



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0032147

(43) 공개일자 2015년03월25일

(51) 국제특허분류(Int. Cl.)

G06F 17/00 (2006.01) G06F 17/20 (2006.01)

(21) 출원번호 10-2014-0008526

(22) 출원일자 2014년01월23일

심사청구일자 2014년01월23일

(30) 우선권주장

1020130111355 2013년09월16일 대한민국(KR)

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

이정현

서울특별시 영등포구 도신로4길 14, 120동 1905호 (대림동, 현대1차아파트)

이상근

서울특별시 강남구 선릉로 221, 205동 1104호 (도곡동, 도곡테크아파트)

장원준

경기도 안산시 단원구 선부로 166, 128동 502호 (선부동, 공작한양아파트)

(74) 대리인

특허법인엠에이피에스

전체 청구항 수 : 총 17 항

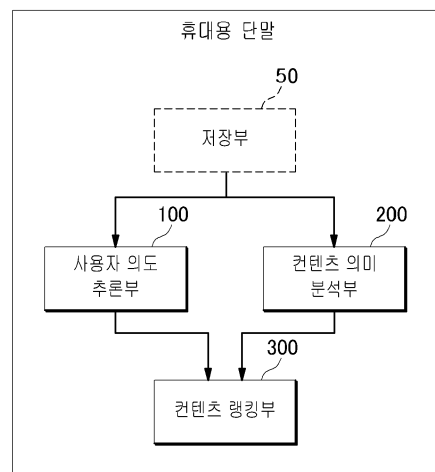
(54) 발명의 명칭 사용자 의도 추론에 기반한 휴대용 단말 장치 및 이를 이용한 콘텐츠 추천 방법

(57) 요약

휴대용 단말은, 특정 카테고리를 나타내는 명사 및 명사와 연관된 동사를 포함하는 주제 분류 트리가 저장되고, 주제 분류 트리를 이용하여, 휴대용 단말에 저장된 텍스트 데이터를 분석하고 사용자 의도를 명사 및 동사 형태로 추론하며, 주제 분류 트리를 이용하여, 인터넷에서 수집한 콘텐츠를 분석하고 콘텐츠 의미를 명사 및 동사 형태로 추론하고, 사용자 의도 및 콘텐츠 의미 사이의 유사도를 계산하여 가장 유사도가 높은 콘텐츠를 랭킹하고 추천한다.

대표도 - 도1

10



특허청구의 범위

청구항 1

휴대용 단말에 있어서,

특정 카테고리를 나타내는 명사 및 상기 명사와 연관된 동사를 포함하는 주제 분류 트리가 저장된 저장부;

상기 주제 분류 트리를 이용하여, 상기 휴대용 단말에 저장된 텍스트 데이터를 분석하고 사용자의 의도를 명사 및 동사 형태로 추론하는 사용자 의도 추론부;

상기 주제 분류 트리를 이용하여, 인터넷에서 수집한 콘텐츠를 분석하고 상기 콘텐츠의 의미를 명사 및 동사 형태로 추론하는 콘텐츠 의미 분석부; 및

상기 사용자의 의도 및 상기 콘텐츠의 의미 사이의 유사도를 계산하여 가장 유사도가 높은 콘텐츠를 랭킹하고 추천하는 콘텐츠 랭킹부를 포함하는 휴대용 단말.

청구항 2

제 1 항에 있어서,

상기 명사 및 동사 형태는 명사와 동사가 쌍을 이루는 형태인 휴대용 단말.

청구항 3

제 1 항에 있어서,

상기 사용자 의도 추론부는,

상기 휴대용 단말에 저장된 데이터로부터 텍스트 데이터를 추출하는 사용자 데이터 추출부;

추출한 텍스트 데이터를 단어로 구분하여 단어 벡터를 생성하는 사용자 데이터 분석부; 및

상기 주제 분류 트리로부터 상기 단어 벡터와 연관성이 높은 카테고리를 나타내는 명사 및 행동을 나타내는 동사를 선택하는 사용자 데이터 분류부를 포함하되,

상기 사용자 데이터 분석부는,

추출한 텍스트 데이터를 단어로 구분하고, 상기 추출한 텍스트 데이터가 포함하는 각 단어의 빈도수를 기반으로 각 단어의 가중치 값을 계산하여 단어 벡터를 생성하는 휴대용 단말.

청구항 4

제 3 항에 있어서,

상기 사용자 데이터 분석부는,

상기 휴대용 단말로부터 추출된 텍스트 데이터가 포함하는 단어의 빈도수, 학습될 때 사용된 전체 문서 수, 및 학습될 때 사용된 전체 문서 중에서 단어가 발생한 문서의 빈도수에 기반하여 상기 단어의 가중치 값을 계산하는 휴대용 단말.

청구항 5

제 3 항에 있어서,

상기 사용자 데이터 분류부는,

상기 주제 분류 트리의 카테고리에 있는 중심 벡터와 상기 사용자 데이터 분석부에서 생성한 단어 벡터 간 유사도를 계산하여 유사도가 가장 높은 카테고리에 해당하는 명사를 선택하고,

선택된 명사와 연관된 동사 및 상기 단어 벡터 간의 연관성을 계산하여 연관성이 가장 높은 동사를 선택하는 휴대용 단말.

청구항 6

제 5 항에 있어서,

상기 사용자 데이터 분류부는,

상기 유사도를 계산하는 경우,

상기 주제 분류 트리의 학습 데이터를 구성하는 단어 집합, 상기 사용자 데이터 분석부에서 생성한 단어 벡터가 포함하는 단어의 가중치 값, 및 상기 중심 벡터가 포함하는 단어의 가중치 값을 기반으로 계산하고,

상기 연관성을 계산하는 경우,

상기 사용자 데이터 분석부에서 생성한 단어 벡터가 포함하는 동사의 가중치 값, 및 상기 동사와 상기 중심벡터 간의 유사도를 기반으로 계산하거나, 또는

상기 사용자 데이터 분석부에서 생성한 단어 벡터, 및 상기 주제 분류 트리가 생성되는 과정에서 해당 카테고리 중심 벡터에 연관된 동사가 추출될 때 사용된 학습 데이터의 단어 벡터 간의 유사도를 기반으로 계산하는 휴대용 단말.

청구항 7

제 1 항에 있어서,

상기 콘텐츠 의미 분석부는,

인터넷에서 콘텐츠를 수집하여 텍스트 데이터를 추출하는 콘텐츠 수집부;

추출한 텍스트 데이터를 단어로 구분하여 단어 벡터를 생성하는 콘텐츠 분석부; 및

상기 주제 분류 트리로부터 상기 단어 벡터와 연관성이 높은 카테고리를 나타내는 명사 및 행동을 나타내는 동사를 선택하는 콘텐츠 분류부를 포함하되,

상기 콘텐츠 분석부는,

추출한 텍스트 데이터를 단어로 구분하고, 상기 콘텐츠가 포함하는 각 단어의 빈도수를 기반으로 각 단어의 가중치 값을 계산하여 단어 벡터를 생성하는 휴대용 단말.

청구항 8

제 7 항에 있어서,

상기 콘텐츠 분석부는,

상기 콘텐츠로부터 추출된 텍스트 데이터가 포함하는 단어의 빈도수, 학습될 때 사용된 전체 문서 수, 및 학습될 때 사용된 전체 문서 중에서 단어가 발생한 문서의 빈도수에 기반하여 상기 단어의 가중치 값을 계산하는 휴대용 단말.

청구항 9

제 7 항에 있어서,

상기 콘텐츠 분류부는,

상기 주제 분류 트리의 카테고리에 있는 중심 벡터와 상기 콘텐츠 분석부에서 생성한 단어 벡터 간 유사도를 계산하여 유사도가 가장 높은 카테고리에 해당하는 명사를 선택하고,

선택된 명사와 연관된 동사 및 상기 단어 벡터 간의 연관성을 계산하여 연관성이 가장 높은 동사를 선택하는 휴대용 단말.

청구항 10

제 9 항에 있어서,

상기 콘텐츠 분류부는,

상기 유사도를 계산하는 경우,

상기 주제 분류 트리의 학습 데이터를 구성하는 단어 집합, 상기 컨텐츠 분석부에서 생성한 단어 벡터가 포함하는 단어의 가중치 값, 및 상기 중심 벡터가 포함하는 단어의 가중치 값을 기반으로 계산하고,

상기 연관성을 계산하는 경우,

상기 컨텐츠 분석부에서 생성한 단어 벡터가 포함하는 동사의 가중치 값, 및 상기 동사와 상기 중심벡터 간의 유사도를 기반으로 계산하거나, 또는

상기 컨텐츠 분석부에서 생성한 단어 벡터, 및 상기 주제 분류 트리가 생성되는 과정에서 해당 카테고리 중심 벡터에 연관된 동사가 추출될 때 사용된 학습 데이터의 단어 벡터 간의 유사도를 기반으로 계산하는 휴대용 단말.

청구항 11

제 1 항에 있어서,

상기 컨텐츠 랭킹부는,

상기 사용자 의도 추론부에서 추론된 명사 및 상기 컨텐츠 의미 분석부에서 추론된 명사 간의 유사도를 계산하고,

상기 사용자 의도 추론부에서 추론된 동사 및 상기 컨텐츠 의미 분석부에서 추론된 동사 간의 유사도를 계산하고,

상기 명사 간의 유사도 및 상기 동사 간의 유사도를 합산하여 컨텐츠를 랭킹하고 추천하는 휴대용 단말.

청구항 12

제 11 항에 있어서,

상기 컨텐츠 랭킹부는,

상기 동사 간의 유사도를 계산하는 경우,

상기 사용자 의도 추론부로부터 선택된 명사, 및 상기 컨텐츠 의미 분석부로부터 선택된 명사가 동일한 경우, 기 계산된 동사 간의 유사도를 상기 동사 간의 유사도로서 산출하는 휴대용 단말.

청구항 13

제 11 항에 있어서,

상기 컨텐츠 랭킹부는,

상기 명사 간 유사도 및 상기 동사 간 유사도의 선형 조합에 대한 가중치 값을 이용하여 상기 명사 간의 유사도 및 상기 동사 간의 유사도를 합산하는 휴대용 단말.

청구항 14

휴대용 단말을 통해 컨텐츠를 추천하는 방법에 있어서,

상기 휴대용 단말에 저장된 텍스트 데이터를 분석하고 사용자의 의도를 명사 및 동사 형태로 추론하는 단계;

인터넷에서 수집한 컨텐츠를 분석하고 상기 컨텐츠의 의미를 명사 및 동사 형태로 추론하는 단계; 및

상기 사용자의 의도 및 상기 컨텐츠의 의미 사이의 유사도를 계산하여 컨텐츠를 랭킹하고 추천하는 단계를 포함하는 컨텐츠 추천 방법.

청구항 15

제 14 항에 있어서,

상기 사용자 의도를 명사 및 동사 형태로 추론하는 단계는,

상기 휴대용 단말에 저장된 데이터로부터 텍스트 데이터를 추출하는 단계;

추출한 텍스트 데이터를 단어로 구분하고, 상기 추출한 텍스트 데이터가 포함하는 각 단어의 빈도수를 기반으로 각 단어의 가중치 값을 계산하여 단어 벡터를 생성하는 단계; 및

기 저장된 명사 및 동사 형태의 주제 분류 트리로부터 상기 단어 벡터와 연관성이 높은 카테고리를 나타내는 명사 및 행동을 나타내는 동사를 선택하는 단계를 포함하는 콘텐츠 추천 방법.

청구항 16

제 14 항에 있어서,

상기 콘텐츠 의미를 명사 및 동사 형태로 추론하는 단계는,

인터넷에서 콘텐츠를 수집하여 텍스트 데이터를 추출하는 단계;

추출한 텍스트 데이터를 단어로 구분하고, 상기 콘텐츠가 포함하는 각 단어의 빈도수를 기반으로 각 단어의 가중치 값을 계산하여 단어 벡터를 생성하는 단계; 및

기 저장된 명사 및 동사 형태의 주제 분류 트리로부터 상기 단어 벡터와 연관성이 높은 카테고리를 나타내는 명사 및 행동을 나타내는 동사를 선택하는 단계를 포함하는 콘텐츠 추천 방법.

청구항 17

제 14 항에 있어서,

상기 콘텐츠를 랭킹하고 추천하는 단계는,

상기 사용자 의도로서 추론된 명사 및 상기 콘텐츠 의미로서 추론된 명사 간의 유사도, 및 상기 사용자 의도로서 추론된 동사 및 상기 콘텐츠 의미로서 추론된 동사 간의 유사도를 계산하는 단계; 및

상기 명사 간의 유사도 및 상기 동사 간의 유사도를 합산하여 콘텐츠를 랭킹하고 추천하는 단계를 포함하는 콘텐츠 추천 방법.

명세서

기술 분야

[0001]

본 발명은 명사 및 동사형태로 사용자 의도를 추론하여 콘텐츠를 추천하는 휴대용 단말 및 방법에 관한 것으로서, 보다 상세하게는 휴대용 단말 또는 인터넷으로부터 수집한 데이터를 명사 및 동사 형태로 분석하여 사용자의 의도에 부합하는 디지털 콘텐츠를 랭킹하고 추천하는 휴대용 단말 및 방법에 관한 것이다.

배경 기술

[0002]

스마트폰의 보급으로 인하여 이동통신 기술이 급격하게 발달함에 따라, 휴대용 단말을 통한 다양한 서비스가 제공되고 있다. 특히 정보 폭주(Information Explosion) 시대에 사용자의 경험과 서비스의 효과를 극대화하기 위하여, 사용자의 관심 분야를 모바일 기기상에서 자동으로 파악하여 개인화된 콘텐츠를 추천하는 서비스가 이용되고 있는데, 해당 서비스에 사용되는 종래 기술로서, 1) 사용자 단말에서 자동으로 사용자의 의도를 추론하는 기술 및 2) 추론된 사용자의 의도를 기반으로 사용자 단말에서 콘텐츠를 추천하는 기술이 있어왔다.

[0003]

사용자 단말에서 자동으로 사용자의 의도를 추론하는 기술로서, 한국 등록특허공보 제10-1095069호(발명의 명칭: 사용자 관심 주제를 추출하는 휴대용 통신 단말기 및 그 방법)는 사용자 단말 상에서 사용자의 관심 사항을 자동으로 파악하기 위한 기술에 대하여 개시하고 있다. 더욱 상세하게는 단말기 내부의 사용자의 데이터(SMS, Web Usage 등)를 추출하고, 공개된 디렉토리 데이터를 가공하여 생성된 주제 분류 트리, 미리 학습된 분류기를 사용하여 사용자의 데이터를 자동 분류함으로써, 사용자의 관심사항을 주제 분류 트리 내의 명사 형태의 카테고리 명으로 추론하는 기술에 대하여 기재하고 있다.

[0004]

이때, 주제 분류 트리란, 모든 웹 페이지의 주제를 트리 형태로 구성한 것으로서, 각 주제들은 "is-a" 관계를 가지고 연결되는데, 모든 주제들을 정의하고 그들의 관계를 연결하기 위해선 각 분야의 다수의 전문가들의 장기간에 걸친 노력이 필요하다. 따라서, 생성된 주제 분류 트리의 주제들 중 웹 페이지와 광고의 내용에 가장 적합

한 주제를 자동으로 결정하기 위해선 사람이 손수 정의해 놓은 많은 양의 학습 데이터가 필요하다.

[0005] 또한, 추론된 사용자의 의도를 기반으로 사용자 단말에서 콘텐츠를 추천하는 기술로서, 한국 공개특허공보 제 10-2012-0026682호(발명의 명칭: 사용자 단말에서 인터넷 서비스 제공 방법 및 장치)는 주제 분류 트리 내의 명사 형태의 카테고리 명으로 추론된 사용자의 의도가 주어졌을 때, 사용자 의도에 연관된 콘텐츠를 자동으로 추천하기 위한 기술에 대하여 개시하고 있다. 더욱 상세하게는, 후보 콘텐츠를 웹으로부터 수집하고, 공개된 디렉토리 데이터를 가공하여 생성된 주제 분류 트리, 미리 학습된 분류기를 사용하여 콘텐츠 데이터를 자동 분류함으로써, 콘텐츠의 의미를 주제 분류 트리 내의 명사 형태의 카테고리 명으로 파악한 후, 사용자의 의도와 콘텐츠 의미 사이의 의미론적 유사도를 계산하여, 사용자의 의도와 의미가 유사한 정도에 따라 콘텐츠를 랭킹하여 콘텐츠를 추천하는 기술에 대하여 기재하고 있다.

[0006] 상술한 종래기술에서는 사용자의 의도와 콘텐츠의 의미를 파악함에 있어서, 공개된 디렉토리 데이터를 가공하여 생성된 주제 분류 트리 내의 명사를 이용하고 있다. 예를 들어, 주제 분류 트리 내의 "Vehicle" 이란 카테고리로 사용자의 의도를 추론한 경우, 같은 카테고리로 의미가 추론된 콘텐츠가 사용자에게 추천된다. 하지만, "Vehicle" 에 대한 관심을 가지고 있는 사용자들도 "Vehicle" 을 "Rent" 하려는 사용자와 "Buy" 하려는 사용자로 서로 다른 행위를 수행할 수 있다. 이 경우, 사용자의 의도는 (명사, 동사) 형태인 (Vehicle, Rent) 및 (Vehicle, Buy) 로 추론하는 것이 명사 형태인 Vehicle 로만 추론하는 것보다 더 정확하다. 따라서 이 경우, 해당 사용자들이 원하는 행동에 따라, "Vehicle"에 대한 콘텐츠도 "Rent" 에 관한 콘텐츠와 "Buy"에 대한 콘텐츠를 나누어서 추천해 줄 필요가 있다. 결과적으로 종래 기술은 사용자가 원하는 행동에 대한 의도를 정확하게 추론하지 못하며, 이에 맞는 사용자 의도에 부합하는 콘텐츠를 추천할 수 없는 한계점이 있다.

[0007] 또한, 지식 베이스 확장 기술로서, 한국 공개특허공보 제 10-2013-0084923호(발명의 명칭: 지식 베이스 확장 장치 및 방법)에는 공개된 디렉토리 데이터를 가공하여 생성된 주제 분류 트리의 각 카테고리별로 연관된 데이터를 여러 지식 베이스로부터 수집하여 관련 문서를 생성하고, 관련 문서에서 행동을 표현하는 후보 동사를 추출한 후, 후보 동사들 중에서 각 카테고리와의 연관성이 높은 행위를 나타내는 동사를 선별하여 주제 분류 트리를 확장하는 기술이 개시되어 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 본 발명의 일 실시예에 따른 휴대용 단말은 주제 분류 트리를 이용하여 사용자의 의도를 카테고리화 행위, 즉, 명사 및 동사(일 예로서, 명사와 동사 쌍) 형태로 추론한다.

[0009] 또한, 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 본 발명의 일 실시예에 따른 추천 방법은 명사 및 동사 형태로 추론된 사용자의 의도와 콘텐츠의 의미 사이의 의미론적 유사도를 계산하여, 콘텐츠를 랭킹 및 추천한다.

과제의 해결 수단

[0010] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 1 측면에 따른 휴대용 단말은, 특정 카테고리를 나타내는 명사 및 명사와 연관된 동사를 포함하는 주제 분류 트리가 저장된 저장부; 주제 분류 트리를 이용하여, 휴대용 단말에 저장된 텍스트 데이터를 분석하고 사용자의 의도를 명사 및 동사 형태로 추론하는 사용자 의도 추론부; 주제 분류 트리를 이용하여, 인터넷에서 수집한 콘텐츠를 분석하고 콘텐츠의 의미를 명사 및 동사 형태로 추론하는 콘텐츠 의미 분석부; 및 사용자 의도 및 콘텐츠 의미 사이의 유사도를 계산하여 가장 유사도가 높은 콘텐츠를 랭킹하고 추천하는 콘텐츠 랭킹부를 포함한다.

[0011] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 2 측면에 따른 휴대용 단말을 통해 콘텐츠를 추천하는 방법은, 휴대용 단말에 저장된 텍스트 데이터를 분석하고 사용자의 의도를 명사 및 동사 형태로 추론하는 단계; 인터넷에서 수집한 콘텐츠를 분석하고 콘텐츠의 의미를 명사 및 동사 형태로 추론하는 단계; 및 사용자의 의도 및 콘텐츠의 의미 사이의 유사도를 계산하여 콘텐츠를 랭킹하고 추천하는 단계를 포함한다.

발명의 효과

[0012] 전술한 본 발명의 과제 해결 수단에 의하면 본 발명의 일 실시예에 따른 휴대용 단말은 사용자 데이터로부터 사

용자의 의도를 명사 및 동사 형태로 추론함으로써, 사용자의 의도를 명사 형태로만 추론하는 종래 기술과 달리 사용자가 원하는 행동까지도 파악하여 보다 정확하게 사용자 의도에 따른 콘텐츠를 제공할 수 있다.

[0013]

또한, 본 발명의 일 실시예에 따른 콘텐츠 추천 방법은 인터넷으로부터 콘텐츠를 수집하여 콘텐츠의 의미를 명사 및 동사형태로 파악한 후, 명사 및 동사 형태로 추론된 사용자의 의도와 의미론적 유사도를 계산하여 콘텐츠를 랭킹 및 추천함으로써, 명사 형태만을 이용하여 의미론적 유사도를 계산하여 콘텐츠를 랭킹 및 추천하는 종래 기술과 달리 사용자의 의도와 콘텐츠를 의미적으로 더 정확히 연관시켜 추천할 수 있다.

도면의 간단한 설명

[0014]

도 1은 본 발명의 일 실시예에 따른 휴대용 단말의 블록도를 나타낸 도면이다.

도 2는 저장부에 저장된 명사 및 동사형태의 주제 분류 트리에 대한 도면이다.

도 3은 본 발명의 일 실시예에 따른 사용자 의도 추론부의 블록도이다.

도 4는 본 발명의 일 실시예에 따른 콘텐츠 의미 분석부의 블록도이다.

도 5는 본 발명의 일 실시예에 따른 콘텐츠 랭킹부의 블록도이다.

도 6은 본 발명의 일 실시예에 따른 콘텐츠 추천 방법에 대한 순서도이다.

도 7은 저장된 명사 및 동사 형태의 주제 분류 트리에 기반하여, 단어 벡터와 연관성이 높은 카테고리를 나타내는 명사 및 행동을 나타내는 동사를 선택하는 단계에 대한 도면이다.

도 8은 본 발명의 실시예에 따른 휴대용 단말의 프로토타입 어플리케이션인 EPE Demo의 실행 화면을 나타낸다

도 9는 실험을 통해 측정된 EPE Demo 어플리케이션의 명사 및 동사 형태의 사용자 의도 추론 성능의 측정결과를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0015]

아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0016]

명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0017]

명세서 전체에서, "명사 및 동사"는 후술할 주제 분류 트리의 카테고리(Category)를 나타내는 명사와 카테고리 와 연관된 행동(Action)을 나타내는 동사를 모두 포함하는 것으로 일 예로서, 명사와 동사 쌍, 즉, (명사, 동사)와 같은 형태를 가질 수 있다.

[0018]

도 1은 본 발명의 일 실시예에 따른 휴대용 단말의 블록도를 나타낸 도면이다.

[0019]

휴대용 단말(10)은 사용자의 의도에 적합한 콘텐츠를 랭킹하여 추천해 주는 단말로서, 네트워크를 통해 공개된 디렉토리, 인터넷에 접속할 수 있는 휴대성과 이동성이 보장되는 무선 통신 장치이고, 통신 모듈, 프로세서, 디스플레이 등을 기본적으로 포함하며, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), WiBro(Wireless Broadband Internet) 단말, 스마트 폰(Smart Phone), 태블릿PC, 슬레이트PC 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다.

[0020]

휴대용 단말(10)에 저장된 텍스트 데이터 및 인터넷에서 수집된 콘텐츠를 분석하여 사용자 의도와 연관된 콘텐츠를 랭킹하고 추천하기 위해, 본 발명의 일 실시예에 따른 휴대용 단말(10)은 공개된 디렉토리로부터 수집된 데이터가 가공된 하나 이상의 중심 벡터가 트리 형태로 저장된 주제 분류 트리가 저장된 저장부(50), 휴대용 단

말에 저장된 텍스트 데이터를 분석하여 사용자 의도를 명사 및 동사 형태로 추론하는 사용자 의도 추론부(100), 인터넷에서 수집된 콘텐츠를 분석하여 콘텐츠 의미를 명사 및 동사 형태로 추론하는 콘텐츠 의미 분석부(200), 및 사용자 의도 및 상기 콘텐츠 의미 사이의 유사도를 계산하여 콘텐츠를 랭킹하고 추천하는 콘텐츠 랭킹부(300)를 포함한다. 이때, "명사 및 동사" 형태는 일 예로서, (명사, 동사) 형태, 즉 명사와 동사 쌍 형태를 가질 수 있되, 동사는 명사에 대하여 사용자가 수행하고자 하는 행위를 나타내며, 명사는 동사와 의미적 연관성을 갖는다. 일 예로서, 명사와 동사 쌍의 명사는 동사의 목적어 역할을 수행한다.

[0021] 따라서, 휴대용 단말은 특정 카테고리를 나타내는 명사 및 해당 명사와 연관된 행동을 나타내는 동사를 포함하는 주제 분류 트리를 이용하여, 사용자 의도를 추론하고, 디지털 콘텐츠를 추천한다.

[0022] 우선, 저장부(50)에는 하나 이상의 중심 벡터가 카테고리화 되어 저장되어 있다. 즉, 저장부(50)에는 명사 및 동사 형태의 주제 분류 트리가 저장되어 있다.

[0023] 도 2는 본 발명에 적용되는 명사 및 동사형태의 주제 분류 트리에 대한 도면이다.

[0024] 명사 및 동사 형태의 주제 분류 트리는 공개된 디렉토리 데이터(Open Directory Project)로부터 가공된 명사 형태의 카테고리화 및 각 카테고리화 연관된 하나 이상의 동사를 포함한다. 각 카테고리화 연관된 하나 이상의 동사는 여러 지식 베이스로부터 추출함으로써 생성된다.

[0025] 도 2를 참조하면, 명사 및 동사 형태의 주제 분류 트리는 스마트 단말 내부(sdcard)에 파일시스템이나 데이터베이스(일례로 SQLite) 등의 형태로 기 저장될 수 있다. 또한 명사는 일 예로서 공개된 디렉토리(ODP: Open Directory Project) 사이트(www.dmoz.com)에서 수집될 수 있고, 동사는 Microsoft의 검색 엔진 클릭 로그 데이터, ODP 카테고리 및 웹 사이트 데이터, Wikipedia, 검색 엔진 검색 결과 페이지 등의 외부 지식 베이스들로부터 수집될 수 있다.

[0026] 이때, 저장부(50)는 캐쉬, ROM(Read Only Memory), PROM(Programmable ROM), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM) 및 플래쉬 메모리(Flash memory)와 같은 비휘발성 메모리 소자 또는 RAM(Random Access Memory)과 같은 휘발성 메모리 소자 또는 하드디스크 드라이브(HDD, Hard Disk Drive), CD-ROM과 같은 저장 매체 중 적어도 하나로 구현될 수 있으나 이에 한정되지는 않는다.

[0027] 또한, 도 2는 주제 분류 트리의 일 예를 표현한 도면으로, 카테고리를 나타내는 명사는 계층 구조로 구성되어 있으며, 각 카테고리화에는 연관된 동사들이 존재한다. 도 2의 예제에서는 "Top>Shopping>Vehicle" 카테고리화 "Buy", "Rent", "Drive"의 세 가지 연관된 동사가 존재한다.

[0028] 이때, 후술할 사용자 의도 추론부(100) 및 콘텐츠 의미 분석부(200)는 저장부(50)에 저장된 데이터를 이용하여 사용자 의도 및 콘텐츠 의미를 추론할 수 있다. 사용자 의도 및 콘텐츠 의미를 추론하는 방법은 도 3 내지 도 5와 관련하여 후술한다.

[0029] 도 3은 본 발명의 일 실시예에 따른 사용자 의도 추론부의 블록도이다.

[0030] 사용자 의도 추론부(100)는 주제 분류 트리를 이용하여, 휴대용 단말에 저장된 텍스트 데이터를 분석하고 사용자 의도를 명사 및 동사 형태로 추론하는 구성으로서, 휴대용 단말에 저장된 데이터로부터 텍스트 데이터를 추출하는 사용자 데이터 추출부(110), 추출한 텍스트 데이터를 단어로 구분하여 단어 벡터를 생성하는 사용자 데이터 분석부(120), 및 생성된 단어 벡터와 연관성이 높은 주제 분류 트리의 카테고리 및 동사를 선택하는 사용자 데이터 분류부(130)를 포함할 수 있다. 이하, 각 구성요소에 대해 설명한다.

[0031] 먼저, 사용자 데이터 추출부(110)는 문자 메시지, 웹 기록, 일정 등의 휴대용 단말 내부의 사용자 데이터의 텍스트 데이터를 추출하는 역할을 수행한다.

[0032] 이어서, 사용자 데이터 분석부(120)는 추출된 텍스트 데이터를 분석하여 벡터 스페이스 모델(Vector Space Model)에 기반한 단어 벡터를 생성하는 역할을 수행한다.

[0033] 사용자 데이터 분석부(120)는 추출된 텍스트 데이터를 개별적인 단어로 구분하고, 각 단어들이 나타나는 빈도수, 즉 텍스트 데이터가 포함하는 각 단어의 빈도수를 기반으로 각 단어의 가중치 값을 계산하여 단어 벡터를 생성한다.

[0034] 보다 상세하게는, 사용자 데이터 분석부(120) 추출된 텍스트 데이터를 개별적인 단어로 구분하고, 각 단어의 가중치 값을 계산 함으로써 단어 벡터를 생성한다. 단어 벡터의 각 단어에 대한 가중치 값은 아래 수학적 1을 이용하여 계산할 수 있다.

수학식 1

$$w_t = (1 + \log tf_t) \times (\log \frac{N}{df_t})$$

[0035]

[0036]

이때, t 는 단어, W_t 는 단어 t 의 가중치 값, tf_t 는 추출된 텍스트 데이터가 포함하는 단어 t 의 빈도수, N 은 사용자 데이터 분류부(130)가 학습될 때 사용된 전체 문서 수, df_t 는 사용자 데이터 분류부(130)가 학습될 때 사용된 전체 문서 중에서 단어 t 가 발생한 문서의 빈도수를 의미한다. 여기서 학습 데이터는 공개 디렉토리에서 각 카테고리마다 분류된 페이지(사이트)의 제목(title), 설명(description) 및 페이지 본문을 포함한다. 마지막으로, 사용자 데이터 분류부(130)는 추출된 사용자 데이터로부터 생성된 단어 벡터와 명사 및 동사 형태로 구성된 주제 분류 트리의 학습 데이터로부터 생성된 중심 벡터를 사용하여, 사용자의 의도를 주제 분류 트리의 카테고리 및 카테고리들과 연관된 하나의 행동 즉, 카테고리 및 행동 형태로 분류한다.

[0037]

다시 말하면, 사용자 데이터 분류부(130)는 명사 및 동사 형태로 구성된 주제 분류 트리의 카테고리에 있는 중심 벡터와 사용자 데이터 분석부(120)에서 생성한 단어 벡터 간 유사도를 계산하여 유사도가 가장 높은 중심 벡터를 지니는 카테고리(명사)를 먼저 선택한다. 이후, 선택된 해당 카테고리들과 연관된 동사들 중에서 상기 단어 벡터와 연관성이 가장 높은 행동을 나타내는 동사를 하나 선택한다. 이때, 단어 벡터와 중심 벡터 간의 유사도는 수학식 2를 이용하여 계산될 수 있고, 카테고리들과 연관된 동사들 중에서 상기 단어 벡터와의 연관성은 수학식 3 또는 수학식 4를 이용하여 계산될 수 있다.

수학식 2

$$\cos(u, c) = \frac{\vec{u} \cdot \vec{c}}{|\vec{u}| |\vec{c}|} = \frac{\sum_{i \in V} u_i \times c_i}{\sqrt{\sum_{i \in V} u_i^2} \sqrt{\sum_{i \in V} c_i^2}}$$

[0038]

[0039]

이때, $\cos(u, c)$ 는 단어 벡터($\vec{u}$)와 카테고리의 중심 벡터($\vec{c}$)의 유사도, $\vec{u}$ 는 사용자 데이터 분석부(120)에서 생성한 단어 벡터, $\vec{c}$ 는 명사 및 동사 형태로 구성된 주제 분류 트리의 카테고리에 있는 데이터 (예를 들어, 웹 페이지의 제목, 설명 및 페이지 본문)로부터 구성된 중심 벡터, V 는 명사 및 동사 형태로 구성된 주제 분류 트리의 학습 데이터를 구성하는 단어 집합, u_i 는 단어 벡터 $\vec{u}$ 가 포함하는 i 번째 단어의 가중치 값, c_i 는 중심 벡터($\vec{c}$)가 포함하는 i 번째 단어의 가중치 값을 의미한다.

수학식 3

$$r(u, v) = \begin{cases} u_v \times cv(v \in u) \\ 0 & (v \notin u) \end{cases}$$

이때, $r(u, v)$ 는 단어 벡터 $\vec{u}$ 와 동사 v 의 연관성, u_v 는 단어 벡터 $\vec{u}$ 가 포함하는 동사 v 의 가중치 값, cv 는 동사 v 와 주제 분류 트리의 카테고리의 중심 벡터 $\vec{c}$ 의 유사도로서 명사 및 동사 형태의 주제 분류 트리를 생성할 때 계산된다.

수학식 4

$$r(u, v) = \cos(\vec{u}, \vec{cd_v})$$

이때, $r(u, v)$ 는 단어 벡터 $\vec{u}$ 와 동사 v 의 연관성, $\vec{u}$ 는 사용자 데이터 분석부(120)에서 생성한 단어 벡터, $\vec{cd_v}$ 는 동사 v 가 명사 및 동사형태의 주제 분류 트리가 생성되는 과정에서 해당 카테고리 중심 벡터 $\vec{c}$ 에 연관된 동사로 추출될 때 사용된 학습 데이터의 단어 벡터, $\cos(\vec{u}, \vec{cd_v})$ 는 단어 벡터($\vec{u}$)와 학습 데이터의 단어 벡터($\vec{cd_v}$) 간의 유사도를 의미한다.

다시 도 1로 돌아가면, 휴대용 단말(10)는 인터넷에서 수집된 콘텐츠를 분석하여 콘텐츠 의미를 명사 및 동사 형태로 추론하는 콘텐츠 의미 분석부(200)를 포함한다.

도 4는 본 발명의 일 실시예에 따른 콘텐츠 의미 분석부의 블록도이다.

콘텐츠 의미 분석부(200)는 주제 분류 트리를 이용하여, 인터넷에서 수집한 콘텐츠를 분석하고 콘텐츠 의미를 명사 및 동사 형태로 추론하는 구성으로서, 인터넷에서 콘텐츠를 수집하여 텍스트 데이터를 추출하는 콘텐츠 수집부(210), 추출한 텍스트 데이터를 단어로 구분하여 단어 벡터를 생성하는 콘텐츠 분석부(220), 및 명사 및 동사 형태의 주제 분류 트리를 사용하여 콘텐츠 수집부에서 생성된 단어 벡터와 연관성이 높은 주제 분류 트리의 카테고리를 나타내는 명사와 동사를 선택하는 콘텐츠 분류부(230)를 포함한다.

먼저, 콘텐츠 수집부(210)는 인터넷으로부터 뉴스, App 정보 등의 콘텐츠를 수집하여, 텍스트 데이터를 추출하는 역할을 수행한다.

이어서, 콘텐츠 분석부(220)는 콘텐츠 수집부(210)로부터 추출한 텍스트 데이터를 단어로 구분하여 단어 벡터를 생성한다. 즉, 앞서 설명한 사용자 데이터 분석부와 같이, 추출한 텍스트 데이터를 단어로 구분하고, 텍스트 데이터가 포함하는 각 단어의 빈도수를 기반으로 각 단어의 가중치 값을 계산함으로써 단어 벡터를 생성한다. 이때, 단어의 가중치 값은 사용자 데이터 분석부와 마찬가지로 수학식 1을 이용하여 계산될 수 있고 수학식 1에 대한 설명은 상술한 바 이하 생략한다. 단, tf_t 는 콘텐츠로부터 추출된 텍스트 데이터가 포함하는 단어 t 의 빈도수, N 은 콘텐츠 분류부(230)가 학습될 때 사용된 전체 문서 수, df_t 는 콘텐츠 분류부(230)가 학습될 때 사

용된 전체 문서 중에서 단어 t 가 발생한 문서의 빈도수를 의미한다.

[0049] 마지막으로, 콘텐츠 분류부(230)는 추출된 콘텐츠 데이터로부터 생성된 단어 벡터와 명사 및 동사 형태로 구성된 주제 분류 트리의 학습 데이터로부터 생성된 중심 벡터를 사용하여, 콘텐츠의 의미를 주제 분류 트리의 카테고리리를 나타내는 명사 및 카테고리리와 연관된 행동을 나타내는 하나의 동사로 분류한다. 즉, 콘텐츠 분류부(230)는 명사 및 동사(일 예로서, 명사와 동사 쌍 또는 (명사, 동사)) 형태로 구성된 주제 분류 트리의 카테고리에 있는 중심 벡터와 콘텐츠 분석부(220)에서 추출한 단어 벡터 간 유사도를 계산하여 유사도가 가장 높은 중심 벡터를 지니는 카테고리를 먼저 선택한다. 그런 다음, 선택된 해당 카테고리리와 연관된 동사들 중에서 상기 단어 벡터와의 연관성이 가장 높은 동사를 하나 선택한다. 이때, 중심 벡터 및 단어 벡터 간의 유사도는 수학적 2를 이용하여 계산될 수 있고, 카테고리리와 연관된 동사들 중에서 상기 단어 벡터와의 연관성은 수학적 3 또는 수학적 4를 이용하여 계산될 수 있다. 수학적 2, 수학적 3, 및 수학적 4에 대한 설명은 상술한 바 이하

생략한다. 단, u 는 콘텐츠 분석부(220)에서 생성한 단어 벡터를 의미한다.

[0050] 다시 도 1로 돌아가면 본 발명의 일 실시예에 따른 휴대용 단말은 콘텐츠 랭킹부(300)를 포함할 수 있는데, 콘텐츠 랭킹부(300)는 사용자 의도 및 콘텐츠 의미 사이의 유사도를 계산하여 콘텐츠를 랭킹하고 추천하는 역할을 수행한다.

[0051] 도 5는 본 발명의 일 실시예에 따른 콘텐츠 랭킹부의 블록도이다.

[0052] 콘텐츠 랭킹부(300)는 명사 및 동사 형태로 추론된 사용자 의도와 명사 및 동사 형태로 파악된 콘텐츠 의미 사이의 연관성을 계산하여, 사용자의 의도와 가장 연관성이 높은 콘텐츠를 사용자에게 추천해주는 역할을 수행한다.

[0053] 이때, 콘텐츠란 인터넷에서 제공되는 뉴스, 애플리케이션 정보 등 유무선 통신망을 통해 제공되는 디지털 정보나 그러한 내용물을 의미하는 것으로 관용적인 의미와 동일하다.

[0054] 이러한 역할을 수행하기 위해 본 발명의 일 실시예에 따른 콘텐츠 랭킹부(300)는 사용자 의도 및 콘텐츠 의미의 명사 간 유사도를 계산하는 명사 유사도 계산부(310), 사용자 의도 및 콘텐츠 의미의 동사 간 유사도를 계산하는 동사 유사도 계산부(320), 및 명사 간 유사도 및 동사 간 유사도를 합산하여 콘텐츠를 랭킹하고 추천하는 콘텐츠 추천부(330)를 포함할 수 있다.

[0055] 명사 유사도 계산부(310)는 사용자 의도를 나타내는 것으로 선택된 명사와 콘텐츠 의미를 나타내는 것으로 선택된 명사 간의 유사도를 계산하는 역할을 수행한다. 명사 유사도 계산부는 등록특허공보 제 10-1083274호(발명의 명칭: 유사도 그래프를 활용한 문맥 광고 시스템)에 기재된 유사도 산출부(232)를 사용할 수 있다.

[0056] 이때, 본 발명의 일 실시예에 따른 명사 유사도 계산부는 분류 트리의 각 카테고리마다 할당되어 있는 학습 데이터를 이용하여 산출한다. 벡터 스페이스 모델에 따라 각 카테고리의 학습 데이터에 기초하여, 각 카테고리를 대표하는 중심 벡터를 생성하고, 각 중심 벡터의 코사인 유사도(cosine similarity)를 산출하여, 각 카테고리 간의 유사도로서 사용한다. 이때, 코사인 유사도는 수학적 5를 기초로 산출될 수 있다.

수학적 5

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}.$$

[0057]

[0058] 이어서 동사 유사도 계산부(320)는 사용자 의도 추론부에서 선택된 동사 및 콘텐츠 의미 분석부에서 선택된 동사 간의 유사도를 계산하는 역할을 수행한다. 본 발명의 일 실시예에 따르면 동사 유사도 계산부(320)는 아래 수학적 6과 같이 동사 간 유사도를 계산할 수 있다.

수학식 6

$$ActionScore(v_u, v_s) = \begin{cases} score(v_u, v_s) & (c_u = c_s) \\ 0 & (c_u \neq c_s) \end{cases}$$

[0059]

[0060]

이때, c_u 는 사용자 의도 추론부로부터 선택된 주제 분류 트리의 카테고리를 나타내는 명사, c_s 는 콘텐츠 의미 분석부로부터 선택된 주제 분류 트리의 카테고리를 나타내는 명사, v_u 는 사용자 의도 추론부로부터 선택된 행동을 나타내는 동사, 및 v_s 는 콘텐츠 의미 분석부로부터 선택된 행동을 나타내는 동사, $ActionScore(v_u, v_s)$ 는 최종 동사(v_u, v_s) 간 유사도, $score(v_u, v_s)$ 는 기 계산된 동사(v_u, v_s) 간 유사도를 의미한다.

[0061]

이때, $score(v_u, v_s)$ 은 예를 들어 1) 여러 지식 베이스로부터 각 동사가 추출될 때 사용된 학습 데이터를 이용하여 유사도를 계산하는 방법과 2) WordNet 과 ConceptNet에서 제공하는 단어 사이의 유사도를 이용하여 계산될 수 있다.

[0062]

이어서, 콘텐츠 추천부(330)는 명사 유사도 계산부(310)에서 계산한 명사 간 유사도 및 동사 유사도 계산부(320)에서 계산한 동사 간 유사도를 합산하여 콘텐츠를 랭킹하고 추천한다. 이때, 명사간 유사도 및 동사 간 유사도는 아래 수학식 7과 같이 합산될 수 있다.

수학식 7

$$NVScore(u, s) = (1 - \beta) GraphScore(c_u, c_s) + (\beta) ActionScore(v_u, v_s)$$

[0063]

[0064]

이때, $NVScore(u, s)$ 는 합산된 사용자 의도 및 콘텐츠 의미의 유사도, $GraphScore(c_u, c_s)$ 는 사용자 의도 추론부로부터 선택된 주제 분류 트리의 카테고리를 나타내는 명사(c_u) 및 콘텐츠 의미 분석부로부터 선택된 주제 분류 트리의 카테고리를 나타내는 명사(c_s) 간 유사도, $ActionScore(v_u, v_s)$ 는 사용자 의도 추론부로부터 선택된 행동을 나타내는 동사(v_u) 및 콘텐츠 의미 분석부로부터 선택된 행동을 나타내는 동사(v_s) 간 유사도, β 는 명사 간 유사도 및 동사 간 유사도의 선형 조합에 대한 가중치 값을 의미한다.

[0065]

한편, 도 1 내지 도 5에서 도시된 각각의 구성요소는 일종의 '모듈'로 구성될 수 있다. 상기 '모듈'은 소프트웨어 또는 Field Programmable Gate Array(FPGA) 또는 주문형 반도체(ASIC, Application Specific Integrated Circuit)와 같은 하드웨어 구성요소를 의미하며, 모듈은 어떤 역할들을 수행한다. 그렇지만 모듈은 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. 모듈은 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 실행시키도록 구성될 수도 있다. 구성요소들과 모듈들에서 제공되는 기능은 더 작은 수의 구성요소들 및 모듈들로 결합되거나 추가적인 구성요소들과 모듈들로 더 분리될 수 있다.

[0066]

도 6은 본 발명의 일 실시예에 따른 콘텐츠 추천 방법에 대한 순서도이다.

[0067]

먼저, 휴대용 단말에는 공개된 디렉토리 데이터가 가공된 하나 이상의 중심 벡터가 주제 분류 트리 형태로 저장되어 있다. 즉, 명사 및 동사 형태의 주제 분류 트리가 기 저장되어 있다. 명사 및 동사 형태의 주제 분류 트리에 대한 상세한 설명은 도 2와 관련하여 상술한 바 이하 생략한다.

[0068]

다음으로 휴대용 단말은 주제 분류 트리를 이용하여, 휴대용 단말에 저장된 텍스트 데이터를 분석하고 사용자 의도를 명사 및 동사(일 예로서, 명사와 동사 쌍, 또는 (명사, 동사)) 형태로 추론한다(S1100). 보다 상세하게는, 휴대용 단말에 저장된 데이터로부터 텍스트 데이터를 추출하고, 추출한 텍스트 데이터를 단어로 구분하여 단어 벡터를 생성하고, 명사 및 동사 형태의 주제 분류 트리를 사용하여 단어 벡터와 연관성이 높은 카테고리를 나타내는 명사 및 행동을 나타내는 동사를 선택한다. 이때, 동사는 카테고리를 나타내는 명사와 연관적인 의미를 갖는 행동을 나타내는 것으로, 일 예로서 명사는 동사의 목적어일 수 있다. 또한, 명사 및 동사는 일 예로서 명사와 동사 쌍 형태를 가질 수 있다.

- [0069] 도 7은 저장된 명사 및 동사 형태의 주제 분류 트리에 기반하여, 단어 벡터와 연관성이 높은 카테고리를 나타내는 명사 및 행동을 나타내는 동사를 선택하는 단계에 대한 도면이다.
- [0070] 먼저 카테고리를 나타내는 명사를 선택하기 위해, 단어 벡터와 연관성이 높은 카테고리를 선택하는데, 도 7에 따르면 단어 벡터(u) 간 유사도가 가장 높은 카테고리를 나타내는 명사(Vehicle)를 선택한다.
- [0071] 다음으로 행동을 나타내는 동사를 선택하기 위해, 선택된 카테고리 및 연관된 하나 이상의 동사 중 단어 벡터와 가장 연관성이 높은 동사를 선택한다. 도 7에 따르면 선택된 카테고리(Vehicle)가 포함하는 하나 이상의 동사(Buy, Rent, Drive) 중 단어 벡터(u)의 동사와 가장 연관성이 높은 동사(Rent)를 선택한다.
- [0072] 더욱 상세한 사용자 의도를 추론하는 단계(S1100)에 대한 설명은 도 3과 관련하여 상술한 바 이하 생략한다.
- [0073] 이어서, 휴대용 단말은 주제 분류 트리를 이용하여 인터넷에서 수집된 콘텐츠를 분석하고, 콘텐츠 의미를 명사 및 동사 형태로 추론한다(S1200). 보다 상세하게는, 휴대용 단말은 인터넷에서 콘텐츠를 수집하여 텍스트 데이터를 추출하고, 추출한 텍스트 데이터를 단어로 구분하여 단어 벡터를 생성하고, 저장된 명사 및 동사 형태의 주제 분류 트리를 사용하여 단어 벡터와 연관성이 높은 카테고리를 나타내는 명사 및 행동을 나타내는 동사를 선택한다. 이 단계에 대한 상세한 설명은 도 4와 관련하여 상술한 바 이하 생략한다.
- [0074] 이어서, 휴대용 단말은 사용자 의도 및 콘텐츠 의미 사이의 유사도를 계산하여 콘텐츠를 랭킹하고 추천한다(S1300). 보다 상세하게는, 휴대용 단말은 사용자 의도 및 콘텐츠 의미의 명사 간 유사도를 계산하고, 사용자 의도 및 콘텐츠 의미의 동사 간 유사도를 계산하고, 명사 간 유사도 및 동사 간 유사도를 합산하여 콘텐츠를 랭킹하고 추천한다. 이 단계에 대한 상세한 설명은 도 5와 관련하여 상술한 바 이하 생략한다.
- [0075] 한편, 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 메커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.
- [0076] 상술한 본 발명에 따른 콘텐츠 추천 방법은 컴퓨터 또는 휴대용 단말이 읽을 수 있는 기록 매체에 컴퓨터 또는 휴대용 단말이 읽을 수 있는 코드로서 구현되는 것이 가능하다. 컴퓨터 또는 휴대용 단말이 읽을 수 있는 기록 매체로는 컴퓨터 또는 휴대용 단말 시스템에 의하여 해독될 수 있는 데이터가 저장된 모든 종류의 기록 매체를 포함한다. 예를 들어, ROM(Read Only Memory), RAM(Random Access Memory), 자기 테이프, 자기 디스크, 플래쉬 메모리, 광 데이터 저장장치 등이 있을 수 있다. 또한, 컴퓨터 또는 휴대용 단말이 읽을 수 있는 기록매체는 통신망으로 연결된 컴퓨터 또는 휴대용 단말 시스템에 분산되어, 분산방식으로 읽을 수 있는 코드로서 저장되고 실행될 수 있다.
- [0077] 도 8은 본 발명의 실시예에 따른 휴대용 단말의 프로토타입 어플리케이션인 EPE Demo의 실행 화면을 나타낸다. EPE Demo 어플리케이션은 윈도우 8 폰 환경에서 개발되었다.
- [0078] 도 8에 따르면, EPE Demo 어플리케이션에서는 휴대용 단말 내에서 추출한 사용자 데이터가 표출되고, 휴대용 단말 내에서 추출된 사용자 데이터로부터 추론한 명사 및 동사 형태의 사용자 의도가 표출된다.
- [0079] 도 9는 실험을 통해 측정된 EPE Demo 어플리케이션의 명사 및 동사 형태의 사용자 의도 추론 성능의 측정결과를 나타낸다.
- [0080] 도 9에 따르면, 15가지의 시나리오(애완동물, 교육, 요리, 건강, 영화, 음악, 야구, 축구, 여행, 뉴스, 요가, 금융, 점술, 게임, 소셜 네트워크)에 대한 테스트 사용자 데이터에 대하여 Top-k 개의 명사 및 동사 형태의 사용자 의도를 추론한 결과의 정확도가 정밀지표(Precision at k의 지표)로 측정되었다. 측정 결과 Precision at 1 은 0.778의 성능을, Precision at 2는 0.711의 성능을, Precision at 3은 0.667의 성능을 보였다.
- [0081] 이와같이 본 발명의 일 실시예에 따른 휴대용 단말 및 방법은 인터넷으로부터 콘텐츠를 수집하여 콘텐츠의 의미를 명사 및 동사형태로 파악한 후, 명사 및 동사 형태로 추론된 사용자의 의도와 의미론적 유사도를 계산하여 콘텐츠를 랭킹 및 추천함으로써, 명사 형태만을 이용하여 의미론적 유사도를 계산하여 콘텐츠를 랭킹 및 추천하

는 종래 기술과 달리 사용자가 원하는 행동과 의미적으로 연관된 콘텐츠를 추천할 수 있다.

[0082]

전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0083]

본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

[0084]

10: 휴대용 단말

50: 저장부

100: 사용자 의도 추론부

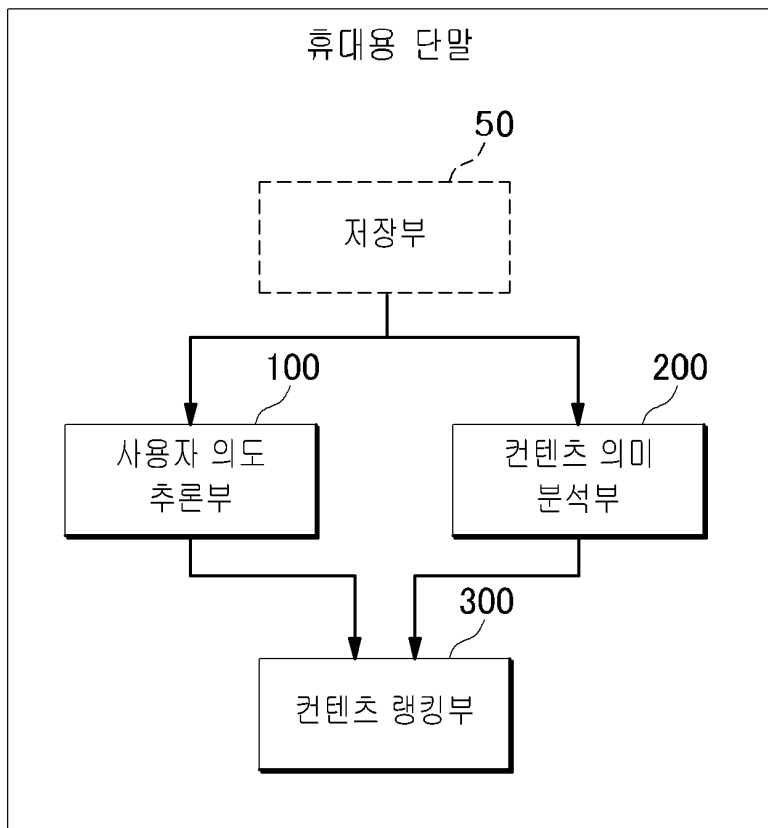
200: 콘텐츠 의미 분석부

300: 콘텐츠 랭킹부

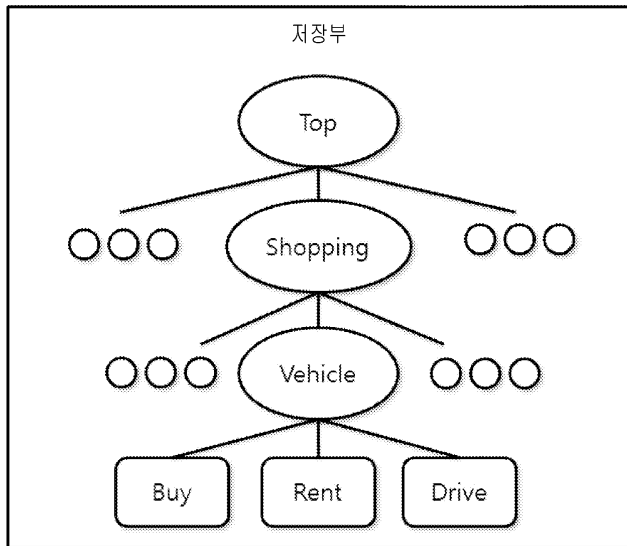
도면

도면1

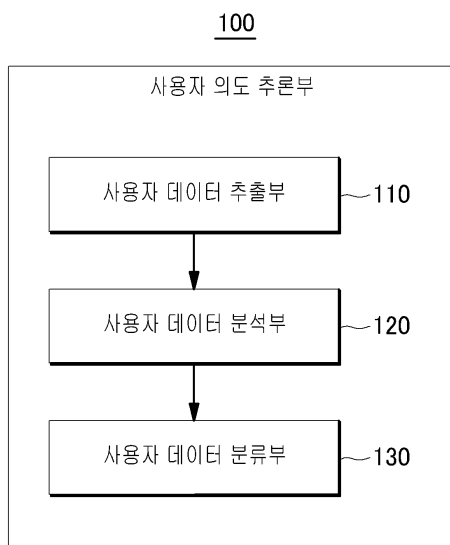
10



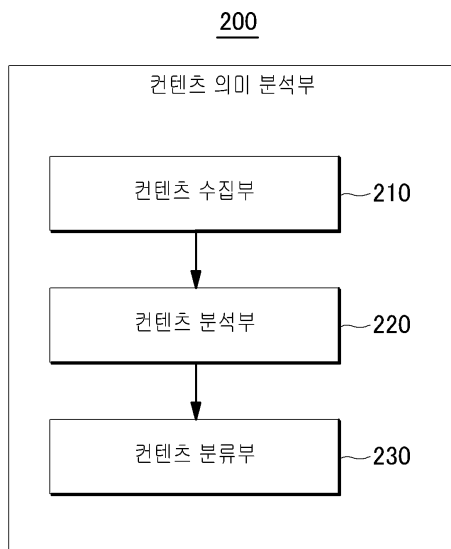
도면2



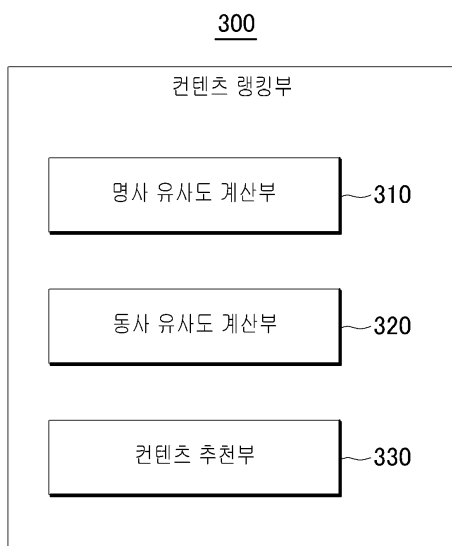
도면3



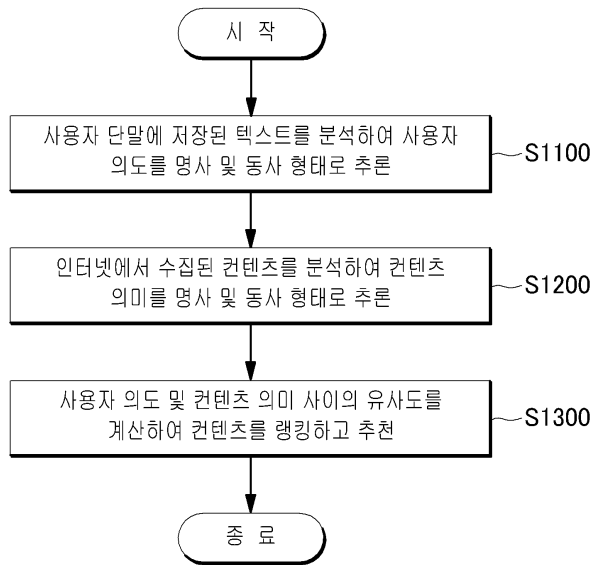
도면4



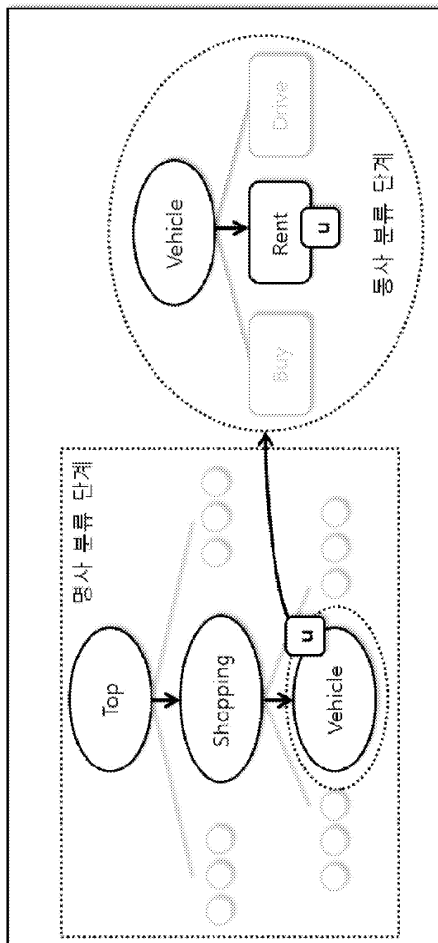
도면5



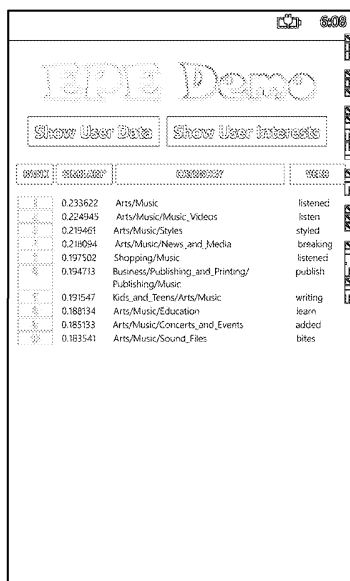
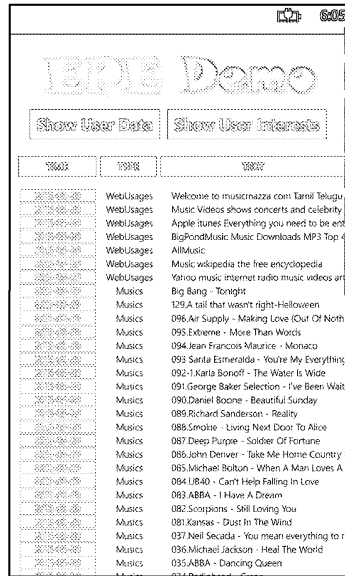
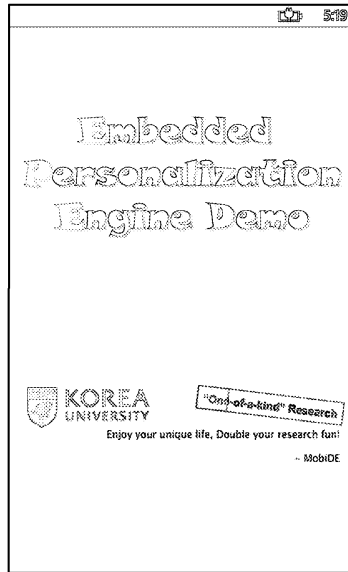
도면6



도면7



도면8



도면9

