



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0054309
(43) 공개일자 2016년05월16일

(51) 국제특허분류(Int. Cl.)
G06F 17/00 (2006.01) G06F 17/30 (2006.01)
(21) 출원번호 10-2014-0153744
(22) 출원일자 2014년11월06일
심사청구일자 2014년11월06일

(71) 출원인
아주대학교산학협력단
경기도 수원시 영통구 월드컵로 206 (원천동)
(72) 발명자
이경원
서울특별시 성북구 길음로 118, 407동 704호 (길
음동, 대림e편한세상아파트)
김기남
서울특별시 성동구 상원길 63, 107동 1301호 (성
수동1가, 쌍용아파트)
(74) 대리인
특허법인충정

전체 청구항 수 : 총 16 항

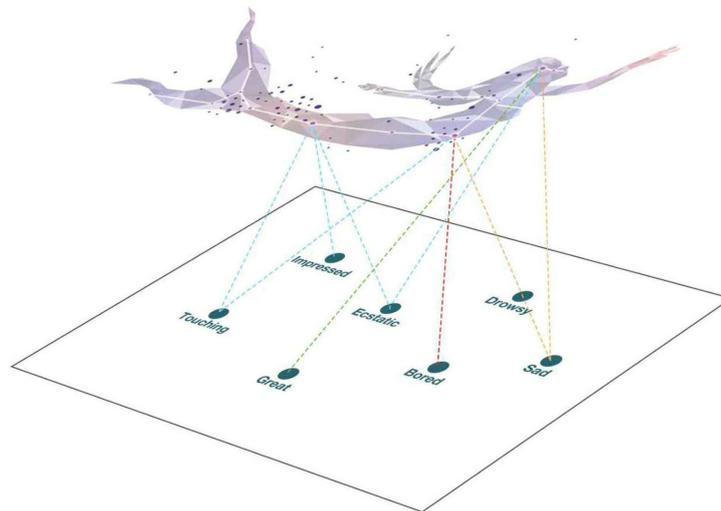
(54) 발명의 명칭 레퍼런스 의미 지도를 이용한 데이터 시각화 방법 및 시스템

(57) 요약

본 발명은 객체에 대한 속성값(노드)들을 레퍼런스 의미 지도와 결합하여 시각화하는 방법 및 시스템에 관한 기술로, 보다 상세하게는 객체에 대한 속성값에 대한 네트워크 그래프와 미리 저장된 레퍼런스 의미 지도를 결합하여, 네트워크를 이루는 속성값의 위치 및 의미 거리 기반으로 객체의 위치를 결정하여 시각화하는 기술에 관한 것이다.

본 발명은 동일한 속성값을 가지고 있는 객체들에 대한 네트워크 그래프와 2차원 의미 지도를 결합하여 레퍼런스 의미 지도 대비 속성값의 위치 및 의미 거리 기반으로 객체들의 위치를 결정하여 시각화 하는 것을 목적으로 한다.

대표도 - 도8



(72) 발명자

하효지

서울시 영등포구 국제금융로7길 1, A동 707호(여의도동 수정아파트)

최한민

경기도 용인시 기흥구 이현로30번길 7, 104동 702호 (보정동, 솔피마을현대홈타운)

황원주

경기도 안양시 동안구 학의로 20, 123동 903호 (비산동, 관악현대아파트)

강한성

서울특별시 도봉구 노해로 216, 403호 (쌍문동)

이 발명을 지원한 국가연구개발사업

과제고유번호 S-2013-A0403-00010

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 인문사회기초연구

연구과제명 상황별 감정어휘 분포맵을 이용한 영화추천 시스템의 시각화(2년차)

기여율 1/1

주관기관 아주대학교

연구기간 2012.05.01 ~ 2014.04.30

명세서

청구범위

청구항 1

객체에 대하여 적어도 하나 이상의 속성값을 추출하는 단계; 및

상기 추출된 속성값을, 상기 추출된 속성값들 간의 의미 거리 기반으로, 미리 설정된 래퍼런스 맵을 기준으로 하여 배치하는 단계;

를 포함하고,

상기 래퍼런스 맵은 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성된 것을 특징으로 하는 데이터 시각화 방법.

청구항 2

객체에 대하여 적어도 하나 이상의 속성값을 추출하는 단계;

상기 추출된 하나 이상의 속성값 각각을 미리 결정된 래퍼런스 맵을 기준으로 하여 배치하는 단계; 및

상기 객체의 위치를 상기 배치된 하나 이상의 속성값 각각과의 가중치에 기반한 거리를 반영하여 결정하는 단계

를 포함하고,

상기 래퍼런스 맵은 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성된 것을 특징으로 하는 데이터 시각화 방법.

청구항 3

제1항에 있어서,

상기 객체로부터 추출된 상기 하나 이상의 속성값들 각각을 하나의 노드에 대응시키고, 상기 객체로부터 추출된 상기 하나 이상의 속성값들 각각에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화하는 단계

를 더 포함하는 데이터 시각화 방법.

청구항 4

제2항에 있어서,

복수의 객체들 각각에 대하여 상기 적어도 하나 이상의 속성값을 추출하는 단계, 상기 추출된 하나 이상의 속성값 각각을 배치하는 단계, 및 상기 상기 객체의 위치를 결정하는 단계를 실행한 후,

상기 복수의 객체들 각각을 하나의 노드에 대응시키고, 상기 복수의 객체들 중 기준 속성값이 공통적으로 추출된 객체들에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화하는 단계

를 더 포함하는 데이터 시각화 방법.

청구항 5

제2항에 있어서,

상기 객체의 위치를 결정하는 단계는

상기 객체에 대하여 상기 적어도 하나 이상의 속성값이 추출되는 빈도수를 계수하는 단계; 및

상기 계수된 빈도수를 반영하여 상기 속성값 각각과의 가중치를 결정하는 단계

를 포함하는 데이터 시각화 방법.

청구항 6

제1항 또는 제2항 중 어느 한 항에 있어서,
 상기 속성값을 추출하는 단계는
 상기 객체에 대한 코멘트 데이터를 수집하는 단계; 및
 상기 수집된 코멘트 데이터에서 상기 적어도 하나 이상의 속성값을 추출하는 단계;
 를 포함하는 것을 특징으로 하는 데이터 시각화 방법.

청구항 7

제6항에 있어서,
 상기 수집된 코멘트 데이터에서 상기 적어도 하나 이상의 속성값을 추출하는 단계는
 상기 속성값을 유니그램 또는 바이그램 방식으로 추출하는 것을 특징으로 하는 데이터 시각화 방법.

청구항 8

제1항 또는 제2항 중 어느 한 항에 있어서,
 상기 레퍼런스 맵은 다차원척도 분석(Multi-Dimensional Scaling: MDS)을 이용하여 형성되는 것을 특징으로 하는 데이터 시각화 방법.

청구항 9

객체[평가 및 의견을 제시할 대상]에 대하여 적어도 하나 이상의 속성값[표현요소, 감정어휘, 표현어휘, 평가어휘, 가치의견 등등을 의미]을 추출하는 속성값 추출부; 및
 상기 추출된 속성값을, 상기 추출된 속성값들 간의 의미 거리 기반으로, 미리 설정된 레퍼런스 맵을 기준으로 하여 배치하는 속성값 배치부;
 를 포함하고,
 상기 레퍼런스 맵은 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성된 것을 특징으로 하는 데이터 시각화 시스템.

청구항 10

객체에 대하여 적어도 하나 이상의 속성값을 추출하는 속성값 추출부;
 상기 추출된 하나 이상의 속성값 각각을 미리 결정된 레퍼런스 맵을 기준으로 하여 배치하는 속성값 배치부; 및
 상기 객체의 위치를 상기 배치된 하나 이상의 속성값 각각과의 가중치에 기반한 거리를 반영하여 결정하는 위치 결정부;
 를 포함하고,
 상기 레퍼런스 맵은 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성된 것을 특징으로 하는 데이터 시각화 시스템.

청구항 11

제9항에 있어서,
 상기 객체로부터 추출된 상기 하나 이상의 속성값들 각각을 하나의 노드에 대응시키고, 상기 객체로부터 추출된 상기 하나 이상의 속성값들 각각에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화하는 네트워크 그래프 생성부;
 를 더 포함하는 데이터 시각화 시스템.

청구항 12

제10항에 있어서,

상기 복수의 객체들 각각을 하나의 노드에 대응시키고, 상기 복수의 객체들 중 기준 속성값이 공통적으로 추출된 객체들에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화하는 네트워크 그래프 생성부; 를 더 포함하는 데이터 시각화 시스템.

청구항 13

제10항에 있어서,
 상기 위치 결정부는,
 상기 객체에 대하여 상기 적어도 하나 이상의 속성값이 추출되는 빈도수를 계수하는 빈도수 계수부; 및
 상기 계수된 빈도수를 반영하여 상기 속성값 각각과의 가중치를 결정하는 가중치 결정부
 를 포함하는 데이터 시각화 시스템.

청구항 14

제9항 또는 제10항 중 어느 한 항에 있어서,
 상기 객체에 대한 코멘트 데이터를 수집하는 코멘트 데이터 수집부;
 를 더 포함하고,
 상기 속성값 추출부는
 상기 수집된 코멘트 데이터에서 상기 적어도 하나 이상의 속성값을 추출하는 것을 특징으로 하는 데이터 시각화 방법.

청구항 15

제14항에 있어서,
 상기 속성값 추출부는,
 상기 속성값을 유니그램 또는 바이그램 방식으로 추출하는 것을 특징으로 하는 데이터 시각화 시스템.

청구항 16

제9항 또는 제10항 중 어느 한 항에 있어서,
 상기 레퍼런스 맵은 다차원척도 분석(Multi-Dimensional Scaling: MDS)을 이용하여 형성되는 것을 특징으로 하는 데이터 시각화 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 객체에 대한 속성값을 시각화하는 데이터 시각화 방법 및 시스템에 관한 기술로, 보다 상세하게는 객체에 대한 속성값을 레퍼런스 의미 지도를 이용하여 시각화하고, 레퍼런스 의미 지도를 이용하여 객체의 의미 지도 상의 위치를 결정하는 기술에 관한 것이다.

[0002] 본 발명은 교육부 및 한국연구재단의 인문사회기초연구사업의 일환으로 수행한 연구로부터 도출된 것이다[과제 관리번호: S-2013-A0403-00010, 과제명: 상황별 감정어휘 분포맵을 이용한 영화추천 시스템의 시각화].

배경 기술

[0003] 일반적으로, 영화, 음악, 문학 작품, 등 콘텐츠를 소비하거나, 상품, 또는 서비스를 이용한 사용자들은 콘텐츠, 상품, 또는 서비스(이하 "객체"라 함)를 이용한 감정 또는 의견을 코멘트(리뷰)형식으로 표현하게 되고, 아직 객체를 이용하지 못한 사용자들이나 객체에 대한 정보를 얻고자 하는 사용자들은 객체를 미리 이용해본 사용자들이 남긴 코멘트(리뷰)를 참고하여 정보를 얻게 된다.

[0004] 사용자는 객체에 대한 정보를 얻고자 하는데, 객체에 대한 코멘트 데이터는 텍스트에 기반하여 구성되어 있으며

로, 사용자가 코멘트 데이터를 참고하여 객체에 대한 정보를 얻는 데 걸리는 시간이 상당히 길게 소요되는 문제점이 있었다. 특히, 객체에 대한 코멘트 데이터의 양이 방대하거나, 다수의 사용자가 코멘트를 남겼거나, 오랜 시간 동안 코멘트 데이터가 누적된 경우에는 사용자가 코멘트 데이터의 내용을 읽는 것만으로도 상당한 노력이 필요하였다.

- [0005] 따라서 이러한 문제점에 대응하고자 코멘트 데이터 상의 어휘를 기반으로 하여 코멘트 또는 객체를 검색하고, 사용자로 하여금 코멘트 및 객체에 대한 탐색 시간을 단축하도록 할 수 있는 기술에 대한 연구가 진행되었다.
- [0006] 이러한 콘텐츠에 대한 코멘트 정보를 이용하여 콘텐츠를 검색하는 방법의 일 예가 한국등록특허 제10-0917784호 "콘텐츠에 대한 코멘트를 기반으로 한 집단 감성 정보 검색방법 및 시스템"에 기술되어 있다.
- [0007] 상기 선행기술은 인터넷 상의 각종 콘텐츠에 달린 코멘트를 수집하여 검색용 데이터베이스(이하 DB라고 한다)를 작성하고 이 검색용DB를 이용하여 감성적인 질의에 대해 객관적이고 신뢰할 수 있는 순위 결과를 보여 주는 검색 방법 및 시스템을 제공하는데 그 목적이 있다. 특히 감성적인 단어가 포함된 질의에 대하여 감성적인 단어가 코멘트 상에 나타나는 빈도를 반영하여 객체의 추천 우선 순위를 조정하는 기술이다.
- [0008] 그러나 위의 선행기술은 객체에 대한 코멘트로부터 감성 단어를 검색해 내는 기술에 대해서는 언급하고 있지만, 객체에 대하여 사용자에게 기대되는 전반적인 감정 또는 의견이 효과적으로 도시되는 것은 아니고, 다수의 코멘트가 존재하는 객체가 우선적으로 추천되는 등의 한계가 있다.
- [0009] 이는 위의 선행기술이 텍스트 기반의 감정단어 검색을 채택하였기 때문에 생기는 한계로서, 하나의 객체(콘텐츠, 상품 또는 서비스)에 대하여 기대되는 전반적인 감정 또는 의견을 효과적으로 보여줄 수 있는 기술의 개발이 요구된다.
- [0010] 한편, 복수의 객체들을 대표하는 노드들을 시각화하는 종래의 네트워크 그래프는 객체를 대표하는 노드(속성값을 가짐)끼리의 연결을 통해 각 노드의 관계를 살펴보고, 유사한 노드끼리 이루는 군집을 바탕으로 사회관계나 현상을 분석하는데 사용된다.
- [0011] 도 1은 대표적인 네트워크 그래프 생성 알고리즘인 포스-디렉티드 알고리즘(Force-Directed Algorithm)을 통해 영화 '레미제라블'에 등장하는 등장인물간의 관계를 네트워크로 나타낸 것이다.
- [0012] 도 1에 도시된 것처럼 좌측 그래프와 우측 그래프는 동일한 데이터를 이용하여 나타낸 그래프임에도 불구하고 서로 다른 형태를 나타내고 있다. 이는 네트워크 그래프가 각 노드간의 연결을 통해 최종 위치가 결정되기 때문에 각 노드와 군집은 그래프가 생성 될 때마다 다른 위치에 나타나게 된다. 이에 따라 데이터가 추가되거나 수정 될 때에 노드의 위치 변화가 많기 때문에 네트워크를 분석함에 있어서 어려움이 발생하게 된다.
- [0013] 따라서 이러한 문제점에 대응하고자 도출된 기술이 같은 노드의 종류가 겹치지 않고 관련성이 있는 것은 근방에 배치하여 네트워크 그래프를 시각화하는 방법이다.
- [0014] 이러한 네트워크 그래프를 시각화하는 방법의 일 예가 일본공개특허 제2014-142900호 "그래프 시각화 표시 장치 및 방법 및 프로그램"에 기술되어 있다.
- [0015] 상기 선행기술은 복수 노드와 노드 사이를 잇는 엣지로 구성되는 그래프 데이터를 가시화하기 위한 그래프 시각화 표시장치 및 방법 및 프로그램을 제공하는데 그 목적이 있다.
- [0016] 그러나 위의 선행기술에서도 데이터가 추가되거나 수정될 때에 노드의 위치 변화에 따른 기준이 없어서 노드들의 위치가 기존의 위치에서 많이 벗어나기 때문에 네트워크 그래프를 분석함에 있어서 문제점이 있다.
- [0017] 따라서, 객체에 대한 데이터가 추가되거나 수정되어도 서로 연관성이 있는 특성을 가지는 노드는 기존의 위치와 비슷한 위치에 나타나도록 하여 직관적으로 파악이 가능한 네트워크 그래프를 시각화하는 기술의 개발이 필요하다.

선행기술문헌

특허문헌

- [0018] (특허문헌 0001) 한국등록특허 제10-0917784호 (등록일 2009.09.10)

(특허문헌 0002) 일본공개특허 제2014-142900호 (공개일 2014.08.07)

발명의 내용

해결하려는 과제

- [0019] 본 발명은 상기와 같은 종래 기술의 문제점을 해결하고자 도출된 것으로서, 하나의 객체(콘텐츠 상품 또는 서비스)에 대하여 나타나는 속성값들에 대한 네트워크 그래프와 2차원 의미지도와 결합하여 데이터가 추가되거나 변경되어 새로운 그래프를 생성하여도 연관성이 있는 특성을 가지는, 해당 속성값을 가지는 노드는 기존의 위치와 비슷한 위치에 나타나도록 하여 직관적인 네트워크 분석이 이루어지도록 하는 것을 목적으로 한다.
- [0020] 또한 본 발명은 기존의 객체 정보가 제공하는 제작회사, 가격 등 객관적인 정보뿐만 아니라 사용자가 객체를 이용하고 표현하는 감정 또는 의견을 분석하여 객체를 새로이 이용하려는 사용자에게 객체 선택의 기준으로 삼을 수 있는 정보를 제공하는 것을 목적으로 한다.
- [0021] 본 발명은 하나의 객체에 대하여 감정 또는 의견이 표현된 복수의 요소들을 의미 거리(semantic distance) 기반으로 시각화함으로써, 객체에 대하여 표현된 감정 또는 의견의 전체적인 분포를 직관적으로 시각화할 수 있는 방법 및 시스템을 제공하는 것을 목적으로 한다.
- [0022] 본 발명은 복수의 표현 요소(expression element)들을 상대적인 의미 거리 기반으로 시각화함으로써, 객체에 대하여 표현된 복수의 표현 요소들 간의 상대적인 거리 및 분포를 직관적으로 인식할 수 있는 수단을 제공하는 것을 목적으로 한다.
- [0023] 또한 본 발명은 텍스트에 한정하지 않고, 이모티콘이나 아이콘 등 감정 또는 의견을 표현할 수 있는 다양한 비언어적 요소를 모두 반영하여 의미 거리 기반의 분포를 시각화할 수 있는 수단을 제공하는 것을 목적으로 한다. 또한, 텍스트라는 제약에서 벗어나 자유로울 수 있기 때문에, 다양한 외국어로 표현된 의견 또는 감정까지도 망라하여 하나의 프레임 안에서 시각화할 수 있는 수단을 제공할 수도 있다.
- [0024] 또한 본 발명은 동일한 속성값을 가지고 있는 객체들에 대한 네트워크 그래프와 레퍼런스 의미지도를 결합하여 레퍼런스 의미지도 대비 속성값의 위치 및 의미 거리 기반으로 객체들의 위치를 결정하여 시각화하는 것을 목적으로 한다.

과제의 해결 수단

- [0025] 상기와 같은 목적을 달성하기 위하여, 본 발명의 일 실시예에 따른 데이터 시각화 방법은 객체에 대하여 적어도 하나 이상의 속성값을 추출하는 단계 및 상기 추출된 속성값을, 상기 추출된 속성값들 간의 의미 거리 기반으로, 미리 설정된 레퍼런스 맵을 기준으로 하여 배치하는 단계를 포함한다.
- [0026] 이때, 레퍼런스 맵은 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성된 것을 특징으로 한다.
- [0027] 또한, 객체로부터 추출된 상기 하나 이상의 속성값들 각각을 하나의 노드에 대응시키고, 상기 객체로부터 추출된 상기 하나 이상의 속성값들 각각에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화하는 단계를 더 포함할 수 있다.
- [0028] 본 발명의 일 실시예에 따른 객체의 위치를 결정하여 시각화하는 방법은 객체에 대하여 적어도 하나 이상의 속성값을 추출하는 단계 상기 추출된 하나 이상의 속성값 각각을 미리 결정된 레퍼런스 맵을 기준으로 하여 배치하는 단계 및 상기 객체의 위치를 상기 배치된 하나 이상의 속성값 각각과의 가중치에 기반한 거리를 반영하여 결정하는 단계를 포함한다.
- [0029] 이때, 레퍼런스 맵은 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성된 것을 특징으로 하며, 다차원척도 분석(Multi-Dimensional Scaling: MDS)을 이용하여 형성되는 것을 특징으로 한다.
- [0030] 또한, 복수의 객체들 각각에 대하여 상기 적어도 하나 이상의 속성값을 추출하는 단계, 상기 추출된 하나 이상의 속성값 각각을 배치하는 단계, 및 상기 상기 객체의 위치를 결정하는 단계를 실행한 후, 상기 복수의 객체들 각각을 하나의 노드에 대응시키고, 상기 복수의 객체들 중 기준 속성값이 공통적으로 추출된 객체들에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화하는 단계를 더 포함할 수 있다.

- [0031] 또한, 상기 객체의 위치를 결정하는 단계는 상기 객체에 대하여 상기 적어도 하나 이상의 속성값이 추출되는 빈도수를 계수하는 단계 및 상기 계수된 빈도수를 반영하여 상기 속성값 각각과의 가중치를 결정하는 단계를 포함한다.
- [0032] 또한, 상기 속성값을 추출하는 단계는 상기 객체에 대한 코멘트 데이터를 수집하는 단계 및 상기 수집된 코멘트 데이터에서 상기 적어도 하나 이상의 속성값을 추출하는 단계를 포함하는 것을 특징으로 한다.
- [0033] 이때, 상기 수집된 코멘트 데이터에서 상기 적어도 하나 이상의 속성값을 추출하는 단계는 상기 속성값을 유니그램 또는 바이그램 방식으로 추출하는 것을 특징으로 한다.
- [0034] 본 발명의 일 실시예에 따른 데이터를 시각화 하는 시스템은 객체에 대하여 적어도 하나 이상의 속성값을 추출하는 속성값 추출부 및 상기 추출된 속성값을, 상기 추출된 속성값들 간의 의미 거리 기반으로, 미리 설정된 레퍼런스 맵을 기준으로 하여 배치하는 속성값 배치부를 포함하고, 상기 객체로부터 추출된 상기 하나 이상의 속성값들 각각을 하나의 노드에 대응시키고, 상기 객체로부터 추출된 상기 하나 이상의 속성값들 각각에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화하는 네트워크 그래프 생성부를 포함한다. 또한, 객체에 대한 코멘트 데이터를 수집하는 코멘트 데이터 수집부를 더 포함한다.
- [0035] 본 발명의 일 실시예에 따른 객체의 위치를 결정하여 시각화 하는 시스템은 객체에 대하여 적어도 하나 이상의 속성값을 추출하는 속성값 추출부 상기 추출된 하나 이상의 속성값 각각을 미리 결정된 레퍼런스 맵을 기준으로 하여 배치하는 속성값 배치부 및 상기 객체의 위치를 상기 배치된 하나 이상의 속성값 각각과의 가중치에 기반한 거리를 반영하여 결정하는 위치 결정부를 포함하고, 상기 복수의 객체들 각각을 하나의 노드에 대응시키고, 상기 복수의 객체들 중 기준 속성값이 공통적으로 추출된 객체들에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화하는 네트워크 그래프 생성부를 포함한다.
- [0036] 또한, 위치 결정부는 상기 객체에 대하여 상기 적어도 하나 이상의 속성값이 추출되는 빈도수를 계수하는 빈도수 계수부 및 상기 계수된 빈도수를 반영하여 상기 속성값 각각과의 가중치를 결정하는 가중치 결정부를 포함한다.

발명의 효과

- [0037] 본 발명에 따르면 레퍼런스 의미지도를 기준으로 네트워크 그래프를 재배치하여 네트워크 그래프를 구성하는 속성값이 변화하거나 추가되는 경우 네트워크 그래프의 변화가 적어 직관적으로 파악이 가능한 효과가 있다.
- [0038] 또한, 본 발명은 별도의 프로그램 설치 없이 브라우저상의 웹 페이지를 통해 제공할 수 있으므로 속성값이 갱신될 때마다 개발자가 새로운 데이터 관리나 배포의 절차 없이 실시간으로 분석 결과를 사용자는 제공 받을 수 있다.
- [0039] 또한, 본 발명은 정부 또는 공공기관이 어떠한 정책이나 계획을 발표하여 사람들이 이에 대하여 인터넷을 통해 의사를 표현하는 경우, 정책에 대한 여론의 반응을 직관적으로 확인할 수도 있다.
- [0040] 또한, 인터넷에서 발생하는 기업에서 발생하는 각종 사고 또는 외부에서 발생하는 기업에 대한 여론을 수집하고 이를 분석한 여론 반응의 변화를 실시간으로 파악할 수도 있으며, 기업은 이러한 정보를 이용하여 사내 위기관리 프로토콜을 통해 대응할 수도 있다.
- [0041] 또한, 다수의 객체들에 대하여 감정 또는 의견이 표현된 복수의 요소들을 의미 거리(semantic distance) 기반으로 시각화함으로써, 객체에 대하여 표현된 감정 또는 의견의 전체적인 분포를 직관적으로 사용자에게 제공할 수 있다.
- [0042] 또한, 하나의 객체에 대하여 대표적으로 표현된 감정 또는 의견을 시각화할 수도 있지만, 복수의 표현 요소들을 상대적인 의미 거리 기반으로 시각화함으로써, 객체에 대하여 표현된 복수의 표현 요소들 간의 상대적인 거리 및 분포를 직관적으로 인식할 수 있다.
- [0043] 또한, 텍스트에 한정하지 않고, 이모티콘이나 아이콘 등 감정을 표현할 수 있는 다양한 비언어적 요소를 모두 반영하여 의미 거리 기반의 분포를 시각화할 수 있으며, 텍스트라는 제약에서 벗어나 자유로울 수 있기 때문에, 다양한 외국어로 표현된 의견 또는 감정까지도 망라하여 하나의 프레임 안에서 시각화할 수도 있다.

도면의 간단한 설명

- [0044] 도 1은 종래의 네트워크 그래프를 나타낸 도면이다.
- 도 2는 본 발명의 일 실시예에 따른 감정어휘 분포맵 제작을 위해 선정된 감정어휘를 나타낸 도면이다.
- 도 3은 도 2에 도시된 각 감정어휘의 TF-IDF 스코어의 최대치를 나타낸 도면이다.
- 도 4는 도 2에 도시된 각 감정어휘들 중 최종 선정된 36개의 감정어휘를 나타낸 도면이다.
- 도 5는 본 발명의 일 실시예에 따른 감정어휘 분포맵(의미지도)을 나타낸 도면이다.
- 도 6 내지 도 7은 본 발명의 일 실시예에 따른 복수의 영화를 네트워크 그래프의 형태로 나타낸 도면이다.
- 도 8은 본 발명의 일 실시예에 따른 레퍼런스 의미지도와 네트워크 그래프의 결합관계를 나타낸 도면이다.
- 도 9 내지 도 10은 본 발명의 일 실시예에 따른 레퍼런스 의미지도와 결합된 네트워크 그래프의 노드들을 나타낸 도면이다.
- 도 11은 본 발명의 일 실시예에 따른 객체에 대한 속성값들에 대한 데이터를 시각화 하는 방법을 나타낸 순서도이다.
- 도 12는 본 발명의 일 실시예에 따른 객체의 위치를 결정하여 시각화 하는 방법을 나타낸 순서도이다.
- 도 13은 본 발명의 일 실시예에 따른 객체의 위치를 속성값 각각의 가중치에 기반한 거리를 반영하여 결정하는 단계를 보다 상세하게 나타낸 도면이다.
- 도 14는 본 발명의 일 실시예에 따른 속성값을 추출하는 단계를 보다 상세하게 나타낸 순서도이다.
- 도 15는 본 발명의 일 실시예에 따른 데이터를 시각화 하는 시스템을 나타내는 도면이다.
- 도 16은 본 발명의 일 실시예에 따른 위치 결정부를 보다 상세하게 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0045] 상기 목적 외에 본 발명의 다른 목적 및 특징들은 첨부 도면을 참조한 실시 예에 대한 설명을 통하여 명백히 드러나게 될 것이다.
- [0046] 본 발명의 바람직한 실시예를 첨부된 도면들을 참조하여 상세히 설명한다. 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.
- [0047] 그러나, 본 발명이 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.
- [0048] 도 2는 본 발명의 일 실시예에 따른 감정어휘 분포맵 제작을 위해 선정된 감정어휘를 나타낸 도면이다.
- [0049] 본 발명은 객체의 코멘트 데이터에서 수집된 속성값을 이용하여 네트워크 그래프를 생성하고, 생성된 네트워크 그래프를 레퍼런스 의미지도에 대응하여 재배치하는 시각화 방법 및 시스템에 관한 것으로서, 객체는 사용자가 선택한 영화, 상품, 소셜, 게임, 여행 등 사람의 감정이 포함되는 객체를 의미하며, 사람의 감정이 포함되는 객체에 대한 코멘트 또는 리뷰로부터 나타나는 감정을 네트워크 그래프를 통하여 시각화 할 수 있다.
- [0050] 본 발명의 일 실시예로서, 객체는 영화로 한정하여 영화에 대한 속성값을 이용한 네트워크 그래프 시각화 방법 및 시스템으로도 설명할 수 있다.
- [0051] 영화의 코멘트 데이터는 사용자에게 따라서 구축된 웹 서비스 통해 수집된 데이터를 이용할 수 있으며, 또는 대형 포털 및 동호회 게시판에 축적되는 코멘트 데이터를 프로그램을 이용하여 개별적으로 수집할 수도 있다.
- [0052] 본 발명의 일 실시예로서, 영화에 대한 코멘트 데이터로부터 사용자의 감정을 포함하고 있는 감정어휘 수집을 자동화하기 위하여 데이터를 수집할 수 있는 웹 크롤러를 이용할 수 있으며, 크롤러는 대형 포털(네이버, 다음 등) 영화 홈페이지에서 특정 영화의 댓글과 코멘트들을 정제되지 않은 데이터 형태로 수집하고, 수집된 데이터를 연구에 사용 가능한 데이터로 가공할 수 있으며, 정제된 데이터를 분석하여 감정어휘를 추출할 수 있다. 이에 따라 크롤러를 통하여 수집되는 감정어휘는 영화를 보는 상황과 연결시켜 추후에 사용자의 이용 동기에 맞는 영화를 추천할 수도 있다.
- [0053] 영화에 나타난 감정어휘의 빈도를 시각화하기 위해서는 2차원 평면상에 각 감정어의 위치를 지정해야 한다. 이

를 위해 감정어간의 상관관계를 이용하여 2차원 상의 위치 좌표를 도출할 수 있다. 감정어휘의 분포맵을 제작하기 위하여 한덕웅, 강혜자(2000)의 한국어 정서 용어들의 적절성과 경험 빈도에 대한 연구를 참고하여 834개의 정서용어 중에서 영화를 봤을 때 느낄 수 있는 감정어휘만을 분류하였다. 이때, 아주대학교의 국어국문학과박사 전문가 1명과 본 발명의 발명자 2명이 함께 서로 의견취합이 가능한 감정어휘만을 골라 최종 100개의 감정어휘를 선별하였다.

- [0054] 또한, 전문가 분석을 통한 감정어휘 선별작업 이외에도 영화를 시청하였을 때, 사용자들이 가장 많이 느끼는 감정어휘를 선별하기 위해 선정된 100개의 감정어휘를 토대로 최종감정어휘 선정을 위한 서베이를 실시하였다. 서베이는 아주대학교의 미디어학과 학생 30명을 대상으로 영화를 봤을 때 느낄 수 있는 감정에 대한 간단한 개념 설명을 거친 뒤에, 전문가 분석을 통해 얻어진 100개의 감정어휘에 대해서 영화를 보는 상황일 때 해당 감정어휘를 느낄 수 있는 정도가 어떻게 되는지를 조사하였다. 실제 설문에서는 ‘여러분이 지금까지 보신 여러 장르의 영화 스토리를 생각하신 뒤 해당 영화를 봤을 때 다음에 제시된 감정어휘들을 느끼는 정도가 어떠한 지 답하시기 바랍니다.’ 와 같이 시작하였으며, 각 감정어휘에 대해서 리커트 7점 척도(Likert-type scale)로 응답하게 하여 1점은 ‘전혀 관련 없다.’ 를 의미하고 7점은 ‘매우 관련 있다.’ 를 의미하도록 질문하였다.
- [0055] 본 연구에서는 사용자의 이용 동기를 이용해 영화를 추천한다는 취지에 부합하고자 영화를 봤을 때 가장 잘 느낄 수 있는 감정어휘를 수집하기 위해 전문가 분석 및 사용자 서베이를 실시하였으며, 사용자들이 설문한 리커트 7점 척도 정보를 토대로 관련성이 높은 감정어휘를 선별하기 위해 평균분석을 통해서 평균이 상대적으로 낮은 감정어휘(4.00 ‘보통이다.’ 를 뜻하는 수치 이하) 32개를 추가적으로 제거하여 영화 추천에 적합한 68개의 감정어휘를 선별하였다.
- [0056] 도 2는 이렇게 선별된 영화 추천에 적합한 68개의 감정어휘를 나타낸 도면이다.
- [0057] 도 3은 도 2에 도시된 각 감정어휘의 TF-IDF 스코어의 최대치를 나타낸 도면이다.
- [0058] 도 3은 도 2에 설명된 68개의 감정어휘에 실제 영화데이터를 비교하여 영향력이 미미한 감정어휘를 추가로 제거하기 위하여, 영화의 코멘트 또는 리뷰에 나타나는 각 감정어휘의 TF-IDF 스코어를 도출하고, 각 감정어휘에 나타날 수 있는 TF-IDF 스코어의 최대치를 나타낸 도면이다.
- [0059] 이때, TF(단어 빈도수, term frequency)는 특정한 단어가 문서 내에 얼마나 자주 등장하는지 나타내는 값을 의미하며, DF(Document Frequency)는 특정 단어가 나타난 문서의 수를 의미하며, 이 값의 역수를 IDF(inverse document frequency)라고 한다.
- [0060] 도 4는 도 3에 도시된 각 감정어휘들 중 최종 선정된 36개의 감정어휘를 나타낸 도면이다.
- [0061] 도 3에 도시된 도면은 TF-IDF 스코어가 도출된 각 감정어휘들 중에서 '경악하다'의 경우 모든 영화에서 TF-IDF 스코어의 비율이 0.8% 이하로 나타났으며, 반면에 '달콤하다'의 경우에는 적어도 한 개의 영화에서는 TF-IDF 스코어의 비율이 42%에 달하는 것을 의미한다.
- [0062] 이때, 도 3은 TF-IDF 스코어의 비율이 10% 미만인 감정어휘를 제거하고 최종적으로 선택된 36개의 감정어휘를 나타낸 도면이다.
- [0063] 도 5는 본 발명의 일 실시예에 따른 감정어휘 분포맵(의미지도)을 나타낸 도면이다.
- [0064] 도 4에 도시한 최종 근집화된 36개의 감정어휘를 2차원 평면에 각 감정어휘 간의 의미 거리를 도출하기 위하여 36개의 감정어휘를 바탕으로 유사하거나 상이한 감정어휘 간의 거리도를 측정하여 상관관계를 분석한 다음 다차원척도 분석(Multi-Dimensional Scaling: MDS)을 이용할 수 있다.
- [0065] 이때, 다차원척도 분석이란 개체들 간의 상대적인 거리를 계산하여 사람이 인지 할 수 있는 평면상에 상대적인 거리로 나타내는 통계와 연관된 기술로 정보 시각화에서는 데이터 내의 유사성 및 비유사성을 측정하기 위한 배경기술이다.
- [0066] 다차원척도법의 장점은 상대적인 거리만을 알고 있는 개체들의 의미 지도를 작성할 수 있으며 물리적인 거리뿐만 아니라 심리적인 거리에 근거하여서도 의미지도를 작성할 수 있다.
- [0067] 본 발명의 일 실시예에 따른 다차원척도 분석을 위해 경기도 및 서울 소재 대학교 20대 남학생 11명, 여학생 9명으로 총 20명을 실험 대상으로 하여 36개의 감정어휘에 대해 의미상 거리 서베이를 실시하였으며, 서베이는 가로축 세로축 36개의 감정어휘를 배치한 설문지를 만들고(68x68), 감정어휘간의 거리가 가장 가깝다고 느껴지면 3점, 가장 멀다고 느껴지면 -3점을 주는 방식의 리커트 척도를 이용하여 체크하는 형식으로 구성하였다. 20

명이 기록한 데이터를 바탕으로 다양한 네트워크 분석기법이 활용 가능한 UCINET 프로그램을 사용 하였고, 이에 따라 영화 36개 감정 어휘 간의 의미상의 거리에 선정된 68개의 감정어휘를 기반한 Metric MDS를 도 5에 나타내었다.

- [0068] 그 결과, X축의 양(+)의 방향으로는 대표어 “Happy”, “Surprise” 와 관련된 감정어휘가 분포되었으며, X축의 음(-)의 방향으로는 대표어 “Anger”, “Disgust” 와 관련된 감정어휘가 분포되었다. 그리고 Y축의 양(+)의 방향으로는 대표어 “Fear”, “Surprise” 와 관련된 감정어휘가 분포되었으며, Y축의 음(-)의 방향으로는 대표어 “Sad”, “Boring” 과 관련된 감정어휘가 분포되었다.
- [0069] 이에 따라, 감정어휘의 성격 상 X축의 양(+)의 방향은 긍정적인 감정어휘들이 분포되었고, X축의 음(-)의 방향으로는 부정적인 감정어휘들이 분포됨을 알 수 있다.
- [0070] 또한, Y축의 양(+)의 방향은 동적인(감정을 느낄 때 비교적 큰 제스처를 취할 수 있는) 감정어휘들이 분포되었고, Y축의 음(-)의 방향으로는 정적인(감정을 느낄 때 비교적 작은 제스처를 취할 수 있는) 감정어휘들이 분포됨을 알 수 있다.
- [0071] 그리고 ‘Happy’, ‘Sad’, ‘Anger’, ‘Fear’, ‘Disgust’, ‘Boring’ 대표어와 관련된 어휘들은 각각의 단어가 뚜렷하게 군집이 되는 것을 볼 수 있는데, 대표어 ‘Surprise’ 에 대해서는 ‘Happy’ 대표어 군집과 ‘Fear’ 대표어 군집에 나뉘어서 분포한다는 것을 알 수 있다. 이는 사용자가 영화를 봤을 때 ‘벽찬 기쁨으로 인해 놀라운 감정이 생기는 경우’ 와 ‘갑작스럽게 등장하는 공포로 인해서 놀라운 감정이 생기는 경우’ 가 지배적이기 때문인 것으로 해석할 수 있다.
- [0072] 도 6 내지 도 7은 본 발명의 일 실시예에 따른 복수의 영화를 네트워크 그래프의 형태로 나타낸 도면이다.
- [0073] 도 6은 도 2 내지 도 5에서 설명한 의미지도와 연결될 네트워크 그래프를 나타내었다. 도 6에 도시된 네트워크 그래프를 구성하기 위하여 300개의 영화를 대상으로 각 영화의 코멘트 데이터에 나타나는 감정어휘의 빈도수를 계수하였으며, 같은 감정어휘를 가진 영화를 유사한 영화라 가정하였다.
- [0074] 도 7은 도 6에 도시된 영화의 수를 400개로 증가 하였을 때 나타나는 네트워크 그래프를 나타낸 것이다.
- [0075] 도 6과 도 7을 비교해보면 데이터(영화)가 증가함에 따라 클러스터의 위치가 변화하는 것을 볼 수 있다.
- [0076] 도 8은 본 발명의 일 실시예에 따른 레퍼런스 의미지도와 네트워크 그래프의 결합관계를 나타낸 도면이다.
- [0077] 도 8은 도 2 내지 도 5에서 설명한 의미지도와 도 7 내지 도 8에서 설명한 네트워크 그래프를 연결한 형태를 나타내고 있다.
- [0078] 레퍼런스 의미지도를 기준으로 영화와 관련된 네트워크 지도를 연결할 수 있으며, 연결된 네트워크 지도는 의미지도에 분포된 감정어휘에 대응하여 이미지를 삽입할 수도 있다.
- [0079] 예를 들어, 2차원 의미지도를 기준으로 네트워크 그래프를 재배치 하였을 때, 의미지도상 긍정적인고, 동적인 부분에 네트워크 그래프의 노드들이 배치되는 빈도수가 많을 때, 긍정적인고, 동적인 이미지를 가지고 있는 그림, 사진, 별자리 등의 메타포어를 네트워크 그래프 대신 삽입하여 시각화할 수 있다.
- [0080] 또한, 그림, 사진, 별자리 등의 삽입된 이미지 속에 세부적으로 이미지를 추가 삽입할 수 있으며, 사용자가 이미지를 선택하면 세부적인 네트워크 그래프를 보여줄 수도 있고, 또는 세부적인 네트워크 그래프에 대응하는 삽입된 이미지를 사용자에게 보여줄 수도 있다.
- [0081] 이 때 도 8에 도시된 그림, 사진, 별자리 등의 삽입된 이미지(메타포어, metaphor)는 유사한 의미를 가지는 객체들의 군에 대응하는 일종의 인덱스로서 기능할 수도 있다. 즉, 지나치게 많은 수의 객체들(콘텐츠들)이 동시에 디스플레이되는 경우, 가시성이 떨어질 우려가 있으므로 이들 객체들을 어떤 메타포어에 대응하도록 그룹핑하고, 메타포어를 선택하면 그 메타포어 내의 객체들만이 선별적으로 디스플레이될 수 있다. 이 때 하나의 객체는 하나의 메타포어에만 대응할 필요는 없고, 의미 상 연관성이 있으면 둘 이상의 메타포어에 동시에 속할 수도 있다.
- [0082] 메타포어는 다수 객체를 포함하는 그룹인데, 시각화 과정에서는 다수 노드의 네트워크 그래프에 대한 그룹으로 표현될 수도 있다. 이 때, 메타포어는 해당 네트워크 그래프 그룹의 의미에 대한 직관적인 느낌을 전달하기 위하여 선택되는 것으로, 그림, 캐릭터, 널리 알려진 도형, 널리 알려진 영화의 장면 등이 채용될 수 있다.
- [0083] 메타포어는 표현요소(감정어휘, 의견을 반영하는 어휘 등)의 의미에 따라서 계층적으로 설정할 수도 있다. 즉,

상위의 개념을 반영하는 메타포어와 그보다 하위의 개념을 반영하는 메타포어가 존재할 수 있으며, 이 때 사용자는 상위 개념의 메타포어를 탐색한 뒤 세부적으로 하위의 개념을 반영하는 메타포어를 선택하여 객체의 수를 선별적으로 줄어나갈 수 있다.

- [0084] 도 8에서는 메타포어를 통하여 사용자에게 객체의 수를 제한하여 디스플레이하는 과정이 도시되었지만, 본 발명의 사상은 여기에 국한되지 않는다. 예를 들어, 메타포어 대신에 사용자의 검색 결과에 따라서 디스플레이되는 노드의 범위를 제한하는 기법도 가능하다.
- [0085] 도 9 내지 도 10은 본 발명의 일 실시예에 따른 레퍼런스 의미지도와 결합된 네트워크 그래프의 노드들을 나타낸 도면이다.
- [0086] 도 9는 도 6에 도시된 300개의 영화를 대상으로 나타나는 네트워크 그래프와 레퍼런스 의미지도를 연결하였을 때 나타나는 최종 네트워크 그래프의 형태를 나타낸 것이다.
- [0087] 도 10은 도 7에 도시된 400개의 영화를 대상으로 나타나는 네트워크 그래프와 레퍼런스 의미지도를 연결하였을 때 나타나는 최종 네트워크 그래프의 형태를 나타낸 것이다.
- [0088] 이와 같이 대상 데이터(영화)가 증가함에 따라 네트워크 그래프는 복잡해질 수 있지만 색상으로 구분이 가능한 같은 속성을 가진 노드가 항상 감정어휘상 비슷한 지점에 위치하게 됨으로써 그래프의 분석이 더욱 편리해질 수 있다.
- [0089] 도 11은 본 발명의 일 실시예에 따른 객체에 대한 속성값들에 대한 데이터를 시각화 하는 방법을 나타낸 순서도이다.
- [0090] 하나의 객체에 대한 속성값들을 레퍼런스 의미지도(Semantic Map)와 결합하여 시각화 하는 방법은 객체에 대하여 적어도 하나 이상의 속성값을 추출하고(S1110), 추출된 속성값을, 추출된 속성값들 간의 의미 거리 기반으로, 미리 설정된 레퍼런스 맵을 기준으로 하여 배치한다(S1120).
- [0091] 이때, 레퍼런스 맵은 다차원척도 분석(Multi-Dimensional Scaling: MDS)을 이용하여 형성되며, 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성될 수 있으며, 의미 거리는 도 5에서 설명한 것처럼 속성값마다의 특성에 따라 정해지게 된다.
- [0092] 또한, 객체는 평가 및 의견을 제시할 대상을 의미하며, 속성값은 표현요소, 감정어휘, 표현어휘, 평가어휘, 가치의견 등 사람의 감정을 나타내는 단어, 문단, 문장 등과 같은 글씨와 캐릭터, 이모티콘 등의 이미지를 모두 포함한다.
- [0093] 이후, 객체로부터 추출된 하나 이상의 속성값들 각각을 하나의 노드에 대응시키고, 객체로부터 추출된 하나 이상의 속성값들 각각에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화 한다(S1130). 즉, 네트워크 그래프는 하나의 객체에서 추출된 속성값들을 연결하게 된다.
- [0094] 이때, 네트워크 그래프는 레퍼런스 맵의 의미 거리에 기반하여 배치된 모양 또는 속성값들의 빈도수에 따라 이미지를 삽입하여 시각화할 수도 있다.
- [0095] 예를 들어, 속성값들의 분포가 레퍼런스 맵에서 긍정적이고 동적인 의미를 가지는 위치에 집중되어 있는 경우, 긍정적이고 동적인 이미지를 가지는 캐릭터, 자연환경, 별자리 등의 이미지(메타포어)를 삽입하여 사용자에게 제공할 수도 있다.
- [0096] 이에 따라, 사용자는 제공받은 이미지만으로도 객체에 대한 정보를 획득할 수도 있으며, 사용자가 이미지를 선택하는 경우, 세부적인 계층화된(hierarchical) 네트워크 그래프 및 각각 하나의 노드에 대응되는 속성값에 대한 정보를 제공받을 수도 있으며, 또는 전체적인 이미지 안에 계층화된 세부적인 이미지를 사용자에게 제공할 수도 있다. 예를 들어, 긍정적이고 동적인 이미지를 사용자가 선택하는 경우, 세부적으로 계층화된 긍정적인 이미지와 동적인 이미지(메타포어)를 각각 사용자에게 제공할 수도 있다.
- [0097] 도 12는 본 발명의 일 실시예에 따른 객체의 위치를 결정하여 시각화 하는 방법을 나타낸 순서도이다.
- [0098] 2차원 의미지도(레퍼런스 맵) 대비 속성값의 위치 및 의미 거리 기반으로 객체의 위치를 결정하여 시각화 하는 방법은 객체에 대하여 적어도 하나 이상의 속성값을 추출하고(S1210), 추출된 하나 이상의 속성값 각각을 미리 결정된 레퍼런스 맵을 기준으로 하여 배치한다(S1220). 이후, 객체의 위치를 배치된 하나 이상의 속성값 각각과의 가중치에 기반한 거리를 반영하여 결정한다(S1230).

- [0099] 이후, 복수의 객체들 각각을 하나의 노드에 대응시키고, 복수의 객체들 중 기준 속성값이 공통적으로 추출된 객체들에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화한다(S1240). 즉, 네트워크 그래프는 동일한 속성값을 가지고 있는 객체들을 연결하게 된다.
- [0100] 이때, 네트워크 그래프는 레퍼런스 맵의 의미 거리에 기반하여 배치된 모양 또는 속성값들의 빈도수에 따라 이미지를 삽입하여 시각화할 수도 있다.
- [0101] 이에 따라, 사용자는 제공받은 이미지만으로도 객체에 대한 정보를 획득할 수도 있으며, 사용자가 이미지를 선택하는 경우, 세부적인 네트워크 그래프 및 각각 하나의 노드에 대응되는 속성값에 대한 정보를 제공받을 수도 있으며, 또는 전체적인 이미지 안에 세부적인 이미지를 사용자에게 제공할 수도 있다. 예를 들어, 긍정적이고 동적인 이미지를 사용자가 선택하는 경우, 세부적으로 긍정적인 이미지와 동적인 이미지를 각각 사용자에게 제공할 수도 있다.
- [0102] 또한, 레퍼런스 맵은 다차원척도 분석(Multi-Dimensional Scaling: MDS)을 이용하여 형성되며, 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성될 수 있으며, 의미 거리는 도 5에서 설명한 것처럼 속성값마다의 특성에 따라 정해지게 된다.
- [0103] 도 13은 본 발명의 일 실시예에 따른 객체의 위치를 속성값 각각의 가중치에 기반한 거리를 반영하여 결정하는 단계를 보다 상세하게 나타낸 도면이다.
- [0104] 객체의 위치를 배치된 하나 이상의 속성값 각각과의 가중치에 기반한 거리를 반영하여 결정할 때(S1230), 객체에 대하여 적어도 하나 이상의 속성값이 추출되는 빈도수를 계수하고(S1231), 계수된 빈도수를 반영하여 속성값 각각과의 가중치를 결정한다(S1232).
- [0105] 이에 따라, 레퍼런스 맵에서 의미거리를 기반으로 네트워크 그래프를 생성할 때, 속성값들의 빈도수에 따라 객체의 위치를 결정할 수 있다. 예를 들어, 객체에 대한 속성값들의 빈도수가 레퍼런스 맵에서 긍정적인 부분의 빈도수가 높으면, 객체의 위치는 긍정적인 부분으로 결정할 수 있다.
- [0106] 도 14는 본 발명의 일 실시예에 따른 속성값을 추출하는 단계를 보다 상세하게 나타낸 순서도이다.
- [0107] 객체에 대하여 적어도 하나 이상의 속성값을 추출할 때(S1110, S1210), 객체에 대한 코멘트 데이터를 수집하고(S1410), 수집된 코멘트 데이터에서 적어도 하나 이상의 속성값을 추출한다(S1420).
- [0108] 이때, 속성값을 코멘트 데이터의 음절/형태소/단어가 하나인 유니그램(Unigram) 또는 음절/형태소/단어가 2개인 바이그램(Bigram) 방식으로 추출할 수도 있으며, 마찬가지로 음절/형태소/단어가 3개인 트라이그램(Trigram) 방식으로 추출할 수도 있다. 이처럼 속성값을 코멘트 데이터로부터 추출하는 과정에서는 자연어 처리 알고리즘을 활용할 수도 있다.
- [0109] 이때, 객체는 사용자가 선택한 영화, 상품, 소설, 게임, 여행 등 사람의 감정이 포함되는 객체를 의미하며, 사람의 감정이 포함되는 객체에 대한 코멘트 또는 리뷰를 코멘트 데이터라고 한다.
- [0110] 코멘트 데이터는 사용자에게 따라서 구축된 웹 서비스 통해 수집된 데이터를 이용할 수 있으며, 또는 대형 포털 및 동호회 게시판에 축적되는 코멘트 데이터를 프로그램을 이용하여 개별적으로 수집할 수도 있다.
- [0111] 본 발명의 일 실시예로서, 객체(영화)에 대한 코멘트 데이터로부터 사용자의 감정을 포함하고 있는 감정어휘 수집을 자동화하기 위하여 데이터를 수집할 수 있는 웹 크롤러를 이용할 수 있으며, 크롤러는 대형 포털(네이버, 다음 등) 영화 홈페이지에서 특정 영화의 댓글과 코멘트들을 정제되지 않은 데이터 형태로 수집하고, 수집된 데이터를 연구에 사용 가능한 데이터로 가공할 수 있으며, 정제된 데이터를 분석하여 감정어휘를 추출할 수 있다. 이에 따라 크롤러를 통하여 수집되는 감정어휘는 영화를 보는 상황과 연결시켜 추후에 사용자의 이용 동기에 맞는 영화를 추천할 수도 있다.
- [0112] 도 15는 본 발명의 일 실시예에 따른 데이터를 시각화 하는 시스템을 나타내는 도면이다.
- [0113] 데이터 시각화 시스템(1500)은 하나의 객체에 대한 속성값들을 레퍼런스 의미지도와 결합하여 시각화 할 수도 있으며, 레퍼런스 의미지도 대비 속성값의 위치 및 의미 거리 기반으로 객체의 위치를 결정하여 시각화 할 수도 있다.
- [0114] 이러한 데이터 시각화 시스템(1500)은 코멘트 데이터 수집부(1510), 속성값 추출부(1520), 속성값 배치부(1530), 위치 결정부(1540), 네트워크 그래프 생성부(1550)를 포함한다.

- [0115] 코멘트 수집부(1510)는 객체에 대한 코멘트 데이터를 수집하고, 속성값 추출부(1520)는 수집된 코멘트 데이터에서 적어도 하나 이상의 속성값을 추출한다.
- [0116] 이때, 속성값을 코멘트 데이터의 음절이 하나인 유니그램(Unigram) 또는 음절이 2개인 바이그램(Bigram) 방식으로 추출할 수도 있으며, 음절이 3개인 트라이그램(Trigram) 방식으로 추출할 수도 있다.
- [0117] 이때, 객체는 사용자가 선택한 영화, 상품, 소설, 게임, 여행 등 사람의 감정이 포함되는 객체를 의미하며, 사람의 감정이 포함되는 객체에 대한 코멘트 또는 리뷰를 코멘트 데이터라고 한다.
- [0118] 코멘트 데이터는 사용자에게 따라서 구축된 웹 서비스 통해 수집된 데이터를 이용할 수 있으며, 또는 대형 포털 및 동호회 게시판에 축적되는 코멘트 데이터를 프로그램을 이용하여 개별적으로 수집할 수도 있다.
- [0119] 본 발명의 일 실시예로서, 객체(영화)에 대한 코멘트 데이터로부터 사용자의 감정을 포함하고 있는 감정어휘 수집을 자동화하기 위하여 데이터를 수집할 수 있는 웹 크롤러를 이용할 수 있으며, 크롤러는 대형 포털(네이버, 다음 등) 영화 홈페이지에서 특정 영화의 댓글과 코멘트들을 정제되지 않은 데이터 형태로 수집하고, 수집된 데이터를 연구에 사용 가능한 데이터로 가공할 수 있으며, 정제된 데이터를 분석하여 감정어휘를 추출할 수 있다. 이에 따라 크롤러를 통하여 수집되는 감정어휘는 영화를 보는 상황과 연결시켜 추후에 사용자의 이용 동기에 맞는 영화를 추천할 수도 있다.
- [0120] 데이터 시각화 시스템(1500)이 하나의 객체에 대한 속성값들을 레퍼런스 의미지도(Semantic Map)와 결합하여 시각화 하는 경우, 속성값 추출부(1520)는 객체에 대하여 적어도 하나 이상의 속성값을 추출하고, 속성값 배치부(1530)는 추출된 속성값을, 추출된 속성값들 간의 의미 거리 기반으로, 미리 설정된 레퍼런스 맵을 기준으로 하여 배치한다.
- [0121] 이후, 네트워크 그래프 생성부(1550)는 객체로부터 추출된 하나 이상의 속성값들 각각을 하나의 노드에 대응시키고, 객체로부터 추출된 하나 이상의 속성값들 각각에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화 한다. 즉, 네트워크 그래프는 하나의 객체에서 추출된 속성값들을 연결하게 된다.
- [0122] 데이터 시각화 시스템(1500)이 레퍼런스 의미지도(레퍼런스 맵) 대비 속성값의 위치 및 의미 거리 기반으로 객체의 위치를 결정하여 시각화 하는 경우,
- [0123] 속성값 추출부(1520)는 객체에 대하여 적어도 하나 이상의 속성값을 추출하고, 속성값 배치부(1530)는 추출된 하나 이상의 속성값 각각을 미리 결정된 레퍼런스 맵을 기준으로 하여 배치한다. 이후, 위치 결정부(1540)는 객체의 위치를 배치된 하나 이상의 속성값 각각과의 가중치에 기반한 거리를 반영하여 결정한다.
- [0124] 이후, 네트워크 그래프 생성부(1550)는 복수의 객체들 각각을 하나의 노드에 대응시키고, 복수의 객체들 중 기준 속성값이 공통적으로 추출된 객체들에 대응하는 노드를 아크(arc)로 연결하여 네트워크 그래프를 시각화한다(S1240). 즉, 네트워크 그래프는 동일한 속성값을 가지고 있는 객체들을 연결하게 된다.
- [0125] 이때, 레퍼런스 맵은 다차원척도 분석(Multi-Dimensional Scaling: MDS)을 이용하여 형성되며, 미리 설정된 복수의 속성값들 간의 의미 거리에 기반하여 형성될 수 있으며, 의미 거리는 도 5에서 설명한 것처럼 속성값마다의 특성에 따라 정해지게 된다.
- [0126] 또한, 객체는 평가 및 의견을 제시할 대상을 의미하며, 속성값은 표현요소, 감정어휘, 표현어휘, 평가어휘, 가치의견 등 사람의 감정을 나타내는 단어, 문단, 문장등과 같은 글씨와 캐릭터, 이모티콘 등의 이미지를 모두 포함한다.
- [0127] 이때, 네트워크 그래프는 레퍼런스 맵의 의미 거리에 기반하여 배치된 모양 또는 속성값들의 빈도수에 따라 이미지를 삽입하여 시각화할 수도 있다.
- [0128] 예를 들어, 속성값들의 분포가 레퍼런스 맵에서 긍정적이고 동적인 의미를 가지는 위치에 집중되어 있는 경우, 긍정적이고 동적인 이미지를 가지는 캐릭터, 자연환경, 별자리 등의 이미지를 삽입하여 사용자에게 제공할 수도 있다.
- [0129] 이에 따라, 사용자는 제공받은 이미지만으로도 객체에 대한 정보를 획득할 수도 있으며, 사용자가 이미지를 선택하는 경우, 세부적인 네트워크 그래프 및 각각 하나의 노드에 대응되는 속성값에 대한 정보를 제공받을 수도 있으며, 또는 전체적인 이미지 안에 세부적인 이미지를 사용자에게 제공할 수도 있다. 예를 들어, 긍정적이고 동적인 이미지를 사용자가 선택하는 경우, 세부적으로 긍정적인 이미지와 동적인 이미지를 각각 사용자에게 제

공할 수도 있다.

- [0130] 도 16은 본 발명의 일 실시예에 따른 위치 결정부를 보다 상세하게 나타낸 도면이다.
- [0131] 위치 결정부(1540)는 빈도수 계수부(1541)와 가중치 결정부(1542)를 포함한다. 빈도수 계수부(1541)는 객체에 대하여 적어도 하나 이상의 속성값이 추출되는 빈도수를 계수하고, 가중치 결정부(1542)는 계수된 빈도수를 반영하여 속성값 각각과의 가중치를 결정한다.
- [0132] 이에 따라, 레퍼런스 맵에서 의미거리를 기반으로 네트워크 그래프를 생성할 때, 속성값들의 빈도수에 따라 객체의 위치를 결정할 수 있다. 예를 들어, 객체에 대한 속성값들의 빈도수가 레퍼런스 맵에서 긍정적인 부분의 빈도수가 높으면, 객체의 위치는 '긍정적인 부분'으로 결정할 수 있다.
- [0133] 본 발명의 상술한 실시예에 따르면, 하나의 대표적인 표현요소에 대응하는 메타포어, 그리고 그 메타포어에 대응하는 객체들 각각이 의미지도 상에서 의미 거리에 기반하여 위치된다. 이렇게 시각화된 의미지도는 각 객체들 간의 유사성 뿐만 아니라 각 객체들 간의 미묘한 차이점도 반영하여, 사용자가 직관적으로 인식할 수 있도록 한다. 이 때 사용자는 상위 개념의 메타포어 또는 검색 메뉴를 이용하여 원하는 객체 범위를 특정하고, 계층화된(hierarchical) 하위 개념의 메타포어 또는 검색 메뉴를 이용하여 시각화되는 객체의 범위를 좁혀 나갈 수 있다. 이 때, 사용자가 하나의 노드를 선택하는 경우 하나의 노드(객체)에 대응하는 표현요소들이 상세하게 시각화될 수도 있다. 이 때 하나의 노드에 대응하는 표현요소라 함은, 하나의 노드(콘텐츠) 내에 포함되는 표현요소들일 수도 있고, 노드(콘텐츠)에 대한 사용자 리뷰 내에 포함되는 표현요소들일 수도 있다. 또한, 복수의 노드들을 선택하여, 각 선택된 노드들에 대응하는 표현요소들이 상세하게 시각화될 수도 있으며, 선택된 노드에 대응하는 표현요소들을 하나의 집합으로 간주한다면, 선택된 노드들 간의 집합 연산(교집합, 합집합, 차집합)을 통하여 도출된 표현요소들이 시각화될 수도 있다.
- [0134] 또한 객체에 대응하는 표현요소의 분포가 변화하는 경우, 이를 반영하여 각 객체의 위치가 재조정될 수 있으며, 신규한 표현요소가 추가되는 경우에도 re-draw될 수 있다. 이 같은 변화는 실시간적으로 반영될 수도 있으며, 주기적으로 반영될 수도 있고, 사용자가 미리 설정한 조건에 따라 반영될 수도 있고, 사용자의 명령이 입력되었을 때에 인터랙티브하게 반영될 수도 있다.
- [0135] 또한 본 발명의 또 다른 실시예에서는 시간에 따른 변화를 시계열적인 버전(레이어)으로 관리할 수도 있으며, 이 경우에는 시간에 따른 변화가 비교되어 시각화될 수도 있고, 최근 신규하게 등장하였거나 최근 집중적으로 참조된 객체 또는 최근 집중적으로 증가하였거나 신규하게 등장한 표현요소를 반영하여 시각화될 수도 있다. 이 때 시계열적으로 관심 노드의 의미 지도 상의 위치 이동을 시각화하는 메뉴가 제공될 수도 있다.
- [0136] 한편 본 발명의 또 다른 실시예에서는, 이처럼 시각화된 이후에 사용자가 특정 객체에 대한 선택을 통하여 이차적인 액션, 즉, 구매 또는 광고의 관람을 선택할 수 있는 메뉴를 추가적으로 제공할 수 있다. 예를 들어, 특정 영화를 선택한 경우에는 VOD 서비스 메뉴가 추가적으로 제공되거나, 상품의 경우에는 구매 버튼 또는 구매 사이트로 연결되는 윈도우가 디스플레이될 수 있다. 즉, 특정 객체의 선택으로 인하여 광고 링크로 연결되거나, 직접 구매를 선택할 수 있는 메뉴가 제공될 수 있다.
- [0137] 본 발명의 일 실시 예에 따른 데이터 시각화 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0138] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.

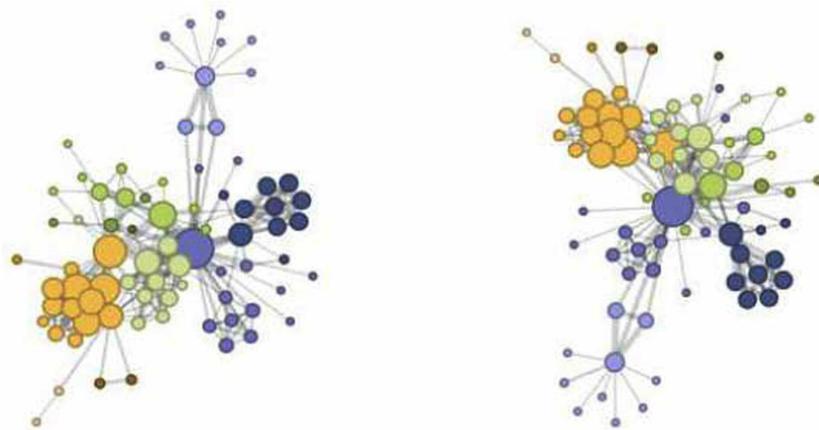
[0139] 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

부호의 설명

- [0140] 1500: 데이터 시각화 시스템
- 1510: 코멘트 데이터 수집부
- 1520: 속성값 추출부
- 1530: 속성값 배치부
- 1540: 위치 결정부
- 1550: 네트워크 그래프 생성부

도면

도면1



도면2

68개의 감정어휘					
기련하다	가슴아프다	감격스럽다	감동적이다	거북하다	겁나다
격분하다	경악하다	경이롭다	경쾌하다	공포스럽다	광강하다
구역질나다	그립다	기겁하다	기쁘다	꼼짝하다	나른하다
놀리다	늘어지다	달콤하다	담담하다	대단하다	더럽다
두렵다	등골이 서늘하다	무료하다	무섭다	무시무시하다	분노하다
분하다	불결하다	불쌍하다	불쾌하다	불편하다	서글프다
서럽다	섬뜩하다	소름끼치다	속상하다	슬프다	신경질나다
신나다	쓸쓸하다	안타깝다	암울하다	애석하다	애절하다
역겹다	오싹하다	우울하다	유쾌하다	으스스하다	잔인하다
재미있다	즐겁다	증오스럽다	지루하다	칭그럽다	차분하다
통쾌하다	평온하다	행복하다	혐오스럽다	환상적이다	활기 있다
황홀하다	흥겹다				

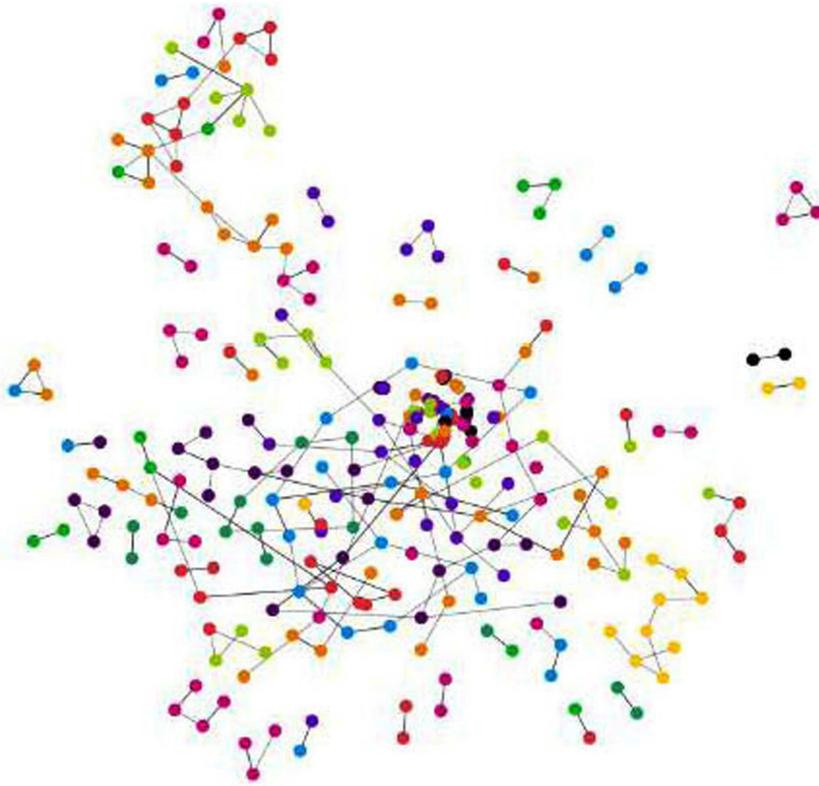
도면3

경쾌하다	가련하다	경악하다	서럽다	담담하다	으시시하다	끔찍하다	섬뜩하다
0.002525	0.005773	0.008493	0.016623	0.016993	0.021914	0.023694	0.025571
기겁하다	늘어지다	두렵다	속상하다	증오스럽다	겁나다	불쌍하다	기쁘다
0.027224	0.027751	0.03335	0.034018	0.034087	0.035402	0.042074	0.051313
징그럽다	애절하다	우울하다	분하다	차분하다	불쾌하다	그립다	거북하다
0.052592	0.053775	0.055982	0.068981	0.079125	0.079163	0.083	0.0838
더럽다	신경질나다	혐오스럽다	암울하다	흥겁다	무료하다	불편하다	역겹다
0.085487	0.08675	0.090305	0.094547	0.09533	0.096808	0.097098	0.097621
구역질나다	감격스럽다	서글프다	경이롭다	굉장하다	무시무시하다	쉴쉴하다	애석하다
0.104709	0.104717	0.112113	0.113636	0.121043	0.121759	0.123827	0.125668
분노하다	안타깝다	활기있다	가슴아프다	평온하다	격분하다	불결하다	나른하다
0.130046	0.130744	0.132016	0.140839	0.14449	0.148861	0.152184	0.168187
통쾌하다	오싹하다	신나다	유쾌하다	행복하다	재미있다	대단하다	잔인하다
0.170107	0.175868	0.186403	0.187977	0.194248	0.211447	0.234945	0.247608
지루하다	등골이서늘하다	무섭다	환상적이다	공포스럽다	즐겁다	감동적이다	황홀하다
0.249294	0.256714	0.268716	0.275393	0.27846	0.326984	0.341209	0.352601
놀라다	소름끼치다	달콤하다	슬프다				
0.394018	0.407559	0.422425	0.433089				

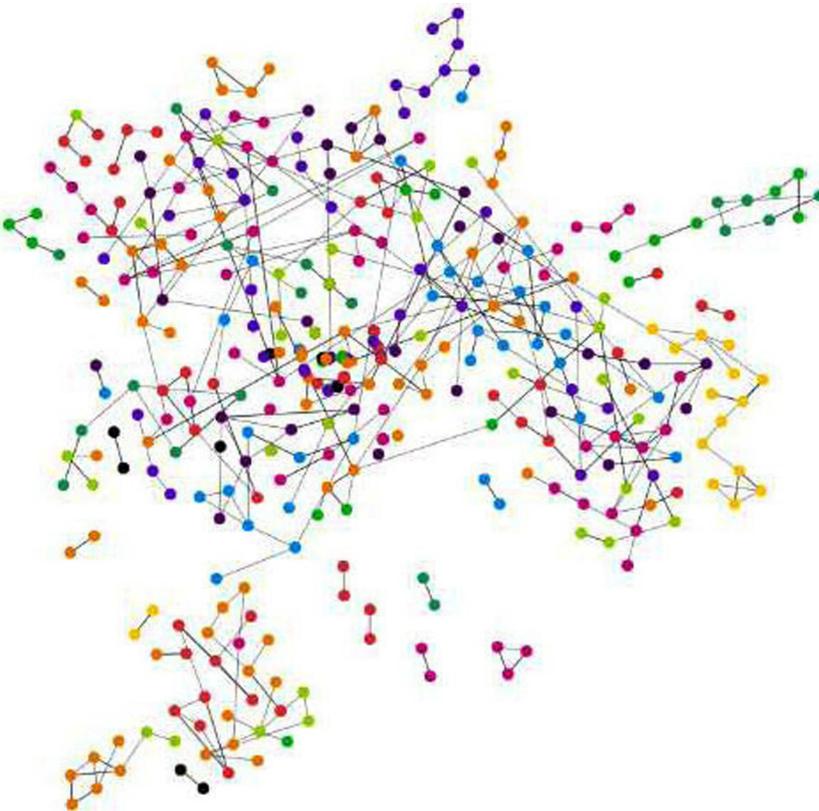
도면4

구역질나다	감격스럽다	서글프다	경이롭다	굉장하다	무시무시하다
0.104709195	0.104716879	0.112112533	0.113635779	0.121043081	0.121758791
쉴쉴하다	애석하다	분노하다	안타깝다	활기있다	가슴아프다
0.12382667	0.12566764	0.130046071	0.130743697	0.132015568	0.140838894
평온하다	격분하다	불결하다	나른하다	통쾌하다	오싹하다
0.144490291	0.148861125	0.152184309	0.168186738	0.170106636	0.175867531
신나다	유쾌하다	행복하다	재미있다	대단하다	잔인하다
0.186403236	0.18797709	0.194247664	0.211447121	0.234945365	0.247608174
지루하다	등골이서늘하다	무섭다	환상적이다	공포스럽다	즐겁다
0.249293951	0.256713888	0.26871569	0.275393151	0.278460331	0.326983704
감동적이다	황홀하다	놀라다	소름끼치다	달콤하다	슬프다
0.341209046	0.352600559	0.39401762	0.407559275	0.422425395	0.433089246

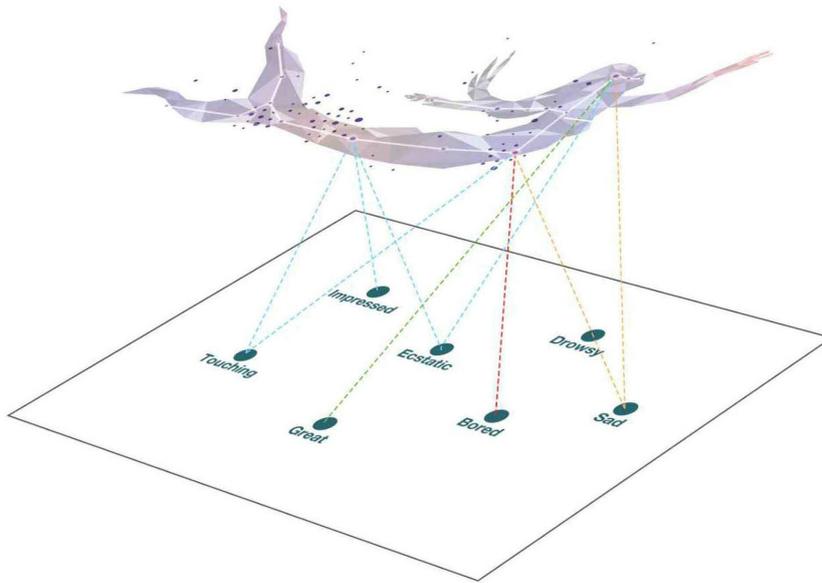
도면6



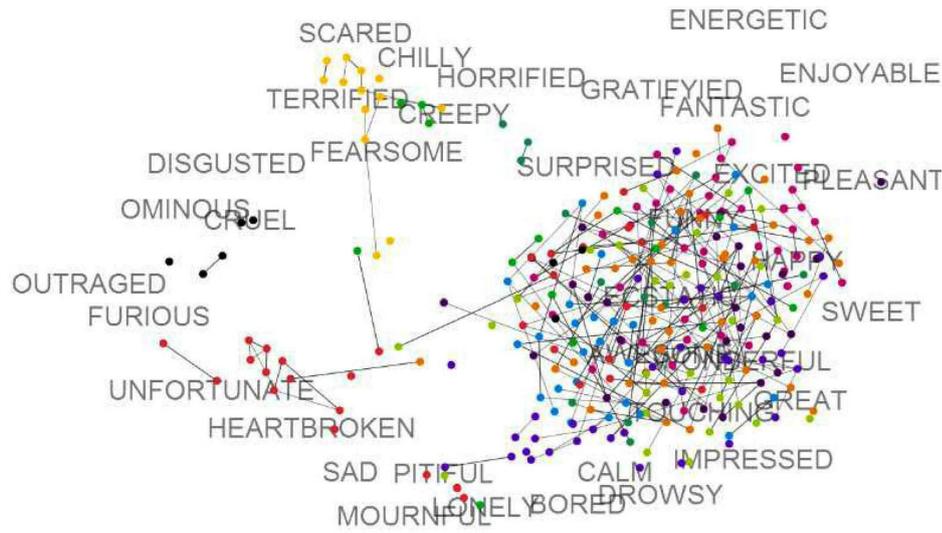
도면7



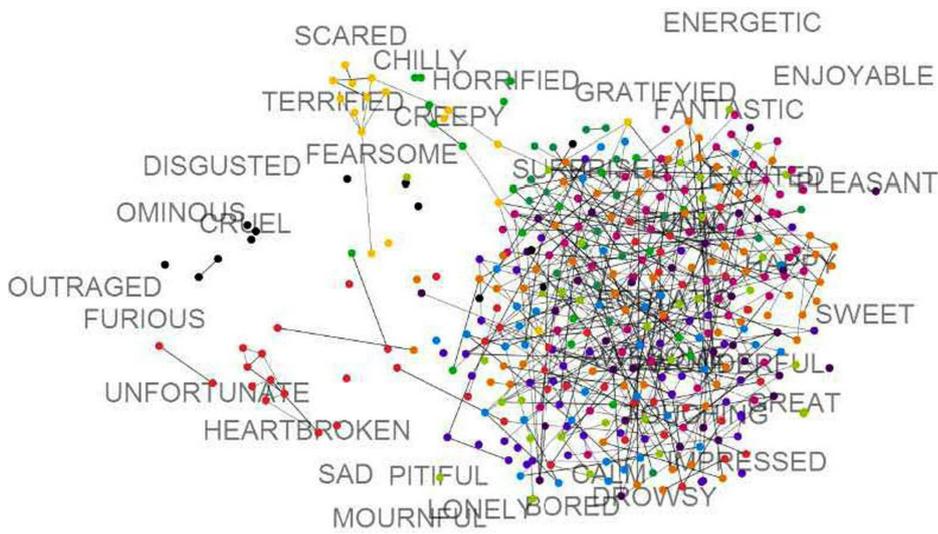
도면8



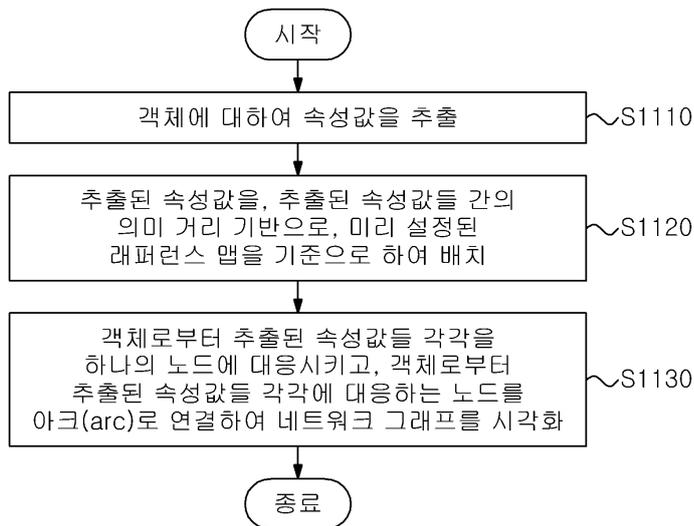
도면9



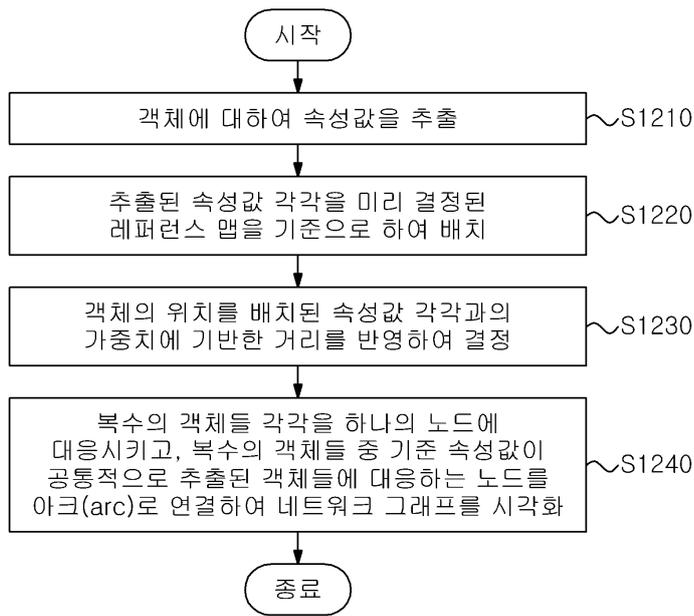
도면10



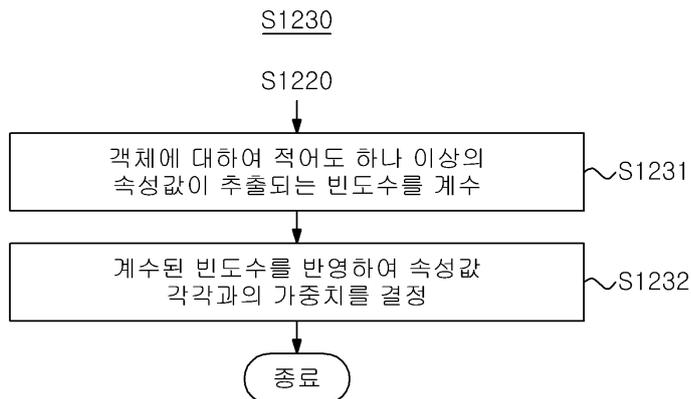
도면11



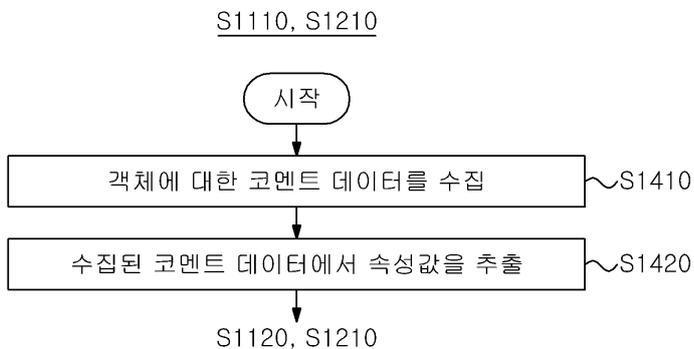
도면12



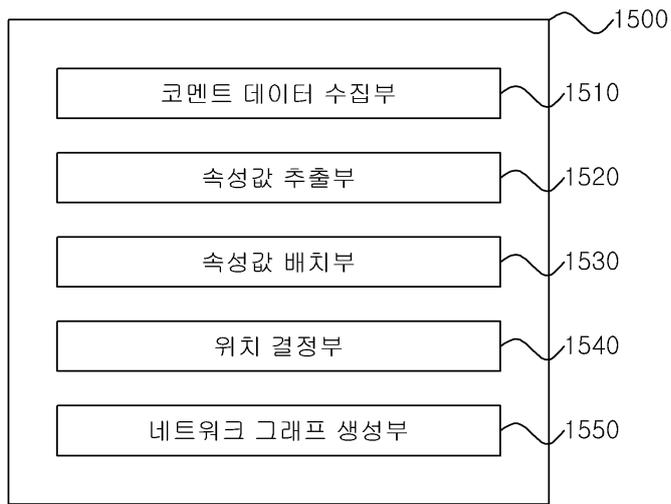
도면13



도면14



도면15



도면16

