

관계형 데이터베이스 시스템을 위한 지능적인 질의처리 시스템

An Intelligent Query Processing System for Relational Database System

김대수 · 김창석*

Dae Su Kim and Chang Suk Kim*

한신대학교 전자계산학과
*부산외국어대학교 컴퓨터공학과

요 약

본 논문에서는 관계형 데이터베이스 시스템을 위한 새로운 지능적인 질의처리 시스템을 제안하였다. 애매한 질의와 관련된 기존의 연구 결과를 분석하여 그 문제점들을 해결할 수 있는 새로운 질의처리 시스템이 개발되었는데, 지능적 질의처리기를 비롯한 각 모듈들의 역할을 정의하였으며, 구문분석기, 질의변환 모듈, 추론 엔진, 의미 데이터베이스 모듈, 결과합성기 등의 알고리즘을 제시하였다. 또한 제안된 질의처리 시스템에 기반한 실제 작동의 과정을 한 예를 들어 검증한 결과 타당성 있는 결과가 입증되었으며, 사용자의 일반적인 질의뿐만 아니라 애매한 질의에도 지능적으로 대처하여 순서화된 결과를 보여주고 있다.

ABSTRACT

In this paper, we propose a new intelligent query processing system for relational database systems. By analyzing previous research results related with fuzzy queries, a new intelligent query processing system is developed and the role of each module including intelligent query processor is defined and some algorithms for parser, query translation module, inference engine, semantic DB and result composer are suggested. By applying a typical example to the proposed intelligent query processing system, reasonable results for the ambiguous query are drawn, and therefore it shows a promising model returning ordered result for both the ambiguous queries and general queries.

1. 서 론

최근 들어 컴퓨터나 인터넷을 통한 정보검색에 대한 관심이 급증하고 있으며 사용자 측면에서 보다 편리하게 이용할 수 있는 지능적인 질의처리 시스템 개발에 대한 요구가 대두되고 있다.

기존의 관계형 데이터베이스 시스템(RDBMS)에서 정보검색은 사전 지식이나 경험이 있는 사람에게 있어서는 자기가 원하는 정보를 키워드(keyword)를 이용하여 쉽게 검색할 수 있다. 그러나 이 경우에도 정확하게 매치(match)되는 키워드가 없거나 모를 경우는 원활한 자료 검색이 어려운 실정이다.

예를 들어 "키가 매우 크고 수학 성적이 뛰어난 학생들의 이름을 출력하십시오."라는 질문에 대해 현재의

관계형 데이터베이스로는 검색이 전혀 불가능한 실정이다. 그러나 애매한(fuzzy) 개념을 사용할 경우에는 그 관계를 나타내는 적절한 연산자(operator)들에 따라 어느 정도 오차를 가진 정보들을 DB로부터 검색이 가능하다[3-4, 6-8]. 따라서 사용자들에게 사용하기 편리하고 인간 중심의 지능적인 정보검색을 가능하게 해주는 애매한 질의 처리에도 적절히 대응할 수 있는 지능적인 질의 처리 시스템의 도입이 절실히 요구되고 있다. 따라서 이를 위한 지능적인 모델링에 대한 이론적인 연구와 더불어 이것을 원만하게 실현할 수 있는 시스템의 구현이 필수적으로 요청되고 있는 실정이다.

애매한 질의를 원만하게 처리할 수 있는 시스템이 아직까지도 기초적인 연구에 국한되고 실용화되지

못한 주요 이유로는 아직까지 애매한 질의를 처리할 수 있는 시스템의 성능이 극히 낮다는 점, 시스템간의 호환성 문제, 그리고 사용자의 인식 부족 문제를 들 수 있다[9, 11~13]. 그러나 가장 근본적인 문제점으로 는 현재 일반적으로 쓰이고 있는 RDBMS를 활용할 수 없다는 점과 애매한 질의를 편리하게 연결해 줄 수 있는 지능적인 사용자 인터페이스, 구문분석기, 추론 엔진과 의미 데이터베이스를 활용한 질의 변환 모듈 등이 제대로 연구되지 않았다는 점일 것이다[14, 15]. 따라서 이 논문에서는 위에서 지적한 문제점들과 요구 사항들을 수용하여 일반적인 질의뿐만 아니라 애매한 질의까지도 처리할 수 있는 지능적인 질의 처리 시스템을 제안하고 이들의 연관관계를 분석하여 실제 예를 통하여 처리 과정을 분석함으로써 새롭게 제안한 모델의 타당성을 입증하며, 또한 가능한 문제점들을 도출한다.

2. 관련 연구

애매한 질의를 처리하기 위한 연구들이 지난 20여 년 동안 수행되어 왔다. 실생활에서 누구나 사용할 수 있는 질의의 경우에는 애매하거나 부정확한 질의가 흔히 있게 마련이다. 이러한 질의를 기존의 RDBMS에 적용시킬 때 모든 경우에 매치되지 않을 경우에는 질의에 대한 응답을 할 수가 없다. 여기서는 애매한 질의와 관련된 시스템에 관한 지금까지의 연구를 간략히 고찰하고 그 문제점들을 살펴본다.

2.1 ARES와 VAGUE

Ichikawa[10]는 "similar to" (\approx)라는 새로운 연산자를 도입한 ARES라는 시스템을 발표했다. 이것은 질의에 $A \approx a$ (ambiguous selection A는 임의의 도메인, a는 상수)와 $A1 \approx A2$ (ambiguous join)이라는 연산을 지원한다. 이런 애매한 연산은 유사 관계(similarity relation)를 이용하여 기존의 관계 연산으로 변환된다.

ARES의 단점은 애매한 연산을 명확한 연산으로 변환할 때 이용되는 유사 관계가 정규화되어 있지 않아 질의의 의미를 여러 사용자가 공감하지 못하는 데 있다. 즉 보통 퍼지 집합에서는 $\mu \in [0,1]$ 이므로 어떤 사용자라도 μ 의 값을 보고 그 의미를 짐작할 수 있는 것과 대조적이다.

Motro[11]는 ARES와 같은 "similar to" (\approx) 연산자를 사용하지만 정규화된 유사 관계를 사용하는 VAGUE를 개발했다. VAGUE는 데이터 사이의 유사한 정도를 거리(metrics)로 표현한다. 그러나 VAGUE도 오직 하나의 \approx 연산자밖에 지원하지 않는다. 또한

$A \approx a$ 를 $A = a \pm r$ (여기서 r은 radius로 데이터간의 간격을 의미함)로 변환되어 결과의 부정확성을 도출할 수도 있다.

2.2 DEDUCE2

Chang[5]이 개발한 DEDUCE2는 연역(deduction) DBMS를 확장하여 부울조건(Boolean condition) F1과 선택적인 불명확 조건 (optional imprecise condition) F2로 구성된 질의를 처리한다. F2는 "old", "around 45"와 같은 항(term)이 포함되며 AND, OR 연산자에 의해 여러 항들이 접속될 수 있다. 만약 F가 불명확한 항(imprecise term) T1과 T2를 AND 연산한다면 해당 튜플의 순위는 $\max(R1, R2)$ 가 되고, OR 연산을 한다면 $\min(R1, R2)$ 이 된다. 예를 들어 릴레이션 직원(이름, 봉급, 나이, 거주지)이 있다고 하자.

"서울에 살면서 봉급이 고소득(well-paid)이고, 나이가 40세 가량(around 40)인 사람을 찾아라"라는 질의가 주어지고, 이것의 결과가 다음과 같은 3개의 튜플로 출력되었다고 가정하자.

- 튜플 1 <kim, 11000, 38, 서울>
- 튜플 2 <Lee, 30000, 37, 서울>
- 튜플 3 <Cho, 10000, 39, 서울>

여기서 튜플 (봉급, 나이)을 기준으로 위의 세 튜플의 순서를 매기면 튜플1 (2, 2), 튜플2 (1, 3), 튜플3 (3, 1)이 된다. 이것을 봉급 AND 나이를 구하기 위해서 MAX 연산을 취하면 튜플1 (2), 튜플2 (3), 튜플3 (3)이 되어 kim이 주어진 질의를 가장 잘 만족 만족시키며, Lee와 Cho는 두 번째가 된다. 그러나 일반적인 관점에서는 튜플2가 더욱 질의를 만족함을 알 수 있다. 결국 DEDUCE2는 튜플의 순서 매김을 순위로 정함으로써 사람의 판단과 다른 결과를 가져오기도 한다.

2.3 SQL^f

Bosc[1, 2]는 표준 SQL을 확장하여 퍼지 질의를 처리할 수 있는 SQL^f 질의처리를 개발하였다. 이것은 상용 DB언어 (SQL)가 불명확한 질의를 포함할 수 있도록 확장하여 정의하였다. 예를 들어 직원 릴레이션에서 "고소득(well-paid)이고 젊은 사람(young) 10명을 찾아라"라는 질의는 다음과 같이 할 수 있다.

- Select 10 직원. 이름
- from 직원
- where 직원. 나이="young" AND 직원.봉급="well-paid";

Bosc의 접근 방법은 이런 질의를 처리할 수 있는 DBMS를 직접 설계하는 것으로서 퍼지 질의를 가장 적합하게 처리할 수 있는 장점이 있다. 그러나 전세

계적으로 볼 때 상용 관계형 DBMS가 널리 보급되어 있는 현실을 고려하면 기존 DBMS들과의 호환성이 없으므로 프로토타입 개발로 머무를 가능성이 높다.

(Intelligent Query Processor)의 구조를 제안하고 각 모듈의 기능을 설명한다. 또한 질의의 변환 과정과 처리 예를 보인다.

3. 지능적 질의처리기

본 장에서는 새롭게 제안된 일반화된 퍼지 관계형 데이터 모델에 기반을 둔 퍼지 질의의 처리 방법을 기술한다. 먼저 기존 관계형 데이터베이스 시스템의 전처리기로 사용 가능한 지능적 질의처리기인 IQP

3.1 지능적 질의처리기의 구조

IQP는 기존의 관계형 데이터베이스 관리 시스템에 부착하여 전처리기로 사용 가능하다. 즉, IQP를 기존의 관계형 데이터베이스 관리 시스템에다 접속하면 퍼지 관계형 데이터베이스 관리 시스템이 되고, 분리하면 원래대로 관계형 데이터베이스 관리 시스템으

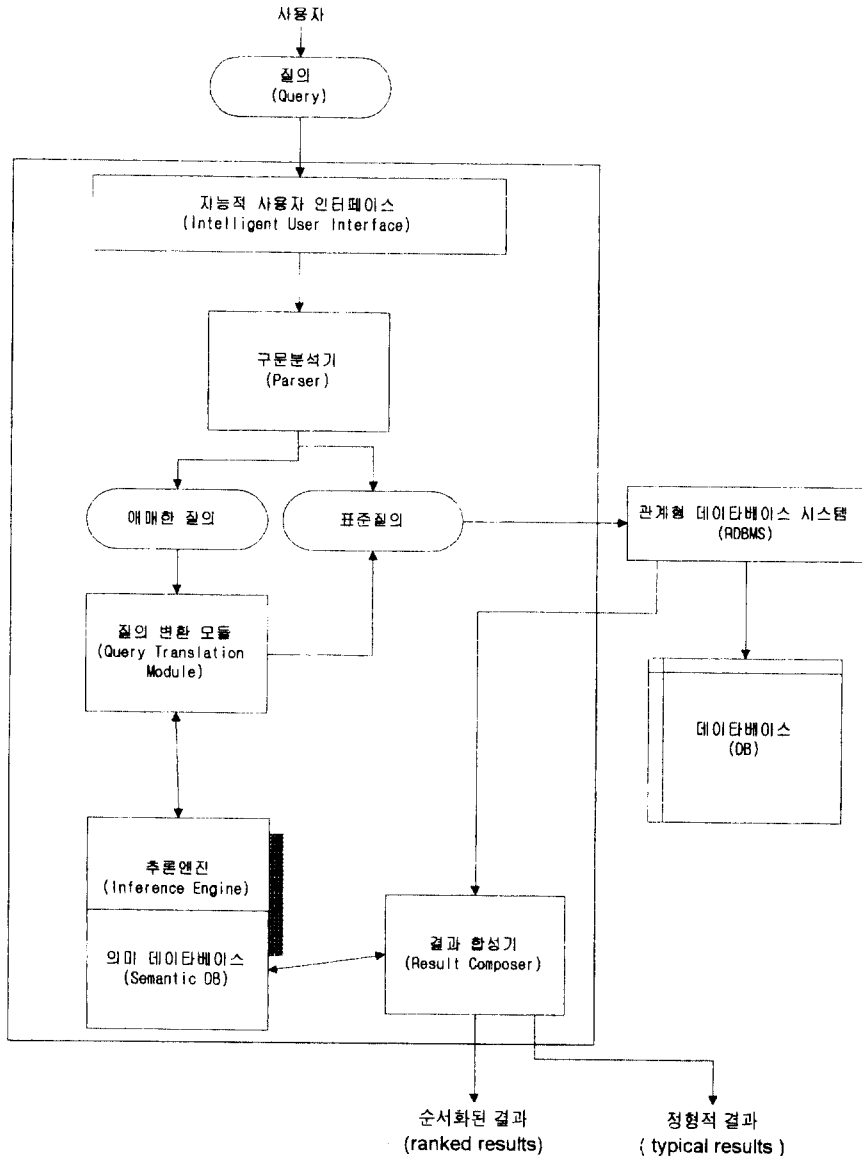


그림 1. 지능적 질의처리기 (Intelligent Query Processor)의 구조

로 사용 가능하다. 이러한 접근 방법은 퍼지 질의를 본 논문에서는 2차원적 데이터의 표현으로 보고 순수 데이터와 소속척도 데이터를 분리하여 처리하는 개념에 입각한 것이다. 결과적으로 이 접근 방법은 기존 관계형 데이터베이스 관리 시스템과의 호환성과 이미 구축된 관계형 데이터베이스와의 호환성을 가진다.

그림 1은 IQP의 구조와 기존 관계형 데이터베이스 시스템과의 관계를 블록 구성도로 나타낸 것이다. IQP는 지능적 사용자 인터페이스, 구문 분석기(parser), 질의 변환기(query translator), 추론 엔진 및 의미 데이터베이스, 결과 합성기(result composer)로 구성된다.

퍼지 질의가 처리되는 과정을 보면, GUI를 기반으로 하는 지능적 사용자 인터페이스를 통해 사용자가 질의를 입력하면 구문 분석기는 입력된 질의(query)을 읽어 표준 SQL 질의인 경우에는 그대로 관계형 DBMS에 보내고 애매한 질의인 경우에는 표준 SQL로 변화되어야 한다. 이때 퍼지 관련 토큰인 퍼지 상수와 퍼지 비교 연산자는 특별히 표시하여 질의 변환기가 인식 할 수 있게 한다. 질의 변환기는 퍼지 상수, 퍼지 비교 연산자를 의미 관계를 참조하여 표준 SQL의 상수나 비교 연산자로 변환한다. 소속 척도를 가진 2차원 데이터를 소속척도 개념이 없는 1차원 데이터로 변환하는 것이다. 기존의 관계형 데이터베이스 관리 시스템은 표준 SQL로 변환된 질의를 처리한 후 그 결과를 다시 IQP의 결과 합성기로 보낸다. 결과합성기는 관계형 데이터베이스 관리 시스템에서 출력된 결과들을 의미 관계를 이용하여 순서화된 결과로 합성한 후 최종 결과로 출력한다. 즉, 1차원 데이터를 의미 데이터베이스를 이용하여 다시 2차원 데이터로 환원한다.

추론 엔진 및 의미 데이터 베이스 관리기는 저장된 의미 데이터베이스에 접근하여 질의변환기나 결과합성기에 필요한 의미 데이터를 검색하거나, 퍼지상수나 근접관계를 생성한다. IQP를 구성하는 각각 구성요소들의 기능을 서술하면 다음과 같다.

3.2 구문분석기

구문분석기는 입력된 질의를 구문 분석하여 오류를 검사하고, 파스트리(parse tree)를 생성한다. 파스트리는 각 질의문의 의미를 적절한 자료구조체의 연결형태로 표현한 것으로 각 구조체 내에 질의문으로부터 추출한 정보를 저장한다. 구문 분석기는 어휘분석 모듈과 구문검사 모듈로 구성된다.

3.2.1 어휘분석 모듈

입력 문자열로부터 질의문의 각 문자를 토큰으로

분리한 후 토큰의 종류와 토큰 값을 돌려준다. 토큰의 종류에는 식별자(identifier), 키워드(keyword), 숫자, 인용 문자열, 연산에 쓰이는 각종 부호 그리고 구두점 등이 있다.

3.2.2 구문검사 모듈

어휘분석 모듈의 출력인 토큰을 읽어 들어 질의 구문에 맞는지를 확인하여 맞지 않으면 오류 처리한다. 오류가 없으면 각 문의 의미를 파스트리로 표현한다. 구문분석기의 수행절차를 요약하면 다음과 같다.

○ Procedure 구문분석

- 입력 : 질의문의 문자열
- 출력 : 문법 검사가 된 표준 질의, 또는 애매한 질의에 대한 파스트리

- 1) 질의문의 문자열을 읽어 들인다.
- 2) 어휘분석 모듈에게 토큰을 요구한다.
- 3) 어휘분석 모듈은 한 문자씩 검사하여 토큰으로 분리한 후 구문검사 모듈에게 토큰을 넘긴다.
- 4) 구문검사 모듈은 토큰이 해당 문의 문법에 맞는 지 검사한다.
- 5) 표준 질의와 애매한 질의를 구별한다. 표준 질의이면 그대로 출력한다.
- 6) 애매한 질의인 경우 토큰이 가지는 정보(식별자, 문자열, 연산자 등)를 추출하여 구성할 파스트리의 구조체 안에 저장한다.
- 7) 오류가 없으면 파스트리를 만들고, 오류이면 오류번호를 돌려준다.

3.3 질의 변환 모듈

질의 변환기는 파스트리를 입력으로 받아들여 의미 데이터베이스 관리 및 생성기를 호출하여 퍼지 상수나 퍼지 비교 연산자를 포함한 프레디킷트를 표준 SQL의 프레디킷트로 변환한다. 바뀐 파스트리는 다시 SQL 문자열로 바꾸어 일반 관계형 데이터베이스 시스템이 처리할 수 있는 형태로 만든다. 질의 변환기는 질의변환 모듈과 SQL 생성 모듈로 구성된다.

3.3.1 질의 변환 모듈

질의 변환 모듈은 파스트리에서 퍼지 상수와 퍼지 비교 연산자가 포함된 프레디킷트를 찾아 보통상수와 보통 비교 연산자로 구성된 프레디킷트로 변환하는데 변환 유형은 표 1과 같다. 표 1에서 변환 유형은 3가지 변환타입과 변환 불필요 및 변환 의미가 없는 무의미 타입으로 분류할 수 있다. 여기서 퍼지상수 간의 대소 비교와 유사 비교는 의미가 없는 것으로 간주한다. 또한 보통상수의 비교는 IQP에서 수행하지 않고 관계형 데이터베이스 관리 시스템에서 수행하므로 변환 불필요 타입으로 분류한다. 그의 변환타입-I과 변환타입-II

표 1. X θ Y에 대한 퍼지 질의 변환 유형

도메인 의 연속성		θ	Y	
			= \neq	< > $\leq \geq \approx ! \approx \ll \gg$
연 속	보통상수	변환 불필요		변환타입-III
	퍼지상수	변환타입-I	변환무의미	
불 연 속	보통상수	변환불필요		변환타입-III
	퍼지상수	변환타입-II	변환무의미	

및 변환타입-III의 변환과정은 다음과 같다.

• 변환타입-I

변환타입-I이란 프레디킷트가 '속성명 θ 퍼지상수' 형태이며, 퍼지상수는 연속 도메인이고 θ 는 = 또는 \neq 보통 비교연산자이다. 질의변환에서 고려할 사항으로는 FLEVEL로 표현되는 α 수준 집합이다. 변환타입-I 변환과정은 해당 퍼지상수를 의미 데이터베이스 관리 및 생성기를 통해 α 수준이상의 데이터 값만 읽어온다. 이때 정의된 퍼지 상수는 볼록 퍼지집합(convex fuzzy set)이므로 α 수준과 소속함수와 교차하는 곳의 데이터가 최소 데이터 값과 최대 데이터 값이 된다. 이것을 해당 속성명으로 범위 프레디킷트로 만들면 변환이 끝난다. 이 과정을 요약하면 다음과 같다.

○Procedure 변환타입-I 변환

• 입력 : '속성명 θ 퍼지상수' 프레디킷트(도메인이 연속공간)

• 출력 : 최소값 \leq 속성명 \leq 최대값 조건절

- 1) 입력 퍼지상수가 의미 데이터베이스에 접근한다.
- 2) 의미 데이터베이스 중 α 수준 이상만 선정한다.
- 3) 선정된 데이터 값 중 최소값과 최대값을 구한다.
- 4) '최소값 \leq 속성명 \leq 최대값' 조건절을 만든다.

○변환타입-II

변환타입-II란 프레디킷트가 '속성명 θ 퍼지상수' 형태이며, 퍼지상수는 불연속 도메인이고 θ 는 = 또는 \neq 보통 비교연산자이다. 도메인의 성격이 불연속적이라는 것이 변환타입-I과 다른 점이다. 변환타입-I과 마찬가지로 질의변환에 고려할 사항으로는 FLEVEL로 표현되는 α 수준 집합이다. 변환타입-II 변환과정은 테이블 형태로 저장되어 있는 해당 퍼지상수를 의미 데이터베이스 관리 및 생성기를 통해 α 수준이상의 데이터 값만 읽어온다. 이것을 해당 속성명과 θ 로 프레디킷트를 만들고 OR 연산자로 연결한다. 이 과정을 요약하면 다음과 같다.

○Procedure 변환타입-II 변환

• 입력 : '속성명 θ 퍼지상수' 프레디킷트(도메인이 불연속공간)

• 출력 : '속성명 θ 보통상수 OR... 속성명 θ 보통상수' 조건절

- 1) 입력 퍼지상수로 의미 데이터베이스에 접근한다.
- 2) 의미 데이터베이스 중 α 수준 이상만 선정한다.
- 3) 선정된 데이터 값을 '속성명 θ 보통상수' 형태로 OR 연산을 한다.
- 4) '속성명 θ 보통상수 OR... 속성명 θ 보통상수' 조건절을 만든다.

○변환타입-III

변환타입-III란 프레디킷트가 '속성명 θ 보통상수' 형태이며, 보통상수는 연속 또는 불연속 도메인이고 θ 는 \approx 또는 $! \approx$ 퍼지 비교 연산자이다. 여기서도 질의 변환시에 FLEVEL을 허용하지 않으므로 α 수준 집합은 고려할 필요가 없다. 변환타입-III 변환과정은 행렬 형태로 저장되어 있는 근접관계에 접근하여 속성명 θ 보통 상수 조건에 만족하는 데이터 값을 읽어온다. 변환타입-II 변환과정과 마찬가지로 해당 속성명으로 OR하여 프레디킷트로 만든다. 이 과정을 요약하면 다음과 같다.

○Procedure 변환타입-III 변환

• 입력 : '속성명 θ 보통상수' 프레디킷트(θ 는 퍼지 비교 연산자)

• 출력 : '속성명 θ 보통상수 OR... 속성명 θ 보통상수' 조건절

- 1) 입력 프레디킷트로 의미 데이터베이스에 접근한다.
- 2) 근접관계에서 입력 퍼지 비교 연산자의 근접 허용치를 만족하는 데이터를 가져온다. 이때 근접 허용치는 미리 시스템에서 정해졌다고 가정하며 사용자가 줄 수가 없다.
- 3) 선정된 데이터 값을 '속성명 θ 보통상수' 형태로 OR 연산을 한다.
- 4) '속성명 θ 보통상수 OR... 속성명 θ 보통상수' 조건절을 만든다.

3.3.2 SQL 생성 모듈

SQL 생성 모듈은 퍼지상수와 퍼지 비교 연산자에

포함되지 않은 파스트리에서 표준 SQL 문자열로 변환한다. 이 과정은 기존에 존재하는 표준 SQL 처리기의 입력형태로 맞추기 위함이다. 여기서 표준 SQL이라 함은 ISO SQL2를 의미한다.

3.4 결과합성기

결과합성기는 관계형 데이터베이스에서 1차원 데이터만으로 변환된 질의를 처리한 결과를 입력으로 받고, 처음 주어진 파스트리를 이용하여 2차원 데이터 즉 퍼지 질의에 대한 결과를 만든다. 입력으로 받아들이는 결과 튜플의 각 속성을 파스트리를 이용하여 소속척도가 필요한 지 또는 필요없는지를 조사한다. 만약 필요하다면 의미 데이터베이스 관리 및 생성기를 호출하여 소속척도 값을 할당한다. AND와 OR의 논리 연산자에 대한 소속척도를 MIN 연산자를 이용하여 계산한다. 입력으로 받아들이는 것이 튜플이 아니고 오류코드와 같은 실행결과 코드인 경우에는 그대로 사용자에게 내보낸다.

○Procedure 결과 합성

(파스트리와 시스템 카탈로그는 전역 변수로 주어진다고 가정함)

- 입력 : 결과 튜플 또는 실행 결과 코드
- 출력 : 소속척도가 포함된 결과 튜플 또는 실행 결과 코드

1) 입력이 실행 결과 코드이면 그대로 출력하고 끝낸다.

2) 입력이 결과 튜플이면 파스트리를 이용하여 의미 데이터베이스에 접근하여 해당 속성 데이터의 소속척도를 읽어온다.

3) AND, OR 논리 연결자에 따라 MIN 연산자를 적용하여 튜플의 소속척도를 구한다.

3.5 추론 엔진 및 의미 데이터베이스

추론 엔진 및 의미 데이터베이스 모듈은 질의변환 모듈과 결과 합성기에 퍼지 상수와 퍼지 비교연산에서 보통상수와 보통비교 연산자로 변환하거나 그 역으로 환원할 때 필요한 데이터와 소속척도 정보를 제공한다. 여기서 퍼지상수나 근접관계와 같은 의미 데이터베이스는 처리의 효율성을 위해 주기억 장치내에 저장하여, 특히 근접관계는 1열배열방식이나 3열배열 방식의 자료구조로 저장한다.

4. 시뮬레이션 및 결과

퍼지 질의 처리 과정은 크게 두 단계로 구분된다. 첫째는 퍼지 질의를 기존의 표준 관계형 질의로 변환하

는 것이고, 둘째는 관계형 데이터베이스 관리 시스템에 의해 산출된 결과를 퍼지 개념을 가지도록 환원하는 것이다. 퍼지 질의 처리과정을 예를 들어 살펴본다.

[예 1] 그림 2와 같은 보통 관계와 의미 관계가 있다. 이것을 이용하여 '젊고 연봉이 많으며 데이터베이스와 유사한 전문분야를 가진 사람을 찾아라'라는 다음과 같은 퍼지 질의가 처리되는 과정을 살펴보자.

○입력 퍼지 질의 형태

```
SELECT 이름, 나이, 연봉, 전문분야
FROM 개인신상
WHERE 나이 = $Young AND 연봉 = $High_salary
AND 전문분야 ≈ "데이터베이스"
WITH FLEVEL(나이) = 0.4 AND FLEVEL(연봉) = 0.5;
```

○퍼지 질의 변환 단계

1) 나이 = \$Young을 변환타입-I 변환 방법을 이용하여 $0 \leq \text{나이} \leq 31$ 로 바꾼다. 이때 0세에서 31세까지의 소속척도는 그림 2(c)의 의미 데이터베이스에 따른다.

2) 연봉 = \$High_salary를 변환타입-II 변환방법을 이용하여 (연봉 = 3,000 OR 연봉 = 3,100 OR 연봉 = 3,200 OR 연봉 = 3,300 OR 연봉 = 3,400 OR 연봉 = 3,500)으로 바꾼다. 여기서도 각 연봉의 소속척도는 그림 2(b)의 의미 데이터베이스에 따른다.

3) 전문분야 ≈ "데이터베이스"는 타입-III의 근접 관계를 이용하여 ≈ 연산자가 정한 임계값을 만족하는 가장 근접한 전문분야로 대체한다. 즉, (전문분야 = "DB" OR 전문분야 = "KR" OR 전문분야 = "KB")로 변환한다. (그림 2(d) 참조)

이런 과정을 거쳐 변환된 입력 질의는 다음과 같다.

○변환된 입력 퍼지 질의 (표준 SQL 형태)

```
SELECT 이름, 나이, 연봉, 전문분야
FROM 개인신상
WHERE (0 ≤ 나이 ≤ 31) AND (연봉 = 3,000 OR 연봉 = 3,100 OR 연봉 = 3,200 OR 연봉 = 3,300 OR 연봉 = 3,400 OR 연봉 = 3,500) AND (전문분야 = "DB" OR 전문분야 = "IR" OR 전문분야 = "KB");
```

위의 변환된 표준 SQL 형태는 1차원적인 데이터만 포함된 질의로 그림 2의 (b), (c), (d)의 의미 데이터베이스가 합쳐져야 원래의 의미가 된다.

○관계형 DBMS로 처리된 결과

변환된 질의를 관계형 데이터베이스 관리시스템으로 처리하면 다음과 같은 튜플을 얻는다.

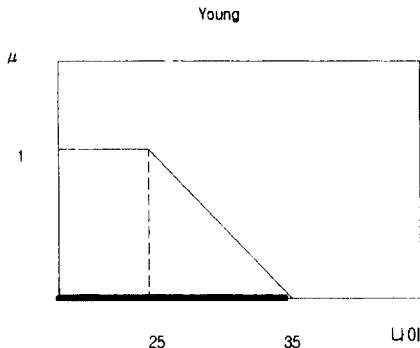
```
<Frank 31 3,200 IR>
<Anna 25 3,500 KB>
<John 20 3,000 IR>
<Steve 29 3,000 KB>
```

이름	나이	연봉(만원)	전문분야
Frank	31	3,200	IR
Susan	26	1,900	DB
Tom	40	2,700	OS
Anna	25	3,500	KB
Margaret	38	2,600	IR
John	20	3,000	IR
Steve	29	3,000	KB
Smith	19	2,500	OS
Jane	31	3,500	DB
Maria	48	5,000	DB

(a) '개인신상' 관계

연봉 (만원)	μ
2,500 이하	0
2,600	0.1
2,700	0.2
2,800	0.3
2,900	0.4
3,000	0.5
3,100	0.6
3,200	0.7
3,300	0.8
3,400	0.9
3,500	1.0

(b) 퍼지상수 'High_salary'



(c) 퍼지상수 'Young'

	DB	AI	IR	OS	KB
DB	1.0	0.7	0.8	0.0	0.9
AI	0.7	1.0	0.8	0.0	0.3
IR	0.8	0.8	1.0	0.0	0.6
OS	0.0	0.0	0.0	1.0	0.0
KB	0.9	0.3	0.6	0.0	1.0

(d) '전문분야' 도메인의 근접관계

그림 2. 보통관계와 의미 데이터베이스의 저장 예

<Jane 31 3,500 DB>

이것은 1차원 데이터만 가지므로 어느 것이 주어진 질의에 가장 적합한지 알 수 없다. 위의 5개의 튜플은 소속척도를 가지는 2차원 데이터로 환원되어야 비로소 퍼지개념을 가진 질의에 대한 결과라고 할 수 있다. 환원과정은 변환 방법의 역으로 생각할 수 있다. 구해진 데이터를 의미 데이터베이스와 매핑하여 소속척도를 구하면된다.

○소속척도를 합성한 결과

여기서는 MIN 연산자를 사용하였는데, Anna(0.8)가 첫번째로 질의를 만족하는 사람이고, John(0.5)과

표 2. 질의 처리 결과

이름	나이	연봉(만원)	전문분야	min
Frank	31(0.4)	3,200(0.7)	IR(0.9)	0.4
Anna	25(1.0)	3,500(1.0)	KB(0.8)	0.8
John	20(1.0)	3,000(0.5)	IR(0.9)	0.5
Steve	29(0.6)	3,000(0.5)	KB(0.8)	0.5
Jane	31(0.4)	3,600(1.0)	DB(1.0)	0.4

Steve(0.5)가 공동으로 두번째로 만족한다. 세번째로 만족하는 사람은 Frank(0.4)와 John(0.4)이다.

5. 결 론

본 연구에서는 관계형 데이터베이스 시스템을 위한 지능적인 질의처리 시스템을 제안하였다. 애매한 질의와 관련된 기존의 연구결과를 분석하여 그 문제점들을 해결할 수 있는 새로운 모델을 제시하였고, 각 모듈의 역할들을 분석하였으며, 실제 예를 통하여 이 모델의 작동을 입증하였다.

본 논문에서 제안된 모델은 다음과 같은 주요 특징을 가진다.

첫째, 애매한 질의를 분석하여 분류할 수 있는 질의 변환 모듈을 새롭게 적용 설계함으로써 기존의 상용 RDBMS를 그대로 활용할 수 있다. 둘째, 추론 엔진 및 의미 데이터베이스를 적용하여 질의변환에 적용함으로써 주어진 질의에 보다 정확한 결과를 도출할 수 있다. 셋째, 애매한 질의를 순수 데이터와 소속척

도의 2차원적인 데이터로 해석하여 질의를 변환하였고, RDBMS에서 나온 결과에 의미 데이터인 소속척도를 합성하여 사용자의 질의에 대응하는 순서화된 결과를 산출할 수 있다.

앞으로는 의미 데이터베이스를 좀 더 체계적으로 구성할 수 있는 방법에 관한 연구가 진행되어야 될 것이며, 모든 질의에 있어서 보다 지능적으로 처리해 줄 수 있는 사용자 인터페이스에 관한 연구가 후후의 연구과제로 진행중이다.

참고문헌

[1] P. Bosc and O. Pivert, "Some Approaches for Relational Databases Flexible Querying," *Journal of Intelligent Information Systems*, 1, pp. 323-354, 1992.

[2] P. Bosc and O. Pivert, "SQLf: A Relational Database Language for Fuzzy Querying," *IEEE Transactions on Fuzzy Systems*, 3, pp. 1-17, 1995.

[3] B. Buckles and F. Petry, "Fuzzy Databases in the New Era," Proceedings of FUZZ-IEEE/IFES'95 Workshop on Fuzzy Database Systems and Information Retrieval, pp. 85-91, 1995.

[4] B. Buckles and F. Petry, "Constructing a Windowing Function to Interface a Database with Expert System Rules," *Journal of Systems Eng. and Electronics*, 5, pp. 45-54, Special Issue on Information Systems Eng., 1996.

[5] C. L. Chang, Decision Support in an Imperfect World. Research Report RJ3421, IBM, San Jose, CA. U. S. A, 1982.

[6] M. Codd and F. Petry, "Fuzzy Querying Binary Relationships in Spatial Databases," *Proceedings of 1995 IEEE International Conference on Cybernetics and Society*, pp. 378-385, 1995.

[7] E. Cox, "Relational Database Querying Using Fuzzy Logic," *AI Expert*, 10, pp. 23-28, 1995.

[8] A. D'Atri and L. Tarantino, "From Browsing to Query," *Data Eng. Bull.*, 12(2), pp. 47-53, 1989.

[9] N. Fuhr, "A Probabilistic Framework for Vague Queries and Imprecise Information in Databases," Proceedings of 6th Int. Conference on Very large Databases, pp. 77-85, 1990.

[10] T. Ichikawa, M. Hirakawa, ARES: A Relational Database with the Capability of performing flexible interpretation of queries, *IEEE Trans. on Soft. Eng.*, 12 (5), pp. 623-634, 1986.

[11] A. Motro, "VAGUE: A User Interface to Relational Databases that Permit Vague Queries," *ACM Trans. on Office Information Systems*, 6, pp. 187-214, 1988.

[12] A. Motro, "A Trio of Database User Interfaces for Handling Vague Retrieval Requests," *Data Eng. Bull.*, 12(2), pp. 54-63, 1989.

[13] Frederick E. Petry, "The Promise of Intelligent Information Systems," *Advances in Databases and Artificial Intelligence: Intelligent Database Technology Vol. 1*, 9-11, JAI Press, Greenwich, CT, 1995.

[14] Frederick E. Petry, *Fuzzy databases-Principles and Applications*, Kluwer Academic Publishers, 1996.

[15] L. A. Zadeh, J. Kacprzyk, *Fuzzy Logic for the Management of Uncertainty*, John Wiley & Sons, Inc., 1992.



김대수 (Dae Su Kim) 종신회원

1977년: 서울대학교 사대 수학과 졸업
 1986년: 미국 Univ. of Mississippi 대학원 전산학과 졸업(M. S)
 1990년: 미국 Univ. of South Carolina, Computer Science(Ph. D)
 1990년: 미국 Intelligent System Lab, Researcher

1991년: 한국전자통신연구소 컴퓨터연구단 선임연구원
 1993년: 한신대학교 전자계산학과 부교수
 주관심분야: 신경망, 퍼지이론, 지능시스템, 인공지능, 퍼지 DB, 패턴인식 등



김창석 (Chang Suk Kim)

1983년: 경북대학교 전자공학과 학사
 1990년: 경북대학교 전자공학과 석사
 1994년: 경북대학교 컴퓨터공학과 박사
 1983년~1994년: 한국전자통신연구소 데이터베이스연구실 선임연구원

1994년~현재: 부산외국어대학교 컴퓨터공학과 조교수, 부산외국어 대학교 전자계산소 소장
 1996년~현재: 한국 퍼지 및 지능시스템 학회 논문지 편집위원
 1996년~현재: 바다(DBMS) 원천코드 사용자 그룹 운영위원
 1994년: 한국전자통신연구소 데이터베이스연구실 선임연구원
 주관심분야: 지능형 데이터베이스 시스템, 웹 데이터베이스, 전역 정보 시스템 (Global Information System), 퍼지 정보 처리 시스템 등