



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2013년03월26일
 (11) 등록번호 10-1247739
 (24) 등록일자 2013년03월20일

(51) 국제특허분류(Int. Cl.)
 G06F 19/00 (2011.01) G06F 17/21 (2006.01)
 G06F 17/10 (2006.01)
 (21) 출원번호 10-2012-0077376
 (22) 출원일자 2012년07월16일
 심사청구일자 2012년07월16일
 (56) 선행기술조사문헌
 KR101102468 B1*
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
 한국과학기술정보연구원
 대전광역시 유성구 대학로 245 (어은동)
 (72) 발명자
 송사광
 대전광역시 서구 청사로 5 하나로아파트 107-205
 이승우
 대전광역시 유성구 어은동 한빛아파트 123동 801호
 (뒷면에 계속)
 (74) 대리인
 특허법인화우

전체 청구항 수 : 총 16 항

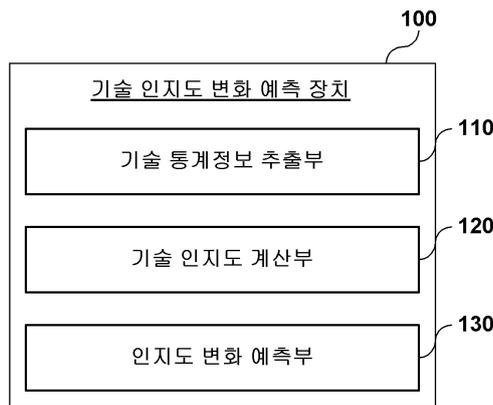
심사관 : 김동성

(54) 발명의 명칭 **기술 인지도 변화 예측 장치 및 방법**

(57) 요약

본 발명은 기술 인지도 변화 예측 장치 및 방법에 관한 것으로, 다중 문헌 제공 장치에 저장된 문서들을 분석하여, 특정 기술에 대한 통계 정보를 추출하는 기술 통계 정보 추출부, 상기 통계 정보를 기반으로 상기 특정 기술의 인지도 값을 문서별로 각각 구하는 기술 인지도 계산부, 상기 문서별로 구해진 인지도 값들을 연도별로 수집 및 평균값을 구하고, 연도별 평균값을 기반으로 상기 특정 기술의 인지도 변화 속도를 예측하는 인지도 변화 예측부를 포함한다.

대표도 - 도1



(72) 발명자

조민희

대전광역시 유성구 하기동 송림마을5단지 503-1702

이진희

대전광역시 유성구 봉명동 604-2 상그릴라플러스빌
416호

정한민

대전광역시 유성구 어은동 99 한빛아파트 111동
1404호

성원경

대전광역시 유성구 엑스포로 448 엑스포아파트
102-604

특허청구의 범위

청구항 1

다중 문헌 제공 장치에 저장된 문서들을 분석하여, 특정 기술에 대한 통계 정보를 추출하는 기술 통계 정보 추출부;

상기 통계 정보를 기반으로 특정 기술이 추출된 문서의 신뢰도 값, 문서내에서 특정 기술이 추출된 위치정보 값, 특정 기술의 문서내 중요도 값, 특정 기술의 문서 전체 내 중요도 값 중 적어도 하나를 구하고, 그 구해진 값들을 이용하여 상기 특정 기술의 인지도 값을 문서별로 각각 구하는 기술 인지도 계산부; 및

상기 문서별로 구해진 인지도 값들을 연도별로 수집 및 평균값을 구하고, 연도별 평균값을 기반으로 상기 특정 기술의 인지도 변화 속도를 예측하는 인지도 변화 예측부;를 포함하되,

상기 특정 기술의 문서내 중요도 값은 문서내에 포함된 전체 기술명 빈도수(TF) 중에 상기 특정 기술의 빈도수(AF)를 통해 구해지는 것을 특징으로 하는 기술 인지도 변화 예측 장치.

청구항 2

제1항에 있어서,

상기 기술 통계정보 추출부는 상기 다중 문헌 제공 장치에 저장된 문서들을 분석하여 상기 특정 기술을 포함하는 문서들을 획득하고, 상기 획득된 문서들을 분석하여, 상기 특정 기술에 대한 문서내 빈도수, 문서 내에 포함된 전체 기술명 빈도수, 문서내 위치 및 위치별 빈도수, 문서 종류별 빈도수, 문헌의 연도 중 적어도 하나를 포함하는 통계정보를 추출하는 것을 특징으로 하는 기술 인지도 변화 예측 장치.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 기술 인지도 계산부는 문서의 신뢰도 값, 문서내에서 특정 기술이 추출된 위치정보 값, 특정 기술의 문서내 중요도 값, 특정 기술의 문서 전체 내 중요도 값을 합하여 상기 특정 기술의 인지도 값을 구하는 것을 특징으로 하는 기술 인지도 변화 예측 장치.

청구항 5

제1항에 있어서,

상기 문서의 신뢰도 값은 사용자에게 의해 설정된 가중치 또는 웹 검색 통계 정보를 기반으로 구해진 가중치를 이용하여 구해진 값인 것을 특징으로 하는 기술 인지도 변화 예측 장치.

청구항 6

삭제

청구항 7

제1항에 있어서,

상기 특정 기술의 문서 전체내 중요도 값은 전체 문서수(N)에서 특정 기술이 포함된 문서수(d)의 역(Inverse) 비율 값으로 구해지는 것을 특징으로 하는 기술 인지도 변화 예측 장치.

청구항 8

제1항에 있어서,

상기 인지도 변화 예측부는 상기 특정 기술에 대해 문서별로 구해진 인지도 값을 연도별로 수집하고, 그 수집된 인지도 값들의 연도별 평균을 구한 후, 상기 연도별 평균값을 기반으로 예측 연도의 인지도 값을 예측하고, 그 예측된 인지도 값을 이용하여 인지도의 변화 속도를 구하는 것을 특징으로 하는 기술 인지도 변화 예측 장치.

청구항 9

제8항에 있어서,

상기 인지도 변화 예측부는 회귀분석 방법, 이동평균분석 방법, 상기 두 방법을 혼합한 방법(ARMA) 중 적어도 하나의 방법을 이용하여 예측 연도의 인지도 값을 예측하는 것을 특징으로 하는 기술 인지도 변화 예측 장치.

청구항 10

제1항에 있어서,

상기 인지도 변화 예측부는 상기 예측된 변화 속도를 기 설정된 변화 방향 단계와 비교하고, 그 비교결과에 따라 인지도 변화 방향을 결정하는 것을 특징으로 하는 기술 인지도 변화 예측 장치.

청구항 11

기술 인지도 변화 예측 장치가 특정 기술의 인지도 변화를 예측하는 방법에 있어서,

- (a) 다중 문헌 제공 장치에 저장된 문서들을 분석하여, 특정 기술에 대한 통계 정보를 추출하는 단계;
- (b) 상기 통계 정보를 기반으로 특정 기술이 추출된 문서의 신뢰도 값, 문서내에서 특정 기술이 추출된 위치정보 값, 특정 기술의 문서내 중요도 값, 특정 기술의 문서 전체 내 중요도 값 중 적어도 하나를 구하고, 그 구해진 값들을 이용하여 상기 특정 기술의 인지도 값을 구하는 단계; 및
- (c) 상기 문서별로 구해진 인지도 값들을 연도별로 수집 및 평균값을 구하고, 연도별 평균값을 기반으로 상기 특정 기술의 인지도 변화 속도를 예측하는 단계;를 포함하되,

상기 특정 기술의 문서내 중요도 값은 문서내에 포함된 전체 기술명 빈도수(TF) 중에 상기 특정 기술의 빈도수(AF)를 통해 구해지는 것을 특징으로 하는 기술 인지도 변화 예측 방법.

청구항 12

제11항에 있어서,

상기 (a) 단계는,

상기 다중 문헌 제공 장치에 저장된 문서들을 분석하여 상기 특정 기술을 포함하는 문서들을 획득하고, 상기 획득된 문서들을 분석하여, 상기 특정 기술에 대한 문서내 빈도수, 문서 내에 포함된 전체 기술명 빈도수, 문서내 위치 및 위치별 빈도수, 문서 종류별 빈도수, 문헌의 연도 중 적어도 하나를 포함하는 통계정보를 추출하는 것을 특징으로 하는 기술 인지도 변화 예측 방법.

청구항 13

삭제

청구항 14

제11항에 있어서,

상기 특정 기술의 인지도 값은 상기 문서의 신뢰도 값, 문서내에서 특정 기술이 추출된 위치정보 값, 특정 기술의 문서 내 중요도 값, 특정 기술의 문서 전체 내 중요도 값을 합하여 구해진 값인 것을 특징으로 하는 기술 인지도 변화 예측 방법.

청구항 15

제11항에 있어서,

상기 문서의 신뢰도 값은 사용자에게 의해 설정된 가중치 또는 웹 검색 통계 정보를 기반으로 구해진 가중치를 이용하여 구해진 값인 것을 특징으로 하는 기술 인지도 변화 예측 방법.

청구항 16

삭제

청구항 17

제11항에 있어서,

상기 특정 기술의 문서 전체내 중요도 값은 전체 문서수(N)에서 특정 기술이 포함된 문서수(d)의 역(Inverse) 비율 값으로 구해지는 것을 특징으로 하는 기술 인지도 변화 예측 방법.

청구항 18

제11항에 있어서,

상기 (c) 단계는,

상기 특정 기술에 대해 문서별로 구해진 인지도 값을 연도별로 수집하고, 그 수집된 인지도 값들의 연도별 평균을 구하는 단계;

상기 연도별 평균값을 기반으로 예측 연도의 인지도 값을 예측하는 단계;

상기 예측된 인지도 값을 이용하여 인지도의 변화 속도를 구하는 단계를 포함하는 것을 특징으로 하는 기술 인지도 변화 예측 방법.

청구항 19

제11항에 있어서,

상기 예측된 인지도 변화 속도를 기 설정된 변화 방향 단계와 비교하고, 그 비교결과에 따라 인지도 변화 방향을 결정하는 단계를 더 포함하는 것을 특징으로 하는 기술 인지도 변화 예측 방법.

청구항 20

기술 인지도 변화 예측 장치에 의해 실행될 때,

다중 문헌 제공 장치에 저장된 문서들을 분석하여, 특정 기술에 대한 통계 정보를 추출하는 단계;

상기 통계 정보를 기반으로 특정 기술이 추출된 문서의 신뢰도 값, 문서내에서 특정 기술이 추출된 위치정보 값, 특정 기술의 문서내 중요도 값, 특정 기술의 문서 전체 내 중요도 값 중 적어도 하나를 구하고, 그 구해진 값들을 이용하여 상기 특정 기술의 인지도 값을 구하는 단계; 및

(c) 상기 문서별로 구해진 인지도 값들을 연도별로 수집 및 평균값을 구하고, 연도별 평균값을 기반으로 상기 특정 기술의 인지도 변화 속도를 예측하는 단계;를 포함하되,

상기 특정 기술의 문서내 중요도 값은 문서내에 포함된 전체 기술명 빈도수(TF) 중에 상기 특정 기술의 빈도수(AF)를 통해 구해지는 것을 특징으로 하는 기술 인지도 변화 예측 방법이 프로그램으로 기록되고 컴퓨터에서 단독 가능한 기록매체.

명세서

기술분야

- [0001] 본 발명은 기술 인지도 변화 예측 장치 및 방법에 관한 것으로, 보다 상세하게는 특정 기술에 대한 통계 정보를 기반으로 특정 기술의 인지도 값을 문서별로 각각 구하고, 문서별로 구해진 인지도 값들을 연도별로 수집 및 평균값을 구한 후, 연도별 평균값을 기반으로 상기 특정 기술의 인지도 변화 속도를 예측하는 기술 인지도 변화 예측 장치 및 방법에 관한 것이다.

배경 기술

- [0002] 지식과 정보가 그 국가의 경쟁력을 좌우하는 지식기반 산업사회로 전환되고, 특히 국가과학기술경쟁력이 국가경쟁력의 원천으로 인식되고 있는 실정이다.
- [0003] 이에 세계 각국들은 미래의 경쟁에 살아남기 위한 핵심기술 및 연구과제를 미리 도출하고 선정하여 집중적인 연구개발을 추진해 나가고자 하고 있다.
- [0004] 이러한 이유로 각국들은 미래 국가 과학기술을 주도할 유망한 연구영역을 찾아내어 기술 라이프 사이클을 제공 하고 있다.
- [0005] 그러나, 종래의 기술 라이프 사이클은 각 단계에 속하는 기술만을 제시할 뿐 각 기술의 인지도, 인지도의 변화 속도 및 변화 방향에 대한 정보를 제공하지 못하는 단점이 있다.

발명의 내용

해결하려는 과제

- [0006] 본 발명은 상기한 문제점을 해결하기 위하여 안출된 것으로, 기술에 대한 인지도, 그 인지도의 변화 속도 및 변화 방향에 대한 정보를 제공할 수 있는 기술 인지도 변화 예측 장치 및 방법을 제공하는데 있다.
- [0007] 본 발명의 다른 목적은 기술에 대한 시장에서의 인지도, 관심도 등을 기반으로 기술에 대한 전문가뿐만 아니라 일반인의 기대를 추정하고, 그 추정의 변화를 감지하여 기술의 인지도/기대치에 대한 미래 예측을 가능하게 하는 기술 인지도 변화 예측 장치 및 방법을 제공하는데 있다.

[0008]

과제의 해결 수단

- [0009] 상기 목적들을 달성하기 위하여 본 발명의 일 측면에 따르면, 다중 문헌 제공 장치에 저장된 문서들을 분석하여, 특정 기술에 대한 통계 정보를 추출하는 기술 통계 정보 추출부, 상기 통계 정보를 기반으로 상기 특정 기술의 인지도 값을 문서별로 각각 구하는 기술 인지도 계산부, 상기 문서별로 구해진 인지도 값들을 연도별로 수집 및 평균값을 구하고, 연도별 평균값을 기반으로 상기 특정 기술의 인지도 변화 속도를 예측하는 인지도 변화 예측부를 포함하는 기술 인지도 변화 예측 장치가 제공된다.

[0010] 상기 기술 통계정보 추출부는 상기 다중 문헌 제공 장치에 저장된 문서들을 분석하여 상기 특정 기술을 포함하는 문서들을 획득하고, 상기 획득된 문서들을 분석하여, 상기 특정 기술에 대한 문서내 빈도수, 문서 내에 포함된 전체 기술명 빈도수, 문서내 위치 및 위치별 빈도수, 문서 종류별 빈도수, 문헌의 연도 중 적어도 하나를 포함하는 통계정보를 추출할 수 있다.

[0011] 상기 기술 인지도 계산부는 상기 통계 정보를 기반으로 특정 기술이 추출된 문서의 신뢰도 값, 문서내에서 특정 기술이 추출된 위치정보 값, 특정 기술의 문서내 중요도 값, 특정 기술의 문서 전체 내 중요도 값을 각각 구하고, 그 구해진 값들을 이용하여 상기 특정 기술의 인지도 값을 구할 수 있다.

[0012] 상기 기술 인지도 계산부는 하기의 수학적식을 이용하여 상기 특정 기술의 인지도 값을 구할 수 있다.

[0013] [수학적식]

$$I = \frac{W_I \cdot W_J \cdot W_K \cdot W_L}{I \cdot J \cdot K \cdot L}$$

[0014]

[0015] 여기서, W_I 는 문서의 신뢰도 값, W_J 는 문서내에서 특정 기술이 추출된 위치 정보 값, W_K 는 특정 기술의 문서 내 중요도 값, W_L 은 특정 기술의 문서 전체 내 중요도 값, I, J, K, L 은 W_I, W_J, W_K, W_L 각각의 비중을 의미함.

[0016] 상기 문서의 신뢰도 값은 사용자에게 의해 설정된 가중치 또는 웹 검색 통계 정보를 기반으로 구해진 가중치를 이용하여 구해진 값일 수 있다.

[0017] 상기 특정 기술의 문서내 중요도 값은 문서내에 포함된 전체 기술명 빈도수(TF) 중에 상기 특정 기술의 빈도수(AF)를 통해 구해진다.

[0018] 상기 특정 기술의 문서 전체내 중요도 값은 전체 문서수(N)에서 특정 기술이 포함된 문서수(d)의 역(Inverse) 비율 값으로 구해진다.

[0019] 상기 인지도 변화 예측부는 상기 특정 기술에 대해 문서별로 구해진 인지도 값을 연도별로 수집하고, 그 수집된 인지도 값들의 연도별 평균을 구한 후, 상기 연도별 평균값을 기반으로 예측 연도의 인지도 값을 예측하고, 그 예측된 인지도 값을 이용하여 인지도의 변화 속도를 구할 수 있다.

[0020] 상기 인지도 변화 예측부는 회귀분석 방법, 이동평균분석 방법, 상기 두 방법을 혼합한 방법(ARMA) 중 적어도 하나의 방법을 이용하여 예측 연도의 인지도 값을 예측할 수 있다.

[0021] 또한, 상기 인지도 변화 예측부는 상기 예측된 변화 속도를 기 설정된 변화 방향 단계와 비교하고, 그 비교결과에 따라 인지도 변화 방향을 결정할 수 있다.

[0022] 본 발명의 다른 측면에 따르면, 기술 인지도 변화 예측 장치가 특정 기술의 인지도 변화를 예측하는 방법에 있어서, (a) 다중 문헌 제공 장치에 저장된 문서들을 분석하여, 특정 기술에 대한 통계 정보를 추출하는 단계, (b) 상기 통계 정보를 기반으로 상기 특정 기술의 인지도 값을 문서별로 각각 구하는 단계, (c) 상기 문서별로 구해진 인지도 값들을 연도별로 수집 및 평균값을 구하고, 연도별 평균값을 기반으로 상기 특정 기술의 인지도 변화 속도를 예측하는 단계를 포함하는 기술 인지도 변화 예측 방법이 제공된다.

[0023] 상기 (a) 단계는, 상기 다중 문헌 제공 장치에 저장된 문서들을 분석하여 상기 특정 기술을 포함하는 문서들을 획득하고, 상기 획득된 문서들을 분석하여, 상기 특정 기술에 대한 문서내 빈도수, 문서 내에 포함된 전체 기술명 빈도수, 문서내 위치 및 위치별 빈도수, 문서 종류별 빈도수, 문헌의 연도 중 적어도 하나를 포함하는 통계 정보를 추출하는 것을 특징으로 할 수 있다.

[0024] 상기 (b) 단계는, 상기 통계 정보를 기반으로 특정 기술이 추출된 문서의 신뢰도 값, 문서내에서 특정 기술이 추출된 위치정보 값, 특정 기술의 문서내 중요도 값, 특정 기술의 문서 전체 내 중요도 값을 각각 구하고, 그 구해진 값들을 이용하여 상기 특정 기술의 인지도 값을 구하는 것을 특징으로 할 수 있다.

[0025] 상기 특정 기술의 인지도 값은 하기의 수학적식을 이용하여 상기 특정 기술의 인지도 값을 구할 수 있다.

[0026] [수학적식]

$$c = \frac{w_I \cdot (-w_J) - w_K \cdot L}{I - K - L}$$

[0027]

[0028] 여기서, W_I 는 문서의 신뢰도 값, W_J 는 문서내에서 특정 기술이 추출된 위치 정보 값, W_K 는 특정 기술의 문서 내 중요도 값, W_L 은 특정 기술의 문서 전체 내 중요도 값, I, J, K, L은 W_I , W_J , W_K , W_L 각각의 비중을 의미함.

[0029] 상기 문서의 신뢰도 값은 사용자에게 의해 설정된 가중치 또는 웹 검색 통계 정보를 기반으로 구해진 가중치를 이용하여 구해진 값일 수 있다.

[0030] 상기 특정 기술의 문서내 중요도 값은 문서내에 포함된 전체 기술명 빈도수(TF) 중에 상기 특정 기술의 빈도수(AF)를 통해 구해지는 것을 특징으로 할 수 있다.

[0031] 상기 특정 기술의 문서 전체내 중요도 값은 전체 문서수(N)에서 특정 기술이 포함된 문서수(d)의 역(Inverse) 비율 값으로 구해질 수 있다.

[0032] 상기 (c) 단계는, 상기 특정 기술에 대해 문서별로 구해진 인지도 값을 연도별로 수집하고, 그 수집된 인지도 값들의 연도별 평균을 구하는 단계, 상기 연도별 평균값을 기반으로 예측 연도의 인지도 값을 예측하는 단계, 상기 예측된 인지도 값을 이용하여 인지도의 변화 속도를 구하는 단계를 포함할 수 있다.

[0033] 상기 기술 인지도 변화 예측 방법은 상기 예측된 변화 속도를 기 설정된 변화 방향 단계와 비교하고, 그 비교결과에 따라 인지도 변화 방향을 결정하는 단계를 더 포함할 수 있다.

[0034] 본 발명의 또 다른 측면에 따르면, 다중 문헌 제공 장치에 저장된 문서들을 분석하여, 특정 기술에 대한 통계 정보를 추출하는 단계, 상기 통계 정보를 기반으로 상기 특정 기술의 인지도 값을 문서별로 각각 구하는 단계, 상기 문서별로 구해진 인지도 값들을 연도별로 수집 및 평균값을 구하고, 연도별 평균값을 기반으로 상기 특정 기술의 인지도 변화 속도를 예측하는 단계를 포함하는 기술 인지도 변화 예측 방법이 컴퓨터로 판독 가능한 기록매체가 제공된다.

발명의 효과

[0035] 본 발명에 따르면, 기술에 대한 인지도, 그 인지도의 변화 속도 및 변화 방향에 대한 정보를 제공할 수 있다.

[0036] 또한, 기술에 대한 시장에서의 인지도, 관심도 등을 기반으로 기술에 대한 전문가뿐만 아니라 일반인의 기대를 추정하고, 그 추정의 변화를 감지하여 기술의 인지도/기대치에 대한 미래 예측을 가능하게 한다.

[0037] 또한, 기술·산업·시장·정책 등에 대한 글로벌 모니터링 체계를 구축하고, 이를 기반으로 이머징 시그널 및 미래기술 전후방 정보를 제공함으로써 혁신형 중소기업의 기술사업화 성공률 및 기술경쟁력 제고, 연구자/혁신형 중소기업의 사전 글로벌 연구동향 파악이 가능하여 중복연구방지 및 연구개발 시행착오를 최소화시킴으로써 핵심역량의 선택과 집중이 가능하여 효율성을 높이고, 이에 따른 기술사업의 성공 가능성 증대시킬 수 있다.

도면의 간단한 설명

[0038] 도 1은 본 발명에 따른 기술 인지도 변화 예측 장치의 구성을 개략적으로 나타낸 도면이다.

도 2는 본 발명에 따른 기술 인지도 변화 예측 장치가 기술 인지도의 변화를 예측하는 방법을 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0039] 본 발명의 기술한 목적과 기술적 구성 및 그에 따른 작용 효과에 관한 자세한 사항은 본 발명의 명세서에 첨부된 도면에 의거한 이하 상세한 설명에 의해 보다 명확하게 이해될 것이다.
- [0040] 도 1은 본 발명에 따른 기술 인지도 변화 예측 장치의 구성을 개략적으로 나타낸 도면이다.
- [0041] 도 1을 참조하면, 기술 인지도 변화 예측 장치(100)는 기술 통계정보 추출부(110), 기술 인지도 계산부(120), 인지도 변화 예측부(130)를 포함한다.
- [0042] 기술 통계 정보 추출부(110)는 문서들을 분석하여 기술명에 대한 통계 정보를 추출한다. 여기서, 문서는 논문, 특허, 보고서, 웹 문서 등을 포함한다. 통계 정보는 문서내에서 추출된 기술명, 문서내 기술명의 빈도수, 문서내에 포함된 전체 기술명 빈도수, 기술명이 추출된 문서내의 위치 및 위치별 빈도수, 문서 종류별 빈도수, 문서내 개별 기술의 통계, 기술명이 추출된 문서의 연도 등을 포함한다.
- [0043] 기술 통계 정보 추출부(110)는 다중 문헌 제공 장치에 저장된 문서들을 분석하여 기술용어 사전에 정의된 기술명을 포함하는 문서들을 획득하고, 그 문서들로부터 기술명을 추출한다. 기술용어 사전에는 의학, 공학, 자연과학 등 다양한 분야의 기술용어들이 저장되어 있고, 다중문헌제공 장치에는 논문, 특허, 보고서 등의 문서가 저장되어 있다.
- [0044] 즉, 기술 통계정보 추출부(110)는 문서에 대해 형태소 분석, 구문 분석, 의미 분석을 수행하여 용어를 인식하고, 인식된 용어가 기술 용어 사전에 등록된 경우, 기술명으로 인식한다. 이때, 기술 통계 정보 추출부(110)는 각 기술명이 추출된 위치를 판단하고, 각 기술명과 위치를 함께 저장한다. 위치는 제목, 부제, 중간 제목, 내용, 초록 등을 포함한다. 또한, 기술 통계 정보 추출부(110)는 기술명이 추출된 문서의 위치별 빈도수를 함께 저장한다.
- [0045] 또한, 기술 통계정보 추출부(110)는 각 문서에서 기술명이 추출된 빈도수를 누적하고, 각 기술명이 추출된 문서들의 연도를 함께 저장한다.
- [0046] 또한, 기술 통계 정보 추출부(110)는 기 저장된 문맥 패턴 정보를 근거로 추출된 기술명 사이의 문맥을 추출하고, 구비된 정규화 사전 데이터베이스를 근거로 추출된 문맥을 정규화하는 방식으로 기술명간의 관계를 인식한다.
- [0047] 기술 통계 정보 추출부(110)는 2개 이상의 기술명이 인식된 경우, 문맥 패턴 정보를 근거로 그 기술명 사이의 문맥을 추출하고, 정규화 사전 데이터베이스를 근거로 추출된 문맥을 정규화하는 방식으로 기술명의 관계를 생성한다.
- [0048] 예를 들어, "notebook consists of monitor, SSD, and mobile processor" 문장의 경우, 기술 통계 정보 추출부(110)는 notebook, monitor, SSD, mobile processor의 기술명을 인식하게 된다.
- [0049] 이때, 2개 이상의 기술명이 인식되었으므로, 기술 통계 정보 추출부(110)는 기 저장된 문맥 패턴 정보 예를 들면, "Be동사 + 과거분사형 + 전치사"에 부합하는 문맥만을 추출한다.
- [0050] 그런 다음 기술 통계 정보 추출부(110)는 정규화 사전을 근거로 해당 문맥의 시제 예를 들면 "was"를 "be"로 통일하고, 유사표현 예를 들면, "similar to"와 "similar with"를 "similar to"로 통일하는 등의 문맥 정규화를 수행하여 기술명간의 관계를 생성한다.
- [0051] 기술 인지도 계산부(120)는 기술 통계 정보 추출부(110)에서 추출된 통계 정보를 기반으로 각 기술의 인지도 값을 구한다. 기술 인지도는 특정 기술이 전문가, 일반인 등에게 얼마만큼 중요하게 인식되고 있는지에 대한 정보를 수치화한 값이다. 이때, 기술 인지도 계산부(120)는 각 기술에 대한 인지도 값을 개별 문서별로 각각 구한다.
- [0052] 기술 인지도 계산부(120)는 통계 정보를 기반으로 특정 기술이 추출된 문서의 신뢰도 값, 문서내에서 특정 기술이 추출된 위치정보 값, 특정 기술의 문서내 중요도 값, 특정 기술의 문서 전체 내 중요도 값을 각각 구하고, 그 구해진 값들을 이용하여 특정 기술의 인지도 값을 구한다.
- [0053] 즉, 기술 인지도 계산부(120)는 수학적 식 1을 이용하여 각 기술에 대한 인지도(W_D)를 문서별로 구한다.

수학식 1

$$W_D = \frac{W_I * I + W_J * J + W_K * K + W_L * L}{I + J + K + L}$$

[0054]

[0055] 여기서, W_I 는 기술명이 추출된 소스 문서의 신뢰도 값, W_J 는 문서내에서 기술명이 추출된 위치 정보 값, W_K 는 기술명의 문서 내 중요도 값, W_L 은 기술명의 문서 전체 내 중요도 값을 의미한다. I, J, K, L은 W_I , W_J , W_K , W_L 각각의 비중을 의미하는 값으로, 그 값들은 미리 설정되어 있다.

[0056] 기술명이 추출된 소스 문서의 신뢰도(W_I)는 문서가 작성된 사이트 또는 데이터베이스의 신뢰도를 의미하고, 뉴욕 타임즈, CNN 등 다양한 사이트별 신뢰도를 가중치로 표현한 수치이다. W_I 가 큰 값을 갖게 되는 경우 소스의 신뢰도는 높은 것이고, 반대로 낮은 값을 갖게 되는 경우 신뢰도가 낮은 것을 의미한다.

[0057] W_I 는 수학식 2를 이용하여 구해진다.

수학식 2

$$W_I = \alpha * W_{IM} + \beta * W_{IA}$$

[0058]

[0059] 여기서, W_{IM} 은 사람의 직관과 경험을 통해 수동적으로 계산된 가중치로, 미리 테이블 형태로 정의된다.

W_{IA} 는 웹 검색 통계 정보를 기반으로 자동으로 얻어진 가중치로, 전처리를 통해 미리 또는 주기적으로 계산된 값일 수 있다. α 는 W_{IM} 의 W_I 에서의 비중, β 는 W_{IA} 의 W_I 에서의 비중을 의미한다.

[0060] 이하, W_{IA} 를 구하는 방법에 대해 살펴보기로 한다.

[0061] 각 리소스에 대한 가중치는 기본적으로 검색엔진의 검색 결과로 얻는다. 특정 리소스, 예를 들어 cnn.com에 대한 가중치(사람들의 선호도/인지도)는 웹 문서 내에서 cnn.com을 포함하는 문서 빈도수의 비율(W_{IAT})에 기반한다. 이 문서 빈도수는 기술명 또는 (기술명+관계명)을 질의어로 검색엔진을 통해 검색했을 때 나오는 검색 결과 문서 중 cnn사이트의 문서수(g), cnn사이트의 문서가 아닌 문서 중 cnn을 참조(링크)하는 문서수(h)를 합쳐서 계산한다. 이때 효율을 위해 상위 n개의 문서에서만 해당 문서 빈도수를 계산한다.

[0062] 즉, 각 질의(기술명 또는 기술명+관계명)에 대해 cnn.com을 포함하는 문서 빈도수의 비율(W_{IAT})은 수학식 3을 이용하여 구한다.

수학식 3

$$W_{IAT} = \frac{a * g + b * h}{n * (a + b)}$$

[0063]

[0064]

여기서, a,b는 g와 h의 비중을 고려한 계수로 이들의 값을 조절함으로써 g와 h의 상대적 중요도를 반영한다. w_{IAT} 는 각 질의어에 대해 한번씩 계산되므로, 전체 또는 일부 질의어 집합의 검색 결과로부터 추출한 후 이를 평균하여 얻는다.

[0065]

즉, 각 질의어 검색을 수행하면 각 리소스(cnn, abc, bbc 등)에 대한 w_{IAT} 가 얻어지고, 질의어를 바꿔 검색을 반복하면서 w_{IAT} 의 값을 수집하여 평균함으로써 모든 질의(기술명 또는 기술명+관계명)에 대한 리소스(cnn)의 w_{IAT} 의 평균값, 즉 선호도/인지도(w_{IA})를 얻게 된다.

[0066]

문서 내에서 기술명이 추출된 위치 정보 값(W_J)은 기술 통계정보 추출부에서 추출된 기술명의 위치에 따라 가중치를 부여한 값일 수 있다. 위치는 제목, 부제, 중간 제목, 문단 등으로 구분할 수 있고, W_J 는 기술명이 추출된 위치별로 값이 미리 정의되어 있다. 예를 들면, 제목은 10, 부제는 8,중간제목은 5 등과 같이 위치별로 값이 정해져 있다. 기술 인지도 계산부는 기술명이 추출된 위치가 다를 때, 각 추출된 문서내의 위치별로 서로 다른 가중치를 부여하여 위치 정보 값(W_J)을 구한다.

[0067]

기술명의 문서내 중요도(W_K)는 문서내의 빈도수 기반 중요도를 의미하는 것으로, 문서내에 포함된 전체 기술명 빈도수(TF) 중에 해당 기술명의 출현 빈도수(AF)를 통해 계산된다.

[0068]

즉, 기술인지도 계산부(120)는 수학식 4을 이용하여 기술명의 문서내 중요도를 구한다.

수학식 4

$$W_k = \frac{AF}{TF}$$

[0069]

[0070]

기술명의 문서 전체내 중요도(W_L)는 문서내의 역문헌 빈도수를 기반으로 한 중요도 값으로, 전체 문헌수(N)에서 특정 기술명이 포함된 문서수(d)의 역(Inverse) 비율 값을 통해 계산된다.

[0071]

즉, 기술 인지도 계산부(120)는 수학식 5를 이용하여 기술명의 문서 전체 내 중요도를 구한다.

수학식 5

$$W_L = \log(N/d)$$

[0072]

[0073]

많은 문서에서 출현하는 기술명은 w_L 이 작아지고, 일부 문서에서만 출현하는 w_L 이 커지는 경향이 있다.

[0074]

상기와 같이 기술 인지도 계산부(120)는 기술이 추출된 문서의 신뢰도, 문서내에서 기술이 추출된 위치정보, 기

술의 문서내 중요도, 기술의 문서 전체 내 중요도를 이용하여 각 기술에 대한 인지도를 구한다. 이때, 기술 인지도 계산부(120)는 각 기술에 대한 인지도를 문서별로 구한다. 예를 들어, 특정 기술이 포함된 문서가 10개인 경우, 기술 인지도 계산부(120)는 특정 기술에 대한 인지도를 10개 구하게 된다.

- [0075] 인지도 변화 예측부(130)는 각 기술에 대해, 문서별로 구해진 기술 인지도 값을 수집하고, 그 수집된 기술 인지도값을 연도별로 평균을 구한 후, 그 연도별 평균값을 기반으로 각 기술의 인지도 변화 속도 및 변화 방향을 예측한다. 이때, 기술 통계정보 추출부에서 추출된 통계 정보에는 연도가 포함되어 있으므로, 인지도 변화 예측부(130)는 문서의 연도를 확인하고, 그 연도를 기초로 기술 인지도값을 연도별로 분리할 수 있다.
- [0076] 즉, 인지도 변화 예측부(130)는 특정 기술에 대해 개별 문서별로 구해진 인지도 값을 연도별로 수집하고, 그 수집된 인지도 값들의 연도별 평균을 구한다. 그런 후, 인지도 변화 예측부(130)는 연도별 평균값을 기반으로 예측 연도의 인지도 값을 예측하고, 그 예측된 인지도 값을 이용하여 인지도의 변화 속도를 구한다. 이때, 인지도 변화 예측부(130)는 회귀분석 방법, 이동평균분석 방법, 또는 이 두 방법을 혼합한 방법(ARMA) 등을 이용하여 예측 연도의 인지도 값을 예측할 수 있다. 여기서, 예측 연도는 인지도 값을 예측하고자 하는 연도(year)를 의미한다. 또한, 인지도의 변화 속도는 전년도 인지도 평균값 대비 예측 연도 인지도 예측값의 기울기를 이용하여 구할 수 있다. 여기서, 전년도는 예측 연도의 이전 연도를 의미하고, 예측 연도의 인지도 값은 연도별 평균값을 기반으로 예측한 인지도 값을 의미한다.
- [0077] 예를 들어, 특정 기술에 대해 2001년부터 2011년까지 개별 문서별로 구해진 인지도 값을 연도별로 각각 수집하고, 2001년, 2002년, ..., 2011년 각 연도별로 인지도 값의 평균을 구한다. 그런 후, 회귀분석 방법, 이동평균 분석 방법, 이 두 방법을 혼합한 방법(ARMA) 등을 이용하여 2012년의 인지도 값을 예측하고, 2011년의 인지도 평균값 대비 2012년의 인지도 예측값에 대한 기울기를 구한다. 그 구해진 기술기가 특정 기술에 대한 인지도 변화속도 일 수 있다. 예를 들어, 2011년의 인지도 평균값이 4, 2012년의 인지도 예측값이 6인 경우, 기울기는 $(6-4)/1=2$ 가 되므로, 기술 인지도 변화 속도는 '2'일 수 있다.
- [0078] 상기와 같이 특정 기술에 대한 인지도 변화 속도가 계산되면, 인지도 변화 예측부(130)는 그 계산된 변화 속도가 속하는 단계를 판단하여 변화 방향을 결정한다. 이때, 인지도 변화 속도에 따른 단계 및 변화방향은 미리 설정되어 있고, 인지도 변화 예측부는 변화속도가 해당하는 단계를 판단하여 변화 방향을 결정할 수 있다.
- [0079] 예를 들어, 0~1 범위에 해당하는 변화속도는 1단계, 1~3 범위에 해당하는 변화속도는 2단계, 3이상에 해당하는 변화속도는 3단계로 변화속도에 따른 단계를 나누고, 1단계는 완만상승, 2단계는 상승, 3단계는 급상승으로 변화 방향을 설정할 수 있다. 기술 인지도 변화 속도가 '2'이면, 2단계에 해당하므로, 상승으로 변화방향을 결정할 수 있다.
- [0080] 또한, 0~-1 범위에 해당하는 변화속도는 1단계, -1~-3 범위에 해당하는 변화속도는 2단계, -3이하에 해당하는 변화속도는 3단계로 변화속도에 따른 단계를 나누고, 1단계는 완만하강, 2단계는 하강, 3단계는 급하강으로 변화 방향을 설정할 수 있다.
- [0081] 도 2는 본 발명에 따른 기술 인지도 변화 예측 장치가 기술 인지도의 변화를 예측하는 방법을 나타낸 흐름도이다.
- [0082] 도 2를 참조하면, 기술 인지도 변화 예측 장치는 문서를 분석하여 기술명을 추출하고(S202), 추출된 기술명에 대한 통계 정보를 추출한다(S204). 즉, 기술 인지도 변화 예측 장치는 다중 문헌 제공 장치에 저장된 문서들을 분석하여 기술용어 사전에 정의된 기술명을 포함하는 문서들을 획득하고, 그 문서들로부터 기술명을 추출한다. 그런 다음 기술 인지도 변화 예측 장치는 각 기술명이 추출된 위치, 연도 등을 함께 추출하여, 문서내에서 추출된 기술명, 문서내 기술명의 빈도수, 문서내에 포함된 전체 기술명 빈도수, 기술명이 추출된 문서내의 위치 및 위치별 빈도수, 문서 종류별 빈도수, 문서내 개별 기술의 통계, 기술명이 추출된 문서의 연도 등을 포함하는 통계 정보를 생성하게 된다.
- [0083] 단계 S204의 수행 후, 기술 인지도 변화 예측 장치는 통계 정보를 기반으로 기술의 인지도 값을 구한다(S206). 즉, 기술 인지도 변화 예측 장치는 각 기술의 통계정보를 기반으로 기술이 추출된 문서의 신뢰도 값, 문서내에서 기술이 추출된 위치정보값, 기술의 문서내 중요도 값, 기술의 문서 전체 내 중요도 값을 각각 구하고, 그 구해진 값들을 이용하여 각 기술에 대한 인지도 값을 개별 문서별로 구한다.
- [0084] 단계 S206의 수행 후, 기술 인지도 변화 예측 장치는 각 기술에 대해, 문서별로 구해진 기술 인지도 값들을 수

도면2

