



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2012년07월03일
(11) 등록번호 10-1157349
(24) 등록일자 2012년06월11일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06F 17/40 (2006.01)
(21) 출원번호 10-2007-7020803
(22) 출원일자(국제) 2006년03월10일
심사청구일자 2010년04월21일
(85) 번역문제출일자 2007년09월11일
(65) 공개번호 10-2007-0110868
(43) 공개일자 2007년11월20일
(86) 국제출원번호 PCT/US2006/008961
(87) 국제공개번호 WO 2006/099331
국제공개일자 2006년09월21일
(30) 우선권주장
11/078,685 2005년03월10일 미국(US)
(56) 선행기술조사문헌
JP08137898 A*
JP2004240769 A
US20040199498 A1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
야후! 인크.
미국, 94089 캘리포니아, 썬니베일, 퍼스트 애브
뉴 701
(72) 발명자
카푸르, 시암
미국 94087 캘리포니아 썬니베일 히더스톤 웨이
1030
파리크호, 지그나슈
인도 구야라트 잠나가 361008 에어로드롬 로드
오피피. 코디야템플 비하인드 메이여 번갈로우
기리바
(74) 대리인
백만기, 양영준

전체 청구항 수 : 총 23 항

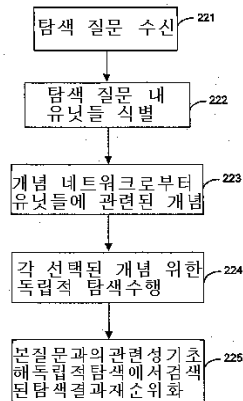
심사관 : 석상문

(54) 발명의 명칭 탐색 결과들의 관련성을 재순위화 및 증가시키기

(57) 요약

인터넷 탐색들의 결과들의 관련성을 재순위화 및 증가시키기 위한 기술들이 제공된다. 탐색 질문이 초기에 개별 유닛들로 파싱된다. 각 유닛은 자연적 개념을 나타내는 하나 이상의 단어들에 해당된다. 탐색 질문 내 상
기 유닛들에 관련된 개념들을 발견하기 위해 개념 네트워크가 분석된다. 상기 개념 네트워크로부터 특정한 개념들이 선택된다. 독립적인 인터넷 탐색들이 상기 선택된 개념들의 각각을 위하여 수행된다. 상기 탐색들로부터의 탐색 결과들은 그런 다음 본래 탐색 질문 내 유닛들과 비교되고, 본래 탐색 질문에 대한 그들의 관련성에 따라 정렬된다.

대표도 - 도2



특허청구의 범위

청구항 1

탐색들로부터 검색되는 탐색 결과들의 관련성(relevance)을 증가시키기 위한 방법으로서,
 네트워크를 통하여, 복수의 사용자로부터 복수의 탐색 질의(search query)를 수신하는 단계;
 상기 복수의 탐색 질의를 분석하는 것에 기초하여 복수의 개념 네트워크(concept network)를 생성하는 단계;
 상기 복수의 개념 네트워크를 생성한 후, 상기 네트워크를 통해 탐색 질의를 수신하는 단계;
 상기 탐색 질의를 수신하는 단계에 대응하여:
 상기 탐색 질의 내의 하나 이상의 유닛들을 식별하는 단계;
 상기 탐색 질의 내의 하나 이상의 유닛들에 관련된, 상기 복수의 개념 네트워크 중의, 개념 네트워크로부터 하나 이상의 개념들을 선택하는 단계 - 상기 하나 이상의 개념들은 상기 하나 이상의 유닛들과 다름 - ;
 상기 개념 네트워크에서의 상기 하나 이상의 개념들에 기초하여, 복수의 탐색 결과를 검색하기 위한 탐색을 수행하는 단계 - 상기 복수의 탐색 결과 중의 각각의 탐색 결과는 순위화(ranking)와 연관됨 -; 및
 상기 탐색을 수행한 후에, 상기 탐색 질의에 대한 상기 복수의 탐색 결과의 연관성에 기초하여 상기 복수의 탐색 결과를 정렬(sorting)하는 단계 - 상기 탐색 질의에 대한 상기 복수의 탐색 결과의 상기 연관성은 상기 탐색 질의 내의 상기 하나 이상의 유닛을 상기 복수의 탐색 결과와 비교하여 결정됨 -
 를 포함하는 탐색 결과 관련성 증가 방법.

청구항 2

제1항에 있어서,
 상기 개념 네트워크로부터의 하나 이상의 개념들을 선택하는 단계는, 상기 복수의 탐색 질의의 서브세트에서 상기 복수의 개념의 서브세트가 얼마나 자주 나타났는지를 결정하는 단계를 더 포함하고,
 상기 하나 이상의 개념들은 상기 개념 네트워크에서의 다른 개념들에 비해 상기 복수의 탐색 질의의 상기 서브세트에서 더 자주 나타났던
 탐색 결과 관련성 증가 방법.

청구항 3

제1항에 있어서,
 상기 하나 이상의 개념들은 적어도 제1 개념 및 제2 개념을 포함하고,
 상기 탐색을 수행하는 단계는 상기 제1 개념에 기초하여 제1 탐색을 수행하는 단계 및 상기 제2 개념에 기초하여 제2 탐색을 수행하는 단계를 포함하며,
 상기 복수의 탐색 결과는 상기 제1 탐색을 수행하는 단계로부터의 탐색 결과들 및 상기 제2 탐색을 수행하는 단계로부터의 탐색 결과들을 포함하는
 탐색 결과 관련성 증가 방법.

청구항 4

제1항에 있어서,
 상기 개념 네트워크로부터의 하나 이상의 개념들을 선택하는 단계는, 상기 하나 이상의 유닛들에 가장 관련된 상기 개념 네트워크에서의 상기 복수의 개념 중에서 미리 결정된 개수의 개념들을 선택하는 단계를 더 포함하는
 탐색 결과 관련성 증가 방법.

청구항 5

제1항에 있어서,

상기 개념 네트워크로부터의 하나 이상의 개념들을 선택하는 단계는, 상기 개념 네트워크에서의 상기 하나 이상의 유닛들에 직접 링크된 개념들 전체를 선택하는 단계를 더 포함하는

탐색 결과 관련성 증가 방법.

청구항 6

제1항에 있어서,

상기 개념 네트워크로부터의 하나 이상의 개념들을 선택하는 단계는, 상기 하나 이상의 유닛들과 동의어(synonym)들인 상기 개념 네트워크에서의 개념들을 선택하는 단계를 더 포함하는

탐색 결과 관련성 증가 방법.

청구항 7

제1항에 있어서,

상기 탐색 질의 내의 하나 이상의 유닛들이 상기 복수의 탐색 결과의 각각에서 얼마나 자주 나타나는지에 기초하여 상기 복수의 탐색 결과의 각각에게 재순위 점수(rerank score)를 할당하는 단계를 더 포함하는 탐색 결과 관련성 증가 방법.

청구항 8

제1항에 있어서,

상기 탐색 질의 내의 하나 이상의 확장 유닛들이 상기 복수의 탐색 결과의 각각에 얼마나 자주 나타나는지에 기초하여 상기 복수의 탐색 결과의 각각에게 재순위 점수를 할당하는 단계를 더 포함하는 탐색 결과 관련성 증가 방법.

청구항 9

제7항 또는 제8항에 있어서,

상기 재순위 점수들은 상기 탐색을 수행하는 탐색 엔진에 의해 할당된 순위 점수에 기초하여 결정되는

탐색 결과 관련성 증가 방법.

청구항 10

제7항 또는 제8항에 있어서,

상기 재순위 점수들의 값들에 기초하여 상기 복수의 탐색 결과를 정렬(sorting)하는 단계를 더 포함하는 탐색 결과 관련성 증가 방법.

청구항 11

제1항에 있어서,

다른 타입들의 질의들로부터 검색되는 탐색 결과들보다 향해 질의(navigational query)들로부터 검색되는 탐색 결과들을 더 높게 순위화하는 단계를 더 포함하는 탐색 결과 관련성 증가 방법.

청구항 12

탐색들로부터 검색되는 탐색 결과들의 관련성을 증가시키기 위한 컴퓨터 시스템으로서,

네트워크를 통하여, 복수의 사용자로부터 복수의 탐색 질의(search query)를 수신하기 위한 코드;

상기 복수의 탐색 질의를 분석하는 것에 기초하여 복수의 개념 네트워크(concept network)를 생성하기 위한 코드;

상기 복수의 개념 네트워크를 생성한 후, 상기 네트워크를 통해 탐색 질의를 수신하기 위한 코드;

상기 탐색 질의를 수신하는 것에 대응하여;

상기 탐색 질의 내의 유닛들을 식별하기 위한 코드;

상기 탐색 질의 내의 상기 유닛들에 관련된 개념 네트워크에서 하나 이상의 개념들을 선택하기 위한 코드 - 상기 하나 이상의 개념들은 상기 하나 이상의 유닛들과 다름 -;

상기 개념 네트워크에서의 상기 하나 이상의 개념들에 기초하여, 복수의 탐색 결과를 검색하기 위한 탐색을 수행하기 위한 코드 - 상기 복수의 탐색 결과 중의 각각의 탐색 결과는 순위화와 연관됨 -;

상기 탐색을 수행한 후에, 상기 탐색 질의에 대한 상기 복수의 탐색 결과의 연관성에 기초하여, 상기 복수의 탐색 결과를 정렬하기 위한 코드 - 상기 탐색 질의에 대한 상기 복수의 탐색 결과의 상기 연관성은 상기 탐색 질의 내의 상기 하나 이상의 유닛을 상기 복수의 탐색 결과와 비교하여 결정됨 - ; 및

상기 코드들을 저장하기 위한 컴퓨터 판독가능 매체

를 포함하는 컴퓨터 시스템.

청구항 13

제12항에 있어서,

상기 개념 네트워크에서의 하나 이상의 개념들을 선택하기 위한 코드는, 상기 복수의 탐색 질의의 서브세트에서 상기 복수의 개념의 서브세트가 얼마나 자주 나타났는지를 결정하기 위한 코드를 포함하고,

상기 하나 이상의 개념들은 상기 개념 네트워크에서의 다른 개념들에 비해 상기 복수의 탐색 질의의 상기 서브세트에서 가장 자주 나타나는 컴퓨터 시스템.

청구항 14

제12항에 있어서,

상기 개념 네트워크에서의 하나 이상의 개념들을 선택하기 위한 코드는, 상기 하나 이상의 유닛들에 가장 관련된 상기 개념 네트워크에서의 개념들로부터의 하나 이상의 개념들 중 미리 결정된 개수의 개념들을 선택하기 위한 코드를 더 포함하는 컴퓨터 시스템.

청구항 15

제12항에 있어서,

상기 개념 네트워크에서의 하나 이상의 개념들을 선택하기 위한 코드는, 상기 개념 네트워크에서의 상기 하나 이상의 유닛들에 직접 링크된 개념들 전체를 선택하기 위한 코드를 더 포함하는

컴퓨터 시스템.

청구항 16

제12항에 있어서,

상기 개념 네트워크에서의 하나 이상의 개념들을 선택하기 위한 코드는, 상기 하나 이상의 유닛들과 동의어들이인 상기 개념 네트워크에서의 개념들을 선택하기 위한 코드를 더 포함하는

컴퓨터 시스템.

청구항 17

제12항에 있어서,

상기 탐색 질의 내의 하나 이상의 유닛들이 상기 복수의 탐색 결과의 각각에서 얼마나 자주 나타나는지에 기초하여 상기 복수의 탐색 결과의 각각에 재순위 점수를 할당하기 위한 코드를 더 포함하는 컴퓨터 시스템.

청구항 18

제12항에 있어서,

상기 탐색 질의 내의 하나 이상의 확장 유닛들이 상기 복수의 탐색 결과의 각각에서 얼마나 자주 나타나는지에 기초하여 상기 복수의 탐색 결과의 각각에게 재순위 점수를 할당하기 위한 코드를 더 포함하는 컴퓨터 시스템.

청구항 19

제17항 또는 제18항에 있어서, 상기 복수의 탐색 결과를 재순위화하기 위한 코드는, 상기 재순위 점수들의 값들에 기초하여 상기 복수의 탐색 결과를 정렬하기 위한 코드를 더 포함하는

컴퓨터 시스템.

청구항 20

제17항 또는 제18항에 있어서,

상기 재순위 점수들은 상기 탐색을 수행하는 탐색 엔진에 의해 할당된 순위 점수에 기초하여 결정되는

컴퓨터 시스템.

청구항 21

하나 이상의 프로세서들에 의해 실행될 때, 제1항 내지 제11항 및 제22항 중 어느 한 항에 기재된 단계들의 수행을 야기하는 컴퓨터 실행가능 명령어들을 갖는(carry) 컴퓨터 판독가능 매체.

청구항 22

제1항에 있어서, 상기 개념 네트워크로부터의 하나 이상의 개념들을 선택하는 단계는 상기 개념 네트워크 내에서 상기 하나 이상의 유닛들에 간접적으로 링크되는 적어도 하나의 개념을 선택하는 단계를 포함하는

탐색 결과 관련성 증가 방법.

청구항 23

제12항에 있어서,

다른 타입들의 질의들로부터 검색되는 탐색 결과들보다 항해 질의(navigational query)들로부터 검색되는 탐색 결과들을 더 높게 순위화하기 위한 코드를 더 포함하는 컴퓨터 시스템.

명세서

기술 분야

[0001]

본 발명은 인터넷 탐색들의 결과들의 관련성을 재순위화 및 증가시키기 위한 기법들에 관한 것으로, 더욱 상세하게는 인터넷 탐색 결과들의 관련성을 증가시키고 개념 네트워크를 이용하여 상기 결과들을 재순위화하기 위한 기법들에 관한 것이다.

배경 기술

[0002]

인터넷의 등장과 월드 와이드 웹(웹)을 통해 사용자에게 이용될 수 있는 다수의 웹 페이지들 및 미디어 콘텐츠에 의해, 사용자들에게 웹으로부터 원하는 정보를 필터링하고 획득하기 위한 합리화된 접근법들을 제공할 필요가 전개되어왔다. 탐색 시스템들 및 프로세스들이 원하는 정보를 획득하기 위한 사용자들의 요구들을 충족시키기 위해 개발되어왔다. 이러한 기술들의 예시들은 야후!, 구글 그리고 다른 사이트들을 통해 접근될 수 있다. 통상적으로, 사용자는 질문을 입력하고, 탐색 프로세스는 상기 질문에 관련된 하나 이상의 탐색 결과들(링크들)을 돌려준다. 돌려받은 탐색 결과들은 사용자가 실제로 찾고 있었던 것에 매우 관련될 수도 있고, 또는 완전히 관련되지 않을 수도 있다. 질문에 대한 탐색 결과들의 관련됨은 부분적으로는 입력된 실제 질문의 기능뿐만 아니라 사용된 탐색 시스템(이하 수집 시스템)의 강건함일 수 있다.

[0003]

사람들은 질문들에 관해 자연스레 생각하지 않는다. 그들은 탐색 엔진들에 질문하거나 도서관 카탈로그들을 찾기 위한 필요에 의하여 부분적으로 우리에게 떠맡겨진 인공적인 구조체이다. 사람들은 단지 단일 단어들에 관해서도 자연스레 생각하지 않는다. 사람들이 생각하는 것은 자연적 개념들에 관한 것이다.

- [0004] 탐색 질문들은 보통 하나 이상의 개념들을 규정하는 여러 단어들을 포함한다. 통상적으로, 탐색 질문 내 단어들의 일부는 다른 것들보다 상기 개념들을 규정하는 것에 더욱 관련된다. 탐색 엔진은 탐색 질문 내 어느 단어가 사용자의 목적에 가장 관련되는지를 알 방법을 갖지 않는다. 결과적으로, 탐색 엔진들은 사용자의 목적에 관련되지 않은 많은 탐색 결과들을 통상 보여준다.
- [0005] 탐색 엔진들은 탐색 질문에 대한 관련성에 따라 탐색 결과들을 순위화한다. 탐색 질문들은 종종, 탐색 엔진이 관련없는 탐색 결과들을 돌려주도록 야기하는 많은 용어들을 포함한다. 종종 탐색 엔진은 탐색 결과들의 목록 내에서 관련없는 탐색 결과들을 높게 순위화한다.
- [0006] 그러므로, 사용자의 본래 목적에 대한 탐색 결과들의 관련성을 증가시키기 위한 인터넷 탐색 기법들을 제공하는 것이 바람직하다.

발명의 상세한 설명

- [0007] 본 발명은 사용자의 목적에 대한 인터넷 탐색 결과들의 관련성을 증가시키기 위한 기법들을 제공한다. 또한, 본 발명은 탐색 질문의 유닛들, 유닛 연관들, 및 유닛 확장들에 대한 탐색 결과들의 관련성을 결정함으로써 탐색 질문의 탐색 결과들을 재순위화하기 위한 기법들을 제공한다.
- [0008] 탐색 질문은 초기에 개별 유닛들로 파싱된다. 각 유닛은 자연적 개념을 나타내는 하나 이상의 단어들에 해당된다. 연관된 유닛들은 탐색 질문에서 함께 있는 둘 이상의 유닛들이나, 새로운 유닛을 형성하기에 충분히 관련되지 않는다. 확장 유닛들은 탐색 질문에서 함께 있는 둘 이상의 유닛들로서 새로운 유닛들을 형성하기에 충분히 관련된다.
- [0009] 본 발명은 탐색 질문 내 유닛들에 관련된 개념들을 발견하기 위한 개념 네트워크를 분석한다. 개념 네트워크는 서로에 관련된 개념들을 링크한다. 질문에서 유닛들에 대한 각자의 관련성들에 기초하여 개념 네트워크로부터 특정한 개념들이 선택된다.
- [0010] 한 실시예에 따르면, 개념들이 사전에 제출된 탐색 질문들에 얼마나 자주 나타났었는지에 기초하여 상기 개념들은 개념 네트워크로부터 선택된다. 사전 탐색 질문들에 더욱 자주 나타났던 개념들이 개념 네트워크로부터 선택된 개념들 중에서 선택된다. 개념 네트워크로부터 선택된 하나 이상의 관련된 개념들을 위해 개별적인 인터넷 탐색들이 수행된다.
- [0011] 각각의 개별 탐색들로부터의 탐색 결과들은 탐색 질문과 비교되고 질문에 대한 그들의 관련성에 따라 정렬된다. 탐색 결과들은 본래 탐색 질문의 유닛들, 연관된 유닛들, 및 확장 유닛들에 대한 그들의 관련성에 기초하여 정렬될 수 있다.
- [0012] 본 발명의 다른 목적들, 특징들, 및 장점들은 하기의 상세한 설명 및 동반된 도면들을 고려함으로써 더 명백해질 것이며, 동일한 참조부호들은 도면들을 통틀어 동일한 특징들을 나타낸다.

실시예

- [0017] 도 1A는 본 발명의 한 실시예에 따라 클라이언트 시스템(120)을 포함하는 정보 검색 및 통신 네트워크(100)에 대한 개략도를 나타낸다. 컴퓨터 네트워크(100)에서, 클라이언트 시스템(120)은 다수의 서버 시스템들(150₁ 내지 150_N)을 갖는 인터넷(140), 또는 다른 통신 네트워크, 예를 들면 임의의 LAN 또는 WAN 접속을 통해 통신할 수 있다. 예를 들면, 클라이언트 시스템(120)은 탐색 결과 서버(160)와 통신할 수 있다. 여기에 기술되는 바와 같이, 클라이언트 시스템(120)은 본 발명에 따라 미디어 콘텐츠, 그리고 웹 페이지들 및 웹 사이트들과 같은 다른 정보를 접근, 수신, 검색 및 디스플레이하기 위하여 임의의 서버 시스템들(150₁ 내지 150_N, 및 160)과 통신하기 위해 구성된다.
- [0018] 도 1A에 도시된 시스템에서 여러 엘리먼트들은 여기에서 설명될 필요가 없는 종래의 잘 알려진 엘리먼트들을 포함한다. 예를 들면, 클라이언트 시스템(120)은 인터넷에 직접적으로 또는 간접적으로 인터페이싱할 수 있는 데스크톱 개인용 컴퓨터, 워크스테이션, 랩톱, PDA, 셀룰러 폰, 또는 임의의 WAP-가능 장치 또는 임의의 다른 컴퓨팅 장치를 포함한다. 클라이언트 시스템(120)은 통상적으로 HTTP 클라이언트, 마이크로소프트사의 인터넷 익스플로러 브라우저, 넷스케이프 내비게이터 브라우저, 모질라 브라우저, 오프라 브라우저, 또는 셀룰러 폰, PDA 또는 다른 무선 장치의 경우에는 WAP-가능 브라우저, 또는 유사 종류와 같은 브라우징 프로그램을 실행시키며, 상기 브라우징 프로그램은 자신에 이용될 수 있는 정보 및 페이지들을 클라이언트 시스템

(120)의 사용자가 인터넷(140)을 통해 접근 및 처리할 수 있도록 한다.

[0019] 또한, 클라이언트 시스템(120)은 서버 시스템들(150₁ 내지 150_N) 또는 다른 서버들에 의해 제공되는 페이지들, 폼들 및 다른 정보와 관련하여 디스플레이(예를 들면, 모니터 스크린, LCD 디스플레이 등) 상에서 브라우저에 의해 제공되는 그래픽 사용자 인터페이스(GUI)와 상호작용하기 위하여 키보드, 마우스, 터치스크린, 펜 또는 유사 종류와 같은 하나 이상의 사용자 인터페이스 장치들(122)을 통상적으로 포함한다. 본 발명은 네트워크들의 특정한 글로벌 인터넷워크로 불리는 인터넷을 사용하기에 적합하다. 그러나, 인터넷 대신에 또는 인터넷에 부가하여 다른 네트워크들, 예를 들면 인트라넷, 엑스트라넷, 가상 사설 네트워크(VPN), 비-TCP/IP 기반 네트워크, 임의의 LAN 또는 WAN 또는 그러한 종류도 사용될 수 있음이 이해되어야 한다.

[0020] 한 실시예에 따르면, 클라이언트 시스템(120)과 그 구성요소들 전부는 인텔 펜티엄 프로세서, AMD 애슬론 프로세서, 또는 유사 종류 또는 다중 프로세서들에 의한 컴퓨터 코드 실행을 포함하는 애플리케이션을 이용하여 운영자에 의해 구성될 수 있다. 여기서 기술되는 바와 같은 데이터 및 미디어 콘텐츠를 통신, 처리 및 디스플레이하기 위해 클라이언트 시스템(120)을 운영 및 구성하기 위한 컴퓨터 코드는 바람직하게는 다운로드되어 하드 디스크에 저장되지만, 전체 프로그램 코드, 또는 그 일부는 또한 임의의 다른 휘발성 또는 비휘발성 메모리 매체 또는 ROM 또는 RAM과 같은 잘 알려진 장치에 저장될 수 있거나, 또는 콤팩트 디스크(CD) 매체, 디지털 비디오 디스크(DVD) 매체, 플로피 디스크, 및 유사 종류와 같이 프로그램 코드를 저장할 수 있는 임의의 미디어에 제공될 수 있다.

[0021] 부가하여, 전체 프로그램 코드, 또는 그 일부는 공지된 인터넷을 통해 소프트웨어 소스, 예를 들면 서버 시스템들(150₁ 내지 150_N) 중의 하나로부터 클라이언트 시스템(120)에 전송 및 다운로드될 수 있거나, 또는 공지된 임의의 통신 매체 및 프로토콜(예를 들면, TCP/IP, HTTP, HTTPS, 이더넷 등)을 이용하여 임의의 다른 공지된 종래 네트워크 접속(예를 들면, 엑스트라넷, VPN, LAN 등)을 통해 전송될 수 있다. 본 발명의 측면들을 구현하기 위한 컴퓨터 코드가 클라이언트 시스템상에서 실행될 수 있는 임의의 프로그래밍 언어, 예를 들면 C, C++, HTML, XML, 자바, 자바스크립트, VBScript와 같은 임의의 스크립팅 언어로 구현될 수 있음이 고려될 것이다. 일부 실시예에서는, 어떠한 코드도 클라이언트 시스템(120)에 다운로드되지 않으며, 필요한 코드는 서버에 의해 실행되고, 또는 클라이언트 시스템(120)에 이미 존재하는 코드가 실행된다.

[0022] 한 실시예에 따르면, 클라이언트 시스템(120) 상에서 실행되는 (모듈(125)로서 도시된) 클라이언트 애플리케이션은 서버 시스템들(150₁ 내지 150_N, 및 160)을 이용하여 통신하고 그들로부터 수신된 데이터 콘텐츠를 처리 및 디스플레이하기 위하여 클라이언트 시스템(120) 및 그 구성요소들을 제어하기 위한 명령어들을 포함한다. 부가하여, 클라이언트 애플리케이션 모듈(125)은 데이터 및 미디어 콘텐츠를 처리하기 위한 다양한 소프트웨어 모듈들을 포함한다. 예를 들면, 애플리케이션 모듈(125)은 탐색 요청들과 탐색 결과 데이터를 처리하기 위한 하나 이상의 탐색 모듈(126), 텍스트 형태의 데이터 및 미디어 콘텐츠와 데이터 프레임들과 활성 윈도우들, 예를 들면 브라우저 윈도우들 그리고 대화 상자들을 렌더링하기 위한 사용자 인터페이스 모듈(127), 및 클라이언트(120) 상에서 실행중인 다양한 애플리케이션들과 인터페이싱하고 통신하기 위한 애플리케이션 인터페이스 모듈(128)을 포함할 수 있다. 또한, 인터페이스 모듈(127)은 클라이언트 시스템(120) 상에 구성되는 기본 브라우저 또는 상이한 브라우저와 같은 브라우저를 포함할 수 있다.

[0023] 한 실시예에 따르면, 탐색 결과 서버(160)가 탐색 결과 데이터 및 미디어 콘텐츠를 클라이언트 시스템(120)에 제공하기 위하여 구성되고, 서버 시스템들(150)은 예를 들면 서버 시스템(160)에 의해 제공된 탐색 결과 페이지들에서 선택되는 링크들에 대한 응답으로 웹 페이지들과 같은 데이터 및 미디어 콘텐츠를 클라이언트 시스템(120)에 제공하기 위하여 구성된다. 한 실시예에서 서버 시스템(160)은 월드 와이드 웹으로부터 정보를 수집하고, 예를 들면 페이지들, 페이지들에 대한 링크들 등을 갖는 하나 이상의 인덱스들을 거주시키기 위한 다양한 수집 기술들을 참조한다. 이러한 수집 기술들은 계층적 구조 내에서 웹 페이지들을 분류하고 순위화하기 위한 인터페이스들 및 수동 또는 반자동 분류 알고리즘들뿐만 아니라, 자동적인 웹 크롤러들, 스파이더들 등을 포함한다. 일정한 측면에서, 서버(160)는 또한 웹 페이지들을 처리 및 순위화하기 위한 탐색 관련 알고리즘들, 예를 들면 구글의 PageRank 알고리즘을 이용하여 구성된다. 서버(160)는 또한 바람직하게도 질문 로그 파일들의 형태로 탐색 질문들을 기록하기 위해 구성된다.

[0024] 한 측면에서, 서버 시스템(160)은 클라이언트 시스템, 특히 탐색 모듈(126)로부터 수신된 다양한 탐색 요청들에 대한 응답으로 데이터를 제공하기 위해 구성된다. 서버 시스템들(150 및 160)은 개별 조직체, 예를 들면 야후! 회사에 의해 사용자들에 제공되는 것과 같은 분산 서버 시스템의 일부일 수 있거나, 또는 이종의 조직체들의 일부일 수 있다. 서버 시스템들(150)과 서버 시스템(160) 각각은 하나 이상의 서버 그리고 연관된 네

이터베이스 시스템을 포함하고, 다중 서버들 및 연관된 데이터베이스 시스템들을 포함할 수 있으며, 단일 블록으로서 도시되어 있을지라도 지리적으로 분산될 수 있다. 예를 들면, 서버 시스템(160)의 모든 서버들은 서로 근접하여(예를 들면, 단일 빌딩 또는 캠퍼스에 위치되는 서버 팜에서) 위치될 수 있거나, 또는 서로 원격인 장소들에 분산될 수 있다(예를 들면, A 도시에 하나 이상의 서버들이 위치되고 B 도시에 하나 이상의 서버들이 위치된다). 여기에서 사용되는 바와 같이, "서버 시스템"이란 용어는 하나 이상의 논리적 및/또는 물리적으로 연결된, 국부적으로 또는 하나 이상의 지리적 장소들에 걸쳐 분산된 서버들을 통상적으로 포함할 것이다. 부가하여, "서버"란 용어는 종래 기술에 공지된 컴퓨터 시스템과 연관된 저장 시스템과 데이터베이스 애플리케이션을 통상적으로 포함한다. "서버" 및 "서버 시스템"이란 용어는 여기에서 교환적으로 사용될 것이다.

[0025] 한 실시예에 따르면, 서버(160)는 클라이언트 시스템(120)으로부터 수신된 탐색 질문들에 대한 응답으로 탐색 결과들을 사용자에게 제공하는 알고리즘들을 포함한다. 본 발명의 실시예에 따르면, 서버 시스템(160)은 하기에 더욱 상세히 논의되는 바와 같이 클라이언트 시스템(120)으로부터 수신된 탐색 질문에 대한 관련성을 증가시키기 위해 구성된다.

[0026] 도 1B는 본 발명의 실시예들을 구현하는 인터넷 탐색 시스템에 대한 개략도를 나타낸다. 탐색 질문(170)이 인터넷 탐색(예를 들면, 웹 탐색)을 개시하기 위해 탐색 엔진(175)에 전송된다. 탐색 엔진(175)은 탐색 코퍼스(corpus)(190)로부터 탐색 질문(170)에 매칭되는 웹 콘텐츠의 위치를 알아낸다. 탐색 코퍼스(190)는 월드 와이드 웹, 인터넷, 인트라넷, 로컬 네트워크들, 및 광대역 네트워크들을 통해 접근될 수 있는 콘텐츠를 나타낸다.

[0027] 탐색 엔진(175)은 탐색 질문(170)에 매칭되는 콘텐츠를 탐색 코퍼스(190)로부터 검색하여 상기 매칭 콘텐츠(즉, 탐색 결과들)를 페이지 어셈블러(180)에 전송한다. 페이지 어셈블러(180)는 탐색 질문에 대한 각각의 관련성에 따라 탐색 결과들을 정렬하고, 사용자에게 용이하게 디스플레이될 수 있는 순서로 상기 결과들을 어셈블링한다. 가장 관련된 탐색 결과들이 탐색 결과 디스플레이 스크린(185)에서 사용자에게 디스플레이된다.

[0028] 본 발명은 사용자의 목적에 대한 인터넷 탐색 결과들의 관련성을 증가시키기 위한 기법들을 제공한다. 도 2는 본 발명의 실시예에 따른 방법론의 예시를 나타낸다. 도 2에 나타난 특정한 단계들이 본 발명의 범위를 제한하도록 의도되지 않음이 이해되어야 한다. 도 2의 방법론에 대한 다수의 수정들이 본 발명의 범위 내에서 이루어진다.

[0029] 사용자는 탐색 질문을 입력함으로써 인터넷 탐색(예를 들면, 웹 탐색)을 개시할 수 있다. 본 발명의 시스템은 도 2에 도시된 바와 같이 단계(221)에서 사용자로부터 탐색 질문을 수신한다. 그런 다음 상기 탐색 질문은 단계(222)에서 유닛들로 파싱된다.

[0030] 탐색 질문들은 유닛들로서 언급되는 구성 부분들로 분해될 수 있다. 질문 처리 엔진은 탐색 질문을 통계적 방법들을 이용하여 유닛들로 분해한다. 한 유닛은 "뉴욕시" 또는 "맹금"과 같은 자연적 개념에 통상적으로 해당되는 하나 이상의 단어 시퀀스들이다. 탐색 질문들로부터 개념 유닛들을 생성하기 위한 기법들의 추가 내용들은 동시계속되고 시암 카푸르(shyam kapur)에 대하여 공동 할당된 2003년 11월 12일자 미국 특허 10/713,576에 논의되어 있으며, 참조로서 통합된다.

[0031] 개념 네트워크는 관련된 개념들 사이의 관계성을 나타내는 구성체이다. 탐색 질문의 각 유닛은 개념 네트워크에서 발견된다. 개념 네트워크는 탐색 질문 유닛들에 관련된 개념들을 식별하기 위해 사용된다. 탐색 질문 유닛이 개념 네트워크에서 발견된 후, 유닛에 관련된 개념 네트워크의 개념들이 단계(223)에서 선택된다.

[0032] 개념 네트워크는 다수의 기법들을 이용하여 개념들에 관련된 링크일 수 있다. 본 발명의 한 실시예에 따르면, 개념 네트워크는 동의어들인 개념들, 더욱 특정한 의미를 갖는 개념들, 더욱 일반적인 의미를 갖는 개념들, 개념의 특정한 실생활 예시들, 및 개념에 유사하게 들리거나 동일한 단어들의 일부를 사용하는 잘 알려진 용어나 명칭들을 링크한다.

[0033] 도 3A에 도시된 예시를 이용하여, 탐색 질문의 유닛이 마천루인 경우, 시스템은 개념 "마천루"를 개념 네트워크에서 발견하고 관련된 개념들을 식별한다. 도 3A는 마천루를 위한 개념 네트워크의 예시를 도시한다. 상기 개념 네트워크에서, 개념 "마천루"는 "빌딩"과 "건물"과 같은 더욱 일반적인 용어들에 링크된다. 마천루는 또한 유사한 용어인 "고층"과 마천루의 유명한 예시인 "엠파이어 스테이트 빌딩"에 링크된다.

[0034] 다른 실시예에 따르면, 사전에 제출된 탐색 질문들이 개념 네트워크에서 관련된 개념들이 얼마나 자주 상기 사전에 제출된 탐색 질문들에 함께 나타났었는지를 결정하기 위해 분석된다. 개념 네트워크는 사전에 제출된

탐색 질문들에서 함께 나타났었던 개념들을 링크함으로써 구성될 수 있다. 단계(223)에서, 사전에 제출된 탐색 질문들에서 가장 자주 함께 나타났었던 관련된 개념들이 개념 네트워크에서 선택된다. 모든 사전에 제출된 탐색 질문들은 분석을 위해 메모리에 저장된다.

[0035] 개념 네트워크는 모든 사용자에게 의해 제출된 질문들에서 함께 나타났었던 개념들에 기초할 수 있다. 다른 예시로서, 개념 네트워크 또는 그의 임의의 일부는 한 특정한 사용자(또는 사용자들의 그룹)에 의해 제출된 탐색 질문들에서 함께 나타났었던 개념들을 링크하는 세션 기반 개념 네트워크일 수 있다. 사전에 제출된 탐색 질문들에서 가장 자주 함께 나타났었던 관련된 개념들이 단계(223)에서 선택된다.

[0036] 도 3B는 세션 기반 개념 네트워크의 예시는 도시한다. 도 3B에서, 주요 개념 "재규어"가 관련된 개념들 "고급 자동차", "XYZ 자동차 회사", 및 "카 레이싱"에 링크되는데, 그 이유는 특정한 사용자가 과거에 자동차 관련 개념들에 "재규어"를 링크시킨 질문들을 제출했기 때문이다. 예를 들어 상이한 사용자는 자신이 재규어 동물들에 관심이 있음을 지시하는 이전 질문들을 제출했었을 수도 있다. 상기 사용자를 위해, 본 발명은 "재규어"를 고양이, 동물원, 또는 사파리와 같은 동물 관련 개념들에 링크하는 상이한 개념 네트워크를 생성한다.

[0037] 본 발명의 다른 실시예에 따르면, 개념 네트워크는 한 명 이상의 사용자에게 의해 특정 시간 기간 동안에 제출된 사전 질문들에서 가장 자주 함께 나타났었던 개념들을 링크할 수 있다. 도 3C는 시간 제한 개념 네트워크의 예시를 나타낸다. 상기 예시에서, 개념 "제인 도우"는 관련된 개념 "제인 도우 라이브 퍼포먼스", "제인 도우 음악 CD들", 및 "기악곡"에 링크된다. 상기 관련된 개념들은 특정한 시간 간격에 걸쳐 사전 탐색 질문들에서 "제인 도우"와 함께 가장 자주 나타났었던 개념들이다. 상기 특정한 시간 간격은 예를 들면 과거 24시간, 과거 일주일, 또는 과거 한 달일 수 있다.

[0038] 도 3C의 예시에서, 개념 네트워크는 과거 24시간 내 최고 인기 탐색 질문들에 기초하여 제인 도우로 명명된 가수에 관련된 개념들에 기초한다. 후속 24 시간 기간 동안에, "제인 도우"를 포함하는 최고 인기 탐색 질문들은 동일한 이름을 갖는 정치인에 관련될 수 있다. 도 3D는 "제인 도우"를 위한 개념 네트워크가 관련된 개념들 "제인 도우 미국 상원의원" 및 "도우 발의권"에 대한 링크를 포함하도록 어떻게 변화되는지를 나타낸다. 개념 네트워크는 최근 질문들에서 유닛 "제인 도우"와 함께 가장 자주 나타났었던 개념들을 포함하도록 갱신된다.

[0039] 본 발명의 실시예에 따르면, 개념 네트워크로부터 가장 근접한 관련된 개념들이 선택된다. 가장 근접한 관련된 개념들은 예를 들면 개념 네트워크에서 주요 개념에 직접 링크된 모든 개념들일 수 있다. 다른 개념들은 직접 링크된 개념들 중의 하나를 통해 간접적으로 주요 개념에 링크될 수 있다. 도 3C는 "기악곡"을 통한 개념들 "제인 도우" 및 "바이올린들" 사이의 간접 접속의 예시를 나타낸다.

[0040] 도 3A 내지 도 3D는 개념 네트워크에서 개념들을 링크하는 관계성들의 몇몇 예시들을 나타낸다. 상기 예시들은 실례적 목적으로 제공될 뿐 본 발명의 범위를 제한하도록 의도되지 않는다. 개념 네트워크에서는 개념들 사이의 많은 다른 관계성들이 가능하다.

[0041] 단계(223)에서 수행되는 선택 프로세스는 임의의 자의적 기준에 기초할 수 있다. 예를 들면, 개념 네트워크에서 상위 다섯 개의 가장 자주 발생하는 관련된 개념들이 단계(223)에서 선택될 수 있다. 다른 예시로서, 가장 자주 발생하는 관련된 개념들 중에서 상위 50% 또는 상위 25%가 단계(223)에서 선택될 수 있다. 많은 다른 선택 기법들이 본 발명에 따라 사용될 수 있다. 여기서 논의되는 예시들은 본 발명의 원리들을 설명하기 위해 제공되며 본 발명의 범위를 제한하기 위해 의도되지 않는다.

[0042] 단계(224)에서, 독립적인 인터넷 탐색들(예를 들면, 웹 탐색들)이 단계(223)에서 선택된 관련된 개념들 중의 하나 이상을 위해 수행된다. 예를 들면, 탐색 질문에 네 개의 유닛들이 있고 단계(223)에서 각 유닛을 위한 한 가지 관련된 개념이 선택되는 경우, 네 개의 독립적인 인터넷 탐색들이 단계(224)에서 수행된다. 한 실시예에 따르면, 많은 수의 관련된 개념들이 단계(223)에서 선택될 때, 인터넷 탐색들은 상기 개념들의 부분집합만을 위해서 수행된다. 예를 들면, 20개의 개념들이 단계(223)에서 선택될 때, 인터넷 탐색들은 탐색 질문의 유닛들 모두에 관련된 상위 5개의 개념들만을 위해 수행된다.

[0043] 탐색 엔진(175)은 단계(223)에서 선택된 개념들을 위해 독립적인 인터넷 탐색들을 수행할 수 있다. 인터넷 탐색들은 임의의 공지된 인터넷 탐색 기법들을 이용하여(예를 들면, 구글 또는 야후! 탐색 기술을 이용하여) 수행될 수 있다.

[0044] 별개 집합들의 탐색 결과들이 단계(224)에서 탐색 엔진(175)에 의해 수행된 개별 인터넷 탐색들 각각을 위해

검색된다. 탐색 엔진(175)은 통상적으로 각 관련된 개념에 대한 그 관련성에 따라 인터넷 탐색들 각각을 위한 탐색 결과들을 정렬한다.

[0045] 단계(225)에서, 본 발명은 단계(224)에서 수행된 인터넷 탐색들로부터 검색된 탐색 결과들을 재순위화한다. 인터넷 탐색들로부터 검색된 탐색 결과들은 전체적인 본래 탐색 질문에 대하여 수행된 인터넷 탐색으로부터 검색되는 탐색 결과들과 그룹핑될 수 있다.

[0046] 탐색 결과들 각각은 본래 탐색 질문의 유닛들, 유닛 연관들, 및 유닛 확장들과 비교된다. 각 탐색 결과에는 본래 탐색 질문에 대한 각각의 관련성에 기초하여 순위 또는 점수가 할당된다.

[0047] 본래 탐색 결과에 대한 탐색 결과들의 관련성은 질문의 유닛들, 연관된 유닛들, 및 확장 유닛들을 탐색 결과들과 비교함으로써 결정될 수 있다. 연관된 유닛들은 탐색 질문에서 함께 있는 둘 이상의 유닛들이나, 새로운 유닛을 형성하기에 충분히 관련되지는 않는다. 확장 유닛들은 탐색 질문에서 함께 있는 둘 이상의 유닛들이고 새로운 유닛을 형성하기에 충분히 관련된다.

[0048] 탐색 결과들은 탐색 질문으로부터의 유닛들, 유닛들의 연관들, 및 유닛 확장들이 얼마나 자주 탐색 결과들에 나타나는지를 결정하기 위해 분석된다. 탐색 결과들에는 유닛들, 연관된 유닛들, 및/또는 확장된 유닛들의 예시들이 탐색 결과들에 나타나는 빈도(또는 상대적 빈도)에 기초하여 새로운 점수가 할당된다. 유닛들, 연관된 유닛들, 및/또는 확장 유닛들의 더욱 많은 예시들을 포함하는 탐색 결과들에 더 높은 점수가 주어진다.

[0049] 본 발명의 실시예에 따르면, 인터넷 탐색들로부터 검색되는 탐색 결과들은 재순위화된다. 재순위화 프로세스에 따르면, 재순위 점수가 각각의 탐색 결과들에 할당된다. 재순위 점수는 예를 들면 탐색 엔진(175)에 의해 할당된 본래 순위 점수를 탐색 결과들에서 질문 유닛들, 연관된 유닛들, 및 확장 유닛들의 빈도수에 기초하여 계산된 새로운 점수와 곱셈함으로써 계산될 수 있다. 그런 다음 탐색 결과들은 재순위 점수에 기초하여 정렬된다.

[0050] 일정 타입들의 탐색 질문들로부터 검색되는 탐색 결과들에는 단계(225)에서 더 높은 점수들이 주어질 수 있다. 예를 들면, 향해 질의들로부터 검색되는 탐색 결과들에는 다른 타입들의 질문들로부터 검색되는 탐색 결과들보다 더 높은 순위가 매겨질 수 있다. 향해 질의들에 주어진 더 높은 점수들은 향해 질의들이 일반적으로 더욱 관련된 탐색 결과들을 검색한다는 인식에 기초한다.

[0051] 일단 재순위 점수들이 본래 탐색 질문에 대한 그 관련성에 기초하여 각각의 탐색 결과들에 할당되면, 상기 탐색 결과들은 최고 재순위 점수들로부터 최저 재순위 점수들까지 단계(225)에서 정렬된다. 최고 재순위 점수들은 본래 탐색 질문에 가장 관련된 콘텐츠를 나타내고, 최저 재순위 점수들은 본래 탐색 질문에 가장 적게 관련된 콘텐츠를 나타낸다.

[0052] 본 발명은 탐색 질문의 유닛들에 관련된 개념들에 매칭하는 콘텐츠를 발견함으로써 인터넷 탐색들로부터 검색되는 탐색 결과들의 관련성을 증가시킨다. 상기 개념들은 상기 논의된 바와 같이 개념 네트워크로부터 선택된다. 상기 탐색 결과들은 그런 다음에 전체 탐색 질문에 기초한 표준 인터넷 탐색으로부터의 탐색 결과들과 결합되고, 탐색 질문에 대한 그들의 관련성에 따라 정렬된다.

[0053] 개념 네트워크를 이용하여 탐색 결과들의 개수를 증가시킴으로써, 탐색 결과들의 적어도 일부가 사용자의 목적이 무엇이든지 탐색 질문 및 사용자의 목적에 용이하게 더 많이 관련되도록 한다. 탐색 질문들이 질문에 대한 관련성에 기초하여 정렬되므로, 가장 관련된 결과들이 먼저 디스플레이된다. 이러한 기법들을 이용하여, 본 발명은 많은 수의 관련된 탐색 결과들을 식별할 수 있다.

[0054] 본 발명이 특정한 실시예들을 참조하여 여기에 기술된 반면에, 어느 정도의 수정, 다양한 변화들, 및 치환들이 본 발명에서 의도된다. 일부 예시들에서, 본 발명의 특징들은 다른 특징들의 상응하는 사용 없이, 전술된 본 발명의 범위에서 벗어남 없이 사용될 수 있다. 그러므로, 본 발명의 필수적 범위 및 정신으로부터 벗어남 없이, 많은 수정들이 개시된 특정한 구성 또는 방법을 적응시키기 위해 이루어질 수 있다. 본 발명이 논의된 특정한 실시예들에 제한되지 않으나, 실시예들 및 청구항들의 범위 내 등가물 모두를 포함할 것이 명백하다.

도면의 간단한 설명

[0013] 도 1A는 본 발명의 실시예들을 구현할 수 있는 인터넷 통신 시스템에 대한 도면,

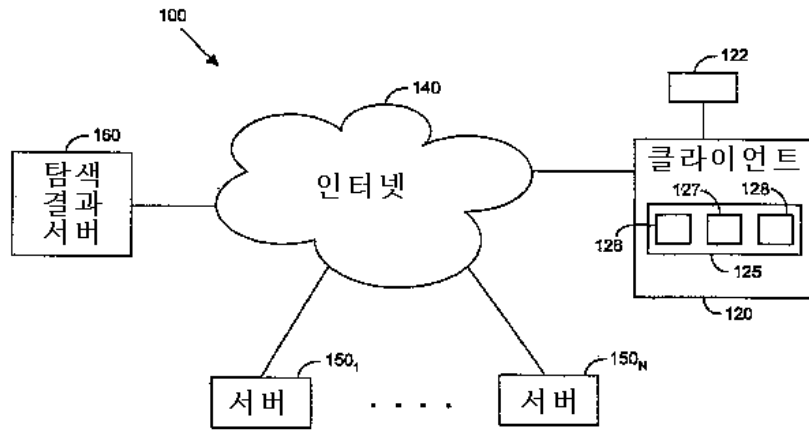
[0014] 도 1B는 본 발명의 실시예들을 구현할 수 있는 인터넷 탐색 시스템에 대한 도면,

[0015] 도 2는 본 발명의 실시예에 따라 인터넷 탐색 결과들의 관련성을 증가시키기 위한 방법론에 대한 흐름도, 및

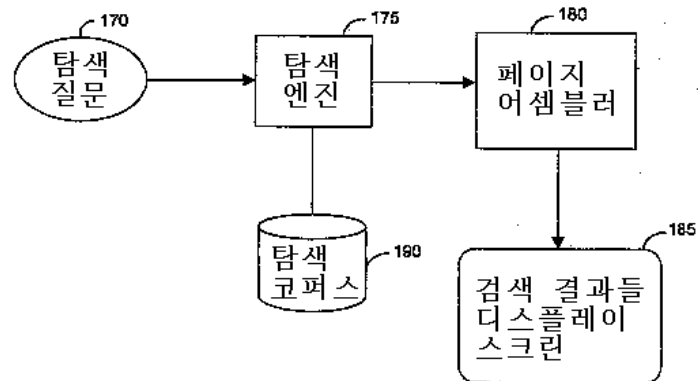
[0016] 도 3A 내지 도 3D는 본 발명에 따라 탐색 질문들을 처리하기 위해 사용될 수 있는 개념 네트워크들의 예시에 대한 도면.

도면

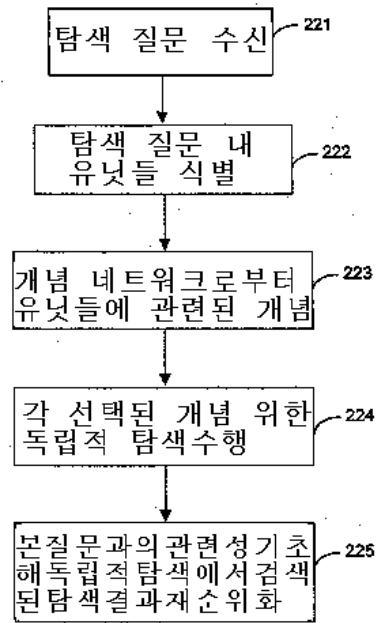
도면1A



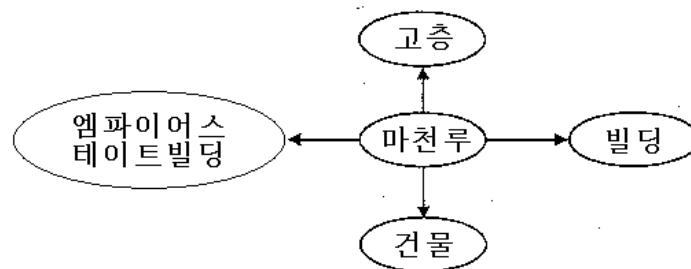
도면1B



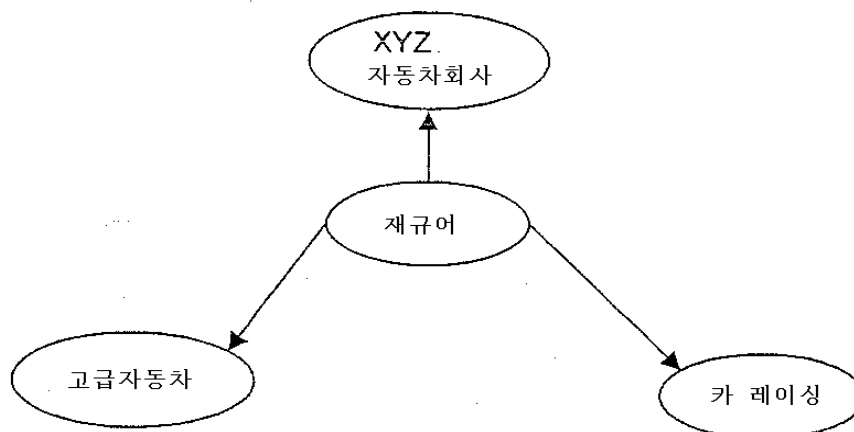
도면2



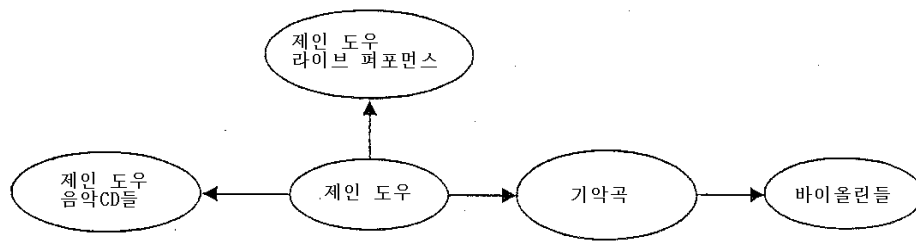
도면3A



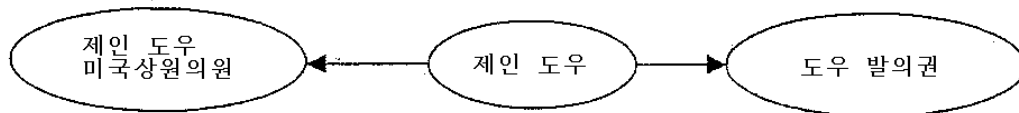
도면3B



도면3C



도면3D



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 제1항

【변경전】

상기 하나 이상의 유닛들을

【변경후】

하나 이상의 유닛들을