



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2008-0080098
(43) 공개일자 2008년09월02일

(51) Int. Cl.
G06F 7/00 (2006.01) *G06F 17/30* (2006.01)
(21) 출원번호 10-2008-7012450
(22) 출원일자 2008년05월23일
심사청구일자 없음
번역문제출일자 2008년05월23일
(86) 국제출원번호 PCT/US2005/045189
국제출원일자 2005년12월13일
(87) 국제공개번호 WO 2007/046829
국제공개일자 2007년04월26일
(30) 우선권주장
11/256,883 2005년10월23일 미국(US)

(71) 출원인
구글 인코포레이티드
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
(72) 발명자
레디 빈두
미국 캘리포니아주 94103 샌프란시스코 #305 하워드 스트리트1097
스파이트 마샬
미국 캘리포니아주 94065 레드우드 시티 시브룩코트 2054
모스버거 님
미국 캘리포니아주 94536 프리몬트 런켈 레인 35706
(74) 대리인
특허법인태평양

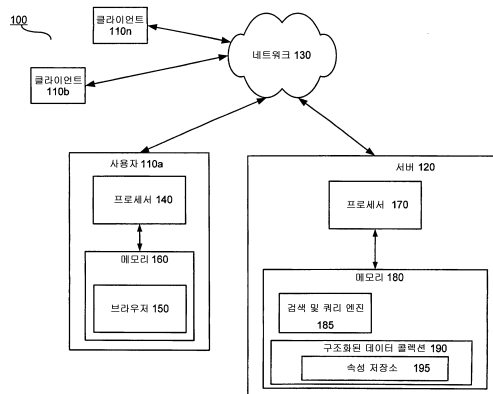
전체 청구항 수 : 총 18 항

(54) 구조화된 데이터에 속성 및 라벨을 추가하는 방법

(57) 요약

본 발명은 라벨 및 속성값을 데이터 컬렉션의 아이템과 관련시키는 방법 및 시스템을 개시한다. 제공자는 속성 및 라벨을 자신의 데이터와 관련시킬 수 있고 또는 속성 및 라벨이 현존하는 데이터에 추가될 수 있다. 바람직한 실시예는 콘텐츠 제공자가 데이터를 업로드하게 하고, 자신들의 고유한 개인화된 라벨 및 속성을 아이템에 첨부하게 하거나 미리 정의된 라벨 및 속성을 사용하게 한다. 제공자는 사용자 인터페이스 또는 벌크 업로드 메커니즘을 사용하여 데이터를 업로드할 수 있다.

대표도 - 도1a



특허청구의 범위

청구항 1

아이템을 구조화하고 검색 가능하게 만드는 방법으로서,
 온라인 인터페이스를 사용하여, 구조화될 데이터 아이템을 수신하는 단계;
 상기 수신된 데이터 아이템과 관련된 속성 이름을 상기 온라인 인터페이스를 통해 수신하는 단계;
 상기 수신된 데이터 아이템과 관련된 속성값을 유도하는 단계; 및
 웹 기반 검색 엔진에 의해 검색 가능한 메모리에, 상기 데이터 아이템을 상기 데이터 아이템 관련 속성 및 속성 값과 함께 저장하는 단계를 포함하는 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 2

청구항 1에 있어서,
 상기 데이터 아이템은 웹 탐색(Crawl)의 결과로서 수신되고, 그 속성값은 상기 웹 탐색의 결과로서 자동적으로 유도되는 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 3

청구항 1에 있어서,
 상기 데이터 아이템은 상기 온라인 인터페이스를 통해 제공자가 다수의 데이터 아이템, 상기 데이터 아이템의 속성 이름 및 속성값을 제공자가 벌크 업로드한 결과로서 수신되는 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 4

청구항 1에 있어서,
 상기 데이터 아이템은 상기 온라인 인터페이스를 통해 제공자가 상기 데이터 아이템, 그 속성 이름 및 속성값을 개별적으로 입력한 결과로서 수신되는 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 5

청구항 1에 있어서,
 상기 속성 이름은 디스플레이된 쿼리(Query) 결과와 함께 디스플레이될 속성인 코어(Core) 속성 이름 집합의 일부인 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 6

청구항 1에 있어서,
 상기 속성 이름은 미리 결정된 속성 이름 집합의 일부인 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 7

청구항 1에 있어서,
 상기 속성 이름은 제공자가 정의한 속성 이름 집합의 일부인 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 8

청구항 1에 있어서,
 제공자는 데이터 아이템의 정보 유형을 위한 속성 이름 및 속성 유형을 정의하고, 상기 정보 유형을 가지는 다수의 데이터 아이템 각각을 위한 속성값을 정의하는 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 9

청구항 1에 있어서,

메모리 내 데이터 구조의 속성을, 유사 데이터 아이템을 위한 속성의 인기(Popularity)에 기초하여 정의하는 단계로서, 상기 데이터 구조는 상기 데이터 아이템과 관련되는 상기 속성을 정의하는 단계를 더 포함하는 아이템을 구조화 및 검색 가능하게 만드는 방법.

청구항 10

아이템을 구조화하고 검색 가능하게 만드는 장치로서,

온라인 인터페이스를 사용하여, 구조화될 데이터 아이템을 수신하는 수단;

상기 수신된 데이터 아이템과 관련된 속성 이름을 상기 온라인 인터페이스를 통해 수신하는 수단;

상기 수신된 데이터 아이템과 관련된 속성값을 유도하는 수단; 및

웹 기반 검색 엔진에 의해 검색 가능한 메모리에 상기 데이터 아이템을 상기 데이터 아이템 관련 속성 및 속성값과 함께 저장하는 수단을 포함하는 아이템을 구조화 및 검색 가능하게 만드는 장치.

청구항 11

청구항 10에 있어서,

상기 데이터 아이템은 웹 탐색의 결과로서 수신되고, 그 속성값은 상기 웹 탐색의 결과로서 자동적으로 유도되는 아이템을 구조화 및 검색 가능하게 만드는 장치.

청구항 12

청구항 10에 있어서,

상기 데이터 아이템은 상기 온라인 인터페이스를 통해 제공자가 다수의 데이터 아이템, 그 속성 이름 및 속성값을 벌크 업로드한 결과로서 수신되는 아이템을 구조화 및 검색 가능하게 만드는 장치.

청구항 13

청구항 10에 있어서,

상기 데이터 아이템은 상기 온라인 인터페이스를 통해 제공자가 상기 데이터 아이템, 그 속성 이름 및 속성값을 개별적으로 입력한 결과로서 수신되는 아이템을 구조화 및 검색 가능하게 만드는 장치.

청구항 14

청구항 10에 있어서,

상기 속성 이름은 디스플레이된 쿼리 결과와 함께 디스플레이될 속성인 코어 속성 이름 집합의 일부인 아이템을 구조화 및 검색 가능하게 만드는 장치.

청구항 15

청구항 10에 있어서,

상기 속성 이름은 제공자가 정의한 속성 이름 집합의 일부인 아이템을 구조화 및 검색 가능하게 만드는 장치.

청구항 16

청구항 1에 있어서,

메모리 내 데이터 구조의 속성을, 유사 데이터 아이템을 위한 속성의 인기에 기초하여 정의하는 수단으로서, 상기 데이터 구조는 상기 데이터 아이템과 관련되는 상기 속성을 정의하는 수단을 더 포함하는 아이템을 구조화 및 검색 가능하게 만드는 장치.

청구항 17

아이템을 구조화하고 검색 가능하게 만드는 장치로서,

온라인 인터페이스를 사용하여, 구조화된 데이터 아이템을 수신하는 사용자 인터페이스;

상기 수신된 데이터 아이템과 관련된 속성 이름을 상기 온라인 인터페이스를 통해 수신하는 사용자 인터페이스;

상기 수신된 데이터 아이템과 관련된 속성값을 유도하는 모듈; 및

웹 기반 검색 엔진에 의해 검색 가능한 메모리에, 상기 데이터 아이템을 상기 데이터 아이템 관련 속성 및 속성 값과 함께 저장하는 모듈을 포함하는 아이템을 구조화 및 검색 가능하게 만드는 장치.

청구항 18

온라인 인터페이스를 사용하여, 구조화된 데이터 아이템을 수신하는 단계;

상기 수신된 데이터 아이템과 관련된 속성 이름을 상기 온라인 인터페이스를 통해 수신하는 단계;

상기 수신된 데이터 아이템과 관련된 속성값을 유도하는 단계; 및

웹 기반 검색 엔진에 의해 검색 가능한 메모리에, 상기 데이터 아이템을 상기 데이터 아이템 관련 속성 및 속성 값과 함께 저장하는 단계를 포함하는 방법을 수행하기 위해 데이터 처리 시스템에 의해 실행될 수 있는 명령어를 저장하는 컴퓨터 판독 가능 매체를 포함하는 컴퓨터 프로그램 제품.

명세서

기술분야

- <1> 본 출원은 35 U.S.C. § 119(e)에 따라 2005년 10월 23일에 출원되고, 발명자 Reddy 등이 발명한 미국 특허출원 제 11/256,883 호, 발명의 명칭 "구조화된 데이터에 속성 및 라벨을 추가하는 방법"에 대해 우선권을 주장한다. 본 출원은 2005년 10월 23일에 출원되고, 발명자 Reddy 등이 발명한 미국 특허 출원 제 11/257,282 호, 발명의 명칭 "구조화된 데이터 검색 방법"과도 관련되고, 이 출원들은 본 출원에 참조로서 포함되어 있다.

배경기술

- <2> 종래 검색 엔진은 월드 와이드 웹 또는 매우 큰 데이터베이스 같은 매우 큰 컬렉션(Collection)을 검색할 수 있다. 검색할 데이터 컬렉션의 크기가 커질수록, 사용자에게 의해 입력된 쿼리 용어와 일치하는 쿼리 결과를 정확하게 얻을 수 없게 되었다. 대신에, 사용자가 검색으로부터 돌아온 대량의 데이터를 분류하도록 돕는 메커니즘(Mechanism)이 제공되는 것이 바람직하다.
- <3> 현재, 다수의 종래 검색 엔진은 쿼리 결과로 돌아온 데이터를 조직화하기 위해 다양한 방법을 사용한다. 그러한 조직 방법의 목표는 사용자가 가장 관심을 갖는 쿼리 결과를 결정하는 것이다. 종래 검색 엔진은 일반적으로 검색 결과를 우선순위화하는 다양한 기술을 사용하지만, 이러한 기술은 사용자가 검색하는 정보의 유형에 대한 가정을 사용해야만 하기 때문에 이상적이지 않다. 예를 들어, 사용자가 "직업(Jobs)"을 입력했다면, 사용자는 직업 공고, 스티브 잡스(Steve Jobs)에 대한 정보, 특정 국가의 직업 통계, 또는 많은 다른 데이터들을 검색하게 된다. 따라서, 종래 검색 엔진을 사용할 때, 사용자는 쿼리 용어로 "직업"이라고만 입력해서는 안된다. 사용자는 아마도 검색을 좁혀줄 추가적인 쿼리 용어도 입력해야 한다. 불행하게도, 사용자는 좁혀줄 용어를 포함하지 않는 적절한 목록을 놓치기도 한다.
- <4> 현재, 월드 와이드 웹에 저장되거나 또는 저장되지 않는 서로 다른 유형의 데이터를 검색하는 것이 매우 어렵다. 종래 검색 엔진은 보통 소수의 소스만으로부터 데이터에 작용한다. 예를 들어, 웹 기반 검색 엔진은 전통적으로 사용자가 월드 와이드 웹 상에서 페이지를 검색하도록 한다. 보통 웹 검색 엔진은 정보 컬렉션을 검색 가능하도록 색인화하는 "후단(Back End)"을 포함한다. 예를 들어, 웹 기반 검색 엔진은 주기적으로 월드 와이드 웹을 탐색하고 탐색된 페이지와 사이트의 색인을 만든다. 다른 검색 엔진은 사용자가 현존하는 데이터베이스를 검색하도록 한다. 그러한 검색 엔진은 미리 결정된 데이터베이스 조직에 의존한다. 예를 들어, 데이터베이스가 알려진 영역 및 속성을 가지고 있으면, 사용자는 그러한 속성 내에서 검색할 수 있다. 예를 들어, XML 데이터베이스는 잘 형성된 XML 입력만을 수용한다. 검색할 데이터가 잘 조직화되어 있지 않다면, XML 데이터베이스는 일반적으로 검색을 위해 데이터를 수용하거나 데이터를 조직할 수 없다.
- <5> 다른 검색 엔진은 사용자가 데이터베이스를 검색하거나 또는 플랫폼(Flat) 조직을 가진 텍스트 문서를 검색하도록 한다. 그러한 검색 엔진은 데이터베이스의 조직 및 그 내부의 문서 조직에 대해 알고 있어야만 한다. 데이터가 저장되는 위치 및 포맷의 다양성은 사용자가 일반적으로 필요한 정보를 찾기 위해서 다중 데이터베이스의 다중

위치에서 검색을 해야만 한다는 것을 의미한다.

<6> 문서 컬렉션은 웹 기반 검색 엔진을 통해 검색 가능하고 따라서 대부분의 사람들에게 쉽게 접속 가능하고, 동시에 다양한 유형의 문서 및 다양한 포맷의 데이터를 포함하는 것이 바람직하다. 추가적으로, 검색 가능한 문서 컬렉션이 사용자가 자신의 검색을 정밀하게 하도록 도울 수 있는 방법으로 조직화되는 것이 바람직하다.

발명의 상세한 설명

<7> 본 발명의 실시예는 라벨 및 속성값과 검색할 데이터 아이템을 관련시킨다. 제공자는 속성 및 라벨을 자신들의 데이터와 관련시킬 수 있거나, 속성 및 라벨을 현존하는 데이터에 추가시킬 수 있다. 바람직한 실시예는 콘텐츠 제공자가 자신의 개인화된 라벨 및 속성을 아이템에 첨부(Attach)하거나 또는 미리 결정된 라벨 및 속성을 사용할 수 있도록 한다. 제공자는 사용자 인터페이스 또는 벌크 업로드 메커니즘을 사용하여 데이터를 업로드(Upload)할 수 있다.

<8> 본 발명의 사상(Teaching)은 이하의 발명의 상세한 설명을 첨부한 도면과 함께 고려함으로써 즉시 이해될 수 있다. 첨부한 도면에서 동일한 참조 번호는 동일한 구성요소를 위해 사용된다.

<9> 도면은 본 발명의 실시예들을 예시의 목적으로만 나타낸다. 당업자는 이하의 발명의 상세한 설명으로부터 본원에 예시된 구조 및 방법의 대체적 실시예가 본원에 설명된 발명의 핵심을 벗어나지 않는 범위에서 개시됨을 직관적으로 인식한다.

실시예

<31> 이하의 단락에서는 본 발명에 따라 구조화된 데이터를 업로드하고 검색하기 위한 시스템의 다양한 실시예를 설명한다.

<32> 도 1a는 본 발명의 바람직한 실시예에 따르는 데이터 처리 시스템을 도시하는 블록 다이어그램(100)이다. 도 1a는 다수의 클라이언트 데이터 처리 시스템(110a, ..., 110n), 네트워크(130), 및 서버 데이터 처리 시스템(120)을 포함한다. 도 1a에서, 예시적 사용자 데이터 처리 시스템(110a)은 프로세서(140), 브라우저(150), 및 메모리(160)를 포함한다. 사용자 데이터 처리 시스템(110) 또는 그 구성요소는 개인용 컴퓨터, 유선 네트워크 컴퓨터, 무선 네트워크 컴퓨터, 이동 전화기 또는 이동 전화기를 포함하는 장비, 휴대용 장비, 초박형 클라이언트 장비, 이들의 조합 등을 포함하는 어떠한 적절한 데이터 처리 시스템도 될 수 있지만 이들에 제한되지는 않는다. 네트워크(130)는 하나 이상의 사용자 데이터 처리 시스템(110)과 서버 데이터 처리 시스템(120) 사이의 통신을 가능하게 하는 어떠한 네트워크도 될 수 있다. 예를 들어, 네트워크(130)는 인터넷, LAN, WAN, 유선 네트워크, 무선 네트워크, 이동 전화 네트워크, 문자 메시지를 송신하는 네트워크, 이들의 조합이 될 수 있지만 이들에 제한되지는 않는다.

<33> 본 발명의 바람직한 실시예에서, 사용자 데이터 처리 시스템(110a)은 사용자가 서버 시스템(120)과 통신할 수 있도록 하고 프로세서(140)에 의해 실행되는 브라우저 소프트웨어(150)를 메모리(160)에 포함한다. 이하에서 상세히 설명될 것처럼, 그러한 브라우저(150)는 사용자가 서버 데이터 처리 시스템(120)과 통신하게 하고, 쿼리 용어를 서버 데이터 처리 시스템(120)에 전송하게 하며, 시스템(120)으로부터 쿼리 결과를 수신하게 한다. 이하에서 더 상세히 설명될 것처럼, 브라우저(150)는 사용자가 쿼리 결과와 관련된 라벨 및 속성을 수신하게 하고, 라벨 및 속성을 사용하여 쿼리 결과를 더 정의하게 한다. 상기한 실시예는 브라우저 기반이지만, 본 발명은 브라우저 기반의 검색에 제한되지 않고 사용자(110)와 서버(120) 사이의 어떠한 적절한 통신 메커니즘도 본 발명의 핵심 및 범위를 벗어나지 않는 범위에서 사용된다.

<34> 본원발명에 설명된 모든 소프트웨어 및 컴퓨터 실행 가능 명령어의 일부는 데이터 처리 시스템의 메모리, CD ROM, 플래시 메모리, 플로피 디스크를 포함하지만 이에 제한되지는 않는 컴퓨터 판독 가능 매체에 컴퓨터 프로그램 제품으로 저장될 수 있거나, 또는 네트워크를 통해 또는 시스템 구성요소들 사이에서 신호로서 송신될 수 있다.

<35> 서버 데이터 처리 시스템(120)은 서버 시스템(120)이 쿼리 용어를 구조화된 데이터 컬렉션(190)에서 검색할 수 있도록 하기 위해 검색 및 쿼리 엔진 소프트웨어(185)를 실행하는 프로세서(170)를 포함한다(검색 및 쿼리 엔진(185)은 "검색 엔진"으로도 불림). 구조화된 데이터의 일례는 필드화된(Fielded) 데이터, 즉 각각이 하나 이상의 데이터 필드(이름, 주소, 상태 등과 같은)를 갖는 데이터 아이템들이다.

<36> 메모리(180)는 속성 저장소(195)도 포함하는데, 이는 구조화된 데이터(190)에서 일부 또는 모든 데이터 아이템

을 위한 속성(및 라벨)을 저장한다. 저장소는 도 5와 관련하여 이하에서 설명한다. 저장소(195)는 구조화된 데이터 컬렉션(190)의 일부로 도시되었지만, 저장소(195)는 데이터 컬렉션(190)으로부터 분리도 된다.

- <37> 큰 검색 엔진 및 큰 데이터 컬렉션은 분산(Distributed) 데이터 처리 시스템, 협동(Cooperating) 데이터 처리 시스템, 네트워크 데이터 처리 시스템 등을 포함하지만 이에 제한되지는 않는 많은 방식으로 저장되지만, 도 1a에서는 검색 엔진(185), 저장소(195), 및 구조화된 데이터 컬렉션(190) 모두가 단일 메모리(180) 내에 도시되었다. 검색 엔진(185)은 소프트웨어, 하드웨어, 펌웨어, 또는 이들의 어떠한 조합도 가능하다.
- <38> 바람직한 실시예에서, 쿼리 용어는 하나 이상의 다수의 사용자 시스템을 통해 사용자에게 의해 입력되고, 네트워크(130)를 통해 서버 데이터 처리 시스템(120)으로 송신된다. 데이터 컬렉션을 수신, 색인화, 및 검색하기 위해 서버(120)에 의해 사용되는 방법의 자세한 내용이 본원발명에서 설명된다.
- <39> 도 1b는 본 발명의 바람직한 실시예에 따르는 다른 데이터 처리 시스템을 도시하는 블록 다이어그램(111)이다. 도 1b에서, 사용자는 자신의 기계(110)에 개인 데이터 컬렉션(190)을 저장한다. 개인 검색 엔진은, 사용자 및 가능하게는 네트워크(130)를 통해 다른 사용자에게 의해 검색 가능하도록, 이 데이터에 접속하고 조직화하는 것이 고려된다. 그러한 시스템은 데이터베이스 및 다른 유형의 데이터 컬렉션이 중앙 검색 엔진에 의해 접속 가능한 검색 가능 문서의 풀(Pool)에 추가될 수 있도록 한다.
- <40> 도 1b의 실시예에서, 데이터 컬렉션(190)은 사용자의 데이터 처리 시스템(110) 또는 엔터프라이즈(Enterprise) 서버(도시되지 않음)에 저장되고, 사용자, 사용자들 중 더 작은 부분 집합, 또는 데이터 컬렉션(190)에 어떻게 접속하는지 알고 있는 모든 사용자와 같이 선택된 그룹의 사람 또는 개인이 이용 가능하도록 만들어진다. 그러한 경우에, 상기한 것처럼 속성 및 라벨을 통해 검색을 필터링하는 능력이 컴퓨터에서 또는 컴퓨터의 로컬 네트워크에서 지역적으로 실행되는 개인 검색 엔진(185)의 일부가 될 수 있다. 예를 들어, 캘리포니아 마운틴 뷰에 있는 구글 사에서 제공하는 구글 데스크탑 검색 툴은 사용자의 데스크탑에서 실행되는 검색 툴이고 사용자의 개인용 컴퓨터 상의 데이터를 색인화한다. 본 발명을 포함하는 구글 데스크탑의 구현은 사용자에게 사용자의 데스크탑에 저장되거나 그로부터 접속 가능한 데이터베이스 및 다른 유형의 데이터 컬렉션을 검색하는 능력을 제공한다.
- <41> 구글 데스크탑의 구현은 사용자에게 유용한 속성 및 라벨로 자신의 데이터를 조직하는 능력도 제공한다. 예를 들어, 대학 도서관은 모든 온라인 컬렉션을 대학의 학생, 교직원, 및 졸업생이 사용 가능하도록 할 수 있다. 그러한 경우에, 정보는 공개적으로 사용 가능한 서버에 있지 않으나, 대학의 서버에 저장되고 대학 데이터 제공자에게 의해 접속이 허락된 사람(및 프로그램)에게만 사용 및 검색이 가능하다. 이 예에서, 대학은 어떤 제공자가 데이터 컬렉션에 추가할 능력을 가지는지도 제어할 수 있다.
- <42> 도 1c는 본 발명의 바람직한 실시예에 따르는 아키텍처 다이어그램(131)이다. 설명된 실시예에서, 제공자는 데이터 및 속성을 시스템에 입력하는 세 가지 방식 중 하나 이상을 사용할 수 있다. 제공자측 전단(Front End, 132)은 제공자가 그러한 목적을 위해 제공되는 사용자 인터페이스를 사용하여 데이터 아이템 및 속성을 입력하도록 한다(예를 들어, 도 6b 참조). 제공자는 데이터 아이템의 벌크 업로드(133)도 수행할 수 있다(예를 들어, 도 8a-8d 참조). 제공자는 특정 URL로부터 아이템을 업로드(134)도 할 수 있다(예를 들어, FTP를 사용함). 검색 및 쿼리 엔진(185)은 데이터 컬렉션(190)의 아이템을 색인화하고, 데이터 컬렉션(190)은 바람직하게는 모든 데이터의 색인(137)을 생산하기 위해 데이터 아이템에 대하여 입력된 속성 및 속성값을 포함한다. 검색 엔진(185)은 사용자가 쿼리를 입력하도록 한다(예를 들어, 도 4a 참조). 시스템은 소프트웨어 프로그램이 검색 엔진(185)을 통해 데이터를 쿼리하도록 하는 API(Application Program Interface)도 포함한다.
- <43> 도 2a는 본 발명의 바람직한 실시예에 따라 검색 가능한 데이터 아이템 컬렉션을 만드는 개관을 도시하는 순서도(200)이다. 도 6a-6e 및 도 8a-8d와 관련하여 이하에 설명된 것처럼, 서버(120)는 데이터 아이템 컬렉션을 수신한다(202). 이 데이터는 표준 웹 탐색의 결과로서 수신되거나 또는 자신의 데이터가 검색 가능하도록 되기를 원하는 하나 이상의 제공자에게 의해 제공될 수 있다. 수신된 데이터 아이템 컬렉션은 이하에 설명될 것처럼 라벨, 속성, 및 속성값을 추출하기 위해 처리되고 그러한 라벨, 속성, 및 속성값은 다양한 정보 유형과 관련된다. 특정 환경에서, 사용자는 입력된 데이터의 일부 또는 전체를 위해 속성 이름 및/또는 속성값을 제공한다. 예를 들어, 사용자가 의학 저널 컬렉션을 보관하기 위해 만든 데이터베이스를 업로드한다고 한다. 사용자는 "저널", "발행 년도", "저널 이름"과 같은 속성이름을 반영하는 값으로 이러한 저널에 대한 속성을 특정할 수 있다. 사용자는 "의학", "치학", "하바드로부터" 등과 같은 저널 각각을 위한 0 개 이상의 라벨을 입력한다. 라벨은 그와 관련된 값을 갖지 않는 특별한 종류의 속성이다(값없는 태그라고도 함). 요소(204)의 상세한 내용을 도 3a과 함께 설명한다.

- <44> 도 2b는 본 발명의 바람직한 실시예에 따라 문서 콜렉션을 검색하고 검색을 걸러내는 개관을 도시하는 순서도 (210)이다. 실시예에서, 사용자는 쿼리 용어(도 4a의 스크린 샷(400)에서 "암 수용체(402)"와 같은 쿼리 용어)를 입력한다(212).
- <45> 특정 실시예에서, 사용자는 속성 이름 및 값을 영역(402)에 입력된 쿼리의 일부로서 입력한다. 예를 들어, 사용자는 영역(402)에 다음의 내용을 입력한다.
- <46> 암 수용체 속성(저널유형:의학)
- <47> 쿼리 결과의 일부 아이템이 저널유형으로 명명된 속성을 가지지만 속성이 코어 속성 집합의 일부분이 아니라는 것을 사용자가 안다면, 사용자는 의학 저널만을 돌려받기를 원한다.
- <48> 도 3b와 관련하여 이하에서 상세히 설명될 것처럼, 시스템은 쿼리 결과를 결정한다(213). 일부 실시예에서, 쿼리 결과는 이 시점에서 디스플레이된다(213). 다른 실시예에서, 쿼리 결과는 아직 디스플레이되지 않고, 대신에 사용자는 쿼리 용어에 특정하는 라벨 및/또는 속성을 선택함으로써 자신의 검색을 더 걸러내도록 요청받는다. 예를 들어, 도 3d에 도시된 것처럼, 사용자는 라벨 및 속성을 특정함으로써 자신의 검색을 걸러낼 수 있다(214).
- <49> 도 3a는 데이터 아이템 콜렉션으로부터 라벨 및 속성을 추출하는 방법을 도시하는 순서도이다. 이 방법은 데이터 콜렉션이 검색될 수 있도록 데이터 콜렉션을 조직하기 위해 사용된 셋업(Set-up) 과정의 일부이다.
- <50> 각각이 정보 유형을 가지는 데이터 아이템이 수신되면, 시스템은 이 정보 유형을 위한 라벨 및 속성을 결정한다(304). 속성은 "저널"과 같은 이름을 가지고, 하나 이상의 가능한 저널 이름의 값을 가지는 이름/값의 쌍(Pair)이다.
- <51> 바람직한 실시예에서, 속성 및 라벨은 데이터 제공자에 의해 특정된다. 따라서, 속성의 결정은 사용자가 제공한 속성 및 라벨을 식별하는 문제에 불과하다.
- <52> 특정한 경우에, 데이터 제공자는 자신의 아이템을 위한 속성 및 라벨을 특정하지 않는다. 예를 들어, 아이템이 웹 탐색자(Crawler)에 의해 위치한 웹 페이지라면, 웹 페이지의 소유자는 자신의 페이지를 위한 속성 또는 라벨을 특정하기 위한 기회를 갖지 못한다. 따라서, 다른 바람직한 실시예에서, 라벨 및 속성은 데이터 콜렉션을 위해 소프트웨어에 의해 유도된다. 라벨 및 속성을 유도하는 단계는 미리 결정된 라벨 및 속성 목록을 위한 잠재적 값이 소프트웨어에 의해 데이터 콜렉션 내에서 발견되는 순수하게 자동화된 과정과 관련된다. 예를 들어, 판매를 위한 아이템 목록에서(예를 들어, 구글의 프루글 시스템), 미리 결정된 기준(Criteria)을 만족시키는 가격의 금액(Amount)은 그 아이템을 위해 "가격" 속성의 값으로 지정된다. 다른 바람직한 실시예에서, 소프트웨어는 소프트웨어가 속성/값 쌍을 제안하고, 제공자에 의해 속성/값 쌍이 수용되거나 거절되는 제공자와의 쌍방향 과정을 수행한다. 다른 바람직한 실시예에서, html 태그가 스캔되고, 발견된 정보는 태그를 가지는 페이지를 위한 속성 값을 유도하기 위해 사용된다. 예를 들어, 페이지가 다음과 같은 html 코멘트를 포함한다고 가정한다.
- <53> <! 현재 가격은 <http://www.todayspricesforbigco.com?id=32423490> 에 있음 !>
- <54> 소프트웨어는 지시된 URL로부터 현재 가격을 획득하고, 그 웹 페이지를 위한 가격 속성값을 만들어낸다.
- <55> 속성 및 라벨이 데이터 아이템과 관련되면(306), 데이터 아이템은 검색될 수 있도록 색인화된다(309). 바람직한 제 1 실시예에서, 속성과 라벨, 및 그 값들이 색인화되지만, 다른 바람직한 실시예에서는 개별적으로 검색되거나 개별적으로 색인화된다.
- <56> 도 5a는 검색 가능한 데이터 콜렉션을 위한 속성 및 라벨을 저장소(195)에 저장하기 위해 사용되는 데이터 포맷(500)을 도시한다. 각 아이템은 그 유형에 적합한 특정 속성 및 라벨과 관련된다. 예를 들어, 직업 공고는, 직업 기능 - 제품 관리; 고용주 - ABC 사; 및 직업 유형 - 전문적인 속성을 가진다. 바람직한 실시예에서 속성 및 라벨은 다음 유형의 값을 가질 수 있다.
- <57> BOOLEAN
- <58> INT
- <59> FLOAT
- <60> URL

<61> STRING

<62> LOCATION

<63> DATE

<64> DATE RANGE

<65> 속성 및 라벨은 저장 장치에 다음의 메타태그(Metatag)에 의해 지시된다.

<66> <start name>

<67> 이름

<68> </end name>

<69> <start value>

<70> 값

<71> </end value>

<72> 따라서, 바람직한 실시예에서, 각 속성은 "저널"인 속성 이름 및 "염증에 대한 저널"인 "저널" 속성에 대한 값과 같은 이름/값 쌍이다(도 5b 참조). 각 라벨은 특정 저널이 의학 저널이라는 것을 지시하는 "의학"과 같은 이름만을 가진다(도 5c 참조). 바람직한 실시예에서, 데이터 아이템의 정보 유형도 라벨 중 하나의 이름이다. 따라서, "이벤트 및 활동"이라는 정보 유형을 가진 데이터 아이템도 같은 이름의 라벨을 가진다. 이런 방식으로, 사용자는 데이터 아이템의 정보 유형과 같은 이름을 가진 라벨을 특정함으로써 특정 정보 유형을 가진 데이터를 검색할 수 있다.

<73> 도 5d는 정보 유형을 그 속성에 매핑시키기 위한 예시적 데이터 구조를 도시한다. 따라서, 데이터 컬렉션(190)에 있는 아이템이 "제품"이라는 정보 유형을 가지면, 아이템의 속성은 "제품"이라는 정보 유형을 위한 속성 및 속성 유형을 포함하는 도 5c의 데이터 구조에 접속함으로써 결정될 수 있다.

<74> 도 5d에 도시된 것처럼, 각 정보 유형은 미리 결정된 속성이다. 속성값은 속성 유형의 값이다. 도 5e는 일부 실질적 값을 도시한다. 따라서, "저널"이라는 정보 유형은 속성 유형이 스트링(String)인 값을 가지는 "저널 이름" 속성 및 널(Null) 값을 가지는 "의학"이라는 라벨을 가진다. 예를 들어, 그러한 속성은 사용자가 특정 저널 제목을 검색할 수 있게 하거나 모든 의학 저널을 검색할 수 있게 한다. 유사하게, "제품"이라는 정보 유형은 판매 가능한 특정 제품의 개수를 지시하고 정수의 속성 유형을 가지는 "NumAvail" 속성을 가진다. 모든 속성은 선택적이다. 제공자는 제안된 속성을 채우거나(Populate) 자신만의 속성을 만드는 것을 선택할 수 있다.

<75> 도 3b는 수신된 쿼리 용어 또는 용어에 대응하여 쿼리 결과를 디스플레이하는 방법(300)을 도시하는 순서도이다. 바람직한 실시예에서, 쿼리 결과는 검색 엔진(185)에 의해 결정된다. 예를 들어, "암 수용체"라는 쿼리(도 4a 참조)는 도 4b에 도시된 것과 같은 속성(404)을 가지는 아이템의 쿼리 결과(406)를 가져온다(312). 상기한 것처럼, 본 발명의 일부 실시예는 이 시점에서 쿼리 결과(406) 결정하지만 디스플레이하지는 않는다.

<76> 쿼리에 대한 쿼리 결과가 결정되면(선택적으로 디스플레이되지 않음), 쿼리 결과에 대한 최소한 일부의 속성 이름 및 라벨이 디스플레이된다(322). 데이터 집합의 데이터 아이템(406)은 특정 정보 유형을 가진다. 처음에 디스플레이되는 속성(404)은 쿼리 결과(406)에서 데이터 아이템의 정보 유형을 위한 속성의 일부 또는 전체이다. 쿼리 결과는 데이터 아이템을 가지고, 각각의 데이터 아이템은 서로 다른 속성을 가진다. 쿼리 결과의 위쪽에 보이는 속성은 쿼리 결과에서 가장 흔하고, 검색자들에 의해서 가장 많이 클릭되거나 걸러진 속성이다. 예를 들어, "주거(Housing)"라는 쿼리는 침실 및 욕실 속성을 가진 많은 아이템을 가지고, 검색자는 항상 주거라는 쿼리에 대해서 "침실" 및 "욕실" 속성에 의해 걸러내기를 한다. 그래서 침실 및 욕실은 검색 결과에서 제일 먼저 보인다.

<77> 도 4b는 쿼리 결과(406), 다수의 속성 및 라벨 이름(404, "저널", "공중보건", "뉴스 소스", "저자")을 도시한다. 각 속성 다음의 숫자는 검색 결과(406)와 관련된 속성을 가진 검색 결과에서 아이템의 개수를 지시한다. 예를 들어, 도 4b에서, 쿼리 결과(406)는 "저널"과 관련된 속성/라벨을 가진 2050 개의 아이템을 포함한다. 따라서, 특정 쿼리 결과와 함께 도시된 속성의 개수 및 주제(Identity)는 쿼리 의존적이고, 나중에 검색 결과를 좁히기 위해 선택되는 속성 및 라벨에 더 의존적이다.

- <78> 도 3c는 주어진 쿼리 결과(406)에 대해 어떤 속성을 디스플레이할 것인지 결정하는 방법을 도시하는 순서도 (340)이다. 최종 사용자가 검색을 수행할 때, q 개의 가장 적절한 결과가 검색 엔진(185)에 의해 결정되고 (341), q 개의 가장 적절한 결과를 위해 n 개의 가장 인기있는 속성이 결정된다(342). 상위 n 개의 속성 이름을 위해, 시스템은 상위 m 개의 속성/라벨 값을 결정한다(344). 그리고나서, 히스토그램을 계산하거나, 적절한 결과 집합에서 매칭(Matching)되는 오퍼(Offer)의 개수를 세어 오퍼 카운트(Count)를 계산한다. q, n, 및 m 값은 모두 설정 가능하다. 이하의 값에 제한되는 것은 아니지만 예시적으로, q는 1,000 개에서 100,000,000 개(q는 특정 쿼리 용어와 매칭되는 모든 결과로 설정될수도 있음), N은 100 개 부근 및 M은 20 개에서 100개의 값이다.
- <79> 바람직한 실시예에서, 속성은 히스토그램이 결정되기 전에 표준화된다(346). 특정 구현에서, 소정량의 데이터 소거(Cleanup) 및 표준화는 데이터가 초기에 데이터 콜렉션(190)에 저장될 때 이루어진다. 설명된 실시예에서, 데이터 표준화는 검색되는 쿼리 용어에 기초하여 즉시(On the fly) 이루어진다(예를 들어, 쿼리 용어가 "자동차"일 때 모든 "브랜드" 속성을 "제작"으로 표준화하는 것이 의미있지만, 쿼리가 "핸드백"이면 모든 제작 속성을 "브랜드"로 표준화하는 것이 의미있음). 다른 실시예는 데이터가 데이터 콜렉션(190)으로 수신될 때 더 많은 표준화를 한다. 데이터 표준화는 바람직한 실시예에서 다음과 같이 이루어진다.
- <80> 1. 보충하기(Stemming) - 예를 들어, 레스토랑 = 레스토랑들
- <81> 2. 축약하기(Abbreviation) - 예를 들어, sz = 크기
- <82> 3. 동등한 단위 - 예를 들어, 무게 = 온스, 파운드 등
- <83> 4. 철자 수정 시도
- <84> 보충하기는 변형 및 오기를 데이터 콜렉션(190)으로 끌어들여 함으로써 제공자가 자신의 속성 이름을 특정하는 시스템에서 특히 유용하다. 예를 들어, 보충하기는 사용자가 "저널", "저널들", "저럴" 속성 이름을 보충된 속성인 "저널들"의 단일 선택으로 필터링하도록 한다.
- <85> 특정 바람직한 실시예에서, 제공자에 의해 추가된 속성은 확인된 유형이다. 예를 들어, URL, DateTime, Number, String, Location, Boolean 속성은 그들이 유효한 값을 가지는지 알아보기 위해 확인된다. 일부 실시예는 각각의 URL 값이 활성인지 알아보기 위해 접속을 확인하는데(Ping), 이는 다양한 구현에서 선택적이다. 바람직한 실시예에서, 위치는 지리적으로 코딩되어 예를 들어, 구글맵과 같은 온라인 지도 서비스에 의해 참조될 수 있다. 특정 실시예에서, 지리적으로 코딩될 수 없는 "위치" 속성은 유효하지 않은 것으로 고려된다.
- <86> 인기있는 속성 및 라벨이 결정되고 디스플레이되면(322)(도 3b 참조), 사용자는 쿼리 결과를 위해 하나 이상의 디스플레이된 라벨 및 속성값을 특정하도록 허용된다(324)(도 3d 참조).
- <87> 도 4c는 사용자가 도 4b로부터 "저널" 속성을 선택하고 자신의 검색을 한정하기 위해서 영역(408)에 저널의 이름을 입력하기 위해 준비하는 예를 도시한다. 쿼리 용어(402)는 이제 "암 수용체 필터 : 저널"이라는 것을 주목한다. "저널" 속성은 나열된 코어 속성(404)에서 사라졌다.
- <88> 유사하게, 도 4d에서, 사용자는 제 2 속성인 "년도(410)"를 선택하고, 사용자가 특정 저널 쿼리 용어에서 검색하고자 하는 년도 또는 범위 또는 년도들을 입력한다. 년도 속성은 "년도의 범위" 속성 유형이다. 쿼리 용어(402)가 이제 "암 수용체 필터 : 저널 필터 : 년도"라는 것을 주목한다. "년도" 속성은 나열된 속성(404)로부터 사라졌다. 사용자가 GO 버튼(411)을 선택하면, 선택된 속성이 필터로 사용되어 검색이 다시 수행되고, 도 4e와 같은 화면이 디스플레이된다. 따라서, 사용자는 디스플레이된 쿼리 결과를 위해 하나 이상의 속성을 선택할 수 있고, 디스플레이된 속성(또는 라벨)에 따라 초기 검색을 필터링할 수 있다. 사용자가 속성값을 빈 칸으로 두면, 모든 속성값이 매칭된다. 예를 들어, 사용자가 저널 속성을 선택하지만 저널 이름을 입력하지 않으면, 모든 저널 속성을 가진 데이터 아이템이 쿼리 결과에 대한 가능한 후보로 선택된다. 저널 속성을 가지지 않은 데이터 아이템은 쿼리 결과에 선택되지 않는다.
- <89> 도 4d는 검색을 좁히기 위해 사용자가 하나 이상의 속성 또는 라벨을 선택하는 예를 도시한다. 설명된 실시예에서, 다중 라벨 및 속성은 속성 및 라벨(404)의 다중 라벨 및 속성을 클릭함으로써 선택된다. 다른 바람직한 실시예에서는 검색창(402)에 라벨 및 속성이 입력되도록 한다. 예를 들어, 가격 속성이 있으면, 사용자는 쿼리 용어로 다음과 같이 입력한다고 한다.
- <90> 속성(가격: 150 달러)
- <91> 이 쿼리는 현재 쿼리 결과에서 가격 속성 및 150 달러 속성값을 가진 현재 쿼리 결과에 데이터 아이템을

찾는다.

- <92> 다른 예로, 사용자가 다음과 같이 입력한다고 한다.
- <93> 속성(가격: 150 달러) 및 라벨(빵상자보다작음)
- <94> 이 쿼리는 가격 속성, 150 달러 속성값, 및 빵상자보다작음이라는 라벨을 가진 데이터 아이템을 현재 쿼리 결과에서 찾는다. 다른 바람직한 실시예는 다른 적절한 사용자 인터페이스 요소를 사용하여 사용자가 논리적으로 속성 및 라벨을 결합하도록 한다.
- <95> 도 4e는 도 4d에서 특정된 대로 특정 년도나 특정 년도들의 특정 저널에 한정되는 쿼리 결과를 도시한다. 사용자는 저널 내에서 검색(412)을 계속하기 원하는지 여부 또는 전체 데이터 아이템 컬렉션을 검색(413, 예를 들어, "구글베이스 전체 검색")을 계속하기 원하는지 여부를 결정하도록 허용된다. 이 예에서, 사용자는 영역(414)에서 다수의 라벨("생명공학", "의료", 및 "사진", 이들 각각은 쿼리 결과(406')에서 아이템(30,15 및 6)과 관련됨)의 선택을 오픈받는다. 이 예에서, 사용자는 영역(416: 날짜, 저자, 공중보전, 인용)에서 속성을 위한 값을 특정하기 위한 선택을 오픈받는다. 사용자는 관련성, 날짜 속성 또는 사용자가 정의하는 어떠한 속성(예를 들어, 가격, 위치 등)에 의해 쿼리 결과(406')를 분류(416)하는 옵션을 오픈받기도 한다.
- <96> 도 4f에서, 사용자는 도 4e의 영역(414)로부터 "날짜" 속성을 선택하고, 날짜(420)를 입력할 기회가 주어진다. 사용자가 드롭다운(Drop-down) 연산자 "기간(Between)"을 선택할 때, 사용자에게 날짜 범위(도시됨)를 선택할 기회가 주어진다. "날짜" 속성은 나열된 속성(418)로부터 사라졌다. 이 예에서, "저자" 속성은 속성(414)로부터 사라졌다. 속성이 더 이상 쿼리 및 쿼리 결과와 관련되지 않으면 속성이 사라진다. 사용자가 현재 저널에 의해 필터링하지 않았다는 사실은 사용자가 제한된 아이템 집합만을 찾아본다는 것을 가정한다. 검색 및 저자 용어가 다시 나타나도록 하려면 사용자는 Go 버튼을 선택하면 된다.
- <97> 도 4g는 사용자가 저자 이름(422)을 선택하는 것을 도시한다. 속성 및 라벨은 쿼리 결과에 기초하고 쿼리 결과는 지속적으로 변화하므로, 쿼리가 걸리지면서 새로운 속성 및 라벨이 보이게 된다. 사용자가 Go 버튼(423)을 누를 때, 다른 검색이 수행되고, 사용자에게 의해 특정된 속성 및 속성값을 반영하기 위해 쿼리 결과가 더 필터링된다.
- <98> 이하의 단락에서는 속성 및 라벨을 사용하여 검색 도중 또는 검색을 좁히는 도중에 속성 저장소(195)로의 접속을 설명한다.
- <99> 저장소(195)를 참조하는 쿼리 및 색인과정은 바람직하게 다음의 연산자를 지원한다.
- <100> 숫자 - 같음, 사이에, 초과, 미만, 숫자 범위 제안
- <101> 스트링 - 이다(Is), 가지다(Has)
- <102> 날짜 - 범위, 이전, 이후, 현재
- <103> 위치 - 이내
- <104> 저장소(195)는 적어도 다음의 방식으로 쿼리될 수 있다.
- <105> -특정 속성 이름-유형 쌍과 매칭되는 모든 아이템을 내게 보낼 것
- <106> -속성값의 값에 기초하여 이 아이템들을 분류할 것
- <107> -이하의 속성 유형에 따른 분류가 지원됨
- <108> -날짜시간
- <109> -Number - Int, Float
- <110> -스트링
- <111> -위치 - 위치를 입력한 사용자로부터의 거리
- <112> 이 쿼리 능력은 사용자가 다음의 속성 쿼리 유형을 입력하도록 한다.
- <113> -특정 이름-유형 쌍을 가진 모든 아이템을 내게 보낼 것
- <114> -주어진 이 아이템을 속성값에 의해 분류할 것(예를 들어, 이벤트_날짜를 가진 모든 아이템을 내게 보낼 것 및

오름차순으로 저장된 아이템을 내게 보낼 것)

- <115> -특정 이름-유형 속성을 위한 value1 및 value2 사이의 모든 아이템을 내게 보낼 것
- <116> 예
- <117> -요리-유형 속성을 가지고 15 분에서 30 분 사이의 값을 가지는 모든 아이템을 내게 보낼 것
- <118> -크기 속성을 가지고 단위 없이 1 및 15의 값을 가지는 모든 아이템을 내게 보낼 것
- <119> -이벤트_날짜 속성을 가지고 오늘 이전의 값을 가지는 모든 아이템을 내게 보낼 것
- <120> -발행_날짜 속성을 가지고 1925 년의 값을 가지는 모든 아이템을 내게 보낼 것
- <121> 이하의 연산자가 지원된다.
- <122> -숫자의 경우 -int, float
- <123> -미만
- <124> -초과
- <125> -사이
- <126> -날짜시간의 경우 -현재
- <127> -이전
- <128> -이후
- <129> -사이
- <130> -아이템 점수
- <131> 현재 아이템이 채점되는 2 개의 주요한 신호가 있다.
- <132> -쿼리 의존 순위 - 주로 IR 점수
- <133> -쿼리 독립 순위 - 페이지 순위 및 아이템 순위의 복합
- <134> 페이지 순위는 제공자의 웹사이트 페이지 순위이다. 페이지 순위는 아이템이 데이터 콜렉션(190)에 호스트(Host)되거나 및/또는 아이템이 다른 아이템에 링크되지 않거나 연결되지 않은 경우에는 존재하지 않는다.
- <135> 아이템 순위는 다수의 인자(Factor)에 의해 결정될 수 있다. 두가지 주요한 신호는 다음과 같다.
- <136> -제공자 특정 신호(예를 들어, 등급)
- <137> -오퍼 특정 신호(예를 들어, desc의 길이, 속성 및 라벨의 개수, 사진 등)
- <138> 아이템 순위는 다음의 신호에 의해 정의될 수 있다.
- <139> -desc 길이
- <140> -제목의 길이
- <141> -라벨의 개수
- <142> -속성의 개수
- <143> -사진
- <144> -오퍼가 스팸으로 보고된 횟수
- <145> -제공자의 등급
- <146> -오퍼의 새로움(Vecency)
- <147> 아이템은 쿼리 의존 순위*쿼리 독립 순위로 채점된다.
- <148> 순위가 기본 분류(Default Sort)이다.

- <149> 바람직한 실시예에서, 특정 파라미터가 시스템에 설정될 수 있다. 이 파라미터는 제공자 당 아이템의 최대 개수를 포함한다. 이는 특정 제공자에 의한 페이지의 혼잡을 방지한다.
- <150> 사용자가 검색을 좁히기 위해 속성 및/또는 라벨을 선택할 때, 시스템은 라벨, 제목, 설명 및 속성값을 검색한다. 속성 이름도 완성된 이름처럼 검색 가능해야만 한다. 구(Phrase)는 자주 발생하는 단어와 비교하여 훨씬 중요하다. 라벨은 제목에 비해 훨씬 중요하고, 제목은 설명보다 훨씬 더 중요하다. 속성값은 라벨과 동일한 정도로 중요하다. 각 제공자에 의한 소매상 모집(Merchant Crowding)은 개인 제공자로부터 아이템의 페이지 수가 검색 결과로 디스플레이되거나 디스플레이되지 않는지 여부를 규제하기 위해 사용자에게 의해 커지거나 커지지 않는다. 수행된 검색에 의존하여, 소매상 모집은 바람직하거나 바람직하지 않게 된다.
- <151> 바람직한 실시예에서, 시스템은 동일하거나 유사한 유형의 다른 아이템과 관련된 속성에 기초하여 새로운 아이템의 특정 유형의 구조를 정의한다(예를 들어, "직업들" 정보 유형의 대부분의 아이템이 직업 기능, 직업 유형 및 고용주의 속성을 가지고 있다면, "직업" 정보 유형의 데이터 아이템을 위한 공통 속성 구조는 직업 유형, 고용주 및 직업 기능을 갖지 않음). 검색자 및 다른 프로그램은 "고용주가 ABC 사이고 직업 유형이 제품 관리인 모든 직업을 내게로 보낼 것"과 같은 쿼리로 데이터 집합을 쿼리할 수 있다.
- <152> 본원발명에 설명된 예는 사람인 사용자를 나타내지만, 본 발명의 다른 실시예는 인공 지능 소프트웨어 프로그램과 같은 사람이 아닌 사용자, 또는 사람일 수도 있고 사람이 아닐 수도 있는 웹을 통해 통신하는 존재(Entity)와 동작하기 위해 설계됨이 이해된다. 사람이 아닌 사용자가 소프트웨어 프로그램이면, 상기한 것처럼 결과 및 속성을 디스플레이할 필요가 없다. 대신에, 그러한 구현은 쿼리 결과를 좁히는데 사용될 수 있는 잠재적 속성과 통신할 뿐이다. 그러한 실시예에서, 사람이 아닌 인공 지능은 무엇으로부터 선택을 해야하는지 다수의 속성에 의해 압도되지 않기 때문에, 더 많은 속성의 옵션이 디스플레이될 수 있다. 그러한 실시예에서, 히스토그램을 결정하는 과정과 같은 방법의 구성 요소는 필요하지 않거나, 속성 선택의 순위를 매기고 가능한 속성 선택의 개수를 제한하지 않기 위해서만 사용된다.
- <153> 주기적으로, 구조화된 데이터(190)에서 다양한 정보 유형을 위한 코어 속성이 업데이트될 필요가 있음이 이해된다. 데이터가 구조화된 데이터 컬렉션에 추가되면서, 처음에는 인기가 없었던 특정 속성이 인기가 있게 된다. 예를 들어, 정수(Integer)의 속성 유형을 가지며 출연진(Cast) 사진의 출처인 텔레비전 쇼의 어떤 시즌을 특정하는 "시즌(Season)" 속성은 "TV 쇼" 정보 유형으로 초기 코어 속성에 의해 처음에는 고려되지 않지만, 더 많은 출연진 사진이 데이터 컬렉션에 추가되면서 시즌 속성은 인기가 있게 될 수 있다. 일부 실시예에서, 코어 속성은 스팸 필터를 통과한 후에 인기 및 시즌에 기초하여 자동 업데이트 되기도 한다.
- <154> 도 3e는 제공자가 제공한 새로운 속성이 정보 유형을 위해 코어 속성에 추가되어야만 하는지를 결정하기 위해 주기적으로 수행되는 방법(350)을 도시한다. 아이템 정보 유형을 위한 속성의 코어 그룹은 제공자가 새로운 정보 유형의 아이템을 추가할 때마다 자동으로 오퍼되는 속성이다. 바람직한 실시예에서, 제공자가 자신의 방식을 디스플레이된 속성에 강제하기 위해 속성을 스팸(Spam)할 가능성을 감소시키도록 코어 속성만이 오퍼된다. 정보 유형 각각을 위해, 본 방법은 그 정보 유형을 위한 가장 인기 있는 사용자가 추가한 속성만을 살피고(352), 그 정보 유형에 대해 가장 인기 있는 속성을 코어 속성으로 승급시킨다(Promote).
- <155> 어떤 속성이 코어 속성으로 승급되는지 결정하기 위해 사용되는 "가장 인기있음"은 서로 다른 실시예에서 다르게 정의된다. 예를 들어, 가장 인기있음은 예를 들어, 일주일 또는 한 달과 같은 미리 결정된 시간 주기에 동안 사용자에게 의해 가장 자주 선택(352)된, 코어 속성에 없는 속성일 수 있다. 다른 예로서, 가장 인기있음은 미리 결정된 시간 주기 동안 쿼리 결과에서 가장 자주 나타나는 데이터 아이템을 가진, 코어 속성에 없는 속성일 수 있다. 또 다른 예로서, 가장 인기있음은 미리 결정된 시간 주기 동안 대규모의 제공자 데이터에 나타나는, 코어 속성에 없는 속성일 수 있다. 가장 인기있음은 코어 속성에 추가되며 검색을 좁히는데 유용한 속성을 만들어 낸다면 어떠한 적절한 방식으로도 결정될 수 있다.
- <156> 예를 들어, 제공자는 블로그에서 언급되는 기사(Article)를 지시하기 위해 기사의 아이템 정보 유형을 위한 "블로그됨(Blogged)" 속성을 추가하기 시작한다. 그러한 속성은 URL 속성 유형을 가지고, 블로그의 URL은 아이템이 어디에서 언급되는지 지시해준다. 유일 제공자 또는 사용자의 임계치 숫자가 정보 유형을 위해 특정한 새로운 속성을 사용하면(354), 속성은 그 정보 유형을 위한 속성의 코어 그룹에 추가된다(356). 바람직한 실시예에서, 임계치 값은 시스템을 사용하는 제공자의 전체 숫자에 기초한다. 이는 2~3 개의 작은 값에서 시작하여 큰 숫자로 증가된다. 인기 있는 라벨을 코어 라벨 집합에 추가하기 위해 라벨의 경우에도 유사한 방법이 수행된다. 바람직한 실시예에서, 승급된 속성은 사람에 의해 또는 적절한 소프트웨어 또는 하드웨어로 구현된 방법에 의해

완전함(Sanity)이 확인된다.

- <157> 상기 단락에서는 구조화된 데이터 콜렉션(190)에 입력되는 데이터를 검색하고 업데이트하기 위한 방식을 일반적으로 설명하였다. 이하의 단락에서는 제공자가 구조화된 데이터 콜렉션(190)에 데이터를 입력하거나 추가할 수 있는 방식에 대해서 설명한다. 바람직한 실시예에서, 제공자는 자신의 데이터를 위한 새로운 속성을 특정한다.
- <158> 도 6a-6e는 제공자가 데이터 콜렉션의 아이템을 어떻게 편집할 수 있는지를 도시하는 예시적 스크린 샷이다. 제공자는 데이터 콜렉션(190)에 콘텐츠를 추가하거나 추가할 수 있는 자이다. 설명된 실시예에서, 데이터 콜렉션(190)은 개인, 비영리 집단, 또는 회사와 같은 하나 이상의 제공자에 의해 소유된 데이터이다. 본 실시예는 그러한 사용자가 웹을 통해 자기들의 고유한 구조화된 데이터 콜렉션(예를 들어, 데이터베이스)을 셋업하고 채우도록 하고, 웹 또는 유사한 네트워크를 통해 검색 가능한 그러한 콜렉션을 만들어내도록 한다. 제공자는 다른 사람에 의해 데이터가 검색되도록 허용하는 수수료 또는 대가로서 중앙 저장소에 데이터를 저장하려고 하는 것이 고려된다. 그러한 경우에, 데이터 콜렉션은 본원발명에 설명된 일부 또는 모든 기능을 포함한 버전의 구글 브라우저 또는 구글 데스크탑 검색 엔진과 같은 웹 또는 네트워크 기반 브라우저를 통해 검색될 수 있다.
- <159> 도 6a-6e는 제공자가 데이터를 편집 및 시스템에 입력하도록 하는 사용자 인터페이스를 도시하는 예시적 스크린 샷이다.
- <160> 도 6a는 제공자가 데이터 콜렉션(190)의 데이터 아이템을 보고 편집할 수 있도록 하는 사용자 인터페이스(600)를 도시한다. 사용자 인터페이스는 데이터 콜렉션(190)에 아이템을 추가하기 위해서도 사용될 수 있다. 영역(602)은 데이터 콜렉션(190)에 있는 아이템의 일부 목록을 포함한다. 이 예에서, 이 목록은 아이템 제목(601), 아이템 유형(605, 정보 유형이라고도 함), 상태(603), 만료 기한, 임프레션(Impression) 횟수(아이템이 디스플레이된 횟수), 대상을 클릭한 횟수, 및 사용자 클릭 비율, 검색 결과에서 아이템이 클릭된 횟수를 포함한다. 예에서, 데이터 콜렉션의 모든 아이템의 부분 집합이 영역(602)에 도시되지만, 제공자는 자신의 개인적 데이터 콜렉션(620)을 검색하거나 전체 데이터 콜렉션(622)을 검색할 수도 있다. 제공자는 비활성 아이템(616) 또는 벌크 업로드 파일(618)을 볼 수도 있다. 데이터 아이템 각각은 관련 "편집" 링크를 가진다. 바람직한 실시예에서, 제공자는 자신의 고유한 데이터 아이템만을 편집할 수 있다. 영역(604)은 제공자가 현존하는 정보 유형(이벤트 및 활동, 주거 등)을 보여주는 드롭다운 메뉴와 같은 선택 장비를 디스플레이하도록 한다. 제공자가 정보 유형을 선택하면, 제공자는 자신의 데이터를 위해 영역(606)에 정보 유형의 설명을 추가할 수 있다.
- <161> 도 6b는 제공자가 데이터 콜렉션(190)의 데이터 아이템을 보고 편집(610)할 수 있도록 하는 사용자 인터페이스를 도시한다. 아이템은 "뉴스 및 기사" 정보 유형을 가진다. 제공자가 도 6a의 영역(602)의 데이터 아이템을 선택하면, 그 아이템의 정보가 영역의 필드(611)에 디스플레이된다. 하지만, 이 예에서, 제공자가 아이템을 선택하지 않았다면, 제공자가 새로운 데이터 아이템을 입력하는 것은 자유이다. 이 예에서, "뉴스 및 기사" 정보 유형(610)은 쿼리 결과에 디스플레이되기 위한 제목, 사진, 설명 및 링크(예를 들어, URL)와 같은 필드를 포함한다.
- <162> 도 6b의 사용자 인터페이스는 제공자가 아이템의 속성 및 라벨도 편집할 수 있도록 한다. 주목할 것은, 정보 유형 각각이 속성과 관련되지만, 특정 유형의 데이터 아이템 모두가 그 정보 유형을 위해 모든 가능한 속성값을 가지는 것은 아니라는 점이다. 이 예에서, 참조 번호 612에 의해 도시된 것처럼, 제공자는 아이템 수량 "1 개"가 사용 가능하거나 존재함을 지시한다. 이 아이템에 대한 저자 또는 뉴스 소스 속성의 값이 특정되지 않는다. 각각의 속성은 "텍스트" 속성 유형을 가진다. 제공자가 개별적인 데이터 아이템의 속성값을 추가하는 것은 자유이다. 제공자는 영역(613)을 사용하여 속성을 추가할 수도 있다. 이 경우, 제공자는 속성 이름 및 속성값을 추가할 수 있다.
- <163> 제공자는 영역(618)의 연락처(Contact) 정보와 관련된 속성값을 제공할 수 있다. 제공자는 영역(619)의 위치 정보와 관련된 속성값을 제공할 수 있다.
- <164> 제공자는 라벨을 영역(619)의 아이템에 추가할 수 있다. 특정 실시예에서, 정보 유형은 기본 속성 이름이다. 이 경우, 정보 유형은 뉴스 및 기사이고, 이것이 라벨이기도 하다.
- <165> 도 6c는 제공자가 데이터 콜렉션(190)의 데이터 아이템을 보고 편집(610)할 수 있도록 하는 도 6b의 사용자 인터페이스를 도시한다. 이 예에서, 제공자는 제공자가 정의한 새로운 속성(613)을 위한 이름 및 값을 추가할 수 있다. 기본 속성 유형은 "텍스트"이지만, 제공자는 숫자 단위, 숫자, 날짜 범위, 큰 텍스트, URL, Boolean, 및 위치와 같은 다른 속성 유형을 선택할 수 있다.
- <166> 도 6d는 제공자가 데이터 콜렉션(190)의 데이터 아이템을 보고 편집(610)할 수 있도록 하는 사용자 인터페이스

를 도시한다. 아이템은 "제품" 정보 유형(630)을 가진다. 제공자가 도 6a의 영역(602)에 있는 데이터 아이টে 선택하면, 그 아이টে의 정보가 영역(611)의 필드에 디스플레이된다. 하지만, 이 예에서, 제공자가 아이টে 선택하지 않았으면, 제공자가 사용자 인터페이스(630)를 사용하여 새로운 아이টে 입력하는 것은 자유이다. 이 예에서, "제품" 정보 유형은 쿼리 결과에 디스플레이되기 위한 제목, 사진, 설명 및 링크(634: 예를 들어, URL)와 같은 필드를 포함한다.

- <167> 도 6d의 사용자 인터페이스는 제공자가 아이টে의 속성 및 라벨도 편집할 수 있도록 한다. 정보 유형 각각이 속성과 관련되지만, 특정 유형의 데이터 아이টে 모두가 그 정보 유형을 위해 모든 가능한 속성값을 가지는 것은 아니라는 점이다. 이 예에서, 참조 번호(632)에 의해 도시된 것처럼, 제공자는 아이টে 당(예를 들어, 파운드 당 또는 다스 당과 비교됨) 가격이 150 달러임을 지시한다. 수량이 "1 개"임이 특정된다. 가격 유형은 제공자가 설정한 가격 유형이다(예를 들어, 최적 오퍼, 흥정 가능, 고정됨 등). 이 아이টে에 대한 가격 옵션, 브랜드, 조건, 및 제품 유형의 어떠한 값도 특정되지 않는다. 그러한 속성 각각은 "텍스트" 속성 유형을 가진다. 이 예에서, 제공자는 자신이 특정한 속성을 위한 속성 유형을 변경할 수 있다. 제공자가 개인적인 데이터 아이টে의 속성값을 추가하는 것은 자유이다. 제공자는 영역(613)을 사용하여 속성을 추가할 수도 있다. 이 경우, 제공자는 속성 이름 및 속성값을 추가할 수 있다.
- <168> 이 실시예에서, 제공자가 추가한 속성은 현재 정보 유형의 모든 아이টে 추가된다. 예를 들어, 이 경우 일단 정의되면, "제품" 유형을 가진 모든 제공자의 아이টে 새로 추가된 속성(613)이 주어진다. 각 아이টে를 위한 값은 보통 개별적으로 추가된다. 특정 실시예는 제공자가 특정한 정보 유형의 모든 아이টে를 위한 값도 특정할 수 있도록 한다. 상기한 것처럼, 새로운 속성이 코어 속성 집합으로 진급하는 것도 가능하다. 다른 실시예에서, 한 명의 제공자가 속성을 추가할 때 그룹의 다른 제공자도 동일한 속성을 가지게 함으로써 정의된 제공자 그룹이 모두 동일한 속성을 가지도록 제공자가 동의할 수 있다.
- <169> 제공자는 영역(618)의 연락처 정보와 관련된 속성값을 제공할 수 있다. 제공자는 영역(619)의 위치 정보와 관련된 속성값을 제공할 수 있다. 제공자는 영역(638)의 지불 방법과 관련된 속성값을 제공할 수 있다.
- <170> 제공자는 영역(616)의 아이টে 라벨을 추가할 수 있다. 특정 실시예에서, 정보 유형은 기본 속성 이름이다. 이 경우, 정보 유형은 제품이고, 이는 라벨이기도 하다. 이 실시예에서, 제공자가 추가하는 라벨은 현재 유형의 아이টে 모두에 추가되지는 않는다(라벨이 정보 유형인 경우는 제외함). 상기한 것처럼, 새로운 라벨이 코어 라벨 집합에 진급하는 것도 가능하다. 다른 실시예에서, 새로운 라벨은 정보 유형의 모든 아이টে 항상 추가된다.
- <171> 도 6e는 제공자가 데이터 콜렉션(190)의 데이터 아이টে를 보고 편집(630)할 수 있도록 하는 도 6d의 사용자 인터페이스를 도시한다. 이 예에서, 연락처, 지불, 및 위치는 제품 정보 유형의 모든 속성이다. 이 속성은 복합 유형(정수 또는 단순한 스트링이 아님)을 가지는 속성이다. 이 예에서, 제공자는 "제품" 정보 유형의 아이টে를 위해 연락처(618)와 관련된 값을 추가할 수 있다. 이 경우, 제공자는 별명, 전화 번호, 이메일 주소(제공자 정보의 데이터베이스로부터 가져온 잠재적 값, 도시되지 않음)의 일부 또는 전부를 특정한다. 이 예에서, 제공자는 "제품" 정보 유형의 아이টে를 위해 지불(638)과 관련된 값을 추가할 수 있다. 이 경우, 제공자는 지불 방법 및 노트의 일부 또는 전부를 특정한다. 이 예에서, 제공자는 "제품" 정보 유형의 아이টে를 위해 위치(619)와 관련된 값을 추가할 수 있다. 이 경우, 제공자는 텍스트 노트(예를 들어, "캘리포니아, 프리몬트")의 일부 또는 전부를 특정한다. 이 실시예에는, 고객이 이 위치 및 배달 반경으로부터 픽업(Pickup)할 수 있는지 여부를 지시하는 체크박스도 있다.
- <172> 이 예에서, 연락처, 지불, 및 위치 값은 각 아이টে마다 별도로 입력된다. 제공자가 추가하는 값은 현재 정보 유형의 아이টে 모두에 추가되지는 않는다. 예를 들어, 이 경우, "제품" 정보 유형의 모든 제공자 아이টে도 6e에 도시된 연락처, 지불, 및 위치 값에 주어지는 것은 아니다. 각 아이টে를 위한 값은 보통 개별적으로 추가된다. 특정 실시예는 제공자가 특정 정보 유형의 모든 아이টে를 위한 값도 특정할 수 있도록 한다. 예를 들어, 지불 정보는 제공자의 모든 "제품"에 대해서 동일하다.
- <173> 후원자(Promoter)는 도 6의 사용자 인터페이스를 통해, 또는 도 7 및 도 8에 도시된 벌크 업로드 방법을 통해 아이টে를 입력할 수 있다.
- <174> 도 7은 벌크 업로드 파일을 등록하기 위한 사용자 인터페이스를 도시하는 예시적 스크린 샷(700)이다. 벌크 업로드 파일은 데이터 콜렉션(190)을 만들거나 그에 추가하기 위해 사용된다. 이 예에서, 모두 동일한 정보 유형인 아이টে의 플랫폼 파일이 추가된다. 이 예에서, 파일 이름(712)은 "지역 조사(Local Inventory)"이다. 제공자는 미리 결정된 정보 유형 또는 개인화된 정보인 데이터 유형(714)을 선택한다. 제공자는 데이터의 텍스트 스트링

을 위한 언어(716)를 선택한다. 제공자가 "벌크 업로드 파일 등록" 버튼(718)을 선택할 때, 파일 이름(712)을 가진 파일이 등록되고 제공자는 파일을 업로드 할 수 있게 된다. 제공자는 웹 기반 업로드 인터페이스를 사용하거나 FTP(File Transfer Protocol) 또는 RSS와 같은 다른 메커니즘을 사용하여 파일을 업로드할 수 있다.

- <175> 도 8a는 벌크 업로드되는 탭제한(Tab-delimited) 파일을 위한 포맷(801)을 도시한다. 다음의 내용은 벌크 업로드 파일을 위한 포맷 요구사항이다.
- <176> -탭제한 일반 텍스트
- <177> -파일의 첫번째 줄은 헤더일 것 - 반드시 속성 이름을 포함하고(이하에 기술함), 탭에 의해 나누어질 것
- <178> -한 줄에 하나의 아이템; 각 속성은 탭에 의해 나누어질 것
- <179> -라인의 마지막에 트레일링(Trailing) 탭을 가지면 안됨
- <180> -파일은 반드시 LATIN1 또는 UTF-8 인코딩으로 저장될 것. LATIN1의 부분집합인 ASCII도 가능함
- <181> -링크와 이미지 URL은 충분히 검증된 것이어야만 함. 즉, 예를 들어, <http://www.example.com/image.gif>와 같이 <http://> 부분을 반드시 포함할 것
- <182> -탭, 복귀(Carriage Return), 또는 줄 내림 문자(New Line Character) - 이들 중 하나라도 속성에 나타나면, 그 아이템은 디스플레이할 수 없음
- <183> -HTML 태그, 주석, 및 확장 문자열(Escape Sequence) - html은 벌크 업로드로부터 제거되지 않지만, 최적의 모습을 위해 HTML을 포함하지도 않음
- <184> 바람직한 실시예에서, 데이터 아이템은 속성도 포함하는 업로드된 파일의 일부이다. 다른 바람직한 실시예에서, 데이터 아이템 및 속성은 어떤 속성값이 어떤 데이터 아이템에 속하는지 명확하게 하기 위해 구성되는 별개의 파일로 업로드된다.
- <185> 도 8b는 벌크 업로드 파일을 만들기 위해 제공자에 의해 사용되는 예시적 방법의 순서도(800)이다. 제공자는 사람이거나, 하드웨어 또는 소프트웨어일 수 있다.
- <186> 요소 802: 스프레드시트(Spreadsheet) 프로그램에서 새 파일을 오픈함
- <187> 설명된 방법은 벌크 업로드 파일을 만들기 위해 마이크로소프트의 엑셀과 같은 스프레드시트 프로그램을 사용한다. 마이크로소프트의 엑셀과 같은 스프레드시트 프로그램을 사용하면 벌크 업로드 파일을 쉽게 만들고 이를 적절한 포맷으로 전환할 수 있다. 적절하게 포맷된 파일을 만들기 위해 다른 방법도 사용될 수 있다.
- <188> 요소 804: 헤더 가로줄(Row)을 만들
- <189> 예로서, 제품 벌크 업로드를 위한 헤더 가로줄은 도 8c의 가로줄(832)처럼 생겼다. 제공자가 제출하고자 하는 아이템의 정보 유형(도 7의 714 참조)에 따라 벌크 업로드의 세로줄 각각을 특정한다. 스프레드시트의 첫 번째 가로줄(832)에서, 제공자가 자신의 아이템을 설명하기 위해 포함시키고자 하는 각 속성의 이름을 입력한다. 이것이 헤더 가로줄이다. 헤더 가로줄의 콘텐츠는 제출된 정보의 정보 유형, 및 제공자가 정의된 정보 유형을 전송했는지 또는 자신이 직접 창조한 정보 유형을 전송했는지 여부에 의존한다.
- <190> 개인화 정보 유형:
- <191> 벌크 업로드는 어떤 유형의 정보도 제출하도록 사용될 수 있다. 제공자가 자신의 고유한 정보 유형을 전송하면, 제공자는 미리 정의된 속성의 어떠한 조합도 사용할 수 있다. 바람직한 실시예에서, 제공자는 미리 정의된 속성을 사용하도록 강력하게 권고된다. 제공자는 개인화 속성을 개수 제한 없이 사용할 수도 있다. 제공자는 자신의 아이템을 가장 잘 설명하는 속성 집합을 골라야 한다.
- <192> 정의된 정보 유형:
- <193> 제공자는 정의된 정보 유형 중 하나를 위해 벌크 업로드를 전송할 수 있다. 제공자는 자신의 벌크 업로드에 정의된 정보 유형을 포함시키도록 강력하게 권고된다. 정의된 정보 유형은 쿼리를 검색하기 위해 더 정확하게 아이템이 매칭되도록 한다. 제공자가 더 많은 정보를 줄수록, 사용자가 아이템을 찾기가 쉬워진다. 바람직한 실시예에서, 제공자는 제공자의 아이템이 완료된 검색의 중요한 부분에 나타날 수 있도록 하는 권고된 속성을 포함시켜야만 한다.

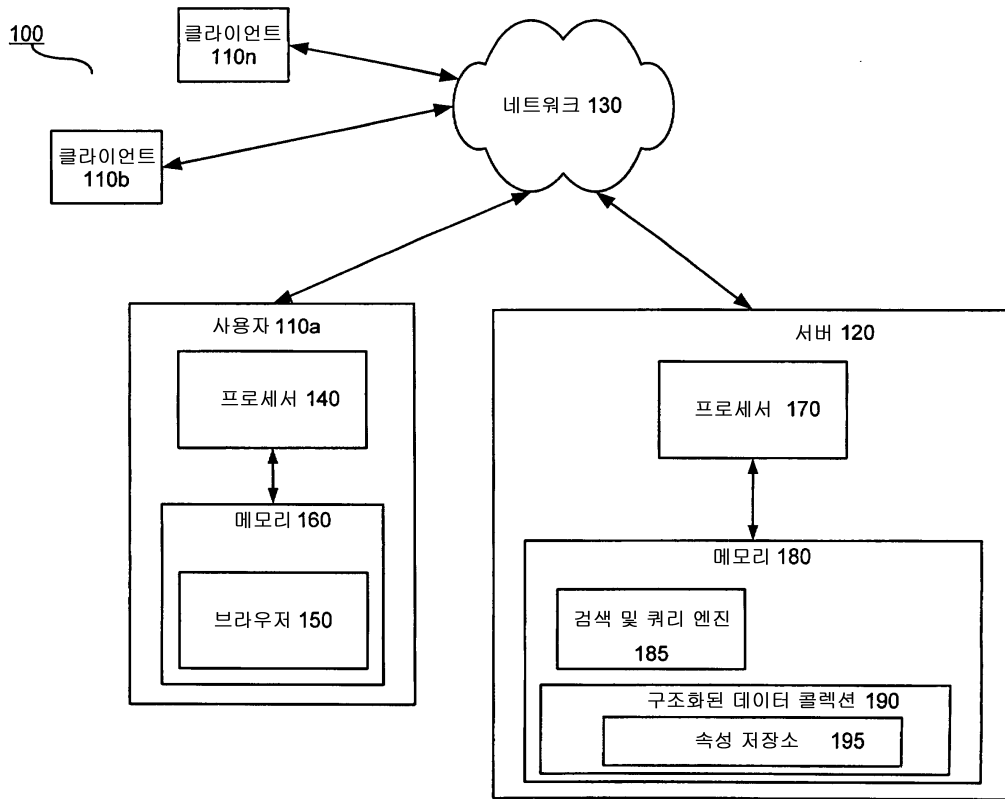
- <194> 요소 806: 아이템 정보를 입력함
 - <195> 각 가로줄(834)에서, 제공자는 자신의 데이터 컬렉션의 아이템을 위한 정보를 입력한다. 각 정보는 그 정보가 속한 세로줄의 헤더를 반영해야만 한다(예를 들어, 제품의 가격은 "가격" 헤더 아래로 가야만 함). 각 가로줄은 가로줄 당 하나의 아이템만 포함한다. 도 8c를 참조한다.
 - <196> 요소 808: 벌크 업로드를 탭제한이 있는 보통(Plain) 텍스트로 전환함
 - <197> 이전에 등록된 파일 이름을 사용하여 스프레드시트를 탭제한이 있는 텍스트(.txt) 파일로 전환한다(도 7 참조). 제공자가 모든 아이템을 스프레드시트에 입력한 후에, 제공자는 탭제한이 있는 텍스트(.txt) 포맷으로 스프레드시트를 저장한다. 등록된 파일 이름은 차기 업로드를 위해 재사용될 수 있다. 업로드된 파일이 등록되지 않은 이름을 가지면, 파일의 아이템은 데이터 컬렉션(190)에 추가되지 않는다. 바람직한 실시예에서, 데이터 컬렉션(190)에 아이템이 유지되는 것을 확실히 하기 위해서 30 일마다 적어도 한 번씩 업데이트된 벌크 업로드가 전송되어야만 한다.
 - <198> 요소 810: 파일을 업로드함
 - <199> 도 8d는 파일을 업로드하기 위한 사용자 인터페이스(840)을 도시한다.
 - <200> 요소 812: 벌크 업로드의 에러(Error)를 확인함
 - <201> 제공자가 벌크 업로드를 전송한 후에, 제공자는 중앙 웹 사이트에 로그인하여 벌크 업로드 상태를 볼 수 있다. 결과가 '성공'으로 나열되면, 벌크 업로드는 변경될 필요가 없다. 실패이면, 제공자는 에러(들)를 어떻게 수정해야 하는지에 관한 정보를 보기 위해 벌크 업로드의 파일 이름을 클릭할 수 있다.
 - <202> 벌크 업로드가 업로드된 후에, 파일은 아이템, 속성, 및 라벨을 데이터 컬렉션(190) 및 도 5의 데이터 구조에 추가하기 위해 처리된다. 업로드가 승인되면, 이후에 동일한 파일 이름의 어떠한 업데이트도 자동으로 처리된다.
- 산업상 이용 가능성**
- <203> 본 발명은 상기한 것처럼 다수의 실시예와 함께 설명되었지만, 다양한 수정이 본 발명의 범위 내에서 만들어질 수 있다. 예를 들어, 특정한 바람직한 실시예는 유효하지 않거나 "스팸" 속성 및 라벨을 탐지하기 위한 방법 및 시스템을 포함한다. 제공자가 데이터 아이템을 검색의 가장 위에 오도록 하는 자신의 데이터에 속성을 추가하는 것은 바람직하지 못하다. 그러한 속성을 피하기 위해 사용된 일부 방법은 블랙리스트링(Blacklisting), 특정 히스토크램 분포 등을 포함한다.
 - <204> 다른 바람직한 실시예에서 디스플레이된 상위 속성 및 라벨은 단순히 속성 키-유형 튜플(Tuple) 및 라벨의 인기 에 기초하여 결정되는 것이 아니고 값의 분포에 기초하여 결정된다(분포가 이산적(Discrete)일수록 더 좋고, 스큐(Skew)가 클수록 더 좋음. 예를 들어, 속성을 위한 5 개의 인기있는 값이 균등하게 분포된 50 개의 값보다 좋음). 예를 들어, 색깔이 속성이고, 상위 색깔이 빨간색, 파란색 및 녹색이면, 이들은 걸러내기를 하기에 좋은 속성이다. 한편, 각각의 색깔이 3 번 발생하는 100 개의 값을 가지는 것은 도움이 되지 않는다.
 - <205> 다른 바람직한 실시예는 속성, 아이템 순위/각 오퍼의 오퍼 순위를 사용하는 제공자의 수에 기초하여 복잡한 신뢰도 점수를 계산한다.
 - <206> 다른 바람직한 실시예는 어떤 속성을 사용자에게 디스플레이할 것인지 결정하기 위해 사용자로부터의 클릭 신호를 사용한다. 속성 및 라벨은 인기 순위로 정의된 무엇인가에 의해 채점된다.
 - <207> $PR = \text{쿼리 결과에서의 인기} * \text{그 특정 쿼리에 대한 CTR}$
 - <208> 다른 바람직한 실시예에서, 사용자의 ALWAYS2 속성이 특정 쿼리를 제한하면 쿼리 결과에 미리 적용된 제한을 보여준다(예를 들어, 아이팟은 90%의 경우에 가격 및 위치에 대해 항상 제한되고, 사용자가 아이팟을 입력할 때 시스템은 미리 가격 및 위치를 제한함).
 - <209> 따라서, 본 발명의 개시는 예시적인 것을 의도한 것이고, 본 발명의 핵심을 한정하지 않으며, 본 발명의 핵심은 이하의 특허 청구범위에 기술된다.

도면의 간단한 설명

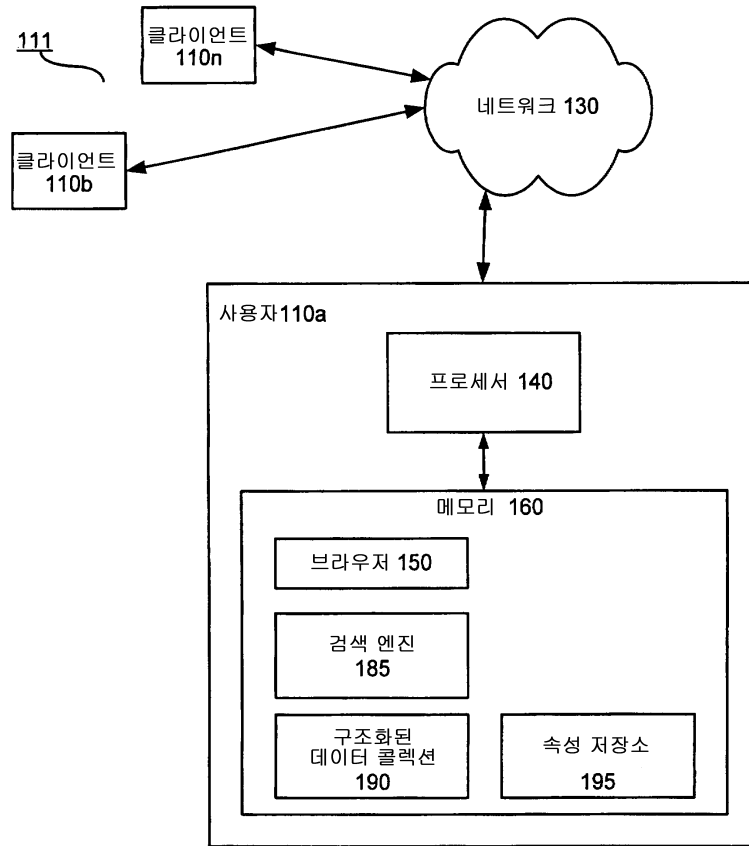
- <10> 도 1a는 본 발명의 바람직한 실시예에 따르는 데이터 처리 시스템을 도시하는 블록 다이어그램(Block Diagram)이다.
- <11> 도 1b는 본 발명의 바람직한 실시예에 따르는 다른 데이터 처리 시스템을 도시하는 블록 다이어그램이다.
- <12> 도 1c는 본 발명의 바람직한 실시예에 따르는 아키텍처(Architecture) 다이어그램이다.
- <13> 도 2a는 본 발명의 바람직한 실시예에 따라 검색 가능한 데이터 아이템 콜렉션을 만드는 개관을 도시하는 순서도이다.
- <14> 도 2b는 본 발명의 바람직한 실시예에 따라 문서 콜렉션을 검색하고 검색을 걸러내는 개관을 도시하는 순서도이다.
- <15> 도 3a는 데이터 아이템 콜렉션으로부터 라벨 및 속성을 추출하는 방법을 도시하는 순서도이다.
- <16> 도 3b는 쿼리 용어를 수신하고 쿼리 결과를 디스플레이하는 방법을 도시하는 순서도이다.
- <17> 도 3c는 주어진 쿼리 결과에 대해 어떤 속성을 디스플레이할 것인지 결정하는 방법을 도시하는 순서도이다.
- <18> 도 3d는 사용자가 디스플레이된 쿼리 결과를 라벨 및/또는 속성값을 사용하여 걸러내도록 하는 방법을 도시하는 순서도이다.
- <19> 도 3e는 제공자가 제공한 새로운 속성이 정보 유형을 위해 코어 속성에 추가되어야만 하는지를 결정하기 위해 주기적으로 수행되는 방법을 도시한다.
- <20> 도 4a는 검색 엔진 및 사용자에 의해 입력된 쿼리 용어의 예시적 스크린 샷(Screen Shot)이다.
- <21> 도 4b는 도 4a의 쿼리로부터 쿼리 결과를 도시하고, 쿼리 용어에 대한 쿼리 결과와 관련된 라벨 및 속성도 도시하는 예시적 스크린 샷이다.
- <22> 도 4c-4g는 추가적인 속성과 라벨, 및 어떻게 사용자가 속성 및/또는 라벨을 사용하여 검색을 좁히는지를 도시하는 예시적 스크린 샷이다.
- <23> 도 5a는 검색 가능한 데이터 콜렉션을 위한 속성 및 라벨을 저장하기 위해 사용되는 데이터 포맷을 도시한다.
- <24> 도 5b는 도 5a의 포맷을 사용하여 저장된 속성의 예를 도시한다.
- <25> 도 5c는 도 5a의 포맷을 사용하여 저장된 라벨의 예를 도시한다.
- <26> 도 5d는 정보 유형을 그 속성에 대응시키기 위한 예시적 데이터 구조를 도시한다.
- <27> 도 5e는 정보 유형을 위한 일부 예시적 속성에 매핑(Mapping)되는 정보 유형의 예를 도시한다.
- <28> 도 6a-6e는 제공자가 데이터를 편집하고 시스템에 입력하도록 하는 사용자 인터페이스를 도시하는 예시적 스크린 샷이다.
- <29> 도 7은 벌크 업로드 파일을 등록하기 위한 사용자 인터페이스를 도시하는 예시적 스크린 샷이다.
- <30> 도 8a-8d는 제공자가 데이터 및 속성값을 벌크 업로드하는 방법을 도시한다.

도면

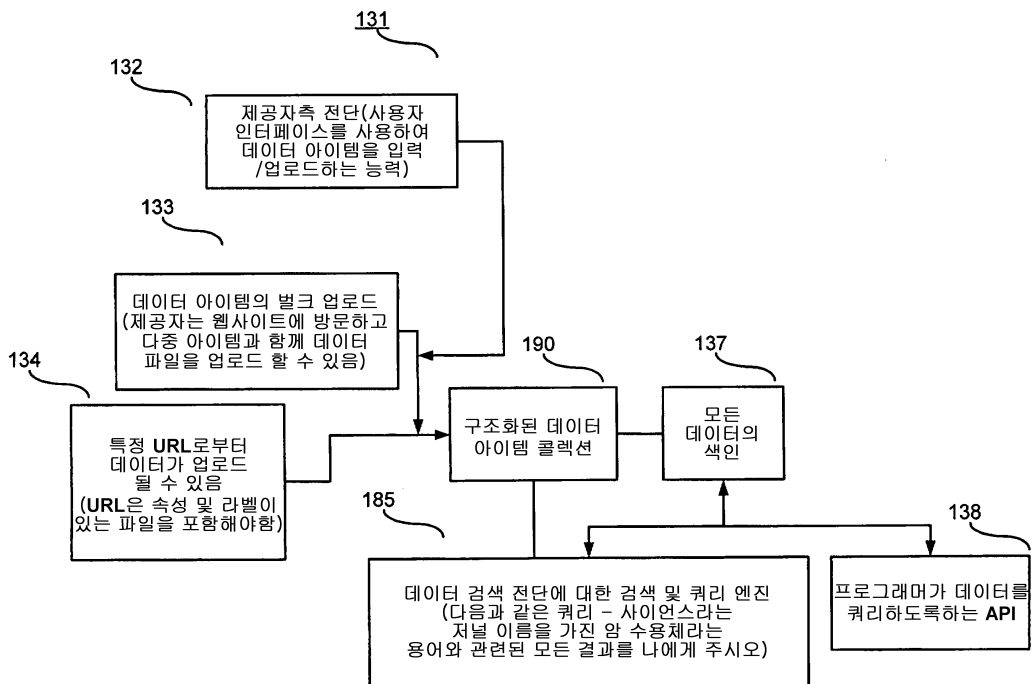
도면1a



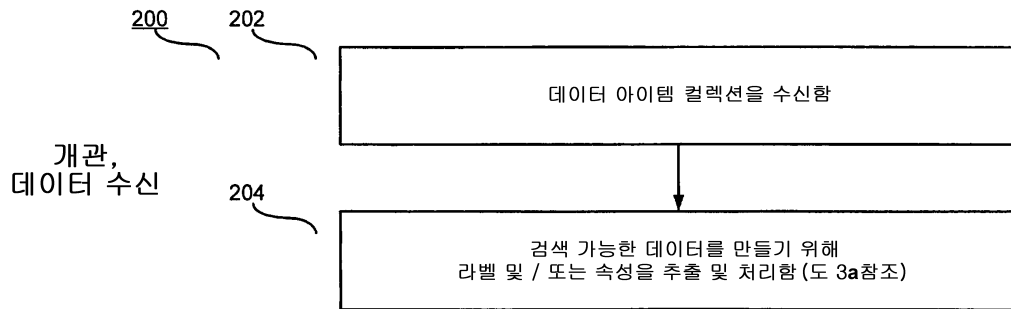
도면1b



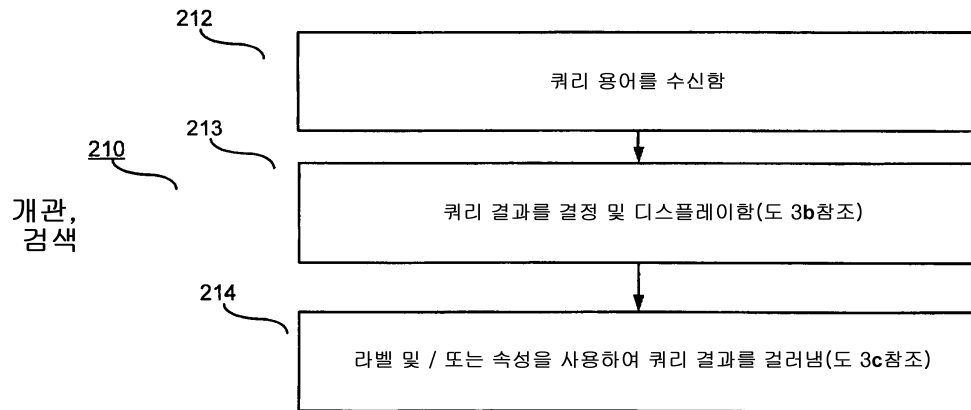
도면1c



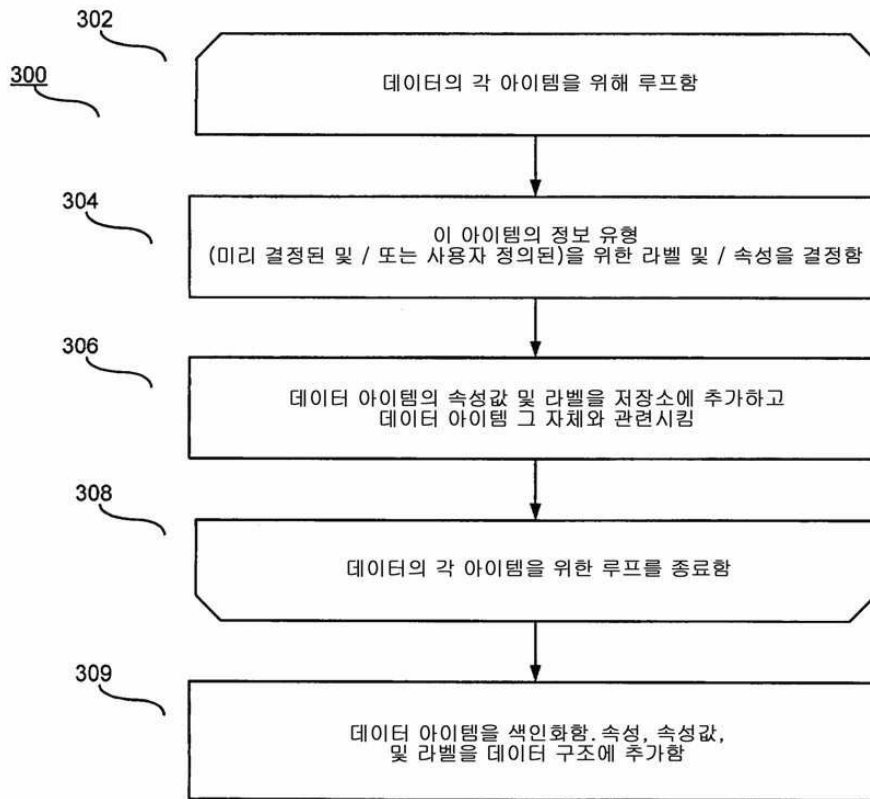
도면2a



도면2b

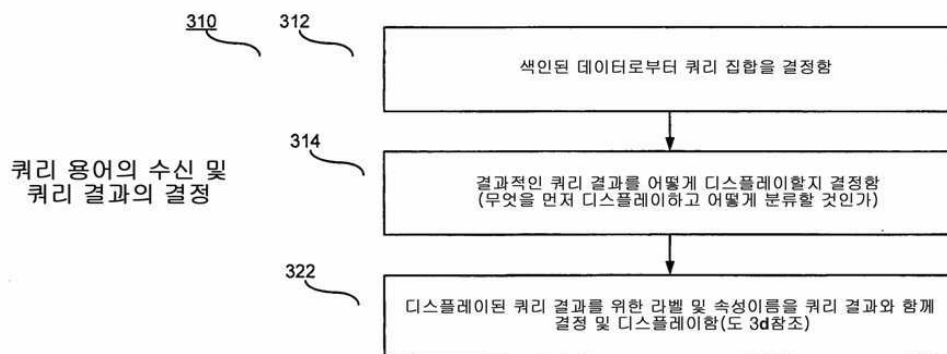


도면3a



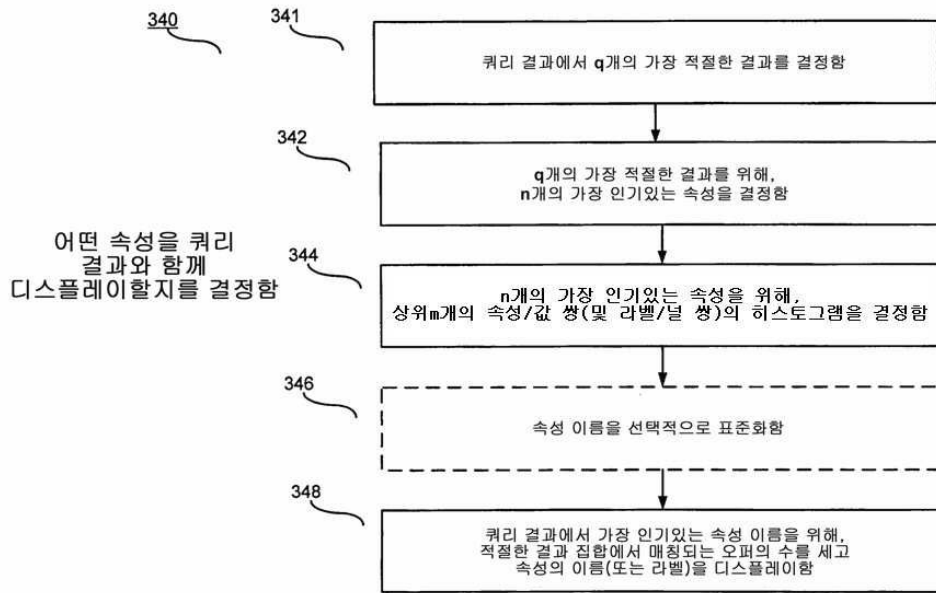
라벨 및 / 또는 속성의 추출 및 처리(셋업)

도면3b

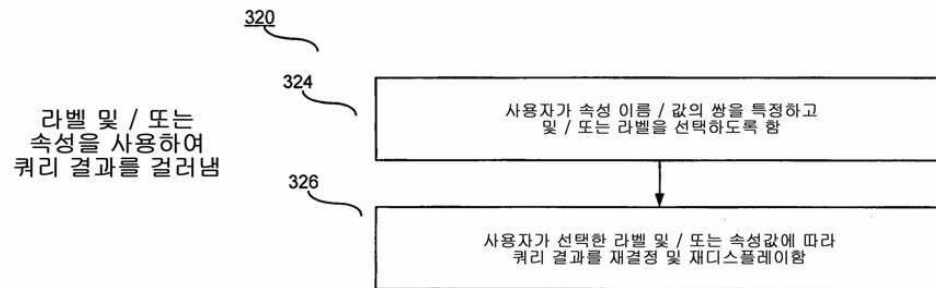


쿼리 용어의 수신 및
쿼리 결과의 결정

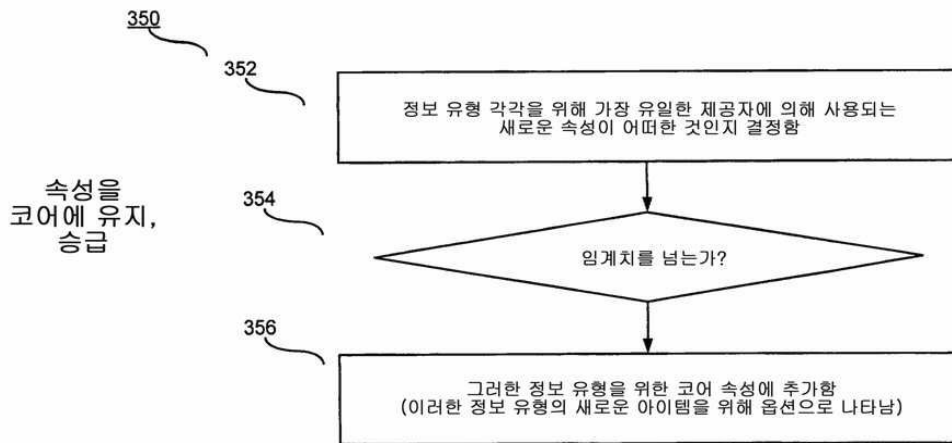
도면3c



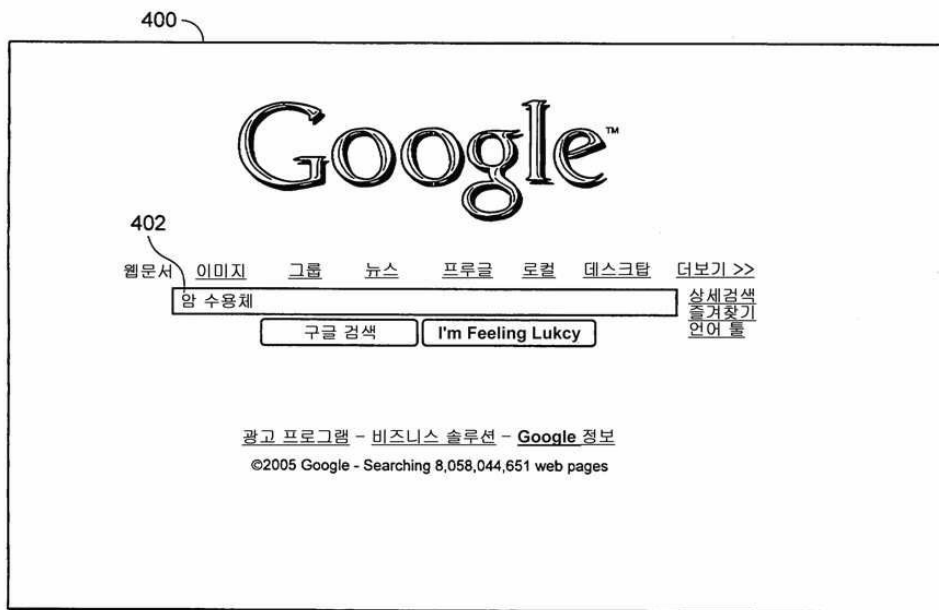
도면3d



도면3e



도면4a



도면4b

402

404

406

필터: [저널 \(2050\)](#) [공중보건\(6654\)](#) [뉴스소스 \(105\)](#) [저자 \(32\)](#) [알 수용체에 대한 3,900,000개의 결과 중 1-10\(0.28초\)](#)

유방암 수용체 주요 구조의 결정
 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ... ,저널 : 사이언스, 1995 - ncbi.nlm.nih.gov
 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 20% 정도 구조가 뒤틀어지는 수용체의 3차원 구조를 밝혀냈다고...
www.sciencedaily.com/releases/2003/02/030213073521.htm - 62k - 캐쉬 - 유사 페이지

유방암 수용체 주요 구조의 결정
 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ... ,저널 : 사이언스, 1995 - ncbi.nlm.nih.gov
 유방암 수용체의 주요 구조가 결정됨. 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 3차원 구조를 밝혀냈다고...
www.hopkinsmedical.org/press/2003/February/030212.htm - 8k - 캐쉬 - 유사 페이지

아르테미스 - 2003년 3월
 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ... ,저널 : 사이언스, 1995 - ncbi.nlm.nih.gov
 유방암 수용체의 주요 구조가 결정됨... 이 발견은 왜 HER2 수용체가 유사한 HER1과 다르게 행동하는지도 설명함...
www.hopkinsmedical.org/breastcenter/artemis/200303/feature14.htm - 12k - 캐쉬 - 유사 페이지
 [더 많은 결과는 www.hopkinsmedicine.org]

유방암 수용체 주요 구조의 결정
 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ... ,저널 : 사이언스, 1995 - ncbi.nlm.nih.gov
 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 20% 정도 구조가 뒤틀어지는 수용체의 3차원 구조를 밝혀냈다고...
www.eurekalert/pub_releases/2003-02/jhmi-sok020703.php - 8k - 캐쉬 - 유사 페이지

도면4c

402

404

408

406

필터: [공중보건\(98\)](#) [뉴스소스 \(50\)](#) [저자 \(32\)](#) [인용 \(20\)](#) [더보기 >>](#)
4,656개의 결과 중 1-10(0.28초)

저널 제목:

검색: [알 수용체>저널](#)

유방암 수용체 주요 구조의 결정
 인용횟수 : 200, 저널 : 사이언스, 1995 - ncbi.nlm.nih.gov, 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 20% 정도 구조가 뒤틀어지는 수용체의 3차원 구조를 밝혀냈다고...
www.sciencedaily.com/releases/2003/02/030213073521.htm - 62k - 캐쉬 - 유사 페이지

유방암 수용체의 주요 구조의 결정
 인용횟수 : 198, 저널 : 사이언스, 1995 - ncbi.nlm.nih.gov, 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨. 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 3차원 구조를 밝혀냈다고...
www.hopkinsmedical.org/press/2003/February/030212.htm - 8k - 캐쉬 - 유사 페이지

아르테미스 - 2003년 3월
 인용횟수 : 198, 저널 : 사이언스, 1995 - ncbi.nlm.nih.gov,
 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨... 이 발견은
 왜 HER2 수용체가 유사한 HER1과 다르게 행동하는지도 설명함...
www.hopkinsmedical.org/breastcenter/artemis/200303/feature14.htm - 12k - 캐쉬 - 유사 페이지
 [더 많은 결과는 www.hopkinsmedical.org]

유방암 수용체 주요 구조의 결정
 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ... ,저널 : 사이언스, 1995 - ncbi.nlm.nih.gov
 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 20% 정도
 구조가 뒤틀어지는 수용체의 3차원 구조를 밝혀냈다고...

도면4d

402

404

408

410

411

406

검색 : 암 수용체 > 자날 > 년도

필터 : 공중보건(98) 뉴스 소스(50) 저자(32) 인용(20) 더보기 >> 4,656개의 결과 중 1-10(0.28초)

자날 제목: []

년도: []

Go

유방암 수용체 주요 구조의 결정
 인용횟수 : 200, 저널 : 사이언스, 2004 - ncbi.nlm.nih.gov, 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 20% 정도 구조가 뒤떨어지는 수용체의 3차원 구조를 밝혀냈다고...
www.sciencedaily.com/releases/2003/02/030213073521.htm - 62k - 캐쉬 - 유사 페이지

유방암 수용체의 주요 구조의 결정
 인용횟수 : 198, 저널 : 사이언스, 2003 - ncbi.nlm.nih.gov, 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨... 이 발견은 왜 HER2 수용체가 유사한 HER1과 다르게 행동하는지도 설명함, ...
www.hopkinsmedical.org/press/2003/February/030212.htm - 8k - 캐쉬 - 유사 페이지

아르테미스 - 2003년 3월
 인용횟수 : 198, 저널 : 사이언스, 2003 - ncbi.nlm.nih.gov, 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨... 이 발견은 왜 HER2 수용체가 유사한 HER1과 다르게 행동하는지도 설명함, ...
www.hopkinsmedical.org/breastcenter/artemis/200303/feature14.htm - 12k - 캐쉬 - 유사 페이지
 [더 많은 결과는 www.hopkinsmedical.org도]

도면4e

414

402

412

413

416

406'

Google 암 수용체 구글 베이스 검색

BASE

자날 내 검색

구글 베이스 전체 검색

여러분의 아이템을 게시하세요

포스트된 아이템 : 자날

암 수용체에 대한 30개의 결과 중 1-10(0.02초)

검색 결과 걸러내기 : 생명공학(30) 의료(15) 사진(6) 날짜(10) 저자(10) 공중보건(10) 인용(6) 더보기...

관련도순 날짜순

유방암 수용체 주요 구조의 결정 - 2005년 8월 25일
 인용횟수 : 198, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨, 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 3차원 구조를 밝혀냈다고...
www.eurekalert/pub_releases/2003-02/jhmi-sok020703.php
 라벨: AIDS 기사 생명공학

아르테미스 - 2003년 3월 - 2005년 8월 23일
 인용횟수 : 198, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨... 이 발견은 왜 HER2 수용체가 유사한 HER1과 다르게 행동하는지도 설명함, ...
www.eurekalert/pub_releases/2003-02/jhmi-sok020703.php
 라벨: AIDS 기사 생명공학

유방암 수용체 주요 구조의 결정 - 2005년 8월 26일
 인용횟수 : 200, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 20% 정도 구조가 뒤떨어지는 수용체의 3차원 구조를 밝혀냈다고...
www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&iist_uids=1871007&dopt=Abstract
 라벨: 암 기사 생명공학

도면4f

414 402 412 413 416

Google BASE 암 수용체 구글 베이스 검색
 ● 저널 내 검색 ○ 구글 베이스 전체 검색

여러분의 아이템을 게시하세요

포스트된 아이템 : 저널 암 수용체에 대한 30개의 결과 중 1-10(0.02초)

420 검색 결과 걸러내기 : 생명공학(30) 의료(15) 사진(6) 날짜(10) 저자(10) 공중보건(10) 인용(6) 위치(3) 더 보기...

날짜 : 기간 [mm/dd/yyyy] - [mm/dd/yyyy] [Go]

관련도순 날짜순

406' 유방암 수용체 주요 구조의 결정 - 2005년 8월 25일
 인용횟수 : 198, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨. 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 3차원 구조를 밝혀냈다고...
 www.eurekalert/pub_releases/2003-02/jhmi-sok020703.php
 라벨: AIDS 기사 생명공학

아르테미스 - 2003년 3월 - 2005년 8월 23일
 인용횟수 : 198, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨... 이 발견은 왜 HER2 수용체가 유사한 HER1과 다르게 행동하는지도 설명함, ...
 www.eurekalert/pub_releases/2003-02/jhmi-sok020703.php
 라벨: AIDS 기사 생명공학

유방암 수용체 주요 구조의 결정 - 2005년 8월 26일
 인용횟수 : 200, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 20% 정도 구조가 뒤물어지는 수용체의 3차원 구조를 밝혀냈다고...
 www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=1871007&dopt=Abstract
 라벨: 암 기사 생명공학

도면4g

402

Google BASE 암 수용체 구글 베이스 검색
 ● 저널 내 검색 ○ 구글 베이스 전체 검색

포스트된 아이템 : 저널>의료 암 수용체에 대한 30개의 결과 중 1-10(0.02초)

404 검색 결과 걸러내기 : 생명공학(30) 의료(15) 사진(6) 날짜(10) 저자(10) 공중보건(10) 인용(6) 위치(3) 더 보기...

420 날짜 : 기간 [mm/dd/yyyy] - [mm/dd/yyyy]

422 저자 : [저자 선택]

423 [Go]

관련도순 날짜순

406' 유방암 수용체 주요 구조의 결정 - 2005년 8월 25일
 인용횟수 : 198, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨. 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 3차원 구조를 밝혀냈다고...
 www.eurekalert/pub_releases/2003-02/jhmi-sok020703.php
 라벨: AIDS 기사 생명공학

아르테미스 - 2003년 3월 - 2005년 8월 23일
 인용횟수 : 198, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 유방암 수용체의 주요 구조가 결정됨... 이 발견은 왜 HER2 수용체가 유사한 HER1과 다르게 행동하는지도 설명함, ...
 www.eurekalert/pub_releases/2003-02/jhmi-sok020703.php
 라벨: AIDS 기사 생명공학

유방암 수용체 주요 구조의 결정 - 2005년 8월 26일
 인용횟수 : 200, 저널 : 사이언스, 1995 저자 : S Markowitz, J Wang, L Myeroff, R Parsons, L Sun, ...
 존스 홉킨스와 생명공학 회사인 Genitope의 과학자 팀은 20% 정도 구조가 뒤물어지는 수용체의 3차원 구조를 밝혀냈다고...

도면5a

500

데이터에 저장되는 속성을 나타내는 방법

```

>
> 속성 및 라벨은 다음의 유형 중 하나임 :

> - BOOLEAN
> - INT
> - FLOAT
> - URL
> - STRING
> - LOCATION
> - DATE
> - DATE RANGE
>
> 속성 및 라벨은 다음의
> 메타 태그에 의해 지시됨 :
>
> <start name>
> 이름
> </end name>
> <start value>
> 값
> </end value>
    
```

도면5b

510

```

<start name> 저널 이름 </end name>
<start value> 염증에 대한 저널 </end
value>
    
```

속성 데이터의 예

도면5c

520

<start name> 의료 </end name> <start value> 널속성 </end value>

라벨 데이터의 예

도면5d

530

유형-속성 매핑

데이터 정보 유형	다른 유형의 데이터 아이템의 이 유형을 위한 속성/ 및 속성 유형
데이터 정보 유형1	속성/유형 #1-1, 속성/유형 #1-2,...
...	
데이터 정보 유형n	속성/유형 #n-1, 속성/유형 #n-2,...

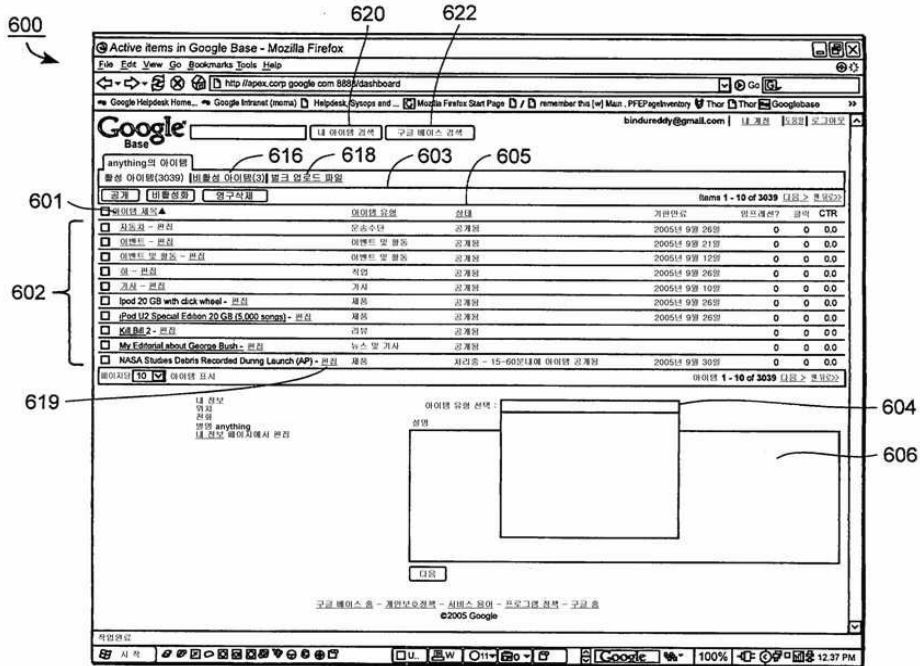
도면5e

540

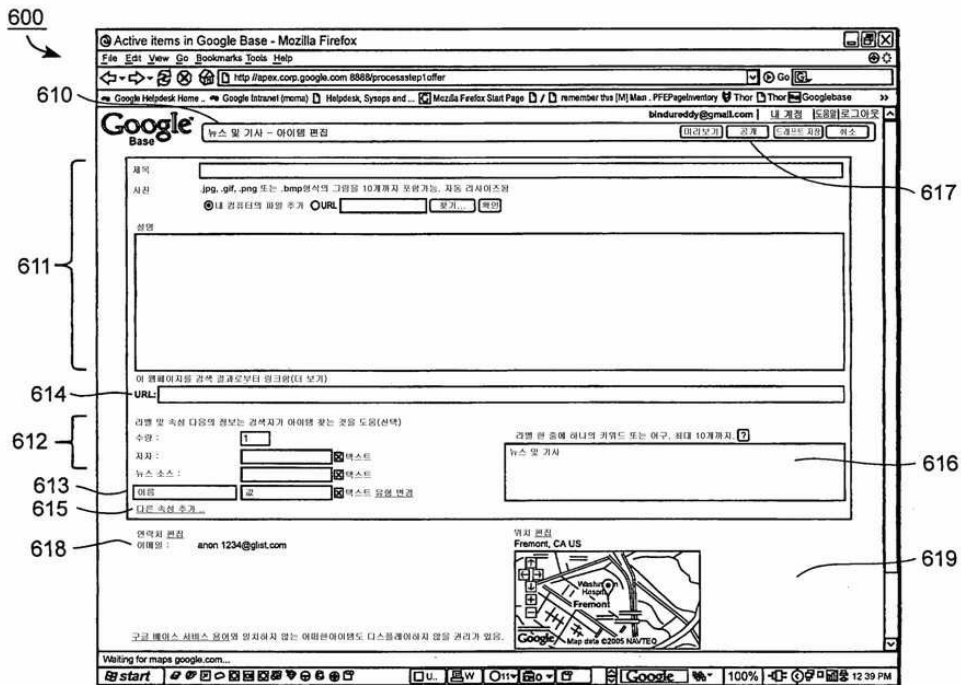
데이터 정보 유형	다른 유형의 데이터 아이템의 이 유형을 위한 속성/ 및 속성 유형
이벤트 및 활동	날짜/날짜, 휴일/ 휴일이름
...	
제품	Numavail/integer

데이터 정보 유형
및 그 속성의 예

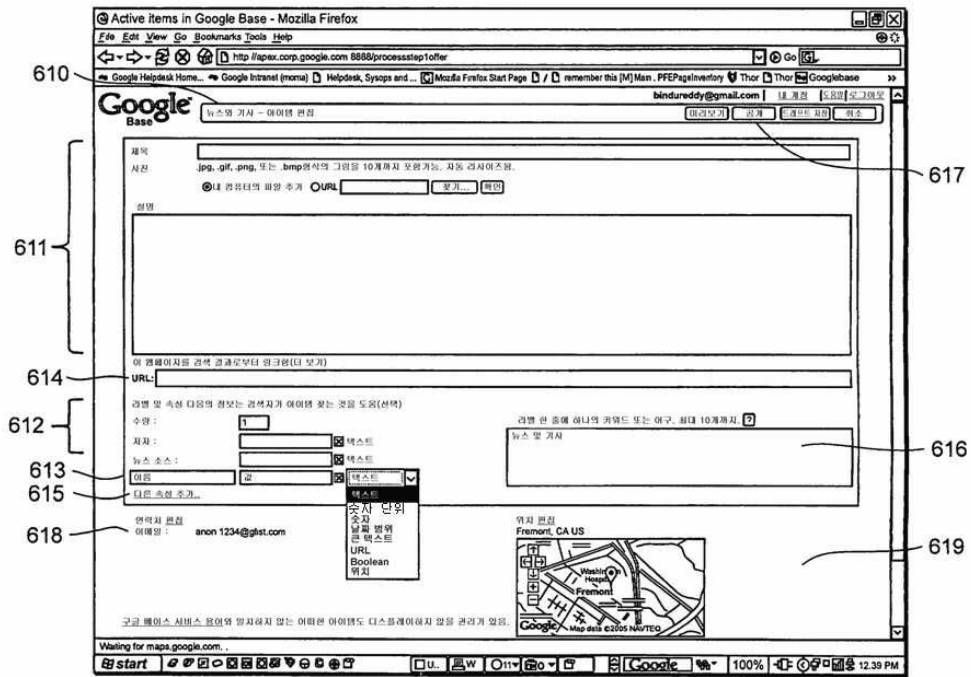
도면6a



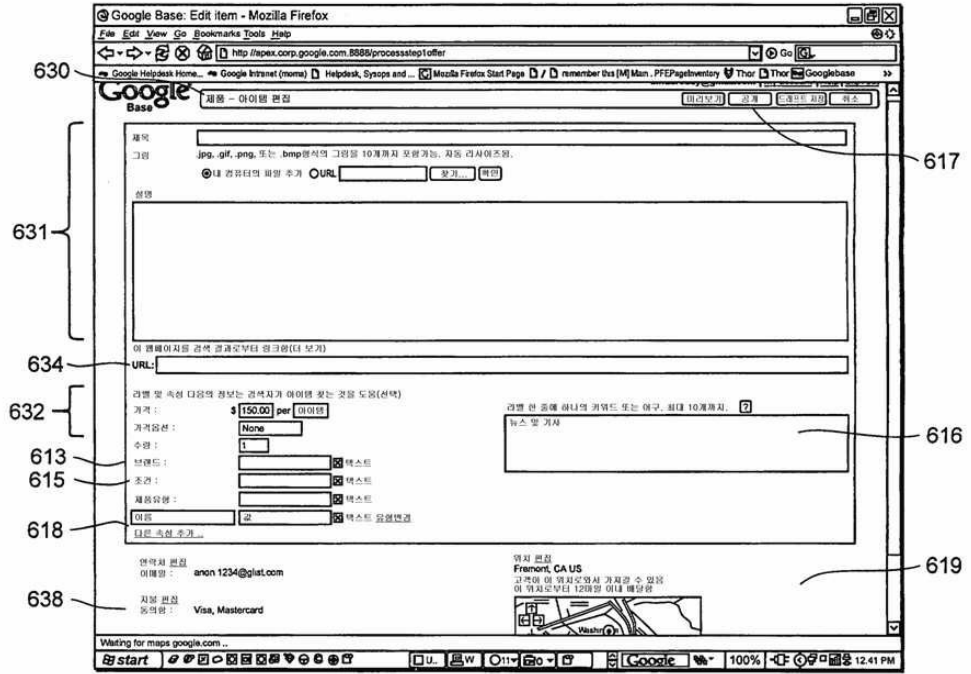
도면6b



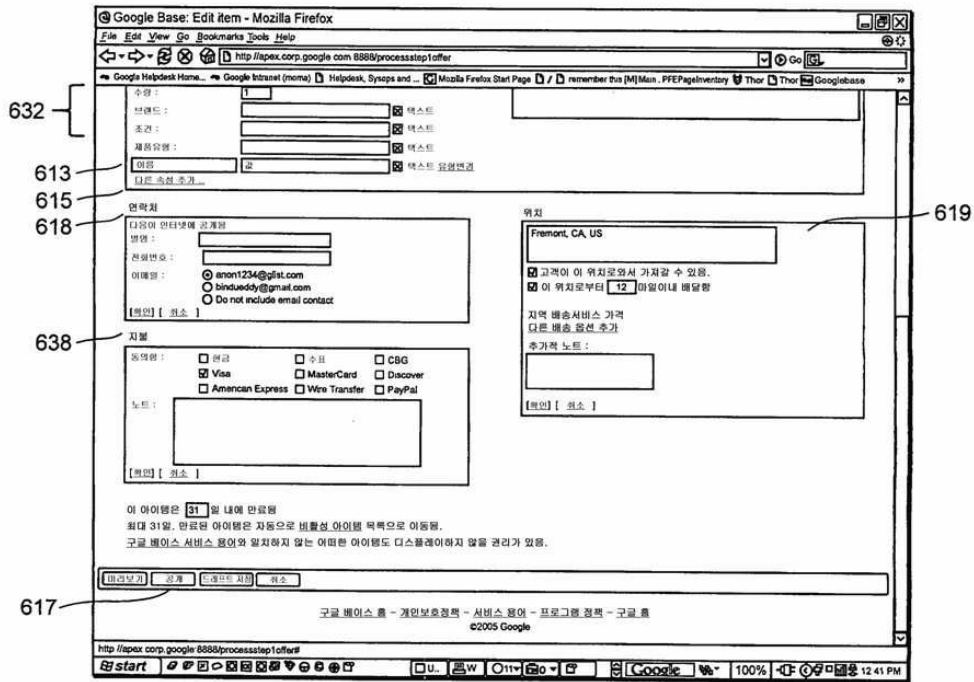
도면6c



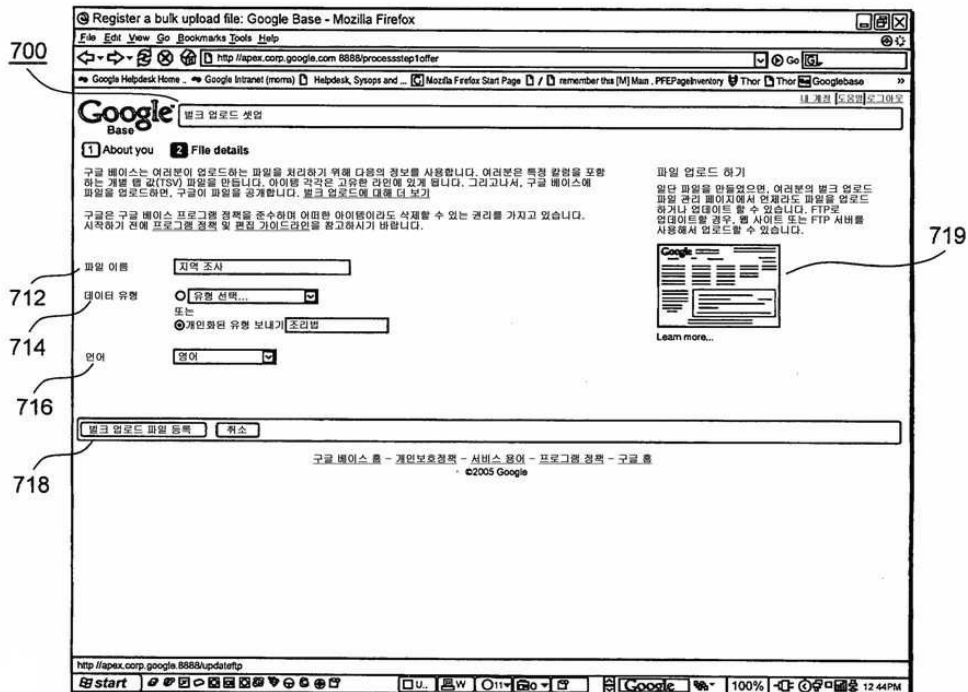
도면6d



도면6e



도면7



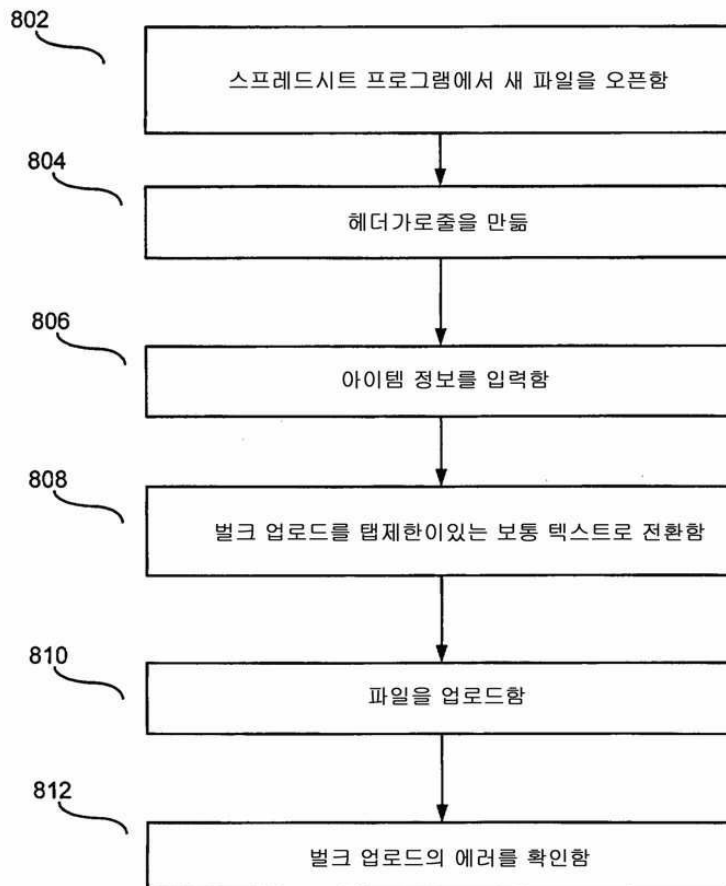
도면8a

801

```
<n names, separated by tabs>  
<item#1 attr#1 value> <tab> ...<tab> <item #1 attr#n value>  
...  
<item#m attr#1 value> <tab> ...<tab> <item #m attr#n value>  
<!-- Comments -->
```

도면8b

800



도면8c

832

834

	A	B	C	D	E	F	G
1	링크	제목	설명	이미지 링크	id	기환만료_날짜	리뷰
2	http://www.example.com	무관 알루미늄	이 멋진 금속 상발이	http://www.examp	1	12/20/2005	보석, 알루미늄
3							
4							
5							
6							
7							
8							
9							
10							

도면8d

841

842

844

846

848

상태	파일 이름	유형	기환만료	업로드연료	최종연료	특색	제어
<input type="checkbox"/>	aboutme.txt		**	**	**	**	**
<input type="checkbox"/>	aboutme.txt	Travel	**	**	**	**	**
<input type="checkbox"/>	localinventory.html	Recipes	**	**	**	**	**
<input type="checkbox"/>	news.txt	Events and activities	**	**	**	**	**
<input type="checkbox"/>	newfed.txt		**	**	**	**	**
<input type="checkbox"/>	p.rst	Products	**	**	**	**	**

FTP 정보
 서버: headwig.google.com
 사용자 이름: 이름
 비밀번호 변경
 FTP에 대한 도움말

파일 업로드 - 업로드 설명 보기

 주의: 파일 업로더는 기존 파일을 완전히 삭제하여 새로운 버전으로 대체할
 이 파일을 업로드하고 저장

• 다른 링크 업로드 설명 등록
 • 특정 업로드를 위한 FTP 계정 만들기 ?

start