



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2025년03월05일

(11) 등록번호 10-2776164

(24) 등록일자 2025년02월27일

- (51) 국제특허분류(Int. Cl.)
H04N 21/466 (2011.01) *G06F 16/35* (2025.01)
G06F 16/435 (2019.01) *G06F 16/735* (2019.01)
G06F 16/901 (2019.01) *G06F 18/00* (2023.01)
G06N 20/00 (2019.01) *G06N 5/00* (2023.01)
G06N 7/00 (2023.01) *H04N 21/482* (2011.01)
H04N 21/8405 (2011.01)
- (52) CPC특허분류
H04N 21/4668 (2013.01)
G06F 16/355 (2025.01)
- (21) 출원번호 10-2020-7037749
- (22) 출원일자(국제) 2019년09월04일
 심사청구일자 2021년12월23일
- (85) 번역문제출일자 2020년12월29일
- (65) 공개번호 10-2021-0054491
- (43) 공개일자 2021년05월13일
- (86) 국제출원번호 PCT/US2019/049599
- (87) 국제공개번호 WO 2020/051249
 국제공개일자 2020년03월12일
- (30) 우선권주장
 201841033185 2018년09월04일 인도(IN)
 (뒷면에 계속)
- (56) 선행기술조사문헌
 Chandra Shekhar Yadav 외 2명, 'Semantic graph based approach for text mining', IEEE, 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques, 2014.02.07. pp.596-601 1부*
 Sashikumar Venkataraman 외 1명, 'A Natural Language Interface for Search and Recommendations of Digital Entertainment Media', IBC 2015 Conference, 2015.01.01., pp.1-11 1부*
 Watchreewan외2명, 'Enhancing comment Feedback classification using text classifiers with word centrality measures', IEEE, 2017 2nd international Conference on Information Technology, 2017.11.02. pp.1-5 1부*
 *는 심사관에 의하여 인용된 문헌
- (73) 특허권자
 로비 가이드스, 인크.
 미국 캘리포니아주 95002 산 호세 골드 스트리트 2160
- (72) 발명자
 춘가펠리 리진
 인도 벵갈루루 560103 파나서 로드 114
 페람바투 벤카타 랍지
 인도 벵갈루루 560103 샌딕 스퀘어 아파트먼트스 플랫 넘버 210
- (74) 대리인
 김태홍, 김진희

전체 청구항 수 : 총 9 항

심사관 : 장우진

(54) 발명의 명칭 검색, 추천, 및 발견을 촉진하기 위해 머신-학습 추출물들 및 시맨틱 그래프들을 이용하여 구조화된 데이터를 생성하는 방법들 및 시스템들

(57) 요약

시맨틱 그래프들과 머신 학습의 조합을 이용하여, 구조화된 데이터를 자동으로 생성하고, 중요한 엔티티들/키워드들을 인식하고, 그리고 더 관련된 검색 결과들 및 추천들을 위해 가중된 연결들을 생성하는 방법들 및 시스템들. 예를 들어, 관련된 엔티티들을 추론함으로써, 메타데이터 결과들이 더 풍부하고 더 의미 있어, 소비자를 위

(뒷면에 계속)

대표도



한 더 빠른 의사-결정 및 콘텐츠 소유자를 위한 더 강한 시정률을 가능하게 한다.

(52) CPC특허분류

G06F 16/435 (2019.01)

G06F 18/217 (2023.01)

G06F 18/29 (2023.01)

G06N 20/00 (2021.08)

G06N 5/01 (2023.01)

G06N 7/01 (2023.01)

H04N 21/4662 (2013.01)

H04N 21/4828 (2013.01)

H04N 21/8405 (2013.01)

(30) 우선권주장

16/289,573 2019년02월28일 미국(US)

16/289,575 2019년02월28일 미국(US)

명세서

청구범위

청구항 1

텍스트 문자열들에서 엔티티들의 관련성들을 자동으로 결정함으로써 콘텐츠 추천들을 제공하는 방법으로서,

사용자 입력 인터페이스에 의해, 미디어 콘텐츠로의 액세스를 요청하는 텍스트 문자열을 수신하는 단계;

제어 회로부에 의해, 상기 텍스트 문자열에서 대명사를 식별하는 단계;

상기 제어 회로부에 의해, 상기 대명사를 명사로 분석(resolving)하여 분석된 텍스트 문자열을 생성하는 단계;

상기 제어 회로부에 의해, 상기 분석된 텍스트 문자열에서 특정의 명사 청크를 식별하는 단계;

상기 제어 회로부에 의해, 복수의 미디어 자산과 연관된 복수의 명사 청크에 대응하는 복수의 노드를 포함하는 시맨틱 그래프의 노드들 사이의 가중된 연결들(weighted connections), 및 상기 복수의 명사 청크에 대한 복수의 스코어(score) — 상기 복수의 스코어 및 상기 가중된 연결들은 상기 복수의 미디어 자산과 관련하여 상기 노드들의 명사 청크들 사이의 문맥적(contextual) 관계를 표시함 —에 기초하여, 트레이닝된 분류기 머신을 생성하는 단계;

상기 제어 회로부에 의해, 근접 중심성 메트릭(closeness centrality metric) 및 매개 중심성 메트릭(betweenness centrality metric)에 기초하여, 상기 시맨틱 그래프에서 상기 복수의 노드에 대응하는 상기 복수의 명사 청크에 대한 상기 복수의 스코어를 각각 생성하는 단계 — 상기 근접 중심성 메트릭은 상기 시맨틱 그래프에서의 개개 노드와 상기 복수의 노드의 다른 노드 각각 사이의 최단 경로의 길이의 총합(sum)의 척도이고, 상기 매개 중심성 메트릭은 상기 시맨틱 그래프의 상기 복수의 노드의 개개 노드의 상기 시맨틱 그래프에서의 중심성의 척도임 —;

트레이닝된 분류기 머신 학습 모델에 상기 분석된 텍스트 문자열의 상기 특정의 명사 청크를 입력하고, 상기 입력에 기초하여, 상기 트레이닝된 분류기 머신 학습 모델로부터의 출력으로서 상기 복수의 미디어 자산과 관련하여 상기 특정의 명사 청크의 개개 컨텍스트의 중요성의 표시를 수신하는 단계;

상기 제어 회로부에 의해, 상기 시맨틱 그래프에 기초하여 트레이닝된 상기 트레이닝된 분류기 머신 학습 모델의 상기 출력에 기초하여, 상기 텍스트 문자열을 통해 요청된 상기 미디어 콘텐츠가 상기 복수의 미디어 자산 중 특정의 미디어 자산에 대응한다고 결정하는 단계; 및

상기 결정하는 단계에 응답하여, 디스플레이 디바이스 상에 상기 수신된 텍스트 문자열에 응답하여 상기 특정의 미디어 자산의 추천을 디스플레이를 위해 생성하는 단계

를 포함하는, 방법.

청구항 2

제 1 항에 있어서,

상기 트레이닝된 분류기 머신 학습 모델을 사용하여 상기 특정의 명사 청크를 프로세싱하는 것에 기초하여, 복수의 엔티티 중 엔티티를 결정하는 것은:

각각의 엔티티를 스코어링하는 단계;

해당 개개 스코어에 기초하여 각각의 엔티티를 순위지정하는 단계; 및

최고 스코어를 가진 엔티티를 선택하는 단계

를 더 포함하는, 방법.

청구항 3

제 1 항에 있어서,

상기 트레이닝된 분류기 머신 학습 모델은 의사결정 트리 분류기 또는 랜덤 포레스트 분류기인, 방법.

청구항 4

제 1 항에 있어서,

상기 수신된 텍스트 문자열에 응답하여, 엔티티를 디스플레이를 위해 생성하는 것은 검색, 추천, 또는 발견 특징에서 상기 엔티티를 디스플레이를 위해 생성하는 것을 포함하는, 방법.

청구항 5

제 1 항에 있어서,

상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 단계는 코레퍼런스 분석(coreference resolution)을 사용하여 상기 대명사를 분석하는 단계를 포함하는, 방법.

청구항 6

제 1 항에 있어서,

상기 분석된 텍스트 문자열에서 상기 특징의 명사 청크를 식별하는 단계는 품사(part-of-speech) 태깅을 사용하여 상기 특징의 명사 청크를 식별하는 단계를 포함하는, 방법.

청구항 7

제 1 항에 있어서,

상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 방법.

청구항 8

개개 하나 이상의 프로세서에 의해 실행될 때, 상기 하나 이상의 프로세서로 하여금, 제 1 항 내지 제 7 항 중 어느 한 항에 기재된 방법을 수행하게 하는 컴퓨터-판독가능 명령어들을 포함하는 매체에 저장된 컴퓨터 프로그램.

청구항 9

텍스트 문자열들에서 엔티티들의 관련성들을 자동으로 결정함으로써 콘텐츠 추천들을 제공하는 시스템으로서,

텍스트 문자열을 수신하도록 구성된 사용자 입력 인터페이스; 및

제 1 항 내지 제 7 항 중 어느 한 항의 방법을 실행하도록 구성된 제어 회로부를 포함하는, 시스템.

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

발명의 설명

배경 기술

- [0001] 오늘날의 소비자들은 다양한 선형 및 스트리밍 서비스들에서 제공하는 영화들, 프로그램들, 뉴스, 및 짧은 형식의 비디오를 포함한 콘텐츠의 바다에서 선택할 수 있는 이점을 갖고 있다. 소비할 수 있는 콘텐츠가 너무 많기 때문에, 소비자들은 이들이 보고 싶은 무언가를 찾기 위해 이 콘텐츠를 필터링하는 것이 어려울 수도 있다. 실제로, 과잉의 가용 콘텐츠는 "쇼-덤핑(show-dumping)"이라는 현상을 일으키며, 이에 의해 소비자들은 이들에 액세스하는데 수반되는 문제들로 인해 프로그램들을 간단히 포기한다. 쇼-덤핑은 콘텐츠 소유자들 및 콘텐츠 소비자들 양자 모두에게 큰 문제를 야기한다. 콘텐츠 소유자들은 콘텐츠를 제작하는데 막대하게 투자하지만 여전히 소비자들이 콘텐츠에 액세스할 수 있도록 하기 위해 고군분투할 수도 있다. 이와 유사하게, 콘텐츠 소비자들은 콘텐츠를 용이하게 이용할 수 있음에도 불구하고 원하는 콘텐츠를 찾을 수 없지만, 찾기도 어렵다.

발명의 내용

- [0002] 이러한 문제를 감안하여, 사용자들로 하여금 이들이 소비하고자 하는 콘텐츠를 더 빠르고 더 쉽게 찾을 수 있게 하는 애플리케이션에 대한 방법들 및 시스템들이 본원에서 설명된다. 이러한 해결책을 제공하기 위해서는, 콘텐츠에 대한 더 깊은 이해가 요구된다. 예를 들어, 너무 많은 콘텐츠가 있고, 구조화된 메타데이터가 거의 없기 때문에, 종래의 검색 및 추천 기법들은 콘텐츠의 양이 증가함에 따라 사용자들을 점차 실망시킨다. 일단 이 문제가 이해되면, 본원에서 설명되는 해결책이 이 문제를 극복하는데 이용될 수 있다.
- [0003] 예를 들어, 종래의 검색 및 추천 시스템들은 통계-기반 모델들에 기초한 엔티티 추출에 의존한다. 예를 들어, 이러한 시스템들에서는, 식별된 용어(예컨대, 미디어 자산에 대한 메타데이터에서 발견된 설명 용어(descriptive term))에, 관련된 용어가 식별된 용어에 대응할 가능성을 나타내는 통계치에 기초하여, 다른 관련된 용어들이 할당된다. 따라서, 입력(예컨대, 사용자 검색 요청)이 수신될 때, 시스템은 입력에서의 용어들을 관련된 용어들과 비교한다. 관련된 용어들 중 하나 이상이 입력에서의 용어에 대응하면, 시스템은 매치(match)를 결정한다.
- [0004] 그러나, 콘텐츠의 양이 증가하고, 그에 따라 그 콘텐츠에 대한 식별된 용어들, 관련된 용어들 등의 양이 기하급수적으로 증가함에 따라, 엔티티 추출을 위한 이들 종래의 통계-기반 모델들은 개개의 사용자의 요구들에 맞는 정확한 검색 결과들을 제공하지 못한다. 예를 들어, 지속적으로 증가하는 데이터의 양들을 프로세싱할 수 있는 더욱 강력한 프로세서들의 존재에도 불구하고, 이들 시스템들은 이들이 종래의 통계-기반 모델 외부의 입력들을 해석하지 못하기 때문에, 전술한 문제들을 여전히 해결하지 못할 것이다. 특히, 이들 시스템들은 주어진 입력의 의미론적(semantic) 이해를 얻지 못하고 이 정보를 이용하여 검색, 추천, 및 발견 프로세스를 진전시키지 못한다.
- [0005] 임계치 레벨에서, 위에서 설명한 바와 같이, 초과 데이터에 의해 과부하가 걸린 시스템에의 (예컨대, 시맨틱 관계들에 관련한) 더 많은 정보의 추가는 단지 기존 문제들을 더욱 악화시키는 것으로 보인다. 그러나, 머신-학습에서의 최근의 진보들은 원하는 결과들을 제공하기 위해 이 증가된 데이터를 효율적으로 이용하는 방법을 제공한다. 구체적으로 설명하면, 4개의 별개의 단계들, 즉, 대명사 분석(resolution), 후보 식별, 시맨틱 그래프 생성, 및 노드 스코어링을 특징으로 하는 특징의 아키텍처의 사용을 통해서, 본원에서 설명되는 시스템들 및 방법들은 정밀도(precision)와 재현율(recall) 사이의 조화 평균이고 검색, 추천, 및 발견 특징들을 제공할 때 성능을 평가하기 위한 통계적 척도로서 사용되는 향상된 F1 스코어(score)를 제공하는 애플리케이션을 제공한다. 즉, 본원에서의 시스템들 및 방법들은 사용자들에게 더 나은 결과들을 제공하기 위해 주어진 텍스트 문자열에서 엔티티의 관련성(relevance)을 자동으로 결정하는 머신-학습 모델을 트레이닝하기 위해 시맨틱 그래프에서의 노드들의 중요성을 레버리지한다. 실질적인 문제로, 머신-학습 방법들과 시맨틱 그래프들을 이 고유한 방법으로 결합하는 것은 절실히 필요한 컨텍스트를 추가하고 소비자 불만을 완화할 수 있을 뿐만 아니라, 콘텐츠 소유자들의 시청률을 강화할 수 있다.
- [0006] 일부 양태들에서, 본원에서 설명되는 방법들 및 시스템들은 검색, 추천, 및 발견 특징들(features)을 제공한다. 예를 들어, 시스템은 데이터 세트를 수집할 수도 있다. 사용자는 외부 데이터 세트에서 텍스트 문자열들을 입력할 수도 있거나, 또는 시스템은 데이터 세트를 팝플레이트하기 위해 웹으로부터 데이터를 능동적으로 수집할 수도 있다. 시스템은 그후 데이터 세트에 대해 대명사 분석을 수행할 수도 있다. 예를 들어, 시스템은 데이터 세

트에서의 텍스트 문자열들 내 각각의 대명사를 식별하여 라벨링할 수도 있다. 시스템은 그후 데이터 세트에 대해 후보 식별을 수행할 수도 있다. 예를 들어, 시스템은 데이터 세트에서의 텍스트 문자열들 내 모든 명사 청크들을 식별하기 위해 데이터 세트 상에 POS(Part-Of-Speech) 태깅을 적용할 수도 있다. 시스템은 그후 복수의 키 엔티티들(key entities) 및 복수의 키 엔티티들 사이의 복수의 연관들(associations)을 식별하는 시맨틱 그래프를 생성할 수도 있다. 시맨틱 그래프는 노드들 사이의 시맨틱 관계들을 나타내는 방향성 에지들에 의해 연결된 데이터 세트로부터의 후보들에 대응하는 노드들을 포함할 수도 있다. 시스템은 그후 사용자 입력 인터페이스에 의해, 사용자 입력을 수신할 수도 있다. 사용자 입력은 텍스트 문자열 또는 발화(utterance)일 수도 있다. 시스템은 그후 시맨틱 그래프를 이용하여 사용자 입력을 프로세싱할 수도 있다. 예를 들어, 시스템은 사용자 입력으로부터의 후보들을 시맨틱 그래프에서의 노드들과 매칭시킬 수도 있다. 종속성 트리를 횡단함으로써, 시스템은 입력의 시맨틱을 학습할 수도 있다. 시스템은 입력에 관련된 관련 정보를 추가로 학습할 수도 있다. 시스템은 그후 프로세싱된 사용자 입력에 기초하여 출력을 생성할 수도 있다. 예를 들어, 출력은 사용자 입력에 대한 답변, 사용자 입력에 기초한 추천, 사용자 입력에 대한 관련 정보, 또는 다른 정보를 포함할 수도 있다.

[0007] 일부 양태들에서, 방법들 및 시스템들은 텍스트 문자열들에서 엔티티들의 관련성들을 자동으로 결정함으로써 콘텐츠 추천을 제공한다. 예를 들어, 시스템은 사용자 입력 인터페이스에 의해, "What was the movie with the iceberg? It sinks the ship"과 같은, 텍스트 문자열을 수신할 수도 있다. 시스템은 그후 제어 회로부에 의해, 텍스트 문자열에서 대명사를 식별할 수도 있다. 예를 들어, 시스템은 "it"을 대명사로서 식별할 수도 있다. 시스템은 그후 제어 회로부에 의해, 대명사를 고유 명사로 분석하여, 분석된 텍스트 문자열을 생성할 수도 있다. 예를 들어, 시스템은 대명사 "it"이 명사 "iceberg"를 지칭하는 것으로 결정하여, 분석된 텍스트 문자열: "What was the movie with the iceberg? The iceberg sinks the ship"을 생성할 수도 있다. 시스템은 그후 제어 회로부에 의해, 분석된 텍스트 문자열에서 명사 청크를 식별할 수도 있다. 예를 들어, 시스템은 명사 "iceberg"를 제 1 명사 청크로서 그리고 명사 "ship"을 제 2 명사 청크로서 식별할 수도 있다. 시스템은 그후 제어 회로부에 의해, 분류기를 이용해 복수의 명사 청크들을 특징으로 하는 시맨틱 그래프에 기초하여 명사 청크를 프로세싱할 수도 있으며, 여기서, 복수의 명사 청크들의 각각은 근접 중심성 메트릭(closeness centrality metric) 및 매개 중심성 메트릭(betweenness centrality metric)에 기초하여 스코어링되며, 여기서, 근접 중심성 메트릭은 시맨틱 그래프에서의 개개 노드와 다른 노드들 각각 사이의 최단 경로의 길이의 총합의 척도이며, 여기서, 매개 중심성 메트릭은 개개 노드의 시맨틱 그래프에서의 중심성의 척도이다. 예를 들어, 시맨틱 그래프는 노드들로서 복수의 명사들을 특징화할 수도 있으며, 여기서, 명사들은 특징의 소스로부터의 데이터 세트 및/또는 특징의 주제의 명사들에 대응한다. 시스템은 그후 제어 회로부에 의해, 명사 청크를 프로세싱하는 것에 기초하여 분류기를 이용해 엔티티를 결정할 수도 있다. 예를 들어, 시스템은 시맨틱 그래프의 각각의 노드에 대해 스코어를 결정함으로써 엔티티(예컨대, 명사, 엔티티, 미디어 콘텐츠의 제목, 컴퓨터-생성된 쿼리, 등)를 결정할 수도 있다. 시스템은 그후 최고 스코어를 가진 노드를 결정하고 그 노드에 대응하는 엔티티를 리트리빙(retrieving)할 수도 있다. 시스템은 그후 수신된 텍스트 문자열에 응답하여, 디스플레이 디바이스 상에서, 엔티티를 디스플레이를 위해 생성할 수도 있다. 예를 들어, 시스템은 사용자에게 대한 컴퓨터-생성된 응답에 엔티티를 포함시킬 수도 있다. 컴퓨터-생성된 응답은 엔티티에 대응하는 미디어 콘텐츠를 특징으로 하는 검색 결과들의 리스트를 포함할 수도 있다.

[0008] 일 실시형태에 대해 본원에서 설명되는 방법들 및 시스템들은 본원에서 설명된 바와 같은 다른 실시형태들과 결합될 수도 있다는 점에 유의해야 한다.

도면의 간단한 설명

[0009] 본 개시물의 상기 및 다른 목적들 및 이점들은 첨부 도면들과 함께 취해진, 다음의 상세한 설명을 고려하면 명백해질 것이며, 도면 중 유사한 참조 부호들은 전체에 걸쳐 유사한 부재들을 지칭한다.

도 1은 본 개시물의 일부 실시형태들에 따른, 사용자 인터페이스의 예시적인 예를 나타낸다.

도 2는 본 개시물의 일부 실시형태들에 따른, 사용자 인터페이스의 다른 예시적인 예를 나타낸다.

도 3은 본 개시물의 일부 실시형태들에 따른, 예시적인 사용자 장비 디바이스의 블록도이다.

도 4는 본 개시물의 일부 실시형태들에 따른, 예시적인 미디어 시스템의 블록도이다.

도 5는 본 개시물의 일부 실시형태들에 따른, 예시적인 모델에 대한 결과들을 특징으로 하는 테이블을 나타낸다.

도 6은 본 개시물의 일부 실시형태들에 따른, 검색, 추천, 및 발견 특징들을 제공하는데 사용되는 아키텍처의 예시적인 예이다.

도 7은 본 개시물의 일부 실시형태들에 따른, 예시적인 시맨틱 그래프를 나타낸다.

도 8 내지 도 10은 본 개시물의 일부 실시형태들에 따른, 추출된 엔티티들 및 역할들의 예시적인 예들을 나타낸다.

도 11은 본 개시물의 일부 실시형태들에 따른, 사용자 인터페이스의 예시적인 예를 나타낸다.

도 12는 본 개시물의 일부 실시형태들에 따른, 사용자 인터페이스의 다른 예시적인 예를 나타낸다.

도 13은 본 개시물의 일부 실시형태들에 따른, 사용자 인터페이스의 또 다른 예시적인 예를 나타낸다.

도 14는 본 개시물의 일부 실시형태들에 따른, 검색, 추천, 및 발견 특징들을 제공하는데 사용되는 프로세스의 예시적인 플로우차트를 도시한다.

도 15는 본 개시물의 일부 실시형태들에 따른, 엔티티를 결정하는데 사용되는 프로세스의 예시적인 플로우차트를 도시한다.

도 16은 본 개시물의 일부 실시형태들에 따른, 검색, 추천, 및 발견 특징들을 제공하는데 사용되는 아키텍처의 예시적인 예를 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0010] 시맨틱 그래프들과 머신 학습의 조합을 이용하여, 구조화된 데이터를 자동으로 생성하고, 중요한 엔티티들/키워드들을 인식하고, 그리고 더 관련성 있는 검색 결과들 및 추천들을 생성하는 가중된 연결들(weighted connections)을 생성하기 위한 방법들 및 시스템들이 본원에서 설명된다. 예를 들어, 관련된 엔티티들을 추론함으로써, 메타데이터 결과들이 더 풍부하고 더 의미 있어, 소비자를 위한 더 빠른 의사-결정 및 콘텐츠 소유자를 위한 더 강한 시청률을 가능하게 한다.
- [0011] 본원에서 지칭될 때, 시맨틱 그래프는 개념들 사이의 시맨틱 관계들을 나타내는 네트워크일 수도 있다. 특히, 본원에서 설명되는 시맨틱 그래프는 상이한 품사들 사이의 시맨틱 관계들을 나타낼 수도 있다. 예를 들어, 이 네트워크에서, 시맨틱 그래프는 개념들에 대응하는 정점들 및 개념들 사이의 시맨틱 관계들을 나타내는 에지들로 구성될 수도 있다.
- [0012] 예를 들어, 시맨틱 그래프에서, 개념들은 8개의 품사들(예컨대, 등위 접속사들, 종속 접속사들, 접속 부사들, 상관 접속사들, 및/또는 감탄사들을 포함한, 명사들, 동사들, 형용사들, 부사들, 전치사들, 접속사들)의 각각을 포함할 수도 있다. 이들 품사들, 및 시맨틱 그래프(즉, 개념)의 각각의 단어에 대한 품사를 표시하는 메타데이터가, (예컨대, 그래프에서의 노드들을 나타내는) 단어들을 함께 결합하여 해석가능한 문장들을 만드는 방법을 결정하기 위해 시스템에 의해 사용된다. 그후, 이들 단어들 사이의 결합들(joins)이 (예컨대, 사용자에 의해) 시스템에 제기된 쿼리를 해석할 뿐만 아니라 쿼리에 대한 응답을 생성하기 위해, 순위지정된다.
- [0013] 도 1은 방법들 및 시스템들의 애플리케이션을 예시한다. 도 1에서, 사용자 인터페이스(100)가 디스플레이 디바이스 상에 디스플레이된다. 사용자 인터페이스(100)는 (예컨대, 사용자 입력 인터페이스에의 사용자 입력을 통해서) 텍스트 문자열(102)을 수신한다. 이에 응답하여, 시스템은 프로그램 추천(104)을 디스플레이를 위해 생성한다. 다음 예는 시맨틱 그래프들로부터의 키워드들이 콘텐츠의 더 깊은 이해를 보여주고 더 풍부한 검색 경험을 제공하는 방법을 예시한다. 예를 들어, 텍스트 문자열(102)("movie where a person falls in love with an operating system")에 대해, 시스템은 시맨틱 그래프를 통해서 영화 "Her"에 대응하는 프로그램 추천(102)을 반환한다. 이 실시형태에서, 시맨틱 그래프는 미디어 콘텐츠의 플롯 세부 사항들로부터 키워드들 및 설명들을 포함하는 데이터 세트에 기초하여 구축된다. 데이터 세트는 임의의 데이터 소스로부터 또/또는 임의의 특징의 주제에 기초하여 임의 종류의 데이터를 포함할 수 있다는 점에 유의해야 한다. 도 1에서, 시스템은 텍스트 문자열(102)에서의 단어들 "love" 및 "operating system"이 매우 관련성있고 문맥적(contextual) 키워드들인 것으로 결정한다. 시스템은 시맨틱 키워드들을 "Good_Keyword"로 플래깅하고 이들 키워드들을 검색 시스템에서 더 높은 가중치로 인덱싱한다.
- [0014] 도 2는 방법들 및 시스템들의 다른 애플리케이션을 예시한다. 도 2에서, 사용자 인터페이스(200)는 디스플레이 디바이스 상에 디스플레이된다. 사용자 인터페이스(200)는 영화 "Argo"에 대응하는 텍스트 문자열(202)을 (예컨대,

대, 사용자 입력 인터페이스로의 사용자 입력을 통해서) 수신한다. 예를 들어, 사용자 요청에 응답하여, 시스템은 "Argo"와 유사한 특성들을 공유하는 다른 콘텐츠를 추천할 수도 있다. 이에 응답하여, 시스템은 프로그램 추천들(204 및 206)을 디스플레이를 위해 생성한다. 추가적으로, 시스템은 유사한 영화들의 각각에 대한 스코어들을 생성한다. 예를 들어, 프로그램 추천(204)은 스코어(208)를 포함한다. 추가적으로 또는 대안적으로, 시스템은 프로그램 추천에 대응하는 프로그램들에 액세스하기 위한 링크들을 생성할 수도 있다. 예를 들어, 도 2는 프로그램 추천(204)에 대응하는 프로그램에 액세스하기 위한 링크인, 링크(210)를 포함한다.

[0015] 도 2에서, 엔티티들(예컨대, 프로그램 추천들(204 및 206))은 시맨틱 개념들로서 간주되며, 엔티티들의 유사성들이 추천들에서 사용된다. 예를 들어, 영화에서, "Argo", "CIA", "thriller", 및 "war"는 중요한 주제, 장르, 및 테마 개념들이다. 시스템은 이들 중 하나 이상을 레버리지하고 "Fair Game", 및 "Syriana"와 같은 유사한 영화들을 추천한다. 예를 들어, 본원에서 설명되는 시맨틱 그래프는 검색 결과들을 향상시키기 위해, 비구조화된 텍스트(예컨대, 미디어 콘텐츠에 대한 메타데이터)에서 가장 중요한 노드들에 가중치를 부여한다. 이에 반해, 용어 빈도-역 문서 빈도("TF-IDF", term frequency-inverse document frequency)와 같은 통계적 방법들에 의해 유도된 모델들로부터 추출된 키워드들은 컨텍스트의 엘리먼트들을 무관한 컨텍스트 엘리먼트들과 구분하지 않는다. TF-IDF는 컬렉션 또는 코퍼스(corpus)에서 단어가 문서에 얼마나 중요한지를 반영하기 위한 수치적인 통계치이다. 이는 정보 리트리빙, 텍스트 마이닝, 및 사용자 모델링의 검색들에서 가장 인자로서 종종 이용된다. TF-IDF 값은 단어가 문서에서 나타나는 횟수에 비례하여 증가하고 그 단어를 포함하는 코퍼스에서 문서들의 개수만큼 오프셋되는데, 이는 일부 단어들이 일반적으로 더 빈번하게 나타나는 사실을 조정하는 것을 돕는다. 이러한 경우, "사랑(love)"과 같은 포괄적인 용어는 높은 용어 및 문서 빈도를 가지며, 전통적인 TF-IDF-기반 모델들은 우수한 가중치 키워드로 간주되지 않을 것이다. 시맨틱 그래프 접근법은, 이에 반해, 컨텍스트의 중요성에 기초하여 키워드의 관련성을 측정함으로써 종래의 통계치들을 향상시킨다. 컨텍스트의 중요성의 결정은 아래에서 설명하는 바와 같이, 시맨틱 그래프에서의 키워드의 위치 및 그 키워드와 다른 개념들 사이의 연결들에 기초한다.

[0016] 시맨틱 그래프 특징들이 영화들 및 텔레비전 프로그램들과 같은 미디어 자산들뿐만 아니라, 뉴스 기사들, 짧은 형식의 콘텐츠, 심지어 시상식들과 같은 일회용 이벤트들까지, 다양한 콘텐츠에 적용될 수 있다는 점에 유의해야 한다. 실제로, 시맨틱 그래프 특징은 임의의 미디어 자산에 적용될 수도 있다. 본원에서 지칭될 때, 용어들 "미디어 자산(media asset)" 및 "콘텐츠(content)"는 전자적으로 소비가능한 사용자 자산, 예컨대 텔레비전 프로그램들뿐만 아니라, 유료 시청 프로그램들, (주문형 비디오(VOD) 시스템들에서와 같이) 온-디맨드 프로그램들, 인터넷 콘텐츠(예컨대, 스트리밍 콘텐츠, 다운로드가능한 콘텐츠, 웹캐스트들, 등), 비디오 클립들, 오디오, 콘텐츠 정보, 사진들, 회전 이미지들, 문서들, 플레이리스트들, 웹사이트들, 물품들, 책들, 전자책들, 블로그들, 채팅 세션들, 소셜 미디어, 애플리케이션들, 게임들, 및/또는 임의의 다른 미디어 또는 멀티미디어 및/또는 이들의 조합을 의미하는 것으로 이해되어야 한다. 안내 애플리케이션들은 또한 사용자들로 하여금 콘텐츠를 탐색하고 위치추적할 수 있게 한다. 본원에서 지칭될 때, 용어 "멀티미디어"는 위에서 설명된 적어도 2개의 상이한 콘텐츠 형태들, 예를 들어, 텍스트, 오디오, 이미지들, 비디오, 또는 대화형 콘텐츠 형태들을 이용하는 콘텐츠를 의미하는 것으로 이해되어야 한다. 콘텐츠는 사용자 장비 디바이스들에 의해 리코딩되거나, 플레이되거나, 디스플레이되거나, 또는 액세스될 수도 있지만, 또한 라이브(live) 공연의 부분일 수 있다.

[0017] 이들 미디어 자산들 중 임의의 것에 대해, 시맨틱 그래프로부터 결정된 정보는 콘텐츠의 발견을 향상시키는 데 적용될 수 있으며 소비자들을 위한 관련된 결과들 및 의미 있는 추천들을 생성할 수 있다. 추가적으로 또는 대안적으로, 시맨틱 그래프들은 시스템에 의해, 유행하는 주제 식별에 사용될 수도 있다. 예를 들어, 시스템은 구글 뉴스(Google News)와 같은 비구조화된 소스들로부터 유행하는 주제들을 추출할 수도 있다. 예를 들어, 뉴스 기사로부터, 시스템은 가장 관련된 엔티티들을 하이라이트하고 덧없는 멘션(fleeting mention)의 시끄러운 엔티티들(noisy entities)을 억제할 수도 있으며, 시맨틱 그래프의 노드-스코어링 메커니즘이 가장 관련된 엔티티들을 평가할 수도 있다.

[0018] 추가적으로 또는 대안적으로, 시맨틱 그래프들은 시스템에 의해, 명명된 엔티티 추출에 사용될 수도 있다. 예를 들어, 시스템은 텍스트에서의 명명된 엔티티들을 위치추적하여, 사람들의 이름들, 조직들, 위치들, 시간 표현들, 수량들, 금전적 가치들, 퍼센티지들 등과 같은, 미리 정의된 카테고리들로 분류할 수도 있다. 시스템은 그후 콘텐츠 발견을 위해 비구조화된 텍스트(예컨대, 뉴스 기사, 콘텐츠 설명)로부터 문맥상 중요한 엔티티들 또는 키워드들을 자동으로 추출할 수도 있다.

[0019] 추가적으로 또는 대안적으로, 시맨틱 그래프들은 시맨틱 그래프로부터의 노드 스코어에 기초한 콘텐츠에서 중요하고 중요하지 않은 캐스트 멤버들 및 역할들의 분류인 역할 중요성을 위해 시스템에 의해 사용될 수도 있다.

예를 들어, 도 8 및 도 9에서, 높은 스코어를 얻기 위해 결정된 중요한 역할들이 도시된다. 이들 중요한 역할들은 도 1 내지 도 2의 디스플레이들에 디스플레이될 수도 있다.

[0020] 또한, 시스템은 시맨틱 그래프들을 머신 학습과 함께 이용하여 콘텐츠의 더 깊은 이해를 얻고, 컨텍스트에 기초하여 관련된 엔티티들/키워드들을 빨리 식별하고, 그리고 종종 "검색 및 발견" 방법들을 소모하는 것 이상으로 엔터테인먼트 발견을 확장할 수도 있다는 점에 유의해야 한다. 따라서, 뷰어들은 정확한 제목 또는 캐릭터를 기억하는데 더 이상 얽매이지 않고, 대신 자연어를 이용하여 이들이 관심있는 콘텐츠를 찾을 수 있다. 문맥상 관련된, 음성-기반(voice-powered) 검색 결과들 및 추천들을 위한 이 기반(foundation)은 올바른 콘텐츠를 빨리 찾으려는 소비자들의 욕구를 만족시키고 콘텐츠 소유자들로 하여금 이들의 롱테일(long-tail) 카탈로그들의 시청률을 증가시킬 수 있게 한다.

[0021] 도 3은 본원에서 설명된 검색, 추천, 및 발견 특징들을 제공할 수도 있는, 예시적인 사용자 장비 디바이스(300)의 일반적인 실시형태들을 나타낸다. 예를 들어, 사용자 장비 디바이스(300)는 스마트폰 디바이스 또는 리모트 컨트롤일 수도 있다. 다른 예에서, 사용자 장비 시스템(301)은 사용자 텔레비전 장비 시스템일 수도 있다. 이러한 경우, 디바이스들은 이들의 메모리에 시맨틱 그래프를 저장하고/하거나 요청을 프로세싱하기 위해 시맨틱 그래프에 액세스할 수도 있다. 사용자 텔레비전 장비 시스템(301)은 셋-탑 박스(316)를 포함할 수도 있다. 셋-탑 박스(316)는 스피커(314) 및 디스플레이(312)에 통신 가능하게 접속될 수도 있다. 일부 실시형태들에서, 디스플레이(312)는 텔레비전 디스플레이 또는 컴퓨터 디스플레이일 수도 있다. 일부 실시형태들에서, 셋-탑 박스(316)는 사용자 인터페이스 입력(310)에 통신 가능하게 접속될 수도 있다. 일부 실시형태들에서, 사용자 인터페이스 입력(310)은 원격 제어 디바이스일 수도 있다. 셋-탑 박스(316)는 하나 이상의 회로 보드들을 포함할 수도 있다. 일부 실시형태들에서, 회로 보드들은 프로세싱 회로부, 제어 회로부, 및 스토리지(예컨대, RAM, ROM, 하드 디스크, 착탈식 디스크, 등)를 포함할 수도 있다. 일부 실시형태들에서, 회로 보드들은 입력/출력 경로를 포함할 수도 있다. 사용자 장비 디바이스들의 좀더 구체적인 구현예들은 아래에서 도 4와 관련하여 설명된다. 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301) 각각은 입력/출력(이하 I/O) 경로(302)를 통해서 콘텐츠 및 데이터를 수신할 수도 있다. I/O 경로(302)는 콘텐츠(예컨대, 브로드캐스트 프로그래밍, 온-디맨드 프로그래밍, 인터넷 콘텐츠, 근거리 네트워크(LAN) 또는 광역 네트워크(WAN)를 통해 이용가능한 콘텐츠, 및/또는 다른 콘텐츠) 및 데이터를 프로세싱 회로부(306) 및 스토리지(308)를 포함하는 제어 회로부(304)에 제공할 수도 있다. 제어 회로부(304)는 I/O 경로(302)를 이용하여 지령들, 요청들, 및 다른 적합한 데이터를 전송 및 수신하는데 이용될 수도 있다. I/O 경로(302)는 제어 회로부(304)(및 구체적으로 프로세싱 회로부(306))를 (아래에서 설명하는) 하나 이상의 통신 경로들에 접속할 수도 있다. I/O 기능들은 이들 통신 경로들 중 하나 이상에 의해 제공될 수도 있지만 도면을 과도하게 복잡하게 만드는 것을 피하기 위해 도 3에 단일 경로로서 도시된다.

[0022] 제어 회로부(304)는 프로세싱 회로부(306)와 같은 임의의 적합한 프로세싱 회로부에 기초할 수도 있다. 본원에서 지칭될 때, 프로세싱 회로부는 하나 이상의 마이크로프로세서들, 마이크로제어기들, 디지털 신호 프로세서들, 프로그래밍가능 로직 디바이스들, 필드-프로그래밍가능 게이트 어레이들(FPGAs), 주문형 집적 회로들(ASICs) 등에 기초한 회로부를 의미하는 것으로 이해되어야 하며, 멀티-코어 프로세서(예컨대, 이중-코어, 쿼드-코어, 헥사-코어, 또는 임의의 적당한 수의 코어들) 또는 슈퍼컴퓨터를 포함할 수도 있다. 일부 실시형태들에서, 프로세싱 회로부는 다수의 별개의 프로세서들 또는 프로세싱 유닛들, 예를 들어, 여러 개의 동일한 유형의 프로세싱 유닛들(예컨대, 2개의 인텔(Intel) 코어 i7 프로세서들) 또는 다수의 상이한 프로세서들(예컨대, 인텔 코어 i5 프로세서 및 인텔 코어 i7 프로세서)에 걸쳐서 분산될 수도 있다. 일부 실시형태들에서, 제어 회로부(304)는 메모리(예컨대, 스토리지(308))에 저장된 애플리케이션에 대한 명령어들을 실행한다. 구체적으로 설명하면, 제어 회로부(304)는 애플리케이션에 의해 위에서 그리고 아래에서 설명된 기능들을 수행하도록 지시받을 수도 있다. 예를 들어, 애플리케이션은 명령어들을 제어 회로부(304)에 제공하여 미디어 안내 디스플레이들을 생성할 수도 있다. 일부 구현예들에서, 제어 회로부(304)에 의해 수행되는 임의의 액션은 애플리케이션으로부터 수신된 명령어들에 기초할 수도 있다.

[0023] 클라이언트/서버-기반 실시형태들에서, 제어 회로부(304)는 안내 애플리케이션 서버 또는 다른 네트워크들 또는 서버들과 통신하기에 적합한 통신 회로부를 포함할 수도 있다. 위에서 언급된 기능을 실행하기 위한 명령어들은 안내 애플리케이션 서버 상에 저장될 수도 있다. 통신 회로부는 케이블 모뎀, 종합 정보 통신망(ISDN) 모뎀, 디지털 가입자 회선(DSL) 모뎀, 전화기 모뎀, 이더넷 카드, 또는 다른 장비와의 통신을 위한 무선 모뎀, 또는 임의의 다른 적합한 통신 회로부를 포함할 수도 있다. 이러한 통신들은 인터넷 또는 임의의 다른 적합한 통신 네트워크들 또는 경로들(도 4와 관련하여 좀더 자세하게 설명됨)을 수반할 수도 있다. 게다가, 통신 회로부는 사용자 장비 디바이스들의 피어-투-피어 통신, 또는 (아래에 좀더 자세히 설명되는) 서로 멀리 떨어진 위치들에서

사용자 장비 디바이스들의 통신을 가능하게 하는 회로부를 포함할 수도 있다.

[0024] 메모리는 제어 회로부(304)의 부분인 스토리지(308)로서 제공되는 전자 저장 디바이스일 수도 있다. 본원에서 지칭될 때, 어구 "전자 저장 디바이스" 또는 "저장 디바이스"는 전자 데이터, 컴퓨터 소프트웨어, 또는 펌웨어를 저장하는 임의의 디바이스, 예컨대 랜덤-액세스 메모리, 판독-전용 메모리, 하드 드라이브들, 광학 드라이브들, 디지털 비디오 디스크(DVD) 리코더들, 콤팩트 디스크(CD) 리코더들, BLU-RAY 디스크(BD) 리코더들, BLU-RAY 3D 디스크 리코더들, 디지털 비디오 리코더들(개인 비디오 리코더들로 종종 불리는 DVR들, 또는 PVR들), 고체 상태 디바이스들, 양자 저장 디바이스들, 게이밍 콘솔들, 게이밍 미디어, 또는 임의의 다른 적합한 고정 또는 착탈식 저장 디바이스들, 및/또는 이들의 임의의 조합을 의미하는 것으로 이해되어야 한다. 스토리지(308)는 위에서 설명된 미디어 안내 데이터뿐만 아니라 본원에서 설명되는 다양한 유형들의 콘텐츠를 저장하는데 이용될 수도 있다. 비휘발성 메모리는 또한 (예컨대, 부트업 루틴 및 다른 명령어들을 시동하는데) 이용될 수도 있다. 도 4와 관련하여 설명된, 클라우드-기반 스토리지는 스토리지(308)를 보충하거나 또는 스토리지(308) 대신에 사용될 수도 있다.

[0025] 제어 회로부(304)는 비디오-생성 회로부 및 동조(tuning) 회로부, 예컨대 하나 이상의 아날로그 튜너들, 하나 이상의 MPEG-2 디코더들 또는 다른 디지털 디코딩 회로부, 고선명 튜너들, 또는 임의의 다른 적합한 동조 또는 비디오 회로들 또는 이러한 회로들의 조합들을 포함할 수도 있다. (예컨대, 저장을 위해 오버-디-에어, 아날로그, 또는 디지털 신호들을 MPEG 신호들로 변환하기 위한) 인코딩 회로부가 또한 제공될 수도 있다. 제어 회로부(304)는 또한 콘텐츠를 사용자 장비(300)의 원하는 출력 포맷으로 상향변환 및 하향변환하는 스케일러 회로부를 포함할 수도 있다. 회로부(304)는 또한 디지털 신호와 아날로그 신호 사이에 변환하는 디지털-대-아날로그 변환기 회로부 및 아날로그-대-디지털 변환기 회로부를 포함할 수도 있다. 동조 및 인코딩 회로부는 콘텐츠를 수신, 디스플레이, 또는 리코딩하기 위해 사용자 장비 디바이스에 의해 이용될 수도 있다. 동조 및 인코딩 회로부는 또한 안내 데이터를 수신하는데 이용될 수도 있다. 예를 들어, 동조, 비디오 생성, 인코딩, 디코딩, 암호화, 복호화, 스케일러, 및 아날로그/디지털 회로부를 포함하는, 본원에서 설명되는 회로부는 하나 이상의 범용 또는 특수 프로세서들 상에서 실행하는 소프트웨어를 이용하여 구현될 수도 있다. 다수의 튜너들이 동시적인 동조 기능들(예컨대, 시청 및 리코딩 기능들, 화면 속 화면(PIP) 기능들, 다중 튜너 리코딩, 등)을 처리하기 위해 제공될 수도 있다. 스토리지(308)가 사용자 장비(300)로부터 별개의 디바이스로서 제공되면, 동조 및 (다수의 튜너들을 포함하는) 인코딩 회로부는 스토리지(308)와 연관될 수도 있다.

[0026] 사용자는 사용자 입력 인터페이스(310)를 이용하여 명령어들을 제어 회로부(304)로 전송할 수도 있다. 사용자 입력 인터페이스(310)는 원격 제어, 마우스, 트랙볼, 키패드, 키보드, 터치 스크린, 터치패드, 스타일러스 입력, 조이스틱, 음성 인식 인터페이스, 또는 다른 사용자 입력 인터페이스들과 같은, 임의의 적합한 사용자 인터페이스일 수도 있다. 디스플레이(312)는 독립형 디바이스로서 제공되거나 또는 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301) 각각의 다른 엘리먼트들과 통합될 수도 있다. 예를 들어, 디스플레이(312)는 터치스크린 또는 터치-감지 디스플레이일 수도 있다. 이러한 상황들에서, 사용자 입력 인터페이스(310)는 디스플레이(312)와 통합되거나 또는 결합될 수도 있다. 디스플레이(312)는 모니터, 텔레비전, 모바일 디바이스용 액정 디스플레이(LCD), 비정질 실리콘 디스플레이, 저온 폴리 실리콘 디스플레이, 전자 잉크 디스플레이, 전기영동 디스플레이, 능동 매트릭스 디스플레이, 전기-습윤 디스플레이, 전기유체 디스플레이, 음극선관 디스플레이, 발광 다이오드 디스플레이, 일렉트로루미네센트 디스플레이, 플라즈마 디스플레이 패널, 고성능 어드레싱 디스플레이, 박막 트랜지스터 디스플레이, 유기 발광 다이오드 디스플레이, 표면-전도 전자-에미터 디스플레이(SED), 레이저 텔레비전, 탄소 나노튜브들, 양자점 디스플레이, 간섭 변조기 디스플레이, 또는 시각적 이미지들을 디스플레이하기 위한 임의의 다른 적합한 장비 중 하나 이상일 수도 있다. 일부 실시형태들에서, 디스플레이(312)는 HDTV 가능할 수도 있다. 일부 실시형태들에서, 디스플레이(312)는 3D 디스플레이일 수도 있으며, 대화형 애플리케이션 및 임의의 적합한 콘텐츠는 3D로 디스플레이될 수도 있다. 비디오 카드 또는 그래픽 카드는 출력을 디스플레이(312)에 생성할 수도 있다. 비디오 카드는 3D 장면들 및 2D 그래픽의 가속된 렌더링, MPEG-2/MPEG-4 디코딩, TV 출력, 또는 다수의 모니터들을 연결하는 기능과 같은, 다양한 기능들을 제공할 수도 있다. 비디오 카드는 제어 회로부(304)와 관련하여 위에서 설명된 임의의 프로세싱 회로부일 수도 있다. 비디오 카드는 제어 회로부(304)와 통합될 수도 있다. 스피커들(314)은 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301) 각각의 다른 엘리먼트들과 통합되어 제공되거나 또는 독립형 유닛들일 수도 있다. 디스플레이(312) 상에 디스플레이된 비디오들 및 다른 콘텐츠의 오디오 성분은 스피커들(314)을 통해서 플레이될 수도 있다. 일부 실시형태들에서, 오디오는 수신기(미도시)에 분배될 수도 있으며, 수신기는 스피커들(314)을 통해서 오디오를 프로세싱하여 출력한다.

- [0027] 안내 애플리케이션은 임의의 적합한 아키텍처를 이용하여 구현될 수도 있다. 예를 들어, 안내 애플리케이션은 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301) 각각 상에 전체적으로 구현된 독립형 애플리케이션일 수도 있다. 이러한 접근법에서, 애플리케이션의 명령어들은 (예컨대, 스토리지(308)에) 로컬로 저장되며, 애플리케이션에 사용하기 위한 데이터는 주기적으로 (예컨대, 대역외 공급원으로부터, 인터넷 리소스로부터, 또는 다른 적합한 접근법을 이용하여) 다운로드된다. 제어 회로부(304)는 스토리지(308)로부터 애플리케이션의 명령어들을 리트리빙하고 명령어들을 프로세싱하여 본원에서 설명된 디스플레이들 중 임의의 디스플레이를 생성할 수도 있다. 프로세싱된 명령어들에 기초하여, 제어 회로부(304)는 입력이 입력 인터페이스(310)로부터 수신될 때 수행할 액션을 결정할 수도 있다. 예를 들어, 디스플레이 상에서의 커서의 상/하 이동은, 상/하 버튼이 선택되었음을 입력 인터페이스(310)가 표시할 때, 프로세싱된 명령어들에 의해 표시될 수도 있다.
- [0028] 일부 실시형태들에서, 애플리케이션은 클라이언트/서버-기반 애플리케이션이다. 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301)의 각각 상에 구현된 썩(thick) 또는 썬(thin) 클라이언트에 사용하기 위한 데이터는 요청들을 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301) 각각에 원격인 서버로 발행함으로써 온-디맨드로 리트리빙된다. 클라이언트/서버-기반 안내 애플리케이션의 일 예에서, 제어 회로부(304)는 원격 서버에 의해 제공되는 웹 페이지들을 해석하는 웹 브라우저를 실행한다. 예를 들어, 원격 서버는 저장 디바이스에 애플리케이션을 위한 명령어들을 저장할 수도 있다. 원격 서버는 저장된 명령어들을 회로부(예컨대, 제어 회로부(304))를 이용하여 프로세싱하고 상기 및 하기에 설명하는 디스플레이들을 생성할 수도 있다. 클라이언트 디바이스는 원격 서버에 의해 생성된 디스플레이들을 수신할 수도 있으며 디스플레이들의 콘텐츠를 장비 디바이스(300) 상에 로컬로 디스플레이할 수도 있다. 이렇게, 명령어들의 프로세싱은 서버에 의해 원격으로 수행되는 반면, 결과적인 디스플레이들은 장비 디바이스(300) 상에 로컬로 제공된다. 장비 디바이스(300)는 사용자로부터의 입력들을 입력 인터페이스(310)를 통해서 수신하고 대응하는 디스플레이들을 프로세싱하여 생성하기 위해 이들 입력들을 원격 서버로 송신할 수도 있다. 예를 들어, 장비 디바이스(300)는 상/하 버튼이 입력 인터페이스(310)를 통해서 선택되었다는 것을 표시하는 통신을 원격 서버로 송신할 수도 있다. 원격 서버는 그 입력에 따라서 명령어들을 프로세싱하고 그 입력에 대응하는 애플리케이션의 디스플레이(예컨대, 커서를 상/하 이동시키는 디스플레이)를 생성할 수도 있다. 생성된 디스플레이는 그후 사용자에게의 프리젠테이션을 위해 장비 디바이스(300)로 송신된다.
- [0029] 일부 실시형태들에서, 애플리케이션은 다운로드되어 해석되거나, 또는 아니면(제어 회로부(304)에 의해 실행되는) 해석기 또는 가상 머신에 의해 실행된다. 일부 실시형태들에서, 안내 애플리케이션은 ETV 바이너리 인터체인지 포맷(EBIF)으로 인코딩되고, 제어 회로부(304)에 의해 적합한 피딩(feeding)의 부분으로서 수신되고, 제어 회로부(304) 상에서 실행하는 사용자 에이전트에 의해 해석될 수도 있다. 예를 들어, 안내 애플리케이션은 EBIF 애플리케이션일 수도 있다. 일부 실시형태들에서, 안내 애플리케이션은 로컬 가상 머신에 의해 수신되어 실행되는 일련의 자바-기반 파일들 또는 제어 회로부(304)에 의해 실행되는 다른 적합한 미들웨어에 의해 정의될 수도 있다. 이러한 실시형태들 중 일부(예컨대, MPEG-2 또는 다른 디지털 미디어 인코딩 방식들을 채용하는 것들)에서, 안내 애플리케이션은 예를 들어, 프로그램의 MPEG 오디오 및 비디오 패킷들과 함께 MPEG-2 오브젝트 캐러셀로 인코딩 및 송신될 수도 있다.
- [0030] 도 3의 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301) 각각은 사용자 텔레비전 장비(402), 사용자 컴퓨터 장비(404), 무선 사용자 통신 디바이스(406), 또는 콘텐츠에 액세스하기에 적합한 임의의 다른 유형의 사용자 장비, 예컨대 비-포터블 게이밍 머신으로서 도 4의 시스템(400)에서 구현될 수 있다. 간결성을 위해, 이들 디바이스들은 사용자 장비 또는 사용자 장비 디바이스들로서 일괄하여 본원에서 지칭될 수도 있으며, 위에서 설명된 사용자 장비 디바이스들과 실질적으로 유사할 수도 있다. 애플리케이션이 상부에 구현될 수도 있는 사용자 장비 디바이스들은 독립형 디바이스로서 기능할 수도 있거나 또는 디바이스들의 네트워크의 부분일 수도 있다. 디바이스들의 다양한 네트워크 구성들이 구현될 수도 있으며 아래에서 좀더 자세하게 설명된다.
- [0031] 도 3과 관련하여 위에서 설명되는 시스템 특징들 중 적어도 일부를 이용하는 사용자 장비 디바이스는 사용자 텔레비전 장비(402), 사용자 컴퓨터 장비(404), 또는 무선 사용자 통신 디바이스(406)로만 분류되지 않을 수도 있다. 예를 들어, 사용자 텔레비전 장비(402)는 일부 사용자 컴퓨터 장비(404)와 같이, 인터넷 콘텐츠에의 액세스를 가능하게 하는 인터넷-이용가능형일 수도 있는 반면, 사용자 컴퓨터 장비(404)는 일부 텔레비전 장비(402)와 같이, 텔레비전 프로그래밍에의 액세스를 가능하게 하는 튜너를 포함할 수도 있다. 애플리케이션은 다양한 상이한 유형들의 사용자 장비 상에서 동일한 레이아웃을 가질 수도 있거나 또는 사용자 장비의 디스플레이 능력들에 맞게 조정될 수도 있다. 예를 들어, 사용자 컴퓨터 장비(404) 상에서, 안내 애플리케이션은 웹 브라우저에 의해 액세스되는 웹사이트로서 제공될 수도 있다. 다른 예에서, 안내 애플리케이션은 무선 사용자 통신 디바이스들

(406)에 대해 축소될 수도 있다.

- [0032] 시스템(400)에서, 전형적으로 각각의 유형의 사용자 장비 디바이스 중 둘 이상이 있지만, 도면을 과도하게 복잡하게 만드는 것을 피하기 위해 단지 각각 하나만이 도 4에 도시된다. 게다가, 각각의 사용자는 복수의 유형의 사용자 장비 디바이스 및 또한 각각의 유형의 사용자 장비 디바이스 중 하나 이상을 이용할 수도 있다.
- [0033] 일부 실시형태들에서, 사용자 장비 디바이스(예컨대, 사용자 텔레비전 장비(402), 사용자 컴퓨터 장비(404), 무선 사용자 통신 디바이스(406))는 "제 2 스크린 디바이스"로서 지칭될 수도 있다. 예를 들어, 제 2 스크린 디바이스는 제 1 사용자 장비 디바이스 상에 제시된 콘텐츠를 보충할 수도 있다. 제 2 스크린 디바이스 상에 제시된 콘텐츠는 제 1 디바이스 상에 제시된 콘텐츠를 보충하는 임의의 적합한 콘텐츠일 수도 있다. 일부 실시형태들에서, 제 2 스크린 디바이스는 제 1 디바이스의 세팅들 및 디스플레이 선호사항들을 조정하기 위한 인터페이스를 제공한다. 일부 실시형태들에서, 제 2 스크린 디바이스는 다른 제 2 스크린 디바이스들과 상호작용하기 위해 또는 소셜 네트워크와 상호작용하기 위해 구성된다. 제 2 스크린 디바이스는 제 1 디바이스와는 동일한 방에, 제 1 디바이스와는 상이한 방이지만 동일한 주택 또는 빌딩에, 또는 제 1 디바이스와는 상이한 빌딩에 위치될 수 있다.
- [0034] 사용자는 또한 가정용 디바이스들 및 원격 디바이스들에 걸쳐서 일관된 애플리케이션 세팅들을 유지하기 위해 다양한 세팅들을 설정할 수도 있다. 세팅들은 본원에서 설명되는 것들뿐만 아니라, 채널 및 프로그램 즐겨찾기들, 안내 애플리케이션이 프로그래밍 추천들을 만드는데 이용되는 프로그래밍 선호사항들, 디스플레이 선호사항들, 및 다른 원하는 안내 세팅들을 포함한다. 예를 들어, 사용자가 이들의 사무실의 이들의 개인용 컴퓨터 상에 예를 들어, 웹사이트 www.Tivo.com에 대한 즐겨찾기로서 채널을 설정하면, 동일한 채널이 원할 경우, 사용자의 모바일 디바이스들뿐만 아니라, 사용자의 가정용 디바이스들(예컨대, 사용자 텔레비전 장비 및 사용자 컴퓨터 장비) 상에 즐겨찾기로서 나타날 것이다. 따라서, 하나의 사용자 장비 디바이스 상에서 이루어지는 변화들은 이들이 동일한 또는 상이한 유형들의 사용자 장비 디바이스들인지 여부에 관계없이, 다른 사용자 장비 디바이스에 대한 안내 경험을 변경할 수 있다. 게다가, 이루어진 변화들은 사용자에게 의해 입력된 세팅들뿐만 아니라, 안내 애플리케이션에 의해 모니터링된 사용자 활동에 기초할 수도 있다.
- [0035] 사용자 장비 디바이스들은 통신 네트워크(414)에 커플링될 수도 있다. 즉, 사용자 텔레비전 장비(402), 사용자 컴퓨터 장비(404), 및 무선 사용자 통신 디바이스(406)는 각각 통신 경로들(408, 410, 및 412)을 통해서 통신 네트워크(414)에 커플링된다. 통신 네트워크(414)는 인터넷, 모바일 폰 네트워크, 모바일 음성 또는 데이터 네트워크(예컨대, 4G 또는 LTE 네트워크), 케이블 네트워크, 공중 교환 전화 네트워크, 또는 다른 유형들의 통신 네트워크 또는 통신 네트워크들의 조합들을 포함한 하나 이상의 네트워크들일 수도 있다. 경로들(408, 410, 및 412)은 위성 경로, 광섬유 경로, 케이블 경로, 인터넷 통신들을 지원하는 경로(예컨대, IPTV), (예컨대, 브로드캐스트 또는 다른 무선 신호들에 대한) 자유-공간 접속들, 또는 임의의 다른 적합한 유선 또는 무선 통신 경로 또는 이러한 경로들의 조합과 같은, 하나 이상의 통신 경로들을 별개로 또는 함께 포함할 수도 있다. 경로(412)는 도 4에 나타난 예시적인 실시형태에서 무선 경로임을 표시하기 위해 점선들로 도시되며, 경로들(408 및 410)은 (원할 경우, 이들 경로들은 무선 경로들일 수도 있지만) 이들이 유선 경로들임을 표시하기 위해 실선들로 도시된다. 사용자 장비 디바이스들과의 통신들은 이들 통신 경로들 중 하나 이상에 의해 제공될 수도 있지만 도면을 과도하게 복잡하게 만드는 것을 피하기 위해 도 4에서 각각의 디바이스로의 단일 경로로서 도시된다.
- [0036] 통신 경로들이 사용자 장비 디바이스들 사이에 도시되지 않지만, 이들 디바이스들은 경로들(408, 410, 및 412)과 관련하여 위에서 설명한 것들과 같은 통신 경로들뿐만 아니라, 다른 단거리 점-대-점 통신 경로들, 예컨대 USB 케이블들, IEEE 1394 케이블들, 무선 경로들(예컨대, 블루투스, 적외선, IEEE 402-11x, 등), 또는 유선 또는 무선 경로들을 통한 다른 단거리 통신을 통해서 서로 직접 통신할 수도 있다. 블루투스는 Bluetooth SIG, INC가 소유한 인증 마크이다. 사용자 장비 디바이스들은 또한 통신 네트워크(414)를 경유해 간접 경로를 통해서 서로 직접 통신할 수도 있다.
- [0037] 시스템(400)은 원격 네트워크(424)를 포함한다. 원격 네트워크(424)는 콘텐츠 전달을 위한 복수의 서버들 및 디바이스들을 포함하는 클라우드-기반 네트워크일 수도 있다. 예를 들어, 원격 네트워크(424)는 원본(origin) 서버(417) 및 에지 서버(419)를 포함할 수도 있다. 예를 들어, 콘텐츠 전달 네트워크(CDN)는 하나 이상의 원본 서버들의 부하를 줄이기 위해 에지 서버들에게 전략적 위치들에 콘텐츠를 저장(캐시)하게 할 수도 있다. 이미지들, HTML 및 자바스크립트 파일들(및 잠재적으로 다른 콘텐츠)와 같은 정적 자산들을, 요청하는 클라이언트 머신에 가능한 한 가깝게 이동시킴으로써, 에지 서버 캐시는 웹 리소스를 로드하는데 걸리는 시간의 양을 감소시킬 수 있다. 시스템(400)은 각각 통신 경로들(420 및 422)을 통해서 통신 네트워크(414)에 커플링된 콘텐츠

소스(416) 및 미디어 안내 데이터 소스(418)를 포함한다. 경로들(420 및 422)은 경로들(408, 410, 및 412)과 관련하여 위에서 설명된 통신 경로들 중 임의의 경로를 포함할 수도 있다. 콘텐츠 소스(416) 및 미디어 안내 데이터 소스(418)와의 통신들은 하나 이상의 통신 경로들을 통해서 교환될 수도 있지만 도면을 과도하게 복잡하게 만드는 것을 피하기 위해 도 4에 경로들(420 및 422)로서 도시된다. 게다가, 각각의 콘텐츠 소스(416) 및 미디어 안내 데이터 소스(418) 중 둘 이상이 있을 수도 있지만, 도면을 과도하게 복잡하게 만드는 것을 피하기 위해 단지 각각 하나만이 도 4에 도시된다. (이들 소스들의 각각의 상이한 유형들이 이하에서 설명된다.) 원할 경우, 콘텐츠 소스(416) 및 미디어 안내 데이터 소스(418)는 하나의 소스 디바이스로서 통합될 수도 있다. 소스들(416 및 418)과 사용자 장비 디바이스들(402, 404, 및 406) 사이의 통신들이 통신 네트워크(414)를 통하는 것으로 도시되지만, 일부 실시형태들에서, 소스들(416 및 418)은 경로들(408, 410, 및 412)과 관련하여 위에서 설명한 것들과 같은 통신 경로들(미도시)을 통해서 사용자 장비 디바이스들(402, 404, 및 406)과 직접 통신할 수도 있다.

[0038] 콘텐츠 소스(416)는 텔레비전 배포 시설, 케이블 시스템 헤드엔드, 위성 배포 시설, 프로그래밍 소스들(예컨대, 텔레비전 방송사들, 예컨대 NBC, ABC, HBO 등), 중간 배포 설비 및/또는 서버들, 인터넷 제공자들, 온-디맨드 미디어 서버들, 및 다른 콘텐츠 제공자들을 포함한 하나 이상의 유형들의 콘텐츠 배포 장비를 포함할 수도 있다. NBC는 National Broadcasting Company, Inc.가 소유한 상표이고, ABC는 American Broadcasting Company, Inc.가 소유한 상표이고, HBO는 Home Box Office, Inc가 소유한 상표이다. 콘텐츠 소스(416)는 콘텐츠의 발신자(예컨대, 텔레비전 방송사, 웹캐스트 제공자 등)일 수도 있거나 또는 콘텐츠의 발신자(예컨대, 온-디맨드 콘텐츠 제공자, 다운로드를 위한 브로드캐스트 프로그램들의 콘텐츠의 인터넷 제공자 등)가 아닐 수도 있다. 콘텐츠 소스(416)는 케이블 소스들, 위성 제공자들, 온-디맨드 제공자들, 인터넷 제공자들, 오버더탑 콘텐츠 제공자들, 또는 다른 콘텐츠의 제공자들을 포함할 수도 있다. 콘텐츠 소스(416)는 또한 사용자 장비 디바이스들 중 임의의 디바이스로부터 멀리 있는 위치에 (사용자에 의해 선택된 비디오 콘텐츠를 포함하는) 상이한 유형들의 콘텐츠를 저장하는데 사용되는 원격 미디어 서버를 포함할 수도 있다. 콘텐츠의 원격 저장 및 원격 저장된 콘텐츠를 사용자 장비에 제공하기 위한 시스템들 및 방법들은 본원에서 전체적으로 참고로 포함되는, 2010년 7월 20일에 허여된 Ellis 등의 미국 특허 제 7,761,892호와 관련하여 더욱더 자세히 설명된다.

[0039] 미디어 안내 데이터 소스(418)는 위에서 설명된 미디어 안내 데이터와 같은 미디어 안내 데이터를 제공할 수도 있다. 미디어 안내 데이터는 임의의 적합한 접근법을 이용하여 사용자 장비 디바이스들에 제공될 수도 있다. 일부 실시형태들에서, 안내 애플리케이션은 데이터 피드(예컨대, 연속적인 피드 또는 트리클 피드)을 통해서 프로그램 가이드 데이터를 수신하는 독립형 대화형 텔레비전 프로그램 가이드일 수도 있다. 프로그램 스케줄 데이터 및 다른 안내 데이터는 대역내 디지털 신호를 이용하여, 대역외 디지털 신호를 이용하여, 또는 임의의 다른 적합한 데이터 송신 기법에 의해 텔레비전 채널 측파대 상에서 사용자 장비에 제공될 수도 있다. 프로그램 스케줄 데이터 및 다른 미디어 안내 데이터는 다수의 아날로그 또는 디지털 텔레비전 채널들 상에서 사용자 장비에 제공될 수도 있다.

[0040] 일부 실시형태들에서, 미디어 안내 데이터 소스(418)로부터의 안내 데이터는 클라이언트/서버 접근법을 이용하여 사용자들의 장비에 제공될 수도 있다. 예를 들어, 사용자 장비 디바이스는 서버로부터 미디어 안내 데이터를 풀(pull)할 수도 있거나, 또는 서버는 미디어 안내 데이터를 사용자 장비 디바이스로 푸시할 수도 있다. 일부 실시형태들에서, 사용자의 장비 상에 상주하는 안내 애플리케이션 클라이언트는 필요할 경우, 예컨대, 안내 데이터가 오래된 경우 또는 사용자 장비 디바이스가 데이터를 수신하라는 요청을 사용자로부터 수신하는 경우, 안내 데이터를 획득하기 위해, 소스(418)와의 세션들을 개시할 수도 있다. 미디어 안내는 임의의 적합한 빈도로(예컨대, 연속적으로, 매일, 사용자-규정된 시간 기간에, 시스템-규정된 시간 기간에, 사용자 장비로부터의 요청에 응답하여, 등등으로) 사용자 장비에 제공될 수도 있다. 미디어 안내 데이터 소스(418)는 사용자 장비 디바이스들(402, 404, 및 406)에 애플리케이션 자체 또는 애플리케이션에 대한 소프트웨어 업데이트들을 제공할 수도 있다.

[0041] 일부 실시형태들에서, 미디어 안내 데이터는 뷰어 데이터를 포함할 수도 있다. 예를 들어, 뷰어 데이터는 현재 및/또는 과거 사용자 활동 정보(예컨대, 사용자가 전형적으로 시청하는 콘텐츠, 사용자가 콘텐츠를 시청하는 일중 시간대, 사용자가 소셜 네트워크와 상호작용하는지 여부, 정보를 게시하기 위해 사용자가 소셜 네트워크와 상호작용하는 시간, 사용자가 전형적으로 시청하는 콘텐츠의 유형들(예컨대, 유료 TV 또는 무료 TV), 분위기, 두뇌 활동 정보, 등)을 포함할 수도 있다. 미디어 안내 데이터는 또한 가입 데이터를 포함할 수도 있다. 예를 들어, 가입 데이터는 주어진 사용자가 가입하는 소스들 또는 서비스들 및/또는 주어진 사용자가 이전에 가입되었지만 나중에 액세스를 종료한 소스들 또는 서비스들(예컨대, 사용자가 프리미엄 채널들에 가입하는지 여부, 사용자가 프리미엄 레벨의 서비스들을 추가하였는지 여부, 사용자가 인터넷 속도를 높였는지 여부)을 식별할 수

도 있다. 일부 실시형태들에서, 뷰어 데이터 및/또는 가입 데이터는 1년 이상의 기간 동안 주어진 사용자의 패턴들을 식별할 수도 있다. 미디어 안내 데이터는 주어진 사용자가 서비스/소스에의 액세스를 종료할 가능성을 표시하는 스코어를 생성하는데 사용되는 모델(예컨대, 생존자 모델)을 포함할 수도 있다. 예를 들어, 애플리케이션은 주어진 사용자가 특정의 서비스 또는 소스에의 액세스를 종료할지 여부의 가능성을 표시하는 값 또는 스코어를 생성하기 위해 모델을 이용하여 가입 데이터로 뷰어 데이터를 프로세싱할 수도 있다. 특히, 더 높은 스코어는 사용자가 특정의 서비스 또는 소스에의 액세스를 종료할 것이라는 더 높은 신뢰도를 표시할 수도 있다. 스코어에 기초하여, 애플리케이션은 사용자가 액세스를 종료할 가능성이 있는 것으로 스코어에 의해 표시된 특정의 서비스 또는 소스를 유지하도록 사용자를 유도하는 프로모션들을 생성할 수도 있다.

[0042]

애플리케이션들은 예를 들어, 사용자 장비 디바이스들 상에 구현되는 독립형 애플리케이션들일 수도 있다. 예를 들어, 애플리케이션은 스토리지(308)에 저장되어 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301) 각각의 제어 회로부(304)에 의해 실행될 수도 있는 소프트웨어 또는 실행가능한 명령어들의 세트로서 구현될 수도 있다. 일부 실시형태들에서, 애플리케이션들은 단지 클라이언트 애플리케이션이 사용자 장비 디바이스 상에 상주하고 서버 애플리케이션이 원격 서버 상에 상주하는 클라이언트/서버 애플리케이션들일 수도 있다. 예를 들어, 애플리케이션들은 사용자 장비 디바이스(300) 및 사용자 장비 시스템(301) 각각의 제어 회로부(304) 상에 클라이언트 애플리케이션으로서 부분적으로, 그리고, 원격 서버 상에 원격 서버의 제어 회로부 상에서 실행하는 서버 애플리케이션(예컨대, 미디어 안내 데이터 소스(418))으로서 부분적으로 구현될 수도 있다. 원격 서버(예컨대, 미디어 안내 데이터 소스(418))의 제어 회로부에 의해 실행될 때, 애플리케이션은 안내 애플리케이션 디스플레이들을 생성하고 생성된 디스플레이들을 사용자 장비 디바이스들로 송신하도록 제어 회로부에게 명령할 수도 있다. 서버 애플리케이션은 사용자 장비 상에 저장을 위한 데이터를 송신하도록 미디어 안내 데이터 소스(418)의 제어 회로부에게 명령할 수도 있다. 클라이언트 애플리케이션은 안내 애플리케이션 디스플레이들을 생성하도록 수신 사용자 장비의 제어 회로부에게 명령할 수도 있다.

[0043]

사용자 장비 디바이스들(402, 404, 및 406)로 전달된 콘텐츠 및/또는 미디어 안내 데이터는 오버더탑(OTT) 콘텐츠일 수도 있다. OTT 콘텐츠 전달은 위에서 설명된 임의의 사용자 장비 디바이스를 포함한, 인터넷-이용가능 사용자 디바이스들로 하여금, 케이블 또는 위성 접속들을 통해서 수신되는 콘텐츠에 추가하여, 위에서 설명된 임의의 콘텐츠를 포함한, 인터넷을 통해서 전송되는 콘텐츠를 수신할 수 있게 한다. OTT 콘텐츠는 인터넷 서비스 제공자(ISP)에 의해 제공되는 인터넷 접속을 통해서 배포되지만, 제 3 자가 콘텐츠를 배포한다. ISP는 콘텐츠의 시청 능력들, 저작권들, 또는 재배포에 대해 책임을 지지 않을 수도 있으며, 단지 OTT 콘텐츠 제공자에 의해 제공되는 IP 패킷들만 전송할 수도 있다. OTT 콘텐츠 제공자들의 예들은 IP 패킷들을 통해서 오디오 및 비디오를 제공하는, YOUTUBE, NETFLIX, 및 HULU를 포함한다. YouTube는 Google Inc.가 소유한 상표이고, Netflix는 Netflix Inc.가 소유한 상표이며, Hulu는 Hulu, LLC가 소유한 상표이다. OTT 콘텐츠 제공자들은 추가적으로 또는 대안적으로, 위에서 설명된 미디어 안내 데이터를 제공할 수도 있다. 콘텐츠 및/또는 미디어 안내 데이터에 추가하여, OTT 콘텐츠의 제공자들은 애플리케이션들(예컨대, 웹-기반 애플리케이션들 또는 클라우드-기반 애플리케이션들)을 배포할 수 있거나, 또는 콘텐츠는 사용자 장비 디바이스 상에 저장된 애플리케이션들에 의해 디스플레이될 수 있다.

[0044]

미디어 안내 시스템(400)은 사용자 장비 디바이스들 및 콘텐츠 및 안내 데이터의 소스들이 콘텐츠에 액세스하고 미디어 안내를 제공하려는 목적을 위해 서로 통신할 수도 있는 다수의 접근법들, 또는 네트워크 구성들을 제시하기 위한 것이다. 본원에서 설명하는 실시형태들은 이들 접근법들의 임의의 하나 또는 서브세트에, 또는 콘텐츠를 배포하고 미디어 안내를 제공하기 위한 다른 접근법들을 채용하는 시스템에 적용될 수도 있다. 다음 4개의 접근법들은 도 4의 일반적인 예의 특정의 예시들을 제공한다.

[0045]

하나의 접근법에서, 사용자 장비 디바이스들은 홈 네트워크 내에서 서로 통신할 수도 있다. 사용자 장비 디바이스들은 위에서 설명된 단거리 점-대-점 통신 방식들을 통해서, 허브 또는 홈 네트워크 상에 제공되는 다른 유사한 디바이스를 경유한 간접 경로들을 통해서, 또는 통신 네트워크(414)를 통해서, 서로 직접 통신할 수 있다. 단일 홈에서의 다수의 개인들 각각은 홈 네트워크 상에서 상이한 사용자 장비 디바이스들을 운영할 수도 있다. 그 결과, 다양한 미디어 안내 정보 또는 세팅들이 상이한 사용자 장비 디바이스들 사이에 통신되는 것이 바람직할 수도 있다. 예를 들어, 본원에 전체적으로 참고로 포함되는, 2005년 7월 11일에 출원된, Ellis 등의, 미국 특허 공보 번호 제 2005/0251827호에 더욱더 자세히 설명된 바와 같이, 사용자들이 홈 네트워크 내 상이한 사용자 장비 디바이스들 상에서 일관된 애플리케이션 세팅들을 유지하는 것이 바람직할 수도 있다. 홈 네트워크에서의 상이한 유형들의 사용자 장비 디바이스들은 또한 콘텐츠를 송신하기 위해 서로 통신할 수도 있다. 예를 들어, 사용자는 콘텐츠를 사용자 컴퓨터 장비로부터 포터블 비디오 플레이어 또는 포터블 뮤직 플레이어로 송신

할 수도 있다.

- [0046] 제 2 접근법에서, 사용자들은 이들이 콘텐츠에 액세스하여 미디어 안내를 획득하는 다수의 유형들의 사용자 장비를 가질 수도 있다. 예를 들어, 일부 사용자들은 가정용 및 모바일 디바이스들에 의해 액세스되는 홈 네트워크들을 가질 수도 있다. 사용자들은 원격 디바이스 상에 구현된 애플리케이션을 통해서 가정용 디바이스들을 제어할 수도 있다. 예를 들어, 사용자들은 이들의 사무실에 있는 개인용 컴퓨터, 또는 모바일 디바이스, 예컨대 PDA 또는 웹-이용가능 모바일 전화기를 통해서 웹사이트 상에서 온라인 애플리케이션에 액세스할 수도 있다. 사용자는 사용자의 가정용 장비를 제어하기 위해 온라인 안내 애플리케이션 상에서 다양한 세팅들(예컨대, 리코딩들, 알람들(reminders), 또는 다른 세팅들)을 설정할 수도 있다. 온라인 가이드는 사용자의 장비를 직접 제어하거나, 또는 사용자의 가정용 장비 상의 애플리케이션과 통신하여 제어할 수도 있다. 서로 멀리 떨어져 있는 위치들에 있는 사용자 장비 디바이스들이 통신하는 다양한 시스템들 및 방법들은 예를 들어, 본원에서 전체적으로 참고로 포함되는, 2011년 10월 25일에 허여된, Ellis 등의, 미국 특허 제 8,046,801호에 설명되어 있다.
- [0047] 제 3 접근법에서, 홈 내부 및 외부의 사용자 장비 디바이스들의 사용자들은 콘텐츠에 액세스하기 위해 이들의 애플리케이션을 이용하여, 콘텐츠 소스(416)와 직접 통신할 수 있다. 구체적으로 설명하면, 홈 내에서, 사용자 텔레비전 장비(402) 및 사용자 컴퓨터 장비(404)의 사용자들은 애플리케이션에 액세스하여 원하는 콘텐츠를 탐색하고 위치추적할 수도 있다. 사용자들은 또한 액세스 무선 사용자 통신 디바이스들(406)을 이용하여 홈 외부의 애플리케이션에 액세스하여 원하는 콘텐츠를 탐색하고 위치추적할 수도 있다.
- [0048] 제 4 접근법에서, 사용자 장비 디바이스들은 클라우드 서비스들에 액세스하기 위해 클라우드 컴퓨팅 환경에서 동작할 수도 있다. 클라우드 컴퓨팅 환경에서, 콘텐츠 공유, 저장 또는 배포를 위한 다양한 유형들의 컴퓨팅 서비스들(예컨대, 비디오 공유 사이트들 또는 소셜 네트워킹 사이트들)은 "클라우드"로서 지칭되는, 네트워크-액세스가능한 컴퓨팅 및 저장 리소스들의 컬렉션에 의해 제공된다. 예를 들어, 클라우드는 통신 네트워크(414)를 통해 인터넷과 같은 네트워크를 통해서 접속된 다양한 유형들의 사용자들 및 디바이스들에 클라우드-기반 서비스들을 제공하는, 중앙집중식으로 또는 분산된 위치들에 위치될 수도 있는, 서버 컴퓨팅 디바이스들의 컬렉션을 포함할 수 있다. 이들 클라우드 리소스들은 하나 이상의 콘텐츠 소스들(416) 및 하나 이상의 미디어 안내 데이터 소스들(418)을 포함할 수도 있다. 추가적으로 또는 대안적으로, 원격 컴퓨팅 사이트들은 다른 사용자 장비 디바이스들, 예컨대 사용자 텔레비전 장비(402), 사용자 컴퓨터 장비(404), 및 무선 사용자 통신 디바이스(406)를 포함할 수도 있다. 예를 들어, 다른 사용자 장비 디바이스들은 비디오 또는 스트리밍된 비디오의 저장된 복사본에의 액세스를 제공할 수도 있다. 이러한 실시형태들에서, 사용자 장비 디바이스들은 중앙 서버와 통신함이 없이 피어-투-피어 방법으로 동작할 수도 있다.
- [0049] 클라우드는 다른 예들 중에서, 콘텐츠 저장, 콘텐츠 공유, 또는 소셜 네트워킹 서비스들과 같은, 서비스들에의 액세스뿐만 아니라, 사용자 장비 디바이스들에 대해 위에서 설명된 임의의 콘텐츠에의 액세스를 제공한다. 서비스들은 클라우드 컴퓨팅 서비스 제공자들을 통해서, 또는 온라인 서비스들의 다른 제공자들을 통해서 클라우드에서 제공될 수 있다. 예를 들어, 클라우드-기반 서비스들은 콘텐츠 저장 서비스, 콘텐츠 공유 사이트, 소셜 네트워킹 사이트, 또는 접속된 디바이스들 상에서 다른 사람들이 볼 수 있도록 하기 위해 사용자-소싱된 콘텐츠가 배포되는 다른 서비스들을 포함할 수 있다. 이들 클라우드-기반 서비스들은 콘텐츠를 로컬로 저장하고 저장된 콘텐츠에 로컬로 액세스하는 대신, 사용자 장비 디바이스로 하여금, 클라우드에 콘텐츠를 저장하게 하고 클라우드로부터 콘텐츠를 수신하게 할 수도 있다.
- [0050] 사용자는 캠코더들, 비디오 모드를 가진 디지털 카메라들, 오디오 리코더들, 모바일 폰들, 및 핸드헬드 컴퓨팅 디바이스들과 같은 다양한 콘텐츠 캡처 디바이스들을 이용하여 콘텐츠를 리코딩할 수도 있다. 사용자는 예를 들어, 콘텐츠 캡처 특징을 가지는 사용자 컴퓨터 장비(404) 또는 무선 사용자 통신 디바이스(406)로부터 직접 클라우드 상의 콘텐츠 저장 서비스에 콘텐츠를 업로드할 수 있다. 대안적으로, 사용자는 먼저 콘텐츠를 사용자 컴퓨터 장비(404)와 같은 사용자 장비 디바이스로 전송할 수 있다. 콘텐츠를 저장하는 사용자 장비 디바이스는 통신 네트워크(414) 상의 데이터 송신 서비스를 이용하여 콘텐츠를 클라우드에 업로드한다. 일부 실시형태들에서, 사용자 장비 디바이스 자체는 클라우드 리소스이고, 다른 사용자 장비 디바이스들은 사용자가 콘텐츠를 저장한 사용자 장비 디바이스로부터 직접 콘텐츠에 액세스할 수 있다.
- [0051] 클라우드 리소스들은 예를 들어, 웹 브라우저, 애플리케이션, 데스크탑 애플리케이션, 모바일 애플리케이션, 및/또는 이들의 액세스 애플리케이션들의 임의의 조합을 이용하여 사용자 장비 디바이스에 의해 액세스될 수도 있다. 사용자 장비 디바이스는 애플리케이션 전달을 위해 클라우드 컴퓨팅에 의존하는 클라우드 클라이언트일 수도 있거나, 또는 사용자 장비 디바이스는 클라우드 리소스들에 대한 액세스 없이 일부 기능을 가질 수도 있다.

예를 들어, 사용자 장비 디바이스 상에서 실행하는 일부 애플리케이션들은 클라우드 애플리케이션들, 즉, 인터넷을 통한 서비스로서 배포된 애플리케이션들일 수도 있는 반면, 다른 애플리케이션들은 사용자 장비 디바이스 상에 저장되어 실행될 수도 있다. 일부 실시형태들에서, 사용자 디바이스는 다수의 클라우드 리소스들로부터 동시에 콘텐츠를 수신할 수도 있다. 예를 들어, 사용자 디바이스는 제 2 클라우드 리소스로부터 콘텐츠를 다운로드하는 동안 하나의 클라우드 리소스로부터 오디오를 스트리밍할 수 있다. 또는, 사용자 디바이스는 좀더 효율적인 다운로드를 위해 다수의 클라우드 리소스들로부터 콘텐츠를 다운로드할 수 있다. 일부 실시형태들에서, 사용자 장비 디바이스들은 도 3과 관련하여 설명된 프로세싱 회로부에 의해 수행되는 프로세싱 동작들과 같은 프로세싱 동작들에 대해 클라우드 리소스들을 이용할 수 있다.

[0052] 본원에서 설명되는 방법들 및 시스템들은 시맨틱 그래프들과 머신 학습의 조합을 이용하여, 구조화된 데이터를 자동으로 생성하고, 중요한 엔티티들/키워드들을 인식하고, 그리고 더 관련된 검색 결과들 및 추천들을 생성하는 가중된 연결들을 생성한다. 더 관련된 검색 결과들 및 추천들이 얻어지는 레이트의 예가 도 5에 도시된다. 도 5는 상위 10,000개의 영화들의 수동으로 선별된 리스트의 테스트 분할된 예시적인 모델에 대한 결과들 테이블(테이블(500))이다. 테이블에 포함된 것은 그래프 특징들을 사용하거나 사용하지 않고 의사결정 트리 분류기가 실행되었을 때의 정밀도(precision), 재현율(recall), 및 F1 스코어들이다. F1 스코어는 (아래에서 설명되는 바와 같이) 정밀도 및 재현율을 고려하여 수행된 테스트의 정확도의 척도이다. 정밀도는 올바른 긍정적인 결과들의 개수를 분류기에 의해 반환된 모든 긍정적인 결과들의 개수로 나눈 값이다. 재현율은 올바른 긍정적인 결과들의 개수를 모든 관련된 샘플들(긍정적으로 식별된 모든 샘플들)의 개수로 나눈 값이다. 그후 정밀도 및 재현율의 조화 평균이 취해져 F1 스코어를 생성한다. F1 스코어들은 (완전한 정밀도 및 재현율을 표시하는) 1부터 0까지의 범위이다. 나타낸 바와 같이, 그래프 특징이 없는 모델이 고 품질 엔티티들과 저 품질 엔티티 사이를 구분할 수 없기 때문에, 재현율은 그래프 특징들이 없는 모델에서 더 높고, 정밀도는 예상대로 낮다. 따라서, 본원에서 시맨틱 그래프들을 이용함으로써, 검색, 추천, 및 발견 특징들은 더 높은 정밀도 및 F1 스코어들을 갖는 결과들을 얻을 수 있다. 예를 들어, 시맨틱 그래프를 이용하여, 시스템은 더 관련된 요청들을 반환할 뿐만 아니라, 주어진 키워드에 관련된 엔티티들의 유니버스를 결정하기 위해, 엔티티들(예컨대, 영화 내 또는 영화에 관한 키워드, 영화 내 오브젝트, 주요 플롯 지점, 등)을 순위지정할 수 있다. 엔티티들은 시맨틱 그래프에서의 노드들에 대응할 수도 있으며, 이들 노드들의 각각은 더 높거나 또는 낮게 평가될 수도 있다.

[0053] 도 5에서, 시스템은 그의 결과들을 엔티티들의 수동으로 선별된 리스트와 비교하여 모델의 정밀도 및 재현율을 측정한다. 시스템은 머신-생성 엔티티들의 총 개수(K)에 대한 수동으로 선별된 리스트(N)에 매칭하는 머신-생성 엔티티들의 개수의 비율로서 정밀도를 정의한다.

[0054]
$$\text{정밀도} = \frac{N}{K}$$

[0055] 재현율은 시스템에 의해, 수동으로-선별된 엔티티들의 개수(M)에 대한 모델에 의해 추출되는 수동으로-선별된 엔티티들(N)의 비율로서 측정된다.

[0056]
$$\text{재현율} = \frac{N}{M}$$

[0057] 도 6은 본원에서 설명되는 검색, 추천, 및 발견 특징들을 제공하는데 사용되는 아키텍처의 예시적인 예이다. 도 6에 나타낸 바와 같이, 시스템은 데이터 세트를 수집하고 키 엔티티들 및 이들의 연관들을 식별하는 시맨틱 그래프를 생성한다. 시맨틱 그래프 및 데이터 세트로부터의 특징들은 머신-학습 모델을 통과하여 가장 문맥상 중요한 엔티티들을 추론한다. 프로세스는 4개의 스테이지들: 대명사 분석, 후보 식별, 시맨틱 그래프의 생성, 및 사용자 입력의 프로세싱을 수반한다.

[0058] 단계 602에서, 시스템은 데이터 세트를 수집한다. 예를 들어, 사용자는 기지의 데이터 세트로부터 텍스트 문자열들을 입력할 수도 있다. 추가적으로 또는 대안적으로, 시스템은 웹 크롤러를 이용하여 데이터를 수집하여 데이터 세트를 팝ULATE할 수도 있다. 일부 실시형태들에서, 시맨틱 그래프를 구축하기 위해, 시스템은 특정의 데이터 세트에 대해 트레이닝된다. 데이터 세트는 시스템이 수신할 가능성이 있는 입력들에 기초하여 선택된다. 특히, 시스템은 전형적인 대화의 사용자 톤을 반영하는 데이터에 대해 트레이닝된다. 적합한 톤을 특징으로 하는 대화를 획득하기 위해, 선택된 데이터 세트는 사용자 협업에 기초한 콘텐츠 및 사용자-생성된/수정된 콘텐츠와 같은 특정의 기준들을 특징으로 하는 데이터 세트들에 기초한다. 일부 실시형태들에서, 콘텐츠는 데이터 수집을 용이하게 하기 위해, 단순화된 마크업 언어들을 특징으로 하는 포럼들에서 추가로 선택된다. 예를 들어, 시스템은 위키(wiki) 웹사이트로부터 데이터를 가져올 수도 있다. 이들 소스들로부터의 데이터를 이용함으로써,

시스템은 사용자들로부터의 요청들의 전형적인 톤을 반영하도록 모델의 트레이닝을 향상시킬 수 있다.

- [0059] 추가적으로 또는 대안적으로, 시스템은 사용자 쿼리들의 전형적인 대화의 콘텐츠를 반영하는 데이터에 대해 트레이닝된다. 특히, 시스템은 위키 플롯 섹션들, 시놉시스 섹션들, 플롯 섹션들의 카테고리 참조들, 및 플롯으로부터의 명사 청크들로부터 그의 데이터 세트를 가져올 수도 있다. 데이터의 이들 특정의 유형들을 이용함으로써, 시스템은 사용자들로부터의 요청들의 전형적인 콘텐츠를 반영하도록 모델의 트레이닝을 향상시킬 수 있다.
- [0060] 데이터 세트는 그후 트레이닝 모델을 구축하기 위해 트레이닝 데이터 대 검증 데이터의 70:30 비로 분할될 수 있다. 예를 들어, 모델은 트레이닝 데이터 세트에 대해 트레이닝될 수도 있다. 트레이닝 데이터 세트는 검색 결과들 및 추천들에 대한, 중요한 엔티티들/키워드들을 인식하고 가중된 연결들을 생성하는 것과 같은, 모델의 파라미터들(예컨대, 시맨틱 그래프에서의 노드들 사이의 연결들의 가중치들)을 나타낸다. 모델(예컨대, 신경망 또는 나이트 베이스 분류기)은 그후 지도 학습 방법(예컨대, 경사 하강법 또는 확률적 경사 하강법)을 이용하여 트레이닝 데이터 세트에 대해 트레이닝된다. 예를 들어, 시스템은 추론된 엔티티들이 주어진 검색 요청에 관련 되는지 여부를 결정할 수도 있다. 모델이 트레이닝 데이터 세트에 대해 트레이닝되어 결과들을 생성함에 따라, 시스템은 결과들을 실제 결과(또는, 목표 결과들)와 비교할 수 있다. 비교의 실제 결과 및 사용중인 특정의 학습 알고리즘에 기초하여, 모델의 파라미터들이 조정된다. 반복 프로세스를 통해서, 시스템은 트레이닝된 모델에 피팅하여, 검색 결과들 및 추천들에 대한 사용자 검색 쿼리들에서 발견될 수도 있는 중요한 엔티티들/키워드들을 예측한다.
- [0061] 단계 604에서, 시스템은 대명사 분석을 수행한다. 대명사 분석은 풍부하고 정확한 시맨틱 그래프들에 필요한 엔티티 관계들을 식별하는데 중요하다. 프로세스의 이 단계에서, 시스템은 텍스트 문자열의 문장들에 걸쳐서 모든 대명사들을 분석한다. 예를 들어, 시스템은 대명사가 지칭하는 명사 또는 고유 명사(예컨대, "명사 청크")를 결정하는데 이용가능한 단-대-단 신경 코레퍼런스 분석의 Python 구현을 이용할 수도 있다.
- [0062] 코레퍼런스는 텍스트에서의 2개 이상의 표현들(예컨대, 대명사들, 어구들, 오브젝트들, 등)이 동일한 것(예컨대, 고유 명사)을 지칭할 때 발생한다. 예를 들어, 텍스트 문자열 "Bill said he would come"에서, 고유 명사 "Bill" 및 대명사 "he"는 동일한 사람 - Bill을 지칭한다. 코레퍼런스는 선택스의 분야에서 바인딩(binding) 현상들의 기초가 되는 메인 개념이다. 일부 실시형태들에서, 시스템은 대명사들을 분석하기 위한 신경망을 개발할 수도 있다. 예를 들어, 시스템은, 제어 회로부(304)를 통해서, 텍스트 문자열 "John helped Mary. He is doctor"을 수신할 수도 있다. 시스템은 대명사들을 분석하여, 분석된 텍스트 문자열, "John helped Mary. John is doctor"를 생성할 수도 있다.
- [0063] 종래의 시스템들에서, 시스템은 먼저 입력 문서를 검토하여 엔티티들(예컨대, 대명사들)의 멘션들을 검출한다. 시스템은 그후 각각의 대명사 클러스터가 동일한 고유 명사에 대응하도록 엔티티들(예컨대, 대명사들)을 클러스터링한다. 이들 단계들을 수행하기 위해, 시스템은 검출 및 클러스터링을 위해, 파서 및 사전-프로세싱에 의존할 수도 있다. 단-대-단 신경 코레퍼런싱(coreferencing)에서, 시스템은 엔티티들 사이의 모든 스캔들을 고려하고, 엔티티들 사이의 스캔들을 순위지정하고, 검색 공간들을 정리하기 위한 인수분해된(factored) 모델을 생성할 것이다. 시스템은 그후 주어진 대명사가 지칭하는 명사 청크를 높은 확률로 검출할 수도 있다.
- [0064] 스캔 순위에 대해, 시스템은 입력 문서에서의 각각의 스캔을 프로세싱하고 모든 스캔에 선행항목을 할당할 것이다. 일부의 경우, 시스템은 암시적인 스캔들을 생성한다. 결과적인 클러스터는 시스템으로 하여금 3개의 유형들의 스캔들: i) 이전 멘션을 갖지 않는 스캔들; ii) 이전 링크들을 갖지 않는 멘션들; 및 iii) 예측된 코레퍼런스 링크를 갖는 스캔들을 식별하게 할 것이다. 각각의 스캔에 대해, 시스템은 독립적인 결정을 행하여, 2개의 스캔들 사이의 코레퍼런스의 가능성을 결정하는 쌍방식(pairwise) 코레퍼런스 스코어를 적용할 것이다. 시스템은 그후 최고 스코어를 가진 쌍에 기초하여 선행항목을 결정할 것이다. 단-대-단 코레퍼런싱에 관한 추가적인 설명은 Proceedings of Empirical Methods in Natural Language Processing(EMNLP 2017)에서, 페이지 188-197, 2017년, Lee 등의, End-to-end Neural Coreference Resolution에서 찾을 수 있으며, 이는 본원에 전체적으로 참고로 포함된다.
- [0065] 단계 606에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 후보 식별을 수행한다. 예를 들어, 시스템은 프로세싱된 텍스트에 대해 POS(Part-Of-Speech) 태깅을 적용하여, 모든 명사 청크들을 시맨틱 그래프에서의 노드들로서 식별할 수도 있다. 문법적 태깅 또는 단어-카테고리 명확화라고도 불리는, 품사 태깅(POS 태깅 또는 PoS 태깅 또는 POST)은 해당 정의 및 해당 컨텍스트(즉, 어구, 문장, 또는 단락에서 근접 및 관련된 단어들과의 관계) 둘 다에 기초하여, 텍스트(코퍼스) 내 단어를 특정의 품사에 대응하는 것으로 표시하는 프로세스이다.

- [0066] 예를 들어, 시맨틱 그래프를 구축하기 위해, 시스템은 텍스트에서의 각각의 단어에 대해 단어-카테고리를 결정할 수도 있다. 단어-카테고리들은 8개의 품사들 (예컨대, 등위 접속사들, 종속 접속사들, 접속 부사들, 상관 접속사들, 및/또는 감탄사들을 포함한, 명사들, 동사들, 형용사들, 부사들, 전치사들, 접속사들의 각각을 포함할 수도 있다. 이들 품사들, 및 시맨틱 그래프(즉, 개념)의 각각의 단어에 대한 품사를 표시하는 메타데이터는 (예컨대, 그래프에서의 노드들을 표현하는) 단어들이 함께 결합되어 해석가능한 문장들을 만드는 방법을 결정하기 위해 시스템에 의해 사용된다. 일부 실시형태들에서, POS 태깅은 설명 태그들의 세트에 따라서, 별개의 용어들 뿐만 아니라 숨겨진 품사들을 연결하는 알고리즘들을 이용하여, 컴퓨터 언어학의 컨텍스트에서 이루어진다. POS-태깅 알고리즘들은 2개의 독특한 그룹들: 규칙-기반 및 확률론으로 나누어진다. 규칙-기반 POS 태깅의 경우, 시스템은 일련의 수동 규칙들을 통해서 수동으로 구축된다. 예를 들어, 시스템은 태깅된 단어에 선행하는 단어가 if-then 문들을 통해서 특정의 방법으로 태깅된다는 것을 표시하는 규칙을 포함할 수도 있다. 통계적(또는, 확률론적) 품사 태깅은 각각의 단어가 알려져 있으며 가능한 태그들의 유한한 세트를 갖고 있다고 가정한다. 이들 태그들은 사전 또는 형태학적 분석으로부터 도출될 수 있다. 예를 들어, 단어가 복수의 가능한 태그를 가질 때, 시스템은 통계적 방법들을 이용하여 품사 태그들의 시퀀스를 결정할 수도 있다. 시스템은 또한 규칙-기반과 확률론을 결합하는 하이브리드 접근법을 이용할 수도 있다. 마지막으로, 일부 실시형태들에서, POS 태깅이 수동으로 수행될 수도 있다는 점에 유의해야 한다.
- [0067] POS 태깅을 수행하기 위해, 시스템은 고급 자연어 프로세싱을 위해 소프트웨어 라이브러리를 이용할 수도 있다. 일부 실시형태들에서, 시스템은 고급 자연어 프로세싱을 위한 SpaCy, Python 라이브러리를 이용하여, 해당 POS 태깅 능력을 통해 식별을 강화할 수도 있다. POS 태깅에 더해서, 시스템은 비-과도적 토큰화, 명명된 엔티티 인식, 다수의 언어들에 대한 통계적 모델들, 사전-트레이닝된 단어 벡터들, 라벨링된 종속성 파싱, 선택스-기반 문장 세그멘테이션, 텍스트 분류, 선택스 및 명명된 엔티티들에 대한 빌트인 시각화 장치들, 및/또는 심층 학습 통합과 같은, 추가적인 특징들을 이용할 수도 있다.
- [0068] 단계 608에서, 시스템은 시맨틱 그래프를 생성한다. 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스이다. 시스템은 시맨틱 그래프를 지식 표현의 형태로 이용한다. 이는 개념들 및/또는 엔티티들을 나타낼 수도 있는 노드들, 및 개념들 및/또는 엔티티들 사이의 시맨틱 관계들을 나타내는 에지들로 구성되는, 방향성(예컨대, 에지들이 그들과 연관된 방향을 갖는 경우, 에지들에 의해 연결된 정점들의 세트로 구성되는 그래프) 및/또는 비방향성 그래프이다. 아래에 설명된, 도 7은 예시적인 시맨틱 그래프를 제공한다. 예를 들어, 텍스트 문자열("Jack is doctor".)에 나타나는 후보들(예컨대, "Jack", "doctor", 등)의 각각에 대해, 시맨틱 그래프는 이들 용어들 사이의 관계들을 표시할 수도 있다. 이러한 예에서, 후보들은 시맨틱 그래프들에서의 정점들을 나타낼 수도 있는 반면, 후보들(예컨대, "is") 사이의 관계는 시맨틱 그래프의 에지들에 의해 나타내어진다. 더욱이, 제 2 텍스트 문자열(예컨대, "He has an office on First Street".)에서, 시맨틱 그래프는 용어 "he"와 "Jack" 사이의 관계들을 표시할 수도 있다. 이들 용어들 사이의 관계들은 시맨틱 네트워크를 생성하는(POS 태깅에 기초하여 생성된) 종속성 트리들을 인터리빙하면서 종속성 트리를 횡단함으로써 발견될 수도 있다. 예를 들어, 일부 실시형태들에서, 시스템은 연결들이 동사들을 통해서 이루어지는 것으로 결정하고 비방향성 그래프(즉, 에지들이 방향을 갖지 않는 그래프)가 이들 에지들을 이용하여 생성된다. 시맨틱 그래프(700)에서, "Jack" 및 "doctor"가 "is"에 의해 연결된다. 종속성 트리에서, "is"가 용어 "Jack"과 "doctor"를 연결한다.
- [0069] 일부 실시형태들에서, 종속성 트리는 일부 컨텍스트-프리 문법(context-free grammar)에 따라 문자열의 구문 구조를 나타낼 수도 있다. 종속성 트리는 구성 문법들(어구 구조 문법들)의 구성 관계 또는 종속성 문법들의 종속성 관계에 기초하여 구성될 수도 있다. 종속성 트리는 프로그래밍 언어들과 같은 컴퓨터 언어들, 프로세싱 동안뿐만 아니라 자연어들에서의 문장들에 대해서도 생성될 수 있다.
- [0070] 일부 실시형태들에서, 시스템은 의사결정 트리 분류기 및 랜덤 포레스트 분류기를 이용해 트레이닝한다. 의사결정 트리 분류기는 플로우차트형 구조로서, 각각의 내부(넢-리프) 노드는 속성에 대한 테스트를 나타내고, 각각의 분기부는 테스트의 결과를 나타내고, 각각의 리프(또는, 터미널) 노드는 클래스 라벨을 보유한다. 트리에서의 최상부 노드는 루트 노드이다. 랜덤 포레스트 분류기는 트레이닝 시간에 다수의 의사결정 트리들을 구성하고 개개의 트리들의 클래스들의 모드(분류) 또는 평균 예측(회귀)인 클래스를 출력함으로써 동작할 수도 있다. 랜덤 결정 포레스트들은 이들의 트레이닝 세트에 오버피팅하는 의사결정 트리의 습관을 수정한다. 시스템이 임의의 결정-트리 알고리즘들을 구현할 수도 있다는 점에 유의해야 한다.
- [0071] 시맨틱 그래프는 그래프의 노드들에 의해 정의된다. 각각의 노드는 그의 중심성에 의해 추가로 정의된다. 중심성의 4개의 유형들은 연결(degree), 근접, 매개, 및 내향연결(indegree)을 포함한다. 연결 및 내향연결의 중심

성과는 반대로, 시맨틱 그래프는 그것의 근접성(closeness) 및 매개성(betweenness)에 의해 정의된다. 예를 들어, 연산 중에, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 근접 중심성 및 매개 중심성에 기초하여 그래프 특징들을 결정한다. 근접 중심성과 관련하여, 노드의 근접 중심성(또는, 근접성)은 노드와 그래프에서의 모든 다른 노드들 사이의 최단 경로들의 길이의 총합으로서 계산되는, 네트워크의 중심성을 측정한다. 따라서, 노드는 중심에 있을수록, 모든 다른 노드들에 더 가까워진다. 노드 $C(x)$ 의 근접 중심성은 다음으로 표시된다:

$$C(x) = \frac{N}{\sum_y d(y, x)}$$

[0072]

여기서, $d(y, x)$ 는 노드 x 와 y 사이의 거리이고 N 은 노드들의 개수이다.

[0073]

매개 중심성과 관련하여, "매개" 중심성은 최단 경로들에 기초한 그래프에서의 중심성의 척도이다. 연결된 그래프에서의 노드들의 모든 쌍에 대해, (비가중된 그래프들의 경우) 그 경로가 통과하는 에지들의 개수 또는 (가중된 그래프들의 경우) 에지들의 가중치들의 총합이 최소화되도록 노드들 사이에 적어도 하나의 최단 경로가 존재한다. 각각의 노드에 대한 매개 중심성은 정점을 통과하는 이들 최단 경로들의 개수이다. 매개 중심성 $g(v)$ 는 다음으로 표시된다:

$$g(v) = \sum_{s, t \in V} \frac{\sigma(s, t/v)}{\sigma(s, t)}$$

[0075]

여기서, V 는 노드들의 세트이고, $\sigma(s, t)$ 는 최단 (s, t) -경로들의 개수이고, $\sigma(s, t/v)$ 는 s, t 가 아닌 일부 노드 v 를 통과하는 이들 경로들의 개수이고, 여기서, if $s == t$ 이면, $\sigma(s, t) = 1$ 이고, $v \in s$ 이면, $\sigma(s, t/v) = 0$ 이다.

[0076]

시맨틱 그래프가 트레이닝된 후, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 사용자 입력들을 분석하여 사용자 응답들을 식별하기 위해 시맨틱 그래프를 이용하기 시작할 수도 있다. 예를 들어, 이 프로세스는 도 14와 관련하여 아래에서 설명된다. 단계 610에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 사용자 입력을 수신한다. 사용자 입력은 사용자 입력 인터페이스(예컨대, 사용자 입력 인터페이스(310)를 통해서 수신된 사용자 발화 또는 텍스트 문자열일 수도 있다. 시스템은 사용자 발화들에 대해 음성-대-텍스트 프로세싱과 같은 동작들을 수행하여, 그 발화들에 대응하는 텍스트 문자열들을 획득할 수도 있다. 시스템은 추가적인 프로세싱을 위해 사용자 입력을 컴포넌트들로, (예컨대, 후보들로 그리고 8개의 품사들로) 추가로 세분화할 수도 있다.

[0077]

단계 612에서, 시스템은 단계 608에서 생성된 시맨틱 그래프를 이용하여 사용자 입력을 프로세싱한다. 시스템은 사용자 입력으로부터의 후보들을 시맨틱 그래프에서의 노드들에 매칭시킬 수도 있다. 예를 들어, 사용자 입력이 텍스트 문자열 "Jack is doctor. He has an office on First Street"이면, 시스템은 후보들 "Jack", "doctor", "office", 및 "First Street"를 시맨틱 그래프에서의 노드들에 매칭할 수도 있다. 더욱이, 후보들(예컨대, "is") 사이의 관계는 시맨틱 그래프의 에지들에 의해 표시된다. 이들 관계들은 "is", "has", 및 "on"과 같은 단어들로 표시될 수도 있다. 시맨틱 그래프는 추가적으로 용어 "he"와 "Jack" 사이의 관계들을 표시할 수도 있다. 이들 용어들 사이의 관계들은 종속성 트리를 횡단함으로써 발견될 수도 있다. 시맨틱 그래프(700)에서, "Jack" 및 "doctor"는 "is"에 의해 연결된다. 종속성 트리에서, "is"가 용어 "Jack"과 "doctor"를 연결한다.

[0078]

단계 614에서, 시스템은 프로세싱된 사용자 입력에 기초하여 출력을 생성한다. 시스템은 단계 612에서 완료된 프로세싱을 이용하여, 사용자 입력(예컨대, 단계 612에서 수신된 사용자 입력)의 컴포넌트들에 관련된 엔티티들을 결정할 수도 있다. 시스템은 사용자 입력을 나타내는 노드들에 밀접하게 연결된 노드들을 결정하기 위해 시맨틱 그래프를 횡단할 수도 있다. 예를 들어, 시스템은 사용자 입력 노드들 사이의 간극을 연결하는 노드들을 식별할 수도 있다. 시스템은 노드들을 연결하는 에지들뿐만 아니라 식별된 노드들을 포함하는 출력을 구성할 수도 있다. 출력은 사용자 입력에서 제기된 질문에 대한 답변을 포함할 수도 있거나 또는 사용자 입력에 따라 확장되는 추가 정보를 포함할 수도 있다. 출력은 스테이트먼트의 형태, 추가적인 리소스들로의 링크, 또는 임의의 다른 형태의 출력일 수도 있다.

[0079]

도 7은 예시적인 시맨틱 그래프(700)를 제공한다. 예를 들어, 텍스트 문자열("Jack wanted to learn more about Mary".)에 나타나는 후보들의 각각에 대해, 시스템은 텍스트 문자열에서의 단어들이 spaCy를 이용하여 생성되는 종속성 트리를 횡단함으로써 연결되어 있는지 여부를 체크한다. 일부 실시형태들에서, 시스템은 연결들이 동사들을 통해서 행해진다고 결정하고, 비방향성 그래프가 이들 에지들을 이용하여 생성된다. 시맨틱 그래프(700)에서, "Jack" 및 "Mary"는 동사들 "wanted" 및 "learn"에 의해 연결된다.

[0080]

- [0081] 일부 실시형태들에서, 종속성 트리는 일부 컨텍스트-프리 문법에 따라서 문자열의 구문 구조를 나타낼 수도 있다. 종속성 트리는 구성 문법들(어구 구조 문법들)의 구성 관계 또는 종속성 문법들의 종속성 관계에 기초하여 구성될 수도 있다. 종속성 트리는 프로그래밍 언어들과 같은 컴퓨터 언어들의 프로세싱 동안뿐만 아니라 자연어들에서의 문장들에 대해 생성될 수도 있다.
- [0082] 종속성 트리는 텍스트 문자열에서의 각각의 후보에 대한 품사 태그들을 포함한다. 예를 들어, "Jack"은 고유 명사임을 표시하는 "PROPN"으로서 라벨링된다. 종속성 트리는 아크들을 이용하여 종속성 트리에서의 단어들을 연결한다. 각각의 아크는 종속성을 표시하는 "머리(head)" 및 "자식(child)"을 갖는다, 즉, 자식은 머리에 종속된다. 도 7에서, 예를 들어, "wanted" 및 "learn"은 아크에 의해 연결되고; "wanted"는 머리이지만 "learn"은 자식이므로, "wanted"에 종속된다. 아크들은 추가적으로 수정을 표시한다, 즉, 자식은 머리를 수정한다. 예를 들어, "more"은 "learn"의 자식이고, 이는 "more"가 "learn"을 수정한다는 것을 표시한다. 종속성 트리에서의 각각의 단어는 정확히 하나의 머리를 갖는다. 각각의 단어는 자식들을 갖지 않는 것을 포함하여, 임의 개수의 자식들을 가질 수도 있다.
- [0083] 각각의 아크는 자식을 머리에 연결하는 구문 관계의 유형을 표시하는 라벨을 할당받을 수도 있다. 예를 들어, 도 7에서, "wanted"는 "xcomp"로 라벨링된 아크에 의해 "learn"에 연결되는데, 이는 "learn"이 "wanted"의 개방향 보완 요소라는 것을 표시한다.
- [0084] 문자열의 의미는 따라서 단어들 사이의 구문 관계들을 표시하는 아크들 및 품사 태그들로 구분된다. 도 7에서의 종속성 트리와 같은 종속성 트리를 횡단하면 문자열에서의 단어들이 연결되는 방법을 알 수 있다.
- [0085] 도 8 내지 도 10은 시스템에 의해 추출된 엔티티들 및 역할들의 예시적인 예들을 나타낸다. 낮은-스코어 노드들은 용이한 표현을 위해 제거되었다. 도 8은 영화 "Pulp Fiction"에 대응한다. 시스템은 엔티티 "Briefcase"가 (플롯을 주도하는 McGuffin이기 때문에) 높은 스코어를 갖는다고 결정하고, 이는 TF-IDF와 같은 통계적 모델들로 표현화하기 어려울 것이다. "Briefcase"와 같은 포괄적인 용어의 TF-IDF 스코어는 매우 낮을 것이며, 통계적 모델은 영화의 컨텍스트에서 어구의 시맨틱 관련성을 파악하지 못한다. 도 9는 영화 "Dr. Strangelove"에 대응한다. 시스템은 "Russia", "CRM-114", 및 "Water Fluoridation"과 같은 중요한 엔티티들을 식별하였으며, 이들 모두는 종래의 모델들에 의해서는 추출되지 않았을 것이다. 또한, 영화의 플롯에 필수적인 역할들이 더 높은 스코어들을 받는다는 것이 관찰된다.
- [0086] 도 10은 뉴스 기사 "Sending Tesla Roadster to Mars"에 적용되는 시스템들의 예시적인 예이다. 시스템은 "Tesla Roadster", "Elon Musk", "Mars" 및 "Starman"과 같은 엔티티들을 성공적으로 추출하면서 "잡음" - "Kevin Anderson", "bio threat", "Harry Potter" 및 "bacteria"와 같은 중요하지 않은 키워드들을 제거하였다.
- [0087] 도 11은 도 1에서 설명된 용도에 대응하는 방법들 및 시스템의 애플리케이션을 예시한다. 도 11에서, 사용자 인터페이스(1100)는 디스플레이 디바이스 상에 디스플레이된다. 사용자 인터페이스(1100)는 텍스트 문자열을 (예컨대, 사용자 입력 인터페이스로의 사용자 입력을 통해서) 수신한다. 이에 응답하여, 시스템은 프로그램 추천을 디스플레이를 위해 생성한다. 다음 예는 시맨틱 그래프들로부터의 키워드들이 콘텐츠의 더 깊은 이해를 보여주고 더 풍부한 검색 경험을 제공하는 방법을 예시한다.
- [0088] 본원에서 설명되는 방법들 및 시스템들이 미디어 안내를 제공하기 위한 애플리케이션에서 구현될 수도 있다는 점에 유의해야 한다. 예를 들어, 임의의 주어진 콘텐츠 전달 시스템에서 사용자들에게 이용가능한 콘텐츠의 양이 상당할 수 있다. 그 결과, 다수의 사용자들은 사용자들로 하여금 콘텐츠 선택들을 효율적으로 탐색하게 하고 이들이 원할 수도 있는 콘텐츠를 용이하게 식별하게 하는 인터페이스를 통한 미디어 안내의 형태를 원한다. 이러한 안내를 제공하는 애플리케이션은 대화형 미디어 안내 애플리케이션 또는, 종종, 미디어 안내 애플리케이션 또는 안내 애플리케이션으로서 본원에서 지칭된다.
- [0089] 대화형 미디어 안내 애플리케이션들은 이들이 안내를 제공하는 콘텐츠에 따라서 다양한 형태들을 취할 수도 있다. 미디어 안내 애플리케이션의 하나의 전형적인 유형은 대화형 텔레비전 프로그램 가이드이다. (전자 프로그램 가이드들로서 종종 지칭되는)대화형 텔레비전 프로그램 가이드들이 특히, 사용자들로 하여금 다수의 유형들의 콘텐츠 또는 미디어 자산들을 탐색하게 하여 위치추적하게 하는 널리 공지된 안내 애플리케이션들이다. 대화형 미디어 안내 애플리케이션들은 사용자로 하여금 콘텐츠를 탐색하고 위치추적하여 선택가능하게 하는 그래픽 사용자 인터페이스 스크린들을 생성할 수도 있다.
- [0090] 본원에서 설명되는 실시형태들 중 임의의 실시형태를 수행하기 위한 미디어 안내 애플리케이션 및/또는 임의의

명령어들은 컴퓨터-판독가능 미디어 상에 인코딩될 수도 있다. 컴퓨터-판독가능 미디어는 데이터를 저장하는 것이 가능한 임의의 미디어를 포함한다. 컴퓨터 판독가능 미디어는 전기 또는 전자기 신호들을 전파하는 것을 포함하지만 이에 한정되지 않는, 일시성일 수도 있거나, 또는 휘발성 및 비-휘발성 컴퓨터 메모리 또는 저장 디바이스들, 예컨대 하드 디스크, 플로피 디스크, USB 드라이브, DVD, CD, 미디어 카드들, 레지스터 메모리, 프로세서 캐시들, 랜덤 액세스 메모리("RAM"), 등을 포함하지만 이에 한정되지 않는, 비일시적일 수도 있다.

[0091] 인터넷, 모바일 컴퓨팅, 및 고속 무선 네트워크들의 출현으로, 사용자들은 이들이 종래 사용하지 않았던 사용자 장비 디바이스들 상의 미디어에 액세스하고 있다. 본원에서 지칭될 때, 어구 "사용자 장비 디바이스", "사용자 장비", "사용자 디바이스", "전자 디바이스", "전자 장비", "미디어 장비 디바이스", 또는 "미디어 디바이스"는 텔레비전, 스마트 TV, 셋-탑 박스, 위성 텔레비전을 처리하기 위한 통합된 수신기 디코더(IRD), 디지털 저장 디바이스, 디지털 미디어 수신기(DMR), 디지털 미디어 어댑터(DMA), 스트리밍 미디어 디바이스, DVD 플레이어, DVD 리코더, 접속된 DVD, 로컬 미디어 서버, BLU-RAY 플레이어, BLU-RAY 리코더, 개인용 컴퓨터(PC), 랩탑 컴퓨터, 태블릿 컴퓨터, WebTV 박스, 개인용 컴퓨터 텔레비전(PC/TV), PC 미디어 서버, PC 미디어 센터, 핸드-헬드 컴퓨터, 고정 전화기, 개인 휴대정보 단말기(PDA), 모바일 전화기, 포터블 비디오 플레이어, 포터블 뮤직 플레이어, 포터블 게이밍 머신, 스마트폰, 또는 임의의 다른 텔레비전 장비, 컴퓨팅 장비, 또는 무선 디바이스, 및/또는 이들의 조합과 같은, 위에서 설명된 콘텐츠에 액세스하는 임의의 디바이스를 의미하는 것으로 이해되어야 한다. 일부 실시형태들에서, 사용자 장비 디바이스는 전면 스크린 및 후면 스크린, 다수의 전면 스크린들, 또는 다수의 각진 스크린들을 가질 수도 있다. 일부 실시형태들에서, 사용자 장비 디바이스는 전면 카메라 및/또는 후면 카메라를 가질 수도 있다. 이들 사용자 장비 디바이스들 상에서, 사용자들은 텔레비전을 통해서 이용가능한 동일한 콘텐츠를 탐색하여 위치추적할 수도 있다. 그 결과, 미디어 안내는 이들 디바이스들 상에서 역시 이용가능할 수도 있다. 제공된 안내는 단지 텔레비전을 통해서만 이용가능한 콘텐츠, 하나 이상의 다른 유형들의 사용자 장비 디바이스들을 통해서만 이용가능한 콘텐츠, 또는 텔레비전 및 하나 이상의 다른 유형들의 사용자 장비 디바이스들 양자를 통해서 이용가능한 콘텐츠에 대한 것일 수도 있다. 미디어 안내 애플리케이션들은 온라인 애플리케이션들로서 제공되거나(즉, 웹사이트 상에 제공되거나), 또는 사용자 장비 디바이스들 상의 독립형 애플리케이션들 또는 클라이언트들로서 제공될 수도 있다. 미디어 안내 애플리케이션들을 구현할 수도 있는 다양한 디바이스들 및 플랫폼들이 아래에서 좀더 자세하게 설명된다.

[0092] 미디어 안내 애플리케이션의 기능들 중 하나는 미디어 안내 데이터를 사용자에게 제공하는 것이다. 본원에서 지칭될 때, 어구 "미디어 안내 데이터" 또는 "안내 데이터"는 안내 애플리케이션을 동작시킬 때에 사용되는 콘텐츠 또는 데이터에 관련된 임의의 데이터를 의미하는 것으로 이해되어야 한다. 예를 들어, 안내 데이터는 프로그램 정보, 안내 애플리케이션 세팅들, 사용자 선호사항들, 사용자 프로필 정보, 미디어 리스트들, 미디어-관련 정보(예컨대, 브로드캐스트 시간들, 브로드캐스트 채널들, 제목들, 설명들, 등급들 정보(예컨대, 부모 통제 등급들, 비평가의 등급들, 등), 장르 또는 카테고리 정보, 행위자 정보, 방송사들 또는 제공자들의 로고들에 대한 로고 데이터, 등), 미디어 포맷(예컨대, 표준 해상도, 고선명, 3D, 등), 광고 정보(예컨대, 텍스트, 이미지들, 미디어 클립들, 등), 온-디맨드 정보, 블로그들, 웹사이트들, 및 사용자가 원하는 콘텐츠 선택들을 탐색하고 찾는 데 도움이 되는 임의의 다른 유형의 안내 데이터를 포함할 수도 있다.

[0093] 본원에서 설명되는 기법들 방법들 및 시스템은 다수의 유형들의 사용자 인터페이스들 및 애플리케이션들에 적용될 수도 있다는 점에 유의해야 한다. 이들 기법들을 구현하기 위한 2개의 예시적인 미디어 안내 애플리케이션들이 도 12 - 도 13에 도시된다. 도 12 내지 도 13은 미디어 안내 데이터를 제공하는 데 사용될 수도 있는 예시적인 디스플레이 스크린들을 나타낸다. 도 12 내지 도 13에 나타난 디스플레이 스크린들은 임의의 적합한 사용자 장비 디바이스 또는 플랫폼 상에 구현될 수도 있다. 도 12 내지 도 13의 디스플레이들이 풀 스크린 디스플레이들로서 예시되지만, 이들은 또한 디스플레이중인 콘텐츠 상에 전체적으로 또는 부분적으로 오버레이될 수도 있다. 사용자는 디스플레이 스크린에 제공되는 선택가능한 옵션(예컨대, 메뉴 옵션, 리스트들 옵션, 아이콘, 하이퍼링크, 등)을 선택함으로써 또는 원격 제어 또는 다른 사용자 입력 인터페이스 또는 디바이스 상에서 전용 버튼(예컨대, 가이드 버튼)을 누름으로써, 콘텐츠 정보에 액세스하려는 의사를 표시할 수도 있다. 사용자의 표시에 응답하여, 미디어 안내 애플리케이션은 디스플레이 스크린에 여러 방법들 중 하나로, 예컨대 그리드에서의 시간 및 채널에 의해, 시간에 의해, 채널에 의해, 소스에 의해, 콘텐츠 유형에 의해, 카테고리(예컨대, 영화들, 스포츠, 뉴스, 칠드런, 또는 프로그래밍의 다른 카테고리들), 또는 다른 미리 정의된, 사용자-정의된, 또는 다른 조직 기준들에 의해 조직된 미디어 안내 데이터를 제공할 수도 있다.

[0094] 도 12는 단일 디스플레이에서 상이한 유형들의 콘텐츠에의 액세스를 또한 가능하게 하는 시간 및 채널별로 배열된 프로그램 리스트들 디스플레이(1200)의 예시적인 그리드를 나타낸다. 디스플레이(1200)는 (1) 채널/콘텐츠

유형 식별자들(1204)의 칼럼(여기서, (칼럼 내 셀인) 각각의 채널/콘텐츠 유형 식별자가 이용가능한 상이한 채널 또는 콘텐츠 유형을 식별함); 및 (2) 시간 식별자들(1206)의 로우(여기서, (로우 내 셀인) 각각의 시간 식별자가 프로그래밍의 시간 블록을 식별함)을 갖는, 그리드(1202)를 포함할 수도 있다. 그리드(1202)는 또한 프로그램 리스트(1208)와 같은 프로그램 리스트들의 셀들을 포함하며, 여기서 각각의 리스트는 리스트의 연관된 채널 및 시간에 제공된 프로그램의 제목을 제공한다. 사용자 입력 디바이스로, 사용자는 하이라이트 영역(1210)을 이동시킴으로써 프로그램 리스트들을 선택할 수 있다. 하이라이트 영역(1210)에 의해 선택된 프로그램 리스트에 관련된 정보는 프로그램 정보 영역(1212)에 제공될 수도 있다. 영역(1212)은 예를 들어, 프로그램 제목, 프로그램 설명, 프로그램이 제공되는 시간(해당되는 경우), 프로그램이 있는 채널(해당되는 경우), 프로그램의 등급, 및 다른 원하는 정보를 포함할 수도 있다.

[0095] 선형 프로그래밍(예컨대, 미리 결정된 시간에 복수의 사용자 장비 디바이스들로 송신되도록 스케줄링되고 스케줄에 따라서 제공되는 콘텐츠)에의 액세스에 추가하여, 미디어 안내 애플리케이션은 또한 비선형 프로그래밍(예컨대, 임의의 시간에 사용자 장비 디바이스에 액세스가능하고 스케줄에 따라서 제공되지 않는 콘텐츠)에의 액세스를 제공한다. 비선형 프로그래밍은 온-디맨드 콘텐츠(예컨대, VOD), 인터넷 콘텐츠(예컨대, 스트리밍 미디어, 다운로드가능한 미디어, 등), 로컬로 저장된 콘텐츠(예컨대, 위에서 설명된 임의의 사용자 장비 디바이스 또는 다른 저장 디바이스 상에 저장된 콘텐츠), 또는 다른 시간-독립적인 콘텐츠를 포함한, 상이한 콘텐츠 소스들로부터의 콘텐츠를 포함할 수도 있다. 온-디맨드 콘텐츠는 특정의 콘텐츠 제공자(예컨대, "Sopranos" 및 "Curb Your Enthusiasm"을 제공하는 HBO ON DEMAND)에 의해 제공되는 영화들 또는 임의의 다른 콘텐츠를 포함할 수도 있다. HBO ON DEMAND는 Time Warner Company L.P. 등이 소유한 서비스 마크이고, THE SOPRANOS 및 CURB YOUR ENTHUSIASM은 Home Box Office, Inc가 소유한 상표들이다. 인터넷 콘텐츠는 채팅 세션 또는 웹캐스트와 같은 웹 이벤트들, 또는 인터넷 웹사이트 또는 다른 인터넷 액세스(예컨대, FTP)를 통해 스트리밍 콘텐츠 또는 다운로드 가능한 콘텐츠로서 온디맨드로 이용가능한 콘텐츠를 포함할 수도 있다.

[0096] 그리드(1202)는 온-디맨드 리스트(1214), 리코딩된 콘텐츠 리스트(1216), 및 인터넷 콘텐츠 리스트(1218)를 포함한, 비선형 프로그래밍을 위한 미디어 안내 데이터를 제공할 수도 있다. 상이한 유형들의 콘텐츠 소스들로부터의 콘텐츠에 대한 미디어 안내 데이터를 결합하는 디스플레이는 "혼합된-미디어" 디스플레이로서 종종 지칭된다. 디스플레이(1200)와는 상이한, 디스플레이될 수도 있는 미디어 안내 데이터의 유형들의 다양한 치환들은 사용자 선택 또는 안내 애플리케이션 정의(예컨대, 리코딩 및 브로드캐스트 리스트들만 표시, 온-디맨드 및 브로드캐스트 리스트들만 표시 등)에 기초할 수도 있다. 예시된 바와 같이, 리스트들(1214, 1216, 및 1218)은 이들 리스트들의 선택이 각각 온-디맨드 리스트들, 리코딩된 리스트들, 또는 인터넷 리스트들을 담당하는 디스플레이에의 액세스를 제공할 수도 있다는 것을 표시하기 위해, 그리드(1202)에 디스플레이된 전체 시간 블록을 포괄하는 것으로 도시된다. 일부 실시형태들에서, 이들 콘텐츠 유형들에 대한 리스트들은 그리드(1202)에 직접 포함될 수도 있다. 추가적인 미디어 안내 데이터는 사용자가 네비게이션 아이콘들(1220) 중 하나를 선택하는 것에 응답하여 디스플레이될 수도 있다. (사용자 입력 디바이스 상에서 화살표 키를 누르는 것은 네비게이션 아이콘들(1220)을 선택하는 것과 유사한 방법으로 디스플레이에 영향을 미칠 수도 있다.)

[0097] 디스플레이(1200)는 또한 비디오 영역(1222), 광고(1224), 및 옵션들 영역(1226)을 포함할 수도 있다. 비디오 영역(1222)은 사용자로 하여금 현재 이용가능하거나, 이용가능하거나, 또는 사용자에게 이용가능하였던 프로그램들을 보게 하고 및/또는 미리 볼 수 있게 할 수도 있다. 비디오 영역(1222)의 콘텐츠는 그리드(1202)에 디스플레이된 리스트들 중 하나에 대응하거나 또는 이와 독립적일 수도 있다. 비디오 영역을 포함하는 그리드 디스플레이들은 종종 PIG(picture-in-guide) 디스플레이들로서 지칭된다. PIG 디스플레이들 및 이들의 기능들은 2003년 5월 13일에 허여된, Satterfield 등의 미국 특허 제 6,564,378호, 및 2001년 5월 29일에 허여된, Yuen 등의 미국 특허 제 6,239,794호에 더욱더 자세히 설명되어 있으며, 이들은 본원에서 그 전체가 참고로 포함된다. PIG 디스플레이들은 본원에서 설명하는 실시형태들의 다른 미디어 안내 애플리케이션 디스플레이 스크린들에 포함될 수도 있다.

[0098] 광고(1224)는 (예컨대, 가입 프로그래밍을 위한) 뷰어의 액세스 권한들에 따라서, 현재 시청가능하거나, 미래에 시청가능하거나, 또는 시청가능하지 않을 수도 있는 콘텐츠에 대한 광고를 제공할 수도 있으며, 그리드(1202)에서의 콘텐츠 리스트들 중 하나 이상에 대응하거나 또는 관련되지 않을 수도 있다. 광고(1224)는 또한 그리드(1202)에 디스플레이된 콘텐츠에 관련되거나 또는 관련되지 않은 제품들 또는 서비스들에 대한 것일 수도 있다. 광고(1224)는 선택가능하며, 콘텐츠에 관한 추가적인 정보를 제공하고; 제품 또는 서비스에 관한 정보를 제공하고; 콘텐츠, 제품, 또는 서비스의 구매를 가능하게 하고; 광고에 관련된 콘텐츠를 제공하는; 등등을 행할 수도 있다. 광고(1224)는 사용자의 프로파일/선호사항들, 모니터링된 사용자 활동, 제공된 디스플레이의 유형에 기초

하여, 또는 다른 적합한 타겟팅된 광고에 기초하여 타겟팅될 수도 있다.

[0099] 광고(1224)는 직사각형 또는 배너-형상으로 도시되지만, 광고들은 안내 애플리케이션 디스플레이에서 임의의 적합한 사이즈, 형상, 및 위치로 제공될 수도 있다. 예를 들어, 광고(1224)는 그리드(1202)에 수평으로 근접한 직사각형 형상으로서 제공될 수도 있다. 이는 패널 광고로서 종종 지칭된다. 게다가, 광고들은 콘텐츠 또는 안내 애플리케이션 디스플레이 상에 오버레이되거나 또는 디스플레이 내에 삽입될 수도 있다. 광고들은 또한 텍스트, 이미지들, 회전 이미지들, 비디오 클립들, 또는 위에서 설명된 다른 유형들의 콘텐츠를 포함할 수도 있다. 광고들은 안내 애플리케이션을 가지는 사용자 장비 디바이스에, 사용자 장비에 접속된 데이터베이스에, (스트리밍 미디어 서버들을 포함하는)원격 위치에, 또는 다른 저장 수단 상에, 또는 이들 위치들의 조합에 저장될 수도 있다. 미디어 안내 애플리케이션에서 광고들을 제공하는 것은 예를 들어, 2003년 1월 17일에 출원된, Knudson 등의 미국 특허 출원 공개 번호 제 2003/0110499호; 2004년 6월 29일에 허여된, Ward, III 등의 미국 특허 제 6,756,997호; 및 2002년 5월 14일에 허여된, Schein 등의 미국 특허 제 6,388,714호에 더욱더 자세히 설명되어 있으며, 이들은 본원에서 그 전체가 참고로 포함된다. 광고들이 본원에서 설명하는 실시형태들의 다른 미디어 안내 애플리케이션 디스플레이 스크린들에 포함될 수도 있음을 알 수 있을 것이다.

[0100] 옵션들 영역(1226)은 사용자로 하여금 상이한 유형들의 콘텐츠, 미디어 안내 애플리케이션 디스플레이들, 및/또는 미디어 안내 애플리케이션 특징들에 액세스가능하게 할 수도 있다. 옵션들 영역(1226)은 디스플레이(1200) (및 본원에서 설명되는 다른 디스플레이 스크린들)의 부분일 수도 있거나, 또는 온스크린 옵션을 선택하거나 또는 사용자 입력 디바이스 상에서 전용 또는 할당가능한 버튼을 누름으로써, 사용자에게 의해 호출될 수도 있다. 옵션들 영역(1226) 내 선택가능한 옵션들은 그리드(1202)에서의 프로그램 리스트들에 관련된 특징들과 관련된 수도 있거나, 또는 메인 메뉴 디스플레이로부터 이용가능한 옵션들을 포함할 수도 있다. 프로그램 리스트들에 관련된 특징들은 다른 방송 시간들 또는 프로그램을 수신하는 방법들을 검색하는 것, 프로그램을 리코딩하는 것, 프로그램의 시리즈 리코딩을 가능하게 하는 것, 프로그램 및/또는 채널을 즐겨찾기로서 설정하는 것, 프로그램을 구매하는 것, 또는 다른 특징들을 포함할 수도 있다. 메인 메뉴 디스플레이로부터 이용가능한 옵션들은 검색 옵션들, VOD 옵션들, 부모 통제 옵션들, 인터넷 옵션들, 클라우드-기반 옵션들, 디바이스 동기화 옵션들, 제 2 스크린 디바이스 옵션들, 다양한 유형들의 미디어 안내 데이터 디스플레이들에 액세스하는 옵션들, 프리미엄 서비스에 가입하는 옵션들, 사용자의 프로파일을 편집하는 옵션들, 브라우저 오버레이에 액세스하는 옵션들, 또는 다른 옵션들을 포함할 수도 있다.

[0101] 미디어 안내 애플리케이션은 사용자의 선호사항들에 기초하여 개인화될 수도 있다. 개인화된 미디어 안내 애플리케이션은 사용자로 하여금 디스플레이들 및 특징들을 커스터마이징하여 미디어 안내 애플리케이션으로 개인화된 "경험"을 생성가능하게 한다. 이 개인화된 경험은 사용자로 하여금 이들 커스터마이제이션들을 입력하게 함으로써, 또/또는 다양한 사용자 선호사항들을 결정하기 위해 사용자 활동을 모니터링하는 미디어 안내 애플리케이션에 의해, 생성될 수도 있다. 사용자들은 로그인하거나 또는 아니면 안내 애플리케이션에 자신들을 식별시킴으로써 이들의 개인화된 안내 애플리케이션에 액세스할 수도 있다. 미디어 안내 애플리케이션의 커스터마이제이션은 사용자 프로파일에 따라서 이루어질 수도 있다. 커스터마이제이션들은 다양한 프리젠테이션 방식들(예컨대, 디스플레이들의 칼라 방식, 텍스트의 폰트 사이즈, 등), 디스플레이된 콘텐츠 리스트들의 양태들(예컨대, HDTV 단독 또는 3D 프로그래밍 단독, 즐겨찾기 채널 선택들에 기초한 사용자-규정된 브로드캐스트 채널들, 채널들의 디스플레이를 재순서정렬하는 것, 추천된 콘텐츠, 등), 원하는 리코딩 특징들(예컨대, 특징의 사용자들에 대한 리코딩 또는 시리즈 리코딩들, 리코딩 품질, 등), 부모 통제 세팅들, 인터넷 콘텐츠의 커스터마이징된 프리젠테이션(예컨대, 소셜 미디어 콘텐츠의 프리젠테이션, email, 전자적으로 전달된 물품들, 등) 및 다른 원하는 커스터마이제이션들을 포함할 수도 있다.

[0102] 미디어 안내 애플리케이션은 사용자로 하여금, 사용자 프로파일 정보를 제공하게 할 수도 있거나 또는 사용자 프로파일 정보를 자동으로 컴파일할 수도 있다. 미디어 안내 애플리케이션은 예를 들어, 콘텐츠 사용자 액세스들 및/또는 사용자가 안내 애플리케이션과 가질 수도 있는 다른 상호작용들을 모니터링할 수도 있다. 추가적으로, 미디어 안내 애플리케이션은 (예컨대, www.Tivo.com과 같은, 사용자가 액세스하는 인터넷 상의 다른 웹사이트들로부터, 사용자가 액세스하는 다른 미디어 안내 애플리케이션들로부터, 사용자가 액세스하는 다른 대화형 애플리케이션들로부터, 사용자의 다른 사용자 장비 디바이스로부터, 등등으로부터) 특징의 사용자에게 관련된 다른 사용자 프로파일들의 모두 또는 부분을 획득하거나, 또/또는 미디어 안내 애플리케이션이 액세스할 수도 있는 다른 소스들로부터 사용자에게 관한 정보를 획득할 수도 있다. 그 결과, 사용자에게 사용자의 상이한 사용자 장비 디바이스들에 걸쳐서 통합된 안내 애플리케이션 경험이 제공될 수 있다. 추가적인 개인화된 미디어 안내 애플리케이션 특징들은 2005년 7월 11일에 출원된, Ellis 등의 미국 특허 출원 공개 번호 제 2005/0251827호;

2007년 1월 16일에 허여된 Boyer 등의 미국 특허 제 7,165,098호; 및 2002년 2월 21일에 출원된, Ellis 등의 미국 특허 출원 공개 번호 제 2002/0174430호에 더욱더 자세히 설명되어 있으며, 이들은 본원에 그 전체가 참고로 포함된다.

- [0103] 미디어 안내를 제공하는 다른 디스플레이 배열이 도 13에 도시된다 비디오 모자이크 디스플레이(1300)는 콘텐츠 유형, 장르, 및/또는 다른 조직 기준들에 기초하여 조직화된 콘텐츠 정보에 대한 선택가능한 옵션들(1302)을 포함한다. 디스플레이(1300)에서, 텔레비전 리스트들 옵션(1304)이 선택되어, 리스트들(1306, 1308, 1310, 및 1312)을 브로드캐스트 프로그램 리스트들로서 제공한다. 디스플레이(1300)에서, 리스트들은 표지 아트, 콘텐츠로부터의 정지 이미지들, 비디오 클립 미리보기들, 콘텐츠로부터의 라이브 비디오, 또는 리스트에서의 미디어 안내 데이터에 의해 설명되는 콘텐츠를 사용자에게 표시하는 다른 유형들의 콘텐츠를 포함하는 그래픽 이미지들을 제공할 수도 있다. 그래픽 리스트들의 각각에는 또한 리스트와 연관된 콘텐츠에 관한 추가적인 정보를 제공하는 텍스트가 함께 제공될 수도 있다. 예를 들어, 리스트(1308)는 미디어 부분(1314) 및 텍스트 부분(1316)을 포함하는 복수의 부분을 포함할 수도 있다. 미디어 부분(1314) 및/또는 텍스트 부분(1316)은 콘텐츠를 풀-스크린으로 시청하거나 또는 미디어 부분(1314)에 디스플레이된 콘텐츠에 관련된 정보를 시청하도록(예컨대, 비디오가 디스플레이되는 채널에 대한 리스트들을 시청하도록) 선택가능할 수도 있다.
- [0104] 디스플레이(1300)에서의 리스트들은 상이한 사이즈들이며 (즉, 리스트(1306)는 리스트들(1308, 1310, 및 1312)보다 크고), 그러나 원할 경우, 모든 리스트들은 동일한 사이즈일 수도 있다. 리스트들은 상이한 사이즈들이일 수도 있거나, 또는 사용자에게 관심도들을 나타내거나 또는 콘텐츠 제공자가 원하는 대로 또는 사용자 선호사항들에 기초하여 특정의 콘텐츠를 강조하기 위해 그래픽으로 강조될 수도 있다. 콘텐츠 리스트들을 그래픽으로 강조하기 위한 다양한 시스템들 및 방법들은 예를 들어, 2009년 11월 12일에 출원된, Yates의 미국 특허 출원 공개 번호 제 2010/0153885호에 설명되어 있으며, 이는 본원에 전체적으로 참고로 포함된다.
- [0105] 도 14는 본원에서 설명되는 검색, 추천, 및 발견 특징들에 기초하여 엔티티를 생성하는 프로세스의 일 실시형태를 도시한다. 프로세스(1400)의 각각의 단계가 (예컨대, 애플리케이션에 의해 제어 회로부(304)에 명령하는 방법으로) 제어 회로부(304) 또는 도 3 내지 도 4에 나타난 임의의 다른 시스템 컴포넌트들에 의해 수행될 수 있다는 점에 유의해야 한다. 제어 회로부(304)는 사용자 장비(예컨대, 콘텐츠(402), 시스템 제어기(404), 및/또는 무선 통신 디바이스(406)를 소비하는 수단의 기능의 일부 또는 모두를 가질 수도 있는 디바이스)일 수도 있거나, 또는 통신 네트워크(414)에 의해 사용자 장비로부터 분리된 또는 양자의 조합을 통해서 분산된 원격 서버의 부분일 수도 있다.
- [0106] 단계 1402에서, 시스템은 텍스트 문자열을 수신한다. 텍스트 문자열은 사용자 입력 인터페이스(310)를 통해서 수신될 수도 있다. 텍스트 문자열은 사용자로부터 또는 다른 전자 디바이스로부터 수신될 수도 있다.
- [0107] 단계 1404에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 텍스트 문자열에서 대명사들을 식별한다. 일부 실시형태들에서, POS 태깅은 설명 태그들의 세트에 따라서, 별개의 용어들뿐만 아니라 숨겨진 품사들을 연결하는 알고리즘들을 이용하여, 컴퓨터 언어학의 컨텍스트에서 이루어진다. POS-태깅 알고리즘들은 2개의 독특한 그룹들: 규칙-기반 및 확률론으로 나누어진다. 규칙-기반 POS 태깅에 대해, 시스템은 일련의 수동 규칙들을 통해서 수동으로 구축된다. 예를 들어, 시스템은 태깅된 단어에 선행하는 단어가 if-then 문들을 통해서 특정의 방법으로 태깅된다는 것을 표시하는 규칙을 포함할 수도 있다. 통계적(또는, 확률론적) 품사 태깅은 각각의 단어가 알려져 있으며 가능한 태그들의 유한한 세트를 갖고 있다고 가정한다. 이들 태그들은 사전 또는 형태학적 분석으로부터 유도될 수 있다. 예를 들어, 단어가 복수의 가능한 태그를 가질 때, 시스템은 통계적 방법들을 이용하여 품사 태그들의 시퀀스를 결정할 수도 있다. 시스템은 또한 규칙-기반과 확률론을 결합하는 하이브리드 접근법을 이용할 수도 있다. 마지막으로, 일부 실시형태들에서, POS 태깅이 수동으로 수행될 수도 있다는 점에 유의해야 한다.
- [0108] POS 태깅을 수행하기 위해, 시스템은 고급 자연어 프로세싱을 위해 소프트웨어 라이브러리를 이용할 수도 있다. 일부 실시형태들에서, 시스템은 고급 자연어 프로세싱을 위한 SpaCy, Python 라이브러리를 이용하여, 해당 POS 태깅 능력을 통해 식별을 강화할 수도 있다. POS 태깅에 추가하여, 시스템은 비-과외적 토큰화, 명명된 엔티티 인식, 다수의 언어들에 대한 통계적 모델들, 사전-트레이닝된 단어 벡터들, 라벨링된 종속성 파싱, 선택스-기반 문장 세그멘테이션, 텍스트 분류, 선택스 및 명명된 엔티티들에 대한 내장 시각화 장치들, 및/또는 심층 학습 통합과 같은, 추가적인 특징들을 이용할 수도 있다.
- [0109] 단계 1406에서, 시스템은 대명사 분석을 수행한다. 구체적으로 설명하면, 시스템은 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성한다. 대명사 분석은 풍부하고 정확한 시맨틱 그래프들에 필요한 엔티티 관계들을

식별하는데 중요하다. 프로세스의 이 단계에서, 시스템은 텍스트 문자열의 문장들에 걸쳐서 모든 대명사들을 분석한다. 예를 들어, 시스템은 대명사가 지칭하는 명사 또는 고유 명사(예컨대, "명사 체크")를 결정하는데 이용 가능한 단-대-단 신경 코퍼레이션 분석의 Python 구현을 이용할 수도 있다. 단-대-단 신경 코퍼레이션에서, 시스템은 엔티티들 사이의 모든 스캔들을 고려하고, 엔티티들 사이의 스캔들을 순위지정하고, 검색 공간들을 정리하기 위한 인수분해된(factored) 모델을 생성할 것이다. 시스템은 그후 주어진 대명사가 지칭하는 명사 체크를 높은 확률로 검출할 수도 있다.

[0110] 스캔 순위에 대해, 시스템은 입력 문서에서의 각각의 스캔을 프로세싱하고 모든 스캔에 선행항목(antecedent)을 할당할 것이다. 일부의 경우, 시스템은 암시적인 스캔들을 생성한다. 결과적인 클러스터는 시스템으로 하여금 3개의 유형들의 스캔들: i) 이전 멘션을 갖지 않는 스캔들; ii) 이전 링크들을 갖지 않는 멘션들; 및 iii) 예측된 코퍼레이션 링크를 갖는 스캔들을 식별하게 할 것이다. 각각의 스캔에 대해, 시스템은 독립적인 결정을 행하여, 2개의 스캔들 사이의 코퍼레이션의 가능성을 결정하는 쌍방식(pairwise) 코퍼레이션 스코어를 적용할 것이다. 시스템은 그후 최고 스코어를 가진 쌍에 기초하여 선행항목을 결정할 것이다.

[0111] 단계 1408에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 분석된 텍스트 문자열에서 명사 체크를 식별한다. 예를 들어, 시스템은 모든 명사 체크들을 도 6과 관련하여 위에서 설명된 바와 같은 시맨틱 그래프에서의 노드들로서 식별하기 위해 프로세싱된 텍스트 상에 POS(Part-Of-Speech) 태깅을 적용할 수도 있다. 일부 실시형태들에서, POS 태깅이 수동으로 수행될 수도 있다는 점에 유의해야 한다.

[0112] 단계 1410에서, 시스템은 복수의 노드들을 특징으로 하는 시맨틱 그래프에 기초하여 분류기를 이용해, 식별된 명사 체크를 프로세싱한다. 위에서 도 6과 관련하여 더욱더 자세하게 설명한 바와 같이, 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스이다. 시스템은 시맨틱 그래프를 지식 표현의 형태로 이용한다. 이는 개념들 및/또는 엔티티들을 나타낼 수도 있는 노드들, 및 개념들 및/또는 엔티티들 사이의 시맨틱 관계들을 나타내는 에지들로 구성되는 방향성 및/또는 비방향성 그래프이다. 예시적인 시맨틱 그래프가 위에서 도 7과 관련하여 설명되었다.

[0113] 예를 들어, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 텍스트 특징들을 결정할 수도 있다. 텍스트 특징들은 spaCy를 이용하여 추출된 후보 시스템의 POS 태그; 데이터 세트의 플롯에 걸쳐서 계산된 후보의 TF-IDF(용어 빈도-역 문서 빈도) 값; 텍스트 블러브(blurb)에서의 후보의 대문자; 후보가 메타데이터에서 다른 데이터 소스(예컨대, 웹사이트)에 대한 링크를 갖고 있는지 여부(그렇지 않으면 거짓으로 설정됨); 그 후보가 관련된 주제에 대한 카테고리로서 언급되는지 여부(그렇지 않으면 거짓으로 설정됨); 그 후보가 데이터 소스에서 제 1 단락 및/또는 두드러진 위치에서 언급되는지 여부(그렇지 않으면 거짓으로 설정됨); 제 1 라인을 이용하여 태깅된 후보의 유형 및/또는 페이지의 유형 및 7개의 유형들 - (임의의 후보에 대한 디폴트 유형과 함께) 프로그램들, 사람, 허구, 장소, 조직, 스포츠, 및 어구로 태깅된 카테고리들을 포함할 수도 있다.

[0114] 일부 실시형태들에서, 고 품질 정보가 통계적 패턴 학습과 같은 수단을 통해 패턴들 및 경향들의 고안을 통해서 유도된다. 텍스트 특징들을 결정하는 단계는 입력 텍스트를 구조화하고 일반적으로 일부 파생된 언어 특징들의 추가 및 나머지의 제거와 함께, 과잉 및 데이터베이스에의 후속 삽입), 구조화된 데이터 내 패턴들을 유도하고, 최종적으로, 출력을 평가 및 해석하는 프로세스를 포함할 수도 있다. 텍스트 특징들에서의 "고 품질"은 관련성, 신규성, 및 흥미도의 일부 조합을 지칭할 수도 있다는 점에 유의해야 한다. 전형적인 텍스트 특징들은 텍스트 분류, 텍스트 클러스터링, 개념/엔티티 추출, 세분화된 분류법들의 생성, 정서 분석, 문서 요약, 및 엔티티-관계 모델링(즉, 명명된 엔티티들 사이의 학습 관계들)을 포함할 수도 있다. 일부 실시형태들에서, 텍스트 분석은 정보 리트리빙, 단어 빈도 분포들을 연구하기 위한 어휘 분석, 패턴 인식, 태깅/주석, 정보 추출, 링크 및 연관 분석을 포함한 데이터-마이닝 기법들, 시각화, 및 예측 해석학을 수반한다.

[0115] 시스템은 그후 (예컨대, 제어 회로부(304)를 통해서) 노드들을 스코어링할 수도 있다. 다수의 접속된 컴포넌트들을 가진 일부 실시형태들에서, 시스템은 이들 특징들을 각각의 접속된 컴포넌트 상에서 별개로 계산한다. 시스템은 상기 도 6의 프로세스로부터 나온 모델을 이용할 수도 있다.

[0116] 단계 1412에서, 시스템은 단계 1410에서 분류기를 이용해 명사 체크를 프로세싱하는 것에 기초하여, (예컨대, 제어 회로부(304)를 통해서) 엔티티들을 결정한다. 분류기를 이용해 명사 체크를 프로세싱하는 것에 기초하여 엔티티들을 결정하는 예시적인 프로세스가 도 6과 관련하여 위에서 설명되었다. 단계 1414에서, 시스템은 수신된 텍스트 문자열에 응답하여 엔티티를 (예컨대, 디스플레이 디바이스(312) 상에) 디스플레이를 위해 생성한다.

[0117] 이 실시형태가 이 설명에서의 임의의 다른 실시형태와 결합될 수 있고 프로세스(1400)가 이 실시형태에서 프로

세스(1400)를 예시하는데 사용되는 디바이스들 또는 제어 컴포넌트들에 한정되지 않는다는 점에 유의해야 한다.

- [0118] 도 15는 본원에서 설명하는 바와 같이, 분류기를 이용해 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 프로세스의 일 실시형태를 도시한다. 프로세스(1500)의 각각의 단계가 (예컨대, 애플리케이션에 의해 제어 회로부(304)에 명령하는 방법으로) 제어 회로부(304) 또는 도 3 내지 도 4에 나타난 임의의 다른 시스템 컴포넌트들에 의해 수행될 수도 있다는 점에 유의해야 한다. 제어 회로부(304)는 사용자 장비(예컨대, 콘텐츠(402), 시스템 제어기(404), 및/또는 무선 통신 디바이스(406)를 소비하는 수단의 기능의 일부 또는 모두를 가질 수도 있는 디바이스)일 수도 있거나, 또는 통신 네트워크(414)에 의해 사용자 장비로부터 분리된, 또는 양자의 조합을 통해서 분산된 원격 서버의 일부일 수도 있다.
- [0119] 단계 1502에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 스코어를 각각의 엔티티에 할당한다. 예를 들어, 시맨틱 그래프들은 시맨틱 그래프로부터의 노드 스코어에 기초한 콘텐츠에서 중요하고 중요하지 않은 캐스트 멤버들 및 역할들의 분류인 역할 중요성을 위해 시스템에 의해 사용될 수도 있다. 예를 들어, 도 8 및 도 9에서, 높은 스코어를 얻도록 결정된 중요한 역할들이 도시된다.
- [0120] 단계 1504에서, 시스템은 해당 개개 스코어에 기초하여 엔티티들의 각각을 순위지정한다. 단계 1506에서, 최고 스코어를 갖는 엔티티가 수신된 텍스트 문자열에 대응하도록 결정된다. 스코어링 및 순위지정 메커니즘을 설명하는 구체적인 예들은 위에서 도 6과 관련하여 더욱더 자세하게 설명되었다.
- [0121] 이 실시형태가 이 설명에서의 임의의 다른 실시형태와 결합될 수 있고 프로세스(1500)가 이 실시형태에서 프로세스(1500)를 예시하는데 사용되는 디바이스들 또는 제어 컴포넌트들에 한정되지 않는다는 점에 유의해야 한다.
- [0122] 도 16은 본원에서 설명되는 검색, 추천, 및 발견 특징들을 제공하는데 사용되는 아키텍처의 예시적인 예이다. 도 16에 나타난 바와 같이, 시스템은 텍스트 문자열을 입력으로서 수신하여 이를 키 엔티티들 및 이들의 연관들을 식별하는 시맨틱 그래프로 변환한다. 시맨틱 그래프 및 텍스트 문자열로부터의 특징들은 머신-학습 모델을 통과하여 가장 문맥상 중요한 엔티티들을 추론한다. 프로세스는 4개의 스테이지들:
- [0123] 대명사 분석, 후보 식별, 시맨틱 그래프의 생성, 및 노드 스코어링을 수반한다.
- [0124] 단계 1602에서, 시스템은 텍스트 문자열을 수신한다. 텍스트 문자열은 사용자 입력 인터페이스(310)를 통해서 수신될 수도 있다. 텍스트 문자열은 사용자로부터 또는 다른 전자 디바이스로부터 수신될 수도 있다.
- [0125] 단계 1604에서, 시스템은 대명사 분석을 수행한다. 대명사 분석은 풍부하고 정확한 시맨틱 그래프들에 필요한 엔티티 관계들을 식별하는데 중요하다. 프로세스의 이 단계에서, 시스템은 텍스트 문자열의 문장들에 걸쳐서 모든 대명사들을 분석한다. 예를 들어, 시스템은 대명사가 지칭하는 명사 또는 고유 명사(예컨대, "명사 청크들")를 결정하는데 이용가능한 단-대-단 신경 코렐레이션 분석의 Python 구현을 이용할 수도 있다.
- [0126] 예를 들어, 코렐레이션은 텍스트에서의 2개 이상의 표현들이 동일한 사람 또는 사물을 지칭할 때; 이들이 동일한 지시대상물을 가질 때 발생한다. 예를 들어, 텍스트 문자열 "Bill said he would come"에서; 고유 명사 "Bill" 및 대명사 "he"는 동일한 사람, 즉 "Bill"을 지칭한다. 코렐레이션은 선택스의 분야에서 바인딩(binding) 현상들의 기초가 되는 메인 개념이다. 바인딩의 이론은 문장들 및 텍스트들에서의 코렐레이션 표현들 사이에 존재하는 구문 관계를 분석한다. 일부 실시형태들에서, 시스템은 대명사들을 분석하기 위한 신경망을 개발할 수도 있다. 예를 들어, 시스템은, 제어 회로부(304)를 통해서, 텍스트 문자열 "John helped Mary. He is a doctor"를 수신할 수도 있다. 시스템은 대명사들을 분석하여, 분석된 텍스트 문자열, "John helped Mary. John is doctor"를 생성할 수도 있다.
- [0127] 단계 1606에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 후보 식별을 수행한다. 예를 들어, 시스템은 프로세싱된 텍스트에 대해 POS(Part-Of-Speech) 태깅을 적용하여, 모든 명사 청크들을 시맨틱 그래프에서의 노드들로서 식별할 수도 있다. 문법적 태깅 또는 단어-카테고리 명확화라고도 불리는, 품사 태깅(POS 태깅 또는 PoS 태깅 또는 POST)은 그의 정의 및 그의 컨텍스트 양자-즉, 어구, 문장, 또는 단락에서 근접 및 관련된 단어들과의 관계에 기초하여, 텍스트(코퍼스) 내 단어를 특정의 품사에 대응하는 것으로 표시하는 프로세스이다. 예를 들어, 애플리케이션은 텍스트 문자열에서의 단어들을 명사들, 동사들, 형용사들, 부사들, 등으로 식별할 수도 있다. 일부 실시형태들에서, POS 태깅은 설명 태그들의 세트에 따라서, 별개의 용어들뿐만 아니라 숨겨진 품사들을 연결하는 알고리즘들을 이용하여, 컴퓨터 언어학의 컨텍스트에서 이루어진다. POS-태깅 알고리즘들은 2개의 독특한 그룹들: 규칙-기반 및 확률론으로 나누어진다. 최초이자 가장 널리 사용되는 영어 POS-태거들 중 하나인, E. Brill의 태거는 규칙-기반 알고리즘들을 채용한다. 일부 실시형태들에서, POS 태깅이 수동으로 수행될

수도 있다는 점에 유의해야 한다.

- [0128] 일부 실시형태들에서, 시스템은 고급 자연어 프로세싱을 위한 SpaCy, Python 라이브러리를 이용하여, 해당 POS 태깅 능력을 통해 식별을 강화할 수도 있다. 따라서, 시스템은 그것의 풍부한 구조를 레버리지하여, 플롯, 시놉시스 및 카테고리 멘션들로부터 링크들과 같은 더 많은 후보들을 식별한다.
- [0129] 단계 1608에서, 시스템은 시맨틱 그래프를 생성한다. 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스이다. 시스템은 시맨틱 그래프를 지식 표현의 형태로 이용한다. 이는 개념들 및/또는 엔티티들을 나타낼 수도 있는 노드들, 및 개념들 및/또는 엔티티들 사이의 시맨틱 관계들을 나타내는 에지들로 구성되는 방향성 및/또는 비방향성 그래프이다. 위에서 설명된 도 7은 예시적인 시맨틱 그래프를 제공한다.
- [0130] 단계 1610에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 근접 중심성 및 매개 중심성에 기초하여 그래프 특징들을 결정한다. 근접 중심성과 관련하여, 노드의 근접 중심성(또는, 근접성)은 (예컨대, 도 6에서 설명된 바와 같이) 그래프에서의 노드와 모든 다른 노드들 사이의 최단 경로들의 길이의 총합으로서 계산되는, 네트워크의 중심성을 측정한다.
- [0131] 데이터 세트는 트레이닝 및 테스트 세트들을 70:30의 비율로 분할될 수도 있다. 예를 들어, 시스템은 데이터 소스(예컨대, 웹사이트)로부터 (예컨대, 인기에 기초하여) 10,000개의 미디어 콘텐츠 리스트들을 취하고, 미디어 콘텐츠에 대한 메타데이터(예컨대, 플롯들 설명들)로부터 엔티티들/키워드들에 대한 후보들을 추출하고, 그리고 이들을 수동으로 대조하여 데이터 세트에서 긍정적인(모두 수락) 및 부정적인(모두 거부) 라벨들을 생성할 수도 있다. 트레이닝 세트는 모델을 구축하는데 사용되며, 테스트 세트는 벤치마킹을 위해 평가 및 사용된다. 시스템은 머신 학습을 이용하여, 예시적인 입력-출력 쌍들(예컨대, 트레이닝 데이터)에 기초하여 입력을 출력에 맵핑하는 함수를 생성한다. 이는 트레이닝 예들의 세트로 구성되는 라벨링된 트레이닝 데이터로부터 함수를 추론한다. 지도 학습에서, 각각의 예는 입력 오브젝트(전형적으로 벡터) 및 원하는 출력 값(또한, 감시(supervisory) 신호라고도 함)으로 구성된 쌍이다. 시스템들 학습 알고리즘은 트레이닝 데이터를 분석하여, 새로운 예들을 맵핑하는데 사용될 수 있는 추론된 함수를 생성한다. 학습된 알고리즘은 그후 보이지 않는 인스턴스들(예컨대, 텍스트 문자열 내 사용자 쿼리들)에 대한 클래스 라벨들을 정확하게 결정하는데 이용될 수 있다.
- [0132] 단계 1612에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 텍스트 특징들을 결정한다. 텍스트 특징들은 spaCy를 이용하여 추출된 후보 시스템의 POS 태그; 데이터 세트의 플롯에 걸쳐서 계산된 후보의 TF-IDF(용어 빈도-역 문서 빈도) 값; 텍스트 블러브(blurb)에서의 후보의 대문자; 그 후보가 메타데이터에서 다른 데이터 소스(예컨대, 웹사이트)에 대한 링크를 갖고 있는지 여부(그렇지 않으면 거짓으로 설정됨); 그 후보가 관련된 주제에 대한 카테고리로서 언급되는지 여부(그렇지 않으면 거짓으로 설정됨); 그 후보가 데이터 소스에서 제 1 단락 및/또는 두드러진 위치에서 언급되는지 여부(그렇지 않으면 거짓으로 설정됨); 제 1 라인을 이용하여 태깅된 후보의 유형 및/또는 페이지의 유형 및 7개의 유형들 - (임의의 후보에 대한 디폴트 유형과 함께) 프로그램들, 사람, 허구, 장소, 조직, 스포츠, 및 어구로 태깅된 카테고리들을 포함할 수도 있다.
- [0133] 일부 실시형태들에서, 고 품질 정보가 통계적 패턴 학습과 같은 수단을 통해 패턴들 및 경향들의 고안을 통해서 유도된다. 텍스트 특징들을 결정하는 단계는 입력 텍스트를 구조화하고 (일반적으로 일부 파생된 언어 특징들의 추가 및 나머지의 제거와 함께, 파싱 및 데이터베이스에의 후속 삽입), 구조화된 데이터 내 패턴들을 유도하고, 최종적으로, 출력을 평가 및 해석하는 프로세스를 포함할 수도 있다. 텍스트 특징들에서의 "고 품질"은 관련성, 신규성, 및 흥미도의 일부 조합을 지칭할 수도 있다는 점에 유의해야 한다. 전형적인 텍스트 특징들은 텍스트 분류, 텍스트 클러스터링, 개념/엔티티 추출, 세분화된 분류법들의 생성, 정서 분석, 문서 요약, 및 엔티티 관계 모델링(즉, 명명된 엔티티들 사이의 학습 관계들)을 포함할 수도 있다. 일부 실시형태들에서, 텍스트 분석은 정보 리트리빙, 단어 빈도 분포들을 연구하기 위한 어휘 분석, 패턴 인식, 태깅/주석, 정보 추출, 링크 및 연관 분석을 포함하는 데이터 마이닝 기법들, 시각화, 및 예측 해석학을 수반한다.
- [0134] 단계 1614에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) 노드들을 스코어링한다. 다수의 접속된 컴포넌트들을 갖는 일부 실시형태들에서, 시스템은 이들 특징들을 각각의 접속된 컴포넌트 상에서 별개로 계산한다. 일부 실시형태들에서, 시스템은 위에서 설명된 9개(7개의 텍스트 특징들 및 2개의 그래프 특징들)를 이용하여, 이들을 정규화하고, 수동으로-선별된 데이터에 걸쳐서 분류기를 트레이닝하고 이 모델을 이용하여 엔티티들을 예측한다. 특히 구체적인 구현예에서 분류를 구현하는 알고리즘은 분류기로서 알려져 있다. 분류 및 클러스터링은 주어진 입력 값에의 일종의 출력 값의 할당인 패턴 인식의 좀더 일반적인 문제의 예들이다. 다른 예들은 실수 값의 출력을 각각의 입력에 할당하는 회귀; 클래스를 일련의 값들의 각각의 멤버에 할당하는 시퀀스 라벨링(예를 들어, 품사를 입력 문장에서의 각각의 단어에 할당하는 품사 태깅); 파스(parse) 트리를 입력 문장에 할

당하여 문장의 구문 구조를 기술하는 파싱; 등이다.

- [0135] 일부 실시형태들에서, 시스템은 의사결정 트리 분류기 및 랜덤 포레스트 분류기를 이용해 트레이닝한다. 의사결정 트리 분류기는 플로우차트형 구조이며, 여기서, 각각의 내부(년-리프) 노드는 속성에 대한 테스트를 나타내고, 각각의 분기부는 테스트의 결과를 나타내고, 각각의 리프(또는, 터미널) 노드는 클래스 라벨을 보유한다. 트리에서의 최상부 노드는 루트 노드이다. 랜덤 포레스트 분류기는 트레이닝 시간에 다수의 의사결정 트리들을 구성하고 개개의 트리들의 클래스들의 모드(분류) 또는 평균 예측(회귀)인 클래스를 출력함으로써 동작할 수도 있다. 무작위 결정 포레스트들은 이들의 트레이닝 세트에 오버피팅하는 의사결정 트리의 습관을 수정한다. 시스템이 임의의 결정-트리 알고리즘들을 구현할 수도 있다는 점에 유의해야 한다. 단계 614에서, 시스템은 (예컨대, 제어 회로부(304)를 통해서) (예컨대, 도 9 내지 도 10과 관련하여 도시 및 설명된 바와 같이) 엔티티들을 결정한다.
- [0136] 본 개시물의 위에서 설명한 실시형태들은 예시의 목적을 위해 그리고 제한없이 제시되며, 본 개시물은 다음의 청구범위에 의해서만 제한된다. 더욱이, 임의의 일 실시형태에서 설명된 특징들 및 제한들이 본원에서의 임의의 다른 실시형태에 적용될 수도 있으며 일 실시형태에 관련된 플로우차트들 또는 예들이 임의의 다른 실시형태와 적합한 방법으로 결합되거나, 상이한 순서들로 이루어지거나, 또는 병렬로 이루어질 수도 있다는 점에 유의해야 한다. 게다가, 본원에서 설명되는 시스템들 및 방법들은 실시간으로 수행될 수도 있다. 또한, 시스템들 및/또는 위에서 설명한 방법들이 다른 시스템들 및/또는 방법들에 적용될 수도 있거나, 또는 다른 시스템들 및/또는 방법들에 따라서 사용될 수도 있다는 점에 유의해야 한다.
- [0137] 본 명세서에서는 다음을 포함하지만 이에 한정되지 않는 실시형태들을 개시한다:
- [0138] 1. 검색, 추천 및 발견 특징들을 제공하는 방법으로서,
- [0139] 제어 회로부에 의해, 데이터 세트를 수집하는 단계;
- [0140] 상기 제어 회로부에 의해, 상기 데이터 세트에 걸쳐서 대명사 분석을 수행하는 단계;
- [0141] 상기 제어 회로부에 의해, 상기 데이터 세트에 걸쳐서 후보 식별을 수행하는 단계;
- [0142] 상기 제어 회로부에 의해, 복수의 키 엔티티들 및 상기 복수의 키 엔티티들 사이의 복수의 연관들을 식별하는 시맨틱 그래프를 생성하는 단계;
- [0143] 사용자 입력 인터페이스에 의해, 사용자 입력을 수신하는 단계;
- [0144] 상기 제어 회로부에 의해, 상기 시맨틱 그래프를 이용하여 상기 사용자 입력을 프로세싱하는 단계; 및
- [0145] 상기 제어 회로부에 의해, 프로세싱된 상기 사용자 입력에 기초하여 출력을 생성하는 단계를 포함하는, 방법.
- [0146] 2. 항목 1에 있어서,
- [0147] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 방법.
- [0148] 3. 항목 1에 있어서,
- [0149] 상기 데이터 세트는 검증 데이터에 대한 트레이닝 데이터의 비로 분할되며, 상기 트레이닝 데이터는 상기 시맨틱 그래프에 대해 상기 제어 회로부를 트레이닝하는데 사용되는, 방법.
- [0150] 4. 항목 1에 있어서,
- [0151] 상기 대명사 분석을 수행하는 단계는 코렐퍼런스 분석을 이용하여 상기 대명사를 분석하는 단계를 포함하는, 방법.
- [0152] 5. 항목 1에 있어서,
- [0153] 상기 후보 식별은 문법적 태깅 및 단어-카테고리 명확화를 포함하는, 방법.
- [0154] 6. 항목 1에 있어서,
- [0155] 상기 사용자 입력은 사용자로부터 또는 전자 디바이스로부터 직접 수신되는, 방법.
- [0156] 7. 항목 1에 있어서,

- [0157] 상기 사용자 입력을 프로세싱하는 단계는 상기 사용자 입력으로부터의 복수의 후보들을 상기 시맨틱 그래프에서의 복수의 노드들과 매칭시키는 단계를 포함하는, 방법.
- [0158] 8. 항목 1에 있어서,
- [0159] 상기 사용자 입력으로부터의 복수의 후보들 사이의 복수의 관계들은 종속성 트리를 횡단함으로써 식별되는, 방법.
- [0160] 9. 항목 1에 있어서,
- [0161] 상기 출력은 상기 사용자 입력에 기초한 검색 결과 또는 추천을 포함하는, 방법.
- [0162] 10. 항목 1에 있어서,
- [0163] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 방법.
- [0164] 11. 검색, 추천 및 발견 특징들을 제공하는 시스템으로서,
- [0165] 메모리; 및
- [0166] 제어 회로부를 포함하되,
- [0167] 상기 제어 회로부는,
- [0168] 데이터 세트를 수집하고;
- [0169] 상기 데이터 세트에 걸쳐서 대명사 분석을 수행하고;
- [0170] 상기 데이터 세트에 걸쳐서 후보 식별을 수행하고;
- [0171] 복수의 키 엔티티들 및 상기 복수의 키 엔티티들 사이의 복수의 연관들을 식별하는 시맨틱 그래프를 생성하고;
- [0172] 사용자 입력을 수신하고;
- [0173] 상기 시맨틱 그래프를 이용하여 상기 사용자 입력을 프로세싱하고; 그리고
- [0174] 프로세싱된 상기 사용자 입력에 기초하여 출력을 생성하도록 구성되는, 시스템.
- [0175] 12. 항목 11에 있어서,
- [0176] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 시스템.
- [0177] 13. 항목 11에 있어서,
- [0178] 상기 데이터 세트는 검증 데이터에 대한 트레이닝 데이터의 비로 분할되며, 상기 트레이닝 데이터는 상기 시맨틱 그래프에 대해 상기 제어 회로부를 트레이닝하는데 사용되는, 시스템.
- [0179] 14. 항목 11에 있어서,
- [0180] 상기 대명사 분석을 수행하는 단계는 코레퍼런스 분석을 이용하여 상기 대명사를 분석하는 단계를 포함하는, 시스템.
- [0181] 15. 항목 11에 있어서,
- [0182] 상기 후보 식별은 문법적 태깅 및 단어-카테고리 명확화를 포함하는, 시스템.
- [0183] 16. 항목 11에 있어서,
- [0184] 상기 사용자 입력은 사용자로부터 또는 전자 디바이스로부터 직접 수신되는, 시스템.
- [0185] 17. 항목 11에 있어서,
- [0186] 상기 사용자 입력을 프로세싱하는 단계는 상기 사용자 입력으로부터의 복수의 후보들을 상기 시맨틱 그래프에서의 복수의 노드들과 매칭시키는 단계를 포함하는, 시스템.
- [0187] 18. 항목 11에 있어서,
- [0188] 상기 사용자 입력으로부터의 복수의 후보들 사이의 복수의 관계들은 종속성 트리를 횡단함으로써 식별되는, 시

시스템.

- [0189] 19. 항목 11에 있어서,
- [0190] 상기 출력은 상기 사용자 입력에 기초한 검색 결과 또는 추천을 포함하는, 시스템.
- [0191] 20. 항목 11에 있어서,
- [0192] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 시스템.
- [0193] 21. 검색, 추천 및 발견 특징들을 제공하는 시스템으로서,
- [0194] 데이터 세트를 수집하는 수단;
- [0195] 상기 데이터 세트에 걸쳐서 대명사 분석을 수행하는 수단;
- [0196] 상기 데이터 세트에 걸쳐서 후보 식별을 수행하는 수단;
- [0197] 복수의 키 엔티티들 및 상기 복수의 키 엔티티들 사이의 복수의 연관들을 식별하는 시맨틱 그래프를 생성하는 수단;
- [0198] 사용자 입력을 수신하는 수단;
- [0199] 상기 시맨틱 그래프를 이용하여 상기 사용자 입력을 프로세싱하는 수단; 및
- [0200] 프로세싱된 상기 사용자 입력에 기초하여 출력을 생성하는 수단을 포함하는, 시스템.
- [0201] 22. 항목 21에 있어서,
- [0202] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 시스템.
- [0203] 23. 항목 21에 있어서,
- [0204] 상기 데이터 세트는 검증 데이터에 대한 트레이닝 데이터의 비로 분할되며, 상기 트레이닝 데이터는 상기 시맨틱 그래프에 대해 상기 제어 회로부를 트레이닝하는데 사용되는, 시스템.
- [0205] 24. 항목 21에 있어서,
- [0206] 상기 대명사 분석을 수행하는 것은 코렐러런스 분석을 이용하여 상기 대명사를 분석하는 것을 포함하는, 시스템.
- [0207] 25. 항목 21에 있어서,
- [0208] 상기 후보 식별은 문법적 태깅 및 단어-카테고리 명확화를 포함하는, 시스템.
- [0209] 26. 항목 21에 있어서,
- [0210] 상기 사용자 입력은 사용자로부터 또는 전자 디바이스로부터 직접 수신되는, 시스템.
- [0211] 27. 항목 21에 있어서,
- [0212] 상기 사용자 입력을 프로세싱하는 것은 상기 사용자 입력으로부터의 복수의 후보들을 상기 시맨틱 그래프에서의 복수의 노드들과 매칭시키는 것을 포함하는, 시스템.
- [0213] 28. 항목 21에 있어서,
- [0214] 상기 사용자 입력으로부터의 복수의 후보들 사이의 복수의 관계들은 종속성 트리를 횡단함으로써 식별되는, 시스템.
- [0215] 29. 항목 21에 있어서,
- [0216] 상기 출력은 상기 사용자 입력에 기초한 검색 결과 또는 추천을 포함하는, 시스템.
- [0217] 30. 항목 21에 있어서,
- [0218] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 시스템.

- [0219] 31. 검색, 추천 및 발견 특징들을 제공하는 방법으로서,
- [0220] 데이터 세트를 수집하는 단계;
- [0221] 상기 데이터 세트에 걸쳐서 대명사 분석을 수행하는 단계;
- [0222] 상기 데이터 세트에 걸쳐서 후보 식별을 수행하는 단계;
- [0223] 복수의 키 엔티티들 및 상기 복수의 키 엔티티들 사이의 복수의 연관들을 식별하는 시맨틱 그래프를 생성하는 단계;
- [0224] 사용자 입력을 수신하는 단계;
- [0225] 상기 시맨틱 그래프를 이용하여 상기 사용자 입력을 프로세싱하는 단계; 및
- [0226] 프로세싱된 상기 사용자 입력에 기초하여 출력을 생성하는 단계를 포함하는, 방법.
- [0227] 32. 항목 31에 있어서,
- [0228] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 방법.
- [0229] 33. 항목 31 또는 32에 있어서,
- [0230] 상기 데이터 세트는 검증 데이터에 대한 트레이닝 데이터의 비로 분할되며, 상기 트레이닝 데이터는 상기 시맨틱 그래프에 대해 상기 제어 회로부를 트레이닝하는데 사용되는, 방법.
- [0231] 34. 항목 31 내지 33에 있어서,
- [0232] 상기 대명사 분석을 수행하는 단계는 코렐퍼런스 분석을 이용하여 상기 대명사를 분석하는 단계를 포함하는, 방법.
- [0233] 35. 항목 31 내지 34에 있어서,
- [0234] 상기 후보 식별은 문법적 태깅 및 단어-카테고리 명확화를 포함하는, 방법.
- [0235] 36. 항목 31 내지 35에 있어서,
- [0236] 상기 사용자 입력은 사용자로부터 또는 전자 디바이스로부터 직접 수신되는, 방법.
- [0237] 37. 항목 31 내지 36에 있어서,
- [0238] 상기 사용자 입력을 프로세싱하는 단계는 상기 사용자 입력으로부터의 복수의 후보들을 상기 시맨틱 그래프에서의 복수의 노드들과 매칭시키는 단계를 포함하는, 방법.
- [0239] 38. 항목 31 내지 37에 있어서,
- [0240] 상기 사용자 입력으로부터의 복수의 후보들 사이의 복수의 관계들은 종속성 트리를 횡단함으로써 식별되는, 방법.
- [0241] 39. 항목 31 내지 38에 있어서,
- [0242] 상기 출력은 상기 사용자 입력에 기초한 검색 결과 또는 추천을 포함하는, 방법.
- [0243] 40. 항목 31 내지 39에 있어서,
- [0244] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 방법.
- [0245] 41. 검색, 추천 및 발견 특징들을 제공하기 위한 명령어들을 기록한 비일시적 컴퓨터-판독가능 매체로서,
- [0246] 상기 명령어들은,
- [0247] 데이터 세트를 수집하는 명령어;
- [0248] 상기 데이터 세트에 걸쳐서 대명사 분석을 수행하는 명령어;
- [0249] 상기 데이터 세트에 걸쳐서 후보 식별을 수행하는 명령어;
- [0250] 복수의 키 엔티티들 및 상기 복수의 키 엔티티들 사이의 복수의 연관들을 식별하는 시맨틱 그래프를 생성하는

명령어;

- [0251] 사용자 입력을 수신하는 명령어;
- [0252] 상기 시맨틱 그래프를 이용하여 상기 사용자 입력을 프로세싱하는 명령어; 및
- [0253] 프로세싱된 상기 사용자 입력에 기초하여 출력을 생성하는 명령어를 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0254] 42. 항목 41에 있어서,
- [0255] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 비일시적 컴퓨터-판독가능 매체.
- [0256] 43. 항목 41에 있어서,
- [0257] 상기 데이터 세트는 검증 데이터에 대한 트레이닝 데이터의 비로 분할되며, 상기 트레이닝 데이터는 상기 시맨틱 그래프에 대해 상기 제어 회로부를 트레이닝하는데 사용되는, 비일시적 컴퓨터-판독가능 매체.
- [0258] 44. 항목 41에 있어서,
- [0259] 상기 대명사 분석을 수행하는 것은 코레퍼런스 분석을 이용하여 상기 대명사를 분석하는 것을 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0260] 45. 항목 41에 있어서,
- [0261] 상기 후보 식별은 문법적 태깅 및 단어-카테고리 명확화를 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0262] 46. 항목 41에 있어서,
- [0263] 상기 사용자 입력은 사용자로부터 또는 전자 디바이스로부터 직접 수신되는, 비일시적 컴퓨터-판독가능 매체.
- [0264] 47. 항목 41에 있어서,
- [0265] 상기 사용자 입력을 프로세싱하는 것은 상기 사용자 입력으로부터의 복수의 후보들을 상기 시맨틱 그래프에서의 복수의 노드들과 매칭시키는 것을 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0266] 48. 항목 41에 있어서,
- [0267] 상기 사용자 입력으로부터의 복수의 후보들 사이의 복수의 관계들은 종속성 트리를 횡단함으로써 식별되는, 비일시적 컴퓨터-판독가능 매체.
- [0268] 49. 항목 41에 있어서,
- [0269] 상기 출력은 상기 사용자 입력에 기초한 검색 결과 또는 추천을 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0270] 50. 항목 41에 있어서,
- [0271] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 비일시적 컴퓨터-판독가능 매체.
- [0272] 51. 텍스트 문자열들에서 엔티티들의 관련성들을 자동으로 결정함으로써 콘텐츠 추천들을 제공하는 방법으로서,
- [0273] 사용자 입력 인터페이스에 의해, 텍스트 문자열을 수신하는 단계;
- [0274] 제어 회로부에 의해, 상기 텍스트 문자열에서 대명사를 식별하는 단계;
- [0275] 상기 제어 회로부에 의해, 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 단계;
- [0276] 상기 제어 회로부에 의해, 상기 분석된 텍스트 문자열에서 명사 청크를 식별하는 단계;
- [0277] 상기 제어 회로부에 의해, 복수의 노드들을 특징으로 하는 시맨틱 그래프에 기초하여 분류기를 이용해 상기 명사 청크를 프로세싱하는 단계로서, 상기 복수의 노드들의 각각은 근접 중심성 메트릭 및 매개 중심성 메트릭에 기초하여 스코어링되며, 상기 근접 중심성 메트릭은 상기 시맨틱 그래프에서의 개개 노드와 다른 노드들의 각각 사이의 최단 경로의 길이의 총합의 척도이고, 상기 매개 중심성 메트릭은 개개 노드의 상기 시맨틱 그래프에서의 중심성의 척도인, 상기 명사 청크를 프로세싱하는 단계;
- [0278] 상기 제어 회로부에 의해, 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정

하는 단계; 및

- [0279] 디스플레이 디바이스 상에서, 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 단계를 포함하는, 방법.
- [0280] 52. 항목 51에 있어서,
- [0281] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 방법.
- [0282] 53. 항목 51에 있어서,
- [0283] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 단계는,
- [0284] 각각의 엔티티를 스코어링하는 단계;
- [0285] 해당 개개 스코어에 기초하여 각각의 엔티티를 순위지정하는 단계; 및
- [0286] 최고 스코어를 가진 엔티티를 선택하는 단계를 포함하는, 방법.
- [0287] 54. 항목 53에 있어서,
- [0288] 각각의 엔티티는 7개의 텍스트 특징들 및 2개의 그래프 특징들에 대해 스코어링되는, 방법.
- [0289] 55. 항목 51에 있어서,
- [0290] 상기 분류기는 의사결정 트리 분류기 또는 랜덤 포레스트 분류기인, 방법.
- [0291] 56. 항목 51에 있어서,
- [0292] 상기 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 단계는 검색, 추천, 또는 발견 특징에서 상기 엔티티를 디스플레이를 위해 생성하는 단계를 포함하는, 방법.
- [0293] 57. 항목 51에 있어서,
- [0294] 상기 텍스트 문자열은 사용자로부터 또는 전자 디바이스로부터 수신되는, 방법.
- [0295] 58. 항목 51에 있어서,
- [0296] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 단계는 코레퍼런스 분석을 이용하여 상기 대명사를 분석하는 단계를 포함하는, 방법.
- [0297] 59. 항목 51에 있어서,
- [0298] 상기 분석된 텍스트 문자열에서 상기 명사 청크를 식별하는 단계는 품사 태깅을 이용하여 상기 명사 청크를 식별하는 단계를 포함하는, 방법.
- [0299] 60. 항목 51에 있어서,
- [0300] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 방법.
- [0301] 61. 텍스트 문자열들에서 엔티티들의 관련성들을 자동으로 결정함으로써 콘텐츠 추천을 제공하는 시스템으로서,
- [0302] 메모리; 및
- [0303] 제어 회로부를 포함하며,
- [0304] 상기 제어 회로부는,
- [0305] 텍스트 문자열을 수신하고;
- [0306] 상기 텍스트 문자열에서 대명사를 식별하고;
- [0307] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하고;
- [0308] 상기 분석된 텍스트 문자열에서 명사 청크를 식별하고;
- [0309] 복수의 노드들을 특징으로 하는 시맨틱 그래프에 기초하여 분류기를 이용해 상기 명사 청크를 프로세싱하고 - 상기 복수의 노드들의 각각은 근접 중심성 메트릭 및 매개 중심성 메트릭에 기초하여 스코어링되며, 상기 근접

중심성 메트릭은 상기 시맨틱 그래프에서의 개개 노드와 다른 노드들의 각각 사이의 최단 경로의 길이의 총합의 척도이고, 상기 매개 중심성 메트릭은 개개 노드의 상기 시맨틱 그래프에서의 중심성의 척도임 -;

- [0310] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하고; 그리고
- [0311] 수신된 텍스트 문자열에 응답하여 상기 엔티티를 디스플레이를 위해 생성하도록 구성되는, 시스템.
- [0312] 62. 항목 61에 있어서,
- [0313] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 시스템.
- [0314] 63. 항목 61에 있어서,
- [0315] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 것은,
- [0316] 각각의 엔티티를 스코어링하는 것;
- [0317] 해당 개개 스코어에 기초하여 각각의 엔티티를 순위지정하는 것; 및
- [0318] 최고 스코어를 가진 엔티티를 선택하는 것을 포함하는, 시스템.
- [0319] 64. 항목 63에 있어서,
- [0320] 각각의 엔티티는 7개의 텍스트 특징들 및 2개의 그래프 특징들에 대해 스코어링되는, 시스템.
- [0321] 65. 항목 61에 있어서,
- [0322] 상기 분류기는 의사결정 트리 분류기 또는 랜덤 포레스트 분류기인, 시스템.
- [0323] 66. 항목 61에 있어서,
- [0324] 상기 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 것은 검색, 추천, 또는 발견 특징에서 상기 엔티티를 디스플레이를 위해 생성하는 것을 포함하는, 시스템.
- [0325] 67. 항목 61에 있어서,
- [0326] 상기 텍스트 문자열은 사용자로부터 또는 전자 디바이스로부터 수신되는, 시스템.
- [0327] 68. 항목 61에 있어서,
- [0328] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 것은 코레퍼런스 분석을 이용하여 상기 대명사를 분석하는 것을 포함하는, 시스템.
- [0329] 69. 항목 61에 있어서,
- [0330] 상기 분석된 텍스트 문자열에서 상기 명사 청크를 식별하는 것은 품사 태깅을 이용하여 상기 명사 청크를 식별하는 것을 포함하는, 시스템.
- [0331] 70. 항목 61에 있어서,
- [0332] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 시스템.
- [0333] 71. 텍스트 문자열들에서 엔티티들의 관련성들을 자동으로 결정함으로써 콘텐츠 추천을 제공하는 시스템으로서,
- [0334] 텍스트 문자열을 수신하는 수단;
- [0335] 상기 텍스트 문자열에서 대명사를 식별하는 수단;
- [0336] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 수단;
- [0337] 상기 분석된 텍스트 문자열에서 명사 청크를 식별하는 수단;
- [0338] 복수의 노드들을 특징으로 하는 시맨틱 그래프에 기초하여 분류기를 이용해 상기 명사 청크를 프로세싱하는 수단으로서, 상기 복수의 노드들의 각각은 근접 중심성 메트릭 및 매개 중심성 메트릭에 기초하여 스코어링되며, 상기 근접 중심성 메트릭은 상기 시맨틱 그래프에서의 개개 노드와 다른 노드들의 각각 사이의 최단 경로의 길이의 총합의 척도이고, 상기 매개 중심성 메트릭은 개개 노드의 상기 시맨틱 그래프에서의 중심성의 척도인, 상

기 명사 청크를 프로세싱하는 수단;

- [0339] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 수단; 및
- [0340] 디스플레이 디바이스 상에서, 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 수단을 포함하는, 시스템.
- [0341] 72. 항목 71에 있어서,
- [0342] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 시스템.
- [0343] 73. 항목 71에 있어서,
- [0344] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 것은,
- [0345] 각각의 엔티티를 스코어링하는 것;
- [0346] 해당 개개 스코어에 기초하여 각각의 엔티티를 순위지정하는 것; 및
- [0347] 최고 스코어를 가진 엔티티를 선택하는 것을 포함하는, 시스템.
- [0348] 74. 항목 73에 있어서,
- [0349] 각각의 엔티티는 7개의 텍스트 특징들 및 2개의 그래프 특징들에 대해 스코어링되는, 시스템.
- [0350] 75. 항목 71에 있어서,
- [0351] 상기 분류기는 의사결정 트리 분류기 또는 랜덤 포레스트 분류기인, 시스템.
- [0352] 76. 항목 71에 있어서,
- [0353] 상기 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 것은 검색, 추천, 또는 발견 특징에서 상기 엔티티를 디스플레이를 위해 생성하는 것을 포함하는, 시스템.
- [0354] 77. 항목 71에 있어서,
- [0355] 상기 텍스트 문자열은 사용자로부터 또는 전자 디바이스로부터 수신되는, 시스템.
- [0356] 78. 항목 71에 있어서,
- [0357] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 것은 코레퍼런스 분석을 이용하여 상기 대명사를 분석하는 것을 포함하는, 시스템.
- [0358] 79. 항목 71에 있어서,
- [0359] 상기 분석된 텍스트 문자열에서 상기 명사 청크를 식별하는 것은 품사 태깅을 이용하여 상기 명사 청크를 식별하는 것을 포함하는, 시스템.
- [0360] 80. 항목 71에 있어서,
- [0361] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 시스템.
- [0362] 81. 텍스트 문자열들에서 엔티티들의 관련성들을 자동으로 결정함으로써 콘텐츠 추천을 제공하는 방법으로서,
- [0363] 텍스트 문자열을 수신하는 단계;
- [0364] 상기 텍스트 문자열에서 대명사를 식별하는 단계;
- [0365] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 단계;
- [0366] 상기 분석된 텍스트 문자열에서 명사 청크를 식별하는 단계;
- [0367] 복수의 노드들을 특징으로 하는 시맨틱 그래프에 기초하여 분류기를 이용해 상기 명사 청크를 프로세싱하는 단계로서, 상기 복수의 노드들의 각각은 근접 중심성 메트릭 및 매개 중심성 메트릭에 기초하여 스코어링되며, 상기 근접 중심성 메트릭은 상기 시맨틱 그래프에서의 개개 노드와 다른 노드들의 각각 사이의 최단 경로의 길이의 총합의 척도이고, 상기 매개 중심성 메트릭은 개개 노드의 상기 시맨틱 그래프에서의 중심성의 척도인, 상기

명사 청크를 프로세싱하는 단계;

- [0368] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 단계; 및
- [0369] 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 단계를 포함하는, 방법.
- [0370] 82. 항목 81에 있어서,
- [0371] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 방법.
- [0372] 83. 항목 81 또는 82에 있어서,
- [0373] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 단계는,
- [0374] 각각의 엔티티를 스코어링하는 단계;
- [0375] 해당 개개 스코어에 기초하여 각각의 엔티티를 순위지정하는 단계; 및
- [0376] 최고 스코어를 가진 엔티티를 선택하는 단계를 포함하는, 방법.
- [0377] 84. 항목 83에 있어서,
- [0378] 각각의 엔티티는 7개의 텍스트 특징들 및 2개의 그래프 특징들에 대해 스코어링되는, 방법.
- [0379] 85. 항목 81 내지 84 중 어느 하나의 항목에 있어서,
- [0380] 상기 분류기는 의사결정 트리 분류기 또는 랜덤 포레스트 분류기인, 방법.
- [0381] 86. 항목 81 내지 85 중 어느 하나의 항목에 있어서,
- [0382] 상기 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 단계는 검색, 추천, 또는 발견 특징에서 상기 엔티티를 디스플레이를 위해 생성하는 단계를 포함하는, 방법.
- [0383] 87. 항목 81 내지 86 중 어느 하나의 항목에 있어서,
- [0384] 상기 텍스트 문자열은 사용자로부터 또는 전자 디바이스로부터 수신되는, 방법.
- [0385] 88. 항목 81 내지 87 중 어느 하나의 항목에 있어서,
- [0386] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 단계는 코레퍼런스 분석을 이용하여 상기 대명사를 분석하는 단계를 포함하는, 방법.
- [0387] 89. 항목 81 내지 88 중 어느 하나의 항목에 있어서,
- [0388] 상기 분석된 텍스트 문자열에서 상기 명사 청크를 식별하는 단계는 품사 태깅을 이용하여 상기 명사 청크를 식별하는 단계를 포함하는, 방법.
- [0389] 90. 항목 81 내지 89 중 어느 하나의 항목에 있어서,
- [0390] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 방법.
- [0391] 91. 텍스트 문자열들에서 엔티티들의 관련성들을 자동으로 결정함으로써 콘텐츠 추천을 제공하기 위한 명령어들을 기록한 비일시적 컴퓨터-판독가능 매체로서,
- [0392] 상기 명령어들은,
- [0393] 텍스트 문자열을 수신하는 명령어;
- [0394] 상기 텍스트 문자열에서 대명사를 식별하는 명령어;
- [0395] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 명령어;
- [0396] 상기 분석된 텍스트 문자열에서 명사 청크를 식별하는 명령어;
- [0397] 복수의 노드들을 특징으로 하는 시맨틱 그래프에 기초하여 분류기를 이용해 상기 명사 청크를 프로세싱하는 명령어로서, 상기 복수의 노드들의 각각은 근접 중심성 메트릭 및 매개 중심성 메트릭에 기초하여 스코어링되며, 상기 근접 중심성 메트릭은 상기 시맨틱 그래프에서의 개개 노드와 다른 노드들 각각 사이의 최단 경로의 길이

의 총합의 척도이고, 상기 매개 중심성 메트릭은 개개 노드의 상기 시맨틱 그래프에서의 중심성의 척도인, 상기 명사 청크를 프로세싱하는 명령어;

- [0398] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 명령어; 및
- [0399] 디스플레이 디바이스 상에서, 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 명령어를 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0400] 92. 항목 91에 있어서,
- [0401] 상기 시맨틱 그래프는 복수의 노드들을 포함하며, 상기 복수의 노드들의 각각은 엔티티들의 데이터 세트로부터의 엔티티에 대응하는, 비일시적 컴퓨터-판독가능 매체.
- [0402] 93. 항목 91에 있어서,
- [0403] 상기 분류기를 이용해 상기 명사 청크를 프로세싱하는 것에 기초하여 엔티티를 결정하는 것은,
- [0404] 각각의 엔티티를 스코어링하는 것;
- [0405] 해당 개개 스코어에 기초하여 각각의 엔티티를 순위지정하는 것; 및
- [0406] 최고 스코어를 가진 엔티티를 선택하는 것을 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0407] 94. 항목 93에 있어서,
- [0408] 각각의 엔티티는 7개의 텍스트 특징들 및 2개의 그래프 특징들에 대해 스코어링되는, 비일시적 컴퓨터-판독가능 매체.
- [0409] 95. 항목 91에 있어서,
- [0410] 상기 분류기는 의사결정 트리 분류기 또는 랜덤 포레스트 분류기인, 비일시적 컴퓨터-판독가능 매체.
- [0411] 96. 항목 91에 있어서,
- [0412] 상기 수신된 텍스트 문자열에 응답하여, 상기 엔티티를 디스플레이를 위해 생성하는 것은 검색, 추천, 또는 발견 특징에서 상기 엔티티를 디스플레이를 위해 생성하는 것을 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0413] 97. 항목 91에 있어서,
- [0414] 상기 텍스트 문자열은 사용자로부터 또는 전자 디바이스로부터 수신되는, 비일시적 컴퓨터-판독가능 매체.
- [0415] 98. 항목 91에 있어서,
- [0416] 상기 대명사를 명사로 분석하여 분석된 텍스트 문자열을 생성하는 것은 코레퍼런스 분석을 이용하여 상기 대명사를 분석하는 것을 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0417] 99. 항목 91에 있어서,
- [0418] 상기 분석된 텍스트 문자열에서 상기 명사 청크를 식별하는 것은 품사 태깅을 이용하여 상기 명사 청크를 식별하는 것을 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0419] 100. 항목 91에 있어서,
- [0420] 상기 시맨틱 그래프는 네트워크에서의 개념들 사이의 시맨틱 관계들을 나타내는 지식 베이스인, 비일시적 컴퓨터-판독가능 매체.
- [0421] 101. 검색 쿼리에 대한 결과들을 제공하는 컴퓨터-구현 방법으로서,
- [0422] 복수의 엔티티들에 대한 참조들을 포함하는 검색 쿼리를 수신하는 단계;
- [0423] 제어 회로부를 이용하여, 상기 검색 쿼리에 대해 대명사 분석을 수행하는 단계;
- [0424] 상기 제어 회로부를 이용하여, 복수의 에지들에 의해 커플링된 복수의 노드들을 포함하는 그래프를 이용하여 상기 검색 쿼리를 프로세싱하는 단계로서, 상기 복수의 노드들의 서브세트는 상기 검색 쿼리의 분석된 대명사들과 연관된 상기 복수의 엔티티들을 나타내는, 상기 검색 쿼리를 프로세싱하는 단계;
- [0425] 상기 제어 회로부를 이용하여, 근접 메트릭에 기초하여 상기 서브세트의 2개의 노드들에 커플링되는 상기 그래

프의 노드를 식별하는 단계로서, 상기 근접 메트릭은 상기 노드와 상기 2개의 노드들 중 하나 사이의 거리에 반비례하는, 상기 식별하는 단계; 및

- [0426] 상기 그래프의 상기 노드에 의해 표현되는 엔티티에 대한 참조를 상기 검색 쿼리에 대한 결과로서 제공되게 하는 단계를 포함하는, 검색 쿼리에 대한 결과들을 제공하는 컴퓨터-구현 방법.
- [0427] 102. 항목 101에 있어서,
- [0428] 상기 복수의 엔티티들은 제 1 복수의 엔티티들이고, 상기 그래프를 이용하여 상기 검색 쿼리를 프로세싱하는 단계는 상기 제 1 복수의 엔티티들을 상기 복수의 노드들에 의해 표현되는 제 2 복수의 엔티티들에 매칭하는 단계를 포함하는, 검색 쿼리에 대한 결과들을 제공하는 컴퓨터-구현 방법.
- [0429] 103. 항목 101 또는 102에 있어서,
- [0430] 대화의 사용자 톤을 반영하는 복수의 데이터 세트들을 선택하는 단계를 더 포함하는, 검색 쿼리에 대한 결과들을 제공하는 컴퓨터-구현 방법.
- [0431] 104. 항목 103에 있어서,
- [0432] 품사(POS) 태깅을 이용하여 상기 복수의 데이터 세트들을 프로세싱하는 단계를 더 포함하는, 검색 쿼리에 대한 결과들을 제공하는 컴퓨터-구현 방법.
- [0433] 105. 항목 101 내지 104 중 어느 하나의 항목에 있어서,
- [0434] 상기 근접 메트릭에 기초하여 상기 서브세트의 2개의 노드들에 커플링된 상기 그래프의 상기 노드를 식별하는 단계는,
- [0435] 상기 개개 노드와 상기 복수의 노드들의 다른 노드 사이의 거리에 기초하여 상기 그래프의 상기 복수의 노드들의 개개 노드를 스코어링하는 단계; 및
- [0436] 상기 노드가 상기 그래프의 상기 스코어링된 복수의 노드들의 최고 스코어를 갖는 것으로 결정하는 단계를 포함하는, 검색 쿼리에 대한 결과들을 제공하는 컴퓨터-구현 방법.
- [0437] 106. 검색 쿼리에 대한 결과들을 제공하는 시스템으로서,
- [0438] 복수의 엔티티들에 대한 참조들을 포함하는 검색 쿼리를 수신하는 수단;
- [0439] 상기 검색 쿼리에 대한 대명사 분석을 수행하는 수단;
- [0440] 복수의 에지들에 의해 커플링된 복수의 노드들을 포함하는 그래프를 이용하여 상기 검색 쿼리를 프로세싱하는 수단으로서, 상기 복수의 노드들의 서브세트는 상기 검색 쿼리의 분석된 대명사들과 연관된 상기 복수의 엔티티들을 나타내는, 상기 검색 쿼리를 프로세싱하는 수단;
- [0441] 근접 메트릭에 기초하여 상기 서브세트의 2개의 노드들에 커플링되는 상기 그래프의 노드를 식별하는 수단으로서, 상기 근접 메트릭은 상기 노드와 상기 2개의 노드들 중 하나 사이의 거리에 반비례하는, 상기 식별하는 수단; 및
- [0442] 상기 그래프의 상기 노드에 의해 표현되는 엔티티에 대한 참조를 상기 검색 쿼리에 대한 결과로서 제공되게 하는 수단을 포함하는, 검색 쿼리에 대한 결과들을 제공하는 시스템.
- [0443] 107. 항목 106에 있어서,
- [0444] 상기 복수의 엔티티들은 제 1 복수의 엔티티들이고, 상기 그래프를 이용하여 상기 검색 쿼리를 프로세싱하는 것은 상기 제 1 복수의 엔티티들을 상기 복수의 노드들에 의해 표현되는 제 2 복수의 엔티티들에 매칭하는 것을 포함하는, 검색 쿼리에 대한 결과들을 제공하는 시스템.
- [0445] 108. 항목 106 또는 107에 있어서,
- [0446] 대화의 사용자 톤을 반영하는 복수의 데이터 세트들을 선택하는 수단을 더 포함하는, 검색 쿼리에 대한 결과들을 제공하는 시스템.
- [0447] 109. 항목 108에 있어서,
- [0448] 품사(POS) 태깅을 이용하여 상기 복수의 데이터 세트들을 프로세싱하는 수단을 더 포함하는, 검색 쿼리에 대한

결과들을 제공하는 시스템.

- [0449] 110. 항목 101 내지 109 중 어느 하나의 항목에 있어서,
- [0450] 상기 근접 메트릭에 기초하여 상기 서브세트의 2개의 노드들에 커플링된 상기 그래프의 상기 노드를 식별하는 것은,
- [0451] 상기 개개 노드와 상기 복수의 노드들의 다른 노드 사이의 거리에 기초하여 상기 그래프의 상기 복수의 노드들의 개개 노드를 스코어링하는 것; 및
- [0452] 상기 노드가 상기 그래프의 상기 스코어링된 복수의 노드들의 최고 스코어를 갖는 것으로 결정하는 것을 포함하는, 검색 쿼리에 대한 결과들을 제공하는 시스템.
- [0453] 111. 제어 회로부에 의해 실행될 때 방법을 수행하는 명령어들이 인코딩된 비일시적 컴퓨터-판독가능 매체로서, 상기 방법은,
- [0454] 복수의 엔티티들에 대한 참조들을 포함하는 검색 쿼리를 수신하는 단계;
- [0455] 제어 회로부를 이용하여, 상기 검색 쿼리에 대해 대명사 분석을 수행하는 단계;
- [0456] 상기 제어 회로부를 이용하여, 복수의 예지들에 의해 커플링된 복수의 노드들을 포함하는 그래프를 이용하여 상기 검색 쿼리를 프로세싱하는 단계로서, 상기 복수의 노드들의 서브세트는 상기 검색 쿼리의 분석된 대명사들과 연관된 상기 복수의 엔티티들을 나타내는, 상기 검색 쿼리를 프로세싱하는 단계;
- [0457] 상기 제어 회로부를 이용하여, 근접 메트릭에 기초하여 상기 서브세트의 2개의 노드들에 커플링되는 상기 그래프의 노드를 식별하는 단계로서, 상기 근접 메트릭은 상기 노드와 상기 2개의 노드들 중 하나 사이의 거리에 반 비례하는, 상기 식별하는 단계; 및
- [0458] 상기 그래프의 상기 노드에 의해 표현되는 엔티티에 대한 참조를 상기 검색 쿼리에 대한 결과로서 제공되게 하는 단계를 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0459] 112. 항목 111에 있어서,
- [0460] 상기 복수의 엔티티들은 제 1 복수의 엔티티들이고, 상기 그래프를 이용하여 상기 검색 쿼리를 프로세싱하는 단계는 상기 제 1 복수의 엔티티들을 상기 복수의 노드들에 의해 표현되는 제 2 복수의 엔티티들에 매칭하는 단계를 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0461] 113. 항목 111 또는 112에 있어서,
- [0462] 대화의 사용자 톤을 반영하는 복수의 데이터 세트들을 선택하는 단계를 더 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0463] 114. 항목 113에 있어서,
- [0464] 품사(POS) 태깅을 이용하여 상기 복수의 데이터 세트들을 프로세싱하는 단계를 더 포함하는, 비일시적 컴퓨터-판독가능 매체.
- [0465] 115. 항목 111 내지 114 중 어느 하나의 항목에 있어서,
- [0466] 상기 근접 메트릭에 기초하여 상기 서브세트의 2개의 노드들에 커플링된 상기 그래프의 상기 노드를 식별하는 단계는,
- [0467] 상기 개개 노드와 상기 복수의 노드들의 다른 노드 사이의 거리에 기초하여 상기 그래프의 상기 복수의 노드들의 개개 노드를 스코어링하는 단계; 및
- [0468] 상기 노드가 상기 그래프의 상기 스코어링된 복수의 노드들의 최고 스코어를 갖는 것으로 결정하는 단계를 포함하는, 비일시적 컴퓨터-판독가능 매체.

100

Home

movie where a person falls in love with his operating system

Prod ▾ Dev ▾ China ▾ Datatest ▾ Nuance ▾ English (US) ▾

🔍

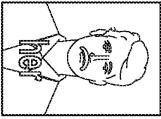
🔊 RPR=1

is_debug=1&XPID=pkg

Conversation

it does not appear to be available anywhere.

Answers



Her (2013) 

A lonely writer falls for his computer's operating system in this quirky and endearing...

Spike Jonze

WIKI3685907 | Relevance = 5349 | Release: 2013-10-13

104

Input Interpretations

Detailed URL | Report Issue | Teach me Boss

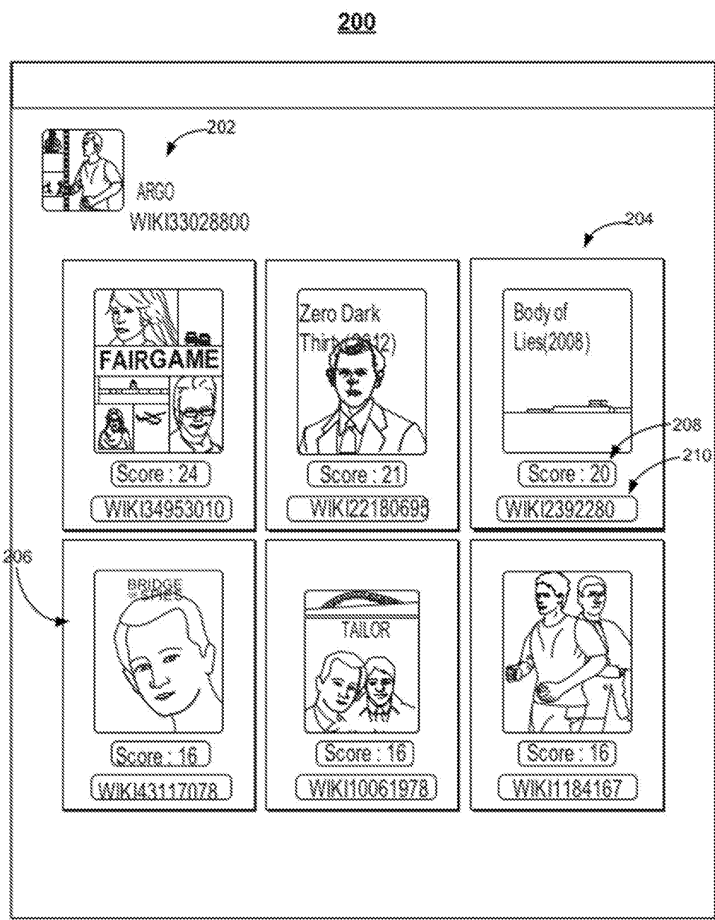
Intent: movie#na; Action: {}

love	Romance - genre - WIKI285141
Good Keyword1	
DF	300168
negation_flag	0
operating system	
Good Keyword1	
DF	19072
negation_flag	0
movie	
movies - filter - FL1	

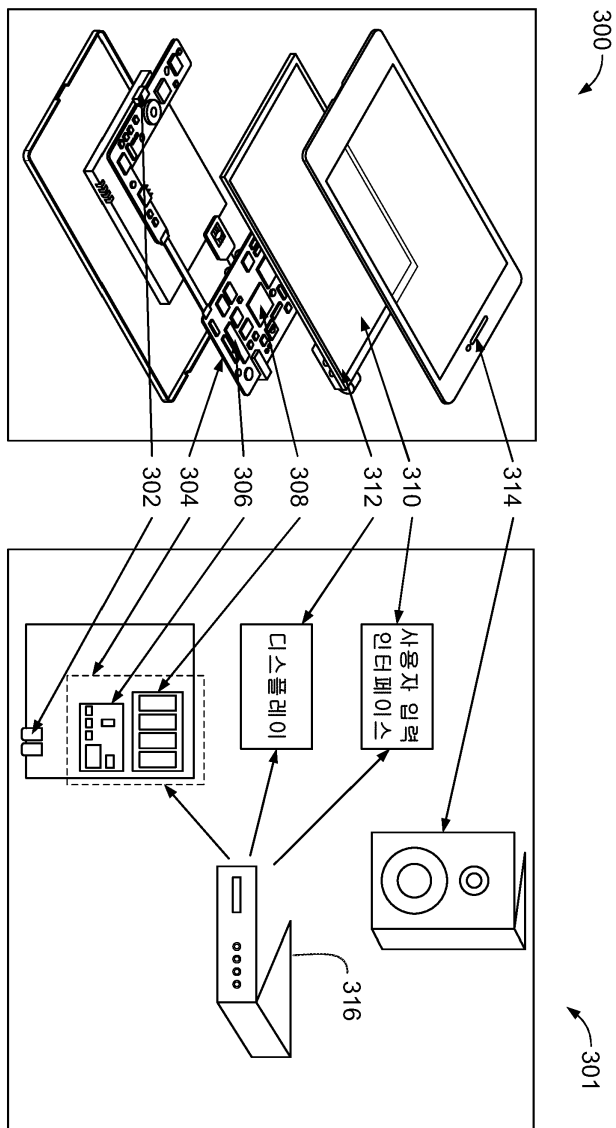
도면

도면1

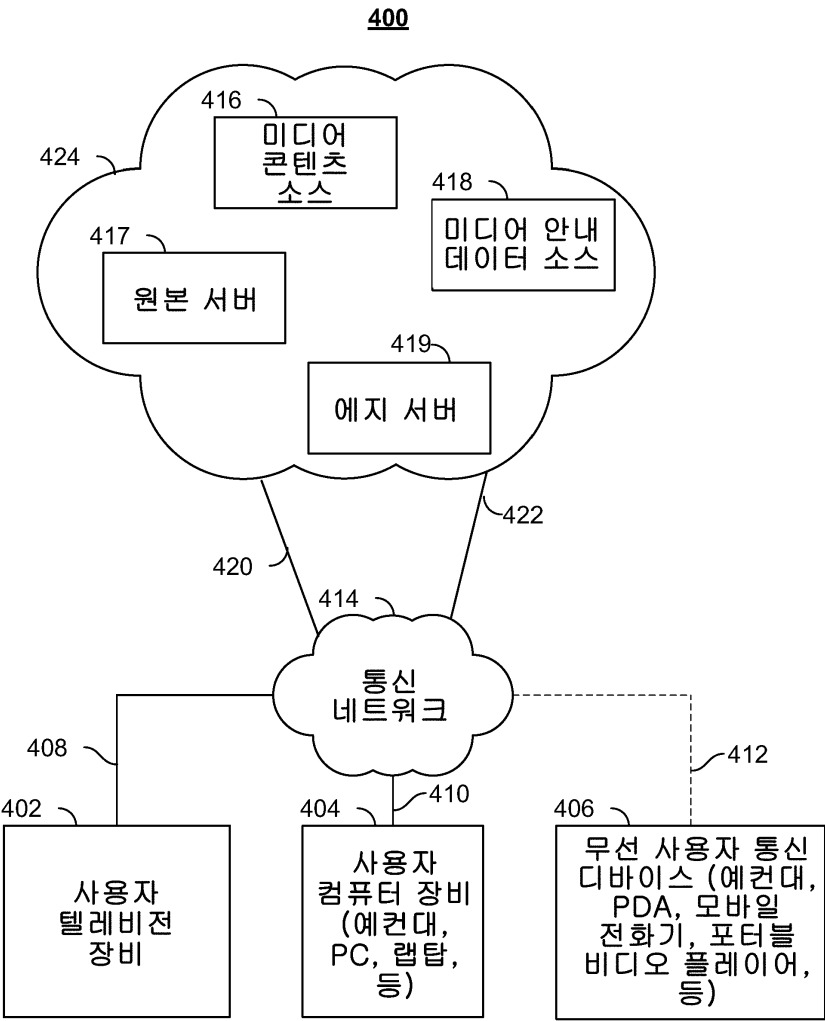
도면2



도면3



도면4

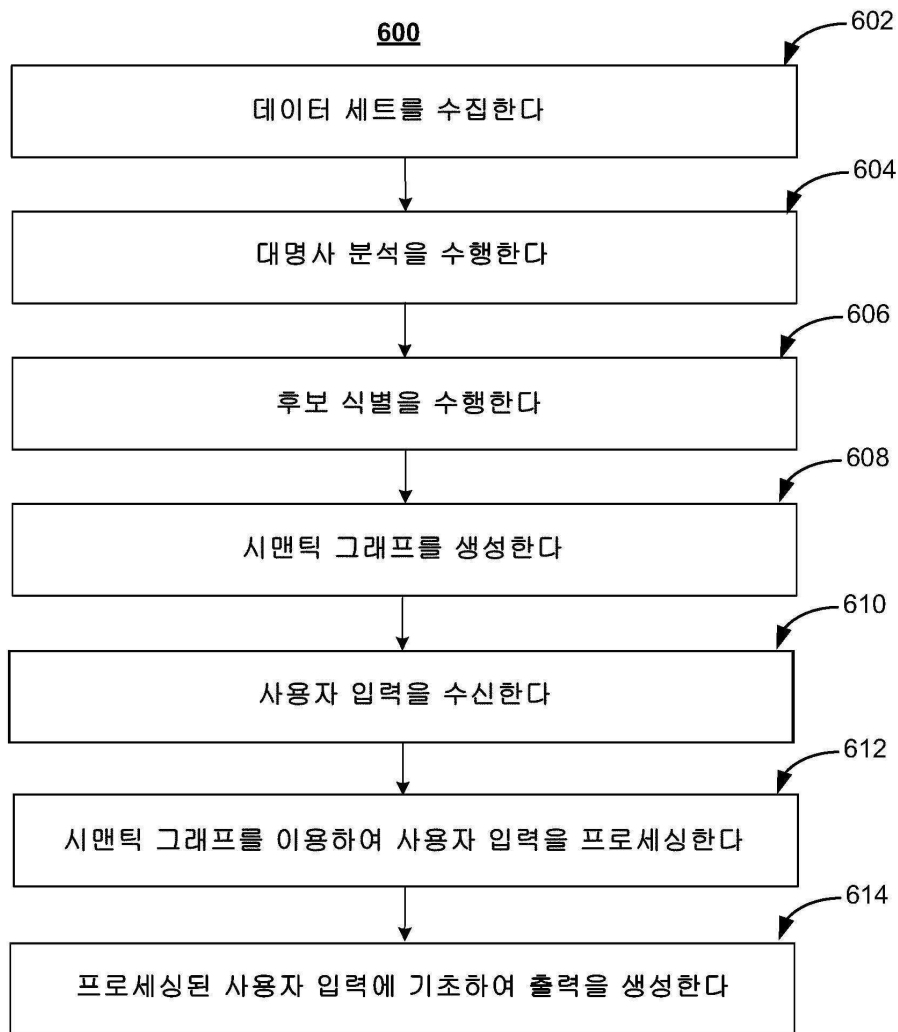


도면5

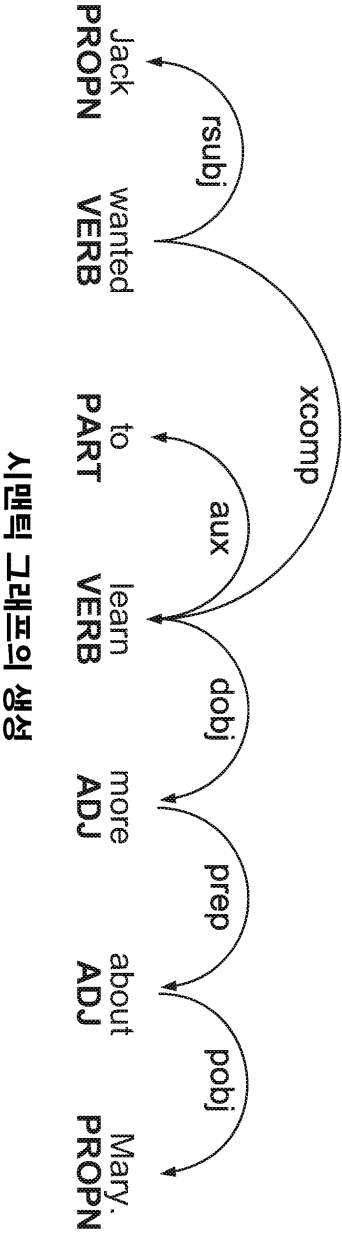
500

특징들	정밀도	재현율	F1-스코어
그래프 특징들 사용 안함	0.611	0.887	0.723
그래프 특징들 사용함	0.879	0.82	0.848

도면6

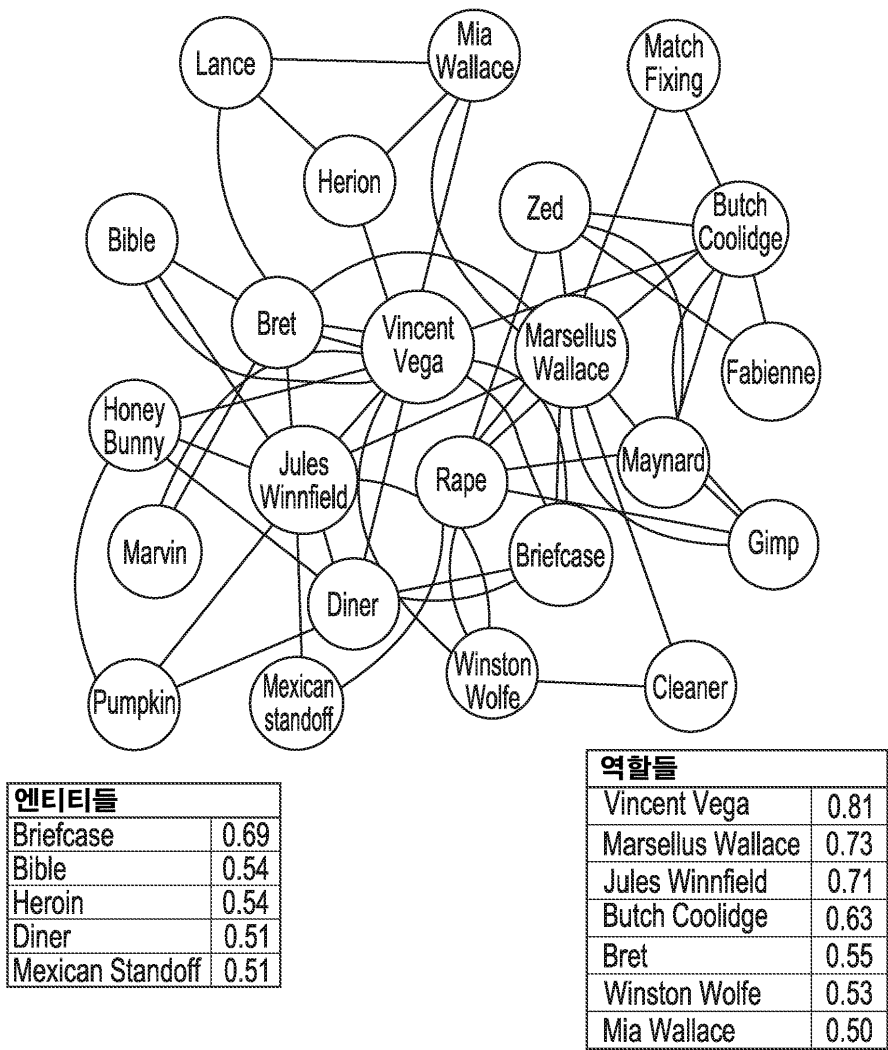


도면7

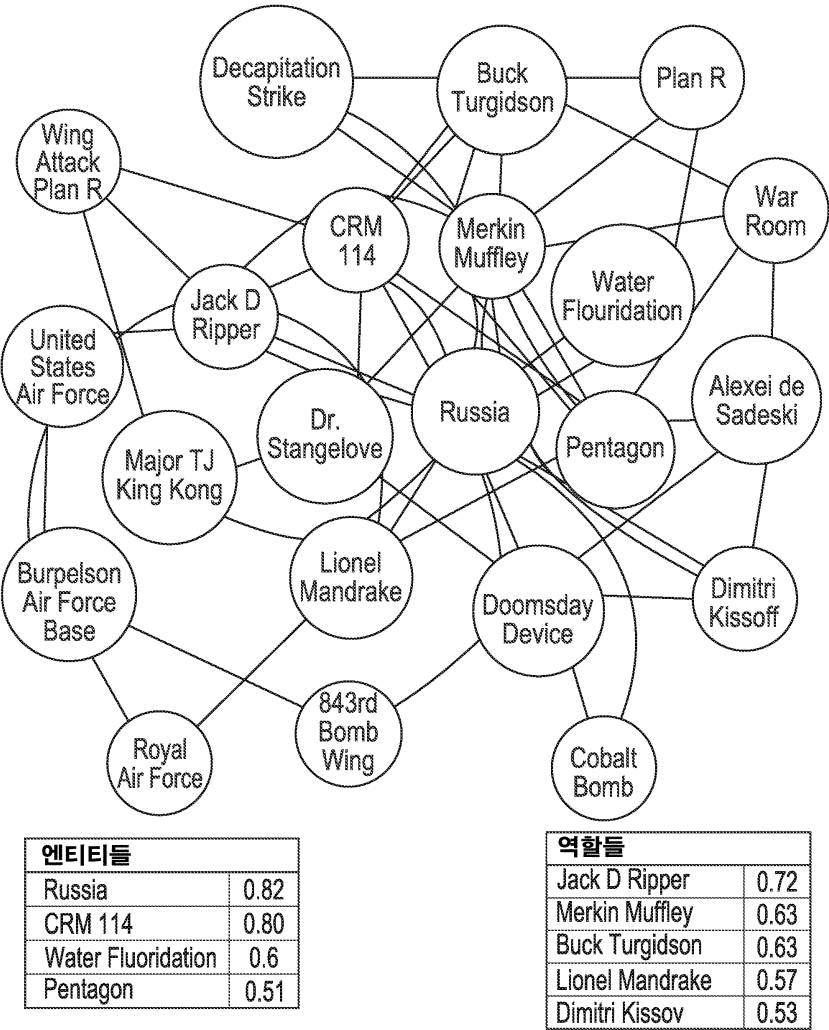


700

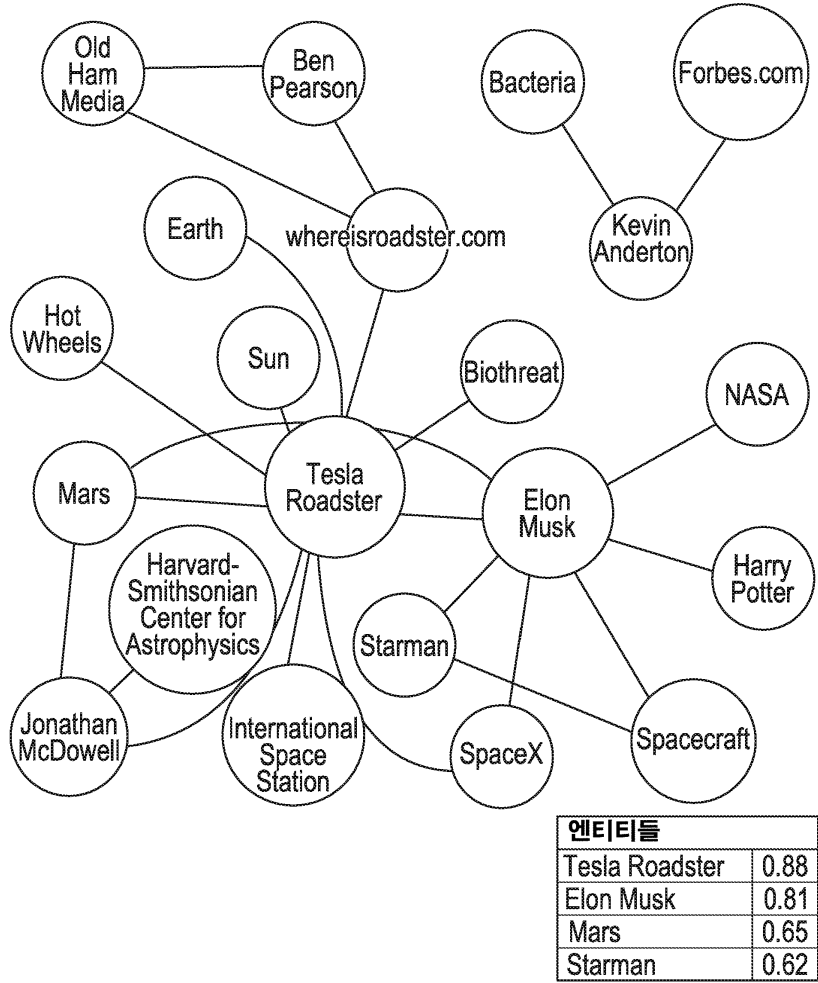
도면8



도면9



도면10



1100

Home

movie where they kill hitler in a cinema theatre

Prod Dev China Datatest Nuance English (US) RPR=1 is_debug=1&XPID=plg Conversation

Ingolourous Bastards is available tonight at 07:30 PM EDT on SHOWTIME. Would you like to record it?

Answers

Ingolourous Bastards (2009)

This thrilling war epic is about an American military officer who leads a special squa..

Quentin Tarantino

WIKI633062 | Relevance = 7957 | Release: 2009-5-20

SHOHD-SE (2018-07-03 23:30) SHOHD-DVR

Input Interpretations

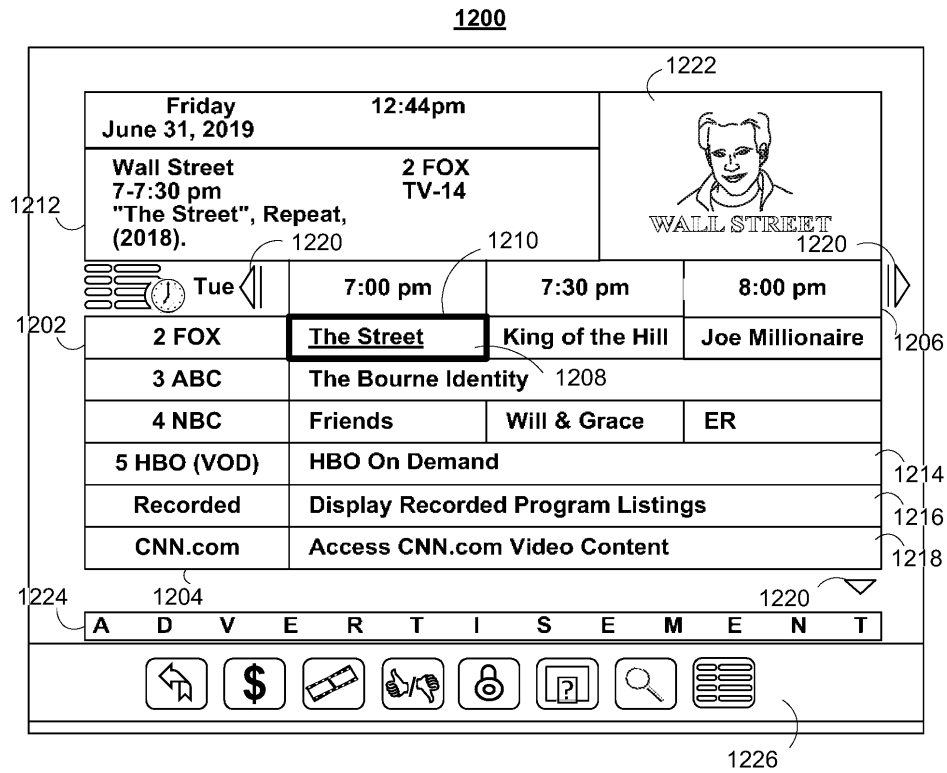
Detailed URL | Report Issue | Teach me Boss

Intent: movie#keyword; Action: {}

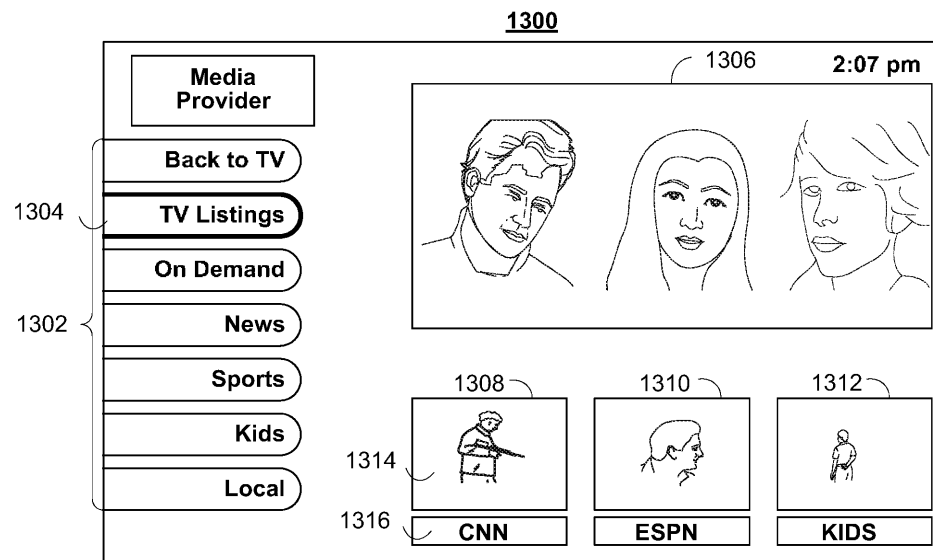
hitler	Adolf Hitler - person - WIKI2731583
Good_Keyword1	
DF	17938
negation_flag	0

cinema	
Good_Keyword1	
DF	83040
negation_flag	0

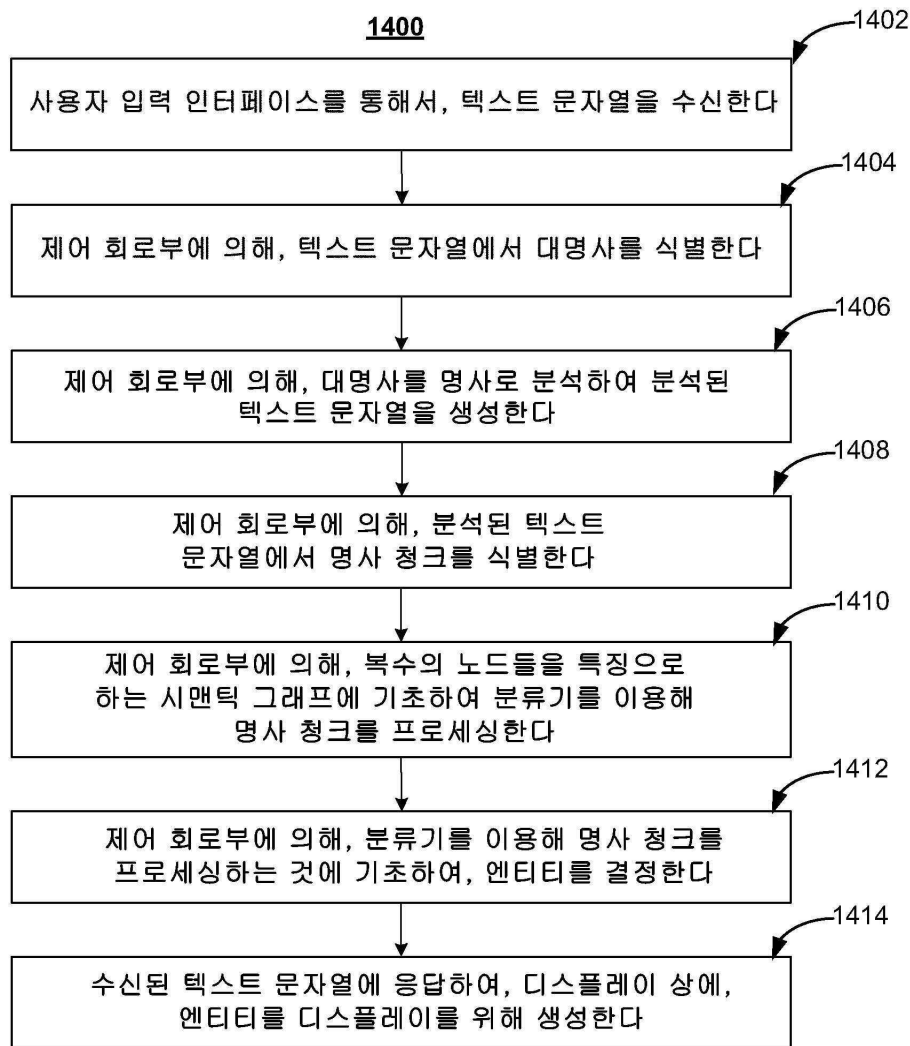
도면12



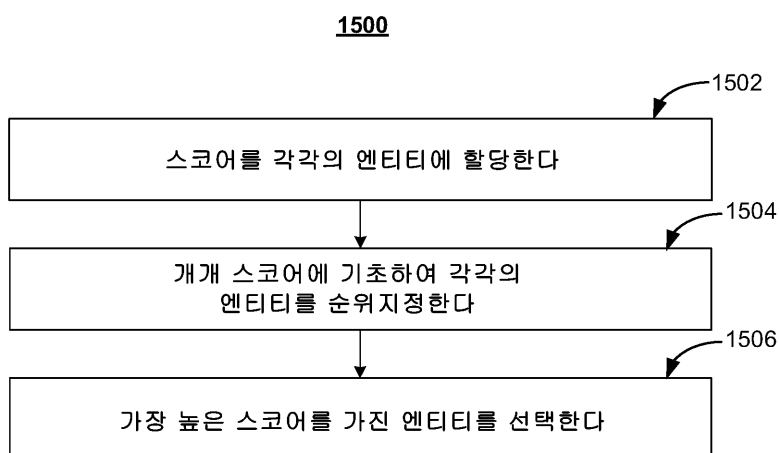
도면13



도면14



도면15



도면16

