

데이터베이스의 퍼지 질의를 위한 시각적 인터페이스에 관한 연구

IMPLEMENTATION OF VISUAL INTERFACE FOR FUZZY QUERY ON DATABASE

피수영 * 정홍 † 정환목 ‡

요약

For the retrieval of database through the queries expressed by the linguistic variables which are vague, this paper implemented fuzzy query system that can process fuzzy queries. First of all, the system extracts linguistic variables, hedges, and operators from the query expressions, generates fuzzy sets by allocating a domain of variables and membership functions, and modifies the sets by applying the hedges. Secondly, it reads tuples from the database, calculates the membership degrees of them using the fuzzy sets, performs fuzzy operations among attributes having membership degrees, and obtains the final degree, called suitability factor. And finally, the system sorts the tuples by the suitability factor and outputs the results. This paper carried out an experiment with a real database, and evaluated it very practical compared to other systems.

1. 서론

퍼지 이론을 기본으로 한 퍼지 데이터베이스의 연구는 기존의 데이터베이스에서 모호한 질의를 처리하기 위한 방법과 모호한 자료를 저장 및 처리하기 위해 데이터 모델을 확장하는 방법으로 나눌 수 있다[8]. 전자의 연구로는 퍼지집합의 소속도(membership degree)를 이용한 Tahani[9]의 연구와 속성의 정의역에 대한 거리 계산을 사용하는 Motro의 VAGUE 모델[7] 등이 있다. 후자의 방법으로는 유사관계(similarity relation)를 확장한 Zemankova의 FRDB 모델[11]과 Buckles와 Petry[5]의 연구가 있고, 애매성을 가능성 분포로 표현하여 관계 모델을 확장한 Umamo의 FREEDOM-0 모델[10] 등이 있다.

모호한 질의 처리를 위한 최근 연구들의 공통적인 처리 방식은 유사관계나 언어변수(linguistic variable)에 해당하는 소속함수(membership function)의 값을 사용자가 시스템 내부에 지식베이스 형태로 미리 저장해 두고, 데이터베이스의 튜플(tuple)을 읽어 들여 지

* 영진전문대학 전자계산과

† 계명대학교 전자계산학과

‡ 대구효성가톨릭대학교 전자정보공학부

식베이스의 유사관계나 소속함수의 값과 비교하여 소속도를 구한다[2,3,4,9]. 이러한 연구들의 문제점은 유사관계와 소속도를 미리 준비하여야 하는 불편함이 있고, 변경에 대한 융통성이 결여되어 있다. 이러한 문제를 해결하기 위해 구현한 본 시스템은 연속형 퍼지 집합에 대한 소속함수를 적용하여 별도로 유사관계 베이스(similarity relation base)나 소속도 베이스(membership degree base)를 참조하지 않고도 소속도를 구할 수 있도록 하고, 시각적인 인터페이스(visual interface)를 제공하여 사용자가 쉽고 융통성 있는 질의를 처리할 수 있도록 한다.

연구 내용은 구현하고자 하는 시스템의 처리 과정을 상세히 분석하여 질의 결과에 대해 신뢰성이 높고, 융통성 있는 정보를 검색할 수 있는 시스템을 구현하고, 그 실용성을 실험 및 평가한다.

2. 퍼지질의 시스템의 구현

2.1. 질의 분석

퍼지 질의는 관계형 데이터베이스 언어인 SQL 형식과 같은 질의문을 사용한다. 그 이유는 SQL이 현재 가장 많이 사용되고 있고, 배우기 쉬우면서 사용하기 쉽기 때문이다.

```
SELECT Fields(필드명) FROM Table(테이블명)
WHERE Predicate(술어조건)
```

- . SELECT 다음의 Fields(필드명)는 결과로 출력될 필드들의 이름인데, 각각 콤마(,)로 구분하여 표현한다.
- . FROM 다음의 Table(테이블명)은 검색될 테이블 이름이다.
- . WHERE 다음 Predicate(술어조건)은 Field_{name1} = "Condition1 Operand1 Field_{name2} = "Condition2" ... 로 구성된다. Field_{name} 은 검색하고자 하는 필드의 이름이고, Condition은 언어변수인데 일반적으로 헤지와 언어변수로 구성된다. Operand는 AND와 OR, NOT 연산을 사용한다.

예를 들어 "나이가 약간 어리고, 키가 매우 큰 사람의 이름과 나이, 키를 검색하시오"라는 질의가 있다면, 아래와 같은 SQL 형식의 질의문으로 입력한다.

```
SELECT 이름,나이,키 FROM 사람
WHERE 나이="약간 어린" AND 키="매우 큰"
```

SQL 질의문을 토큰 단위로 분리한 후, 분리된 토큰을 분석하여 언어변수와 헤지 및 연산자를 추출한다.

2.2. 퍼지집합의 생성

퍼지집합은 "매우 큰 키", "약간 중년" 등과 같이 경계가 확실치 않은 애매모호한 성질을 표현한 집합이다. 퍼지집합의 형태가 결정되면 퍼지집합을 생성해야 하는데, 이를 위하

여 퍼지집합에 대한 정의역을 결정해야 한다. 퍼지집합의 형태와 정의역이 설정되면 퍼지 집합이 생성된다. 다퍼지집합에는 연속형 퍼지집합과 이산형 퍼지집합이 있다[6]

- 연속형 함수

- 선형함수: 퍼지집합의 모양이 선형(linear)으로서, 소속함수가 0이 되는 값을 a, 1이 되는 값을 b라 할 때, 다음과 같이 정의된다.

$$L(x; a, b) = \begin{cases} 0 & : x \leq a \\ (x - a)/(b - a) & : a < x < b \\ 1 & : x \geq b \end{cases}$$

- S형(sigmoid)함수: 퍼지집합의 모양이 비선형(nonlinear)으로서, 소속함수가 0이 되는 값을 a, 1이 되는 값을 c, 0.5가 되는 변곡점을 b라 할 때, 다음과 같이 정의된다.

$$S(x; a, b, c) = \begin{cases} 0 & : x \leq a \\ 2((x - a)/(c - a))^2 & : a < x \leq b \\ 1 - 2((x - c)/(c - a))^2 & : b < x < c \\ 1 & : x \geq c \end{cases}$$

- 범종(bell)형함수: 퍼지집합의 모양이 범종 모양으로서, PI형, Beta형, Gauss형 등이 있는데, 가장 많이 활용되는 것이 PI형이다. PI형은 증가 S형함수와 감소 S형함수를 합친 형태로 다음과 같이 정의된다.

$$\Pi(x; b, c) = \begin{cases} S(x; c - b, c - b/2, c) & : x \leq c \\ 1 - S(x; c, c + b/2, c + b) & : x > c \end{cases}$$

c: 정의역의 중앙값

b: 중앙값으로부터 소속도가 0이되는 곳까지의 폭

- 삼각형함수: 퍼지집합의 모양이 삼각형으로서, b를 중앙값이라 할 때, 다음과 같이 증가 선형함수와 감소 선형함수의 결합 형태이다.

$$T(x; a, b, c) = \begin{cases} L(x; a, b) & : x \leq b \\ 1 - L(x; b, c) & : x > b \end{cases}$$

이와 같이 소속함수의 표현에는 여러가지가 있고 응용에 따라 새로운 것을 만들어 사용할 수도 있다.

- 이산형 함수

- 소속함수를 연속형 함수로 표시할수 없는 퍼지집합은 다음과 같이 요소와 소속도의 집합으로 표현한다.

$$A = \{x, \mu_A(x)\}$$

예를 들어, “학력”이라는 이산형 퍼지집합은 (대학원졸, 1), (대학졸, 0.8), (전문대졸, 0.7), (고졸, 0.5), (중졸, 0.3), (초등졸, 0.1)과 같이 표현된다.

2.3. 헤지의 적용

언어변수의 결정시 추출된 헤지를 적용하여 퍼지집합을 변형한다. 헤지는 언어변수를 수식하는 형용사나 부사로서, 퍼지집합의 성질을 변화시킨다. 즉, 헤지를 퍼지집합에 적용하면 퍼지집합의 모양이 바뀌어 새로운 퍼지집합을 생성한다[1]. 헤지에는 부정 헤지, 제한 헤지, 강화 헤지, 약화 헤지, 대비 강화 헤지, 대비 약화 헤지 등이 있다[6].

- 부정(negation) 헤지: not과 같은 것으로, 퍼지집합을 반대 모양으로 바꾼다.

$$\mu_{\text{not}A}(x) = 1 - \mu_A(x)$$

- 제한(restriction) 헤지: below, above 등과 같은 것으로서, 퍼지집합의 영역을 변경시킨다.
- 강화(intensification) 헤지: very, extremely 등과 같은 것으로, 퍼지집합의 모양을 날카롭게 강화시킨다.

$$\mu_{\text{very}A}(x) = \mu_A^2(x)$$

$$\mu_{\text{extremely}A}(x) = \mu_A^3(x)$$

- 약화(dilation) 헤지: somewhat, slightly 등과 같은 것으로, 퍼지집합의 모양을 둔하게 약화시킨다.

$$\mu_{\text{somewhat}A}(x) = \mu_A^{0.5}(x)$$

$$\mu_{\text{slightly}A}(x) = \mu_A^{0.3}(x)$$

- 대비 강화(contrast intensification) 헤지: positively와 같은 것으로, 퍼지집합의 퍼지정도를 약하게, 즉 크리스프(crisp)하게 한다.

$$\begin{aligned} \mu_{\text{positively}A}(x) &= \min(1, 2(\mu_A^2(x))) && : \mu_A(x) \geq 0.5 \\ &= \max(0, 1 - 2(1 - \mu_A(x))^2) && : \text{else} \end{aligned}$$

- 대비 확산(contrast diffusion) 헤지: generally와 같은 것으로, 퍼지집합의 퍼지정도를 강하게, 즉 더 모호하게 한다.

$$\begin{aligned} \mu_{\text{generally}A}(x) &= 0.7(\mu_A^{0.5}(x)) && : \mu_A(x) \geq 0.5 \\ &= 1 - 0.7(1 - \mu_A(x))^{0.5} && : \text{else} \end{aligned}$$

2.4. 퍼지 연산

퍼지 연산에는 교집합, 합집합, 여집합 등이 있는데, 언어변수 결정시 추출된 연산자에 의해 결정되고 소속도 수준에서 계산되어진다. “AND” 연산자는 소속도 중에서 Min값을 취하고, “OR” 연산자는 Max값을 취한다. 그리고 “NOT” 연산자는 소속도의 보수값을 취한다[3,10]. 연산시 우선 순위를 고려하여 후위 표기식(postfix)로 변환하여 처리한다.

- . 퍼지 교집합: $\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))$
- . 퍼지 합집합: $\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))$
- . 퍼지 여집합: $\mu_{A^c}(x) = 1 - \mu_A(x)$

일반적으로, 퍼지 여집합보다 부정 해지를 많이 사용한다.

3. 실험 및 평가

3.1. 퍼지질의 시스템의 실험

실험에 사용할 관계 데이터베이스 “사람”은 [표-1]과 같이 이름, 나이, 키, 학력이라는 속성을 가지고 있다.

이름	나이	키	학력
김 유 신	43	174	대학졸
강 감 찬	46	176	대학졸
양 만 춘	50	175	고졸
울지문덕	48	176	대학졸
이 순 신	40	178	전문대졸

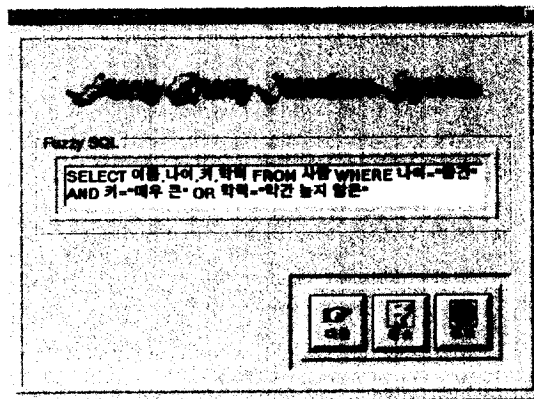
[표-1] “사람” 데이터베이스

“중간 나이이고, 매우 큰키 또는 학력이 약간 높지 않은 사람의 이름과 나이, 키, 학력을 검색하시오” 라는 질의를 처리하고자 한다. 이 질의를 다음과 같이 SQL 질의로 입력한다.

SELECT 이름,나이,키,학력 FROM 사람

WHERE 나이=“중간” AND 키=“매우 큰” OR 학력=“약간 높지 않은”

[그림-1]은 SQL 질의를 실제 입력한 화면이다.

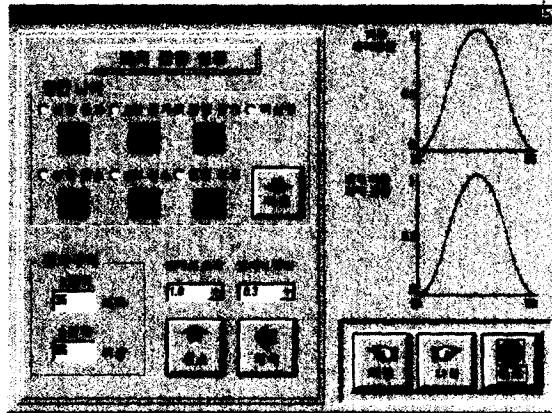


[그림-1] 퍼지 SQL 질의 화면

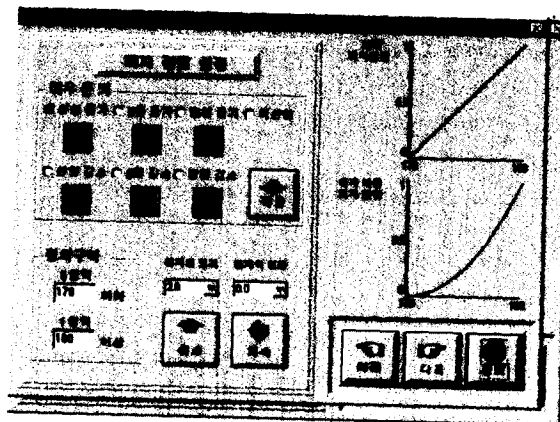
퍼지집합 생성 화면에서 퍼지집합의 형과 정의역, 해지의 크기 및 임계치를 [표-2]와 같은 자료로 입력한다. 이때, 퍼지집합의 형과 정의역은 언어변수에 의해 결정되나, 언어변수가 테이블에 존재하지 않을 경우 사용자가 직접 입력하여야 한다. 해지의 크기도 자동적으로 결정되나, 사용자의 주관에 따라 변경할수도 있다.

퍼지집합	중간나이	매우 큰키	약간 높지않은 학력
형	벨형	증가선형	이산형
정의역	35 - 55	170 - 180	-
임계치	0.3	0	0.4

[표-2] 퍼지집합의 정의



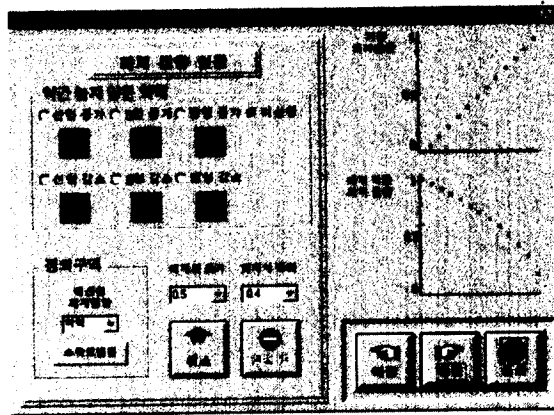
[그림-2] “중간 나이”에 대한 퍼지집합 생성



[그림-3] “매우 큰 키”에 대한 퍼지집합 생성

[그림-2]는 “중간 나이”에 대한 퍼지집합 생성 화면인데, 오른쪽에 기본 퍼지집합 그래프와 헤지가 적용된 퍼지집합 그래프를 보여준다. “중간 나이”라는 퍼지집합에서는 헤지가 적용되지 않았으므로 두 그래프의 모양이 같다.

[그림-3]은 “매우 큰 키”에 대한 퍼지집합 생성 화면인데, very 강화 헤지에 의해 자동적으로 헤지의 크기가 2.0이 된다. 헤지의 크기는 사용자가 변경할 수도 있다. [그림-4]는 “약간 높지 않은 학력”에 대한 이산형 퍼지집합 생성 화면인데, somewhat 약화 헤지에 의해 자동적으로 헤지의 크기가 0.5가 된다. <소속도 변경> 버튼을 클릭하여 소속도를 변경할 수도 있다.



[그림-4] “약간 높지 않은 학력”에 대한 퍼지집합 생성

“중간 나이”와 “큰 키”, “높은 학력”이라는 언어변수에 대한 소속도는 [표-3]과 같은데, “중간 나이”는 헤지가 적용되지 않아 소속도가 변형되지 않고, “매우 큰 키”와 “약간 높지 않은 학력”과 같이 헤지를 적용된 것은 [표-4]와 같이 소속도가 변형된다.

이름	중간 나이	큰 키	높은 학력
김 유 신	0.92	0.40	0.80
강 감 찬	0.98	0.60	0.80
양 만 춘	0.50	0.50	0.60
울 지 문 덕	0.82	0.60	0.75
이 순 신	0.50	0.80	0.65

[표-3] 소속도

이름	중간 나이	매우 큰키	약간 높지않은 학력
김유신	0.9200	0.1600	0.4472
강감찬	0.9800	0.3600	0.4472
양만춘	0.5000	0.2500	0.6325
울지문덕	0.8200	0.3600	0.5000
이순신	0.50000	0.640	0.5916

[표-4] 헤지가 적용된 소속도

헤지가 적용된 소속도를 이용한 퍼지 연산 결과는 다음과 같다.

$$\begin{aligned}
 \text{김유신} &= (0.9200 \text{ AND } 0.1600) \text{ OR } 0.4472 \\
 &= \text{Max}(\text{Min}(0.9200, 0.1600), 0.4472) \\
 &= 0.4472
 \end{aligned}$$

위의 계산 방식에 의해 “김유신”과 “강감찬” 튜플은 0.4472의 적합도를 가지며, “양만춘” 튜플은 0.6325의 적합도를 가진다. “울지문덕” 튜플은 0.5000이고, “이순신” 튜플은 0.5916의 적합도를 가진다. SQL질의에 대한 퍼지 연산 결과를 최대 검색 수를 5로 제한하여 적합도의 크기 순서에 따라 화면에 출력한 결과는 [그림-5]와 같다.

결과				튜플				
이름	나이	키	학력	이름	나이	키	학력	적합도
김유신	0.92	0.16	0.4472	1. 김유신	0.92	0.16	0.4472	0.4472
강감찬	0.98	0.36	0.4472	2. 강감찬	0.98	0.36	0.4472	0.4472
양만춘	0.50	0.25	0.6325	3. 양만춘	0.50	0.25	0.6325	0.6325
울지문덕	0.82	0.36	0.5000	4. 울지문덕	0.82	0.36	0.5000	0.5000
이순신	0.50	0.64	0.5916	5. 이순신	0.50	0.64	0.5916	0.5916

[그림-5] 퍼지 질의에 대한 출력 화면

3.2. 퍼지질의 시스템의 평가

본 연구에서 구현한 퍼지 질의시스템을 질의 처리과정과 실험 결과에 의해 기존의 불리언 (Boolean) 질의시스템, 그리고 기존의 퍼지 데이터베이스 시스템과 비교한 장점을 기술하고, 본 시스템의 문제점과 그에 대한 해결 방안을 분석한다.

- 본 시스템과 불리언 질의시스템을 비교할 때 장점은 다음과 같다.

- . 모호한 질의에 대한 검색이 가능하다.
- . 질의문에 언어변수를 사용하여 자연스러운 표현이 가능하다.
- . 해지를 적용하여 섬세하고 융통성 있는 검색이 가능하다.
- . 원하는 결과가 검색되지 않는 경우, 정의역을 원하는 결과가 나올 때까지 반복 변경 입력하여 원하는 결과를 얻을 수 있다.
- . 특이한 경우를 제외하고 검색의 실패가 없다.
- . 검색 순위 및 적합도까지 제공한다.
- . 본 시스템이 다른 퍼지 질의시스템에 비해 나아진 점은 아래와 같다.
- . 소속 테이블이나 유사 테이블을 사용하지 않으므로 자원을 줄일 수 있다.
- . 다양한 해지의 적용으로 인간의 언어에 가까운 질의가 가능하다.
- . 강화 해지와 약화 해지의 크기에 융통성을 줄 수 있다.
- . 소속 함수를 적용하여 소속도를 구하므로 신뢰성이 높다.
- . 임계치에 의해 일정 수준 이하의 자료를 제외시킬 수 있다.
- . 퍼지집합을 생성할 때 사용자의 주관에 따른 검색이 가능하다.
- . 실행 과정을 시각적으로 보여 줌으로써 사용자의 이해도를 높인다.
- 본 시스템의 문제점과 그에 대한 해결 방안
- . 퍼지에 대한 기본 지식이 없는 사용자라면 퍼지집합 설정(소속함수의 선택, 정의역 범위 결정, 해지의 사용)에 따른 혼란과 어려움이 있을 수 있다. 이는 사용자에게 대한 기본적인 퍼지 이론과 시스템 운영에 대한 사전 교육으로 해결할 수 있다. 그리고, 시각적인 표현에 의해 쉽게 익숙해 진다.
- . 퍼지집합의 생성에 사용자의 주관적인 개입이 많았으므로 사용자에게 따라 질의한 결과가 달라 질 수 있다. 이것은 객관적인 표준 퍼지집합을 많이 만들어 두고 사용자가 선택하여 사용하게 함으로써 객관화할 수 있다.
- . 모든 튜플들의 해당 속성들에 대한 소속도를 계산해야 하므로 연산 시간이 오래 걸려 검색 속도가 느려진다. 그러나 질의 시스템의 전반적인 운용면에서 보면 연산 속도의 비중이 그리 크지 않으므로 크게 문제가 되지 않을 것이다.

4. 결론

본 연구는 기존의 퍼지 데이터베이스 시스템을 고찰하여 문제점을 분석하고, 미비한 점을 보완하여 사용자 인터페이스를 중심으로 시각적 퍼지 질의 시스템을 구현하였다.

본 논문에서 구현한 시스템은 질의문에 언어변수와 다양한 헤지를 적용하여 인간의 언어와 유사한 자연스러운 언어 표현과 섬세하고 융통성 있는 질의를 할 수 있다. 퍼지집합에 소속함수를 사용하여 소속도 구하고, 임계치를 적용하여 일정 수준 이하의 자료를 제외시켜 신뢰성을 향상시켰다. 또한, 실행 과정에서 시각적 인터페이스를 제공하여 사용자의 이해도를 높이고, 퍼지집합을 설정할 때 사용자의 주관에 따른 검색이 가능하도록 하였다. 특이한 경우를 제외하고는 검색의 실패가 없으며, 원하는 결과가 나올 때까지 정의역을 반복 입력할 수 있고, 결과를 적합도 순으로 정렬하여 사용자의 의사결정에 도움을 주며, 원하는 검색 대상 수 만큼 출력한다.

이와 같은 검색 시스템을 쉽게 사용하기 위해서는 SQL형식의 질의가 아닌, 자연어로 기술된 질의로부터 언어변수, 헤지, 연산자 등을 추출할 수 있는 자연어 질의 처리 모듈을 개발하는 것이 바람직하다. 또한, 초보자의 경우 소속함수의 선택, 정의역 범위 결정, 헤지의 사용에 따른 어려움이 있을 수 있으므로, 퍼지집합 별로 정의역과 소속함수를 많이 준비해 두는 것이 좋다. 이것을 위해서는 각종 언어변수에 대한 객관적인 표준 자료의 수집이 선행되어야 한다. 퍼지 시스템이란 인간적인 요소가 많이 포함되는 시스템이므로 이를 운용하는데 있어서도 사용자의 개입이 필요할 때가 많으며 결과도 근사적인 해가 도출되므로 근사해를 사용자가 면밀히 검토해 보고 의사 결정에 반영해야 할 것이다.

본 연구에서는 언어변수들에 대해 각기 퍼지성을 부여하여 모호한 질의를 처리하도록 하였는데, 조금 더 폭 넓은 처리를 위해서는 언어변수들 간의 관련성도 고려하면 좋을 것이다. 즉, 언어변수들 간에 유사성까지 고려한 검색이 가능하도록 하는 시스템으로 확장하는 연구가 필요하다. 그리고, 모호한 자료를 포함한 확장된 퍼지 데이터베이스에 대한 연구가 필요하다.

참고문헌

1. 류성렬, C언어에 의한 실용 퍼지, 세화, 1995
2. 박양규, "지식 기반형 퍼지 질의 시스템에 관한 연구", 석사학위논문, 1991
3. 이현주, "지식 기반형 퍼지 질의 응답 시스템", 석사학위논문, 1989
4. 조경달, "모호성의 다중 수준 정의에 기반을 둔 퍼지 질의 시스템의 설계 및 구현", 석사학위논문, 1992
5. Buckles, B. and Petry, F. "A Fuzzy Representation of Data for Relational Database", Fuzzy Sets for intelligent Systems, D.Dubois ed., 1993
6. Earl Cox, *Fuzzy Systems Handbook*, Prentice-Hall, 1994

7. Motro, A. "VAGUE : A User Interface to relational Databases that permit value queries", *ACM trans. on Office Information Systems*, Vol.6, No.3, pp187-214, 1988
8. Petry and Frederick, *Fuzzy databases*, Kluwer Academic, 1996
9. Tahani, V., "A Conceptual Framework for Fuzzy Query Processing - A Step toward Intelligent Database Systems", *Information Processing & Management*, Vol.13, pp. 289-303, 1977
10. Umano, M., "FREEDOM-0 : A Fuzzy Database System", in Gupta and Sanchez eds., *Fuzzy Information and Decision Processes*, pp.339-347, 1982
11. Zemankova-Leech, M., and Kandel, A., "FRDB : Fuzzy Relational Data Bases - A Key to Expert Systems", Verlag, 1984

[1996년 11월 접수, 1997년 3월 개정]