



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0088205
(43) 공개일자 2014년07월09일

(51) 국제특허분류(Int. Cl.)
H04N 21/45 (2011.01) *H04N 21/47* (2011.01)
(21) 출원번호 10-2014-7014841
(22) 출원일자(국제) 2012년09월13일
심사청구일자 없음
(85) 번역문제출일자 2014년05월30일
(86) 국제출원번호 PCT/US2012/055222
(87) 국제공개번호 WO 2013/066503
국제공개일자 2013년05월10일
(30) 우선권주장
13/467,339 2012년05월09일 미국(US)
61/553,221 2011년10월30일 미국(US)

(71) 출원인
구글 인코포레이티드
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(72) 발명자
글로와티 그레체고로츠
폴란드 프셰미슬 피엘-37-700 오스트로 516
브르조조우스키 미카엘
폴란드 바르샤바 피엘-01-451 울. 로만스카 1 엠 49
월가스 마신
폴란드 크라코프 피엘-30-611 울. 위슬로코프 13 엠 58
(74) 대리인
박장원

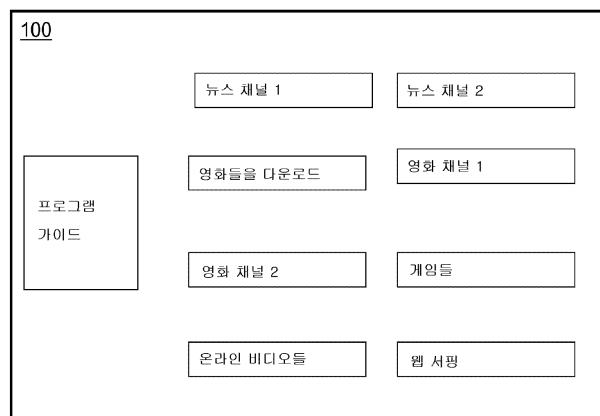
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **매체 프로그램들 사이의 유사성 계산**

(57) 요약

매체 프로그램들을 관련시키거나 계산하기 위한 시스템 및 방법에 제공된다. 방법은: 적어도 하나의 처리 회로를 사용하여, 제1 매체 프로그램에 대한 제1 메타데이터 및 제2 매체 프로그램에 대한 제2 메타데이터를 획득하는 단계 - 상기 제1 메타데이터는 복수의 제1 필드들로 구성되고, 상기 제2 메타데이터는 복수의 제2 필드들로 구성되고 - 와; 적어도 하나의 처리 회로를 사용하여, 상기 복수의 제1 필드들의 각각으로부터 복수의 제1 토큰들 및 상기 복수의 제2 필드들의 각각으로부터 복수의 제2 토큰들을 추출하는 단계와; 상기 제1 및 제2 토큰들의 각각에 가중치 팩터를 할당하는 단계와; 상기 복수의 제1 필드들과 상기 복수의 제2 필드들 사이의 상기 제1 및 제2 토큰들을 교차-상관시키는 단계와; 그리고 상기 교차-상관시키는 단계에 기초하여 상기 제1 및 제2 매체 프로그램 사이의 유사성 스코어를 계산시키는 단계를 포함한다.

대표도



특허청구의 범위

청구항 1

컴퓨터로 구현되는, 매체 프로그램들을 관련(associate)시키는 방법에 있어서,

(A) 적어도 하나의 처리 회로를 사용하여, 제1 매체 프로그램에 대한 제1 메타데이터 및 제2 매체 프로그램에 대한 제2 메타데이터를 획득하는 단계 - 상기 제1 메타데이터는 복수의 제1 필드들로 구성되고, 상기 제2 메타데이터는 복수의 제2 필드들로 구성되고 - 와;

(B) 적어도 하나의 처리 회로를 사용하여, 상기 복수의 제1 필드들의 각각으로부터 복수의 제1 토큰들을 추출하고 상기 복수의 제2 필드들의 각각으로부터 복수의 제2 토큰들을 추출하는 단계와;

(C) 상기 제1 및 제2 토큰들의 각각에 가중치 팩터를 할당하는 단계와;

(D) 상기 복수의 제1 필드들과 상기 복수의 제2 필드들 사이의 상기 제1 및 제2 토큰들을 교차-상관(cross-correlating)시키는 단계와; 그리고

(E) 상기 교차-상관시키는 단계에 기초하여 상기 제1 및 제2 매체 프로그램 사이의 유사성 스코어(similarity score)를 계산시키는 단계를 포함하는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 2

제 1 항에 있어서, (D)는:

(F) 상기 제1 토큰들 및 제1 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제1 필드들 중 하나를 나타내는 제1 벡터를 구성하고 상기 제2 토큰들 및 제2 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제2 필드들 중 하나를 나타내는 제2 벡터를 구성하는 단계와;

(G) 상기 복수의 제1 필드들 중 하나와 상기 복수의 제2 필드들 중 하나 사이의 유사성을 나타내는 상기 제1 및 제2 벡터의 내적(dot product)을 계산하는 단계와; 그리고

(H) 상기 복수의 제1 필드들 및 복수의 제2 필드들 모두에 대해 (F) 및 (G)를 수행하는 단계를 포함하고,

상기 유사성 스코어는 (H)로부터 획득된 복수의 내적들을 합산하는 것에 기초하여 계산되는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 3

제 1 항에 있어서,

상기 복수의 제1 필드들 중 하나 및 상기 복수의 제2 필드들 중 하나는 각각 제1 및 제2 매체 프로그램들의 캐스트(cast)들을 나타내고, 상기 복수의 제1 토큰들 및 상기 복수의 제2 토큰들은 각각 상기 제1 및 제2 매체 프로그램에서 복수의 배우들(actors)을 각각 나타내고, 상기 방법은 상기 유사성 스코어에 기초하여 사용자에게 매체 프로그램에 대한 제안을 전달하는 것을 더 포함하는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 4

제 1 항에 있어서,

각각의 필드 내의 상기 제1 및 제2 토큰들의 각각의 제1 발생 빈도를 계산하는 단계와; 그리고

서로 다른 필드들에 걸쳐 상기 제1 및 제2 토큰들의 각각의 제2 발생 빈도를 계산하는 단계를 더 포함하고,

각각의 토큰의 가중치 팩터는 상기 계산된 제1 및 제2 발생 빈도에 기초하여 할당되는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 5

제 1 항에 있어서,

선택된 토큰들에 가중치를 제공하기 위해 상기 가중치 팩터들 중 적어도 하나를 조정하는 단계를 더 포함하는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 6

제 1 항에 있어서,

(D)는 가중치들의 매트릭스를 사용하여 상기 제1 및 제2 메타데이터 사이의 서로 다른 필드들을 교차-상관 시키는 것을 포함하는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 7

제 1 항에 있어서,

복수의 매체 프로그램들의 각각과 상기 제1 매체 프로그램 사이의 유사성 스코어를 계산하는 단계와; 그리고 상기 유사성 스코어에 기초하여 상기 복수의 매체 프로그램들의 순위를 매기는 단계를 더 포함하는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 8

제 7 항에 있어서,

상기 제1 매체 프로그램에 대한 등급(rating)을 사용자로부터 수신하는 단계와; 그리고 상기 등급에 기초하여 상기 복수의 매체 프로그램들에 대한 사용자의 등급을 예측하는 단계를 더 포함하는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 9

제 1 항에 있어서,

상기 제1 및 제2 메타데이터는 실질적으로 균등한(uniform) 데이터베이스로부터 획득되는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 10

제 1 항에 있어서,

상기 제1 및 제2 메타데이터는 복수의 이질적인(heterogeneous) 데이터베이스로부터 획득되는 것을 특징으로 하는 매체 프로그램들을 관련시키는 방법.

청구항 11

명령어들을 저장하는 비-일시적인 컴퓨터-관독가능한 매체에 있어서, 상기 명령어들은 실행될 때 적어도 하나의 프로세서로 하여금 복수의 동작들을 수행하도록 하고, 상기 동작들은:

(A) 제1 매체 프로그램에 대한 제1 메타데이터 및 제2 매체 프로그램에 대한 제2 메타데이터를 획득하는 것 - 상기 제1 메타데이터는 복수의 제1 필드들로 구성되고, 상기 제2 메타데이터는 복수의 제2 필드들로 구성되고 - 과;

(B) 상기 복수의 제1 필드들의 각각으로부터 복수의 제1 토큰들 및 상기 복수의 제2 필드들의 각각으로부터 복수의 제2 토큰들을 추출하는 것과;

(C) 상기 제1 및 제2 토큰들의 각각에 가중치 팩터를 할당하는 것과;

(D) 상기 복수의 제1 필드들과 상기 복수의 제2 필드들 사이의 상기 제1 및 제2 토큰들을 교차-상관시키는 것과; 그리고

(E) 상기 교차-상관시키는 것에 기초하여 상기 제1 및 제2 매체 프로그램 사이의 유사성 스코어를 계산하는 것을 포함하는 것을 특징으로 하는 비-일시적인 컴퓨터-관독가능한 매체.

청구항 12

제 11 항에 있어서, (D)는:

(F) 상기 제1 토큰들 및 제1 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제1 필드들 중 하나를 나타내는 제1 백터를 구성하고, 상기 제2 토큰들 및 제2 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제2 필드들 중 하나를 나타내는 제2 백터를 구성하는 것과;

(G) 상기 복수의 제1 필드들 중 하나와 상기 복수의 제2 필드들 중 하나 사이의 유사성을 나타내는 상기 제1 및 제2 백터의 내적을 계산하는 것과; 그리고

(H) 상기 복수의 제1 필드들 및 복수의 제2 필드들 모두에 대해 (F) 및 (G)를 수행하는 것을 포함하고, 상기 유사성 스코어는 (H)로부터 획득된 복수의 내적들을 합산하는 것에 기초하여 계산되는 것을 특징으로 하는 비-일시적인 컴퓨터-판독가능한 매체.

청구항 13

제 12 항에 있어서,

상기 복수의 제1 필드들 및 상기 복수의 제2 필드들은 각각 상기 제1 및 제2 매체 프로그램의 캐스트, 설명 (description), 자막 또는 타이틀을 나타내는 것을 특징으로 하는 비-일시적인 컴퓨터-판독가능한 매체.

청구항 14

제 11 항에 있어서, 상기 명령어들은

각각의 필드 내의 상기 제1 및 제2 토큰들의 각각의 제1 발생 빈도를 계산하는 것과; 그리고

서로 다른 필드들에 대한 상기 제1 및 제2 토큰들의 각각의 제2 발생 빈도를 계산하는 것을 더 포함하고,

각각의 토큰의 가중치 팩터는 상기 계산된 제1 및 제2 발생 빈도에 기초하여 할당되는 것을 특징으로 하는 비-일시적인 컴퓨터-판독가능한 매체.

청구항 15

제 11 항에 있어서,

선택된 토큰들에 가중치를 제공하기 위해 상기 가중치 팩터들 중 적어도 하나를 조정하는 것을 더 포함하는 것을 특징으로 하는 비-일시적인 컴퓨터-판독가능한 매체.

청구항 16

제 11 항에 있어서,

(D)는 가중치들의 매트릭스를 사용하여 상기 제1 및 제2 메타데이터 사이의 서로 다른 필드들을 교차-상관 시키는 것을 포함하는 것을 특징으로 하는 비-일시적인 컴퓨터-판독가능한 매체.

청구항 17

적어도 하나의 처리 회로를 포함하는 시스템에 있어서,

(A) 제1 매체 프로그램에 대한 제1 메타데이터 및 제2 매체 프로그램에 대한 제2 메타데이터를 획득하고, 상기 제1 메타데이터는 복수의 제1 필드들로 구성되고, 상기 제2 메타데이터는 복수의 제2 필드들로 구성되고;

(B) 상기 복수의 제1 필드들의 각각으로부터 복수의 제1 토큰들 및 상기 복수의 제2 필드들의 각각으로부터 복수의 제2 토큰들을 추출하고;

(C) 상기 제1 및 제2 토큰들의 각각에 가중치 팩터를 할당하고;

(D) 상기 복수의 제1 필드들과 상기 복수의 제2 필드들 사이의 상기 제1 및 제2 토큰들을 교차-상관시키고;

(E) 상기 교차-상관시키는 것에 기초하여 상기 제1 및 제2 매체 프로그램 사이의 유사성 스코어를 계산하도록 된 것을 특징으로 하는 적어도 하나의 처리 회로를 포함하는 시스템.

청구항 18

제 17 항에 있어서,

상기 적어도 하나의 처리 회로는;

(F) 상기 제1 토큰들 및 제1 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제1 필드들 중 하나를 나타내는 제1 벡터, 및 상기 제2 토큰들 및 제2 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제2 필드들 중 하나를 나타내는 제2 벡터를 구성하고;

(G) 상기 복수의 제1 필드들 중 하나와 상기 복수의 제2 필드들 중 하나 사이의 유사성을 나타내는 상기 제1 및 제2 벡터의 내적을 계산하고;

(H) 상기 복수의 제1 필드들 및 복수의 제2 필드들 모두에 대해 (F) 및 (G)를 수행하도록 또한 구성되고;

상기 유사성 스코어는 (H)로부터 획득된 복수의 내적들을 합산하는 것에 기초하여 계산되는 것을 특징으로 하는 적어도 하나의 처리 회로를 포함하는 시스템.

청구항 19

제 17 항에 있어서,

상기 적어도 하나의 처리 회로는 스마트 TV 모듈을 통해 매체 프로그램의 사용자 요청을 수신하도록 또한 구성되는 것을 특징으로 하는 적어도 하나의 처리 회로를 포함하는 시스템.

청구항 20

제 19 항에 있어서,

상기 적어도 하나의 처리 회로는, 유사한 프로그램들 각각의 계산된 유사성 스코어들에 기초하여 상기 스마트 TV 모듈 상에 복수의 유사한 프로그램들을 디스플레이하도록 또한 구성되는 것을 특징으로 하는 적어도 하나의 처리 회로를 포함하는 시스템.

명세서

기술분야

[0001] 본 출원은, 2011년 10월 30일에 출원된 미국 가특허 출원 번호 61/553,221의 이득을 청구하는, 2012년 5월 9일에 출원된 미국 특허 출원 번호 13/467,339의 우선권을 주장한다. 미국 특허 출원 번호 13/467,339 및 미국 가특허 출원 번호 61/553,221 모두 참조로써 본 명세서에 포함된다.

배경기술

[0002] 본 발명은 일반적으로 매체 프로그램들, 특히, 텔레비전 프로그램들 사이의 관계들을 결정하는 것에 관한 것이다.

발명의 내용

해결하려는 과제

[0003] 인터넷은 비디오 파일들, 오디오 파일들, 사진들, 사업 및 개인 연락처 정보, 제품 정보, 지도들 및 뉴스 기사들과 같은 매우 다양한 소스들로의 접근을 제공한다. 이러한 소스들의 접근은 광고 기회들과 같은 사업 기회들을 만들어 낼 수 있다. 인터넷 초창기에서는 텍스트 데이터가 일반적이었지만, 비디오 및 오디오 파일들과 같은 매체 프로그램들은 대역폭의 증가로 인해 더 인기를 얻고 있다.

과제의 해결 수단

[0004] 일 양상에서, 컴퓨터로 구현되는, 매체 프로그램들을 관련(associate)시키는 방법이 제공된다. 상기 방법은: (A) 적어도 하나의 처리 회로를 사용하여, 제1 매체 프로그램에 대한 제1 메타데이터 및 제2 매체 프로그램에 대한 제2 메타데이터를 획득하는 단계 - 상기 제1 메타데이터는 복수의 제1 필드들로 구성되고, 상기 제2 메타데이터는 복수의 제2 필드들로 구성되고 - 와; (B) 적어도 하나의 처리 회로를 사용하여, 상기 복수의 제1 필드들의 각각으로부터 복수의 제1 토큰들을 추출하고 상기 복수의 제2 필드들의 각각으로부터 복수의 제2 토큰들을

추출하는 단계와; (C) 상기 제1 및 제2 토큰들의 각각에 가중치 팩터를 할당하는 단계와; (D) 상기 복수의 제1 필드들과 상기 복수의 제2 필드들 사이의 상기 제1 및 제2 토큰들을 교차-상관(cross-correlating)시키는 단계와; 그리고 (E) 상기 교차-상관시키는 단계에 기초하여 상기 제1 및 제2 매체 프로그램 사이의 유사성 스코어(similarity score)를 계산시키는 단계를 포함할 수 있다.

- [0005] 일 실시예에서, (D)는 (F) 상기 제1 토큰들 및 제1 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제1 필드들 중 하나를 나타내는 제1 벡터를 구성하고 상기 제2 토큰들 및 제2 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제2 필드들 중 하나를 나타내는 제2 벡터를 구성하는 단계와; (G) 상기 복수의 제1 필드들 중 하나와 상기 복수의 제2 필드들 중 하나 사이의 유사성을 나타내는 상기 제1 및 제2 벡터의 내적(dot product)을 계산하는 단계와; 그리고 (H) 상기 복수의 제1 필드들 및 복수의 제2 필드들 모두에 대해 (F) 및 (G)를 수행하는 단계를 포함하고, 상기 유사성 스코어는 (H)로부터 획득된 복수의 내적들을 합산하는 것에 기초하여 계산된다.
- [0006] 일 실시예에서, 상기 복수의 제1 필드들 중 하나 및 상기 복수의 제2 필드들 중 하나는 각각 제1 및 제2 매체 프로그램들의 캐스트(cast)들을 나타내고, 상기 복수의 제1 토큰들 및 상기 복수의 제2 토큰들은 각각 상기 제1 및 제2 매체 프로그램에서 복수의 배우들(actors)을 각각 나타낸다. 상기 방법은 상기 유사성 스코어에 기초하여 사용자에게 매체 프로그램에 대한 제안을 전달하는 것을 더 포함할 수 있다.
- [0007] 일 실시예에서, 상기 방법은 각각의 필드 내의 상기 제1 및 제2 토큰들의 각각의 제1 발생 빈도를 계산하는 단계와; 그리고 서로 다른 필드들에 걸쳐 상기 제1 및 제2 토큰들의 각각의 제2 발생 빈도를 계산하는 단계를 포함하며, 각각의 토큰의 가중치 팩터는 상기 계산된 제1 및 제2 발생 빈도에 기초하여 할당될 수 있다.
- [0008] 일 실시예에서, 상기 방법은 선택된 토큰들에 가중치를 제공하기 위해 상기 가중치 팩터들 중 적어도 하나를 조정하는 단계를 포함한다.
- [0009] 일 실시예에서, 상기 방법은 가중치들의 매트릭스를 사용하여 상기 제1 및 제2 메타데이터 사이의 서로 다른 필드들을 교차-상관시키는 것을 포함한다.
- [0010] 일 실시예에서, 상기 방법은 복수의 매체 프로그램들의 각각과 상기 제1 매체 프로그램 사이의 유사성 스코어를 계산하는 단계를 포함한다. 일 실시예에서, 상기 방법은 상기 유사성 스코어에 기초하여 상기 복수의 매체 프로그램들의 순위를 매기는 단계를 더 포함한다.
- [0011] 일 실시예에서, 상기 방법은 상기 제1 매체 프로그램에 대한 등급(rating)을 사용자로부터 수신하는 단계와; 그리고 상기 등급에 기초하여 상기 복수의 매체 프로그램들에 대한 사용자의 등급을 예측하는 단계를 포함한다.
- [0012] 일 실시예에서, 상기 제1 및 제2 메타데이터는 실질적으로 균등한(uniform) 데이터베이스로부터 획득된다.
- [0013] 다른 실시예에서, 상기 제1 및 제2 메타데이터는 복수의 이질적인(heterogeneous) 데이터베이스로부터 획득된다.
- [0014] 다른 양상에서, 명령어들을 저장하는 비-일시적인 컴퓨터-판독가능한 매체가 제공된다. 상기 명령어들은 실행될 때 적어도 하나의 프로세서로 하여금 복수의 동작들을 수행하도록 할 수 있고, 상기 동작들은: (A) 제1 매체 프로그램에 대한 제1 메타데이터 및 제2 매체 프로그램에 대한 제2 메타데이터를 획득하는 것 - 상기 제1 메타데이터는 복수의 제1 필드들로 구성되고, 상기 제2 메타데이터는 복수의 제2 필드들로 구성되고 - 과; (B) 상기 복수의 제1 필드들의 각각으로부터 복수의 제1 토큰들 및 상기 복수의 제2 필드들의 각각으로부터 복수의 제2 토큰들을 추출하는 것과; (C) 상기 제1 및 제2 토큰들의 각각에 가중치 팩터를 할당하는 것과; (D) 상기 복수의 제1 필드들과 상기 복수의 제2 필드들 사이의 상기 제1 및 제2 토큰들을 교차-상관시키는 것과; 그리고 (E) 상기 교차-상관시키는 것에 기초하여 상기 제1 및 제2 매체 프로그램 사이의 유사성 스코어를 계산하는 것을 포함한다.
- [0015] 일 실시예에서, (D)는 (F) 상기 제1 토큰들 및 제1 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제1 필드들 중 하나를 나타내는 제1 벡터를 구성하고 상기 제2 토큰들 및 제2 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제2 필드들 중 하나를 나타내는 제2 벡터를 구성하는 것과; (G) 상기 복수의 제1 필드들 중 하나와 상기 복수의 제2 필드들 중 하나 사이의 유사성을 나타내는 상기 제1 및 제2 벡터의 내적을 계산하는 것과; 그리고 (H) 상기 복수의 제1 필드들 및 복수의 제2 필드들 모두에 대해 (F) 및 (G)를 수행하는 것을 포함하고, 상기 유사성 스코어는 (H)로부터 획득된 복수의 내적들을 합산하는 것에 기초하여 계산된다.
- [0016] 일 실시예에서, 상기 복수의 제1 필드들 및 상기 복수의 제2 필드들은 각각 상기 제1 및 제2 매체 프로그램의

캐스트, 설명(description), 자막 또는 타이틀을 나타낸다.

- [0017] 일 실시예에서, 상기 명령어들은 각각의 필드 내의 상기 제1 및 제2 토큰들의 각각의 제1 발생 빈도를 계산하는 것과; 그리고 서로 다른 필드들에 대한 상기 제1 및 제2 토큰들의 각각의 제2 발생 빈도를 계산하는 것을 더 포함하고, 각각의 토큰의 가중치 팩터는 상기 계산된 제1 및 제2 발생 빈도에 기초하여 할당된다.
- [0018] 일 실시예에서, 상기 명령어들은 선택된 토큰들에 가중치를 제공하기 위해 상기 가중치 팩터들 중 적어도 하나를 조정하는 것을 포함한다.
- [0019] 일 실시예에서, 상기 명령어들은 가중치들의 매트릭스를 사용하여 상기 제1 및 제2 메타데이터 사이의 서로 다른 필드들을 교차-상관 시키는 것을 포함한다.
- [0020] 다른 양상에서, 적어도 하나의 처리 회로를 포함하는 시스템이 제공된다. 상기 적어도 하나의 처리 회로는 (A) 제1 매체 프로그램에 대한 제1 메타데이터 및 제2 매체 프로그램에 대한 제2 메타데이터를 획득하고, 상기 제1 메타데이터는 복수의 제1 필드들로 구성되고, 상기 제2 메타데이터는 복수의 제2 필드들로 구성되고; (B) 상기 복수의 제1 필드들의 각각으로부터 복수의 제1 토큰들 및 상기 복수의 제2 필드들의 각각으로부터 복수의 제2 토큰들을 추출하고; (C) 상기 제1 및 제2 토큰들의 각각에 가중치 팩터를 할당하고; (D) 상기 복수의 제1 필드들과 상기 복수의 제2 필드들 사이의 상기 제1 및 제2 토큰들을 교차-상관시키고; 그리고 (E) 상기 교차-상관시키는 것에 기초하여 상기 제1 및 제2 매체 프로그램 사이의 유사성 스코어를 계산하도록 구성될 수 있다.
- [0021] 일 실시예에서, 상기 적어도 하나의 처리 회로는 (F) 상기 제1 토큰들 및 제1 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제1 필드들 중 하나를 나타내는 제1 벡터, 및 상기 제2 토큰들 및 제2 토큰들 각각의 가중치 팩터들에 기초하여 상기 복수의 제2 필드들 중 하나를 나타내는 제2 벡터를 구성하고; (G) 상기 복수의 제1 필드들 중 하나와 상기 복수의 제2 필드들 중 하나 사이의 유사성을 나타내는 상기 제1 및 제2 벡터의 내적을 계산하고; (H) 상기 복수의 제1 필드들 및 복수의 제2 필드들 모두에 대해 (F) 및 (G)를 수행하도록 또한 구성되고; 상기 유사성 스코어는 (H)로부터 획득된 복수의 내적들을 합산하는 것에 기초하여 계산된다.
- [0022] 일 실시예에서, 상기 적어도 하나의 처리 회로는 스마트 TV 모듈을 통해 매체 프로그램의 사용자 요청을 수신하도록 구성된다.
- [0023] 일 실시예에서, 상기 적어도 하나의 처리 회로는, 유사한 프로그램들 각각으로 계산된 유사성 스코어들에 기초하여 상기 스마트 TV 모듈 상에 복수의 유사한 프로그램들을 디스플레이하도록 구성된다.
- [0024] 이전의 개념들 및 아래에서 더 상세하게 논의되는 추가 개념들의 모든 조합들은 (그와 같은 개념들이 상호 불일치하지 않으면) 본 명세서에서 개시된 본원 발명의 부분으로서 고려된다. 특히, 본 명세서의 말미에서 나타나고 있는 청구된 본 발명의 모든 조합들은 본 명세서에서 개시된 발명의 일 부분으로서 고려된다. 또한 참조로서 포함되는 임의의 개시에서 표시될 수 있고 본 명세서에서 명시적으로 사용되는 용어는, 본 명세서에서 개시되는 특정 개념들과 가장 일치하는 의미에 부합됨을 인식해야 한다.
- [0025] 본 명세서의 앞선 그리고 다른 양상들, 실시예들 및 특징들이 첨부된 도면들과 결합하여 아래의 상세한 설명으로부터 더 충분히 이해될 수 있다.

도면의 간단한 설명

- [0026] 당업자는, 본 명세서에서 서술된 도면들이 단지 설명의 목적임을 이해할 것이다. 일부 예들에서, 본 발명의 여러 양상들은 본 발명의 이해를 용이하게 하도록 과장되거나 확장되어 도시될 수 있음을 이해해야 한다. 도면들에서, 동일한 참조 번호들은 일반적으로 여러 도면들을 통해 유사한 특징들, 기능적으로 유사하고 그리고/또는 구조적으로 유사한 요소들 언급한다. 도면들은 본 명세서의 원리들을 설명하는 대신에 크기를 조정하거나, 강조할 필요는 없다. 도면들은 본 발명의 범위를 어떤 방식으로든 제한하도록 의도되지 않는다.
- 도 1은 스마트 TV 인터페이스의 예를 나타내는 블록도이다.
- 도 2는 매체 프로그램들 사이의 유사성들을 비교하는 예시적인 방법을 나타내는 흐름도이다.
- 도 3은 벡터 표현을 사용하여 서로 다른 매체 프로그램들의 필드들 사이의 상관관계들을 계산하는 예시적인 방법을 나타내는 다이어그램이다.
- 도 4는 교차-상관시키는 서로 다른 필드들에 대한 가중치 팩터들의 매트릭스의 예를 나타내는 다이어그램이다.

도 5는 사용자에게 프로그램들을 추천하는 예시적인 방법을 나타내는 흐름도이다.

도 6은 프로그램의 인기를 예측하는 예시적인 방법을 나타내는 흐름도이다.

도 7은 여러 동작들 중 적어도 일부를 수행하기 위해 사용될 수 있는 시스템을 나타내는 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0027] 매체 프로그램들을 관련시키는 방법들 및 시스템들에 관련된 여러 개념들 및 실시예들의 상세한 설명이 아래에서 서술된다. 위에서 도입되고 아래에서 더 상세하게 논의되는 여러 개념들은 임의의 많은 방식으로 구현될 수 있고, 이는 개시된 개념들이 어떤 특정 실시 방식으로 제한되지 않음이 인식되어야 한다. 특정 실시예들 및 애플리케이션들의 예시들은 주로 설명의 목적으로 제공된다.

[0028] **개관**

[0029] 본 명세서에서 개시되는 실시예들은 인터넷을 통해 매체 프로그램들을 관련시키기 위해 사용될 수 있다. 비디오 및 오디오 프로그램들과 같은 그러한 매체 프로그램들은 인터넷 텔레비전(TV) 또는 웹 TV 상에서, 또는 스마트 TV 시스템으로 디스플레이될 수 있다.

[0030] 인터넷 TV 또는 웹 TV는 통상적으로 인터넷을 통해 매체 프로그램들을 스트리밍하거나 다운로드하도록 개인용 컴퓨터 또는 모바일 디바이스를 통상적으로 사용할 수 있다. 스마트 TV 시스템에서, 스마트 TV 모듈(또는 접속된 TV 모듈, 하이브리드 TV 모듈 등)은 종래의 TV 세트를 인터넷과 통합시킬 수 있고, 사용자들에게 더 통합적인 경험을 전달할 수 있다. 사용자는 스마트 TV 상에서 보고, 인터넷을 통해 게임들을 하고, 웹 페이지들을 브라우징하기 위해 요구된 콘텐츠를 선택할 수 있고, 상기 콘텐츠가 종래 TV의 방송에서 오는지 또는 상기 인터넷으로부터 오는지에 대해서 반드시 알 필요는 없다.

[0031] 상기 스마트 TV는 (예를 들어, 케이블, 위성, 무선 또는 다른 신호들을 통해 수신되는) 더 전통적인 TV 프로그래밍 소스들과 인터넷 접속을 통합시키기 위해 구성되는 처리 회로를 포함할 수 있다. 상기 스마트 TV 모듈은 물리적으로 TV 세트에 통합될 수 있거나 셋-톱 박스, 블루-레이 또는 디지털 매체 플레이어, 게임 콘솔, 호텔 TV 시스템, 및 다른 컴패니언 디바이스(companion device)와 같은 개별 디바이스를 포함할 수 있다.

[0032] 스마트 TV 모듈은 사용자들로 하여금 웹 상에서, 로컬 케이블 TV 채널에서, 위성 TV 채널에서, 또는 로컬 저장 디바이스 상에 저장되는, 비디오들, 영화들, 사진들 및 다른 콘텐츠를 검색하고 발견할 수 있도록 구성될 수 있다. 셋-톱 박스(STB) 또는 셋-톱 유닛(STU)은 튜너를 포함하고 TV 세트 및 신호의 외부 소스에 접속할 수 있고, 이후 TV 스크린 또는 디스플레이 디바이스 상에서 디스플레이되는 콘텐츠를 상기 신호로 변환시키는 정보 기기 장치를 포함할 수 있다.

[0033] 도 1에서 도시된 것처럼, 스마트 TV 모듈은 홈 스크린(100), 또는 웹 브라우저와 같은 복수의 서로 다른 애플리케이션들 및 복수의 스트리밍 매체 서비스들(예를 들어, Netflix, Vudu, Hulu 등), 접속된 케이블 또는 위성 매체 소스, 다른 웹 "채널들(channels)" 등을 위한 아이콘들을 포함하는 톱 레벨 스크린을 제공하도록 구성될 수 있다. 상기 스마트 TV 모듈은 또한 전자 프로그래밍 가이드를 사용자에게 제공하기 위해 구성될 수 있다. 스마트 TV 모듈에 대한 컴패니언 디바이스는, 사용자로 하여금 상기 스마트 TV 모듈들을 제어하도록 사용자에게 이용가능한 프로그램들에 대한 추가적인 정보를 제공하기 위해 모바일 컴퓨팅 디바이스 상에서 동작가능할 수 있다. 키보드와 같은 입력 디바이스는, 사용자로 하여금 종래의 개인용 컴퓨터에서처럼 길이가 긴 명령어들을 입력하도록 포함될 수 있다.

[0034] 텍스트 데이터 사이의 상관관계들은, 키워드 기반 검색들과 같은 인터넷에서의 다수의 애플리케이션과 문맥 및 타겟 광고를 찾아내고 있다. 오디오 및 비디오 프로그램들과 매체 프로그램들은 상관되거나 수량화되기는 더 어렵다. 본 발명에서 개시된 실시예들은, TV 프로그램들, 영화들, 시리즈들, 쇼들, 음악 및 다른 데이터와 같은 비-텍스트 데이터를 포함하는 매체 프로그램들 사이에서 유사성(similarity)들이 계산되는 것을 허용한다.

[0035] **매체 프로그램들에 대한 메타데이터**

[0036] 하나의 실시예에서, 유사성은 이러한 프로그램들에 대한 이용가능한 메타데이터를 비교하는 것에 기초하여 2개의 프로그램들 사이에서 계산된다. 예시적인 메타데이터는 캐스트의 리스트, 촬영팀, 프로그램 소요시간, 카테

고리들, 장르, 자막 또는 부제들, 시놉시스들, 등급들 등을 포함할 수 있다. 일부 예시들에서, 인터넷 영화 데이터베이스(IMDb) 또는 영화 리뷰 웹 사이트들과 같은 웹 소스들, TMS(Tribune Media Services), 또는 다른 이용가능한 소스들로부터 획득될 수 있다.

[0037] 상기 메타데이터는 여러 형태들로 이용가능할 수 있고, 여러 필드들로 구성될 수 있다. 예를 들어, 영화에 대한 특정 소스들로부터의 메타데이터는 프로그램 제목, 설명, 줄거리, 캐스트, 촬영팀, 장르 등과 같은 필드들을 포함할 수 있다. 다른 예에서, 메타데이터는 제목, 주요 콘텐츠, 캐스트, 카테고리들, 잡다한 것 같은 필드들을 포함하는 웹 소스, 동일한 웹 소스 내의 다른 기사(article)들로의 내부 링크들 또는 다른 기사들로부터의 내부 링크들, 및 다른 웹 소스들로 외부 링크들로부터 획득될 수 있다.

[0038] 일 실시예에서, 서로 다른 프로그램들과 같은 메타데이터는 동일한 웹 소스 내의 내부 링크들을 통해 획득될 수 있고, 그 결과 상대적으로 균등한 데이터 소스를 제공한다. 다른 실시예에서, 데이터가 실질적으로 균등한 표준 없이 복수의 웹 소스들/사이트들로부터 온 것 같이, 데이터 품질 및 형태가 더 이질적일 수 있지만, 외부 링크들은 데이터 관련(data association)에 포함될 수 있다. 또 다른 실시예에서, 데이터 마이닝(data mining)은 검색 엔진을 통해 제공되고 WWW(world wide web)을 통해 행해질 수 있다.

[0039] **예시적인 방법**

[0040] 일 실시예에서, 2개의 프로그램들을 비교하는 방법(200)은, 도 2의 흐름도에서 도시된 것처럼 제공될 수 있다. 상기 방법(200)은 다음의 동작들을 포함할 수 있다.

[0041] 동작 202에서, 제1 매체 프로그램에 대한 제1 메타데이터 및 제2 매체 프로그램에 대한 제2 메타데이터는 처리 회로를 사용하는 분석을 위해 서버에서 수신될 수 있다. 일 예에서, 제1 및 제2 프로그램들은 2개의 영화들이 될 수 있고, 상기 제1 및 제2 메타데이터는 이러한 2개의 영화들에 대한 대응하는 2개의 기사들/리뷰들일 수 있다. 상기 메타데이터 각각은 복수의 필드들, 예를 들어, 주요 콘텐츠 또는 주요 설명, 캐스트, 제작, 줄거리, 사운드트랙 설명, 배급, 리셉션, 흥행 성과, 수상 등으로 체계화된다.

[0042] 다음, 동작 204에서, 복수의 토큰들은 각각의 프로그램의 메타데이터 내의 각각의 필드로부터 추출될 수 있다. 토큰은 가장 빈번하게 사용되는 용어들일 수 있다. 각각의 필드의 텍스트는 단일 단어, 그리고/또는 2개 이상의 단어들을 포함하는 구(phrase)들로서 인식될 수 있다. 예를 들어, 2개의 인접한 단어들 "뉴(New)" 및 "욕(York)"은 하나의 구 "뉴욕(New York)"으로 분석되고, 하나의 토큰으로서 인식될 수 있다.

[0043] 일 예에서, "캐스트(cast)"의 필드에서, 더 잘 알려진 배우들 각각은 토큰으로서 처리될 수 있다. 일 예에서, 사람의 이름들을 포함하는 필드에 대해, 이름과 성은 단일 토큰으로 결합될 수 있다. 설명을 위해, 단어들이 한 번보다 많게, 바람직하게는 미리결정된 수의 경우들보다 많게 텍스트에서 함께 나타나면, 이러한 단어들은 구들로 결합될 수 있다.

[0044] 일 실시예에서, 아래에서 서술되는 계산 동작들의 속도를 높이기 위해, 키워드들은 분석되는 프로그램들의 세트에 대해 고유한 번호들과 같은 고유한 식별자들로서 세계적으로 전환될 수 있다.

[0045] 각각에 필드에 대해, 알고리즘은 텍스트로부터 토큰들을 추출하기 위해 채택될 수 있다. 서로 다른 알고리즘들은, 토큰화를 위해 주요 콘텐츠, 타이틀, 캐스트 및 사소한 정보와 같은 다른 필드들에 대해 사용될 수 있다. 숫자 관련 팩터는 프로그램 유사성들을 계산함에 있어 관련성의 레벨을 나타내는 각각의 토큰에 대해 주어질 수 있다.

[0046] 메타데이터 소스들의 주요 콘텐츠(예를 들어, TMS 또는 다른 소스들로부터 획득된 것과 같은 영화들의 설명들)은 종종 많은 양의 텍스트 데이터를 포함할 수 있고, 문장들, 문장의 말단들(sentence terminals), 그리고/또는 클러스터(cluster)들로 분할될 수 있다.

[0047] 말단들은 상기 텍스트에서 발생하는 단어들, 그리고/또는 말단들이 유행어들을 형성하면 함께 결합된 단어들이 될 수 있다. 하나의 실시예에서, 모든 말단들은 토큰들로서 포함될 수 있다. 그것들의 숫자와 관련된 값은, 예를 들어, 1로서 주어질 수 있다.

[0048] 클러스터들은 상기 텍스트에 주제별로 접속되는 단어들의 그룹들일 수 있다. 예를 들어, 영화 설명의 콘텐츠는 단어 클러스터들을 사용하여 결정될 수 있다. 일반적으로, 단어 클러스터는 동일한 또는 유사한 생각을 전달하는 단어들의 세트가 될 수 있다. 단어 클러스터는 하나의 실시예에 따른 동의어들일 수 있다. 그와 같은 클러스

터는 동일한 주제에 속한 영화들을 식별하기 위해 사용될 수 있지만, 그렇게 하기 위해 서로 다른 용어들을 사용한다. 일부 경우들에서, 단어 클러스터는 관련되지만, 서로 다른 의미들을 갖는 단어들을 포함할 수 있다. 일부 실시예들에서, 텍스트 설명의 특성은 서로 다른 단어 클러스터들의 세트일 수 있다. 각각의 클러스터에 대해, 상기 매체 프로그램으로의 접속의 정도를 나타내는 숫자 값이 또한 주어질 수 있다. 일 예에서, 특정 임계치를 넘는 숫자 값을 갖는 그와 같은 클러스터들은 토큰들로서 포함될 수 있다.

- [0049] 프로그램 타이틀들은 문장들 또는 구들로 그룹화하는 충분한 텍스트를 가지지 못할 수 있다. 따라서, 상기 타이틀들로부터의 텍스트 데이터는 개별적인 단어들로 (예를 들어, 여백에 의해) 분할될 수 있고, 그리고 각각의 단어는 토큰으로 포함되고, 1이라는 관련 숫자 값으로 주어질 수 있는데, 이는 타이틀들이 일반적으로 매체 프로그램들에 중요하기 때문이다.
- [0050] 일 실시예에서, (TMS 캐스트 데이터로부터 획득된 것과 같은) 상기 캐스트 리스트 상의 매(every) 아이템은 토큰이 될 수 있다. 상기 캐스트 리스트는 중요성에 따라 구성될 수 있다. 따라서, 일 예에서, 주연인 여배우의 관련 스코어가 1로 설정될 수 있다. 상기 리스트에서 다른 배우들은, 예를 들어, 고정된 팩터에 의해 이전보다 더 낮은 스코어들을 받을 수 있다.
- [0051] 상기 토큰화는 수학적 모델들을 통해 자동적으로 하나 이상의 처리 회로 부분들을 사용하여 수행될 수 있다. 상기 수학적 모델들은 단어 클러스터들에 기초한 확률적 토픽 모델(probabilistic topic model)들을 포함할 수 있다.
- [0052] 제1 예에서, 확률적 토픽 모델은 TMS "설명(description)" 데이터 또는 다른 소스들로부터의 다른 영화 설명들에 적용될 수 있다. 상기 확률적 토픽 모델은 큰 수(예를 들어, 12백만)의 알려진 용어들 및 큰 수(예를 들어, 1백만)의 개념들, 그리고 그것들 사이의 확률적 관계들을 포함할 수 있다. 상기 TMS "설명"으로부터의 텍스트들은 토큰화될 일반적인 토픽 모델에 입력될 수 있다.
- [0053] 일 실시예에서, 하기의 프로그램의 메타데이터 필드들: 타이틀, 설명, 캐스트, 장르 및 감독이 토큰화된다. 매 토큰은 프로그램과의 관련성을 나타내기 위해 0과 1사이의 숫자 가중치 팩터가 할당될 수 있다. 상기 숫자 가중치 팩터들은 아래에서 서술된 계산들에서 사용될 수 있다.
- [0054] 상기 메타데이터 내의 다른 필드는 상기 소스에 의해 제공된 카테고리로서 정의될 수 있다. 일 실시예에서, 각 카테고리는 해석을 위한 일부 미리결정된 규칙들을 통해 전달된다. 일부 카테고리들은 더 넓은 카테고리들에 변경될 수 있고, 그들의 가중치 팩터들이 조정될 수 있다. 일부 카테고리들은 변경되지 않고, 1의 가중치 팩터가 주어질 수 있다.
- [0055] 매 토큰은 필드 내에서 가중치가 할당될 수 있다. 상기 가중치는 상기 토큰이 발생하는 프로그램의 관련성 및 전체 수에 의존할 수 있다. 일부 경우들에서, 토큰은 하나의 필드 내에서 다수회 발생할 수 있다. 하나의 실시예에 따라, 가중치 팩터들은 서브-선형 함수에 따라 누적된다. 예를 들어, 상기 누적된 가중치 팩터는 개별적인 관련 스코어들의 합보다 훨씬 작을 수 있다. 일 실시예에서, 토큰들은 서로 다른 필드들에 걸쳐 누적되지 않는다.
- [0056] 동작 206에서, 더 적은 의미를 갖는 토큰들은 가외치(outlier)들로서 제거될 수 있다. 이러한 가외치들은 필드에서 빈번하게 중첩하여 발생하거나 상기 메타데이터에서 전체적으로 발생하는 토큰들을 포함할 수 있다.
- [0057] 가외치임에도 불구하고, 일반적으로 토큰들의 발생하는 수 또는 빈도는 상기 프로그램의 관련성에 대한 속성들을 나타낼 수 있다. 일 실시예에서, 수정된 TF-IDF(term frequency-inverse document frequency) 방법이 예시적인 동작 208에서 토큰들의 가중치 팩터들을 계산하기 위해 채택될 수 있다. 상기 TF는 각 필드 내의 토큰의 발생들의 빈도를 나타내고, 상기 IDF는 전체 문서에서와 같이, 서로 다른 필드들에 걸쳐 토큰이 발생하는 빈도를 나타낼 수 있다.
- [0058] 구체적으로, 토큰 각각에 대해, TF(term frequency)가 계산된다. 서로 다른 필드 타입들에 대한 서로 다른 계산 방법들이 채택될 수 있다. 일 예에서, TF는 $TF = 1 + \log(x)$ 로서 계산되고, x는 상기 메타데이터의 특정 필드에서의 토큰이 발생하는 수이다.
- [0059] IDF(Inverse Document Frequency)는 또한 필드들 또는 문서들에 걸쳐 계산될 수 있다. 이 빈도는 각각의 필드-토큰 쌍에 대해 전반적일 수 있다. 일 실시예에서, 상기 IDF를 계산하는 방법은 각각의 필드 타입(예를 들어, 타이틀, 캐스트 등)에 대해 또한 개별적으로 설정된다. $IDF = 1 - \log(count)/\log(max_count+1)$ 이고, 상기 카운트는 상기 토큰이 발생한 문서들/필드들의 수이고, max_count 는 이러한 수들의 최대치이다.

- [0060] 일 실시예에서, 토큰의 가중치 벡터는, TF 및 IDF의 곱셈과 같이 계산된 TF 및 IDF에 기초한다.
- [0061] 일 실시예에서, 필드 각각은 i 번째 컴포넌트는 토큰 i (일반적인 지수)의 가중치가 될 수 있다. 이러한 벡터는 1의 길이를 갖도록 정규화될 수 있다. 2개의 필드들 사이의 유사성은 2개의 정규화된 벡터들의 내적(dot product)(또는 스칼라 곱(scalar product), 내적(inner product))과 같은 2개의 벡터들 사이의 상관 관계에 의해 표시될 수 있다.
- [0062] 상기 다차원 공간에서의 벡터들 및 그것들 사이의 내적들은 단지 다수의 컴포넌트들(토큰들)의 필드들 사이의 관계들 또는 상관관계들의 한가지 타입의 수학적 표현이다. 다른 수학적인 방법들 및 공식은 그와 같은 관계들을 계산하기 위해 사용될 수 있다. 일반적으로, 복수의 제1 필드들 및 복수의 제2 필드들 사이의 제1 및 제2 토큰들의 교차 상관관계들이 수행될 수 있고, 상기 제1 및 제2 매체 프로그램들 사이의 유사성 스코어는 교차 상관관계들에 기초할 수 있다.
- [0063] 구체적으로, 도 3에 도시된 것처럼, 알려진 각각의 토큰의 가중치 벡터로 토큰화 처리를 한 후, 복수의 필드 벡터들은 각각의 프로그램에 대해 구성될 수 있다. 각각의 필드 벡터는 필드를 나타낼 수 있고 상기 필드 및 그것들의 가중치 벡터들로부터 추출된 상기 토큰들에 기초하여 구성될 수 있다.
- [0064] 예를 들어, 필드 벡터 A는 제1 영화에 대한 "캐스트(cast)" 필드에 대해 구성될 수 있다. 이러한 예에서, 캐스트 벡터 A는, 제1 배우의 이름을 나타내는 토큰 1, 제2 배우의 이름을 나타내는 토큰 2, 및 제3 배우의 이름을 나타내는 토큰 3과 같은, 3개의 토큰 축들에 컴포넌트들을 가질 수 있다. 일부 실시예들에서, 벡터는 더 많거나 더 적은 컴포넌트들을 가질 수 있다. 필드 벡터 B는 상기 제2 영화에 대한 "캐스트(cast)" 필드에 대해 구성될 수 있고, 여기서 토큰들 1', 2' 및 3'은 제2 영화의 캐스트에서 3명의 배우들을 나타낸다.
- [0065] 토큰 1 및 토큰 1'의 정렬은, 상기 제1 배우는 상기 2개의 영화들에서 공통임을 나타낸다. 서로 다른 배우들은 직교 축들에 의해 표현된다. 도 3에서 도시된 예에서 필드 벡터들 A 및 B에 대한 좌표 시스템들이 정렬되지만, 많은 다른 경우들에서 상기 좌표 시스템들은 다차원 벡터 공간에서 서로 완전히 직교하고, 이는 어떤 공통의 배우들도 비교되고 있는 2개의 프로그램들에서 전혀 존재하지 않음을 나타낸다.
- [0066] 상기 벡터들은 2개의 토큰 리스트들에 대해 정규화 벡터들을 사용하여 정규화될 수 있다. 하나의 실시예에서, 각각의 벡터는, 예를 들어 상기 벡터의 기준(norm)을 사용하여 분할함으로써 벡터의 길이들로 정규화될 수 있다.
- [0067] 일부 실시예들에서, 상기 정규화 벡터는, 예를 들어, 더 많은 토큰들로 이러한 벡터들을 신장시키고, 더 짧은 텍스트들을 축진시키고, 그리고/또는 특정 토큰들을 축진시키기 위해 하나 이상의 선택된 컴포넌트들의 척을 곱함으로써 수정된다.
- [0068] 일반적으로, 상기 벡터들은 0 ~ 1의 범위로 정규화될 수 있다. 일 예에서, 0 ~ 1(예를 들어, 1/3)의 지수 멱(exponential power)은, 예를 들어, TF-IDF 계산들에서 선택한 토큰들에 주어질 수 있다. 당업자들은, 다른 정규화들이 상기 처리들에서 사용될 수 있음을 인식할 것이다.
- [0069] 다음, 동작 210에서, 2개의 벡터들 A 및 B의 내적이 계산될 수 있다. 도 3에서의 필드 벡터들 A 및 B의 정렬된 축들은, 상기 내적이 높은 값을 가질 것을 나타내고, 이는 (적어도 캐스트의 측면에서) 2개의 영화들 사이에서 높은 정도의 유사성들을 나타낸다. 일부 다른 예들에서, 2개의 영화들에서 공통된 배우들은 없고, 상기 캐스트 벡터들에서의 모든 컴포넌트들은 수직이며, 상기 내적은 0이고, 이는 (적어도 상기 캐스트의 측면에서) 상기 2개의 영화들 사이에서 어떠한 유사성들도 없음을 나타낸다.
- [0070] 다음, 동작 214에서, 모든 필드 쌍들로부터의 내적들은 유사성 스코어를 계산하기 위해 합산될 수 있다.
- [0071] 일부 실시예에서, 복수의 설정들이 계산에서 포함될 수 있다. 그와 같은 설정들은, 예를 들어, (동일한 타입들 사이의 비교들을 포함하는) 여러 필드 타입들로부터의 토큰들의 비교가 얼마나 많이 최종 스코어에 영향을 미치는지를 나타내는 가중치들의 매트릭스(실수(real number)들), 및 각각의 필드 타입에 대해 사용될 적절한 TF 및 IDF 함수들을 포함할 수 있다.
- [0072] 2개의 프로그램들 사이의 유사성들은 개별적인 필드들의 쌍들의 유사성들에 기초할 수 있다. 하나의 실시예에서, 모든 필드들의 쌍들(이들중 많은 것이 0과 같을 수 있다)에 대해 스코어 승수들을 정의하는 매트릭스가 제공된다. 예를 들어, 프로그램 X의 "설명(description)" 필드는 프로그램 Y의 "설명" 필드와 직접 비교될 수 있고, 상기 대응하는 매트릭스 컴포넌트는 1의 값을 가질 수 있다. 또한, 상기 "설명"은 "타이틀" 및 "줄거리" 등과 같은 다른 필드들과 부분적으로 비교될 수 있고, 대응하는 매트릭스 컴포넌트들은 비-제로(non-

zero)이다. 상기 비-제로 어크로스-필드 매트릭스 컴포넌트들은, 유사성 스코어들을 계산함에 있어 일정 정도의 필드 믹싱이 유용할 수 있음을 나타낸다.

[0073] 하지만, 일부 필드 믹싱은 비교 계산들에서 유용하지 않을 수 있다. 상기 설명은 일반적으로 "장르"와 비교될 수 없고, 따라서 상기 대응하는 매트릭스 컴포넌트들은 단지 0일 수 있다.

[0074] 예시적인 매트릭스 $[M]$ 이 도 3에서 도시되고, 상기 매트릭스 컴포넌트들의 수치값들이 통계적인 샘플링 처리에서의 결과들 또는 경험적으로 제공된 결과들을 을 최적화하기 위해 여러 결합들을 테스트함으로써 도출된다. 도시된 예에서, $[M]$ 은 대칭 매트릭스가 될 수 있다.

[0075] 따라서, 하나의 실시예에서, 2개의 매체 프로그램들 사이의 최종 유사성 스코어는 도 3에서 도시된 것과 유사한 매트릭스의 형태의 적절한 어크로스-필드 가중치들을 사용하여 전체 합으로 계산된다. 상기 토큰들은 필드 타입들의 각각의 쌍에 대한 미리정의된 가중치들의 매트릭스에 따라 믹스될 수 있다. 따라서, 서로 다른 필드들 사

이의 유사성들이 상기 계산들에서 고려될 수 있다. 수학적으로, 상기 최종 유사성 스코어는 $S = v_A [M] v_B$ 으로 표현될 수 있다.

[0076] 일부 구현들에서, 조정/정규화는 TF 및 IDF의 계산에 적용될 수 있다. 상기 TF-IDF 방법들은, 토큰 추출을 조정하고 가중치들의 ($[M]$ 과 같은) 미리정의된 매트릭스를 사용하여 서로 다른 타입들의 텍스트(예를 들어, 설명들, 캐스트, 촬영팀, 줄거리 등)에 대해 적절하게 결과들을 믹싱하여 프로그램-관련된 메타데이터에 적용될 수 있다.

[0077] 일 실시예에서, 보너스 스코어는, 예를 들어, 적어도 하나의 토큰을 공통으로 갖는 양의 유사성 스코어를 갖는 필드들의 매(every) 쌍에 대해 주어진다.

[0078] **예시적인 애플리케이션들**

[0079] 본 명세서에서 개시된 구현들은, 예를 들어, 도 1에서 도시된 스마트 TV 플랫폼의 일부로서 관련된 프로그램들을 결정하기 위해 사용될 수 있다.

[0080] 도 5는 프로그램들을 사용자에게 추천하는 예시적인 방법(500)을 나타내는 흐름도이다. 동작 502에서, 사용자가 매체 프로그램을 요청하거나 스마트 TV 플랫폼 상에서 상기 매체 프로그램의 검색을 행할 때, 서버는 요청을 수신한다. 캐스트들 및 해설들과 같은 프로그램의 메타데이터의 복수의 필드들에 기초하여, 유사한 프로그램들이 그것들 각각의 메타데이터에 기초하여 발견될 수 있다. 일 실시예에서, 하나의 프로그램에 관련되는 모든 프로그램들이 동작 504에서 데이터베이스 내의 모든 프로그램들에 대한 프로그램의 유사성 스코어들을 계산함으로써 발견될 수 있다. 가장 높은 스코어들을 갖는 프로그램들은 다시 되돌려질 수 있고, 동작 506에서 유사성 스코어들에 따라 순위가 매겨진다. 동작 508에서, 관련된 프로그램들은 요청된 프로그램에 따른 순위에 따른 순서로 디스플레이될 수 있다. 예를 들어, 사용자가 스마트 TV 모듈 상에서 영화를 먼저 선택한 후, 더욱 유사한 영화 타이틀들이 유사성 스코어들에 기초하여 나타날 것이다. 본 명세서에서 개시된 실시예들에 따라 추천된 영화들은, 예를 들어, 일부 공급자들(예를 들어, 아마존, 넷플릭스)에 의해 사용되는 공동의 필터링 기법들과 비교하여 사용자의 정확한 흥미를 더 정확하게 반영할 수 있다.

[0081] 도 6은 프로그램의 인기를 예측하는 예시적인 방법(600)을 나타내는 흐름도이다. 동작 602에서, 서버는 프로그램에 대해 사용자에게 의해 주어진 별점 평가(예를 들어, 얼마나 많은 사용자가 상기 프로그램을 좋아하는지 또는 싫어하는지에 기초하여 사용자에게 의해 주어진, 1개에서 5개까지의 별과 같은 별들의 수 또는 다른 랭킹)을 수신할 수 있다. 상기 사용자는 데이터의 수집을 거절할 수 있거나 또는 사용자에게 프로그램들을 추천하는데 유용할 수 있는 인구통계학적 데이터와 같은 추가 데이터를 제공할 수 있다. 상기 평가와 관련된 사용자 ID들 또는 다른 사용자 데이터는 익명으로 되고 사용자 이름들과 연결되지 않는다. 동작 604에서, 상기 서버는 상기에서 서술되는 계산 방법들을 사용하여 사용자가 보지 못한 모든 유사한 프로그램들에 대한 사용자의 등급들을 예측할 수 있다. 동작 606에서, 상기 예측된 등급들의 통계들은 사용자들의 샘플에서 수행될 수 있다. 동작 608에서, 상기 시스템은, 또한, 신작과 같은 프로그램들이 유사한 영화들의 인기에 근거하여 인기를 얻을 수 있을지 관련해 매체 프로그램 공급자들에게 분석을 제공할 수 있다.

[0082] TV 시리즈들에 대해, 시즌제가 아닌 TV-시리즈 레벨 또는 에피소드 레벨에서 동작하는 대단위 모델(coarse-grained model)이 구축될 수 있다. 상기 방법들 및 시스템들은 동일한 TV 시리즈들의 시즌들 또는 에피소드들

사이의 유사성/차이를 비교하는 것과 같은 정밀한 모델(finer-grain model)들에 적용할 수 있다.

[0083] 예시적인 시스템들

[0084] 도 7은 상기에서 서술되는 여러 동작들의 적어도 일부에서 수행하기 위해 사용될 수 있는 시스템을 나타내는 블록도이다. 프로세서(710)는 서버 프로세서일 수 있고, 하나 이상의 마이크로프로세서들 또는 처리 회로들을 포함할 수 있다. 버스(740)는 시스템 버스를 포함할 수 있다. 저장 디바이스(720)는 ROM(read only memory) 그리고/또는 RAM(random access memory)와 같은 시스템 메모리를 포함할 수 있다. 저장 디바이스(720)는, 예를 들어, 하드 디스크로부터 판독하고 하드 디스크에 기입하는 하드 디스크 드라이브, (예를 들어, 착탈가능한) 마그네틱 디스크로부터 판독하거나 상기 마그네틱 디스크에 기입하는 마그네틱 디스크 드라이브, 콤팩트 디스크 또는 다른 (자기) 광학 매체와 같은 착탈가능한 (자기) 광학 디스크로부터 판독하거나 상기 광학 디스크에 기입하는 광학 디스크 드라이브, 또는 임의의 다른 타입의 저장 매체를 포함하는 임의의 적합한 저장 장치를 포함할 수 있다.

[0085] 사용자는, 키보드 및 포인팅 디바이스(예를 들어, 마우스)와 같은 입력 디바이스(732)를 통해, 스마트 TV 모듈, 개인용 컴퓨터, 또는 모바일 디바이스와 같은 사용자 단말에 명령어 및 정보를 입력할 수 있다. 마이크로폰, 터치입력 인터페이스, 조이스틱, 게임 패드, 위성 방송 수신 안테나, 스캐너 등과 같은 다른 입력 디바이스들은 또한 (또는 대안적으로) 포함될 수 있다. 이러한 그리고 다른 입력 디바이스들은 적합한 인터페이스(730)를 통해 상기 네트워크(예를 들어, 상기 인터넷) 또는 시스템 버스(740)에 연결된 프로세서(710)에 접속될 수 있다.

[0086] 출력 디바이스(734)는 적합한 인터페이스를 통해 네트워크(740)에 접속될 수 있는 모니터 또는 다른 타입들의 디스플레이 디바이스들을 포함할 수 있다. 모니터에 더하여(또는 모니터 대신), 사용자 단말은, 예를 들어, 스피커들 및 프린터들과 같은 다른 (주변) 출력 디바이스들(미도시)을 포함할 수 있다. 일부 경우들에서, 출력 디바이스(734)는 하나 이상의 시각 출력, 촉각 출력(haptic output), 또는 청각 출력을 제공하는 컴포넌트를 포함할 수 있다.

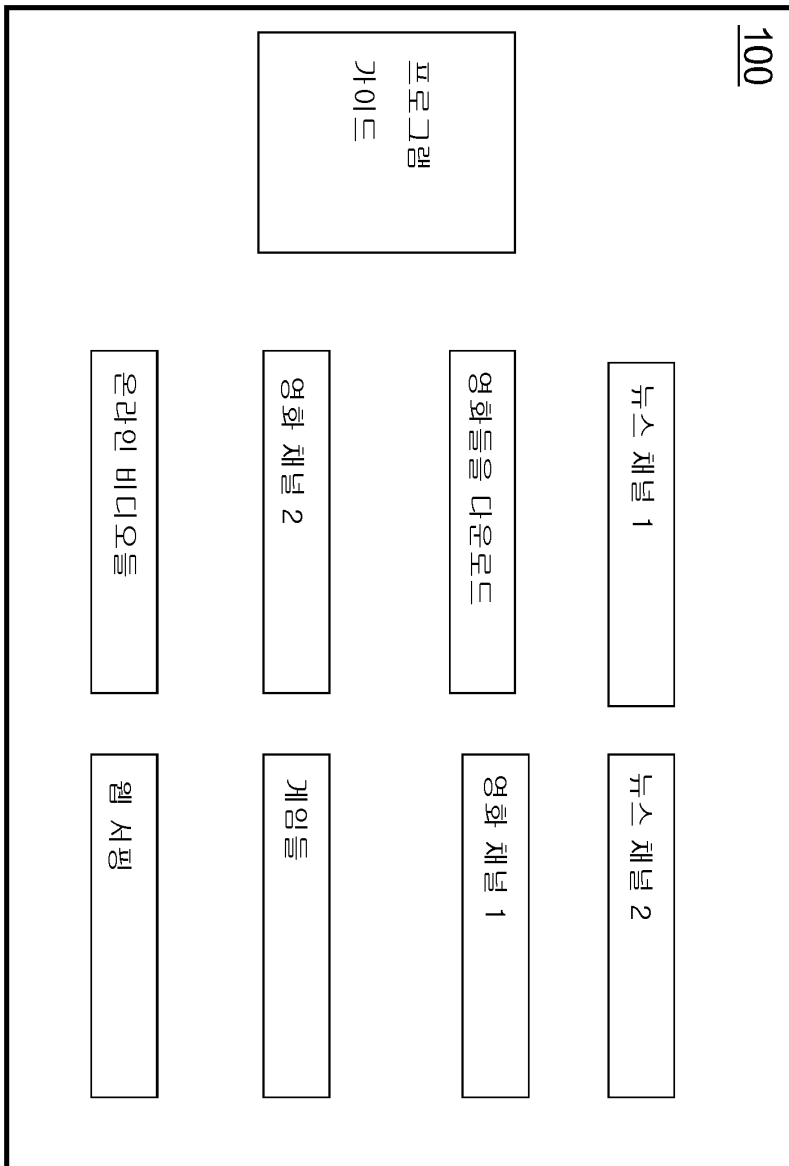
[0087] 여러 발명의 실시예들이 본 발명에서 서술되고 설명되지만, 당업자들은 기능을 수행하고 그리고/또는 본 명세서에서 서술되는 결과들 그리고/또는 하나 이상의 이점들을 얻기 위한 다양한 여러 수단 그리고/또는 구조들을 용이하게 고안할 것이고, 그리고 그와 같은 변형들 그리고/또는 수정들의 각각은 본 명세서에서 서술되는 발명의 실시예들의 범위 내에 있는 것으로 간주된다. 더 일반적으로, 당업자들은, 본 명세서에서 서술된 모든 파라미터들, 치수들, 재료들 및 구성들은 예시적인 것임을 의미하고, 실제 파라미터들, 치수들, 재료들 그리고/또는 구성들은 본 발명의 교시들이 사용되는 특정 애플리케이션 또는 애플리케이션들에 의존할 것임을 용이하게 이해할 것이다. 당업자들은 통상적인 과정만을 이용하여 본 발명에서 서술된 특정 발명의 실시예들과 동일한 많은 균등물들을 인식하거나 알 수 있다. 따라서, 이전의 실시예들은 단지 예로서 표시되고, 첨부된 청구범위 및 그것의 균등물의 범위 내에서, 발명의 실시예들은 구체적으로 서술되고 청구되는 것과 다르게 실행될 수 있음이 이해될 것이다. 본 발명의 실시예들은 본 명세서에서 서술되고 있는 각각의 개별적인 프레임, 시스템, 물품, 재료, 키트 그리고/또는 방법에 관한 것이다. 또한, 그와 같은 특징들, 시스템들, 물품들, 재료들, 키트들 그리고/또는 방법들이 상호 불일치하지 않으면, 2개 이상의 그와 같은 특징들, 시스템들, 물품들, 재료들, 키트들 그리고/또는 방법들의 임의의 조합은 본 발명의 범위 내에 포함된다.

[0088] 본 발명의 상기에서 서술된 실시예들은 임의의 다수의 방식으로 구현될 수 있다. 예를 들어, 일부 실시예들은 하드웨어, 소프트웨어 또는 그것의 조합을 사용하여 구현될 수 있다. 실시예의 임의의 양상이 소프트웨어에서 적어도 부분적으로 구현될 때, 단일 디바이스 또는 컴퓨터에 제공되거나 다수의 디바이스들/컴퓨터들 사이에서 분산되든지 상관없이 소프트웨어 코드는 임의의 적합한 프로세서 또는 프로세서들의 집합 상에서 실행될 수 있다.

[0089] 청구범위는 그 취지를 명시하지 않는 한 서술된 순서 또는 요소들을 제한하도록 판독되지 않아야 한다. 형식 및 세부사항들의 여러 변형들이 첨부된 청구범위들로 벗어나지 않고 당업자에 의해 행해질 수 있음이 이해되어야 한다. 하기의 청구범위 및 청구범위의 균등물의 정신 및 범위 내에 있는 모든 실시예들이 청구된다.

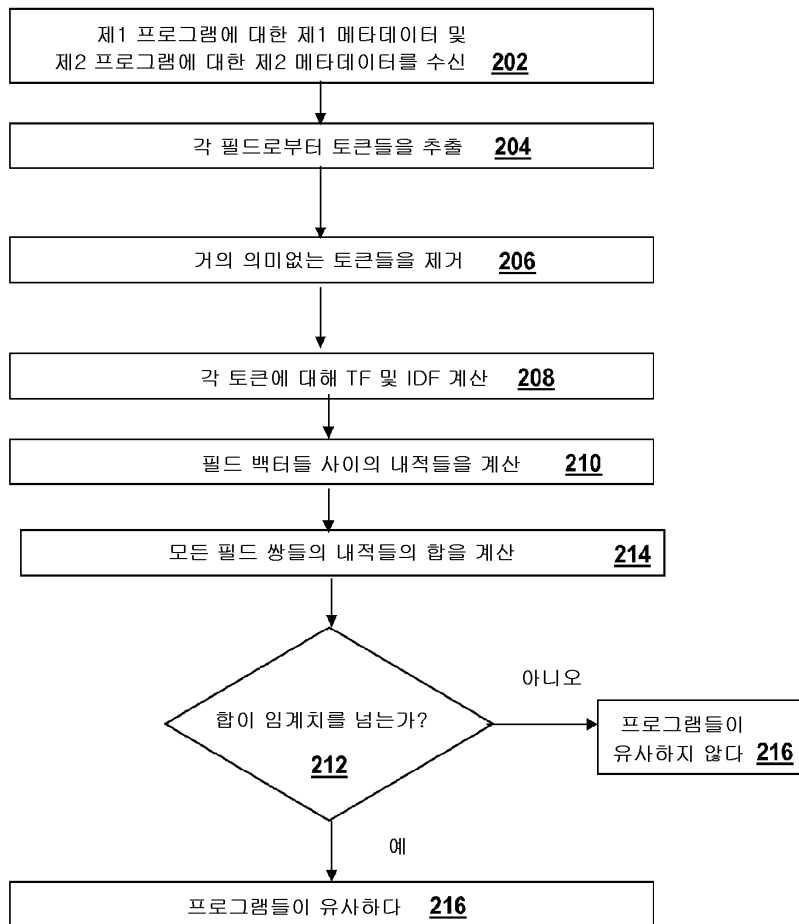
도면

도면1

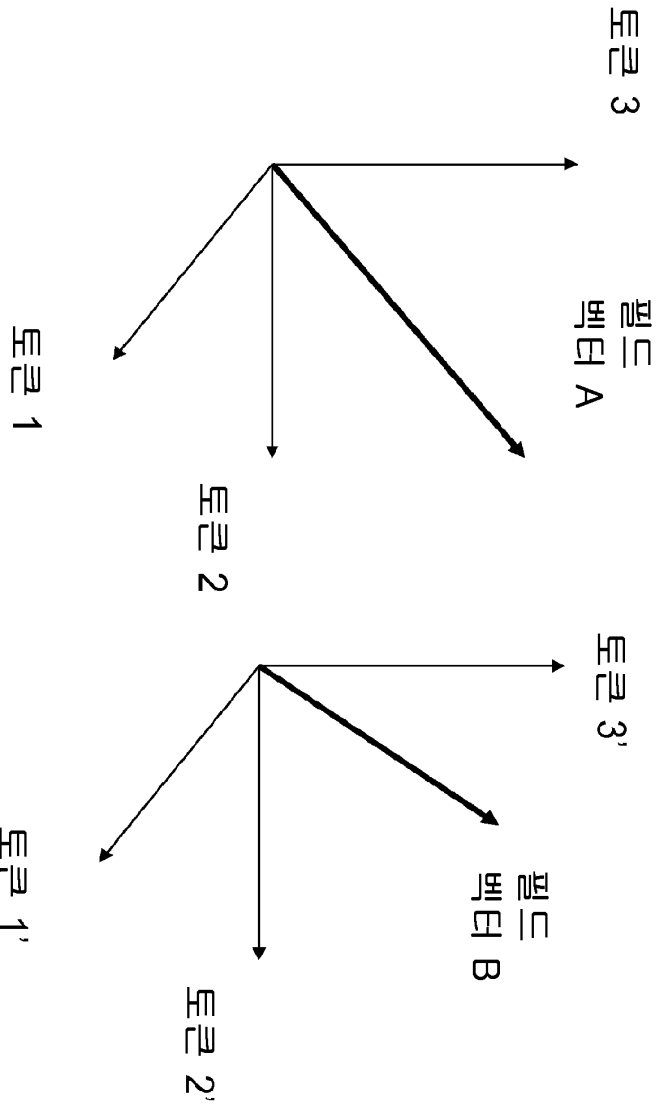


도면2

200 ↘



도면3

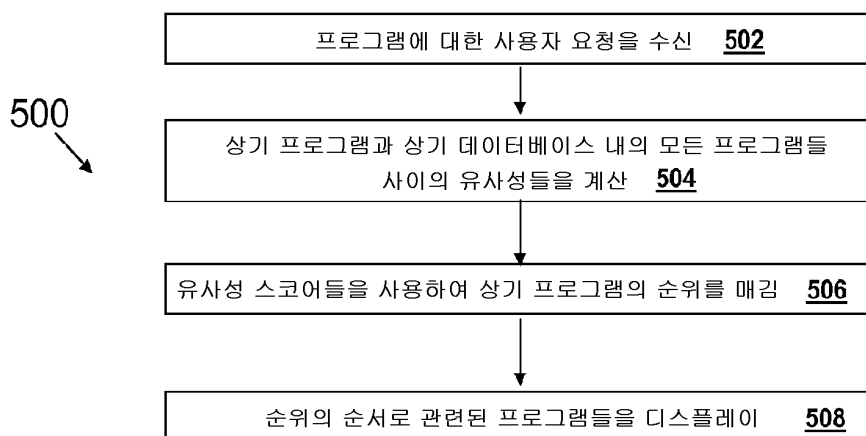


도면4

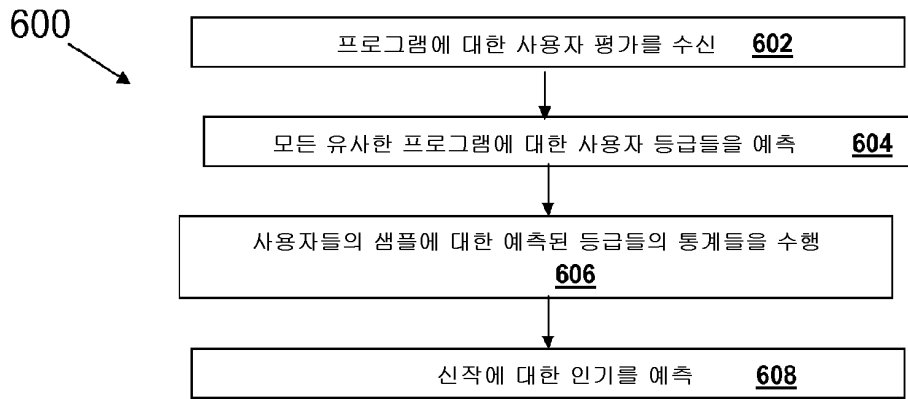
	X	Y				
설명	1	0.8	0.6	0.9	0	
타이틀	0.8	1	0	0.5	0	
캐스팅	0.6	0	1	0.4	0	
위키 플롯	0.9	0.5	0.4	1	0	
장르	0	0	0	0	1	

[M]

도면5



도면6



도면7

