모델이 부정확해지는지 파악**

## 취약점

**직업 정보가 없는 사용자의 성격, 감성적 정보의 입력  
or 부정적 문장(사용자가 뭘 싫어하는지)**

# 모델 변화 과정

## 지표기반 직업 추천 도입

-> 모델의 취약점인  
직업적 정보가 없는 사용자의 성격, 감성적 정보가 들어간 입력

직업과 관련된 텍스트 정보가 있어야 기대 수준으로 작동  
(게임을 만들고 싶다, 데이터 분석을 하고 싶다)

그러나 이 서비스를 사용할 사람들은  
감성적 텍스트를 입력하는 경향이 있다.  
(나는 이런 성격이야, 나는 ~를 좋아해)

# 직업 유사도 결과

직업 유사도 결과 요약

	0	1	2	3	4
count	280.000000	280.000000	280.000000	280.000000	278.000000
mean	0.601916	0.585805	0.580741	0.574238	0.566952
std	0.057505	0.062120	0.063041	0.065182	0.066092
min	0.471185	0.413373	0.403016	0.384809	0.369986
25%	0.561930	0.540801	0.533707	0.524647	0.514869
50%	0.597455	0.581102	0.576703	0.570648	0.563661
75%	0.642410	0.622482	0.619819	0.616918	0.613587
max	0.754574	0.785294	0.780048	0.801467	0.782600

# 강좌 유사도 결과

	0	1	2	3	4	5	6	7	8	9
count	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000	537.000000
mean	0.602023	0.598871	0.596174	0.600917	0.598713	0.596602	0.596189	0.595541	0.593501	0.592313
std	0.066316	0.067529	0.070847	0.067862	0.069404	0.071637	0.070308	0.073038	0.074668	0.075324
min	0.401559	0.407029	0.369484	0.389298	0.372408	0.339081	0.312004	0.274558	0.325949	0.310074
25%	0.560550	0.554478	0.550277	0.554803	0.560002	0.551612	0.545638	0.550546	0.541242	0.544577
50%	0.603742	0.600695	0.601392	0.603220	0.602353	0.602384	0.602506	0.602248	0.598418	0.596331
75%	0.647790	0.647077	0.650322	0.649032	0.645120	0.646890	0.641898	0.646930	0.648736	0.641485
max	0.763517	0.774201	0.805451	0.805156	0.778780	0.809799	0.810940	0.810521	0.781862	0.806586

## 임베딩 보조지표

work embedding metric

sil : -0.00928 dbi : 3.41456 chi : 10.51341

lecture embedding metric

sil : 0.40087 dbi : 29855.90890 chi : 9.83588

# 주 지표

직업 추천지표 결과

	<b>precision</b>	<b>recall</b>	<b>f1</b>
count	280.000000	280.000000	280.000000
mean	0.494286	0.247143	0.329524
std	0.287311	0.143656	0.191541
min	0.000000	0.000000	0.000000
25%	0.200000	0.100000	0.133333
50%	0.600000	0.300000	0.400000
75%	0.800000	0.400000	0.533333
max	1.000000	0.500000	0.666667

강좌 추천지표 결과

	<b>precision</b>	<b>recall</b>	<b>f1</b>
count	537.000000	537.000000	537.000000
mean	0.034264	0.009304	0.014375
std	0.060618	0.018036	0.027006
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.100000	0.017241	0.029412
max	0.300000	0.107143	0.157895

# 한계점

1. 사용자 **입력의 질**이 좋아야 함

→ 감성적 내용보다는 **건조하게 구체적으로**, **유의미한 텍스트**가 들어갈 수 있게 **자세히** 입력해야 함.

2. **부정의 표현이 끈임**

→ **A가 싫어** → **but, A와 관련한 것** 찾아줌

3. **지표점수**의 부정확성

→ **모든 직업이 비슷한 벡터**로 표현됨

→ **비슷한 입력에 대해 같은 직업만 추천**할 가능성

4. **선 후수 과목의 부재**

→ **유사도로만** 강의 추천,

따라서 **선, 후수과목의 관계성은 고려하지 않음**

5. **이상값** 처리

→ [0.8, 0.1, 0.1, 0.1, 0.1] 같이 **극단적인 추천**이  
있을 때, 상대적으로 **정상 추천인 0.8의 것이 끈임**

6. **검증한계**

-> 1과 같이 **평가지표를 위한 적절한 테스트셋을 만들기 어렵다**

# 기여점

1. 사용자가 원하는 정보를 찾기 위한 **소요시간이 짧음**

→ 고용24 등에서 제공하는 **직업심리**

**검사보다 정보 입력 요구량이 적음**

2. **자연어 입력**으로 진로 탐색 가능

→ 전용 검사가 필요한 기존의 방법보단

**접근성이 좋고 심리적 부담감이 적음**

3. **공신력 있는 데이터**를 기반으로 함

→ 단순히 인터넷 검색으로 접하는

**무작위의 데이터보단 신뢰 가능**

4. **직업과 강좌 모두 검색**

→ **한번의 검사로 일률적으로** 진로 탐색

# Q&A