



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0052320
(43) 공개일자 2019년05월16일

(51) 국제특허분류(Int. Cl.)
G06F 16/00 (2019.01) G06Q 50/10 (2012.01)
G06Q 50/20 (2012.01)
(52) CPC특허분류
G06F 16/9535 (2019.01)
G06Q 50/10 (2013.01)
(21) 출원번호 10-2017-0147920
(22) 출원일자 2017년11월08일
심사청구일자 2017년11월08일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
황인준
서울특별시 성북구 낙산길 255, 105동 1504호(보문동3가, 이편한세상보문)
김용성
서울특별시 노원구 동일로227길 26, 1509동 203호(상계동, 상계주공15단지아파트)
(74) 대리인
송인호, 윤형근, 최영중, 최관락

전체 청구항 수 : 총 10 항

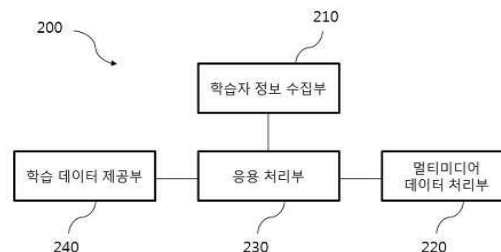
(54) 발명의 명칭 개인화된 콘텐츠 제공 장치

(57) 요약

본 발명은 IoT 기반의 개인화된 학습 콘텐츠를 제공할 수 있는 개인화된 콘텐츠 제공 장치에 관한 것이다.

본 발명의 일실시예에 의한 콘텐츠 제공 장치는 학습자 단말기로부터 사전 테스트 결과를 포함하는 학습자 정보를 수집하는 학습자 정보 수집부; 트위터 뉴스 및 멀티미디어 콘텐츠를 수집하고, 시맨틱 데이터(Semantic Data)와 비시맨틱 데이터(Non-Semantic Data)를 통합하는 멀티미디어 데이터 처리부; 상기 학습자 정보에 근거하여 학습자의 수준을 판단하고, 상기 학습자 단말기로부터 수신한 쿼리 결과를 계산하는 응용 처리부; 및 상기 계산된 쿼리 결과에 근거하여 학습 데이터를 상기 학습자 단말기로 제공하는 학습 데이터 제공부를 포함할 수 있다.

대표도 - 도2



(52) CPC특허분류

G06Q 50/20 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 20150005900032002

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 정보통신방송연구개발사업

연구과제명 빅데이터 처리 고도화 핵심기술개발 사업 총괄 및 고성능 컴퓨팅 기술을 활용한 성능 가속
화 기술 개발

기 여 율 1/1

주관기관 울산과학기술원

연구기간 2017.06.01 ~ 2018.03.31

명세서

청구범위

청구항 1

학습자 단말기로부터 사전 테스트 결과를 포함하는 학습자 정보를 수집하는 학습자 정보 수집부;

트위터 뉴스 및 멀티미디어 콘텐츠를 수집하고, 시맨틱 데이터(Semantic Data)와 비시맨틱 데이터(Non-Semantic Data)를 통합하는 멀티미디어 데이터 처리부;

상기 학습자 정보에 근거하여 학습자의 수준을 판단하고, 상기 학습자 단말기로부터 수신한 쿼리 결과를 계산하는 응용 처리부; 및

상기 계산된 쿼리 결과에 근거하여 학습 데이터를 상기 학습자 단말기로 제공하는 학습 데이터 제공부를 포함하는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 2

제1항에 있어서, 상기 멀티 미디어 데이터 처리부는

시맨틱 데이터를 처리하는 시맨틱 데이터 처리부; 및

비시맨틱 데이터를 처리하는 비시맨틱 데이터 처리부를 포함하는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 3

제2항에 있어서, 상기 비시맨틱 데이터 처리부는

컨벌루션 필터를 이용하여 상기 수집된 트위터 뉴스의 주제를 분류하는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 4

제2항에 있어서, 상기 비시맨틱 데이터 처리부는

비디오 리소스의 메타데이터를 추출하는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 5

제2항에 있어서, 상기 비시맨틱 데이터 처리부는

상기 비디오 리소스의 스크립트를 수집하고, 상기 수집된 스크립트로부터 부족한 메타데이터를 보강하는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 6

제2항에 있어서, 상기 비시맨틱 데이터 처리부는

상기 수집된 스크립트로부터 지문 및 감탄사 중 어느 하나를 제거하는 전처리를 수행하는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 7

제5항에 있어서, 상기 비시맨틱 데이터 처리부는

단어의 길이, 문장의 길이 및 단어 사용의 복잡성 중 적어도 하나를 이용하여 상기 수집된 스크립트의 가독성을 계산하는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 8

제5항에 있어서, 상기 시맨틱 데이터 처리부는

상기 비시맨틱 데이터 처리부에서 추출되거나 보강된 메타데이터를 시맨틱 데이터 형식으로 변환하는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 9

제8항에 있어서, 상기 시맨틱 데이터 처리부는

상기 변환된 시맨틱 데이터를 외부 링크드 데이터(Linked Data)와 인터링크 시키는 것을 특징으로 하는 콘텐츠 제공 장치.

청구항 10

제9항에 있어서, 상기 시맨틱 데이터 처리부는

상기 인터링크된 시맨틱 데이터와 외부 링크드 데이터를 최종 메타데이터 세트로 구성하는 것을 특징으로 하는 콘텐츠 제공 장치.

발명의 설명

기술 분야

[0001] 본 발명은 개인화된 콘텐츠 제공 장치에 관한 것으로, 보다 상세하게는 IoT 기반의 개인화된 학습 콘텐츠를 제공할 수 있는 개인화된 콘텐츠 제공 장치에 관한 것이다.

배경 기술

[0002] 최근 IoT(Internet of Things)라는 용어가 이슈가 되고 있으며 이와 관련된 많은 연구가 이루어지고 있다. 이런 IoT는 Transportation, Healthcare, agriculture, electricity 등 다양한 분야에서 활용이 되고 있다.

[0003] 최근 교육 분야에도 IoT가 점점 많이 활용되고 있는 추세이다. 지금까지의 교육 분야에서는 e-learning, m-learning, u-learning과 같은 다양한 컴퓨터 기반의 학습 방법이 주목 받아왔으나, 최근 IoT라는 용어가 각광 받음으로 인해서 기존의 교육방법들은 IoT를 기반으로 하는 방법으로 새롭게 변화하고 있다. 이러한 IoT 기반의 교육은 학습자가 학습 성과를 이루기 위해서 "Things"와 상호작용 하는 것이다. 여기서 말하는 "Things"는 물리적인 객체(e.g 다양한 디바이스)뿐만 아니라 가상의 객체(e.g 웹의 콘텐츠)도 포함한다. 이러한 상호작용을 통해 학습자는 수업에 흥미를 갖고 적극적으로 참여를 할 수 있게 되고, 학습할 이슈에 대한 이해를 높여 학습 성과를 향상시킬 수 있다. 관련된 선행문헌으로 대한민국 등록특허 제10-1631671호가 있다.

[0004] 따라서 신문(뉴스)과 같은 미디어를 교육에 활용하는 Newspaper in Education (NIE) 수업에 IoT 기술을 활용하는 방법에 대한 연구가 필요한 실정이다.

발명의 내용

해결하려는 과제

[0005] 본 발명의 목적은 학습자들 수준에 적합한 온라인 뉴스 및 동영상 학습 자료들을 IoT 기반의 기술을 통해 학습자에게 제공할 수 있는 개인화된 콘텐츠 제공 장치를 제공하는 데 있다.

과제의 해결 수단

[0006] 상기 목적을 달성하기 위해 본 발명의 일실시예에 의하면, 학습자 단말기로부터 사전 테스트 결과를 포함하는 학습자 정보를 수집하는 학습자 정보 수집부; 트위터 뉴스 및 멀티미디어 콘텐츠를 수집하고, 시맨틱 데이터(Semantic Data)와 비시맨틱 데이터(Non-Semantic Data)를 통합하는 멀티미디어 데이터 처리부; 상기 학습자 정보에 근거하여 학습자의 수준을 판단하고, 상기 학습자 단말기로부터 수신한 쿼리 결과를 계산하는 응용 처리부; 및 상기 계산된 쿼리 결과에 근거하여 학습 데이터를 상기 학습자 단말기로 제공하는 학습 데이터 제공부를 포함하는 개인화된 콘텐츠 제공 장치가 개시된다.

발명의 효과

[0007] 본 발명의 일실시예에 의한 개인화된 콘텐츠 제공 장치는 학습자 정보를 통해 이에 적합한 IoT 기반의 맞춤형

학습 콘텐츠를 제공할 수 있다.

- [0008] 본 발명의 일실시예에 의하면, 비 시맨틱(non-semantic) 데이터들(트위터 뉴스, 관련 동영상 등)뿐만 아니라 시맨틱(semantic) 데이터들도 학습자가 별다른 노력 없이 접근할 수 있다.
- [0009] 본 발명의 일실시예에 의하면, 트위터 뉴스 및 이와 연관된 수많은 비디오 및 링크드 데이터 중에서 교육 목적에 적합한 자료들을 선택하여 학습자에게 제공할 수 있다.
- [0010] 본 발명의 일실시예에 의하면, 수집된 교육용 멀티미디어 데이터의 메타 데이터를 활용하여 사용자의 선호도 및 수준에 적합한 맞춤형 멀티미디어 데이터를 추천 및 제공할 수 있다.

도면의 간단한 설명

- [0011] 도 1은 본 발명의 일실시예와 관련된 IoT 기반으로 개인화된 학습 콘텐츠를 제공하는 학습 콘텐츠 제공 시스템의 구성을 나타낸다.
- 도 2는 본 발명의 일실시예와 관련된 개인화된 콘텐츠를 제공할 수 있는 콘텐츠 제공의 블록도이다.
- 도 3은 도 2에 도시된 멀티미디어 데이터 처리부의 블록도이다.
- 도 4는 도 3의 비시맨틱 데이터 처리부가 비시맨틱 데이터를 처리하는 과정을 나타내는 흐름도이다.
- 도 5는 도 3의 시맨틱 데이터 처리부가 시맨틱 데이터를 처리하는 과정을 나타내는 흐름도이다.
- 도 6은 도 2에 도시된 응용 처리부의 블록도이다.

발명을 실시하기 위한 구체적인 내용

- [0012] 이하, 본 발명의 일실시예와 관련된 개인화된 콘텐츠 제공 장치에 대해 도면을 참조하여 설명하도록 하겠다.
- [0013] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.
- [0014] IoT 기술은 다양한 수업 방법에 활용될 수 있으며, 그 중 하나는 신문(뉴스)과 같은 미디어를 교육에 활용하는 Newspaper in Education (NIE) 수업 방법에 적용하는 것이다. NIE는 신문(뉴스)을 다양한 과목의 교육용 자료로 활용하는 것이다. 실제로 NIE 수업을 할 때에는 다양한 뉴스(신문)가 학습자에게 제공이 되고, 학습자는 이를 활용하여 토론, 협동학습 등을 통해 특정 주제에 대해서 학습한다. 이를 통해서 학습자들은 비판적 사고 능력, 협동 학습 능력 등 다양한 능력을 키울 수 있다. 최근 온라인 뉴스가 발달하면서, NIE의 매체로 온라인 뉴스를 활용하는 경우가 많아졌다. 뉴스 회사들은 이러한 온라인 뉴스들을 사용자에게 빠르게 전달하기 위해서 트위터와 같은 다양한 채널을 활용한다. 사용자는 뉴스 회사가 게시한 뉴스 트윗을 읽고 이에 대한 관심("Favorite", "Reply" 등)을 표명하거나 트위터의 기능("Retweet")을 통해 다른 사람들에게 전파한다. 따라서 트위터 뉴스의 이러한 기능을 통해 발생하는 데이터를 분석한다면, 대중의 관심사 등을 쉽게 파악할 수 있다. 즉, 트위터 뉴스는 대중들에게 인기가 있는 최신 뉴스를 쉽게 확인할 수 있는 효과적인 교육 매체로 활용할 수 있다. NIE 수업에서 학습 효율 향상을 위해서는 트위터 뉴스뿐만 아니라 연관된 다양한 종류의 멀티미디어 자료들을 제시할 필요가 있다.
- [0015] 이하, 실시예에서는 IoT 기술을 Newspaper in Education (NIE) 수업에 적용하여, 학습자 수준에 적합한 학습 콘텐츠를 제공하는 방법에 대해 설명하도록 하겠다.
- [0016] 도 1은 본 발명의 일실시예와 관련된 IoT 기반으로 개인화된 학습 콘텐츠를 제공하는 학습 콘텐츠 제공 시스템의 구성을 나타낸다.
- [0017] 도시된 바와 같이, 개인화된 학습 콘텐츠를 제공하는 학습 콘텐츠 제공 시스템은 학습자 단말기(100), 콘텐츠 제공 장치(200), 교수자 단말기(300)를 포함할 수 있다.
- [0018] 상기 학습자 단말기(100)는 학습자가 입력한 쿼리(query)를 상기 콘텐츠 제공 장치(200)에 전송하고, 상기 콘텐츠 제공 장치(200)로부터 학습자의 수준에 적합한 학습 데이터를 제공 또는 추천 받을 수 있다. 상기 학습자 단말기(100)로 하나의 장치가 사용될 수 있지만, 복수의 장치가 사용될 수도 있다. 예를 들어, 상기 학습자 단말

기(100)로 IoT 기반 연구에서 널리 사용되는 라즈베리 파이와 모바일 디바이스(예: 스마트 폰이나 스마트패드 등)의 두 가지 장치가 사용될 수 있다. 상기 라즈베리 파이는 교수자가 제시한 QR 코드를 인식할 수 있다. 상기 QR코드는 학습자의 수준을 확인할 수 있는 사전 테스트 문제가 링크되어 있으며 학습자가 이를 해결하게 되면 레벨이 쉽게 판단될 수 있게 된다. 또한, 모바일 디바이스(예: 스마트 폰, 스마트패드 등)는 최근 대부분의 사람들이 사용하고 있는 모바일 장치이다. 이를 통해 학습자가 원하는 데이터를 효율적으로 검색하고 브라우징하기 위해 UI(User Interface)를 제작할 수 있다. 즉, 사용자 인터페이스는 특정 키워드, 날짜 및 인기를 기반으로 뉴스, 비디오 및 기타 멀티미디어 콘텐츠를 검색 및 정렬하고 쿼리 결과를 적절하게 시각화하도록 설계되었다.

- [0019] 상기 콘텐츠 제공 장치(200)는 NIE(Newspaper in Education) 수업에서 학습 효율 향상을 위해서는 트위터 뉴스뿐만 아니라 연관된 다양한 종류의 멀티미디어 자료들을 상기 학습자 단말기(100)로 제공할 수 있다.
- [0020] 상기 교수자 단말기(300)는 교수자가 제작한 상기 QR 코드를 상기 콘텐츠 제공 장치(200)로 전송할 수 있다. 상기 교수자 단말기(300)는 콘텐츠 제공 장치(200)에 접속하여 문서를 수정 및 보완한 학습자를 추적할 수 있다. 이를 통해 교수자는 학습자를 보다 쉽게 평가할 수 있다
- [0021] 도 2는 본 발명의 일실시예와 관련된 개인화된 콘텐츠를 제공할 수 있는 콘텐츠 제공의 블록도이다.
- [0022] 도시된 바와 같이, 콘텐츠 제공 장치(200)는 학습자 정보 수집부(210), 멀티미디어 데이터 처리부(220), 응용 처리부(230), 학습 데이터 제공부(240)를 포함할 수 있다.
- [0023] 상기 학습자 정보 수집부(210)는 학습자 단말기(100)로부터 다양한 종류의 학습자 정보를 수집할 수 있다.
- [0024] 일반적으로 스마트 장치가 교실에서 사용될 때, 학습자는 여러 장치를 통해 다양한 종류의 데이터를 생성할 수 있다. 예를 들어, 본 발명의 일실시예에서는 학습자 정보 수집부(210)는 라즈베리 파이와 스마트 폰(패드)의 두 가지 장치를 사용하여 학습자의 데이터를 수집할 수 있다. 첫 번째로, 라즈베리 파이는 이미 IoT 기반 연구에서 널리 사용되고 있다. 이 장치는 매우 다양한 센서 모듈을 부착하여 여러 가지 응용 기능이 쉽게 구현될 수 있기 때문에 본 발명의 일실시예에 적합할 수 있다. 상기 라즈베리 파이의 카메라 모듈 v2를 사용하여 교수자가 제시한 QR 코드가 인식될 수 있다. 상기 QR코드는 학습자의 수준을 확인할 수 있는 사전 테스트 문제가 링크되어 있으며 학습자가 이를 해결하게 되면 사용자 레벨이 쉽게 판단될 수 있다. 두 번째로, 스마트 폰 또는 스마트패드 등의 모바일 장치를 통해 학습자가 원하는 데이터를 효율적으로 검색하고 브라우징하기 위해 UI(User Interface)를 제작할 수 있다. 즉, 사용자 인터페이스는 특정 키워드, 날짜 및 인기를 기반으로 뉴스, 비디오 및 기타 멀티미디어 콘텐츠를 검색 및 정렬하고 쿼리 결과를 적절하게 시각화하도록 설계되었다.
- [0025] 이 두 장치가 생성한 모든 데이터는 후술할 응용 처리부(230)에 실시간으로 저장되며, 저장된 데이터는 학습자의 다양한 데이터에 기반한 선택을 통해 학습자에게 제공될 수 있다.
- [0026] 상기 멀티미디어 데이터 처리부(220)는 NIE 수업을 위한 다양한 트위터 뉴스 및 멀티미디어 콘텐츠를 수집할 수 있다. 멀티미디어 콘텐츠에는 다양한 비디오 리소스(예: YouTube 동영상, TED 비디오, Vimeo 비디오) 및 교육 Linked Data가 포함될 수 있다. 콘텐츠의 효율적인 검색을 지원하기 위해 동영상 및 트위터 뉴스의 메타 데이터가 수집되고 누락된 메타 데이터는 다양한 처리를 통해 강화될 수 있다. 각 리소스의 최종 메타 데이터는 클라우드 서버에 저장되고, 나중에 콘텐츠를 탐색하는 데 사용될 수 있다.
- [0027] 상기 응용 처리부(230)는 사용자에게 다양한 도구와 인프라를 제공하여 모바일 및 웹 응용 프로그램을 구축할 수 있게 할 수 있다. 상기 응용 처리부(230)는 복잡한 사용자 요청을 처리할 수 있다. 예를 들어, 상기 응용 처리부(230)는 사용자 수준 판단, 쿼리 결과 계산 등과 같은 많은 작업을 처리할 수 있다.
- [0028] 상기 학습 데이터 제공부(240)는 학습자 단말기(100)로부터 수신한 쿼리에 적합한 관련 뉴스, 비디오 및 linked data를 제공할 수 있다. 학습 데이터 제공부(240)는 사용자에게 의해 검색된 쿼리 결과를 보다 효과적으로 시각화할 수 있다. 또한, 상기 학습 데이터 제공부(240)는 학습 효과를 높이기 위해 다양한 협동 학습 도구를 사용할 수 있다. 예를 들어, 학습 데이터 제공부(240)는 클라우드 기반 Google 문서 도구를 사용하여 학습자 간의 공동 작업을 지원할 수 있다. Google 문서 도구는 슬라이드, 단어 문서 및 스프레드 시트와 같은 기능을 제공하며 이러한 문서는 클라우드 기반이므로 학습자는 이를 협동 학습용 도구로 사용할 수 있다. 이를 통해 학습자는 문서를 실시간으로 공유 할 수 있으므로 모든 팀 구성원이 문서를 수정하고 보완 할 수 있다. 또한 교수자가 문서를 수정 및 보완한 학습자를 추적 할 수 있으므로, 학습자를 보다 쉽게 평가할 수 있다. 마지막으로, 응용 처리부(230) 내에 채팅 기능을 생성함으로써 학습자는 아이디어를 공유하거나 문제를 해결하기 위해 같은 팀의 다른

구성원과 의견을 나눌 수 있다.

- [0029] 도 3은 도 2에 도시된 멀티미디어 데이터 처리부의 블록도이다.
- [0030] 도시된 바와 같이, 상기 멀티미디어 데이터 처리부(220)는 비시맨틱 데이터 처리부(221) 및 시맨틱 데이터 처리부(222)를 포함할 수 있다.
- [0031] 상기 비시맨틱 데이터 처리부(221)는 트위터 뉴스나 비디오 리소스 등의 비시맨틱 데이터를 수집하여, 메타데이터를 추출 및 보강하여 시맨틱 데이터 처리부(222)로 그 결과를 전송할 수 있다.
- [0032] 상기 시맨틱 데이터 처리부(222)는 상기 비시맨틱 데이터 처리부(221)에서 수신한 메타데이터를 시맨틱 형식으로 변환하여 데이터를 처리할 수 있다.
- [0033] 이하에서는 비시맨틱 데이터 처리부(221)가 비시맨틱 데이터 처리 과정에 대해 상세하게 설명하도록 하겠다.
- [0034] 도 4는 도 3의 비시맨틱 데이터 처리부가 비시맨틱 데이터를 처리하는 과정을 나타내는 흐름도이다.
- [0035] 먼저, 트위터 뉴스 데이터를 수집할 수 있다(S410).
- [0036] 그리고 딥러닝 기법을 활용하여 트위터 뉴스 토픽을 분류할 수 있다(S420). 상기 딥러닝 기법을 활용한 모델은 Convolutional Neural Network(CNN)와 Attention network의 두 개의 네트워크로 구성된다. 첫 번째로, CNN은 트윗의 컨텍스트를 인식하는 주요 네트워크이다. 이 네트워크는 Convolutional layer와 Fully-connected layer로 구성된다. Convolutional layer는 컨벌루션 필터를 사용하여 Tweet matrix(T)로부터 피처를 추출하고, Fully-connected layer는 차원 c 의 벡터 V_{cnn} 을 출력으로 생성할 수 있다. 여기서 c 는 분류할 주제의 수를 나타낸다.
- [0037] 두 번째로, Attention network는 트윗의 각 단어에 초점을 맞춘다. 이 네트워크는 Fully-connected layer와 softmax 레이어로 구성된다. 이 Fully-connected layer의 출력은 단어가 각 주제와 얼마나 관련이 있는지 나타낸다. Tweet matrix(T)의 각 단어 벡터 w_k 는 Fully-connected layer의 입력이 되며, Fully-connected layer가 생성한 결과는 벡터 V_{word}^c 이다. 모든 V_{word} 는 단어 가중치 행렬($W_{word}^{c \times 1}$)과 연결된다. 그리고 max-pooling을 사용하여 각 클래스의 최대 값을 추출하고 이 값을 softmax layer에 의한 클래스의 확률 분포로 변환한다. 결과적으로, Attention network는 벡터 V_{att}^c 로 구성된다. 이 두 개의 분리된 네트워크는 c 차원 벡터를 생성한다. 이 두 벡터 V_{cnn} 과 V_{att} 를 병합하려면 element-wise multiplication이 적용되고 결과는 softmax 계층의 입력이 된다.
- [0038] 그리고 주제가 분류된 후에, 상기 트위터 뉴스의 메타데이터를 추출 및 보강할 수 있다(S430, S460).
- [0039] 트위터 뉴스에는 트위터 API에서 기본적으로 제공하는 포스팅 시간, retweet 수, favorite 수, comment 수 등과 같은 다양한 메타데이터가 있다. 비시맨틱 데이터 처리부(221)는 분류된 모든 뉴스 트윗에 대해 이러한 메타데이터를 수집할 수 있다. 또한, 뉴스 트윗에는 원본 뉴스로 연결되는 링크가 있다. 트위터 뉴스는 140자의 짧은 글자수로 이루어져 있기 때문에, 해당 뉴스의 자세한 내용을 확인하기 위해서 학습자는 뉴스 원문을 읽어봐야 한다. 따라서, 이러한 링크를 통해 원본 뉴스를 수집할 수 있는데 이 방법을 dereferencing 방법이라고 한다. 이 방법을 통해 우리가 수집한 트위터 뉴스와 연결되는 원본 뉴스 텍스트들을 모두 수집한다. 또한, 이 수집된 데이터들은 전처리, readability 측정, 키워드 추출 등과 같은 방법을 거쳐 처리하고 트위터 뉴스의 메타데이터로 보강될 수 있다. 상기 전처리, readability 측정, 키워드 추출 방법에 대해서는 후술하도록 하겠다.
- [0040] 상기 비시맨틱 데이터 처리부(221)는 비디오 리소스로부터 교육용 비디오를 수집하여 메타데이터를 추출할 수 있다(S440, S450). 예를 들어, 우리는 교육용 비디오를 제공하는 웹사이트 중, 가장 대중적으로 많이 활용되며, API 및 크롤링을 통해 빠르고 정확하게 메타데이터를 수집할 수 있는 TED talk, YouTube, Vimeo 비디오를 선택할 수 있다. 교육용 비디오 리소스의 다양한 메타 데이터를 추출하게 되면 여러 이점이 있다. 우선 메타데이터 기반으로 학습자에게 최적의 비디오를 제공해줄 수 있고, Linked data 시스템 구축시 학습자가 원하는 메타데이터 기반의 검색을 빠르고 정확하게 할 수 있다. 따라서 다음과 같은 방법을 통해 교육용 동영상 및 메타데이터를 수집한다.
- [0041] 첫 번째로, 비시맨틱 데이터 처리부(221)는 TED talk 메타데이터를 수집할 수 있다. TED 웹사이트에는 API가 제공되지 않으므로, 비시맨틱 데이터 처리부(221)는 TED 웹사이트 크롤링을 통해 메타데이터를 수집할 수 있다. 또한 TED는 전체 비디오의 양이 많지 않으므로, 웹사이트에 존재하는 전체 비디오의 메타데이터를 수집한다. 두

번째로, 비시맨틱 데이터 처리부(221)는 교육용 Youtube 채널인 Youtube EDU 비디오 메타데이터를 수집한다. 이를 위해서 비시맨틱 데이터 처리부(221)는 Youtube API를 활용하여 education 카테고리안에 있는 동영상들의 비디오 ID 및 메타데이터를 수집할 수 있다. 세 번째로, 비시맨틱 데이터 처리부(221)는 Vimeo API를 활용하여 Vimeo 비디오 메타데이터를 수집할 수 있다.

[0042] 상기 비시맨틱 데이터 처리부(221)는 수집된 교육용 비디오로부터 추출된 메타데이터 외에 부족한 메타데이터를 보강할 수 있다(S460).

[0043] 상기 수집된 메타데이터는 직접 활용할 수 있는 것들과 추가적인 처리가 필요한 것들(readability, keywords)이 있다. 상기 추가적인 처리가 필요한 메타데이터 추출을 위해서 우선 비디오 리소스의 스크립트(캡션)을 수집할 수 있다. 그리고 수집된 각 리소스 텍스트에 대해서 readability 계산, keywords 추출 등을 통해 동영상 리소스의 부족한 메타데이터들을 보강할 수 있다. 이를 위해서 다음과 같은 과정이 수행될 수 있다.

[0044] 먼저, 전처리 과정이 수행될 수 있다.

[0045] 상기 수집된 스크립트 데이터는 원래 텍스트 형태로 출판된 것이 아니기 때문에, 텍스트 데이터에서 볼 수 없는 다양한 특성이 있을 수 있다. 따라서 이런 것들에 대한 전처리를 수행해야만 더욱 정확한 메타데이터가 수집될 수 있다. 이를 위해서, 첫 번째로, TED talk 비디오 스크립트에 대한 전처리가 수행될 수 있다. TED talk Script는 등장인물의 행동이나 배경음악 등이 기재된 '지문'이 존재하는 것이 특징이다. 이러한 지문은 보통 '(내용)'의 형태로 표현된다. TED 강의의 script는 배경음악, 효과음 혹은 관객들의 박수와 함성 등이 '(내용)' 형태의 지문으로 표현되어있다. 따라서 이것들이 모두 제거될 수 있다. 또한 TED script는 일반적인 텍스트 문서에서는 볼 수 없는 'Wow'나 'Oops' 등의 감탄사가 그대로 표현되어있다. 따라서 COCA(Corpus of Contemporary American English)가 제공하는 감탄사 리스트를 이용해 script 내의 감탄사가 모두 제거될 수 있다. 두 번째로, YouTube 비디오 캡션에 대한 전처리가 수행될 수 있다. YouTube는 음성 인식 기술을 사용하여 자동으로 비디오 캡션을 만든다. 유튜브 캡션 또한 TED talk과 유사한 감탄사 제거의 전처리 과정을 거친다. 또한 유튜브 캡션은 각 캡션별로 시간이 표시되어 있기 때문에 이것이 추가적으로 제거될 수 있다. 마지막으로, Vimeo 비디오 캡션에 대한 전처리가 수행될 수 있다. Vimeo 메타데이터 종류 중 캡션이 존재하지만 실제로 수집할 수 있는 캡션이 있는 데이터는 많지 않으므로, 우선적으로 캡션이 있는 데이터만이 수집될 수 있다.

[0046] 다음으로 Readability가 계산될 수 있다.

[0047] 보다 개인화된 비디오 및 뉴스 추출을 위해서 가독성(Readability)이 계산될 수 있다. 가독성은 인쇄물이 얼마나 쉽게 읽히는가 하는 능력의 정보를 말한다. 이는 단어의 길이, 문장의 길이, 단어 사용의 복잡성 등을 반영한 수학 공식에 의해 측정될 수 있다. 지금까지 다양한 readability 측정 formula들이 개발되었는데 본 실시예에서는 이들 중 Coleman-Liau Index, Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning Fog Index, SMOG, Automated Readability Index가 활용될 수 있다. 그리고 이들을 통합한 새로운 formula를 통해 최종 readability가 계산될 수 있다. 이하에서는 TED, Youtube, Vimeo의 스크립트의 readability가 다음의 공식을 활용하여 계산된다.

수학식 1

$$\alpha_1 = 0.0588L - 0.296S - 15.8$$

L: the average number of letters per 100 words

S: the average number of sentences per 100 words

[0048]

수학식 2

$$\alpha_2 = 206.835 - 1.015 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left(\frac{\text{total syllables}}{\text{total words}} \right)$$

[0049]

수학식 3

$$\alpha_3 = 0.39 \left(\frac{\text{total words}}{\text{total sentences}} \right) + 11.8 \left(\frac{\text{total syllables}}{\text{total words}} \right) - 15.59$$

수학식 4

$$\alpha_4 = 0.4 \left[\left(\frac{\text{words}}{\text{sentences}} \right) + 100 \left(\frac{\text{complex words}}{\text{words}} \right) \right]$$

수학식 5

$$\alpha_5 = 1.0430 \sqrt{\text{Number of polysyllables} \times \frac{30}{\text{Number of sentence}}} + 3.1291$$

수학식 6

$$\alpha_6 = 4.71 \left(\frac{\text{characters}}{\text{words}} \right) + 0.5 \left(\frac{\text{words}}{\text{sentences}} \right) - 21.43$$

상기 수학식 1은 Coleman-Liau Index을 나타내고, 상기 수학식 2는 Flesch Reading Ease을 나타내고, 상기 수학식 3은 Flesch-Kincaid Grade Level을 나타내고, 상기 수학식 4는 Gunning Fog Index을 나타내고, 상기 수학식 5는 SMOG를 나타내고, 상기 수학식 6은 Automated Readability Index를 나타낸다.

여기서 $\sigma_1 \sim$ 내지 σ_6 는 Readability를 의미한다.

여기에서 Flesch Reading Ease 산출 값은 추가적으로 normalization 작업을 거친 후 최종 값을 산출하며 다른 공식들은 산출된 값을 그대로 활용한다. 우리는 리소스들의 readability를 위의 공식들을 활용하여 각각 계산하고, 이를 평균을 내어 최종 readability값으로 추출한다.

다음으로 Keywords 추출 과정이 수행될 수 있다.

키워드 추출 작업은 문서를 가장 잘 설명하는 단어들을 문서에서 자동으로 식별하는 것이다. 본 실시예에서는 다양한 키워드 추출 방법들 중 널리 사용되고 있는 Textrank 알고리즘이 활용될 수 있다. 따라서 주어진 캡션(스크립트)을 하나의 문서로 가정하고, 이것에 대하여 Textrank 알고리즘을 활용해 키워드가 추출될 수 있다.

다음으로 부족한 메타데이터가 추가될 수 있다.

TED의 대부분 강의는 유튜브 TED 채널을 통해 공개되기 때문에, 이를 통해서 TED 웹사이트 자체에서 제공하는 메타데이터의 부족함을 보충할 수 있다. 따라서 TED 웹사이트 크롤링을 통해 얻은 TED talk의 제목을 기반으로 동일한 유튜브 영상을 찾고 이를 통해 부족한 메타데이터가 수집될 수 있다.

다음으로 시맨틱 데이터 처리 과정이 수행될 수 있다.

학습자에게 좀 더 가치 있는 정보를 제공하기 위해서는 뉴스 및 연관 비디오 이외에 다양한 learning resource 들을 추가적으로 제공할 필요가 있다. 많은 교육 기관과 대학교가 교육용 리소스들을 링크드 데이터 형식으로 제공한다. 하지만, 기존의 non-semantic 교육용 데이터에서는 시맨틱 교육용 데이터와의 접점을 찾기 어려웠다.

따라서 본 실시예에서는 시맨틱 데이터도 비시맨틱(non-semantic) 데이터처럼 학습자가 쉽게 활용할 수 있도록 다음과 같은 방법이 활용될 수 있다. 즉, 수집된 뉴스 및 메타데이터를 시맨틱 데이터 형식(예: RDF)으로 변경하고 이를 외부의 교육용 Linked Open Data와의 interlinking을 한 결과를 클라우드 저장소에 저장한다면, 트위터 뉴스와 외부의 교육용 semantic data와 쉽게 연결이 될 것이다. 이를 통해 콘텐츠를 더욱 풍부하게 할 수 있으며, 교육용 시맨틱 데이터에 대한 접근이 쉬워질 것이다.

- [0063] 트위터 뉴스와 외부 교육용 linked data와의 연계를 위해서는 트위터 뉴스 및 메타데이터를 RDF 형식으로 변경해야 한다.
- [0064] 도 5는 도 3의 시맨틱 데이터 처리부가 시맨틱 데이터를 처리하는 과정을 나타내는 흐름도이다.
- [0065] 먼저, 트위터 뉴스 및 메타데이터를 우리의 저장소에 저장한다.
- [0066] 그리고 S460 단계에서 수집된 메타데이터 데이터를 기반으로 온톨로지 스키마를 설계하고 설계된 스키마에 맞추어 인스턴스를 추가한다(S510).
- [0067] 이것들을 RDF 형식으로 변환하여, SPARQL endpoint 또는 RDF dump를 통해 이들을 사용할 수 있도록 한다(S520).
- [0068] 일반적으로 인터링킹 툴은 매칭 알고리즘을 활용하여 SPARQL Endpoints 또는 RDF 덤프를 통해 두 데이터 셋간의 유사성이 발견될 수 있다. 그 중 많이 활용되는 툴이 SILK와 LIMES이며 이것들은 매우 잘 업데이트 및 관리된다. 본 실시예에서는 인터링킹을 위해서 SILK 툴이 활용될 수 있다.
- [0070] 학습자에게 제공되는 교육 자료들은 웹 상의 다양한 정보와 연결될 때 더욱 풍부하고 유용한 자료가 될 수 있다. 이를 위해서 우리는 외부의 다양한 교육용 Linked data 들을 파악해야 하고 이를 트위터 뉴스와 연결시키는 인터링킹이 수행될 수 있다(S530).
- [0071] 교육용으로 출판된 데이터셋을 분석하기 위해서 Open University의 데이터 셋이 활용될 수 있다. 이 플랫폼에는 약 370만개의 triple, 131개의 class, 다양한 주제를 가지는 3,000개 이상의 audio-video resources, 700개 이상의 course 들이 있다. 또한 OU에서 운영하는 Openlearn은 free educational resource들을 제공해주며, 약 800개의 수준별 free course가 8개의 카테고리로 분류되어 있다.
- [0072] 뉴스 데이터와 Open university의 데이터(비디오, 오디오, 강의자료, Openlearn)를 인터링킹하며, DBpedia는 각종 용어들을 보다 자세히 정의하기 위해서 인터링킹이 수행될 수 있다. 즉, 이 두 가지 Linked data를 우리의 트위터 뉴스와 인터링킹한다면 더욱 풍부한 데이터셋이 만들어 질 수 있다(S540). 이 경우, 학습자는 특정 주제에 대해 더욱 심도 있는 학습을 할 수 있다.
- [0073] 이하에서 인터링킹 과정과 최종 메타데이터 구성 및 저장 과정에 대해 구체적인 방법에 대해 설명하도록 하겠다.
- [0074] 인터링킹을 위해서 source resource와 이와 연결되는 target resource를 정의해야 한다. source resource는 우리가 앞장에서 제작한 트위터 뉴스이며, 이와 연결되는 target 리소스는 DBpedia 및 Open University linked 데이터이다. 두 데이터의 연결을 위해서는 이들을 연결할 수 있는 property를 정의하고, 연결할 리소스들의 class를 선정해야 한다.
- [0075] 첫 번째로, 우리의 데이터셋에서는 연결이 가능한 entity가 없기 때문에 이것을 직접 제작해야 한다. 또한 이 과정을 통해 dbpedia와의 인터링킹도 같이 진행한다. 이를 위해서, 일례로 source resource에 해당하는 트위터 뉴스안의 URL을 역참조하여 뉴스 웹사이트에서 원본 뉴스의 데이터를 수집할 수 있다. 그리고 수집한 원본 뉴스 텍스트에 textrank 알고리즘을 적용하여 term들의 textrank score를 계산한다. 이때, 계산된 textrank score가 높은 단어일수록 문서에서 중요한 term으로 간주한다. 또한 우리는 각 문서별 textrank score가 높은 키워드를 순차적으로 추출한다. 추출된 키워드들은 dbpedia resource인지 아닌지 파악을 해야 인터링킹이 가능하다. dbpedia linked data의 class 및 resource들은 각각 <http://dbpedia.org/ontology/>, <http://dbpedia.org/page/> 형식의 URI를 가지고 있다. 즉, 우리가 뉴스 텍스트를 textrank 알고리즘으로 처리한 term을 위의 URI 뒷부분에 입력하고 유효한 주소인지 검증한다. 즉, 추출된 term들이 dbpedia의 class 및 resource name과 매칭이 되는 키워드들만 추출한다. 처리 결과 유효한 주소로 검증이 되면 이 주소를 우리의 온톨로지 dc:subject property에 추가하여 dbpedia dataset과 인터링킹을 한다. 하지만 모든 키워드들을 리소스의

subject로 넣게 되면 중요성이 낮은 키워드들도 추가가 되고, 필요 없는 자료들이 인터링킹될 확률이 높아진다. 따라서 textRank score의 threshold(0.0015) 값을 정하고 그 이하의 키워드들은 모두 제거를 한다.

- [0076] 두 번째로, Target resource 연결을 위하여 이들의 특징을 파악해야 한다. Open University Linked data는 다양한 property로 구성되어 있으며, 이들은 해당 resource를 잘 설명할 수 있는 자료들로 구성되어 있다. Property 중 subject는 해당 리소스의 주제를 Open university에 의해 자체적으로 선별하였다. 또한, 이것들은 모두 dbpedia resource와 인터링킹이 되어 있다. 본 실시예에서는 source resource와 target resource의 특징이 파악될 수 있으며, 이를 기반으로 두 데이터간의 interlinking을 진행될 수 있다.
- [0077] 인터링킹을 하게 되면 트위터 뉴스와 연관된 linked data가 검색이 되며, 이 결과는 우선적으로database에 저장될 수 있다. 이렇게 되면 추출된 시맨틱 데이터를 비시맨틱 데이터 형식으로 활용할 수 있다는 장점을 갖는다. 또한, database에 저장된 non-semantic data는 추후 linked data로 출판할 수 있도록 RDF/XML 파일 형식으로도 저장하고 이를 웹에 업로드할 수 있다. 이를 위해서 온톨로지 스키마를 구성하고, 이를 활용해 최종 결과를 RDF/XML 파일로 저장한다. 이를 통해 학습자 및 교수자가 시맨틱 데이터 형식도 웹을 통해 쉽게 접근할 수 있다.
- [0078] 한편, 본 발명의 일 실시예에 의하면, 상기 학습 데이터 제공부(240)는 메타데이터를 활용하여 학습자에게 적합한 교육 자료를 추천할 수 있다.
- [0079] 한편, 수집된 비디오 및 메타데이터를 모두 학습자들에게 제공할 수 없다. 왜냐하면, 수집한 비디오의 양이 매우 많기 때문에 학습자가 적절한 콘텐츠를 선택할 수 없으며, 학습 주제와 연관이 없는 비디오들을 제공할 경우 학습자의 흥미 및 학습 효율이 감소할 것이기 때문이다.
- [0080] 이하에서는 수집한 메타데이터를 활용하여 이들 중 뉴스 트윗 및 이와 가장 적합한 비디오들을 골라서 제공하는 방법에 대해 설명하도록 한다. 쿼리는 학습자의 모바일 디바이스의 UI를 통해서 전달되며, 계산된 결과 또한 그곳으로 전달이 된다.
- [0081] 먼저, 학습 데이터 제공부(204)의 뉴스 트윗 선택기(미도시)는 상기 수집 및 분류한 뉴스 트윗 중 적절한 것들을 선택할 수 있다.
- [0082] 학습자는 시스템에 쿼리를 통해 원하는 뉴스를 찾는다. 일반적으로 사용자의 쿼리에는 기간, 키워드 및 원하는 주제가 포함될 수 있다. 그러한 쿼리가 주어지면, 관련 뉴스 트윗이 검색되고 하기 7개의 수학적식에 의해 스코어링 될 수 있다. 하기의 수학적식에 의해 최신의 인기있는 뉴스 트윗이 더욱 높은 점수를 얻게 될 수 있다.

수학식 7

$$Score_{date} = 1 - \left| \frac{\min(\alpha, D_{current} - D_{upload})}{\alpha} \right|$$

[0083]

수학식 8

$$Score_{date_final} = \frac{Score_{date} - \frac{1}{N} \sum_{i=1}^N (Score_{date})_i}{\sqrt{\frac{1}{N} \sum_{i=1}^N (Score_{date} - \frac{1}{N} \sum_{i=1}^N (Score_{date})_i)^2}}$$

[0084]

수학식 9

$$Score_{retweet} = \frac{(count(retweet, tweet) - \frac{1}{N} \sum_{i=1}^N (count(retweet, tweet))_i)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (count(retweet, tweet) - \frac{1}{N} \sum_{i=1}^N (count(retweet, tweet))_i)^2}}$$

수학식 10

$$Score_{favorite} = \frac{(count(favorite, tweet) - \frac{1}{N} \sum_{i=1}^N (count(favorite, tweet))_i)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (count(favorite, tweet) - \frac{1}{N} \sum_{i=1}^N (count(favorite, tweet))_i)^2}}$$

수학식 11

$$Score_{readability} = \frac{1}{|\beta - readability(text)| + 1}$$

수학식 12

$$Score_{readability_final} = \frac{Score_{readability} - \frac{1}{N} \sum_{i=1}^N (Score_{readability})_i}{\sqrt{\frac{1}{N} \sum_{i=1}^N (Score_{readability} - \frac{1}{N} \sum_{i=1}^N (Score_{readability})_i)^2}}$$

수학식 7에 의해 최신 뉴스를 선택하는 score가 계산될 수 있다. 여기에서 $Score_{date}$ 는 뉴스 트윗이 생성된 날짜와 현재 날짜를 비교하여 최신 뉴스가 더 높은 점수를 산출하도록 한다. 또한 α 는 원하는 기간을 나타낸다. 수학식 9 및 수학식 10에서 $Score_{retweet}$ 과 $Score_{favorite}$ 는 각각 "retweets"과 "favorite"의 수에 비례하는 점수이다. 각 메타데이터들은 값의 기준이 다르기 때문에 이것들을 단순 비교하기는 어렵다. 따라서 각 데이터들을 rescaling하는 작업이 필요하다. 따라서 우리는 수학식 8, 수학식 9, 수학식 10을 활용해서 이러한 값들을 standardization한다. 여기서 $count(retweet, tweet)$ 와 $count(favorite, tweet)$ 는 하나의 트윗에서 "retweets"과 "favorite"의 수를 나타낸다. 수학식 11은 사용자가 입력한 readability인 β 값과 뉴스 텍스트의 readability를 비교하는 공식이다. 즉, 두 값 간의 차이가 많이 날수록 낮은 점수를 부여하고, 차이가 적을수록 높은 점수를 부여하는 것이다. 식 수학식 12는 수학식 11에서 계산한 값들을 standardization하는 공식이다.

수학식 13

$$Score_{final_tweet} = \begin{cases} Score_{readability_final} + Score_{date_final} + \left(\frac{Score_{retweet} + Score_{favorite}}{2} \right) * \gamma, & \text{if } \frac{Score_{retweet} + Score_{favorite}}{2} \geq 0 \\ Score_{readability_final} + Score_{date_final} + \left(\frac{Score_{retweet} + Score_{favorite}}{2} \right) * \frac{1}{\gamma}, & \text{if } \frac{Score_{retweet} + Score_{favorite}}{2} < 0 \end{cases}$$

마지막으로, 수학식 13에서는 위의 수식의 결과값들을 활용해 최종 score를 산출한다. 이 수식에서는 γ factor를 사용하는데, 이것은 트윗의 인기를 반영하는 factor이다. 여기서, γ 값이 증가할수록 인기가 높은 뉴스 트윗들을 선택할 수 있다. 최종적으로, α , β , γ 들은 모두 학습자들의 쿼리에 의해 결정된다. 즉, 학습자는 위에서 제시한 다양한 factor를 변경하면서 본인에게 적절한 뉴스를 찾을 수 있고, 이를 통해 개인화된 뉴스 확

용 수업이 가능해진다.

[0092] 학습 데이터 제공부(204)의 비디오 콘텐츠 선택기(미도시)는 수집한 다양한 비디오 리소스들은 학습자에게 제공되기 위하여 두 가지 방식으로 데이터 처리를 수행할 수 있다. 첫 번째는, 뉴스와 연관된 비디오들을 유사도 기반으로 선택한다. 이를 위해서 트위터 뉴스의 URL을 역참조하여 얻은 원본 뉴스 텍스트와 각 비디오들의 캡션 텍스트에 대하여 tf-idf값을 추출하고, 이를 기반으로 유사도를 계산한다. 이때, 우리는 뉴스 및 비디오 각각을 하나의 document로 가정하고, NVS (News-Video Similarity)를 계산한다. NVS 공식은 다음의 수학식 14를 활용할 수 있다.

수학식 14

$$NVS(d_i, d_j) = Similarity(v_i, v_j) = \frac{v_i \cdot v_j}{|v_i| \times |v_j|}$$

[0093]

[0094] 상기 수학식 14는 코사인 유사도 equation으로, 상기 수학식 14에서 v_i 와 v_j 는 각 도큐먼트의 feature vector이고, 이것은 각 문서에서 추출한 tf-idf 값을 이용하였다. 수학식 14를 활용하여 원본 뉴스의 텍스트와 비디오 내의 스크립트 텍스트와의 유사도를 비교하고, 유사도가 높은 순으로 나열한다. 우리는 유사도가 특정 threshold 값 이상이 되는 비디오들만을 추출하여 이들을 뉴스와 연관이 있는 비디오로 정의하고 이를 뉴스 메타데이터로 저장한다. 하지만 1차적으로 선택된 비디오의 양이 적지 않고 사용자의 수준 및 선호도를 반영하지 않은 결과이기 때문에, 좀 더 정제할 필요가 있다. 따라서 우리는 아래의 formula를 통해서 사용자 수준 및 선호도를 반영하여 연관 비디오들을 추출하려고 한다.

수학식 15

$$Score_{view} = \frac{(count(view, video) - \frac{1}{N} \sum_{i=1}^N (count(view, video))_i)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (count(view, video) - \frac{1}{N} \sum_{i=1}^N (count(view, video))_i)^2}}$$

[0096]

수학식 16

$$Score_{like} = \frac{(count(like, video) - \frac{1}{N} \sum_{i=1}^N (count(like, video))_i)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (count(like, video) - \frac{1}{N} \sum_{i=1}^N (count(like, video))_i)^2}}$$

[0097]

수학식 17

$$Score_{dislike} = \frac{(count(dislike, video) - \frac{1}{N} \sum_{i=1}^N (count(dislike, video))_i)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (count(dislike, video) - \frac{1}{N} \sum_{i=1}^N (count(dislike, video))_i)^2}}$$

[0098]

수학식 18

$$Score_{comment} = \frac{(count(comment, video) - \frac{1}{N} \sum_{i=1}^N (count(comment, video))_i)}{\sqrt{\frac{1}{N} \sum_{i=1}^N (count(comment, video) - \frac{1}{N} \sum_{i=1}^N (count(comment, video))_i)^2}}$$

[0099]

[0100]

학습자에게 가장 적합한 비디오들을 제공해주기 위해서 동영상의 메타데이터가 활용될 수 있다. 활용 가능한 메타 데이터 중 view 수, like 수, unlike 수, comment 수, 생성 날짜, readability 를 중점적으로 활용될 수 있다. 그리고 이 값들은 활용해 뉴스 트윗을 처리할 때와 같이 모두 rescaling 작업을 한다. 이 값들 중 생성 날짜, readability 관련 수식은 앞에서 활용한 수학식 8과 수학식 12가 활용될 수 있다. 수학식 15 내지 수학식 18은 비디오 view 수, like수, dislike수, comment 수를 각각 standardization하는 수식이며, count(view, video)는 하나의 비디오가 조회된 횟수를 나타낸다. 최종 score는 standardization되어 score_{view}, score_{like}, score_{dislike}, score_{comment}로 각각 산출된다.

수학식 19

$$Score_{final_video} = \begin{cases} Score_{readability_final} + Score_{date_final} + \left(\frac{Score_{view} + Score_{like} + Score_{comment} - Score_{dislike}}{4} \right) * \gamma, & \text{if } \frac{Score_{view} + Score_{like} + Score_{comment} - Score_{dislike}}{4} \geq 0 \\ Score_{readability_final} + Score_{date_final} + \left(\frac{Score_{view} + Score_{like} + Score_{comment} - Score_{dislike}}{4} \right) * \frac{1}{\gamma}, & \text{if } \frac{Score_{view} + Score_{like} + Score_{comment} - Score_{dislike}}{4} < 0 \end{cases}$$

[0101]

[0102]

마지막으로, 수학식 19에서는 뉴스 트윗 처리기와 마찬가지로, α, β, γ 값에 따라서 최종 score를 계산한다. 이를 통해, 학습자는 본인 수준에 적합하며, 최신의 인기 있는 동영상을 선택하고 이를 통해 효율적인 학습을 할 수 있게 된다.

[0103]

도 6은 도 2에 도시된 응용 처리부의 블록도이다.

[0104]

도시된 바와 같이, 응용 처리부(230)는 연산부(231) 및 저장부(232)를 포함할 수 있다. 상기 연산부(231)는 학습자 정보를 통해 학습자의 수준 판단 및 쿼리 결과 계산 등과 같은 많은 작업에 대한 연산 처리를 수행할 수 있다.

[0105]

상기 저장부(232)에는 데이터베이스 및 이미지가 저장될 수 있다. 상기 저장부(232)는 데이터베이스 이외의 다양한 기능을 제공할 수 있다. 예를 들어, 교수자는 저장부(232)를 통해 학습자가 생성한 다양한 데이터와 수집 및 처리된 멀티미디어 데이터에 대한 통계를 쉽게 볼 수 있다.

[0106]

전술한 바와 같이, 본 발명의 일실시예에 의한 개인화된 콘텐츠 제공 장치는 학습자 정보를 통해 이에 적합한 IoT 기반의 맞춤형 학습 콘텐츠를 제공할 수 있다.

[0107]

본 발명의 일실시예에 의하면, 비 시맨틱(non-semantic) 데이터들(트위터 뉴스, 관련 동영상 등)뿐만 아니라 시맨틱(semantic) 데이터들도 학습자가 별다른 노력 없이 접근할 수 있다.

[0108]

본 발명의 일실시예에 의하면, 트위터 뉴스 및 이와 연관된 수많은 비디오 및 링크드 데이터 중에서 교육 목적에 적합한 자료들을 선택하여 학습자에게 제공할 수 있다.

[0109]

본 발명의 일실시예에 의하면, 수집된 교육용 멀티미디어 데이터의 메타 데이터를 활용하여 사용자의 선호도 및 수준에 적합한 맞춤형 멀티미디어 데이터를 추천 및 제공할 수 있다.

[0110]

상술한 개인화된 콘텐츠 제공 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터로 판독 가능한 기록 매체에 기록될 수 있다. 이때, 컴퓨터로 판독 가능한 기록매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 한편, 기록매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지된 것일 수도 있다.

[0111]

컴퓨터로 판독 가능한 기록매체에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM, DVD와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광

매체(Magneto-Optical Media), 및 롬(ROM), 램(RAM), 플래시 메모리, SSD (Solid State Drive)와 같은 메모리 저장장치 등 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.

[0112] 한편, 이러한 기록매체는 프로그램 명령, 데이터 구조 등을 지정하는 신호를 전송하는 반송파를 포함하는 광 또는 금속선, 도파관 등의 전송 매체일 수도 있다.

[0113] 또한, 프로그램 명령에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0114] 상기와 같이 설명된 개인화된 콘텐츠 제공 장치 및 방법은 상기 설명된 실시예들의 구성과 방법이 한정되게 적용될 수 있는 것이 아니라, 상기 실시예들은 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.

부호의 설명

[0115]

100: 학습자 단말기

200: 콘텐츠 제공 장치

210: 학습자 정보 수집부

220: 멀티미디어 데이터 처리부

221: 비시맨틱 데이터 처리부

222: 시맨틱 데이터 처리부

230: 응용 처리부

231: 연산부

232: 저장부

240: 학습 데이터 제공부

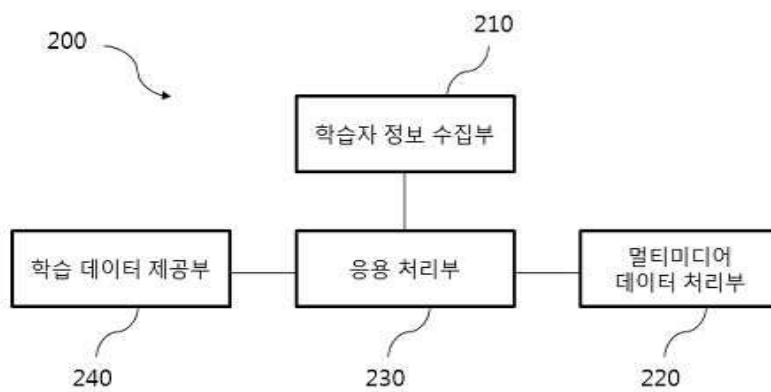
300: 교수자 단말기

도면

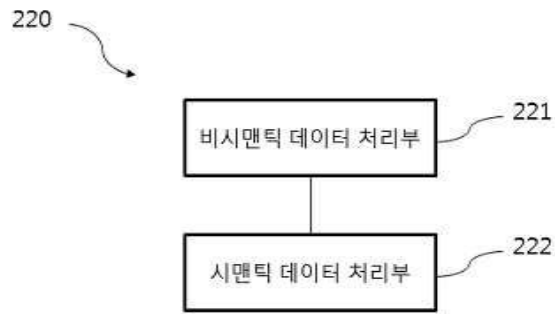
도면1



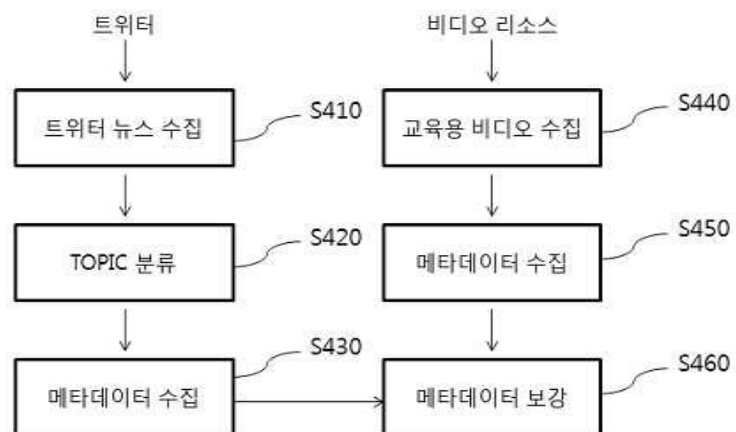
도면2



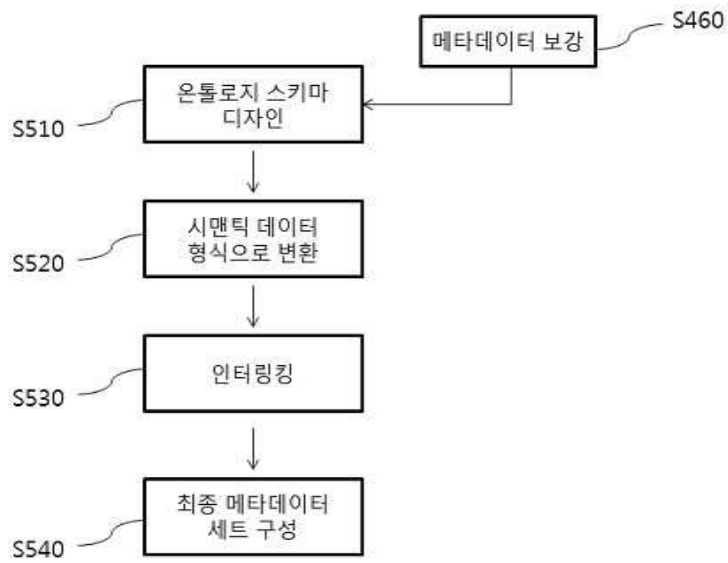
도면3



도면4



도면5



도면6

