



**(19) 대한민국특허청(KR)**  
**(12) 등록특허공보(B1)**

(45) 공고일자 2012년04월12일  
(11) 등록번호 10-1130533  
(24) 등록일자 2012년03월19일

(51) 국제특허분류(Int. Cl.)  
G06F 17/30 (2006.01) G06F 17/00 (2006.01)  
G06F 15/16 (2006.01)  
(21) 출원번호 10-2005-0040008  
(22) 출원일자 2005년05월13일  
심사청구일자 2010년05월06일  
(65) 공개번호 10-2006-0047858  
(43) 공개일자 2006년05월18일  
(30) 우선권주장  
10/846,949 2004년05월14일 미국(US)  
(56) 선행기술조사문헌  
JP2002073607 A  
JP2002278989 A  
JP2002507794 A

(73) 특허권자  
마이크로소프트 코포레이션  
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원  
마이크로소프트 웨이  
(72) 발명자  
즈항, 벤유  
미국 98052 워싱턴주 레드몬드 원 마이크로소프트  
웨이마이크로소프트 코포레이션 내  
주에, 구이-룽  
미국 98052 워싱턴주 레드몬드 원 마이크로소프트  
웨이마이크로소프트 코포레이션 내  
(뒷면에 계속)  
(74) 대리인  
제일특허법인

전체 청구항 수 : 총 12 항

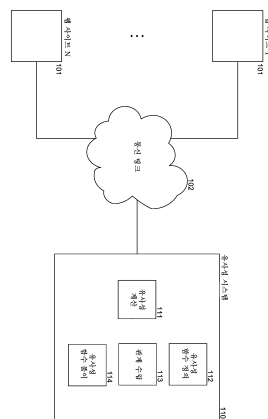
심사관 : 이명진

**(54) 발명의 명칭 이종 관계에 기초하여 객체들의 유사성을 결정하기 위한방법 및 시스템**

**(57) 요약**

동일한 유형 및 상이한 유형들의 객체들과의 관계들, 및 다른 객체들로의 객체들의 유사성들에 기초하여 객체들의 유사성을 측정하기 위한 방법 및 시스템이 제공된다. 일 실시예에서, 유사성 시스템은 객체의 각 유형에 대하여 유형 내 및 유형 간 유사성 함수들을 정의한다. 유사성 시스템은 특정 유형에 대한 유형 내 및 유형 간 유사성 함수들을 그 유형에 대한 전체 유사성 함수로 결합할 수 있다. 유사성 함수들을 정의한 후, 유사성 시스템은 객체들에 대한 속성 값들을 수집하는데, 그것은 유형 내 관계들이라 불리는 동일한 유형의 객체들 사이의 관계 데이터, 유형 간 관계들이라 불리는 상이한 유형들의 객체들 사이의 관계들을 포함할 수 있다. 객체들에 대한 속성 값들을 수집한 후, 유사성 시스템은 유사성들이 해에 수렴할 때까지 객체들에 대한 유사성들을 반복적으로 계산함으로써 유형 내 및 유형 간 유사성 함수들을 푼다.

**대표도 - 도1**



(72) 발명자

**쟁, 후아-준**

미국 98052 워싱턴주 레드몬드 원 마이크로소프트  
웨이마이크로소프트 코포레이션 내

**마, 웨이-잉**

미국 98052 워싱턴주 레드몬드 원 마이크로소프트  
웨이마이크로소프트 코포레이션 내

**첸, 즈헝**

미국 98052 워싱턴주 레드몬드 원 마이크로소프트  
웨이마이크로소프트 코포레이션 내

**특허청구의 범위**

**청구항 1**

컴퓨터 시스템에서 객체들 사이의 유사성의 측정값들(measurements of similarity)을 생성하기 위한 컴퓨터 구현 방법으로서,

각 객체는 복수의 유형들 중 하나를 가지고, 하나의 유형은 동일한 유형의 객체들 간의 유형 내 관계(intra-type relationship)를 가지며, 한 쌍의 유형은 상기 한 쌍 중 제1 유형의 하나의 객체와 상기 한 쌍 중 제2 유형의 다른 객체 간의 유형 간 관계(inter-type relationship)를 가지며,

상기 방법은,

각 유형에 대해,

상기 유형의 객체들 간의 유사성이 유형 내 관계에 기초하는 경우, 이러한 관계 각각에 대하여 상기 유형의 객체들 사이의 유사성을 측정하는 유형 내 유사성 함수(intra-type similarity function)를 정의하는 단계;

상기 유형의 객체들 간의 유사성이 다른 유형의 객체와의 유형 간 관계에 기초하는 경우, 이러한 관계 각각에 대하여, 상기 다른 유형의 객체들의 유사성에 기초하여 상기 유형의 객체들 사이의 유사성을 측정하는 유형 간 유사성 함수(inter-type similarity function)를 정의하는 단계 - 상기 유형 간 유사성 함수는 다른 유형의 객체들의 쌍들 사이의 유사성의 가중화 평균(weighted average)인 상기 유형의 제1 및 제2 객체들에 대한 유사성을 생성하고, 상기 쌍의 하나의 객체는 상기 제1 객체와의 관계를 가지고 상기 쌍의 다른 객체는 상기 제2 객체와의 관계를 가지며, 상기 유형 간 유사성 함수는 다음과 같이 정의되고,

$$S_{O_1}[a,b] = \frac{C}{|M_R(a)||M_R(b)|} \sum_{i=1}^{|M_R(a)|} \sum_{j=1}^{|M_R(b)|} S_{O_2}[M_R^i(a), M_R^j(b)]$$

여기서  $S_{O_1}$  은 유형  $O_1$ 의 객체들 a와 b 사이의 유사성을 나타내고,  $S_{O_2}$  는 다른 유형의 객체들 i와 j 사이의 유사성을 나타내며, R은 유사성이 기초하는 유형 간 관계를 나타내고, C는 가중치 팩터임 -; 및

상기 유형에 대한 임의의 유형 내 유사성 함수들 및 임의의 유형 간 유사성 함수들에 기초하여, 상기 유형의 객체들 사이의 유사성을 측정하는 전체 유사성 함수(combined similarity function)를 정의하는 단계 - 상기 전체 유사성 함수는 다음과 같이 정의되고,

$$S[a,b] = \alpha S_{intra}[a,b] + \beta S_{inter}[a,b]$$

여기서  $S_{intra}$  및  $S_{inter}$ 는 유형 내 유사성 함수들 및 유형 간 유사성 함수들로부터 유도된 유사성들을 나타내고 a 및  $\beta$ 는  $\alpha + \beta = 1$ 을 갖는 유사성들에 대한 가중치들을 나타냄 - 를 포함하고,

상기 방법은,

각 관계에 대하여, 그 관계를 갖는 객체들 a 및 b를 특정하는 데이터가 제공되는 단계;

상기 정의된 유사성 함수들을 반복적으로 계산함으로써 상기 정의된 유사성 함수들을 푸는 단계 - 상기 유사성 함수들은, 하나의 반복으로부터 다음 반복까지의 유사성들에 기초하는 차이 측정값이 차이 임계값보다 작은 때에 풀림 -; 및

상기 정의된 유사성 함수들을 동시에 푸는 것에 기초하여 유사성들을 저장하는 단계

를 포함하는 컴퓨터 구현 방법.

**청구항 2**

제1항에 있어서,

질의들에 대한 전체 유사성 함수는 질의들에 대한 내용 유사성 함수 및 질의들에 대한 클릭 유사성 함수를 조합

함으로써 정의되는, 컴퓨터 구현 방법.

**청구항 3**

제1항에 있어서,

웹 페이지들에 대한 전체 유사성 함수는 웹 페이지들에 대한 인커밍 링크 유사성 함수, 웹 페이지들에 대한 아웃고잉 링크 유사성 함수 및 웹 페이지들에 대한 클릭 유사성 함수를 조합함으로써 정의되는, 컴퓨터 구현 방법.

**청구항 4**

제2항에 있어서, 각각의 상기 내용 유사성 함수 및 상기 클릭 유사성 함수는 가중치가 부여되는, 컴퓨터 구현 방법.

**청구항 5**

제3항에 있어서, 각각의 상기 인커밍 링크 유사성 함수, 상기 아웃고잉 링크 유사성 함수 및 상기 클릭 유사성 함수는 가중치가 부여되는, 컴퓨터 구현 방법.

**청구항 6**

제4항 또는 제5항에 있어서,

상기 가중치의 합은 1인, 컴퓨터 구현 방법.

**청구항 7**

제1항 내지 제5항 중 어느 한 항의 컴퓨터 구현 방법에 의해, 질의들 간의, 그리고 웹 페이지들 간의 유사성의 측정값들을 생성하기 위하여 컴퓨터 시스템을 제어하기 위한 명령어들을 포함하는 컴퓨터 판독가능 기록 매체.

**청구항 8**

질의들 및 웹 페이지들에 대한 유사성 스코어(similarity score)를 계산하기 위한 컴퓨터 시스템으로서, 상기 시스템은,

질의들 a 및 b 에 대한 다음의 유사성 함수

$$S_{QC}[a,b] = \frac{|Keyword(a) \cap Keyword(b)|}{|Keyword(a) \cup Keyword(b)|}$$

를 구현하는 컴포넌트;

질의들 a 및 b 에 대한 다음의 클릭 유사성 함수

$$S_{QCT}[a,b] = \frac{C_{CT}}{|M_{CT}(a)||M_{CT}(b)|} \sum_{i=1}^{|M_{CT}(a)|} \sum_{j=1}^{|M_{CT}(b)|} S_{PCT}[M_{CT}^i(a), M_{CT}^j(b)]$$

를 구현하는 컴포넌트 - 여기서 S<sub>PCT</sub>는 클릭에 기초한 웹 페이지들의 유사성 함수를 나타내고, M<sub>CT</sub>(a)는 질의 로 그들로부터 식별되는 질의 a로부터 웹 페이지들로의 클릭을 나타내고, C<sub>CT</sub>는 가중치 팩터를 나타냄 -;

다음의 인커밍 링크 유사성 함수

$$S_{IL}[A,B] = \frac{C_{IL}}{|M_{IL}(A)||M_{IL}(B)|} \sum_{i=1}^{|M_{IL}(A)|} \sum_{j=1}^{|M_{IL}(B)|} S_{IL}[M_{IL}^i(A), M_{IL}^j(B)];$$

를 구현하는 컴포넌트 및

다음의 아웃고잉 링크 유사성 함수

$$S_{OL}[A, B] = \frac{C_{OL}}{|M_{OL}(A)||M_{OL}(B)|} \sum_{i=1}^{|M_{OL}(A)|} \sum_{j=1}^{|M_{OL}(B)|} S_{IL}[M_{OL}^i(A), M_{OL}^j(B)]$$

를 구현하는 컴포넌트 - 여기서, A 및 B는 웹 페이지들을 나타내고, C<sub>OL</sub> 및 C<sub>IL</sub>은 가중치 팩터를 나타내며, S<sub>OL</sub> 및 S<sub>IL</sub>은 아웃고잉 및 인커밍 링크들에 기초한 유사성 함수들을 나타내고, M<sub>OL</sub>(A)는 웹 페이지 A로부터의 아웃고잉 링크들의 목적지 웹 페이지들을 나타내며, M<sub>IL</sub>(A)는 웹 페이지 A로의 인커밍 링크들의 웹 페이지들의 소스를 나타냄 -;

웹 페이지들에 대한 다음의 클릭 유사성 함수

$$S_{CT}[A, B] = \frac{C_{CT}}{|M_{CT}(A)||M_{CT}(B)|} \sum_{i=1}^{|M_{CT}(A)|} \sum_{j=1}^{|M_{CT}(B)|} S_{QCT}[M_{CT}^i(A), M_{CT}^j(B)]$$

를 구현하는 컴포넌트 - 여기서, M<sub>CT</sub>(A)는 웹 페이지 A를 액세스하기 위하여 사용자들이 클릭하는 질의들을 나타냄 -; 및

상기 유사성 함수들을 구현하는 컴포넌트들을 반복적으로 실행하여 상기 유사성 함수들을 푸는 컴포넌트 - 상기 유사성 함수들은, 하나의 반복으로부터 다음 반복까지의 유사성들에 기초하는 차이 측정값이 차이 임계값보다 작은 때에 풀림 -

를 포함하는 컴퓨터 시스템.

**청구항 9**

제8항에 있어서, 웹 페이지들의 쌍들에 대한 전체 유사성 함수는 웹 페이지들의 쌍들에 대한 유사성 함수들의 선형 조합으로 정의되는, 컴퓨터 시스템.

**청구항 10**

제8항에 있어서, 질의들의 쌍들에 대한 전체 유사성 함수는 질의들의 쌍들에 대한 유사성 함수들의 선형 조합으로 정의되는, 컴퓨터 시스템.

**청구항 11**

제9항 또는 제10항에 있어서, 각각의 유사성 함수는 가중치가 부여되는, 컴퓨터 시스템.

**청구항 12**

제11항에 있어서, 상기 가중치의 합은 1인, 컴퓨터 시스템.

**청구항 13**

삭제

**청구항 14**

삭제

**청구항 15**

삭제

**청구항 16**

삭제

**청구항 17**

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

**명세서**

**발명의 상세한 설명**

**발명의 목적**

**발명이 속하는 기술 및 그 분야의 종래기술**

[0014] 개시된 기술은 일반적으로 객체들의 유사성을 결정하는 것, 특히 객체 관계에 기초하여 유사성을 결정하는 것에 관련된 것이다.

[0015] 구글(Google) 및 오버츄어(Overture)와 같은 많은 검색 엔진 서비스가 인터넷을 통해 액세스가능한 정보에 대한 검색을 제공한다. 이러한 검색 엔진 서비스들은 사용자들에게 관심있을 수 있는 웹 페이지와 같은 디스플레이 페이지들을 사용자들이 검색하도록 허용한다. 사용자가 검색 용어를 포함하는 검색 요청("질의"로도 언급됨)을 제출한 후, 검색 엔진 서비스는 그러한 검색 용어들에 관련될 수 있는 웹 페이지들을 식별한다. 관련 웹 페이지들을 신속히 식별하기 위하여, 검색 엔진 서비스는 웹 페이지에 대한 키워드의 매핑을 유지할 수 있다. 검색 엔진 서비스는 각 웹 페이지의 키워드들을 추출하기 위하여 웹(즉, 월드 와이드 웹)을 "크롤링(crawling)"함으로써 이러한 매핑을 생성할 수 있다. 웹을 크롤링하기 위하여, 검색 엔진 서비스는 루트 웹 페이지들의 리스트를 사용할 수 있으며, 그러한 루트 웹 페이지들을 통하여 액세스가능한 모든 웹 페이지들을 식별할 수 있다. 임의의 특정 웹 페이지의 키워드들은 헤드라인의 단어들, 웹 페이지의 메타데이터에서 제공된 단어들, 하이라이트된 단어들 등을 식별하는 것과 같은 다양한 공지된 정보 검색 기술을 사용하여 추출될 수 있다. 검색 엔진 서비스는 각 매치의 근접성, 웹 페이지 인기도(예를 들어, 구글의 PageRank) 등에 기초하여 각 웹 페이지가 검색 요청에 얼마나 관련있는지를 나타내는 관련성 스코어(relevance score)를 계산할 수 있다. 그 후 검색 엔진 서비스는 웹 페이지로의 링크들을 웹 페이지들의 관련성에 기초한 순서로 사용자에게 디스플레이한다. 검색 엔진들은 더욱 일반적으로 임의의 컬렉션의 문서들에 있는 정보에 대한 검색을 제공할 수 있다. 예를 들어, 문서들의 컬렉션들은 모든 미국 특허, 모든 연방 법원 의견, 회사의 모든 보관 문서 등을 포함할 수 있다.

**발명이 이루고자 하는 기술적 과제**

[0016] 검색 엔진 서비스들은 웹 페이지나 질의와 같은 다양한 객체들 사이의 유사성을 측정할 필요가 있을 수 있다. 예를 들어, 검색 엔진 서비스는 대화식 질의 확장(interactive query expansion)을 허용할 수 있는데, 그것은 질의 용어들과 기타 용어들 사이의 유사성 계산을 요구한다. 또다른 예로서, 검색 엔진 서비스는 웹 페이지를 통해 네비게이션하는 사용자를 돕기 위하여 웹 페이지들을 유사한 웹 페이지의 클러스터들로 그룹화하기를 원할 수 있다. 객체들의 유사성을 결정하기 위한 전형적인 알고리즘들은 일반적으로 객체들에 관련된 특성 벡터(feature vector)를 사용하고, 그 후 유사성의 표시로서 특성 벡터들 사이의 거리를 계산한다. 예를 들어, 웹 페이지들은 유사성을 계산하기 위하여 사용되는 키워드, 내용 등을 포함하는 특성들을 가질 수 있다. 유사성을 결정할 때 대부분의 알고리즘이 객체들에 관련된 특성들에 전적으로 의존한다. 예를 들어, 웹 페이지들 사이의 유사성은 웹 페이지들의 내용에 전적으로 기초할 수 있다. 그러나, 소수의 알고리즘들은 이종의 객체(heterogeneous object)들에 기초한 특성들을 고려한다. 예를 들어, 한 알고리즘은 질의들이 동일한 용어들을

포함하거나 사용자들이 동일한 웹 페이지를 선택하게 한다면 질의들이 유사한 사용자 클릭 데이터(click-through data)를 사용한다. 따라서, 그러한 질의들에 대한 특성 벡터는 사용자들에 의해 선택된 질의 결과의 웹 페이지 상의 정보를 포함할 것이다.

[0017] 그러나, 이러한 기술들은 하나의 유형의 객체들 사이의 유사성을 계산할 때, 관련될 수 있는 또다른 유형의 객체들 사이의 유사성을 고려하지 않는다. 즉, 하나의 유형의 객체들에 대한 유사성 측정은 또다른 유형의 객체들에 대한 유사성 측정에 관련될 수 있다. 예를 들어, 하나의 질의는 사용자들이 선택하거나 클릭하는 결과들의 웹 페이지들 사이의 유사성에 부분적으로 기초하여 또다른 질의에 유사할 수 있다. 역으로, 웹 페이지들은 결과로 웹 페이지들을 리턴하는 질의들 사이의 유사성에 부분적으로 기초하여 또다른 웹 페이지에 유사할 수 있다. 이종의 객체들 사이의 관계들을 고려하는 객체들의 유사성을 측정하기 위한 기술을 가지는 것이 바람직할 것이다.

**발명의 구성 및 작용**

[0018] 동일한 유형의 객체들과 상이한 유형들 사이의 관계, 및 그러한 객체들의 기타 객체들에의 유사성에 기초하여 객체들의 유사성을 측정하기 위한 방법 및 시스템이 제공된다. 일 실시예에서, 유사성 시스템은 객체의 각 유형에 대한 유형 내(intra-type) 및 유형 간(inter-type) 유사성 함수를 정의한다. 유사성 시스템은 특정 유형에 대한 유형 내 및 유형 간 유사성 함수를 그 유형에 대한 전체 유사성 함수로 결합할 수 있다. 유사성 함수들을 정의한 후에, 유사성 시스템은 객체들에 대한 속성 값들을 수집하며, 그것은 유형 내 관계(intra-type relationship)로 불리는 동일한 유형의 객체들 사이의 관계 데이터, 및 유형 간 관계(inter-type relationship)로 불리는 상이한 유형의 객체들 사이의 관계를 포함할 수 있다. 객체들에 대한 속성 값들을 수집한 후, 유사성 시스템은 유사성이 해(solution)에 수렴할 때까지 객체들에 대한 유사성들을 반복적으로 계산함으로써 유형 내 및 유형 간 유사성 함수를 푼다.

[0019] 동일한 유형의 객체들과 상이한 유형들 사이의 관계, 및 그러한 객체들의 기타 객체들에의 유사성에 기초하여 객체들의 유사성을 측정하기 위한 방법 및 시스템이 제공된다. 일 실시예에서, 유사성 함수는 객체의 각 유형에 대한 유형 내 및 유형 간 유사성 함수들을 정의한다. 유형 내 유사성 함수는 동일한 유형의 객체들 사이의 유사성을 측정한다. 예를 들어, 질의들 사이의 유형 내 유사성 함수는 질의들의 검색 용어들이 질의들을 제출하는 사용자들의 속성에 기초하여 얼마나 근접하게 조화하는지에 기초할 수 있다. 객체들 사이의 유형 내 유사성은 또한 동일한 유형의 기타 객체들의 유사성에 의존할 수 있다. 예를 들어, 두 질의들은 그들 각각이 제 3의 질의에 높은 유사성을 가지는 경우 서로에게 더욱 유사할 수 있다. 기타 객체들 사이의 유사성에 기초한 객체들 사이의 그러한 유형 내 유사성은 재귀적인 함수를 정의한다. 유형 간 유사성 함수는 또다른 유형의 객체들의 유사성을 포함하는 속성들에 기초하여 하나의 유형의 두 객체들 사이의 유사성을 측정한다. 예를 들어, 사용자들이 클릭하는 하나의 질의의 결과의 웹 페이지가 사용자가 클릭하는 다른 질의의 결과의 웹 페이지에 유사한 경우 두 질의가 더욱 유사할 수 있다. 기타 유형의 객체들의 유사성은 또한 한 유형의 객체들의 유사성에 의존할 수 있다. 게다가, 한 유형의 객체들의 유사성이 또다른 유형의 객체들의 유사성에 의존할 수 있고 그 반대도 마찬가지이기 때문에, 유형 간 유사성 함수들은 상이한 유형들 사이에 재귀적이다.

[0020] 객체의 유형은 객체들의 상이한 속성들에 기초하여 객체들에 대해 정의된 다양한 유사성 정의들을 가질 수 있다. 예를 들어, 웹 페이지는 웹 페이지들의 내용에 기초한 유형 내 유사성 및 웹 페이지들 사이의 링크들에 기초한 또다른 유형 내 유사성을 가질 수 있다. 유사성 시스템은 특정 유형에 대한 유형 내 및 유형 간 유사성 함수들을 그 유형에 대한 전체 유사성 함수로 결합할 수 있다. 일 실시예에서, 유사성 시스템은 그 유형의 객체들 사이의 전체 유사성을 표현하는 데에 있어서 감지된 정확성에 기초하여 각 유형 내 및 유형 간 유사성 함수에 적용된 가중치들을 갖는 선형 방정식에 의하여 유형 내 및 유형 간 유사성 함수들을 결합한다. 예를 들어, 높은 정확성을 갖는 유형 내 유사성 함수는 높은 가중치를 부여받을 수 있고, 낮은 정확성을 갖는 유형 내 유사성 함수는 낮은 가중치를 부여받을 수 있다.

[0021] 유사성 함수들을 정의한 후에, 유사성 시스템은 객체들에 대한 속성 값들을 수집하는데, 그것은 유형 내 관계들이라 불리는 동일한 유형의 객체들 사이의 관계 데이터 및 유형 간 관계들이라 불리는 상이한 유형들의 객체들 사이의 관계들을 포함할 수 있다. 예를 들어, 웹 페이지는 웹 페이지의 키워드들에 대응하는 비관계 기반(non-relationship-based) 속성 값들을 가질 수 있다. 웹 페이지는 또한 웹 페이지들 사이의 인커밍 및 아웃고잉 링크들에 기초한 유형 내 관계를 가질 수 있다. 웹 페이지는 질의 결과들로부터 웹 페이지들로의 클릭에 기초한 질의들과의 유형 간 관계를 가질 수 있다.

[0022] 객체들에 대한 속성 값들을 수집한 후에, 유사성 시스템은 유사성들이 해에 수렴할 때까지 객체들에 대한 유사

성들을 반복적으로 계산함으로써 유형 내 및 유형 간 유사성 함수들을 푼다. 유사성 시스템은 유사성 함수들의 재귀적 본질 때문에 반복적인 접근법을 사용한다. 유사성 시스템은 초기화된 유사성들을 가지고 시작하여 그 후 새로운 유사성들을 부여하기 위해 초기 유사성들에 기초하여 객체의 각 유형에 대하여 유사성 함수를 계산한다. 유사성 시스템은 유사성들이 해에 수렴하는지를 결정하기 위하여 새로운 유사성(new similarity)들과 지난 유사성(old similarity)들 사이의 차이를 측정한다. 만약 그렇다면, 새로운 유사성들은 해를 나타낸다. 만약 그렇지 않다면, 유사성 시스템은 지난 유사성들이 되는 새로운 유사성들을 가지고 반복한다. 따라서 유사성 시스템은 또다른 유형의 객체들의 유사성들에 기초하여 그리고 상이한 유형들의 객체들 사이의 유사성들에 기초하여 하나의 유형의 객체들의 유사성들을 계산한다.

[0023] 다음으로, 검색 엔진의 문맥에서 유사성 시스템의 프로세싱의 예가 제공된다. 유사성 시스템은 객체들(예를 들어, 웹 페이지들 및 질의들) 및 노드  $V$ 가 검색 엔진의 객체들을 나타내고 에지  $E$ 가 객체들 사이의 관계들을 나타내는 방향 그래프  $G=(V,E)$ 로서 검색 엔진에 의해 사용된 관계들(예를 들어, 인커밍 링크들 및 클릭들)을 모델링한다. 노드들  $V$ 는  $Q$ 가 질의들을 나타내고  $P$ 가 웹 페이지들을 나타내는 두 부분집합  $Q=\{q^1, q^2, \dots, q^m\}$  및  $P=\{p^1, p^2, \dots, p^n\}$ 으로 나누어질 수 있다. 이러한 웹 페이지들과 질의들 사이의 관계들은 인커밍 링크 관계(IL), 아웃고잉 링크 관계(OL), 및 클릭 관계(CT)를 포함할 수 있다. 그래프에서 노드  $v$ 에 대하여,  $M_R(v)$ 는 노드  $v$ 와의 관계  $R$ 를 가지는 인접 노드들의 집합을 나타낸다. 예를 들어,  $M_{IL}(v)$ 는 웹 페이지  $v$ 로의 인커밍 링크들의 소스인 웹 페이지들의 집합을 나타낸다.  $M_R^i(v)$ 는 집합에서  $i^{th}$  웹 페이지를 나타낸다. 유사성 시스템은 객체들 사이의 유사성을 나타내기 위하여 유사성 행렬  $S$ 를 사용하고,  $S[a,b]$ 는 객체들  $a$ 와  $b$  사이의 유사성을 나타낸다.

[0024] 유사성 시스템은 하나의 유형의 객체들이 부분적으로 또다른 유형의 상호관련된 객체들의 유사성에 기초하여 유사하다는 원리에 기초한다. 만약 하나의 유형의 두 객체들이 또다른 유형의 동일한 객체와의 관계를 가진다면, 두 객체들은 어느 정도 유사하다. 또한 만약 동일한 유형의 두 객체들이 또다른 유형의 두 가지 상이하지만 유사한 객체와의 관계를 가진다면, 두 객체들은 어느 정도 유사하다. 유사성 시스템은 다음 방정식들을 가지고 이러한 원리를 표현한다.

[0025] 
$$S_{\alpha}[a,b] = \frac{C}{|M_R(a)||M_R(b)|} \sum_{i=1}^{|M_R(a)|} \sum_{j=1}^{|M_R(b)|} S_{\alpha_i}[M_R^i(a), M_R^j(b)] \quad (1)$$

[0026] 여기서  $S_{\alpha}$  은 유형  $O_1$ 의 객체들  $a$ 와  $b$  사이의 유사성을 나타내고,  $S_{\alpha_i}$  는 또다른 유형의 객체들  $i$ 와  $j$  사이의 유사성을 나타내며,  $R$ 은 유사성이 기초하는 유형 간 관계를 나타내고,  $C$ 는 가중치 팩터이다. 만약  $a$ 가  $b$ 와 동일하다면,  $S_{\alpha}[a,b]$  는 1로 정의되는데, 즉, 객체와 그 자신 사이의 유사성은 최대 유사성 1을 정의한다. 만약  $a$ 와  $b$  둘 모두  $O_2$ 에 있는 동일한 객체  $A$ 에 관련된다면,  $S_{\alpha}[A,A]$  는 1이며, 그것은  $S_{\alpha}[a,b]$ 에 최대 기여를 부여한다. 만약  $a$  또는  $b$ 가 임의의 이웃(neighbor)을 가지지 않는다면, 즉,  $O_2$ 에 있는 객체에 아무런 관계를 갖지 않는다면,  $|M_R(a)|$  또는  $|M_R(b)|$ 는 0과 같을 것이다. 그러한 경우에, 유사성 시스템은 0에 의한 나눗셈을 방지하면서,  $S_{\alpha}[a,b]$  를 0으로 설정한다. 예로서,  $O_1$ 이 객체들  $a$  및  $b$ 를 포함하고,  $O_2$ 가 객체들  $A, B$  및  $C$ 를 포함하며,  $a$ 는  $A$  및  $B$ 에 관련되고  $b$ 는  $B$  및  $C$ 에 관련된다고 가정한다. 만약  $S_{\alpha}[A,B]$ 가 .7이고,  $S_{\alpha}[B,C]$ 가 .7이며,  $S_{\alpha}[A,C]$  는 .49이고 가중치 팩터는 .7이라면, 방정식 1을 적용함으로써,  $S_{\alpha}[a,b]$  는 .5이다(예를 들어,  $.7/4 * (.7 + .49 + 1.0 + .7)$ ).

[0027] 유사성 시스템은 유형 내 유사성 함수들 및 유형 간 유사성 함수들로부터 유도된 유사성들의 조합에 기초하여 객체들의 유형의 전체 유사성을 정의한다. 일 실시예에서, 유사성 시스템은 다음 방정식에 의해 표현된 바와 같이 유형 내 유사성 함수들 및 유형 간 유사성 함수들의 유사성들의 선형 조합을 사용한다.

[0028] 
$$S[a,b] = \alpha S_{micro}[a,b] + \beta S_{macro}[a,b] \quad (2)$$

[0029] 여기서  $S_{intra}$  및  $S_{inter}$ 는 유형 내 유사성 함수들 및 유형 간 유사성 함수들로부터 유도된 유사성들을 나타내고  $\alpha$  및  $\beta$ 는  $\alpha+\beta=1$ 을 갖는 유사성들에 대한 가중치들을 나타낸다.  $\alpha$  및  $\beta$ 에 상이한 값들을 할당함으로써, 유사성 시스템은 상이한 유사성 함수들의 기여들을 전체 유사성으로 조정할 수 있다. 상기에 기술된 바와 같이, 방정식 2는 하나의 객체의 유사성이 또다른 객체의 유사성에 기초하여 정의될 수 있기 때문에 재귀적으로 정의될 수 있으며, 그것은 차례로 하나의 객체의 유사성에 기초하여 정의될 수 있다. 일 실시예에서, 유사성 시스템은 수렴(즉,  $\|S^i - S^{i-1}\| < \epsilon$ , 여기서  $\epsilon$ 은 차이 임계값)할 때까지 유사성들을 계산함으로써 유사성 함수들을 반복적으로 푼다.

[0030] 검색 엔진 문맥에서, 유사성 시스템은 유형 내 유사성 함수를 정의하기 위하여 질의의 내용만을 사용할 수 있다. 내용에 기초한 유형 내 유사성 함수는 다음 방정식에 의해 정의될 수 있다.

$$S_{QC}[a, b] = \frac{|Keyword(a) \cap Keyword(b)|}{|Keyword(a) \cup Keyword(b)|} \quad (3)$$

[0032] 여기서  $a$  및  $b$ 는 질의들이며,  $S_{QC}$ 는 내용에 기초한 질의들의 내용 유사성 행렬이다. 예로서, 질의들  $a$  및  $b$ 가 공통으로 하나의 키워드를 갖는 두 검색 용어(또는 키워드들)를 가지는 경우, 그들의 유사성 값은 .33(즉, 1/3)일 것이다. 유사성 시스템은 다음 방정식에 의해 웹 페이지들로의 클릭 관계에 기초하여 질의에 대한 유형 간 유사성 함수를 정의할 수 있다.

$$S_{QCT}[a, b] = \frac{C_{CT}}{|M_{CT}(a)||M_{CT}(b)|} \sum_{i=1}^{|M_{CT}(a)|} \sum_{j=1}^{|M_{CT}(b)|} S_{PCT}[M_{CT}^i(a), M_{CT}^j(b)] \quad (4)$$

[0034] 여기서,  $S_{QCT}$ 는 클릭에 기초한 질의들의 유사성 행렬을 나타내고,  $S_{PCT}$ 는 클릭에 기초한 웹 페이지들의 유사성 행렬을 나타내며,  $M_{CT}(a)$ 는 질의 로그들로부터 식별되는 질의  $a$ 로부터 웹 페이지들로의 클릭을 나타내고,  $C_{CT}$ 는 가중치 팩터를 나타낸다. 유사성 시스템은 방정식 (3) 및 (4)를 다음 방정식에 의해 표현되는 질의들에 대한 전체 유사성 함수로 결합한다.

$$S_Q[a, b] = \alpha S_{QC}[a, b] + \beta S_{QCT}[a, b] \quad (5)$$

[0036] 여기서  $S_Q$ 는 질의들의 전체 유사성 행렬을 나타낸다.

[0037] 유사성 시스템은 인커밍 링크들 및 아웃고잉 링크들의 유형 내 관계들, 및 웹 페이지들로의 클릭으로 귀결하는 질의들과의 유형 간 관계에 기초하여 웹 페이지들의 유사성을 표현한다. 유사성 시스템은 동일한 웹 페이지(또는 유사한 웹 페이지들)에 의해 링크될 때 두 웹 페이지가 유사할 수 있다는 점을 반영하기 위하여 인커밍 링크 관계에 기초하여 유형 내 유사성 함수를 정의한다. 유사성 시스템은 또한 두 웹 페이지가 동일한 웹 페이지(또는 유사한 웹 페이지들)로 링크될 때 두 웹 페이지가 유사할 수 있다는 점을 반영하기 위하여 아웃고잉 링크 관계에 기초하여 유형 간 유사성 함수를 정의한다. 유사성 시스템은 다음 방정식들에 의해 아웃고잉 및 인커밍 링크 관계들에 기초한 웹 페이지들에 대한 유형 내 유사성 함수들을 표현한다.

$$S_{OL}[A, B] = \frac{C_{OL}}{|M_{OL}(A)||M_{OL}(B)|} \sum_{i=1}^{|M_{OL}(A)|} \sum_{j=1}^{|M_{OL}(B)|} S_{IL}[M_{OL}^i(A), M_{OL}^j(B)]$$

$$S_{IL}[A, B] = \frac{C_{IL}}{|M_{IL}(A)||M_{IL}(B)|} \sum_{i=1}^{|M_{IL}(A)|} \sum_{j=1}^{|M_{IL}(B)|} S_{OL}[M_{IL}^i(A), M_{IL}^j(B)] \quad (6), (7)$$

[0039] 여기서,  $A$  및  $B$ 는 웹 페이지들을 나타내고,  $C_{OL}$  및  $C_{IL}$ 은 가중치 팩터들을 나타내며,  $S_{OL}$  및  $S_{IL}$ 은 아웃고잉 및 인커밍 링크들에 기초한 유사성 행렬들을 나타내고,  $M_{OL}(A)$ 는 웹 페이지  $A$ 로부터의 아웃고잉 링크들의 목적지 웹 페이지들을 나타내며,  $M_{IL}(A)$ 는 웹 페이지  $A$ 로의 인커밍 링크들의 웹 페이지들의 소스를 나타낸다. 유사성 시스템은 다음 방정식에 의하여 클릭 관계에 기초한 웹 페이지들에 대한 유형 간 유사성 함수를 나타낸다.

$$S_{PCT}[A, B] = \frac{C_{CT}}{|M_{CT}(A)||M_{CT}(B)|} \sum_{i=1}^{|M_{CT}(A)||M_{CT}(B)|} \sum_{j=1}^{|M_{CT}(A)||M_{CT}(B)|} S_{QCT}[M_{CT}^i(A), M_{CT}^j(B)] \quad (8)$$

[0040]

[0041] 여기서  $M_{CT}(A)$ 는 웹 페이지 A를 액세스하기 위하여 사용자들이 클릭하는 질의들을 나타낸다. 방정식 8이 방정식 4의 용어(즉,  $S_{QCT}$ )로 정의되고 반대로 마찬가지로 방정식들의 쌍이 재귀적 함수를 정의한다. 유사성 시스템은 유형 내 유사성 함수들 및 유형 간 유사성 함수들의 선형 조합으로서 웹 페이지들에 대한 전체 유사성 함수를 정의하는데, 그것은 다음 방정식에 의해 표현된다.

$$S_P[A, B] = \alpha' S_{OL}[A, B] + \beta' S_{IL}[A, B] + \gamma' S_{PCT}[A, B] \quad (9)$$

[0042] 여기서  $S_P$ 는 웹 페이지들에 대한 유사성 행렬을 나타내고,  $\alpha'$ ,  $\beta'$  및  $\gamma'$ 은  $\alpha' + \beta' + \gamma' = 1$ 인 가중치들이다.

[0044] 따라서 유사성 시스템은 이중 객체들과 그들의 유형 간 관계들을 통합하기 위하여 통합 프레임워크를 사용한다. 전체 유사성 함수들이 재귀적이기 때문에, 유사성 시스템은 유사성 함수들을 동시에 그리고 반복적으로 푼다. 유사성 함수들은 다음 방정식들에 의해 표현된다.

$$\begin{aligned} S_{QC}[a, b] &= \frac{Keyword(a) \cap Keyword(b)}{Keyword(a) \cup Keyword(b)} \\ S_{QCT}[a, b] &= \frac{C_{CT}}{|M_{CT}(a)||M_{CT}(b)|} \sum_{i=1}^{|M_{CT}(a)||M_{CT}(b)|} \sum_{j=1}^{|M_{CT}(a)||M_{CT}(b)|} S_P[M_{CT}^i(a), M_{CT}^j(b)] \\ S_Q[a, b] &= \alpha S_{QC}[a, b] + \beta S_{QCT}[a, b] \\ S_{OL}[A, B] &= \frac{C_{PC}}{|M_{OL}(A)||M_{OL}(B)|} \sum_{i=1}^{|M_{OL}(A)||M_{OL}(B)|} \sum_{j=1}^{|M_{OL}(A)||M_{OL}(B)|} S_P[M_{OL}^i(A), M_{OL}^j(B)] \end{aligned} \quad (10)$$

[0045]

$$\begin{aligned} S_{IL}[A, B] &= \frac{C_{PR}}{|M_{IL}(A)||M_{IL}(B)|} \sum_{i=1}^{|M_{IL}(A)||M_{IL}(B)|} \sum_{j=1}^{|M_{IL}(A)||M_{IL}(B)|} S_P[M_{IL}^i(A), M_{IL}^j(B)] \\ S_{PCT}[A, B] &= \frac{C_{CT}}{|M_{CT}(A)||M_{CT}(B)|} \sum_{i=1}^{|M_{CT}(A)||M_{CT}(B)|} \sum_{j=1}^{|M_{CT}(A)||M_{CT}(B)|} S_Q[M_{CT}^i(A), M_{CT}^j(B)] \\ S_P[A, B] &= \alpha' S_{OL}[A, B] + \beta' S_{IL}[A, B] + \gamma' S_{PCT}[A, B] \end{aligned}$$

[0046]

[0047] 방정식 10에서 볼 수 있는 것과 같이, 임의의 두 질의 사이의 유형 간 유사성은 웹 페이지들의 유사성에 의해 영향받는데, 유형 내 및 유형 간 유사성 둘 모두에 의해 영향받는다. 웹 페이지들 사이의 유형 간 유사성이 질의들의 유사성, 즉, 유형 내 및 유형 간 모두에 의해 영향받기 때문에, 방정식 10은 재귀적 관계들을 정의한다. 따라서, 웹 페이지들 및 질의들의 유사성들은 상호적으로 서로에게 향하며 안정된 상태로 수렴한다.

[0048]

도 1은 일 실시예에서 유사성 시스템의 컴포넌트들을 도시하는 블록도이다. 웹 사이트들(101)은 통신 링크(102)를 통해 유사성 시스템(110)에 연결된다. 유사성 시스템은 유사성 계산 컴포넌트(111), 유사성 함수 정의 컴포넌트(112), 관계 수립 컴포넌트(113), 및 유사성 함수 풀이 컴포넌트(114)를 포함한다. 유사성 계산 컴포넌트는 다른 유형들의 객체들의 유형 간 관계들 및 유사성들에 기초하여 객체들 사이의 유사성들을 계산한다. 유사성 계산 컴포넌트는 유사성 함수 정의 컴포넌트, 관계 수립 컴포넌트, 및 유사성 함수 풀이 컴포넌트를 인보크한다. 유사성 함수 정의 컴포넌트는 객체들의 유형들, 객체들 사이의 관계들, 및 각 유형의 객체에 대한 다양한 유사성 함수들을 정의하기 위하여 사용자와 상호작용할 수 있다. 관계 수립 컴포넌트는 수집된 데이터에 기초하여 관계 데이터를 생성한다. 예를 들어, 수집된 데이터는 질의들, 질의 결과들의 웹 페이지들, 및 질의 로그들을 포함할 수 있다. 유사성 함수 풀이 컴포넌트는 유사성 행렬들의 유사성들이 해에 수렴할 때까지 갱신된 유사성 행렬들을 생성하기 위하여 정의된 유사성 함수들을 반복적으로 계산한다.

[0049]

유사성 시스템이 구현되는 컴퓨팅 장치는 중앙 처리 장치, 메모리, 입력 장치들(예를 들어, 키보드 및 포인팅 장치들), 출력 장치들(예를 들어, 디스플레이 장치들), 및 저장 장치들(예를 들어, 디스크 드라이브들)을 포함할 수 있다. 메모리 및 저장 장치들은 유사성 시스템을 구현하는 명령어들을 포함할 수 있는 컴퓨터 판독가능 매체이다. 또한, 데이터 구조들 및 메시지 구조들은 통신 링크상의 신호와 같은 데이터 전송 매체를 통해 저장되거나 전송될 수 있다. 인터넷, LAN, WAN, 또는 점대점 전화 연결과 같은 다양한 통신 링크들이 사용될 수 있다.

[0050]

유사성 시스템은 다양한 운영 환경에서 구현될 수 있다. 사용에 적합할 수 있는 다양한 공지된 컴퓨팅 시스템,

환경, 및 구성은 개인용 컴퓨터, 서버 컴퓨터, 포켓형 또는 랩탑 장치, 멀티프로세서 시스템, 마이크로프로세서 기반 시스템, 프로그래밍 가능한 가전기기, 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터, 상기 시스템 또는 장치들 중 임의의 것을 포함하는 분산 컴퓨팅 환경 등을 포함한다.

[0051] 유사성 시스템은 하나 이상의 컴퓨터 또는 기타 장치들에 의해 실행되는 프로그램 모듈과 같은 컴퓨터 실행가능 명령어들의 일반적인 문맥으로 기술될 수 있다. 일반적으로, 프로그램 모듈은 특정 태스크를 수행하거나 특정 추상 데이터 유형을 구현하는 루틴, 프로그램, 객체, 컴포넌트, 데이터 구조 등을 포함한다. 전형적으로, 프로그램 모듈의 기능은 다양한 실시예에서 원하는 대로 결합되거나 분산될 수 있다.

[0052] 도 2는 일 실시예에서 유사성 함수 정의 컴포넌트의 프로세싱을 도시하는 흐름도이다. 블록들(201-209)에서, 컴포넌트는 객체의 각 유형을 선택하고 그 유형의 객체들에 대한 유형 내 및 유형 간 유사성 함수들을 정의하면서 루핑한다. 일 실시예에서, 컴포넌트는 객체들 사이의 유형 내 및 유형 간 관계들을 정의하기 위하여 사용자들과 상호작용할 수 있다. 컴포넌트는 또한 질의의 검색 용어들에 기초한 유사성과 같이, 객체들 사이의 유사성에 재귀적으로 기초하지 않는 유사성 함수들을 정의할 수 있다. 블록(201)에서, 컴포넌트는 객체의 다음 유형을 선택한다. 결정 블록(202)에서, 만약 객체들의 모든 유형들이 이미 선택되었다면, 컴포넌트는 리턴하고, 그렇지 않으면 컴포넌트는 블록(203)에서 계속된다. 블록(203)에서, 컴포넌트는 선택된 유형에 대한 다음 유형 내 관계를 선택한다. 결정 블록(204)에서, 만약 모든 유형 내 관계가 이미 선택되었다면, 컴포넌트는 블록(206)에서 계속되고, 그렇지 않으면 컴포넌트는 블록(205)에서 계속된다. 블록(205)에서, 컴포넌트는 선택된 유형 및 관계에 대한 유형 내 유사성 함수를 정의한다. 그 후 컴포넌트는 다음 유형 내 관계를 선택하기 위하여 블록(203)으로 루핑한다. 블록(206)에서, 컴포넌트는 선택된 유형에 대한 다음 유형 간 관계를 선택한다. 결정 블록(207)에서, 만약 모든 유형 간 관계가 이미 선택되었다면, 컴포넌트는 블록(209)에서 계속되고, 그렇지 않으면 컴포넌트는 블록(208)에서 계속된다. 블록(208)에서, 컴포넌트는 선택된 유형 및 관계에 대한 유형 간 유사성 함수를 정의한다. 그 후 컴포넌트는 다음 유형 간 관계를 선택하기 위하여 블록(206)으로 루핑한다. 블록(209)에서, 컴포넌트는 선택된 유형에 대한 정의된 유형 내 유사성 함수들과 유형 간 유사성 함수들을 결합함으로써 전체 유사성 함수를 정의한다. 컴포넌트는 결합된 유사성 함수들 각각에 가중치 팩터들을 적용할 수 있다. 그 후 컴포넌트는 객체의 다음 유형을 선택하기 위하여 블록(201)으로 루핑한다.

[0053] 도 3은 일 실시예에서 관계 수립 컴포넌트의 프로세싱을 도시하는 흐름도이다. 컴포넌트는 수집된 데이터를 프로세싱하고 관계 데이터를 생성한다. 블록들(301-308)에서, 컴포넌트는 객체의 각 유형을 선택하고 객체의 그 유형에 대한 관계 데이터를 생성하면서 루핑한다. 블록(301)에서, 컴포넌트는 객체의 다음 유형을 선택한다. 결정 블록(302)에서, 만약 모든 유형들이 이미 선택되었다면 컴포넌트는 리턴하고, 그렇지 않으면 컴포넌트는 블록(303)에서 계속된다. 블록(303)에서, 컴포넌트는 선택된 유형에 대한 다음 유형 내 관계를 선택한다. 결정 블록(304)에서, 만약 모든 유형 내 관계들이 이미 선택되었다면 컴포넌트는 블록(306)에서 계속되고, 그렇지 않으면 컴포넌트는 블록(305)에서 계속된다. 블록(305)에서, 컴포넌트는 선택된 유형 및 선택된 유형 내 관계에 대한 관계 데이터의 구성요소들을 설정한다. 그 후 컴포넌트는 다음 유형 내 관계를 선택하기 위하여 블록(303)으로 루핑한다. 블록(306)에서, 컴포넌트는 선택된 유형에 대한 다음 유형 간 관계를 선택한다. 결정 블록(307)에서, 만약 모든 유형 간 관계들이 이미 선택되었다면, 컴포넌트는 객체의 다음 유형을 선택하기 위하여 블록(301)으로 루핑하고, 그렇지 않으면 컴포넌트는 블록(308)에서 계속된다. 블록(308)에서, 컴포넌트는 선택된 유형 및 선택된 유형 간 관계에 대한 관계 데이터의 구성요소들을 설정한다. 그 후 컴포넌트는 선택된 유형에 대한 다음 유형 간 관계를 선택하기 위하여 블록(306)으로 루핑한다.

[0054] 도 4는 일 실시예에서 유사성 함수 풀이 컴포넌트의 프로세싱을 도시하는 흐름도이다. 블록(401)에서, 컴포넌트는 유사성 행렬들을 초기화한다. 예를 들어, 컴포넌트는 최대 유사성을 나타내기 위하여 대각선의 유사성 값들을 1로 설정하고 나머지 유사성 값들을 난수로 설정할 수 있다. 블록(402)에서, 컴포넌트는 적어도 한 번의 반복이 수행되도록 차이 값을 매우 큰 수로 설정한다. 블록들(403-408)에서, 컴포넌트는 유사성 값들이 해에 수렴할 때까지 여러 번의 반복이 유사성 행렬들을 갱신하도록 전체 유사성 함수들을 계산하면서 루핑한다. 블록(403)에서, 컴포넌트는 다음 반복을 선택한다. 결정 블록(404)에서, 유형들의 유사성들에 대한 차이 값들의 합이 차이 임계값보다 작으면 해는 수렴하며 컴포넌트는 리턴하고, 그렇지 않으면 컴포넌트는 블록(405)에서 계속된다. 블록(405)에서, 컴포넌트는 객체의 다음 유형을 선택한다. 결정 블록(406)에서, 만약 모든 유형들이 이미 선택되었다면, 컴포넌트는 블록(408)에서 계속되고, 그렇지 않으면 컴포넌트는 블록(407)에서 계속된다. 블록(407)에서, 컴포넌트는 선택된 유형에 대한 유사성 행렬을 갱신하기 위하여 선택된 유형에 대한 유사성 함수를 계산하고, 그 후 다음 유형을 선택하기 위하여 블록(405)으로 루핑한다. 블록(408)에서, 컴포넌트는 선택된 유형에 대한 이번 반복의 유사성 값들과 이전 반복의 유사성 값들 사이의 차이를 계산한다. 그 후 컴포넌트

는 다음 반복을 시작하기 위하여 블록(403)으로 루핑한다.

[0055] 도 5는 일 실시예에서 유사성 계산 컴포넌트의 프로세싱을 도시하는 흐름도이다. 컴포넌트는 객체의 유형을 전달받고 그 유형에 대한 유사성 행렬들을 갱신한다. 블록(501)에서, 컴포넌트는 전달된 유형에 대한 다음 유형 내 유사성 함수를 선택한다. 결정 블록(502)에서, 만약 모든 유형 내 유사성 함수들이 이미 선택되었다면 컴포넌트는 블록(504)에서 계속되고, 그렇지 않으면 컴포넌트는 블록(503)에서 계속된다. 블록(503)에서, 컴포넌트는 전달받은 유형의 각 객체에 대하여 새로운 유사성 값을 계산한다. 그 후 컴포넌트는 다음 유형 내 유사성 함수를 선택하기 위하여 블록(501)으로 루핑한다. 블록(504)에서, 컴포넌트는 전달받은 유형에 대한 다음 유형 간 유사성 함수를 선택한다. 결정 블록(505)에서, 만약 모든 유형 간 유사성 함수들이 이미 선택되었다면 컴포넌트는 블록(507)에서 계속되고, 그렇지 않으면 컴포넌트는 블록(506)에서 계속된다. 블록(506)에서, 컴포넌트는 선택된 유형 간 유사성 함수를 사용하여 전달받은 유형의 각 객체에 대한 새로운 유사성 값들을 계산한다. 그 후 컴포넌트는 다음 유형 간 유사성 함수를 선택하기 위하여 블록(504)으로 루핑한다. 블록(507)에서, 컴포넌트는 현재의 반복에 대해 전달받은 유형에 대한 전체 유사성을 생성하기 위하여 가중치들을 사용하여 행렬들을 결합한다. 그 후 컴포넌트는 리턴한다.

[0056] 본 분야에서 숙련된 기술을 가진 자들은 예시의 목적을 위하여 유사성 시스템의 특정 실시예들이 기술되었지만 본 발명의 취지 및 범위를 벗어나지 않고서 다양한 변경이 행해질 수 있다는 점을 잘 알 것이다. 따라서, 본 발명은 첨부된 특허청구범위에 의해서를 제외하면 제한되지 않는다.

**발명의 효과**

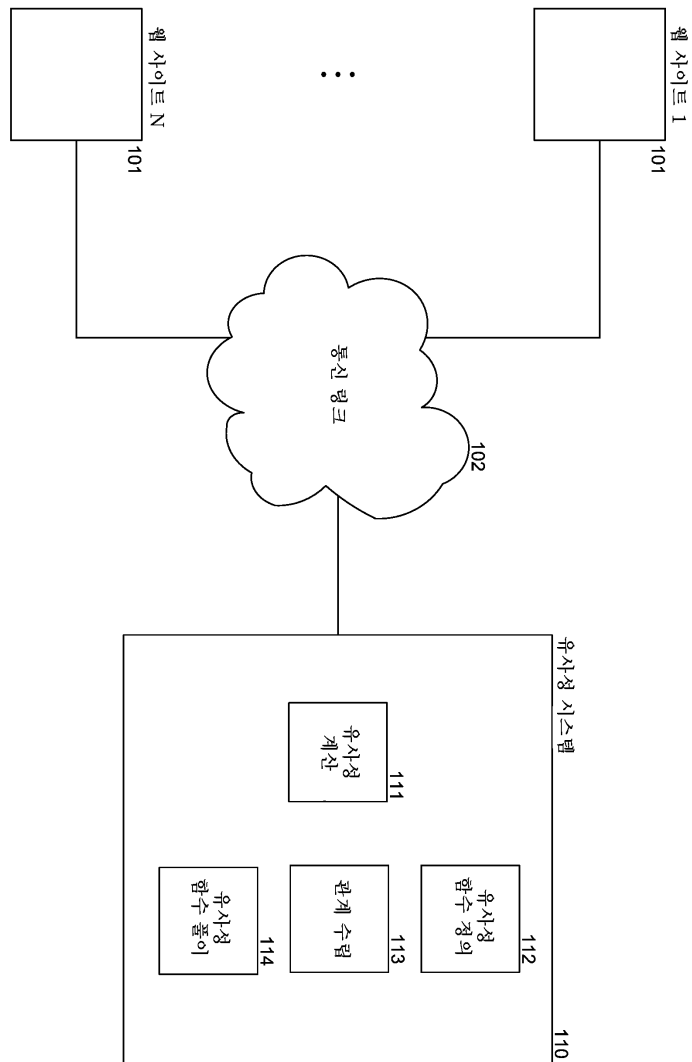
[0057] 동일한 유형의 객체들과 상이한 유형들 사이의 관계, 및 그러한 객체들의 기타 객체들에의 유사성에 기초하여 객체들의 유사성을 측정하기 위한 방법 및 시스템이 제공된다.

**도면의 간단한 설명**

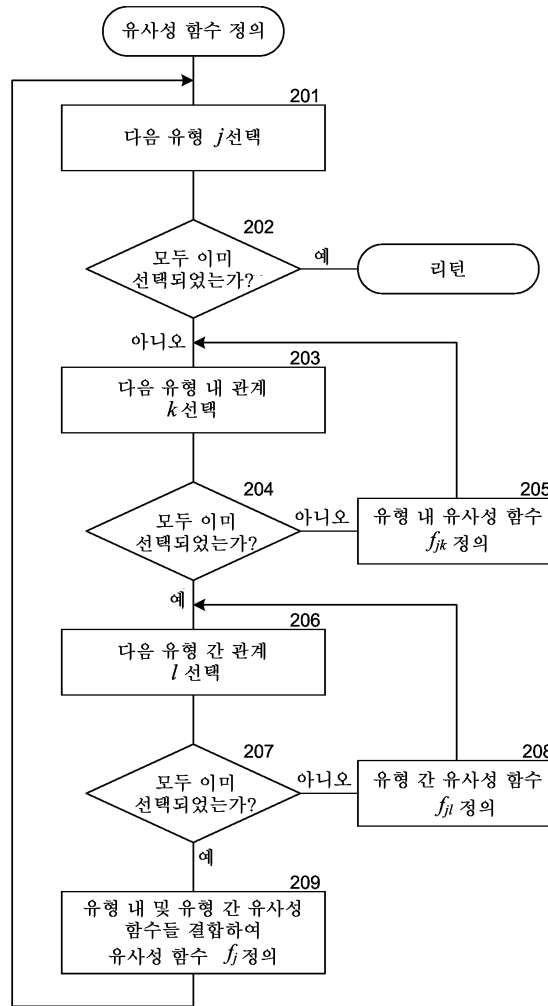
- [0001] 도 1은 일 실시예에서 유사성 시스템의 컴포넌트들을 도시하는 블록도.
- [0002] 도 2는 일 실시예에서 유사성 함수 정의 컴포넌트의 프로세싱을 도시하는 흐름도.
- [0003] 도 3은 일 실시예에서 관계 수립 컴포넌트의 프로세싱을 도시하는 흐름도.
- [0004] 도 4는 일 실시예에서 유사성 함수 풀이 컴포넌트의 프로세싱을 도시하는 흐름도.
- [0005] 도 5는 일 실시예에서 유사성 함수 계산 컴포넌트의 프로세싱을 도시하는 흐름도.
- [0006] <도면의 주요부분에 대한 부호의 설명>
- [0007] 101: 웹사이트
- [0008] 102: 통신 링크
- [0009] 110: 유사성 시스템
- [0010] 111: 유사성 계산
- [0011] 112: 유사성 함수 정의
- [0012] 113: 관계 수립
- [0013] 114: 유사성 함수 풀이

도면

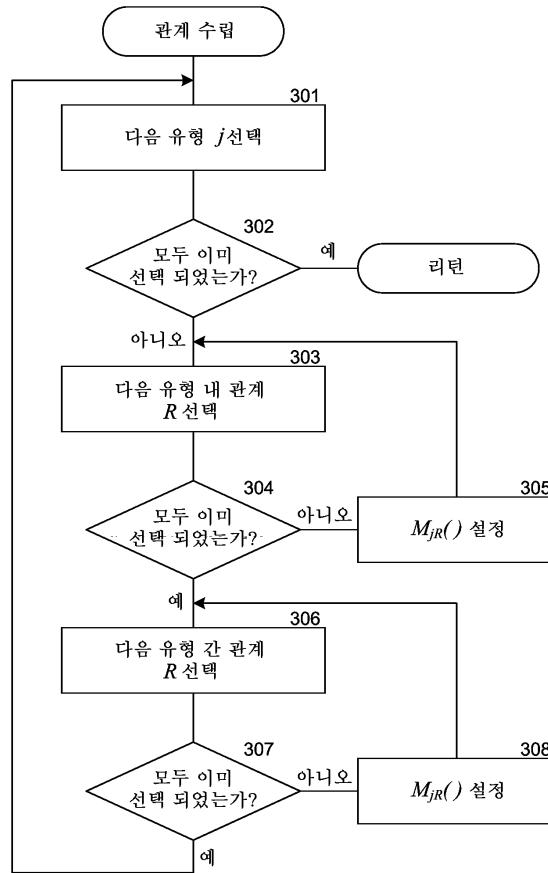
도면1



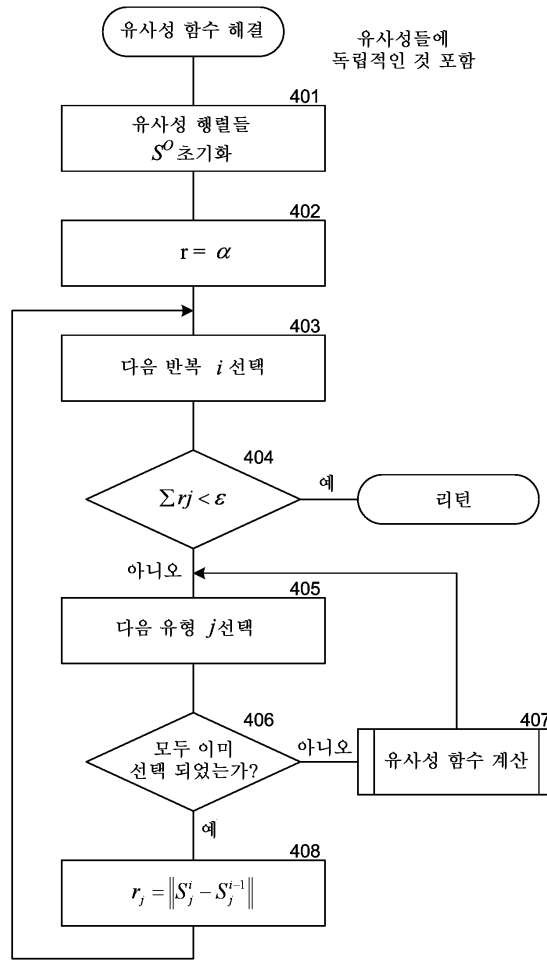
도면2



도면3



도면4



도면5

