

(19)대한민국특허청(KR)  
(12) 등록특허공보(B1)

|  |                                     |  |
|--|-------------------------------------|--|
| (51) Int. Cl. <sup>7</sup><br>G06F 17/30 | (45) 공고일자<br>(11) 등록번호<br>(24) 등록일자 | 2005년11월03일<br>10-0525616<br>2005년10월26일 |
|--|-------------------------------------|--|

|                        |                                |                        |                                |
|------------------------|--------------------------------|------------------------|--------------------------------|
| (21) 출원번호<br>(22) 출원일자 | 10-2004-0026074<br>2004년04월16일 | (65) 공개번호<br>(43) 공개일자 | 10-2005-0100873<br>2005년10월20일 |
|------------------------|--------------------------------|------------------------|--------------------------------|

(73) 특허권자                   엔에이치엔(주)  
  서울특별시 강남구 역삼동 737 스타타워빌딩 34층

(72) 발명자                    최재걸  
  서울특별시관악구봉천6동100-270푸른솔403호

  문상준  
  서울특별시송파구거여1동178-443층

  최병엽  
  서울특별시중랑구목2동목동아이파크101동1602호

  이준호  
  서울특별시송파구송파2동가락삼익맨션208동905호

(74) 대리인                    천성진

심사관 : 이영수

(54) 연관 검색 쿼리 추출 방법 및 시스템

요약

본 발명은 상호 연관된 검색 쿼리(query)를 추출하는 방법 및 시스템에 관한 것으로서, 더욱 상세하게는 각 검색 쿼리가 입력된 검색 세션(session)의 수와 상기 검색 쿼리를 포함하는 검색 쿼리의 쌍이 입력된 검색 세션의 수를 측정하고, 이를 이용하여 검색 쿼리 간 연관 여부를 판단할 수 있는 연관 검색 쿼리 추출 방법 및 시스템에 관한 것이다.

본 발명에 따른 연관 검색 쿼리 추출 방법은 검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지하는 단계(상기 레코드는 소정의 시간 간격 마다 생성되어 상기 데이터베이스에 기록됨), 상기 데이터베이스를 참조하여 상기 시간 간격 당 설정된 총 검색 세션의 수를 카운팅(counting)하여 총 검색 세션 수 정보를 생성하는 단계, 상기 데이터베이스를 참조하여 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하는 단계, 상기 데이터베이스를 참조하여 상기 시간 간격 당 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성하는 단계, 상기 데이터베이스를 참조하여 상기 시간 간격 당 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제3 검색 세션 수 정보를 생성하는 단계, 상기 제1 검색 세션 수 정보 및 상기 제3 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성하는 단계, 상기 총 검색 세션 수 정보, 상기 제1 검색 세션 수 정보, 상기 제2 검색

세션 수 정보, 및 상기 제3 검색 세션 수 정보를 이용하여 상관 관계(correlation) 정보를 생성하는 단계, 및 상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 단계를 포함하는 것을 특징으로 한다.

본 발명에 따르면, 사용자로부터 입력 받은 검색 쿼리에 관한 데이터를 효과적으로 수집, 분석하여 검색 쿼리 간 연관 여부를 자동적으로 판단함으로써 정확한 연관 검색 쿼리를 신속하게 추출하여 보다 질 높은 서비스를 사용자에게 제공할 수 있는 연관 검색 쿼리 추출 방법 및 시스템이 제공된다.

## 대표도

도 2

## 색인어

검색 쿼리, 연관, 조건부 확률 정보, 상관 관계 정보, 연관 지수 정보

## 명세서

### 도면의 간단한 설명

도 1은 본 발명의 일실시예에 따른 연관 검색 쿼리 추출 시스템의 네트워크 연결을 도시한 도면이다.

도 2는 본 발명의 일실시예에 따른 연관 검색 쿼리 추출 방법을 도시한 흐름도이다.

도 3은 본 발명의 일실시예에 있어서, 데이터베이스를 유지하는 과정을 도시한 흐름도이다.

도 4는 본 발명의 일실시예에 있어서, 데이터베이스에 포함된 레코드의 일례를 도시한 도면이다.

도 5는 본 발명의 일실시예에 있어서, 검색 세션 및 검색 쿼리를 숫자로 매핑한 레코드의 일례를 도시한 도면이다.

도 6은 본 발명의 일실시예에 있어서, 검색 세션 수를 카운팅하는 데 사용되는 해쉬 트리 자료구조의 일례를 도시한 도면이다.

도 7은 본 발명의 일실시예에 있어서, 연관 지수 정보를 이용하여 연관 검색 쿼리 목록을 제공하기 위한 과정을 도시한 흐름도이다.

도 8은 본 발명의 일실시예에 있어서, 연관 지수 정보가 기록된 제2 데이터베이스의 일례를 도시한 도면이다.

도 9는 본 발명의 일실시예에 있어서, 누적된 연관 지수 정보를 이용하여 연관 여부를 판단하는 과정을 도시한 흐름도이다.

도 10은 본 발명의 일실시예에 있어서, 누적 연관 지수 정보 및 기록 시점을 기록한 제2 데이터베이스의 일례를 도시한 도면이다.

도 11은 본 발명의 또 다른 실시예에 따른 연관 검색 쿼리 추출 시스템을 도시한 블록도이다.

도 12는 본 발명에 따른 연관 검색 쿼리 추출 방법을 수행하는 데 채용될 수 있는 범용 컴퓨터 장치의 내부 블록도이다.

<도면의 주요 부분에 대한 부호의 설명>

1100 : 연관 검색 쿼리 추출 시스템

1101 : 데이터베이스 1102 : 데이터베이스 관리 수단

1103 : 카운터 수단 1104 : 조건부 확률 정보 생성 수단

1105 : 상관 관계 정보 생성 수단 1106 : 연관 여부 판단 수단

## 발명의 상세한 설명

### 발명의 목적

#### 발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 상호 연관된 검색 쿼리(query)를 추출하는 방법 및 시스템에 관한 것으로서, 더욱 상세하게는 각 검색 쿼리가 입력된 검색 세션(session)의 수와 상기 검색 쿼리를 포함하는 검색 쿼리의 쌍이 입력된 검색 세션의 수를 측정하고, 이를 이용하여 검색 쿼리 간 연관 여부를 판단할 수 있는 연관 검색 쿼리 추출 방법 및 시스템에 관한 것이다.

일반적으로 검색 서비스를 제공하는 검색 서비스 시스템은 사용자로부터 검색쿼리가 입력되면 상기 검색 쿼리에 대응하는 검색 결과(예를 들면, 상기 검색 쿼리를 포함하는 웹 사이트, 상기 검색 쿼리를 포함하는 기사, 상기 검색 쿼리를 포함하는 파일명을 갖는 이미지 등)를 사용자에게 제공한다.

한편, 최근의 검색 서비스 시스템은 사용자가 원하는 정보를 보다 빠르고 정확하게 찾을 수 있도록 하기 위해 사용자로부터 입력 받은 검색 쿼리와 연관 관계가 있는 검색 쿼리를 추출하여 사용자에게 제공하는 연관 검색 쿼리 제공 서비스를 제공하고 있다. 즉, 검색 서비스 시스템은 입력된 검색 쿼리에 따라 매우 다른 검색 결과를 사용자에게 제공하는 것이 보통이다. 예를 들어, 검색자가 "자동차"를 입력하여 얻은 검색 결과와 "승용차"를 입력하여 얻게 되는 검색 결과는 서로 상이하다. 따라서, 검색자는 자신이 원하는 정보를 얻기 위하여 자신이 원하는 정보와 좀 더 관련성이 높은 검색 쿼리를 입력하려고 하지만, 검색자 스스로 이러한 검색 쿼리를 떠올리기가 어려운 경우들이 있다. 따라서, 최근의 검색 서비스 시스템은 사용자가 입력한 검색 쿼리 및 통계 정보를 이용하여 상기 입력된 검색 쿼리와 관련성이 있는 검색 쿼리들을 사용자에게 제공함으로써, 사용자가 다른 검색 쿼리를 이용하여 검색할 수 있도록 하는 서비스를 제공하고 있다.

여기서, 상기 연관 관계가 있는 검색 쿼리라 함은, 예를 들어 사용자로부터 입력 받은 검색 쿼리의 상위 개념 또는 하위 개념에 해당하는 검색 쿼리("일본어"를 입력 받은 경우 "외국어", 또는 역으로 "외국어"를 입력 받은 경우 "일본어"), 사용자로부터 입력 받은 검색 쿼리와 동의어 관계에 있는 검색 쿼리("책방"을 입력 받은 경우 "서점"), 사용자로부터 입력 받은 검색 쿼리와 유의어 관계에 있는 검색 쿼리("꼬리"를 입력 받은 경우 "꼬지"), 사용자로부터 입력 받은 검색 쿼리와 관계어 관계에 있는 검색 쿼리("see"를 입력 받은 경우 "saw, seen, seeing") 등 의미적으로 관련이 있는 검색 쿼리를 의미할 수 있다. 그러나, 상기 연관 관계가 있는 검색 쿼리는 단순히 의미적 관련이 있는 검색 쿼리만을 의미하는 것은 아니며, 예를 들어 사용자로부터 "박찬호"를 입력 받은 경우 그의 직업인 "야구", 그가 속한 리그인 "메이저 리그", 그의 출신 대학인 "한양대", 그가 속해 있는 소속팀인 "텍사스 레인저스", 같은 메이저 리그에 속한 한국인 야구 선수인 "김병현" 등 다양한 관점에서의 관련이 있는 검색 쿼리를 의미할 수 있다.

그런데, 종래 기술에 따른 연관 검색 쿼리 제공 서비스에 의하면, 서비스 운영자가 하나의 검색 쿼리와 연관성이 있는 다른 검색 쿼리를 일일이 분류하여 저장해야 했기 때문에, 이를 위해 시간적, 경제적 손실을 감수해야만 하는 문제점이 있었다.

또한, 보다 적은 시간과 비용으로 연관 검색 쿼리를 추출하기 위해, 용어들 간의 관련성을 동시에 발생할 확률로 정의한 동시 발생 분류 방법, 문서들을 분류한 후에 각 그룹에서만 주로 나타나는 용어들을 관련어로 정의하는 문서 분류 방법, 어학적 지식과 문서에서의 동시 발생 특성을 이용하여 용어들 간의 관계를 파악하는 문법 분류 방법이 등장하였으나, 상기 방법들은 주로 용어들 간의 의미적 관계는 고려하지 않고 통계적인 관계만 고려함으로써, 사용자가 추출된 연관 검색 쿼리 간의 연관 관계를 납득할 수 없는 경우가 발생하는 문제점이 있었다.

이러한 문제를 해결하기 위해, 한국특허등록 제 10-0372078 호는 사용자로부터 입력 받은 질의어 로그를 이용하여 관련어집을 자동으로 생성하고, 생성된 관련어집을 이용하여 의미적으로 관련성이 있는 관련어를 검색할 수 있도록 하는 관련어 검색 방법에 대해 개시하고 있다.

상기 관련어 검색 방법은 사용자가 입력한 질의어들 중 일정한 세션 내에 이루어진 질의어들 중 일부를 관련어로 등록하여 관련어집을 생성하고, 관련어 검색 장치에 질의어를 입력하면, 상기 질의어의 관련어들을 식별하여 정렬하고 이를 사용자에게 제공한다.

그러나, 상기 관련어 검색 방법은 일정한 세션 내에 이루어진 질의어 로그를 이용하여 질의어 간 연관 여부를 판단할 수는 있지만, 아래와 같은 종래 기술의 문제를 그대로 안고 있다.

즉, 상기 관련어 검색 방법을 이용하는 경우에도 1) 같은 세션에서 1회 이상 출현한 용어라면 이를 모두 관련어로 등록함으로 인해 사용자가 추출된 관련어 간의 연관 관계를 납득할 수 없는 경우가 발생한다는 점, 2) 일정 기간 동안 누적된 데이터를 체계적으로 분석하여 관련어 등록을 수행하는 것이 아님으로 인해 우연히 같은 세션에서 출현한 용어 모두가 관련어로 등록된다는 점, 3) 통상 사용자로부터 입력 되는 것은 질의어가 아닌 검색 쿼리임에도 불구하고, 질의어로부터 용어를 추출하는 방법을 채택함으로 인해 검색 쿼리 간 연관 관계를 파악하기 위해 불필요한 시간이 소비된다는 점, 4) 같은 세션에서 1회 이상 출현한 용어라면 이를 모두 관련어로 등록함으로 인해 너무 많은 관련어가 존재할 수 있다는 점 등의 문제점이 여전히 존재한다.

결국, 상기 관련어 검색 방법에 의하더라도, 사용자로부터 입력 받은 검색 쿼리와 연관 관계가 없는 검색 쿼리가 추출될 가능성이 여전히 높고, 이로 인해 질 높은 서비스를 제공 받아 원하는 정보를 보다 빠르고 정확하게 찾기를 원하는 사용자의 욕구를 충족시킬 수 없는 단점이 있다.

이에, 사용자로부터 입력 받은 검색 쿼리에 관한 데이터를 효과적으로 수집하고, 상기 수집된 데이터를 체계적으로 분석하고 이를 이용하여 검색 쿼리 간 연관 여부를 정확하게 판단함으로써 사용자에게 보다 질 높은 서비스를 제공할 수 있는 새로운 기술의 출현이 요청되어 왔다.

### 발명이 이루고자 하는 기술적 과제

본 발명은 상술한 바와 같은 종래 기술의 문제점을 해결하기 위해 안출된 것으로서, 사용자로부터 입력 받은 검색 쿼리에 관한 데이터를 효과적으로 수집, 분석하여 검색 쿼리 간 연관 여부를 자동적으로 판단할 수 있는 시스템을 구축함으로써, 하나의 검색 쿼리와 연관성이 있는 다른 검색 쿼리를 일일이 분류하여 저장함으로써 서비스 운영자에게 야기되는 시간적, 경제적 손실을 줄일 수 있는 연관 검색 쿼리 추출 방법 및 시스템을 제공하는 것을 그 목적으로 한다.

또한, 본 발명에 따른 연관 검색 쿼리 추출 방법 및 시스템은 검색 쿼리 간 연관 지수 정보를 체계적으로 기록하는 데이터 베이스를 유지함으로써, 사용자로부터 검색 쿼리가 입력된 경우 상기 연관 지수 정보를 이용하여 연관 정도가 더 높은 검색 쿼리를 우선적으로 상기 사용자에게 제공하는 것을 그 목적으로 한다.

또한, 본 발명에 따른 연관 검색 쿼리 추출 방법 및 시스템은 사용자들로부터 입력된 검색 데이터 중 체계적인 전처리 과정을 거친 유용한 데이터만을 추출하여 적절한 수의 연관 검색 쿼리를 유지함으로써, 진정으로 의미 있는 연관 검색 쿼리를 추출하여 보다 질 높은 연관 검색 쿼리 서비스를 사용자에게 제공하는 것을 그 목적으로 한다.

또한, 본 발명에 따른 연관 검색 쿼리 추출 방법 및 시스템은 일정 기간 동안 누적된 데이터를 종합하여 검색 쿼리 간 연관 여부를 판단함으로써, 장기간에 걸쳐 일정한 연관 관계를 유지하고 있는 검색 쿼리를 사용자에게 제공하여 보다 정확한 연관 검색 쿼리 서비스를 제공 받기를 원하는 사용자의 욕구를 충족시키는 것을 그 목적으로 한다.

### 발명의 구성 및 작용

상기의 목적을 달성하고, 상술한 종래 기술의 문제점을 해결하기 위하여, 본 발명에 따른 연관 검색 쿼리 추출 방법은 검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지하는 단계(상기 레코드는 소정의 시간 간격 마다 생성되어 상기 데이터베이스에 기록됨), 상기 데이터베이스를 참조하여 상기 시간 간격 당 설정된 총 검색 세션의 수를 카운팅(counting)하여 총 검색 세션 수 정보를 생성하는 단계, 상기 데이터베이스를 참조하여 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하는 단계, 상기 데이터베이스를 참조하여 상기 시간 간격 당 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성하는 단계, 상기 데이터베이스를 참조하여 상기 시간 간격 당 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제3 검색 세션 수 정보를 생성하는 단계, 상기 제1 검색 세션 수 정보 및 상기 제3 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성하는 단계, 상기

총 검색 세션 수 정보, 상기 제1 검색 세션 수 정보, 상기 제2 검색 세션 수 정보, 및 상기 제3 검색 세션 수 정보를 이용하여 상관 관계(correlation) 정보를 생성하는 단계, 및 상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 단계를 포함하는 것을 특징으로 한다.

본 발명의 일측에 따르면, 상기 조건부 확률 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 상기 단계는 상기 조건부 확률 정보가 소정의 수치 이상인 경우에 한하여 상기 연관 여부를 판단하고, 상기 수치는 상기 제1 검색 세션 수 정보를 변수로 하는 소정의 함수에 기초하여 변동되는 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다.

또한, 본 발명의 다른 일측에 따르면, 상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 상기 단계는 상기 조건부 확률 정보 및 상기 상관 관계 정보를 이용하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 지수 정보를 생성하는 단계 및 상기 연관 지수 정보에 기초하여 상기 연관 여부를 판단하는 단계를 포함하고, 상기 연관 여부 판단 결과, 연관된 것으로 판단된 경우 상기 연관 지수 정보를 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리와 연관하여 제2 데이터베이스에 기록하는 단계, 사용자 단말기로부터 제3 검색 쿼리를 수신하는 단계, 상기 제2 데이터베이스를 참조하여 상기 제3 검색 쿼리와 연관된 하나 이상의 제4 검색 쿼리를 추출하는 단계, 상기 연관 지수 정보에 따라 상기 추출된 제4 검색 쿼리를 소팅(sorting)하여 연관 검색 쿼리 목록을 생성하는 단계, 및 상기 생성된 연관 검색 쿼리 목록을 상기 사용자 단말기로 제공하는 단계를 더 포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다.

또한, 본 발명에 따른 연관 검색 쿼리 추출 시스템은 검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스, 상기 레코드를 소정의 시간 간격 마다 생성하여 상기 데이터베이스에 기록하는 데이터베이스 관리 수단, 상기 데이터베이스를 참조하여, 상기 시간 간격 당 설정된 총 검색 세션의 수를 카운팅(counting)하여 총 검색 세션 수 정보를 생성하고, 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하고, 상기 시간 간격 당 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성하며, 상기 시간 간격 당 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제3 검색 세션 수 정보를 생성하는 카운터 수단, 상기 제1 검색 세션 수 정보 및 상기 제3 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성하는 조건부 확률 정보 생성 수단, 상기 총 검색 세션 수 정보, 상기 제1 검색 세션 수 정보, 상기 제2 검색 세션 수 정보, 및 상기 제3 검색 세션 수 정보를 이용하여 상관 관계(correlation) 정보를 생성하는 상관 관계 정보 생성 수단, 및 상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 연관 여부 판단 수단을 포함하는 것을 특징으로 한다.

이하, 첨부된 도면을 참조하여 본 발명에 따른 연관 검색 쿼리 추출 방법 및 시스템을 상세히 설명하기로 한다.

도 1은 본 발명의 일실시예에 따른 연관 검색 쿼리 추출 시스템의 네트워크 연결을 도시한 도면이다. 사용자들은 사용자 단말기(110a 또는 110b)를 이용하여 유무선 통신망을 통해 연관 검색 쿼리 추출 시스템(100)에 접속하여 검색 쿼리를 입력한다. 연관 검색 쿼리 추출 시스템(100)은 상기 검색 쿼리에 대응하는 연관 검색 쿼리를 사용자 단말기(110a 또는 110b)로 전송한다. 또한, 본 발명의 또 다른 실시예에 따른 연관 검색 쿼리 추출 시스템(100)은 상기 연관 검색 쿼리의 연관 지수 정보에 따른 순위 정보 등을 사용자 단말기(110a 또는 110b)에 더 전송할 수 있다. 본 발명의 연관 검색 쿼리 추출 시스템은 인터넷 검색 서비스 시스템에 통합되어 운영될 수 있고, 따라서, 사용자가 인터넷 검색 서비스 시스템에 접속하여 검색 쿼리를 입력한 경우, 상기 검색 쿼리에 대한 검색 결과를 사용자에게 제공할 때에 상기 검색 쿼리의 연관 검색 쿼리도 상기 사용자에게 제공될 수 있다.

도 2는 본 발명의 일실시예에 따른 연관 검색 쿼리 추출 방법을 도시한 흐름도이다. 본 실시예에 따른 연관 검색 쿼리 추출 방법은 소정의 연관 검색 쿼리 추출 시스템에서 제공된다.

단계(201)에서 본 발명에 따른 연관 검색 쿼리 추출 시스템은 검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지한다.

본 발명의 일실시예에 따르면, 상기 검색 세션은 상기 사용자 단말기로 검색창이 최초로 제공될 때 설정되고, 소정의 시간 동안 상기 사용자 단말기로부터 데이터 전송이 없을 때 종료되며, 상기 검색 세션의 종료 후 상기 사용자 단말기로부터 새로운 검색 쿼리를 수신한 경우에는 새로운 검색 세션을 시작하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다.

본 실시예에 있어서, 상기 검색 세션은 종래의 검색 세션과는 다른 방식으로 설정된 새로운 방식의 검색 세션이다. 종래의 검색 세션은 단순히 임의의 사용자가 소정의 시간 동안 수행한 검색 활동을 의미하는 것으로, 검색 세션이 설정되어 종료되기까지의 시간이 사전에 결정되어 있었다. 예를 들어, 상기 시간이 10분이라고 하면, 사용자가 10분 동안 검색 활동을 수행한 순간 종래의 검색 세션은 종료되고, 상기 종료 후 10분 동안의 검색 활동은 새로운 검색 세션에서 수행된 것으로 보는 것이 종래의 검색 세션 설정 방식이다.

이에 반해, 본 실시예에 있어서의 검색 세션은 사용자 단말기로 검색창이 최초로 제공될 때 설정되고, 소정의 시간 동안 상기 사용자 단말기로부터 데이터 전송이 없을 때 종료되는 것을 특징으로 한다. 예를 들어, 상기 시간이 5분이라고 하면, 사용자가 사용자 단말기를 통하여 검색 서비스 웹 페이지에 접속하여 검색창이 최초로 열릴 때 상기 검색 세션이 시작되고, 상기 사용자가 검색 쿼리를 입력하거나 검색 결과를 선택하는 등의 검색 활동을 수행한 "최종 시각"으로부터 5분 동안 검색 활동을 전혀 수행하지 않고 대기하는 경우 상기 검색 세션이 종료된다. 즉, 사용자의 최종 검색 활동 시각으로부터 5분 동안 사용자가 아무 액션을 취하지 않고 대기하는 경우 검색 세션이 종료되도록 하는 것이다. 또한, 상기 검색 세션의 종료 후 사용자 단말기로부터 새로운 검색 쿼리를 수신한 경우에는 새로운 검색 세션을 시작한다.

검색 쿼리를 이용한 검색 결과가 사용자를 만족시키지 못할 경우, 상기 사용자는 상기 검색 쿼리와 관련된 다른 검색 쿼리를 입력하여 검색 활동을 수행하는 것이 일반적이므로, 본 실시예에 따른 연관 검색 쿼리 추출 시스템은 최종 검색 활동 시각으로부터 소정의 시간 동안 사용자가 아무 액션을 취하지 않고 대기한 것을 이미 특정 주제와 관련된 검색 활동을 마친 것으로 판단한다. 따라서, 상기 연관 검색 쿼리 추출 시스템은 상기 시간이 경과한 후 새로운 검색 쿼리가 입력된 경우, 새로운 특정 주제에 관한 검색 쿼리를 입력한 것으로 보고, 새로운 세션을 시작한다.

한편, 아무리 긴 시간이 경과하더라도 사용자로부터 계속적인 검색 활동이 수행되는 경우, 본 실시예에 따른 연관 검색 쿼리 추출 시스템은 상기 사용자가 하나의 의도에 따라 특정 주제와 관련된 다양한 검색 쿼리를 입력하고 있다고 판단한다. 따라서, 상기 연관 검색 쿼리 추출 시스템은 이러한 경우 새로운 검색 세션을 시작하지 않고, 하나의 검색 세션을 유지한다.

검색 세션의 시작에서 종료까지 시간을 고정적으로 정해 놓고 그 검색 세션에서 등장한 검색 쿼리를 모두 연관 검색 쿼리로 등록하는 종래의 방식과 비교할 때, 본 실시예에 의하면 검색 세션을 상기와 같은 새로운 방식으로 정의함으로써, 동일한 검색 세션에서 수신한 검색 쿼리 간 연관성을 매우 높은 정도로 신뢰할 수 있는 효과를 얻을 수 있다.

또한, 본 발명에 따르면, 동일한 검색 세션에서 수신한 검색 쿼리를 모두 연관 검색 쿼리로 등록하는 것이 아니라, 후술하듯이 검색 쿼리의 쌍이 등장한 검색 세션의 수를 카운팅하여 상기 검색 쿼리 간 연관 여부를 판단할 수 있는 하나의 인자로 활용함으로써, 검색 쿼리 간 연관 여부를 보다 정확하게 판단할 수 있는 연관 검색 쿼리 추출 방법 및 시스템이 제공된다.

이하, 상기와 같은 검색 세션의 새로운 정의를 이용하여, 검색 세션 및 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지하는 단계(201)의 일실시예에 대하여 상세히 설명한다. 도 3은 본 실시예에 따라 데이터베이스를 유지하는 과정을 도시한 흐름도이다.

단계(301)에서 본 실시예에 따른 연관 검색 쿼리 추출 시스템은 제1 검색 세션과 연관된 제1 검색 세션 식별자를 생성하여 상기 데이터베이스에 기록한다. 상기 연관 검색 쿼리 추출 시스템은 사용자로부터 검색 활동이 이루어질 때마다 상기 제1 검색 세션 식별자 및 각각의 검색 시간에 관한 시각 정보를 사용자 단말기로 송신하고, 상기 사용자 단말기는 수신된 제1 검색 세션 식별자 및 상기 시각 정보를 쿠키(cookie)의 형태로 상기 사용자 단말기 내 소정의 위치에 저장할 수 있다.

상기 사용자로부터 최종적인 검색 활동이 이루어진 경우, 단계(302)에서 상기 연관 검색 쿼리 추출 시스템은 상기 제1 검색 세션 식별자 및 최종 검색 시각에 관한 제1 시각 정보를 사용자 단말기로 송신하고, 상기 사용자 단말기는 상기 제1 검색 세션 식별자 및 상기 제1 시각 정보를 역시 쿠키(cookie)의 형태로 상기 사용자 단말기 내 소정의 위치에 저장할 수 있다.

단계(303)에서 상기 연관 검색 쿼리 추출 시스템은 상기 사용자 단말기로부터 검색 쿼리를 수신하고, 단계(304)에서 상기 검색 쿼리가 수신된 제2 시각 정보 및 상기 제1 시각 정보를 비교한다.

단계(305)에서 상기 비교 결과 양 시각 정보 간 격차가 소정의 시간을 초과한 것으로 판단된 경우, 상기 연관 검색 쿼리 추출 시스템은 단계(306)에서 제2 검색 세션과 연관된 제2 검색 세션 식별자를 생성하고 단계(307)에서 상기 제2 검색 세션 식별자 및 상기 수신된 검색 쿼리에 관한 레코드를 상기 데이터베이스에 기록한다.

한편, 단계(305)에서 양 시각 정보 간 격차가 소정의 시간 이하인 것으로 판단된 경우, 상기 연관 검색 쿼리 추출 시스템은 단계(308)에서 상기 제1 검색 세션 식별자와 연관하여 상기 수신된 검색 쿼리에 관한 레코드를 상기 데이터베이스에 기록한다.

본 실시예에 따른 연관 검색 쿼리 추출 방법에 의하면, 상기와 같은 체계적인 과정을 통하여 데이터베이스를 유지함으로써 검색 세션을 효율적으로 관리할 수 있고, 이렇게 효율적으로 관리된 검색 세션을 이용할 경우, 동일한 검색 세션에서 수신된 검색 쿼리 간 연관성을 매우 높은 정도로 신뢰할 수 있는 효과를 얻을 수 있다.

도 4는 본 발명의 일실시예에 있어서, 데이터베이스에 포함된 레코드의 일례를 도시한 도면이다. 도 4에 도시한 것과 같이, 상기 레코드는 검색 세션 식별자(401) 및 상기 검색 세션 식별자(401)와 연관된 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 정보를 포함할 수 있다. 도 4에 도시된 도면 부호(402)를 참조하면, "sessionId"이라는 검색 세션 식별자 및 "sessionId"이라는 검색 세션 식별자가 할당된 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리인 "박찬호", "메이저리그", "야구" 등이 레코드에 기록되어 있음을 알 수 있다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 상기 레코드를 소정의 시간 간격 마다 생성하여 상기 데이터베이스에 기록한다. 상기 시간 간격은 서비스 운영자에 의해서 "하루", "이틀", "일주일" 등과 같이 사전에 결정되어 있을 수 있고, 상기 서비스 운영자는 기존의 시간 간격을 다른 시간 간격으로 변경할 수도 있다. 본 실시예에 의하면, 일정한 시간 간격에 따라 데이터를 수합하여 검색 쿼리 간 연관 여부를 판단함에 있어 상기 데이터를 활용할 수 있게 되고, 시간의 경과에 따라 변경될 수 있는 검색 쿼리 간 연관 여부를 지속적으로 체크할 수 있게 된다. 예를 들어, 하루 간격으로 상기 레코드를 생성하는 경우, 이틀 전 레코드에는 "박찬호"와 "메이저리그"라는 검색 쿼리가 동시에 수신된 검색 세션이 다수 기록되어 있어 이에 기초하여 어제의 연관 검색 쿼리 추출 시스템은 "박찬호"와 "메이저리그"를 연관 검색 쿼리로 판단했을 수 있지만, 어제 레코드에는 "박찬호"와 "메이저리그"가 동시에 수신된 검색 세션이 거의 기록되지 않아 오늘의 연관 검색 쿼리 추출 시스템은 이에 기초하여 "박찬호"와 "메이저리그"를 연관 검색 쿼리가 아닌 것으로 판단할 수 있다. 따라서, 가장 최근의 데이터를 바탕으로 추출된 연관 검색 쿼리를 사용자에게 제공할 수 있는 장점이 있다.

본 발명의 일실시예에 따른 연관 검색 쿼리 추출 시스템은 단계(201)에서 상기 검색 세션 또는 상기 수신된 검색 쿼리를 숫자로 매핑(mapping)하고, 상기 매핑된 숫자를 이용하여 상기 레코드를 생성할 수 있다.

도 5는 본 실시예에 있어서, 검색 세션 및 검색 쿼리를 숫자로 매핑한 레코드의 일례를 도시한 도면이다. 도 5에 도시된 도면 부호(501)를 참조하면, 도 4에 도시된 도면 부호(402)와 비교하여 "sessionId"이라는 검색 세션 식별자에 "56"이라는 숫자가 매핑되었고, "박찬호", "메이저리그", "야구"의 검색 쿼리에 "18759", "18760", "18761"이라는 숫자가 매핑되었음을 알 수 있다.

문자열로 이루어진 데이터를 이용하여 상호 연관된 검색 쿼리를 추출하는 것과 비교하여, 본 실시예에 따라 숫자로 매핑된 데이터를 이용하여 본 발명에 따른 각 단계를 수행하는 경우, 상기 데이터를 데이터베이스에 기록할 때 메모리를 더 적게 사용하게 되므로 메모리가 절약될 뿐 아니라, 문자열에 비해 그 처리가 매우 간단하므로 처리속도 향상을 꾀할 수 있는 효과를 얻을 수 있다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 단계(201)에서 특정 검색 세션 동안 소정의 수를 초과하는 검색 쿼리가 수신된 경우, 상기 특정 검색 세션 및 상기 특정 검색 세션 동안 수신된 검색 쿼리에 관한 레코드는 상기 데이터베이스에 포함시키지 않는 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다. 본 실시예에 있어서, 상기 연관 검색 쿼리 추출 시스템은 하나의 검색 세션에서 발생한 검색 쿼리의 수를 카운팅하여 상기 검색 쿼리의 수가 소정의 수를 초과하는 경우, 그 검색 세션에서 수신한 데이터는 일반적인 방식이 아닌 예상치 못한 방식으로 들어온 데이터로 판단할 수 있다. 하나의 검색 세션에서 수신한 검색 쿼리의 수가 너무 많은 경우, 그 검색 쿼리 모두가 연관 검색 쿼리일 가능성은 매우 낮기 때문에, 이러한 데이터는 데이터베이스에 기록하지 않음으로써 보다 정확한 연관 검색 쿼리를 추출해 낼 수 있는 효과를 얻기 위함이다.

단계(202)에서 상기 연관 검색 쿼리 추출 시스템은 상기 데이터베이스를 참조하여 상기 시간 간격 당 설정된 총 검색 세션의 수를 카운팅(counting)하여 총 검색 세션 수 정보를 생성한다. 예를 들어, 상기 시간 간격이 하루인 경우, 상기 연관 검색 쿼리 추출 시스템은 하루 동안 설정된 검색 세션의 총 수를 카운팅할 수 있다.

단계(203)에서 상기 연관 검색 쿼리 추출 시스템은 상기 데이터베이스를 참조하여 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하고, 단계(204)에서 상기 데이터베이스를 참

조하여 상기 시간 간격 당 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성한다. 예를 들어, 상기 시간 간격이 하루인 경우, 상기 연관 검색 쿼리 추출 시스템은 하루 동안 "박찬호"라는 검색 쿼리가 수신된 검색 세션의 수를 카운팅하고, "메이저리그"라는 검색 쿼리가 수신된 검색 세션의 수를 카운팅할 수 있다.

단계(205)에서 상기 연관 검색 쿼리 추출 시스템은 상기 데이터베이스를 참조하여 상기 시간 간격 당 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 모두 수신된 검색 세션의 수를 카운팅하여 제3 검색 세션 수 정보를 생성한다. 예를 들어, 상기 시간 간격이 하루인 경우, 상기 연관 검색 쿼리 추출 시스템은 하루 동안 "박찬호"라는 검색 쿼리 및 "메이저리그"라는 검색 쿼리가 모두 수신된 검색 세션의 수를 카운팅할 수 있다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 단계(205)에서 상기 제1 검색 세션 수 정보 및 상기 제2 검색 세션 수 정보가 소정의 수 이상인 경우에 한하여 상기 제3 검색 세션 수 정보를 생성하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다. 즉, 본 실시예에 있어서, 상기 연관 검색 쿼리 추출 시스템은 상기 제1 검색 세션 수 정보 또는 상기 제2 검색 세션 수 정보가 소정의 수에 미치지 못하는 경우 상기 제3 검색 세션 수 정보를 생성하지 않을 수 있다. 각각의 검색 쿼리가 수신된 검색 세션의 수가 너무 적은 경우, 이러한 검색 쿼리들은 서로 연관 검색 쿼리가 될 가능성이 매우 낮기 때문에, 소정의 수 이상 등장하지 못하는 검색 쿼리들은 연관 검색 쿼리를 판단하기 위한 데이터로 생성하지 않음으로써, 상기 연관 검색 쿼리 추출 시스템의 수행 속도 향상에 크게 기여하는 효과를 얻을 수 있다.

본 발명의 일실시예에 따르면, 단계(205)에서 검색 세션 수를 카운팅함에 있어서 해쉬 트리(Hash-tree) 자료구조를 이용하는 연관 검색 쿼리 추출 방법이 제공된다.

해쉬 트리(Hash-tree) 자료구조라 함은 데이터를 저장하고 찾는 데 사용되는 자료 구조의 한 종류로서, 찾고자 하는 문자열을 특정한 함수(Hash function)로 처리하여 얻은 값을 이용하여 데이터의 위치를 찾는 방법으로 알려져 있다. 해쉬 트리(Hash-tree) 자료구조는 데이터를 찾는 속도에 데이터의 개수가 거의 영향을 주지 않는 특성을 지니고 있어, 이를 이용할 경우 효율적이고 빠르게 데이터의 위치를 찾을 수 있을 뿐만 아니라 시스템 내 메모리를 크게 절약할 수 있다.

도 6은 본 실시예에 있어서, 검색 세션 수를 카운팅하는 데 사용되는 해쉬 트리 자료구조의 일례를 도시한 도면이다. 도 6에서는 "대장균"이라는 검색 쿼리와 "이영애"라는 검색 쿼리로 이루어진 검색 쿼리의 쌍이 수신된 검색 세션이 상기 데이터베이스에 존재할 경우, 상기 검색 세션의 수를 해쉬 트리 자료구조를 사용하여 카운팅하는 일례가 도시되어 있다.

단계(206)에서 상기 연관 검색 쿼리 추출 시스템은 상기 제1 검색 세션 수 정보 및 상기 제3 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성한다.

상기 조건부 확률 정보는 검색 쿼리 간 연관 여부를 평가하는 하나의 인자로서 활용될 수 있다. 예를 들어, "대장균"이라는 검색 쿼리 및 "이영애"라는 검색 쿼리 간 연관 여부는, "대장균"이 등장한 검색 세션의 수 중 "이영애"가 등장한 검색 세션의 수가 얼마나 되는가에 관한 확률 정보를 하나의 인자로 활용하여 판단될 수 있다. 즉, "대장균"이 등장한 검색 세션에 "이영애"가 다수 등장했다면, 양 검색 쿼리 간 연관 여부를 판단함에 있어 강한 영향을 미칠 수 있는 하나의 인자로서 활용될 수 있는 것이다.

다음은 상기 조건부 확률 정보를 생성하는 데 이용될 수 있는 수식의 일례를 나타낸 것이다.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

<수식 1. 조건부 확률 정보>

수식 1에서 보는 것과 같이, "A" 검색 쿼리와 "B" 검색 쿼리 모두가 같은 검색 세션에 등장할 확률을 "B" 검색 쿼리가 검색 세션에 등장할 확률로 나눔으로써, "B" 검색 쿼리가 등장한 세션에 "A" 검색 쿼리도 등장할 확률 정보를 생성할 수 있다.

단계(207)에서 상기 연관 검색 쿼리 추출 시스템은 상기 총 검색 세션 수 정보, 상기 제1 검색 세션 수 정보, 상기 제2 검색 세션 수 정보, 및 상기 제3 검색 세션 수 정보를 이용하여 상관 관계(correlation) 정보를 생성한다.

상기 상관 관계 정보는 검색 쿼리 간 연관 여부를 평가하는 또 하나의 인자로서 활용될 수 있다. 사용자로부터 입력되는 빈도가 매우 큰 검색 쿼리의 경우 실제적인 연관성은 없어도 상기 생성된 조건부 확률 정보의 값이 높을 수 있기 때문에, 상기 상관 관계 정보를 검색 쿼리 간 연관 여부를 평가하는 또 하나의 인자로서 활용하는 것은 보다 정확하게 연관 여부를 판단하는 데 크게 기여할 수 있다.

확률 이론 중에는 독립성 판단이라는 것이 존재하는데, 상기 상관 관계 정보는 상기 독립성 판단에 활용된다. 즉, 상기 상관 관계 정보가 1에 가까운 값을 갖는 경우 이를 양 검색 쿼리가 연관되어 있지 않고 독립적이라고 판단할 수 있는 강한 인자로서 활용할 수 있고, 상기 상관 관계 정보가 1보다 상당히 큰 값을 갖는 경우 이를 양 검색 쿼리가 상호 연관되어 있다고 판단할 수 있는 강한 인자로서 활용할 수 있다.

다음은 상기 상관 관계 정보를 생성하는 데 이용될 수 있는 수식의 일례를 나타낸 것이다.

$$\text{상관관계 계수} = \frac{P(A \cap B)}{P(A)P(B)}$$

<수식 2. 상관 관계 정보>

수식 2에서 보는 것과 같이, "A" 검색 쿼리와 "B" 검색 쿼리가 모두 같은 검색 세션에 등장할 확률을 "A" 검색 쿼리가 검색 세션에 등장할 확률과 "B" 검색 쿼리가 검색 세션에 등장할 확률을 곱한 값으로 나누어줌으로써, 상관 관계 정보를 생성할 수 있다. 또한, 상기 수식을 전개할 경우 상기 상관 관계 정보는 "A" 검색 쿼리 및 "B" 검색 쿼리가 모두 등장한 검색 세션의 수에 총 검색 세션 수를 곱한 값을 "A" 검색 쿼리가 등장한 검색 세션의 수와 "B" 검색 쿼리가 등장한 검색 세션의 수로 나누어준 값으로 산정된다. 따라서, 상기 연관 검색 쿼리 추출 시스템은 상기 총 검색 세션 수 정보, 상기 제1 검색 세션 수 정보, 상기 제2 검색 세션 수 정보, 및 상기 제3 검색 세션 수 정보를 이용하여 상관 관계(correlation) 정보를 생성할 수 있게 된다.

단계(208)에서 상기 연관 검색 쿼리 추출 시스템은 상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단한다. 예를 들어, 양 정보에 소정의 수치를 곱하고 이를 더하여 일정 지수를 산정하고, 상기 지수를 이용하여 연관 여부를 판단할 수 있다. 또한, 양 정보에 소정의 수치를 곱하고 이를 다시 곱하여 일정 지수를 산정하고, 상기 지수를 이용하여 연관 여부를 판단하는 것도 가능하다. 상기 방법 이외에도, 상기 조건부 확률 정보 및 상기 상관 관계 정보를 이용하여 연관 여부를 판단하는 방법에는 다양한 실시예가 존재할 수 있음은 본 발명이 속하는 기술분야의 당업자에게 있어 자명하다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 단계(208)에서 상기 조건부 확률 정보가 소정의 수치 이상인 경우에 한하여 상기 연관 여부를 판단하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다.

본 실시예에 있어서, 상기 연관 검색 쿼리 추출 시스템은 상기 조건부 확률 정보가 소정의 수치에 미달될 정도로 매우 낮은 수치에 해당하는 경우, 이미 검색 쿼리 간 연관성이 매우 낮다고 판단하여 연관 여부를 판단하지 않음으로써, 불필요한 메모리의 소요를 줄이고 상기 시스템의 수행 속도를 향상시키는 효과를 얻을 수 있다.

본 발명의 일실시예에 따르면, 상기 실시예에서의 상기 소정의 수치는 상기 제1 검색 세션 수 정보를 변수로 하는 소정의 함수에 기초하여 변동되는 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다.

상기 제1 검색 세션 수 정보가 매우 낮은 수치에 해당하는 경우, 상기 조건부 확률 정보가 정상적으로 구해지지 않아, 고정되어 있는 소정의 수치를 훨씬 초과할 가능성이 있기 때문이다. 예를 들어, "A" 검색 쿼리 및 "B" 검색 쿼리가 모두 등장한 검색 세션의 수가 "1"이고, "A" 검색 쿼리가 등장한 검색 세션의 수가 "5"인 경우, 조건부 확률 정보가 "1/5"로 계산되어 매우 높은 수치를 기록할 수 있게 된다. 이 경우, 실제 "A" 검색 쿼리 및 "B" 검색 쿼리 간 연관 정도는 낮음에도 불구하고, 양 검색 쿼리 간 연관성이 인정되어 정확하지 않은 연관 검색 쿼리를 사용자에게 제공할 가능성이 있다. 따라서, 본 실시예와 같이, 상기 소정의 수치를 제1 검색 세션 수 정보에 따라서 변동해야 할 필요성이 있고, 이로써 보다 정확한 연관 검색 쿼리를 사용자에게 제공할 수 있게 된다.

본 발명의 일실시예에 따르면, 상기 실시예에서의 상기 소정의 함수는 100 퍼센트(값으로는 "1")를 상기 제1 검색 세션 수 정보의 제곱근으로 나눈 퍼센티지 값을 함수값으로 갖는 함수인 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다. 본 실시예에 따르면, 상기 함수는 "100/√(root)제1 검색 세션 수 정보"로 표현될 수 있다.

또한, 다른 실시예에 따르면, 상기 소정의 함수는 아래와 같은 일반적인 수식으로 표현될 수도 있다.

$$y(\%) = a * x^{-b}$$

<수식 3. 함수>

수식 3에서, y는 상기 소정의 수치이고, x는 상기 제1 검색 세션 수 정보이다. 또한, 수식 3에 있어서, 상수 a와 차수 b는 상기 수치를 구하기 위한 최선의 식을 유도하는 실험 과정을 통해서 구해질 수 있다. 예를 들어, a는 80, b는 1의 값이 사용될 수 있다.

상기와 같은 실시예들에 의하면, 상기 조건부 확률 정보가 소정의 수치 이상인 경우에 한하여 상기 연관 여부를 판단하는 것을 특징으로 하는 본 발명의 일실시예에 따른 연관 검색 쿼리 추출 방법에 있어서, 상기 제1 검색 세션 수 정보가 낮은 수치에 해당할 수록 상기 소정의 수치는 높아지므로 상기 연관 검색 쿼리 추출 시스템은 연관 여부를 판단하지 않는 경우가 증가하게 되고, 이로써 불필요한 메모리의 소요를 줄이고 상기 시스템의 수행 속도를 향상시키는 효과를 얻을 수 있다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 단계(208)에서 상기 상관 관계 정보가 소정의 수치 이상인 경우에 한하여 상기 연관 여부를 판단하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법이 제공된다.

본 실시예에 있어서, 상기 연관 검색 쿼리 추출 시스템은 상기 상관 관계 정보가 소정의 수치에 미달될 정도로 매우 낮은 수치에 해당하는 경우(예를 들어, "1"에 거의 근접한 경우), 이미 검색 쿼리 간 연관성이 매우 낮다고 판단하여 연관 여부를 판단하지 않음으로써, 불필요한 메모리의 소요를 줄이고 상기 시스템의 수행 속도를 향상시키는 효과를 얻을 수 있다.

본 발명의 일실시예에 따른 연관 검색 쿼리 추출 방법에 의하면, 상기 연관 검색 쿼리 추출 시스템은 양 검색 쿼리가 연관된 것으로 판단된 경우 이를 기록하고, 토글 오류 검사를 수행함으로써 보다 정확한 연관 검색 쿼리를 추출할 수 있는데, 이하 본 실시예에 대하여 설명한다.

일반적으로 키보드에 있어서 토글키는 하나의 키로 두 가지 이상의 기능을 할 수 있는 키를 의미하는 용어이다. 이러한 토글키의 대표적인 예로는 "Insert", "한/영", "Caps Lock", "Num Lock", "Scroll Lock" 등이 있다.

본 명세서에서 사용되는 "토글 오류 검사"는 상기 토글키 중 "한/영" 변환키와 연관된 검사에 관한 것이다. 예를 들어, 한글 자판을 통하여 "다음"을 입력하고자 하는 사용자가, "한/영" 변환키의 설정에 따라 영어 자판을 통한 "ekdma"를 입력하는 경우가 있을 수 있다. 상기 예와 같이, 실제적으로 사용자가 "한/영" 변환키의 설정에 따라서 한글로 된 검색 쿼리를 영어 자판으로 입력하거나, 또는 영어로 된 검색 쿼리를 한글 자판으로 입력하게 되는 경우는 빈번히 발생할 수 있다. 이 경우, 사용자는 정확한 검색 쿼리를 다시 입력하게 되는데, 이로 인하여 원래의 검색 쿼리와 토글 오류로 인한 검색 쿼리가 동일한 검색 세션에서 등장하게 되어, 양 검색 쿼리 간 연관성이 인정되는 경우가 발생할 수 있다. 상기 예의 경우, "다음"의 토글 오류인 "ekdma"이 "다음"의 연관 검색 쿼리로 지정될 가능성이 있다.

토글 오류로 인해 실제적으로 아무 연관이 없는 검색 쿼리가 연관 검색 쿼리로 지정되는 경우, 이를 기록하기 위해 불필요한 메모리가 소요되고, 그로 인해 상기 시스템의 수행 속도가 저하되는 문제점이 발생할 수 있고, 또한 부정확한 연관 검색 쿼리가 사용자에게 제공되어 서비스의 신뢰도가 낮아지는 문제점 등도 발생할 수 있다. 그러나, 본 실시예에 따른 연관 검색 쿼리 추출 시스템은 상기와 같은 토글 오류 검사를 수행함으로써 상술한 문제점을 해결할 수 있다.

본 실시예에 따른 연관 검색 쿼리 추출 시스템은 단계(208)에서 연관 여부 판단 결과 연관된 것으로 판단된 경우, 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리를 연관 검색 쿼리로 지정하여 제2 데이터베이스에 기록할 수 있다.

또한, 상기 연관 검색 쿼리 추출 시스템은 상기 제2 데이터베이스를 참조하여, 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리에 대한 토글 오류 검사를 수행할 수 있다. 본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 형태소 분석을 이용하여 상기 토글 오류 검사를 수행할 수 있다.

상기 검사 결과, 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 토글 오류 관계에 있는 경우, 상기 연관 검색 쿼리 추출 시스템은 상기 제2 데이터베이스로부터 상기 연관 검색 쿼리 지정과 연관된 기록을 삭제할 수 있다.

본 발명의 일실시예에 따른 연관 검색 쿼리 추출 방법에 의하면, 상기 연관 검색 쿼리 추출 시스템은 검색 쿼리 간 연관 지수 정보를 생성하고 이에 기초하여 연관 여부를 판단하여 연관 검색 쿼리 목록을 사용자에게 제공할 수 있다.

도 7은 본 실시예에 있어서, 연관 지수 정보를 이용하여 연관 검색 쿼리 목록을 제공하기 위한 과정을 도시한 흐름도이다. 본 실시예에 따른 연관 검색 쿼리 추출 방법에 의하면, 단계(208)은 단계(701) 및 단계(702)를 포함할 수 있다.

단계(701)에서 상기 연관 검색 쿼리 추출 시스템은 상기 조건부 확률 정보 및 상기 상관 관계 정보를 이용하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 지수 정보를 생성한다. 예를 들어, 상기 연관 지수 정보는 양 정보에 소정의 수치를 곱하고 이를 더하거나, 양 정보에 소정의 수치를 곱하고 이를 다시 곱하는 방법으로 생성될 수 있다. 상기 방법 이외에도, 상기 조건부 확률 정보 및 상기 상관 관계 정보를 이용하여 상기 연관 지수 정보를 생성하는 방법에는 다양한 실시예가 존재할 수 있음은 본 발명이 속하는 기술분야의 당업자에게 있어 자명하다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 상기 조건부 확률 정보와 연관된 제1 가중치 정보를 생성하고, 상기 상관 관계 정보와 연관된 제2 가중치 정보를 생성하여, 상기 제1 가중치 정보 및 상기 제2 가중치 정보를 이용하여 상기 연관 지수 정보를 생성할 수 있는 연관 검색 쿼리 추출 방법이 제공된다. 이 경우, 상기 제1 가중치 정보 또는 상기 제2 가중치 정보는 소정의 기준에 따라 변동될 수 있다.

본 실시예에 의하면, 상기 조건부 확률 정보가 상기 연관 지수 정보에 미치는 영향 또는 상기 상관 관계 정보가 상기 연관 지수 정보에 미치는 영향을 상황에 따라 적절히 조정할 수 있게 되어, 보다 정확한 연관 지수 정보를 생성할 수 있는 효과를 얻을 수 있다. 또한, 상기 연관 검색 쿼리 추출 시스템은 상기 가중치 정보를 변동할 수도 있다. 예를 들어, 제1 검색 세션 수 정보 또는 제2 검색 세션 수 정보가 너무 낮은 수치에 해당할 경우, 상기 조건부 확률 정보 또는 상기 상관 관계 정보가 매우 커질 수 있고, 이 경우 상기 연관 검색 쿼리 추출 시스템은 상기 조건부 확률 정보 또는 상기 상관 관계 정보의 영향을 조정함으로써, 보다 정확한 연관 지수 정보를 생성할 수 있게 된다.

단계(702)에서 상기 연관 검색 쿼리 추출 시스템은 상기 연관 지수 정보에 기초하여 상기 연관 여부를 판단한다. 예를 들어, 상기 연관 지수 정보가 소정의 수치 이상인 경우 양 검색 쿼리 간 연관성을 인정할 수 있다.

본 발명의 일실시예에 따른 연관 검색 쿼리 추출 방법에 의하면, 상기 연관 검색 쿼리 추출 시스템은 단계(703) 내지 단계(707)을 더 수행할 수 있다.

단계(703)에서 상기 연관 검색 쿼리 추출 시스템은 상기 연관 여부 판단 결과, 연관된 것으로 판단된 경우 상기 연관 지수 정보를 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리와 연관하여 제2 데이터베이스에 기록한다.

도 8은 본 실시예에 있어서, 연관 지수 정보가 기록된 제2 데이터베이스의 일례를 도시한 도면이다. 도 8에 도시한 것과 같이, 검색 쿼리의 쌍과 연관하여 연관 지수 정보가 상기 제2 데이터베이스에 기록되어 있다.

단계(704)에서 상기 연관 검색 쿼리 추출 시스템은 사용자 단말기로부터 제3 검색 쿼리를 수신하고, 단계(705)에서 상기 제2 데이터베이스를 참조하여 상기 제3 검색 쿼리와 연관된 하나 이상의 제4 검색 쿼리를 추출한다.

단계(706)에서 상기 연관 검색 쿼리 추출 시스템은 상기 연관 지수 정보에 따라 상기 추출된 제4 검색 쿼리를 소팅(sorting)하여 연관 검색 쿼리 목록을 생성한다. 예를 들어, 연관 지수 정보가 소정의 수치 이상인 추출된 검색 쿼리만을 소팅하여 상기 연관 검색 쿼리 목록을 생성하는 방법을 이용할 수 있고, 또는 연관 지수 정보의 오름차순 또는 내림차순으로 추출된 검색 쿼리를 소팅하여 상기 연관 검색 쿼리 목록을 생성하는 방법을 이용할 수 있다. 상기 생성 방법 이외에도, 연관 지수 정보에 따라 추출된 검색 쿼리를 소팅(sorting)하여 연관 검색 쿼리 목록을 생성하는 방법에는 다양한 실시예가 존재할 수 있음은 본 발명이 속하는 기술분야의 당업자에게 있어 자명하다.

단계(707)에서 상기 연관 검색 쿼리 추출 시스템은 상기 생성된 연관 검색 쿼리 목록을 상기 사용자 단말기로 제공한다.

본 실시예에 의하면, 사용자는 연관 정도가 높은 검색 쿼리만을 제공 받을 수 있고, 또한 연관 정도가 높거나 낮은 순서로 검색 쿼리를 제공 받을 수도 있게 되어, 제공된 연관 검색 쿼리를 이용하여 자신이 찾고자 하는 정보를 더 신속하게 찾을 수 있는 장점이 있다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 검색 쿼리 간 연관 지수 정보를 일정한 시간 간격마다 생성하고 이를 소정의 데이터베이스에 누적 기록하며, 상기 누적 기록된 연관 지수 정보를 이용하여 검색 쿼리 간 연관 여부를 판단하는 연관 검색 쿼리 추출 방법이 제공된다.

도 9는 본 실시예에 있어서, 누적된 연관 지수 정보를 이용하여 연관 여부를 판단하는 과정을 도시한 흐름도이다. 본 실시예에 의하면, 단계(208)은 단계(901) 내지 단계(905)를 포함할 수 있다.

단계(901)에서 상기 연관 검색 쿼리 추출 시스템은 상기 조건부 확률 정보 및 상기 상관 관계 정보를 이용하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 지수 정보를 제2 시간 간격마다 생성한다. 이 때, 상기 제2 시간 간격은 검색 세션 및 검색 쿼리에 관한 레코드를 생성하는 상기 소정의 시간 간격과 동일하거나 또는 다를 수 있다.

단계(902)에서 상기 연관 검색 쿼리 추출 시스템은 상기 생성된 연관 지수 정보를 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리와 연관하여 상기 제2 시간 간격마다 제2 데이터베이스에 누적 기록한다.

도 10은 본 실시예에 있어서, 누적 연관 지수 정보 및 그 기록 시점을 기록한 제2 데이터베이스의 일례를 도시한 도면이다. 도 10에 도시한 것과 같이, 상기 제2 데이터베이스에는 검색 쿼리의 쌍, 상기 검색 쿼리의 쌍에 해당하는 누적 연관 지수 정보, 상기 누적 연관 지수 정보가 기록된 각각의 기록 시점이 기록되어 있다. 도 10에서 상기 제2 시간 간격은 "1일"로 되어 있지만, 이는 하나의 예시에 불과하고 다양한 시간 간격마다 상기 연관 지수 정보를 생성 및 기록할 수 있다.

단계(903)에서 상기 연관 검색 쿼리 추출 시스템은 상기 제2 데이터베이스를 참조하여 제1 누적 연관 지수 정보 및 제2 누적 연관 지수 정보를 추출하고, 단계(904)에서 상기 제1 누적 연관 지수 정보 및 상기 제2 누적 연관 지수 정보를 이용하여 제2 연관 지수 정보를 생성한다. 예를 들어, 상기 제2 연관 지수 정보는 양 정보에 소정의 수치를 곱하고 이를 더하거나, 양 정보에 소정의 수치를 곱하고 이를 다시 곱하는 방법으로 생성될 수 있다. 상기 방법 이외에도, 상기 제1 누적 연관 지수 정보 및 상기 제2 누적 연관 지수 정보를 이용하여 상기 제2 연관 지수 정보를 생성하는 방법에는 다양한 실시예가 존재할 수 있음은 본 발명이 속하는 기술분야의 당업자에게 있어 자명하다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 상기 제1 누적 연관 지수 정보와 연관된 제1 가중치 정보를 생성하고, 상기 제2 누적 연관 지수 정보와 연관된 제2 가중치 정보를 생성하여, 상기 제1 가중치 정보 및 상기 제2 가중치 정보를 이용하여 상기 제2 연관 지수 정보를 생성할 수 있는 연관 검색 쿼리 추출 방법이 제공된다. 예를 들어, 상기 제2 시간 간격이 "1일"이라고 할 경우, 상기 제1 누적 연관지수 정보 및 상기 제2 누적 연관 지수 정보는 매일 생겨나는 결과물로서 하루만의 상태를 표현하므로, 상기 매일 생겨나는 결과물을 합성하는 방법을 택하는 것이다. 예를 들어, 7일 전 결과물부터 오늘의 결과물까지를 합성하여 최종 결과물로서의 연관 지수 정보를 만들어 낼 수 있다.

본 실시예에 의하면, 장기간에 걸쳐 일정한 연관 관계를 유지하고 있는 검색 쿼리를 사용자에게 제공하여 보다 정확한 연관 검색 쿼리 서비스를 제공 받기를 원하는 사용자의 욕구를 충족시킬 수 있는 효과를 얻을 수 있다.

본 실시예에 있어서, 상기 제1 가중치 정보는 상기 제1 누적 연관 지수 정보가 상기 제2 데이터베이스에 기록된 시점에 기초하여 생성되고, 상기 제2 가중치 정보는 상기 제2 누적 연관 지수 정보가 상기 제2 데이터베이스에 기록된 시점에 기초하여 생성될 수 있다. 상기 가중치 정보를 생성할 때, 선형적(linear) 방식을 사용할 수도 있고, 비선형적(non-linear) 방식을 사용할 수도 있다.

본 실시예에 의하면 최근에 생성되어 기록된 누적 연관 지수 정보에 더 가중치를 두어 연관 지수 정보를 생성함으로써, 보다 최근의 연관 관계를 반영한 연관 검색 쿼리를 사용자에게 제공할 수 있게 되어 보다 정확한 연관 검색 쿼리 서비스를 제공 받기를 원하는 사용자의 욕구를 충족시킬 수 있는 효과를 얻을 수 있다.

단계(905)에서 상기 연관 검색 쿼리 추출 시스템은 상기 제2 연관 지수 정보에 기초하여 상기 연관 여부를 판단한다. 예를 들어, 상기 제2 연관 지수 정보가 소정의 수치 이상인 경우 양 검색 쿼리 간 연관성을 인정할 수 있다.

본 발명의 일실시예에 따르면, 상기 연관 검색 쿼리 추출 시스템은 검색 쿼리 간 연관 여부를 판단하여 기록하고, 이를 이용하여 연관 검색 쿼리를 사용자에게 제공할 수 있다. 본 실시예에 따른 연관 검색 쿼리 추출 방법은 단계(209) 내지 단계(212)를 더 포함할 수 있다.

단계(209)에서 상기 연관 검색 쿼리 추출 시스템은 단계(208)에서의 연관 여부 판단 결과, 연관된 것으로 판단된 경우 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리를 연관 검색 쿼리로 지정하여 제2 데이터베이스에 기록한다.

단계(210)에서 상기 연관 검색 쿼리 추출 시스템은 사용자 단말기로부터 제3 검색 쿼리를 수신하고, 단계(211)에서 상기 제2 데이터베이스를 참조하여 상기 제3 검색 쿼리와 연관된 하나 이상의 제4 검색 쿼리를 추출한다.

단계(212)에서 상기 연관 검색 쿼리 추출 시스템은 상기 추출된 제4 검색 쿼리를 상기 사용자 단말기로 제공한다.

본 발명에 따른 연관 검색 쿼리 추출 방법에 의하면, 상기와 같은 일련의 과정을 통하여, 진정으로 의미 있는 연관 검색 쿼리를 추출하여 보다 질 높은 연관 검색 쿼리 서비스를 사용자에게 제공할 수 있는 효과를 얻을 수 있다.

또한, 본 발명의 실시예들은 다양한 컴퓨터로 구현되는 동작을 수행하기 위한 프로그램 명령을 포함하는 컴퓨터 판독 가능 매체를 포함한다. 상기 컴퓨터 판독 가능 매체에 기록되는 프로그램은 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체는 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

이하, 본 발명의 또 다른 실시예에 따른 연관 검색 쿼리 추출 시스템에 대하여 설명한다. 도 11은 본 실시예에 따른 연관 검색 쿼리 추출 시스템을 도시한 블록도이다.

본 실시예에 따른 연관 검색 쿼리 추출 시스템(1100)은 데이터베이스(1101), 데이터베이스 관리 수단(1102), 카운터 수단(1103), 조건부 확률 정보 생성 수단(1104), 상관 관계 정보 생성 수단(1105), 및 연관 여부 판단 수단(1106)을 포함한다.

데이터베이스(1101)는 검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함한다.

상기 검색 세션은 상기 사용자 단말기로 검색창이 최초로 제공될 때 설정되고, 소정의 시간 동안 상기 사용자 단말기로부터 데이터 전송이 없을 때 종료될 수 있다. 또한, 상기 검색 세션의 종료 후 사용자 단말기로부터 새로운 검색 쿼리를 수신한 경우에는 새로운 검색 세션을 시작한다.

도 3은 데이터베이스(1101)를 유지하는 과정에 관한 일실시예를 도시한 흐름도이고, 도 4는 데이터베이스(1101)에 포함된 레코드의 일례를 도시한 도면이다. 도 3 및 도 4에 관하여는 본 명세서에서 이미 설명한 바 있으므로, 자세한 설명을 생략한다.

본 발명의 일실시예에 따르면, 연관 검색 쿼리 추출 시스템(1100)은 상기 검색 세션 또는 상기 수신된 검색 쿼리를 숫자로 매핑(mapping)하고, 상기 매핑된 숫자를 이용하여 생성된 레코드를 포함하는 데이터베이스(1101)를 유지할 수 있다. 도 5는 검색 세션 및 검색 쿼리를 숫자로 매핑한 레코드의 일례를 도시한 도면이다. 도 5에 관하여는 본 명세서에서 이미 설명한 바 있으므로, 자세한 설명을 생략한다.

데이터베이스 관리 수단(1102)은 상기 레코드를 소정의 시간 간격 마다 생성하여 데이터베이스(1101)에 기록한다.

카운터 수단(1103)은 데이터베이스(1101)를 참조하여, 상기 시간 간격 당 설정된 총 검색 세션의 수를 카운팅(counting)하여 총 검색 세션 수 정보를 생성하고, 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하고, 상기 시간 간격 당 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성하며, 상기 시간 간격 당 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제3 검색 세션 수 정보를 생성한다.

본 발명의 일실시예에 따르면, 카운터 수단(1103)은 검색 세션 수를 카운팅함에 있어서 해쉬 트리(Hash-tree) 자료구조를 이용할 수 있다.

해쉬 트리(Hash-tree) 자료구조라 함은 데이터를 저장하고 찾는 데 사용되는 자료 구조의 한 종류로서, 찾고자 하는 문자열을 특정한 함수(Hash function)로 처리하여 얻은 값을 이용하여 데이터의 위치를 찾는 방법으로 알려져 있다. 도 6은 본 실시예에 있어서, 검색 세션 수를 카운팅하는 데 사용되는 해쉬 트리 자료구조의 일례를 도시한 도면이다. 도 6에 관하여는 본 명세서에서 이미 설명한 바 있으므로, 자세한 설명을 생략한다.

조건부 확률 정보 생성 수단(1104)은 상기 제1 검색 세션 수 정보 및 상기 제3 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성한다.

상관 관계 정보 생성 수단(1105)은 상기 총 검색 세션 수 정보, 상기 제1 검색 세션 수 정보, 상기 제2 검색 세션 수 정보, 및 상기 제3 검색 세션 수 정보를 이용하여 상관 관계(correlation) 정보를 생성한다.

연관 여부 판단 수단(1106)은 상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단한다.

본 발명의 또 다른 실시예에 따른 연관 검색 쿼리 추출 시스템(1120)은 소정의 기록 수단(1107)을 더 포함할 수 있다. 기록 수단(1107)은 상기 제1 검색 세션 수 정보를 변수로 하는 소정의 함수를 기록할 수 있다.

본 실시예에 있어서, 연관 여부 판단 수단(1106)은 상기 조건부 확률 정보가 소정의 수치 이상인 경우에 한하여 상기 연관 여부를 판단하고, 상기 수치는 기록 수단(1107)에 기록된 상기 함수에 기초하여 변동될 수 있다.

본 실시예는 상기 제1 검색 세션 수 정보가 매우 낮은 수치에 해당하는 경우, 상기 조건부 확률 정보가 정상적으로 구해지지 않아, 고정되어 있는 소정의 수치를 훨씬 초과하는 경우에 대비하기 위한 것이다. 예를 들어, "A" 검색 쿼리 및 "B" 검색 쿼리가 모두 등장한 검색 세션의 수가 "1"이고, "A" 검색 쿼리가 등장한 검색 세션의 수가 "5"인 경우, 조건부 확률 정보가 "1/5"로 계산되어 매우 높은 수치를 기록할 수 있게 된다. 이 경우, 실제 "A" 검색 쿼리 및 "B" 검색 쿼리 간 연관 정도는 낮음에도 불구하고, 양 검색 쿼리 간 연관성이 인정되어 정확하지 않은 연관 검색 쿼리를 사용자에게 제공할 가능성이 있다. 따라서, 본 실시예와 같이, 상기 소정의 수치를 제1 검색 세션 수 정보에 따라서 변동해야 할 필요성이 있고, 이로써 보다 정확한 연관 검색 쿼리를 사용자에게 제공할 수 있게 된다.

본 발명의 또 다른 실시예에 따르면, 검색 쿼리 간 연관 여부를 판단하여 기록하고, 이를 이용하여 연관 검색 쿼리를 사용자에게 제공할 수 있는 연관 검색 쿼리 추출 시스템이 제공된다. 본 실시예에 따른 연관 검색 쿼리 추출 시스템은 상기 실시예에 따른 연관 검색 쿼리 추출 시스템(1100)에 추가하여 도면 부호(1130)에 해당하는 소정의 장치들을 더 포함할 수 있다.

도면 부호(1130)에는 제2 데이터베이스(1108), 제2 데이터베이스 관리 수단(1109), 검색 쿼리 수신 수단(1110), 검색 쿼리 추출 수단(1111), 검색 쿼리 제공 수단(1112)이 도시되어 있다.

제2 데이터베이스(1108)는 연관 검색 쿼리에 관한 레코드를 포함한다. 도 8은 제2 데이터베이스(1108)의 일례를 도시한 도면이다. 도 8에 관하여는 본 명세서에서 이미 설명한 바 있으므로, 자세한 설명을 생략한다.

제2 데이터베이스 관리 수단(1109)은 연관 여부 판단 수단(1106)이 상기 연관 여부 판단 결과, 연관된 것으로 판단한 경우 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리를 연관 검색 쿼리로 지정하여 제2 데이터베이스(1108)에 기록한다.

검색 쿼리 수신 수단(1110)은 사용자 단말기로부터 제3 검색 쿼리를 수신하고, 검색 쿼리 추출 수단(1111)은 제2 데이터베이스(1108)를 참조하여 상기 제3 검색 쿼리와 연관된 하나 이상의 제4 검색 쿼리를 추출한다.

검색 쿼리 제공 수단(1112)은 상기 추출된 제4 검색 쿼리를 상기 사용자 단말기로 제공한다.

본 실시예에 따른 연관 검색 쿼리 추출 시스템에 의하면, 상기와 같은 일련의 과정을 통하여, 진정으로 의미 있는 연관 검색 쿼리를 추출하여 보다 질 높은 연관 검색 쿼리 서비스를 사용자에게 제공할 수 있는 효과를 얻을 수 있다.

도 12는 본 발명에 따른 연관 검색 쿼리 추출 방법을 수행하는 데 채용될 수 있는 범용 컴퓨터 장치의 내부 블록도이다.

컴퓨터 장치(1200)는 램(RAM: Random Access Memory)(1020)과 롬(ROM: Read Only Memory)(1230)을 포함하는 주기억장치와 연결되는 하나 이상의 프로세서(1210)를 포함한다. 프로세서(1210)는 중앙처리장치(CPU)로 불리기도 한다. 본 기술분야에서 널리 알려져 있는 바와 같이, 롬(1230)은 데이터(data)와 명령(instruction)을 단방향성으로 CPU에 전송하는 역할을 하며, 램(1220)은 통상적으로 데이터와 명령을 양방향성으로 전송하는 데 사용된다. 램(1220) 및 롬(1230)은 컴퓨터 판독 가능 매체의 어떠한 적절한 형태를 포함할 수 있다. 대용량 기억장치(Mass Storage)(1240)는 양방향성으로 프로세서(1210)와 연결되어 추가적인 데이터 저장 능력을 제공하며, 상기된 컴퓨터 판독 가능 기록 매체 중 어떠한 것일 수 있다. 대용량 기억장치(1240)는 프로그램, 데이터 등을 저장하는데 사용되며, 통상적으로 주기억장치보다 속도가 느린 하드 디스크와 같은 보조기억장치이다. CD 롬(1260)과 같은 특정 대용량 기억장치가 사용될 수도 있다. 프로세서(1210)는 비디오 모니터, 트랙볼, 마우스, 키보드, 마이크로폰, 터치스크린 형 디스플레이, 카드 판독기, 자기 또는 종이 테이프 판독기, 음성 또는 필기 인식기, 조이스틱, 또는 기타 공지된 컴퓨터 입출력장치와 같은 하나 이상의 입출력 인터페이스(1250)와 연결된다. 마지막으로, 프로세서(1210)는 네트워크 인터페이스(1270)를 통하여 유선 또는 무선 통신 네트워크에 연결될 수 있다. 이러한 네트워크 연결을 통하여 상기된 방법의 절차를 수행할 수 있다. 상기된 장치 및 도구는 컴퓨터 하드웨어 및 소프트웨어 기술 분야의 당업자에게 잘 알려져 있다.

상기된 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있다.

이상과 같이 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상의 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.

그러므로, 본 발명의 범위는 설명된 실시예에 국한되어 정해져서는 안되며, 후술하는 특허 청구의 범위뿐 아니라 이 특허 청구의 범위와 균등한 것들에 의해 정해져야 한다.

### 발명의 효과

본 발명에 따른 연관 검색 쿼리 추출 방법 및 시스템에 의하면, 사용자로부터 입력 받은 검색 쿼리에 관한 데이터를 효과적으로 수집, 분석하여 검색 쿼리 간 연관 여부를 자동적으로 판단할 수 있는 시스템을 구축함으로써, 하나의 검색 쿼리와 연관성이 있는 다른 검색 쿼리를 일일이 분류하여 저장함으로써 서비스 운영자에게 야기되는 시간적, 경제적 손실을 줄일 수 있는 효과를 얻을 수 있다.

또한, 본 발명에 따른 연관 검색 쿼리 추출 방법 및 시스템에 의하면, 검색 쿼리 간 연관 지수 정보를 체계적으로 기록하는 데이터베이스를 유지함으로써, 사용자로부터 검색 쿼리가 입력된 경우 상기 연관 지수 정보를 이용하여 연관 정도가 더 높은 검색 쿼리를 우선적으로 상기 사용자에게 제공할 수 있는 효과를 얻을 수 있다.

또한, 본 발명에 따른 연관 검색 쿼리 추출 방법 및 시스템에 의하면, 사용자들로부터 입력된 검색 데이터 중 체계적인 전처리 과정을 거친 유용한 데이터만을 추출하여 적절한 수의 연관 검색 쿼리를 유지함으로써, 진정으로 의미 있는 연관 검색 쿼리를 추출하여 보다 질 높은 연관 검색 쿼리 서비스를 사용자에게 제공할 수 있는 효과를 얻을 수 있다.

또한, 본 발명에 따른 연관 검색 쿼리 추출 방법 및 시스템에 의하면, 일정 기간 동안 누적된 데이터를 종합하여 검색 쿼리 간 연관 여부를 판단함으로써, 장기간에 걸쳐 일정한 연관 관계를 유지하고 있는 검색 쿼리를 사용자에게 제공하여 보다 정확한 연관 검색 쿼리 서비스를 제공 받기를 원하는 사용자의 욕구를 충족시킬 수 있는 효과를 얻을 수 있다.

### (57) 청구의 범위

#### 청구항 1.

삭제

#### 청구항 2.

삭제

#### 청구항 3.

상호 연관된 검색 쿼리(query)를 추출하는 방법에 있어서,

검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지하는 단계 - 상기 레코드는 소정의 시간 간격 마다 생성되어 상기 데이터베이스에 기록됨 -;

상기 데이터베이스를 참조하여 상기 시간 간격 당 설정된 총 검색 세션의 수를 카운팅(counting)하여 총 검색 세션 수 정보를 생성하는 단계;

상기 데이터베이스를 참조하여 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하는 단계;

상기 데이터베이스를 참조하여 상기 시간 간격 당 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성하는 단계;

상기 데이터베이스를 참조하여 상기 시간 간격 당 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제3 검색 세션 수 정보를 생성하는 단계;

상기 제1 검색 세션 수 정보 및 상기 제3 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성하는 단계;

상기 총 검색 세션 수 정보, 상기 제1 검색 세션 수 정보, 상기 제2 검색 세션 수 정보, 및 상기 제3 검색 세션 수 정보를 이용하여 상관 관계(correlation) 정보를 생성하는 단계; 및

상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 단계

를 포함하고,

상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 상기 단계는, 상기 조건부 확률 정보가 소정의 수치 이상인 경우 상기 연관 여부를 판단하고,

상기 수치는 상기 제1 검색 세션 수 정보의 상승에 따라 감소하는 소정의 함수에 기초하여 변동되며,

상기 검색 세션은 상기 사용자 단말기로 검색창이 제공될 때 설정되고, 소정의 시간 동안 상기 사용자 단말기로부터 데이터 전송이 없을 때 종료되며, 상기 검색 세션의 종료 후 상기 사용자 단말기로부터 새로운 검색 쿼리를 수신한 경우 새로운 검색 세션을 시작하는 것

을 특징으로 하는 연관 검색 쿼리 추출 방법.

#### 청구항 4.

제3항에 있어서,

검색 세션 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지하는 상기 단계는,

제1 검색 세션과 연관된 제1 검색 세션 식별자를 생성하여 상기 데이터베이스에 기록하는 단계;

상기 제1 검색 세션 식별자 및 최종 검색 시각에 관한 제1 시각 정보를 사용자 단말기로 송신하는 단계;

상기 사용자 단말기로부터 검색 쿼리를 수신하는 단계;

상기 검색 쿼리가 수신된 제2 시각 정보 및 상기 제1 시각 정보를 비교하는 단계; 및

상기 비교 결과, 상기 제1 시각 정보 및 상기 제2 시각 정보 간 격차가 소정의 시간을 초과하는 경우 제2 검색 세션과 연관된 제2 검색 세션 식별자를 생성하여 상기 제2 검색 세션 식별자 및 상기 수신된 검색 쿼리에 관한 레코드를 상기 데이터베이스에 기록하고, 상기 격차가 소정의 시간 이하인 경우 상기 제1 검색 세션 식별자와 연관하여 상기 수신된 검색 쿼리에 관한 레코드를 상기 데이터베이스에 기록하는 단계

를 포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

### 청구항 5.

제3항에 있어서,

상기 검색 세션 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지하는 상기 단계는,

상기 검색 세션 또는 상기 수신된 검색 쿼리를 숫자로 매핑(mapping)하는 단계; 및

상기 매핑된 숫자를 이용하여 상기 레코드를 생성하는 단계

를 포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

### 청구항 6.

제3항에 있어서,

상기 검색 세션 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지하는 상기 단계는,

특정 검색 세션 동안 소정의 수를 초과하는 검색 쿼리가 수신된 경우, 상기 특정 검색 세션 및 상기 특정 검색 세션 동안 수신된 검색 쿼리에 관한 레코드는 상기 데이터베이스에 포함시키지 않는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

### 청구항 7.

제3항에 있어서,

상기 데이터베이스를 참조하여 상기 시간 간격 당 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제3 검색 세션 수 정보를 생성하는 상기 단계는,

상기 제1 검색 세션 수 정보 및 상기 제2 검색 세션 수 정보가 소정의 수 이상인 경우에 한하여 상기 제3 검색 세션 수 정보를 생성하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

### 청구항 8.

삭제

### 청구항 9.

삭제

### 청구항 10.

제3항에 있어서,

상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 상기 단계는,

상기 상관 관계 정보가 소정의 수치 이상인 경우에 한하여 상기 연관 여부를 판단하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

### 청구항 11.

제3항에 있어서,

상기 연관 여부 판단 결과, 연관된 것으로 판단된 경우 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리를 연관 검색 쿼리로 지정하여 제2 데이터베이스에 기록하는 단계;

상기 제2 데이터베이스를 참조하여, 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리에 대한 토글 오류 검사를 수행하는 단계; 및

상기 검사 결과, 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 토글 오류 관계에 있는 경우 상기 제2 데이터베이스로부터 상기 연관 검색 쿼리 지정과 연관된 기록을 삭제하는 단계

를 더 포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

### 청구항 12.

제3항에 있어서,

상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 상기 단계는,

상기 조건부 확률 정보 및 상기 상관 관계 정보를 이용하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 지수 정보를 생성하는 단계; 및

상기 연관 지수 정보에 기초하여 상기 연관 여부를 판단하는 단계

를 포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

### 청구항 13.

제12항에 있어서,

상기 연관 여부 판단 결과, 연관된 것으로 판단된 경우 상기 연관 지수 정보를 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리와 연관하여 제2 데이터베이스에 기록하는 단계;

사용자 단말기로부터 제3 검색 쿼리를 수신하는 단계;

상기 제2 데이터베이스를 참조하여 상기 제3 검색 쿼리와 연관된 하나 이상의 제4 검색 쿼리를 추출하는 단계;

상기 연관 지수 정보에 따라 상기 추출된 제4 검색 쿼리를 소팅(sorting)하여 연관 검색 쿼리 목록을 생성하는 단계; 및

상기 생성된 연관 검색 쿼리 목록을 상기 사용자 단말기로 제공하는 단계를 더 포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

#### 청구항 14.

제12항에 있어서,

상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 지수 정보를 생성하는 상기 단계는,

상기 조건부 확률 정보와 연관된 제1 가중치 정보를 생성하는 단계;

상기 상관 관계 정보와 연관된 제2 가중치 정보를 생성하는 단계; 및

상기 제1 가중치 정보 및 상기 제2 가중치 정보를 이용하여 상기 연관 지수 정보를 생성하는 단계를

포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

#### 청구항 15.

제14항에 있어서,

상기 제1 가중치 정보 또는 상기 제2 가중치 정보는 소정의 기준에 따라 변동되는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

#### 청구항 16.

제3항에 있어서,

상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 상기 단계는,

상기 조건부 확률 정보 및 상기 상관 관계 정보를 이용하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 지수 정보를 제2 시간 간격 마다 생성하는 단계;

상기 생성된 연관 지수 정보를 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리와 연관하여 상기 제2 시간 간격 마다 제2 데이터베이스에 누적 기록하는 단계;

상기 제2 데이터베이스를 참조하여 제1 누적 연관 지수 정보 및 제2 누적 연관 지수 정보를 추출하는 단계;

상기 제1 누적 연관 지수 정보 및 상기 제2 누적 연관 지수 정보를 이용하여 제2 연관 지수 정보를 생성하는 단계; 및

상기 제2 연관 지수 정보에 기초하여 상기 연관 여부를 판단하는 단계를

포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

#### 청구항 17.

제16항에 있어서,

상기 제1 누적 연관 지수 정보 및 상기 제2 누적 연관 지수 정보를 이용하여 제2 연관 지수 정보를 생성하는 상기 단계는,

상기 제1 누적 연관 지수 정보와 연관된 제1 가중치 정보를 생성하는 단계;

상기 제2 누적 연관 지수 정보와 연관된 제2 가중치 정보를 생성하는 단계; 및

상기 제1 가중치 정보 및 상기 제2 가중치 정보를 이용하여 상기 제2 연관 지수 정보를 생성하는 단계

를 포함하고,

상기 제1 가중치 정보는 상기 제1 누적 연관 지수 정보가 상기 제2 데이터베이스에 기록된 시점에 기초하여 생성되고, 상기 제2 가중치 정보는 상기 제2 누적 연관 지수 정보가 상기 제2 데이터베이스에 기록된 시점에 기초하여 생성되는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

## 청구항 18.

상호 연관된 검색 쿼리(query)를 추출하는 방법에 있어서,

검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스를 유지하는 단계 - 상기 레코드는 소정의 시간 간격 마다 생성되어 상기 데이터베이스에 기록됨 -;

상기 데이터베이스를 참조하여 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하는 단계;

상기 데이터베이스를 참조하여 상기 시간 간격 당 상기 제1 검색 쿼리 및 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성하는 단계;

상기 제1 검색 세션 수 정보 및 상기 제2 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성하는 단계; 및

상기 조건부 확률 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 단계

를 포함하고,

상기 조건부 확률 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 상기 단계는, 상기 조건부 확률 정보가 소정의 수치 이상인 경우 상기 연관 여부를 판단하고,

상기 수치는 상기 제1 검색 세션 수 정보의 상승에 따라 감소하는 소정의 함수에 기초하여 변동되며,

상기 검색 세션은 상기 사용자 단말기로 검색창이 제공될 때 설정되고, 소정의 시간 동안 상기 사용자 단말기로부터 데이터 전송이 없을 때 종료되며, 상기 검색 세션의 종료 후 상기 사용자 단말기로부터 새로운 검색 쿼리를 수신한 경우 새로운 검색 세션을 시작하는 것

을 특징으로 하는 연관 검색 쿼리 추출 방법.

## 청구항 19.

삭제

## 청구항 20.

삭제

## 청구항 21.

제3항 또는 제18항에 있어서,

상기 연관 여부의 판단 결과, 연관된 것으로 판단된 경우 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리를 연관 검색 쿼리로 지정하여 제2 데이터베이스에 기록하는 단계;

사용자 단말기로부터 제3 검색 쿼리를 수신하는 단계;

상기 제2 데이터베이스를 참조하여 상기 제3 검색 쿼리와 연관된 하나 이상의 제4 검색 쿼리를 추출하는 단계; 및

상기 추출된 제4 검색 쿼리를 상기 사용자 단말기로 제공하는 단계

를 더 포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 방법.

## 청구항 22.

제3항 내지 제7항, 제10항 내지 제18항 중 어느 한 항의 방법을 실행시키기 위한 프로그램을 기록한 컴퓨터 판독 가능한 기록 매체.

## 청구항 23.

상호 연관된 검색 쿼리(query)를 추출하는 시스템에 있어서,

검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스;

상기 레코드를 소정의 시간 간격마다 생성하여 상기 데이터베이스에 기록하는 데이터베이스 관리 수단;

상기 데이터베이스를 참조하여, 상기 시간 간격 당 설정된 총 검색 세션의 수를 카운팅(counting)하여 총 검색 세션 수 정보를 생성하고, 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하고, 상기 시간 간격 당 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성하며, 상기 시간 간격 당 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제3 검색 세션 수 정보를 생성하는 카운터 수단;

상기 제1 검색 세션 수 정보 및 상기 제3 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성하는 조건부 확률 정보 생성 수단;

상기 총 검색 세션 수 정보, 상기 제1 검색 세션 수 정보, 상기 제2 검색 세션 수 정보, 및 상기 제3 검색 세션 수 정보를 이용하여 상관 관계(correlation) 정보를 생성하는 상관 관계 정보 생성 수단; 및

상기 조건부 확률 정보 및 상기 상관 관계 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 연관 여부 판단 수단

을 포함하고,

상기 연관 여부 판단 수단은 상기 조건부 확률 정보가 소정의 수치 이상인 경우 상기 연관 여부를 판단하고,

상기 수치는 상기 제1 검색 세션 수 정보의 상승에 따라 감소하는 소정의 함수에 기초하여 변동되며,

상기 검색 세션은 상기 사용자 단말기로 검색창이 제공될 때 설정되고, 소정의 시간 동안 상기 사용자 단말기로부터 데이터 전송이 없을 때 종료되며, 상기 검색 세션의 종료 후 상기 사용자 단말기로부터 새로운 검색 쿼리를 수신한 경우 새로운 검색 세션을 시작하는 것

을 특징으로 하는 연관 검색 쿼리 추출 시스템.

#### 청구항 24.

상호 연관된 검색 쿼리(query)를 추출하는 시스템에 있어서,

검색 세션(session) 및 상기 검색 세션 동안 사용자 단말기로부터 수신된 검색 쿼리에 관한 레코드를 포함하는 데이터베이스;

상기 레코드를 소정의 시간 간격마다 생성하여 상기 데이터베이스에 기록하는 데이터베이스 관리 수단;

상기 데이터베이스를 참조하여, 상기 시간 간격 당 제1 검색 쿼리가 수신된 검색 세션의 수를 카운팅(counting)하여 제1 검색 세션 수 정보를 생성하고, 상기 시간 간격 당 상기 제1 검색 쿼리 및 제2 검색 쿼리가 수신된 검색 세션의 수를 카운팅하여 제2 검색 세션 수 정보를 생성하는 카운터 수단;

상기 제1 검색 세션 수 정보 및 상기 제2 검색 세션 수 정보를 이용하여 조건부 확률(conditional probability) 정보를 생성하는 조건부 확률 정보 생성 수단; 및

상기 조건부 확률 정보에 기초하여 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리 간 연관 여부를 판단하는 연관 여부 판단 수단

을 포함하고,

상기 연관 여부 판단 수단은 상기 조건부 확률 정보가 소정의 수치 이상인 경우 상기 연관 여부를 판단하고,

상기 수치는 상기 제1 검색 세션 수 정보의 상승에 따라 감소하는 소정의 함수에 기초하여 변동되며,

상기 검색 세션은 상기 사용자 단말기로 검색창이 제공될 때 설정되고, 소정의 시간 동안 상기 사용자 단말기로부터 데이터 전송이 없을 때 종료되며, 상기 검색 세션의 종료 후 상기 사용자 단말기로부터 새로운 검색 쿼리를 수신한 경우 새로운 검색 세션을 시작하는 것

을 특징으로 하는 연관 검색 쿼리 추출 시스템.

#### 청구항 25.

삭제

#### 청구항 26.

삭제

#### 청구항 27.

제23항 또는 제24항에 있어서,

연관 검색 쿼리에 관한 레코드를 포함하는 제2 데이터베이스;

상기 연관 여부 판단 결과, 연관된 것으로 판단된 경우 상기 제1 검색 쿼리 및 상기 제2 검색 쿼리를 연관 검색 쿼리로 지정하여 상기 제2 데이터베이스에 기록하는 제2 데이터베이스 관리 수단;

사용자 단말기로부터 제3 검색 쿼리를 수신하는 검색 쿼리 수신 수단;

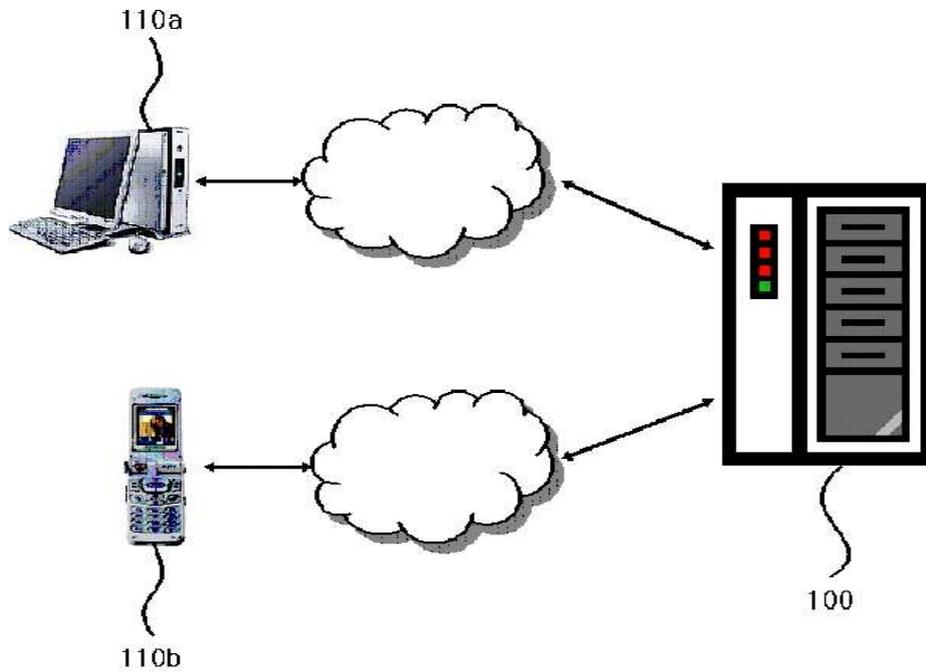
상기 제2 데이터베이스를 참조하여 상기 제3 검색 쿼리와 연관된 하나 이상의 제4 검색 쿼리를 추출하는 검색 쿼리 추출 수단; 및

상기 추출된 제4 검색 쿼리를 상기 사용자 단말기로 제공하는 검색 쿼리 제공 수단

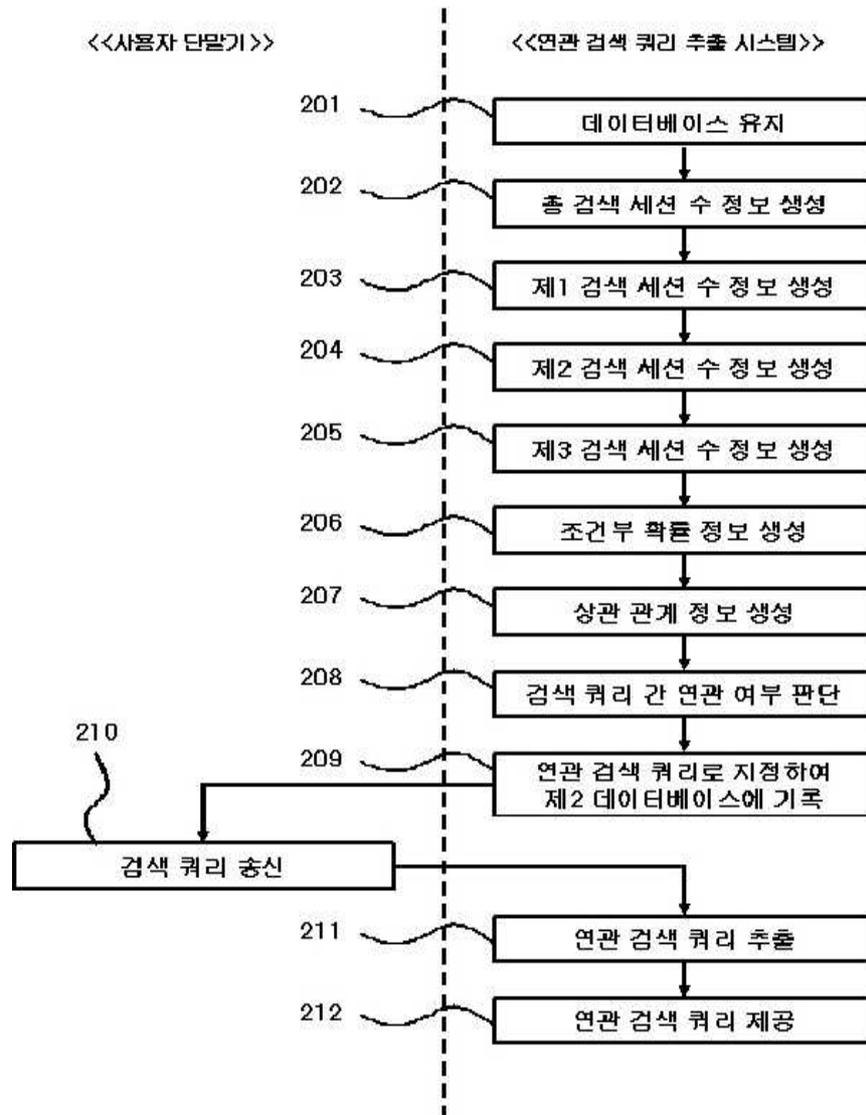
을 더 포함하는 것을 특징으로 하는 연관 검색 쿼리 추출 시스템.

도면

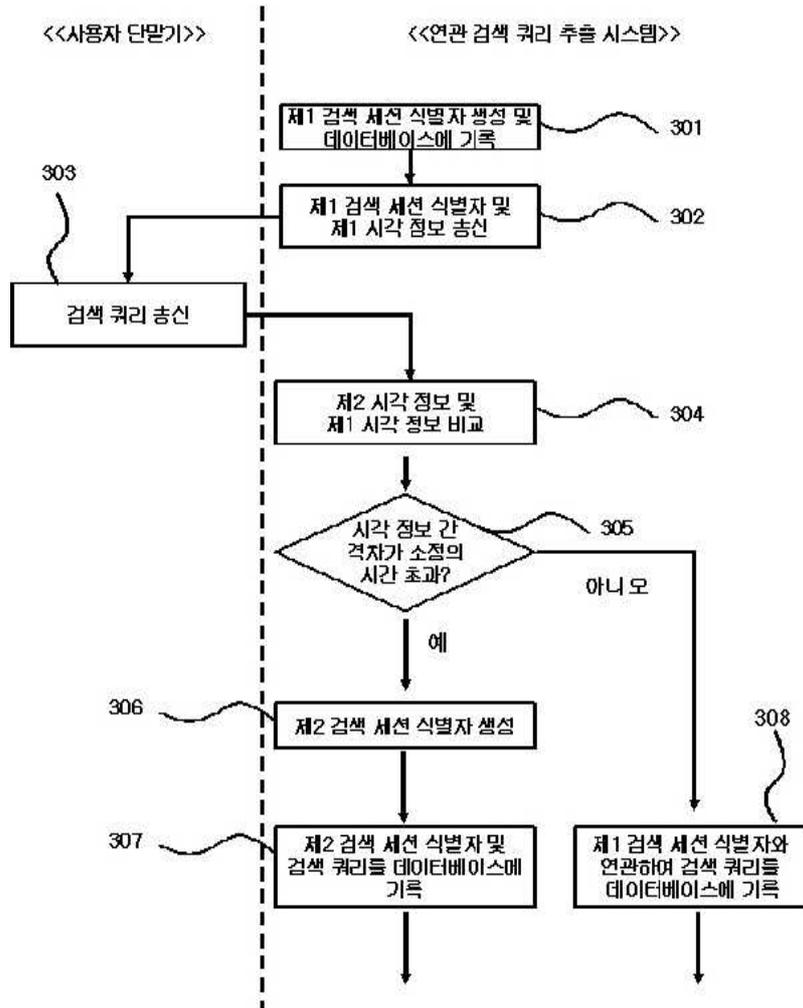
도면1



도면2



도면3



도면4

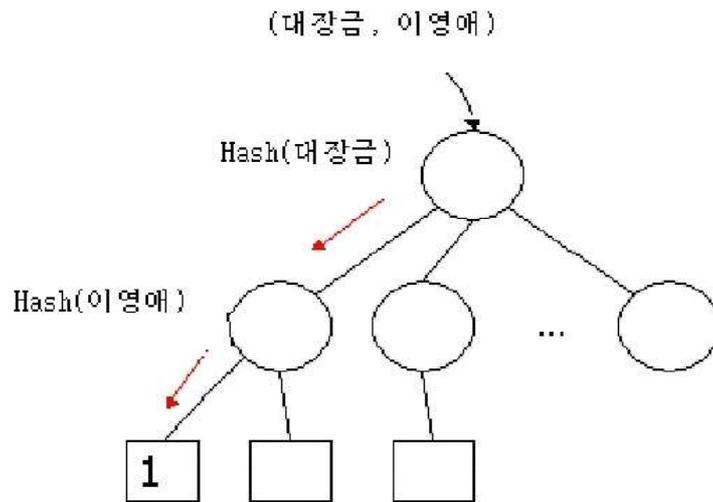
| 검색 세션 식별자<br>(401) | 검색 쿼리        |       |       |       |
|--------------------|--------------|-------|-------|-------|
|                    | session ID 1 | 박찬호   | 메이저리그 | 야구    |
| session ID 2       | 영화           | 박스오피스 | 서울극장  | ----  |
| session ID 3       | 차범근          | 차두리   | 축구    | ----  |
| -----              | -----        | ----- | ----- | ----- |

도면5

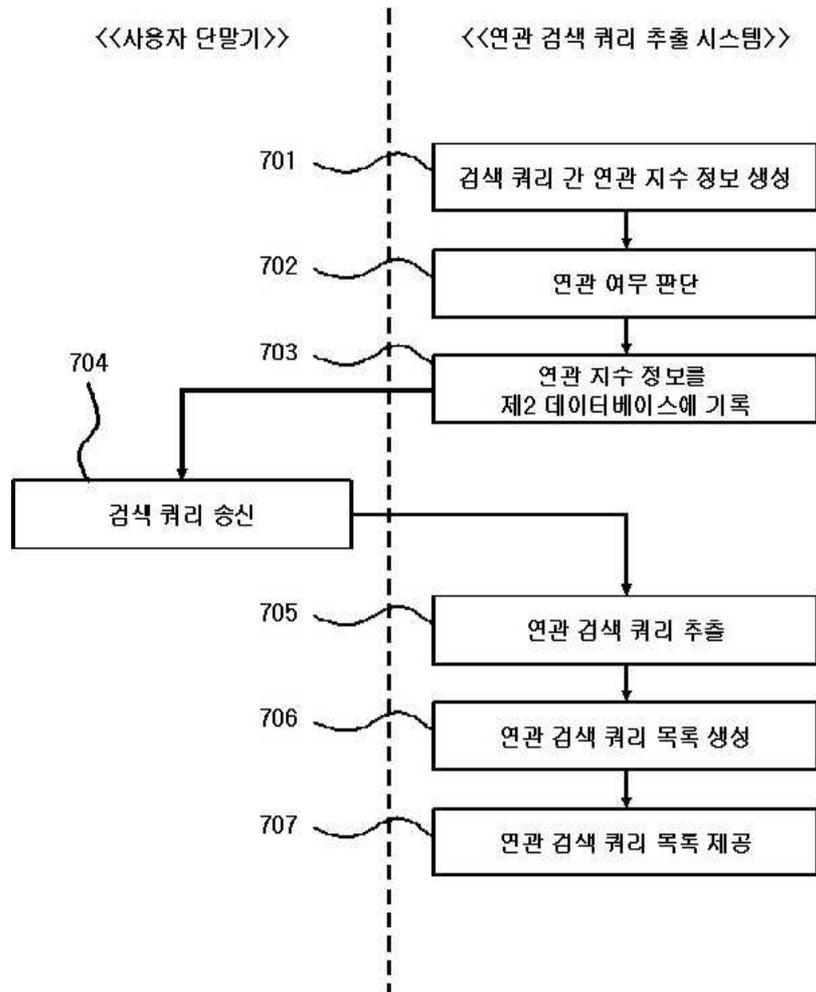
| 검색 세션 식별자 | 검색 쿼리 |       |       |       |
|-----------|-------|-------|-------|-------|
| 56        | 18758 | 18760 | 18761 | ----- |
| 448       | 19700 | 19701 | 19702 | ----- |
| 14501     | 20101 | 20102 | 20103 | ----- |
| -----     | ----- | ----- | ----- | ----- |

501

도면6



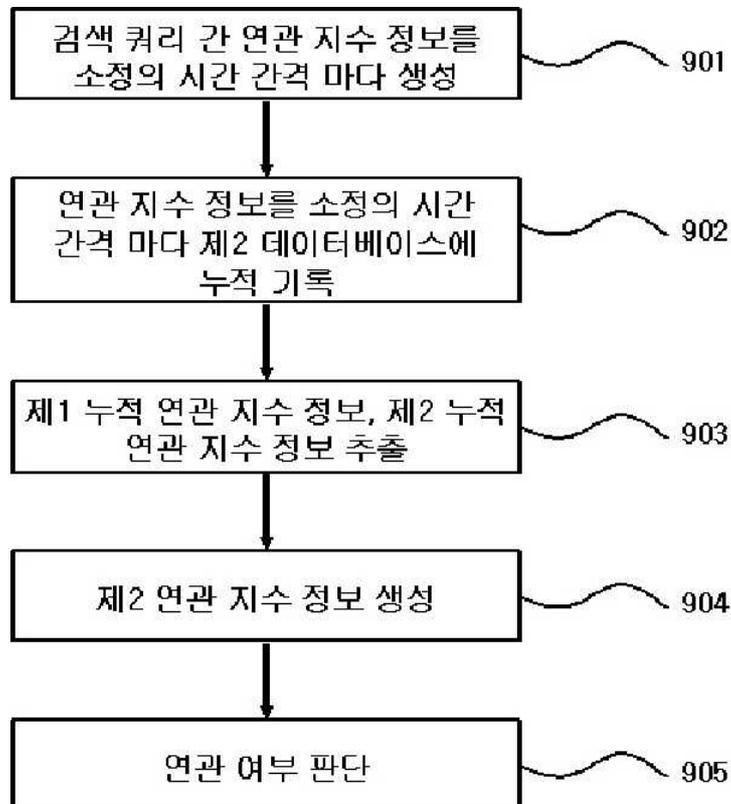
도면7



도면8

| 검색 쿼리 |       | 연관 지수 정보 |
|-------|-------|----------|
| A     | B     | 50       |
|       | C     | 45       |
|       | D     | 40       |
|       | ----- | -----    |
| B     | A     | 50       |
|       | C     | 30       |
|       | D     | 10       |
|       | ----- | -----    |
| ----- | ----- | -----    |

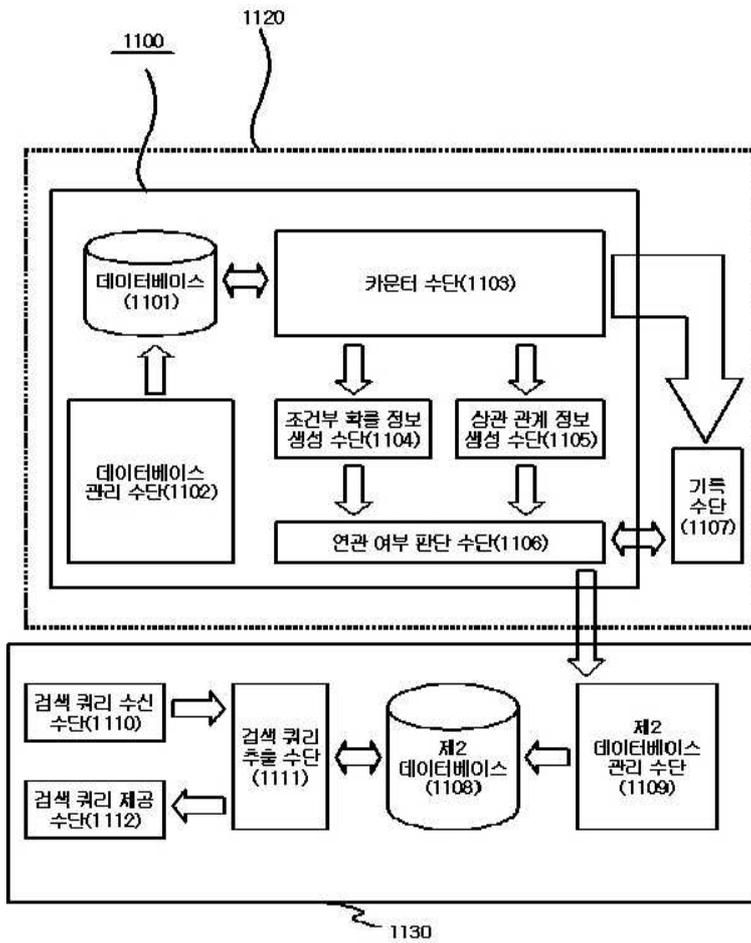
도면9



도면10

| 기록 시점 |     | 2004. 4. 1        | 2004. 4. 2        | 2004. 4. 3        | ----- |
|-------|-----|-------------------|-------------------|-------------------|-------|
| 검색 쿼리 |     | 제1 누적 연관<br>지수 정보 | 제2 누적 연관<br>지수 정보 | 제3 누적 연관<br>지수 정보 | ----- |
| A     | B   | 50                | 70                | 40                | ----- |
|       | C   | 45                | 30                | 15                | ----- |
|       | D   | 40                | 20                | 10                | ----- |
|       | --- | -----             | -----             | -----             | ----- |
| B     | A   | 50                | 70                | 40                | ----- |
|       | C   | 30                | 60                | 30                | ----- |
|       | D   | 10                | 40                | 20                | ----- |
|       | --- | -----             | -----             | -----             | ----- |
| ---   | --- | -----             | -----             | -----             | ----- |

도면11



도면12

