



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2013년09월24일
(11) 등록번호 10-1311050
(24) 등록일자 2013년09월13일

(51) 국제특허분류(Int. Cl.)
G06F 17/27 (2006.01) G06F 17/30 (2006.01)
(21) 출원번호 10-2008-7006775
(22) 출원일자(국제) 2006년09월20일
심사청구일자 2011년09월09일
(85) 번역문제출일자 2008년03월20일
(65) 공개번호 10-2008-0046670
(43) 공개일자 2008년05월27일
(86) 국제출원번호 PCT/US2006/037206
(87) 국제공개번호 WO 2007/035919
국제공개일자 2007년03월29일
(30) 우선권주장
11/231,955 2005년09월21일 미국(US)
(56) 선행기술조사문헌
US6272507 B1
US20020026390 A1
US20030046389 A1
US20040260695 A1

(73) 특허권자
마이크로소프트 코포레이션
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원
마이크로소프트 웨이
(72) 발명자
마이러존, 드미트리
미국 98052-6399 워싱턴주 레드몬드 원 마이크로
소프트 웨이
자라고자, 휴고
미국 98052-6399 워싱턴주 레드몬드 원 마이크로
소프트 웨이
(뒷면에 계속)
(74) 대리인
제일특허법인

전체 청구항 수 : 총 14 항

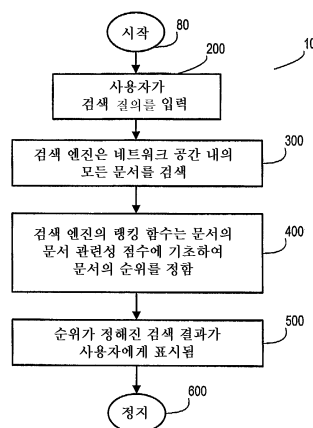
심사관 : 박상현

(54) 발명의 명칭 문서 사용 통계치를 사용한 랭킹 함수

(57) 요약

네트워크상의 문서에 대한 문서 관련성 점수를 제공하는 방법이 개시된다. 네트워크상의 문서에 대한 문서 관련성 점수를 제공하는 방법을 실행하는 컴퓨터 실행가능 명령어가 그 위에 저장되어 있는 컴퓨터 판독가능 매체가 또한 개시된다. 더욱이, 최소한 하나의 애플리케이션 모듈을 포함하는 컴퓨팅 시스템이 개시되는데, 최소한 하나의 애플리케이션 모듈은 네트워크상의 문서에 대한 문서 관련성 점수를 제공하는 방법을 실행하는 애플리케이션 코드를 포함한다.

대표도 - 도1



(72) 발명자

펠토넨, 카일

미국 98052-6399 워싱턴주 레드몬드 원 마이크로소
프트 웨이

데브루이네, 앤드류

미국 98052-6399 워싱턴주 레드몬드 원 마이크로소
프트 웨이

특허청구의 범위

청구항 1

컴퓨터 실행가능 명령어들이 저장되어 있는 컴퓨터 판독가능 매체에 있어서,

상기 컴퓨터 실행가능 명령어들은 네트워크상의 문서들의 순위를 정하는 명령어를 포함하고, 상기 컴퓨터 실행가능 명령어들은 하나 이상의 질의-독립 성분들(query-independent components)을 포함하는 랭킹 함수(ranking function) -- 상기 랭킹 함수는 상기 문서들의 사용자 질의를 위해 작동됨 -- 를 이용하고, 적어도 하나의 질의-독립 성분은 상기 네트워크상의 하나 이상의 문서들에 대한 서버-생성, 서버-저장 사용 데이터를 고려한 사용 파라미터를 포함하고, 상기 사용 데이터는 상기 문서들로의 네트워크 액세스(access)를 사용자들에게 제공하고 관리하는 네트워크 저장 시스템에 의해 생성 및 저장되고, 상기 사용 데이터는 문서의 사용 데이터가 상기 네트워크 저장 시스템을 통한 다수의 사용자들에 의한 상기 문서의 질의-독립적이고 대화식인(interactive) 사용을 반영하도록 상기 네트워크 저장 시스템을 통한 상기 네트워크상의 하나 이상의 문서와의 실제 사용자 상호작용의 측정을 포함하고, 상기 컴퓨터 실행가능 명령어들은 또한 상기 랭킹 함수에 의해 생성되고 순서대로 문서들의 순위를 정하는 데 사용되는 점수를 상기 네트워크상의 문서들의 순위를 정하기 위해 할당하고,

상기 컴퓨터 실행가능 명령어들은 사용자에게 의해 입력된 검색 문자열(string)을 포함하는 검색 조회(inquiry)를 받아들이고, 여러 문서를 포함하는 검색 결과를 생성하기 위해 상기 네트워크상의 문서들에 대한 검색을 행하고, 상기 문서들이 각각의 사용 데이터 및 검색 문자열과의 관련성에 따라 순위가 정해지도록 순위가 정해진 검색 결과를 생성하기 위해 상기 랭킹 함수를 이용하여 상기 여러 문서들의 검색 결과의 순위를 정하고, 상기 순위가 정해진 검색 결과를 사용자에게 제공하는 명령어를 더 포함하는

컴퓨터 판독가능 매체.

청구항 2

제1항에 있어서,

문서의 사용 값은 (i) 서버에 의해 유지된 실제 사용 데이터에 기초한 실제 사용 값, 또는 (ii) 실제 사용 데이터에 기초하지 않는 디폴트 사용 값을 포함하는

컴퓨터 판독가능 매체.

청구항 3

제1항에 있어서,

상기 적어도 하나의 질의-독립 성분은 다음 식으로 표현되고:

$$QID(doc) = w_u \frac{k_u U}{k_u + U}$$

여기에서, U는 실제 사용 값 또는 디폴트 사용 값을 나타내고,

w_u 및 k_u 는 사용 값의 조정(tuning) 파라미터를 나타내는

컴퓨터 판독가능 매체.

청구항 4

제1항에 있어서,

상기 적어도 하나의 질의-독립 성분은 (i) 상기 사용 파라미터 및 (ii) 클릭 횟수(click distance) 또는 바이어스된(biased) 클릭 횟수 파라미터를 포함하는

컴퓨터 판독가능 매체.

청구항 5

제1항에 있어서,

상기 적어도 하나의 질의-독립 성분은 사용 파라미터 및 URL 깊이 파라미터를 포함하는 컴퓨터 판독가능 매체.

청구항 6

제1항에 있어서,

상기 각 문서에 대한 점수는 다음 식을 사용하여 생성되고:

$$\text{점수} = \sum \frac{wtf' (k_1 + 1)}{k_1 + wtf'} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ev}} + b_{ud} UD} + w_u \frac{k_u U}{k_u + U}$$

여기에서, wtf'는 가중된 용어(weighted term) 빈도를 나타내고,

N은 네트워크상의 문서들의 수를 나타내며,

n은 질의 용어를 포함하는 문서들의 수를 나타내고,

w_{cd}는 질의-독립 성분의 가중치를 나타내며,

b_{cd}는 클릭 횟수의 가중치를 나타내고,

b_{ud}는 URL 깊이의 가중치를 나타내며,

CD는 문서에 대한 계산된 클릭 횟수 또는 할당된 바이어스된 클릭 횟수를 나타내고,

k_{ev}는 에지 가중치들(edge weights)에 관련된 조정 상수를 나타내며,

UD는 URL 깊이를 나타내고,

U는 실제 사용 값 또는 디폴트 사용 값을 나타내며,

w_u 및 k_u는 사용 값의 조정 파라미터를 나타내고,

k_{cd} 및 k₁은 상수인

컴퓨터 판독가능 매체.

청구항 7

제1항에 있어서,

관리자가 랭킹 함수에 의해 생성된 랭킹 결과들을 수동으로 조정할 수 있게 하는 컴퓨터 실행가능 명령어들을 더 포함하는

컴퓨터 판독가능 매체.

청구항 8

컴퓨터 시스템상에서 사용할 수 있는 적어도 하나의 애플리케이션 모듈을 포함하는 컴퓨팅 시스템에 있어서,

상기 적어도 하나의 애플리케이션 모듈은 제1항의 컴퓨터 판독가능 매체로부터 그 모듈에 로드된 애플리케이션 코드를 포함하는

컴퓨팅 시스템.

청구항 9

네트워크상의 문서들의 순위를 정하는 명령어를 포함하고 하나 이상의 질의-독립 성분을 포함하는 랭킹 함수를

사용하는 컴퓨터 실행가능 명령어를 저장한 컴퓨터 판독가능 매체로서,

상기 컴퓨터 실행가능 명령어는, 적어도 하나의 질의-독립 성분은 네트워크상의 하나 이상의 문서들에 대한 서버-생성, 서버-저장 사용 데이터를 고려한 사용 파라미터를 포함하고, 상기 사용 데이터는 상기 네트워크상의 하나 이상의 문서들과의 실제 사용자 상호작용의 측정을 포함하고,

상기 컴퓨터 실행가능 명령어는 또한 상기 네트워크상의 문서들의 순위를 정하기 위해 상기 랭킹 함수에 의해 생성되고 순서대로 문서들의 순위를 정하기 위해 사용되는 점수를 할당하고, 상기 사용 데이터는 각각의 문서의 사용값을 포함하고, 상기 문서의 사용값은 문서 또는 문서 집합을 포함하는 하나의 폴더의 하나 이상의 사용-관련 속성들에 의존하고, 상기 하나 이상의 사용-관련 속성들은 주어진 기간 내에 사용자들에 의한 문서 또는 폴더 열어보기의 총 수, 주어진 기간 내에 사용자 당 문서 또는 폴더 열어보기의 평균 수, 주어진 기간 내에 특정 문서 또는 폴더에 소비한 총 시간, 주어진 기간 내에 특정 문서 또는 폴더에 소비한 평균 시간을 포함하며, 상기 주어진 기간은 지난주, 지난달, 지난해, 문서 또는 폴더의 수명 또는 임의의 기타 기간을 포함하는

컴퓨터 판독가능 매체.

청구항 10

제9항에 있어서,

상기 하나 이상의 사용 관련 속성들이 주어진 기간 내에 사용자들에 의한 문서 또는 폴더 열어보기의 총 수를 포함하는

컴퓨터 판독가능 매체.

청구항 11

제9항에 있어서,

상기 하나 이상의 사용 관련 속성들이 주어진 기간 내에 사용자 당 문서 또는 폴더 열어보기의 평균 수를 포함하는

컴퓨터 판독가능 매체.

청구항 12

제9항에 있어서,

상기 하나 이상의 사용 관련 속성들이 주어진 기간 내에 특정 문서 또는 폴더에 소비한 총 시간을 포함하는

컴퓨터 판독가능 매체.

청구항 13

제9항에 있어서,

상기 하나 이상의 사용 관련 속성들이 주어진 기간 내에 특정 문서 또는 폴더에 소비한 평균 시간을 포함하는

컴퓨터 판독가능 매체.

청구항 14

네트워크상에서 문서들의 순위를 정하는 컴퓨터로 구현된 방법으로서,

하나 이상의 질의-독립 성분들을 포함하는 랭킹 함수 -- 적어도 하나의 질의-독립 성분은 상기 네트워크상의 하나 이상의 문서들에 대한 서버-생성, 서버-저장 사용 데이터를 고려한 사용 파라미터를 포함하고, 상기 사용 데이터는 상기 네트워크상의 하나 이상의 문서들과의 실제 사용자 상호작용의 측정을 포함함 -- 를 제공하는 단계; 및

네트워크상의 문서들의 순위를 정하기 위해 상기 랭킹 함수에 의해 생성되고 순서대로 문서들의 순위를 정하기 위해 사용되는 점수를 할당하는 단계 -- 상기 사용 데이터는 각각의 문서의 사용값을 포함하고, 상기 문서의 사용값은 문서 또는 문서 집합을 포함하는 하나의 폴더의 하나 이상의 사용-관련 속성들에 의존하고, 상기 하나 이상의 사용-관련 속성들은 주어진 기간 내에 사용자들에 의한 문서 또는 폴더 열어보기의 총 수, 주어진 기간

내에 사용자 당 문서 또는 폴더 열어보기의 평균 수, 주어진 기간 내에 특정 문서 또는 폴더에 소비한 총 시간, 주어진 기간 내에 특정 문서 또는 폴더에 소비한 평균 시간을 포함하며, 상기 주어진 기간은 지난주, 지난달, 지난해, 문서 또는 폴더의 수명 또는 임의의 기타 기간을 포함함 --

를 포함하는 방법.

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

명세서

기술분야

[0001] 본 발명은 네트워크상의 문서에 대한 문서 관련성 점수를 생성하고, 이 문서 관련성 점수를 사용하여, 검색 질의의 검색 결과의 순위를 정하는 랭킹 함수에 관한 것이다.

배경기술

[0002] 주어진 검색 질의에 대한 문서의 관련성에 따라 문서의 순위를 정하는 랭킹 함수가 공지되어 있다. 공지된 랭킹 함수를 사용하여 검색 엔진에 의해 생성된 검색 결과에 비해 주어진 검색 질의에 대한 양호한 검색 결과를 제공하는 랭킹 함수를 개발하기 위한 노력이 본 분야에서 계속되고 있다.

발명의 상세한 설명

[0003] 여기에서 설명된 것은 특히, 네트워크상의 주어진 문서에 대한 문서 관련성 점수를 결정하는 다양한 기술이다. 문서 관련성 점수는 하나 이상의 질의-독립 성분을 포함하는 랭킹 함수를 통해 생성되는데, 적어도 하나의 질의-독립 성분은 네트워크상의 하나 이상의 문서에 대해 웹 서버상에 유지되고 저장된 실제 문서 사용 데이터를 고려한 사용 파라미터를 포함한다. 랭킹 함수는 다수의 문서의 문서 관련성 점수에 기초하여 다수의 문서를 순서대로(전형적으로, 내림 순서로) 순위를 정하기 위해 검색 엔진에 의해 사용될 수 있다.

[0004] 이 요약은 아래의 "상세한 설명" 부분에 설명되는 하나 이상의 선택된 개념들을 단순화된 형태로 독자에게 소개하기 위해 제공된다. 이 요약은 청구된 주제의 중요한 및/또는 요구된 특징을 확인하고자 하는 것이 아니다.

실시예

[0009] 여기에 개시된 방법 및 프로세스의 원리의 이해를 돕기 위해, 특정 실시예가 다음에 설명되고, 특정 언어가 특정 실시예를 설명하기 위해 사용된다. 그렇지만, 개시된 방법 및 프로세스의 범위를 특정 언어의 사용에 의해 제한하고자 하는 것이 아니라는 것을 이해할 것이다. 설명되는 개시된 방법 및 프로세스의 원리에 관한 변경, 또 다른 변형 및 그러한 또 다른 애플리케이션은 개시된 방법 및 프로세스가 속하는 본 분야에 숙련된 기술자에

게 통상적으로 가능한 일이라 생각된다.

- [0010] 네트워크상의 문서에 대한 문서 관련성 점수를 결정하는 방법이 개시된다. 각각의 문서 관련성 점수는 바람직하게, 하나 이상의 질의-독립 성분(예를 들어, 주어진 검색 질의 또는 검색 질의 용어에 의존하지 않는 함수 성분), 하나 이상의 질의-종속 성분(예를 들어, 주어진 검색 질의 또는 검색 질의 용어의 상세에 의존하는 함수 성분), 또는 그 조합을 포함하는 랭킹 함수를 사용하여 계산된다. 랭킹 함수에 의해 결정된 문서 관련성 점수는 각각의 문서 관련성 점수에 따라 네트워크 공간(예를 들어, 회사 인트라넷 공간) 내에서 문서의 순위를 정하기 위해 사용될 수 있다. 개시된 방법이 사용될 수 있는 예시적인 검색 프로세스는 도 1에 예시적인 프로세스(10)로 도시된다.
- [0011] 도 1은 프로세스 단계(80)에서 시작하는 예시적인 검색 프로세스(10)를 도시한 것으로, 사용자는 검색 질의를 입력한다. 단계(80)로부터, 예시적인 검색 프로세스(10)는 단계(200)로 진행하고, 검색 엔진은 검색 질의의 하나 이상의 용어를 찾기 위해 네트워크 공간 내의 모든 문서를 검색한다. 단계(200)로부터, 예시적인 검색 프로세스(10)는 단계(300)로 진행하고, 검색 엔진의 랭킹 함수는 각 문서의 관련성 점수에 기초하여 네트워크 공간 내에서 문서들의 순위를 정하는데, 문서 관련성 점수는 하나 이상의 질의-독립 성분, 하나 이상의 질의-종속 성분, 또는 그 조합에 기초한다. 단계(300)로부터, 예시적인 검색 프로세스(10)는 단계(400)로 진행하고, 순위가 정해진 검색 결과는 전형적으로 내림 순서로 사용자에게 표시되어, 사용자는 검색 질의에 가장 관련있는 네트워크 공간 내의 문서를 식별한다.
- [0012] 아래에 더욱 상세하게 설명되는 바와 같이, 문서 관련성 점수를 결정하는 몇몇 예시적인 방법에서, 문서 관련성 점수를 결정하기 위해 사용된 랭킹 함수의 적어도 하나의 질의-독립 성분은 하나 이상의 사용자에게 의한 네트워크 공간 내의 하나 이상의 문서의 실제 사용에 관련된 "문서 사용 데이터" 또는 "문서 사용 통계치"를 고려한다. 문서 사용 데이터 및/또는 통계치는 주어진 검색 엔진에서 분리된 웹 서버상의 애플리케이션 코드에 의해 생성되어 저장된다. 예를 들어, 문서 사용 데이터는 사용자가 URL을 요청할 때마다, 서버가 사용 카운터를 갱신하도록, 웹 사이트에 의해 유지될 수 있다. 사용 카운터는 지난주, 지난달, 지난해, 또는 주어진 문서 또는 문서 집합의 수명과 같은 주어진 시간 간격 동안 얻어진 문서 관련 데이터를 유지할 수 있다. 애플리케이션 코드는 (i) 특정 애플리케이션 프로그래밍 인터페이스(API)를 통하거나, (ii) 웹 서비스 요청을 통하거나, 또는 (iii) 웹 사이트상의 모든 URL에 대한 사용 데이터를 반환하는 관리 웹 페이지를 요청함으로써, 웹 사이트로부터 사용 데이터를 얻기 위해 사용될 수 있다.
- [0013] 특정 웹 사이트는 로컬 또는 원격 저장 시스템 내에 사용 데이터를 저장할 뿐만 아니라, 네트워크 공간 내에서 사용 데이터를 생성하여 유지하기 위해 사용될 수 있다. 네트워크 공간 내에서 문서의 사용 데이터를 생성하여 유지하고 저장하기 적합한 웹 사이트는 WINDOWS® SHAREPOINT® Services 사이트를 포함하는데, 이것에 제한되는 것은 아니다.
- [0014] 문서 관련성 점수를 결정하는 개시된 방법은 더 나아가, 하나 이상의 추가 질의-독립 성분을 포함하는 랭킹 함수를 이용할 수 있다. 적합한 추가 질의-독립 성분은 2004년 8월 30일자로 출원된 "SYSTEM AND METHOD FOR RANKING SEARCH RESULTS USING CLICK DISTANCE"라는 발명의 명칭의 U.S. 특허 출원 제10/955,983호에서 설명된 네트워크 공간 내의 각 문서에 대한 클릭 횟수(click distance)를 고려한 질의-독립 성분, 2005년 8월 15일자로 출원된 "RANKING FUNCTIONS USING A BIASED CLICK DISTANCE OF A DOCUMENT ON A NETWORK"라는 발명의 명칭의 U.S. 특허 출원 제11/206,286호에서 설명된 네트워크 공간 내의 각 문서에 대한 바이어스된 클릭 횟수를 고려한 질의-독립 성분, 및 2004년 8월 30일자로 출원된 "SYSTEM AND METHOD FOR RANKING SEARCH RESULTS USING CLICK DISTANCE"라는 발명의 명칭의 U.S. 특허 출원 제10/955,983호에서 설명된 네트워크 공간 내의 각 문서에 대한 URL을 고려한 질의-독립 성분을 포함하는데, 이것에 제한되는 것은 아니다. 본 특허 출원의 양수인에게 양도된 상술된 U.S. 특허 출원들의 각각의 주제는 여기에서 전체적으로 참조로 사용된다.
- [0015] 또 다른 예시적인 실시예에서, 문서 관련성 점수를 결정하는 개시된 방법은 상술된 문서 사용 파라미터, 및 상술된 추가 질의-독립 성분들의 하나 이상을 포함하는 적어도 하나의 질의-독립 성분을 포함하는 랭킹 함수를 이용한다.
- [0016] 문서 관련성 점수는 네트워크 공간 내의 문서의 순위를 정하기 위해 사용될 수 있다. 예를 들어, 네트워크상의 문서의 순위를 정하는 방법은 상술된 방법을 사용하여 네트워크상의 각 문서에 대한 문서 관련성 점수를 결정하는 단계; 및 각 문서의 문서 관련성 점수에 기초하여 원하는 순서로(전형적으로, 내림 순서로) 문서를 순위를 정하는 단계를 포함할 수 있다.

- [0017] 문서 관련성 점수는 또한 검색 질의의 검색 결과의 순위를 정하기 위해 사용될 수 있다. 예를 들어, 검색 질의의 검색 결과의 순위를 정하는 방법은 상술된 방법을 사용하여 검색 질의의 검색 결과 내의 각 문서에 대한 문서 관련성 점수를 결정하는 단계; 및 각 문서의 문서 관련성 점수에 기초하여 원하는 순서로(전형적으로, 내림 순서로) 문서를 순위를 정하는 단계를 포함할 수 있다.
- [0018] 여기에서 개시된 방법을 사용하는 애플리케이션 프로그램은 각종 하드웨어 컴포넌트를 포함하는 각종 컴퓨터 시스템상에서 로드되어 실행될 수 있다. 여기에 개시된 방법을 실시하는 예시적인 컴퓨터 시스템 및 예시적인 운영 환경이 아래에 설명된다.
- [0019] 예시적인 운영 환경
- [0020] 도 2는 여기에서 개시된 방법이 구현될 수 있는 적합한 컴퓨팅 시스템 환경(100)의 예를 도시한 것이다. 컴퓨팅 시스템 환경(100)은 적합한 컴퓨팅 환경의 일례에 불과하며, 여기에 개시된 방법의 용도 또는 기능성의 범위에 관해 어떤 제한을 암시하고자 하는 것이 아니다. 컴퓨팅 환경(100)이 예시적인 운영 환경(100)에 도시된 컴포넌트들 중 임의의 하나 또는 그 컴포넌트들의 임의의 조합과 관련하여 어떤 의존성 또는 요구사항을 갖는 것으로 해석되어서는 안 된다.
- [0021] 여기에 개시된 방법은 많은 기타 범용 또는 특수 목적의 컴퓨팅 시스템 환경 또는 구성에서 동작할 수 있다. 여기에 개시된 방법에서 사용하는 데 적합할 수 있는 잘 알려진 컴퓨팅 시스템, 환경 및/또는 구성의 예로는 퍼스널 컴퓨터, 서버 컴퓨터, 핸드-헬드 또는 랩톱 장치, 멀티프로세서 시스템, 마이크로프로세서 기반 시스템, 셋톱 박스, 프로그램가능한 가전제품, 네트워크 PC, 미니컴퓨터, 메인프레임 컴퓨터, 상기 시스템들이나 장치들 중 임의의 것을 포함하는 분산 컴퓨팅 환경, 기타 등등이 있지만 이에 제한되는 것은 아니다.
- [0022] 여기에 개시된 방법 및 프로세스는 일반적으로 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터 실행가능 명령어와 관련하여 기술될 것이다. 일반적으로, 프로그램 모듈은 특정 태스크를 수행하거나 특정 추상 데이터 유형을 구현하는 루틴, 프로그램, 개체, 컴포넌트, 데이터 구조 등을 포함한다. 여기에 개시된 방법 및 프로세스는 또한 통신 네트워크를 통해 연결되어 있는 원격 처리 장치들에 의해 태스크가 수행되는 분산 컴퓨팅 환경에서 실시되도록 설계된다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 메모리 저장 장치를 비롯한 로컬 및 원격 컴퓨터 저장 매체 둘 다에 위치할 수 있다.
- [0023] 도 2와 관련하여, 여기에 개시된 방법 및 프로세스를 구현하는 예시적인 시스템은 컴퓨터(110) 형태의 범용 컴퓨팅 장치를 포함한다. 컴퓨터(110)의 컴포넌트들은 처리 장치(120), 시스템 메모리(130), 및 시스템 메모리(130)를 비롯한(시스템 메모리에 제한되는 것은 아님) 각종 시스템 컴포넌트들을 처리 장치(120)에 연결하는 시스템 버스(121)를 포함하지만 이에 제한되는 것은 아니다. 시스템 버스(121)는 메모리 버스 또는 메모리 컨트롤러, 주변 장치 버스 및 각종 버스 아키텍처 중 임의의 것을 이용하는 로컬 버스를 비롯한 몇몇 유형의 버스 구조 중 어느 것이라도 될 수 있다. 예로서, 이러한 아키텍처는 ISA(industry standard architecture) 버스, MCA(micro channel architecture) 버스, EISA(Enhanced ISA) 버스, VESA(video electronics standard association) 로컬 버스, 그리고 메자닌 버스(mezzanine bus)로도 알려진 PCI(peripheral component interconnect) 버스 등을 포함하지만 이에 제한되는 것은 아니다.
- [0024] 컴퓨터(110)는 통상적으로 각종 컴퓨터 판독가능 매체를 포함한다. 컴퓨터(110)에 의해 액세스 가능한 매체는 그 어떤 것이든지 컴퓨터 판독가능 매체가 될 수 있고, 이러한 컴퓨터 판독가능 매체는 휘발성 및 비휘발성 매체, 이동식 및 비이동식 매체를 포함한다. 예로서, 컴퓨터 판독가능 매체는 컴퓨터 저장 매체 및 통신 매체를 포함하지만 이에 제한되는 것은 아니다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보를 저장하는 임의의 방법 또는 기술로 구현되는 휘발성 및 비휘발성, 이동식 및 비이동식 매체를 포함한다. 컴퓨터 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 기타 메모리 기술, CD-ROM, DVD(digital versatile disk) 또는 기타 광 디스크 저장 장치, 자기 카세트, 자기 테이프, 자기 디스크 저장 장치 또는 기타 자기 저장 장치, 또는 컴퓨터(110)에 의해 액세스되고 원하는 정보를 저장할 수 있는 임의의 기타 매체를 포함하지만 이에 제한되는 것은 아니다. 통신 매체는 통상적으로 반송파(carrier wave) 또는 기타 전송 메커니즘(transport mechanism)과 같은 피변조 데이터 신호(modulated data signal)에 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터 등을 구현하고 모든 정보 전달 매체를 포함한다. "피변조 데이터 신호"라는 용어는, 신호 내에 정보를 인코딩하도록 그 신호의 특성들 중 하나 이상을 설정 또는 변경시킨 신호를 의미한다. 예로서, 통신 매체는 유선 네트워크 또는 직접 배선 접속(direct-wired connection)과 같은 유선 매체, 그리고 음향, RF, 적외선, 기타 무선 매체와 같은 무선 매체를 포함하지만 이에 제한되는 것은 아니다. 상술된 매체들의 모든 조합이 또한 여기에서 사용된 컴퓨터 판독가능 매체의 영역 안에

포함되는 것으로 한다.

- [0025] 시스템 메모리(130)는 판독 전용 메모리(ROM)(131) 및 랜덤 액세스 메모리(RAM)(132)와 같은 휘발성 및/또는 비휘발성 메모리 형태의 컴퓨터 저장 매체를 포함한다. 시동 중과 같은 때에, 컴퓨터(110) 내의 구성요소들 사이의 정보 전송을 돕는 기본 루틴을 포함하는 기본 입/출력 시스템(BIOS)(133)은 통상적으로 ROM(131)에 저장되어 있다. RAM(132)은 통상적으로 처리 장치(120)가 즉시 액세스 할 수 있고 및/또는 현재 동작시키고 있는 데이터 및/또는 프로그램 모듈을 포함한다. 예로서, 도 2는 운영 체제(134), 애플리케이션 프로그램(135), 기타 프로그램 모듈(136) 및 프로그램 데이터(137)를 도시하고 있지만 이에 제한되는 것은 아니다.
- [0026] 컴퓨터(110)는 또한 기타 이동식/비이동식, 휘발성/비휘발성 컴퓨터 저장매체를 포함한다. 단지 예로서, 도 2는 비이동식·비휘발성 자기 매체에 기록을 하거나 그로부터 판독을 하는 하드 디스크 드라이브(141), 이동식·비휘발성 자기 디스크(152)에 기록을 하거나 그로부터 판독을 하는 자기 디스크 드라이브(151), 및 CD-ROM 또는 기타 광 매체 등의 이동식·비휘발성 광 디스크(156)에 기록을 하거나 그로부터 판독을 하는 광 디스크 드라이브(155)를 포함한다. 예시적인 운영 환경에서 사용될 수 있는 기타 이동식/비이동식, 휘발성/비휘발성 컴퓨터 기억 매체로는 자기 테이프 카세트, 플래시 메모리 카드, DVD, 디지털 비디오 테이프, 고상(solid state) RAM, 고상 ROM 등이 있지만 이에 제한되는 것은 아니다. 하드 디스크 드라이브(141)는 통상적으로 인터페이스(140)와 같은 비이동식 메모리 인터페이스를 통해 시스템 버스(121)에 접속되고, 자기 디스크 드라이브(151) 및 광 디스크 드라이브(155)는 통상적으로 인터페이스(150)와 같은 이동식 메모리 인터페이스에 의해 시스템 버스(121)에 접속된다.
- [0027] 위에서 설명되고 도 2에 도시된 드라이브들 및 이들과 관련된 컴퓨터 저장 매체는, 컴퓨터(110)에 대한 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 및 기타 데이터를 저장한다. 도 2에서, 예를 들어, 하드 디스크 드라이브(141)는 운영 체제(144), 애플리케이션 프로그램(145), 기타 프로그램 모듈(146), 및 프로그램 데이터(147)를 저장하는 것으로 도시되어 있다. 여기서 주의할 점은 이들 컴포넌트가 운영 체제(134), 애플리케이션 프로그램(135), 기타 프로그램 모듈(136), 및 프로그램 데이터(137)와 동일하거나 그와 다를 수 있다는 것이다. 이에 관해, 운영 체제(144), 애플리케이션 프로그램(145), 기타 프로그램 모듈(146) 및 프로그램 데이터(147)에 다른 번호가 부여되어 있다는 것은 적어도 이들이 다른 사본(copy)이라는 것을 나타내기 위한 것이다.
- [0028] 사용자는 키보드(162), 및 마우스, 트랙볼(trackball) 또는 터치 패드라 보통 칭해지는 포인팅 장치(161) 등의 입력 장치를 통해 명령 및 정보를 컴퓨터(110)에 입력할 수 있다. 다른 입력 장치(도시 생략)로는 마이크, 조이스틱, 게임 패드, 위성 안테나, 스캐너 등을 포함할 수 있다. 이들 및 기타 입력 장치는 종종 시스템 버스(121)에 결합된 사용자 입력 인터페이스(160)를 통해 처리 장치(120)에 접속되지만, 병렬 포트, 게임 포트 또는 USB(universal serial bus) 등의 다른 인터페이스 및 버스 구조에 의해 접속될 수도 있다. 모니터(191) 또는 다른 유형의 디스플레이 장치도 비디오 인터페이스(190) 등의 인터페이스를 통해 시스템 버스(121)에 접속된다. 모니터(191) 외에, 컴퓨터(110)는 또한 스피커(197) 및 프린터(196) 등의 기타 주변 출력 장치를 포함할 수 있고, 이들은 출력 주변장치 인터페이스(195)를 통해 접속될 수 있다.
- [0029] 컴퓨터(110)는 원격 컴퓨터(180)와 같은 하나 이상의 원격 컴퓨터로의 논리적 접속을 사용하여 네트워크화된 환경에서 동작할 수 있다. 도 2에는 메모리 저장 장치(181)만이 도시되어 있지만, 원격 컴퓨터(180)는 또 하나의 퍼스널 컴퓨터, 서버, 라우터, 네트워크 PC, 피어 장치 또는 기타 통상의 네트워크 노드일 수 있고, 통상적으로 컴퓨터(110)와 관련하여 상술된 구성요소들의 대부분 또는 그 전부를 포함한다. 도 2에 도시된 논리적 접속으로는 LAN(171) 및 WAN(173)이 있지만, 기타 네트워크를 포함할 수도 있다. 이러한 네트워크 환경은 사무실, 전사적 컴퓨터 네트워크(enterprise-wide computer network), 인트라넷, 및 인터넷에서 일반적인 것이다.
- [0030] LAN 네트워킹 환경에서 사용될 때, 컴퓨터(110)는 네트워크 인터페이스 또는 어댑터(170)를 통해 LAN(171)에 접속된다. WAN 네트워킹 환경에서 사용될 때, 컴퓨터(110)는 통상적으로 인터넷과 같은 WAN(173)을 통해 통신을 설정하기 위한 모뎀(172) 또는 기타 수단을 포함한다. 내장형 또는 외장형일 수 있는 모뎀(172)은 사용자 입력 인터페이스(160) 또는 기타 적절한 메커니즘을 통해 시스템 버스(121)에 접속될 수 있다. 네트워크화된 환경에서, 컴퓨터(110) 또는 그의 일부와 관련하여 기술된 프로그램 모듈은 원격 메모리 저장 장치에 저장될 수 있다. 예로서, 도 2는 원격 애플리케이션 프로그램(185)이 메모리 장치(181)에 있는 것으로 도시하고 있지만 이에 제한되는 것은 아니다. 도시된 네트워크 접속은 예시적인 것이며 이 컴퓨터들 사이에 통신 링크를 설정하는 기타 수단이 사용될 수 있다는 것을 이해할 것이다.
- [0031] 여기에 개시된 방법 및 프로세스는 서버 시스템 소프트웨어 애플리케이션(예를 들어, WINDOWS SERVER SYSTEMTM 소프트웨어 애플리케이션), 검색 랭킹 애플리케이션, 및 네트워크 공간 내의 문서의 사용 데이터를 생성하여 유

지하고 저장하는 애플리케이션(예를 들어, WINDOWS[®] SHAREPOINT[®] Services 애플리케이션)을 포함하는(이에 제한되지는 않음) 하나 이상의 애플리케이션 프로그램을 사용하여 구현될 수 있는데, 이들 중 어느 하나는 예시적인 시스템(100) 내에 애플리케이션 프로그램(135), 애플리케이션 프로그램(145) 및 원격 애플리케이션 프로그램(185)으로 표시된 다양한 애플리케이션 프로그램들 중의 하나일 수 있다.

[0032] 상술된 바와 같이, 본 분야에 숙련된 기술자들은 주어진 문서에 대한 문서 관련성 점수를 생성하는 개시된 방법이 핸드헬드 장치, 멀티프로세서 시스템, 마이크로프로세서 기반 또는 프로그램가능 소비자 전자제품, 네트워크화된 퍼스널 컴퓨터, 미니컴퓨터, 메인프레임 컴퓨터 등을 포함한 기타 컴퓨터 시스템 구성에서 구현될 수 있다는 것을 이해할 것이다. 주어진 문서에 대한 문서 관련성 점수를 생성하는 개시된 방법은 또한 통신 네트워크를 통해 연결되는 원격 처리 장치에 의해 태스크가 수행되는 분산 컴퓨팅 환경에서 실시될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 로컬 및 원격 메모리 저장 장치 내에 위치할 수 있다.

[0033] 예시적인 실시예의 구현

[0034] 상술된 바와 같이, 네트워크상의 문서에 대한 문서 관련성 점수를 결정하는 방법이 제공된다. 개시된 방법은 네트워크상의 각 문서의 문서 사용 값을 고려하는 랭킹 함수를 이용하여 네트워크상의 문서의 순위를 정할 수 있다.

[0035] 네트워크상의 문서에 대한 문서 관련성 점수를 결정하는 개시된 방법은 여러 단계를 포함할 수 있다. 한 예시적인 실시예에서, 네트워크상의 문서에 대한 문서 관련성 점수를 결정하는 방법은 N개의 문서를 포함하는 네트워크상의 하나 이상의 문서에 실제 사용 값(U_A)을 할당하는 단계 - 실제 사용 값(U_A)은 서버상에 유지되고 저장된 실제 사용 데이터에 기초함-; N개 미만의 문서에 실제 사용 값(U_A)이 할당되면, 문서와 관련된 실제 사용 데이터를 갖지 않는 문서에 디폴트 사용 값(U_D)을 할당하는 단계; 및 네트워크상의 주어진 문서의 문서 관련성 점수를 결정하기 위해 각 문서에 대한 사용 값(즉, U_A 또는 U_D)을 사용하는 단계를 포함한다.

[0036] 여기에서 사용된 바와 같이, "실제 사용 데이터"라는 용어는 한 명 이상의 사용자에게 의한 문서의 "사용"과 관련된 한 가지 이상의 유형의 데이터를 나타낸다. 주어진 문서 또는 문서 집합에 대한 실제 사용 데이터의 유형은 주어진 기간 내에 모든 사용자에게 의한 문서 열람의 수, 주어진 기간 내에 사용자 당 문서 열람의 평균 수, 주어진 기간 내에 특정 문서에 소비한 총 시간, 주어진 기간 내에 특정 문서에 소비한 평균 시간 등을 포함할 수 있는데, 이에 제한되는 것은 아니다. 주어진 기간은, 예를 들어 지난주, 지난달, 지난해, 문서의 수명, 또는 임의의 기타 원하는 기간일 수 있다.

[0037] 네트워크 공간 내의 문서들에 대한 문서 사용 데이터 또는 통계치를 생성하여 유지하고 저장하는 단계는 일반적으로 컴퓨팅 시스템상에 있는 애플리케이션 코드에 의해 실행될 수 있다. 문서 사용 데이터는 주어진 검색 질의 또는 검색 엔진에 관계없이 생성되어 유지되고 저장되며, 전형적으로 문서(또는 페이지)를 유지하여 문서(또는) 페이지가 사용자에게 이용가능하게 하는 서버상의 애플리케이션 코드에 의해 생성되어 유지되고 저장된다. 문서 사용 데이터 또는 통계치를 생성하여 유지하고 저장하기 적합한 애플리케이션 프로그램은 WINDOWS[®] SHAREPOINT[®] Service 및 기타 유사한 애플리케이션 프로그램을 포함하지만 이에 제한되는 것은 아니다.

[0038] 유사한 기능을 실행하는 기타 웹 사이트뿐만 아니라 이들 서비스 사이트상에 저장되고 유지된 문서 사용 데이터는 상술된 애플리케이션 코드를 사용하여 액세스될 수 있다. 예를 들어, 문서 사용 데이터는 (i) 특정 애플리케이션 프로그래밍 인터페이스(API)를 통하거나, (ii) 웹 서비스 요청을 통하거나, 또는 (iii) 웹 사이트상의 모든 URL에 대한 사용 데이터를 반환하는 관리 웹 페이지를 요청함으로써, 주어진 웹 사이트(예를 들어, WINDOWS[®] SHAREPOINT[®] Service 사이트)로부터 액세스될 수 있다.

[0039] 네트워크상의 문서에 대한 문서 관련성 점수를 결정하는 개시된 방법은 실제 문서 사용을 위해 네트워크 공간 내에서 하나 이상의 문서를 모니터링하는 단계; 하나 이상의 문서에 대한 실제 문서 사용 데이터를 로컬 또는 원격 데이터 저장 파일 내에 저장하는 단계; 문서 또는 문서를 포함하는 폴더에 대한 실제 사용 데이터에 기초하여 문서에 대한 실제 사용 값(U_A)을 계산하는 단계; 하나 이상의 문서에 대한 실제 사용 값(U_A)을 로컬 또는 원격 데이터 저장 파일 내에 저장하는 단계; 저장된 문서 사용 데이터 또는 실제 사용 값(U_A)을 로컬 또는 원격 데이터 저장 파일로부터 요청하는 단계(예를 들어, 사용자에게 의한 특정 검색 질의 후 검색 엔진으로부터의 그러한 데이터에 대한 요청); 하나 이상의 문서에 대한 실제 문서 사용 데이터 또는 실제 사용 값(U_A)을 로컬 또는

원격 데이터 저장 파일로부터 검색하는 단계; 및 선택적으로, 문서에 대한 문서 관련성 점수를 결정하기 위해 문서 사용 값(즉, 실제 또는 디폴트)을 하나 이상의 추가 문서 속성과 병합하는 단계를 포함하는(이에 제한되지는 않음) 다수의 추가 단계를 포함할 수 있다.

[0040] 도 3은 네트워크상의 문서에 대한 실제 또는 디폴트 사용 값을 제공한 다음에 시스템 관리자에 의해 그 값을 선택적으로 떨어뜨리고/높이는 절차를 행하는 예시적인 방법의 예시적인 단계를 도시한 논리 흐름도를 나타낸다. 도 3에 도시된 바와 같이, 예시적인 방법(401)은 블록(402)에서 시작하여 단계(403)로 진행한다. 단계(403)에서, 네트워크상의 제1 문서는 실제 사용 데이터를 찾기 위해 탐색된다.

[0041] 실제 사용 데이터를 찾기 위해 제1 문서를 탐색하는 단계(단계(403))는 제1 문서가 그와 관련된 임의의 실제 사용 데이터를 갖는지 판정하여, 제1 문서가 그와 관련된 실제 사용 데이터를 가지면, 실제 사용 데이터를 검색할 수 있는 크롤러(crawler) 애플리케이션을 사용하여 실행될 수 있다. 네트워크상의 문서에 대한 실제 또는 디폴트 사용 값을 제공하는 개시된 방법에서 사용하기 적합한 크롤러 애플리케이션은 U.S. 특허 제6,463,455호 및 제6,631,369호에서 설명된 크롤러 애플리케이션을 포함하지만 이에 제한되는 것은 아니고, 이들 특허의 주제는 여기에서 전체적으로 참조로 사용된다.

[0042] 상술된 바와 같이, 실제 사용 데이터는 네트워크상의 하나 이상의 문서에 대한 실제 사용 데이터를 저장하는 하나 이상의 파일로부터 얻어질 수 있다. 실제 사용 데이터는 문서 구성요소로서 문서와 함께 저장될 수 있고, 또는 실제 문서에서 분리된 데이터 저장 파일에 저장될 수 있다. 적합한 원격 저장 시스템은 임의의 기타 유사한 원격 저장 시스템뿐만 아니라, 마이크로소프트사(워싱턴 레드몬드 소재)에서 시판되는 WINDOWS[®] SHAREPOINT[®] Service(WSS) 제품을 포함하지만 이에 제한되는 것은 아니다. 예를 들어, WSS 원격 저장 시스템은 예를 들어, 모든 사용자에게 걸쳐있는 주어진 네트워크상의 모든 문서에 대한 요청의 수를 포함하는 실제 사용자 데이터를 기록하고, 지난주, 지난달, 지난해 또는 문서의 전체 수명, 또는 임의의 기타 기간 동안에 문서 당 클릭 수의 통계치를 생성한다. 더욱이, 상술된 바와 같이, 여기에 개시된 방법은 WSS 원격 저장 시스템에 제한되는 것이 아니라, 개시된 방법에서 WSS 원격 저장 시스템 또는 임의의 기타 유사한 문서 데이터 시스템을 이용할 수 있다는 것을 알 수 있을 것이다.

[0043] 일단 문서가 탐색되면, 예시적인 방법(401)은 판정 블록(404)으로 진행한다. 판정 블록(404)에서, 문서가 그와 관련된 실제 사용 데이터를 갖는지의 여부에 관해 애플리케이션 코드에 의해 판정이 이루어진다. 문서가 그와 관련된 실제 사용 데이터를 갖는다고 판정이 이루어지면, 예시적인 방법(401)은 단계(405)로 진행하여, 실제 사용에 기초한 사용 값(U_A)이 문서에 할당된다. 실제 사용 값(U_A)은 문서와 관련된 실제 사용 데이터의 하나 이상의 성분을 사용하여 결정될 수 있다. 예를 들어, 몇몇 실시예에서, 문서에 할당된 실제 사용 값(U_A)은 문서를 열어보는 사용자들의 수에만 관련될 수 있다. 다른 실시예에서, 문서에 할당된 실제 사용 값(U_A)은 주어진 기간 내에 모든 사용자에 의한 문서 열어보기의 수, 주어진 기간 내의 사용자 당 문서 열어보기의 평균 수, 주어진 기간 내에 특정 문서에 소비한 총 시간, 주어진 기간 내에 특정 문서에 소비한 평균 시간, 또는 상기 기준들 중 임의의 기준의 조합에 관련될 수 있는데, 주어진 기간은 지난주, 지난달, 지난해, 문서의 수명, 또는 임의의 기타 원하는 기간을 포함한다.

[0044] 몇몇 경우에, 주어진 문서와 관련된 실제 사용 데이터는 문서가 주어진 기간 동안에 사용되지 않았거나 열려보게 되지 않았음을 암시한다. 그러한 경우에, 문서는 그 기간 동안에 무사용을 나타내기 위해 0과 동일한 사용 값(U_A)이 할당될 수 있지만; 전형적으로, 실제 사용 또는 무 실제 사용에 기초한 사용 값(U_A)은 0 이외의 다른 수가 할당된다.

[0045] 더욱이, 몇몇 경우에, 실제 사용 데이터는 개별 문서와 반대인 문서 집합과 관련될 수 있다. 예를 들어, 폴더는 문서 집합을 포함할 수 있고, 관련된 서버는 폴더 내의 개별 문서가 아니라, 폴더를 액세스하는 것(즉, 폴더의 사용)에 관련된 사용자 데이터만을 추적할 수 있다. 이 실시예에서, 폴더와 관련된 실제 사용 데이터가 있으면, 사용 값(U_A)은 폴더의 실제 사용 데이터에 기초하여 폴더 내의 각 문서에 제공될 수 있다. 전형적으로, 각 사용 값(U_A)은 폴더 내의 각 문서에 대해 동일하겠지만; 원한다면, 상이한 사용 값(U_A)이 폴더 내의 상이한 문서에 할당될 수도 있다.

[0046] 단계(405)로부터, 예시적인 방법(401)은 아래에 설명되는 판정 블록(406)으로 진행한다.

[0047] 판정 블록(404)으로 돌아가서, 문서가 그와 관련된 실제 사용 데이터를 갖지 않는다고 판정이 이루어지면, 예시

적인 방법(401)은 단계(407)로 진행하여, 디폴트 사용 값(U_D)이 문서에 할당된다. 예를 들어, 디폴트 사용 값(U_D)은 문서 사용 데이터를 유지하지 않는 웹 사이트의 일부분인 문서에 할당될 수 있다. 문서에 할당된 디폴트 사용 값(U_D)은 실제 사용 데이터를 갖는 문서에 관련하여 실제 사용 데이터를 갖지 않는 문서에 초기 중요도를 제공하기 위해 사용될 수 있다. 예를 들어, 주어진 문서에 대한 높은 사용 값이 네트워크 내의 문서의 상대 중요도를 나타내는 경우, 그 문서에 낮은 디폴트 사용 값(U_D)을 할당하는 것은 네트워크상의 그외 다른 문서에 비해 그 문서의 중요도를 떨어뜨린다.

[0048] 주어진 문서에 대한 높은 사용 값이 네트워크 내의 문서의 상대 중요도를 나타내는 한 예시적인 실시예에서, 그 문서에 할당된 디폴트 사용 값(U_D)은 네트워크상의 그외 다른 문서에 할당된 실제 사용 값(U_A)에 상대적일 수 있다. 예를 들어, 문서의 상대 중요도를 낮추기 위해, 디폴트 사용 값(U_D)이 그 문서에 할당될 수 있는데, 이때 디폴트 사용 값(U_D)은 상술된 네트워크상의 그외 다른 문서에 할당된 임의의 실제 사용 값(U_A)보다 작다. 문서의 상대 중요도를 증가시키고 싶은 경우에, 디폴트 사용 값(U_D)이 그 문서에 할당될 수 있는데, 이때 디폴트 사용 값(U_D)은 네트워크상의 그외 다른 문서에 할당된 임의의 실제 사용 값(U_A)보다 크거나, 또는 네트워크상의 그외 다른 문서의 일부 문서에 할당된 실제 사용 값(U_A)의 일부 사용 값보다 크다.

[0049] 다른 실시예에서, 디폴트 사용 값(U_D)은 할당된 실제 사용 값(U_A)을 갖는 문서에 비해 실제 사용 데이터가 없는 문서에 평균 상대 중요도가 주어지도록, 실제 사용 데이터가 없는 문서에 할당될 수 있다. 예를 들어, 이 실시예에서, 실제 사용 데이터가 없는 문서에 대한 디폴트 사용 값(U_D)은 최소 할당 실제 사용 값(U_{Amin})에서 최대 할당 실제 사용 값(U_{Amax})까지의 범위에 있을 수 있고, 또는 최소 할당 실제 사용 값(U_{Amin})과 최대 할당 실제 사용 값(U_{Amax}) 사이의 특정 범위 내에 있을 수 있다. 이 실시예에서, 실제 사용 데이터가 없는 문서는 관련된 실제 사용 데이터를 갖는 문서에 비해, 중간 사용을 암시하는 평균 상대 중요도가 제공된다.

[0050] 단계(407)로부터, 예시적인 방법(401)은 판정 블록(406)으로 진행한다. 판정 블록(406)에서, 네트워크상의 모든 문서가 실제(U_A) 또는 디폴트(U_D) 사용 값을 갖는지에 관해 애플리케이션 코드에 의해 판정이 이루어진다. 네트워크상의 모든 문서가 실제(U_A) 또는 디폴트(U_D) 사용 값을 갖지 않는다고 판정이 이루어지면, 예시적인 방법(401)은 단계(408)로 진행하여, 실제 사용 데이터를 찾기 위해 다음 문서가 탐색된다. 단계(408)로부터, 예시적인 방법(401)은 판정 블록(404)으로 되돌아가서 상술된 바와 같이 처리해간다.

[0051] 판정 블록(406)으로 돌아가서, 네트워크상의 모든 문서가 실제(U_A) 또는 디폴트(U_D) 사용 값을 갖는다고 애플리케이션 코드에 의해 판정이 이루어지면, 예시적인 방법(401)은 판정 블록(409)으로 진행한다. 판정 블록(409)에서, 네트워크 공간 내에서 주어진 문서의 중요도를 더욱 세밀하게 나타내기 위해 임의의 실제(U_A) 또는 디폴트(U_D) 사용 값을 떨어뜨릴 것인지에 관해 시스템 관리자에 의해 판정이 이루어진다. 네트워크 공간 내에서 하나 이상의 문서의 중요도를 더욱 세밀하게 나타내기 위해 하나 이상의 실제(U_A) 또는 디폴트(U_D) 사용 값을 떨어뜨리는 것으로 판정이 이루어지면, 예시적인 방법(401)은 단계(410)로 진행하여, 하나 이상의 문서(또는 URL)의 실제(U_A) 또는 디폴트(U_D) 사용 값이 음으로 또는 양으로 조정된다. 단계(410)로부터, 예시적인 방법(401)은 아래에 설명되는 단계(411)로 진행한다.

[0052] 판정 단계(409)로 돌아가서, 하나 이상의 실제(U_A) 또는 디폴트(U_D) 사용 값을 떨어뜨리지(또는 높이지) 않는 것으로 판정이 이루어지면, 예시적인 방법(401)은 바로 단계(411)로 진행한다. 단계(411)에서, 실제(U_A) 및 디폴트(U_D) 사용 값은 네트워크 공간 내에서 각 문서에 대한 전체 문서 관련성 점수를 결정하기 위해 랭킹 함수에서 이용된다. 단계(411)로부터, 예시적인 방법(401)은 종료 블록(412)으로 진행한다.

[0053] 일단 모든 실제(U_A) 및 디폴트(U_D) 사용 값이 결정되고, 원하는 경우에, 선택적으로 떨어졌으면(또는 선택적으로 높아졌으면), 각 문서에 대한 실제(U_A) 또는 디폴트(U_D) 사용 값은 각 문서에 대한 문서 관련성 점수를 제공하기 위해 랭킹 함수에서 파라미터로서 사용될 수 있다. 그러한 문서 관련성 점수는 검색 질의의 검색 결과의 순위를 정하기 위해 사용될 수 있다. 문서 사용 값 파라미터를 포함하는 랭킹 함수를 사용하여 생성된 검색 결과의 순위를 정하는 예시적인 방법은 도 4에 도시된다.

[0054] 도 4는 예시적인 방법(20)의 예시적인 단계를 도시한 논리 흐름도를 제공하는 것으로, 예시적인 방법(20)은 사용 값 파라미터를 포함하는 랭킹 함수를 사용하여 생성하는 검색 결과의 순위를 정하는 방법을 포함한다. 도 4에 도시된 바와 같이, 예시적인 방법(20)은 블록(201)에서 시작하고, 단계(202)로 진행한다. 단계(202)에서, 사용자는 검색 질의를 입력함으로써 검색을 요청한다. 단계(202) 이전에, 네트워크상의 문서들의 각각에 대한 실제 또는 디폴트 사용 값은 미리 계산되어 있다. 단계(202)로부터, 예시적인 방법(20)은 단계(203)로 진행한다.

[0055] 단계(203)에서, 네트워크상의 각 문서에 대한 실제 또는 디폴트 값은 색인(index)에 저장된 각 문서에 대한 임의의 기타 문서 통계치(예를 들어, 기타 질의-독립 통계치)와 병합된다. 기타 문서 통계치와의 실제 또는 디폴트 사용 값의 병합은 랭킹에 관련된 모든 정보가 함께 묶여 생성되기 때문에 질의 응답 시간을 더 빠르게 할 수 있다. 따라서, 색인에 열거된 각 문서는 병합 후 관련된 실제 또는 디폴트 사용 값을 갖는다. 일단 병합이 완료되면, 예시적인 방법(20)은 단계(204)로 진행한다.

[0056] 단계(204)에서, 사용 파라미터를 비롯하여 주어진 문서에 대한 질의-독립 문서 통계치는 랭킹 함수의 성분으로서 제공된다. 질의-종속 데이터는 전형적으로 랭킹 함수의 분리된 성분으로서, 또한 주어진 문서에 대해 제공된다. 랭킹 함수의 질의-종속 데이터 또는 내용-관련 부분은 실제 검색 용어 및 주어진 문서의 내용에 의존한다.

[0057] 한 실시예에서, 랭킹 함수는 사용 파라미터를 포함하는 적어도 하나의 질의-독립(QID) 성분을 포함한다. 한 실시예에서, 질의-독립(QID) 성분은 다음 수학식으로 표현될 수 있다:

수학식 1

$$QID(doc)=w_u \frac{k_u U}{k_u + U}$$

[0058]

[0059] 여기에서, U는 실제 사용 값 또는 디폴트 사용 값을 나타내고;

[0060] w_u 및 k_u 는 사용 값의 조정(tuning) 파라미터를 나타낸다. 다른 실시예에서, 질의-독립(QID) 성분은 다음 수학식으로 표현될 수 있다:

수학식 2

$$QID(doc)=w_u U+k_u$$

[0061]

[0062] 여기에서, U는 실제 사용 값 또는 디폴트 사용 값을 나타내고;

[0063] w_u 및 k_u 는 사용 값의 조정 파라미터를 나타낸다. 또 다른 실시예에서, 질의-독립(QID) 성분은 다음 수학식으로 표현될 수 있다:

수학식 3

$$QID(doc)=w_u [1+\exp(-k_u U-B)]+C$$

[0064]

[0065] 여기에서, U는 실제 사용 값 또는 디폴트 사용 값을 나타내고;

[0066] w_u , k_u , B 및 C는 사용 값의 조정 파라미터(즉, 스칼라 상수)를 나타낸다.

[0067] 다른 실시예에서, 랭킹 함수는 상술된 질의-독립(QID) 성분과 다음과 같은 적어도 하나의 질의-종속(QD) 성분의 합을 포함한다:

$$\text{점수} = QD(doc, query) + QID(doc)$$

[0068]

[0069] QD 성분은 임의의 문서 점수화 함수일 수 있다. 한 실시예에서, QD 성분은 2004년 3월 18일자로 출원된 "FIELD WEIGHTING IN TEXT DOCUMENT SEARCHING"이라는 발명의 명칭의 U.S. 특허 출원 제10/804,326호에 설명된 필드 가중 점수화 함수에 대응하고, 그 주제는 여기에서 전체적으로 참조로 사용된다. U.S. 특허 출원 제10/804,326

호에서 제공된 바와 같이, 필드 가중 점수화 함수의 표현으로서 사용될 수 있는 한 수학식은 다음과 같다:

$$QD(doc, query) = \sum \frac{wtf'(k_1 + 1)}{k_1 + wtf'} \times \log\left(\frac{N}{n}\right)$$

[0070]

[0071] 여기에서, wtf'는 모든 필드(예를 들어, 문서의 제목, 본문 등)에 걸쳐 가중치에 의해 곱해져서 각 필드의 길이 및 대응하는 평균 길이에 따라 정규화된 검색 질의 내의 주어진 용어들의 가중된 용어 빈도 또는 용어 빈도들의 합을 나타내고,

[0072] N은 네트워크상의 문서들의 수를 나타내며,

[0073] n은 질의 용어를 포함하는 문서들의 수를 나타내고,

[0074] k₁은 조정가능 상수이다.

[0075] 상기 항들 및 수학식은 U.S. 특허 출원 제10/804,326호에서 더욱 상세하게 설명되고, 그 주제는 여기에서 참조로 사용된다.

[0076] 몇몇 실시예에서, 랭킹 함수는 (i) 2004년 8월 30일자로 출원된 "SYSTEM AND METHOD FOR RANKING SEARCH RESULTS USING CLICK DISTANCE"라는 발명의 명칭의 U.S. 특허 출원 제10/955,983호에 개시된 방법에 의해 결정된 클릭 횟수 값, (ii) 2005년 8월 15일자로 출원된 "RANKING FUNCTIONS USING A BIASED CLICK DISTANCE OF A DOCUMENT ON A NETWORK"라는 발명의 명칭의 U.S. 특허 출원 제11/206,286호에 개시된 방법에 의해 결정된 바이어스된 클릭 횟수 값 - 이들 출원의 주제는 여기에서 전체적으로 참조로 사용됨-, (iii) 문서의 URL 길이, 또는 (iv) (i) 또는 (ii)와 (iii)의 조합을 고려한 QID 성분을 더 포함할 수 있다. 예를 들어, 이 선택적인 추가 QID 성분은 다음과 같은 함수를 포함할 수 있다:

$$QID(doc) = w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} \div \left(k_{cd} + \frac{b_{cd} + b_{ud}}{b_{cd} + b_{ud}} \right)$$

[0077]

[0078] 여기에서, w_{cd}는 클릭 횟수 또는 바이어스된 클릭 횟수 파라미터를 포함하는 성분과 같은 질의-독립 성분의 가중치를 나타내고,

[0079] b_{cd}는 URL 길이에 관한 클릭 횟수 또는 바이어스된 클릭 횟수의 가중치를 나타내며,

[0080] b_{ud}는 URL 길이의 가중치를 나타내고,

[0081] CD는 문서에 대한 계산된 또는 할당된 클릭 횟수 또는 바이어스된 클릭 횟수를 나타내며,

[0082] k_{ew}는 기타 조정 파라미터와 유사하게 랭킹 함수의 정밀도를 최적화함으로써 결정되는 조정 상수를 나타내고(즉, k_{ew}는 모든 에지(edge)가 동일한 에지 가중치 값을 가질 때의 에지 가중치 값을 나타낼 수 있고, 또는 k_{ew}는 에지 가중치 값들이 서로 다를 때의 평균 에지 값을 나타낼 수 있음),

[0083] UD는 URL 길이를 나타내며,

[0084] k_{cd}는 클릭 횟수 포화 상수이다.

[0085] 가중된 항들(w_{cd}, b_{cd} 및 b_{ud})은 각각의 그들의 관련 항(즉, 각각, 클릭 횟수 또는 바이어스된 클릭 횟수 파라미터, 주어진 문서에 대한 클릭 횟수 또는 바이어스된 클릭 횟수 값, 및 주어진 문서의 URL 길이를 포함하는 성분)의 중요도를 규정하는 것을 도와서, 궁극적으로 점수화 함수의 결과를 규정하는 것을 돕는다.

[0086] URL 길이(UD)는 클릭 횟수 또는 바이어스된 클릭 횟수 값이 점수화 함수에 미칠 수 있는 영향을 완화시키기 위해 상기-언급된 질의-독립 성분으로의 선택적인 추가항이다. 예를 들어, 몇몇 경우에, 그다지 중요하지 않은(즉, 큰 URL 길이를 갖는) 문서는 짧은 클릭 횟수 또는 바이어스된 클릭 횟수 값을 가질 수 있다. URL 길이는 문서의 URL 내의 슬래시의 수로 표시된다. 예를 들어, www.example.com\d1\d2\d3\d4.htm은 4개의 슬래시를

포함하므로, 4의 URL 깊이를 가질 것이다. 그러나, 이 문서는 비교적 낮은 클릭 횟수 또는 바이어스된 클릭 횟수 값을 제공하는 메인 페이지 www.example.com으로부터의 직접적인 링크를 가질 수 있다. 상기 언급된 함수에 URL 깊이 항을 포함하고, 클릭 횟수 또는 바이어스된 클릭 횟수 값에 대해 URL 깊이 항을 가중화하는 것은 네트워크 내에서 문서의 중요도를 더욱 정확하게 반영하도록 비교적 높은 클릭 횟수 또는 바이어스된 클릭 횟수 값을 보상한다. 네트워크에 따라, 3 이상의 URL 깊이는 깊은 링크로 고려될 수 있다.

[0087] 한 실시예에서, 주어진 문서에 대한 문서 관련성 점수를 결정하기 위해 사용된 랭킹 함수는 다음과 같은 함수를 포함한다:

$$\text{점수} = \sum \frac{w_{ff'}(k_1+1)}{k_1 + w_{ff'}} \times \log\left(\frac{N}{n}\right) + w_{cd} \frac{k_{cd}}{b_{cd} \frac{CD}{k_{ew}} + b_{ud} UD} + w_u \frac{k_u U}{k_u + U}$$

[0088]

[0089] 여기에서, 항들은 상술된 바와 같다.

[0090] 그외 다른 실시예에서는, URL 깊이가 랭킹 함수에서 제거될 수 있고, 또는 기타 성분이 질의-종속 성분, 질의-독립 성분, 또는 이 둘 다의 정확도를 개선하기 위해 랭킹 함수에 추가될 수 있다. 더욱이, 사용 파라미터를 포함하는 상술된 질의-독립 성분은 검색 결과의 랭킹을 개선하기 위해 그외 다른 랭킹 함수(도시 생략)에 포함될 수 있다.

[0091] 단계(204)에서 일단 주어진 문서에 대한 문서 통계치가 랭킹 함수에 제공되면, 예시적인 방법(20)은 단계(205)로 진행한다. 단계(205)에서, 문서 관련성 점수가 주어진 문서에 대해 결정되어, 메모리 내에 저장되고, 주어진 문서와 관련된다. 단계(205)로부터, 예시적인 방법(20)은 판정 블록(206)으로 진행한다.

[0092] 판정 블록(206)에서, 문서 관련성 점수가 네트워크 내의 각 문서에 대해 계산되었는지 애플리케이션 코드에 의해 판정이 이루어진다. 문서 관련성 점수가 네트워크 내에서 각 문서에 대해 계산되지 않았다고 판정이 이루어지면, 예시적인 방법(20)은 단계(204)로 돌아가서, 상술된 바와 같이 진행한다. 문서 관련성 점수가 네트워크 내에서 각 문서에 대해 계산되었다고 판정이 이루어지면, 예시적인 방법(20)은 단계(207)로 진행한다.

[0093] 단계(207)에서, 다양한 문서를 포함하는 질의의 검색 결과들은 그들의 관련된 문서 관련성 점수에 따라 순위가 정해진다. 이렇게 얻어진 문서 관련성 점수는 네트워크 내의 각 문서의 실제 또는 디폴트 사용 값을 고려한다. 일단 검색 결과의 순위가 정해지면, 예시적인 방법(20)은 단계(208)로 진행하여, 순위가 정해진 결과가 사용자에게 표시된다. 단계(208)로부터, 예시적인 방법(20)은 단계(209)로 진행하여, 최고 순위 결과를 사용자가 선택하여 열어보게 된다. 단계(209)로부터, 예시적인 방법(20)은 단계(210)로 진행하여, 예시적인 방법(20)이 종료된다.

[0094] 네트워크 내의 문서들에 대한 문서 관련성 점수를 생성하고, 이 문서 관련성 점수를 사용하여, 검색 질의의 검색 결과의 순위를 정하는 상술된 방법 이외에, 상술된 방법을 실행하는 컴퓨터 실행가능 명령어가 그 위에 저장되어 있는 컴퓨터 판독가능 매체가 또한 여기에서 개시된다.

[0095] 컴퓨터 시스템이 또한 여기에서 개시된다. 예시적인 컴퓨팅 시스템은 컴퓨팅 시스템상에서 사용할 수 있는 적어도 하나의 애플리케이션 모듈을 포함하는데, 적어도 하나의 애플리케이션 모듈은 그 위에 로드된 애플리케이션 코드를 포함하고, 애플리케이션 코드는 네트워크 내의 문서들에 대한 문서 관련성 점수를 생성하는 방법을 실행한다. 애플리케이션 코드는 상술된 바와 같이, 네트워크 내의 문서들에 대한 문서 관련성 점수를 생성하고, 이 문서 관련성 점수를 사용하여, 검색 질의의 검색 결과의 순위를 정하는 컴퓨터 실행가능 명령어를 그 위에 갖는 임의의 상술된 컴퓨터 판독가능 매체를 사용하여 컴퓨팅 시스템상으로 로드될 수 있다.

[0096] 명세서가 그 특정 실시예와 관련하여 상세하게 설명되었지만, 본 분야에 숙련된 기술자들은 상기 설명을 이해하게 되면 이들 실시예에 대한 변경, 변형 및 등가물을 용이하게 생각해낼 수 있다는 것을 이해할 것이다. 따라서, 개시된 방법, 컴퓨터 판독가능 매체 및 컴퓨팅 시스템의 범위는 첨부된 청구범위 및 그 등가물의 범위로 평가되어야 한다.

도면의 간단한 설명

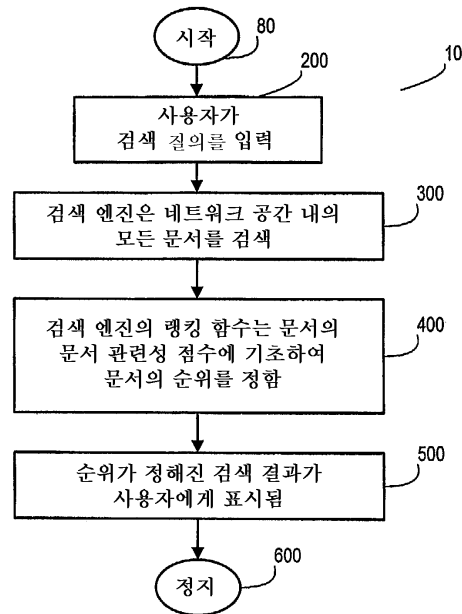
[0005] 도 1은 사용자에게 의해 입력된 검색 질의에 응답하여 순위화 검색 결과를 생성하는 방법의 예시적인 단계를 도시

한 예시적인 논리 흐름도.

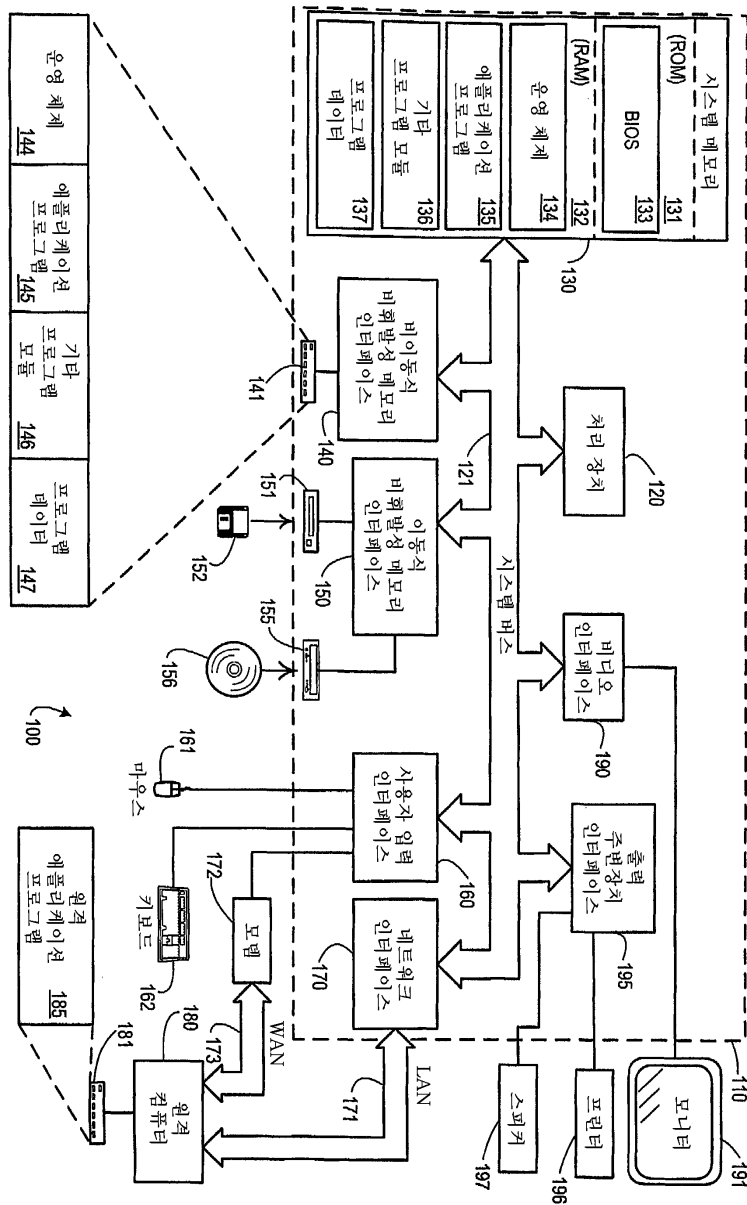
- [0006] 도 2는 여기에 개시된 방법 및 프로세스의 구현을 위한 예시적인 운영 환경의 주요 컴포넌트들의 일부를 도시한 블록도.
- [0007] 도 3은 네트워크상의 문서에 대한 문서 관련성 점수를 결정하는 예시적인 방법의 예시적인 단계를 도시한 논리 흐름도.
- [0008] 도 4는 문서 사용 파라미터를 포함하는 랭킹 함수를 사용하여 생성된 검색 결과의 순위를 정하는 방법의 예시적인 단계를 도시한 논리 흐름도.

도면

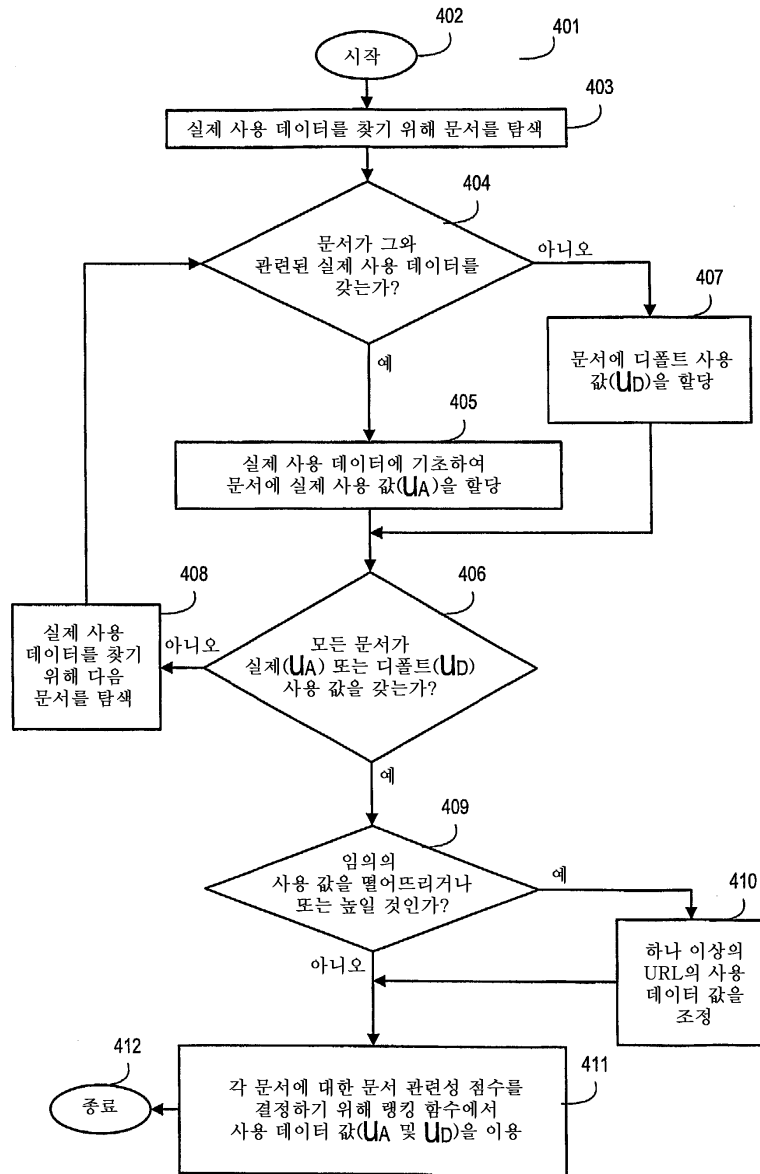
도면1



도면2



도면3



도면4

