



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0157119
(43) 공개일자 2022년11월29일

(51) 국제특허분류(Int. Cl.)

G06Q 50/20 (2012.01) G06F 16/33 (2019.01)
G06F 16/335 (2019.01) G06F 16/34 (2019.01)
G06F 16/9535 (2019.01) G06N 20/00 (2019.01)

(52) CPC특허분류

G06Q 50/20 (2013.01)
G06F 16/3334 (2019.01)

(21) 출원번호 10-2021-0064780

(22) 출원일자 2021년05월20일
심사청구일자 2021년05월20일

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

문기범

서울특별시 도봉구 덕릉로60카길 38, 2층(창동)

이진숙

서울특별시 종로구 난계로29가길 24,710호(승인동, 종로 동광모닝스카이)

(뒷면에 계속)

(74) 대리인

윤귀상

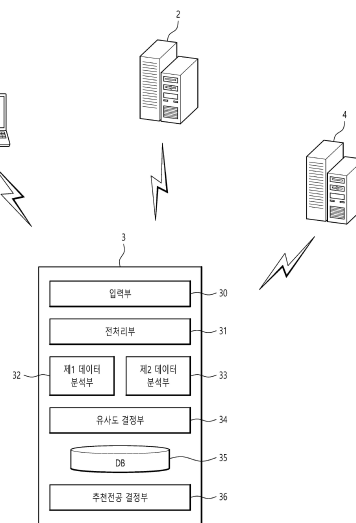
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 사용자 관심 키워드 기반 전공 추천 시스템과 방법 및 이를 위한 컴퓨터 프로그램

(57) 요약

사용자 관심 키워드에 기반한 전공 추천 시스템은, 전공별 강의내용 정보를 생성하고, 전공별 과목명 키워드 정보를 생성하며, 전공별 강의 내용 정보, 전공별 과목명 키워드 정보 및 사용자 관심 키워드 정보를 이용하여 전공별 유사도 정보를 생성하여 제2전공 추천 정보를 제공할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06F 16/3338 (2019.01)

G06F 16/3344 (2019.01)

G06F 16/335 (2019.01)

G06F 16/345 (2019.01)

G06F 16/9535 (2019.01)

G06N 20/00 (2021.08)

(72) 발명자

이수장

서울특별시 마포구 월드컵북로 501,914동302호 (상
암동, 상암월드컵파크9단지)

한수연

서울특별시 용산구 백범로81길 10,202호(원효로1
가)

권혜정

서울특별시 강남구 선릉로69길 19,102동 203호(역
삼동, 역삼래미안아파트)

한재호

서울특별시 노원구 공릉로58길 161,103동 104호(하
계동, 코오롱마들마을아파트)

김규태

경기도 안양시 동안구 시민대로327번길 55,108동
2105호 (관양동, 평촌더샵센트럴시티)

명세서

청구범위

청구항 1

사용자 관심 키워드에 기반한 전공 추천 시스템으로서,

사용자에 의해 관심 키워드 정보를 입력 받는 입력부;

하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하고 상기 강의내용 정보에 전공을 임베딩하여 전공별 강의내용 정보를 생성하도록 구성된 제1 데이터 분석부;

과목명 키워드 정보를 생성하고 상기 과목명 키워드 정보에 전공을 임베딩하여 전공별 과목명 키워드 정보를 생성하도록 구성된 제2 데이터 분석부;

상기 전공별 강의내용 정보 및 상기 전공별 과목명 키워드 정보를 이용하여 상기 사용자로부터 입력된 관심 키워드 정보와 각 전공간 유사도를 결정하도록 구성된 유사도 결정부; 및

상기 전공간 유사도 정보에 기초하여 상기 사용자에게 타 학과 전공인 제2 전공에 대한 추천 정보를 생성하도록 구성된 추천전공 결정부;를 포함하는 사용자 관심 키워드에 기반한 전공 추천 시스템.

청구항 2

제1항에 있어서,

상기 과목의 교수요목 및 학문분야별 출간논문 키워드 중 하나 이상으로부터 상기 텍스트 데이터를 추출하도록 구성된 전처리부를 더 포함하는 사용자 관심 키워드에 기반한 전공 추천 시스템.

청구항 3

제1항에 있어서,

상기 제1 데이터 분석부는, 상기 텍스트 데이터에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 과목 각각에 대한 강의내용을 학습하고, 상기 강의내용에 대한 학습 결과에 기초하여 상기 강의내용 정보를 생성하며,

전공별 커리큘럼 정보와 상기 생성된 강의내용 정보를 이용하여 전공별 강의내용 정보를 생성하도록 구성된 사용자 관심 키워드에 기반한 전공 추천 시스템.

청구항 4

제1항에 있어서,

상기 제2 데이터 분석부는, 커리큘럼 데이터로부터 필수이수 과목명을 추출하고 상기 필수이수 과목명으로부터 텍스트를 추출하며, 상기 필수이수 과목명과 상기 필수이수 과목명으로부터 추출된 텍스트를 병합하고

상기 병합된 텍스트에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 병합된 텍스트를 학습하고, 상기 병합된 텍스트에 대한 학습 결과에 기초하여 상기 전공별 과목명 키워드 정보를 생성하도록 구성된 사용자 관심 키워드에 기반한 전공 추천 시스템.

청구항 5

제1항에 있어서,

상기 유사도 결정부는, 상기 전공별 강의내용 정보 및 상기 전공별 과목명 키워드 정보에 각각 상응하는 매트릭스와 상기 입력된 사용자 관심 키워드 정보 매트릭스 간의 코사인 유사도를 이용하여 각 전공간 유사도 정보를 생성하도록 구성된 사용자 관심 키워드에 기반한 전공 추천 시스템.

청구항 6

제1항에 있어서,

상기 추천 전공 결정부는, 전공별 특성에 상응하는 복수 개의 차원에 의해 정의되며 전공 간의 유사도가 높을수록 전공들이 서로 가까이 위치하는 전공지도를 이용하여 상기 전공 추천 정보를 상기 사용자에게 제공하도록 더 구성된 키워드에 기반한 전공 추천 시스템.

청구항 7

제1항에 있어서,

상기 추천정보 결정부는,

제2 전공 관련 상세 정보 및 추천 활동 정보를 생성하도록 더 구성된 사용자 관심 키워드에 기반한 전공 추천 시스템.

청구항 8

하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하는 단계;

상기 생성된 강의내용 정보에 전공을 임베딩하여 전공별 강의내용 정보를 생성하는 단계;

과목명 키워드 정보를 생성하는 단계;

생성된 과목명 키워드 정보에 전공을 임베딩하여 전공별 과목명 키워드 정보를 생성하는 단계;

사용자가 관심 키워드 정보를 입력하는 단계;

전공별 강의내용 정보 및 전공별 과목명 키워드 정보를 이용하여 상기 사용자로부터 입력된 관심 키워드 정보와 각 전공간 유사도를 결정하는 단계; 및

상기 각 전공간 유사도 정보에 기초하여 상기 사용자에게 타 학과 전공인 제2 전공에 대한 추천 정보를 생성하는 단계를 포함하는 사용자 관심 키워드에 기반한 전공 추천 방법.

청구항 9

제 8항에 있어서,

상기 강의내용 정보를 생성하는 단계는,

상기 과목의 교수요목 및 학문분야별 출간논문 키워드 중 하나 이상으로부터 상기 텍스트 데이터를 추출하여 데이터를 분석하는 사용자 관심 키워드에 기반한 전공 추천 방법.

청구항 10

제 8항에 있어서,

상기 강의내용 정보를 생성하는 단계는,

상기 텍스트 데이터에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 과목 각각에 대한 강의내용을 학습하는 단계; 및

상기 강의내용에 대한 학습 결과에 기초하여 상기 강의내용 정보를 생성하는 단계를 포함하고

상기 전공별 강의내용 정보를 생성하는 단계는,

전공별 커리큘럼 정보와 상기 생성된 강의내용 정보를 이용하여 전공별 강의 내용 정보를 생성하는 단계를 포함하는 사용자 관심 키워드에 기반한 전공 추천 방법.

청구항 11

제 8항에 있어서,

상기 과목명 키워드 정보를 생성하는 단계는,

커리큘럼 데이터로부터 필수이수 과목명을 추출하는 단계; 상기 추출된 필수이수 과목명으로부터 텍스트를 추출하는 단계; 상기 추출된 필수이수 과목명과 상기 필수 이수 과목명으로부터 추출된 텍스트를 병합하는 단계;

상기 병합된 텍스트에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 병합된 텍스트를 학습하는 단계; 및

상기 병합된 텍스트에 대한 학습 결과에 기초하여 상기 과목명 키워드 정보를 생성하는 단계;를 포함하는 사용자 관심 키워드에 기반한 전공 추천 방법.

청구항 12

제 8항에 있어서,

상기 유사도 정보를 생성하는 단계는,

상기 전공별 강의내용 정보 및 상기 전공별 과목명 키워드 정보에 각각 상응하는 매트릭스와 상기 입력된 사용자 관심 키워드 정보 매트릭스 간의 코사인 유사도를 산출하는 단계를 포함하는 사용자 관심 키워드에 기반한 전공 추천 방법.

청구항 13

제 8항에 있어서,

제2 전공 추천 정보를 생성하는 단계는,

전공별 특성에 상응하는 복수 개의 차원에 의해 정의되며 전공 간의 유사도가 높을수록 전공들이 서로 가까이 위치하는 전공지도를 이용하여 상기 전공 추천 정보를 상기 사용자에게 제공하는 단계를 더 포함하는 사용자 관심 키워드에 기반한 전공 추천 방법.

청구항 14

제 8항에 있어서,

제2 전공 추천 정보를 생성하는 단계는,

제2 전공 관련 상세 정보 및 추천 활동 정보를 생성하는 단계를 더 포함하는 사용자 관심 키워드에 기반한 전공 추천 방법.

청구항 15

하드웨어와 결합되어 제8항에 따른 사용자 관심 키워드에 기반한 전공 추천 방법을 실행하도록 컴퓨터로 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 사용자 관심 키워드에 기반한 전공 추천 시스템과 방법 및 이를 위한 컴퓨터 프로그램에 관한 것으로서, 보다 상세하게는, 사용자 관심 키워드와 각 전공간 유사도를 산출하여 전공을 추천해주는 사용자 관심 키워드에 기반한 전공 추천 시스템과 방법 및 이를 위한 컴퓨터 프로그램에 관한 것이다.

배경 기술

[0002] 최근 많은 전문가들은 인공지능(Artificial Intelligence; AI) 기술이 우리 사회에 혁명적인 변화를 일으킬 것으로 전망하고 있다. 특히, 코로나바이러스 감염증-19(COVID-19)로 인해 사회 각 분야의 디지털화가 빠르게 진행되면서 많은 양의 데이터가 축적되고 있으며, AI 기술의 도입과 확산이 가속화되고 있다.

[0003] 교육 영역에서도 AI 도입에 관한 논의가 활발하게 이루어지고 있으며, 대한민국 정부에서는 AI기술을 활용한 개인 맞춤형 교육을 교육의 발전 방향으로 상정 및 추진하고 있다.

[0004] 종래의 전공 추천 시스템은 사용자의 적성 또는 성격을 검사하기 위한 설문지 문항별로 응답 정보를 입력하고 응답 정보에 기초하여 전공을 선택하기 위한 사용자에게 전공을 추천해주는 방식으로 대부분 이루어지고 있었다.

[0005] 더욱이, 문항별로 사용자가 일일이 설문지를 체크해야 하는 불편함이 있어 많은 시간이 소요되며 추천 전공 범위가 좁고 정확하지 않다는 한계점이 있었다.

[0006] 따라서 정확하고 간단한 방식의 전공 추천 시스템과 사용자에게 전체 전공에 대하여 추천 순위별로 한번에 조망할 수 있도록 시각화하여 추천전공을 제공하는 전공 추천 시스템이 요구되고 있었다.

발명의 내용

해결하려는 과제

[0007] 본 발명이 해결하고자 하는 기술적 과제는 과목의 교수요목, 학문분야별 출간논문 키워드와 같은 비정형 데이터를 기초로 강의내용 정보를 생성할 수 있으며, 강의내용 차원과 과목명 키워드 정보를 이용하여 사용자 관심 키워드 정보와 각 전공 간의 유사도를 도출하여 추천 전공을 제공할 수 있는 사용자 관심 키워드 기반 전공 추천 시스템과 방법 및 이를 위한 컴퓨터 프로그램을 제공하는 것을 목적으로 한다.

[0008] 또한, 본 발명이 해결하고자 하는 기술적 과제는, 사용자에게 추천하는 전공을 전공지도 형태로 시각화하여 제공할 수 있는 사용자 관심 키워드 기반 전공 추천 시스템과 방법 및 이를 위한 컴퓨터 프로그램을 제공하는 것을 목적으로 한다.

[0009] 또한, 본 발명이 해결하고자 하는 기술적 과제는, 사용자가 관심 키워드 정보를 입력하거나 전공지도에서 관심 전공을 선택하면 전공 관련 상세 정보 및 추천활동을 제공하는 것을 목적으로 한다.

[0010] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0011] 본 발명의 일측면에 따른 사용자 관심 키워드에 기반한 전공 추천 시스템은, 사용자에게 의해 관심 키워드 정보를 입력 받는 입력부; 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하고 상기 강의내용 정보에 전공을 임베딩하여 전공별 강의내용 정보를 생성하도록 구성된 제1 데이터 분석부; 과목명 키워드 정보를 생성하고 상기 과목명 키워드 정보에 전공을 임베딩하여 전공별

과목명 키워드 정보를 생성하도록 구성된 제2 데이터 분석부; 상기 전공별 강의 내용 정보, 상기 전공별 과목명 키워드 정보를 이용하여 상기 사용자로부터 입력된 관심 키워드 정보와 각 전공간 유사도를 결정하도록 구성된 유사도 결정부; 및 상기 각 전공간 유사도 정보에 기초하여 상기 사용자에게 타 학과 전공인 제2 전공에 대한 추천 정보를 생성하도록 구성된 추천전공 결정부;를 포함할 수 있다.

[0012] 일 실시예에 따른 전공 추천 시스템은, 상기 과목의 교수요목 및 학문분야별 출간논문 키워드 중 하나 이상으로부터 상기 텍스트 데이터를 추출하도록 구성된 전처리부를 더 포함할 수 있다.

[0013] 일 실시예에서, 상기 제 1 데이터 분석부는, 상기 텍스트 데이터에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 과목 각각에 대한 강의내용을 학습하고, 상기 강의내용에 대한 학습 결과에 기초하여 상기 강의내용 정보를 생성하며, 전공별 커리큘럼 정보와 상기 생성된 강의내용 정보를 이용하여 전공별 강의내용 정보를 생성하도록 더 구성될 수 있다.

[0014] 일 실시예에서, 상기 제2 데이터 분석부는, 커리큘럼 데이터로부터 필수이수 과목명을 추출하고 상기 필수이수 과목명으로부터 텍스트를 추출하며, 상기 필수이수 과목명과 상기 필수이수 과목명으로부터 추출된 텍스트를 병합하고, 상기 병합된 텍스트에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 병합된 텍스트를 학습하고, 상기 병합된 텍스트에 대한 학습 결과에 기초하여 상기 전공별 과목명 키워드 정보를 생성하도록 더 구성될 수 있다.

[0015] 일 실시예에서, 상기 유사도 결정부는, 상기 전공별 강의내용 정보 및 상기 전공별 과목명 키워드 정보에 각각 상응하는 매트릭스와 상기 입력된 사용자 관심 키워드 정보 매트릭스 간의 코사인 유사도를 이용하여 각 전공간 유사도 정보를 생성하도록 더 구성될 수 있다.

[0016] 일 실시예에서, 상기 추천 전공 결정부는, 전공별 특성에 상응하는 복수 개의 차원에 의해 정의되며 전공 간의 유사도가 높을수록 전공들이 서로 가까이 위치하는 전공지도를 이용하여 상기 전공 추천 정보를 상기 사용자에게 제공하도록 더 구성될 수 있다.

[0017] 일 실시예에서, 상기 추천정보 결정부는, 제 2전공 관련 상세 정보 및 추천 활동 정보를 생성하도록 더 구성될 수 있다.

[0018] 본 발명의 일측면에 따른 사용자 관심 키워드에 기반한 전공 추천 방법은,

[0019] 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 강의내용 정보를 생성하는 단계; 상기 생성된 강의내용 정보에 전공을 임베딩하여 전공별 강의내용 정보를 생성하는 단계; 과목명 키워드 정보를 생성하는 단계; 생성된 과목명 키워드 정보에 전공을 임베딩하여 전공별 과목명 키워드 정보를 생성하는 단계; 사용자가 관심 키워드 정보를 입력하는 단계; 상기 전공별 강의 내용 정보, 상기 전공별 과목명 키워드 정보를 이용하여 상기 사용자로부터 입력된 관심 키워드 정보와 각 전공간 유사도를 결정하는 단계; 및 상기 각 유사도 정보에 기초하여 상기 사용자에게 타 학과 전공인 제2 전공에 대한 추천 정보를 생성하는 단계를 포함할 수 있다.

[0020] 일 실시예에 따른 전공 추천 방법은, 상기 강의내용 정보를 생성하는 단계는, 상기 과목의 교수요목 및 학문분야별 출간논문 키워드 중 하나 이상으로부터 상기 텍스트 데이터를 추출하여 데이터를 분석하는 것일 수 있다.

[0021] 일 실시예에서, 상기 강의내용 정보를 생성하는 단계는, 상기 텍스트 데이터에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 과목 각각에 대한 강의내용을 학습하는 단계; 및 상기 강의내용에 대한 학습 결과에 기초하여 상기 강의내용 정보를 생성하는 단계를 포함하고 상기 전공별 강의내용 정보를 생성하는 단계는, 전공별 커리큘럼 정보와 상기 생성된 강의내용 정보를 이용하여 전공별 강의 내용 정보를 생성하는 단계를 포함할 수 있다.

[0022] 일 실시예에서, 상기 과목명 키워드 정보를 생성하는 단계는, 커리큘럼 데이터로부터 필수이수 과목명을 추출하는 단계; 상기 추출된 필수이수 과목명으로부터 텍스트를 추출하는 단계; 상기 추출된 필수이수 과목명과 상기 필수이수 과목명으로부터 추출된 텍스트를 병합하는 단계; 상기 병합된 텍스트에 대한 단어 빈도-역 문서 빈도 기반의 분석에 의하여 상기 병합된 텍스트를 학습하는 단계; 및 상기 병합된 텍스트에 대한 학습 결과에 기초하여 상기 과목명 키워드 정보를 생성하는 단계를 포함할 수 있다.

[0023] 일 실시예에서, 상기 유사도 정보를 생성하는 단계는, 상기 전공별 강의내용 정보 및 상기 전공별 과목명 키워드 정보에 각각 상응하는 매트릭스와 상기 입력된 사용자 관심 키워드 정보 매트릭스 간의 코사인 유사도를 산

출하는 단계를 포함할 수 있다.

[0024] 일 실시예에서, 제 2전공 추천 정보를 생성하는 단계는, 전공별 특성에 상응하는 복수 개의 차원에 의해 정의되며 전공 간의 유사도가 높을수록 전공들이 서로 가까이 위치하는 전공지도를 이용하여 상기 전공 추천 정보를 상기 사용자에게 제공하는 단계를 더 포함할 수 있다.

[0025] 일 실시예에서, 제 2 전공 추천 정보를 생성하는 단계는, 제 2전공 관련 상세 정보 및 추천 활동 정보를 생성하는 단계를 더 포함할 수 있다.

[0026] 본 발명의 일 측면에 따른 컴퓨터 프로그램은 하드웨어와 결합되어 전술한 실시예에 따른 사용자 관심 키워드에 기반한 전공 추천 방법을 실행하기 위한 것으로서 컴퓨터로 판독 가능한 기록매체에 저장된 것일 수 있다.

발명의 효과

[0027] 본 발명에 따른 사용자 관심 키워드 기반 전공 추천 시스템과 방법은 다음과 같은 효과를 제공할 수 있다.

[0028] 본 발명의 일측면에 따르면, 강의내용 차원과 과목명 키워드 정보를 동시에 고려하여 전공 간의 유사도를 판별함으로써 종래 기술, 예컨대, 강의내용 또는 과목명 키워드 정보 중 어느 하나만을 사용하는 모델에 비해 훨씬 높은 정확도로 전공을 추천할 수 있으며, 학생별 관심사에 부합하는 폭 넓고 다양한 전공을 추천함으로써 종래 기술에 비해 추천의 범위와 다양성 측면에서도 더 양호한 결과를 도출할 수 있는 이점이 있다.

[0029] 본 발명의 다른 일측면은, 추천 전공을 전공지도 형태로 제공함으로써 지도상의 거리에 따라 전공 간의 유사도를 시각화하여 사용자에게 제공할 수 있는 이점이 있다.

[0030] 본 발명의 또 다른 일측면은, 사용자가 전공을 선택하면 전공 관련 상세 정보 및 추천 활동을 제공하여 타 학과의 전공에 진입하기 전에 해당 전공을 탐색해볼 수 있는 기회를 제공할 수 있는 이점이 있다.

[0031] 본 발명의 효과들은 이상에서 언급한 효과들로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0032] 도 1은 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 장치를 나타낸 개념도이다.

도 2는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템의 각 단계를 나타내는 순서도이다.

도 3a, 3b 및 3c는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의해 강의내용 정보 및 전공별 강의내용 정보를 생성하는 과정의 개념도이다.

도 4a, 4b 및 4c는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의해 전공별 과목명 키워드 정보를 생성하는 과정의 개념도이다.

도 5는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의해 사용자가 관심 키워드 정보를 입력하는 과정의 개념도이다.

도 6은 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의한 추천 정확도를 재현율과 정밀도 곡선을 통하여 종래 기술과 비교하여 나타낸 그래프이다

도 7a, 7b 및 7c는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의하여 제공되는 전공지도의 예시를 나타내는 개념도이다.

도 8a 및 8b는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의하여 제공되는 전공 관련 상세 정보의 예시를 나타내는 개념도이다.

도 9a 및 9b는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의하여 제공되는 추천 활동 정보의 예시를 나타내는 개념도이다.

발명을 실시하기 위한 구체적인 내용

[0033] 이하에서, 도면을 참조하여 본 발명의 실시예들에 대하여 상세히 살펴본다.

[0034] 도 1은 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 장치를 나타낸 개념도이다.

- [0035] 도 1을 참조하면, 사용자 관심 키워드 기반 전공 추천 시스템(3)은 유선 또는 무선 네트워크를 통하여 수강생의 사용자 장치(1), 교육기관 서버(2) 또는 학술정보 제공 서버(4) 등과 통신하면서 동작하도록 구성될 수 있다. 유선 또는 무선 네트워크를 통한 통신 방법은 객체와 객체가 네트워킹 할 수 있는 모든 통신 방법을 포함할 수 있으며, 유선 통신, 무선 통신, 3G, 4G, 혹은 그 이외의 방법으로 제한되지 않는다.
- [0036] 예를 들어, 유선 또는 무선 네트워크는 LAN(Local Area Network), MAN(Metropolitan Area Network), GSM(Global System for Mobile Network), EDGE(Enhanced Data GSM Environment), HSDPA(High Speed Downlink Packet Access), W-CDMA(Wideband Code Division Multiple Access), CDMA(Code Division Multiple Access), TDMA(Time Division Multiple Access), 블루투스(Bluetooth), 지그비(Zigbee), 와이-파이(Wi-Fi), VoIP(Voice over Internet Protocol), LTE Advanced, IEEE802.16m, WirelessMAN-Advanced, HSPA+, 3GPP Long Term Evolution (LTE), Mobile WiMAX (IEEE 802.16e), UMB (formerly EV-DO Rev. C), Flash-OFDM, iBurst and MBWA (IEEE 802.20) systems, HIPERMAN, Beam-Division Multiple Access (BDMA), Wi-MAX(World Interoperability for Microwave Access) 및 초음파 활용 통신으로 이루어진 군으로부터 선택되는 하나 이상의 통신 방법에 의한 통신 네트워크를 지칭할 수 있으나, 이에 한정되는 것은 아니다.
- [0037] 일 실시예에서, 전공 추천 시스템(3)은 제1 데이터 분석부(32), 제2 데이터 분석부(33), 유사도 결정부(34) 및 추천전공 결정부(36)를 포함한다. 일 실시예에서, 전공 추천 시스템(3)은 전처리부(31)를 더 포함할 수 있다. 또한 일 실시예에서, 전공 추천 시스템(3)은 데이터베이스(database; DB)(35)를 더 포함할 수 있다.
- [0038] 본 명세서에 기재된 시스템, 장치 및 서버는 전적으로 하드웨어이거나, 또는 부분적으로 하드웨어이고 부분적으로 소프트웨어인 측면을 가질 수 있다. 예컨대, 본 명세서의 시스템, 장치 및 서버와, 이들에 포함된 각 부(unit)는 특정 형식 및 내용의 데이터를 처리하거나 또는 전자통신 방식으로 주고받기 위한 하드웨어 및 이에 관련된 소프트웨어를 통칭할 수 있다. 본 명세서에서 "부", "모듈", "장치", "단말기", "서버" 또는 "시스템" 등의 용어는 하드웨어 및 해당 하드웨어에 의해 구동되는 소프트웨어의 조합을 지칭하는 것으로 의도된다. 예를 들어, 하드웨어는 CPU 또는 다른 프로세서(processor)를 포함하는 데이터 처리 기기일 수 있다. 또한, 하드웨어에 의해 구동되는 소프트웨어는 실행중인 프로세스, 객체(object), 실행파일(executable), 실행 스레드(thread of execution), 프로그램(program) 등을 지칭할 수 있다.
- [0039] 또한, 본 실시예에 따른 전공 추천 시스템(3)을 구성하는 각각의 요소는 반드시 서로 물리적으로 구분되는 별개의 장치를 지칭하는 것으로 의도되지 않는다. 즉, 도 1의 입력부(30), 전처리부(31), 제1 데이터 분석부(32), 제2 데이터 분석부(33), 유사도 결정부(34), DB(35) 및 추천전공 결정부(36) 등은 전공 추천 시스템(3)을 구성하는 하드웨어를 해당 하드웨어에 의해 수행되는 동작에 따라 기능적으로 구분한 것일 뿐, 반드시 각각의 부가 서로 독립적으로 구비되어야 하는 것이 아니다. 물론, 실시예에 따라서는 전공 추천 시스템(3)의 각 부 중 하나 이상이 서로 물리적으로 구분되는 별개의 장치로 구현되는 것도 가능하다.
- [0040] 전공 추천 시스템(3)은 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 강의내용 정보를 생성하고, 상기 강의내용 정보에 전공을 임베딩하여 전공별 강의내용 정보를 생성하며, 과목명 키워드 정보를 생성하고, 상기 과목명 키워드 정보에 전공을 임베딩하여 전공별 과목명 키워드 정보를 생성할 수 있다. 또한, 전공 추천 시스템(3)은 입력부(30)로부터 사용자의 관심 키워드 정보를 수신하고, 유사도 결정부(34)는 상기 전공별 강의내용 정보, 상기 전공별 과목명 키워드 정보를 이용하여 상기 사용자로부터 입력된 관심 키워드 정보와 각 전공 간 유사도를 결정하도록 구성될 수 있다. 상기 입력된 관심 키워드 정보와 각 전공 간 유사도 정보를 기초로 전공 추천 정보를 생성하여 사용자 장치(1)에 제공할 수 있다.
- [0041] 사용자가 관심 키워드 정보를 전송하고 전공 추천 정보를 수신하는 것이 가능하도록, 전공 추천 시스템(3)은 사용자 장치(1) 상에서 실행되는 애플리케이션(또는, 앱(app))과 통신함으로써 애플리케이션의 기능 수행을 가능하게 하는 애플리케이션 서비스 서버의 기능을 수행할 수 있다. 또는, 다른 실시예에서 전공 추천 시스템(3)은 사용자 장치(1) 상에서 실행되는 웹 브라우저에 의하여 접속 가능한 웹 페이지를 제공하는 웹 서버 등의 형태로 구현될 수도 있다.
- [0042] 도 1에 도시된 실시예에서 사용자 장치(1)는 노트북 컴퓨터의 형태로 도시되었다. 그러나 이는 예시적인 것으로서, 사용자 장치(1)의 종류는 도면에 도시된 것으로 한정되는 것은 아니다. 예를 들어, 전공을 추천받으려 하는 사용자는 스마트폰(smartphone)과 같은 이동 통신 단말기, 개인용 컴퓨터(personal computer), 노트북 컴퓨터, PDA(personal digital assistant), 태블릿(tablet), IPTV(Internet Protocol Television) 등을 위한 셋톱박스(set-top box) 등 임의의 컴퓨팅 장치를 이용하여 전공 추천 시스템(3)이 제공하는 기능을 사용할 수 있다.

- [0043] 본 발명은 과목에 연관된 텍스트 데이터를 추출하고 추출된 텍스트 데이터를 분석하여 과목별 강의내용 정보를 생성하며 생성된 강의내용 정보에 전공을 임베딩하여 전공별 강의내용 정보를 생성할 수 있다. 생성된 전공별 강의내용은 전공 간의 유사도를 계산하기 위한 기초가 될 수 있다. 이 때, 전처리부(31)는 전공별 강의내용 정보의 생성의 기반이 되는 과목별 강의내용 정보를 생성하기 위한 텍스트 데이터를 추출하기 위한 구성일 수 있다. 전처리부(31)는 과목의 교수요목 또는 학습목표, 학과별 커리큘럼 등으로부터 텍스트 데이터를 추출할 수 있다. 이상의 동작을 위하여, 전처리부(31)는 각 과목의 강의를 제공하는 대학 등 교육기관 서버(2)로부터 정보를 수집하도록 구성될 수 있다. 또한, 전처리부(31)는 강의내용 정보를 생성함에 있어서 과목에 해당하는 학문 분야의 키워드를 결정하기 위하여 학문분야별 출간논문 데이터로부터 텍스트 데이터를 추출할 수 있다. 이를 위하여, 전처리부(31)는 연구 논문 데이터를 제공하는 학술정보 제공 서버(4)와 통신하도록 구성될 수 있다.
- [0044] 제1 데이터 분석부(32)는, 하나 이상의 과목에 연관된 텍스트 데이터를 분석함으로써 각 과목에 대한 강의내용 정보를 생성하고 생성된 강의내용 정보에 기초하여 전공별 강의내용 정보를 생성할 수 있다. 일 실시예에서, 제1 데이터 분석부(32)는 텍스트 데이터에 대한 단어 빈도-역 문서 빈도(Term Frequency-Inverse Document Frequency; TF-IDF) 기반의 분석에 의하여 과목의 강의내용을 학습하고, 강의내용에 대한 학습 결과에 기초하여 강의내용 정보를 생성할 수 있다. 강의내용 정보가 생성되면 생성된 강의내용 정보에 전공을 임베딩하여 전공별 강의내용 정보를 생성할 수 있다. 제1 데이터 분석부(32)의 구체적인 동작에 대해서는 도 3을 참조하여 상세히 후술한다.
- [0045] 제2 데이터 분석부(33)는, 필수이수 과목명 데이터를 분석함으로써 상기 하나 이상의 과목 각각에 대한 과목명 키워드 정보를 생성하고 생성된 과목명 키워드 정보에 기초하여 전공별 과목명 키워드 정보를 생성할 수 있다. 일 실시예에서, 과목명 키워드 정보를 생성하면, 제2 데이터 분석부는 과목명 키워드 정보에 전공을 임베딩하여 전공별 과목명 키워드 정보를 생성할 수 있다. 이 때, 제2 데이터 분석부는 커리큘럼 데이터로부터 필수이수 과목명을 추출할 수 있다.
- [0046] 유사도 결정부(34)는, 제1 데이터 분석부(32)에 의하여 생성된 전공별 강의내용 정보 및 제2 데이터 분석부(33)에 의하여 생성된 전공별 과목명 키워드 정보를 이용하여 입력부로부터 수신된 사용자 관심 키워드 정보와 각 전공간 유사도 정보를 생성할 수 있다. 예를 들어, 유사도 결정부(34)는 전공별 학문영역 데이터가 임베딩(embedding)된 전공별 강의내용 매트릭스, 전공별 과목명 키워드 데이터가 임베딩된 전공별 과목명 키워드 매트릭스 각각과 입력된 사용자 관심 키워드 정보 매트릭스 간의 코사인 유사도를 이용하여 전공별 유사도를 결정할 수 있다. 유사도를 결정하는 구체적인 과정에 대해서는 도 2를 참조하여 상세히 후술한다.
- [0047] 일 실시예에서, 제1 데이터 분석부(32)에 의하여 생성된 강의내용 정보/전공별 강의내용 정보와 제2 데이터 분석부(33)에 의하여 생성된 과목명 키워드 정보/전공별 과목명 키워드정보, 또는 유사도 결정부(34)에 의하여 생성된 유사도 정보는 전공 추천 시스템(3)의 DB(35)에 저장될 수 있다.
- [0048] 추천전공 결정부(36)는 생성된 전공 추천 정보를 사용자 장치(1)에 제공할 수 있다. 이때, 전공 추천 정보는 전공 간의 유사도를 나타내는 과목지도의 형태로 시각화 될 수도 있다. 전공지도는 전공별 특성에 상응하는 복수 개의 차원에 의해 정의되며 전공 간의 유사도가 높을수록 전공들이 서로 가까이 위치하는 평면 또는 입체 지도의 형태를 가질 수 있다. 이때, 제2전공별 학문영역 및 과목명특성을 한정된 수의 차원(예컨대, 2차원)으로 축소하기 위하여 UMAP(Uniform Manifold Approximation and Projection) 알고리즘이 이용될 수도 있다. 전공지도의 형태에 대해서는 도 8a, 8b 및 8c를 참조하여 상세히 후술한다. 이 때, 제2전공이란 사용자가 본인의 전공을 이수하면서 타 학과의 전공을 제2전공으로 선택해 동시에 이수하여 졸업시 2개의 학위를 수여하는 제도를 말한다. 제2전공에는 다른 학과의 전공을 이수하는 이중전공, 두 개 이상의 학과가 참여해 구성된 융합전공, 학생이 자신이 원하는 커리큘럼을 새롭게 구성한 학생설계전공이 있다.
- [0049] 추천전공 결정부(36)는 생성된 제 2 전공 관련 상세 정보 및 제 2 전공 관련 추천 활동 정보를 사용자 장치(1)에 제공할 수 있다. 예를 들어, 사용자가 제 2전공을 이수하기 전에 미리 수강해 볼만한 과목을 추천하는 과목 추천 정보, 제 2전공 관련 도서를 추천해주는 도서 추천 정보일 수 있다. 추천 정보와 추천 활동 정보에 대해서는 도 9a, 9b를 참조하여 상세히 후술한다.
- [0050] 도 2는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템의 각 단계를 나타내는 순서도이다. 실시예들에 따른 전공 추천 방법은, 실시예들에 따른 전공 추천 시스템을 이용하여 수행될 수 있다. 이하에서는, 설명의 편의를 위하여 도 1 및 도 2를 참조하여 본 실시예에 따른 전공 추천 방법에 대하여 설명한다.
- [0051] 본 발명의 실시예들은 전공을 추천받고자 하는 대상 사용자가 관심있는 전공과 유사한 전공을 추천하는 사용자

관심 키워드에 기반한 전공 추천 모델을 이용하여 구현된다. 더 구체적으로는, 실시예들은 각 과목별 데이터를 강의내용 차원과 과목명 차원에서 특성 매트릭스에 임베딩하는 방식으로 과목별 강의내용 정보와 과목명 키워드 정보를 산출하고, 전공을 임베딩하여 전공별 강의내용 정보 및 전공별 과목명 키워드 정보를 생성해 이를 활용하여 전공간 유사도를 산출할 수 있다. 이후, 사용자의 선호 전공과 유사도가 높은 전공을 추천하는 방식으로 전공 추천 정보를 사용자에게 제공할 수 있다.

[0052] 구체적으로, 전공 추천 시스템(3)의 입력부(30)는, 전공을 추천받고자 하는 대상 사용자에게 적합한 추천전공을 결정하기 위하여 대상 사용자의 관심 키워드 정보를 수신할 수 있다. 이때, 관심 키워드 정보란, 사용자가 학문 분야별 출간논문, 교수요목, 과목명에서 추출한 키워드 목록 내에서 입력부(30)을 통해 자유롭게 입력하는 정보일 수 있다. 모든 개설과목이 인문계열, 사회계열, 교육계열, 자연계열, 공학계열, 의학계열, 예체능계열, 융합계열로 구분되는 8개 계열 카테고리에 할당되는데, 각 계열 카테고리별 대표 키워드 목록에서 제공하는 키워드 아이콘을 클릭하여 입력하는 정보일 수 있으며 하나 이상 입력되는 정보일 수 있다. 이 때, 각 계열 카테고리별 대표 키워드 목록은 인문계열, 사회계열, 교육계열, 자연계열, 공학계열, 의학계열, 예체능계열, 융합계열로 구분되는 8개 계열 카테고리에 포함된 모든 개설과목의 학문영역 임베딩 벡터의 평균을 통해 각 계열의 학문영역 임베딩 벡터를 산출하고, 다음으로는 각 교수요목 키워드와 각 계열의 학문영역 벡터 간 코사인 유사도를 산출하여 각 계열과 가장 유사한 교수요목 키워드 200개를 추출해 산출될 수 있다.

[0053] 도 3은 일 실시예에 따른 전공별 강의내용 정보를 생성하는 과정의 개념도이다.

[0054] 전공 추천 시스템(3)의 전처리부(31)가 연구논문 데이터로부터 학문영역별 키워드를 텍스트 데이터 형태로 추출할 수 있다. 또한, 전처리부(31)는 각 과목의 교수요목이나 학습목표, 학과별 커리큘럼 데이터 등 과목정보로부터 형태소 분석을 통하여 키워드에 해당하는 텍스트 데이터를 추출할 수 있다. 또한, 제1 데이터 분석부(32)에 의해 생성되는 각 매트릭스에서 각 키워드의 중요도를 반영하기 위하여 TF-IDF 값이 이용될 수도 있다.

[0055] 추천 대상 전공들에 해당하는 각 과목의 학문영역 특성을 임베딩하기 위하여 연구논문 데이터를 이용할 수 있다. 예컨대, 연구논문 데이터의 각 논문은 통상 5 내지 7개 정도의 키워드를 포함하고 있으며, 투고된 저널의 특성에 따라 통상 1 내지 3개 정도의 학문 분야에 할당되어 있다. 일 실시예에서, 연구논문 데이터는 학술정보 제공 서버(4)로부터 획득될 수도 있다.

[0056] 본 실시예에서는, 논문이 포함된 학문분야에 따라 논문에 포함된 키워드의 출현빈도를 집계하고, 집계된 데이터를 바탕으로 학문분야 키워드의 문서-단어 매트릭스(301)를 구축하였다. 문서-단어 매트릭스(301)의 각 행은 키워드에 대응되며, 각 열은 학문영역에 대응된다. 일 실시예에서는, 기록된 키워드가 일정 횟수 미만인 학문영역, 또는 모든 학문영역을 총합하여 일정 횟수 미만 등장한 키워드는 배제하고 전술한 문서-단어 매트릭스(301)를 구성할 수도 있다. 그러나 이는 예시적인 것으로서, 다른 실시예에서는 모든 학문영역과 모든 키워드를 포함하도록 문서-단어 매트릭스(301)를 구성할 수도 있다.

[0057] 이때, 문서-단어 매트릭스(301)의 각 성분은 연구논문의 키워드와 학문분야의 조합별 TF-IDF 값을 포함할 수 있다. TF-IDF 값은 키워드의 전체 등장 빈도와 특정 학문분야 내 등장 빈도를 고려해 하나의 키워드가 특정 학문분야 내에서 얼마나 중요한 키워드인지 나타내는 수치이며, 학문분야 전반에 걸쳐 높은 빈도로 등장하는 일반적인 키워드의 개별 분야 내 중요성을 적절하게 조절하기 위하여 사용된다. 일 실시예에서는, 전체 키워드-학문분야 조합별 TF-IDF 값이 0 내지 1 사이가 되도록 TF-IDF 값을 정규화할 수도 있다.

[0058] 본 발명은 과목별 강의내용 정보에 전공이 임베딩된 전공별 강의내용 정보를 기초로 전공 추천 정보를 제공하는 발명인바, 과목별 강의내용 정보를 생성하기 위해 과목정보가 이용될 수 있다. 이때, 과목정보란 각 과목의 강의에서 학습하는 내용을 나타내는 것으로서, 예를 들어 두 강의에서 다루는 내용이 학문적으로 유사할수록 두 과목은 서로 유사한 것으로 간주할 수 있다. 본 실시예에서는, 과목의 학습내용을 분석하기 위하여 교수자가 입력한 교수요목 또는 학습목표, 교육기관 서버에서 제공하는 커리큘럼 정보 등을 과목정보로 이용할 수 있다. 일 실시예에서, 교수요목, 학습목표, 커리큘럼 정보 등은 교육기관 서버(2)로부터 획득될 수도 있다. 본 실시예에서는, 먼저 비정형 데이터인 과목정보 데이터에 대한 형태소 분석을 통하여 과목정보 키워드를 추출할 수 있다. 이때, 일 실시예에서는 예컨대 [기말, 페이지, 제출, 공부, 교시, 필요] 등 강의별 교육 내용과 밀접한 관련이 없는 단어들은 미리 설정된 불용어로 분류하여 키워드로부터 배제할 수도 있다. 과목정보 데이터로부터 추출된 키워드들의 매트릭스(302)는 전치(transpose)를 통하여 키워드-과목 매트릭스(303)로 변환될 수 있으며, 키워드-과목 매트릭스(303)의 각 행은 각 과목에 상응하고 각 열은 각 키워드에 상응한다.

[0059] 일 실시예에서는, 학문분야 키워드 매트릭스(301)와 마찬가지로 특정 키워드가 한 과목 내에서 얼마나 많이 등

장했는지(TF)와, 얼마나 많은 과목에서 등장했는지(DF)를 고려해 키워드-과목 쌍별 TF-IDF 값을 산출할 수도 있으며, TF-IDF값이 높을수록 특정 키워드가 해당 과목을 대표하는 중요 키워드일 수 있다는 것을 의미한다. 또한 일 실시예에서, 각 과목정보 키워드의 과목별 TF-IDF 값은 0 내지 1 사이의 값으로 정규화 될 수도 있다.

[0060] 다음으로, 학문분야 키워드 매트릭스(301)와 과목정보 키워드 매트릭스(303)를 이용하여 강의내용 매트릭스(310)를 생성할 수 있다. 이때 강의내용 매트릭스(310)의 각 행은 각각의 과목에 해당하며, 각 열은 각각의 학문영역에 해당된다. 또한, 강의내용 매트릭스(310)의 각 성분은 과목정보 키워드의 학문영역 특성을 TF-IDF 가중치를 적용하여 평균을 구한 값으로 산출될 수 있다. 즉, 과목별로 중요한 키워드의 학문영역 특성이 그 과목의 학문영역 특성을 산출할 때 더 크게 반영된다.

[0061] 예를 들어, 학문분야 m에서 키워드 i의 TF-IDF 값을 하기 수학적 식 1의 F로 나타내고, 과목정보 j에서 키워드 i의 TF-IDF 값을 하기 수학적 식 2의 C로 나타내며, 과목정보 j의 키워드 수를 N_j 로 나타낼 경우, 학문분야 m에 속하는 과목정보 j의 가중치가 반영된 TF-IDF 값 F'은 하기 수학적 식 3과 같이 산출될 수 있다.

수학적 식 1

$$F = K = [k_{i,m}]$$

[0062]

수학적 식 2

$$C = [c_{i,j}]$$

[0063]

수학적 식 3

$$F' = [f'_{j,m}] = \frac{(C^T \times K)}{N_j}$$

[0064]

[0065] 다음으로, 제 1데이터 분석부는 전공별 커리큘럼에 포함된 모든 과목의 학문영역 벡터의 평균으로 전공을 학문분야 차원에 임베딩할 수 있다. 구체적으로, 전공별 커리큘럼 매트릭스(400)와 강의내용 매트릭스(310)을 이용하여 전공별 강의내용 매트릭스(410)를 생성할 수 있다. 이 때, 전공별 커리큘럼 매트릭스의 각 행은 복수의 전공들에 해당하며, 각 열은 각 전공의 커리큘럼에 해당된다. 전공별 강의내용 매트릭스의 각 행은 복수의 전공들에 해당하며, 각 열은 각각의 학문영역에 해당된다.

[0066] 도 4는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의해 전공별 과목명 키워드 정보를 생성하는 과정의 개념도이다.

[0067] 도 4를 참조하면, 제 2 데이터 분석부(33)는 과목명 키워드 정보를 생성하고 상기 과목명 키워드 정보에 전공을 임베딩하여 전공별 과목명 키워드 정보를 생성할 수 있으며, 이 때 전공별 과목명 키워드 정보는 전공별 과목명 키워드 매트릭스를 의미할 수 있다.

[0068] 도 4a는 필수 이수 과목명 리스트를 추출한 개념도이며, 도 4b는 필수 이수 과목명 리스트로부터 추출한 텍스트와 필수 이수 과목명 리스트를 병합하는 과정의 개념도이다.

[0069] 먼저, 제 2데이터 분석부(33)는 커리큘럼 데이터에서 필수 이수 과목명 리스트를 추출하고, 추출된 필수 이수 과목명 리스트로부터 텍스트를 추출하며, 상기 필수이수 과목명과 상기 필수이수 과목명으로부터 추출된 텍스트를 병합할 수 있다. 예를 들면, "19세기 유럽 사회사" 라는 과목명에서는 "19세기", "세기", "유럽", "유럽사회", "유럽사회사", "사회사", "사회" 라는 텍스트가 추출될 수 있다.

[0070] 일 실시예에서는, 과목명 키워드가 한 전공 내에서 얼마나 많이 등장했는지(TF)와, 얼마나 많은 전공에서 등장

했는지(DF)를 고려해 전공-과목명 키워드 쌍별 TF-IDF 값을 산출할 수 있으며, TF-IDF값이 높을수록 특정 과목 키워드가 특정 전공 커리큘럼을 대표하는 중요 키워드일 수 있다는 것을 의미한다. 또한 일 실시예에서, 전공-과목명 키워드 조합별 TF-IDF 값은 0 내지 1 사이의 값으로 정규화될 수도 있다.

[0071] 다음으로, 전공-과목명 키워드 조합별 TF-IDF를 0 내지 1사이로 정규화한 값을 임베딩하여 전공별 과목명 키워드 매트릭스(500)를 생성할 수 있다. 이때 전공별 과목명 키워드 매트릭스의 각 행은 전공에 해당하며, 각 열은 과목명 키워드에 해당된다.

[0072] 전공 추천 시스템(3)의 유사도 결정부(34)는, 전술한 것과 같이 생성된 전공별 강의내용 매트릭스(410)와 전공별 과목명 키워드 매트릭스(500)를 이용하여 사용자가 입력한 관심 키워드와 각 전공간 유사도를 산출할 수 있다(170). 일 실시예에서, 각 전공간 유사도 정보는 전공별 강의내용 매트릭스와 전공별 과목명 키워드 매트릭스 각각을 기반으로 하여 입력된 사용자 관심 키워드 정보간에 코사인 유사도에 의하여 산출될 수도 있다. 예를 들어, 사용자로부터 입력된 관심 키워드 정보와 상기 입력된 관심 키워드 정보(또는 전공별 과목명 키워드 매트릭스)를 각각 A,B라고 할 경우, 두 매트릭스 사이의 코사인 유사도는 하기 수학적 식 4와 같이 산출될 수 있다.

수학적 식 4

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

[0073]

[0074] 이 때, 상기 코사인 유사도로부터 산출된 사용자로부터 입력된 관심 키워드와의 평균 거리가 짧은 순서대로 제2 전공 추천목록을 산출할 수 있다.

[0075] 도 5는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의해 사용자가 계열 카테고리별 대표 키워드 목록화면에서 관심 키워드 정보를 입력하는 과정의 개념도이다.

[0076] 추천전공 결정부(36)는 이상과 같이 결정된 추천전공을 포함하는 전공 추천 정보를 사용자 장치(1)에 제공할 수 있다(200). 일 실시예에서, 추천전공 결정부(36)는 추천 제2전공 리스트와 그 순위가 출력될 수 있다.

[0077] 본 발명자들은, 이상에서 설명한 본 발명의 실시예들에 따른 전공 추천 모델의 성능을 검증하기 위해 추천 모델이 추천한 제2전공과 학생들이 실제 예상한 제2전공 결과를 비교하였다. 이를 위해 각 전공별로 할당된 학생을 모집해 관심 키워드 기반 제2전공 추천 결과 테스트 데이터 세트를 구축했다. 학생들은 자신의 관심 키워드 목록과 이 키워드를 입력했을 때 예상되는 제2전공 추천 결과를 함께 입력했다. 테스트 데이터 세트는 총 1,761건이었다. 학생들이 한 번에 입력한 키워드의 평균 개수와 표준편차는 각각 2.07, 0.96이었으며, 예상되는 추천 제2전공의 평균 개수와 표준편차는 3.55, 1.82였다.

[0078] 성능의 검증에는 추천 모델 성능 검증에 널리 활용되는 정밀도(precision rate)와 재현율(recall rate)을 사용하였다. 정밀도는 추천 제 2전공 수 대비 제 2전공 추천 성공 횟수를 통해 [정밀도 = (제 2전공 추천 성공 횟수) / (추천 제2전공 수)]로 산출하였으며, 이때 추천 성공은 추천 성공은 추천된 제2전공이 학생이 예상한 제2전공 목록에 포함되어 있는 경우로 정의하였다. 또한 재현율은 예상 제2전공 수에서 제2전공 추천 성공 횟수의 비율을 통해 [재현율 = (제2전공 추천 성공 횟수) / (예상 제2전공 수)]로 산출하였다.

[0079] 도 6은 사용자 관심 키워드 기반 전공 추천 시스템에 의한 추천 정확도를 재현율과 정밀도 곡선을 통하여 종래 기술과 비교하여 나타낸 그래프이다.

[0080] 본 실시예와 대비되는 가상의 종래 기술은 활용한 데이터에 따라 "교수요목 키워드" 만을 사용하는 모델 및 "과목명 키워드" 만을 사용하는 모델로 구분하였다.

[0081] 도 6에 도시되는 바와 같이, 본 실시예에 따른 전공 추천 방법이 "교수요목 키워드" 또는 "과목명 키워드" 중 어느 하나만을 사용하는 모델에 비해 성능이 우수하였다. 또한, 추천 제2전공 수에 따라 정밀도와 재현율 사이의 트레이드 오프(trade-off)가 확인되었다. 즉, 많은 제2전공을 추천할수록 추천된 제2전공 중에서 학생이 실제로 예상한 제2전공이 있을 확률이 높아지지만, 추천된 제2전공이 학생이 예상한 제2전공일 확률이 낮았다. 이에 기초하여, 추천 제 2전공의 수를 1개에서부터 10개까지 변화해 가며 추천 모델의 성능을 평가하였다. 1개의

제2전공을 추천하는 경우, 추천된 제2전공이 이 학생이 예상한 결과 중 하나일 확률은 약 67%였으며, 10개의 제2전공을 추천하는 경우, 학생이 예상한 제2전공 목록의 약 72%가 추천된 제2전공 목록에 포함되어 있었다. 추천 제2전공의 수를 1개에서부터 10개까지 변화시켜 본 발명을 구현할 수 있으나, 이는 예시적인 것으로서, 추천 제2전공의 수는 실시예들에 따라 상이할 수 있다.

[0082] 도 7a, 7b 및 7c는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의하여 제공되는 전공지도의 예시를 나타내는 개념도이다.

[0083] 도 7a, 7b 및 7c에 도시된 제2전공 지도는, 제2전공별 강의내용 매트릭스 및 제2전공별 과목명 키워드 매트릭스를 McInnes, L., Healy, J. 및 Melville, J. 공저의 논문 "Umap: Uniform manifold approximation and projection for dimension reduction" (arXiv preprint arXiv: 1802.03426, 2018)에 개시된 UMAP 알고리즘을 이용하여 2차원으로 축소한 후 시각화한 결과물을 나타내는 것으로서, 제2전공 지도 상에서 거리가 가까울수록 전공 간의 특성이 유사하며, 멀리 떨어져 있을수록 전공 간의 특성이 상이하다는 것을 의미한다.

[0084] 도 7a는 전공지도에서 제2전공 주관 대학에 따라 노드(node)가 서로 상이한 해칭을 갖도록 구분하여 나타낸 것으로서, 동일한 대학에서 주관한 과목이 서로 인접하여 위치하는 것을 통하여 본 발명의 실시예에 따른 전공 추천 방법에 의한 전공간 유사도 계산이 정확하다는 것을 알 수 있다. 예를 들어, 우측 상단에 위치한 수학교육과와 수학과는 주관 대학이 각각 사범대학과 이과대학으로 다르지만, 서로 가깝게 위치해 있다. 또, 통계학 지식과 수학적 지식을 주로 다루게 되는 금융공학융합전공은 두 전공과 모두 인접해 위치해 있다. 도 7a에서는 흑백 지도상에서 해칭을 통하여 각 노드를 구분하였으나, 과목지도를 컬러로 제공하는 경우 각 노드의 색상을 통하여 제2전공 주관 대학을 구분할 수도 있다.

[0085] 또한, 도 7b는 학문영역 특성 데이터를 활용하여 특정 학문 영역의 특성이 강한 제2전공을 밝게 나타낸 것으로, 경영학과, 금융공학융합전공, 보험과 위험 관리 학생설계전공 등 회계학(accounting) 특성이 높은 전공을 밝게 나타낼 수 있다. 그러나, 그 외 다른 학문영역에 대해서도 동일한 방식의 시각화를 적용할 수 있다. 또한, 다른 실시예에서는 학문영역이 아닌 과목명 키워드 매트릭스에 기초하여 노드의 밝기를 구분함으로써, 전공지도를 구성하는 것도 가능하다. 이와 같이 UMAP 알고리즘을 사용하여 차원이 축소된 전공지도 형태로 제2전공추천 정보를 제공하는 경우 전역 구조(global structure)와 위상적 구조(topological structure)를 저차원에서도 효과적으로 포착할 수 있도록 정보의 제공이 가능한 이점이 있다.

[0086] 도 7c는 대학교에서 개설된 전공들을 전체적으로 조망할 수 있도록 마련된 제2전공 지도 서비스 화면으로서, 서로 인접하여 위치한 전공일수록 유사하고 멀리 위치한 전공일수록 이질적인 전공일 수 있다.

[0087] 도 8a 및 8b는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의하여 제공되는 전공 관련 상세 정보의 예시를 나타내는 개념도이다. 사용자가 제2전공 지도에서 관심있는 제2전공을 클릭하거나 또는 계열 카테고리별 대표 키워드 목록이 마련되어 있는 관심 키워드 입력 화면에 키워드를 입력하면, 추천정보 결정부는 추천된 제2전공 결과값을 제공할 수 있다. 그 다음, 사용자가 추천된 제2전공을 클릭하면, 제2전공 기본정보, 교과과정표, 제1전공 구성비율, 타전공/사용자 전공과의 유사도, 제2전공생 희망 진로 유형, 졸업생 진로유형의 상세정보를 제공할 수 있다. 이 때, 사용자의 전공과 타 전공과의 유사도는 코사인 유사도를 이용해 산출할 수 있다.

[0088] 도 9a 및 9b는 일 실시예에 따른 사용자 관심 키워드 기반 전공 추천 시스템에 의하여 제공되는 추천 활동 정보의 예시를 나타내는 개념도이다. 제2전공 관련 추천 활동 정보는 과목 추천과 도서 추천일 수 있다.

[0089] 과목 추천은 사용자가 관심있는 제2전공을 이수하기 전에 해당 제2전공의 특성을 탐색하기 위해 미리 수강해 볼만한 과목을 추천해주는 것이다. 이를 위해, 제2전공을 이수하는 학생들이 해당 전공에 진입하기 전 수강한 빈도가 가장 높은 커리큘럼 내 전공과목을 순서대로 추출한다. 제2전공 관련 과목추천은 교양과목을 추천(이하, 맛보기 교양과목)해주는 것과 해당 제2전공을 이수하기 위해 수강해야 할 전공과목 추천(이하, 맛보기 교과과정)해주는 것으로 구분된다. 맛보기 교양과목은 제2전공을 이수한 학생이 해당 전공을 이수하지 않는 학생에 비해 가장 많이 수강하는 과목을 통해 산출한 과목이며 이때, 맛보기 교양과목의 교양과목은 해당 제2전공 커리큘럼에 포함되지 않은 교양과목을 의미한다. 이 때, 맛보기 교양과목, 맛보기 교과과정은 제2전공-과목 쌍의 TF-IDF 값을 산출해 제2전공별로 중요도가 높은 과목을 추천해 주는 것일 수 있다.

[0090] 도서 추천은 사용자가 관심있는 제2전공과 관련된 도서를 추천해주는 것이다. 이를 위해, 해당 제2전공을 이수한 학생들의 도서대출이력과 강의정보 데이터를 활용할 수 있다. 해당 제2전공을 이수한 학생들이 많은 비중으로 대출하는 도서를 도서대출이력-강의정보 데이터 쌍의 TF-IDF 값을 산출해 추천해주는 것일 수 있다. 추천도

서는 제 2전공 커리큘럼에 포함된 강의에 지정된 교과서 및 참고도서 목록에 관한 것일 수 있다. 이 때, 참고도서는 특정 학기 해당 제 2전공 교과과정 강의에서 참고도서로 지정한 도서를 의미할 수 있다.

[0091] 한편, 이상에서 설명한 전공별 상세정보 및 추천활동 정보는 지난 10년간의 데이터를 통계 분석하여 제공한 것일 수 있다.

[0092] 한편, 이상에서 설명한 실시예들에 따른 사용자가 입력한 관심 키워드 정보와 각 전공간 유사도 정보에 기반한 전공 추천 방법에 의한 동작은 적어도 부분적으로 컴퓨터 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 실시예들에 따른 전공별 유사도에 기반한 전공 추천 방법에 의한 동작을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터에 의하여 읽혀 질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 또한, 본 실시예를 구현하기 위한 기능적인 프로그램, 코드 및 코드 세그먼트(segment)들은 본 실시예가 속하는 기술 분야의 통상의 기술자에 의해 용이하게 이해될 수 있을 것이다.

[0093] 이상에서 살펴본 본 발명은 도면에 도시된 실시예들을 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시예의 변형이 가능하다는 점을 이해할 것이다. 그러나, 이와 같은 변형은 본 발명의 기술적 보호범위 내에 있다고 보아야 한다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

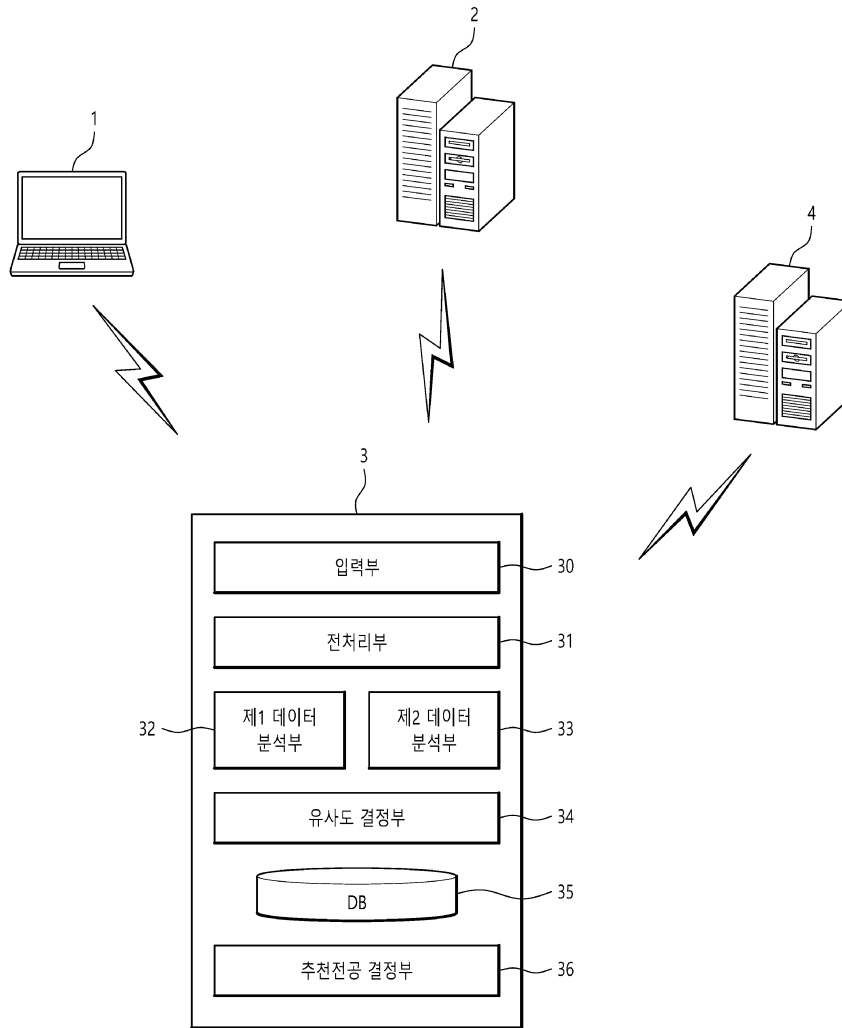
부호의 설명

- [0094]
- 1: 사용자 장치
 - 3: 전공 추천 시스템
 - 30: 입력부
 - 31: 전처리부
 - 32: 제1 데이터 분석부
 - 33: 제2 데이터 분석부
 - 34: 유사도 결정부
 - 35: DB
 - 36: 추천전공 결정부
 - 100: 학문분야 키워드 분석 단계
 - 110: 과목정보 키워드 분석 단계
 - 120: 강의내용 정보 생성 단계
 - 130: 전공별 강의내용 정보 생성 단계
 - 140: 전공별 필수이수 과목명 추출 단계
 - 150: 과목명 키워드 정보 생성 단계
 - 160: 전공별 과목명 키워드 정보 생성 단계
 - 170: 유사도 산출 단계
 - 180: 추천 전공 결정 단계
 - 190: 전공 추천 정보 제공 단계
 - 301: 학문분야 키워드의 문서-단어 매트릭스
 - 302: 키워드 매트릭스

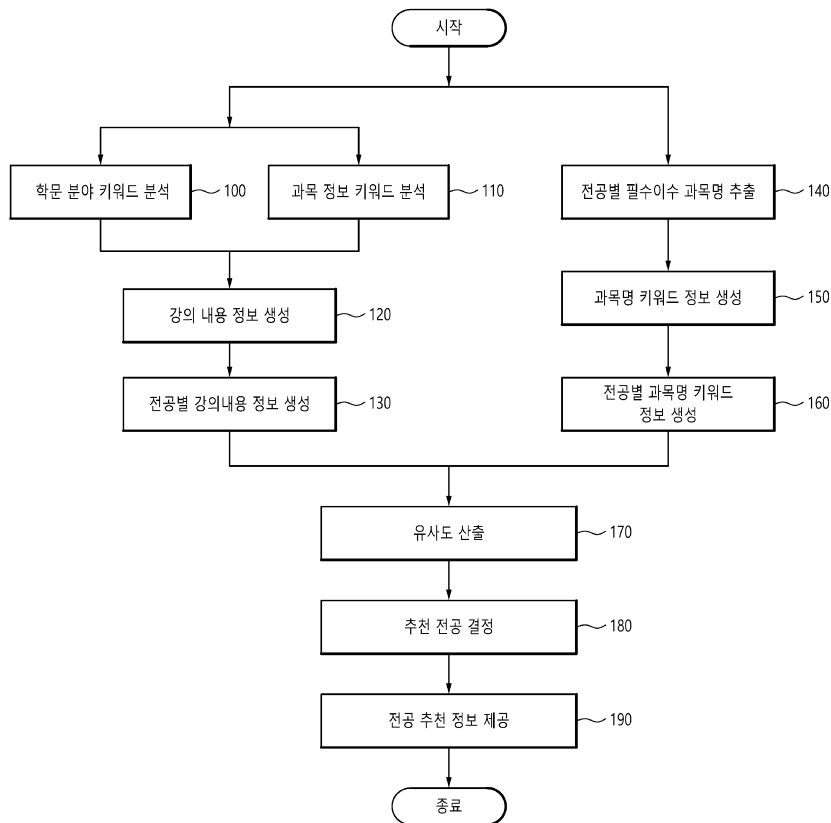
- 303: 과목정보 키워드 매트릭스
- 310: 강의내용 매트릭스
- 400: 전공별 커리큘럼 매트릭스
- 410: 전공별 강의내용 매트릭스
- 500: 전공별 과목명 키워드 매트릭스

도면

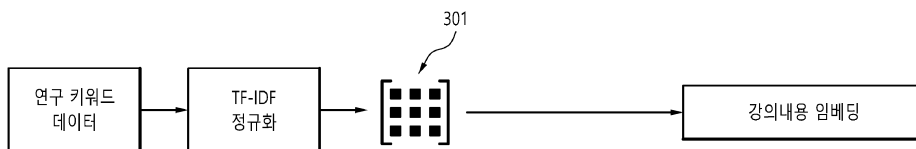
도면1



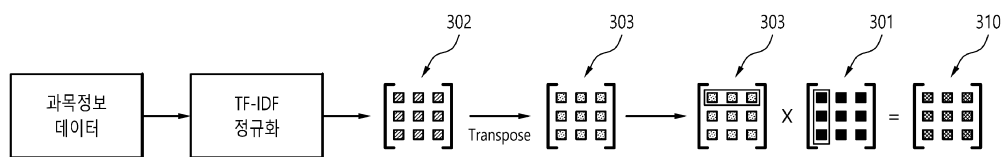
도면2



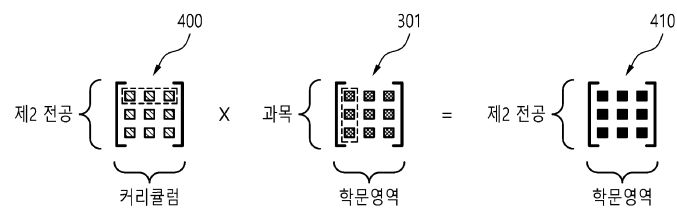
도면3a



도면3b



도면3c



도면4a

	index	mmajor_cd	cour_cd	cour_nm
0	25	0142	BUSS246	경영과학
1	602	0142	BUSS203	경영수학
2	657	0142	JURA158	기업법 I
3	1367	0142	ECON202	거시경제이론
4	1621	0142	ECON203	경제원론II
...	...	...	...	...
126	1746	0142	BUSS311	조직행동론
127	1747	0142	BUSS213	중급회계 I
128	17545	0142	BUSS164	경영과사회
129	17580	0142	BUSS410	기술혁신관리·성공적Entrepreneurship의 기초
130	17631	0142	BUSS450	재무금융론신조류

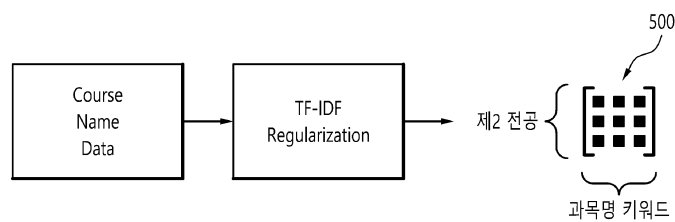
131 rows x 4 columns

도면4b

	index	mmajor_cd	cour_cd	cour_nm	keyword
0	25	0142	BUSS246	경영과학	경영
1	25	0142	BUSS246	경영과학	과학
2	602	0142	BUSS203	경영수학	경영
3	602	0142	BUSS203	경영수학	경영수학
4	602	0142	BUSS203	경영수학	수
...	...	...	...	...	...
1004	17631	0142	BUSS450	재무금융론신조류	금융
1005	17631	0142	BUSS450	재무금융론신조류	금융론
1006	17631	0142	BUSS450	재무금융론신조류	신조류
1007	17631	0142	BUSS450	재무금융론신조류	재무
1008	17631	0142	BUSS450	재무금융론신조류	조류

541 rows x 6 columns

도면4c



도면5

키워드로 알아보는 나만의 제2전공 추천!

관심있는 키워드를 검색해보세요 (최대 5개까지 입력 가능)

ALL

전체

인문계열

인문계열

사회계열

사회계열

교육계열

교육계열

자연계열

자연계열

공학계열

공학계열

의약계열

의약계열

예체능계열

예체능계열

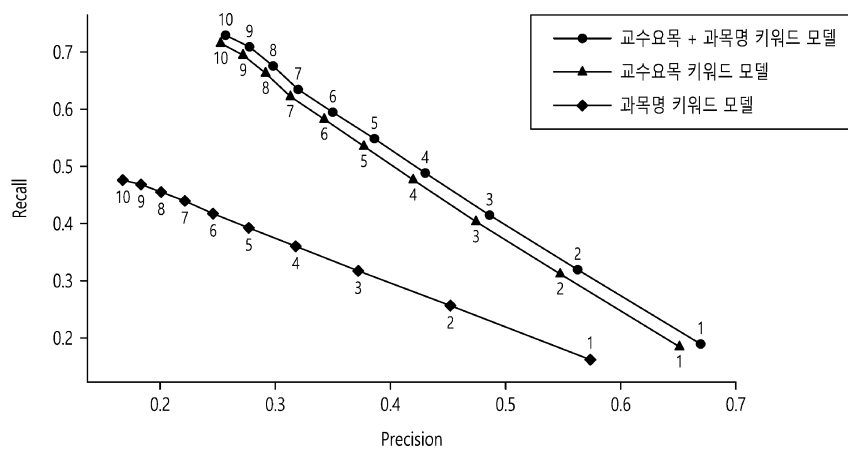
융합계열

융합계열

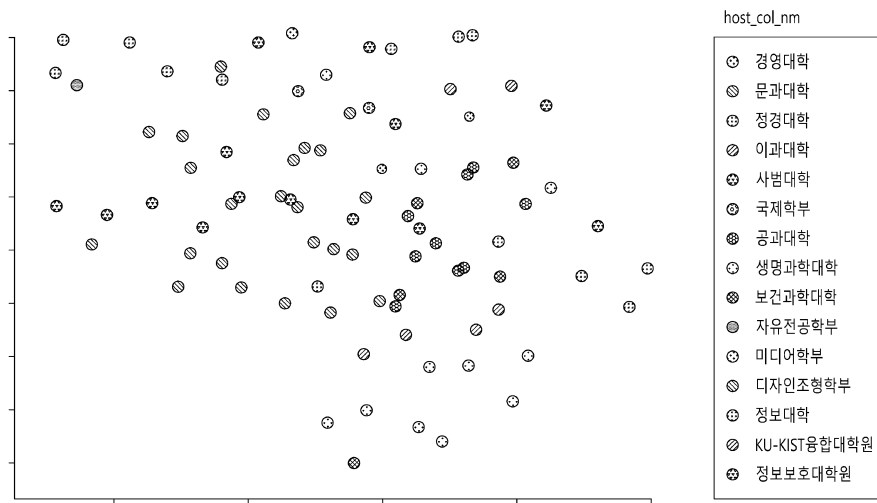
인기키워드
상세키워드

경제 (85) 경영 (73) 역사 (68) 마케팅 (64) 통계 (62) 환경 (60) 언어 (59) 교육 (52) 빅데이터 (51) 문학 (48) 컴퓨터 (45)
금융 (43) 수학 (40) 정치 (40) 법 (37) 심리 (37) 영화 (37) 인공지능 (36) 사회 (35) 의료 (34) 문화 (33) 코딩 (29)
글쓰기 (28) 반도체 (27) 외교 (27) 패션 (27) 디자인 (26) 상담 (26) 글로벌 (25) 뇌 (25) 다문화 (24) 창업 (24) 기계 (23)
보건 (23) 분석 (22) 예술 (22) 주식 (22) 회계 (22) 건강 (21) 식품 (21) 정책 (21) 철학 (21) 공학 (20) 광고 (20) 논리 (19)
관계 (18) 행정 (18) 3D (17) 기술 (17) 기업 (17)

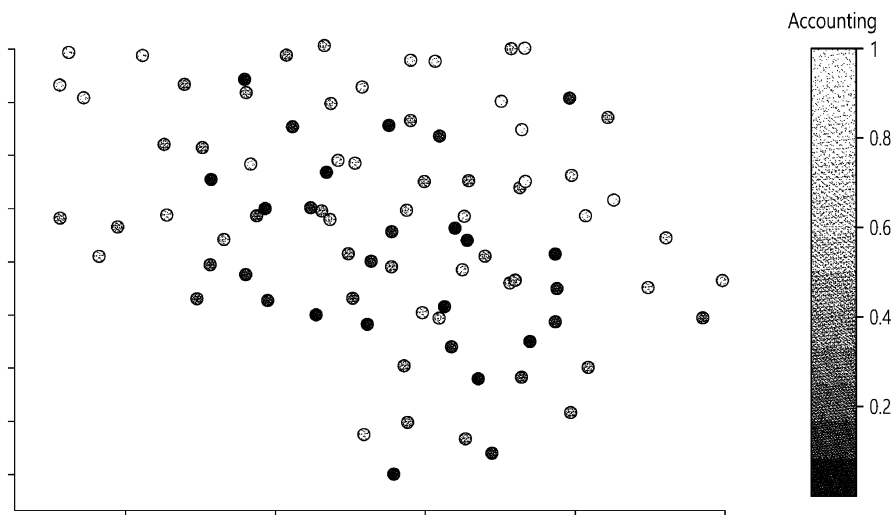
도면6



도면7a



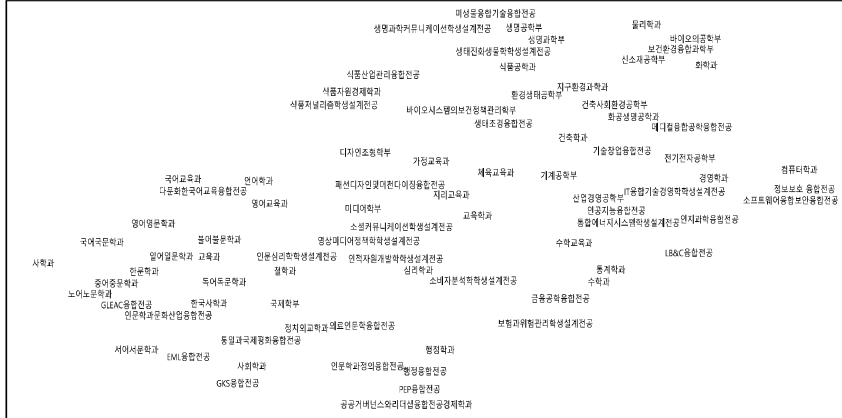
도면7b



도면7c

제2전공 전체를 조망하고 싶을 때!

탐색할 이중전공이나 융합전공을 선택하세요.

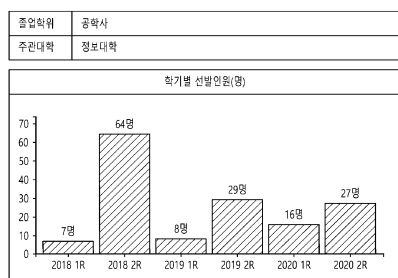


도면 8a

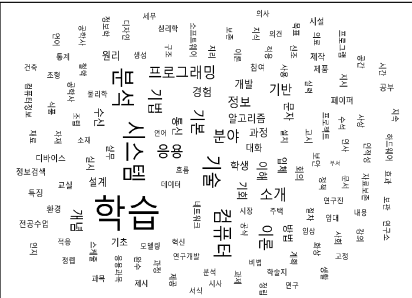
제2전공 상세 정보 A-Z !

컴퓨터학과 ▼ 이증전공

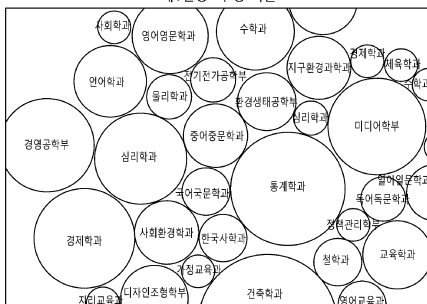
제2전공 기본정보



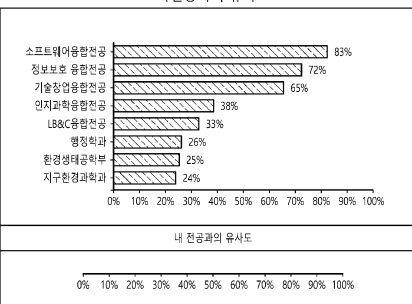
교육과정 워드클라우드



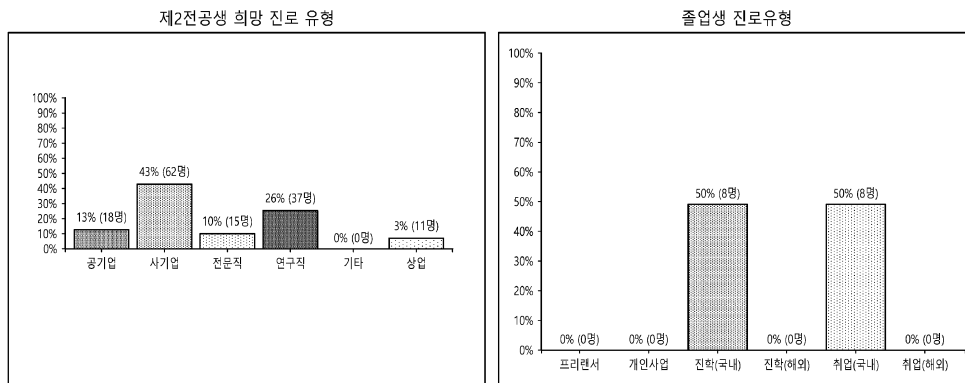
제1전공 구성 비율



타전공과의 유사도



도면8b



도면9a

관심있는 제2전공을 이수하기 전에 어떤 활동을 하면 좋을까?

GLEAC융합전공

"GLEAC융합전공" 을(를) 이수하기 전에 이런 활동을 하면 좋아요.)

② 맞보기 교양과목

1 GESO143 중국의대외정책과동아시아 국제관계 □ 사회의이해 ☞ 3학점	2 GEFC090 문학으로본동아시아문명 □ 세계와문화 ☞ 3학점
3 CHIN105 중국문학산책 □ 선택교양 ☞ 3학점	4 KORE119 한국문화와세계문학 □ 선택교양 ☞ 3학점
5 GEFC016 그림을통해본일본고전문학과문화 □ 세계와문화 ☞ 3학점	6 GEHI080 한국문화유산의이해 □ 역사와탐구 ☞ 3학점

② 맞보기 교과과정

1 GLEA201 동아시아학의이해 □ 전공필수 ☞ 3학점	2 GLEA301 동아시아비교문학 □ 전공선택 ☞ 3학점
3 KORE201 한국문학의이해 □ 전공필수 ☞ 3학점	4 GLEA202 동아시아시대외리더십 □ 전공필수 ☞ 3학점
5 HOKA211 한국문화의기원 □ 전공선택 ☞ 3학점	6 CHIN241 현대중국사회의이해 □ 전공선택 ☞ 3학점

도면9b

