

질의 완화를 이용한 지능적인 질의 응답 시스템

황 혜 정[†] · 김 교 정^{**} · 윤 용 익^{***} · 윤 석 환^{****}

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 시스템을 제안한다. 본 논문에서 제안한 지능적인 질의 응답 시스템은 하이브리드 지식베이스(Hybrid Knowledge Base)를 이용하여 질의 완화 과정을 수행한다. 질의 완화에 사용되는 하이브리드 지식베이스는 구조적인 접근을 위하여 시멘틱 리스트와 규칙 기반의 지식베이스로 구성된다. 또한, 본 논문은 하이브리드 지식베이스를 기반으로 초기의 질의어를 이용하여 질의 재형성을 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 것이다. 사용자는 자신이 원하는 정보를 얻기 위하여 질의어를 작성하는 반면, 질의의 결과가 항상 사용자의 요구를 만족시켜 줄 만하지는 못하다. 현재의 질의 처리 과정은 질의를 간결하게 만들도록 요구하기 때문에 사

용자가 문제와 관련한 응용분야의 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 의도를 분석하고 질의와 데이터에 관한 지식을 활용하여 질의와 연관 있는 정

† 준 회 원 : 숙명여자대학교 대학원 전산학과
** 준 회 원 : 숙명여자대학교 전산학과 교수
*** 종신회원 : 숙명여자대학교 전산학과 교수
**** 정 회 원 : 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].

협력적 질의 응답은 내포적 질의 응답(intensional query answering), 근사적 질의 응답(approximate query answering), 개념적 질의 응답(conceptual query answering) 등 세 가지의 유형으로 구분된다[11]. 내포적 질의 응답은 사용자의 질의 결과를 그대로 보여주는 것 대신에 질의 결과와 관련된 간결한 기술이나 유용한 정보를 제시하는 것이다[5]. 근사적 질의 응답은 질의에 대한 정확한 답이 없는 경우에 질의의 조건을 완화하여 근사적으로 만족하는 유사한 값을 얻는 것을 말한다[4, 15]. 개념적 질의 응답은 사용자가 데이터베이스 스키마의 세부 조건에 대한 지식이 적어 질의의 값이 보다 상위개념으로 제시되거나 너무 구체적인 하위 개념으로 제시되는 경우에 답을 찾아가는 방법이다[4]. 즉, 개념적 질의 응답은 사용자의 질의를 해석하여 데이터베이스에 구축되어 있는 데이터의 개념과 일치하도록 조정해서 원하는 답을 찾아나가는 방법을 의미한다.

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 다단계 지식의 추상화 값을 표현한 통합 지식 추상화 계층(Composite 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 시멘틱 리스트와 규칙 기반의 지식베이스로 구성되었다. 하이브리드 지식베이스를 이용하여 초기의 질의어를 새로운 질의어로 재형성하는 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 시스템인 IQAS를 설명하고, 기본적인 알고리즘 및 전체적인 질의 처리 알고리즘을 설명한다. 5장에서는 성능 평가

를 설명하고 6장에서는 구현 예를 설명한다. 마지막으로 7장에서는 결론 및 향후 과제에 대해 설명한다.

2. 관련연구

Motro는 모호한 질의를 다루기 위해 VAGUE 라고 불리는 시스템을 개발하였다[12]. VAGUE의 주된 원리는 데이터 매트릭스(data metrics)[12]를 가지고 관계형 데이터 모델을 확장하는 것인데, 데이터 값들 간의 유사성을 결정하기 위하여 값간의 거리(distance)를 정의하였다. 모호한 질의를 표현하기 위하여 similar-to라는 연산자를 도입하였는데 similar-to가 규정한 값간의 유사성은 미리 정의된 거리 안에 포함되는 데이터 값들을 유사 데이터 규정하여 결과로 제시한다. VAGUE는 현재 사용중인 데이터베이스 시스템이 수정될 수 없는 경우에 기존 시스템을 그대로 사용하여 실현될 수 있는 점이 장점인 반면, 모든 값의 유사성이 거리로 측정될 수 없기 때문에 질의 처리가 비효율적이다.

Jiawei는 지능적인 질의 처리 기능을 향상시키기 위해서 일반화 규칙들과 개념 계층, 지식 발견 도구 등을 사용하여 지능적인 질의 응답을 제공하는 프레임 워크를 제안하였다[8]. 서로 다른 층간의 관계를 표현하는 개념 계층[2, 8]을 사용하여 지능적으로 질의를 처리하였다. Jiawei가 제안한 프레임 워크는 데이터 요약(data summarization), 질의 변형(query transformation), 지식발견(knowledge discovery), 개념 클러스터링(concept clustering), 질의 재작성(query rewriting), 지연 평가(lazy evaluation) 등을 통해서 질의 응답을 지원한다. Jiawei의 연구는 다양한 방법을 이용한 지식 발견을 통해 지능적인 질의 응답을 지원함으로써 질의 처리 기능을 향상시킨 점은 주목할 만 하나 임의의 단계의 개념으로 제시되는 질의어의 유형을 분류하는 과정이 체계적이지 못하기 때문에 질의 유형에 따른 처리 과정이 비효율적이다.

Chu와 Chen은 협력적인 질의 응답을 지원하기 위하여 다단계 지식 표현이 가능한 형태 추상화 계층(type abstraction hierarchy)을 제안하였다[3, 5]. 협력적인 질의 응답 과정은 초기 질의어의 범위를 완화시키고 가장 가까운 범위에 초점을 맞추으로써 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)

된 것이라 할 수 있는 형태들의 집합의 부분적인 순서를 의미한다[14]. Chu와 Chen은 기존의 형태 계층의 일반화, 세분화 관점을 통합시켜주는 추상화 관점(abstraction view)을 제공하기 위하여 형태 계층을 추가하여 형태 추상화 계층을 제안하였다. 질의 처리를 위해 세 가지 테이블을 사용했다. 형태 계층 테이블(type hierarchy table)은 형태간의 하위형태(subtyping)관계를 서술하였다. 속성 형태 테이블(attribute type table)은 속성 이름과 형태 이름간의 관계를 정의하여 속성이 같은 형태가 다르게 이름 지어질 수 있도록 하였다. 추상화 사상 테이블(abstract mapping table)은 상위형태(supertype)와 하위형태간의 대응되는 인스턴스값의 짝을 연결하기 위해 정의하였다. 형태 추상화 계층을 통해서 협력적인 질의 응답을 지원하기 위한 체계적인 완화 과정을 정의하였으나 전체 층의 관계를 하나의 튜플 형태로 표현하였기 때문에 지식이 추가되거나 삭제 될 경우에 지식을 표현하는 전체 스키마가 변경될 수 있다는 단점이 있다.

Huh 와 Moon은 데이터베이스에 저장된 지식을 추상화하여 질의 조건을 완화하는 방법으로 지식 추상화[13] 계층(Knowledge Abstraction Hierarchy : KAH)을 제안하였다[17, 18]. 지식 추상화 계층은 데이터 추상화를 사용하여 데이터와 데이터베이스에 있는 지식간의 다단계 표현을 해주는 지식 표현 프레임 워크로 제안되었는데 하나의 계층 내에서 유사한 데이터 값들을 추상화 값과 세분화 값간의 관계로 정의하였다. 또한, 데이터 값들이 속하는 영역의 도메인과 연계시키기 위해 상위 도메인(superdomain)과 하위 도메인(subdomain)간의 관계를 도메인 추상화 관계로 정의하였다. 지식 추상화 계층은 관계형 프레임 워크를 기반으로 지식을 설계하였는데 하나의 값 추상화 관계를 표현하는 계층을 한 단계간의 관계로 표현함으로써 데이터 의존도(data dependency)를 제거하였다. 그러나, 계층 구조를 한 단계의 관계로 표현한 관계형 프레임 워크로 설계함에 따라 한 지식의 추가, 삭제 등의 변경이 관련된 지식을 표현하는 모든 지식에 영향을 미치고 여러 층에 걸친 추상화 값을 얻는 과정이 복잡하다. 또한, 지식 추상화 계층에서 하나의 층을 모두 같은 도메인으로 관리하기 때문에 하나의 지식이 삭제 될 경우에는 층을 유지하기 위해 가짜의(dummy) 지식을 유지해야 하는 복잡한 지식 관리가 필요하다.

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 구축된 하이브리드 지식베이스를 설명한다.

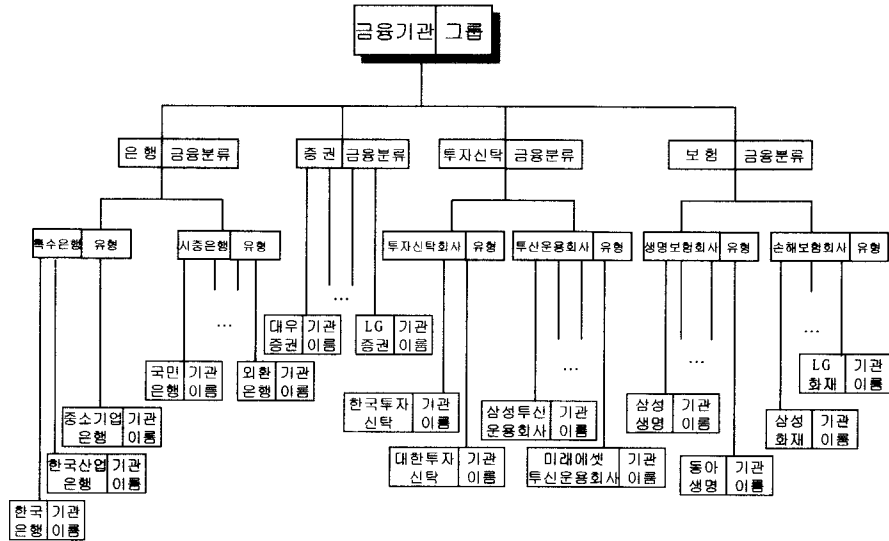
3.1 CKAH

본 논문은 값과 도메인을 하나의 통합 지식 정보로 유지하고 다단계 지식으로 표현하여 통합 지식 추상화 계층인 CKAH를 제안한다. 본 논문에서 사용하는 도메인의 의미는 속성이 가질 수 있는 값의 범위로 정의한다. 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 속성이 가질 수 있는 값의 범위인 하나의 도메인을 갖는다.
- 가정 2. 하나의 CKAH내에서 중복된 속성 값이 존재할 수 있다.
- 가정 3. 도메인은 여러 속성의 도메인이 될 수 있다.
- 가정 4. 하나의 도메인은 한 CKAH에 속하고 하나의 추상화 도메인을 갖는다.

CKAH는 지식 추상화 계층과는 달리, 부모 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 가장 세분화된 값을 갖는다. 하나의 추상화 값은 하위 층에 존재하는 여러 개의 세분화 값을 갖지만 여러 개의 세분화 값은 오직 하나의 추상화 값만을 갖는다. CKAH는 협력적 질의 응답을 제공할 뿐만 아니라 지식베이스의 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 추상화 관계가 함께 규정되어 있다.



(그림 1) 금융기관에 대한 CKAH의 예

3.2 하이브리드 지식베이스

기존에는 규칙, 시멘틱 넷(semantic net), 프레임(frame)등이 주된 지식 표현 방법이다[7,9]. 이러한 형태의 지식 표현의 변형된 구조나 두개 이상의 통합 형태를 갖는 하이브리드 지식 구조가 지식 표현의 한 형태로 널리 연구되고 있다[1,7,9]. 본 연구에서는 빠른 지식의 검색을 추구하고 용이한 지식관리를 위하여 시멘틱 리스트(semantic list)와 규칙 기반의 지식 표현을 통합하는 하이브리드 지식베이스를 제안한다.

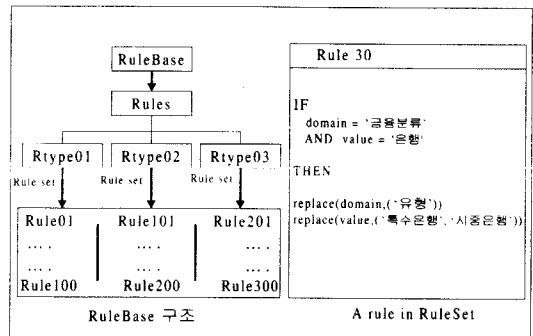
시멘틱 리스트는 CKAH의 추상화 개념간의 관계를 표현하기 위한 트리 형태의 지식 표현으로 각 노드가 포인터로 연결되어 있고 노드의 구조는 다음과 같다.

데이터		Rule구분자	PLINK (부모)	CLINK (자손)	LLINK (형제)	RLINK (형제)
DOMAIN	VALUE					

(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, 임의의 층까지의 추상화 값은 부모 포인

터를 이용해 직접적으로 찾아간다. 전체 지식베이스는 모든 CKAH의 지식으로 구성되어 있으며 하나의 CKAH 내의 지식들은 유사한 종류의 지식들로 분류함으로써 모듈화(modularization)[16]를 부여하고 시멘틱 리스트의 규칙 구분자를 통해 접근하도록 함으로써 검색의 시간을 단축하였다. 다음은 규칙 기반의 지식베이스의 규칙의 구성을 나타내고 있다.



(그림 3) 유사지식을 포함한 규칙들간에 모듈화 구조

(그림 3)에서 표현된 Rtype01, Rtype02, Rtype03 등

의 규칙 집합들은 한 CKAH의 지식들로 구성되어 있고 시멘틱 리스트의 노드에서 규칙 구분자가 이 규칙의 집합을 나타내고 있으므로 시멘틱 리스트에서 추상화 개념을 얻는 경우에 관련된 세분화 규칙을 포함하는 규칙의 집합의 구분자를 통해 관련있는 유사한 규칙에 빠른 접근이 가능하다.

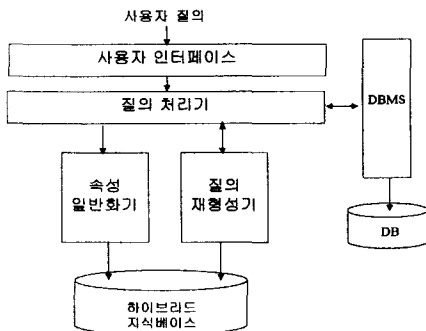
따라서, 추상화 개념을 추출함으로써 일반화의 정보를 얻기 위해서는 시멘틱 리스트에 의해 한 개체의 추상화 값과 슈퍼 도메인(즉, 상위 층의 도메인)을 추출하고 관련된 규칙 구분자의 정보를 통해 해당되는 CKAH의 지식을 갖고 있는 지식베이스에 접근하여 세분화 개념을 얻는 과정을 수행한다.

4. 지능적인 질의 응답 시스템

본 장에서는 속성을 기반으로 구축된 데이터베이스 상에서 체계화된 질의 완화를 지원하는 지능적인 질의 응답 시스템, IQAS를 설명한다. IQAS는 저축상담을 목적으로 고객이 원하는 저축 상품을 컨설팅 하는 지능적인 질의 응답 시스템이다. 고객이 원하는 저축상품과 일치하는 상품이 없는 경우에는 가장 유사한 저축상품을 결과로 제시하도록 근접 질의를 수행하고 데이터베이스와 관련된 지식이 없는 사용자들을 지원하기 위해 개념적 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, 속성 일반화기, 질의 재형성기, 질의 재형성기, 하이브리드 지식베이스로 구성된다.



(그림 4) IQAS의 전체 시스템 구조

질의어는 사용자 인터페이스를 통해 입력되고 입력된 질의는 질의 처리 과정을 수행함으로써 결과가 제시된다. 사용자가 입력한 초기 질의의 수행 결과가 데이터베이스 상에 존재하지 않는 경우에는 질의 완화 과정이 실행된다. 질의 완화 과정은 속성의 일반화기와 질의 재형성기에 의해 하이브리드 지식베이스의 지식을 이용하여 이루어진다. 속성 일반화기는 사용자에게 완화될 속성을 선택하도록 하여 한 속성을 입력받고 그 속성의 값에 대한 추상화 값을 탐색함으로써 완화된 조건의 속성 값을 질의 재형성기에게 제시한다. 질의 재형성기는 속성 일반화기가 제시한 속성의 추상화 값을 실제적으로 데이터베이스에 저장된 데이터 값으로 재작성 될 수 있도록 세분화 값을 탐색하여 질의를 변형한다. 마지막 단계로, 재형성된 질의는 피드백(feedback)과정으로 다시 질의 처리기에 의해 질의 처리 과정을 수행한다.

4.2 기본적인 지식 탐색 연산자

IQAS 에서 지식을 탐색하는데 필요한 기본적인 연산자(primitive operator)는 다음과 같다.

4.2.1 추상화 값의 탐색 연산자

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

Algorithm : Search_Abstract
// 한 노드의 상위 도메인과 상위 속성 값을 찾는 알고리즘 //
// input // var X : char
// output // R_domain, R_value, Id : char
R_domain = X ↑.PPLINK ↑.DOMAIN
R_value = X ↑.PPLINK ↑.VALUE
Id = X ↑.PPLINK ↑.ID
RETURN(R_domain, R_value, Id)

(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 층은 질의어에 표현된 속성이 속하는 도메인의 층이다. 질의 완화를 위하여 사용되는 지식은 추상화 값을 표현하기 위하여 지식의 추상화 관계를 다단계 계층 구조로 표현한 형태를 기반으로 하여 얻어진 개념이기 때문에 데이터베이스 상에서 사용자가 유용하게 사용할 실제적인 데이터의 추출은 질의어에서 사용한 속성과 같은 층의 값을 갖는 데이터이다. 따라서, 질의어로 작성된 속성이 CKAH상에서 속하는 도메인의 층에서부터 임의의 층까지 n-레벨의 추상화 값을 얻었다면, 역으로 추상화 값을 데이터베이스의 인스턴스로 존재하는 값, 즉 질의어에 나타나는 속성의 도메인 층까지 세분화 시켜야 초기 질의어를 완화된 조건을 포함한 질의어로 재형성 할 수 있다. IQAS에서는 규칙 기반의 지식베이스를 이용하여 세분화 값을 얻는다. 다음 연산자는 규칙 기반의 지식베이스에서 세분화 값을 탐색하는 규칙 탐색 연산자이다. 규칙 탐색 연산자를 사용하여 추상화 도메인과 추상화 값인 R_domain, R_value의 세분화 값을 찾는다. 도메인 R_domain과 속성 값 R_value 그리고 규칙 구분자 Id는 시멘틱 리스트를 통해 얻어진 추상화 지식이다. 따라서, 세분화 값의 탐색 연산자는 도메인 R_domain과 속성 값 R_value 의 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. 서론에서 설명하였듯이, 근사적 질의 응답은 사용자가 입력한 질의어의 해당 데이터가 없는 경우에 질의의 범위를 넓혀주는 과정이고 개념적 질의 응답은 데이터베이스에 대한 사용자의 지식이 적은 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 없다. 따라서, 일반 사용자와는 무관하게 지식의 검색하는데 있어서 자동적으로 근사적 질의 유형인지 개념적 질의 유형인지를 분석하는 과정이 선행되어야 한다.

본 논문에서 제안한 CKAH의 통합 지식의 한 개체

는 도메인과 값으로 이루어져 있다. 이 때, 한 개체의 통합 지식은 같은 층이 속하는 도메인과 값으로 구성된다. 시멘틱 리스트상에서 직접적으로 한 노드로 접근하기 위해 해쉬 함수를 이용하는데 빠른 질의 처리를 위해 해쉬 함수를 사용하는 것 이외에 질의 유형을 분류하기 위하여 해쉬 함수를 사용한다. 시멘틱 리스트를 이용한 지식 추출의 선행단계로 값과 도메인의 복합키를 이용한 해쉬 함수값을 통해 직접적인 노드 주소를 알아내는데 이 때는 복합키에 의해 해쉬 함수 결과값이 제시되는 경우이다. 즉, 같은 층이 속하는 도메인과 값으로 구성된 복합키를 이용한 해쉬 함수의 결과가 직접적인 노드 주소를 나타내지만 만일, 도메인과 값이 다른 층의 개념이라면 해쉬 함수의 결과는 주소의 값이 없는(null) 것으로 제시된다. 따라서, 해쉬 함수에 의한 통합 지식의 주소지 사상(mapping) 결과는 질의 유형 분류에 유용하게 사용된다. 분류된 질의 유형에 따라 질의를 처리하는 연산자가 필요하다. 먼저, 근사적 질의를 처리하는 연산자와 개념적 질의를 처리하는 연산자를 정의하고 개념적 질의를 처리하는 과정에서 필요한 기본적인 연산자를 설명한다. 두 가지 유형의 질의를 분석하는 과정은 4.3절의 전체적인 질의 처리 알고리즘중 속성 일반화 알고리즘에서 기술한다.

- 근사적 질의 응답(approximate query answering)
 근사적 질의 응답은 질의어에서 작성된 속성이 속하는 도메인과 속성 값을 이용하여 상위 층의 도메인과 추상화 값을 찾는 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, R_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 층에 있는 경우는 사용자가 질의어로 입력한 속성 값의 개념이 데이터베이스 상의 개념 보다 일반적인 개념을 입력한 것이다. 따라서, 값이 더 높은 층에 있는 경우에는 도메인의 층에서부터 값으로 세분화 개념을 결과로 제시하게 된다. 속성 값이 속성이 의미 하는 개념보다 세분화 값으로 입력되었다면 사용자가 너무 구체적인 개념으로 질의한 것이다. 따라서, 데이터베이스 상의 개념보다 세분화 값으로 질의한 것이므로 속성 값을 속성이 속하는 도메인 층까지의 개념으로 값을 추상화 시켜 그 추상화 값을 결과로 제시한다. 다음은 개념적인 질의 처리 알고리즘을 설명한다.

```

Algorithm : Conceptual_Query
// 초기 질의에 대한 개념적 질의를 수행하는 알고리즘 //
// input // var Domain, Value : char
// output // Result(*) : char

T↑ = Head
Domain# = 0
Vdomain# = 0
Domain_temp = null
call Level_Check(T, Domain, Value, Domain#, Vdomain#,
                Domain_temp)
IF Domain# = 0 & Vdomain# = 0 & Domain_temp = null
THEN RETURN(null)
IF Domain# > Vdomain# THEN
    LOOP UNTIL(R_domain = Domain)
    X=READ(Domain_temp, Value)
    Search_Abstract(X, R_domain, R_value, Id)
    ENDOLOOP
ELSE
    Search_RuleBase(Domain_temp, Value, Id, Domain, Result(*))
RETURN(Result(*))
    
```

(그림 7) 개념적 질의 응답 연산자 알고리즘

개념적인 질의는 도메인과 속성 값의 층이 같은 층의 개념인지 다른 층의 개념인지를 파악하여 분류되는데 사용자가 질의한 속성과 속성 값만을 보고 층의 차이를 파악할 수는 없다. 따라서, 도메인과 속성 값의 층이 다른 경우에 층간의 차이를 계산하기 위하여 전위 탐색(preorder search)을 하며 층의 차이를 파악하고 도메인과 속성 값 중 어느 것이 상위 층인가를 살펴본다. 도메인과 속성 값의 복합기를 이용하여 해쉬 함수에 의한 직접적인 주소 접근이 안되는 경우 개념적 질의 유형으로 분류되는데 도메인과 속성 값을 분

리하여 각각을 전위 탐색하며 탐색 순서를 부여한다. 도메인이나 속성의 값이 일치하는 개념을 각각 찾았을 때 도메인과 속성의 탐색 순서를 비교하여 어느 개념이 상위인가를 파악한다. 그리고, 도메인과 속성 값의 층을 비교하여 기준점, 즉 해당 속성 층을 기준으로 개념을 추상화 또는 세분화시켜 가는 과정이 질의 완화 과정이다. (그림 8)은 전위 탐색을 하며 도메인과 속성 값 간의 층의 차이를 파악하는 알고리즘을 설명한다.

```

Algorithm : Level_Check
// 개념적 질의 수행을 하기에 앞서 속성과 속성 값의 상대적인
// 층의 차이를 파악하는 알고리즘 //
// input // var T : char
//          Domain, Value : char
// output // Domain_temp : char
//          Domain#, Vdomain# : integer

IF T ^= null THEN BEGIN
    count = count + 1
    IF T↑.DOMAIN = Domain THEN Domain# = count
    IF T↑.VALUE = Value THEN BEGIN
        Vdomain# = count
        Domain_temp = T↑.DOMAIN
    END
    IF Domain# > 0 & Vdomain# > 0 THEN
        RETURN(Domain#, Vdomain#, Domain_temp)
        // 속성과 속성 값의 단계 파악 //
    END
    ELSE RETURN(0, 0, null)
    call Level_Check(T↑.CLINK, Domain, Value, Domain#, Vdomain#,
                    Domain_temp)
    call Level_Check(T↑.LLINK, Domain, Value, Domain#, Vdomain#,
                    Domain_temp)
    
```

(그림 8) 도메인과 속성 값 간의 층의 차이를 파악하는 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

는 유형의 질의인가를 검사하여 해당되는 유형별 질의로 속성에 의해서 속성이 속하는 계층과 도메인을 찾고, 속성과 속성 값을 검색키(key)로 하여 계층 내에서 노드를 검색한다. 검색키로 검색이 되지 않는 경우에는 속성의 층과 속성 값의 층 차이 때문에 생기는 것이므로 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 언어진 일반화의 개념을 다시 세분화 시켜서 초기 질의어를 재형성하는 과정이다.

```

Algorithm : Query Processing
// 질의가 입력되어 실행되는 알고리즘 //
// input : query //
// output : query result //
get user query
perform DB search about query
IF exist(query result) THEN display(query result)
Attribute_list = ∅ // 완화한 속성의 리스트 //
request input(Attribute) // 완화할 속성을 사용자가 지정 //
WHILE ( Attribute ~= null & Attribute ∉ Attribute_list ) DO
    call Attribute_Generalization(Attribute, Value, Value_list(*))
    Attribute_list = Attribute_list ∪ Attribute
    IF Value_list(*) ~= null THEN BEGIN // 완화 속성 값이 존재 //
        call Query_Reformulation(query, Attribute, Value_list(*),
            Reformulated_query)
        perform DB search about query
        IF ^exist(query result) THEN request input(Attribute)
    END
END // end of while //
IF exist(query result) THEN display(query result)
ELSE display(fail)
    
```

(그림 9) 질의 처리 알고리즘

```

Algorithm : Attribute_Generalization
// 한 속성에 대해 그 속성 값을 완화하는 알고리즘 //
// input // var Attribute, Value : char
// output // Value_list(*) : char

call Find_Hierarchy(Attribute, domain, hierarchy)
X=READ(domain, Value)
IF X=null THEN call Conceptual_Query(domain, Value, Value_list(*))
ELSE call Approximate_Query(X, Value_list(*))
RETURN(Value_list(*))
    
```

(그림 10) 속성 일반화 알고리즘

(그림 10)의 속성 일반화 알고리즘에서 Find_Hierarchy는 속성을 통해 계층과 도메인을 추출한다. READ(Domain, Value)는 도메인과 속성 값의 복합키를 사용하여 해쉬 함수를 이용하는 부분이다.

(그림 11)은 질의 재형성 알고리즘을 설명하고 있다.

```

Algorithm : Query Reformulation

// 완화된 속성 값에 의해 초기 속성의 값을 재 작성하는 알고리즘 //
// input // var Query, Attribute, Value_list(*) : char
// output // Reformulated_Query : char

Replace values of Attribute with Value_list(*) in Query
RETURN(Reformulated_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 관점에서 사용자의 평가 기준으로 관련 있는 결과들을 미리 선택한 후, 근사적 질의 응답으로 제시된 결과들을 기반으로 recall 과 precision을 계산하여 성능 평가를 수행하였다. recall 과 precision은 다음과 같이 정의된다.

$$recall = \frac{\text{검색된 관련 있는 저축 상품의 수}}{\text{관련 있는 저축 상품 결과의 수}}$$

$$precision = \frac{\text{검색된 관련 있는 저축 상품의 수}}{\text{검색된 총 저축 상품 결과의 수}}$$

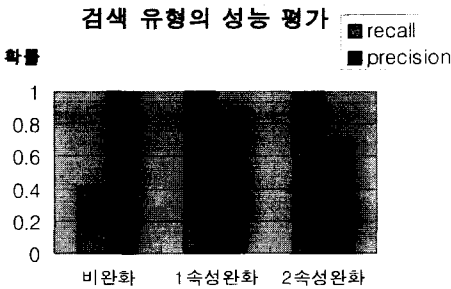
성능 평가를 하기 위하여 먼저, 정확하게 일치하는 데이터만을 결과로 제시하는 일반적인 검색과 질의 재형성을 통한 질의 피드백 과정을 수행하여 근사적 질의 결과를 나타낸 검색 결과를 비교하여 측정하였다.

총 24개 기관의 120 여개의 저축상품 중 저축상품명, 투자기간, 수익률에 따라 내역이 유사하여 관련 있다고 정의한 상품 32개를 기반으로 질의어와 정확하게 일치하는 내용만을 검색한 결과와 한 개의 질의 속성을 완화한 검색 결과(저축상품명을 선정), 두 개의 질의 속성을 완화한 검색을 수행한 결과(저축상품명, 투자기간을 선정)가 <표 1>과 같다.

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

검색 유형	Retrieved	Relevant	Recall	Precision
일치 검색	13	13	0.41	1
1개 속성 완화 검색	38	32	1	0.84
2개 속성 완화 검색	47	32	1	0.68

<표 1>의 성능 평가에 대한 차트는 (그림 12)와 같다. 즉, 질의 완화를 이용한 근사적 질의 수행 결과는 관련 있는 상품을 검색하는 비율이 높아진다.



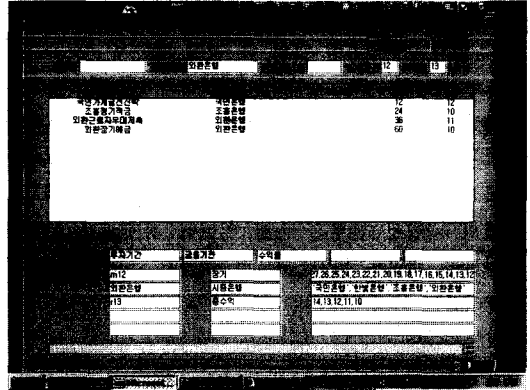
(그림 12) 검색 유형의 성능 평가

6. 구 현

본 논문에서 제안한 지능적인 질의 응답 시스템은 AIX 4.1.3 하에 Oracle 8i, VisualCafe 3.0, JDK 1.2.1 을 이용하여 구현되었다. 상품명, 금융기관, 수익률, 투자기간, 안정성, 저축유형, 세금 유형 등을 화면에서 입력하여 질의할 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 완화화과정을 통해 속성 값의 범위를 완화한 경우에 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 질의할 수 있다.

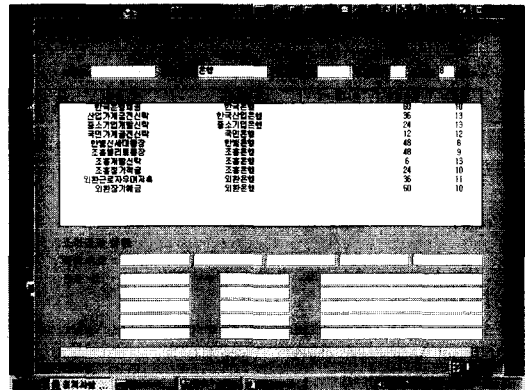
다음은 근사적 질의 응답을 수행한 결과 화면이다. 사용자의 질의어와 정확하게 일치하는 결과가 없어 질

의어와 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 질의가 입력되어 시스템 내부적으로 좀더 구체적인 값으로 질의가 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, 질의를 재형성하고 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 '금융분류'인



(그림 14) 개념적 질의 응답 결과 화면

데 속성 '금융기관'의 도메인 '기관이름' (속성과 속성이 속하는 도메인은 테이블로 관리)과 층의 차이를 비교해 볼 때(알고리즘 Level_Check에서 수행) '금융분류'가 '기관이름'보다 상위 도메인임을 알 수 있다. 즉, 속성 값이 너무 일반적인 값으로 입력되었다는 의미이므로 현재 속성 '금융기관'이 속하는 도메인 '기관이름'의 층까지의 단계 만큼 속성 값을 구체화 시켜야 한다. 따라서, 속성 값 '은행'의 세분화 값을 규칙 기반의 지식베이스를 통해 얻어내고 이 값으로 질의를 재형성하여 질의 처리 과정을 재 수행한다.

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 웹 기반 하에서 사용자에게 금융 상품을 컨설팅하고 정확하게 사용자의 질의와 일치하는 자료가 없는 경우에는 유사한 자료를 제시한다. 본 연구에서는 값간의 추상화와 세분화 개념을 표현하는 다단계 지식의 계층 관계를 표현하기 위해 CKAH를 제안하였다. CKAH를 이용하여 속성과 값 및 각 개념간의 관계를 하나의 통합체로 유지함으로써 지식베이스를 단순하고 용이하게 관리되도록 하였다. 시멘틱 리스트와 규칙기반의 지식 표현 구조를 통합한 하이브리드 지식베이스의 지식 표현 구조를 제안하고 하이브리드 지식베이스의 지식을 사용한 질의 완화를 시도하였다. 시멘틱 리스트에 추상화 지식을 표현하고 규칙기반의 지식베이스에 세분화 지식을 분류하여 표현함으로써 지식 검색의 유연성을 부여하였다. 규칙기반의 지식베이스는 결과가 사용자에게 설명하기 용이한 장점을 이용하여 완화 과정을 설명하였다.

질의 완화에 있어서 지식베이스는 필수적인 요소이다. 질의완화를 통한 사용자 질의를 재 작성하는 과정은 지식베이스와 연계되어 이루어진다. 규칙기반의 지식베이스는 상업적인 도메인 상에서 많이 적용하는 지식표현기법으로 인간의 추론체계와 유사하여 지식의 유추과정을 이해하기가 쉽다는 장점이 있는 반면, 지식이 많아질수록 지식을 분류, 검색하기가 어려워지는 단점을 갖고 있다. 많은 양의 지식을 분류하기 위하여 시멘틱 리스트 지식을 통해 지식 구분자를 탐색하여, 해당 계층의 지식만을 모듈화 한 규칙베이스를 탐색한다.

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

참 고 문 헌

- [1] E. Baralis, S. Ceri, S. Paraboschi, "Modularization Technique for Active Rules Design," ACM Transactions on Database Systems, Vol.21, No.1, pp.1-29, March 1996.
- [2] Y. Cai, N. Cercone databases, "Data-Driven Discovery of Quantitative rules in relational databases," in IEEE transactions on Knowledge and data engineering, Vol.5, No.1, pp.29-40, 1993.
- [3] W. Chu, Q. Chen, "A Structured Approach for Cooperative Query Answering," IEEE transaction on knowledge and data engineering, Vol.6, No.5, OCT, pp.738-749, 1994.
- [4] W. Chu, Q. Chen, R. Lee, "Providing Cooperative Answers via Knowledge-Based Type Abstraction and Refinement," in Proc. of the 5th International Symposium on Methodologies for Intelligent Systems, Knoxville, TE, 1990.
- [5] W. Chu, H. Yang, and G. Chow, "A Cooperative Database System(CoBase) for Query Relaxation," in Proc. of the Third International Conference Query Answering," IEEE Transactions on Knowledge and Data Engineering, Vol.6, No.5, pp.738-749, October 1994.
- [6] F. Cuppens and R. Demolombe, "Cooperative Answering: A Methodologies to Provide Intelligent Access to Databases," in Proc. of 2nd Int. Conference on Expert Database Systems, pp.621-643, 1989.
- [7] S. Dutta, "Knowledge processing and applied artificial intelligence," Butterworth Heinemann, 1993.
- [8] J. Han, Y. Huang, N. Cercone, Y. Fu, "Intelligent Query Answering by Knowledge Discovery Technique," Vol.8, No.3, pp.373-390, June 1996.
- [9] P. H. Winston, Artificial Intelligence, 3rd Edition, 1993.
- [10] A. Motro, "FLEX: A Tolerant and Cooperative User Interface to Databases," IEEE Transactions on Knowledge and Data Engineering, Vol.2, No.2,

pp.231-246, June 1990.

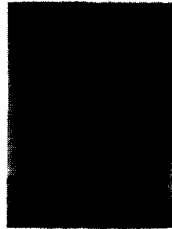
- [11] A. Motro, "Intensional Answers to Database Queries," IEEE Transactions on Knowledge and Data Engineering, Vol.6, No.3, pp.444-454, June 1994.
- [12] A. Motro, "VAGUE: A User Interface to Relational Databases that Permits Vague Queries," ACM Transactions on Office Information Systems, Vol.6, No.3, pp.187-214, July 1988.
- [13] R. Sebesta, "Concepts of Programming Language, 2nd," The Benjamin/Cummings Publishing Company, 1993.
- [14] J. D. Ulman, Database and Knowledge-Based Systems, Vol.1, Computer Science Press, 1987.
- [15] S. V. Vrbsky and W.S. Liu, "APPROXIMATE- A Query Processor that Produces Monotonically Improving Approximate Answers," IEEE Transactions on Knowledge and Data Engineering, Vol.5, No.6, December 1993.
- [16] 최옥현, 소명옥, 하장승, 류길수, "하이브리드 지식 표현을 이용한 선박 엔진 고장 진단 시스템의 개발에 관한 연구", '98한국 전문가 시스템학회 추계학술대회, pp.363-371, 1998.
- [17] 허순영, 문개현, "값과 도메인의 추상화에 관한 지식을 이용한 협력적 질의 응답에 관한 연구", 석사학위논문, 한국과학기술원, 1996.
- [18] 허순영, 신완규, "확장된 지식 추상화 계층을 이용한 관계형 데이터베이스에서의 협력적 응답", 석사학위논문, 한국과학기술원, 1996.



황혜정

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(전신 산업연구원) 책임연구원
 관심분야 : 질의 완화, 멀티미디어



김교정

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



윤용익

e-mail : yiyoon@cs.sookmyung.ac.kr
 1983년 동국 대학교 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



윤석환

e-mail : yoonsh@mail.iita.re.kr
 1982년 아주대학교 산업공학과 (공학사)
 1984년 건국대학교 산업공학과 (공학석사)
 1996년 아주대학교 산업공학과 (공학박사)

1992년 품질관리기술사 취득
 1986년~현재 한국전자통신연구원/정보통신연구원 책임연구원
 1995년~현재 한국정보처리학회지 학회지 편집위원장
 관심분야 : 그룹웨어, 생산정보시스템, 지식관리시스템, S/W 공학, 개발방법론