



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년04월07일
(11) 등록번호 10-1027848
(24) 등록일자 2011년03월31일

(51) Int. Cl.

G06F 17/26 (2006.01) G06F 17/30 (2006.01)

(21) 출원번호 10-2004-0065838

(22) 출원일자 2004년08월20일

심사청구일자 2009년08월12일

(65) 공개번호 10-2005-0035066

(43) 공개일자 2005년04월15일

(30) 우선권주장

10/683,807 2003년10월10일 미국(US)

(56) 선행기술조사문헌

JP11338864 A

KR1019990047854 A

KR1020010105241 A

전체 청구항 수 : 총 24 항

(73) 특허권자

마이크로소프트 코포레이션

미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이

(72) 발명자

라트나파크히, 애드웨이트

미국 98052 워싱턴주 레드몬드 원 마이크로소프트
웨이 마이크로소프트 코포레이션 내

고로드니스키, 보리스

미국 98052 워싱턴주 레드몬드 원 마이크로소프트
웨이 마이크로소프트 코포레이션 내

(뒷면에 계속)

(74) 대리인

주성민, 이중희, 백만기

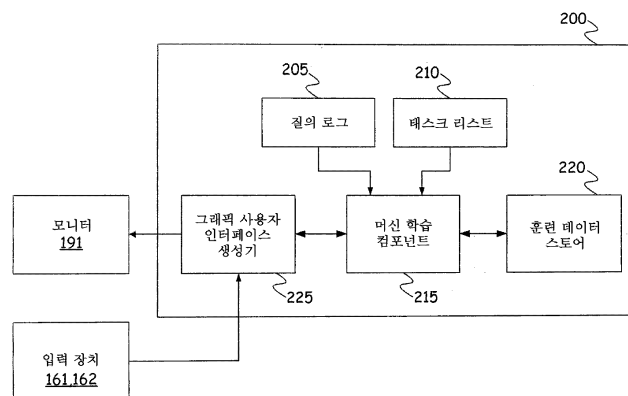
심사관 : 유병철

(54) 컴퓨터를 이용한 질의-태스크 벌크 매핑 시스템 및 방법

(57) 요약

주석 시스템은 검색 컴포넌트를 훈련시키기 위한 훈련 데이터를 획득하기 위해 사용자가 많은 수의 질의를 태스크에 매핑하는 것을 돕는다. 주석 시스템은 검색 엔진에 이전에 전송되었던 대량의 질의를 갖는 질의 로그를 포함한다. 복수의 가능한 태스크를 갖는 태스크 리스트가 저장된다. 머신 학습 컴포넌트는 질의 로그 데이터 및 태스크 리스트 데이터를 처리한다. 질의 로그에 대응하는 복수의 질의 엔트리 각각에 대해, 머신 학습 컴포넌트는 훈련 데이터의 함수로서 잠재적인 질의-태스크 매핑에 대한 최선의 추측 태스크를 제안한다. 그래픽 사용자 인터페이스 생성 컴포넌트는 질의 로그 내의 복수의 질의 엔트리를 디스플레이하되 디스플레이된 복수의 질의 엔트리 각각을 그의 대응하는 제안된 최선의 추측 태스크와 관련시키는 방식으로 디스플레이하도록 구성되어 있다.

대표도



(72) 발명자

나란조, 펠리페루이스

미국 98052 워싱턴주 레드몬드 원 마이크로소프트
웨이 마이크로소프트 코포레이션 내

라그노, 로버트존

미국 98052 워싱턴주 레드몬드 원 마이크로소프트
웨이 마이크로소프트 코포레이션 내

특허청구의 범위

청구항 1

검색 컴포넌트(search component)를 훈련(training)시키기 위한 훈련 데이터를 갱신하기 위해 사용자가 질의들을 질의 결과들에 벌크 매핑(bulk mapping)하는 것을 돕기 위한 주석 시스템(annotating system)으로서,

검색 엔진에 이전에 전송되었던 질의들을 포함하는 질의 로그(205);

복수의 가능한 질의 결과를 포함하는 질의 결과 리스트(210);

상기 질의 로그로부터의 질의를 상기 질의 결과 리스트로부터의 질의 결과에 매핑하는 질의-질의 결과 매핑에 대한 예를 나타내는 훈련 데이터를 포함하는 훈련 데이터 스토어(220);

상기 훈련 데이터의 함수로서 최선의 추측 질의-질의 결과 매핑(query-to-query result mapping)을 제안하는 머신 학습 컴포넌트(machine learning component; 215); 및

상기 제안된 최선의 추측 질의-질의 결과 매핑들을 나타내는 방식으로 상기 질의 결과 리스트 내의 복수의 질의 결과들 중 적어도 일부 및 상기 질의 로그 내의 복수의 질의 중 적어도 일부를 디스플레이하도록 구성된 그래픽 사용자 인터페이스 생성 컴포넌트(225)를 포함하고,

상기 그래픽 사용자 인터페이스 생성 컴포넌트(225)는 또한 사용자로부터, 사용자가 특정의 질의를 특정의 질의 결과에 매핑하고자 함을 나타내는 제1 유형의 입력을 수신하도록 구성되고,

상기 그래픽 사용자 인터페이스를 통해 상기 제1 유형의 입력의 수신 시에, 상기 머신 학습 컴포넌트는 특정의 질의 엔트리를 상기 특정의 질의 결과에 매핑함으로써 상기 훈련 데이터를 갱신하는 주석 시스템.

청구항 2

제1항에 있어서,

상기 그래픽 사용자 인터페이스 생성 컴포넌트(225)는 상기 질의 로그(205) 내의 복수의 질의를 디스플레이하되 그 디스플레이된 복수의 질의 각각을 그의 대응하는 제안된 최선의 추측 질의 결과와 관련시키는 방식으로 디스플레이하도록 구성되는 주석 시스템.

청구항 3

제1항 또는 제2항에 있어서,

상기 그래픽 사용자 인터페이스 생성 컴포넌트(225)는 또한 사용자가 잠재적인 매핑을 위해 상기 특정의 질의에 대한 그 다음 최선의 추측 질의 결과들의 리스트를 보고자 할 때, 사용자로부터 제2 유형의 입력을 수신하도록 구성되는 주석 시스템.

청구항 4

제1항 또는 제2항에 있어서,

상기 질의 로그(205)에 포함된 상기 복수의 질의 각각이 질의 묶음(query bundle)이고,

각각의 질의 묶음은 함께 묶음으로 되어 있는(bundled) 복수의 단일 질의를 나타내는 주석 시스템.

청구항 5

제4항에 있어서,

상기 그래픽 사용자 인터페이스를 통해 상기 제1 유형의 입력을 수신하면, 상기 머신 학습 컴포넌트(215)는 상기 질의 묶음으로 표현된 상기 복수의 단일 질의 각각을 상기 제안된 최선의 추측 질의 결과에 매핑함으로써 상기 훈련 데이터를 갱신하는 주석 시스템.

청구항 6

제4항에 있어서,

각각의 질의 묶음에 대한 상기 제안된 최선의 추측은 상기 질의 묶음으로 표현된 상기 복수의 단일 질의 각각에 대한 최선의 추측들의 가중 평균인 주석 시스템.

청구항 7

제4항에 있어서,

상기 그래픽 사용자 인터페이스 생성 컴포넌트(225)는 또한 사용자가 특정의 질의 묶음으로 표현된 단일 질의들의 리스트를 보고자 할 때, 사용자로부터 제2 유형의 입력을 수신하도록 구성되어 있으며,

상기 제2 유형의 입력의 수신에 응답하여, 상기 그래픽 사용자 인터페이스 생성 컴포넌트(225)는 상기 질의 묶음으로 표현된 단일 질의들의 리스트를 디스플레이하도록 구성되는 주석 시스템.

청구항 8

제1항에 있어서,

상기 그래픽 사용자 인터페이스 생성 컴포넌트(225)는 또한 상기 질의 결과 리스트를 디스플레이하도록 구성되는 주석 시스템.

청구항 9

제8항에 있어서,

상기 그래픽 사용자 인터페이스 생성 컴포넌트(225)는 또한 사용자가 특정의 질의 엔트리를 상기 디스플레이된 질의 결과 리스트 중의 상기 제안된 최선의 추측 질의 결과와 다른, 특정의 질의 결과에 매핑하고자 할 때, 사용자로부터 제2 유형의 입력을 수신하도록 구성되어 있으며,

상기 그래픽 사용자 인터페이스를 통해 제2 유형의 입력을 수신하면, 상기 머신 학습 컴포넌트(215)는 상기 특정의 질의 엔트리를 상기 디스플레이된 질의 결과 리스트로부터의 상기 특정의 질의 결과에 매핑함으로써 상기 훈련 데이터를 갱신하도록 구성되는 주석 시스템.

청구항 10

제1항에 있어서,

상기 특정의 질의 엔트리를 특정의 질의 결과에 매핑함으로써 상기 훈련 데이터를 갱신하면, 상기 머신 학습 컴포넌트(215)는 상기 갱신된 훈련 데이터의 함수로서 상기 복수의 질의들 중 나머지 질의들 각각에 대한 최선의 추측 질의 결과를 자동적으로 갱신하도록 구성되는 주석 시스템.

청구항 11

제10항에 있어서,

상기 머신 학습 컴포넌트(215)는 나이브 베이스 분류자(Naive Bayes classifier)인 주석 시스템.

청구항 12

제1항에 있어서,

상기 머신 학습 컴포넌트(215)는 상기 질의 결과 리스트 내의 복수의 가능한 질의 결과 각각에 대해, 상기 훈련 데이터의 함수로서 질의-질의 결과 매핑에 대해 가능성 있는 상기 질의 로그(205)로부터의 질의들의 리스트를 제안하도록 구성되는 주석 시스템.

청구항 13

검색 컴포넌트를 훈련시키기 위한 훈련 데이터를 갱신하기 위해 사용자가 질의들을 질의 결과들에 벌크 매핑하는 것을 돕는 방법으로서,

검색 엔진에 이전에 전송되었던 질의들을 포함하는 질의 로그를 획득하는 단계;

복수의 가능한 질의 결과를 포함하는 질의 결과 리스트를 획득하는 단계(402);

상기 질의 로그로부터의 질의를 상기 질의 결과 리스트로부터의 질의 결과에 매핑하는 질의-질의 결과 매핑에 대한 예를 나타내는 훈련 데이터를 획득하는 단계(404);

상기 질의 로그에 포함되는 질의 각각에 대해, 잠재적인 질의-질의 결과 매핑을 위한 최선의 추측 질의 결과를 결정하는 단계(406) - 상기 최선의 추측 질의 결과는 머신 학습 컴포넌트를 사용하여 상기 훈련 데이터의 함수로서 결정됨 - ; 및

상기 질의 로그 내의 복수의 질의 중 적어도 일부를 디스플레이하되 상기 디스플레이된 질의 각각을 그의 대응하는 제안된 최선의 추측 질의 결과와 관련시키는 방식으로 디스플레이하는 단계(408);

사용자로부터, 사용자가 특정의 질의를 특정의 질의 결과에 매핑하고자 함을 나타내는 제1 유형의 입력을 수신하는 단계(452), 및

상기 제1 유형의 입력을 수신하면 특정의 질의 엔트리를 상기 특정의 질의 결과에 매핑함으로써 상기 머신 학습 컴포넌트를 사용하여 상기 훈련 데이터를 갱신하는 단계(454)

를 포함하는, 벌크 매핑을 돕는 방법.

청구항 14

제13항에 있어서,

사용자가 잠재적인 매핑을 위해 상기 특정의 질의에 대한 그 다음 최선의 추측 질의 결과들의 리스트를 보고자 할 때, 사용자로부터 제2 유형의 입력을 수신하는 단계(502), 및

상기 머신 학습 컴포넌트를 사용하여 상기 훈련 데이터의 함수들로서 결정된, 상기 특정 질의에 대한 그 다음 최선의 추측 질의 결과들의 리스트를 디스플레이하는 단계(504)

를 더 포함하는, 벌크 매핑을 돕는 방법.

청구항 15

제13항에 있어서,

상기 질의 로그에 포함되는 복수의 질의 각각은 질의 묶음이고,

각각의 질의 묶음은 함께 묶음으로 되어 있는 복수의 단일 질의를 나타내는, 벌크 매핑을 돕는 방법.

청구항 16

제15항에 있어서,

상기 제1 유형의 입력을 수신하면, 상기 질의 묶음으로 표현된 복수의 단일 질의 각각을 상기 제안된 최선의 추측 질의 결과에 매핑함으로써 상기 머신 학습 컴포넌트를 사용하여 상기 훈련 데이터를 갱신하는 단계

를 더 포함하는, 벌크 매핑을 돕는 방법.

청구항 17

제15항에 있어서,

사용자가 특정의 질의 묶음으로 표현된 단일 질의들의 리스트를 보고자 할 때, 사용자로부터 제2 유형의 입력을 수신하는 단계(552), 및

상기 제2 유형의 입력의 수신에 응답하여 상기 질의 묶음으로 표현된 단일 질의들의 리스트를 디스플레이하는 단계(554)

를 더 포함하는, 벌크 매핑을 돕는 방법.

청구항 18

제13항에 있어서,

태스크 리스트를 디스플레이하는 단계를 더 포함하는, 벌크 매핑을 돕는 방법.

청구항 19

제18항에 있어서,

사용자가 상기 특정의 질의를 상기 디스플레이된 질의 결과 리스트 중의 상기 제안된 최선의 추측 질의 결과와 다른, 특정의 질의 결과에 매핑하고자 할 때, 사용자로부터 제2 유형의 입력을 수신하는 단계(602), 및

상기 제2 유형의 입력을 수신하면, 상기 특정의 질의를 디스플레이된 태스크 리스트로부터의 상기 특정의 질의 결과에 매핑함으로써 상기 머신 학습 컴포넌트를 사용하여 상기 훈련 데이터를 갱신하는 단계(604)

를 더 포함하는, 벌크 매핑을 돕는 방법.

청구항 20

제13항에 있어서,

상기 특정의 질의를 질의 결과에 매핑함으로써 상기 훈련 데이터를 갱신하면, 상기 갱신된 훈련 데이터의 함수로서 상기 복수의 질의들 중 나머지 질의들 각각에 대한 최선의 추측 질의 결과를 상기 머신 학습 컴포넌트를 사용하여 자동적으로 갱신하는 단계를 더 포함하는, 벌크 매핑을 돕는 방법.

청구항 21

제20항에 있어서,

상기 머신 학습 컴포넌트는 나이브 베이스 분류자(Naive Bayes classifier)인, 벌크 매핑을 돕는 방법.

청구항 22

제13항에 있어서,

상기 질의 로그에 포함되는 복수의 질의 각각에 대해 최선의 추측 질의 결과를 결정하는 단계는 상기 질의 결과 리스트 내의 각각의 질의 결과에 대해, 상기 머신 학습 컴포넌트를 사용하여 상기 훈련 데이터의 함수로서 질의-질의 결과 매핑에 대해 가능성 있는 상기 질의 로그로부터의 질의들의 리스트를 결정하는 단계를 더 포함하는, 벌크 매핑을 돕는 방법.

청구항 23

제13항에 있어서,

상기 머신 학습 컴포넌트는 분류자 모델을 이용하고,

상기 분류자 모델을 서버로부터 다운로드하는 단계

를 더 포함하는, 벌크 매핑을 돕는 방법..

청구항 24

제13항의 단계들을 구현하기 위한 컴퓨터 실행가능한 명령어들을 포함하는 컴퓨터 판독가능 기록 매체.

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

- [0016] 본 발명은 사용자 인터페이스의 머신 학습 컴포넌트를 훈련시키는 것에 관한 것이다. 보다 상세하게는, 본 발명은 질의를 태스크로 매핑함으로써 훈련 데이터를 획득하는 것에 관한 것이다.
- [0017] 자연스런 사용자 인터페이스(natural user interface, NUI)는 자연어 질의를 받아들이며, 그에 응답하여 의도한 질의 주제에 대응할 가능성이 높은 결과의 리스트를 반환한다. 이 결과들은 일반적으로 질의에 대한 대답이기를 바라는 태스크, 문서, 파일, 이메일, 또는 기타 항목(이 모두를 본 명세서에서는 총괄하여 태스크라고 함)을 포함한다. 질의에 대한 결과를 생성하는 유망한 기술은 머신 학습 기술이다. 머신 학습 알고리즘은 통계적 데이터를 사용하여 특정의 질의에 대해 원하는 결과를 예측한다. 머신 학습 알고리즘을 사용하여, 통계적 데이터는 초기 훈련 이후에 사용자로부터의 추가 질의에 대한 결과에 기초하여 끊임없이 또는 빈번히 갱신될 수 있다.
- [0018] 자연스런 사용자 인터페이스가 의미있는 결과를 제공할 것으로 기대하면서 이 인터페이스와 함께 사용하기 위해 임의의 머신 학습 알고리즘이 제공될 수 있기 전에, 이 알고리즘은 정확한 주석달린 데이터(annotated data)로 "훈련"되어야만 한다. 즉, 이 알고리즘은 대규모의 질의-태스크 매핑 리스트로부터의 통계를 나타내는 훈련 데이터를 필요로 한다. 자연스런 사용자 인터페이스 및 대응하는 머신 학습 알고리즘이 고객에게 배포되어질 때, 그 배포 이전에 머신 학습 알고리즘이 정확한 주석달린 데이터로 훈련되어야만 하는 것이 더욱 필수적이다. 예를 들어, 머신 학습 알고리즘의 한가지 출력 유형이 "프린터 설치" 또는 "프린터 문제 해결" 등의 태스크 리스트인 경우, 머신 학습 알고리즘은 이들 태스크가 그에 대한 원하는 결과인 것인 자연어 질의의 예들을 나타내는 데이터를 필요로 한다.
- [0019] 자연스런 사용자 인터페이스의 머신 학습 알고리즘의 정확성을 향상시키기 위해, 훈련 데이터는 아주 대규모의 질의-태스크 매핑에 리스트를 나타내야만 한다. 종래에는, 많은 수의 질의-태스크 매핑이 검색 엔진으로 전송되는 아주 많은 수의 실제 질의를 포함하는 질의 로그(query log)를 획득함으로써 생성되었다. 예를 들어, 질의 로그는 일반적으로 10,000개 이상의 질의를 포함한다. 이어서, 사용자나 저작자는 이들 질의를 하나씩 거쳐가면서 수작업으로 그에 주석을 붙인다(그와 특정 태스크를 관련시킨다).
- [0020] 태스크에 대한 질의에 주석을 붙이는 통상의 방법은 스프레드 시트 데이터베이스의 첫번째 열에 각각의 질의를 나타내고 이 스프레드 시트의 두번째 열의 동일 행에 그의 대응하는 의도한 태스크를 나타내는 것이다. 따라서, 이 과정은 많은 수고를 필요로 하고 시간이 많이 걸린다. 게다가, 그로부터 선택하게 될 대규모 잠재적 태스크 리스트가 주어지면, 어느 태스크에 특정의 질의로 주석을 붙일 것인지 선택하는 것은 아주 성가신 일이 된다.
- 발명이 이루고자 하는 기술적 과제**
- [0021] 따라서, 훈련 데이터를 획득하기 위해 보다 빠르고 보다 정확한 질의-태스크 매핑을 용이하게 하는 데 사용될 수 있는 시스템 또는 방법은 기술 분야에 상당한 개선이 될 것이다.

발명의 구성 및 작용

- [0022] 주석 시스템은 검색 컴포넌트를 훈련시키기 위한 훈련 데이터를 획득하기 위해 사용자가 많은 수의 질의를 태스크에 매핑하는 것을 돕는다. 이 주석 시스템은 검색 엔진으로 이전에 전송되었던 대량의 질의를 갖는 질의 로그를 포함한다. 복수의 가능한 태스크를 갖는 태스크 리스트가 저장된다. 머신 학습 컴포넌트는 질의 로그 데이터 및 태스크 리스트 데이터를 처리한다. 질의 로그에 대응하는 복수의 질의 엔트리 각각에 대해, 머신 학습 컴포넌트는 훈련 데이터의 함수로서 잠재적인 질의-태스크 매핑에 대한 최선의 추측 태스크를 제안한다. 그래픽 사용자 인터페이스 생성 컴포넌트는 질의 로그 내의 복수의 질의 엔트리를 디스플레이하되 그 디스플레이된 복수의 질의 엔트리 각각을 그의 대응하는 제안된 최선의 추측 태스크와 관련시키도록 디스플레이하게 구성되어 있다.
- [0023] <실시예>
- [0024] 검색 컴포넌트를 훈련시키기 위한 훈련 데이터를 획득하기 위해 사용자가 질의 로그로부터의 질의를 태스크에 벌크 매핑하는 것을 돕는 주석 시스템을 참조하여 본 발명에 대해 기술된다. 본 발명은 또한 사용자가 훈련 데이터를 획득 및/또는 갱신하기 위해 많은 수의 질의를 대응하는 태스크에 매핑하는 것을 돕는 방법 및 이 방법을 구현하는 컴퓨터 판독가능 매체를 포함한다.
- [0025] 도 1은 본 발명이 구현될 수 있는 적절한 컴퓨팅 시스템 환경(100)의 예를 나타낸다. 컴퓨팅 시스템 환경(100)은 단지 적절한 컴퓨팅 환경의 일 예이며 본 발명의 사용 또는 기능의 범위에 제한을 가하도록 의도된 것은 아니다. 컴퓨팅 환경(100)은 예시적인 오퍼레이팅 환경(100)에 도시된 컴포넌트들 중의 임의의 하나 또는 조합에 관하여 임의의 종속성(dependency) 또는 요구사항(requirement)을 갖는 것으로 해석되어서는 안된다.
- [0026] 본 발명은 많은 다른 범용 또는 특수목적 컴퓨팅 시스템 환경들 또는 구성들과 함께 동작될 수 있다. 본 발명과 함께 사용하기에 적합할 수 있는 잘 알려진 컴퓨팅 시스템, 환경, 및/또는 구성의 예로는, 퍼스널 컴퓨터, 서버 컴퓨터, 핸드헬드(hand-held) 또는 랩탑 장치, 멀티프로세서 시스템, 마이크로프로세서-기반 시스템, 셋탑 박스(set top box), 프로그램가능한 가전제품(programmable consumer electronics), 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터, 상기의 시스템 또는 장치 중의 임의의 것을 포함하는 분산형 컴퓨팅 환경 등이 포함될 수 있지만, 이에 한정되지 않는다.
- [0027] 본 발명은 컴퓨터에 의해 실행되는, 프로그램 모듈과 같은 컴퓨터 실행가능한 명령과 일반적으로 관련하여 기술될 수 있다. 일반적으로, 프로그램 모듈은 특정 태스크를 수행하거나 특정 추상 데이터 유형을 구현하는 루틴, 프로그램, 오브젝트, 컴포넌트, 데이터 구조 등을 포함한다. 본 발명은 또한 통신 네트워크를 통해 링크된 원격 프로세싱 장치에 의해 태스크를 수행하는 분산형 컴퓨팅 환경에서 실행될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 메모리 저장 장치를 포함하는 국부 및 원격 컴퓨터 저장 매체 내에 위치할 수 있다.
- [0028] 도 1을 참조하면, 본 발명을 구현하기 위한 예시적인 시스템은 컴퓨터(110)의 형태의 범용 컴퓨팅 장치를 포함한다. 컴퓨터(110)의 컴포넌트들로는, 프로세싱 유닛(120), 시스템 메모리(130), 및 시스템 메모리를 포함하는 다양한 시스템 컴포넌트를 프로세싱 유닛(120)에 연결시키는 시스템 버스(121)가 포함될 수 있지만, 이에 한정되는 것은 아니다. 시스템 버스(121)는 다양한 버스 아키텍처 중의 임의의 것을 사용하는 로컬 버스, 주변 버스, 및 메모리 버스 또는 메모리 컨트롤러를 포함하는 몇가지 유형의 버스 구조 중의 임의의 것일 수 있다. 예로서, 이러한 아키텍처는 산업 표준 아키텍처(ISA) 버스, 마이크로 채널 아키텍처(MCA) 버스, 인헨스드 ISA(Enhanced ISA; EISA) 버스, 비디오 일렉트로닉스 표준 어소시에이션(VESA) 로컬 버스, 및 (메자닌(Mezzanine) 버스로도 알려진) 주변 컴포넌트 상호접속(PCI) 버스를 포함하지만, 이에 한정되는 것은 아니다.
- [0029] 컴퓨터(110)는 통상적으로 다양한 컴퓨터 판독가능 매체를 포함한다. 컴퓨터 판독가능 매체는 컴퓨터(110)에 의해 액세스될 수 있는 임의의 이용가능한 매체일 수 있으며, 휘발성 및 비휘발성 매체, 분리형(removable) 및 비분리형(non-removable) 매체를 둘다 포함한다. 예로서, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체 및 통신 매체를 포함할 수 있지만, 이에 한정되는 것은 아니다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령, 데이터 구조, 프로그램 모듈 또는 다른 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현되는 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 둘다 포함한다. 컴퓨터 저장 매체는 RAM, ROM, EEPROM, 플래쉬 메모리 또는 기타 메모리 기술, CD-ROM, DVD(digital versatile disk) 또는 기타 광학 디스크 저장장치, 자기 카세트, 자기 테이프, 자기 디스크 저장장치 또는 기타 자기 저장장치, 또는 컴퓨터(110)에 의해 액세스될 수 있고 원하는 정보를 저장하는 데 사용될 수 있는 임의의 기타 매체를 포함할 수 있지만, 이에 한정되지 않는다. 통신 매체는 통상적으로 반송파 또는 기타 전송 메카니즘 등의 변조된 데이터 신호에 컴퓨터 판독가능 명령, 데이터 구조, 프로그램 모듈, 또는 다른 데이터를 구현하며, 임의의 정보 전달 매체를 포함한다. "변조된 데이터 신호"라는 용어는 신호 내에 정보를 인코딩하도록 설정되거나 변환된 특성을 하나 또는 그 이상을 갖는 신호를 의미

한다. 예로서, 통신 매체는 유선 네트워크 또는 직접 유선 접속 등의 유선 매체와, 음향, RF, 적외선 및 기타 무선 매체 등의 무선 매체를 포함하지만, 이에 한정되지 않는다. 상술한 것들 중의 임의의 조합이 컴퓨터 판독가능 매체의 범위 내에 포함되어야 한다.

[0030] 시스템 메모리(130)는 ROM(131) 및 RAM(132) 등의 휘발성 및/또는 비휘발성 메모리의 형태의 컴퓨터 저장 매체를 포함한다. 시동중과 같은 때에 컴퓨터(110) 내의 구성요소들간에 정보를 전송하는 것을 돕는 기본 루틴을 포함하는 기본 입출력 시스템(133; BIOS)은 일반적으로 ROM(131)에 저장된다. RAM(132)은 일반적으로 프로세싱 유닛(120)에 즉시 액세스될 수 있고 및/또는 프로세싱 유닛(120)에 의해 현재 작동되는 프로그램 모듈 및/또는 데이터를 포함한다. 예로서, (한정하고자 하는 것은 아님) 도 1은 오퍼레이팅 시스템(134), 애플리케이션 프로그램(135), 기타 프로그램 모듈(136), 및 프로그램 데이터(137)를 도시한다.

[0031] 컴퓨터(110)는 또한 다른 분리형/비분리형, 휘발성/비휘발성 컴퓨터 저장 매체를 포함할 수 있다. 단지 예로서, 도 1에는 비분리형 비휘발성 자기 매체로부터 판독하거나 그 자기 매체에 기록하는 하드 디스크 드라이브(140), 분리형 비휘발성 자기 디스크(152)로부터 판독하거나 그 자기 디스크에 기록하는 자기 디스크 드라이브(151), 및 CD-ROM 또는 기타 광학 매체 등의 분리형 비휘발성 광학 디스크(156)로부터 판독하거나 그 광학 디스크에 기록하는 광학 디스크 드라이브(155)가 도시되어 있다. 예시적인 오퍼레이팅 환경에서 사용될 수 있는 다른 분리형/비분리형, 휘발성/비휘발성 컴퓨터 저장 매체는 자기 테이프 카세트, 플래쉬 메모리 카드, DVD(Digital versatile disk), 디지털 비디오 테이프, 고체 RAM, 고체 ROM 등을 포함하지만 이에 한정되지 않는다. 하드 디스크 드라이브(141)는 일반적으로 인터페이스(140)와 같은 비분리형 메모리 인터페이스를 통해 시스템 버스(121)에 접속되고, 자기 디스크 드라이브(151) 및 광학 디스크 드라이브(155)는 일반적으로 인터페이스(150)와 같은 분리형 메모리 인터페이스에 의해 시스템 버스(121)에 접속된다.

[0032] 앞서 기술되고 도 1에 도시된 드라이브 및 그 관련 컴퓨터 저장 매체는 컴퓨터(110)를 위한 컴퓨터 판독가능 명령, 데이터 구조, 프로그램 모듈 및 기타 데이터의 저장을 제공한다. 도 1에서, 예를 들어, 하드 디스크 드라이브(141)는 오퍼레이팅 시스템(144), 애플리케이션 프로그램(145), 기타 프로그램 모듈(146), 및 프로그램 데이터(147)를 저장하는 것으로 도시된다. 이들 컴포넌트는 오퍼레이팅 시스템(134), 애플리케이션 프로그램(135), 기타 프로그램 모듈(136), 및 프로그램 데이터(137)와 동일할 수도 있고 다를 수도 있다. 오퍼레이팅 시스템(144), 애플리케이션 프로그램(145), 다른 프로그램 모듈(146), 및 프로그램 데이터(147)는 최소한 다른 복사본(different copies)임을 나타내기 위하여 다른 번호를 부여하였다.

[0033] 사용자는 일반적으로 키보드(162), 마이크로폰(163), 및 마우스, 트랙볼, 또는 터치 패드와 같은 포인팅 장치(161)와 같은 입력 장치를 통해 컴퓨터(110)에 명령 및 정보를 입력할 수 있다. (도시되지 않은) 기타 입력 장치는 마이크로폰, 조이스틱, 게임 패드, 위성 안테나, 스캐너 등을 포함할 수 있다. 이들 입력 장치 및 그외의 입력 장치는 시스템 버스에 연결된 사용자 입력 인터페이스(160)를 통해 종종 프로세싱 유닛(120)에 접속되지만, 병렬 포트, 게임 포트 또는 유니버설 시리얼 포트(USB)와 같은 기타 인터페이스 및 버스 구조에 의해 접속될 수 있다. 모니터(191) 또는 다른 유형의 디스플레이 장치는 또한 비디오 인터페이스(190) 등의 인터페이스를 통해 시스템 버스(121)에 접속된다. 모니터외에도, 컴퓨터는 또한 출력 주변 인터페이스(195)를 통해 접속될 수 있는 스피커(197) 및 프린터(196) 등의 기타 주변 출력 장치를 포함할 수 있다.

[0034] 컴퓨터(110)는 원격 컴퓨터(180)와 같은 하나 이상의 원격 컴퓨터로의 논리적 접속을 이용한 네트워크 환경에서 동작할 수 있다. 원격 컴퓨터(180)는 퍼스널 컴퓨터, 휴대용 장치, 서버, 라우터, 네트워크 PC, 피어(peer) 장치, 또는 기타 공통 네트워크 노드일 수 있으며, 컴퓨터(110)에 관하여 상술한 구성요소 중 다수 또는 모든 구성요소를 일반적으로 포함할 수 있다. 도시된 논리적 접속은 근거리 통신망(LAN; 171) 및 원격 통신망(WAN; 173)을 포함하지만, 그 외의 네트워크를 포함할 수도 있다. 이러한 네트워크 환경은 사무실, 기업 광역 컴퓨터 네트워크(enterprise-wide computer network), 인트라넷, 및 인터넷에서 일반적인 것이다.

[0035] LAN 네트워크 환경에서 사용되는 경우, 컴퓨터(110)는 네트워크 인터페이스 또는 어댑터(170)를 통해 LAN(171)에 접속된다. WAN 네트워크 환경에서 사용되는 경우, 컴퓨터(110)는 일반적으로 인터넷 등의 WAN(173)을 통해 통신을 구축하기 위한 모뎀(172) 또는 기타 수단을 포함한다. 내장형 또는 외장형일 수 있는 모뎀(172)은 사용자 입력 인터페이스(160) 또는 기타 적절한 메카니즘을 통해 시스템 버스(121)에 접속될 수 있다. 네트워크 환경에서, 컴퓨터(110)에 관하여 도시된 프로그램 모듈 또는 그 일부분은 원격 메모리 저장 장치에 저장될 수 있다. 예로서 (한정하고자 하는 것은 아님), 도 1은 원격 컴퓨터(180)에 상주하는 원격 애플리케이션 프로그램(185)을 도시한다. 도시된 네트워크 접속은 예시적인 것이며, 컴퓨터들간의 통신 링크를 구축하는 그 외의 수단이 사용될 수 있다.

- [0036] 도 2는 도 1에 도시한 것 같은 프로세싱 환경에서 구현될 수 있는 본 발명에 따른 주식 시스템(200)을 나타내고 있다. 주식 시스템(200)은 검색 컴포넌트를 훈련시키기 위한 훈련 데이터를 획득하기 위해 사용자가 질의를 태스크에 벌크 매핑하는 것을 돕는다. 시스템(200)은 검색 엔진에 이전에 전송되었던 질의를 나타내는 데이터를 갖는 질의 로그(205)를 포함하거나 다른 실시예들에서는 이를 획득한다. 일반적으로 질의 로그는 많은 수의, 예를 들어 10,000개 이상의 질의를 나타내는 데이터를 포함하게 된다. 그렇지만, 더 작은 질의 로그도 사용될 수 있다. 많은 수의 가능한 태스크를 나타내는 데이터를 갖는 태스크 리스트(210)가 포함되어 있다. 훈련 데이터(220)를 획득하고 연속적으로 갱신하기 위해, 시스템(200)은 사용자가 질의 로그(205) 내의 각각의 질의를 태스크 리스트(210) 내의 태스크로 매핑하는 것을 돕는다.
- [0037] 머신 학습 컴포넌트 또는 분류자(classifier)(215)는 질의 로그(205)로부터 질의 엔트리를 검색하고 태스크 리스트(210)로부터 태스크의 리스트를 검색한다. 본 명세서에 정의되는 바와 같이, 질의 엔트리는 예를 들어 질의 로그로부터의 개별적인 질의일 수 있다. 다른 대안에서, 질의 엔트리는 함께 묶음으로 되어 있는 질의 로그(205) 내의 다수의 질의를 나타내는 질의 묶음(query bundle)일 수 있다. 질의 묶음은 원시 질의 로그(raw query log)의 질의 로그 클러스터링(query log clustering)을 통해 생성된다. 질의 묶음의 예는 질의 내의 여러가지 단어의 사소한 철자 오류를 포함하는 일련의 유사한 질의이다. 각각의 개별적인 질의를 한번에 하나씩 태스크에 매핑하는 대신에 질의 묶음을 태스크에 매핑하는 것은 상당한 시간 절감을 제공한다.
- [0038] 질의 로그에 대응하는 복수의 질의 엔트리 각각에 대해, 머신 학습 컴포넌트는 훈련 데이터(220)를 사용하여 잠재적인 질의-태스크 매핑에 대한 최선의 추측 태스크를 제안한다. 이 추측은 이전에 매핑된 질의로부터의 통계적 정보에 기초하여 머신 학습 컴포넌트에 의해 생성된다. 초기에, 훈련 데이터(220)에 통계적으로 의미있는 데이터가 적은 경우, 최선의 추측은 원하는 만큼 정확하지 않을 것이다. 그렇지만, 더 많은 질의가 매핑되므로, 추측의 정확도가 향상된다. 이하에서 보다 상세히 기술하는 바와 같이, 사용자 또는 저작자는 항상 시스템으로부터의 추측보다 우선하는 선택을 갖는다.
- [0039] 실제로, 저작자가 추측을 받아들일 때마다, 그 저작자는 어떻게 질의를 태스크에 매핑할지를 시스템에 "가르쳐 준다". 특정의 질의 엔트리를 태스크에 매핑함으로써 훈련 데이터를 갱신할 때, 머신 학습 컴포넌트(215)는 갱신된 훈련 데이터(220)의 함수로서 나머지 질의 엔트리 각각에 대한 최선의 추측 태스크를 자동적으로 갱신하도록 구성되어 있다. 충분한 데이터에 주석이 붙여진 경우, 그 데이터는 자연스런 사용자 인터페이스 플랫폼의 검색 컴포넌트에서 배포를 위한 머신 학습 알고리즘을 훈련시키는 데 사용될 수 있다. 그의 최선의 추측을 이와 같은 방식으로 계속하여 갱신할 수 있는 머신 학습 컴포넌트 또는 분류자 유형의 예는 나이브 베이스 분류자(Naive Bayes classifier)이다.
- [0040] 시스템(200)에서, 그래픽 사용자 인터페이스(GUI) 생성 컴포넌트(225)는 디스플레이(모니터(191) 등)와 입력 장치(장치(161, 162) 등)를 통해 머신 학습 컴포넌트(215)와 사용자 사이에서 인터페이스하기 위해 포함되어 있다. 컴포넌트(225)는 질의 로그 내의 복수의 질의 엔트리를 디스플레이하되 디스플레이된 복수의 질의 엔트리 각각을 그의 대응하는 제안된 최선의 추측 태스크와 관련시키도록 디스플레이하게 구성되어 있다. 컴포넌트(225)는 또한 매핑 프로세스를 진행시키기 위해 사용자로부터 여러가지 유형의 입력을 수신하도록 구성되어 있다. 예시적인 실시예에 따른 컴포넌트(225)에 의해 생성된 GUI가 도 3 내지 도 7에 도시되어 있다.
- [0041] 이제 도 3을 참조하면, 본 발명의 어떤 실시예들에 따라 모니터(191) 상에 디스플레이될 수 있는 GUI(300)가 도시되어 있다. GUI(300)는 태스크 리스트(302), 질의 엔트리 리스트(질의 묶음 리스트로 나타내어져 있음)(304), 각 질의 묶음 내의 질의의 수를 나타낸 열(306), 질의의 발생 빈도를 나타낸 열(308), 머신 학습 컴포넌트(215)에 의해 생성된 각각의 질의 엔트리에 대한 최선의 추측(310), 및 제안된 최선의 추측이 정확할 확률을 나타내는 열(312)을 포함한다. 태스크 리스트(302)는 질의-태스크 매핑을 위해 최선의 추측 태스크가 저작자에 의해 선택되지 않은 경우 질의 엔트리가 그에 매핑되는 태스크의 리스트이다. 또한, 질의 엔트리(304)가 질의 묶음일 때, 각각의 질의 묶음에 대해 제안된 최선의 추측 태스크(310)는 질의 묶음에 의해 나타내어진 복수의 질의 각각에 대한 최선의 추측의 가중 평균일 수 있다.
- [0042] 이제 도 4를 참조하면, GUI 생성 컴포넌트(225)는 또한 사용자가 특정의 질의 엔트리를 그의 대응하는 제안된 최선의 추측 태스크에 매핑하기를 원할 때 사용자로부터 제1 유형의 입력을 수신하도록 구성되어 있다. 도 4에 도시된 예에서, 사용자는 질의 묶음 "패스워드"로 나타내어진 질의들 모두를 머신 학습 컴포넌트에 의해 생성된 "패스워드 변경"이라는 최선의 추측 태스크로 매핑한다. 이 매핑이 계속되도록 지시하는 데 사용되는 제1 유형의 입력은 예를 들어 커서(340)를 이 질의 엔트리 상에 위치한 상태로 마우스형 포인팅 장치의 우측 클릭일 수 있다. 줄(350)은 이 질의 엔트리 및 관련 통계와 최선의 추측 태스크의 선택 또는 하이라이트를 나타낸다.

GUI(300)를 통해 제1 유형의 입력의 수신 시에, 머신 학습 컴포넌트(215)는 특정의 질의 엔트리를 제안된 최선의 추측 태스크에 매핑함으로써 훈련 데이터를 갱신한다. 전술한 바와 같이, 이어서 나머지 질의 엔트리에 대한 최선의 추측 태스크는 갱신된 훈련 데이터에 기초하여 갱신된다.

[0043] 이제 도 5를 참조하면, 질의 엔트리를 제안된 최선의 추측 태스크에 매핑하는 대신에 사용자가 통계적으로 결정된 그 다음 최선의 추측 태스크의 리스트를 검토하기를 원하는 경우, 이것은 적절한 입력을 제공함으로써 달성될 수 있다. 예를 들어, 일 실시예에서, 커서(340)를 제안된 최선의 추측 "패스워드 변경" 태스크 상에 위치시킨 상태에서의 마우스를 클릭한 결과, 그 다음 최선의 추측의 리스트를 갖는 드롭 다운 메뉴(360)가 디스플레이된다. 이어서 사용자는 메뉴(360) 내의 원하는 태스크 상에서 클릭함으로써 문제의 질의 엔트리를 그 다음 최선의 추측 태스크 중 임의의 것에 매핑할 수 있다.

[0044] 이제 도 6을 참조하면, 사용자는 원하는 경우, 질의 묶음 형태의 질의 엔트리로 나타내어진 개별적인 질의를 볼 수 있다. 예를 들어, GUI 생성 컴포넌트(225)는 어떤 실시예들에서 사용자가 특정의 질의 묶음으로 나타내어진 질의의 리스트를 갖는 드롭 다운 메뉴(370)를 보기를 원하는 경우 사용자로부터 제2 유형의 입력을 수신하도록 구성되어 있다. 그 입력에 응답하여, GUI 생성 컴포넌트는 메뉴(370)를 디스플레이한다. 제2 유형의 입력의 예는 커서를 "패스워드" 질의 묶음 상에 위치시킨 상태에서의 더블 클릭이다. 메뉴(370)에 열거된 개별적인 질의는 이어서 원하는 경우 한 번에 하나씩 태스크에 매핑될 수 있다.

[0045] 이제 도 7을 참조하면, 제안된 최선의 추측 태스크 중 어느 것도 질의 엔트리에 대한 매핑에 적절하지 않은 경우, 사용자는 적절한 입력을 제공함으로써 태스크 리스트(302)로부터의 질의 엔트리를 태스크에 매핑할 수 있다. 예로서, 일 실시예에서, 적절한 입력은 질의 엔트리를 선택하고(예를 들어, 줄(380)에 의해 선택된 것으로 표시된 "cd") 이어서 리스트(302)로부터의 태스크 상에서 클릭하는 것을 포함한다.

[0046] 이제 도 8 내지 도 12를 참조하면, 본 발명의 방법의 어떤 실시예의 여러가지 단계를 나타낸 흐름도(400, 450, 500, 550, 600)가 도시되어 있다. 이들 흐름도는 도 1 내지 도 7을 참조하여 전술한 본 발명의 어떤 특징들을 요약한 것이다. 도 8의 흐름도(400)에 도시한 바와 같이, 검색 컴포넌트를 훈련시키기 위한 훈련 데이터를 획득하기 위해 사용자가 질의를 태스크에 벌크 매핑하는 것을 돕는 방법은 검색 엔진에 이전에 전송되었던 질의들로 된 질의 로그를 획득하는 단계를 포함한다. 이것은 블록 402로 나타내어져 있다. 블록 404에 나타난 바와 같이, 본 방법은 또한 가능한 태스크의 태스크 리스트를 획득하는 단계를 포함한다. 블록 406에서, 본 방법은 질의 로그에 대응하는 복수의 질의 엔트리의 각각에 대하여 잠재적인 질의-태스크 매핑을 위한 최선의 추측 태스크를 결정하는 단계를 포함한다. 최선의 추측 태스크는 머신 학습 컴포넌트를 사용하여 훈련 데이터의 함수로서 결정된다. 블록 408에서, 본 방법은 질의 로그 내의 복수의 질의 엔트리를 디스플레이하되 디스플레이된 복수의 질의 엔트리의 각각을 그의 대응하는 제안된 최선의 추측 태스크와 관련시키도록 디스플레이하는 단계를 포함하는 것으로 도시되어 있다.

[0047] 이제 도 9를 참조하면, 본 방법은 또한 블록 452에 나타난 바와 같이 사용자가 특정의 질의 엔트리를 그의 대응하는 제안된 최선의 추측 태스크에 매핑하기를 원하는 경우 사용자로부터 제1 유형의 입력을 수신하는 단계, 및 블록 454로 나타난 바와 같이 머신 학습 컴포넌트를 사용하여 훈련 데이터를 갱신하는 단계를 포함할 수 있다.

[0048] 이제 도 10을 참조하면, 어떤 실시예들에서, 본 발명의 방법은 블록 502로 나타난 바와 같이 사용자가 잠재적인 매핑을 위해 특정의 질의 엔트리에 대한 그 다음 최선의 추측 태스크의 리스트를 보고자 하는 경우 그 사용자로부터 제2 유형의 입력을 수신하는 단계를 더 포함한다. 이들 실시예에서, 본 방법은 이어서 블록 504로 나타난 바와 같이 특정의 질의 엔트리에 대해 머신 학습 컴포넌트를 사용하여 훈련 데이터의 함수로서 결정된 그 다음 최선의 추측 태스크의 리스트를 디스플레이하는 단계를 포함한다.

[0049] 어떤 실시예들에서, 본 발명의 방법은 도 11의 흐름도(550)에 나타난 추가의 단계들을 포함한다. 블록 552에 나타난 바와 같이, 이들 실시예에서, 본 방법은 사용자가 특정의 질의 묶음에 의해 나타내어진 질의의 리스트를 보고자 하는 경우 사용자로부터 제2 유형의 입력을 수신하는 단계를 더 포함한다. 블록 554에 나타난 바와 같이, 질의 묶음으로 나타난 질의의 리스트는 이어서 제2 유형의 입력의 수신에 응답하여 디스플레이된다.

[0050] 이제 도 12를 참조하면, 사용자가 특정의 질의 엔트리를 태스크 리스트(302)로부터의 태스크에 매핑할 수 있도록 해주기 위해 어떤 실시예들에서 포함될 수 있는 본 방법의 추가의 단계들이 도시되어 있다. 예를 들어, 흐름도(600)의 블록 602에 나타난 바와 같이, 본 방법은 사용자가 특정의 질의 엔트리를, 제안된 최선의 추측 태스크와는 다른 디스플레이된 태스크 리스트로부터의 태스크에 매핑하고자 하는 경우 사용자로부터 제2 유형의 입력을 수신하는 추가의 단계를 포함할 수 있다. 블록 604의 단계에 나타난 바와 같이, 이어서 훈련 데이터가

머신 학습 컴포넌트를 사용하여 특정의 질의 엔트리를 디스플레이된 태스크 리스트로부터의 태스크에 매핑함으로써 갱신된다.

[0051] 도 3 내지 도 7에 도시한 GUI 실시예는 본 발명의 방법을 "질의-중심" 동작으로서 구현한다. 즉, 이들 실시예에서, 사용자는 최선의 추측 태스크를 복수의 질의 각각에 할당한다. 그렇지만, 본 발명은 최선의 추측 질의가 특정의 태스크에 할당되는 것인 "태스크-중심" 동작으로서 구현될 수도 있다.

[0052] 예를 들어, 도 13을 참조하면, 본 발명의 어떤 실시예들에 따라 모니터(191)상에 디스플레이될 수 있는 GUI(700)가 도시되어 있다. GUI(700)는 계층적 태스크 리스트(702)를 포함한다. 질의-태스크 매핑을 수행하는 사용자 또는 저작자는 태스크와 함께 매핑하기 위한 제안된 질의들로 된 리스트를 보기 위해 계층적 태스크 리스트(702)로부터 태스크를 선택한다. 이것이 도 14에 도시되어 있다.

[0053] GUI 생성 컴포넌트(225)는 사용자가 태스크 리스트(702)로부터의 특정의 태스크를 대응하는 제안된 최선의 추측 질의 중 하나 이상으로 매핑하기를 원하는 경우, 사용자로부터 제1 유형의 입력을 수신하도록 구성되어 있다. 예를 들어, 커서(740)를 갖는 마우스 형태의 포인팅 장치의 클릭은 이러한 입력을 제공하는 데 사용될 수 있다. 줄(750)은 이 태스크의 선택 또는 하이라이트를 나타낸다. 이 입력에 응답하여, 시스템 또는 틀은 이어서 머신 학습 컴포넌트에 의해 제안된 질의들로 된 리스트(704)를 보여준다. 질의들은 선택된 태스크에 매핑할 가능성의 순서로 디스플레이된다. 이 프로세스는 사용자가 질의로부터 가능한 태스크로의 매핑과는 반대로 태스크로부터 가능한 질의로 매핑할 수 있게 해준다. 다른 측면에서 보면, 결과로서 얻어지는 추측의 시간에 따른 향상을 포함하여, 이 프로세스는 전술한 "질의-중심" 동작과 동일한 거동을 갖는다.

[0054] 본 발명의 실시예들에서, 2명 이상의 사용자가 질의-태스크 매핑 시스템을 사용하여 질의를 태스크에 매핑할 수 있다. 예를 들어, 도 15에 도시한 컴퓨팅 환경(800)을 생각해보자. 컴퓨팅 환경(800)은 중앙 서버(802) 및 다수의 사용자 컴퓨터(810)(컴퓨터(810-1, 810-2, 810-3)가 도시되어 있음)를 포함한다. 중앙 서버 컴퓨터는 분류자 모델(classifier model)(806)을 확립하는 데 사용되는 훈련 데이터를 저장하는 훈련 데이터 스토어(804)를 포함한다. 예시된 시스템 또는 프로세싱 환경(800)에서, 사용자에게 추측을 제공하기 위해 머신 학습 컴포넌트에 의해 사용되는 분류자 모델(806)의 생성은 모든 사용자에게 의해 행해지는 일련의 매핑에 의해 훈련된다. 이들 매핑은 중앙 데이터베이스 훈련 데이터 스토어(804)에 저장된다. 분류자 모델(806)은 이들 매핑을 사용하여 주기적으로 훈련된다. 훈련 데이터 스토어(804) 및 분류자 모델(806) 둘다가 중앙 서버(802) 상에 저장되어 있는 것으로 도시되어 있지만, 다른 실시예들에서 분류자 모델(806)은 다른 곳에 저장되고, 이어서 중앙 서버(802)에 계속 연결되어 있을 수 있으며, 이 서버로부터 사용자는 매핑 세션 사이에 갱신된 모델을 다운로드할 수 있다.

[0055] 모델의 다운로드된 복사본(로컬 분류자 복사본(814-1 내지 814-3)으로 도시됨)은 각각의 사용자의 컴퓨터 상에 로컬적으로 존속된다. 사용자는 이어서 모든 사용자로부터 수집된 매핑 데이터를 사용하여 적어도 주기적으로 훈련된 모델을 사용할 수 있다. 특정의 사용자가 매핑 작업을 계속함에 따라, 모델의 로컬 복사본(814)은 그 개별적인 사용자의 매핑에 적합하게 된다. 이것은 각각의 사용자 컴퓨터 상에 로컬 매핑 데이터 스토어(812-1 내지 812-3)를 포함하는 것에 의해 나타내어져 있다. 따라서, 로컬 분류자 모델이 그의 가장 최근의 매핑을 고려하기 때문에 사용자는 이득을 본다. 이것은 원래 존속된 모델(806)보다 추측의 정확도를 향상시킨다. 게다가, 사용자는 추측이 최근의 매핑에 주문화되는 정도를 조정할 수 있다. 주기적으로, 로컬 사용자 컴퓨터에서 행해진 매핑에 대응하는 매핑 데이터(812-1 내지 812-3)는 분류자 모델(806)의 장래의 갱신을 위해 중앙 서버(802) 상의 훈련 데이터 스토어(804)로 제공된다.

[0056] 본 발명의 다중 사용자 컴퓨터 특징은 또한 중앙 서버 및 로컬 컴퓨터 둘다를 사용하여 질의를 태스크에 벌크 매핑하는 일반적인 방법을 설명하는 도 16에 나타난 흐름도(850)에 나타내어져 있다. 도 16의 블록 852에 도시한 바와 같이, 본 방법은 중앙 서버로부터 사용자 컴퓨터로 분류자 모델의 복사본을 다운로드하는 단계를 포함한다. 블록 854에서, 본 방법은 다운로드된 분류자 모델을 사용하여 생성된 추측에 기초하여 사용자 컴퓨터에서 질의-태스크 매핑을 행하는 단계를 포함하는 것으로 도시되어 있다. 블록 856에서, 본 방법은 로컬 매핑에 기초하여 로컬 훈련 데이터 스토어를 생성 또는 갱신하는 단계를 포함하는 것으로 도시되어 있다. 즉, 사용자의 컴퓨터 상에 저장된 분류자 모델의 로컬 버전을 갱신하기 위해 로컬 매핑이 사용된다. 마지막으로, 블록 858에 도시된 바와 같이, 본 방법은 로컬 매핑 데이터를 사용하여 중앙 서버의 훈련 데이터 스토어를 갱신하는 단계를 포함한다. 따라서, 중앙 서버에 저장되거나 그로 다운로드된 분류자 모델(806)은 개별적인 사용자 컴퓨터 각각에서 수행되는 매핑을 포함하는 훈련 데이터 스토어(804)를 사용하여 갱신된다.

[0057] 조밀하지 않은 질의-태스크 매핑 데이터가 이용가능한 경우, 분류자 모델(806)은 다른 정보 소스를 사용하여 사

전 훈련될 수 있다. 예를 들어, 이러한 사전 훈련은 종종 매핑 프로세스의 초기 단계에서 사용된다. 이어서 이 모델은 로컬 컴퓨터로부터의 모든 사용자에게 의해 행해진 일련의 매핑을 사용하여 훈련될 수 있다. 이들 정보 소스 각각의 상대적 영향은 조정될 수 있다. 이러한 사전 훈련은 모델 생성 프로세스에 언제라도 적용될 수 있다.

[0058] 표준 사전 훈련 기술은 태스크와 관련된 데이터로부터 일련의 특징을 추출하는 것이다. 이어서 모델은 특징과 태스크 사이의 연관을 바탕으로 훈련될 수 있다. 그 효과는 질의를 태스크에 매핑하는 것에 의해 발생하는 것과 유사하다. 가장 간단한 형태의 이러한 사전 훈련에서, 태스크 이름은 질의로서 기능하며, 그 질의에 대해 태스크로의 매핑이 도입된다(비록 이 질의는 명시적으로 존속되지 않음). 사전 훈련은 또한 다른 소스로부터의 텍스트 매핑을 사용하여 달성될 수 있다. 이들 매핑은 훈련 데이터 스토어에 노출되거나 존속될 필요가 없다.

[0059] 로컬 컴퓨터(810-1 내지 810-3) 중 하나에서의 사용자 세션 동안, 모델의 추측은 또한 그것이 생성될 때에 영향을 받을 수 있다. 예를 들어, 질의가 이미 태스크에 매핑되어 있는 경우 그 질의는 추측으로서 나타나지 않는다. 이것은 다수의 사용자에게 의해 중복 매핑이 일어나지 않도록 할 수 있다. 사용자는 또한 제공된 추측을 제약하기 위해 명시적으로 필터를 적용할 수 있다.

[0060] 본 발명이 특정의 실시예를 참조하여 기술되어 있지만, 당업자라면 본 발명의 사상 및 범위를 벗어나지 않고 형태 및 세부 사항에 있어서 변경이 행해질 수 있음을 잘 알 것이다.

발명의 효과

[0061] 본 발명의 주석 시스템을 이용하여 사용자가 많은 수의 질의를 태스크에 매핑하는 것을 도와 검색 컴포넌트를 훈련시키기 위한 훈련 데이터를 획득할 수 있다.

도면의 간단한 설명

[0001] 도 1은 본 발명이 사용될 수 있는 한 예시적인 환경의 블록도.

[0002] 도 2는 본 발명에 따른 주석 시스템의 한 예시적인 실시예를 나타낸 블록도.

[0003] 도 3 내지 도 7은 사용자가 질의-태스크 매핑을 효과적으로 수행하도록 돕기 위해 도 2에 도시된 시스템을 사용하여 생성될 수 있는 그래픽 사용자 인터페이스를 개략적으로 나타낸 도면.

[0004] 도 8 내지 도 12는 본 발명의 방법의 실시예를 나타낸 흐름도.

[0005] 도 13 및 도 14는 도 3 내지 도 7에 나타난 질의 중심 동작과는 반대로 사용자가 질의-태스크 매핑을 태스크 중심 동작으로서 효과적으로 수행하도록 돕기 위해 도 2에 도시된 시스템을 사용하여 생성될 수 있는 그래픽 사용자 인터페이스를 개략적으로 나타낸 도면.

[0006] 도 15는 본 발명이 사용될 수 있는 한 예시적인 실시예를 나타낸 블록도.

[0007] 도 16은 본 발명의 방법의 실시예를 나타낸 흐름도.

[0008] <도면의 주요부분에 대한 부호의 설명>

[0009] 161, 162: 입력 장치

[0010] 191: 모니터

[0011] 205: 질의 로그

[0012] 210: 태스크 리스트

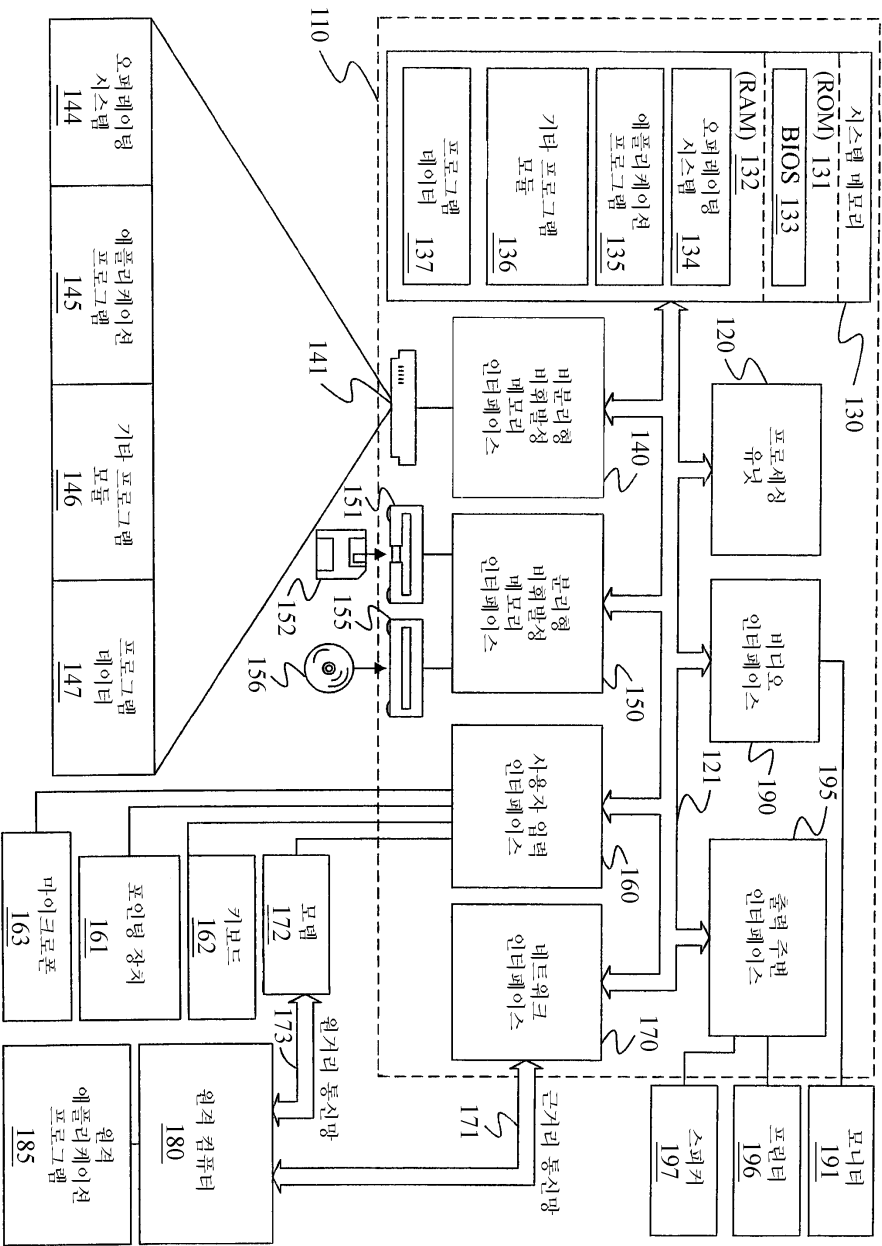
[0013] 215: 머신 학습 컴포넌트

[0014] 220: 훈련 데이터 스토어

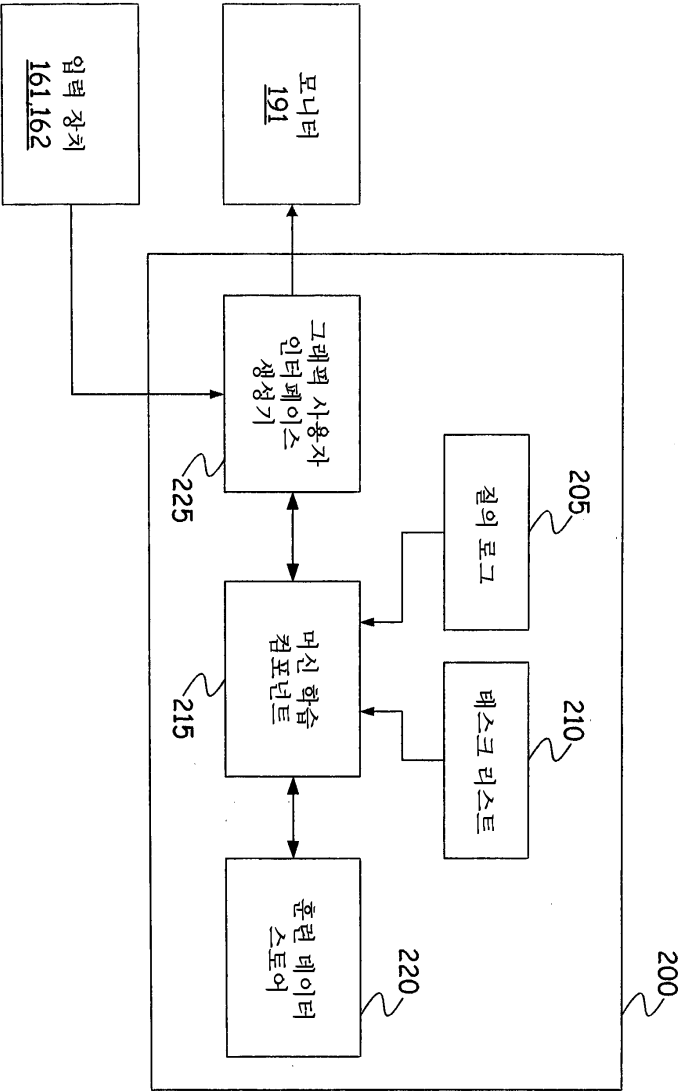
[0015] 225: 그래픽 사용자 인터페이스 생성기

도면

도면1



도면2



도면3

302		304		306		308		310		312	
태스크 리스트		질의 묶음		질의의 수		빈도		최선의 예측		최선의 예측 확률	
access shared printer		firewall		101		19449		install firewall		67	
access system info		register		293		11892		save picture		1	
add address to address		netmeeting		112		11647		save picture		1	
add comment date and		printer		196		8239		configure printer		26	
add contact		boot disk		49		8269		copy file to floppy disk		15	
add icon to desktop		format		172		7657		format paragraph in email		25	
⋮		cookies		81		7652		save picture		1	
backup data to tape		password		100		7611		change password		28	
burn file to CD		startup		229		7294		troubleshoot startup comput		24	
burn music to CD		email		45		7331		e-mail picture		6	
capture screen content		dvd		13		7361		save picture		1	
change font in document		administrator		309		6851		administer user account		1	
change password		startup disk		97		6811		troubleshoot startup comput		29	
change user account		activation		194		6398		learn about op system activat		10	
change volume sound to		backup		35		6350		learn about backup		30	
check for e-mail messa		ntfs		10		6338		save picture		1	
close program all		partition		200		5837		create partition		44	
compress file and fold		cd		24		5308		save picture		1	
configure display setti		⋮		⋮		⋮		⋮		⋮	
configure Internet Exp		compatibility		277		4570		save picture		1	
configure modem		copy cd		49		4675		copy music		39	
configure modem for fa		remote desktop		52		4670		learn about desktop		27	

302		304		306		308		310		312	
태스크 리스트		질의 목록		질의의 수		빈도		최신의 추측		최신의 추측 확률	
access shared printer		firewall		101		19449		install firewall		67	
access system info		register		293		11892		save picture		1	
add address to address		netmeeting		112		11647		save picture		1	
add comment date and		printer		196		8239		configure printer		26	
add contact		boot disk		49		8269		copy file to floppy disk		15	
add icon to desktop		format		172		7657		format paragraph in email		25	
::		cookies		81		7652		save picture		1	
backup data to tape		password		100		7611		change password		28	
burn file to CD		startup 340		229		7294		troubleshoot startup comput		24	
burn music to CD		email		45		7331		e-mail picture		6	
capture screen content		dvd		13		7361		save picture		1	
change font in documen		administrator		309		6851		administer user account		1	
change password		startup disk		97		6811		troubleshoot startup comput		29	
change user account		activation		194		6398		learn about op system activat		10	
change volume sound to		backup		35		6350		learn about backup		30	
check for e-mail messa		ntfs		10		6338		save picture		1	
close program all		partition		200		5837		create partition		44	
compress file and fold		cd		24		5308		save picture		1	
configure display setti		::		::		::		::		::	
configure Internet Exp		compatibility		277		4570		save picture		1	
configure modem		copy cd		49		4675		copy music		39	
configure modem for fa		remote desktop		52		4670		learn about desktop		27	
::											

도면4

도면5

302		304		306		308		310		312	
태스크 리스트		질의 목록		질의의 수		빈도		최신의 추측		최신의 추측 확률	
access shared printer	firewall	101	19449	install firewall	67						
access system info	register	293	11892	save picture	1						
add address to address	netmeeting	112	11647	save picture	1						
add comment date and	printer	196	8239	configure printer	26						
add contact	boot disk	49	8269	copy file to floppy disk	15						
add icon to desktop	format	172	7657	format paragraph in email	25						
⋮	cookies	81	7652	save picture	1						
backup data to tape	password	100	7611	change password	28						
burn file to CD	startup	229	7294	troubleshoot startup comput	24						
burn music to CD	email	45									
capture screen content	dvd	13									
change font in document	administrator	309									
change password	startup disk	97									
change user account	activation	194									
change volume sound to	backup	35									
check for e-mail messa	ntfs	10									
close program all	partition	200									
compress file and fold	cd	24									
configure display setti	⋮	⋮									
configure Internet Exp	compatibility	277									
configure modem	copy cd	49									
configure modem for fa	remote desktop	52									
⋮											

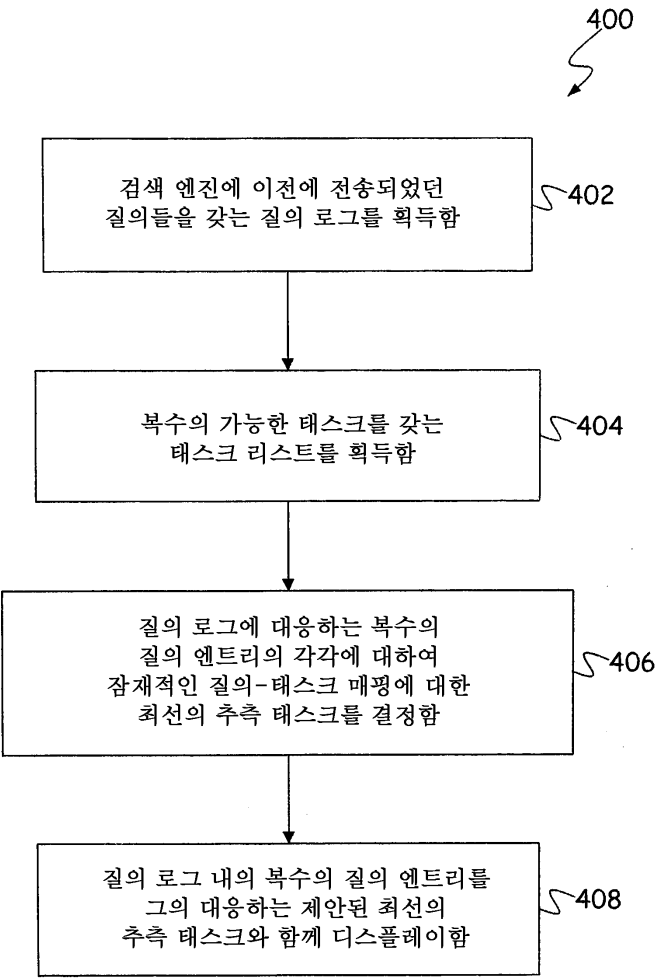
도면6

302		304		306		308		310		312	
데스크 리스트		질의 묶음		질의의 수		빈도		최신의 주석		최신의 주석의 확률	
access shared printer		firewall		101		19449		install firewall		67	
access system info		register		293		11892		save picture		1	
add address to address		netmeeting		112		11647		save picture		1	
add comment date and		printer		196		8239		configure printer		26	
add contact		boot disk		49		8269		copy file to floppy disk		15	
add icon to desktop		format		340		172		format paragraph in email		25	
⋮		cookies		81		7652		save picture		1	
backup data to tape		password		100		7611		change password		28	
burn file to CD		startup		229		7294		troubleshoot startup comput		24	
burn music to CD		Queries in Query Bundle									
capture screen content		adn		How do I change my password?				account		1	
change font in document		sta		How to set password?				rtup comput		29	
change password		act		Can I change the password?				ystem activat		10	
change user account		bac		Can I keep anyone from modifying the document?				kup		30	
change volume sound to		nfi		⋮						1	
check for e-mail messa		par		⋮						44	
close program all		cd		How do I remove the password?						1	
compress file and fold		⋮		Do I need a network password?						⋮	
configure display setti		⋮								1	
configure Internet Exp		conf								39	
configure modem		cop								27	
configure modem for fa		refine account									
⋮		refine account									
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											
⋮											

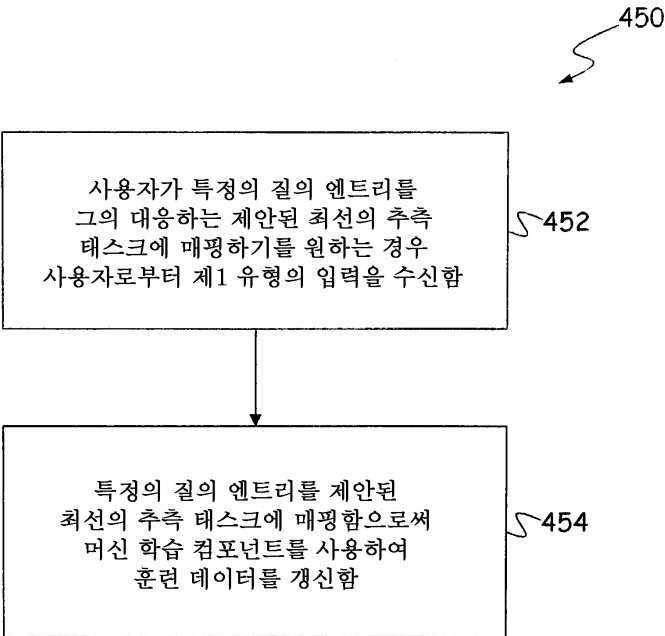
도면7

302		304		306		308		310		312	
메스크 리스트		질의 묶음		질의의 수		빈도		최선의 주측		최선의 주측 확률	
access shared printer		firewall		101		19449		install firewall		67	
access system info		register		293		11892		save picture		1	
add address to address		netmeeting		112		11647		save picture		1	
add comment date and		printer		196		8239		configure printer		26	
add contact		boot disk		49		8269		copy file to floppy disk		15	
add icon to desktop		format		172		7657		format paragraph in email		25	
⋮		cookies		81		7652		save picture		1	
backup data to tape		password		100		7611		change password		28	
burn file to CD		startup		229		7294		troubleshoot startup comput		24	
burn music to CD		email		45		7331		e-mail picture		6	
⋮		dvd		13		7361		save picture		1	
capture screen content		administrator		309		6851		administer user account		1	
change font in document		startup disk		97		6811		troubleshoot startup comput		29	
change password		activation		194		6398		learn about op system activat		10	
change user account		backup		35		6350		learn about backup		30	
change volume sound to		ntfs		10		6338		save picture		1	
check for e-mail messa		partition		200		5837		create partition		44	
close program all		cd		24		5308		save picture		1	
compress file and fold		⋮		⋮		⋮		⋮		⋮	
configure display setti		380		277		4570		save picture		1	
configure Internet Exp		compatibility		49		4675		copy music		39	
configure modem		copy cd		52		4670		learn about desktop		27	
configure modem for fa		remote desktop									
⋮											

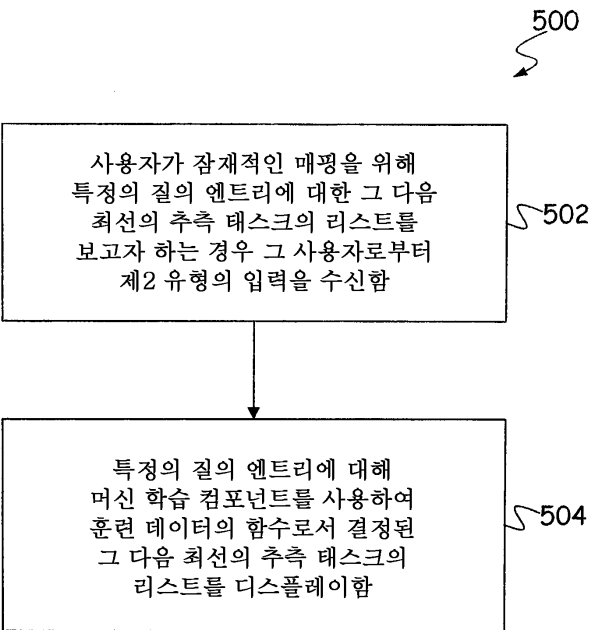
도면8



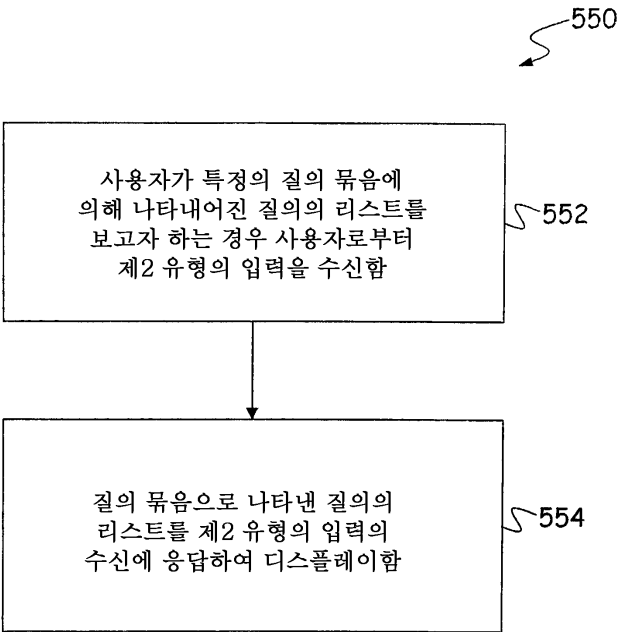
도면9



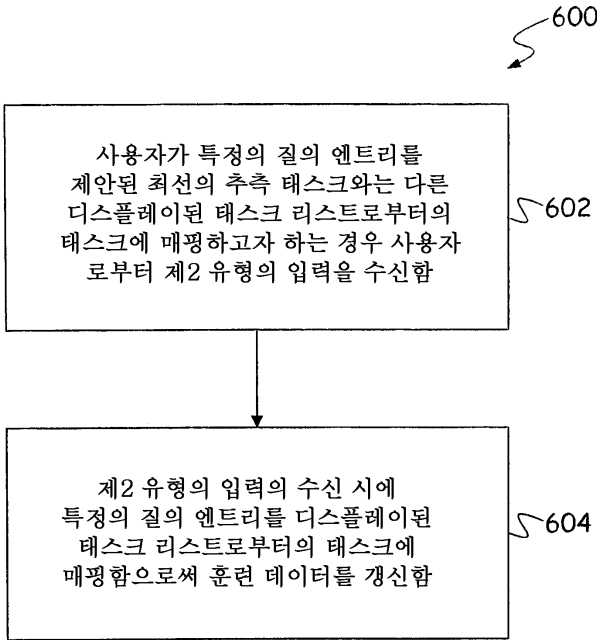
도면10



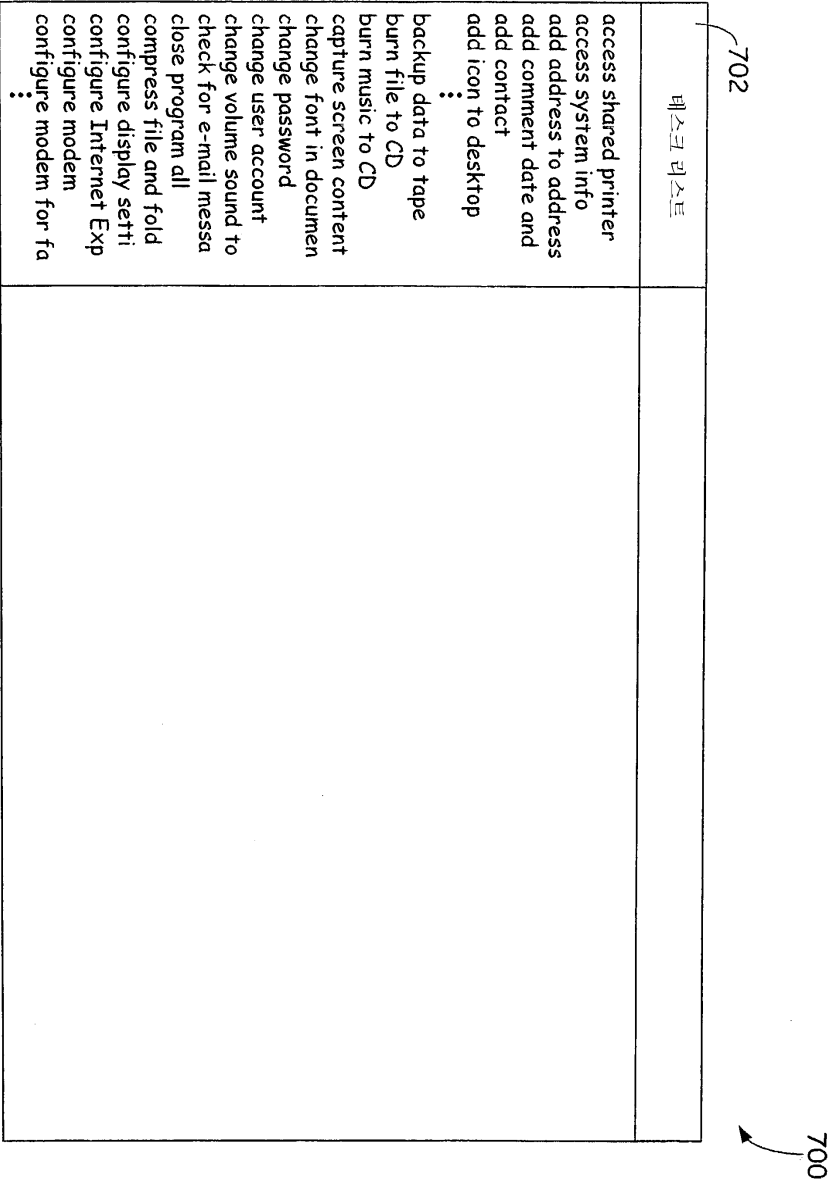
도면11



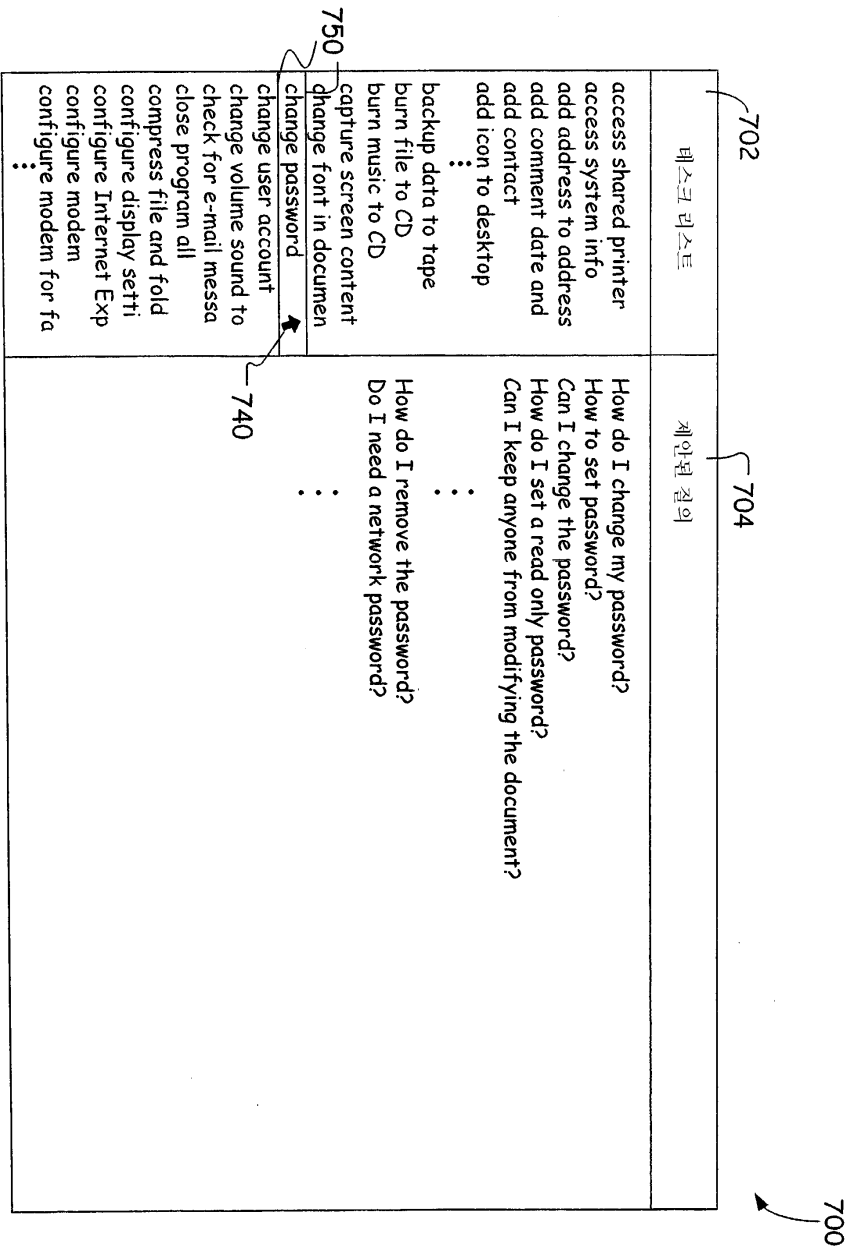
도면12



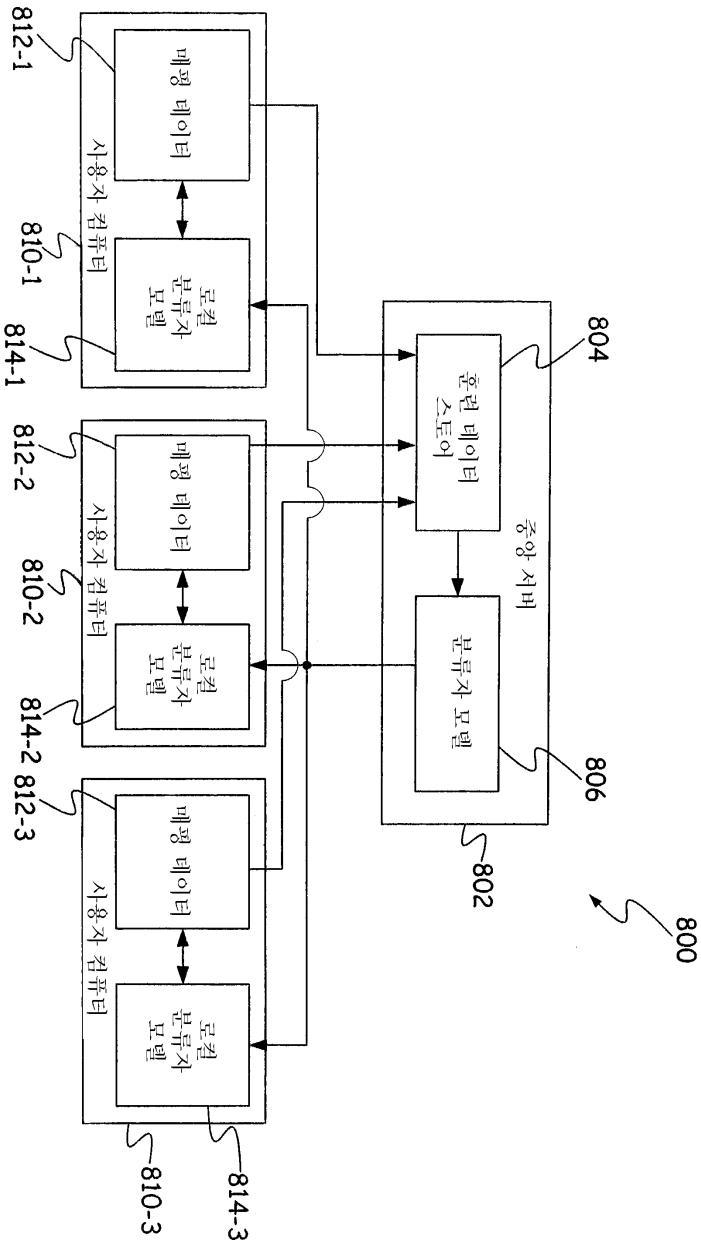
도면13



도면14



도면15



도면16

