



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2016년03월24일

(11) 등록번호 10-1606030

(24) 등록일자 2016년03월18일

(51) 국제특허분류(Int. Cl.)

G06F 17/30 (2006.01)

(21) 출원번호 10-2014-0170274

(22) 출원일자 2014년12월02일

심사청구일자 2014년12월02일

(56) 선행기술조사문헌

시맨틱 콘텐츠 검색을 위한 질의 확장 시스템” 이
무훈 외1, 디지털정책연구 저널 지식맵 / v.10
no.10, pp.307-312, 2012.11.30*

KR1020090125559 A*

KR101093989 B1

KR1020100025731 A

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

윤은일

서울특별시 성북구 종암로5길 7 (종암동)

양홍모

충청북도 청주시 상당구 수영로 350 (금천동, 장
자마을1단지부영아파트) 104동 302동

(뒷면에 계속)

(74) 대리인

정부연

전체 청구항 수 : 총 14 항

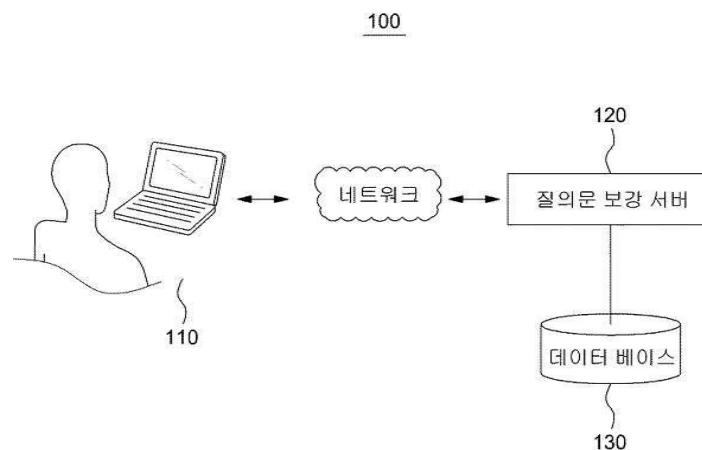
심사관 : 김병균

(54) 발명의 명칭 질의문 보장 방법, 이를 수행하는 질의문 보장 서버 및 이를 저장하는 기록매체

(57) 요약

질의문 보장 방법은 사용자로부터 질의문을 수신하는 단계, 상기 질의문에 포함되어 있는 형태소를 분석하여 키워드 집합을 생성하는 단계, 상기 키워드 집합에 포함되어 있는 적어도 하나의 키워드 각각에 대한 최신 이슈 키워드를 데이터베이스로부터 로드하는 단계 및 상기 최신 이슈 키워드를 상기 질의문에 결합하여 보장 질의문을 생성하는 단계를 포함한다.

대표도 - 도1



(72) 발명자

이강인

충청북도 청주시 흥덕구 모충로12번길 4 (개신동, 송학아파트)103동 905호

김동규

경기도 김포시 전원로 28, 114동 1304호 (장기동, 전원마을월드1단지아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호 NIPA-2014-H0502-14-3008

부처명 미래창조과학부

연구관리전문기관 정보통신산업진흥원

연구사업명 IT/SW 창의연구과정(대학주도형 기술개발)

연구과제명 비정형 빅데이터 분석을 통한 스트림 근접 패턴 마이닝 기반의 실시간 이슈 검색 기술 개

발

기 여 율 1/1

주관기관 세종대학교 산학협력단

연구기간 2014.06.01 ~ 2015.05.31

명세서

청구범위

청구항 1

- (a) 사용자 단말로부터 질의문(query)을 수신하는 단계;
- (b) 상기 질의문에 포함되어 있는 형태소를 분석하여 키워드(keyword) 집합을 생성하는 단계;
- (c) 상기 키워드 집합에 포함되어 있는 적어도 하나의 키워드 각각에 대한 최신 이슈(issue) 정보를 데이터베이스로부터 수신하는 단계; 및
- (d) 상기 최신 이슈 정보를 상기 질의문에 반영하여 보강(expansion) 질의문을 생성하고, 상기 보강 질의문에 포함되는 키워드 각각에 대한 가중치를 산출하는 단계를 포함하는 질의문 보강 방법.

청구항 2

제1항에 있어서, 상기 (d) 단계는

- (d-1) 상기 최신 이슈 정보에 포함되어 있는 최신 이슈 키워드 각각에 대한 빈도율을 산출하는 단계; 및
- (d-2) 상기 빈도율을 기초로 상기 최신 이슈 키워드를 정렬하는 단계를 포함하는 것을 특징으로 하는 질의문 보강 방법.

청구항 3

제2항에 있어서, 상기 빈도율은

상기 키워드 집합에 포함되어 있는 키워드의 개수, 상기 최신 이슈 정보의 개수 및 상기 최신 이슈 정보 중에서 특정 최신 이슈 키워드가 포함되어 있는 최신 이슈 정보의 개수를 기초로 산출되는 것을 특징으로 하는 질의문 보강 방법.

청구항 4

제2항에 있어서, 상기 (d-2) 단계는

상기 최신 이슈 키워드 중에서 상기 빈도율이 기 설정된 특정 기준보다 작은 최신 이슈 키워드를 제외시키는 단계를 포함하는 것을 특징으로 하는 질의문 보강 방법.

청구항 5

제4항에 있어서, 상기 (d) 단계는

상기 정렬된 최신 이슈 키워드를 상기 질의문에 결합하는 단계를 더 포함하는 것을 특징으로 하는 질의문 보강 방법.

청구항 6

제5항에 있어서, 상기 (d) 단계는

상기 질의문에 결합된 최신 이슈 키워드 및 상기 질의문에 있는 키워드 중에서 중복되는 중복 키워드를 삭제하

는 단계를 포함하는 것을 특징으로 하는 질의문 보강 방법.

청구항 7

삭제

청구항 8

제1항에 있어서,

(e) 상기 보강 질의문을 기초로 문서 정보를 수신하고 상기 문서 정보 각각에 대한 랭킹 점수를 산출하는 단계를 더 포함하는 것을 특징으로 하는 질의문 보강 방법.

청구항 9

제8항에 있어서, 상기 (e) 단계는

상기 산출된 랭킹 점수를 기초로 상기 문서 정보를 정렬하는 단계를 포함하는 것을 특징으로 하는 질의문 보강 방법.

청구항 10

제8항에 있어서, 상기 랭킹 점수는

상기 보강 질의문에 포함되어 있는 키워드에 대한 가중치, 특정 문서 정보 내에서 특정 키워드의 빈도수 및 전체 문서 정보에 대한 상기 특정 키워드의 역문헌 빈도수를 기초로 산출되는 것을 특징으로 하는 질의문 보강 방법.

청구항 11

제1항에 있어서, 상기 최신 이슈 정보는

이슈 정보 문서에 포함되어 있는 키워드 중에서 기 설정된 특정 기준 이상 상기 사용자 또는 다른 사용자에게 의하여 질의된 키워드의 집합에 해당하는 것을 특징으로 하는 질의문 보강 방법.

청구항 12

제11항에 있어서, 상기 이슈 정보 문서는

정보 문서 각각의 생성 일자를 기초로 상기 생성 일자가 기 설정된 특정 기준 이후인 정보 문서에 해당하는 것을 특징으로 하는 질의문 보강 방법.

청구항 13

사용자 단말로부터 질의문(query)을 수신하는 질의문 수신부;

상기 질의문에 포함되어 있는 형태소를 분석하여 키워드(keyword) 집합을 생성하는 키워드 집합 생성부;

상기 키워드 집합에 포함되어 있는 적어도 하나의 키워드 각각에 대한 최신 이슈(issue) 정보를 데이터베이스로부터 수신하는 최신 이슈 정보 수신부; 및

상기 최신 이슈 정보를 상기 질의문에 반영하여 보강(expansion) 질의문을 생성하고, 상기 보강 질의문에 포함되는 키워드 각각에 대한 가중치를 산출하는 보강 질의문 생성부를 포함하는 질의문 보강 서버.

청구항 14

제13항에 있어서,

상기 보강 질의문을 기초로 문서 정보를 수신하고 상기 문서 정보 각각에 대한 랭킹 점수를 산출하는 문서 정보 처리부를 더 포함하는 것을 특징으로 하는 질의문 보강 서버.

청구항 15

사용자 단말로부터 질의문(query)을 수신하는 기능;

상기 질의문에 포함되어 있는 형태소를 분석하여 키워드(keyword) 집합을 생성하는 기능;

상기 키워드 집합에 포함되어 있는 적어도 하나의 키워드 각각에 대한 최신 이슈(issue) 정보를 데이터베이스로부터 수신하는 기능; 및

상기 최신 이슈 정보를 상기 질의문에 반영하여 보강(expansion) 질의문을 생성하고, 상기 보강 질의문에 포함되는 키워드 각각에 대한 가중치를 산출하는 기능을 포함하는 질의문 보강 방법에 관한 컴퓨터 프로그램을 기록한 기록매체.

발명의 설명

기술 분야

[0001]

본 발명은 질의문 보강 기술에 관한 것으로서, 보다 구체적으로, 사용자로부터 수신한 질의문을 분석하여 질의문에 포함되어 있는 키워드 각각에 대한 최신 이슈 키워드를 로드하여 보강 질의문을 생성할 수 있는 질의문 보강 방법, 이를 수행하는 질의문 보강 서버 및 이를 저장하는 기록매체에 관한 것이다.

배경 기술

[0002]

일반적으로, 수많은 문서들 중에서 사용자가 검출하고자 하는 특정 문서와 보다 관련성이 높은 문서를 검출하기 위하여 사용자로부터 입력 받은 질의문에 대한 연관 단어 목록이 제공될 수 있다. 즉, 사용자는 자신이 입력한 질의문에 대한 결과를 보고, 검출하고자 하는 특정 문서가 검출되지 않는 경우에는 제공된 연관 단어 목록 중 하나를 선택하여, 자신이 입력한 질의문에 해당 단어를 포함하는 새로운 질의문으로 다시 검출을 실시할 수 있다.

[0003]

한국등록특허 제10-0851042호는 확장 광고 키워드들을 추천하는 방법 및 그 시스템에 관한 것으로서, 광고주로부터 광고하고자 하는 광고 정보를 입력 받고, 광고 정보로부터 키워드를 추출하고 분석 하여, 확장 키워드를 복수의 영역으로 분리하여 광고주에게 추천하는 방법을 제공한다.

[0004]

한국등록특허 제10-1248187호는 확장 검색어 선정 시스템 및 확장 검색어 선정 방법에 관한 것으로서, 사용자로부터 전송 받은 키워드와 연관된 관련 단어들을 화면의 일부분에 출력하여 사용자로부터 특정 단어의 선택을 입력 받고, 사용자로부터 선택된 특정 단어와 연관된 관련 단어들을 화면의 일부분에 출력하여 사용자로부터 또 다른 특정 단어의 선택을 입력 받는 과정을 통하여 매 단계마다 요청되는 키워드에 대응하는 관련 단어들을 단계별로 제공하여 사용자가 복수의 검색어를 선정할 수 있게 한다.

선행기술문헌

특허문헌

[0005]

(특허문헌 0001) 한국등록특허 제10-0851042호

(특허문헌 0002) 한국등록특허 제10-1248187호

발명의 내용

해결하려는 과제

- [0006] 본 발명의 일 실시예는 질의문에 포함되어 있는 키워드 집합과 연관된 최신 이슈 정보를 질의문에 결합하여 보강 질의문을 생성할 수 있는 질의문 보강 방법을 제공하고자 한다.
- [0007] 본 발명의 일 실시예는 수집된 문서들의 생성 시간을 고려한 최근 이슈 정보를 데이터베이스로서 관리할 수 있는 질의문 보강 방법을 제공하고자 한다.
- [0008] 본 발명의 일 실시예는 최신 이슈 정보에 해당하는 키워드일수록 높은 가중치를 부여하고, 이를 기초로 문서를 검색할 수 있는 질의문 보강 방법을 제공하고자 한다.
- [0009] 본 발명의 일 실시예는 질의문을 기초로 검색된 복수의 문서들에 포함되어 있는 각각의 키워드에 대한 이슈 점수를 산출하여 복수의 문서들에 순위를 매길 수 있는 질의문 보강 방법을 제공하고자 한다.

과제의 해결 수단

- [0010] 실시예들 중에서, 질의문 보강 방법은 (a) 사용자로부터 질의문을 수신하는 단계, (b) 상기 질의문에 포함되어 있는 형태소를 분석하여 키워드 집합을 생성하는 단계, (c) 상기 키워드 집합에 포함되어 있는 적어도 하나의 키워드 각각에 대한 최신 이슈 정보를 데이터베이스로부터 수신하는 단계 및 (d) 상기 최신 이슈 정보를 상기 질의문에 반영하여 보강 질의문을 생성하는 단계를 포함한다.
- [0011] 일 실시예에서, 상기 (d) 단계는 (d-1) 상기 최신 이슈 정보에 포함되어 있는 최신 이슈 키워드 각각에 대한 빈도율을 산출하는 단계 및 (d-2) 상기 빈도율을 기초로 상기 최신 이슈 키워드를 정렬하는 단계를 포함할 수 있다.
- [0012] 일 실시예에서, 상기 빈도율은 상기 키워드 집합에 포함되어 있는 키워드의 개수, 상기 최신 이슈 정보의 개수 및 상기 최신 이슈 정보 중에서 특정 최신 이슈 키워드가 포함되어 있는 최신 이슈 정보의 개수를 기초로 산출될 수 있다.
- [0013] 일 실시예에서, 상기 (d-2) 단계는 상기 최신 이슈 키워드 중에서 상기 빈도율이 기 설정된 특정 기준보다 작은 최신 이슈 키워드를 제외시키는 단계를 포함할 수 있다.
- [0014] 일 실시예에서, 상기 (d) 단계는 상기 정렬된 최신 이슈 키워드를 상기 질의문에 결합하는 단계를 더 포함할 수 있다.
- [0015] 일 실시예에서, 상기 (d) 단계는 상기 정렬된 최신 이슈 키워드 중에서 중복되는 중복 키워드를 삭제하는 단계를 포함할 수 있다.
- [0016] 일 실시예에서, 상기 (d) 단계는 상기 보강 질의문에 포함되는 키워드 각각에 대한 가중치를 산출하는 단계를 포함할 수 있다.
- [0017] 일 실시예에서, (e) 상기 보강 질의문을 기초로 문서 정보를 수신하고 상기 문서 정보 각각에 대한 랭킹 점수를 산출하는 단계를 더 포함할 수 있다.
- [0018] 일 실시예에서, 상기 (e) 단계는 상기 산출된 랭킹 점수를 기초로 상기 문서 정보를 정렬하는 단계를 포함할 수 있다.
- [0019] 일 실시예에서, 상기 랭킹 점수는 상기 보강 질의문에 포함되어 있는 키워드에 대한 가중치, 특정 문서 정보 내에서 특정 키워드의 빈도수 및 전체 문서 정보에 대한 상기 특정 키워드의 역문헌 빈도수를 기초로 산출될 수 있다.
- [0020] 일 실시예에서, 상기 최신 이슈 정보는 이슈 정보 문서에 포함되어 있는 키워드 중에서 기 설정된 특정 기준 이상 상기 사용자 또는 다른 사용자에 의하여 질의된 키워드의 집합에 해당할 수 있다.

- [0021] 일 실시예에서, 상기 이슈 정보 문서는 정보 문서 각각의 생성 일자를 기초로 상기 생성 일자가 기 설정된 특정 기준 이후인 정보 문서에 해당할 수 있다.
- [0022] 실시예들 중에서, 질의문 보강 서버는 사용자로부터 질의문을 수신하는 질의문 수신부, 상기 질의문에 포함되어 있는 형태소를 분석하여 키워드 집합을 생성하는 키워드 집합 생성부, 상기 키워드 집합에 포함되어 있는 적어도 하나의 키워드 각각에 대한 최신 이슈 정보를 데이터베이스로부터 수신하는 최신 이슈 정보 수신부 및 상기 최신 이슈 정보를 상기 질의문에 반영하여 보강 질의문을 생성하는 보강 질의문 생성부를 포함한다.
- [0023] 일 실시예에서, 상기 보강 질의문을 기초로 문서 정보를 수신하고 상기 문서 정보 각각에 대한 랭킹 점수를 산출하는 문서 정보 처리부를 더 포함할 수 있다.
- [0024] 실시예들 중에서, 질의문 보강 방법에 관한 컴퓨터 프로그램을 기록한 기록매체는 사용자로부터 질의문을 수신하는 기능, 상기 질의문에 포함되어 있는 형태소를 분석하여 키워드 집합을 생성하는 기능, 상기 키워드 집합에 포함되어 있는 적어도 하나의 키워드 각각에 대한 최신 이슈 정보를 데이터베이스로부터 수신하는 기능 및 상기 최신 이슈 정보를 상기 질의문에 반영하여 보강 질의문을 생성하는 기능을 포함하는 질의문 보강 방법에 관한 컴퓨터 프로그램을 기록한다.

발명의 효과

- [0025] 본 발명의 일 실시예에 따른 질의문 보강 방법은 질의문에 포함되어 있는 키워드 집합과 연관된 최신 이슈 정보를 질의문에 결합하여 보강 질의문을 생성할 수 있다.
- [0026] 본 발명의 일 실시예에 따른 질의문 보강 방법은 수집된 문서들의 생성 시간을 고려한 최근 이슈 정보를 데이터베이스로서 관리할 수 있다.
- [0027] 본 발명의 일 실시예에 따른 질의문 보강 방법은 최신 이슈 정보에 해당하는 키워드일수록 높은 가중치를 부여하고, 이를 기초로 문서를 검출할 수 있다.
- [0028] 본 발명의 일 실시예에 따른 질의문 보강 방법은 질의문을 기초로 검출된 복수의 문서들에 포함되어 있는 각각의 키워드에 대한 이슈 점수를 산출하여 복수의 문서들에 순위를 매길 수 있다.

도면의 간단한 설명

- [0029] 도 1은 본 발명의 일 실시예에 따른 질의문 보강 시스템에 대한 블록도이다.
- 도 2는 도 1의 질의문 보강 서버에 대한 블록도이다.
- 도 3은 최신 이슈 정보에 대한 예시이다.
- 도 4는 도 1의 질의문 보강 시스템에서 수행되는 질의문 보강 방법에 대한 흐름도이다.
- 도 5는 도 4의 질의문 보강 방법에 대한 보다 구체적인 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0030] 본 발명에 관한 설명은 구조적 내지 기능적 설명을 위한 실시예에 불과하므로, 본 발명의 권리범위는 본문에 설명된 실시예에 의하여 제한되는 것으로 해석되어서는 아니 된다. 즉, 실시예는 다양한 변경이 가능하고 여러 가지 형태를 가질 수 있으므로 본 발명의 권리범위는 기술적 사상을 실현할 수 있는 균등물들을 포함하는 것으로 이해되어야 한다. 또한, 본 발명에서 제시된 목적 또는 효과는 특정 실시예가 이를 전부 포함하여야 한다거나 그러한 효과만을 포함하여야 한다는 의미는 아니므로, 본 발명의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.
- [0031] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.
- [0032] "제1", "제2" 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위한 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.

- [0033] 어떤 구성요소가 다른 구성요소에 "연결되어"있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결될 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어"있다고 언급된 때에는 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 한편, 구성요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [0034] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함하다"또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0035] 각 단계들에 있어 식별부호(예를 들어, a, b, c 등)는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 단계들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.
- [0036] 본 발명은 컴퓨터가 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현될 수 있고, 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있으며, 또한, 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한, 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.
- [0037] 여기서 사용되는 모든 용어들은 다르게 정의되지 않는 한, 본 발명이 속하는 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한 이상적이거나 과도하게 형식적인 의미를 지니는 것으로 해석될 수 없다.
- [0038] 도 1은 본 발명의 일 실시예에 따른 질의문 보강 시스템에 대한 블록도이다.
- [0039] 도 1을 참조하면, 질의문 보강 시스템(100)은 사용자 단말(110), 질의문 보강 서버(120) 및 데이터베이스(130)를 포함한다. 여기에서, 사용자 단말(110) 및 질의문 보강 서버(120)는 네트워크를 통하여 연결될 수 있다.
- [0040] 사용자 단말(110)은 질의문 보강 서버(120)와 연결되어 질의문(query) 보강(expansion) 방법을 수행하는데 대한 데이터를 송수신할 수 있는 컴퓨팅 장치에 해당할 수 있다. 예를 들어, 사용자 단말(110)은 데스크톱(Desktop), 노트북(Laptop), 스마트 폰(Smart phone) 또는 태블릿 PC(Tablet Personal Computer)에 해당할 수 있다. 사용자 단말(110)은 질의문 보강 서버(120)에 질의문을 전송할 수 있고, 질의문 보강 서버(120)로부터 문서 정보를 수신할 수 있다.
- [0041] 질의문 보강 서버(120)는 사용자 단말(110)과 연결되어 특정 작업을 실행할 수 있는 컴퓨팅 장치에 해당할 수 있다. 일 실시예에서, 질의문 보강 서버(120)는 사용자 단말(110)로부터 수신한 질의문을 기초로 보강 질의문을 생성할 수 있고, 생성된 보강 질의문을 기초로 문서 정보를 검색하여 사용자 단말(110)에 제공할 수 있다. 일 실시예에서, 질의문 보강 서버(120)는 데이터베이스(130)와 연결되어 데이터베이스(130)로부터 최신 이슈 정보를 수신할 수 있다.
- [0042] 데이터베이스(130)는 최신 이슈 정보를 관리한다. 일 실시예에서, 데이터베이스(130)는 이슈 정보 문서의 생성 시간을 고려하여 특정 기준을 만족하는 이슈 정보 문서만을 최신 이슈 정보로서 관리할 수 있다. 일 실시예에서, 데이터베이스(130)는 질의문 보강 서버(120)에 특정 키워드와 연관된 최신 이슈 정보를 제공할 수 있다.
- [0043] 도 2는 도 1의 질의문 보강 서버에 대한 블록도이다.
- [0044] 도 2를 참조하면, 질의문 보강 서버(120)는 질의문 수신부(210), 키워드 집합 생성부(220), 최신 이슈 정보 수신부(230), 보강 질의문 생성부(240) 및 제어부(250)를 포함한다.

- [0045] 질의문 수신부(210)는 사용자 단말(110)로부터 질의문(query)을 수신한다. 일 실시예에서, 질의문 수신부(210)는 사용자 단말(110)로부터 기 설정된 형식으로 질의문을 수신하거나 또는 다양항 형식의 질의문을 수신할 수 있다. 예를 들어, 질의문 수신부(210)는 사용자 단말(110)로부터 입력된 문장 형태 또는 단어 형태의 질의문을 수신할 수 있다.
- [0046] 키워드 집합 생성부(220)는 질의문 수신부(210)를 통하여 수신한 질의문에 포함되어 있는 형태소를 분석하여 키워드 집합을 생성한다. 여기에서, 형태소는 의미를 가지는 가장 작은 말의 단위로서, 형태소를 분석하는 것은 주어진 언어 문장에서 구조를 파악하고 문장 분할, 분석, 추출 또는 원형 복을 을 거쳐 의미를 가지는 최소 단위인 형태소를 추출해 내는 것에 해당한다. 예를 들어, 사용자 단말(110)로부터 수신한 질의문이 {올해의 프로야구 일정}에 해당하는 경우에 키워드 집합 생성부(220)는 {올해}, {프로}, {야구} 및 {일정}에 해당하는 키워드 집합을 생성할 수 있다.
- [0047] 최신 이슈 정보 수신부(230)는 키워드(keyword) 집합에 포함되어 있는 적어도 하나의 키워드 각각에 대한 최신 이슈(issue) 정보를 데이터베이스(130)로부터 수신한다. 여기에서, 최신 이슈 정보는 이슈 정보 문서에 포함되어 있는 키워드 중에서 기 설정된 특정 기준 이상 사용자 또는 다른 사용자에게 의하여 질의된 키워드의 집합에 해당할 수 있다. 또한, 여기에서, 이슈 정보 문서는 정보 문서 각각의 생성 일자를 기초로 생성 일자가 기 설정된 특정 기준 이후인 정보 문서에 해당할 수 있다.
- [0048] 예를 들어, 도 3을 참조하면, 키워드 집합 생성부(220)는 사용자 단말(110)로부터 수신한 질의문에서 {여름(310)} 및 {여행(320)}에 해당하는 키워드 집합을 생성할 수 있고, 최신 이슈 정보 수신부(230)는 {여름(310)} 및 {여행(320)} 각각에 대한 최신 이슈 정보를 데이터베이스(130)로부터 수신할 수 있다. {여름(310)}에 대한 최신 이슈 정보는 {제1 최신 이슈 정보(311), 제2 최신 이슈 정보(312), 제3 최신 이슈 정보(313)}를 포함하고, 제1 최신 이슈 정보(311)는 {제1 최신 이슈 키워드(311a), 제2 최신 이슈 키워드(311b), 제3 최신 이슈 키워드(311c) 및 제4 최신 이슈 키워드(311d)}를 포함할 수 있다. 즉, 키워드 {여름(310)}에 대한 최신 이슈 정보는 {(날씨, 장마, 비, 더위), (수박, 과일 참외), {장마, 비, 식중독, 음식}}에 해당하고, 키워드 {여행(320)}에 대한 최신 이슈 정보는 {(여름, 비, 날씨), (중국, 홍콩, 날씨)}에 해당한다.
- [0049] 보강 질의문 생성부(240)는 최신 이슈 정보를 질의문에 반영하여 보강 질의문(expansion query)을 생성한다. 일 실시예에서, 보강 질의문 생성부(240)는 최신 이슈 정보에 포함되어 있는 키워드 각각에 대한 빈도율을 산출할 수 있다. 여기에서, 빈도율은 키워드 집합에 포함되어 있는 키워드의 개수, 최신 이슈 정보의 개수 및 최신 이슈 정보 중에서 특정 키워드가 포함되어 있는 최신 이슈 정보의 개수를 기초로 산출될 수 있다. 보다 구체적으로, 보강 질의문 생성부(240)는 아래의 [수식1]을 통하여 특정 최신 이슈 정보 내에서 특정 최신 이슈 키워드의 빈도율을 산출할 수 있고, [수식2]를 통하여 키워드 집합의 전체 키워드에 대한 특정 최신 이슈 키워드의 빈도율을 산출할 수 있다.
- [0050] [수식1]
- [0051]
$$\text{sup}_k(i) = \frac{\sum_{t \in S_k} c(t, i)}{n}, c(t, i) = \begin{cases} 1, i \in t \\ 0, \text{otherwise} \end{cases}$$
- [0052] [수식2]
- [0053]
$$\text{sup}(i) = \frac{\sum_{k=1}^m \text{sup}_k(i)}{m}$$
- [0054] k : 키워드
- [0055] t : 키워드에 대한 최신 이슈 정보
- [0056] i : 최신 이슈 키워드
- [0057] S_k : 키워드 k에 대한 n개의 이슈 정보들(예를 들어, S_k = {t₁, t₂, ..., t_n})
- [0058] sup_k(i) : 키워드 k에 대한 최신 이슈 정보 내에서의 최신 이슈 키워드 i의 빈도율

- [0059] $\text{sup}(i)$: 전체 키워드에 대한 최신 이슈 키워드 i 의 빈도율
- [0060] 일 실시예에서, 보강 질의문 생성부(240)는 빈도율을 기초로 최신 이슈 키워드를 정렬할 수 있고, 예를 들어, 빈도율을 기초로 내림차순으로 정렬할 수 있다. 또한, 보강 질의문 생성부(240)는 빈도율이 기 설정된 특정 기준보다 작은 최신 이슈 키워드를 제외하고 정렬할 수 있다.
- [0061] 일 실시예에서, 보강 질의문 생성부(240)는 정렬된 최신 이슈 키워드를 질의문에 결합할 수 있고, 예를 들어, 정렬된 순서로 순차적으로 질의문에 결합할 수 있다. 일 실시예에서, 보강 질의문 생성부(240)는 최신 이슈 키워드를 질의문에 결합한 결과로서, 질의문에 있는 키워드 및 최신 이슈 키워드 간에 중복되는 중복 키워드가 있는 경우에는 해당 중복 키워드 중 하나를 삭제할 수 있다.
- [0062] 일 실시예에서, 보강 질의문 생성부(240)는 보강 질의문에 포함되는 키워드 각각에 대한 가중치를 산출할 수 있다. 보다 구체적으로, 보강 질의문 생성부(240)는 아래의 [수식3]을 통하여 키워드 각각에 대한 가중치를 산출할 수 있다. 일 실시예에서, 보강 질의문 생성부(240)는 보강 질의문에 포함되어 있는 키워드 중에서, 질의문에 결합된 최신 이슈 키워드보다 사용자 단말(110)로부터 수신한 질의문에 포함되어 있는 키워드에 높은 가중치를 부여할 수 있다.
- [0063] [수식3]
- [0064]
$$\text{weight}(w_b) = \frac{\text{sup}(w_b)}{\log_2 b + 1}$$
- [0065] w : 보강 질의문
- [0066] w_b : 보강 질의문 w 내에서 b 번째 위치한 키워드
- [0067] $\text{sup}(w_b)$: 보강 질의문 w 내에서 b 번째 위치한 키워드의 전체 키워드에 대한 빈도율
- [0068] $\text{weight}(w_b)$: 보강 질의문 w 내에서 b 번째 위치한 키워드의 가중치
- [0069] 일 실시예에서, 질의문 보강 서버(120)는 문서 정보 처리부를 더 포함할 수 있고, 여기에서, 문서 정보 처리부는 보강 질의문을 기초로 문서 정보를 수신하고 문서 정보 각각에 대한 랭킹 점수를 산출할 수 있다. 보다 구체적으로, 랭킹 점수는 보강 질의문에 포함되어 있는 키워드에 대한 가중치, 특정 문서 정보 내에서 특정 키워드의 빈도수 및 전체 문서 정보에 대한 특정 키워드의 역문헌 빈도수를 기초로 산출될 수 있고, 질의문 보강 서버(120)는 아래의 [수식4]를 통하여 문서 정보 각각에 대한 랭킹 점수를 산출할 수 있다.
- [0070] [수식4]
- [0071]
$$\text{rankingscore}(W, d) = \sum_{w_b \in W} \text{weight}(w_b) \times \text{tf}(w_b, d) \times \text{idf}(w_b)$$
- [0072] W : 보강 질의문 내의 모든 키워드 집합
- [0073] b : 키워드의 위치
- [0074] w_b : 보강 질의문 w 내에서 b 번째 위치한 키워드
- [0075] $\text{tf}(w_b, d)$: 문서 정보 d 내에서 w_b 의 빈도수
- [0076] $\text{idf}(w_b)$: 전체 문서 정보에서 w_b 의 역문헌빈도수(inverse document frequency)
- [0077] 일 실시예에서, 문서 정보 처리부는 랭킹 점수를 기초로 문서 정보를 정렬할 수 있고, 예를 들어, 문서 정보 처리부는 랭킹 점수를 기초로 내림차순으로 정렬할 수 있다.
- [0078] 일 실시예에서, 질의문 보강 서버(120)는 문서 정보 제공부를 더 포함할 수 있고, 여기에서, 문서 정보 제공부는 문서 정보 처리부를 통하여 처리된 문서 정보를 사용자 단말(110)에 제공할 수 있다. 일 실시예에서, 문서

정보 제공부는 문서 정보 각각의 랭킹 점수를 기초로 정렬된 순서로 사용자 단말(110)에 제공할 수 있다.

- [0079] 일 실시예에서, 문서 정보 제공부는 최신 이슈 키워드 중에서 상위 빈도율을 가지는 최신 이슈 키워드를 관련 이슈 정보로서 사용자 단말(110)에 제공할 수 있다. 보다 구체적으로, 문서 정보 제공부는 최신 이슈 정보 수신부를 통하여 수신된 최신 이슈 정보에 포함되어 있는 최신 이슈 키워드 중에서 사용자 단말(110)로부터 수신한 질의문에 포함되어 있는 키워드와 중복되는 중복 키워드를 제거한 나머지 키워드들을 빈도율 내림차순으로 정렬하여 사용자 단말(110)에 순차적으로 제공할 수 있다.
- [0080] 제어부(250)는 질의문 보강 서버(120)는 질의문 수신부(210), 키워드 집합 생성부(220), 최신 이슈 정보 수신부(230) 및 보강 질의문 생성부(240)의 동작 및 데이터의 흐름을 제어한다.
- [0081] 도 4는 도 1의 질의문 보강 시스템에서 수행되는 질의문 보강 방법에 대한 흐름도이다.
- [0082] 질의문 수신부(210)는 사용자 단말(110)로부터 질의문을 수신한다(단계 S401). 일 실시예에서, 질의문 수신부(210)는 기 설정된 형태에 적합한 질의문 또는 다양한 형태의 질의문을 수신할 수 있다.
- [0083] 키워드 집합 생성부(220)는 질의문 수신부(210)를 통하여 수신된 질의문의 형태소 분석을 통하여 질의문에 포함되어 있는 키워드의 집합을 생성한다(단계 S402).
- [0084] 최신 이슈 정보 수신부(230)는 키워드 집합 생성부(220)를 통하여 생성된 키워드 집합에 포함되어 있는 키워드 각각에 대한 최신 이슈 정보를 데이터베이스(130)로부터 수신한다(단계 S403). 여기에서, 데이터베이스(130)는 정보 문서 중에서 정보 문서의 생성 날짜 및 생성 시간이 기 설정된 특정 시기 보다 늦게 생성된 정보 문서만을 이슈 정보로서 관리할 수 있고, 이슈 정보에 포함되어 있는 키워드 중에서 기 설정된 특정 기준 이상 사용자 또는 다른 사용자에게 의하여 질의된 키워드만을 최신 이슈 정보로서 관리할 수 있다.
- [0085] 보강 질의문 생성부(240)는 최신 이슈 정보 수신부(230)를 통하여 수신된 최신 이슈 정보에 포함되어 있는 최신 이슈 키워드 각각에 대한 빈도율을 산출하여 특정 기준 이상의 빈도율을 가지는 최신 이슈 키워드를 질의문에 결합하여 보강 질의문을 생성한다(단계 S404). 보다 구체적으로, 보강 질의문 생성부(240)는 최신 이슈 키워드들 간에 중복되는 키워드를 제거하고 질의문에 포함되어 있는 키워드 및 질의문에 결합된 최신 이슈 키워드들 간에 중복되는 중복 키워드를 제거하여 보강 질의문을 생성할 수 있다.
- [0086] 도 5는 도 4의 질의문 보강 방법에 대한 보다 구체적인 흐름도이다.
- [0087] 질의문 수신부(210)는 사용자로부터 사용자 단말(110)을 통하여 질의문을 수신하고(단계 S501), 키워드 집합 생성부(220)는 질의문에 포함되어 있는 키워드에 대한 키워드 집합을 생성한다(단계 S502).
- [0088] 최신 이슈 정보 수신부(230)는 키워드 집합에 포함되어 있는 키워드 각각에 대한 최신 이슈 정보를 수신하고(단계 S503), 최신 이슈 키워드 각각에 대한 빈도율을 산출하여 정렬한다(단계 S504).
- [0089] 보강 질의문 생성부(240)는 최신 이슈 키워드를 질의문에 결합하여 보강 질의문을 생성하고(단계 S505), 보강 질의문 내에서 중복되는 중복 키워드가 있는 경우에는 중복 키워드 중 하나를 삭제하며(단계 S506), 보강 질의문에 포함되어 있는 키워드 각각에 대한 가중치를 산출한다(단계 S507). 일 실시예에서, 가중치는 새로 결합된 최신 이슈 키워드 보다 질의문에 포함되어 있던 키워드에 더 높이 부여될 수 있다.
- [0090] 문서 정보 처리부는 보강 질의문 생성부(240)를 통하여 생성된 보강 질의문을 기초로 문서 정보를 추출하여 수신할 수 있다(단계 S508). 일 실시예에서, 문서 정보 처리부는 수신된 문서 정보 각각에 대한 랭킹 점수를 산출하고 산출된 랭킹 점수를 기초로 문서 정보를 정렬할 수 있다(단계 S509).
- [0091] 문서 정보 제공부는 문서 정보 처리부를 통하여 처리된 문서 정보를 사용자에게 제공할 수 있고(단계 S510), 이에 더하여, 최신 이슈 키워드 중에서 상위 빈도율을 가지는 최신 이슈 키워드를 관련 이슈 정보로서 제공할 수 있다.
- [0092] 상기에서는 본 출원의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 출원을 다양하게 수정

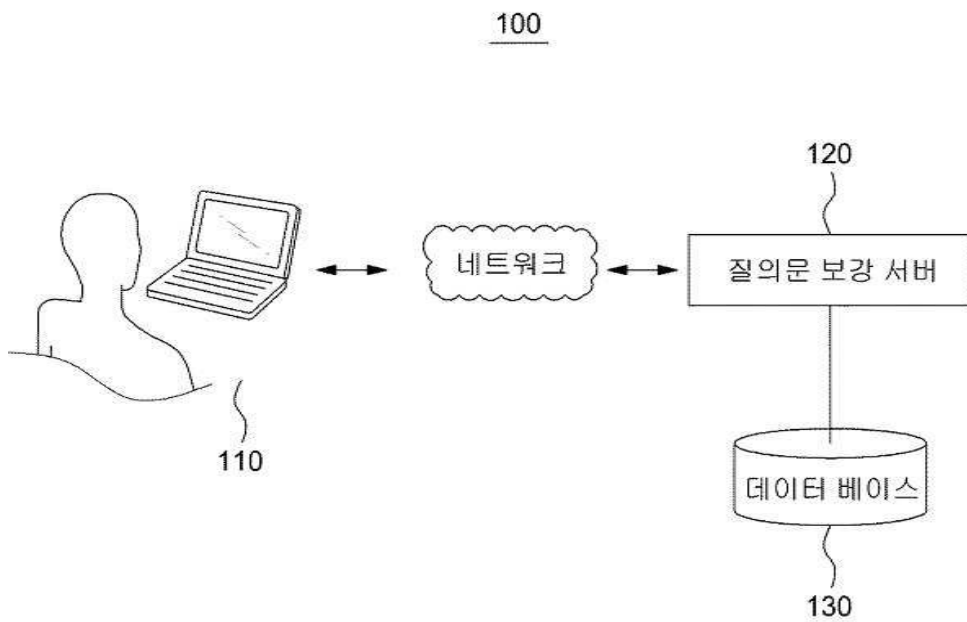
및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

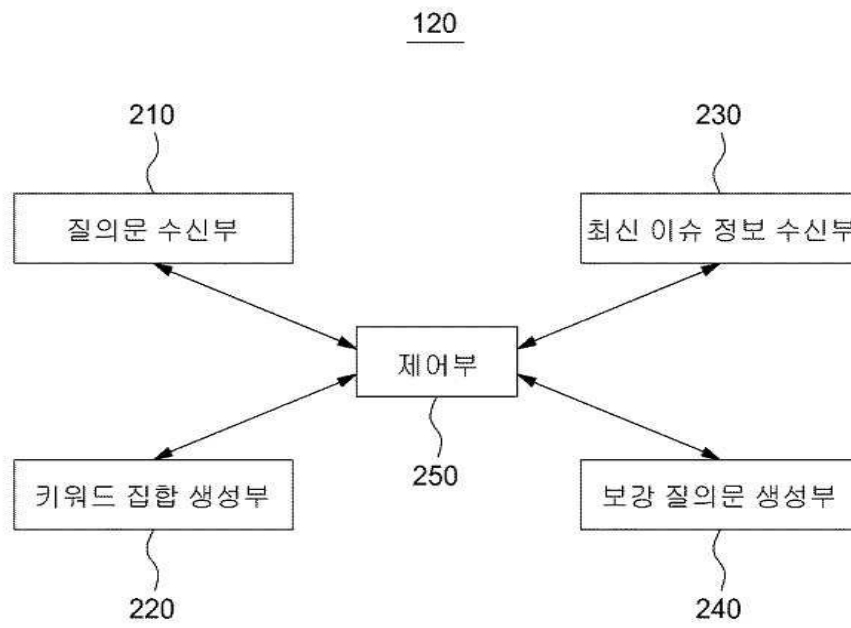
- 100: 질의문 보강 시스템
- 110: 사용자 단말
- 120: 질의문 보강 서버
- 130: 데이터베이스
- 210: 질의문 수신부
- 220: 키워드 집합 생성부
- 230: 최신 이슈 정보 수신부
- 240: 보강 질의문 생성부
- 250: 제어부

도면

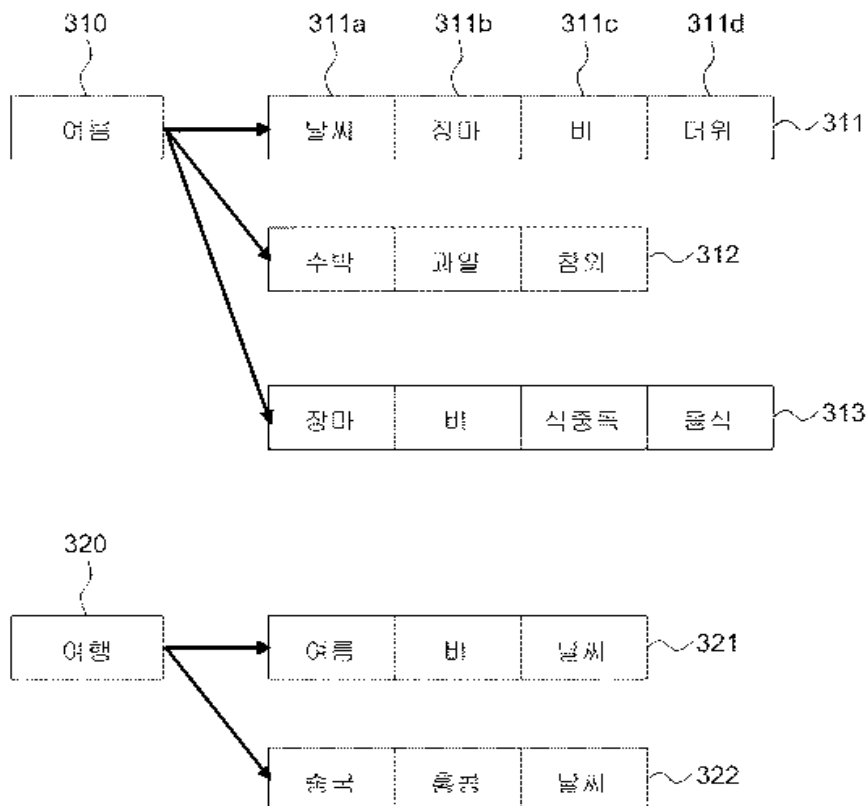
도면1



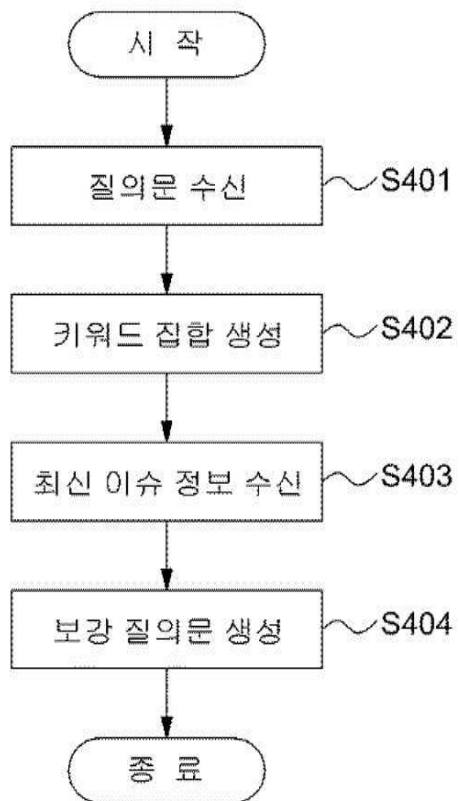
도면2



도면3



도면4



도면5

