



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2016년10월18일
(11) 등록번호 10-1667346
(24) 등록일자 2016년10월12일

- (51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06K 9/00 (2006.01)
- (21) 출원번호 10-2012-7006113
- (22) 출원일자(국제) 2012년08월05일
심사청구일자 2015년08월04일
- (85) 번역문제출일자 2012년03월07일
- (65) 공개번호 10-2012-0058538
- (43) 공개일자 2012년06월07일
- (86) 국제출원번호 PCT/US2010/044603
- (87) 국제공개번호 WO 2011/017557
국제공개일자 2011년02월10일
- (30) 우선권주장
12/850,483 2010년08월04일 미국(US)
(뒷면에 계속)
- (56) 선행기술조사문헌
KR1020080074955 A*
US20080267504 A1*
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
구글 인코포레이티드
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이
1600 (우:94043)
- (72) 발명자
페트로우, 데이비드
미국 11231 뉴욕 브룩클린 데그라우 스트리트 350
아파트먼트 #2
- (74) 대리인
박장원

전체 청구항 수 : 총 18 항

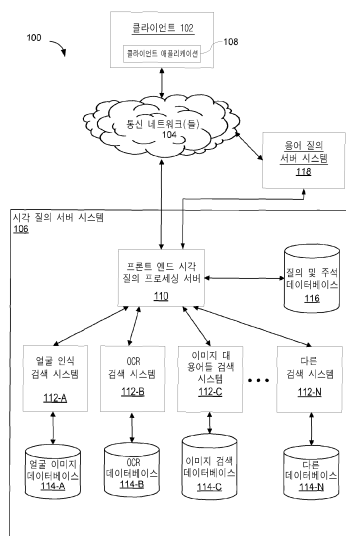
심사관 : 홍경아

(54) 발명의 명칭 시각 질의에 응답하기 위한 아키텍처

(57) 요약

사진, 스크린 샷, 스캔된 이미지, 비디오 프레임, 또는 콘텐츠 저작 애플리케이션에 의해 생성된 이미지와 같은 시각 질의는 시각 질의 검색 시스템에 제출된다. 검색 시스템은 각각이 별개의 시각 질의 검색 프로세스를 구현하는 복수의 병렬 검색 시스템들에 이를 송신함으로써 시각 질의를 프로세싱한다. 이러한 병렬 검색 시스템들은 (뒷면에 계속)

대표도 - 도1



광학식 문자 인식(OCR), 얼굴 인식, 물건 인식, 바 코드 인식, 랜드마크 인식, 지리-위치 정보에 의해 지원되는 장소 인식, 유사 이미지 인식, 오브젝트-또는-오브젝트-카테고리 인식, 명명 엔티티 인식, 및 색 인식을 포함할 수 있다(이에 제한되지 않음). 그 다음에 적어도 하나의 검색 결과는 클라이언트 시스템에 송신된다. 임의의 실시예들에서, 시각 질의가 텍스트 엘리먼트 및 비-텍스트 엘리먼트를 포함하는 이미지일 때, 적어도 하나의 검색 결과는 텍스트 엘리먼트에 대한 광학식 문자 인식 결과 및 비-텍스트 엘리먼트에 대한 적어도 하나의 이미지-매칭 결과를 포함한다.

(30) 우선권주장

61/232,397 2009년08월07일 미국(US)

61/266,116 2009년12월02일 미국(US)

명세서

청구범위

청구항 1

시각 질의(visual query)를 프로세싱하는 컴퓨터-구현된 방법으로서,

하나 이상의 프로세서들 및 상기 하나 이상의 프로세서들에 의한 실행을 위해 하나 이상의 프로그램들을 저장하는 메모리를 포함하는 컴퓨팅 디바이스에서,

2 이상의 오브젝트들을 가지는 시각 질의를 획득하는 단계, 상기 2 이상의 오브젝트들은:

(i) 제1 오브젝트 타입을 가지는 제1 오브젝트 및 (ii) 상기 제1 타입과 별개인 제2 오브젝트 타입을 가지는 제2 오브젝트를 포함하며, 상기 제1 및 제2 오브젝트 타입들은 OCR 문자들, 사람의 얼굴, 비-인간 오브젝트(non-human object) 및 바코드를 포함하는 그룹으로부터 선택되며;

상기 시각 질의를 제1 영역 및 제2 영역을 포함하는 2 이상의 영역들로 분할(partitioning)하는 단계, 상기 제1 영역은 상기 제1 오브젝트를 포함하며 그리고 상기 제2 영역은 상기 제2 오브젝트를 포함하며;

시각 질의 검색 결과들을 동시에 획득함으로써 상기 시각 질의를 프로세싱하는 단계, 상기 시각 질의 검색 결과들은:

(i)상기 제1 오브젝트에 따라 생성된 결과들의 제1 세트 및

(ii)상기 제2 오브젝트에 따라 생성된 결과들의 제2 세트를 포함하며;

(i)상기 결과들의 제1 세트 및 (ii)상기 결과들의 제2 세트를 사용자에게 동시에 디스플레이하기 위해 포매팅(formatting)하는 단계;

상기 결과들의 제1 세트 또는 상기 결과들의 제2 세트에서 특정 결과의 하나 이상의 사용자 주석들을 획득하는 단계, 상기 하나 이상의 사용자 주석들은 상기 시각 쿼리에 대한 각각의 검색 결과의 관련성 또는 관련성의 흠결을 표시하는 사용자에게 의해 취해진 액션을 표시하며;

제2 시각 쿼리를 획득하는 단계; 및

상기 제2 시각 쿼리를 획득함에 응답하여, 상기 하나 이상의 사용자 주석들에서 적어도 하나의 주석에 기초하여 제2 복수의 검색 결과들을 획득하는 단계를 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하는 컴퓨터-구현된 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

청구항 1에 있어서,

상기 시각 질의는 사진, 스크린 샷, 스캔된 이미지, 비디오 프레임 및 복수의 비디오 프레임들로 구성되는 그룹으로부터 선택되는 것을 특징으로 하는 시각 질의를 프로세싱하는 컴퓨터-구현된 방법.

청구항 5

삭제

청구항 6

삭제

청구항 7

삭제

청구항 8

삭제

청구항 9

시각 질의를 프로세싱하기 위한 컴퓨터 시스템으로서,

프로그램들을 실행하기 위한 하나 이상의 중앙 프로세싱 유닛들;

상기 하나 이상의 중앙 프로세싱 유닛들에 의해 실행될 하나 이상의 프로그램들을 저장하는 메모리를 포함하고, 상기 하나 이상의 프로그램들은:

2 이상의 오브젝트들을 가지는 시각 질의를 획득하고, 상기 2 이상의 오브젝트들은:

(i) 제1 오브젝트 타입을 가지는 제1 오브젝트 및 (ii) 상기 제1 타입과 별개인 제2 오브젝트 타입을 가지는 제2 오브젝트를 포함하며, 상기 제1 및 제2 오브젝트 타입들은 OCR 문자들, 사람의 얼굴, 비-인간 오브젝트(non-human object) 및 바코드를 포함하는 그룹으로부터 선택되며;

상기 시각 질의를 제1 영역 및 제2 영역을 포함하는 2 이상의 영역들로 분할(partitioning)하고, 상기 제1 영역은 상기 제1 오브젝트를 포함하며 그리고 상기 제2 영역은 상기 제2 오브젝트를 포함하며;

시각 질의 검색 결과들을 동시에 획득함으로써 상기 시각 질의를 프로세싱하고, 상기 시각 질의 검색 결과들은:

(i)상기 제1 오브젝트에 따라 생성된 결과들의 제1 세트 및

(ii)상기 제2 오브젝트에 따라 생성된 결과들의 제2 세트를 포함하며;

(i)상기 결과들의 제1 세트 및 (ii)상기 결과들의 제2 세트를 사용자에게 동시에 디스플레이하기 위해 포매팅(formatting)하고;

상기 결과들의 제1 세트 또는 상기 결과들의 제2 세트에서 특정 결과의 하나 이상의 사용자 주석들을 획득하고, 상기 하나 이상의 사용자 주석들은 상기 시각 쿼리에 대한 각각의 검색 결과의 관련성 또는 관련성의 흠결을 표시하는 사용자에게 의해 취해진 액션을 표시하며;

제2 시각 쿼리를 획득하고; 및

상기 제2 시각 쿼리를 획득함에 응답하여, 상기 하나 이상의 사용자 주석들에서 적어도 하나의 주석에 기초하여 제2 복수의 검색 결과들을 획득하기 위한 명령들을 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 시스템.

청구항 10

삭제

청구항 11

삭제

청구항 12

청구항 9에 있어서,

상기 시각 질의는 사진, 스크린 샷, 스캔된 이미지, 비디오 프레임 및 복수의 비디오 프레임들로 구성되는 그룹으로부터 선택되는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 시스템.

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

컴퓨터에 의한 실행을 위해 구성된 하나 이상의 프로그램들을 저장하는, 시각 질의를 프로세싱하기 위한 컴퓨터 판독가능한 저장 매체로서, 상기 하나 이상의 프로그램들은:

2 이상의 오브젝트들을 가지는 시각 질의를 획득하고, 상기 2 이상의 오브젝트들은:

(i) 제1 오브젝트 타입을 가지는 제1 오브젝트 및 (ii) 상기 제1 타입과 별개인 제2 오브젝트 타입을 가지는 제2 오브젝트를 포함하며, 상기 제1 및 제2 오브젝트 타입들은 OCR 문자들, 사람의 얼굴, 비-인간 오브젝트(non-human object) 및 바코드를 포함하는 그룹으로부터 선택되며;

상기 시각 질의를 제1 영역 및 제2 영역을 포함하는 2 이상의 영역들로 분할(partitioning)하고, 상기 제1 영역은 상기 제1 오브젝트를 포함하며 그리고 상기 제2 영역은 상기 제2 오브젝트를 포함하며;

시각 질의 검색 결과들을 동시에 획득함으로써 상기 시각 질의를 프로세싱하고, 상기 시각 질의 검색 결과들은:

(i)상기 제1 오브젝트에 따라 생성된 결과들의 제1 세트 및

(ii)상기 제2 오브젝트에 따라 생성된 결과들의 제2 세트를 포함하며;

(i)상기 결과들의 제1 세트 및 (ii)상기 결과들의 제2 세트를 사용자에게 동시에 디스플레이하기 위해 포매팅(formatting)하고;

상기 결과들의 제1 세트 또는 상기 결과들의 제2 세트에서 특정 결과의 하나 이상의 사용자 주석들을 획득하고, 상기 하나 이상의 사용자 주석들은 상기 시각 쿼리에 대한 각각의 검색 결과의 관련성 또는 관련성의 흠결을 표시하는 사용자에게 의해 취해진 액션을 표시하며;

제2 시각 쿼리를 획득하고; 및

상기 제2 시각 쿼리를 획득함에 응답하여, 상기 하나 이상의 사용자 주석들에서 적어도 하나의 주석에 기초하여 제2 복수의 검색 결과들을 획득하기 위한 명령들을 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 판독가능한 저장 매체.

청구항 18

삭제

청구항 19

삭제

청구항 20

청구항 17에 있어서,

상기 시각 질의는 사진, 스크린 샷, 스캔된 이미지, 비디오 프레임 및 복수의 비디오 프레임들로 구성되는 그룹

으로부터 선택되는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 판독가능한 저장 매체.

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

청구항 1에 있어서,

상기 하나 이상의 사용자 주석들은 상기 시각 쿼리의 사용자-표시 부분 및 상기 시각 쿼리의 상기 부분에 디스플레이되는 오브젝트를 식별하는 대응하는 사용자-제공 정보를 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하는 컴퓨터-구현된 방법.

청구항 26

청구항 25에 있어서,

상기 대응하는 사용자-제공 정보는 상기 제1 오브젝트 또는 상기 제2 오브젝트를 식별하는 텍스트를 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하는 컴퓨터-구현된 방법.

청구항 27

청구항 1에 있어서,

상기 결과들의 제1 세트 및 상기 결과들의 제2 세트는 상기 사용자에게 디스플레이되는 서로 다른 그룹들로 그룹화되는 것을 특징으로 하는 시각 질의를 프로세싱하는 컴퓨터-구현된 방법.

청구항 28

청구항 1에 있어서,

상기 하나 이상의 사용자 주석들은 상기 시각 쿼리의 사용자 설명, 상기 시각 쿼리 부분의 사용자 설명, 사용자 리뷰 및 각각의 검색 결과의 사용자 수정으로 구성되는 그룹으로부터 선택되는 것을 특징으로 하는 시각 질의를 프로세싱하는 컴퓨터-구현된 방법.

청구항 29

청구항 9에 있어서,

상기 하나 이상의 사용자 주석들은 상기 시각 쿼리의 사용자-표시 부분 및 상기 시각 쿼리의 상기 부분에 디스플레이되는 오브젝트를 식별하는 대응하는 사용자-제공 정보를 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 시스템.

청구항 30

청구항 29에 있어서,

상기 대응하는 사용자-제공 정보는 상기 제1 오브젝트 또는 상기 제2 오브젝트를 식별하는 텍스트를 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 시스템.

청구항 31

청구항 9에 있어서,

상기 결과들의 제1 세트 및 상기 결과들의 제2 세트는 상기 사용자에게 디스플레이되는 서로 다른 그룹들로 그룹화되는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 시스템.

청구항 32

청구항 9에 있어서,

상기 하나 이상의 사용자 주석들은 상기 시각 쿼리의 사용자 설명, 상기 시각 쿼리 부분의 사용자 설명, 사용자 리뷰 및 각각의 검색 결과의 사용자 수정으로 구성되는 그룹으로부터 선택되는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 시스템.

청구항 33

청구항 17에 있어서,

상기 하나 이상의 사용자 주석들은 상기 시각 쿼리의 사용자-표시 부분 및 상기 시각 쿼리의 상기 부분에 디스플레이되는 오브젝트를 식별하는 대응하는 사용자-제공 정보를 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 판독가능한 저장 매체.

청구항 34

청구항 33에 있어서,

상기 대응하는 사용자-제공 정보는 상기 제1 오브젝트 또는 상기 제2 오브젝트를 식별하는 텍스트를 포함하는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 판독가능한 저장 매체.

청구항 35

청구항 17에 있어서,

상기 결과들의 제1 세트 및 상기 결과들의 제2 세트는 상기 사용자에게 디스플레이되는 서로 다른 그룹들로 그룹화되는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 판독가능한 저장 매체.

청구항 36

청구항 17에 있어서,

상기 하나 이상의 사용자 주석들은 상기 시각 쿼리의 사용자 설명, 상기 시각 쿼리 부분의 사용자 설명, 사용자 리뷰 및 각각의 검색 결과의 사용자 수정으로 구성되는 그룹으로부터 선택되는 것을 특징으로 하는 시각 질의를 프로세싱하기 위한 컴퓨터 판독가능한 저장 매체.

발명의 설명

기술 분야

[0001] 개시된 실시예들은 일반적으로 시각 질의를 프로세싱하기 위한 복수의 병렬 검색 시스템들을 포함하는 서버 시스템 아키텍처에 관한 것이다.

배경 기술

[0002] 사용자가 검색 엔진에 단어 또는 구문을 입력하고 다양한 결과들을 수신하는, 텍스트 기반 또는 용어-기반 검색은 검색을 위한 유용한 툴이다. 그러나, 용어 기반 질의들은 사용자가 관련 용어를 입력할 수 있음을 요구한다. 때때로, 사용자는 이미지에 관한 정보를 알고자할 수 있다. 예를 들어, 사용자는 사진에서 사람의 명칭을 알기를 원할 수 있거나 사용자는 픽처에서 꽃 또는 새의 명칭을 알기를 원할 수 있다. 따라서, 시각 질의를 수신하고 검색 결과들을 제공할 수 있는 시스템이 바람직할 수 있다.

발명의 내용

- [0003] 임의의 실시예들에 따라, 서버 시스템에서 시각 질의를 프로세싱하는 컴퓨터-구현된 방법이 존재한다. 시각 질의는 클라이언트 시스템으로부터 수신된다. 시각 질의는 동시 프로세싱을 위해 복수의 병렬 검색 시스템들에 시각 질의를 송신함으로써 프로세싱된다. 복수의 검색 시스템들의 각각은 복수의 시각 질의 검색 프로세스들 중 별개의 시각 질의 검색 프로세스를 구현한다. 복수의 시각 질의 검색 프로세스들은 적어도, 광학식 문자 인식(OCR), 얼굴 인식, 및 OCR 및 얼굴 인식 외에 제 1 이미지에 의한 질의(query-by-image) 프로세스를 포함한다. 복수의 검색 결과들은 복수의 병렬 검색 시스템들 중 하나 이상으로부터 수신된다. 복수의 검색 결과들 중 적어도 하나는 클라이언트 시스템에 송신된다.
- [0004] 임의의 실시예들에서, 방법은 수신된 검색 결과들 중 적어도 둘이 미리 정의된 기준을 만족시킬 때, 미리 정의된 기준을 만족하는 수신된 검색 결과들을 순위화하는 단계 및 클라이언트 시스템에 순위화된 검색 결과들 중 적어도 하나의 검색 결과를 송신하는 단계를 더 포함한다.
- [0005] 임의의 실시예들에서, 제 1 이미지에 의한 질의 프로세스는 물건 인식, 바 코드 인식, 오브젝트-또는-오브젝트-카테고리 인식, 명명 엔터티 인식 또는 색 인식이다.
- [0006] 임의의 실시예들에서, 시각 질의는 사진, 스크린 샷, 스캔된 이미지 또는 비디오 프레임이다. 클라이언트 시스템은 모바일 디바이스, 데스크톱 디바이스 또는 다른 디바이스일 수 있다.
- [0007] 임의의 실시예들에서, 시각 질의는 검색 애플리케이션, 브라우저 애플리케이션에 대한 검색 엔진 플러그-인 또는 브라우저 애플리케이션에 대한 검색 엔진 확장과 같은, 클라이언트 시스템에 의해 실행되는 클라이언트 애플리케이션으로부터 수신된다. 임의의 실시예들에서, 시각 질의는 클라이언트 시스템에 의해 실행되는 콘텐츠 저작(authoring) 애플리케이션으로부터 수신된다.
- [0008] 시각 질의가 텍스트 엘리먼트 및 비-텍스트 엘리먼트를 포함하는 이미지일 때, 임의의 실시예들에서, 검색 결과는 텍스트 엘리먼트에 대한 광학식 문자 인식 결과 및 비-텍스트 엘리먼트에 대한 적어도 하나의 이미지-매칭 결과를 포함한다.
- [0009] 임의의 실시예들에서, 시각 질의가 텍스트 엘리먼트 및 비-텍스트 엘리먼트를 포함하는 이미지일 때, 검색 결과는 광학식 문자 인식 프로세스에 의해 산출되는 검색 결과에 링크를 가진 텍스트 엘리먼트에 대한 제 1 시각 식별자 및 이미지-매칭 프로세스에 의해 산출되는 검색 결과에 링크를 가진 비-텍스트 엘리먼트에 대한 제 2 시각 식별자를 포함하는 상호작용 결과 문서를 포함한다.
- [0010] 임의의 실시예들에서, 방법은 복수의 검색 결과들 중 적어도 둘을 복합 검색 결과로 결합하는 단계를 더 포함한다.
- [0011] 임의의 실시예들에 따라, 검색 엔진 시스템은 시각 질의를 프로세싱하기 위해 제공된다. 시스템은 프로그램들을 실행하기 위한 하나 이상의 중앙 프로세싱 유닛들 및 상기 하나 이상의 중앙 프로세싱 유닛들에 의해 실행될 하나 이상의 프로그램들을 저장하는 메모리를 포함한다. 하나 이상의 프로그램들은 이하를 수행하기 위한 명령들을 포함한다. 시각 질의가 클라이언트 시스템으로부터 수신된다. 시각 질의는 동시 프로세싱을 위해 복수의 병렬 검색 시스템들에 시각 질의를 송신함으로써 프로세싱된다. 복수의 검색 시스템들의 각각은 복수의 시각 질의 검색 프로세스들 중 별개의 시각 질의 검색 프로세스를 구현한다. 복수의 시각 질의 검색 프로세스들은 적어도, 광학식 문자 인식(OCR), 얼굴 인식, 및 OCR 및 얼굴 인식 외에 제 1 이미지에 의한 질의 프로세스를 포함한다. 복수의 검색 결과들은 복수의 병렬 검색 시스템들 중 하나 이상으로부터 수신된다. 복수의 검색 결과들 중 적어도 하나는 클라이언트 시스템에 송신된다. 이러한 시스템은 상기 논의된 추가적인 옵션들을 실행하기 위한 프로그램 명령들을 더 포함할 수 있다.
- [0012] 임의의 실시예들에 따라, 시각 질의를 프로세싱하기 위한 컴퓨터 판독가능한 저장 매체 시스템이 제공된다. 컴퓨터 판독가능한 저장 매체는 컴퓨터에 의해 실행하기 위해 구성된 하나 이상의 프로그램들을 저장하고, 하나 이상의 프로그램들은 이하를 수행하기 위한 명령들을 포함한다. 시각 질의가 클라이언트 시스템으로부터 수신된다. 시각 질의는 동시 프로세싱을 위해 복수의 병렬 검색 시스템들에 시각 질의를 송신함으로써 프로세싱된다. 복수의 검색 시스템들의 각각은 복수의 시각 질의 검색 프로세스들 중 별개의 시각 질의 검색 프로세스를 구현한다. 복수의 시각 질의 검색 프로세스들은 적어도, 광학식 문자 인식(OCR), 얼굴 인식 및 OCR 및 얼굴 인식 외에 제 1 이미지에 의한 질의 프로세스를 포함한다. 복수의 검색 결과들은 복수의 병렬 검색 시스템들 중 하나 이상으로부터 수신된다. 복수의 검색 결과들 중 적어도 하나는 클라이언트 시스템에 송신된다. 이러한 컴퓨터 판독가능한 저장 매체는 상기 논의된 추가적인 옵션들을 실행하기 위한 프로그램 명령들을 더 포함할 수 있다.

도면의 간단한 설명

- [0013] 도 1은 시각 질의 서버 시스템을 포함하는 컴퓨터 네트워크를 도시하는 블록도이다.
 - 도 2는 임의의 실시예들에 따라, 시각 질의에 응답하기 위한 프로세스를 도시하는 흐름도이다.
 - 도 3은 임의의 실시예들에 따라, 상호작용 결과 문서를 이용하여 시각 질의에 응답하기 위한 프로세스를 도시하는 흐름도이다.
 - 도 4는 임의의 실시예들에 따라, 클라이언트 및 시각 질의 서버 시스템 사이에서 통신들을 도시하는 흐름도이다.
 - 도 5는 임의의 실시예들에 따라, 클라이언트 시스템을 도시하는 블록도이다.
 - 도 6은 임의의 실시예들에 따라, 프론트 엔드 시각 질의 프로세싱 서버 시스템을 도시하는 블록도이다.
 - 도 7은 임의의 실시예들에 따라, 시각 질의를 프로세싱하기 위해 이용되는 병렬 검색 시스템들 중 포괄적인 시스템을 도시하는 블록도이다.
 - 도 8은 임의의 실시예들에 따라, 시각 질의를 프로세싱하기 위해 이용되는 OCR 검색 시스템을 도시하는 블록도이다.
 - 도 9는 임의의 실시예들에 따라, 시각 질의를 프로세싱하기 위해 이용되는 얼굴 인식 검색 시스템을 도시하는 블록도이다.
 - 도 10은 임의의 실시예들에 따라, 시각 질의를 프로세싱하기 위해 이용되는 이미지 대 용어들 검색 시스템을 도시하는 블록도이다.
 - 도 11은 임의의 실시예들에 따라, 예시적인 시각 질의의 스크린 샷을 가진 클라이언트 시스템을 도시한다.
 - 도 12a 및 12b 각각은 임의의 실시예들에 따라, 바운딩 박스들을 가진 상호작용 결과 문서의 스크린 샷을 가진 클라이언트 시스템을 도시한다.
 - 도 13은 임의의 실시예들에 따라, 타입에 의해 코딩된 상호작용 결과 문서의 스크린 샷을 가진 클라이언트 시스템을 도시한다.
 - 도 14는 임의의 실시예들에 따라, 라벨들을 가진 상호작용 결과 문서의 스크린 샷을 가진 클라이언트 시스템을 도시한다.
 - 도 15는 임의의 실시예들에 따라, 결과 리스트와 동시에 디스플레이되는 상호작용 결과 문서 및 시각 질의의 스크린 샷을 도시한다.
- 동일한 참조 번호들은 도면들 전체를 통해 대응하는 부분들을 지칭한다.

발명을 실시하기 위한 구체적인 내용

- [0014] 이제 첨부된 도면들에서 설명되는 실시예들에 대한 참조가 상세히 이루어질 것이다. 아래의 상세한 설명에서, 많은 특정 세부사항들이 본 발명의 완전한 이해를 제공하기 위해 설명된다. 하지만, 본 발명이 이러한 특정 세부사항들 없이 실시될 수 있음이 당업자에게 명백할 것이다. 다른 예시들에서, 공지된 방법들, 절차들, 컴포넌트들, 회로들 및 네트워크들이 실시예들의 양상들을 필요이상으로 불분명하지 않도록 상세히 설명되지 않았다.
- [0015] 용어들 제 1, 제 2 등이 다양한 엘리먼트들을 설명하기 위해 본 명세서에 존재할 수 있지만, 이러한 엘리먼트들은 이러한 용어들에 의해 제한되지 않아야 함이 또한 이해될 것이다. 이러한 용어들은 하나의 엘리먼트를 다른 엘리먼트와 구별하기 위해서만 사용된다. 본 발명의 범위로부터 벗어남이 없이 예를 들어, 제 1 컨택트는 제 2 컨택트로 지칭될 수 있고, 유사하게, 제 2 컨택트는 제 1 컨택트로 지칭될 수 있다. 제 1 컨택트 및 제 2 컨택트는 둘 다 컨택트들이지만, 동일한 컨택트가 아니다.
- [0016] 본 명세서에서 본 발명의 설명에서 사용되는 용어는 특정 실시예들을 설명하기 위한 목적일 뿐, 본 발명의 제한으로 의도되지 않는다. 본 발명의 설명 및 첨부된 청구항들에서 사용되는 바와 같이, 단수 형태들은 문맥이 명백하게 달리 나타내지 않는 한 또한 복수 형태들을 포함하고자 의도된다. 또한 본 명세서에서 사용되는 용어

"및/또는"은 연관된 리스트된 아이템들 중 하나 이상 중 임의의 그리고 모든 가능한 조합들을 지칭하고 포함한다고 이해될 것이다. 또한, 용어들 "포함하는(comprises 및/또는 comprising)"은 이 상세에서 사용될 때, 언급된 특징들, 정수들, 단계들, 동작들, 엘리먼트들, 및/또는 컴포넌트들을 식별하지만, 하나 이상의 다른 특징들, 정수들, 단계들, 동작들, 엘리먼트들, 컴포넌트들, 및/또는 이들의 그룹들의 존재 또는 부가를 못하게 하는 것이 아니다.

- [0017] 본 명세서에서 사용되는, 용어 "~할 경우(if)"는 문맥에 따라 "때(when)" 또는 "~할 시(upon)" 또는 "결정에 응답하여" 또는 "검출에 응답하여"를 의미하도록 해석될 수 있다. 유사하게, 구문 "결정될 경우" 또는 "검출될 경우"는 문맥에 따라 "결정 시", 또는 "결정에 응답하여" 또는 "(언급된 조건 또는 이벤트)검출 시" 또는 "(언급된 조건 또는 이벤트)검출에 응답하여"를 의미하도록 해석될 수 있다.
- [0018] 도 1은 임의의 실시예들에 따른 시각 질의 서버 시스템을 포함하는 컴퓨터 네트워크를 도시하는 블록도이다. 컴퓨터 네트워크(100)는 하나 이상의 클라이언트 시스템들(102) 및 시각 질의 서버 시스템(106)을 포함한다. 하나 이상의 통신 네트워크들(104)은 이러한 컴포넌트들을 상호접속한다. 통신 네트워크(104)는 로컬 영역 네트워크들(LAN), 광역 네트워크들(WAN), 무선 네트워크들, 유선 네트워크들, 인터넷 또는 이러한 네트워크들의 조합을 포함하는, 다양한 네트워크들 중 임의의 네트워크일 수 있다.
- [0019] 클라이언트 시스템(102)은 시각 질의(예를 들어, 도 11의 시각 질의(1102))를 수신하기 위해, 클라이언트 시스템에 의해 실행되는 클라이언트 애플리케이션(108)을 포함한다. 시각 질의는 검색 엔진 또는 검색 시스템에 질 의로서 제출되는 이미지이다. 제한이 아닌, 시각 질의들의 예시들은 사진들, 스캔된 문서들 및 이미지들 및 그림들을 포함한다. 임의의 실시예들에서, 클라이언트 애플리케이션(108)은 검색 애플리케이션, 브라우저 애플리케이션에 대한 검색 엔진 플러그-인 및 브라우저 애플리케이션에 대한 검색 엔진 확장으로 구성되는 세트로부터 선택된다. 임의의 실시예들에서, 클라이언트 애플리케이션(108)은 "잡식성(omnivorous)" 검색 박스이고, 이는 사용자가 시각 질의로서 사용될 임의의 포맷의 이미지를 검색 박스로 드래그 앤 드롭할 수 있도록 한다.
- [0020] 클라이언트 시스템(102)은 시각 질의 서버 시스템(106)에 질의들을 송신하고 시각 질의 서버 시스템(106)으로부터 데이터를 수신한다. 클라이언트 시스템(102)은 임의의 컴퓨터 또는 시각 질의 서버 시스템(106)과 통신할 수 있는 다른 디바이스일 수 있다. 제한이 아닌, 예시들은 데스크톱 및 노트북 컴퓨터들, 메인프레임 컴퓨터들, 서버 컴퓨터들, 휴대 전화들 및 개인 휴대 정보 단말들과 같은 모바일 디바이스들, 네트워크 단말들 및 셋-톱 박스들을 포함한다.
- [0021] 시각 질의 서버 시스템(106)은 프론트 엔드 시각 질의 프로세싱 서버(110)를 포함한다. 프론트 엔드 서버(110)는 클라이언트(102)로부터 시각 질의를 수신하고 동시 프로세싱을 위해 복수의 병렬 검색 시스템들(112)에 시각 질의를 송신한다. 검색 시스템들(112) 각각은 별개의 시각 질의 검색 프로세스를 구현하고 자신들의 별개의 검색 프로세스에 의해 시각 질의를 프로세스할 필요에 따라 자신들의 대응하는 데이터베이스들(114)에 액세스한다. 예를 들어, 얼굴 인식 검색 시스템(112-A)은 이미지 질의에 대한 얼굴 매칭들을 찾기 위해 얼굴 이미지 데이터베이스(114-A)에 액세스할 것이다. 도 9와 관련하여 더 상세히 설명될 바와 같이, 시각 질의가 얼굴을 포함하는 경우, 얼굴 인식 검색 시스템(112-A)은 얼굴 이미지 데이터베이스(114-A)로부터 하나 이상의 검색 결과들(예를 들어, 명칭들, 매칭 얼굴들 등)을 리턴할 것이다. 다른 예시에서, 광학식 문자 인식(OCR) 검색 시스템(112-B)은 시각 질의에서 임의의 인식가능한 텍스트를 하나 이상의 검색 결과들로서 리턴하기 위한 텍스트로 변환한다. 광학식 문자 인식(OCR) 검색 시스템(112-B)에서, OCR 데이터베이스(114-B)는 도 8에 관련하여 더 상세히 설명되는 특정 폰트들 또는 텍스트 패턴들을 인식하기 위해 액세스될 수 있다.
- [0022] 임의의 수의 병렬 검색 시스템들(112)이 사용될 수 있다. 임의의 실시예들은 얼굴 인식 검색 시스템(112-A), OCR 검색 시스템(112-B), 이미지-대-용어들 검색 시스템(112-C)(오브젝트 또는 오브젝트 카테고리를 인식할 수 있음), 물건 인식 검색 시스템(책 커버들 및 CD들과 같은 2-D 이미지들을 인식하도록 구성될 수 있고 또한 가구와 같은 3-D 이미지들을 인식하도록 구성될 수 있음), 바 코드 인식 검색 시스템(1D 및 2D 스타일 바 코드들을 인식함), 명명 엔티티 인식 검색 시스템, 랜드마크 인식(에펠 타워와 같은 특정 유명한 랜드마크들을 인식하도록 구성될 수 있고, 또한 빌보드들과 같은 특유한 이미지들의 코퍼스(corpus)를 인식하도록 구성될 수 있음), 클라이언트 시스템(102) 또는 모바일 전화 네트워크의 GPS 수신기에 의해 제공되는 지리-위치 정보에 의해 도움을 받는 장소 인식, 색 인식 검색 시스템 및 유사 이미지 검색 시스템(시각 질의에 유사한 이미지들을 검색하고 식별함)을 포함한다. 추가의 검색 시스템들은 시스템(112-N)에 의해 도 1에 제시된, 부가적인 병렬 검색 시스템들로서 부가될 수 있다. OCR 검색 시스템을 제외한 모든 검색 시스템들은 이미지-매칭 프로세스를 수행하는 검색 시스템들로서 본 명세서에서 집합적으로 정의된다. OCR 검색 시스템을 포함하는 모든 검색 시스템들은 이

미지에 의한 질의 검색 시스템들로서 집합적으로 지칭된다. 임의의 실시예들에서, 시각 질의 서버 시스템(106)은 얼굴 인식 검색 시스템(112-A), OCR 검색 시스템(112-B), 및 적어도 하나의 다른 이미지에 의한 질의 검색 시스템(112)을 포함한다.

[0023] 병렬 검색 시스템들(112) 각각은 개별적으로 시각 검색 질의를 프로세싱하고 자신들의 결과들을 프론트 엔드 서버 시스템(110)에 리턴한다. 임의의 실시예들에서, 프론트 엔드 서버(100)는, 결과들을 복합 문서로 통합하는 것, 디스플레이할 결과들의 서브셋을 선택하는 것, 및 도 6에 관하여 더 상세히 설명될 바와 같이 결과들을 순위화하는 것 중 하나 이상과 같이 검색 결과들에서 하나 이상의 분석들을 수행할 수 있다. 프론트 엔드 서버(110)는 클라이언트 시스템(102)에 검색 결과들을 통신한다.

[0024] 클라이언트 시스템(102)은 사용자에 하나 이상의 검색 결과들을 제시한다. 결과들은 디스플레이상에, 오디오 스피커에 의해 또는 사용자에 정보를 통신하기 위해 사용되는 임의의 다른 수단에 의해 제시될 수 있다. 사용자는 다양한 방식으로 검색 결과들과 상호작용할 수 있다. 임의의 실시예들에서, 사용자의 선택들, 주석들, 및 검색 결과들과의 다른 상호작용들은 시각 질의 서버 시스템(106)에 전송되고 질의 및 주석 데이터베이스(116)에 시각 질의와 함께 기록된다. 질의 및 주석 데이터베이스 내의 정보는 시각 질의 결과들을 향상시키기 위해 사용될 수 있다. 임의의 실시예들에서, 질의 및 주석 데이터베이스(116)로부터의 정보는 병렬 검색 시스템들(112)에 주기적으로 푸싱되고, 정보의 임의의 관련 부분들을 병렬 검색 시스템들(112)의 각각의 개별 데이터베이스들(114)로 통합한다.

[0025] 컴퓨터 네트워크(100)는 선택적으로 용어 질의들에 응답하여 검색들을 수행하기 위한 용어 질의 서버 시스템(118)을 포함한다. 용어 질의는 이미지를 포함하는 시각 질의에 반대되는, 하나 이상의 용어들을 포함하는 질의이다. 용어 질의 서버 시스템(118)은 시각 질의 서버 시스템(106)에서 다양한 검색 엔진들에 의해 산출되는 정보를 보충하는 검색 결과들을 발생시키기 위해 사용될 수 있다. 용어 질의 서버 시스템(118)으로부터 리턴되는 결과들은 임의의 포맷을 포함할 수 있다. 용어 질의 서버 시스템(118)은 텍스트 문서들, 이미지들, 비디오 등을 포함할 수 있다. 용어 질의 서버 시스템(118)이 도 1에 분리된 시스템으로 도시되었지만, 선택적으로 시각 질의 서버 시스템(106)은 용어 질의 서버 시스템(118)을 포함할 수 있다.

[0026] 시각 질의 서버 시스템(106)의 동작에 관한 부가적인 정보는 도 2-4에서 흐름도들에 관하여 아래 제공된다.

[0027] 도 2는 본 발명의 특정 실시예들에 따라, 시각 질의에 응답하기 위한 시각 질의 서버 시스템 방법을 도시하는 흐름도이다. 도 2에 도시된 동작들의 각각은 컴퓨터 메모리 또는 컴퓨터 판독가능한 저장 매체에 저장되는 명령들에 대응할 수 있다.

[0028] 시각 질의 서버 시스템은 클라이언트 시스템으로부터 시각 질의를 수신한다(202). 예를 들어, 클라이언트 시스템은 데스크톱 컴퓨터 디바이스, 모바일 디바이스 또는 도 1을 참조하여 설명되는 다른 유사한 디바이스일 수 있다(204). 예시적인 클라이언트 시스템에서 예시적인 시각 질의는 도 11에 도시된다.

[0029] 시각 질의는 임의의 적합한 포맷의 이미지 문서이다. 예를 들어, 시각 질의는 사진, 스크린 샷, 스캔된 이미지 또는 프레임 또는 비디오의 다수의 프레임들의 시퀀스일 수 있다(206). 임의의 실시예들에서, 시각 질의는 콘텐츠 저작 프로그램(736, 도 5)에 의해 산출되는 그림이다. 따라서, 임의의 실시예들에서, 사용자는 시각 질의를 "그리는" 반면, 다른 실시예들에서, 사용자는 시각 질의를 스캔하거나 또는 사진 찍는다. 임의의 시각 질의들은 아크로바트(Acrobat), 사진 편집 프로그램, 그림 프로그램 또는 이미지 편집 프로그램과 같은 이미지 발생 애플리케이션을 사용하여 생성된다. 예를 들어, 시각 질의는, 사용자가 자신의 휴대 전화에서 자신의 친구의 사진을 찍고 그 다음에 서버 시스템에 시각 질의로서 사진을 제출하는 것으로부터 올 수 있다. 시각 질의는 또한, 사용자가 잡지의 페이지를 스캔하거나 또는 데스크톱 컴퓨터에서 웹페이지의 스크린 샷을 취하고 그 다음에 서버 시스템에 시각 질의로서 스캔 또는 스크린 샷을 제출하는 것으로부터 올 수 있다. 임의의 실시예들에서, 시각 질의는 브라우저 애플리케이션의 검색 엔진 확장을 통해, 브라우저 애플리케이션에 대한 플러그-인을 통해 또는 클라이언트 시스템(102)에 의해 실행되는 검색 애플리케이션에 의해 서버 시스템(106)에 제출된다. 시각 질의들은 또한 클라이언트 시스템에 의해 원격으로 위치된 서버에 전송될 수 있는 이미지들을 지원 또는 발생시키는 다른 애플리케이션 프로그램들(클라이언트 시스템에 의해 실행됨)에 의해 제출될 수 있다.

[0030] 시각 질의는 텍스트 및 비-텍스트 엘리먼트들의 조합일 수 있다(208). 예를 들어, 질의는 도로 표지판 옆에 서 있는 사람과 같은, 이미지들 및 텍스트를 포함하는 잡지 페이지의 스캔일 수 있다. 시각 질의는 클라이언트 시스템에 임베딩된 카메라에 의해 촬영되는, 사람의 얼굴의 이미지 또는 클라이언트 시스템에 의해 스캔되거나 그렇지 않으면 수신되는 문서를 포함할 수 있다. 시각 질의는 또한 텍스트만을 포함하는 문서의 스캔일 수 있다.

시각 질의는 또한 숲 속의 여러 새들, 사람 및 오브젝트(예를 들어, 차, 공원 벤치 등), 사람 및 동물(예를 들어, 애완동물, 농장 동물, 나비 등)과 같은 많은 별개의 서브젝트들의 이미지일 수 있다. 시각 질의들은 둘 이상의 별개의 엘리먼트들을 가질 수 있다. 예를 들어, 시각 질의는 물건의 바코드 및 이미지 또는 물건 포장의 물건 명칭을 포함할 수 있다. 예를 들어, 시각 질의는 책의 타이틀, 커버 아트 및 바코드를 포함하는 책 커버의 픽처일 수 있다. 임의의 예시들에서, 하나의 시각 질의는 아래에 더 상세히 논의되는 바와 같이, 시각 질의의 상이한 부분들에 대응하는 둘 이상의 별개의 검색 결과들을 산출할 것이다.

[0031] 서버 시스템은 아래와 같은 시각 질의를 프로세싱한다. 프론트 엔드 서버 시스템은 동시 프로세싱을 위해 복수의 병렬 검색 시스템들에 시각 질의를 송신한다(210). 각 검색 시스템은 별개의 시각 질의 검색 프로세스를 구현하고, 즉, 개별 검색 시스템은 자신의 고유의 프로세싱 방식에 의해 시각 질의를 프로세싱한다.

[0032] 임의의 실시예들에서, 프로세싱을 위해 시각 질의가 송신되는 검색 시스템들 중 하나는 광학식 문자 인식(OCR) 검색 시스템이다. 임의의 실시예들에서, 프로세싱을 위해 시각 질의가 송신되는 검색 시스템들 중 하나는 얼굴 인식 검색 시스템이다. 임의의 실시예들에서, 별개의 시각 질의 검색 프로세스들을 실행하는 복수의 검색 시스템들은 적어도, 광학식 문자 인식(OCR), 얼굴 인식, 및 OCR 및 얼굴 인식 외에 다른 이미지에 의한 질의 프로세스를 포함한다(212). 다른 이미지에 의한 질의 프로세스는 물건 인식, 바 코드 인식, 오브젝트-또는-오브젝트-카테고리 인식, 명명 엔티티 인식 및 색 인식을 포함하는(이에 제한되지 않음) 프로세스들의 세트로부터 선택된다(212).

[0033] 임의의 실시예들에서, 명명 엔티티 인식은 OCR 검색 시스템의 포스트 프로세스로서 발생하고, OCR의 텍스트 결과는 유명한 사람들, 위치들, 오브젝트들 등에 대하여 분석되고, 그 다음에 명명 엔티티들로서 식별되는 용어들은 용어 질의 서버 시스템(118, 도 1)에서 검색된다. 다른 실시예들에서, 유명한 랜드마크들, 로고들, 사람들, 앨범 커버들, 상표들 등의 이미지들은 이미지-대-용어들 검색 시스템에 의해 인식된다. 다른 실시예들에서, 이미지-대-용어들 검색 시스템으로부터 분리된 별개의 명명 엔티티 이미지에 의한 질의 프로세스가 이용된다. 오브젝트-또는-오브젝트 카테고리 인식 시스템은 "차"와 같은 포괄적인 결과 타입들을 인식한다. 임의의 실시예들에서, 이 시스템은 물건 브랜드들, 특정 물건 물건 모델들 등을 인식하고, "포르쉐"와 같은 더 상세한 설명들을 제공한다. 검색 시스템들 중 일부는 특별 사용자 특유 검색 시스템들일 수 있다. 예를 들어, 색 인식 및 얼굴 인식의 특정 버전들은 블라인드에 의해 사용되는 특별 검색 시스템들일 수 있다.

[0034] 프론트 엔드 서버 시스템은 병렬 검색 시스템들로부터 결과들을 수신한다(214). 임의의 실시예들에서, 결과들은 검색 스코어를 동반한다. 임의의 시각 질의들에 대하여, 검색 시스템들 중 일부는 관련 결과들을 발견하지 않을 것이다. 예를 들어, 시각 질의가 꽃의 픽처인 경우, 얼굴 인식 검색 시스템 및 바 코드 검색 시스템은 임의의 관련 결과들을 발견하지 않을 것이다. 임의의 실시예들에서, 관련 결과들이 발견되지 않은 경우, 널 또는 제로 검색 스코어가 상기 검색 시스템으로부터 수신된다(216). 임의의 실시예들에서, 프론트 엔드 서버가 미리-정의된 시간의 기간(예를 들어, 0.2, 0.5, 1, 2 또는 5 초들) 후에 검색 시스템으로부터 결과를 수신하지 않은 경우, 타임 아웃된 서버가 널 검색 스코어를 산출한 것처럼 수신된 결과들을 프로세싱할 것이고 다른 검색 시스템들로부터 수신된 결과들을 프로세싱할 것이다.

[0035] 선택적으로, 수신된 검색 결과들 중 적어도 둘이 미리-정의된 기준을 만족시킬 때, 이들은 순위화된다(218). 임의의 실시예들에서, 미리 정의된 기준 중 하나는 무효 결과들을 배제한다. 미리-정의된 기준은 결과들이 무효이지 않다는 것이다. 임의의 실시예들에서, 미리 정의된 기준 중 하나는 미리-정의된 최소 스코어 미만에 있는 숫자 스코어를 가지는 결과들(예를 들어, 관련 팩터들에 대하여)을 배제한다. 선택적으로 복수의 검색 결과들이 필터링된다(220). 임의의 실시예들에서, 결과들은 결과들의 총 수가 미리-정의된 임계치를 초과하는 경우에만 필터링된다. 임의의 실시예들에서, 모든 결과들이 순위화되지만 미리-정의된 최소 스코어 미만에 있는 결과들은 배제된다. 임의의 시각 질의들에 대하여, 결과들의 콘텐츠가 필터링된다. 예를 들어, 결과들의 일부가 사설 정보 또는 개인 보호 정보를 포함하는 경우, 이러한 결과들은 필터링되어 제외된다.

[0036] 선택적으로, 시각 질의 서버 시스템은 복합 검색 결과를 생성한다(222). 이의 일 실시예는 도 3에 관하여 설명된 바와 같이, 상호작용 결과 문서에 둘 이상의 검색 시스템 결과가 임베딩될 때이다. 용어 질의 서버 시스템(118, 도 1)은 용어 검색으로부터의 결과들로 병렬 검색 시스템들 중 하나로부터의 결과들을 보장할 수 있고, 부가적인 결과들은 문서들 또는 정보 소스들에 대한 링크들 또는 텍스트 및/또는 시각 질의에 관련될 수 있는 부가적인 정보를 포함하는 이미지들이다. 그러므로, 예를 들어, 복합 검색 결과는 OCR 결과 및 OCR 문서에서 명명 엔티티에 대한 링크를 포함할 수 있다(224).

[0037] 임의의 실시예들에서, OCR 검색 시스템(112-B, 도 1) 또는 프론트 엔드 시각 질의 프로세싱 서버(110, 도 1)는

텍스트에서 비슷한 관련 단어들을 인식한다. 예를 들어, 유명한 사람들 또는 장소들과 같은 명명 엔티티들을 인식할 수 있다. 명명 엔티티들은 용어 질의 서버 시스템(118, 도 1)에 질의 용어들로서 제출된다. 임의의 실시예들에서, 용어 질의 서버 시스템에 의해 산출되는 용어 질의 결과들은 "링크"로서 시각 질의 결과에 임베딩된다. 임의의 실시예들에서, 용어 질의 결과들은 분리된 링크들로서 리턴된다. 예를 들어, 책 커버의 픽처가 시각 질의인 경우, 오브젝트 인식 검색 시스템이 책에 대한 고 스코어 히트를 산출할 가능성이 높다. 따라서, 책의 타이틀에 대한 용어 질의는 용어 질의 서버 시스템(118)에서 실행될 수 있고 용어 질의 결과들은 시각 질의 결과들과 함께 리턴된다. 임의의 실시예들에서, 용어 질의 결과들은 시각 질의 결과들로부터 자신들을 구별하기 위해 라벨링된 그룹으로 제시된다. 결과들은 개별적으로 검색될 수 있거나, 검색은 특히 관련된 부가적인 검색 결과들을 산출하기 위해 검색 질의에서 모든 인식된 명명 엔티티들을 사용하여 수행될 수 있다. 예를 들어, 시각 질의가 파리(Paris)에 관한 스캐닝된 여행 브로슈어인 경우, 리턴된 결과는 용어 질의 "노트 담"에서 검색을 개시하기 위해 용어 질의 서버 시스템(118)에 대한 링크들을 포함할 수 있다. 유사하게, 복합 검색 결과들은 인식된 유명한 이미지들에 대한 텍스트 검색들로부터의 결과들을 포함한다. 예를 들어, 동일한 여행 브로슈어에서, "에펠 탑" 및 "루브르"와 같은 브로슈어에서 픽처들로 도시된 유명한 목적지들에 대한 용어 질의 결과들에 대한 라이브 링크들이 또한 도시될 수 있다(용어들 "에펠 탑" 및 "루브르"는 브로슈어 자신에서 나타나지 않을 수 있는 경우에도).

[0038] 시각 질의 서버 시스템은 그 다음에 클라이언트 시스템에 적어도 하나의 결과를 송신한다(226). 전형적으로, 시각 질의 프로세싱 서버가 복수의 검색 시스템들 중 적어도 일부로부터 복수의 검색 결과들을 수신하는 경우, 시각 질의 프로세싱 서버는 클라이언트 시스템에 복수의 검색 결과들 중 적어도 하나를 송신할 것이다. 임의의 시각 질의들에 대하여, 오직 하나의 검색 시스템이 관련 결과들을 리턴할 것이다. 예를 들어, 텍스트의 이미지만을 포함하는 시각 질의에서, 오직 OCR 서버의 결과들이 관련될 수 있다. 임의의 시각 질의들에 대하여, 하나의 검색 시스템으로부터의 오직 하나의 결과가 관련될 수 있다. 예를 들어, 스캐닝된 바 코드에 관련된 물건만이 관련될 수 있다. 이러한 예시들에서, 프론트 엔드 시각 프로세싱 서버는 관련 검색 결과(들)만을 리턴할 것이다. 임의의 시각 질의들에 대하여, 복수의 검색 결과들은 클라이언트 시스템에 송신되고, 복수의 검색 결과들은 병렬 검색 시스템들 중 둘 이상으로부터의 검색 결과들을 포함한다(228). 이는 둘 이상의 별개 이미지가 시각 질의에 있을 때 발생할 수 있다. 예를 들어, 시각 질의가 말을 타는 사람의 픽처인 경우, 사람의 얼굴 인식의 결과들은 말에 대한 오브젝트 인식 결과들과 함께 디스플레이될 수 있다. 임의의 실시예들에서, 이미지 검색 시스템에 의한 특정 질의에 대한 모든 결과들은 함께 그룹화되고 제시된다. 예를 들어, 상위 N개의 얼굴 인식 결과들은 제목 "얼굴 인식 결과들" 하에서 디스플레이되고 상위 N개의 오브젝트 인식 결과들은 제목 "오브젝트 인식 결과들" 하에서 함께 디스플레이된다. 대안적으로, 아래에 논의되는 바와 같이, 특정 이미지 검색 시스템으로부터의 검색 결과들은 이미지 영역에 의해 그룹화될 수 있다. 예를 들어, 시각 질의가 두 개의 얼굴들을 포함하는 경우, 두 얼굴 모두는 얼굴 인식 결과들을 산출하고, 각 얼굴에 대한 결과들은 별개의 그룹으로 제시될 수 있다. 임의의 시각 질의들(예들들어, 텍스트 및 하나 이상의 오브젝트들 둘다의 이미지를 포함하는 시각 질의)에 대하여, 검색 결과들은 OCR 결과들 및 하나 이상의 이미지-매칭 결과들 둘 다를 포함할 수 있다(230).

[0039] 임의의 실시예들에서, 사용자는 특정 검색 결과에 관해 더 알게 되기를 원할 수 있다. 예를 들어, 시각 질의가 돌고래의 픽처이고 "이미지 대 용어들" 검색 시스템이 아래의 용어들 "물", "돌고래", "파란색" 및 "물갈퀴"를 리턴하는 경우, 사용자는 "물갈퀴"에서 텍스트 기반 질의 용어 검색을 실행하기를 원할 수 있다. 사용자가 용어 질의(예를 들어, 검색 결과들에서 대응하는 링크를 클릭하거나 그렇지 않으면 선택하는 사용자에게 의해 나타내는)에서 검색을 실행하기를 원할 때, 질의 용어 서버 시스템(118, 도 1)이 액세스되고, 선택된 용어(들) 상에서 검색이 실행된다. 대응하는 검색 용어 결과들이 시각 질의 결과들과 함께 또는 분리하여 클라이언트 시스템상에 디스플레이된다(232). 임의의 실시예들에서, 프론트 엔드 시각 질의 프로세싱 서버(110, 도 1)는 자동으로(즉, 초기 시각 질의 외에 어떤 사용자 커맨드도 수신하지 않고) 시각 질의에 대한 하나 이상의 상위 잠재 결과들을 선택하고, 용어 질의 서버 시스템(118)상에 이 텍스트 결과들을 실행하고 그 다음에 클라이언트 시스템에 적어도 하나의 검색 결과를 송신하는 일부로서 클라이언트 시스템에 시각 질의 결과와 함께 이 용어 질의 결과들을 리턴한다(232). 상기 예시에서, "물갈퀴"가 돌고래의 시각 질의 픽처에 대한 제 1 용어 결과인 경우, 프론트 엔드 서버는 "물갈퀴"에서 용어 질의를 실행하고 클라이언트 시스템에 시각 질의 결과들과 함께 이러한 용어 질의 결과들을 리턴한다. 사용자에게 의해 선택된 것으로 고려되는 용어 결과가 시각 질의로부터 사용자에게 검색 결과들을 송신하기 전에 자동으로 실행되는 이 실시예는 사용자 시간을 절약한다. 임의의 실시예들에서, 이러한 결과들은 상기 설명된 바와 같이 복합 검색 결과로서 디스플레이된다(222). 다른 실시예들에서, 결과들은 복합 검색 결과 대신에 또는 복합 검색 결과 외에 검색 결과 리스트의 부분이다.

- [0040] 도 3은 상호작용 결과 문서를 이용하여 시각 질의에 응답하기 위한 프로세스를 도시하는 흐름도이다. 제 1 세 개의 동작들(202, 210, 214)은 도 2를 참조하여 상기 설명된다. 병렬 검색 시스템들(214)로부터 수신되는 검색 결과들로부터, 상호작용 결과 문서가 생성된다(302).
- [0041] 상호작용 결과 문서의 생성(302)은 이제 상세히 설명될 것이다. 임의의 시각 질의들에 대하여, 상호작용 결과 문서는 시각 질의의 각각의 서브-부분들의 하나 이상의 시각 식별자들을 포함한다. 각 시각 식별자는 검색 결과들 중 적어도 하나에 대한 적어도 하나의 사용자 선택가능한 링크를 가진다. 시각 식별자는 시각 질의의 각각의 서브-부분을 식별한다. 임의의 시각 질의들에 대하여, 상호작용 결과 문서는 하나 이상의 결과들에 대한 하나의 사용자 선택가능한 링크를 가진 오직 하나의 시각 식별자를 가진다. 임의의 실시예들에서, 검색 결과들 중 하나 이상에 대한 각각의 사용자 선택가능한 링크는 활성 영역을 가지고, 활성 영역은 대응하는 시각 식별자와 연관된 시각 질의의 서브-부분에 대응한다.
- [0042] 임의의 실시예들에서, 시각 식별자는 바운딩 박스(304)이다. 임의의 실시예들에서, 바운딩 박스는 도 12a에 도시된 바와 같이 시각 질의의 서브-부분을 둘러싼다. 바운딩 박스는 정사각형 또는 직사각형 박스 형상일 필요가 없고, (예를 들어, 시각 질의의 오브젝트, 시각 질의의 엔티티 또는 시각 질의의 영역에 대해), 원형, 타원형, 등각, 불규칙 또는 도 12에 도시된 바와 같이 임의의 다른 형상일 수 있다. 임의의 시각 질의들에 대하여, 바운딩 박스는 시각 질의의 서브-부분에서 식별가능한 엔티티의 경계의 윤곽을 나타낸다(306). 임의의 실시예들에서, 각 바운딩 박스는 하나 이상의 검색 결과들에 대한 사용자 선택가능한 링크를 포함하고, 사용자 선택가능한 링크는 바운딩 박스에 의해 서라운드되는 시각 질의의 서브-부분에 대응하는 활성 영역을 가진다. 바운딩 박스 내의 공간이(사용자 선택가능한 링크의 활성 영역) 사용자에게 의해 선택될 때, 윤곽이 나타내어진 서브-부분에서 이미지에 대응하는 검색 결과들이 리턴된다.
- [0043] 임의의 실시예들에서, 시각 식별자는 도 14에 도시된 바와 같은 라벨이다(307). 임의의 실시예들에서, 라벨은 시각 질의의 각각의 서브-부분에서 이미지와 연관되는 적어도 하나의 용어를 포함한다. 각 라벨은 각각의 서브-부분에 또는 그 주위에 상호작용 결과 문서로 표시하기 위해 포맷된다. 임의의 실시예들에서 라벨들은 색 코딩된다.
- [0044] 임의의 실시예들에서, 각 개별적인 시각 식별자들은 시각 질의의 각각의 서브-부분에서 인식된 엔티티의 타입에 따라 시각적으로 구별되는 방식으로 표시하기 위해 포맷된다. 예를 들어, 도 13에 도시된 바와 같이, 물건, 사람, 상표 및 두 개의 텍스트 영역들 주위의 바운딩 박스들은 각각 상이하게 색칠된 투명 바운딩 박스들을 표시하는, 별개의 크로스-해칭 패턴들로 표시된다. 임의의 실시예들에서, 시각 식별자들은 오버레이 색, 오버레이 패턴, 라벨 배경 색, 라벨 배경 패턴, 라벨 폰트 색 및 경계 색과 같은 시각적으로 구별되는 방식으로 표시하기 위해 포맷된다.
- [0045] 임의의 실시예들에서, 상호작용 결과 문서의 사용자 선택가능한 링크는 시각 질의의 대응하는 서브-부분에 관련된 하나 이상의 결과들을 포함하는 문서 또는 오브젝트에 대한 링크이다(308). 임의의 실시예들에서, 적어도 하나의 검색 결과는 시각 질의의 대응하는 서브-부분에 관련된 데이터를 포함한다. 따라서, 사용자가 각각의 서브-부분과 연관된 선택가능한 링크를 선택할 때, 사용자는 시각 질의의 각각의 서브-부분에서 인식된 엔티티에 대응하는 검색 결과들로 안내된다.
- [0046] 예를 들어, 시각 질의가 바 코드의 사진이었던 경우, 바 코드가 부착된 포장의 관련된 부분들인 사진의 부분들이 존재할 수 있다. 상호작용 결과 문서는 오직 바 코드 주위만의 바운딩 박스를 포함할 수 있다. 윤곽이 나타내어진 바 코드 바운딩 박스 내에서 사용자가 선택할 때, 바 코드 검색 결과가 디스플레이된다. 바 코드 검색 결과는 하나의 결과, 상기 바 코드에 대응하는 물건의 명칭을 포함할 수 있거나, 또는 바 코드 결과들은 물건이 구입, 리뷰 등이 될 수 있는 다양한 장소들과 같은 여러 결과들을 포함할 수 있다.
- [0047] 임의의 실시예들에서, 개별적인 시각 식별자에 대응하는 시각 질의의 서브-부분이 하나 이상의 용어들을 포함하는 텍스트를 포함할 때, 개별적인 시각 식별자에 대응하는 검색 결과들은 텍스트의 용어들 중 적어도 하나에 대한 용어 질의 검색으로부터의 결과들을 포함한다. 임의의 실시예들에서, 개별적인 시각 식별자에 대응하는 시각 질의의 서브-부분이 미리 정의된 신뢰성(또는 다른) 기준을 만족시키는, 적어도 하나의 매칭(즉, 검색 결과)이 발견되는 사람의 얼굴을 포함할 때, 개별적인 시각 식별자에 대응하는 검색 결과들은 명칭, 핸들, 컨택트 정보, 어카운트 정보, 주소 정보, 선택가능한 서브-부분에 포함된 얼굴을 갖는 사람과 연관된 관련 모바일 디바이스의 현재 위치, 선택가능한 서브-부분에 포함된 얼굴을 갖는 사람의 다른 이미지들 및 그 사람의 얼굴에 대한 잠재 이미지 매칭들 중 하나 이상을 포함한다. 임의의 실시예들에서, 개별적인 시각 식별자에 대응하는 시각 질의의 서브-부분이 미리 정의된 신뢰성(또는 다른) 기준을 만족하는, 적어도 하나의 매칭(즉, 검색 결과)이

발견되는 물건을 포함할 때, 개별적인 시각 식별자에 대응하는 검색 결과들은 물건 정보, 물건 리뷰, 물건의 구입을 개시하기 위한 옵션, 물건에 입찰을 개시하기 위한 옵션, 유사한 물건들의 리스트 및 관련된 물건들의 리스트 중 하나 이상을 포함한다.

- [0048] 선택적으로, 상호 작용 결과 문서에서 개별적인 사용자 선택가능한 링크는 앵커(anchor) 텍스트를 포함하고, 이는 링크를 활성화함이 없이 문서에 디스플레이된다. 앵커 텍스트는 링크가 활성화될 때 획득되는 정보에 관련된 키워드 또는 용어와 같은 정보를 제공한다. 앵커 텍스트는 라벨의 일부로서 디스플레이되거나(307) 또는 바운딩 박스의 부분에 디스플레이되거나(304) 또는 사용자가 1 초와 같은 미리-결정된 시간의 기간에 대한 사용자 선택가능한 링크 위에 커서를 맵들게 할 때 디스플레이되는 부가적인 정보로서 디스플레이될 수 있다.
- [0049] 선택적으로, 상호작용 결과 문서에서 개별적인 사용자 선택가능한 링크는 텍스트-기반 질의(때때로 용어 질의로 본 명세서에서 지칭됨)에 대응하는 문서들 또는 정보를 검색하기 위한 검색 엔진에 대한 링크이다. 링크의 활성화는 검색 엔진에 의한 검색의 실행을 야기하고, 질의 및 검색 엔진은 클라이언트 시스템에 리턴되는 결과들을 가진, 링크에 의해 특정된다(예를 들어, 검색 엔진은 링크에서 URL에 의해 특정되고 텍스트-기반 검색 질의는 링크의 URL 파라미터에 의해 특정됨). 선택적으로, 이 실시예에서 링크는 검색 질의에서 용어들 또는 텍스트를 명시하는 앵커 텍스트를 포함할 수 있다.
- [0050] 임의의 실시예들에서, 시각 질의에 응답하여 산출되는 상호작용 결과 문서는 동일한 검색 시스템으로부터의 결과들에 대응하는 복수의 링크들을 포함할 수 있다. 예를 들어, 시각 질의는 사람들의 그룹의 픽처 또는 이미지일 수 있다. 상호작용 결과 문서는 각 사람 주위에 바운딩 박스들을 포함할 수 있고, 활성화될 때 그룹에서 각 얼굴에 대하여 얼굴 인식 검색 시스템으로부터 결과들을 리턴한다. 임의의 시각 질의들에 대하여, 상호작용 결과 문서에서 복수의 링크들은 둘 이상의 검색 시스템으로부터의 검색 결과들에 대응한다(310). 예를 들어, 사람 및 개의 픽처가 시각 질의로서 제출되었던 경우, 상호작용 결과 문서에서 바운딩 박스들은 사람 및 개를 분리하여 윤곽을 나타낼 수 있다. (상호작용 결과 문서에서) 개인이 선택될 때, 얼굴 인식 검색 시스템으로부터 검색 결과들이 리턴되고, (상호작용 결과 문서에서) 개가 선택될 때, 이미지-대-용어들 검색 시스템으로부터 결과들이 리턴된다. 임의의 시각 질의들에 대하여, 상호작용 결과 문서는 OCR 결과 및 이미지 매칭 결과를 포함한다(312). 예를 들어, 표지판 옆에 서있는 사람의 픽처가 시각 질의로서 제출되는 경우, 상호작용 결과 문서는 표지판의 텍스트에 대하여 그리고 사람에 대하여 시각 식별자들을 포함할 수 있다. 유사하게, 잡지의 스캔이 시각 질의로서 사용되는 경우, 상호작용 결과 문서는 또한 상기 페이지에서 기사의 텍스트에 대한 시각 식별자뿐만 아니라 페이지상에서 광고들에서 사진들 또는 상표들에 대한 시각 식별자들을 포함할 수 있다.
- [0051] 상호작용 결과 문서가 생성된 후, 클라이언트 시스템에 송신된다(314). 임의의 실시예들에서, 상호작용 결과 문서(예를 들어, 문서(1200), 도 15)는 도 2를 참조하여 상기 설명된 바와 같이, 하나 이상의 병렬 검색 시스템들로부터 검색 결과들의 리스트와 함께 송신된다. 임의의 실시예들에서, 상호작용 결과 문서는 도 15에 도시된 바와 같이 하나 이상의 병렬 검색 시스템들로부터 검색 결과들의 리스트 위에 또는 그렇지 않으면 상기 리스트에 인접하여 클라이언트 시스템에서 디스플레이된다(315).
- [0052] 선택적으로, 사용자는 결과 문서에서 시각 식별자를 선택함으로써 결과 문서와 상호작용할 것이다. 서버 시스템은 상호 결과 문서에서 시각 식별자의 사용자 선택에 관한 정보를 클라이언트 시스템으로부터 수신한다(316). 상기 논의된 바와 같이, 임의의 실시예들에서, 링크는 바운딩 박스 내에 활성화 영역을 선택함으로써 활성화된다. 다른 실시예들에서, 링크는 바운딩 박스가 아닌, 시각 질의의 서버-부분의 시각 식별자의 사용자 선택에 의해 활성화된다. 임의의 실시예들에서, 링크된 시각 식별자가 핫 버튼, 서버-부분 근처에 위치한 라벨, 텍스트에서 밑줄쳐진 단어 또는 시각 질의에서 오브젝트 또는 서브젝트의 다른 표시이다.
- [0053] 검색 결과 리스트가 상호작용 결과 문서로 표시되는(315) 실시예들에서, 사용자가 사용자 선택가능한 링크를 선택할 때(316), 선택된 링크에 대응하는 검색 결과 리스트에서 검색 결과가 식별된다. 임의의 실시예들에서, 커서는 선택된 링크에 대응하는 제 1 결과를 점프하거나 또는 자동으로 제 1 결과로 이동할 것이다. 클라이언트(102)의 디스플레이가 상호작용 결과 문서 및 전체 검색 결과 리스트 둘 다를 디스플레이하기에는 너무 작은 임의의 실시예들에서, 상호작용 결과 문서에서 링크를 선택하는 것은 선택된 링크에 대응하는 적어도 하나의 제 1 결과를 디스플레이하도록 검색 결과 리스트를 스크롤 또는 점프하게 한다. 임의의 다른 실시예들에서, 상호작용 결과 문서에서 링크의 사용자 선택에 응답하여 결과 리스트는 결과 리스트의 상부에서 링크에 대응하는 제 1 결과가 디스플레이되도록 다시 순서화된다.
- [0054] 임의의 실시예들에서, 사용자가 사용자 선택가능한 링크를 선택할 때(316), 시각 질의 서버 시스템은 사용자에게 디스플레이하기 위해 클라이언트에 시각 질의의 대응하는 서버-부분에 관련된 결과들의 적어도 서브세트를 송신

한다(318). 임의의 실시예들에서, 사용자는 동시에 다수의 시각 식별자들을 선택할 수 있고 선택된 시각 식별자들 모두에 대한 결과들의 서브세트를 동시에 수신할 것이다. 다른 실시예들에서, 사용자 선택가능한 링크들에 대응하는 검색 결과들은, 상호작용 결과 문서의 하나 이상의 링크들의 사용자 선택에 응답하여 순간적으로 가상으로 사용자에게 검색 결과들을 제공하도록 사용자 선택가능한 링크들 중 일부의 사용자 선택 전에 클라이언트에 전로딩된다.

[0055] 도 4는 클라이언트 및 시각 질의 서버 시스템 사이에서의 통신들을 도시하는 흐름도이다. 클라이언트(102)는 사용자/질의자로부터 시각 질의를 수신한다(402). 임의의 실시예들에서, 시각 질의들은 시각 질의 시스템에 사인 업 또는 "옵트 인(opt in)"되는 사용자들로부터만 수용될 수 있다. 임의의 실시예들에서, 얼굴 인식 매칭들에 대한 검색들은 얼굴 인식 시각 질의 시스템에 대하여 사인 업한 사용자들에 대해서만 수행되는 반면, 시각 질의들의 다른 타입들은 사용자들이 얼굴 인식 부분에 "옵트 인"했는지에 관계없이 누구에 대해서도 수행된다.

[0056] 상기 설명된 바와 같이 시각 질의의 포맷은 많은 형태들을 취할 수 있다. 시각 질의는 시각 질의 문서의 서브-부분들에 위치된 하나 이상의 서브젝트들을 포함할 가능성이 있을 것이다. 임의의 시각 질의들에 대하여, 클라이언트 시스템(102)은 시각 질의에 타입 인식 선-프로세싱을 수행한다(404). 임의의 실시예들에서, 클라이언트 시스템(102)은 이 선-프로세싱 시스템에서 특정 인식가능한 패턴들을 검색한다. 예를 들어, 임의의 시각 질의들에 대하여 클라이언트는 색들을 인식할 수 있다. 임의의 시각 질의들에 대하여, 클라이언트는 특정 서브-부분이 텍스트를 포함할 가능성이 있음을 인식할 수 있다(상기 영역이 빛 공간 등에 의해 서라운드된 작은 어두운 문자들로 이루어지기 때문에). 클라이언트는 임의의 수의 선-프로세싱 타입 인식기들 또는 타입 인식 모듈들을 포함할 수 있다. 임의의 실시예들에서, 클라이언트는 바 코드들을 인식하기 위해 타입 인식 모듈을 가질 것이다(바코드 인식(406)). 직사각형 영역에서 독특한 줄무늬 패턴을 인식함으로써 그렇게 할 수 있다. 임의의 실시예들에서, 클라이언트는 시각 질의의 서브-부분 또는 특정 서브젝트가 얼굴을 포함할 가능성이 있음을 인식하기 위해 타입 인식 모듈을 가질 것이다(얼굴 검출(408)).

[0057] 임의의 실시예들에서, 인식된 "타입"은 확인을 위해 사용자에게 리턴된다. 예를 들어, 클라이언트 시스템(102)은 "바 코드가 당신의 시각 질의에서 발견되었다, 당신은 바 코드 질의 결과들을 수신하는데 관심이 있는가?"라고 말하는 메시지를 리턴할 수 있다. 임의의 실시예들에서, 메시지는 타입이 발견되는 시각 질의의 서브-부분도 나타낼 수 있다. 임의의 실시예들에서, 이 표시는 도 3을 참조하여 설명된 상호작용 결과 문서와 유사하다. 예를 들어, 시각 질의의 서브-부분의 윤곽을 나타낼 수 있고, 서브-부분이 얼굴을 포함할 가능성이 있음을 나타내고 사용자에게 자신들이 얼굴 인식 결과들을 수신하는데 관심이 있는지를 물을 수 있다.

[0058] 클라이언트(102)가 시각 질의의 선택적 선-프로세싱을 수행한 후, 클라이언트는 시각 질의 서버 시스템(106)에, 특히 프론트 엔드 시각 질의 프로세싱 서버(110)에 시각 질의를 송신한다. 임의의 실시예들에서, 선-프로세싱이 관련 결과들을 산출한 경우, 즉, 타입 인식 모듈들 중 하나가 질의 또는 질의의 서브-부분이 특정 타입(얼굴, 텍스트, 바코드 등)일 가능성이 있음을 나타내는, 특정 임계치 초과 결과들을 산출하는 경우, 클라이언트는 선-프로세싱의 결과들에 관한 정보를 패스할 것이다. 예를 들어, 클라이언트는 얼굴 인식 모듈이 시각 질의의 특정 서브-부분이 얼굴을 포함한다고 75% 확신한다고 나타낼 수 있다. 더 일반적으로, 선-프로세싱 결과들은, 존재한다면, 하나 이상의 서브젝트 타입 값들(예를 들어, 바 코드, 얼굴, 텍스트 등)을 포함한다. 선택적으로, 시각 질의 서버 시스템에 송신된 선-프로세싱 결과들은, 선-프로세싱 결과들에서 각 서브젝트 타입에 대하여 서브젝트 타입 값에 대응하는 시각 질의의 서브-부분을 식별하는 정보 및 선-프로세싱 결과들에서 각 서브젝트 타입에 대하여, 시각 질의의 대응하는 서브-부분의 식별 및/또는 서브젝트 타입 값에서 확신의 레벨을 나타내는 확신 값 중 하나 이상을 포함한다.

[0059] 프론트 엔드 서버(110)는 클라이언트 시스템으로부터 시각 질의를 수신한다(202). 수신되는 시각 질의는 상기 논의된 선-프로세싱 정보를 포함할 수 있다. 상기 설명된 바와 같이, 프론트 엔드 서버는 복수의 병렬 검색 시스템들에 시각 질의를 송신한다(210). 프론트 엔드 서버(110)가 서브-부분이 특정 타입의 서브젝트를 포함하는 가능성에 관한 선-프로세싱 정보를 수신한 경우, 프론트 엔드 서버는 병렬 검색 시스템들 중 하나 이상에 따라 이 정보를 패스할 수 있다. 예를 들어, 얼굴 인식 검색 시스템(112-A)이 먼저 시각 질의의 그 서브섹션을 프로세싱할 수 있도록 특정 서브-부분이 얼굴일 가능성이 있다는 정보를 패스할 수 있다. 유사하게, (특정 서브-부분이 얼굴일 가능성이 있다는) 동일 정보의 송신은 그 서브-부분을 무시 또는 다른 서브-부분들을 먼저 분석하기 위해 다른 병렬 검색 시스템들에 의해 사용될 수 있다. 임의의 실시예들에서, 프론트 엔드 서버는 병렬 검색 시스템들에 선-프로세싱 정보를 패스하지 않을 것이지만, 병렬 검색 시스템들로부터 수신되는 결과들을 프로

세싱할 방법을 보장하기 위해 이 정보를 대신 사용할 것이다.

- [0060] 도 2를 참조하여 설명된 바와 같이, 얼마 정도의 시각 질의들에 대하여, 프론트 엔드 서버(110)는 병렬 검색 시스템들로부터 복수의 검색 결과들을 수신한다(214). 프론트 엔드 서버는 그 다음에 다양한 순위화 및 필터링을 수행할 수 있고, 도 2 및 3을 참조하여 설명되는 상호작용 검색 결과 문서를 생성할 수 있다. 프론트 엔드 서버(110)가 서브-부분이 특정 타입의 서브젝트를 포함하는 가능성에 관한 선-프로세싱 정보를 수신한 경우, 선-프로세싱된 인식된 서브젝트 타입에 매칭하는 이러한 결과들에 선호도를 부여함으로써 필터링 및 순서화할 수 있다. 사용자가 특정 타입의 결과가 요청되었다고 나타내는 경우, 프론트 엔드 서버는 결과들을 프로세싱할 때 사용자의 요청들을 고려할 것이다. 예를 들어, 프론트 엔드 서버는 사용자가 바 코드 정보를 요청한 경우에만 모든 다른 결과들을 필터링할 수 있거나 또는 프론트 엔드 서버가 다른 결과들을 리스팅하기 전에 요청된 타입에 포함되는 모든 결과들을 리스팅할 것이다. 상호작용 시각 질의 문서가 리턴된 경우, 오직 상호작용 결과 문서에 나타난 다른 서브젝트들에 대한 관련 검색들을 수행하기 위해 링크들을 제공하면서 서버는 사용자가 관심 있다고 나타난 타입의 결과와 연관된 링크들을 선-검색할 수 있다. 그 다음에 프론트 엔드 서버(110)는 클라이언트 시스템에 검색 결과들을 송신한다(226).
- [0061] 클라이언트(102)는 서버 시스템으로부터 결과들을 수신한다(412). 적용가능할 때, 이러한 결과들은 선-프로세싱 스테이지에서 발견된 결과의 타입에 매칭하는 결과들을 포함할 것이다. 예를 들어, 임의의 실시예들에서, 상기 결과들은 하나 이상의 바 코드 결과들을 포함하거나(414) 또는 하나 이상의 얼굴 인식 결과들을 포함할 것이다(416). 클라이언트의 선-프로세싱 모듈들이 특정 타입의 결과가 선호됐음을 나타냈고 그 결과가 발견된 경우, 상기 타입의 발견된 결과들은 두드러지게 리스팅될 것이다.
- [0062] 선택적으로 사용자는 결과들 중 하나 이상을 선택하거나 또는 주석을 달 것이다(418). 사용자는 하나의 검색 결과를 선택할 수 있고, 특정 타입의 검색 결과를 선택할 수 있고 그리고/또는 상호작용 결과 문서의 부분을 선택할 수 있다(420). 결과의 선택은 리턴된 결과가 질의에 관련되었다는 묵시된 피드백이다. 이러한 피드백 정보는 미래의 질의 프로세싱 동작들에 이용될 수 있다. 주석은 또한 미래의 질의 프로세싱 동작들에서 이용될 수 있는 리턴된 결과에 관한 명시적 피드백을 제공한다. 주석들은 분리된 주석(자유 형태 또는 구조된 형태) 또는 리턴된 결과의 부분들의 (오(mis)-OCR된 단어에 대한 정정과 같은) 정정들의 형태를 취한다.
- [0063] 동일 타입 중 여럿으로부터 "올바른" 결과를 일반적으로 선택하는(예를 들어, 얼굴 인식 서버로부터 정확한 결과를 선택하는), 하나의 검색 결과의 사용자의 선택은 해석들 사이에서의 선택으로 지칭되는 프로세스이다. 여러 상이한 타입들의 리턴된 결과들로부터 관심의 결과 "타입"을 일반적으로 선택하는, 특정 타입의 검색 결과의 사용자의 선택(예를 들어, 또한 동일 페이지상에서 광고들에 대한 시각 결과들보다 오히려 잡지에서 기사의 OCR된 텍스트를 선택)은 의도의 명확화로 지칭되는 프로세스이다. 사용자는 도 8을 참조하여 상세히 설명된 OCR된 문서에서 특정 링크된 단어들(인식된 명명 엔티티들과 같은)을 유사하게 선택할 수 있다.
- [0064] 사용자는 대안적으로 또는 부가적으로 특정 검색 결과들에 주석을 달고자할 수 있다. 이 주석은 자유형 스타일 또는 구조된 포맷으로 이루어질 수 있다(422). 주석들은 결과의 설명들일 수 있거나 또는 결과의 리뷰들일 수 있다. 예를 들어, 주석들은 결과에서 서브젝트(들)의 명칭을 나타낼 수 있거나 또는 주석들은 "이는 좋은 책" 또는 "이 물건은 구입한 해 내에 망가졌다"를 나타낼 수 있다. 주석의 다른 예는 바운딩 박스 내에 서브젝트 또는 오브젝트를 식별하는 사용자-제공된 텍스트 및 시각 질의의 서브-부분 주위에 사용자에 의해 그려진 바운딩 박스이다. 사용자 주석들은 도 5를 참조하여 더 상세히 설명된다.
- [0065] 다른 주석들 및 검색 결과들의 사용자 선택들은 서버 시스템에 송신된다(424). 프론트 엔드 서버(110)는 선택들 및 주석들을 수신하고 추가적으로 이들을 프로세싱한다(426). 정보가 상호작용 결과 문서에서 오브젝트, 서브-영역 또는 용어의 선택인 경우, 상기 선택에 관한 추가의 정보는 적절히 요청될 수 있다. 예를 들어, 선택이 하나의 시각 결과인 경우, 시각 결과에 관한 더 많은 정보가 요청될 수 있다. 선택이 단어인 경우(OCR 서버 또는 이미지-대-용어들 서버로부터), 상기 용어의 텍스트 검색은 용어 질의 서버 시스템(118)에 송신될 수 있다. 선택이 얼굴 이미지 인식 검색 시스템으로부터의 사람인 경우, 상기 사람의 프로파일이 요청될 수 있다. 선택이 상호작용 검색 결과 문서의 특정 부분에 대한 것인 경우, 밀출된 시각 질의 결과들이 요청될 수 있다.
- [0066] 서버 시스템이 주석을 수신하는 경우, 주석은 도 5를 참조하여 설명된, 질의 및 주석 데이터베이스(116)에 저장된다. 그 다음에 주석 데이터베이스(116)로부터의 정보는 도 7 내지 도 10을 참조하여 아래에 논의되는 바와 같이, 병렬 서버 시스템들 중 하나 이상에 대하여 개별적인 주석 데이터베이스들에 주기적으로 카피된다.

- [0067] 도 5는 본 발명의 일 실시예에 따라 클라이언트 시스템(102)을 도시하는 블록도이다. 클라이언트 시스템(102)은 전형적으로 하나 이상의 프로세싱 유닛(CPU)들(702), 하나 이상의 네트워크 또는 다른 통신 인터페이스들(704), 메모리(712) 및 이러한 컴포넌트들을 상호접속시키기 위한 하나 이상의 통신 버스들(714)를 포함한다. 클라이언트 시스템(102)은 사용자 인터페이스(705)를 포함한다. 사용자 인터페이스(705)는 디스플레이 디바이스(706)를 포함하고 선택적으로 키보드, 마우스 또는 다른 입력 버튼들(708)과 같은 입력 수단을 포함한다. 대안적으로 또는 부가적으로 디스플레이 디바이스(706)는 터치 감응 표면(709)을 포함하고, 이 경우 디스플레이(706/709)가 터치 감응 디스플레이이다. 터치 감응 디스플레이(706/709)를 가지는 클라이언트 시스템들에서, 물리 키보드는 선택적이다(예를 들어, 키보드 입력이 필요할 때 소프트 키보드가 디스플레이될 수 있다). 또한, 임의의 클라이언트 시스템들은 키보드를 보충하거나 대체하기 위해 마이크로폰 및 음성 인식을 사용한다. 선택적으로, 클라이언트(102)는 GPS(global positioning satellite) 수신기 또는 클라이언트 시스템의 위치를 결정하기 위한 다른 위치 검출 장치(707)를 포함한다. 임의의 실시예들에서, 클라이언트 시스템(102)의 위치를 나타내는 위치 정보를 수신하기 위해 클라이언트 시스템(102)이 시각 질의 서버 시스템을 제공하도록 요구하는 시각 질의 검색 디바이스들이 제공된다.
- [0068] 클라이언트 시스템(102)은 또한 카메라 또는 스캐너와 같은 이미지 캡처 디바이스(710)를 포함한다. 메모리(712)는 DRAM, SRAM, DDR RAM 또는 다른 랜덤 액세스 솔리드 스테이트 메모리 디바이스들과 같은, 고속 랜덤 액세스 메모리를 포함하고, 하나 이상의 자기 디스크 저장 디바이스들, 광 디스크 저장 디바이스들, 플래시 메모리 디바이스들 또는 다른 비-휘발성 솔리드 스테이트 저장 디바이스들과 같은 비-휘발성 메모리를 포함할 수 있다. 메모리(712)는 선택적으로 CPU(들)(702)로부터 원격으로 위치한 하나 이상의 저장 디바이스들을 포함할 수 있다. 메모리(712) 또는 대안적으로 메모리(712) 내의 비-휘발성 메모리 디바이스(들)는 비-일시적인 컴퓨터 판독가능한 저장 매체를 포함한다. 임의의 실시예들에서, 메모리(712) 또는 메모리(712)의 컴퓨터 판독가능한 저장 매체는 아래의 프로그램들, 모듈들 및 데이터 구조들 또는 이들의 서브셋을 저장한다:
- [0069] · 하드웨어 의존 태스크들을 수행하기 위한 그리고 다양한 기본 시스템 서비스들을 핸들링하기 위한 절차들을 포함하는 운영 시스템(716);
- [0070] · 인터넷, 다른 광역 네트워크들, 로컬 영역 네트워크들, 도시 영역 네트워크들 등과 같은, 하나 이상의 통신 네트워크들 및 하나 이상의 통신 네트워크 인터페이스들(704)(유선 또는 무선)을 통해 다른 컴퓨터들에 클라이언트 시스템(102)을 접속시키기 위해 사용되는 네트워크 통신 모듈(718);
- [0071] · 이미지 캡처 디바이스/카메라(710)에 의해 캡처된 개별적인 이미지를 프로세싱하기 위한 이미지 캡처 모듈(720), 여기서, 개별적인 이미지는 시각 질의 서버 시스템에 시각 질의로서 (예를 들어, 클라이언트 애플리케이션 모듈에 의해) 송신될 수 있다;
- [0072] · 시각 질의 서버 시스템에 시각 질의들을 제출하기 위한 이미지에 의한 질의 제출 모듈(724); 선택적으로 이미지에서 관심 영역의 (터치 감응 디스플레이(706/709)상에서 동작과 같은) 선택을 검출하고 상기 관심 영역을 시각 질의로서 준비하는 관심 영역 선택 모듈(725); 시각 질의의 결과들을 디스플레이하기 위한 결과 브라우저(726); 및 선택적으로, 다양한 포맷들로부터 주석들을 수용할 수 있는, 자유형 주석 텍스트 입력(732) 또는 일 형태로 파일링하는 것과 같은 구조된 주석 텍스트 입력(730)에 대한 선택적 모듈들을 가진 주석 모듈(728) 및 주석을 위한 이미지의 특정 서브-부분을 사용자가 선택하도록 하는 이미지 영역 선택 모듈(734)(때때로 결과 선택 모듈로 본 명세서에서 지칭됨)을 포함하는(이에 제한되지 않음) 이미지에 의한 질의의 다양한 양상들을 핸들링하기 위한 하나 이상의 클라이언트 애플리케이션 모듈들(722);
- [0073] · 사용자가 이미지 캡처 디바이스(710)를 통해 바로 캡처링한 것보다 이미지를 생성하거나 편집함으로써 시각 질의를 저작하도록 하는 선택적 콘텐츠 저작 애플리케이션(들)(736); 선택적으로 하나 또는 이러한 애플리케이션들(736)은 사용자가 시각 질의로서 사용하기 위한 이미지의 서브-부분을 선택할 수 있도록 하는 명령들을 포함할 수 있다;
- [0074] · 시각 질의 서버 시스템에 시각 질의를 송신하기 전에 시각 질의를 선-프로세싱하는 선택적 로컬 이미지 분석 모듈(738). 로컬 이미지 분석은 이미지들의 특정 타입들 또는 이미지 내의 서브-영역들을 인식할 수 있다. 이러한 모듈들(738)에 의해 인식될 수 있는 이미지 타입들의 예시들은 얼굴 타입(시각 질의 내에 인식된 얼굴 이미지), 바 코드 타입(시각 질의 내에 인식된 바 코드), 및 텍스트 타입(시각 질의 내에 인식된 텍스트) 중 하나 이상을 포함한다; 그리고
- [0075] · 이메일 애플리케이션, 전화 애플리케이션, 브라우저 애플리케이션, 맵핑 애플리케이션, 인스턴트 메시징 애플리케이션,

플리케이션, 소셜 네트워킹 애플리케이션 등과 같은 부가적인 선택적 클라이언트 애플리케이션들(740). 임의의 실시예들에서, 적절한 실행가능한 검색 결과에 대응하는 애플리케이션은 실행가능한 검색 결과가 선택될 때 런칭(launch) 또는 액세스될 수 있다.

[0076] 선택적으로, 사용자가 주식에 대한 이미지의 특정 서브-부분을 선택하도록 하는 이미지 영역 선택 모듈(734)은 또한 사용자가 추가로 주식을 달 필요 없이 검색 결과를 "정확한" 히트로서 선택하도록 한다. 예를 들어, 사용자는 상위 N개의 얼굴 인식 매칭들을 제시받을 수 있고 상기 결과 리스트로부터 정확한 사람들 선택할 수 있다. 임의의 검색 질의들에 대하여 둘 이상의 타입의 결과가 표시될 것이고, 사용자는 결과의 타입을 선택할 것이다. 예를 들어, 이미지 질의는 나무 옆에 서있는 사람을 포함할 수 있지만 사람에 관한 결과들만 사용자에게 관심이 있다. 그러므로, 이미지 선택 모듈(734)은 사용자가 어떤 타입의 이미지가 "정확한" 타입인지 - 즉, 사용자가 수신하기를 원하는 타입인지를 나타내도록 한다. 사용자는 또한 자유형 주식 텍스트 입력 모듈(732) 또는 (일 형태로 파일링하기 위한) 주식 텍스트 입력 모듈(730)을 사용하여 개인적 코멘트들 또는 설명 단어들을 부가함으로써 검색 결과에 주석을 달고자 할 수 있다.

[0077] 임의의 실시예들에서, 선택적 로컬 이미지 분석 모듈(738)은 클라이언트 애플리케이션(108, 도 1)의 부분이다. 또한, 임의의 실시예들에서, 선택적 로컬 이미지 분석 모듈(738)은 시각 질의 또는 이들의 부분을 선-프로세싱 또는 카테고리화하기 위해 로컬 이미지 분석을 수행하기 위한 하나 이상의 프로그램들을 포함한다. 예를 들어, 클라이언트 애플리케이션(722)은 검색 엔진에 시각 질의를 제출하기 전에 이미지가 바 코드, 얼굴, 또는 텍스트를 포함함을 인식할 수 있다. 임의의 실시예들에서, 로컬 이미지 분석 모듈(738)이 시각 질의가 특정 타입의 이미지를 포함함을 검출할 때, 모듈은 사용자에게 자신들이 대응하는 타입의 검색 결과에 관심이 있는지 묻는다. 예를 들어, 로컬 이미지 분석 모듈(738)은 자신의 일반적 특성들에 기반하여 얼굴을 검출하고(즉, 어떤 사람의 얼굴인지 결정하지 않고), 시각 질의 서버 시스템에 질의를 송신하기 전에 사용자에게 즉시 피드백을 제공할 수 있다. "얼굴이 검출되었다, 당신은 이 얼굴에 대한 얼굴 인식 매칭들을 얻는데 관심이 있는가"와 같은 결과를 리턴할 수 있다. 이는 시각 질의 서버 시스템(106, 도 1)에 대하여 시간들을 절약할 수 있다. 임의의 시각 질의들에 대하여, 프론트 엔드 시각 질의 프로세싱 서버(110, 도 1)는 오직 로컬 이미지 분석 모듈(738)에 의해 인식되는 이미지의 타입에 대응하는 검색 시스템(112)에 시각 질의를 송신한다. 다른 실시예들에서, 검색 시스템(112)에 대한 시각 질의는 검색 시스템들(112A-N) 모두에 시각 질의를 송신할 수 있지만, 로컬 이미지 분석 모듈(738)에 의해 인식되는 이미지의 타입에 대응하는 검색 시스템(112)으로부터의 결과들을 순위화할 것이다. 임의의 실시예들에서, 로컬 이미지 분석이 시각 질의 서버 시스템의 동작에 영향을 주는 방식은 클라이언트 시스템의 구성 또는 사용자 또는 클라이언트 시스템과 연관된 파라미터들의 구성 또는 프로세싱에 의존한다. 또한, 임의의 특정 시각 질의 및 로컬 이미지 분석에 의해 산출되는 결과들의 실제 콘텐츠는 상이한 시각 질의들이 클라이언트 시스템 및 시각 질의 서버 시스템 둘 다 또는 둘 중 하나에서 상이하게 핸들링되도록 할 수 있다.

[0078] 임의의 실시예들에서, 바 코드 인식은 시각 질의가 로컬 이미지 분석 모듈(738)에서 클라이언트 시스템상에서 수행되는 바 코드를 포함하는지의 분석을 이용하여, 두 단계들로 수행된다. 그 다음에 시각 질의는 시각 질의가 바 코드를 포함할 가능성이 있다고 클라이언트가 결정하는 경우에만 바 코드 검색 시스템에 패스된다. 다른 실시예들에서, 바 코드 검색 시스템은 모든 시각 질의를 프로세싱한다.

[0079] 선택적으로, 클라이언트 시스템(102)은 부가적인 클라이언트 애플리케이션들(740)을 포함한다.

[0080] 도 6은 본 발명의 일 실시예에 따라 프론트 엔드 시각 질의 프로세싱 서버 시스템(110)을 도시하는 블록도이다. 프론트 엔드 서버(110)는 전형적으로 하나 이상의 프로세싱 유닛(CPU)들(802), 하나 이상의 네트워크 또는 다른 통신 인터페이스들(804), 메모리(812), 및 이러한 컴포넌트들을 상호접속시키기 위한 하나 이상의 통신 버스들(814)을 포함한다. 메모리(812)는 DRAM, SRAM, DDR RAM 또는 다른 랜덤 액세스 솔리드 스테이트 메모리 디바이스들과 같은, 고속 랜덤 액세스 메모리를 포함하고, 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들 또는 다른 비-휘발성 솔리드 스테이트 저장 디바이스들과 같은, 비-휘발성 메모리를 포함할 수 있다. 메모리(812)는 CPU(들)(802)로부터 원격으로 위치된 하나 이상의 저장 디바이스들을 선택적으로 포함할 수 있다. 메모리(812) 또는 대안적으로 메모리(812) 내의 비-휘발성 메모리 디바이스(들)은 비-일시적인 컴퓨터 판독가능한 저장 매체를 포함한다. 임의의 실시예들에서, 메모리(812) 또는 메모리(812)의 컴퓨터 판독가능한 저장 매체는 아래의 프로그램들, 모듈들 및 데이터 구조들 또는 이들의 서브셋을 저장한다;

[0081] · 하드웨어 의존 태스크들을 수행하기 위한 그리고 다양한 기본 시스템 서비스들을 핸들링하기 위한 절차들을

포함하는 운영 시스템(816);

- [0082] · 인터넷, 다른 광역 네트워크들, 로컬 영역 네트워크들, 도시 영역 네트워크들 등과 같은 하나 이상의 통신 네트워크들 및 하나 이상의 통신 네트워크 인터페이스들(804)(유선 또는 무선)을 통해 다른 컴퓨터들에 프론트엔드 서버 시스템(110)을 접속시키기 위해 사용되는 네트워크 통신 모듈(818);
- [0083] · 클라이언트 시스템(102)으로부터 인입하는 시각 질의들을 핸들링하고 이들을 둘 이상의 병렬 검색 시스템들에 송신하기 위한 질의 관리자(820); 본 출원의 다른 곳에서 설명된 바와 같이, 임의의 특별 상황들에서, 시각 질의는 예를 들어, 시각 질의가 클라이언트-발생된 명령을 포함할 때(예를 들어, "얼굴 인식 검색만"), 오직 하나의 검색 시스템들에 안내될 수 있다;
- [0084] · 하나 이상의 병렬 검색 시스템들로부터의 결과들을 선택적으로 필터링하고 표시를 위해 클라이언트 시스템(102)에 상위 또는 "관련" 결과들을 송신하기 위한 결과 필터링 모듈(822);
- [0085] · 하나 이상의 병렬 검색 시스템들로부터의 결과들을 선택적으로 순위화하고 표시를 위해 결과들을 포맷화하기 위한 결과 순위화 및 포맷화 모듈(824);
- [0086] · 상호작용 검색 결과 문서를 생성하기 위해 결과 문서 생성 모듈(826)이 적절할 때 사용된다; 모듈(826)은 바운딩 박스 생성 모듈(828) 및 링크 생성 모듈(830)을 포함하는(이에 제한되지 않음) 서브-모듈들을 포함할 수 있다;
- [0087] · 시각 질의의 개별적인 서브-부분들의 시각 식별자들인 라벨들을 생성하기 위한 라벨 생성 모듈(831);
- [0088] · 사용자로부터 주석들을 수신하고 이들을 주석 데이터베이스(116)에 송신하기 위한 주석 모듈(832);
- [0089] · 시각 질의에 응답하여 하나 이상의 실행가능한 검색 결과 엘리먼트들을 발생시키기 위한 실행가능한 검색 결과 모듈(838), 각각은 클라이언트-측 실행을 런칭하도록 구성된다; 실행가능한 검색 결과 엘리먼트들의 예시들은 전화 통화를 개시하고, 이메일 메시지를 개시하고, 주소를 맵핑하고, 레스토랑 예약을 하고, 물건을 구입하기 위한 옵션을 제공하기 위한 버튼들이다; 그리고
- [0090] · 데이터 베이스 자체(834) 및 데이터베이스에 대한 인덱스(836)를 포함하는 질의 및 주석 데이터베이스(116)
- [0091] 결과 순위화 및 포맷화 모듈(824)은 하나 이상의 병렬 검색 시스템들(112-A - 112-N, 도 1)로부터 리턴되는 결과들을 순위화한다. 이미 상기 언급된 바와 같이, 임의의 시각 질의들에 대하여, 하나의 검색 시스템으로부터의 결과들만이 관련될 수 있다. 이러한 예시에서, 하나의 검색 시스템으로부터의 관련 검색 결과들만이 순위화된다. 이러한 예시들에서, 임의의 실시예들에서, 여러 타입들의 검색 결과들이 관련될 수 있다. 이러한 예시들에서, 임의의 실시예들에서, 결과 순위화 및 포맷화 모듈(824)은 덜 관련된 검색 시스템들에 대한 결과들 위에 가장 관련된 결과(예를 들어, 가장 높은 관련 스코어를 가진 결과)를 가지는 검색 시스템으로부터의 모든 결과들을 순위화한다. 다른 실시예들에서, 결과 순위화 및 포맷화 모듈(824)은 남아있는 결과들 위에 각 관련 검색 시스템으로부터의 상위 결과를 순위화한다. 임의의 실시예들에서, 결과 순위화 및 포맷화 모듈(824)은 검색 결과들의 각각에 대하여 계산된 관련성 스코어에 따라 결과들을 순위화한다. 임의의 시각 질의들에 대하여, 보강된 텍스트 질의들은 병렬 시각 검색 시스템들에서 검색에 부가하여 수행된다. 임의의 실시예들에서, 또한 텍스트 질의들이 수행될 때, 자신들의 결과들이 시각 검색 시스템 결과들로부터 시각적으로 별개인 방식으로 표시된다.
- [0092] 결과 순위화 및 포맷화 모듈(824)은 결과들을 또한 포맷한다. 임의의 실시예들에서, 결과들은 리스트 포맷으로 표시된다. 임의의 실시예들에서, 결과들은 상호작용 결과 문서에 의해 표시된다. 임의의 실시예들에서, 상호작용 결과 문서 및 결과들의 리스트 둘 다 표시된다. 임의의 실시예들에서, 질의의 타입은 어떻게 결과들이 표시될지를 좌우한다. 예를 들어, 둘 이상의 검색가능한 서브젝트가 시각 질의에서 검출되는 경우, 그 다음에 상호작용 결과 문서가 산출되는 반면, 오직 하나의 검색가능한 서브젝트가 검출되는 경우, 결과들은 리스트 포맷으로만 디스플레이될 것이다.
- [0093] 결과 문서 생성 모듈(826)은 상호작용 검색 결과 문서를 생성하기 위해 사용된다. 상호작용 검색 결과 문서는 하나 이상의 검출되고 검색된 서브젝트들을 가질 수 있다. 바운딩 박스 생성 모듈(828)은 검색된 서브젝트들 중 하나 이상 주위에 바운딩 박스를 생성한다. 바운딩 박스들은 직사각형 박스들일 수 있거나 또는 서브젝트(들)의 형상(들)의 윤곽을 표시할 수 있다. 링크 생성 모듈(830)은 상호작용 검색 결과 문서에서 자신들의 개별적인 서브젝트와 연관된 검색 결과들에 대한 링크들을 생성한다. 임의의 실시예들에서, 바운딩 박스 영역 내

의 클릭은 링크 생성 모듈에 의해 삽입되는 대응하는 링크를 활성화한다.

- [0094] 질의 및 주석 데이터베이스(116)는 시각 질의 결과들을 개선하기 위해 사용될 수 있는 정보를 포함한다. 임의의 실시예들에서, 사용자는 시각 질의 결과들이 표시된 후에 이미지에 주석을 달 수 있다. 또한, 임의의 실시예들에서, 사용자는 시각 질의 검색 시스템에 이미지를 송신하기 전에 이미지에 주석을 달 수 있다. 선-주석은 결과들을 포커싱 또는 시각 질의 검색과 병렬로 주석된 단어들에 텍스트 기반 검색들을 실행함으로써 시각 질의 프로세싱을 도울 수 있다. 임의의 실시예들에서, 픽처의 주석된 버전들은 잠재 이미지 매칭 히트로서 리턴되도록 공지될 수 있다(예를 들어, 사용자가 예를 들어, 사적이지만 않도록 이미지 및 주석(들)을 지정함으로써 공개를 위한 허가를 줄 때). 예를 들어, 사용자가 꽃의 픽처를 찍고, 상기 꽃에 대해 세부 속(genus) 및 종(species) 정보를 부여함으로써 이미지에 주석을 다는 경우, 사용자는 상기 꽃을 찾는 시각 질의 리서치를 수행하는 누군가에게 상기 이미지가 표시되기를 원할 수 있다. 임의의 실시예들에서, 질의 및 주석 데이터베이스(116)로부터의 정보는 자신들의 각각의 개별 데이터베이스들(114)로 정보의 관련 부분(존재한다면)을 통합하는, 병렬 검색 시스템들(112)에 주기적으로 푸시된다.
- [0095] 도 7은 시각 질의를 프로세싱하도록 이용되는 병렬 검색 시스템들 중 하나를 도시하는 블록도이다. 도 7은 본 발명의 일 실시예에 따라 "포괄적인" 서버 시스템(112-N)을 도시한다. 이 서버 시스템은 시각 질의 검색 서버들(112-N) 중 임의의 하나를 표시하는 경우에만 포괄적이다. 포괄적인 서버 시스템(112-N)은 전형적으로 하나 이상의 프로세싱 유닛(CPU)들(502), 하나 이상의 네트워크 또는 다른 통신 인터페이스들(504), 메모리(512), 및 이러한 컴포넌트들을 상호접속시키기 위한 하나 이상의 통신 버스들(514)을 포함한다. 메모리(512)는 DRAM, SRAM, DDR RAM 또는 다른 랜덤 액세스 솔리드 스테이트 메모리 디바이스들과 같은, 고속 랜덤 액세스 메모리를 포함하고; 그리고 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들 또는 다른 비-휘발성 솔리드 스테이트 저장 디바이스들과 같은, 비-휘발성 메모리를 포함할 수 있다. 메모리(512)는 CPU(들)(502)로부터 원격으로 위치된 하나 이상의 저장 디바이스들을 선택적으로 포함할 수 있다. 메모리(512) 또는 대안적으로 메모리(512) 내의 비-휘발성 메모리 디바이스(들)는 비-일시적인 컴퓨터 관독가능한 저장 매체를 포함한다. 임의의 실시예들에서, 메모리(512) 또는 메모리(512)의 컴퓨터 관독가능한 저장 매체는 아래의 프로그램들, 모듈들 및 데이터 구조들 또는 이들의 서브세트를 저장한다;
 - [0096] · 하드웨어 의존 태스크들을 수행하기 위한 그리고 다양한 기본 시스템 서비스들을 핸들링하기 위한 절차들을 포함하는 운영 시스템(516);
 - [0097] · 인터넷, 다른 광역 네트워크들, 로컬 영역 네트워크들, 도시 영역 네트워크들 등과 같은, 하나 이상의 통신 네트워크들 및 하나 이상의 통신 네트워크 인터페이스들(504)(유선 또는 무선)을 통해 다른 컴퓨터들에 포괄적 서버 시스템(112-N)을 접속시키기 위해 사용되는 네트워크 통신 모듈(518);
 - [0098] · 특정 서버 시스템에 특유한 검색 애플리케이션(520), 예를 들어, 바 코드 검색 애플리케이션, 색 인식 검색 애플리케이션, 물건 인식 검색 애플리케이션, 오브젝트-또는-오브젝트 카테고리 검색 애플리케이션 등일 수 있다;
 - [0099] · 특정 검색 애플리케이션이 인덱스를 이용하는 경우 선택적 인덱스(522);
 - [0100] · 특정 검색 애플리케이션에 관련된 이미지들을 저장하기 위한 선택적 이미지 데이터베이스(524), 존재한다면, 저장된 이미지 데이터는 검색 프로세스 타입에 의존한다;
 - [0101] · 검색 애플리케이션으로부터의 결과들을 순위화하기 위한 선택적 결과 순위화 모듈(526)(때때로 관련 스코어링 모듈로 지칭됨), 순위화 모듈은 검색 애플리케이션으로부터 각 결과에 대한 관련성 스코어를 할당할 수 있고, 결과들이 미리-정의된 최소 스코어에 도달하지 않는 경우, 이 서버 시스템으로부터의 결과들이 관련되지 않음을 나타내는 널 및 제로 값 스코어를 프론트 엔드 시각 질의 프로세싱 서버에 리턴할 수 있다; 그리고
 - [0102] · 주석 정보 중 일부가 특정 검색 애플리케이션에 관련되는지 여부를 결정하는 주석 정보를 주석 데이터베이스(116, 도 1)로부터 수신하고 주석 정보의 임의의 결정된 관련 부분들을 각각의 주석 데이터베이스(530)에 통합하기 위한 주석 모듈(528).
- [0103] 도 8은 본 발명의 일 실시예에 따라 시각 질의를 프로세싱하기 위해 이용되는 OCR 검색 시스템(112-B)을 도시하는 블록도이다. OCR 검색 시스템(112-B)은 전형적으로, 하나 이상의 프로세싱 유닛(CPU)들(602), 하나 이상의 네트워크 또는 다른 통신 인터페이스들(604), 메모리(612) 및 이러한 컴포넌트들을 상호접속시키기 위한 하나 이상의 통신 버스들(614)을 포함한다. 메모리(612)는 DRAM, SRAM, DDR RAM 또는 다른 랜덤 액세스 솔리드 스테이트 메모리 디바이스들과 같은 고속 랜덤 액세스 메모리를 포함하고; 하나 이상의 자기 디스크 저장 디바이스

들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들 또는 다른 비-휘발성 솔리드 스테이트 저장 디바이스들과 같은 비-휘발성 메모리를 포함할 수 있다. 메모리(612)는 CPU(들)(602)로부터 원격으로 위치된 하나 이상의 저장 디바이스들을 선택적으로 포함할 수 있다. 메모리(612) 또는 대안적으로 메모리(612) 내의 비-휘발성 메모리 디바이스(들)는 비-일시적인 컴퓨터 판독가능한 저장 매체를 포함한다. 임의의 실시예들에서, 메모리(612) 또는 메모리(612)의 컴퓨터 판독가능한 저장 매체는 아래의 프로그램들, 모듈들 및 데이터 구조들 또는 이들의 서브셋을 저장한다;

- [0104] · 다양한 기본 시스템 서비스들을 핸들링하고 하드웨어 의존 태스크들을 수행하기 위한 절차들을 포함하는 운영 시스템(616);
- [0105] · 인터넷, 다른 광역 네트워크들, 로컬 영역 네트워크들, 도시 영역 네트워크들 등과 같은, 하나 이상의 통신 네트워크들 및 하나 이상의 통신 네트워크 인터페이스들(604)(무선 또는 유선)을 통해 다른 컴퓨터들에 OCR 검색 시스템(112-B)을 접속시키기 위해 사용되는 네트워크 통신 모듈(618);
- [0106] · 시각 질의에서 텍스트를 인식하기 위해 시도하고 글자들의 이미지들을 문자들로 변환하는 광학식 문자 인식(OCR) 모듈(620);
- [0107] · 특정 폰트들, 텍스트 패턴들 및 글자 인식에 고유한 다른 특성들을 인식하기 위해 OCR 모듈(620)에 의해 사용되는 선택적 OCR 데이터베이스(114-B);
- [0108] · 사전에 대하여 변환된 단어들을 확인하고, 잠재적으로 오-변환된 글자들을 이와 다른 사전 단어에 매칭하는 단어들로 대체함으로써 글자들의 이미지들의 문자들로의 변환을 개선하는 선택적 철자 확인 모듈(622);
- [0109] · 변환된 텍스트 내에서 명명 엔티티들을 검색하고, 용어 질의 서버 시스템(118, 도 1)에 용어 질의 용어들로서 인식된 명명 엔티티들을 송신하고, 인식된 명명 엔티티들과 연관된 OCR된 텍스트에 임베딩된 링크들로서 용어 질의 서버 시스템으로부터의 결과들을 제공하는 선택적 명명 엔티티 인식 모듈(624);
- [0110] · 텍스트 세그먼트들의 데이터베이스에 대하여 변환된 세그먼트들(변환된 문장들 및 문구들과 같은)을 확인하고, 잠재적으로 오-변환된 글자들을 이와 다른 텍스트 매칭 애플리케이션 텍스트 세그먼트에 매칭하는 OCR된 텍스트 세그먼트들로 대체함으로써 글자들의 이미지들의 문자들로의 변환을 개선하는 선택적 텍스트 매칭 애플리케이션(632), 임의의 실시예들에서, 텍스트 매칭 애플리케이션에 의해 발견된 텍스트 세그먼트는 사용자에게 링크로서 제공된다(예를 들어, 사용자가 뉴욕 타임즈의 한 페이지를 스캔한 경우, 텍스트 매칭 애플리케이션은 뉴욕 타임즈 웹사이트에서 전체 포스팅된 기사에 대한 링크를 제공할 수 있다);
- [0111] · 표시를 위해 OCR된 결과들을 포맷하고 명명 엔티티들에 대한 선택적 링크들을 포맷하고 또한 선택적으로 텍스트 매칭 애플리케이션으로부터 임의의 관련된 결과들을 순위화하기 위한 결과 순위화 및 포맷화 모듈(626); 및
- [0112] · 임의의 주식 정보가 OCR 검색 시스템에 관련되는지 여부를 결정하는 주식 정보를 주식 데이터베이스(116, 도 1)로부터 수신하고, 주식 정보의 임의의 결정된 관련 부분들을 각각의 주식 데이터베이스(630)에 통합하기 위한 선택적 주식 모듈(628).
- [0113] 도 9는 본 발명의 일 실시예에 따라 시각 질의를 프로세싱하도록 이용되는 얼굴 인식 검색 시스템(112-A)을 도시하는 블록도이다. 얼굴 인식 검색 시스템(112-A)은 전형적으로 하나 이상의 프로세싱 유닛(CPU)들(902), 하나 이상의 네트워크 또는 다른 통신 인터페이스들(904), 메모리(912) 및 이러한 컴포넌트들을 상호접속시키기 위한 하나 이상의 통신 버스들(914)을 포함한다. 메모리(912)는 DRAM, SRAM, DDR RAM 또는 다른 랜덤 액세스 솔리드 스테이트 메모리 디바이스들과 같은 고속 랜덤 액세스 메모리를 포함하고; 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들 또는 다른 비-휘발성 솔리드 스테이트 저장 디바이스들과 같은 비-휘발성 메모리를 포함할 수 있다. 메모리(912)는 선택적으로 CPU(들)(902)로부터 원격으로 위치된 하나 이상의 저장 디바이스들을 포함할 수 있다. 메모리(912) 또는 대안적으로 메모리(912) 내의 비-휘발성 메모리 디바이스(들)는 비-일시적인 컴퓨터 판독가능한 저장 매체를 포함한다. 임의의 실시예들에서, 메모리(912) 또는 메모리(912)의 컴퓨터 판독가능한 저장 매체는 아래의 프로그램들, 모듈들 및 데이터 구조들 또는 이들의 서브셋을 저장한다:
- [0114] · 다양한 기본 시스템 서비스들을 핸들링하고 하드웨어 의존 태스크들을 수행하기 위한 절차들을 포함하는 운영 시스템(916);
- [0115] · 인터넷, 다른 광역 네트워크들, 로컬 영역 네트워크들, 도시 영역 네트워크들 등과 같은 하나 이상의 통신 네트워크들 및 하나 이상의 통신 네트워크 인터페이스들(904)(유선 또는 무선)을 통해 다른 컴퓨터들에 얼굴 인식

검색 시스템(112-A)을 접속시키기 위해 사용되는 네트워크 통신 모듈(918);

- [0116] · 얼굴 이미지 데이터베이스(114-A)에서 시각 질의에 표시되는 얼굴(들)에 매칭하는 얼굴 이미지들을 검색하기 위한 얼굴 인식 검색 애플리케이션(920)은 얼굴 이미지 데이터베이스(114-A)에서 발견된 각 매칭에 관한 정보에 대하여 소셜 네트워크 데이터베이스(922)를 검색한다.
- [0117] · 복수의 사용자들에 대한 하나 이상의 얼굴 이미지들을 저장하기 위한 얼굴 이미지 데이터베이스(114-A); 선택적으로, 얼굴 이미지 데이터베이스는 가족 구성원들 및 사용자들에 의해 알려지고 얼굴 이미지 데이터베이스(114-A)에 포함된 이미지들에 나타난 것으로 식별된 다른 사람들과 같은 사용자들 외의 사람들에 대한 얼굴 이미지들을 포함한다; 선택적으로, 얼굴 이미지 데이터베이스는 예를 들어, 법률적으로 공공 도메인에 있는 얼굴 이미지들의 벤더들과 같은, 외부 소스들로부터 획득되는 얼굴 이미지들을 포함한다;
- [0118] · 선택적으로, 도 12a를 참조하여 더 상세히 설명되는 바와 같이, 명칭, 주소, 직업, 그룹 멤버십들, 소셜 네트워크 접속들, 모바일 디바이스의 현재 GPS 위치, 공유 선호도들, 관심들, 나이, 고향, 개인적 통계들, 일 정보 등과 같은 소셜 네트워크의 사용자들에 관한 정보를 포함하는 소셜 네트워크 데이터베이스(922);
- [0119] · 얼굴 이미지 데이터베이스(114-A)로부터 잠재 얼굴 매칭들을 순위화(예를 들어, 관련 및/또는 매칭 품질 스코어를 할당)하고 표시를 위해 결과들을 포맷하기 위한 결과 순위화 및 포맷화 모듈(924); 임의의 실시예들에서, 결과들의 순위화 또는 스코어링은 앞서 언급된 소셜 네트워크 데이터베이스로부터 리트리브되는 관련 정보를 이용한다; 임의의 실시예에서, 검색 포맷 결과들은 소셜 네트워크 데이터베이스로부터의 정보의 서브셋뿐만 아니라 잠재 이미지 매칭들을 포함한다; 그리고
- [0120] · 임의의 주석 정보가 얼굴 인식 검색 시스템에 관련되는지 여부를 결정하는 주석 정보를 주석 데이터베이스(116, 도 1)로부터 수신하고 각각의 주석 데이터베이스(928)로 주석 정보의 임의의 결정된 관련 부분들을 저장하기 위한 주석 모듈(926).
- [0121] 도 10은 본 발명의 일 실시예에 따라 시각 질의를 프로세싱하도록 이용된 이미지-대-용어들 검색 시스템(112-C)을 도시하는 블록도이다. 임의의 실시예들에서, 이미지-대-용어들 검색 시스템은 시각 질의에서 오브젝트들을 인식한다(인스턴스 인식). 다른 실시예들에서, 이미지-대-용어들 검색 시스템은 시각 질의에서 오브젝트 카테고리들을 인식한다(타입 인식). 임의의 실시예들에서, 이미지 대 용어들 시스템은 오브젝트들 및 오브젝트-카테고리들 둘 다를 인식한다. 이미지-대-용어들 검색 시스템은 시각 질의에서 이미지들에 대한 잠재 용어 매칭들을 리턴한다. 이미지-대-용어들 검색 시스템(112-C)은 전형적으로 하나 이상의 프로세싱 유닛(CPU)들(1002), 하나 이상의 네트워크 또는 다른 통신 인터페이스들(1004), 메모리(1012), 및 이러한 컴포넌트들을 상호접속시키기 위한 하나 이상의 통신 버스들(1014)을 포함한다. 메모리(1012)는 DRAM, SRAM, DDR RAM 또는 다른 랜덤 액세스 솔리드 스테이트 메모리 디바이스들과 같은 고속 랜덤 액세스 메모리를 포함하고, 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들 또는 다른 비-휘발성 솔리드 스테이트 저장 디바이스들과 같은 비-휘발성 메모리를 포함할 수 있다. 메모리(1012)는 선택적으로 CPU(들)(1002)로부터 원격으로 위치된 하나 이상의 저장 디바이스들을 포함한다. 메모리(1012) 또는 대안적으로 메모리(1012) 내의 비-휘발성 메모리 디바이스(들)는 비-일시적인 컴퓨터 판독가능한 저장 매체를 포함한다. 임의의 실시예들에서, 메모리(1012) 또는 메모리(1012)의 컴퓨터 판독가능한 저장 매체는 아래의 프로그램들, 모듈들 및 데이터 구조들을 저장하거나 또는 이들의 서브셋을 저장한다:
- [0122] · 다양한 기본 시스템 서비스들을 핸들링하고 하드웨어 의존 태스크들을 수행하기 위한 절차들을 포함하는 운영 시스템(1016);
- [0123] · 인터넷, 다른 광역 네트워크들, 로컬 영역 네트워크들, 도시 영역 네트워크들 등과 같은, 하나 이상의 통신 네트워크들 및 하나 이상의 통신 네트워크 인터페이스들(1004)(유선 또는 무선)을 통해 다른 컴퓨터들에 이미지-대-용어들 검색 시스템(112-C)을 접속시키기 위해 사용되는 네트워크 통신 모듈(1018);
- [0124] · 이미지 검색 데이터베이스(114-C)에서 시각 질의의 서브젝트 또는 서브젝트들에 매칭하는 이미지들을 검색하는 이미지-대-용어들 검색 애플리케이션(1020);
- [0125] · 시각 질의의 서브젝트(들)에 유사한 이미지들을 발견하기 위해 검색 애플리케이션(1020)에 의해 검색될 수 있는 이미지 검색 데이터베이스(114-C);
- [0126] · 텍스트 기반 질의 검색 엔진(1006)을 이용하는 이미지들을 검색할 때 사용자들에 의해 사용되는 텍스트 용어들을 저장하는 용어들-대-이미지 역 인덱스(1022);

- [0127] · 잠재 이미지 매칭들을 순위화하고 그리고/또는 용어들-대-이미지 역 인덱스(1022)에 식별되는 잠재 이미지 매칭들과 연관된 용어들을 순위화하기 위한 결과 순위화 및 포맷 모듈(1024); 및
- [0128] · 임의의 주식 정보가 이미지-대 용어 검색 시스템(112-C)에 관련되는지 여부를 결정하는 주식 정보를 주식 데이터베이스(116, 도 1)로부터 수신하고 각각의 주식 데이터베이스(1028)로 주식 정보의 임의의 결정된 관련 부분들을 저장하기 위한 주식 모듈(1026).
- [0129] 도 5-10은 본 명세서에 설명된 실시예들의 구조적 도식보다 컴퓨터 시스템들의 세트에 표현될 수 있는 다양한 특징들의 기능적 설명들로 의도된다. 실제, 그리고 당업자에 의해 인식되는 바와 같이, 분리하여 도시된 아이템들은 결합될 수 있고 임의의 아이템들은 분리될 수 있다. 예를 들어, 이러한 도면들로 분리하여 도시된 임의의 아이템들은 단일 서버들에서 구현될 수 있고 단일 아이템들은 하나 이상의 서버들에 의해 구현될 수 있다. 시각 질의 프로세싱을 구현하기 위해 사용되는 시스템들의 실제 수 및 어떤 특징들이 이들 사이에 할당되는지는 구현에 따라 변할 것이다.
- [0130] 본 명세서에 설명된 방법들의 각각은 비-일시적인 컴퓨터 판독가능한 저장 매체에 저장되고 하나 이상의 서버들 또는 클라이언트들의 하나 이상의 프로세서들에 의해 실행되는 명령들에 의해 감독될 수 있다. 상기 식별된 모듈들 또는 프로그램들(즉, 명령들의 세트들)은 분리된 소프트웨어 프로그램들, 절차들 또는 모듈들로서 구현될 필요가 없고, 그러므로, 이러한 모듈들의 다양한 서브세트들은 다양한 실시예들에서 결합될 수 있거나 또는 달리 말하면 재배열될 수 있다. 도 5-10에 도시된 동작들의 각각은 컴퓨터 메모리 또는 비-일시적인 컴퓨터 판독가능한 저장 매체에 저장된 명령들에 대응할 수 있다.
- [0131] 도 11은 예시적인 시각 질의(1102)의 스크린 샷을 가진 클라이언트 시스템(102)을 도시한다. 도 11에 도시된 클라이언트 시스템(102)은 휴대 전화, 휴대용 음악 플레이어 또는 휴대용 이메일 디바이스와 같은 모바일 디바이스이다. 클라이언트 시스템(102)은 디스플레이(706) 및 이 도면에 도시된 버튼들과 같은 하나 이상의 입력 수단(708)을 포함한다. 임의의 실시예들에서, 디스플레이(706)는 터치 감응 디스플레이(709)이다. 터치 감응 디스플레이(709)를 가지는 실시예들에서, 디스플레이(709)에 디스플레이된 소프트 버튼들은 전자기계식 버튼들(708) 중 일부 또는 전부를 선택적으로 대체할 수 있다. 터치 감응 디스플레이들은 또한 아래에 더 상세히 설명된 바와 같이, 시각 질의 결과들과 상호작용하는 것을 도울 수 있다. 클라이언트 시스템(102)은 또한 카메라(710)와 같은 이미지 캡처 메커니즘을 포함한다.
- [0132] 도 11은 스토어의 선반에서 포장의 비디오 프레임 또는 사진인 비디오 질의(1102)를 도시한다. 본 명세서에 설명된 실시예들에서, 시각 질의는 두 차원들의 각각의 픽셀들에서 시각 질의의 크기에 대응하는 해상도를 가지는 2 차원 이미지이다. 이 실시예에서 시각 질의(1102)는 3 차원 오브젝트들의 2 차원 이미지이다. 시각 질의(1102)는 배경 엘리먼트, 물건 포장(1104) 및 사람의 이미지(1106), 상표의 이미지(1108), 물건의 이미지(1110) 및 다양한 텍스트 엘리먼트들(1112)을 포함하는 포장의 다양한 타입들의 엔티티들을 포함한다.
- [0133] 도 3을 참조하여 설명된 바와 같이, 시각 질의(1102)는 복수의 병렬 검색 시스템들(112A-N)에 시각 질의(1102)를 송신하고, 결과들을 수신하고 상호작용 결과 문서를 생성하는 프론트 엔드 서버(110)에 송신된다.
- [0134] 도면들(12a 및 12b)은 각각 상호작용 결과 문서(1200)의 실시예의 스크린 샷을 가진 클라이언트 시스템(102)을 도시한다. 상호작용 결과 문서(1200)는 검색 결과들의 서브세트에 대한 사용자 선택가능한 링크를 각각 포함하는, 시각 질의(1102)의 각각의 서브-부분들의 하나 이상의 시각 식별자들(1202)을 포함한다. 도면들(12a 및 12b)은 바운딩 박스들(1202)(예를 들어, 바운딩 박스들(1202-1, 1202-2, 1202-3))인 시각 식별자들을 가진 상호작용 결과 문서(1200)를 도시한다. 도면들(12a 및 12b)에 도시된 실시예들에서, 사용자는 자신의 바운딩 박스(1202)에 의해 윤곽이 나타내어진 공간 내에 활성 영역을 탭핑함으로써 특정 서브-부분에 대응하는 검색 결과들의 디스플레이를 활성화한다. 예를 들어, 사용자는 사람의 이미지를 서라운딩하는 바운딩 박스(1306)(도 13)를 탭핑함으로써, 사람의 이미지에 대응하는 검색 결과들을 활성화할 수 있다. 다른 실시예들에서, 선택가능한 링크는 터치 감응 디스플레이보다 마우스 또는 키보드를 사용하여 선택된다. 임의의 실시예들에서, 제 1 대응하는 검색 결과는 사용자가 바운딩 박스(1202)를 프리뷰할 때(즉, 사용자가 한 번 클릭하거나, 한 번 탭핑하거나, 바운딩 박스 위에 포인터를 맴돌 때) 디스플레이된다. 사용자는 사용자가 바운딩 박스를 선택할 때(즉, 사용자가 더블 클릭하거나, 두 번 탭핑하거나 또는 선택을 나타내기 위한 다른 메커니즘을 사용할 때) 복수의 대응하는 검색 결과들의 디스플레이를 활성화한다.
- [0135] 도면들(12a 및 12b)에서 시각 식별자들은 시각 질의의 서브-부분들을 서라운딩하는 바운딩 박스들(1202)이다. 도면 12a는 정사각 또는 직사각형인 바운딩 박스들(1202)을 도시한다. 도면 12b는 음료수 병에 대한 바운딩 박

스(1202-3)와 같은 시각 질의의 서브-부분에서 식별가능한 엔티티의 경계의 윤곽을 나타내는 바운딩 박스(1202)를 도시한다. 임의의 실시예들에서, 각각의 바운딩 박스(1202)는 그 내에 더 작은 바운딩 박스들(1202)을 포함한다. 예를 들어, 도면들(12a 및 12b)에서, 포장(1202-1)을 식별하는 바운딩 박스는 상표(1202-2)를 식별하는 바운딩 박스 및 다른 바운딩 박스들(1202) 모두를 서라운딩한다. 텍스트를 포함하는 임의의 실시예들에서, 또한 텍스트 용어들 중 일부에 대하여 활성 핫 링크들(1204)을 포함한다. 도면 12b는 "Active Drink" 및 "United States"가 핫 링크들(1204)로서 디스플레이되는 실시예를 도시한다. 이러한 용어들에 대응하는 검색 결과들은 용어 질의 서버 시스템(118)으로부터 수신되는 결과들인 반면, 바운딩 박스들에 대응하는 결과들은 이미지 검색 시스템들에 의한 질의로부터의 결과들이다.

[0136] 도 13은 시각 질의에서 인식된 엔티티의 타입에 의해 코딩되는 상호작용 결과 문서(1200)의 스크린 샷을 가진 클라이언트 시스템(102)을 도시한다. 도 11의 시각 질의는 사람의 이미지(1106), 상표의 이미지(1108), 물건의 이미지(1110), 및 다양한 텍스트 엘리먼트들(1112)을 포함한다. 따라서, 도 13에 디스플레이되는 상호작용 결과 문서(1200)는 사람(1306), 상표(1308), 물건(1310) 및 두 텍스트 영역들(1312) 주위의 바운딩 박스들(1202)을 포함한다. 도 13의 바운딩 박스들은 상이하게 색칠된 투명 바운딩 박스들(1202)을 나타내는 분리된 크로스-해칭을 이용하여 각각 표시된다. 임의의 실시예들에서, 바운딩 박스들의 시각 식별자들(및/또는 상호작용 결과 문서(1200)에서 라벨들 또는 다른 시각 식별자들)은 오버레이 컬러, 오버레이 패턴, 라벨 배경 색, 라벨 배경 패턴, 라벨 폰트 색, 및 바운딩 박스 경계 색과 같은 시각적으로 구별되는 방식으로 표시하기 위해 포맷된다. 특정 인식된 엔티티들에 대한 타입 코딩은 도 13에서 바운딩 박스들에 관하여 도시되지만, 타입에 의한 코딩은 또한 라벨들인 시각 식별자들에 적용될 수 있다.

[0137] 도 14는 도 11의 시각 질의(1102)의 각각의 서브-부분들의 시각 식별자들인 라벨들(1402)을 가진 상호작용 결과 문서(1200)의 스크린 샷을 가진 클라이언트 디바이스(102)를 도시한다. 라벨 시각 식별자들(1402) 각각은 대응하는 검색 결과들의 서브세트에 대한 사용자 선택가능한 링크를 포함한다. 임의의 실시예들에서, 선택가능한 링크는 라벨(1402)의 영역 내에 디스플레이되는 설명 텍스트에 의해 식별된다. 임의의 실시예들은 하나의 라벨(1402) 내에 복수의 링크들을 포함한다. 예를 들어, 도 14에서, 음료를 마시는 여성의 이미지 위에 맵도는 라벨은 여성에 대한 얼굴 인식 결과들에 대한 링크 및 상기 특정 픽처(예를 들어, 동일한 픽처를 사용하는 다른 물건들 또는 광고들의 이미지들)에 대한 이미지 인식 결과들에 대한 링크를 포함한다.

[0138] 도 14에서, 라벨들(1402)은 상호작용 결과 문서의 자신들의 각각의 서브-부분들 위에 위치되는 텍스트를 가진 부분적으로 투명한 영역들로서 디스플레이된다. 다른 실시예들에서, 각각의 라벨은 근처에 포지셔닝되지만 상호작용 결과 문서의 자신의 각각의 서브-부분 위에 위치되지 않는다. 임의의 실시예들에서, 라벨들은 도 13을 참조하여 논의되는 동일한 방식으로 타입에 의해 코딩된다. 임의의 실시예들에서, 사용자는 라벨(1302)의 주변 또는 예지들에 의해 윤곽이 나타내어진 공간 내에 활성 영역을 탭핑함으로써 라벨(1302)에 대응하는 특정 서브-부분에 대응하는 검색 결과들의 디스플레이를 활성화한다. 도면들(12a 및 12b)의 바운딩 박스들을 참조하여 상기 논의된 동일한 프리뷰 및 선택 기능들은 또한 라벨들(1402)인 시각 식별자들에 적용한다.

[0139] 도 15는 결과 리스트(1500)와 동시에 디스플레이되는 오리지널 시각 질의(1102) 및 상호작용 결과 문서(1200)의 스크린 샷을 도시한다. 임의의 실시예들에서, 상호작용 결과 문서(1200)는 도 12-14에 도시된 바와 같이 자기 자신에 의해 디스플레이된다. 다른 실시예들에서, 상호작용 결과 문서(1200)는 도 15에 도시된 바와 같이 오리지널 시각 질의와 동시에 디스플레이된다. 임의의 실시예들에서, 시각 질의 결과들의 리스트(1500)는 오리지널 시각 질의(1102) 및/또는 상호작용 결과 문서(1200)와 함께 동시에 디스플레이된다. 클라이언트 시스템의 타입 및 디스플레이(706)상의 빈 공간의 양은 결과들의 리스트(1500)가 상호작용 결과 문서(1200)와 동시에 디스플레이되는지를 결정할 수 있다. 임의의 실시예들에서, 클라이언트 시스템(102)은 (시각 질의 서버 시스템에 제출된 시각 질의에 응답하여) 결과들의 리스트(1500) 및 상호작용 결과 문서(1200) 둘 다를 수신할 뿐만 아니라, 사용자가 상호작용 결과 문서(1200) 아래로 스크롤할 때 결과의 리스트(1500)를 디스플레이한다. 이러한 실시예들의 일부에서, 클라이언트 시스템(102)은, 결과들의 리스트(1500)가 시각 질의에 응답하여 클라이언트 시스템(102)에 의해 수신되고 그 다음에 클라이언트 시스템(102)에서 로컬하게 저장되기 때문에 다시 서버에 질의할 필요가 없이 사용자 선택된 시각 식별자(1202/1402)에 대응하는 결과들을 디스플레이한다.

[0140] 임의의 실시예들에서, 결과들의 리스트(1500)는 카테고리들(1502)로 정리된다. 각 카테고리는 적어도 하나의 결과(1503)를 포함한다. 임의의 실시예들에서, 카테고리 타이틀들은 결과들(1503)로부터 자신들을 구별하도록 강조된다. 카테고리들(1502)은 자신들의 계산된 카테고리 가중치에 따라 순서화된다. 임의의 실시예들에서, 카테고리 가중치는 그 카테고리에서 최상위 N개의 결과들의 가중치들의 조합이다. 따라서, 더 관련된 결과들을 산출할 것 같은 카테고리가 먼저 디스플레이된다. 둘 이상의 카테고리(1502)가 동일한 인식된 엔티티에 대하여

리턴되는 (도 15에 도시된 얼굴 이미지 인식 매칭 및 이미지 매칭과 같은) 실시예들에서, 먼저 디스플레이되는 카테고리는 더 높은 카테고리 가중치를 가진다.

[0141] 도 3에 관하여 설명되는 바와 같이, 임의의 실시예들에서, 상호작용 결과 문서(1200)에서 선택가능한 링크가 클라이언트 시스템(102)의 사용자에게 의해 선택될 때, 커서는 적절한 카테고리(1502)로 또는 상기 카테고리에서의 제 1 결과(1503)로 자동으로 이동할 것이다. 대안적으로, 상호작용 결과문서에서 선택가능한 링크가 클라이언트 시스템(102)의 사용자에게 의해 선택될 때, 결과들의 리스트(1500)는 선택된 링크에 관련된 카테고리 또는 카테고리들이 먼저 디스플레이되도록 재순서화된다. 이는 예를 들어, 선택가능한 링크들을 대응하는 검색 결과들을 식별하는 정보로 코딩하거나, 대응하는 선택가능한 링크들을 나타내거나 또는 대응하는 결과 카테고리들을 나타내도록 검색 결과들을 코딩함으로써 달성된다.

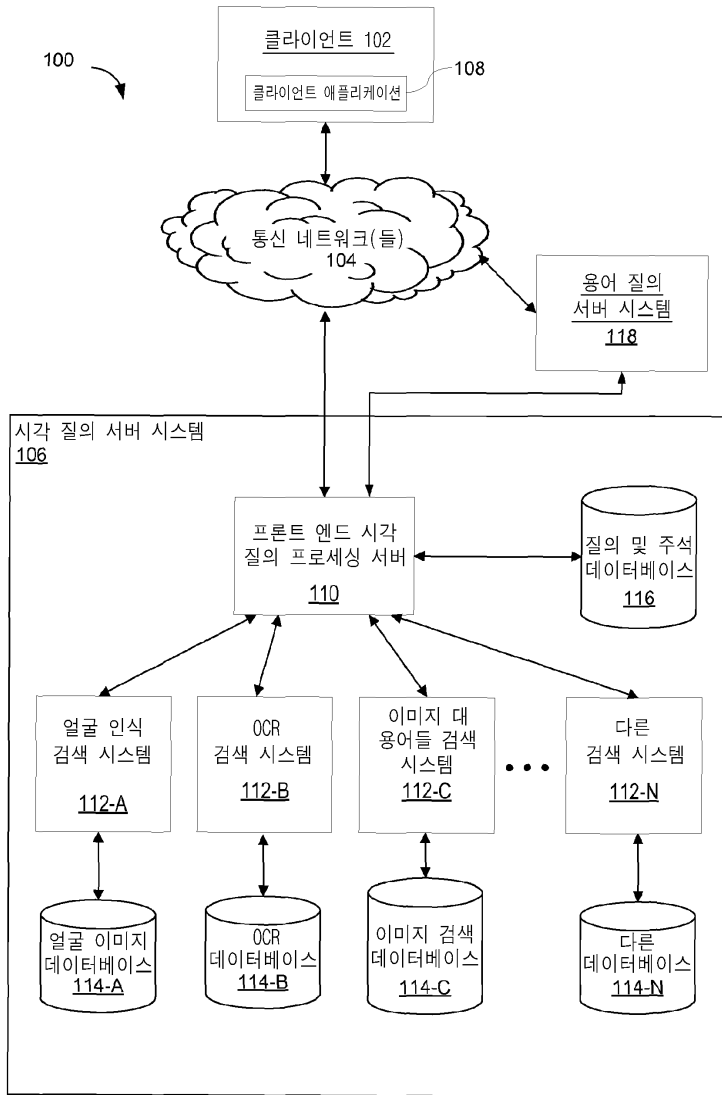
[0142] 임의의 실시예들에서, 검색 결과들의 카테고리들은 이러한 검색 결과들을 산출하는 이미지에 의한 질의 검색 시스템에 대응한다. 예를 들어, 도 15에서, 카테고리들의 일부는 물건 매칭(1506), 로고 매칭(1508), 얼굴 인식 매칭(1510), 이미지 매칭(1512)이다. 오리지널 시각 질의(1102) 및/또는 상호작용 결과 문서(1200)는 질의(1504)와 같은 카테고리 타이틀과 유사하게 디스플레이될 수 있다. 유사하게, 용어 질의 서버에 의해 수행되는 임의의 용어 검색으로부터의 결과들은 또한 웹 결과들(1514)과 같은, 분리된 카테고리로서 디스플레이될 수 있다. 다른 실시예들에서, 시각 질의에서 둘 이상의 엔티티는 동일한 이미지에 의한 질의 검색 시스템으로부터의 결과들을 산출할 것이다. 예를 들어, 시각 질의는 얼굴 인식 검색 시스템으로부터 분리된 결과들을 리턴할 수 있는 두 상이한 얼굴들을 포함할 수 있다. 따라서, 임의의 실시예들에서, 카테고리들(1502)은 검색 시스템에 의해서보다 오히려 인식된 엔티티에 의해 분할된다. 임의의 실시예들에서, 인식된 엔티티의 이미지는, 이미지 검색 시스템에 의한 동일한 질의에 의해 두개의 결과들이 산출되더라도 상기 인식된 엔티티에 대한 결과들이 다른 인식된 엔티티에 대한 결과들과 구별가능하도록, 인식되는 엔티티 카테고리 헤더(1502)에서 디스플레이된다. 예를 들어, 도 15에서, 물건 매칭 카테고리(1506)는 두 개의 엔티티 물건 엔티티들 및 또한 두 개의 엔티티 카테고리들(1502)을 포함한다 - 박스 포장된 물건(1516) 및 병에 담긴 물건(1518), 각각은 복수의 대응하는 검색 결과들(1503)을 가진다. 임의의 실시예들에서, 카테고리들은 인식된 엔티티들 및 이미지에 의한 질의 시스템의 타입에 의해 분할될 수 있다. 예를 들어, 도 15에서, 물건 매칭 카테고리 물건 하에서 관련 결과들을 리턴한 두 개의 분리된 엔티티들이 존재한다.

[0143] 임의의 실시예들에서, 결과들(1503)은 썸네일 이미지들을 포함한다. 예를 들어, 도 15에서 얼굴 인식 매칭 결과들에 대하여 도시된 바와 같이, "여배우 X" 및 "소셜 네트워크 친구 Y"에 대한 얼굴 매칭들의 픽처들의 작은 버전들(또한 썸네일 이미지들로 지칭됨)은 이미지에서 사람의 명칭과 같은 임의의 텍스트 설명과 함께 디스플레이된다.

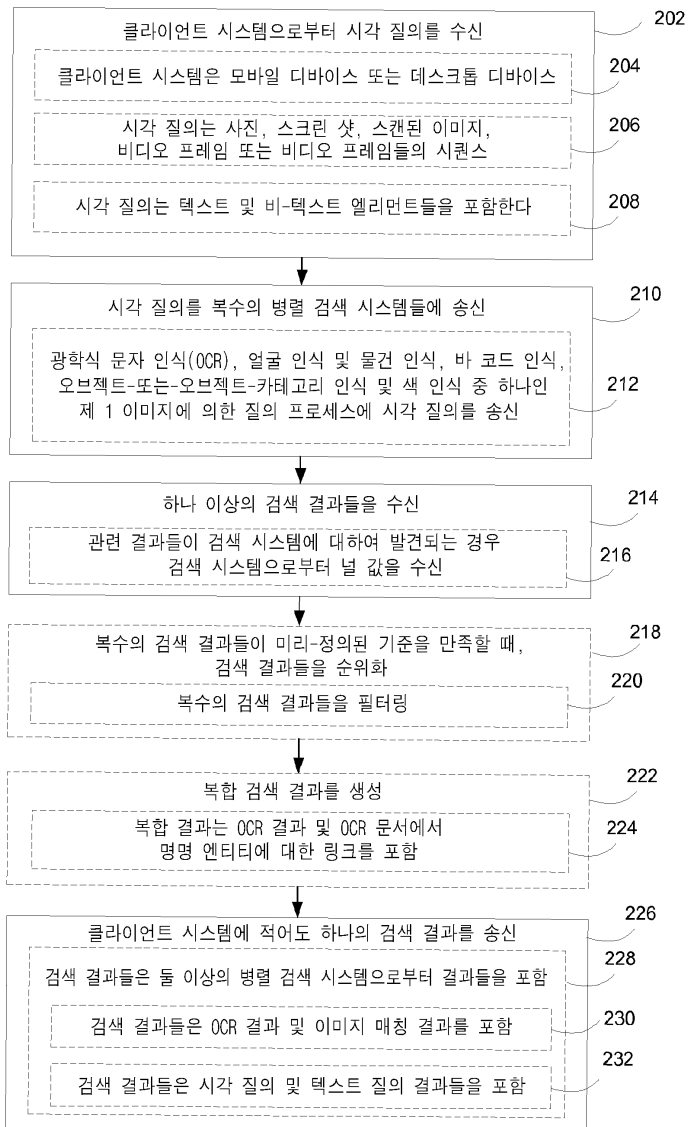
[0144] 설명의 목적을 위해 상기 설명은 특유의 실시예들을 참조하여 설명되었다. 하지만, 상기 예시적인 논의들은 개시된 정확한 형태들로 본 발명을 제한하거나 또는 모두 기재하고자 의도되지 않는다. 많은 수정들 및 변경들이 상기 교시들의 관점에서 가능하다. 실시예들은 본 발명의 원리들 및 이의 실제적 적용들을 가장 잘 설명하기 위해 설명되고 선택되었고 이에 의해 당업자들을 고려되는 특정 사용에 적합한 다양한 수정들을 가진 다양한 실시예들 및 본 발명을 가장 잘 이용하도록 할 수 있다.

도면

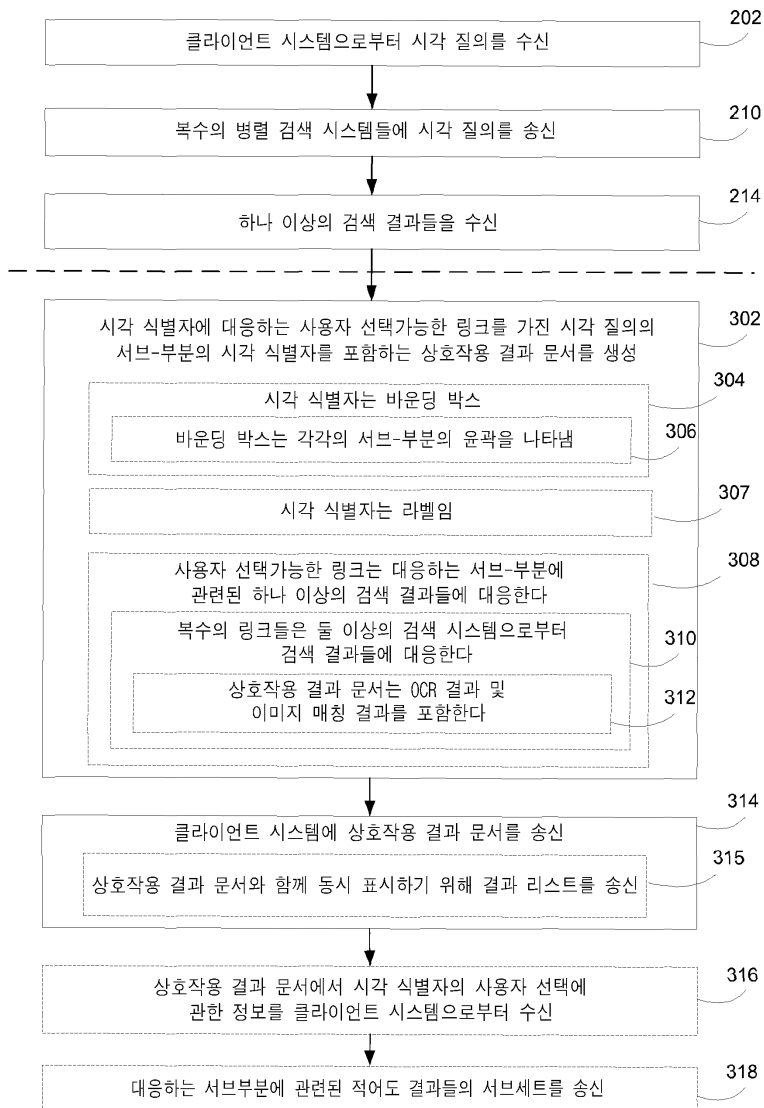
도면1



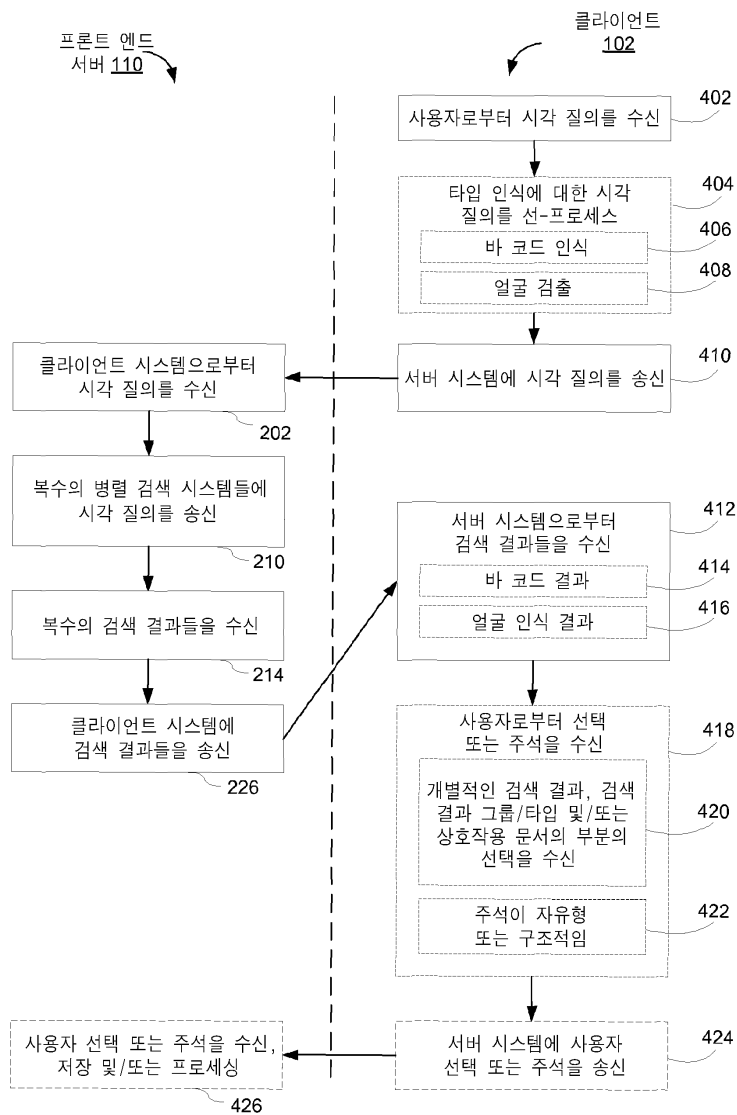
도면2



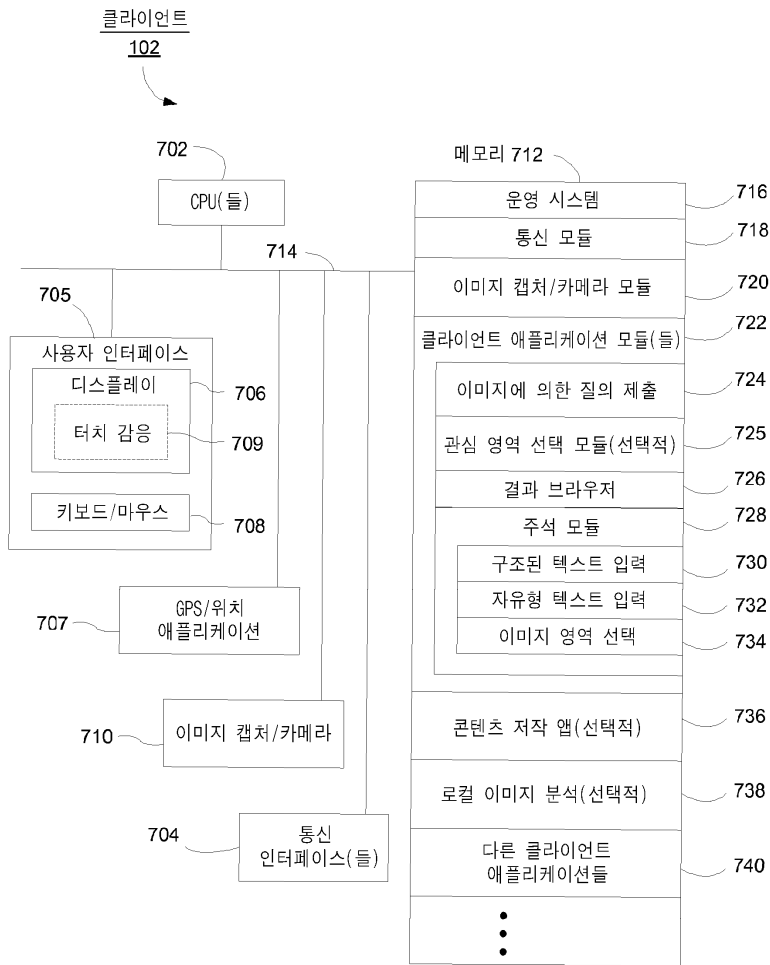
도면3



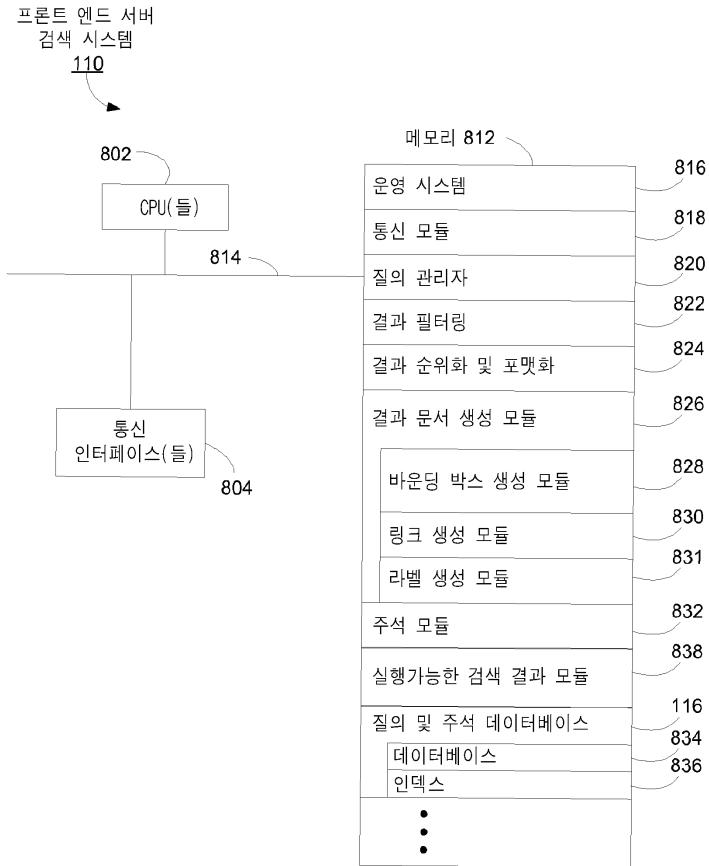
도면4



도면5

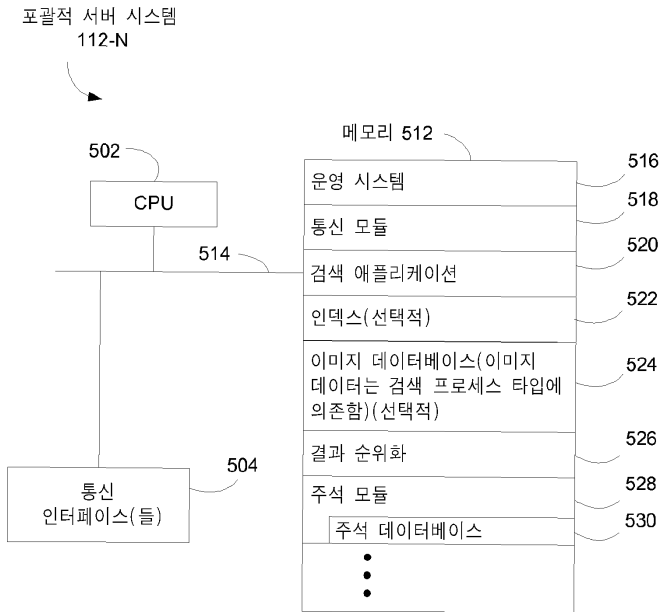


도면6

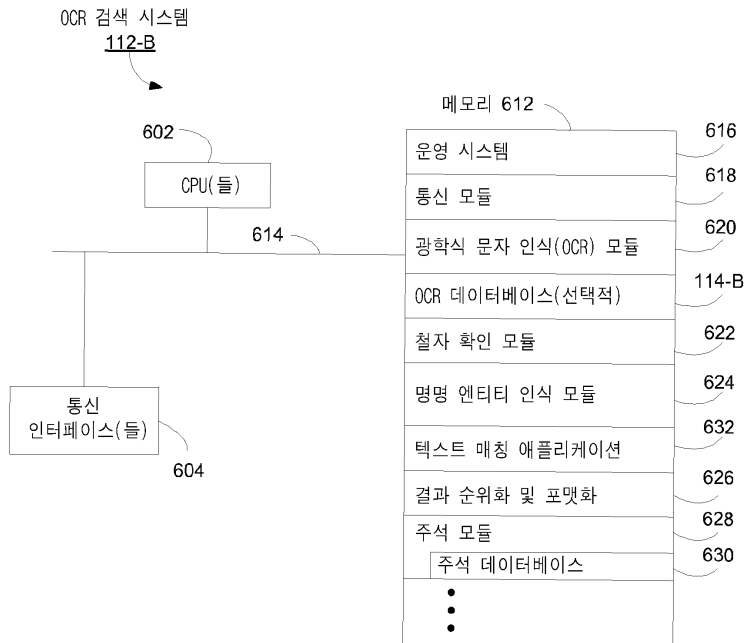


도면7

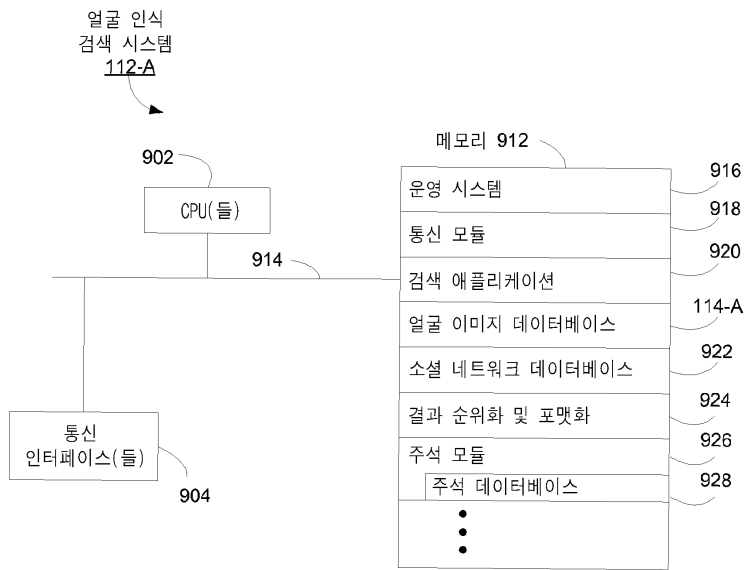
서버
102



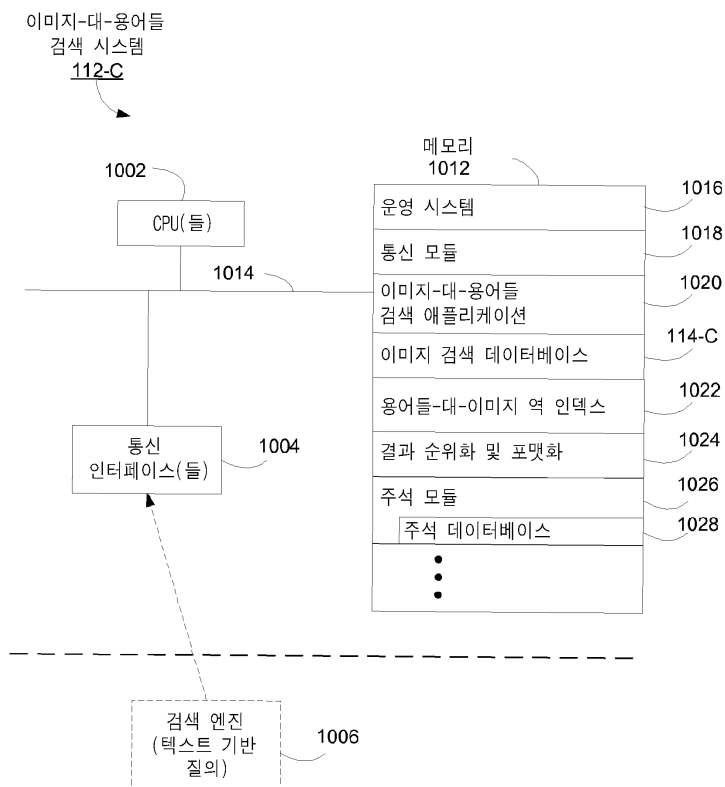
도면8



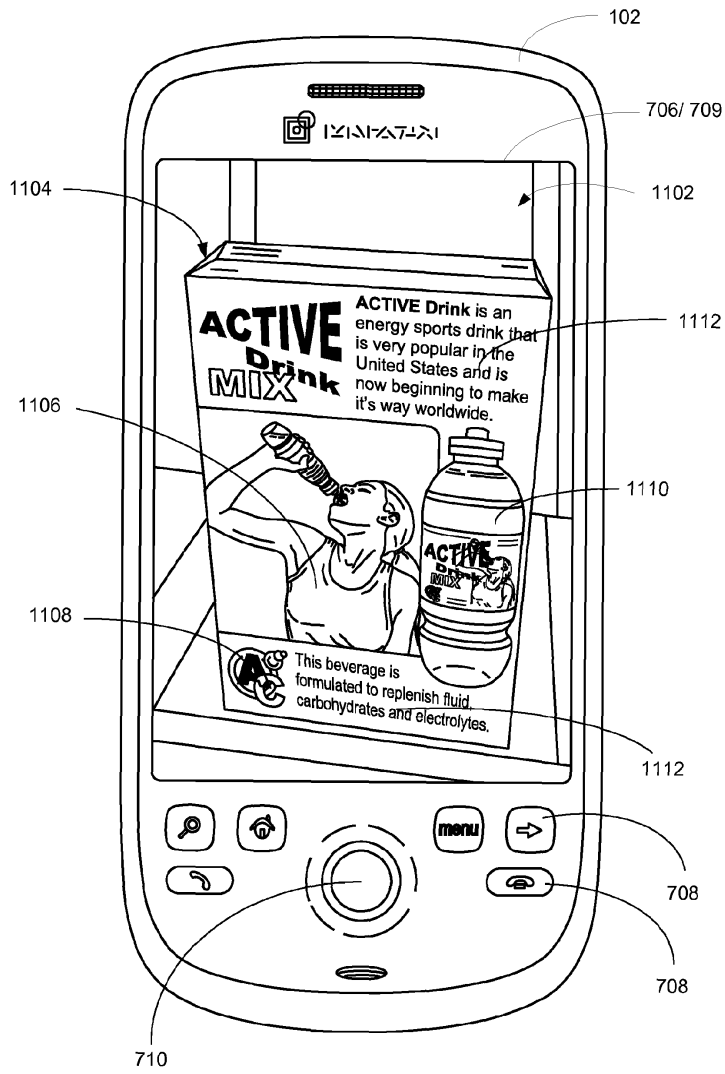
도면9



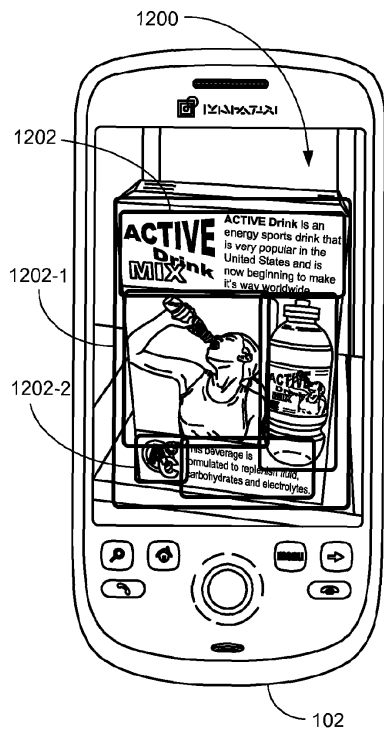
도면10



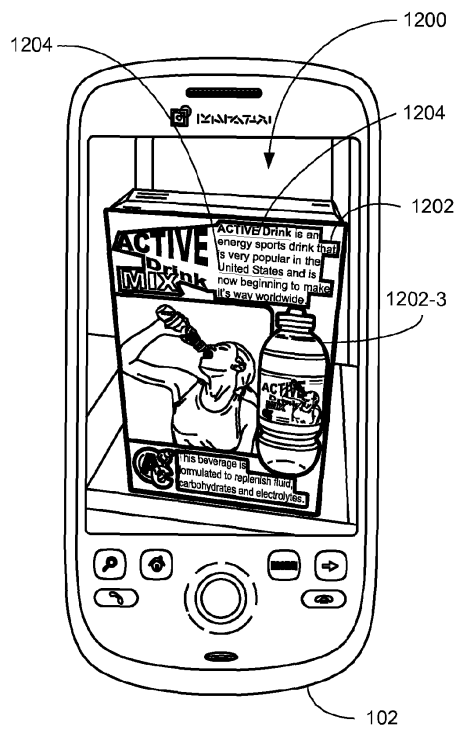
도면11



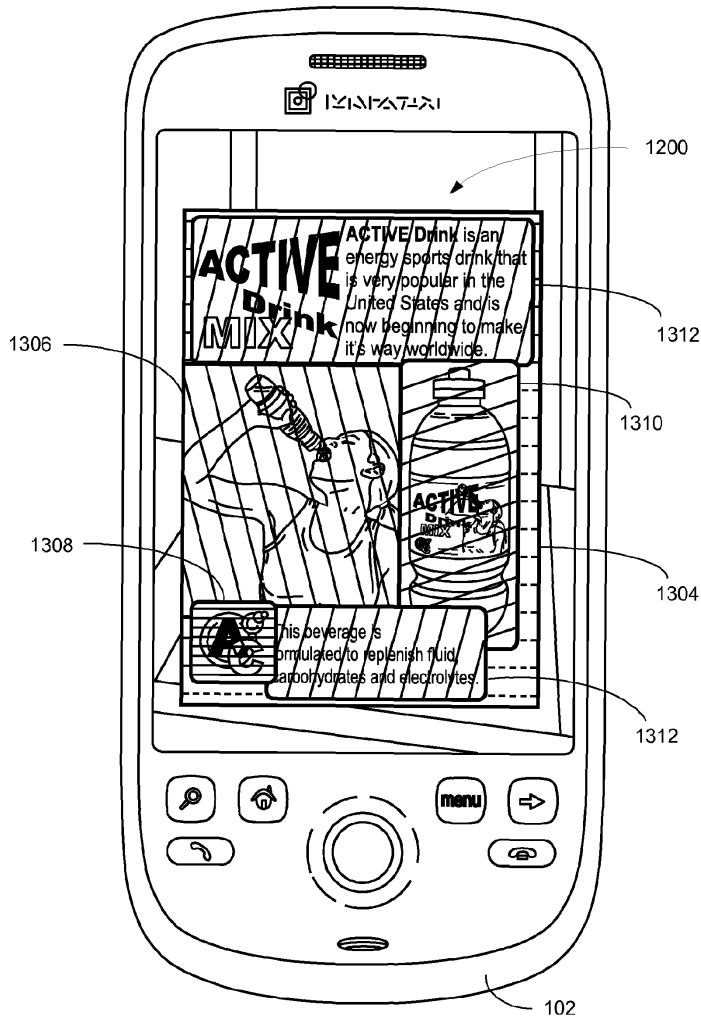
도면12a



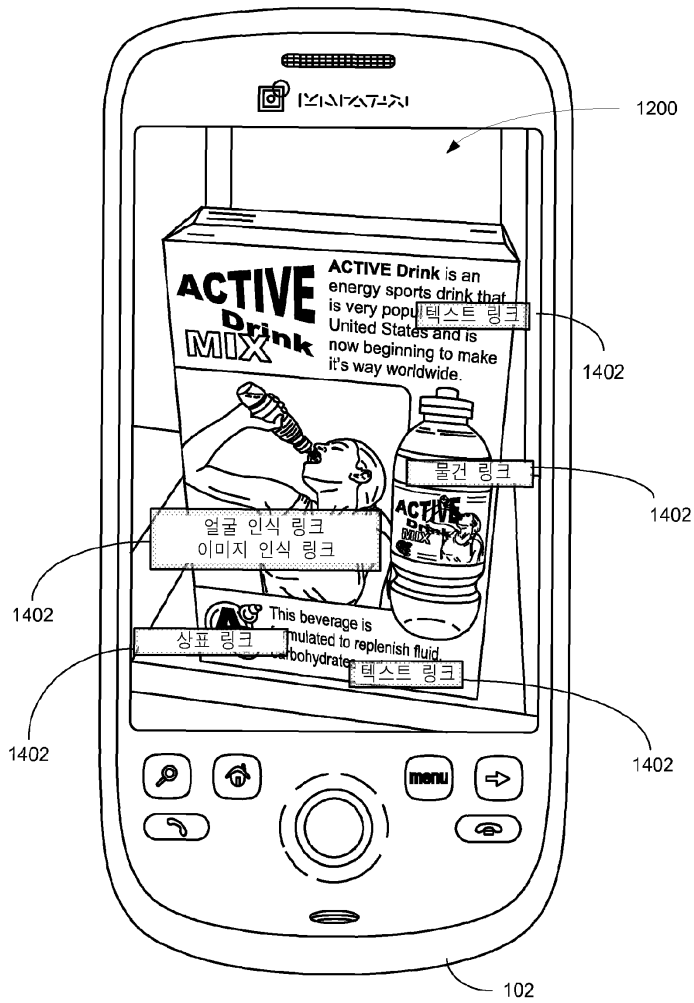
도면12b



도면13



도면14



도면15

