질의 완화를 이용한 지능적인 질의 응답 시스템

황혜정*, 김교정**, 윤용익***, 윤석환****

요 약

협력적 질의 응답은 질의와 데이터에 관한 지식을 이용해서 초기의 질의에 적절한 인접한 정보나 연관된 정보를 제공한다. 본 논문에서는 제시된 질의에 대한 정확한 검색 결과를 보여주지 못하는 경우에 협력적 질의 응답을 지원하기 위하여 지능적인 질의 응답 시스템은 하이브리드 지식베이스(Hybrid Knowledge Base)를 이용하여 질의 완화 과정을 수행한다. 질의 완화에 사용되는 하이브리드 지식베이스는 구조적인 접근을 위하여 시맨틱 리스토와 규칙 기반의 지식베이스로 구성된다. 또한, 본 논문은 하이브리드 지식베이스를 기반으로 초기의 질의 지식을 추출하여 질의 재정성을 하기 위하여 질의 완화 알고리즘을 제안한다.

Intelligent Query Answering System using Query Relaxation

Hye-Jeong Hwang†, Kio-Chung Kim**, Yong-Ik Yoon***, Seok-Hwan Yoon****

ABSTRACT

Cooperative query answering provides neighborhood or associated information relevant to the initial query using the knowledge about the query and data. In this paper, we present an intelligent query answering system for supporting cooperative query answering, provided that it does not show any exact retrieval result for any suggested query. Intelligent query answering system presented in this paper performs query relaxation process using hybrid knowledge base. The hybrid knowledge base which is used for relaxation of queries, composes of semantic list and rule based knowledge base for structural approach. Furthermore, this paper proposes the query relaxation algorithm for query reformulation using initial query on the basis of hybrid knowledge base.

1. 서 론

데이터베이스에서 사용자가 원하는 데이터를 검색하기 위한 일반적인 방법은 질의 언어를 사용하는 것이 다. 사용자는 자신이 원하는 정보를 얻기 위하여 질의 어를 작성하는 반면, 질의의 결과가 항상 사용자의 요구를 만족시켜 줄 만하지는 못하다. 현재의 질의 처리 과정은 질의를 간결하게 만들도록 요구하기 때문에 사용자가 문제와 관련한 동물분야의 지식과 데이터의 구조 등을 완전히 이해해야 한다. 그러나, 제한된 질의 결과가 제시되거나 정확한 결과가 없는 경우에는 이외의 다른 정보는 제공되지 않는다[3]. 따라서 사용자는 스스로가 질의 조건을 완화하여 만족할 만한 응답을 얻을 때까지 질의를 반복해야 한다. 협력적 질의 응답 (cooperative query answering)은 전문가의 문제 해결 과정처럼 주어진 질의의 의미와 데이터베이스에 저장된 지식을 활용하여 정확한 답 뿐만 아니라 유사하거나 관련된 정보를 제공하는 과정을 의미한다[3, 6, 10]. 협력적 질의 응답은 질의의 의미를 분석하고 질의와 데이터에 관한 지식을 활용하여 질의와 관련 있는 정

† 준 편 편: 숙명여자대학교 대학원 전산학과
‡ 준 편 편: 숙명여자대학교 전산학과 교수
§ 준 편 편: 숙명여자대학교 전산학과 교수
¶¶¶ 총 편 편: 한국전자통신연구원 책임연구원
논문접수: 1999년 6월 11일, 심사완료: 1999년 12월 7일
보를 제공하도록 결과를 제시하는 과정으로 구성되어 있다. 협력적 질의 응답을 얻어내기 위한 다양한 연구가 이루어지고 있는데[3,5,8,10,11,15,17,18] 이를 지원하기 위한 인터페이스의 연구 중 관계형 프레임 워크가 주류를 이루고 있다[3,5,10,15,17,18].


본 논문에서는 제시된 질의에 대한 정확한 검색 결과를 보여주지 못하는 경우에 협력적 질의 응답을 지원하기 위한 지능적인 질의 응답 시스템(Intelligent Query Answering System : IQAS)을 제안한다. 또한, IQAS에서 협력적 질의 응답을 지원하기 위하여 다음과 같은 지식의 추상화를 포함한 통합 지식 추상화 계층(Combiner Knowledge Abstraction Hierarchy : CKAH)을 제안한다. IQAS에서는 질의 완화를 수행하기 위해서 CKAH를 바탕으로 지식을 표현한 하이브리드 지식베이스(Hybrid Knowledge Base)를 이용한다. 하이브리드 지식베이스는 구조적인 접근을 위하여 시범적 리스트와 규칙 기반의 지식베이스로 구성되었다. 하이브리드 지식베이스를 이용하여 초기의 질의어를 새로운 질의어로 재정의하는 질의 완화 수행 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 살펴보고 3장에서는 CKAH에 의한 다단계 지식 표현을 설명한다. 또한, 본 연구에서 질의 완화를 수행하는데 이용되는 지식으로 하이브리드 구조의 지식베이스를 소개한다. 4장에서는 지능적인 질의 응답 시스템(sequential IQAS)을 설명하고, 기본적인 알고리즘 및 전체적인 질의 처리 알고리즘을 설명한다. 5장에서는 성능 평가를 설명하고 6장에서는 구현 예를 설명한다. 마지막으로 7장에서는 결론 및 향후 과제에 대해 설명한다.

2. 관련연구

Motro는 모호한 질의를 다루기 위해 VAGUE라고 불리는 시스템을 개발하였다[12]. VAGUE는 주된 원리 는 데이터 메트릭스(data metrics)[12]를 가지고 관계형 데이터 모델을 확장하는 것인데, 데이터 값의 긴 여성이 결정하기 위해서의 거리(distance)를 정의하였다. 모호한 질의를 표현하기 위하여 similarity-to라는 연산자를 도입하였다. similarity-to는 정의된 값의 유사성을 미리 정의된 거리 안에 포함되는 데이터 값들을 유사 데이터로 지정하여 결과로 제시한다. VAGUE는 현재 사용중인 데이터베이스 시스템이 수정될 수 없는 경우에 기존 시스템을 그대로 사용하여 성능을 높이는 것이 장점인 반면, 모든 값의 유사성이 전혀 측정될 수 없기 때문에 질의 처리가 비효율적이다.


Chu와 Chen은 협력적인 질의 응답을 지원하기 위하여 다단계 지식 표현이 가능한 형태 추상화 계층(type abstraction hierarchy)를 제안하였다[3,5]. 협력적인 질의 응답 과정은 초기 연결의 범위를 완화시키고 가장 가까운 범위에 초점을 맞춤으로써 질의의 영역을 넓히는 것이다[3,6]. 형태 계층(type hierarchy)은 상대적으로 높은 위치의 형태가 보다 일반화(generalization)된 것이고 낮은 위치의 형태가 세분화(specialization)
3. CKAH에 의한 다단계 지식 표현

본 논문에서는 협력적 질의 응답을 지원하기 위하여 지식 표현의 프레임 워크로 CKAH를 제안한다. 본 장에서는 CKAH를 설명하고 CKAH를 기반으로 구축 된 하이브리드 지식베이스를 설명한다.

3.1 CKAH

본 논문은 값과 도메인을 하나의 통합 지식 정보로 유지하고 다단계 지식으로 표현하여 통합 지식 추상화 계층의 CKAH를 제안한다. 본 논문에서 사용하는 도
매인의 의미는 속성이 가질 수 있는 값의 범위로 정의
한다. CKAH와 KAH간의 근본적인 차이점은 계층 구조를 가진 지식의 표현 형태이다. CKAH는 도메인과 값을 하나의 단일 개념으로 통합하여 지식을 유지, 관리한다. CKAH는 다음과 같은 가정을 갖는다.

가정 1. 모든 속성은 숫자가 가질 수 있는 값의 범위
인 하나의 도메인을 갖는다.
가정 2. 하나의 CKAH에서 중복된 속성 값이 존재
할 수 있다.
가정 3. 도메인은 여러 속성의 도메인일 수 있다.
가정 4. 하나의 도메인은 한 CKAH에 속하고 하나의
추상화 도메인을 갖는다.

CKAH는 지식 추상화 계층과는 달리, 부분 자식 관계의 도메인 순서만을 고려하기 때문에 지식의 추가, 삭제 등에 의해 전체 지식 구조의 변경이 이루어지지 않는다. 다음 그림은 금융기관에 관한 지식을 CKAH로 표현한 예이다.

CKAH상에서 가장 상위 층의 추상화 값은 계층 내에서 가장 일반적인 값을 갖고 가장 하위 층의 값은 가
장 세분화된 값을 갖는다. 하나의 추상화 값은 하위 층에 존재하는 여러 개의 세분화 값을 갖지만 여러 개
의 세분화 값은 오직 하나의 추상화 값만을 갖는다. CKAH는 협력적 질의 응답을 제공할 뿐만 아니라 지
식베이스의 동적인 관리도 용이하고 단순하다. CKAH는 다단계의 지식 추상화 계층을 표현하고 각 층에서
추상화와 세분화 관계를 갖고 있는 각 지식간의 값의
관계뿐만 아니라 상위 도메인과 하위 도메인간의 추상
화 관계가 함께 구성되어 있다.
32 하이브리드 지식베이스

기존에는 규칙, 시멘틱 네트(semantic net), 프레임(frame) 등이 주된 지식 표현 방법이며[7,9]. 이러한 형태의 지식 표현의 변형된 구조나 두 개 이상의 동합 형태를 갖는 하이브리드 지식 구조가 지식 표현의 한 형태로 널리 연구되고 있다[7,9]. 본 연구에서는 이러한 지식의 검색을 추구하고 용이한 지식관리를 위하여 시멘틱 리스트(semantic list)와 규칙 기반의 지식 표현을 통합하는 하이브리드 지식베이스를 제안한다.

시멘틱 리스트는 CKAH의 추상화 개념간의 관계를 표현하기 위한 트리 형태의 지식 표현으로 각 노드가 포인터로 연결되어 있고 노드의 구조는 다음과 같다.

<table>
<thead>
<tr>
<th>조회</th>
<th>값</th>
<th>PUNK(부정)</th>
<th>CLINK(QB)</th>
<th>LLINK</th>
<th>RLINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>도메인</td>
<td>VALUE</td>
<td>RULE기반</td>
<td>RULE</td>
<td>RULE</td>
<td>RULE</td>
</tr>
</tbody>
</table>

(그림 2) 시멘틱 리스트의 한 노드 구조

질의 처리를 빠르게 하고 질의 유형의 체계적인 분류를 위하여 시멘틱 리스트를 접근 할 때 값과 도메인의 복합키를 이용하여 해쉬 함수를 통해 해당 노드를 찾아 간 후, 일의의 충격자의 추상화 값을 부모 포인

(그림 1) 금융기관에 대한 CKAH의 예

(그림 2) 시멘틱 리스트의 한 노드 구조

(그림 3) 유사지식을 포함한 규칙집간에 모듈화 구조

(그림 3)에서 표현된 Rtype01, Rtype02, Rtype03 등
의 규칙 집합들은 한 CKAH의 지식들로 구성되어 있고 시맨틱 리스트의 노드에서 규칙 구분자가 이 규칙의 집합을 나타내고 있으므로 시맨틱 리스트에서 추상화 개념을 얻는 경우에 관련된 세분화 규칙을 포함하는 규칙의 집합의 구분자를 통해 관련있는 유사한 규칙에 따른 접근이 가능하다.

따라서, 추상화 개념을 추출함으로써 일반화의 정보를 얻기 위해서는 시맨틱 리스트에 의해 한 개념의 추상화 과정의 슈퍼 도메인(즉, 상위 층의 도메인)을 추출하고 관련된 규칙 구분자의 정보를 통해 해당되는 CKAH의 지식을 갖고 있는 지식베이스에 접근하여 세분화 개념을 얻는 과정을 수행한다.

4. 지능적인 질의 응답 시스템

본 장에서는 추상성을 기반으로 구축된 데이터베이스 상에서 체계화된 질의 응답을 지원하는 지능적인 질의 응답 시스템, IQAS를 설명한다. IQAS는 저축성성을 목적으로 고객이 원하는 저축 상품을 권장하는 지능적인 질의 응답 시스템이다. 고객이 원하는 저축상품과 일치하는 상품이 없는 경우에는 가장 유사한 저축상품을 결과로 제시하도록 근접 질의를 수행하고 데이터베이스와 관련된 지식이 없는 사용자들에 지원하기 위해 개발한 질의 응답을 지원한다.

4.1 지능형 질의 응답 시스템의 구조

지능형 질의 응답 시스템인 IQAS의 구조는 (그림 4)와 같다. IQAS는 사용자 인터페이스, 질의 처리기, 상용의 DBMS, 데이터베이스, 독일 일반화기, 질의 재형성기, 하이브리드 지식베이스로 구성된다.

질의이어는 사용자 인터페이스를 통해 입력되고 입력된 질의는 질의 처리 과정을 수행함으로써 결과가 제시된다. 사용자가 입력한 초기 질의의 수행 결과가 데이터베이스 상에서 존재하지 않는 경우에는 질의 완화과정이 수행된다. 질의 완화 과정은 속성의 일반화기와 질의 재형성기에 의해 하이브리드 지식베이스의 지식을 이용하여 이루어진다. 속성 일반화기는 사용자에게 완화될 속성을 선택하도록 하여 한 층을 입력받고 그 속성의 값에 대한 추상화 값을 탐색함으로써 완화된 결과의 속성 값을 질의 재형성기에게 제시한다. 질의 재형성기는 속성 일반화기에서 제시한 속성의 추상화 값을 실제로 데이터베이스에 저장된 데이터 값으로 재작성될 수 있도록 세분화 과정을 향상하여 질의를 변형한다. 마지막 단계로, 재형성된 질의는 피드백(feedback)과정으로 다시 질의 처리기에 의해 질의 처리 과정을 수행한다.

4.2 기본적인 질의 탐색 연산자

IQAS에서는 질의를 탐색하는데 필요한 기본적인 연산자(Primitive operator)는 다음과 같다.

4.2.1 추상화 값의 탐색 연산자

속성이 속하는 도메인과 속성 값에 의해 시맨틱 리스트의 한 노드를 찾아가고 부모 포인터를 통해 상위 층의 추상화 값을 얻는다. 다음은 추상화 값의 탐색 연결자 알고리즘을 설명한다.

<table>
<thead>
<tr>
<th>Algorithm: Search_Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>// 노드의 상위 도메인과 상위 속성 값을 찾는 알고리즘 //</td>
</tr>
<tr>
<td>// input // var X : char</td>
</tr>
<tr>
<td>// output // R_domain, R_value, Id : char</td>
</tr>
<tr>
<td>R_domain = X ? PPLINK 1 DOMAIN</td>
</tr>
<tr>
<td>R_value = X ? PPLINK 1 VALUE</td>
</tr>
<tr>
<td>Id = X ? PPLINK 1 ID</td>
</tr>
<tr>
<td>RETURN(R_domain, R_value, Id)</td>
</tr>
</tbody>
</table>

(그림 5) 추상화 값의 탐색 연산자 알고리즘

4.2.2 세분화 값의 탐색 연산자

규칙기반의 지식베이스에서 도메인과 속성 값을 이용하여 추론함으로써 세분화 값을 해당 결과로 얻게 된다. 추상화 지식이 세분화되어야 할 결과적인 지식
의 충은 질의어에 표현된 속성이 속하는 도메인의 층 이다. 질의 완화를 위하여 사용되는 지식은 추상화 값을 표현하기 위하여 지식의 추상화 관계를 단계가 층 구조로 표현한 형태를 기반으로 하여 언어적 개념이기 때문에 데이터베이스 상에서 사용자가 유용하게 사용할 수 있는 데이터의 추출은 질의어에서 사용한 속성과 같은 층의 값을 갖는 데이터이다. 따라서, 질의어 작성된 속성이 CKAH상에서 속하는 도메인의 층에서부터 임의의 층까지 n-레벨의 추상화 값을 얻었다 면, 역으로 추상화 값을 데이터베이스의 인스턴스로 존재하는 값, 즉 질의어로 나타나는 속성의 도메인 층 까지 세분화 시켜 초기 질의어를 완화된 조건을 포함한 질의어로 재정형 할 수 있다. IQAS에서는 규칙 기반의 지식베이스를 이용하여 세분화 값을 얻는다. 다음 연산자는 규칙 기반의 지식베이스에서 세분화 값 을 탐색하는 규칙 탐색 연산자이다. 규칙 탐색 연산자를 사용하여 추상화 도메인의 추상화 값을 R_domain, R_value의 세분화 값을 찾는다. 도메인 R_domain과 속성 값 R_value 그리고 규칙 구분자 Id는 시스템 리 스트를 통해 얻어진 추상화 지식이다. 따라서, 세분화 값의 탐색 연산자는 도메인 R_domain과 속성 값 R_value의 추상화 지식을 도메인 Domain까지의 세분화 지식을 추출한다.

Search_RuleBase(R_domain, R_value, Id, Domain, Result(*))

4.2.3 질의 유형에 따른 지식 탐색의 분류
IQAS는 세 가지 유형의 협력적 질의 응답 중 근사적 질의 응답과 개념적 질의 응답을 지원한다. 세론에서 설명하였듯이, 근사적 질의 응답은 사용자가 입력한 질의어의 해당 데이터가 없는 경우 질의의 범위를 넘어서는 과정이고 개념적 질의 응답은 데이터베이스에 대한 사용자의 지식이 적절한 경우 사용자가 상위의 개념으로 질의어를 작성하였다면, 질의어에 작성된 속성의 층으로 결과를 데이터를 세분화 시켜서 결과를 제시하는 것이다. 임의의 사용자는 지식베이스에 있는 지식을 검색하는데 협력적 질의 응답의 개념을 알거나, 질의 유형을 분류할 필요가 있다. 따라서, 일반 사용자는 특정하게 지식의 검색하는 데 있어서 자동적으로 근사적 질의 유형인지 개념적 질의 유형인지로 분석하는 과정이 필요할 것이다.

본 논문에서 제안한 CKAH의 통합 지식의 한 개체는 도메인과 값으로 이루어져 있다. 이 때, 한 개체의 통합 지식은 같은 층이 속하는 도메인과 값으로 구성된다. 시스템 리스트상에서 직접적으로 만 도드로 접근하기 위해 실제 함수를 이용하는데, 다른 질의 처리를 위해 실제 함수를 사용하는 것 이외에 질의 유형을 분류하기 위하여 실제 함수를 사용한다. 시스템 리스트를 이용한 지식 추출의 실행단계로 값과 도메인의 복합값을 이용한 실제 함수값을 통해 직점적인 노드 주소를 알 수 있는데 이 때는 복합값에 의해 실제 함수 결과값이 제시되는 것이다. 즉, 같은 층이 속하는 두 도메인과 값으로 구성된 복합값을 이용한 실제 함수의 결과가 직점적인 노드 주소를 나타내지만 만일, 도메인과 값이 다른 층의 개념이라면 실제 함수의 결과는 주소의 값이 없는(null) 것으로 제시된다. 따라서, 실제 함수에 대한 통합 지식의 주소치 사전(mapping) 결과는 질의 유형 분류에 유용하게 사용된다. 분류된 질의 유형에 따라 질의를 처리하는 연산자가 필요하다. 먼저, 근사적 질의를 처리하는 연산자와 개념적 질의를 처리하는 연산자가 필요하다. 각각의 질의를 분류하는 과정은 4.3절의 전체적인 질의 처리 알고리즘 중 속성 단계와 알고리즘에서 기술한다.

군사적 질의 응답(approximate query answering)
군사적 질의 응답은 질의어에서 작성한 속성이 속하는 도메인과 값으로 이용하여 상위 층의 도메인과 추상화 값을 갖는 과정이다.

Algorithm : Approximate_Query
// 초기 질의를 완화하여 근사적 질의를 수행하는 알고리즘 //
// input : var T : char
// output : Result(*) : char
Search_Abstract(T, R_domain, R_value, Id)
Search_RuleBase(R_domain, R_value, Id, Domain, Result(*))
End Approximate_Query

(그림 6) 근사적 질의 응답 연산자 알고리즘

• 개념적 질의 응답(conceptual query answering)
  개념적 질의 응답은 두 가지 경우로 구분된다. 즉, 속성이 더 상위 도메인인지 값이 더 상위 도메인 인지
에 따라 제시되는 결과가 달라진다. 속성 값이 더 높은 측에 있는 경우는 사용자가 절의어로 입력한 속성 값의 개념이 데이터베이스 상의 개념 보다 일반적인 개념을 입력한 것이다. 따라서, 값이 더 높은 측에 있는 경우에는 도메인의 측에서부터 값으로 세분화 개념을 결과로 제시하게 된다. 속성 값이 속성이 의미하는 개념보다 세분화 값으로 입력되었다면 사용자가 너무 구체적인 개념으로 질문한 것이다. 따라서, 데이터베이스 상의 개념보다 세분화 값으로 질문한 것이므로 속성 값의 속성과 속성값은 도메인 측까지의 개념으로 값은 추상화 시켜 그 추상화 값을 결과로 제시한다. 다음은 개념적인 질의 처리 알고리즘을 설명한다.

<table>
<thead>
<tr>
<th>Algorithm : Conceptual_Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>// 초기 절의에 대한 개념적 질의를 수행하는 알고리즘 //</td>
</tr>
<tr>
<td>// input // var Domain, Value : char</td>
</tr>
<tr>
<td>// output // Result(*) : char</td>
</tr>
<tr>
<td>T ! = Head</td>
</tr>
<tr>
<td>Domain# = 0</td>
</tr>
<tr>
<td>Vdomain# = 0</td>
</tr>
<tr>
<td>Domain_temp = null</td>
</tr>
<tr>
<td>call Level_Check(T, Domain, Value, Domain#, Vdomain#, Domain_temp)</td>
</tr>
<tr>
<td>IF Domain# = 0 &amp; Vdomain# = 0 &amp; Domain_temp = null</td>
</tr>
<tr>
<td>THEN RETURN(null)</td>
</tr>
<tr>
<td>IF Domain# &gt; Vdomain# THEN</td>
</tr>
<tr>
<td>LOOP UNTIL(R_domain &gt; Domain)</td>
</tr>
<tr>
<td>= READ(Domain_temp, Value)</td>
</tr>
<tr>
<td>Search_Abstract(X, R_domain, R_value, Id)</td>
</tr>
<tr>
<td>ENDLOOP</td>
</tr>
<tr>
<td>ELSE</td>
</tr>
<tr>
<td>Search_RuleBase(Domain_temp, Value, Id, Domain, Result(*))</td>
</tr>
<tr>
<td>RETURN(Result(*))</td>
</tr>
</tbody>
</table>

(그림 7) 개념적 질의 응답 연산자 알고리즘

4.3 전체적인 질의 처리 알고리즘

사용자가 제시한 질의에 정확하게 부합되는 결과가 없는 경우, 사용자에게 질의어에 나타난 속성들의 완화 순위를 선정하도록 요구하고 사용자가 순위를 정한 속성을 대상으로 속성의 일반화를 수행한다. 하이브리드 지식베이스 중 규칙기반의 지식베이스를 사용하여 재형성된 질의는 다시 질의 처리기로 입력되고 재검색을 수행한다.

속성 일반화 과정에서는 사용자의 질의어에서 속성과 속성 값을 통해 근사적 질의와 개념적 질의간에 어
의 유형의 질의인어를 검색하여 해당되는 유형별 질의 로 속성에 의해서 속성이 속하는 계층과 도메인을 찾고, 속성과 속성 값을 검색(&lt;key&gt;)로 하여 계층 내에서 노드를 검색한다. 검색키로 검색이 되지 않는 경우에는 속성의 종과 속성 값의 차이 때문에 생기는 것 이므로 개념적 질의 유형임을 의미하는 것이다. 따라서, 검색키로 지식의 노드가 검색되는 경우에는 근사적 질의로 분류되고 검색키에 의해 검색되지 않는 경우에는 개념적 질의로 분류되어 지식의 검색을 수행한다.

질의 재형성 과정은 속성 일반화 과정에서 얻어진 일반화의 개념을 다시 세분화 시키면서 초기 질의어로 재형성하는 과정이다.

**Algorithm: Query Reformation**
// 완화된 속성 값에 의해 초기 속성의 값을 제작하는 알고리즘 //
// return var Query, Attribute, Value_list(*) : char
// output Reformed_Query : char

Replace values of Attribute with Value_list(*) in Query
RETURN(Reformed_Query)

(그림 11) 질의 재형성 알고리즘

5. 성능 평가

본 장에서는 본 논문에서 제안한 지능적인 질의 응답 시스템의 근사적 질의에 대한 성능 평가의 결과를 설명한다. 근사적 질의에 대한 성능 평가를 위해 사용자의 관심에서 사용자의 평가 기준으로 관련 있는 결과를 모티가 간선한 후, 근사적 질의 응답으로 제시된 결과들을 기반으로 recall과 precision을 계산하여 성능 평가를 수행하였다.

\[
\text{recall} = \frac{\text{검색된 관련 있는 점측 상품의 수}}{\text{관련 있는 점측 상품 결과의 수}}
\]

\[
\text{precision} = \frac{\text{검색된 관련 있는 점측 상품의 수}}{\text{검색된 총 점측 상품 결과의 수}}
\]

성능 평가를 하기 위하여 먼저, 정확하게 일치하는 데이터만을 결과로 제시하는 일반적인 검색과 질의 재형성을 통한 질의 피드백 과정을 수행하여 근사적 질의 결과를 나타낸 검색 결과를 비교하여 측정하였다.

총 24개 기관의 120여개의 저작상품 중 저작상품 명, 투자기간, 수익률에 따라 내역이 유사하게 관련 있다고 정의한 상품 32개를 기반으로 점데어와 정확하게 일치하는 내용을 검색한 결과와 한 개의 질의 속성을 완화한 검색 결과(저작상품명을 선택), 두 개의 질의 속성을 완화한 검색을 수행한 결과(저작상품명, 투자기간을 선택)가 <표 1>과 같다.
이의와 가장 인접한 정보를 통해 투자기간, 금융기관, 수익률의 내역이 완화되고 질의가 처리된 결과이다.

(그림 13) 근사적 질의 응답 결과 화면

다음은 개념적 질의 응답을 수행한 결과 화면이다. 금융기관의 속성이 의미하는 내용보다 너무 일반적인 의미로 질문이 입력되어 시스템 내부적으로 좀 더 구체적인 값으로 질문이 수정되고 이에 따른 처리 결과가 제시되었다. (그림 14)의 예에서 속성 '금융기관'의 속성 값이 '은행'으로 입력되었는데 데이터베이스에 저장된 금융기관은 구체적인 기관명이므로 '은행'에 해당되는 자료는 없다. 따라서, 질문을 재정의하고 다시 질의 처리 과정을 수행해야 하는데 (그림 14)와 같은 개념적 질의를 처리하는 과정은 다음과 같다. (그림 1)의 CKAH 상에서 볼 때 속성 값 '은행'의 도메인은 '금융기관'인

(그림 14) 개념적 질의 응답 결과 화면

표 1) 3가지 유형의 검색 결과에 대한 성능 평가
(Relevant = 32)

<table>
<thead>
<tr>
<th>검색 유형</th>
<th>Retrieved</th>
<th>Relevant</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>일차 검색</td>
<td>13</td>
<td>13</td>
<td>0.41</td>
<td>1</td>
</tr>
<tr>
<td>1차 측정 완화 검색</td>
<td>32</td>
<td>32</td>
<td>1</td>
<td>0.84</td>
</tr>
<tr>
<td>2차 측정 완화 검색</td>
<td>47</td>
<td>32</td>
<td>1</td>
<td>0.68</td>
</tr>
</tbody>
</table>

(그림 12) 검색 유형의 성능 평가

6. 구현

본 논문에서 제안한 전자적인 질의 응답 시스템은 AXI 413 하에 Oracle 8i, VisualCafe 3.0, JDK 1.2.1 을 이용하여 구성되었다. 상품명, 금융기관, 수익률, 투자기간, 안정성, 저축유형, 세금 유형 등을 화면에서 입력하여 접근할 수 있다. 사용자 질의에 정확하게 일치하는 자료가 없는 경우에는 질의 완화에 의해 가장 근사한 자료가 제시되고 제시된 자료의 근거를 설명한다. 즉, 사용자가 입력한 질의어를 토대로 완화시킨 속성과 완화과정을 통해 속성 값의 범위를 완화한 경위에 관한 설명과 함께 화면 하단에 제시된다. 따라서, 사용자가 요구한 질의와 완전하게 일치하는 자료가 없는 경우에, 현재 화면에 보여준 내용을 사용자가 쉽게 이해할 수 있다. 상품명, 금융기관, 수익률, 투자기간, 안정성, 저축유형, 세금 유형 등을 화면에서 입력하여 철저할 수 있다.

다음은 근사적 질의 응답을 수행한 결과 화면이다. 사용자의 질의어와 정확하게 일치하는 결과가 없어 질
7. 결론 및 향후 연구과제

본 논문에서는 속성을 기반으로 한 데이터베이스 상에서 지능적인 질의 응답 시스템을 제안하였다. 본 연구의 프로토타입 시스템으로 구현된 지능적인 질의 응답 시스템, IQAS는 완 기반 하에서 사용자에게 금융 상품을 설명하고 정확하게 사용자의 질의와 일치하는 자료가 없는 경우에는 유사한 자료를 제시해준다. 본 연구에서는 갭간의 추상화와 세분화 개념을 표현하는 단계적 지식의 계층 관계를 표현하기 위해 CKAH를 제안하였다. CKAH를 이용하여 속성값 간 및 각 개념 간의 관계를 하나의 통합체로 유지함으로써 지식베이스를 단순화하고 용이하게 관리하도록 하였다. 시맨틱 리스트와 규칙기반의 지식 표현 구조를 통합한 하이브리드 지식베이스의 지식 표현 구조를 제안하고 하이브리드 지식베이스의 지식을 사용한 질의 완화를 시도하였다. 시맨틱 리스트에 추상화 지식을 표현하고 규칙 기반의 지식베이스에 세분화 지식을 분류하여 표현함으로써 지식 검색의 유연성을 부여하였다. 규칙 기반의 지식베이스는 결과가 사용자에게 설명하기 용이한 정보를 이용하여 완화 과정을 설명하였다.

질의 완화에 있어서 지식베이스는 필수적인 요소이다. 질의완화를 통한 사용자 질의를 재 작성하는 과정은 지식베이스와 연계되어 이루어진다. 규칙 기반의 지식베이스는 상업적인 도메인 상에서 많이 적용하는 지식표현기법으로 인간의 추론체계와 유사하여 지식의 유사성을 이해하기가 쉽다. 장점이 있는 반면, 지식이 많아질수록 지식을 분류, 검색하기가 어려워지는 단점을 갖고 있다. 많은 양의 지식을 분류하기 위하여 시맨틱 리스트 지식을 통해 지식 구분자를 탐색하여, 해당 계층의 지식만을 모듈화한 지식베이스를 탐색한다.

항후 연구 과제로는 사용자 질의 결과를 이용한 규칙 탐색을 함으로써 내포적인 질의 응답을 지원하는 규칙을 생성하고 이를 규칙기반의 지식베이스에 추가하는 방법 및 엔터프라이즈의 의미 기반의 질의를 하이브리드 지식베이스와 연계시키는 방법이 연구과제로 남아 있다.

참고 문헌

김 교 정
e-mail: kiochkim@sookmyung.ac.kr
1972년 연세대학교 화학과 졸업 (학사)
1983년 Clarkson Univ. 전산학과 (석사)
1991년 Clarkson Univ. 전산학과 (박사)
1986년 - 현재 숙명여자대학교 전산학과 교수
관심분야: 지식베이스, 데이터마이닝 등

윤 용 익
e-mail: yiyoon@cs.sookmyung.ac.kr
1983년 동국대학교 통계학과
1985년 한국과학기술원 전산학과 (석사)
1994년 한국과학기술원 전산학과 (박사)
1985년 - 1997년 한국전자통신연구원 책임연구원
1997년 - 현재 숙명여자대학교 전산학과 교수
관심분야: 분산 시스템, 실시간 시스템, 컴퓨터 통신과 네트워크 등

윤 석 환
e-mail: yoonsh@mail.iit.ac.re.kr
1982년 아주대학교 산업공학과 (학사)
1984년 건국대학교 산업공학과 (공학석사)
1996년 아주대학교 산업공학과 (공학박사)
1992년 평질관리기술사 취득
1986년 - 현재 한국전자통신연구원/정보통신연구관리단 책임연구원
1995년 ~ 현재 한국정보처리학회 회장, 편집위원장
관심분야: 그룹웨어, 생산정보시스템, 지식관리시스템, S/W 공학, 개발방법론

황 해 정
e-mail: hjhwang@cs.sookmyung.ac.kr
1987년 숙명여자대학교 전산학과 (학사)
1989년 숙명여자대학교 전산학과 (석사)
1999년 숙명여자대학교 전산학과 (박사)
1989년 ~ 1990년 산업연구원 연구원
1991년 ~ 1998년 산업기술정보원(전산 산업연구원) 책임연구원
관심분야: 질의 환화, 멀티미디어