

	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2013-0032623 (43) 공개일자 2013년04월02일
(51) 국제특허분류(Int. Cl.) G06F 17/30 (2006.01) G06F 17/10 (2006.01)		(71) 출원인 인하대학교 산학협력단 인천광역시 남구 인하로 100, 인하대학교 (용현동)
(21) 출원번호 10-2011-0096336 (22) 출원일자 2011년09월23일 심사청구일자 2011년09월23일	(72) 발명자 이주홍 경기도 성남시 분당구 돌마로486번길 7, 동아아파트 202동 401호 (서현동, 효자촌) 안찬민 인천광역시 남구 남주길147번길 42, 3층 (주안동)	
		(74) 대리인 특허법인 정안

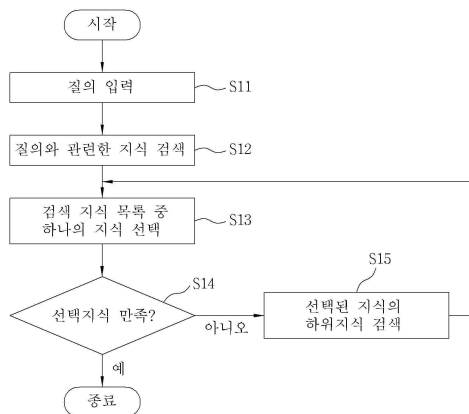
전체 청구항 수 : 총 9 항

(54) 발명의 명칭 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법

(57) 요약

본 발명은 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법에 관한 것으로, 본 발명에 따른 지식 검색 시스템은 사용자에게 의해 질의가 입력되면 그 질의에 대한 지식을 검색하여 추천하고, 그 추천된 지식 중 어느 하나의 지식이 선택되면 그 선택된 지식에 대한 사용자의 만족여부에 따라 상기 선택된 지식의 의미적 포함 정도가 큰 순서대로 지정한 개수만큼의 하위 지식을 추가로 검색하여 추천하므로, 사용자가 검색된 지식에 대해 충분히 만족하지 못하는 경우 검색된 지식 중 사용자가 부분적으로 만족하는 지식을 선택하면 그 선택한 지식을 의미적으로 포함하는 정도가 높은 다른 지식을 검색하여 추천하여 사용자의 요구에 부합하는 결과를 검색하여 추천할 확률을 향상시킬 수 있다.

대표도 - 도3



특허청구의 범위

청구항 1

질문자 단말기로부터 질의를 입력받는 단계와,

상기 질의에 대한 지식을 데이터베이스에서 검색하는 단계와,

상기 지식 검색을 통해 검색된 지식 목록 중 어느 하나의 지식을 선택하는 단계와,

상기 지식이 선택되면 그 선택된 지식에 대한 만족 여부에 따라 상기 선택된 지식과 의미적 포함 정도가 큰 순서대로 지정한 개수만큼의 하위 지식을 추가로 검색하여 추천하는 단계를 포함하는 것을 특징으로 하는 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법.

청구항 2

제1항에 있어서, 상기 지식은,

질의와 질의에 대한 응답들의 집합인 것을 특징으로 하는 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법.

청구항 3

제1항에 있어서, 상기 데이터베이스는,

지식에 대한 질의응답쌍들의 집합으로, 지식과 질의응답쌍의 멤버십 정도를 포함하는 것을 특징으로 하는 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법.

청구항 4

제3항에 있어서, 상기 멤버십 정도는,

질의에 포함된 단어들의 유의어로 질의를 확장하여 상기 유의어와 상기 질의응답쌍 사이의 유사도를 계산하는 확장 유사도와, 상기 지식과 같은 지식의 다른 질의응답쌍 사이의 유사도를 계산하는 보조 유사도를 적용하여 연산되는 것을 특징으로 하는 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법.

청구항 5

제1항에 있어서, 상기 의미적 포함 정도는,

의미적 퍼지 관계 곱연산을 통해 산출되는 것을 특징으로 하는 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법.

청구항 6

제 5항에 있어서, 상기 의미적 퍼지 관계 곱연산은,

퍼지 합의 연산자 값의 평균으로 산출되는 것을 특징으로 하는 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법.

청구항 7

질의를 입력받아 외부로 전송하고, 상기 질의에 대한 지식을 제공받아 화면상에 표시하며 상기 제공받은 지식 중 사용자 만족 지식 존재여부를 출력하는 복수의 질문자 단말기와,

상기 복수의 질문자 단말기와 네트워크를 통해 접속하며 상기 복수의 질문자 단말기로부터 출력되는 질의를 입력받아 등록하고, 그 질의에 대한 지식을 검색하며 상기 사용자 만족 지식 존재여부에 따라 상기 검색된 지식 중 사용자에게 의해 선택된 지식을 의미적 포함정도가 큰 순서대로 지정된 개수만큼의 하위 지식을 추가로 검색하여 추천하는 서버와,

상기 네트워크를 통해 상기 서버에 접속하며 상기 서버에 등록된 질의 중 어느 하나의 질의에 대한 응답을 입력

받아 상기 서버로 전송하는 복수의 답변자 단말기를 포함하는 것을 특징으로 하는 지식 검색 시스템.

청구항 8

제 7항에 있어서, 상기 서버는,

상기 네트워크를 통한 상기 복수의 질문자 단말기 및 상기 복수의 답변자 단말기와의 데이터통신을 위한 통신부와,

상기 통신부를 통해 상기 복수의 질문자 단말기로부터 질의가 입력되면 그 질의에 대한 지식을 검색하여 해당 질문자 단말기로 제공하고, 상기 답변자 단말기로부터 제공되는 특정 질의에 대한 응답이 입력되며 상기 특정 질의를 전송한 해당 질문자 단말기로 입력된 응답을 제공하는 제어부와,

상기 제어부의 제어에 따라 상기 복수의 질문자 단말기로부터 전달되는 질의와 상기 복수의 답변자 단말기로부터 전송되는 응답을 저장하는 데이터베이스를 포함하며,

상기 제어부는 상기 질의에 대한 지식을 검색하여 제공한 후 제공한 지식 중 사용자 만족 지식이 존재하면 지식 추천을 종료하고, 상기 제공한 지식 중 사용자 만족 지식이 존재하지 않으면 사용자에게 의해 선택되는 지식과 의미적 포함 정도가 큰 순서대로 지정된 개수만큼의 하위 지식을 검색하여 해당 질문자 단말기로 제공하는 것을 특징으로 하는 지식 검색 시스템.

청구항 9

제8항에 있어서, 상기 데이터베이스는,

지식에 대한 질의응답쌍들의 집합으로, 상기 지식과 질의응답쌍의 멤버십 정도를 포함하는 것을 특징으로 하는 지식 검색 시스템.

명세서

기술 분야

[0001] 본 발명은 검색된 지식의 내용을 의미적으로 포함하는 정도를 고려하여 지식을 추천하는 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법에 관한 것이다.

배경 기술

[0002] 최근 정보의 양이 많아지면서 원하는 정보를 찾기 위한 검색의 중요성이 증가하고 있다. 지식 검색 시스템은 지식 검색 시스템의 한 형태로서 사용자의 질의에 대해서 전문적인 지식을 가진 전문가와 같은 다른 사용자로부터 응답을 받기 때문에, 문서 검색으로 질의에 대한 응답을 구하는 방법보다 더 적합한 응답을 얻을 수 있다. 따라서 최근 많은 포털 서비스에서 지식 검색 시스템 서비스를 제공하고 있다. 대표적인 지식 검색 시스템에는 Yahoo! Answers, Answer.com, ask.com askville등이 있고, 지식 검색 시스템에는 네이버 지식-iN, Daum 지식, 네이트 지식 등이 있다.

[0003] 지식은 사용자의 질의와 그에 대한 응답들로 구성되고 그 내용은 전문가나 정확한 지식을 가진 다른 사용자가 응답하기 때문에 매우 정확한 정보를 가지고 있다. 그러나 이용자의 수가 많아지고 지식의 수도 매우 많아짐에 따라, 사용자의 질의에 적합한 지식을 시스템에서 찾는 것이 어려워지고 있다. 이는 사용자의 질의 의도를 시스템에 명확하게 전달하는 것이 어렵기 때문이다. 즉, 같은 질의문이라도 사용자에게 따라 원하는 의도가 다를 수 있다. 또한, 사용자가 입력한 질의의 표현이 부족하거나 사용된 단어의 다의성 등으로 인해 질의 의도가 모호할 수도 있다. 또한, 애초에 사용자가 자신이 찾고자 하는 내용을 정확히 알지 못하는 상태에 있기 때문에 사용자가 질의문을 부정확하게 작성할 수 있다. 이러한 이유들로 인해서 종래에 많은 연구들이 수행되었음에도 불구하고 시스템이 사용자가 만족할 수 있는 검색 결과를 찾아주는 데에는 한계가 있을 수 밖에 없다. 따라서 사용자가 검색 결과에 만족하지 못할 때 더 적합한 검색 결과를 얻기 위한 추가적인 검색 기능이 필요하다.

[0004] 이에, 최근에는 이용자들의 수가 상대적으로 매우 많은 웹 포털에 기반한 지식 검색 시스템이 등장하고 활발하게 사용되기 시작하면서 질의에 대한 응답을 사용자로부터 얻는 방법은 새로운 연구 분야로서 주목을 받고 있다. 웹 포털 환경의 지식 검색 시스템에서는 사용자가 질의를 입력하고, 다른 사용자가 입력한 질의에 대해 응답/평가하는 환경을 제공한다. 대표적인 시스템으로 Yahoo! Answers, ask.com, askvillie.com, 네이버 지식-

iN, 네이트 지식, 다음 지식 등이 있다. 이러한 시스템들은 사용자의 질의에 대해 정확한 응답을 할 수 있을 정도의 전문가가 직접 응답을 입력하기 때문에 일반적으로 좋은 응답을 제공할 수 있다.

[0005] 이러한 웹 기반 질의 응답을 이용하는 사용자의 수가 증가하고 지식의 수가 증가하면서 웹 포털 환경의 지식 검색 시스템에 대한 연구는 크게 두 가지 방향으로 이루어졌다. 첫 번째는 질의에 대한 응답에 관한 연구로서 응답을 평가하는 방법, 좋은 응답을 얻는 방법 등 질의에 대한 정확한 응답을 얻는 방법을 연구한다. 두 번째는 이미 완성된 지식을 검색하는 방법에 관한 연구로서 의미적으로 유사한 질의를 검색하는 방법에 관한 연구이다. 이는 시스템에 질의에 부합하는 응답이 있음에도 사용자가 그 응답을 찾지 못할 수도 있고 이용자의 수가 많아지면서 질의 응답의 수도 매우 많아짐에 따라 사용자가 원하는 정보를 찾기 어렵기 때문이다.

[0006] 또한, 웹 포털 환경의 지식 검색 시스템들은 키워드 기반의 검색을 사용하기 때문에 같은 키워드를 갖는다고 하더라도 의도가 다른 지식들이 검색되거나, 의도는 같지만 질의에 사용된 단어가 다른 지식을 함께 보여주지 못하며, 의미와 상관없이 단순히 질의에 속한 카테고리의 질의들을 보여주는 정도로 그 기능이 매우 미약하다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 상기와 같은 문제점을 개선하기 위해 안출된 것으로, 검색된 지식의 내용을 의미적으로 포함하는 정도를 고려하여 지식을 추천하는 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법을 제공하기 위한 것이다.

과제의 해결 수단

[0008] 상기한 과제를 해결하기 위한, 본 발명에 따른 지식 검색 시스템에서 의미적인 퍼지 포함 정도를 사용한 탐색적 지식 추천 방법은 질문자 단말기로부터 질의를 입력받는 단계와, 상기 질의에 대한 지식을 데이터베이스에서 검색하는 단계와, 상기 지식 검색을 통해 검색된 지식 목록 중 어느 하나의 지식을 선택하는 단계와, 상기 지식이 선택되면 그 선택된 지식에 대한 만족 여부에 따라 상기 선택된 지식과 의미적 포함 정도가 큰 순서대로 지정한 개수만큼의 하위 지식을 추가로 검색하여 추천하는 단계를 포함한다.

[0009] 또한, 본 발명에 따른 지식 검색 시스템은 질의를 입력받아 외부로 전송하고, 상기 질의에 대한 지식을 제공받아 화면상에 표시하며 상기 제공받은 지식 중 사용자 만족 지식 존재여부를 출력하는 복수의 질문자 단말기와, 상기 복수의 질문자 단말기와 네트워크를 통해 접속하며 상기 복수의 질문자 단말기로부터 출력되는 질의를 입력받아 등록하고, 그 질의에 대한 지식을 검색하며 상기 사용자 만족 지식 존재여부에 따라 상기 검색된 지식 중 사용자에게 의해 선택된 지식을 의미적 포함정도가 큰 순서대로 지정된 개수만큼의 하위 지식을 추가로 검색하여 추천하는 서버와, 상기 네트워크를 통해 상기 서버에 접속하며 상기 서버에 등록된 질의 중 어느 하나의 질의에 대한 응답을 입력받아 상기 서버로 전송하는 복수의 답변자 단말기를 포함한다.

발명의 효과

[0010] 상기와 같은 본 발명에 따르면, 본 발명은 검색된 지식의 내용을 의미적으로 포함하는 정도가 큰 다른 지식을 추천한다. 따라서, 본 발명은 사용자가 검색된 지식에 대해 충분히 만족하지 못하는 경우 검색된 지식 중 사용자가 부분적으로 만족하는 지식을 선택하면 그 선택한 지식을 의미적으로 포함하는 정도가 높은 다른 지식을 검색하여 추천하여 사용자의 요구에 부합하는 결과를 검색하여 추천할 확률을 향상시킬 수 있다.

도면의 간단한 설명

[0011] 도 1은 본 발명과 관련된 지식 검색 시스템의 전체구성도를 도시한 블록도이다.

도 2는 도 1의 서버를 도시한 블록구성도이다.

도 3은 본 발명과 관련된 지식 검색 시스템의 지식 추천 방법을 도시한 흐름도이다.

도 4는 본 발명과 관련된 지식 검색 시스템에서 질의에 대한 추천 지식을 도시한 일 예이다.

도 5는 도 4의 추천 지식 K_2 에 대한 추천 지식을 도시한 일 예이다.

도 6은 본 발명과 관련된 퍼지 지식 사이의 의미적 퍼지 포함 정도를 도시한 도면이다.

도 7은 본 발명과 관련된 의미적 퍼지 포함 관계를 구성하는 두 퍼지 지식을 도시한 일 예이다.

도 8은 크리스프 개념으로 표현한 규칙을 도시한다.

도 9a 내지 도 9c는 본 발명과 관련된 두 지식이 의미적 포함관계를 가질 경우의 규칙들을 각각 도시한다.

도 10은 본 발명에 따른 퍼지 지식에 대한 T의 멤버십 정도를 도시한다.

도 11은 본 발명에 따른 퍼지 지식에 대한 T의 멤버십 정도 연산과정을 도시한다.

도 12는 본 발명에 따른 퍼지 지식 행렬을 도시한 일 예이다.

도 13은 본 발명에 따른 의미적 포함관계 행렬의 일 예를 도시한다.

도 14는 본 발명에 따른 지식 추천 방법의 평가결과를 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0012] 이하, 첨부된 도면을 참조하여 본 발명과 관련된 실시예를 상세하게 설명한다.
- [0013] 도 1은 본 발명과 관련된 지식 검색 시스템의 전체구성도를 도시한 블록도이다.
- [0014] 도 1에 도시된 바와 같이, 유/무선 통신망인 네트워크를 통해 복수의 질문자 단말기(10), 복수의 답변자 단말기(20), 서버(30) 사이에 질의응답과 관련된 데이터통신이 이루어진다.
- [0015] 상기 복수의 질문자 단말기(10)는 네트워크를 통해 상기 서버(30)에 접속하여 질문자가 입력한 질문 데이터(질의)를 상기 서버(30)로 전송한다. 그리고, 질문자 단말기(10)는 상기 서버(30)로부터 상기 질문자가 문의한 질의에 대한 하나 이상의 답변 데이터(응답)를 실시간으로 제공받아 화면상에 표시한다. 상기 질문자 단말기(10)는 상기 화면상에 표시된 답변 데이터들 중 어느 하나의 답변 데이터가 질문자에 의해 선택되면 그 선택된 답변 데이터의 상세한 정보를 상기 서버(30)로부터 제공받아 화면에 표시하여 질문자가 확인할 수 있게 있다. 상기 질문자 단말기(10)는 상기 선택된 답변 데이터에 대한 질문자의 만족도에 따라 상기 선택된 답변 데이터와 의미적 포함 관계가 높은 답변 데이터의 추가 검색을 상기 서버(30)에 요청하여 제공받는다. 그리고, 상기 질문자 단말기(10)는 상기 서버(30)로부터 제공받은 추가 검색 결과를 화면상에 표시한다.
- [0016] 상기 복수의 답변자 단말기(20)는 네트워크를 통해 상기 서버(30)에 접속하고, 질문자 단말기(10)에 의해 상기 서버(30)에 등록된 복수의 질문 데이터를 상기 서버(30)로부터 제공받아 화면상에 표시하고, 답변자의 지시에 따라 상기 복수의 질문 데이터 중 어느 하나의 질문 데이터를 선택한다. 그리고, 답변자 단말기(20)는 선택한 질문 데이터를 확인한 답변자에 의해 입력되는 답변 데이터를 상기 서버(30)로 전송한다.
- [0017] 상기 서버(30)는 질문자 단말기(10)로부터 전송되는 질문 데이터와 상기 답변자 단말기(20)로부터 전달되는 상기 질문 데이터에 대한 답변 데이터를 데이터베이스로 구축하여 다른 질문자 및 답변자들과 공유할 수 있게 한다.
- [0018] 다음, 상기한 도 1에 도시된 서버의 구성을 도 2를 참조하여 설명한다.
- [0019] 도 2는 도 1의 서버를 도시한 블록구성도이다.
- [0020] 도 2에 도시된 바와 같이, 서버(30)는 통신부(31), 제어부(32), 데이터베이스(33)를 포함한다.
- [0021] 상기 통신부(31)는 복수의 질문자 단말기(10) 및 복수의 답변자 단말기(20)와의 데이터 통신을 위한 것으로, 네트워크를 통해 상기 복수의 질문자 단말기(10) 및 복수의 답변자 단말기(20)의 접속을 제어한다.
- [0022] 상기 통신부(31)는 복수의 질문자 단말기(10)와 복수의 답변자 단말기(20)로부터 입력되는 질문 데이터 및 답변 데이터를 상기 제어부(32)로 출력한다.
- [0023] 상기 제어부(32)는 상기 복수의 질문자 단말기(10)로부터 질문 데이터가 입력되면 상기 질문 데이터에 대한 답변 데이터를 상기 데이터베이스(33)에서 검색하고, 그 검색결과를 상기 통신부(31)를 통해 상기 질문자 단말기(10)로 출력한다.
- [0024] 또한, 상기 제어부(32)는 복수의 답변자 단말기(20)로부터 특정 질문 데이터에 대한 답변 데이터가 상기 통신부(31)를 통해 입력되면 상기 특정 질문 데이터와 그에 해당하는 답변 데이터를 데이터베이스화한다.
- [0025] 데이터베이스(33)는 상기 제어부(32)의 제어에 따라 상기 복수의 질문자 단말기(10)로부터 입력되는 질문 데이

터 및 상기 복수의 답변자 단말기(20)로부터 입력되는 답변 데이터를 저장한다. 상기 데이터베이스(33)는 질의와 그 질의에 대한 응답들의 집합으로 구성된 지식들이 저장되는 지식 데이터베이스와 상기 지식 데이터베이스에 저장된 모든 지식들에 대한 질의응답쌍(퍼지지식)들이 저장되는 질의응답 데이터베이스를 포함한다.

[0026] 상기한 바와 같이 구성된 본 발명에 따른 질의응답 시스템의 지식 추천 방법에도 3을 참조하여 설명한다.

[0027] 도 3은 본 발명과 관련된 지식 검색 시스템의 지식 추천 방법을 도시한 흐름도이다.

[0028] 먼저, 질문자가 질문자 단말기(10)를 통해 질문 데이터를 입력하면 질문자 단말기(10)는 네트워크를 통해 상기 입력된 질문 데이터를 서버(30)로 전송한다(S11).

[0029] 상기 서버(30)는 통신부(31)를 통해 상기 질문 데이터를 수신하고, 그 수신된 질문 데이터에 대한 지식을 지식 데이터베이스에서 검색하여 검색된 지식을 제공한다(S12). 여기서, 상기 서버(30)는 지식 검색 결과 질문자가 만족할 수 있는 검색 지식이 존재하지 않으면 이를 알리는 메시지를 출력한다.

[0030] 상기 검색된 지식들 중 어느 하나의 지식이 선택되면 상기 서버(30)는 상기 선택된 지식에 대한 만족도를 평가한다.

[0031] 상기 만족도를 평가한 결과 질문자가 상기 선택된 지식의 상세한 내용에 만족하면 상기 서버(30)는 지식 추천 과정을 종료한다(S13, S14).

[0032] 한편, 상기 선택된 지식에 만족하지 않으면 상기 서버(30)는 선택된 지식의 하위 지식 검색을 재귀적으로 수행한다(S15). 이때, 상기 서버(30)는 상기 선택된 지식과 의미적 포함관계가 높은 하위지식을 검색하며 상기 의미적 포함 관계값이 높은 순서대로 지정한 개수 만큼의 하위지식을 추천 지식 목록에 저장한다.

[0033] 이하에서는, 지식 검색 시스템의 지식 추천을 위한 지식 검색에 사용되는 이론들을 상세하게 설명한다.

[0034] 지식 검색 시스템의 데이터 단위는 지식 k 이고, 지식 k 는 질의 (Q)와 그에 대한 응답들(A)의 집합으로, 수 1 과 같이 나타낸다.

수학식 1

$$k = \{Q, A_1, A_2, \dots, A_n\}$$

[0036] 그리고, 지식 k 의 퍼지 지식 K_i 는 수 2와 같이 질의응답 쌍 (T)들을 멤버로 가진다.

수학식 2

$$K_i = \left\{ (T_{jp}, \mu_{K_i}(T_{jp})) \left| \begin{array}{l} T_{jp} = (Q_j, A_{jp}) \in TDB, \\ k_j = \{Q_j, A_{j1}, A_{j2}, \dots, A_{jn_j}\} \in KDB, \\ j = 1, 2, \dots, m, \\ p = 1, \dots, n_j \end{array} \right. \right\}$$

$$TDB = \{T_{11}, T_{12}, \dots, T_{1n_1}, T_{21}, T_{22}, \dots, T_{2n_2}, \dots, T_{m1}, T_{m2}, \dots, T_{mn_m}\}$$

$$KDB = \{k_1, k_2, \dots, k_i, \dots, k_m\}$$

[0040] 여기서, $T_{jp} = (Q_j, A_{jp})$ 는 지식 k_j 의 질의와 그에 대한 p 번째 응답의 쌍을 나타낸다. 그리고, KDB는 시스템에 저장된 지식들의 집합인 지식 데이터베이스이고, TDB는 모든 질의응답쌍들의 집합인 질의응답 데이터베이스

스이다. $\mu_{K_i}(T_{jp})$ 는 퍼지 집합 K_i 의 멤버 T_{jp} 의 멤버쉽 정도를 의미한다. 멤버쉽 정도는 지식 k_i 와 멤버 T_{jp} 의 부연 유사도로 계산된다.

[0041] 퍼지 지식 K 가 $K = \{T, \mu_K(T) | T \in TDB\}$ 일 때 멤버쉽 정도가 γ 보다 큰 T 들로 구성된 퍼지 지식을 $K^\gamma = \{T, \mu_K(T) | T \in TDB, \mu_K(T) > \gamma\}$ 로 표기하며 K 의 γ 컷(cut) 이라 한다.

[0042] 예를 들어, 산에 대한 지식을 k_1 이라고 하고, 자동차에 대한 지식을 k_2 라 하고, 산길 교통사고에 대한 지식을 k_3 라고 하자. 시스템에 저장된 모든 지식은 k_1, k_2, k_3 라 하고 각각 $k_1 = \{Q_1, A_{11}, A_{12}, A_{13}\}$, $k_2 = \{Q_2, A_{21}, A_{22}, A_{23}\}$, $k_3 = \{Q_3, A_{31}, A_{32}\}$ 으로 정의된다고 가정하자. 질의응답 쌍 T 는 $T_{11} = \{Q_1, A_{11}\}$, $T_{12} = \{Q_1, A_{12}\}$, $T_{21} = \{Q_2, A_{21}\}$, ..., $T_{32} = \{Q_3, A_{32}\}$ 으로 정의되며 질의응답 쌍의 집합 TDB는 수 3과 같이 나타낼 수 있다. 상기 수 3을 다시 수 4와 같이 나타낼 수 있다.

수학식 3

[0043] $TDB = \{T_{11}, T_{12}, T_{13}, T_{21}, T_{22}, T_{23}, T_{31}, T_{32}\}$

수학식 4

[0044] $TDB = \{T^1, T^2, T^3, T^4, T^5, T^6, T^7, T^8\}$

[0045] 지식 k_1 에 대한 퍼지 지식 K_1 은 TDB의 모든 질의응답 쌍(문서)에 대한 멤버쉽 정도의 집합으로, 아래 수 5와 같이 나타낸다.

수학식 5

[0046] $K_1 = \{(T^1, \mu_{K_1}(T^1)), (T^2, \mu_{K_1}(T^2)), \dots, (T^8, \mu_{K_1}(T^8))\}$

[0047] $K_2 = \{(T^1, \mu_{K_2}(T^1)), (T^2, \mu_{K_2}(T^2)), \dots, (T^8, \mu_{K_2}(T^8))\}$

[0048] $K_3 = \{(T^1, \mu_{K_3}(T^1)), (T^2, \mu_{K_3}(T^2)), \dots, (T^8, \mu_{K_3}(T^8))\}$

[0049] 상기 수 5의 퍼지 지식을 행렬로 나타냈을 때 표 1과 같은 결과를 얻었다고 하면, 지식 K_1 과 다른 문서 T 들의 멤버쉽 정도(유사도) 값은 다음과 같이 해석할 수 있다.

표 1

[0050]

	T1	T2	T3	T4	T5	T6	T7	T8
K1	0.651	1.000	0.547	0.322	0.471	0.216	0.436	0.218
K2	0.188	0.231	0.192	0.902	1.000	0.899	0.648	0.641
K3	0.218	0.517	0.599	0.717	0.819	0.791	0.986	1.000

- [0051] $T^1 \sim T^3$ 까지는 산에 대한 질의응답 쌍이기 때문에 K_1 에 대한 멤버십 정도 값이 크다 (1에 가까운 값). $T^4 \sim T^6$ 까지는 산과 관계 없는 자동차에 대한 내용이기 때문에 멤버십 정도 값이 작다(0에 가까운 값). T^7, T^8 은 산길 교통사고에 대한 질의응답 쌍이기 때문에 산에 대한 내용과 자동차에 대한 내용을 어느 정도 다룬다고 볼 수 있기 때문에 멤버십 값이 $T^1 \sim T^3$ 보다는 작지만 $T^4 \sim T^6$ 보다는 큰 값으로 나타난다.
- [0052] 본 발명에서는 상기 멤버십 정도 계산에 사용되는 부연 유사도 연산을 개선하기 위해 두 가지 방법을 적용한다. 첫번째는, 질의의 단어를 각 단어의 유의어로 질의를 확장하여 확장 질의 유사도를 계산하고, 두번째는, 같은 지식의 이웃한 문장의 유사도 값을 이용하는 보조 유사도를 계산한다. 이 두 방법을 적용한 유사도 계산 방법을 부연 유사도라고 정의하고 지식의 멤버가 가지는 멤버십 정도를 계산하는 데 이용한다.
- [0053] 질의 문장은 몇 개의 단어로 이루어지는데, 하나의 단어는 여러 가지 의미를 가질 수 있다. 그래서, 같은 질의 문을 작성하였지만 질의 의도가 달라질 수 있다. 이러한 단어의 다의성 때문에 단어의 확장이 필요하다. 본 발명에서는 워드넷(WordNet)을 이용하여 영어 단어의 유의어를 검색하였다. 질의의 단어에 대해 유의어 관계를 가진 단어들로 질의를 확장하여 확장 유사도 $esim$ 를 계산한다. 상기 확장 유사도는 수 6과 같이 정의한다.

수학식 6

[0054]
$$esim(K_i, T) = BM25(Q_i, U) + \alpha \cdot BM25(S_i, U)$$

[0055]
$$+ \lambda \cdot \left(\sum_{j=1}^n (BM25(A_{ij}, U) + \beta \cdot BM25(B_{ij}, U)) \right)$$

[0056] 여기서, $BM25(S_i, U)$ 는 $BM25$ 모델을 이용한 S_i 와 U 의 유사도 계산을 의미하며, α , β , λ 는 각각 $0 < \alpha \leq 1$, $0 < \beta \leq 1$, $0 < \lambda \leq 1$ 이다.

[0057] Q_i 는 지식 K_i 의 질의에 사용된 단어들의 집합이고, S_i 는 Q_i 에 속한 단어들의 유의어들의 집합이다. A_{ij} 는 지식 K_i 의 j 번째 응답에 사용된 단어들의 집합이고, B_{ij} 는 A_{ij} 에 속한 단어들의 유의어들의 집합이다. U 는 질의응답 쌍 T 에 사용된 단어들과 그 단어들의 유의어들로 구성된 집합이다. 단어들은 어근 분석(stemming)과 불용어 제거(stopword removal) 작업을 거친다.

[0058] 지식 검색 시스템에서는 지식 k_i 의 응답 A_{in_i} 들은 질의 Q_i 에 대한 다른 사용자들의 응답이다. 그렇기 때문에 질의 Q_i 와 응답 A_{in_i} 의 유사도 값이 0이라고 해도 의미적인 연관 정도는 어느 정도 존재한다고 볼 수 있다.

[0059] 따라서, 본 발명에서는 같은 지식 k_i 에 속한 질의 Q_i 에 대한 응답 A_{in} 들은 다른 지식 k_j 에 존재하는 응답 A_{jn} 들보다 연관 정도가 높다는 가정하에 보조 유사도 방법을 정의한다. 지식 k_l 이 $k_l = \{Q_l, A_{l1}, A_{l2}, \dots, A_{ln}\}$ 이라고 하면 질의응답 쌍의 집합 τ_l 은 $\tau_l = \{T_{l1}, T_{l2}, \dots, T_{ln}\}$ 으로 나타낼 수 있다. 이때 $T_{lp} = \{(Q_l, A_{lp})\}$ 는 τ_l 의 p 번째 질의응답 쌍이고 A_{lp} 는 k_l 의 p 번째 응답이다. 보조 유사도

ssim 은 수 7과 같이 정의한다.

수학식 7

$$ssim(K_i, T_{lp}) = \frac{\sum_{m=1, m \neq p}^n nesim(K_i, T_{lm}) \cdot \frac{1}{|m - p|^2}}{\sum_{m=1, m \neq p}^n \frac{1}{|m - p|^2}}$$

[0060]

수 7은 p 번째 응답의 유사도를 계산하기 위해 같은 지식에 포함된 n 개의 응답들의 유사도를 이용한다. p 번째 응답과 m 번째 응답의 거리가 멀수록 m 번째 응답의 유사도의 영향력이 작아진다. 확장 유사도 $esim$ 을 정규화하는 $nesim$ 은 수 8과 같이 정의한다.

수학식 8

$$nesim(K_i, T_{lx}) = \frac{esim(K_i, T_{lx})}{MAX_{T_{lx} \in \tau_l} (esim(K_i, T_{lx}))}$$

[0062]

확장 유사도와 보조 유사도를 적용한 유사도를 부연 유사도 $Xsim(Q, T_{lp})$ 으로 정의하며 수 9와 같이 나타낸다.

수학식 9

$$Xsim(K_i, T_{lp}) = \frac{nesim(K_i, T_{lp}) + \mu \cdot ssim(K_i, T_{lp})}{1 + \mu}$$

[0064]

여기서, μ 는 유사도 계산을 조절하는 스케일링 요소(scaling factor)이며, μ 이 클수록 부연 유사도의 비율이 높아진다.

[0065]

사용자가 원하는 내용과 연관된 정도를 판단하는 주관적 척도로서 만족도(Satisfaction degree)를 정의한다. 만족도가 높은 지식은 사용자가 원하는 내용과 연관 정도가 높은 지식이다. 즉, 추천되는 지식은 만족도가 높아야 한다. 그런데 지식에 대한 만족도는 사용자에게 따라 기준이 다르다. 그렇기 때문에 지식의 만족도는 정량적으로 표현할 수 없다.

[0066]

하지만, 어떤 지식의 내용이 다른 지식의 내용과 연관된 정도는 유사도를 이용하여 객관적으로 계산할 수 있다. 그리고 만족도가 높은 지식일수록 사용자가 그 지식을 기반으로 또 다른 지식을 추천 받을 때 원하는 정보를 발견할 가능성이 높아진다. 따라서, 본 발명에서는 사용자로 하여금 만족도가 높을 것으로 여겨지는 지식을 선택하도록 하고, 선택한 지식을 의미적으로 포함하는 정도가 큰 지식을 추천 지식으로 보여줌으로써 만족도가 높을 가능성이 높은 지식들로부터 원하는 정보를 찾을 수 있도록 도와주는 방법을 제안한다.

[0067]

예를 들어, 세 명의 사용자 A, B, C가 똑같이 "교통 사고가 났습니다"는 질의를 입력하였으나, 각각 다음과 같이 서로 다른 질의 의도를 가지고 다른 정보를 얻고자 한다고 가정한다.

[0068]

A: 교통 사고 시 보험 처리 방법에 대한 정보

[0069]

B: 교통 사고 시 면허 변동에 대한 정보

[0070]

- [0071] C: 교통 사고 시 차량 수리에 대한 정보
- [0072] 같은 질의를 사용하였기 때문에 A, B, C는 모두 같은 지식을 추천받게 된다. 질의에 대해 3개의 지식 K_1 , K_2 , K_3 가 추천되었다고 가정하면 도 4와 같다.
- [0073] 추천된 지식은 K_1 , K_2 , K_3 로 동일하지만 각 사용자에게 만족도가 높은 지식은 A는 K_1 , B는 K_2 , C는 K_3 로 각각 다르다. 만일 사용자가 K_1 , K_2 , K_3 에서 만족도가 충분히 높은 지식을 발견하지 못한 경우 그 중에서 만족도가 어느 정도 높은 지식을 선택할 수 있다.
- [0074] 사용자 B가 K_1 , K_2 , K_3 의 만족도가 충분히 높지 않다고 판단하고 그 중에서 만족도가 높은 K_2 를 선택하였을 때, K_2 에 대한 추천 지식 K_4 , K_5 , K_6 이 도 5와 같이 검색되었다고 가정한다.
- [0075] 추천되는 지식 중에서 K_6 는 K_2 와 의미적 포함 정도가 가장 크지만 사용자가 원하는 지식과 연관 정도가 없으므로 좋은 추천 지식이 될 수 없다. K_4 와 K_5 는 각각 K_2 와 의미적 포함 정도가 비슷하지만 사용자가 원하는 지식과 연관 정도는 K_4 보다 K_5 가 더 크므로 사용자는 K_5 를 통해 원하는 정보를 얻을 가능성이 높다.
- [0076] 사용자에게 유용한 추천 지식은 사용자가 입력한 질의보다 많은 정보를 찾을 수 있어야 한다. 그리고 추천 지식은 사용자가 찾고자 하는 정보와 연관 정도가 높은 정보를 찾을 수 있어야 한다. 이러한 조건들을 갖는 추천 질의를 찾기 위해, 지식을 퍼지 지식으로 간주하고 두 퍼지 지식의 의미적 퍼지 포함 정도를 고려한다.
- [0077] 도 6에 도시된 바와 같이, 두 퍼지 지식의 의미적 퍼지 포함 정도는 퍼지 지식 K_i , K_j 에 대해 K_i 가 K_j 에 의미적으로 포함되는 정도를 의미하며, 이는 지식 k_i 가 다루는 내용을 k_j 가 의미적으로 잘 설명하는 정도, 자세히 설명하는 정도, 또는 k_j 가 k_i 보다 넓은 범위의 내용을 설명하는 정도라고 말할 수 있다. 의미적 퍼지 포함 정도는 의미적 퍼지 관계 곱 연산으로 계산할 수 있다.
- [0078] 지식 설명의 멤버십 정도 값 $a = \mu_K(T)$, $b = \mu_{K'}(T)$ 는 각각 지식 k 와 k' 에 대해 T 가 설명하는 정도를 나타낸다. 기존의 퍼지 관계 곱 연산에서 $a \rightarrow b$ 의 값은 T 에 의해서 K' 이 K 를 얼마나 의미적으로 포함하느냐를 나타내는 정도를 계산해야 한다. 그런데, a 와 b 의 값이 작은 값 일때 $a \rightarrow b$ 가 의미하는 것은 T 가 k 에서 중요하지 않은 내용인데 k' 에서도 T 가 중요하지 않게 간략히 서술되었다는 뜻이다. 이 경우에 K' 이 K 를 의미적으로 포함하는 정도를 T 로써 판단할 수 있는 근거가 매우 약하다. 따라서 이 경우 $a \rightarrow b$ 값은 작은 값이어야 한다. 그런데 Kleene-Diense는 큰 값으로 계산한다. 이런 의미에서 Kleene-Diense 연산자는 의미적 포함관계를 계산하는데 있어서 적합하지 않다.
- [0079] 새로운 의미적 퍼지 함의 연산자 (semantic fuzzy implication operator) $\rightarrow$ 는 기존의 퍼지 함의 연산자 대신 퍼지 관계 곱에 적용할 수 있으면서, 두 지식의 의미적 포함 관계를 고려하도록 다음과 같이 정의한다.
- [0080] K_a 와 K'_b 에 대한 지식 설명의 멤버십 정도는 각각 $x_a = \mu_{K_a}(T)$ 와 $y_b = \mu_{K'_b}(T)$ 라고 한다. 의미적 퍼지 함의 연산자의 값 $x_a \rightarrow y_b$ 의 의미는 k_a 와 k'_b 가 각각 T 로써 설명되는데 k_a 가 k'_b 에 T 로써 의미적으로 포함되는 정도를 나타낸다. 즉, 질의응답쌍 T 이 k_a 를 설명하는 정도보다 질의응답쌍 T 이 k'_b 를 더 잘 설명되거나 더 많이 설명된다는 정도를 계산한다. $x_a \rightarrow y_b$ 값이 가질 수 있는 상황에 따른 의미는 표 2와 같이 분류할 수 있다.

표 2

y1 (큰 값)	v_2	v_1
y0 (작은 값)	v_4	v_3
y_b x_a	x0 (작은 값)	x1 (큰 값)

[0081]

표 2에서 $v_r = x_a \rightsquigarrow y_b$ 이고 x_0, y_0, x_1, y_1 값은 각각 $x_0 = \mu_{K_0}(T)$, $x_1 = \mu_{K_1}(T)$, $y_0 = \mu_{K'_0}(T)$, $y_1 = \mu_{K'_1}(T)$ 과 같다. 이때 x_0, y_0, x_1, y_1 의 대소 관계는 $x_0 < x_1, y_0 < y_1$ 이라고 하고 각각 $|x_1 - x_0| = |y_1 - y_0|$ 인 관계를 갖는다고 가정한다.

[0083]

x_0 는 x_1 에 비해서 작은 값이다. 값이 작다는 의미는 T가 지식 k_0 를 설명하는데 있어서 별로 중요하지 않은 내용이거나 k_0 를 자세하게 서술한 내용이 아니거나 k_0 를 지나치게 간략하게 서술한 내용임을 의미한다. 다른 관점에서 보면 x_1 은 x_0 에 비해서 큰 값이다. 값이 크다는 의미는 설명문 T가 지식 k_1 를 설명하는데 있어서 중요한 내용이거나 k_1 에 대해 자세히 서술되거나 충분히 서술되었음을 의미한다. y_0 와 y_1 의 경우는 x_0 와 x_1 의 경우와 마찬가지로이다.

[0084]

k_i 가 k_j 에 의미적으로 포함된다면 것은 k_i 에서 설명되는 내용을 k_j 에서도 잘 설명되고, 자세히 설명되고, 넓은 범위에서 설명되고 있음을 의미한다. 의미적으로 포함되는 정도가 높을수록 제공하는 정보의 범위가 넓으므로 사용자의 질의 의도에 대한 답변일 가능성이 높아지므로 좋은 추천 지식이 될 수 있다.

[0085]

두 퍼지 멤버십 값에 대한 의미적 퍼지 함의 연산 값의 크기를 정의하기 위해서는 먼저 표 2에 있는 v_1, v_2, v_3, v_4 의 상대적 크기를 비교하여야 한다. 이를 위해서 먼저 v_1 과 v_2 의 크기를 비교하고 v_3 와 v_4 , v_2 와 v_3 들에 대해서도 각각 크기를 비교하여 전체적인 크기순서를 정한다.

[0086]

우선 v_1 과 v_2 의 크기를 비교하자. v_1 은 T가 k_1 를 설명하는데 있어서 중요한 내용인 경우에 반해서, v_2 는 k_0 을 설명하는데 있어서 덜 중요한 내용인 경우이다. 두 경우 모두 T는 k'_1 에서 중요한 내용이다. 즉 v_1 에서는 k_1 에서 중요한 내용을 k'_1 에서도 중요하게 서술하고 있고 v_2 에서는 k_0 에서 덜 중요한 내용을 k'_1 에서는 중요하게 서술하고 있는 것이다. 따라서 중요한 내용이 자세하고 중요하게 설명되는 경우의 의미적 포함 정도가 덜 중요한 내용을 자세하고 중요하게 설명하는 경우의 의미적 포함 정도 보다 더 크다고 말할 수 있다. 따라서 v_1 이 v_2 보다 T로써 의미적으로 포함되는 정도가 더 크다고 간주할 수 있다.

[0087]

그리고 v_3 와 v_4 의 크기를 비교하자. v_3 와 v_4 공히 T는 k'_0 에 대해서 간략하게 설명하고 있다. 그런데 v_3 는 T는 k_1 에서는 중요한 내용인 경우이고 v_4 는 T는 k_0 에서는 덜 중요한 내용인 경우이다. 즉 k_1 에서 중요한 내용이지만 k_0 에서는 덜 중요한 내용을 k'_0 에서 같은 강도로 설명하고 있다. 따라서 k_1 에서 중요한 내용을 k'_0 에서 의미적으로 포함하는 정도는 k_0 에서 덜 중요한 내용을 k'_0 에서 의미적으로 포함하는 정도보다 더 크다고 할 수 있다. 따라서 v_3 가 v_4 보다 T로써 의미적으로 포함되는 정도가 더 크다고 간주할 수 있다.

[0088] 끝으로 v_2 와 v_3 의 크기를 비교하자. T 는 k_0 에서 덜 중요한 내용인데 k'_1 에서는 중요한 내용으로서 서술되고 있다. 반대로 T 는 k_1 에서는 중요한 내용인데도 k'_0 에서는 중요하지 않게 서술되고 있다. 즉 처음 지식에서 중요하지 않은 내용을 다른 지식에서 중요하게 서술되면 다른 지식은 더 많은 내용을 추가적으로 포함한다고 말할 수 있다. 그러나 처음 지식에 있어서 중요한 내용인데도 불구하고 다른 지식에서 소홀하게 서술하면 다른 지식은 상대적으로 중요한 내용을 누락시킬 수 있다. 따라서 덜 중요한 내용을 중요하게 자세히 서술하는 경우의 의미적 포함 정도가 중요한 내용을 소홀히 서술하는 경우의 의미적 포함 정도 보다 더 크다고 간주할 수 있다. 따라서 v_2 가 v_3 보다 T 로써 의미적으로 포함되는 정도가 더 크다.

[0089] 상술한 내용을 요약하면 표 3과 같다.

표 3

[0090]

x_a	y_b	$v_r = x_a \rightarrow y_b$	$x_a \rightarrow y_b$ 의 의미
x_1 (큰값)	y_1 (큰값)	$v_1 = x_1 \rightarrow y_1$	T 가 k_1 을 설명하는 중요한 내용이며 k'_1 에서도 T 가 중요하고 자세히 서술된 내용이다
x_0 (작은값)	y_1 (큰값)	$v_2 = x_0 \rightarrow y_1$	T 가 k_0 에서 중요하지 않은 내용인데도 불구하고 k'_1 에서 T 가 중요하고 자세히 서술된 내용이다
x_1 (큰값)	y_0 (작은값)	$v_3 = x_1 \rightarrow y_0$	T 가 k_1 에서 중요한 내용인데도 불구하고 k'_0 에서는 T 가 중요하지 않게 간략히 서술되었다.
x_0 (작은값)	y_0 (작은값)	$v_4 = x_0 \rightarrow y_0$	T 가 k_0 에서 중요하지 않은 내용이기 때문에 k'_0 에서도 T 가 중요하지 않게 간략히 서술되었다.

[0091] 최종적으로 $v_1 > v_2$ 이고 $v_2 > v_3$ 이고 $v_3 > v_4$ 이므로 네 값의 대소 관계는 $v_1 > v_2 > v_3 > v_4$ 와 같다고 결론 낼 수 있다. 이를 의미적 퍼지 함의 연산자 대소 관계 규칙이라고 정의한다.

[0092] 정의한 의미적 퍼지 함의 연산자 대소관계 규칙을 바탕으로 의미적 포함 정도를 계산하는 방법을 정의한다. 의미적 퍼지 함의 연산자 대소 규칙을 만족하도록 의미적 퍼지 함의 연산자 크기 순서 표 M_n 을 생성한다. 여기서, n 은 구간의 개수를 의미한다.

표 4

[0093]

2	4	2	1
1	7	5	3
0	9	8	6
b a	0	1	2

[0094] 의미적 퍼지 함의 연산자 크기 순서 표를 구성하는 값은 값이 높은 순서를 의미한다. 상기 의미적 퍼지 함의 연산자 크기 순서 표의 행 a 와 열 b 는 M_n 을 계산하기 위한 식을 정의하기 위해 각각 몇 번째 구간인가를 나타내는 정수 값이다. 상기 의미적 퍼지 함의 연산자 크기 순서 표를 구성하는 각 구간은 표 4의 규칙으로 구성된 의미적 퍼지 함의 연산자 대소관계 규칙을 만족한다. 즉, 표의 모든 구간에서 (높은 값 $\rightarrow$ 높은 값) > (낮은 값 $\rightarrow$ 높은 값) 이다.

은 값) > (높은 값↗낮은 값) > (낮은 값↗낮은 값)의 순서를 만족한다.

[0095] 수 10은 크기가 n 인 의미적 퍼지 함의 연산자 크기 순서 표의 값 $M_n(a, b)$ 를 구하는 식이다.

수학식 10

[0096]
$$M_n(a, b) = \begin{cases} ((n+1)^2 + 1) - M_n^*(a, b) & \text{if } (a+b) \leq n \\ M_n^*(n-a, n-b) & \text{if } (a+b) > n \end{cases}$$

[0097] 여기서, $M_n^*(a, b)$ 은 수 11과 같이 정의한다.

수학식 11

[0098]
$$M_n^*(a, b) = \frac{(a+b)(a+b+1)}{2} + (b+1)$$

[0099] 이 표의 값들을 내림차순으로 정렬하여 0과 1 사이의 일정한 값으로 입력한 표를 $G_n(a, b)$ 이라 하면, $G_n(a, b)$ 은 수 12와 같이 계산할 수 있다.

수학식 12

[0100]
$$G_n(a, b) = \frac{(n+1)^2 - M_n(a, b)}{(n+1)^2 - 1}$$

[0101] 단순한 정수 값인 a, b 대신에 멤버십 값 x_a, y_b 을 적용하기 위해서는 행과 열을 구성하는 값들도 각각 0과 1 사이의 값으로 변환해야 한다. 입력되는 멤버십 값에 근사한 행과 열 값을 찾을 수 있으면 $x_a \rightsquigarrow y_b$ 을 표의 값으로 나타낼 수 있다. 이 표를 의미적 퍼지 함의 값의 표(n=2)라고 하고 표 5와 같이 나타낼 수 있다.

표 5

[0102]

1	0.625	0.875	1.0
0.5	0.25	0.5	0.75
0	0	0.125	0.375
$y_b \diagdown x_a$	0	0.5	1

[0103] 완성된 의미적 퍼지 함의 값의 표는 행과 열 값에 대한 표의 값이 의미적 퍼지 함의 연산자 대소관계 규칙을 만

족하면서 만족하는 0과 1 사이의 값을 일정한 차이로 구성된다. 즉, $x_i = i/n$ 이고 $y_j = j/n$ 이다. 완성된 표

H_n 의 행 y_b 와 열 x_a 에 대한 값은 $x_a \rightarrow y_b$ 로 나타낸다. 이 표의 값은 x_a 와 y_b 에 대한 의미적 퍼지 함의 연산 값이다.

[0104] $x_a \rightarrow y_b$ 의 값을 얻기 위해 x_a , y_b 값을 크기 순서 표 $M_n(a, b)$ 의 a, b 로 근사 시킨다. a, b 는 각각 0부터 n 까지의 자연수로 나타낸다.

[0105] $x_i = i/n$ 는 H_n 의 i 번째 열 값이고, x 가 $x_{i-1} + \frac{1}{2n} < x \leq x_i + \frac{1}{2n}$ 인 범위 안에 포함된다면 x 를 근사시킨다. $\frac{i-1}{n} + \frac{1}{2n} < x \leq \frac{i}{n} + \frac{1}{2n}$ 로부터 $i-1 < \frac{2nx-1}{2} \leq i$ 가 되므로 $i = \left\lceil \frac{2nx-1}{2} \right\rceil$ 이 된다. 따라서 $a = \left\lceil \frac{2nx-1}{2} \right\rceil$ 이고, 마찬가지로 $b = \left\lceil \frac{2ny-1}{2} \right\rceil$ 이다. 이 a, b 를 이용하여 값 $G_n(a, b)$ 를 계산하면 $x_a \rightarrow y_b$ 에 대한 값을 얻을 수 있다.

[0106] 이때, 퍼지 함의 값의 표 값을 얻는 이산적인 함의 연산자 함수인 $H_n(x_i, y_j)$ 를 단조 증가 형태의 함수로 정의할 수 있다. 여기에서 중요한 점은 $G_n(i, j)$ 의 절대값이 아니라 퍼지 함의 연산자 대소 관계 규칙을 만족하는지 여부이다. 단조 증가 함수는 퍼지 함의 연산자 대소 관계 규칙을 만족하기 때문에 어느 것을 사용하더라도 의미적 퍼지 관계 곱 연산에서는 큰 차이가 나지 않는다. 우리는 단순히 $H_n(x_i, y_j) = G_n(i, j)$ 를 이용한다. 따라서 $x \rightarrow y$ 에 대한 의미적 퍼지 함의 값의 표(Semantic fuzzy implication value table) 값을 계산하는 방법은 수 13과 같이 나타낸다.

수식 13

$$x \rightarrow y \approx H_n(x_a, y_b) = H_n\left(\frac{1}{n} \cdot \left\lceil \frac{2nx-1}{2} \right\rceil, \frac{1}{n} \cdot \left\lceil \frac{2ny-1}{2} \right\rceil\right) = G_n\left(\left\lceil \frac{2nx-1}{2} \right\rceil, \left\lceil \frac{2ny-1}{2} \right\rceil\right)$$

[0108] 의미적 퍼지 함의 연산의 예를 들면 표 6과 같다. 표 6은 의미적 퍼지 함의 값의 표 H_6 이다. 입력 값이 $0.285 \rightarrow 0.715$ 라고 하면 계산 과정은 다음과 같다.

표 6

6	1.000	0.563	0.688	0.792	0.875	0.938	0.979	1.000
5	0.833	0.417	0.542	0.667	0.771	0.854	0.917	0.958
4	0.667	0.292	0.396	0.521	0.646	0.750	0.833	0.896
3	0.500	0.188	0.271	0.375	0.500	0.625	0.729	0.813
2	0.333	0.104	0.167	0.250	0.354	0.479	0.604	0.708
1	0.167	0.042	0.083	0.146	0.229	0.333	0.458	0.583
0	0.000	0.000	0.021	0.063	0.125	0.208	0.313	0.438
	y_b/x_a	0.000	0.167	0.333	0.500	0.667	0.833	1.000
b/a		0	1	2	3	4	5	6

[0109]

[0110] $H_6(0.285, 0.715)$ 에서 0.285와 0.715는 각각 $\left\lceil \frac{2nx-1}{2} \right\rceil, \left\lceil \frac{2ny-1}{2} \right\rceil$ 으로 인해 2와 4로 근사할 수 있

으며 $G_6(2,4)$ 가 된다. 이 값은 $G_6(2,4) = \frac{(6+1)^2 - M_6(2,4)}{(6+1)^2 - 1}$ 이다. 이때 $2+4 \leq 6$ 이므로

$M_6(2,4) = ((6+1)^2 + 1) - \left(\frac{(2+4)(2+4+1)}{2} + (4+1) \right) = 24$ 가 되므로

$G_6(2,4) = \frac{49-24}{48} = 0.521$ 이 된다.

[0111] 의미적 퍼지 함의 연산자 값의 표를 구성하는 값은, 표의 크기 n 이 충분히 크다면 표의 크기에 큰 영향을 받지 않는다.

[0112] 다음, 의미적 퍼지관계 곱 연산자와 관련하여 상세하게 설명한다.

[0113] 지식 k_j 가 k_i 를 포함하는 정도를 계산하기 위해, 각 지식의 퍼지집합 K_j 가 K_i 를 의미적으로 포함하는 정도를 수 14를 이용하여 연산한다.

수학식 14

$$K_i \Rightarrow K_j = \frac{V}{|TS_{ij}^Y|} \sum_{T \in TS_{ij}^Y} (\mu_{K_i}(T) \rightarrow \mu_{K_j}(T))$$

[0115] 여기서, $\Rightarrow$ 는 의미적 퍼지 관계 곱 연산자를 의미한다. V 는 두 지식의 의미적 퍼지 포함 정도를 보정하는 파라

미터이고, TS_{ij}^Y 는 K_i^Y 에 속하는 T 집합을 의미하며 $TS_{ij}^Y = \{T | T \in K_i^Y \text{ or } T \in K_j^Y\}$ 로 나타낸다.

[0116] 실제 데이터에서는 멤버십 값이 큰 T 수는 적은데 반해서 멤버십 값이 작은 T 수는 매우 많다. 예를 들어 TDB에는 1000개의 T 있다고 가정하고 그 중 980개는 K_i 에 대해 작은 멤버십 값을 가지고 20개는 큰 멤버십 값을 가진다고 가정한다. 이 경우 TDB의 모든 T 를 이용하여 두 지식의 포함 정도를 계산할 경우, $\mu_{K_i}(T) \rightarrow \mu_{K_j}(T)$ 값이 [낮은값 $\rightarrow$ 낮은값]이 많으므로 지식에 있어서 중요하지 않고 잘 설명하지 않는 T 에 의해 전체 포함 정도 값이 좌우되는 문제가 생긴다. 이를 위해 멤버십 정도 값이 사용자가 지정하는 Y 보다 큰 값을 갖는 T 들로 의미적 포함 정도를 계산한다. 이 기법을 $Y-cut$ 이라 정의한다. 수 15는 수 14에 사용된 TS_{ij}^Y 의 정의이다.

수학식 15

$$TS_{ij}^Y = \{T | T \in K_i^Y \text{ or } T \in K_j^Y\}$$

[0118] 여기서, K^γ 는 수 16과 같다.

수학식 16

[0119]
$$K^\gamma = \{T, \mu_K(T) | T \in TDB, \mu_K(T) > \gamma\}$$

[0120] 두 지식의 포함 정도를 계산하기 위해 멤버쉽 정도 값이 γ 보다 큰 값만을 이용하여 중요한 T 들로 의미적 퍼지 포함 정도를 계산함으로써 전체 포함 정도값의 정확도가 향상된다.

[0121] V 는 의미적 퍼지 포함 정도 값을 보정하는 파라미터로써 수 17과 같이 정의된다.

수학식 17

[0122]
$$V = \frac{V_B(1 + \delta V_C)}{(1 + V_A)}$$

[0123] (단, $0 \leq \delta \leq 1$)

[0124] 두 퍼지 지식의 크기와 포함 정도가 변함에 따라 의미적 포함 정도 값의 변화를 보정한다. 도 7은 두 지식 K_i 와 K_j 가 의미적 포함 관계를 가질 때 K_i 는 A^γ 와 B^γ 로 구성되고, K_j 는 B^γ 와 C^γ 로 구성됨을 나타낸다.

[0125] 도 7에서 $A^\gamma, B^\gamma, C^\gamma$ 는 크리스프 집합이고 A^γ 는 K_i^γ 에 속하는 T 들 중에서 K_j^γ 에 속하지 않는 T 들의 집합이고, B^γ 는 K_i^γ 와 K_j^γ 에 공통으로 속하는 T 들의 집합이고, C^γ 는 K_j^γ 에 속하는 T 들 중에서 K_i^γ 에 속하지 않는 T 들의 집합이다. 특히, K_i^γ 에 속한 T 들과 K_j^γ 에 속한 T 들은 서로 유사한 내용이기 때문에 B^γ 는 K_i^γ 와 K_j^γ 의 유사성을 매개시켜주는 역할을 한다. 조건 $A^\gamma, B^\gamma, C^\gamma$ 를 시각적으로 도시하면 도 8과 같다.

[0126] 각 규칙을 두 지식에 대한 크리스프 집합으로 표현하면 다음 수 18과 같다

수학식 18

[0127]
$$A^\gamma = \{T | T \in K_i^\gamma \text{ and } T \notin K_j^\gamma\}$$

[0128]
$$B^\gamma = \{T | T \in K_i^\gamma \text{ and } T \in K_j^\gamma\}$$

[0129]
$$C^\gamma = \{T | T \notin K_i^\gamma \text{ and } T \in K_j^\gamma\}$$

[0130] 두 퍼지 지식의 크기와 포함 정도가 변함에 따라 의미적 포함 정도 값이 변하는 경우를 규칙으로 정리하면 표 7과 같다.

표 7

[0131]

규칙	내용
Rule 1	A^y 의 범위가 축소될수록, B^y 의 범위가 넓어질수록 K_i^y 가 K_j^y 를 의미적으로 포함하는 정도가 커진다
Rule 2	B^y 의 범위는 공집합이 아니고 C^y 의 범위가 넓어질수록 K_i^y 가 K_j^y 를 의미적으로 포함하는 정도가 커진다
Rule 3	B^y 의 범위는 공집합이면 C^y 의 범위가 크더라도 K_i^y 가 K_j^y 를 의미적으로 포함하는 정도가 크다고 할 수 없다

[0132]

표 7의 각 규칙을 시각적으로 나타내면 도 9a 내지 도 9c와 같다. 도 9a 내지 도 9c에서 점은 문서 T 들을 의미한다.

[0133]

Rule 1은 도 9a와 같이 A^y 가 작아질수록, B^y 가 커질수록 K_j 가 K_i 를 의미적으로 포함하는 정도가 커짐을 의미한다. K_i 에 속한 T 들은 서로 유사한 내용이고 K_j 에 속한 T 들은 서로 유사한 내용이다. 이때 중간 매개 역할을 하는 T 에 의해 K_j 의 범위가 커질수록 의미적으로 포함하는 정도가 커진다.

[0134]

Rule 2은 도 9b와 같이 C^y 가 커질수록 K_j 는 K_i 를 의미적으로 포함하는 정도가 커짐을 의미한다. K_i 에 속한 T 들은 서로 유사한 내용이고 K_j 에 속한 T 들은 서로 유사한 내용이다. 이때 중간 매개 역할을 하는 T 가 많아지기 때문에 K_j 의 범위가 커질수록 의미적으로 포함하는 정도가 커진다.

[0135]

Rule 3는 도 9c에 도시된 바와 같이 B^y 가 매우 작은 값이면 아무리 C^y 의 범위가 넓다 하더라도 공통되는 T 를 가지고 있지 않기 때문에 K_i 와의 의미적 포함 관계를 고려하는 것은 무의미함을 의미한다.

[0136]

이 규칙을 만족하도록 수 17과 같이 파라미터를 정의한다. Rule 1에 의해 V_A 가 작아지고 V_B 가 커질수록 V 값이 커지도록 V_A 를 분모에, V_B 를 분자에 놓는다. Rule 2에 의해 V_C 가 커질수록 V 값이 커지도록 분자에 놓는다. 이때 Rule 3에 의해 V_B 를 반영하도록 식 V 를 구성한다

[0137]

이때, 크리스프 집합에서 V_A, V_B, V_C 는 수 19와 같이 나타낼 수 있다.

수학식 19

[0138]

$$V_A = |A^y|$$

[0139]

$$V_B = |B^y|$$

[0140]

$$V_C = |C^y|$$

[0141] 수 19의 조건 A^Y, B^Y, C^Y 를 퍼지 개념으로 확장하면 수 20과 같이 나타낸다.

수학식 20

[0142]
$$A = \{(T, \mu_A(T)) | T \in TS_{ij}^Y, \mu_A(T) = \min(S_{-w}(a), S_{+w}(b)) \cdot \sqrt{S_{-w}(a) \cdot S_{+w}(b)}\}$$

[0143]
$$B = \{(T, \mu_B(T)) | T \in TS_{ij}^Y, \mu_B(T) = \min(S_{-w}(a), S_{-w}(b)) \cdot \sqrt{S_{-w}(a) \cdot S_{-w}(b)}\}$$

[0144]
$$C = \{(T, \mu_C(T)) | T \in TS_{ij}^Y, \mu_C(T) = \min(S_{+w}(a), S_{-w}(b)) \cdot \sqrt{S_{+w}(a) \cdot S_{-w}(b)}\}$$

[0145] 여기서, $a = f_n(\mu_{K_i}(T)), b = f_n(\mu_{K_j}(T))$ 이다. A 는 a 값이 커질수록, 그리고 b 값이 작아질수록 $\mu_A(T)$ 값이 커진다. B 는 a 값과 b 값이 커질수록 $\mu_B(T)$ 값이 커진다. C 는 a 값이 작아질수록, b 값이 커질수록 $\mu_C(T)$ 값이 커진다.

[0146] $f_n(x)$ 은 $\mu_{K_i}(T)$ 와 $\mu_{K_j}(T)$ 를 각각 -1과 1사이의 퍼지 값으로 변환하는 식으로, 수 21과 같이 나타낸다.

수학식 21

[0147]
$$f_n(x) = \begin{cases} n \frac{1-2\gamma}{-\gamma(1-\gamma)} (x-\gamma) \left(x - \frac{1-\gamma}{1-2\gamma}\right) & (\text{if } \gamma \neq 0.5) \\ n(2x-1) & (\text{if } \gamma = 0.5) \end{cases}$$

[0148] (단, $0 \leq x \leq 1$)

[0149] $f_n(x)$ 값은 각각 1과 γ 와 0일때 $f_n(1) = n, \quad f_n(\gamma) = 0, \quad f_n(0) = -n$ 값을 갖는다.

$S_{\pm w}(f)$ 은 $f_n(x)$ 값이 Rule 1,2,3을 만족시키도록 0과 1 사이의 값으로 변환하는 식으로 수 22와 같이 정의한

다. 이때, $f = f_n(\mu_{K_i}(T))$ or $f_n(\mu_{K_j}(T))$ 이다.

수학식 22

[0150]
$$S_{-w}(f) = \frac{1}{1 + e^{-wf}} \quad S_{+w}(f) = \frac{1}{1 + e^{+wf}}$$

[0151] $\mu_A(T), \mu_B(T), \mu_C(T)$ 는 퍼지 집합으로 각각 다음과 같은 의미를 갖는다. 도 10에 도시된 바와 같이, $\mu_A(T)$ 는 $\mu_{K_i}(T)$ 가 큰 값이고 $\mu_{K_j}(T)$ 가 작은 값일 때 큰 값이다. 여기서, A^Y 는 큰 값을 가지고 B^Y, C^Y 는 낮은 값을 가져야 한다.

[0152] 도 11은 $\mu_A(T) = \min(S_{-w}(a), S_{+w}(b)) \cdot \sqrt{S_{-w}(a) \cdot S_{+w}(b)}$ 을 계산하는 과정을 나타낸 도면이다.

[0153] 도 11의 하단에서 $\mu_{K_i}(T)$ 와 $\mu_{K_j}(T)$ 는 퍼지 지식에 대한 T 의 멤버쉽 정도 값이다. 상기 퍼지 지식에 대한 T 의 멤버쉽 정도 값을 수 23와 같이 함수 $f_n(x)$ 로 n 부터 $-n$ 사이의 값으로 변환하고 $S_{\pm w}(f)$ 로 0과 1 사이의 값으로 나타낸다.

수학식 23

[0154] $f_n: (0,1) \rightarrow (-n, n)$

[0155] $S_{\pm w}: (-n, n) \rightarrow (0,1)$

[0156] 도 11의 상단에서 a 와 b 는 도 10의 규칙에 대응하여 각각 낮은값과 높은 값을 이용하여 계산되는 과정을 보인다. 예를 들어, $\mu_A(T_A)$ 는 도 11에 의하면 a 값이 크고 b 값이 작을 때 계산 결과는 큰 값으로 나타남을 의미한다. $\min(S_{-w}(a), S_{+w}(b))$ 은 어느 한 값이 낮은 값일 경우 전체 계산 결과가 낮아지도록 설정한 부분이다.

[0157] 이를 바탕으로 퍼지 집합에서 V_A, V_B, V_C 는 수 24와 같이 나타낼 수 있다.

수학식 24

[0158]
$$V_A = \sum_{T \in A} \mu_A(T)$$

[0159]
$$V_B = \sum_{T \in B} \mu_B(T)$$

[0160]
$$V_C = \sum_{T \in C} \mu_C(T)$$

[0161] 같은 방법으로 $\mu_B(T)$ 는 $\mu_{K_i}(T)$ 가 큰 값이고 $\mu_{K_j}(T)$ 도 큰 값일 때 큰 값이다. 그리고 $\mu_C(T)$ 는 $\mu_{K_i}(T)$ 가 작은 값이고 $\mu_{K_j}(T)$ 가 큰 값일 때 큰 값이다.

[0162] 의미적 퍼지 관계 연산 값 $K_i \Rightarrow K_j$ 이 클수록 K_i 가 K_j 에 의미적으로 포함되는 정도가 크다는 것을 의미하며, 이는 지식 k_i 가 다루는 내용을 k_j 가 의미적으로 더 잘 설명하고, k_j 가 k_i 보다 다루는 내용을 더 자세히 설명하는 것을 의미한다. 즉 $K_i \Rightarrow K_j$ 이 큰 값이면 k_j 는 k_i 에 있어서 좋은 추천 지식이다.

[0163] 이하, 의미적 포함 정도의 계산과 관련하여 상세하게 설명한다.

[0164] 질의가 입력될 때 마다 KDB 의 모든 지식들에 대한 의미적 포함 정도를 계산하면 검색 시 많은 시간이 소요된다. KDB 에 저장되어 있는 퍼지 지식의 멤버십 정도는 사용자가 질의를 입력하는 것과 무관하게 값이 변하지 않는다. 따라서 새로운 질의가 입력되면 그 질의에 대한 멤버십 정도만 계산하고 기존에 KDB 에 저장된 퍼지 지식을 미리 계산해두면 연산 비용을 줄일 수 있다.

[0165] KDB 에 존재하는 모든 지식에 대한 멤버들의 멤버십 값을 이용하기 위해 퍼지 집합 행렬을 만들 수 있다. 모든 T^r 에 대해 $\mu_{K_i}(T^r) \rightarrow \mu_{K_j}(T^r)$ 연산을 수행하는 것은 의미적 퍼지 관계 곱 연산 $R \Rightarrow R^T$ 이다. 이렇게 만들어진 행렬을 의미적 포함 관계 행렬이라고 하고 L 로 표기한다.

[0166] 예를 들어, 시스템에 산에 대한 지식 k_1, k_2 와, 자동차에 대한 지식 k_3, k_4 와, 산길 교통사고에 대한 지식 k_5, k_6 가 저장되고, 각각 두 개의 응답 A 들을 가진 질의인 경우,

[0167] 지식 $k_1 = \{Q_1, A_{11}, A_{12}\}$, $k_2 = \{Q_2, A_{21}, A_{22}\}$, ..., $k_6 = \{Q_6, A_{61}, A_{62}\}$ 의 퍼지 지식 행렬이 표 8과 같다고 가정한다.

표 8

	T^1	T^2	T^3	T^4	T^5	T^6	T^7	T^8	T^9	T^{10}	T^{11}	T^{12}
K_1	0.9	0.8	0.8	0.9	0.1	0.2	0.2	0.2	0.6	0.7	0.6	0.1
K_2	0.8	0.9	0.9	0.8	0.1	0.1	0.1	0.2	0.5	0.6	0.5	0.4
K_3	0.1	0.1	0.2	0.2	0.8	0.9	0.8	0.8	0.6	0.5	0.6	0.1
K_4	0.1	0.2	0.1	0.1	0.9	0.9	0.8	0.9	0.7	0.6	0.5	0.4
K_5	0.1	0.5	0.6	0.5	0.1	0.4	0.6	0.5	0.7	0.8	0.7	0.9
K_6	0.2	0.5	0.6	0.4	0.2	0.5	0.6	0.6	0.9	0.8	0.8	0.8

[0168]

[0169] 표 8를 참조하면, 지식 K_1 과 다른 문서 T 들의 멤버십 정도(유사도) 값은 다음과 같이 해석할 수 있다. $T^1 \sim T^4$ 까지는 산에 대한 질의응답 쌍이기 때문에 K_1 에 대한 멤버십 정도 값이 크다 (1에 가까운 값). $T^5 \sim T^8$ 까지는 산과 관계 없는 자동차에 대한 내용이기 때문에 멤버십 정도 값이 작다(0에 가까운 값). $T^9 \sim T^{12}$ 까지는 산길 교통사고에 대한 질의응답 쌍이기 때문에 산에 대한 내용과 자동차에 대한 내용을 어느 정도 다룬다고 볼 수 있기 때문에 멤버십 값이 $T^1 \sim T^4$ 보다는 작지만 $T^5 \sim T^8$ 보다는 큰 값으로 나타난다.

[0170] 완성된 퍼지 지식 행렬과 그 전치 행렬을 이용하여 의미적 퍼지 관계 행렬을 도 12와 같이 계산한다

[0171] 완성된 퍼지 관계 곱 행렬에서 지식 K_1 과 의미적 포함 정도가 높은 순서대로 나열하면 $K_6 > K_2 > K_5 > K_4 > K_3$ 이다. 실제로 산에 대한 내용을 다루는 지식들(K_2, K_5, K_6)이 그렇지 않은 지식들(K_3, K_4)보다 먼저 추천된다. 지식 K_1 에 대해서 산과 유사한 정도가 높은 문서(질의응답 쌍) $T^1, T^2, T^3, T^4, T^9, T^{10}, T^{11}, T^{12}$ 들을 이용하여 K_1, K_2, K_5, K_6 을 추천 지식으로 얻을 수 있다. 즉 K 와 T 의 유사도를 이용하여 K 에 대한 T 의 멤버십 정도를 계산하고, 두 퍼지 지식 K_1 과 K_1, K_2, K_5, K_6 의

의미적 퍼지 포함 정도를 의미적 퍼지 관계 곱으로 계산한다.

[0172] 사용자의 질의에 대한 응답을 평가하는 가장 좋은 방법은 질의를 입력한 사용자가 응답 내용에 만족한 정도로 평가하는 것이다. 하지만 응답 내용에 만족하는 정도는 사람에 따라 다르다. 특히 객관적인 응답을 한정할 수 없는 질의일수록 사용자에게 따라 응답의 기준이 달라진다. 따라서, TREC(Text REtrieval Conference)에서는 지식 검색 시스템을 평가하기 위해 표준 데이터 집합을 설정하였다. 이 데이터는 미리 정의한 질의들과 문헌 집합(AQUAINT-2 등)에서 전문가들이 질의에 대한 응답이 존재하는 것으로 판단한 문서들의 집합이다.

[0173] 지식 검색 시스템의 성능은 이 표준 데이터 집합을 이용하여 평가할 수 있다. 즉 지정된 질의를 입력하여 시스템이 찾아낸 응답이 전문가들이 좋은 응답으로 판단한 문장을 문헌들에서 잘 찾아낼수록 좋은 성능을 보이는 것으로 평가할 수 있다. 하지만 TREC 표준 데이터는 질의의 수가 적고(440여개) 질의에 대한 응답이 아닌 문서들의 집합이기 때문에 본 발명에서 제안하는 방법을 평가하는데 이용하기에는 적합하지 않다.

[0174] 본 발명에서 사용된 실험 데이터는 Yahoo! Answers에서 WebCrawler 프로그램을 사용하여 수집한 16,000여 개의 지식들을 KDB에 저장하여 이용하였다. 수집된 데이터는 크게 랜드마크(Landmark), 월드컵(Worldcup), 여행(Tour), 자동차(Car), 모바일(Mobile)의 다섯 가지 Scope 으로 분류하고 각각 관련된 지식들을 DB로 구축하였다. 각 DB는 다음 표 9와 같은 관련 내용을 갖는다.

표 9

Scope	관련 내용
Landmark	관광지, 거주지, 환경 등에 대한 지식
Tour	비자, 환전, 호텔 투숙 등에 대한 지식
Worldcup	월드컵, 축구팀, 축구선수 등에 대한 지식
Car	자동차 수리, 보험, 면허 등에 대한 지식
Mobile	3G, 4G, 아이폰, 갤럭시폰 등에 대한 지식

[0176] 이 중에서 Landmark와 Tour에 저장된 지식들은 상대적으로 지식의 범위가 넓고 Worldcup, Car, 그리고 Mobile은 상대적으로 지식의 범위가 좁다. 지식의 범위가 넓다는 의미는 지식들이 다루는 주제가 다양하다는 뜻으로 특정 주제에 대한 지식의 수는 적다고 볼 수 있다. 반대로 지식의 범위가 좁다는 의미는 지식들이 다루는 주제가 한정되어 있다는 뜻으로 이는 특정 주제에 대한 지식의 수가 많다고 볼 수 있다.

[0177] 본 발명에서 제안하는 방법의 성능을 평가하기 위해 각각의 방법으로 추천된 질의를 평가하여 비교하였다. 사용자는 추천되는 지식 중에서 만족도가 높을 것으로 여겨지는 지식을 찾는다. 만족도가 높은 지식이란, 질의를 입력한 사용자가 찾고자 하는 정보를 얻을 수 있을 것으로 판단되는 지식들을 의미한다. 각 방법에 대한 평가는 추천된 지식들 중에서 사용자를 만족시키는 정도가 높은 지식에 대해 다음 기준과 같이 점수를 부여한다. 본 발명에서는 지식에 부여된 점수의 합을 이용하여 평가하였다. 지식 평가를 위한 배점 기준은 표 10과 같다.

표 10

판정기준	점수
지식의 만족도가 높다. 지식이 원하는 정보에 근접한 내용을 가진다. 추가적인 검색을 통해 또다른 연관 내용을 찾을 수 있을 것으로 판단된다.	2점
지식의 만족도가 낮다. 추가적인 검색을 통해 원하는 정보에 보다 근접한 내용을 찾을 수 있을 것으로 판단된다.	1점
지식의 만족도가 매우 낮다. 질의와 전혀 무관한 지식이다. 추가적인 검색을 시도할 필요가 없다고 판단된다.	0점

[0179] 같은 질의를 입력하더라도 질의에 숨겨진 의도는 사용자에게 따라 그 의도가 달라질 수 있다. 그렇기 때문에 추천된 지식이 사용자의 질의에 연관되는지 여부를 전문가나 실험 참여자가 직접 판정해야 한다.

[0180] 본 발명에서 수행한 실험 순서는 다음과 같다. 지정된 질의를 입력하여 질의에 대해서 각각의 방법으로 추천된

지식들을 정리한다. 실험 참여자들은 추천된 각 지식의 내용을 참조하여 질의 의도에 부합하거나, 어느 정도 부합하는 지식에 대한 점수를 부여하고, 그 점수의 합으로 지식을 평가한다.

[0181] 정보 검색 분야에서 널리 사용되는 평가 방법으로 정확도(Precision)/재현율(Recall)이 있다. 하지만, 본 발명의 실험에서 재현율을 이용한 평가 방법은 사용하기 어렵다. 재현율은 시스템에 몇 개의 정답이 있다는 것을 알고 있고 그 정답이 얼마나 검색되었는지를 평가한다. 하지만 사용자의 의도를 만족하는 지식이 시스템에 얼마나 있는지는 확인할 수 없다.

[0182] 정확도는 검색된 결과에서 정답이 몇 개나 검색되었는지를 평가한다. 본 발명의 실험 방법은 정확도로 볼 수 있다. 이때 제시된 추천 지식 목록에서 사용자의 의도를 만족하는 지식이 나타나는 순서는 고려하지 않는다. 추천 지식 목록은 사용자 질의의 퍼지 지식을 의미적으로 포함하는 정도가 큰 지식들이다. 하지만 반드시 포함하는 정도가 클 수록 사용자를 만족시키는 정도가 높은 것은 아니다.

[0183] 예를 들어, 사용자가 교통 사고 후 차량 수리에 관련된 정보를 얻기 위해, "교통 사고가 났습니다. 어떻게 해야 하나요?" 질의를 입력했다고 가정하자. 제안하는 시스템은 사용자 질의에 대한 추천 질의는 교통 사고 전반에 대해 잘 설명되는 질의를 우선적으로 추천한다. 하지만 교통 사고 전반에 대한 내용은 부분적으로 설명하고 보험에 대해 구체적으로 설명한 질의는 조금 낮은 순위에 추천한다. 하지만 사용자는 낮은 순위에 있는 질의에 대해 만족하는 정도가 더 높을 수 있다. 그래서 질의가 나타나는 순서는 고려하지 않고 평가하였다.

[0184] 본 발명의 성능을 평가하기 위해 사용한 질의는 표 11과 같다. 제안하는 방법의 성능을 평가하기 위해 의미적 퍼지 함의 연산자를 이용한 의미적 퍼지 관계 곱 추천 방법이 다른 유사 질의 검색 방법이나 기존의 퍼지 함의 연산자를 이용한 퍼지 관계 곱 추천 방법에 비해 우수함을 보이고, 추천 과정이 반복될수록 추천되는 질의 중에서 사용자의 의도를 만족시키는 질의가 많아짐을 보인다.

표 11

[0185]

Landmark	Where is the favorite date course? What's the nicest place in USA you ever visited? What is a key to the city worth? Looking for a live in canada The information of school in US
Tour	How do you purchase airplane ticket for someone else Can I cancel a reservation How much money to take to travel around china How can I renew my tourist visa Does anyone a good hotel for faimlies in turkey How to change an airline reservation to a later time
Worldcup	What happens if there is a tie for the Golden Shoe award for player What is difference between football vs soccer What is the most successful club in champions league How many clubs in the UEFA Who deserves the Yashin award in 2010 Who is the most expensive soccer player in this season What is the most expensive clubs in this season When did start Women's worldcup The career of Sir alex ferguson the Manchester United manager
Car	How to repair a flat tire How can I get the driving licence What to do after a car accident The history of driver insurance What to look for when buying a used car What is the minimum age for obtaining a driver's licence

Mobile	Operating System of android phone How can I get the smart phone Usage of Mobile phone What is the recent version of iphone What is different between 4g phone and previous phone How can I upload my application to appstore
--------	---

[0186] 본 발명에서는 의미적으로 포함하는 정도가 높은 질의를 추천하기 위해 의미적 퍼지 함의 연산자를 이용한 의미적 퍼지 관계 곱 추천 방법을 제안하였다(FRP-SFIO). 제안하는 방법의 우수성을 평가하기 위해 벡터 공간에서의 코사인 유사도를 이용한 질의 검색, Liu(M. Liu, Y. Liu, Q. Yang, "Searching semantically similar questions from a large community based question archive," international conference on Natural Language Processing and Knowledge Engineering, pp. 1-8, 2009)가 제안하는 질의 확장 방법, 다른 퍼지 함의 연산자들을 이용한 퍼지 관계 곱 추천 방법과 제안하는 방법을 비교하였다.

[0187] Liu의 방법은 지식을 추천하는 방법은 아니지만 질의의 의미를 확장하여 검색한다는 점에서 본 발명에서 제안하는 방법과 어느 정도 유사점이 있기 때문에 비교 실험에 포함시켰다. 표 12는 다른 추천 방법과 비교하는 실험에 사용된 추천 방법들을 나타낸다.

표 12

약칭	추천 방법
Kernel	Liu가 제안한 방법
FRP-KD	퍼지 관계 곱-Kleene-Dienes퍼지 함의 연산자
FRP-SFIO	퍼지 관계 곱-의미적 퍼지 함의 연산자

[0189] 각 Scope별 평가 결과는 도 13과 같고 각각의 실험 통계량은 표 13과 같다. 실험 결과 상대적으로 지식의 범위가 넓은 Landmark와 Tour는 만족도가 높은 지식이 적게 발견되고 지식의 범위가 작은 Worldcup, Car, Mobile에서는 비교적 많이 발견되었다.

표 13

Scope	통계값	Kernel	FRP-KD	FRP-SFIO
Landmark	평균	3.34	3.75	5.48
	표준편차	1.84	0.35	1.13
	분산	3.38	3.13	1.29
Tour	평균	2.13	3.08	5.25
	표준편차	1.02	1.02	1.14
	분산	1.05	1.04	1.31
Worldcup	평균	8.78	10.67	12.67
	표준편차	4.51	1.58	2.29
	분산	20.32	2.50	5.24
Car	평균	7.83	10.53	13.03
	표준편차	3.31	2.60	2.28
	분산	10.97	6.75	5.18
Mobile	평균	7.33	11.33	14.05
	표준편차	0.82	2.80	2.77
	분산	0.67	7.87	7.69

[0191] 전체적으로 기존의 Liu의 방법에 비해 퍼지 관계 곱 연산을 이용하여 내용을 고려한 지식 추천 방법의 성능이 좋게 나타났다. 그리고 다른 퍼지 함의 연산자를 이용한 방법에 비해 제안하는 의미적 퍼지 함의 연산자를 이용한 방법의 성능이 좋게 나타났다.

[0192] Liu의 방법이 결과가 좋지 않은 이유는 좋은 추천 지식을 얻기에는 사용자의 질의가 다루는 내용과 비슷한 질의의 개수가 부족하여 질의의 확장이 잘 이루어지지 않았기 때문으로 추정된다. 기존의 퍼지 함의 연산자들은 지식 설명의 멤버십 정도 값이 낮은 데이터들 사이의 퍼지 함의 연산값이 높은 것으로 계산한다. 그런데 실제 데

이터에서는 대체로 멤버쉽 정도 값이 낮은 데이터가 높은 데이터에 비해 많이 나타난다. 그러므로 전체 퍼지 함의 연산 값의 평균으로 계산되는 퍼지 관계 곱 연산에서 멤버쉽 정도 값이 낮은 데이터의 영향력이 높아지므로 제안하는 방법에 비해 좋지 않은 결과를 보이는 것으로 추정된다.

[0193] 본 발명에서 제안하는 방법의 우수성을 입증하기 위해 검정을 통해 기존 방법들과 비교한다. 가설 검정은 모집단의 분포의 모양이나 모수 등에 대한 가설을 세우고, 모집단에서 추출한 표본에 기초하여 가설의 채택이나 기각을 결정하는 통계적 기법이다. 즉 모집단에서 추출한 표본의 데이터(표본의 수, 평균, 분산)를 기초로 하여 어떤 주장에 대한 타당성 여부를 결정한다. 통계적 가설검정의 결과는 귀무 가설을 기각시키지 못하거나 귀무 가설을 기각시키는 둘 중 하나이다. 귀무 가설이 참일 때 이를 기각하고 대립가설을 채택하는 오류를 범할 확률의 허용 한계를 유의수준이라고 한다. 유의수준 이내에서 귀무 가설을 기각시킬 수 있다면 그만큼 대립가설이 타당하다고 결론내릴 수 있다. 표본의 수가 많을 때에는 z-검정을 이용하고 표본의 수가 적을 때에는 t-검정을 이용한다.

[0194] 예를 들어, Worldcup Scope 에 대한 실험 결과 FRP-KD방법과 FRP-SFIO 방법을 비교하여 검정하는 방법은 다음과 같다. 표 14는 FRP-KD방법과 FRP-SFIO방법을 비교한 결과와 실험값을 나타낸다. 검정을 위해 귀무 가설 $H_0 : \mu_1 \leq \mu_2$ 를 "FRP-SFIO방법이 FRP-KD방법보다 성능이 우수하지 않다"고 정하고 대립 가설 $H_1 : \mu_1 > \mu_2$ 를 "FRP-SFIO방법이 FRP-KD방법보다 성능이 우수하다"고 하자. 모평균 μ_1 은 FRP-SFIO방법으로 찾아낸 추천 지식들 중 사용자를 만족시키는 지식의 개수의 평균이고 μ_2 는 FRP-KD방법으로 찾아낸 추천 지식들 중 사용자를 만족시키는 지식의 개수의 평균이다.

표 14

[0195]

모평균 μ	평균 $\bar{X}$	표준편차 S	분산 S^2	표본크기 n
$\mu_1 = \text{FRP-SFIO}$	$X_1 = 12.67$	$S_1 = 2.29$	$S_1^2 = 5.24$	$n_1 = 9$
$\mu_2 = \text{FRP-KD}$	$X_2 = 10.67$	$S_2 = 1.58$	$S_2^2 = 2.49$	$n_2 = 9$

[0196] 표 14에서 FRP-SFIO방법의 평균은 $X_1 = 12.67$ 이고 FRP-KD방법의 평균은 $X_2 = 10.67$ 이다. FRP-SFIO방법의 표준 편차는 $S_1 = 2.29$ 이고 FRP-KD방법의 표준 편차는 $S_2 = 1.58$ 이다. 표본의 크기는 $n_1 = n_2 = 9$ 로 같다.

[0197] 검정통계량 t 는 다음 수 25와 같이 계산된다.

수학식 25

$$t = \frac{(X_1 - X_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} = 2.16$$

[0198]

[0199] 자유도 f 는 수 26과 같이 정의된다.

수학식 26

$$f = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left(\frac{S_1^2}{n_1} \right)^2}{(n_1 - 1)} + \frac{\left(\frac{S_2^2}{n_2} \right)^2}{(n_2 - 1)}} = 14.22$$

[0200]

[0201] 귀무가설 H_0 을 기각하고 대립가설 H_1 를 수락하려면 $t > t_{\alpha}(f)$ 이어야 한다.
 $2.16 > t_{0.05}(14) = 1.76$ 이므로 H_0 을 기각하고 H_1 를 수락한다. 표 15는 FRP-SFIO 방법과 FRP-KD 방법의 t-검정을 비교한 표이다.

표 15

[0202]

귀무가설 H_0	대립가설 H_1	검정 통계량 t	자유도 f	$t_{0.05}$	$t \geq t_{\alpha}$
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$t = 2.16$	$f = 14.22$	1.76	H_0 기각

[0203]

즉 FRP-SFIO 방법은 FRP-KD 방법에 비해 유의수준 5% 이내에서 우수하다고 결론내릴 수 있다. 같은 방법으로 각 각의 Scope에 대하여 다른 지식 추천 방법과 제안하는 방법을 비교한 결과는 표 16과 표 17과 같다.

[0204]

검정 결과 제안하는 방법이 기존 방법에 비해 대체로 유의수준 5% 이내에서 우수하다고 결론내릴 수 있다. Car와 Mobile Scope에서는 일부 우수하다고 결론내릴 수 없는 결과가 나타났다. 이는 지식의 범위가 좁기 때문에 데이터 집합에서 만족도가 높은 지식의 수가 많고, 비슷한 지식들이 서로 연관된 정도가 높기 때문에 각 방법으로 추천된 결과가 서로 비슷할 것이기 때문으로 추정된다.

표 16

[0205]

	Landmark	Tour	Worldcup	Car	Mobile
μ_1	FRP-SFIO	FRP-SFIO	FRP-SFIO	FRP-SFIO	FRP-SFIO
μ_2	Kernel	Kernel	Kernel	Kernel	Kernel
X_1	5.48	5.25	12.67	13.03	14.05
X_2	3.34	2.13	8.78	7.83	7.33
S_1	1.13	1.14	2.29	2.28	2.77
S_2	1.84	1.02	4.51	3.31	0.82
n_1	5	6	9	6	6
n_2	5	6	9	6	6
t	2.22	4.97	2.31	3.17	5.69
f	6.66	9.88	11.87	8.86	5.86
$t_{0.05}(f)$	1.94	1.81	1.78	1.81	1.94
검정	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다

표 17

[0206]

	Landmark	Tour	Worldcup	Car	Mobile
μ_1	FRP-SFIO	FRP-SFIO	FRP-SFIO	FRP-SFIO	FRP-SFIO
μ_2	FRP-KD	FRP-KD	FRP-KD	FRP-KD	FRP-KD
X_1	5.48	5.25	12.67	13.03	14.05
X_2	3.75	3.08	10.67	10.53	11.33
S_1	1.13	1.14	2.29	2.28	2.77
S_2	0.35	1.02	1.58	2.6	2.8
n_1	5	6	9	6	6
n_2	5	6	9	6	6
t	3.26	3.46	2.16	1.77	1.69
f	4.77	9.87	14.22	9.83	10
$t_{0.05}(f)$	2.02	1.81	1.76	1.81	1.81
검정	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각하지 못한다	귀무가설을 기각하지 못한다

[0207]

모든 Scope의 질의들을 이용하여 검증한 결과를 나타내며 제안하는 방법이 기존 방법들에 비해 유의수준 5% 이내에서 우수하다고 결론내릴 수 있다.

[0208]

본 발명에서 제안하는 추천 방법은 추천 과정을 반복할수록 추천되는 질의의 성능이 개선된다. 이는 이전 단계의 질의를 의미적으로 포함하는 정도가 큰 질의를 추천하기 때문이다. 이를 입증하기 위한 실험을 다음과 같이 수행하였다.

[0209]

첫 번째 단계에서는 처음 사용자가 입력한 질의에 대한 추천 지식들을 평가하고 두 번째 단계에서는 그 중에서 사용자가 선택한 지식에 대한 추천 지식들을 평가한다. 본 절에서는 두 단계의 추천 과정만 비교하였다. 이는 세 단계 이상으로 추천 과정을 반복할 경우 이전 단계에서 추천된 지식이 다시 추천되는 경우가 많다. 사용자는 만족할 확률이 높은 지식을 주관적으로 선택하기 때문에 이전 단계와 같은 질의를 선택하게 되므로 같은 추천 지식을 반복하여 선택할 수도 있기 때문이다.

[0210]

각 Scope별 평가 결과는 도 14와 같고, 각각 실험 통계량은 표 18과 같다. 실험 결과 첫 번째 단계에서 추천된 지식보다 두 번째 단계에서 추천된 지식이 전체적으로 점수가 높게 나타남을 보인다.

표 18

[0211]

Scope	통계값	FRP-KD1	FRP-KD2	FRP-SFIO1	FRP-SFIO2
Landmark	평균	3.75	6.91	5.48	9.24
	표준편차	0.35	0.79	1.13	1.02
	분산	0.13	0.63	1.29	1.04
Tour	평균	10.67	12.85	12.67	16.66
	표준편차	1.58	2.81	2.29	3.19
	분산	2.50	7.92	5.24	10.18
Worldcup	평균	3.08	5.04	5.25	7.65
	표준편차	1.02	1.20	1.14	1.65
	분산	1.04	1.44	1.31	2.72
Car	평균	10.53	14.37	13.03	15.68
	표준편차	2.60	3.69	2.28	3.55
	분산	6.75	13.59	5.18	12.57
Mobile	평균	11.33	14.22	14.05	18.17
	표준편차	2.80	3.02	2.77	3.49
	분산	7.87	9.12	7.69	12.17

[0212] 표 18에서 FRP-KD1은 퍼지 관계 곱 연산자에서 Kleene-Dienes 퍼지 합의 연산자를 이용한 첫번째 단계의 추천 지식을 의미하고 FRP-KD2는 두번째 단계의 추천 지식을 의미한다. 두 단계의 실험 결과를 비교한 이유는 세 단계 이상의 추천 과정을 거칠 경우 이전 단계에서 추천된 질의가 다시 추천되는 경우가 많았기 때문이다. 이는 실험 데이터가 한정되어 있기 때문에 포함 정도가 높은 데이터가 순서만 바뀌어 같은 질의가 계속 추천되기 때문이다. 한정된 데이터 내에서 만족도가 높은 지식들은 그 수가 정해져 있다. 따라서 추천 과정을 반복하더라도 같은 질의들이 계속 나타나기 때문인 것으로 추론할 수 있다.

[0213] 실험 결과 퍼지 관계 곱 연산 방법에 사용되는 기존의 퍼지 합의 연산자들의 성능은 비슷하게 나타났다. 첫 번째 단계는 사용자가 입력한 질의에 대한 추천이므로 사용자의 질의에 대한 응답은 없는 상태에서 추천된 결과이다. 응답이 없는 대신 다른 질의응답 쌍들을 멤버로 하여 부연 유사도 계산으로 멤버십 정도를 가정하여 추천 지식을 찾기 때문에 추천되는 지식이 사용자의 질의를 포함하는 정도가 두 번째 단계의 추천 지식보다 적을 수 있지만 추천되는 지식이 사용자의 질의를 포함하는 범위는 다양할 수 있다고 추론할 수 있다. 그리고 두 번째 단계에서 추천되는 지식은 질의에 대한 응답이 있으므로 보다 전문적이고 넓은 의미를 가진다고 판단된다.

[0214] Kernel 방법은 질의 추천 방법이 아니고 의미적으로 유사한 질의를 찾는 방법이다. 따라서 접근 방법도 다르고 추천 과정이 반복될수록 사용자를 만족시킬 수 있는 추천 질의의 수가 많아진다고 볼 수 없기 때문에 비교 대상에서 제외하였다. 도 14는 데이터 Scope 에 대하여 추천 방법들의 단계별 추천 지식을 비교한 그래프이다.

[0215] 귀무 가설 ($H_0 : \mu_1 \leq \mu_2$)을 "두번째 단계에서 검색된 추천 지식이 첫번째 단계에서 검색된 추천 지식보다 만족도가 높은 지식이 많지 않다" 라 하고, 대립 가설 ($H_1 : \mu_1 > \mu_2$)을 "두번째 단계에서 검색된 추천 지식이 첫번째 단계에서 검색된 추천 지식보다 만족도가 높은 지식이 많다"고 하자. 두 가설에 대한 각 추천 방법 별 실험 결과에 대한 검정 결과는 표 19 및 표 20과 같다.

[0216] 실험 결과 각각의 퍼지 합의 연산자들을 이용한 퍼지 관계 곱 연산 방법을 이용할 경우 두 번째 단계에서 추천된 지식이 첫 번째 단계보다 높은 점수를 받았다. 하지만 검정 결과 FRP-L 방법과 FRP-KD 방법에서 일부 귀무 가설을 기각할 수 없다는 결과를 얻었다. 하지만 제안하는 방법은 귀무 가설을 기각한다는 결과를 얻었다.

표 19

	Landmark	Tour	Worldcup	Car	Mobile
μ_1	FRP-KD-2	FRP-KD-2	FRP-KD-2	FRP-KD-2	FRP-KD-2
μ_2	FRP-KD-1	FRP-KD-1	FRP-KD-1	FRP-KD-1	FRP-KD-1
X_1	6.91	5.04	12.85	14.37	14.22
X_2	3.75	3.08	10.67	10.53	11.33
S_1	0.79	1.20	2.81	3.69	3.02
S_2	0.35	1.02	1.58	2.60	2.80
n_1	5	6	9	6	6
n_2	5	6	9	6	6
t	8.15	3.04	2.03	2.08	1.71
f	5.54	9.75	12.59	8.98	9.95
$t_{0.05}(f)$	2.02	1.81	1.77	1.81	1.81
검정	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각하지 못한다

표 20

[0218]

	Landmark	Tour	Worldcup	Car	Mobile
μ_1	FRP-SFIO-2	FRP-SFIO-2	FRP-SFIO-2	FRP-SFIO-2	FRP-SFIO-2
μ_2	FRP-SFIO-1	FRP-SFIO-1	FRP-SFIO-1	FRP-SFIO-1	FRP-SFIO-1
X_1	10.48	10.48	10.48	10.48	10.48
X_2	6.23	7.59	7.49	7.98	8.26
S_1	4.27	4.27	4.27	4.27	4.27
S_2	3.86	3.88	4.18	3.98	4.02
n_1	32	32	32	32	32
n_2	32	32	32	32	32
t	4.17	2.83	2.84	2.43	2.14
f	61.39	61.44	61.97	61.69	61.77
$t_{0.05}(f)$	1.94	1.89	1.86	1.86	2.02
검정	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다	귀무가설을 기각한다

[0219]

본 발명에서는 사용자의 지식의 내용을 의미적으로 포함하는 정도가 큰 다른 지식을 추천하는 내용 기반의 탐색적 지식 추천 방법을 제안하였다. 제안하는 방법은 의미적 포함 관계가 보다 높은 지식을 찾아줌으로써 사용자의 의도에 부합하는 결과를 찾아줄 확률을 보다 높일 수 있다.

[0220]

지식의 내용을 의미적으로 포함하는 정도를 나타내기 위해 다른 질의응답 쌍을 멤버로 갖는 퍼지 지식을 정의한다. 멤버십 정도를 계산하기 위해 확장 질의 유사도와 보조 유사도를 적용한 부연 유사도 계산 방법을 정의하였다.

[0221]

의미적 퍼지 포함 정도는 어떤 지식이 다른 지식에 의미적으로 포함되는 정도로서 두 퍼지 지식의 포함 정도로 정의한다. 의미적으로 포함된다는 의미는 어떤 지식에서 설명되는 내용을 다른 지식에서 더 잘 설명되는 정도, 더 자세히 설명되는 정도, 더 넓은 범위의 내용을 포함하는 정도이다. 의미적 퍼지 포함 정도는 의미적 퍼지 관계 곱으로 계산된다. 의미적 퍼지 관계 곱 값이 클 수록 제공하는 정보의 범위가 넓어지기 때문에 좋은 추천 지식이 될 수 있다. 퍼지 관계 곱 연산자는 퍼지 함의 연산자 값의 평균으로 계산된다. 그런데 의미적 퍼지 관계 곱에 기존 퍼지 함의 연산자를 사용할 경우 지식의 의미적 퍼지 포함 정도를 계산하는데 적합하지 않다. 이 문제는 의미적 퍼지 함의 연산자를 정의함으로써 해결하였다.

[0222]

첫 번째 실험을 통해 제안하는 지식 추천 방법으로 추천된 지식이 다른 추천 방법으로 추천된 지식에 비해 사용자를 만족시킬 수 있는 지식을 더 많이 추천할 수 있음을 보였다. 두 번째 실험을 통해 추천 과정이 반복될수록 사용자를 만족시킬 수 있는 지식을 더 많이 추천할 수 있음을 보였다.

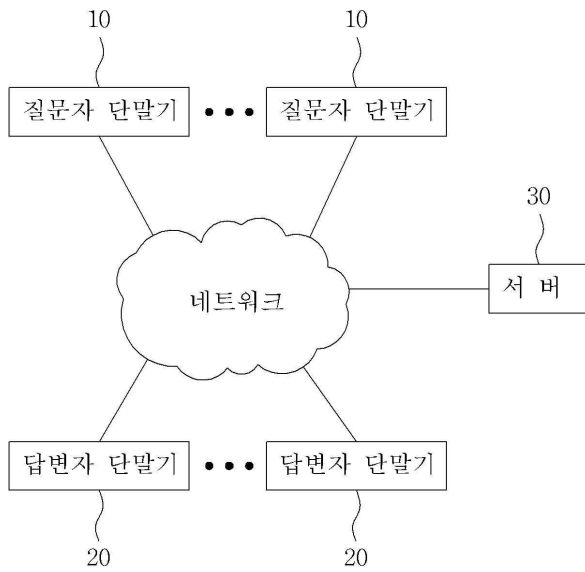
부호의 설명

[0223]

- 10: 질문자 단말기
- 20: 답변자 단말기
- 30: 서버

도면

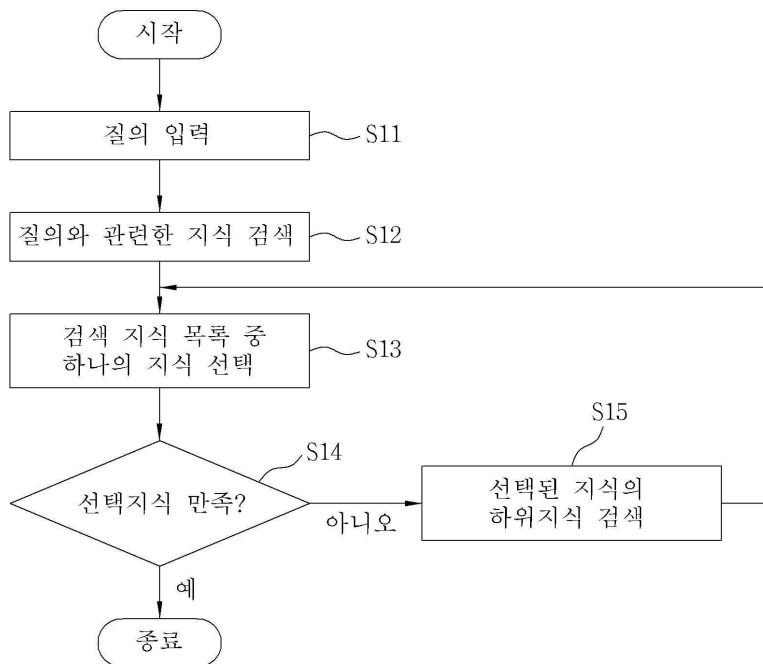
도면1



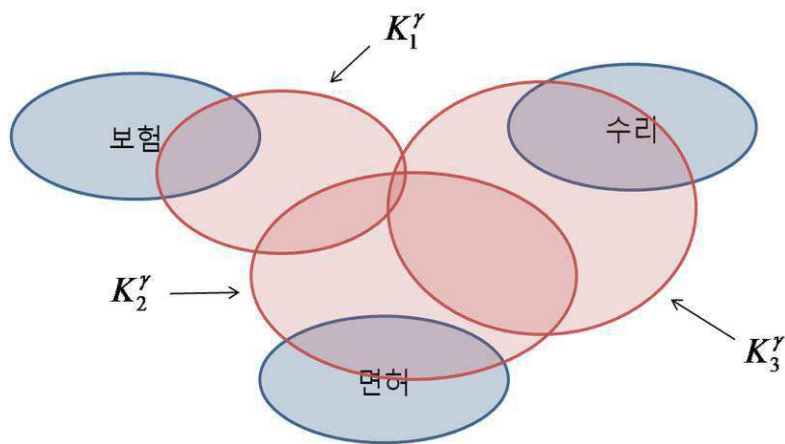
도면2



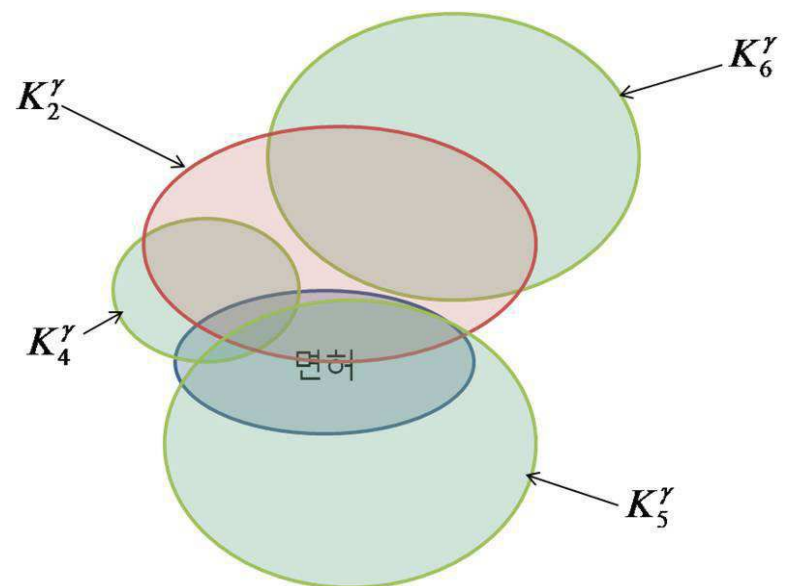
도면3



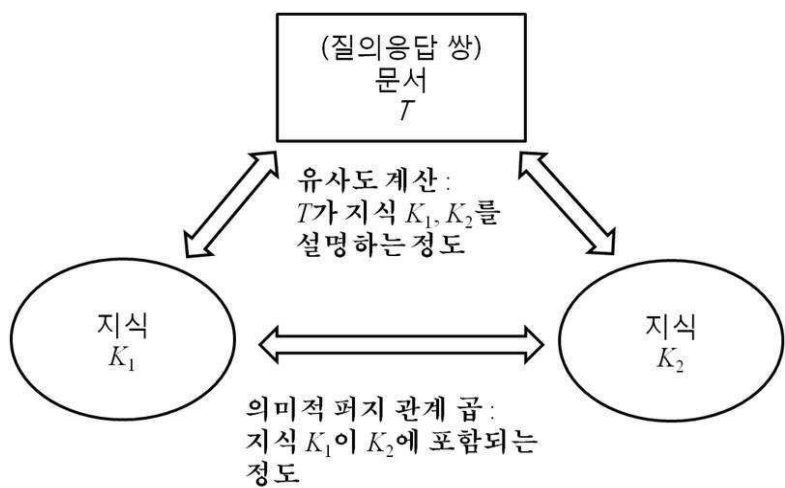
도면4



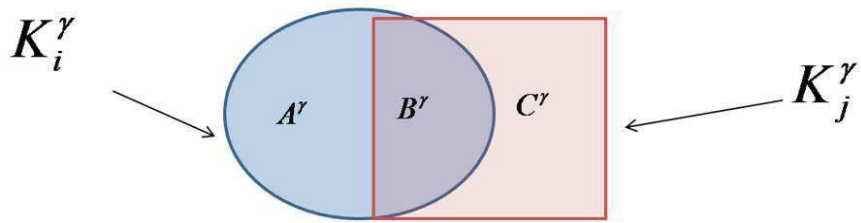
도면5



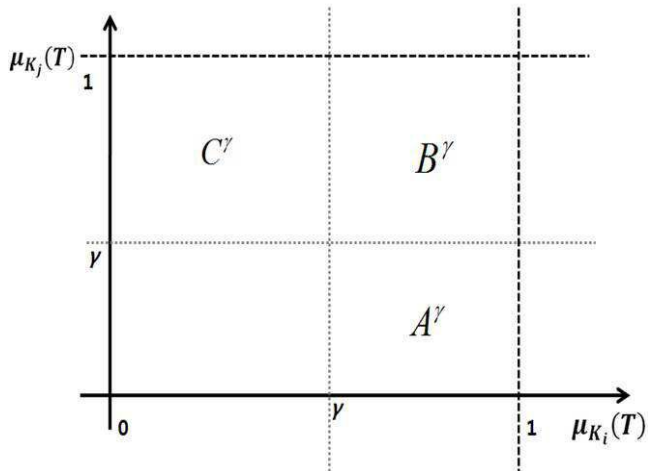
도면6



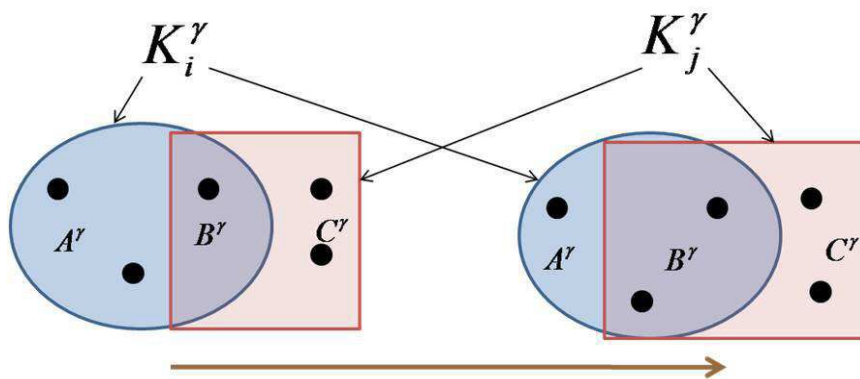
도면7



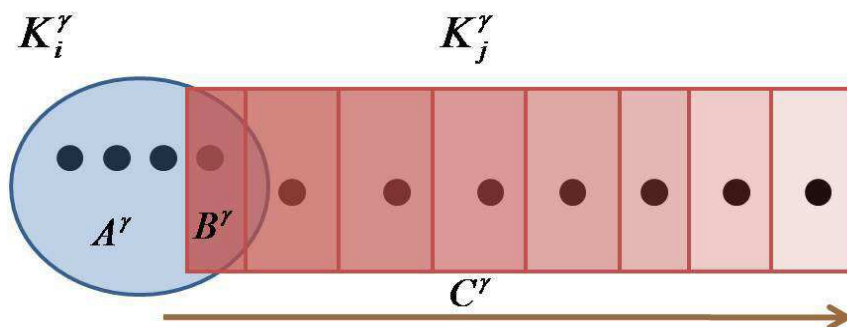
도면8



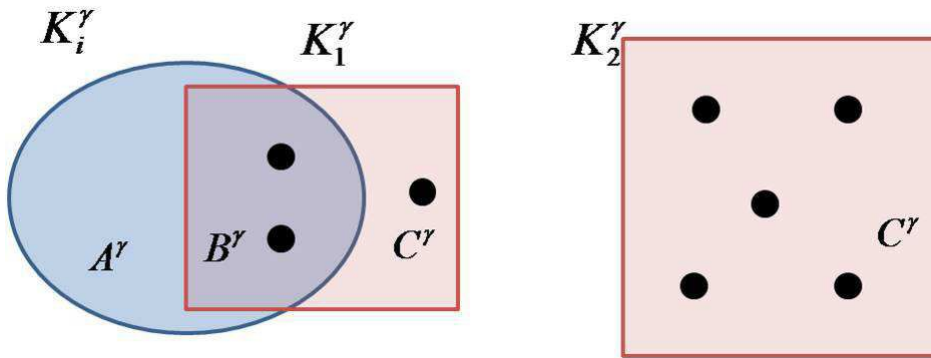
도면9a



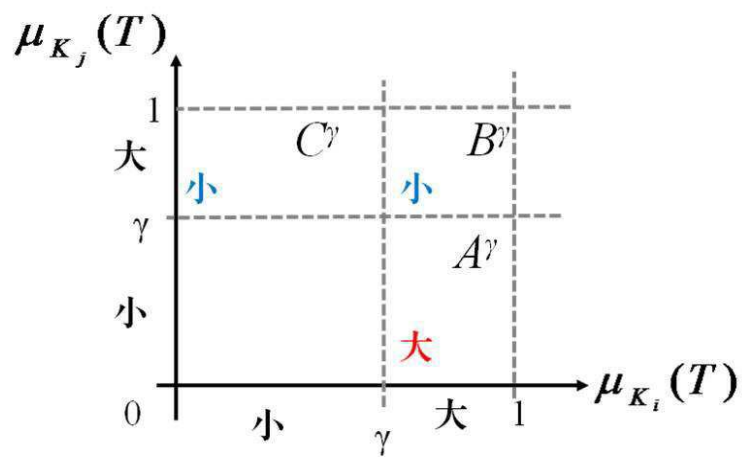
도면9b



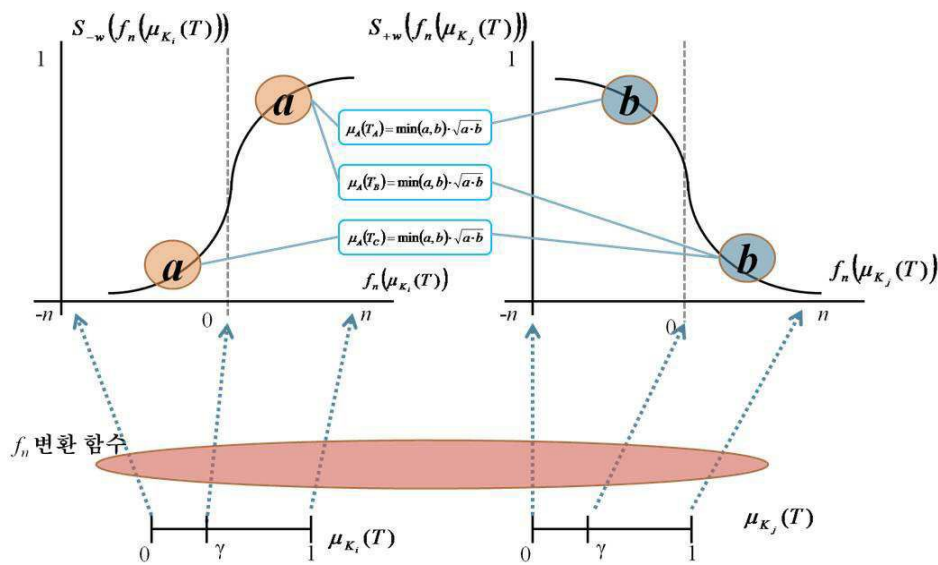
도면9c



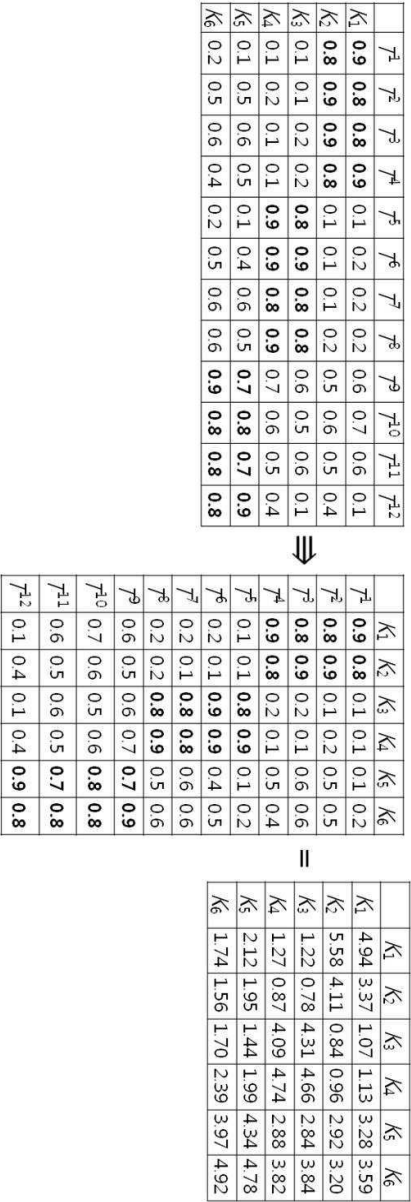
도면10



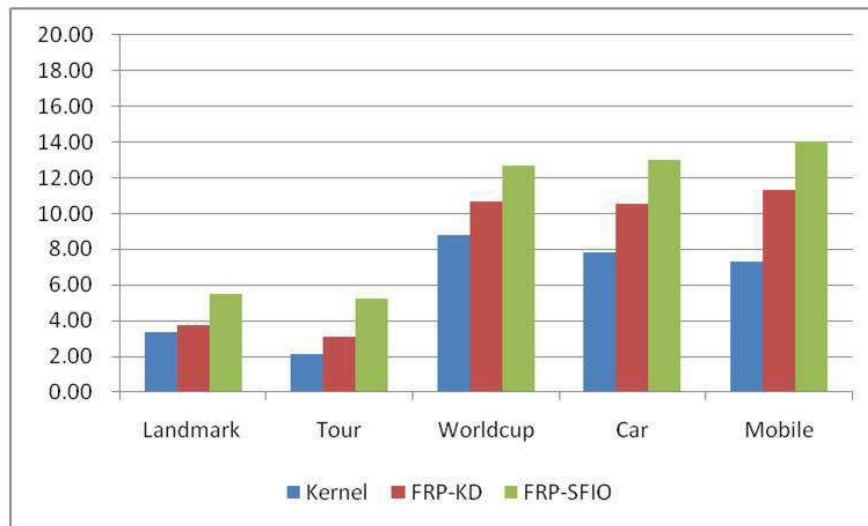
도면11



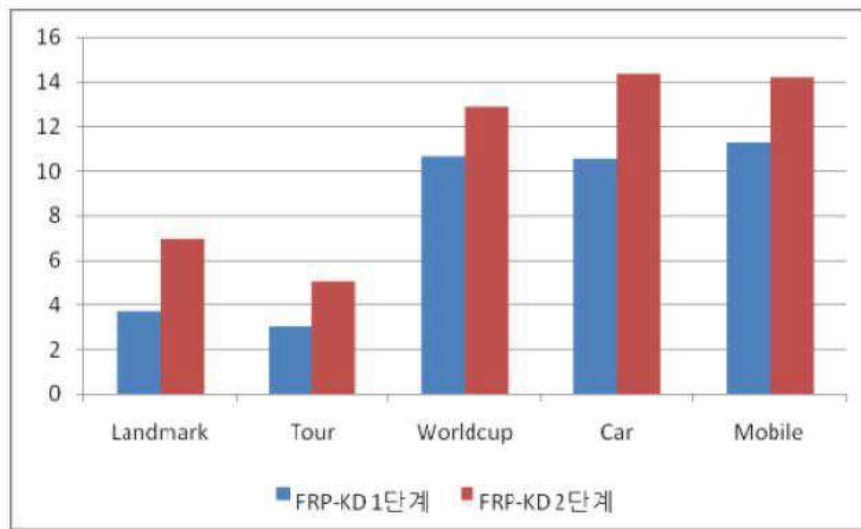
도면12



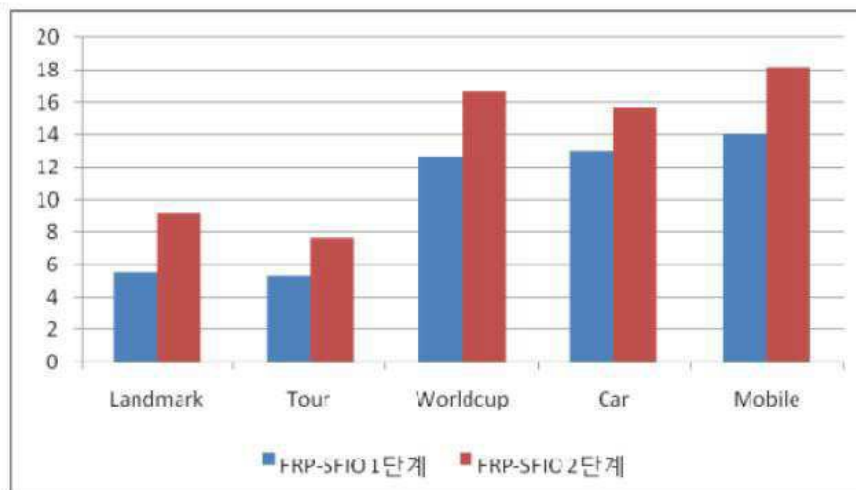
도면13



도면14



(a)



(b)