



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0010652
(43) 공개일자 2016년01월27일

- (51) 국제특허분류(Int. Cl.)
G06F 17/00 (2006.01) G06F 17/30 (2006.01)
G06Q 30/02 (2012.01)
- (52) CPC특허분류
G06F 17/00 (2013.01)
G06F 17/3002 (2013.01)
- (21) 출원번호 10-2016-7000665(분할)
- (22) 출원일자(국제) 2008년04월03일
심사청구일자 2016년01월11일
- (62) 원출원 특허 10-2009-7023002
원출원일자(국제) 2008년04월03일
심사청구일자 2013년03월19일
- (85) 번역문제출일자 2016년01월11일
- (86) 국제출원번호 PCT/US2008/059303
- (87) 국제공개번호 WO 2008/124531
국제공개일자 2008년10월16일
- (30) 우선권주장
11/695,725 2007년04월03일 미국(US)

- (71) 출원인
구글 인코포레이티드
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이
1600 (우:94043)
- (72) 발명자
올드햄 제프리 디.
미국 캘리포니아 95123 산호세 위안도트 드라이브
144
바리안 할 알.
미국 캘리포니아 94526 덴빌 텔 아미고 로드 576
(뒷면에 계속)
- (74) 대리인
박장원

전체 청구항 수 : 총 21 항

(54) 발명의 명칭 불충분한 탐색 콘텐츠 식별

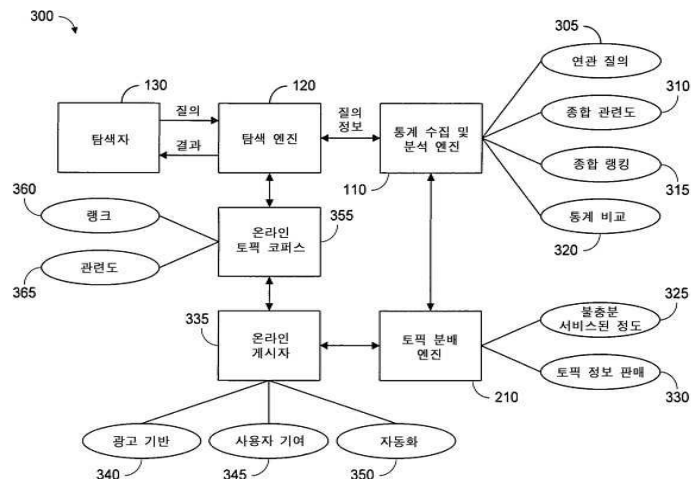
(57) 요약

불충분한 탐색 콘텐츠를 식별하기 위한 시스템 및 방법이 제공된다. 불충분한 탐색 콘텐츠는 예를 들어, 그 콘텐츠에 관계된 탐색 질의들과 연관된 통계에 기초하여 식별될 수 있다.

색인어

토픽, 코퍼스, 탐색, 엔진, 통계, 분석, 수집, 랭크, 관련도, 분배

대표도 - 도3



(52) CPC특허분류

G06F 17/30864 (2013.01)

G06Q 30/02 (2013.01)

(72) 발명자

커티스 매튜 디.

미국 캘리포니아 94024 로스 알토스 글렌 알토 드
라이브 554

로센크란츠 매트

미국 캘리포니아 94110 샌프란시스코 #에이 26번
스트리트 3649

명세서

청구범위

청구항 1

컴퓨터에 의해 구현되는 방법(computer-implemented method)으로서,

데이터 프로세싱 디바이스가 사용자에게 의해 제출된 탐색 질의(search query)를 명시하는(specifying) 데이터를 수신하는 단계와;

데이터 프로세싱 디바이스가 상기 탐색 질의가 속하는 특정한 탐색 토픽(search topic)을 결정하는 단계와;

데이터 프로세싱 디바이스가 상기 탐색 질의가 속하는 상기 특정한 탐색 토픽이 불충분 서비스된 토픽(underserved topic)으로서 분류됨을 결정하는 단계와,

여기서 상기 불충분 서비스된 토픽은 상대적인 토픽 코퍼스 품질 측정치(relative topic corpus quality measure)가 임계값(threshold value)보다 낮은 탐색 토픽이고,

상기 상대적인 토픽 코퍼스 품질 측정치는 상기 탐색 토픽에 속하는 것으로서 식별된 질의들에 대한 질의 탐색량 값(query search volume value)에 상대적인 상기 탐색 토픽에 대한 토픽 코퍼스에 포함된 문서들에 대한 관련도 측정치(relevance measure)이고,

상기 질의 탐색량은 상기 탐색 질의가 수신된 횟수에 비례하는 값이며; 그리고

데이터 프로세싱 장치가 상기 탐색 질의를 수신하는 것에 응답하여, 상기 탐색 질의가 불충분 서비스된 토픽에 속한다는 통지를 표시하도록 하는 데이터를 제공하는 단계를 포함하는 것을 특징으로 하는 컴퓨터에 의해 구현되는 방법.

청구항 2

제 1항에 있어서,

상기 불충분 서비스된 토픽에 대해 새로운 콘텐츠를 수신하는 단계와;

상기 새로운 콘텐츠가 상기 불충분 서비스된 토픽에 대한 상대적인 토픽 코퍼스 품질 측정치를 증가시킴을 결정하는 단계와; 그리고

상기 상대적인 토픽 코퍼스 품질 측정치에 대한 증가에 기초해서 상기 새로운 콘텐츠의 제공자에게 보상하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터에 의해 구현되는 방법.

청구항 3

제 2항에 있어서,

상기 새로운 콘텐츠가 상기 상대적인 토픽 코퍼스 품질 측정치를 증가시킴을 결정하는 단계는 상기 새로운 콘텐츠가 상기 불충분 서비스된 토픽에 대한 토픽 코퍼스에 포함된 문서들에 대한 상기 관련도 측정치를 증가시킴을 결정하는 것을 포함하는 것을 특징으로 하는 컴퓨터에 의해 구현되는 방법.

청구항 4

제 2항에 있어서,

상기 제공자에게 대한 보상의 값(value of the compensation)을 결정하는 단계를 더 포함하며, 상기 보상의 값은 상기 불충분 서비스된 토픽에 대한 수요(demand)와 상기 불충분 서비스된 토픽의 토픽 코퍼스 품질 측정치에 기초해서 결정되는 것을 특징으로 하는 컴퓨터에 의해 구현되는 방법.

청구항 5

제 1항에 있어서,

토픽 코퍼스의 상기 특정한 탐색 토픽에 대한 관련도의 측정치를 획득하는 단계와;
 상기 특정한 탐색 토픽에 대한 토픽 탐색량 값을 획득하는 단계와, 상기 토픽 탐색량 값은 상기 특정한 탐색 토픽에 대응하는 수신된 탐색 질의들의 횟수를 표시하는 값이며;
 상기 관련도의 측정치 및 상기 토픽 탐색량 값에 기초해서 상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치를 계산하는 단계와;
 상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치가 상기 임계값보다 낮음을 결정하는 단계와; 그리고
 상기 특정한 탐색 토픽을 불충분 서비스된 토픽으로서 분류하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터에 의해 구현되는 방법.

청구항 6

제 5항에 있어서,
 상기 특정한 탐색 토픽에 대한 토픽 탐색량의 임계값 내에 있는 토픽 탐색량들을 갖는 다른 탐색 토픽들에 대한 토픽 코퍼스 품질 측정치들을 획득하는 단계와; 그리고
 상기 다른 탐색 토픽들에 대한 토픽 코퍼스 품질 측정치들에 기초해서 상기 임계값을 결정하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터에 의해 구현되는 방법.

청구항 7

제 5항에 있어서,
 토픽 코퍼스의 상기 특정한 탐색 토픽에 대한 관련도의 측정치를 획득하는 단계와;
 상기 특정한 탐색 토픽에 대한 토픽 탐색량 값을 획득하는 단계와, 상기 토픽 탐색량 값은 상기 특정한 탐색 토픽에 대응하는 수신된 탐색 질의들의 횟수를 표시하는 값이며;
 상기 관련도의 측정치 및 상기 토픽 탐색량 값에 기초해서 상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치를 계산하는 단계와;
 상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치의 임계값 내에 있는 토픽 코퍼스 품질 측정치를 갖는 다른 탐색 토픽들에 대한 토픽 탐색량 값들을 획득하는 단계와;
 상기 다른 탐색 토픽들에 대한 탐색량 값들에 기초해서 임계 탐색량 값을 결정하는 단계와;
 상기 특정한 탐색 토픽에 대한 토픽 탐색량이 상기 임계 탐색량 값을 초과함을 결정하는 단계와;
 상기 특정한 탐색 토픽을 불충분 서비스된 토픽으로서 분류하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터에 의해 구현되는 방법.

청구항 8

시스템으로서,
 불충분 서비스된 토픽들과 상기 불충분 서비스된 토픽들에 속하는 것으로서 식별된 탐색 질의들을 표시하는 불충분 서비스된 토픽 데이터(underserved topic data)를 저장하는 데이터 저장소와,
 여기서 불충분 서비스된 토픽은 상대적인 토픽 코퍼스 품질 측정치가 임계값보다 낮은 탐색 토픽이고,
 상기 상대적인 토픽 코퍼스 품질 측정치는 상기 탐색 토픽에 속하는 것으로서 식별된 질의들에 대한 질의 탐색량 값에 상대적인 상기 탐색 토픽에 대한 토픽 코퍼스에 포함된 문서들에 대한 관련도 측정치이고,
 상기 질의 탐색량은 상기 탐색 질의가 수신된 횟수에 비례하는 값이며;
 상기 데이터 저장소와 인터랙션(interaction)하도록 구성된 하나 이상의 컴퓨터들을 포함하며, 상기 하나 이상의 컴퓨터들은:
 사용자에 의해 제출된 탐색 질의를 명시하는 데이터를 수신하는 동작과;
 상기 불충분 서비스된 토픽 데이터에 적어도 부분적으로 기초해서, 상기 탐색 질의가 특정한 불충분 서비스된

토픽에 속함을 결정하는 동작과; 그리고

상기 탐색 질의를 수신함에 응답하여, 상기 탐색 질의가 불충분 서비스된 토픽에 속한다는 통지를 표시하도록 하는 데이터를 제공하는 동작을 포함하는 동작들을 수행하도록 더 구성되는 것을 특징으로 하는 시스템.

청구항 9

제 8항에 있어서,

상기 하나 이상의 컴퓨터들은:

상기 특정한 불충분 서비스된 토픽에 대한 새로운 콘텐츠를 수신하는 동작과;

상기 새로운 콘텐츠가 상기 특정한 불충분 서비스된 토픽에 대한 상기 상대적인 토픽 코퍼스 품질 측정치를 증가시킴을 결정하는 동작과; 그리고

상기 상대적인 토픽 코퍼스 품질 측정치에 대한 증가에 기초해서 상기 새로운 콘텐츠의 제공자에게 보상하는 동작을 포함하는 동작들을 수행하도록 더 구성되는 것을 특징으로 하는 시스템.

청구항 10

제 9항에 있어서,

상기 새로운 콘텐츠가 상기 상대적인 토픽 코퍼스 품질 측정치를 증가시킴을 결정하는 동작은 상기 새로운 콘텐츠가 상기 특정한 불충분 서비스된 토픽에 대한 토픽 코퍼스에 포함된 문서들에 대한 상기 관련도 측정치를 증가시킴을 결정하는 것을 포함하는 것을 특징으로 하는 시스템.

청구항 11

제 9항에 있어서,

상기 하나 이상의 컴퓨터들은 상기 제공자에 대한 보상의 값을 결정하는 동작을 포함하는 동작들을 수행하도록 더 구성되며, 상기 보상의 값은 상기 특정한 불충분 서비스된 토픽에 대한 수요와 상기 특정한 불충분 서비스된 토픽의 토픽 코퍼스 품질 측정치에 기초해서 결정되는 것을 특징으로 하는 시스템.

청구항 12

제 8항에 있어서,

상기 하나 이상의 컴퓨터들은:

특정한 탐색 토픽에 대한 토픽 코퍼스의 관련도의 측정치를 획득하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 탐색량 값을 획득하는 동작과, 상기 토픽 탐색량 값은 상기 특정한 탐색 토픽에 대응하는 수신된 탐색 질의들의 횟수를 표시하는 값이며;

상기 관련도의 측정치 및 상기 토픽 탐색량 값에 기초해서 상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치를 계산하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치가 상기 임계값보다 낮음을 결정하는 동작과; 그리고

상기 특정한 탐색 토픽을 불충분 서비스된 토픽으로서 분류하는 동작을 포함하는 동작들을 수행하도록 더 구성되는 것을 특징으로 하는 시스템.

청구항 13

제 12항에 있어서,

상기 특정한 탐색 토픽에 대한 토픽 탐색량의 임계값 내에 있는 토픽 탐색량들을 갖는 다른 탐색 토픽들에 대한 토픽 코퍼스 품질 측정치들을 획득하는 동작과; 그리고

상기 다른 탐색 토픽들에 대한 토픽 코퍼스 품질 측정치들에 기초해서 상기 임계값을 결정하는 동작을 더 포함하는 것을 특징으로 하는 시스템.

청구항 14

제 12항에 있어서,

상기 특정한 탐색 토픽에 대한 토픽 코퍼스의 관련도의 측정치를 획득하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 탐색량 값을 획득하는 동작과, 상기 토픽 탐색량 값은 상기 특정한 탐색 토픽에 대응하는 수신된 탐색 질의들의 횟수를 표시하는 값이며;

상기 관련도의 측정치 및 상기 토픽 탐색량 값에 기초해서 상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치를 계산하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치의 임계값 내에 있는 토픽 코퍼스 품질 측정치들을 갖는 다른 탐색 토픽들에 대한 토픽 탐색량 값들을 획득하는 동작과;

상기 다른 탐색 토픽들에 대한 탐색량 값들에 기초해서 임계 탐색량 값을 결정하는 동작과;

상기 특정한 탐색 토픽에 대한 상기 토픽 탐색량 값이 상기 임계 탐색량 값을 초과함을 결정하는 동작과; 그리고

상기 특정한 탐색 토픽을 불충분 서비스된 토픽으로서 분류하는 동작을 더 포함하는 것을 특징으로 하는 시스템.

청구항 15

실행시에 하나 이상의 컴퓨터들로 하여금 동작들을 수행하도록 하는 명령어들을 포함하는 컴퓨터 프로그램이 수록된 메모리 디바이스(memory device)로서, 상기 동작들은:

데이터 프로세싱 디바이스가 사용자에게 의해 제출된 탐색 질의를 명시하는 데이터를 수신하는 동작과;

데이터 프로세싱 디바이스가 상기 탐색 질의가 속하는 특정한 탐색 토픽을 결정하는 동작과;

데이터 프로세싱 디바이스가 상기 탐색 질의가 속하는 특정한 탐색 토픽이 불충분 서비스된 토픽으로서 분류됨을 결정하는 동작과,

상기 불충분 서비스된 토픽은 상대적인 토픽 코퍼스 품질 측정치가 임계값보다 낮은 탐색 토픽이고,

상기 상대적인 토픽 코퍼스 품질 측정치는 상기 탐색 토픽에 속하는 것으로서 식별된 질의들에 대한 질의 탐색량 값에 상대적인 상기 탐색 토픽에 대한 토픽 코퍼스에 포함된 문서들의 관련도 측정치이고,

상기 질의 탐색량 값은 상기 탐색 질의가 수신된 횟수에 비례하는 값이며; 그리고

데이터 프로세싱 디바이스가 상기 탐색 질의를 수신함에 응답하여, 상기 탐색 질의가 불충분 서비스된 토픽에 속한다는 통지를 표시하도록 하는 데이터를 제공하는 동작을 포함하는 것을 특징으로 하는 메모리 디바이스.

청구항 16

제 15항에 있어서,

상기 하나 이상의 컴퓨터들로 하여금:

상기 불충분 서비스된 토픽에 대한 새로운 콘텐츠를 수신하는 동작과;

상기 새로운 콘텐츠가 상기 불충분 서비스된 토픽에 대한 상기 상대적인 토픽 코퍼스 품질 측정치를 증가시킴을 결정하는 동작과; 그리고

상기 상대적인 토픽 코퍼스 품질 측정치에 대한 증가에 기초해서 상기 새로운 콘텐츠의 제공자에게 보상하는 동작을 포함하는 동작들을 수행하도록 하는 명령어들을 더 포함하는 것을 특징으로 하는 메모리 디바이스.

청구항 17

제 16항에 있어서,

상기 새로운 콘텐츠가 상기 상대적인 토픽 코퍼스 품질 측정치를 증가시킴을 결정하는 동작은 상기 새로운 콘텐츠가 상기 불충분 서비스된 토픽에 대한 토픽 코퍼스에 포함된 문서들에 대한 관련도 측정치를 증가시킴을 결정

하는 것을 포함하는 것을 특징으로 하는 메모리 디바이스.

청구항 18

제 16항에 있어서,

상기 하나 이상의 컴퓨터들로 하여금:

상기 제공자에 대한 보상의 값을 결정하는 동작을 포함하는 동작들을 수행하도록 하는 명령어들을 더 포함하며, 상기 보상의 값은 상기 불충분 서비스된 토픽에 대한 수요 및 상기 불충분 서비스된 토픽의 토픽 코퍼스 품질 측정치에 기초해서 결정되는 것을 특징으로 하는 메모리 디바이스.

청구항 19

제 15항에 있어서,

상기 하나 이상의 컴퓨터들로 하여금:

상기 특정한 탐색 토픽에 대한 토픽 코퍼스의 관련도의 측정치를 획득하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 탐색량 값을 획득하는 동작과, 상기 토픽 탐색량 값은 상기 특정한 탐색 토픽에 대응하는 수신된 탐색 질의들의 횟수를 표시하는 값이며;

상기 관련도의 측정치 및 상기 토픽 탐색량 값에 기초해서 상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치를 계산하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치가 상기 임계값보다 낮음을 결정하는 동작과; 그리고

상기 특정한 탐색 토픽을 불충분 서비스된 토픽으로서 분류하는 동작을 포함하는 동작들을 수행하도록 하는 명령어들을 더 포함하는 것을 특징으로 하는 메모리 디바이스.

청구항 20

제 19항에 있어서,

상기 하나 이상의 컴퓨터들로 하여금:

상기 특정한 탐색 토픽에 대한 토픽 탐색량의 임계값 내에 있는 토픽 탐색량들을 갖는 다른 탐색 토픽들에 대한 토픽 코퍼스 품질 측정치들을 획득하는 동작과;

상기 다른 탐색 토픽들에 대한 토픽 코퍼스 품질 측정치들에 기초해서 상기 임계값을 결정하는 동작을 포함하는 동작들을 수행하도록 하는 명령어들을 더 포함하는 것을 특징으로 하는 메모리 디바이스.

청구항 21

제 19항에 있어서,

상기 하나 이상의 컴퓨터들로 하여금:

상기 특정한 탐색 토픽에 대한 토픽 코퍼스의 관련도의 측정치를 획득하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 탐색량 값을 획득하는 동작과, 상기 토픽 탐색량 값은 상기 특정한 탐색 토픽에 대응하는 수신된 탐색 질의들의 횟수를 표시하는 값이며;

상기 관련도의 측정치 및 상기 토픽 탐색량 값에 기초해서 상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치를 계산하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 코퍼스 품질 측정치의 임계값 내에 있는 토픽 코퍼스 품질 측정치들을 갖는 다른 탐색 토픽들에 대한 토픽 탐색량 값들을 획득하는 동작과;

상기 다른 탐색 토픽들에 대한 탐색량 값들에 기초해서 임계 탐색량 값을 결정하는 동작과;

상기 특정한 탐색 토픽에 대한 토픽 탐색량이 상기 임계 탐색량 값을 초과함을 결정하는 동작과; 그리고

상기 특정한 탐색 토픽을 불충분 서비스된 토픽으로서 분류하는 동작을 포함하는 동작들을 수행하도록 하는 명

령어들을 더 포함하는 것을 특징으로 하는 메모리 디바이스.

발명의 설명

기술 분야

[0001] 본 명세서는 토픽 코퍼스(topic corpus)의 평가에 관한 것이다.

배경 기술

[0002] 월드와이드웹(www)을 통해 거대한 양의 정보에 쉽게 액세스할 수 있다. 예를 들어, 개인 사용자, 웹 게시자(web publisher), 광고(ad) 회사 등과 같은 다수의 콘텐츠 생성자가 이 정보를 생성할 수 있다. 이들 콘텐츠 생성자는 다양한 목적을 충족시키기 위하여 콘텐츠를 부가할 수 있다. 개인 사용자들은 그 사용자가 개인적으로 관심을 갖는 토픽에 기초하여 콘텐츠를 종종 생성한다. 예를 들어, 개인 사용자는 다른 사람에게 정보를 제공하거나 유사한 관심을 가진 사람들을 만나고자 취미에 관련된 콘텐츠를 생성할 수 있다. 웹 게시자는 예를 들어, 소정 분야의 지식을 나타내거나 사이트에 대한 가입(subscription)을 판매하기 위하여 콘텐츠를 생성할 수 있다. 광고 회사는 예를 들어, 사용자가 사이트를 방문하여 그 사이트가 광고하는 제품을 구매하도록 유도하기 위하여 콘텐츠를 생성할 수 있다.

[0003] 웹사이트의 목적이 방문자를 끌어 들이는 것이면, 다른 사용자들에게 관심이 있는 토픽을 확인하는 것은 어려울 수 있다. 따라서 콘텐츠 생성자는 관심을 계측하기 위하여 다양한 메커니즘을 사용할 수 있다. 예를 들어, 일부 게시자는 여론 조사를 실시하여 사용자들이 관심을 갖는 토픽을 확인할 수 있다. 여론 조사는 부정확하거나 및/또는 응답자/여론 조사자에 의해 조작될 수 있다. 키워드와 관련하여 수행된 총 탐색 수에 기초하여 특정 토픽이 또한 확인될 수 있다.

발명의 내용

[0004] 토픽 코퍼스를 평가하는 시스템과 방법이 개시된다. 일 구현예에서, 시스템은 통계 수집 엔진, 통계 분석 엔진, 비교기 및 토픽 분배 엔진을 포함한다. 통계 수집 엔진은 브라우저들로부터 수신된 하나 이상의 질의와 연관된 질의 통계를 생성하도록 동작할 수 있다. 통계 분석 엔진은 질의들을 하나 이상의 토픽으로 분할하고, 토픽들과 연관된 토픽 통계를 생성하도록 동작할 수 있다. 토픽 통계는 종합 랭크 또는 토픽들과 연관된 토픽 코퍼스의 종합 관련도를 포함할 수 있다. 비교기는 토픽 통계와 질의 통계의 비교에 기초하여 토픽을 식별하도록 동작할 수 있고, 토픽 분배 엔진은 하나 이상의 사용자에게 식별된 토픽을 통지할 수 있다.

[0005] 본 명세서의 방법은, 제1 토픽과 연관된 통계를 결정하는 단계; 제1 토픽이 불충분 서비스되는지를 상기 통계에 기초하여 결정하는 단계; 및 제1 토픽과 연관된 토픽 코퍼스에 포함시키기 위한 부가적 콘텐츠를 상기 결정에 기초하여 요청하는 단계를 포함한다.

[0006] 일 특징으로, 코퍼스의 품질과 탐색 로그의 분석에 기초하여 토픽 코퍼스를 분류하는 방법이 개시된다.

[0007] 식별된 토픽 코퍼스에 기초한 제안들은, 탐색 로그에 근거하여 다른 사용자들이 소망하는 부가적 콘텐츠의 생성을 위해 콘텐츠 생성자들에게 제공될 수 있다.

[0008] 식별된 토픽 코퍼스에 기초한 피드백(이 피드백은 토픽 코퍼스가 고품질이 아님을 나타냄)이 또한 탐색자에게 제공될 수 있어서, 토픽 코퍼스가 부적절할 수 있음을 제공된 질의에 기초하여 탐색자에게 경고할 수 있다.

[0009] 제안들은 콘텐츠 생성자들이 사용자들로부터 얻어지는 가입에 기초한 이윤 또는 콘텐츠와 연관된 웹사이트에 광고를 배치함으로써 얻어지는 광고 수입을 위해 콘텐츠를 제공하도록 만들 수 있다.

도면의 간단한 설명

[0010] 도1은 통계 수집 및 분석 시스템을 포함하는 네트워크 환경의 블록도이다.

도2는 콘텐츠 생성자에게 토픽을 제안하는 시스템을 구비한 네트워크 환경의 블록도이다.

도3은 시스템 내에서 불충분한 웹 콘텐츠를 확인하기 위한 데이터 흐름을 예시하는 블록도이다.

도4는 시스템 내에서 부적절하게 서비스되는 것으로 확인된 토픽에 관련된 웹 콘텐츠를 수집하는 데이터 흐름을 예시하는 블록도이다.

도5-7은 불충분한 웹 콘텐츠를 확인하는 예시적 방법을 나타내는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0011] 본 명세서의 시스템 및 방법은 예를 들어, 탐색 로그의 분석에 기초하여, 불충분한 게시(예를 들어, 웹-기반 게시) 콘텐츠(published content)를 확인하도록 동작할 수 있다. 탐색 로그로부터의 탐색 질의를 분석하여 탐색 질의들 사이의 관계를 발견할 수 있다. 관련된 탐색 질의들은 예를 들어, 하나의 토픽과 같은 상위개념(genus)과 연관된 하위개념(species)일 수 있다. 하나의 토픽에 관련된 것으로 탐색 질의들을 연관시키는 것은 그 관련된 탐색 질의들에 대한 탐색 결과들에 기초한 토픽 코퍼스의 확인을 용이하게 할 수 있다. 토픽 코퍼스의 분석은 예를 들어, 토픽과 연관된 콘텐츠의 품질 확인을 용이하게 할 수 있다. 또한 탐색 질의들을 분석하여 특정 토픽의 인기도를 결정할 수 있다. 인기도는 예를 들어, 토픽에 대하여 수신된 탐색 질의들의 총 수에 기초하여 유도될 수 있다. 토픽 코퍼스 내에서 발견된 콘텐츠의 품질과 탐색 인기도의 비교는, 콘텐츠가 예를 들어, 토픽과 연관된 탐색 질의들의 인기도에 기초하여 부적절하다는 결정을 제공할 수 있다.
- [0012] 도1은 통계 수집 및 분석 시스템을 포함하는 네트워크 환경(100)의 블록도이다. 통계 수집 및 분석 엔진(110)은 탐색 엔진(120)으로부터 통계를 수집할 수 있다. 탐색 엔진은 질의들을 예를 들어, 하나 이상의 컴퓨팅 디바이스(130a-c)로부터 네트워크(140)를 통하여 수신할 수 있다. 탐색 질의들은 구문 분석되고(parsed), 페이지 인덱스 저장부(page index store; 150)에 대해 비교된다. 페이지 인덱스 저장부(150)에 대한 질의의 분석은 예를 들어, 정보 검색(information retrieval; IR) 스코어로 될 수 있다. 정보 검색 스코어는 탐색 질의에 대한 문서(예를 들어, 웹페이지)의 관련도를 측정할 수 있다.
- [0013] 탐색 질의에 관련된 것으로 발견된 문서들은 랭크(rank)를 위해 노드 랭크 엔진(160)에 의해 분석될 수 있다. 일부 실시예에서, 노드 랭킹은 오프라인으로 수행되고 저장될 수 있음을 이해해야 한다. 노드 랭킹은 노드 랭크 저장부로부터 노드 랭크 엔진(160)을 통해 검색될 수 있다. 노드 랭크는 예를 들어, 현재 노드로 연결된 노드들의 수 및 이들 노드들의 평판(reputation)에 기초될 수 있다. 노드 랭킹은 예를 들어, 발명의 명칭이 "링크된 데이터베이스에서 노드 랭킹 방법(Method for Node Ranking in a Linked Database)"이고 본 명세서에 참조로 포함된 미국 특허 제6,285,999호에 설명되어 있다.
- [0014] 탐색자에게 결과들이 제시되는 순서를 결정하기 위하여 IR 스코어와 노드 랭크가 결합될 수 있다. 탐색 엔진(120) 및 질의를 허용하는 것과 같은 다른 서비스들은 서비스되지 않은(unserved) 질의들과 불충분 서비스된(underserved) 질의들을 수집할 수 있다. 서비스 되지 않은 질의들은 예를 들어, 탐색 엔진이 관련 콘텐츠를 발견할 수 없는 임의의 질의들을 포함할 수 있다. 불충분 서비스된 질의들은 예를 들어, 콘텐츠에 대한 수요가 공급이나 토픽 코퍼스에서 발견된 품질을 초과하는 임의의 질의들을 포함할 수 있다. 탐색 엔진 및 그 밖의 그러한 서비스들은 서비스되지 않은 질의들 및 불충분 서비스된 질의들을 선택적으로 필터, 분류 및 수집할 수 있다. 인터넷-기반 범용 탐색 엔진들은 예를 들어, 임의의 토픽에 대한 탐색을 가능하게 하고, 그 탐색 결과들을 사용자에게 제공한다. 질의들은 예를 들어, 텍스트 기반일 수 있고, 연관된 웹페이지들의 목록 형태로 결과들이 제공될 수 있다. 일부 질의들에 대하여, 소수의 결과들 및/또는 오직 저품질의 결과들이 반환된다. 탐색 엔진은 그 질의들을 불충분 서비스된 질의들의 저장소에 수집할 수 있다.
- [0015] 그러한 데이터의 수집은 예를 들어, 탐색 엔진(120)과 같은 인터넷-기반 범용 탐색 엔진들로 제한되지 않는다. 질의들을 처리하고 결과들을 반환하거나 서비스를 제공하는 임의의 시스템이면 족하다. 예를 들어, 많은 소프트웨어 프로그램들의 도움말 기능은 테스트를 입력받고 결과들을 반환한다. 이들 소프트웨어 프로그램들은 상술한 정보를 수집할 수 있다.
- [0016] 일 실시예에 있어서, 통계 수집 및 분석 엔진(110)은 탐색 엔진(120)에 제출되고 있는 질의들에 대한 정보를 수집할 수 있다. 예를 들어, 통계 수집 및 분석 엔진(110)은 질의들이 탐색 엔진(120)에 제출되면 그 질의들을 수신할 수 있다. 통계 수집 및 분석 엔진(110)은 또한 탐색 질의들을 분석하여 하나의 토픽으로 그룹화될 수 있는 질의들을 식별할 수 있다. 이 토픽은 예를 들어, 탐색 질의들에 의해 기술된 하위개념에 대한 상위개념을 기술할 수 있다. 탐색 엔진(120)은 카테고리-기반 페이지들에 상호작용 탐색(interactive search)을 제공할 수 있다. 또 다른 실시예에 있어서, 질의들은 질의와 연관된 가장 특징적 카테고리로 주석이 달릴 수 있다. 또 다른 실시예에 있어서, 탐색 엔진(110)은 탐색 질의를 포함하는 한 세트의 문자열(strings)을 분류체계(taxonomy)에 맵핑할 수 있다. 이 분류체계 카테고리들은 계산되고, 저장된 질의들과 연관될 수 있다. 그 밖의 구현예에서, 질의들은 데이터 저장부에 저장하기 위하여 카테고리들로 결집될 수 있다. 그러한 구현예에서, 질의들 자체는 카테고리의 저장에 기초하여 데이터 저장부에 저장되지 않을 수 있고, 또한, 카테고리 또는 토픽과 연관될

수 있는 질의들의 그룹들을 식별하기 위하여 클러스터(cluster)될 수 있다.

[0017] 또한 통계 수집 및 분석 엔진(110)은 탐색 질의와 연관된 문서들(예를 들어, 토픽 코퍼스)의 관련도(예를 들어, IR 스코어)를 결정하여 질의의 결과들과 연관된 품질을 식별할 수 있다. 일부 실시예에서, 통계 수집 및 분석 엔진(110)은 이 관련도를 평판(예를 들어, 노드 랭킹)과 결합하여, 질의와 연관된 토픽 코퍼스에 연관된 품질을 결정할 수 있다. 일부 실시예에서, 통계 수집 및 분석 엔진(110)은 예를 들어, 탐색의 인기(popularity)와 토픽 코퍼스의 품질을 비교하는 비교기를 포함할 수 있다. 예를 들어, 이러한 비교는 토픽 코퍼스가 적절하게 토픽을 서비스하고 있는지 여부를 결정하기 위하여 사용될 수 있다.

[0018] 질의 시스템에 입력된 불충분 서비스된 질의들에 더하여, 질의들의 의미를 명확히 하거나 불충분 서비스된 토픽들에 대한 정보를 제공하는 것을 지원하는 그 밖의 관련된 데이터가 통계 수집 및 분석 엔진(110)에 의해 수집 및 분석될 수 있다. 언어 분포, 지리적 분포, 인구 분포 및 시간 분포와 같은 질의들과 직접 관련된 특성들(characteristics)은 또한 통계 수집 및 분석 엔진(110)에 의해 수집될 수 있다. 시간 분포와 연관된 질의들은, 예를 들어, 질의가 특정 휴일, 주중 어느 날, 하루 중 어느 시간 부근에서 인기가 있다는 표시(indication)가 될 수 있다. 일부 구현예에서, 질의 빈도가 또한 수집될 수 있고, 예를 들어, 다수의 소스들로부터 질의들이 올 때, 질의들과 연관된 소스에는 주석이 달릴 수 있다. 따라서 통계 수집 및 분석 엔진(110)은 콘텐츠 품질 및/또는 인기를 분석하는데 사용될 수 있는 다양한 정보를 수집하도록 구성될 수 있다.

[0019] 일부 실시예에서, 탐색 엔진(120)은 토픽 코퍼스와 연관된 품질이, 예를 들어, 탐색 로그들에 기초한 유사하게 인기 있는 토픽들에 비하여 낮다는 점을 탐색자에게 통지할 수 있다. 탐색 엔진(120)은 주제에서 탐색자의 관심으로부터 탐색자가 그 주제에 대하여 관심이 있고 따라서 토픽 코퍼스에 포함될 수 있는 정보를 더 가지고 있다고 유추할 수 있다. 탐색 엔진(120)은 토픽 코퍼스에 포함시키기 위한 부가적 콘텐츠를 제공하도록 탐색자에게 요청하기 위한 통지를 사용할 수 있다. 탐색자에게 탐색 결과들과 함께 제공된 통지는, 예를 들어, 토픽이 그 토픽에 대한 요구에 대하여 유사한 비율로 제안됨을 보장하는데 도움을 줄 수 있다.

[0020] 도2는 콘텐츠 생성자(220, 230, 240)에게 토픽을 제안하는 토픽 분배 엔진(210)을 포함하는 네트워크 환경(200)의 블록도이다. 통계 수집 및 분석 엔진(110)은 불충분한 콘텐츠를 가진 영역(예를 들어, 토픽들)을 식별하고, 이들 영역을 토픽 분배 엔진(210)에 알릴 수 있다. 토픽 분배 엔진(210)은 식별된 영역들을 포함하는 토픽들을 콘텐츠 생성자(220, 230, 240)에게 제공할 수 있다. 일 실시예에서, 토픽 제안들은 제안된 토픽에 대하여 지식(knowledge)을 갖는 콘텐츠 생성자(220, 230, 240)에게 제공될 수 있다. 예를 들어, 불충분 서비스된 스포츠 토픽은 스포츠-관련 게시자에게 제안될 수 있다.

[0021] 콘텐츠 생성자(220, 230, 240)는 부가적 콘텐츠를 토픽 코퍼스를 위해 생성하기 위한 다른 메커니즘들을 다양하게 포함할 수 있다. 예를 들어, 콘텐츠 생성자(220, 230, 240)는 웹 게시자(220)를 포함할 수 있다. 웹 게시자(220)는 예를 들어, 고객용 콘텐츠를 생성하도록 운영되는 사업체일 수 있다. 웹 게시자(220)는 예를 들어, 광고 판매 모델(222)에 기초하여 동작할 수 있다. 이 모델에서, 웹 게시자는 연관된 웹사이트에서 무료로 사용할 수 있는 콘텐츠를 생성할 수 있다. 다음, 웹 게시자는 방문자 통계를 수집하고, 그 연관된 웹사이트의 광고 공간(advertising space)을, 그 웹사이트를 보는 방문자의 수에 기초하여 광고주에게 판매할 수 있다.

[0022] 대안적으로는, 웹 게시자(220)는 가입 기반 모델(subscription based model; 224) 상에서 동작한다. 예를 들어, 웹 게시자(220)는 자신이 생성한 콘텐츠에 대한 온라인 액세스를 허용하는 대가로 사용자에게 가입을 판매할 수 있다. 이러한 웹 게시자에는 예를 들어, 신문사 웹사이트, 백과사전 웹사이트, 사전/시소러스(thesaurus) 웹사이트 등이 있다.

[0023] 웹 게시자(220)가 웹페이지들을 생성할 인센티브(incentive)가 있지만, 웹 게시자(220)는 종종 특정 정보에 대한 요구를 인식하지 못하고, 따라서 어느 정보를 가용으로 만들지 모른다. 탐색 제공자(예를 들어, 도1의 탐색 엔진(120))는 매우 다양한 정보 요청에 대하여 액세스하고, 또한 상응하는 결과들의 가용성을 측정할 수 있다. 통계 수집 및 분석 엔진(110)은 용어에 대하여 품질 탐색 결과들이 거의 발견되지 않은 인스턴스(instance)를 추적하고, 토픽 제안 엔진(210)은 탐색자에게 더 많은 정보가 필요함을 제안할 수 있다. 탐색 제공자의 목적은 사용자 만족 및 탐색 제공자에 대한 충성도를 유지하는 것이므로, 탐색 제공자는 통계 수집 엔진(110)과 토픽 분배 엔진(210)을 제공할 인센티브를 갖는다. 고품질의 콘텐츠가 없을 때, 사용자는 탐색 제공자에 대하여 불만을 가지게 된다.

[0024] 탐색 제공자가 게시자-인센티브 시스템을 포함하면, 탐색 제공자는 부가적인 콘텐츠를 고양하는 추가적 인센티브를 갖는다. 예를 들어, 탐색자가 토픽에 관하여 관심(이것은 질의의 입력으로부터 유추될 수 있음)을 표명하

면, 탐색 제공자는 탐색자가 그 주제를 연구(오프라인 및/또는 온라인)하고 그 연구에 기초한 콘텐츠를 생성하여 그 토픽(들)에 대한 웹페이지들을 게시할 것을 요청할 수 있다.

[0025] 탐색 제공자가 탐색자-인센티브 시스템을 포함하면, 탐색자는 부가적인 이점을 받을 수 있다. 게시자 인센티브 시스템은 부가적인 콘텐츠와, 토픽에 대한 질의들 및/또는 게시 후의 문서와 연관된 노드 랭킹(들)을 비교함으로써, 고품질의 정보를 고양하도록 동작할 수 있다. 예를 들어, 게시자 인센티브 시스템은 불충분 서비스된 토픽들에 대한 수요에 따라서 인센티브를 설정할 수 있다. 즉, 수요가 많은 불충분 서비스된 토픽들에 대하여 보상을 높게 설정하고, 수요가 적은 토픽들에 대하여 보상을 낮게 설정하여, 점진적(progressive) 게시자 인센티브 시스템을 제공한다.

[0026] 콘텐츠 생성자(220, 230, 240)는 또한 사용자 기여(contribution) 사이트(230)를 포함할 수 있다. 예를 들어, 위키 사이트와 같은 사용자 기여 사이트에 의해 매우 다양한 범위의 사용자가 콘텐츠를 생성하고 게시할 수 있다. 사용자 기여 사이트(230)는 예를 들어, 사용자들로부터의 제안에 기초하여 스템브 기사(stub article; 235)를 생성할 수 있다. 스템브 기사(235)는, 이 스템브 기사(235)를 생성한 사람들이 갖는 지식 밖의 주제에 대한 지식을 가질 수 있는 다른 사용자들로부터 부가적 기여를 유도하도록 동작할 수 있다. 일부 구현예에서, 토픽 분배 엔진(210)은 통계 수집 및 분석 엔진(110)에 기초하여, 사용자 기여 사이트(230)에 기사 제안을 제공할 수 있다. 사용자 기여 사이트(230)는 그 사이트에 포함시키기 위하여, 기사 제안에 기초하여 스템브 기사(235)를 생성할 수 있다. 스템브 기사(235)의 포함은 또한 탐색자에게 토픽이 탐색되는 빈도와 유사한 빈도로 불충분한 콘텐츠를 통지할 수 있다.

[0027] 콘텐츠 생성자(220, 230, 240)는 또한 자동화된 콘텐츠 생성기(240)를 포함할 수 있다. 자동화된 콘텐츠 생성기는 예를 들어, 다수의 사이트로부터 콘텐츠의 단일 페이지로의 종합화(245)를 제공할 수 있다. 자동화된 콘텐츠 생성기(240)는 예를 들어, 다수의 사이트로부터 콘텐츠를 복사하고, 복사된 콘텐츠를 포함하는 단일의 문서를 생성할 수 있다. 일 구현예에서, 자동화된 콘텐츠 생성기(240)는 특정 사이트들로부터만 콘텐츠를 복사하도록 구성될 수 있다. 이것에 의해, 자동화된 콘텐츠 생성기(240)는 자신이 라이선스를 가진 사이트들/사용자들로부터만의 콘텐츠를 복사할 수 있다. 또한 자동화된 컨트롤러 생성기(240)는 예를 들어, 특정 토픽에 관계된 콘텐츠로의 링크들의 종합화(245)를 제공할 수 있다. 자동화된 콘텐츠 생성기(240)는 새로운 콘텐츠의 생성을 위해 스템브 정보(stub information)를 제공하도록, 웹 게시자(220) 또는 사용자 기여 사이트(230)와 결합될 수 있다.

[0028] 일부 구현예에서, 통계 수집 및 분석 엔진(110)은 다수의 언어에 걸쳐서 콘텐츠 품질을 결정할 수 있다. 이러한 구현예에 있어서, 토픽 분배 엔진(210)은 다양한 언어로부터의 결과들과 연관된 품질을 제공할 수 있다. 품질 결과들은 콘텐츠 생성기(220, 230, 240)에게 특정 언어에서 토픽 코퍼스 품질은 낮지만 다른 언어에서는 적절함을 표시할 수 있다. 콘텐츠 생성기(220, 230, 240)는 그러한 정보를 사용하여, 토픽 코퍼스의 품질이 낮은 것으로 결정된 특정 언어로 부가적 정보를 생성할 수 있다.

[0029] 토픽 분배 엔진(210)은 또한 다양한 인터페이스를 통하여 토픽 제안들을 제공할 수 있다. 예를 들어, 토픽 분배 엔진(210)은 웹 인터페이스를 사용해 부가적 문서들을 보장하는(warranting) 토픽들의 목록을 위키피디아와 같은 정보 제공자들에게 제공할 수 있다.

[0030] 대안적으로, 탐색 제공자(예를 들어, 탐색 엔진(120))는 제안된 토픽들을 탐색하는 개별 서비스를 제공하여, 부가적 정보를 필요로 하는 토픽들에 탐색 엔진(예를 들어, 불충분 서비스된 토픽 탐색 엔진(250))을 제공할 수 있다. 불충분 서비스된 토픽 탐색 엔진(250)은 불충분 서비스된 토픽들 및 관련 데이터의 집합을, 기존의 탐색 엔진 기법을 사용하여 탐색할 수 있다. 이러한 탐색 엔진(250)은 추가적 콘텐츠의 부가를 소망하는 게시자들이 토픽들에 액세스하는 것을 지원할 수 있는데, 게시자의 질의와 불충분 서비스된 토픽이 일치하는 임의의 결과가 있으면 불충분 서비스된 토픽에 대한 정보가 게시자에게 제공되기 때문이다. 불충분 서비스된 토픽들에 대하여 콘텐츠의 기여를 소망하는 탐색자들은 불충분 서비스된 토픽 탐색 엔진(250)을 활용하여 토픽이 적절히 서비스되고 있는지 결정할 수 있다.

[0031] 일부 실시예에서, 불충분 서비스된 토픽들과 연관된 데이터에 기초하여 불충분 서비스된 토픽들을 식별하기 위하여, 질의 확장(query broadening)이 불충분 서비스된 토픽 탐색 엔진(250)과 공동으로 사용될 수 있다. 예를 들어, 탐색 질의의 언어가 불충분 서비스된 토픽과 연관된 데이터의 언어와 관련되면, 질의 확장은 불충분 서비스된 토픽들을 디스플레이할 수 있다. 질의 확장은 일부 실시예에서, 불충분 서비스된 토픽에 대하여 예를 들어, 특정 휴가 기간과 같은 디스플레이 시간을 규정하는데 또한 사용될 수 있다. 일부 구현예에서, 불충분 서비스된 토픽 탐색 엔진(250)은 통상의 탐색 엔진에 비하여 질의 확장에 관대할 수 있는데, 본질적으로 하위 개

념에 대한 정보를 갖는 탐색자는 상위 개념에 관한 정보를 적어도 일부 갖기 때문이다. 예를 들어, 특정 축구 선수 및/또는 축구 통계 값에 관한 지식이 있는 사람은, 비록 그 지식이 특정 선수 또는 통계 값에 한정된 것이더라도, 축구에 대한 일부 지식을 갖는다.

[0032] 다수의 불충분 서비스된 토픽이 탐색과 관련이 있으면, 탐색 엔진(250)은 일치하는 불충분 서비스된 토픽들을 랭킹하고, 결과들을 랭크가 감소하는 순서로 디스플레이하며, 목록을 적절히 절단한다. 다양한 랭킹 함수가 가능하다. 예를 들어, 질의와 보다 잘 일치하는 불충분 서비스된 토픽들은 상위에 랭크될 수 있다. 보다 자주 요청되는 불충분 서비스된 토픽들도 상위에 랭크될 수 있다.

[0033] 토픽 분배 엔진(210)은 또한 다양한 포맷으로 토픽 제안들을 제공할 수 있다. 일 구현예에서, 토픽 분배 엔진(210)은 가장 인기 있는 토픽들에 기초하여 토픽들을 재정렬할 수 있다. 다른 구현예에서, 토픽 분배 엔진(210)은 인기도와 콘텐츠 품질간 불균형이 가장 큰 토픽들에 기초하여 토픽들을 재정렬할 수 있다. 다른 포맷 방식이 또한 사용될 수 있다.

[0034] 다른 구현예에서, 토픽 분배 엔진(210)은 토픽과 연관된 중요도 또는 토픽과 연관된 지리적 관심 영역과 같은 연관된 정보 또는 그 밖의 연관된 정보를 더 포함할 수 있다. 그러한 데이터 및/또는 그 밖의 연관된 데이터는, 게시자가 콘텐츠가 소망되는 영역들에 집중적으로 콘텐츠를 제공하는 것을 지원한다.

[0035] 도3은 부적절하게 게시된 콘텐츠를 식별하기 위하여 시스템에서 사용되는 데이터 흐름을 나타내는 블록도이다. 예를 들어, 웹-기반 게시가 도시된다. 다른 분배 양태들이 가능하다. 탐색자(130)는 질의를 탐색 엔진(120)에 보낸다. 탐색 엔진(120)은 탐색 질의에 대한 문서의 관련도(IR 스코어) 및 그 문서와 연관된 노드 랭크에 기초하여 질의들에 응답한다. 문서의 노드 랭크 및 IR 스코어와 연관된 결합된 스코어는 탐색 결과들의 품질을 평가하는데 사용될 수 있다. 그러한 정보는 통계 수집 및 분석 엔진(110)에 전달될 수 있다. 일부 구현예에서, 시스템은 예를 들어, 사용자가 결과를 클릭했는지 여부, 사용자가 클릭의 결과를 얼마나 오래 동안 검토했는지, 사용자가 이 사이트를 태그 또는 추천했는지 여부 등과 같은 다양한 사용자 행동을 추적하고, 추적된 행동들에 기초하여 탐색의 품질을 평가할 수 있다.

[0036] 통계 수집 및 분석 엔진(110)은 질의 정보를 수집 및 분석할 수 있다. 질의 정보는 질의들과 카테고리들(예를 들어, 토픽들)을 연관시키기 위하여 분석될 수 있다. 질의들은 탐색 결과들, 탐색 용어들의 공통성 또는 그 밖의 다른 공통 인자에 기초하여 서로 연관될 수 있다. 더욱이, 통계 수집 및 분석 엔진은 토픽과 연관된 탐색 질의들로부터의 탐색 결과들과 연관된 종합 랭킹(315) 및 종합 관련도(310)를 결정할 수 있다. 통계 수집 및 분석 엔진(110)은 예를 들어, 토픽에 대한 탐색 결과들의 종합 품질과 토픽의 인기도를 비교하기 위한 통계 비교(320)를 수행하여, 토픽 코퍼스(355)가 토픽을 적절히 서비스하고 있는지 여부를 결정할 수 있다. 이 결정은, 토픽에 대한 탐색 결과들의 품질과, 탐색량(search volume)에 기초한 유사한 인기도를 가진 다른 토픽들에 대한 탐색 결과들의 품질을 비교하는 것에 근거하여, 이루어질 수 있다. 토픽 코퍼스가 토픽을 적절히 서비스하고 있는지 여부를 결정하기 위한 다른 방법들이 또한 사용될 수 있다.

[0037] 통계 수집 및 분석 엔진(110)은 불충분한 콘텐츠를 갖는 토픽들을 토픽 분배 엔진(210)에 전달할 수 있다. 토픽 분배 엔진(210)은 예를 들어, 토픽들을 그 주제에 대한 지식이 있는 콘텐츠 생성기(예를 들어, 온라인 게시자(315))에 분배할 수 있다. 일부 구현예에서, 토픽 분배 엔진(210)은 불충분 서비스된 토픽 탐색 엔진에 기초하여 토픽들을 분배할 수 있고, 이에 의해, 콘텐츠 생성자(예를 들어, 게시자, 사용자 등)는 자신이 익숙한 토픽들을 탐색할 수 있다. 예를 들어, 콘텐츠 생성자가 네트워크 관리에 익숙하면, 콘텐츠 생성자는 네트워크 관리에 관계된 불충분 서비스된 토픽들을 탐색하기 위하여 불충분 서비스된 토픽 탐색 엔진을 사용할 수 있다.

[0038] 토픽 분배 엔진(210)은 예를 들어, 토픽이 불충분 서비스된 정도(325)에 기초하여 토픽들을 재정렬할 수 있다. 예를 들어, 토픽이 불충분 서비스된 정도는 토픽의 인기도와 탐색 결과들의 품질간 불균형에 기초할 수 있다. 대안적으로, 토픽이 불충분 서비스된 정도는 토픽에 관련된 탐색들의 인기도에 기초하거나 또는 토픽과 관계된 탐색 결과들에 연관된 품질의 역(inverse)에 기초하거나, 또는 그 밖의 다른 결정에 의할 수 있다.

[0039] 다른 구현예에서, 토픽 분배 엔진(210)은 토픽들에 관한 정보를 판매할 수 있다(330). 예를 들어, 웹 게시자들이 광고 또는 가입을 판매하면, 웹페이지에 대한 증가하는 방문자들은 수입을 증가시킬 수 있다. 수입에서의 이러한 증가에 기초하여, 더욱 많은 방문자들을 유인하는 토픽들이 웹 게시자들에게 판매될 수 있다. 토픽들을 분배하기 위한 다른 시스템들이 사용될 수 있다.

[0040] 온라인 게시자(335)는 제안된 토픽들을 임의의 적절한 분배 시스템을 사용해 수신할 수 있다. 온라인 게시자(335)에는 많은 게시자들 중 특히, 광고 기반 웹 게시자(340), 사용자 기여 기반 웹 게시자(345), 및/또는 자동

화된 웹 게시자(350)가 포함된다. 온라인 게시자는 예를 들어, 부가적 콘텐츠를 토픽 코퍼스(355)에 제공할 수 있다. 웹 게시자는 예를 들어, 게시자 인센티브 시스템을 통해 부가적 콘텐츠를 제공하도록 장려될 수 있다. 동일하게, 게시자는 부가적 광고 또는 가입 판매를 위한 동기(motivation)에 의해 부가적 콘텐츠를 제공하도록 장려될 수 있다.

[0041] 토픽 코퍼스(355)는 특정 토픽에 관련된 탐색들에 관계하는 모든 문서들을 포함할 수 있다. 온라인 토픽 코퍼스(355)는 예를 들어, 토픽 코퍼스(355)의 각 문서에 대한 관련도(365) 및 랭크(360)를 포함한다. 문서의 관련도와 랭크는 인덱스를 형성하도록 결합될 수 있으며, 탐색 엔진(120)은 인덱스를 사용하여 검색 결과들을 정렬할 수 있다. 부가적 콘텐츠는 새로운 콘텐츠의 관련도 및 랭크에 기초하여 인덱스될 수 있고, 부가적 콘텐츠가 특정 토픽에 대한 탐색 결과들의 품질을 증진시키는지 여부를 결정하기 위하여, 기존의 콘텐츠의 인덱스와 비교될 수 있다. 일부 구현예에서, 온라인 게시자들에게는 토픽과 연관된 콘텐츠의 품질 랭크를 향상시키는 콘텐츠를 부가할 인센티브가 제공될 수 있다.

[0042] 도4는 부적절하게 서비스되는 것으로 식별된 토픽에 관계된 웹 콘텐츠를 모으기 위하여 시스템에서 사용되는 데이터 흐름을 나타내는 블록도이다. 일부 실시예에서, 여러 문서들 사이에 걸친 콘텐츠에 의해 낮은 품질의 결과들이 유발될 수 있다. 그러한 예에서, 문서들로부터의 콘텐츠는 검사되고 단일의 어드레스 또는 사이트에 포함되도록 편집될 수 있다. 도4에서, 종합 엔진(410)은 예를 들어, 토픽 코퍼스(355)에 포함된 정보를 수집하도록 동작할 수 있다. 다음 종합 엔진(410)은 수집된 정보를 어셈블하고 종합 사이트(420)를 생성한다. 종합 사이트(420)는 예를 들어, 토픽 코퍼스(355)에 포함된 문서 각각에 대한 링크를 포함할 수 있다. 일 구현예에서, 종합 사이트(420)는 토픽 코퍼스(355)로부터 복사된 콘텐츠를 포함할 수 있다. 다른 구현예에서, 사용자는 종합 엔진(410)에 의해, 종합 사이트(420) 또는 종합된 정보를 편집하여 토픽에 관련된 새로운 콘텐츠를 생성할 수 있다. 이러한 구현예에 의해, 웹 게시자들 및/또는 개별 사용자들은 토픽을 더욱 효과적으로 연구하고 그 토픽에 연관된 콘텐츠를 제공할 수 있다.

[0043] 도5는 불충분한 웹 콘텐츠를 식별하는 예시적 방법을 나타내는 흐름도이다. 단계(500)에서, 탐색 통계가 수집된다. 탐색 통계는 예를 들어, 수집 엔진(도1의 통계 수집 및 분석 엔진(110))에 의해 수집될 수 있다. 이 수집 엔진은 예를 들어, 탐색 엔진(예를 들어, 도1의 탐색 엔진(120))에 통합될 수 있다. 탐색 통계는 예를 들어, 탐색열, 탐색열에 대한 세분화(refinement), 결과, 관련도, 랭킹 등을 포함하는 탐색 질의에 관한 정보를 포함할 수 있다. 일 구현예에 있어서, 탐색 통계는 온라인 탐색들에 관계된다. 다른 구현예에서, 탐색 통계는 예를 들어, "도움말" 기능, 인터넷 탐색 엔진, 사용자 기여 사이트 탐색 엔진 등을 포함하는 소프트웨어 프로그램이 제공한 탐색들에 관계될 수 있다.

[0044] 단계(510)에서, 질의들은 토픽과 연관될 수 있다. 질의들은 예를 들어, 분석 엔진(예를 들어, 도1의 통계 수집 및 분석 엔진(110))에 의해 토픽과 연관될 수 있다. 일부 실시예에서, 분석 엔진은 탐색 엔진(예를 들어, 도1의 탐색 엔진(120))의 일부가 될 수 있다. 분석 엔진은 예를 들어, 탐색 질의들에 공통 용어를 사용하고, 탐색 질의들의 결과들을 공통성의 관점에서 분석하고, 탐색 질의들과 연관된 세분화를 분석하는 등에 의해 토픽들을 유도할 수 있다.

[0045] 토픽 코퍼스의 품질은 단계(520)에서 탐색 통계에 비교될 수 있다. 토픽 코퍼스 품질과 탐색 통계의 비교는 예를 들어, 분석 엔진(예를 들어, 도1의 통계 수집 및 분석 엔진(110))에 의해 수행될 수 있다. 토픽 코퍼스의 품질은 예를 들어, 그 토픽과 연관된 검색 질의들로부터 결과들의 인덱스에 기초하여 유도될 수 있다. 인덱스는 예를 들어, 검색 질의에 대한 문서의 관련도, 및 그 검색 질의에 대하여 관련이 있는 것으로 발견된 문서의 랭킹에 기초하여 유도될 수 있다. 예를 들어, 유사한 인기도를 갖는 토픽들 사이에서 토픽 코퍼스의 품질을 비교함으로써, 토픽 코퍼스의 품질이 품질의 기대 레벨(예를 들어, 주어진 인기도의 토픽에 대한 임계값)을 충족, 초과 또는 하회하는지 여부를 결정할 수 있다.

[0046] 단계(530)에서, 비교에 기초하여 토픽들이 식별될 수 있다. 토픽들은 분석 엔진(예를 들어, 도1의 분석 엔진(110))에 의해 표식(marked)될 수 있다. 토픽들은 토픽과 연관된 토픽 코퍼스가 유사한 탐색 통계를 가진 토픽들과 연관된 기대 품질을 충족, 초과 또는 하회하는지 여부에 기초하여 표식될 수 있다. 일부 구현예에서, 표식된 토픽들은 불충분 서비스된 것으로 간주될 수 있다. 불충분 서비스된 토픽들은 예를 들어, 연관된 콘텐츠가 토픽과 연관된 탐색 통계에 근거하여 불충분한 토픽들일 수 있다. 따라서 식별된 토픽들은 콘텐츠가 불충분한 인기 토픽들을 분별할 수 있으므로, 콘텐츠 생성자(예를 들어, 콘텐츠 생성자(220, 230, 240))에게 추가적인 콘텐츠를 제공하도록 요청할 수 있다.

[0047] 도6은 불충분한 웹 콘텐츠를 식별하는데 사용되는 예시적 방법을 나타내는 흐름도이다. 단계(600)에서, 탐색 요

청이 수신된다. 탐색 요청은 예를 들어, 탐색 엔진(예를 들어, 도1의 탐색 엔진(120))에 의해 수신될 수 있다. 탐색 요청은 예를 들어, 주제(subject)를 식별하는 탐색 질의를 포함할 수 있다.

- [0048] 단계(610)에서, 탐색 결과들의 품질이 결정된다. 품질은 예를 들어, 탐색 엔진(예를 들어, 도1의 탐색 엔진(120))에 의해 결정될 수 있다. 상술한 바와 같이, 탐색 결과들의 품질은 탐색 질의에 대한 문서들의 관련도 및 탐색 질의에 관련이 있는 문서들의 랭킹에 기초하여 유도될 수 있다. 문서들의 관련도 및 랭킹은 인덱스를 형성하도록 결합될 수 있다. 인덱스는 예를 들어, 탐색 결과들과 연관된 품질 인덱스를 제공할 수 있다.
- [0049] 단계(620)에서, 탐색 결과들의 품질이 낮은지 여부가 결정된다. 예를 들어, 탐색 엔진(예를 들어, 도1의 탐색 엔진(120))이 단계(620)의 결정을 할 수 있다. 탐색 엔진은 예를 들어, 품질과 임계 품질을 비교하여 탐색 결과들과 연관된 품질이 낮은지 여부를 결정할 수 있다. 대안적으로, 탐색 엔진은 탐색 질의들과 연관된 통계를 수집하고, 탐색 결과들의 품질과, 유사한 탐색 통계를 갖는 질의들의 결과 품질의 비교에 기초하여, 탐색 결과의 품질이 낮은지 여부를 결정할 수 있다.
- [0050] 탐색 결과들의 품질이 낮지 않다면, 단계(630)에서 탐색 결과들이 탐색자에게 반환된다. 예를 들어, 탐색 엔진(예를 들어, 도1의 탐색 엔진(120))은 네트워크(예를 들어, 도1의 네트워크(140))를 사용하여 탐색 결과들을 반환할 수 있다. 탐색 결과들은 탐색 결과들을 구성하는 문서들 각각과 연관된 품질 인덱스에 기초하여, 정렬될 수 있다.
- [0051] 하지만 탐색 결과들의 품질이 낮으면, 단계(640)에서 탐색자에게 통보된다. 예를 들어, 탐색 엔진(예를 들어, 도1의 탐색 엔진(120) 또는 도2의 토픽 분배 엔진(210))은 네트워크(예를 들어, 도1의 네트워크(140))를 사용하여 탐색자에게 통보할 수 있다. 탐색 엔진은 탐색과 연관된 결과들이 낮은 품질의 결과들을 산출했음을 탐색자에게 통보할 수 있다.
- [0052] 단계(650)에서, 콘텐츠를 추가하기 위한 요청이 제공된다. 분배 엔진(예를 들어, 도2의 토픽 분배 엔진(210))이 요청을 제공할 수 있다. 이러한 구현예에서, 요청은 통지를 동반할 수 있다. 요청은 예를 들어, 탐색자에게 주제를 연구하고 습득된 지식에 기초하여 부가적 콘텐츠를 제공하거나 또는 탐색자의 기존 지식에 근거하여 부가적 콘텐츠를 제공할 것을 요청할 수 있다.
- [0053] 도7은 불충분한 웹 콘텐츠를 식별하기 위한 예시적 방법을 나타내는 흐름도이다. 단계(700)에서, 제1 토픽과 연관된 통계가 검색된다. 통계는 예를 들어, 통계 수집 엔진(예를 들어, 도1의 통계 수집 및 분석 엔진(110))에 의해 검색될 수 있다. 검색된 통계에는 예를 들어, 탐색 질의들, 탐색 결과들, 탐색 질의들과 연관된 시간과 날짜 정보, 탐색 세션 동안 발생하는 탐색 질의들의 세부사항 등이 포함될 수 있다. 이러한 데이터는 토픽들을 정의하고, 토픽의 인기를 식별하고, 토픽에 대한 계절적 수요(seasonal demand)를 식별하는 등에 사용될 수 있다.
- [0054] 단계(710)에서, 토픽과 연관된 품질이 유도된다. 토픽과 연관된 품질은 예를 들어, 분석 엔진(예를 들어, 도1의 통계 수집 및 분석 엔진(110))에 의해 유도될 수 있다. 품질은 예를 들어, 토픽에 관계된 탐색 질의들을 충족하는 문서들과 연관된 랭킹 및 관련도의 종합에 기초하여 유도될 수 있다. 다른 실시예에서, 품질은 예를 들어, 탐색 결과들에 대한 클릭률 대 세부사항율(refinement rate)에 의해 유도될 수 있다. 토픽 코퍼스와 연관된 품질을 유도하기 위한 그 밖의 정보가 또한 사용될 수 있다.
- [0055] 단계(720)에서 토픽과 연관된 토픽 코퍼스의 품질은 그 토픽과 연관된 탐색량에 비교된다. 예를 들어, 분석 엔진(예를 들어, 도1의 통계 수집 및 분석 엔진(110))이 비교를 수행할 수 있다. 탐색량이 유사한 다른 토픽들에 관계하는 코퍼스 품질에 기초하여 비교가 이루어질 수 있다(예를 들어, 주어진 탐색량에 기초하여 어떤 코퍼스 품질이 기대되는지). 대안적으로, 코퍼스 품질이 유사한 다른 토픽들에 관계하는 탐색량에 기초하여 비교가 이루어질 수 있다(예를 들어, 주어진 코퍼스 품질에 기초하여 어떤 탐색량이 기대되는지).
- [0056] 단계(730)에서, 탐색량이 토픽 코퍼스를 능가하는지 여부가 결정된다. 예를 들어, 분석 엔진이 이 결정을 할 수 있다(예를 들어, 도1의 통계 수집 및 분석 엔진(110))이 비교기를 사용하여 결정함). 코퍼스 품질(예를 들어, 품질 인덱스)과 탐색량(예를 들어, 인기도)의 비교에 기초하여, 탐색량이 토픽 코퍼스를 능가한다. 탐색량이 토픽 코퍼스를 능가하는지 여부를 결정하는 다른 방법들이 사용될 수 있다.
- [0057] 탐색량이 콘텐츠를 능가하면, 단계(740)에서 토픽은 불충분 서비스된 것으로 표시되고, 인덱스된다. 토픽은 예를 들어, 분석 엔진(예를 들어, 도1의 통계 수집 및 분석 엔진(110))에 의해 표시 및 인덱스될 수 있다. 일부 실시예에서, 토픽은 불충분 서비스된 것으로 표시되어서, 불충분 서비스된 토픽들 탐색 인덱스(underserved topics search index) 내로 그 토픽의 포함을 제공한다. 토픽은 예를 들어, 콘텐츠에 의해 그 토픽이 불충분 서

비스된 정도를 표시하기 위하여 인덱스될 수 있다(예를 들어, 토픽과 연관된 콘텐츠의 품질에 기초함). 대안적으로 또는 부가적으로, 토픽은 그 토픽의 수요 정도를 표시하기 위하여 식별될 수 있다(예를 들어, 토픽과 연관된 탐색량에 기초함).

[0058] 단계(750)에서, 다음 토픽이 검색된다. 다음 토픽은 예를 들어, 분석 엔진(예를 들어, 도1의 통계 수집 및 분석 엔진(110))에 의해 검색될 수 있다. 토픽들은 예를 들어, 토픽 상점(topic store)으로부터 검색될 수 있다. 상술한 바와 같이, 토픽들은 예를 들어, 질의 클러스터링에 기초할 수 있다. 단계(720,730-740)는 프로세스가 종료할 때까지 반복될 수 있다.

[0059] 다른 구현예에서, 본 시스템과 방법은 과도하게 서비스된(over-served) 콘텐츠를 식별하는데 사용될 수 있다. 예를 들어, 콘텐츠는 토픽의 인기도에 기초하여 기대되는 것보다, 탐색 결과들과 연관된 매우 높은 품질 인덱스를 가진다. 과도하게 서비스된 콘텐츠가 예를 들어, 웹 게시자에게 통지될 수 있고, 웹 게시자는 그러한 토픽에 대하여 부가적 콘텐츠의 생성을 회피할 수 있다.

[0060] 다른 구현예에 있어서, 다른 유형의 게시자들(예를 들어, 인쇄업자와 같은 오프라인 게시자들)은 불충분 서비스되거나/과도하게 서비스된 토픽들의 식별을 사용할 수 있다. 그러한 게시자들은 불충분 서비스되거나/과도하게 서비스된 토픽들의 식별을 사용하여, 온라인 콘텐츠에 의해 서비스되지 않는 인기 있는 토픽을 찾을 수 있다. 예를 들어, "밀라드 필모어의 전기(biography of Millard Fillmore)"의 탐색이 소수의 결과를 얻으면, 이러한 게시자들은 밀라드 필모어의 삶에 관한 책의 주문을 고려할 수 있다.

[0061] 다른 구현예에서, 제품이 탐색되면, 불충분 서비스되거나/과도하게 서비스된 토픽들의 식별은 그 제품의 관심을 측정하는데 도움을 줄 수 있다. 예를 들어, 통계 수집 및 분석 시스템(예를 들어, 도1의 통계 수집 및 분석 엔진(110))이 "자웃빛 악어가죽 벨트"와 같은 제품에 대한 탐색이 불충분 서비스되었다고 결정하면, 소매상은 그러한 제품의 제공을 고려할 수 있다.

[0062] 본 명세서에 기재된 시스템 및 방법은, 네트워크(예를 들어, 근거리 네트워크(LAN), 광역 네트워크(WAN), 및 인터넷 등), 광섬유 매체, 반송파 웨이브(carrier wave), 무선 네트워크 등을 사용해 전달되는 데이터 신호를, 하나 이상의 데이터 프로세싱 디바이스(예를 들어, 컴퓨팅 디바이스(130), 탐색 엔진(120) 등)와 통신하기 위하여 사용할 수 있다. 데이터 신호는 디바이스로부터 또는 디바이스로 제공되는 본 명세서에 개시된 모든 또는 임의의 데이터를 반송할 수 있다.

[0063] 본 명세서에 기재된 시스템 및 방법은 많은 상이한 유형의 프로세싱 디바이스 상에 프로그램 코드에 의해 구현될 수 있고, 이 프로그램 코드는 하나 이상의 프로세서에 의해 실행될 수 있는 프로그램 인스트럭션들을 포함한다. 이 소프트웨어 프로그램 인스트럭션들은 소스 코드, 목적 코드, 기계 코드, 또는 프로세싱 디바이스가 본 명세서에 기재된 방법을 수행하도록 하는 그 밖의 저장된 데이터를 포함할 수 있다.

[0064] 본 명세서에 기재된 시스템 및 방법은, 방법의 단계를 수행하고 본 명세서에 기재된 시스템을 구현하는 프로세서가 실행 중에 사용하는 인스트럭션들을 담고 있는 컴퓨터 저장 매커니즘들(예를 들어, CD-ROM, 디스켓, RAM, 플래시 메모리, 컴퓨터 하드 드라이브 등)을 포함하는 많은 상이한 유형의 컴퓨터 판독가능 매체 상에 제공될 수 있다.

[0065] 본 명세서에 기재된 컴퓨터 구성요소, 소프트웨어 모듈, 함수들 및 데이터 구조들은 그들의 동작에 필요한 데이터 흐름이 가능하도록, 직접적 또는 간접적으로 상호 연결될 수 있다. 또한, 소프트웨어 인스트럭션들 또는 모듈은 예를 들어, 코드의 서브루틴 유닛, 코드의 소프트웨어 함수 유닛, 개체(객체 지향 패러다임에서), 애플릿, 컴퓨터 스크립트 언어, 또는 그 밖의 유형의 컴퓨터 코드나 펌웨어로서 구현될 수 있음에 주의해야 한다. 소프트웨어 구성요소 및/또는 기능은 상황에 따라서 단일 디바이스에 위치되거나 다수의 디바이스에 걸쳐 분산될 수 있다.

[0066] 본 명세서는 본 발명의 바람직한 실시예를 개시하고, 또한, 본 발명을 설명하고 이 기술분야에서 통상의 지식을 가진 자가 본 발명을 실시 및 사용할 수 있도록 실시예를 제공하고 있다. 본 명세서는 여기에 개시된 용어로 본 발명을 한정하고 있지 않다. 따라서 앞서 개시한 실시예를 참조하여 본 발명이 설명되었지만, 당업자는 본 발명의 사상을 이탈하지 않고 실시예에 변경, 변형 및 수정을 가할 수 있다.

[0067] 상술한 구현 및 그 밖의 구현들은 후술하는 청구항의 범위 내에 있다.

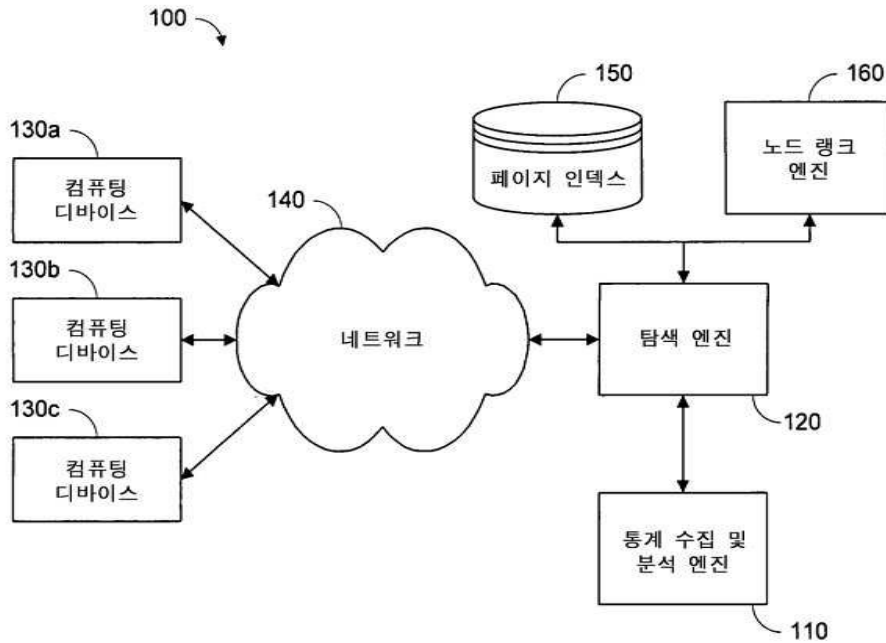
[0068] 산업상 이용가능성

[0069]

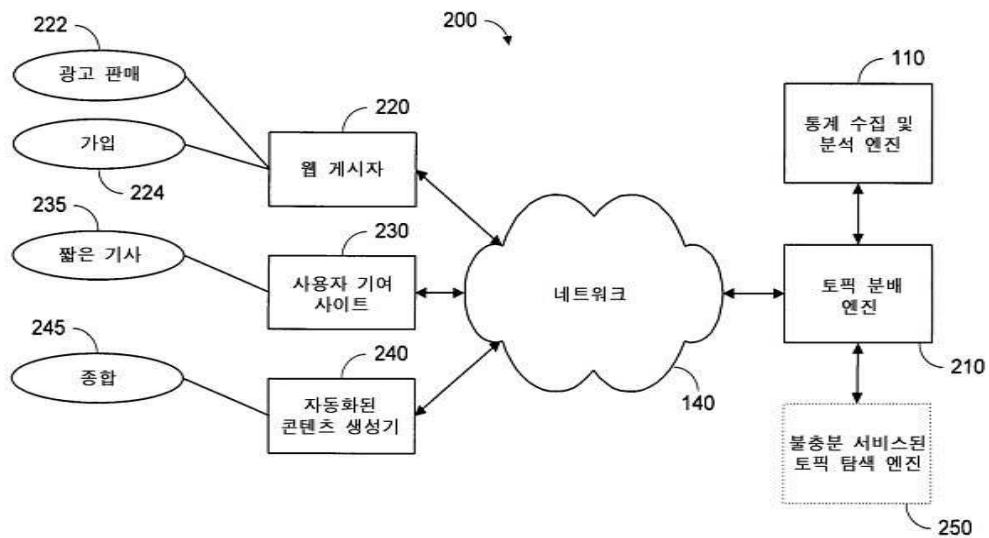
본 발명은 불충분한 탐색 콘텐츠를 식별하기 위한 시스템 및 방법을 제공한다.

도면

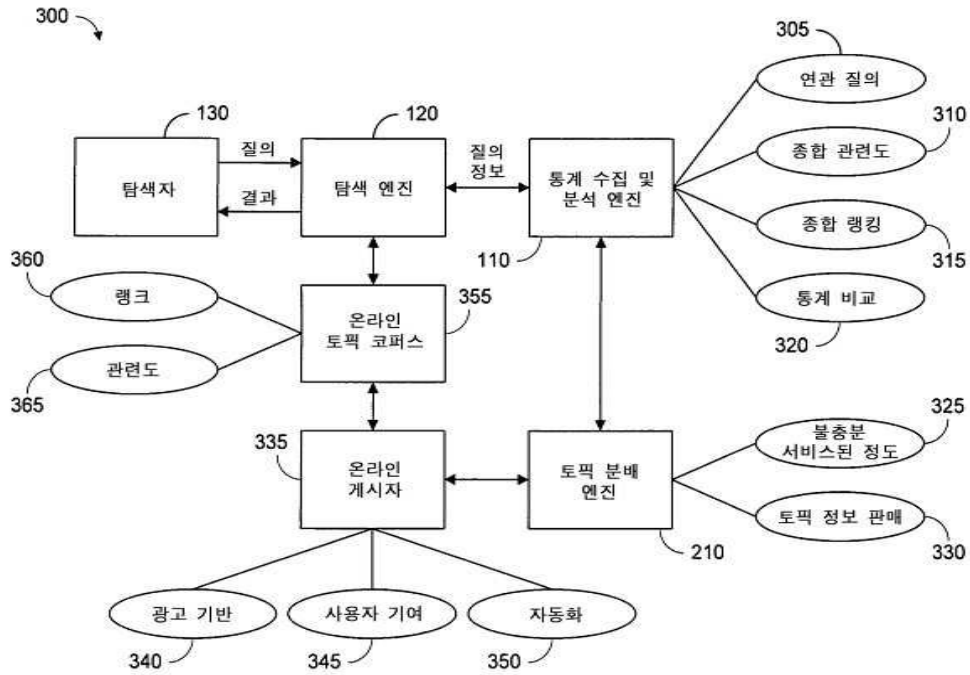
도면1



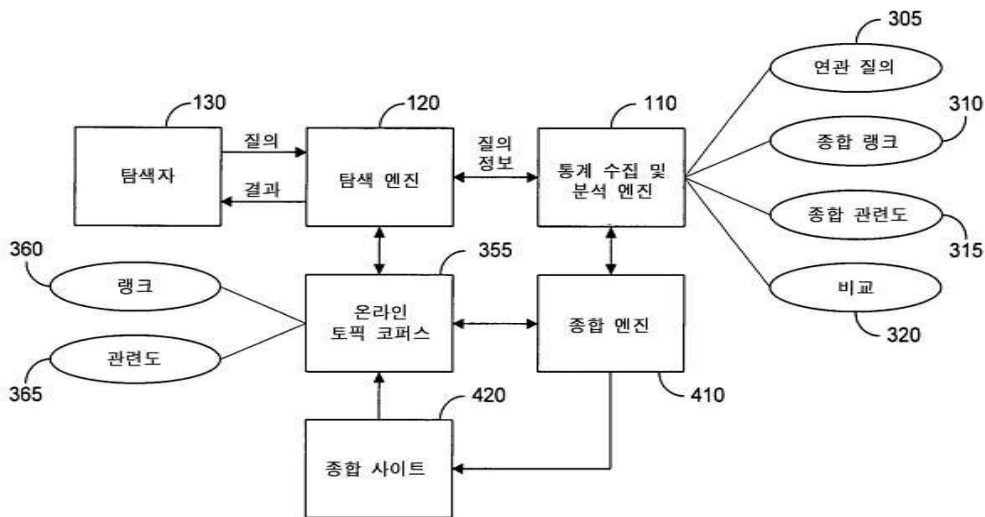
도면2



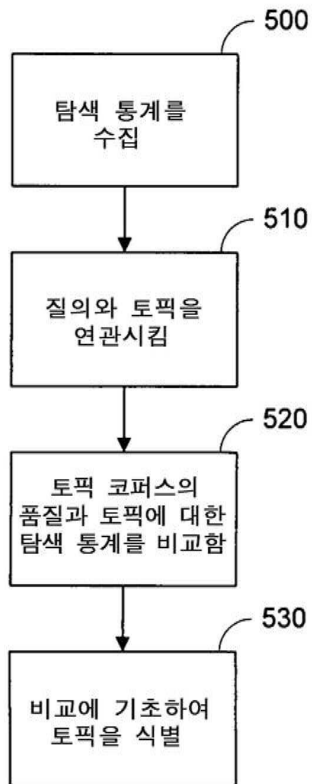
도면3



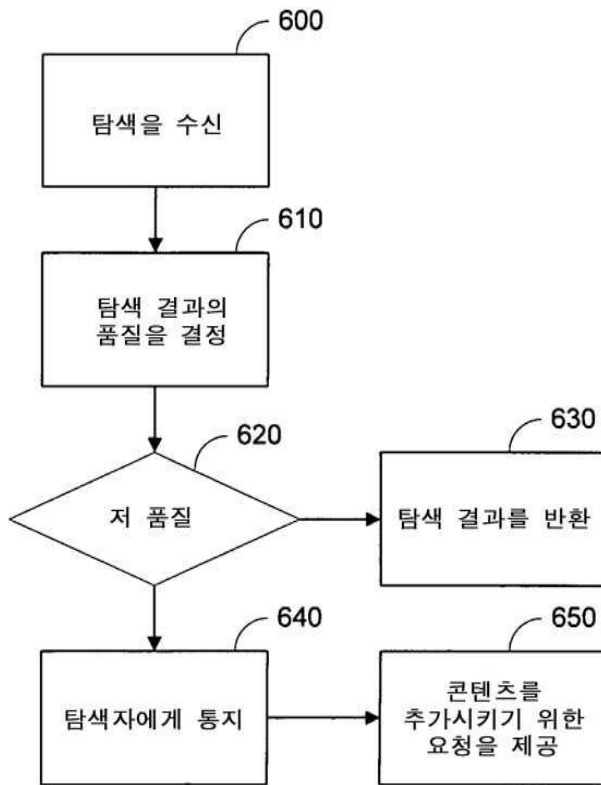
도면4



도면5



도면6



도면7

