

FCM을 이용한 지식기반 데이터베이스 검색 시스템의 구축

Building of Database Retrieval System Based on Knowledge using FCM

박계각* · 서기열* · 천대일* · 양원재**

Gyei-Kark Park, Ki-Yeol Seo, Dea-Il Cheon, and Won-Jea Yang

* 목포해양대학교 해상운송시스템학부

** 한국해양대학교 해사수송과학부

요약

기존의 데이터베이스 검색시스템은 사용자의 검색 조건에 정확히 일치하는 데이터가 데이터베이스 내에 존재할 경우에만 사용자에게 해당 데이터를 제공할 수 있고, 사용자의 검색조건을 정확히 만족하는 데이터가 없을 경우에는 적절한 데이터를 제공 할 수 없는 문제점이 있다. 이러한 문제를 해결하기 위하여 본 논문에서는 FCM의 클러스터증가 및 재초기화 알고리즘을 제안하였고, FCM을 이용하여 데이터베이스 내의 데이터로부터 구축된 지식기반 데이터베이스(KDB)와 구축된 이미지 데이터베이스와 연동을 통하여 사용자의 요구에 가장 근접한 데이터를 제시해주는 검색시스템을 제안하였다. 본 연구에서 제안된 수법을 우체국의 우편주문안내책자를 이용한 선물고르기 DB 검색 시스템에 적용하여 그 유효성을 확인하였다.

ABSTRACT

Conventional database retrieval system have problems of being able to select data out of database only if the data exactly equal to retrieval conditions offered by users. If there are no data in database which exactly equal to user's retrieval conditionals, the system can not provide adequate data. To solve these problems, cluster increase of FCM and re-initialization of algorithm were suggested in this study. And by interlocking knowledge-based database, built with FCM, to image database, new retrieval system was built to provide the data which are most appropriate for the requirement of users. We applied this new retrieval system to gift selection database system in pamphlet of mail order, and confirmed its effectiveness.

Key Words : FCM, KDB, Clustering, Image DB, Linguistic Label.

1. 서 론

최근, 디지털과 컴퓨터 기술 및 광통신 기술, 인터넷 산업의 발전에 힘입어 고도의 정보화 사회가 형성되고 있다. 정보의식의 고양과 인터넷 및 정보기술산업의 활성화에 따라 폭발적으로 증가하는 정보량을 효율적으로 활용하기 위한 데이터베이스 구축 및 검색 기술과 관리 기술 등 많은 정보 관련 기술의 개발이 중요하게 인식되고 있다. 그 중에서 대량의 데이터로부터 원하는 데이터를 검색하는 검색 기술의 개발이 시급한 실정이다. 기존의 데이터베이스 검색시스템은 사용자의 검색 조건에 정확히 일치하는 데이터가 데이터베이스 내에 존재할 경우에만 사용자에게 해당 데이터를 제공할 수 있고, 사용자의 검색조건을 정확히 만족하는 데이터가 없을 경우에는 적절한 데이터를 제공 할 수 없는 문제점이 있다. 이러한 문제점을 해결하기 위하여 사용자의 검색조건을 완화하는 데이터베이스 검색시스템[1] 및 검색조건을 끄지 멤버쉽 함수로 표현하여 그 소속도가 높은 데이터

터를 검색결과로 제시하는 연구사례가 있으나[2], 이들 연구는 각각 휴리스틱한 지식 및 멤버쉽 함수의 선정에 따라 검색되지 않는 조건이 발생한다는 문제점이 있다.

한편, 퍼지클러스터링(FCM:Fuzzy C-means)을 이용하여 데이터베이스 내의 데이터를 복수의 클러스터로 표현하고, 미리 정의된 언어레이블(Linguistic Label)을 이용함으로써, 데이터의 분포상태를 언어적으로 표현하여 사용자에게 제시하는 검색방법의 연구사례가 있다 [3-4]. 그러나 데이터를 분류하는 FCM 알고리즘 적용시 클러스터의 증가 및 재초기화 알고리즘 및 이미지데이터를 취급하는 기법이 제안되어 있지 않다는 문제점이 있다.

따라서, 본 논문에서는 FCM의 클러스터증가 및 재초기화 알고리즘을 제안하고, FCM을 이용하여 데이터베이스 내의 데이터로부터 구축된 지식기반 데이터베이스(KDB:Knowledge-based Database)와 이미지 데이터베이스(Image Database)와의 연동을 통하여, 사용자의 요구에 가장 근접한 데이터를 제시해주는 검색시스템을 제안하고자 한다[5].

본 연구에서 제안된 수법을 우체국 우편주문책자를 이용한 선물고르기 데이터베이스 검색 시스템에 적용하여 그 유효성을 확인하고자 한다.

2. 지식기반 데이터베이스 구축

데이터베이스 내에 있는 데이터간의 패턴, 유사성 등에 관심 있는 지식을 찾아내는 데이터 마이닝(Data Mining)의 기법으로는 통계분석, 클러스터링, 신경망, 규칙추론 및 의사결정 트리 등이 이용되고 있으나[6], 본 논문에서는 매크로한 지식추출의 유효성이 높은 FCM을 이용하여 지식기반 데이터베이스(KDB)를 구축하고자 한다.

데이터들 간의 관계를 추출하여 지식 데이터베이스를 구축하기 위한 FCM과 클러스터 증가 및 재초기화 알고리즘을 기술하고, FCM에 의해 구한 클러스터에 언어레이블(Linguistic Label)을 할당하여 사용자에게 알기 쉽게 데이터를 표현하는 수법을 기술하고자 한다.

2.1 FCM법

Bezdek이 제안한 FCM법은 어떤 개체 X_k 가 오직 한 클러스터에만 속한다고 보는 HCM(Hard C-Means)법에 폐지이론의 특성을 포함시켜, 복수개의 클러스터에 서로 다른 정도로 속한다고 정의하는 클러스터링 방법이다[7]. n 개의 t 차원의 데이터 벡터 $X_k = x_{kp}$, $p=1, 2, \dots, t$, $k=1, 2, \dots, n$ 를 c 개의 클러스터로 분류할 때, 각 클러스터의 중심벡터 V_i , $i=1, 2, \dots, c$ 와 데이터 X_k 와의 비유사도 d_{ik} 를 식(2.1)과 같이 유크리트 거리로 표현한다.

$$d_{ik} = \| X_k - V_i \| \quad (2.1)$$

이 때, 중심벡터 V_i 는 식(2.2)과 같이 표현한다.

$$V_i = \sum_{k=1}^n (U_{ik})^m X_k / \sum_{k=1}^n (U_{ik})^m \quad (2.2)$$

$$U_{ik}^{(l+1)} = 1 / \sum_{j=1}^c (d_{ik}/d_{jk})^{1/(m-1)} \quad (2.3)$$

여기서, U_{ik} 는 X_k 가 클러스터 i 에 속하는 정도를 나타내고, V_i 는 X_k 의 멤버쉽 정도의 m 차원 가중평균이다. FCM법의 알고리즘은 기본적으로는 통상의 C-Means법의 U 와 V 를 개선하기 위한 루틴을 추가한 것으로 다음과 같은 순서로 실행한다.

step 1 : 클러스터 개수 c ($2 \leq c < n$), 가중치 m ($1 < m < \infty$), 수렴판정치 ϵ (threshold), c 개의 분할행렬인 $U^{(l)}$ 의 초기값 $U_{i,j=0}^{(l)}$ 을 적절히 설정한다.

step 2 : 클러스터의 중심벡터 $V_i^{(l)}$ ($i=1, 2, \dots, c$)를 식(2.2)에 의해 $U^{(l)}$ 을 이용하여 구한다.

step 3 : $X_k \neq V_i^{(l)}$ 일 경우, 식(2.3)에 의해 $U_{ik}^{(l+1)}$ 로 개선한다. $X_k = V_i^{(l)}$ 일 경우에는 식(2.4)를 이용하여 개선한다.

$$U_{ik}^{(l+1)} = \begin{cases} 1 & i \in I_k \\ 0 & i \notin I_k \end{cases} \quad (2.4)$$

$$\text{단, } I_k = \{ i \mid 1 \leq i \leq c, d_{ik} = \| X_k - V_i \| = 0 \} \\ \forall k = 1 \sim n$$

step 4 : 식(2.5)와 같이 $U^{(l)}$ 과 $U^{(l+1)}$ 의 차가 주어진 수령 판정치 ϵ 보다 작거나 같으면 종료하고, 그렇지 않으면 step2로 되돌아간다.

$$\| U^{(l+1)} - U^{(l)} \| \leq \epsilon \quad (2.5)$$

2.2 클러스터 증가 및 재초기화 알고리즘

최적의 클러스터 수의 결정은 식(2.6)에 의해 구한 $S(c)$ 를 최소로 하는 클러스터 수 c 로 하면 되지만 해석적인 c 의 결정법은 아직 알려지지 않고 있다. 기존에는 $S(c) \leq S(c+1)$ 을 만족하면, c 를 최적의 클러스터 수를 결정하는 방법이 사용되었으나, $S(c)$ 값의 미묘한 변화로 인한 클러스터 수의 증가로 클러스터링에 불합리한 점이 발생한다.

$$S(c) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m (\| X_k - V_i \|^2 - \| V_i - \bar{x} \|^2) \quad (2.6)$$

본 논문에서는 $S(c)$ 값의 차이가 임계값 M 이하일 경우 즉, 조건 $|S(c+1) - S(c)| \leq M$ 을 추가하여 두 조건 중 하나만 만족하면 해당 c 를 최적의 클러스터 수로 결정하고, 그렇지 않으면 클러스터 수를 1개씩 증가시키는 방식을 제안하고자 한다.

식(2.6)에서 n 은 데이터의 수, X_k 는 k 번째 데이터, \bar{x} 는 데이터의 평균, V_i 는 i 번째 클러스터의 중심벡터, μ_{ik} 는 k 번째 데이터가 i 번째 클러스터에 속하는 정도, m 은 가중치이다. 클러스터가 증가 할 때, U 의 c 분할행렬 U_{ik} 의 초기 설정이 필요하다. 본 논문에서 사용된 U_{ik} 의 재초기화 알고리즘은 다음과 같다.

step 1 : c 개의 각 클러스터의 중심벡터 V_i 에서 해당 클러스터에 속하는 데이터 X_k 까지의 거리를 구한다.

step 2 : 구한 클러스터 중에서 최대의 거리를 갖는 데이터 X_r 을 구한다.

step 3 : $k=r$ 인 경우, $U_{c+1,k}=1$ 로 두고, $k \neq r$ 인 경우, $U_{c+1,k}=0$, $U_{i,r}=0$ ($i=1, 2, \dots, c$)을 할당하여, $U_{c+1,k}$ 의 초기치를 설정한다.

2.3 언어레이블에 의한 클러스터의 표현

2.3.1 언어레이블의 필요성

폐지 클러스터링을 통해서 구해진 데이터들은 각각의 폐지 클러스터로 묶여서 사용자에게 제공된다. 이때, 각각의 데이터는 각 속성별로 수치적인 값으로 표현된다. 제공되는 데이터가 많지 않을 경우에는 문제가 없지만 많은 데이터가 사용자에게 제공된다면 검색에 있어서 불편함과 시간적, 경제적인 손실이 발생한다. 이러한 불편을 없애기 위해서 비슷한 특성을 갖는 데이터는 각각의 특성을 나타내는 언어레이블을 통해서 표현함으로

서 효율적인 검색을 지원할 수 있다.

2.3.2 언어레이블의 선정

언어레이블은 데이터가 갖는 특성에 따라서 적절한 수의 레이블을 상호 간의 관계를 고려하여 선정한다. 그림 1에서 데이터 x_k 는 속성 P_j 에 대해서 언어레이블 $L_{P_j}^s$ 와는 0.2의 관련성이 있고, 언어레이블 $L_{P_j}^3$ 과는 0.7의 관련성이 있다. 이와 같이 각각의 데이터는 복수개의 레이블과 관련이 있을 수 있으므로, 각각의 언어레이블 ($L_{P_j}^1, L_{P_j}^2, \dots, L_{P_j}^s$)의 배치는 서로간의 관계를 고려하여 결정한다.

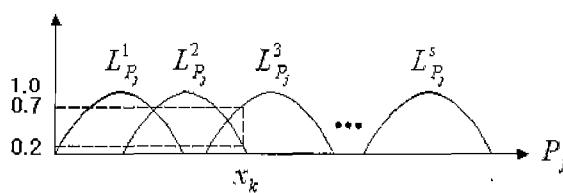


그림 1. 언어레이블을 위한 멤버쉽 함수

Fig. 1 Membership function for linguistic labels.

2.3.3 적정 언어레이블의 할당

결정된 각 퍼지 클러스터를 언어적으로 표현하기 위해 그림 1과 같이 데이터 x_{kp} 의 j 개의 속성별로 적당한 s 개의 언어적 레이블 $L_{P_j}^s$ 를 할당한다. x_{kp} 의 i 번째 클러스터에 속하는 정도 U_{ik} 를 각 속성에 사상시켜, i 번째 클러스터에 대한 속성별 멤버쉽 함수를 구한다. 구해진 속성별 언어레이블의 멤버쉽 함수와 U_{ik} 와의 적합도를 식(2.7)에서 구하여, C_s 를 최소로 하는 언어레이블이 i 번째 클러스터에 할당된다. 식(2.7)에서 클러스터에 속하는 정도가 미소한 데이터로 인하여 C_s 가 증가함을 피하기 위해 적절한 임계치 α 이상의 소속도를 갖는 데이터만을 대상으로 하여 적합도를 구한다.

$$C_s = \sum_{k=1}^n e_k \quad (2.7)$$

$$e_k = \begin{cases} U_{ik} - \mu_{L_i^s}(x_{kp}) & U_{ik} \geq \mu_{L_i^s}(x_{kp}), U_{ik} \geq \alpha \\ 0 & U_{ik} < \mu_{L_i^s}(x_{kp}) \end{cases}$$

U_{ik} : x_k 가 i 번째 클러스터에 속하는 정도

L_i^s : i 번째 속성에 대한 s 번째 언어레이블

$\mu_{L_i^s}(x_{kp})$: i 번째 속성에 대한 k 번째 데이터 x_k 의 s 번째 언어레이블의 멤버쉽 함수

3. 대체응답 알고리즘

클러스터링을 통해서 구한 퍼지 클러스터에 언어적인 레이블을 할당하여 구축한 지식 데이터베이스로부터 사용자의 입력을 검색하여, 대응하는 데이터가 있으면 해

당 데이터를 출력하고, 그렇지 않으면 대체응답을 제공하는 협조적인 응답시스템의 구축을 위한 알고리즘은 다음과 같다.

step 1 : 사용자는 j 번째 정량적 속성에 대해서는 수치 K_j 를 입력하고, k 번째 정성적인 속성에 대해서는 시스템이 제공한 언어레이블 중에서 i 번째 레이블 L_i^s 를 선택한다.

step 2 : 정성적 속성에 대해 입력된 언어레이블 L_i^s 의 중심값 avg_k 를 계산한다.

step 3 : 사용자의 입력 즉, 벡터 $V_{input}(K_j, avg_k)$

$j=1, 2, \dots, n$ $k=1, 2, \dots, m$ 와 퍼지 클러스터링을 통해 생성된 퍼지 클러스터의 중심벡터 V_i 의 유크리트 거리 (Euclidean Distance)를 구한다.

step 4 : 최소의 거리를 갖는 퍼지 클러스터의 언어레이블을 출력하고, 해당 클러스터의 데이터를 출력한다.

4. 데이터베이스 및 검색시스템 구축

데이터베이스의 효율적인 검색을 위해 일반 데이터베이스와 이미지 데이터베이스 그리고 지식기반 데이터베이스로 구분하여 DB 검색시스템을 구축하였다. 지식기반 검색시스템의 구성을 그림 2에 나타내었다. 그림 2에서 보여주는 것처럼 사용자의 요구에 일치하는 데이터가 데이터베이스 내에 존재할 때, 해당 데이터의 내용과 이미지를 제공하고, 만일 사용자의 요구에 일치하는 데이터가 존재하지 않을 때는 지식기반 데이터베이스 (KDB)를 구축하여 사용자의 요구에 가장 근접한 데이터와 해당 데이터의 이미지 정보를 제시하도록 하였다. 일반 DB 검색과 FCM에 의해 구축된 지식기반 DB 검색이 가능하도록 하였으며, 화상 DB와의 연동이 가능하도록 설정하였다.

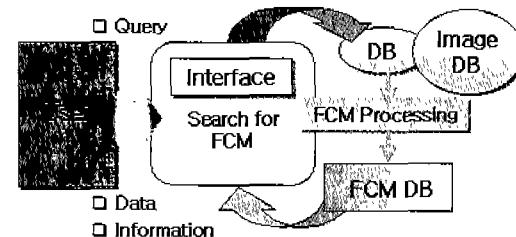


그림 2. 지식기반 데이터베이스 검색시스템의 구성

Fig. 2 Configuration of retrieval system for KDB

4.1 데이터베이스 구축

4.1.1 데이터베이스 구축

데이터베이스 구축을 위해 우체국에서 발매하는 우편주문판매상품 안내서를 바탕으로 데이터베이스를 구축하였다[8]. 우편주문안내책자의 데이터를 M/S 액세스를 이용하여 작성하였고[9], 상품번호, 상품명, 지역, 상품내용, 가격, 공급처 등의 텍스트 데이터를 설정하여 총 312개의 데이터를 입력하였다.

4.1.2 이미지 데이터베이스 구축

이미지 데이터의 설정을 위해 각 데이터의 이미지를 스케닝하여 이미지 파일 포맷으로 생성한 후 데이터베이스 내에 삽입 저장하여 이미지 데이터베이스를 구축하였다. 이미지 파일은 JPG, BMP, TIF의 파일형식을 입력할 수 있도록 설정하였다. 데이터베이스에 이미지를 저장하는 방법으로는 이미지를 따로 파일에 저장하고 그 파일의 파일명만을 데이터베이스에 문자열로 저장하는 방법과 이미지 데이터 자체를 데이터베이스에 저장하는 방법이 있으나 이미지 파일을 따로 관리해야 하는 번거로움이 있기 때문에, 보다 정확하고 효율적인 이미지 검색을 위해 이미지 파일을 해당 레코드에 삽입하는 방식을 취했다. 데이터베이스 내에 삽입을 위해 긴 이진 파일(Long Binary File)로 변환하여 해당 레코드에 삽입하여 데이터베이스를 구축하였다.

4.2 검색을 위한 인터페이스 설정

검색을 위한 인터페이스와 지식기반 데이터베이스(KDB) 알고리즘은 Visual C++ 6.0을 이용하여 구축하였다[10].

검색프로그램이 실행되면 메인 화면이 출력되고 지능 검색과 일반검색으로 나누어 검색하도록 검색버튼 영역을 설정하였다. 먼저 지능검색에는 지식기반 데이터베이스 검색을 설정하였고, 일반검색에는 일반 데이터 검색과 지역별 검색 그리고 항목별 검색을 설정하여 검색하도록 하였다. 사용자가 해당 검색버튼을 누르면 새로운 입력상자가 출력되어 각각의 항목을 선택할 수 있도록 설정하였다. 사용자의 요구에 의해 검색된 데이터를 출력하기 위해 리스트 컨트롤 박스(List Control Box)를 설정하여 검색된 데이터는 그림 4와 같이 리스트 컨트롤 박스에 검색결과를 출력하게 된다. 이 출력된 데이터 중 하나를 선택하게 되면 보다 상세한 텍스트 데이터 정보와 해당 이미지 데이터를 보여준다.

4.3 속성에 대한 언어레이블과 멤버쉽 함수

데이터의 표현 방식은 정량적 속성과 정성적 속성으로 구분하여 설정하였다. 여기에서 ‘나이’나 ‘가격’같은 정량적인 속성은 클러스터링에 사용하기 용이하나 ‘대상’이나 ‘용도’와 같은 정성적 속성은 클러스터링에 사용하기가 쉽지 않으므로 적절히 정량화 하여야 한다. 본 검색시스템에서는 정성적인 속성을 경험적 지식을 토대로 정성적 속성의 각각의 레이블과 데이터와의 관계를 분석하여 정량화 하였다. 각 속성에 대한 언어레이블은 표 1과 같이 설정하였고, 각 속성의 언어레이블에 대한 멤버쉽 함수는 그림 3와 같이 설정하였다.

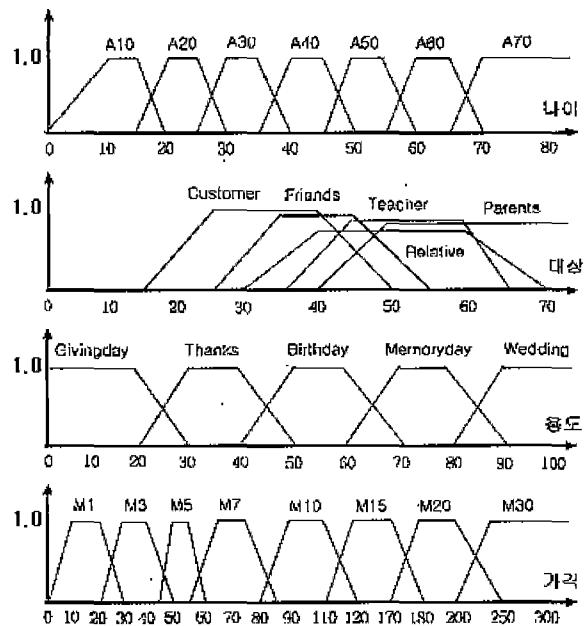


그림 3. 각 속성의 언어레이블 표현

Fig. 3 Expression of each property's label

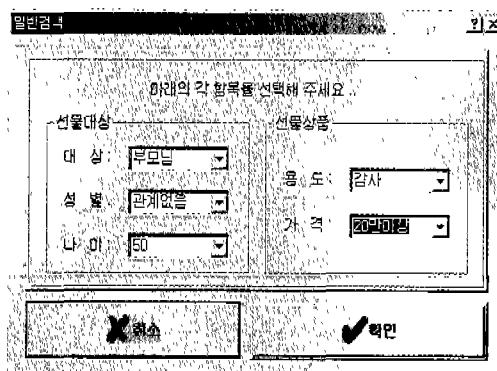
표 1. 속성의 언어레이블 표현

Table 1 linguistic label Expression of property

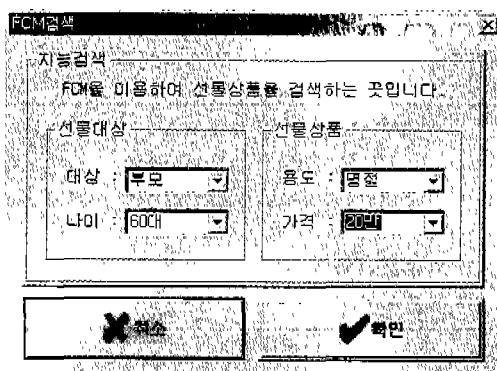
Property	Linguistic Labels
Object	Customer, Friends, Teacher, Relatives, Parents
Age	A10, A20, A30, A40, A50, A60, A70
Use	Givingday, Thanks, Birthday, Memoryday, Wedding
Price	M1, M3, M5, M7, M10, M15, M20, M30

4.4 검색시스템의 활용 예

그림 3은 일반 검색 질의와 지식기반 검색 질의를 비교하기 위한 입력 상자를 나타낸다. 즉, 이용자가 “나이가 60세 정도의 부모님께 명절 선물로 20만원 정도의 선물을 하려고 한다”면 각각 그림 4와 같은 검색결과를 출력한다.



(a) 일반 검색 질의



(b) 지식기반 검색 질의

그림 3 지식기반 검색과 일반 검색 질의 상자

Fig. 3 Dialog box of retrieval system
for KDB and normal DB

그림 4는 사용자의 질의에 의해 검색된 상품을 출력하여 보여주는 예이다.

데이터베이스 접속 시스템																											
파일(F) 작업(W) 검색(S) 레코드(R) 보기(V) 도움말(H)																											
설명																											
지식기반검색은 이용하여 검색하는 품목입니다. 검색을 원하시면 해당 검색버튼을 클릭주세요.																											
<input checked="" type="checkbox"/> 일반검색 <input type="checkbox"/> 지식기반검색 <input type="checkbox"/> 지역별검색 <input type="checkbox"/> 화목별검색 <input type="checkbox"/> 자세히보기(Y) <input type="checkbox"/> 종료(X)																											
선물상품 대상 : 부모님 용도 : 감사 나이 : 50세 가격 : 20만원 검색 확인																											
<table border="1"> <thead> <tr> <th>상품명</th> <th>내용</th> <th>가격</th> </tr> </thead> <tbody> <tr> <td>545003 광양청매살(간)</td> <td>1.2kg</td> <td>220,000</td> </tr> <tr> <td>550101 대수피문어</td> <td>3.8kg(100미리)</td> <td>180,000</td> </tr> <tr> <td>513005 영광방송기풀비</td> <td>5대(100미리)</td> <td>250,000</td> </tr> <tr> <td>513006 영광방송기풀비</td> <td>한대(500미리)</td> <td>350,000</td> </tr> <tr> <td>513105 영광방송기풀비</td> <td>2대(100미리)</td> <td>200,000</td> </tr> <tr> <td>513204 영광방송기풀비</td> <td>3대(100미리)</td> <td>250,000</td> </tr> <tr> <td>513207 영광방송기풀비</td> <td>마든공비 족대(400)</td> <td>260,000</td> </tr> </tbody> </table>				상품명	내용	가격	545003 광양청매살(간)	1.2kg	220,000	550101 대수피문어	3.8kg(100미리)	180,000	513005 영광방송기풀비	5대(100미리)	250,000	513006 영광방송기풀비	한대(500미리)	350,000	513105 영광방송기풀비	2대(100미리)	200,000	513204 영광방송기풀비	3대(100미리)	250,000	513207 영광방송기풀비	마든공비 족대(400)	260,000
상품명	내용	가격																									
545003 광양청매살(간)	1.2kg	220,000																									
550101 대수피문어	3.8kg(100미리)	180,000																									
513005 영광방송기풀비	5대(100미리)	250,000																									
513006 영광방송기풀비	한대(500미리)	350,000																									
513105 영광방송기풀비	2대(100미리)	200,000																									
513204 영광방송기풀비	3대(100미리)	250,000																									
513207 영광방송기풀비	마든공비 족대(400)	260,000																									

(a) 일반 질의에 의한 검색 결과 출력

데이터베이스 접속 시스템																											
파일(F) 작업(W) 검색(S) 레코드(R) 보기(V) 도움말(H)																											
설명																											
지식기반DB를 이용하여 검색하는 프로그램입니다. 검색을 원하시면 해당 검색버튼을 클릭주세요.																											
<input checked="" type="checkbox"/> FDM(D) <input type="checkbox"/> 일반검색(N) <input type="checkbox"/> 지식별검색(P) <input type="checkbox"/> 지역별검색(L) <input type="checkbox"/> 자세히보기(Y) <input type="checkbox"/> 종료(X)																											
<table border="1"> <thead> <tr> <th>상품명</th> <th>내용</th> <th>가격</th> </tr> </thead> <tbody> <tr> <td>545003 광양청매살(간)</td> <td>1.2kg</td> <td>220,000</td> </tr> <tr> <td>550101 대수피문어</td> <td>3.8kg(100미리)</td> <td>180,000</td> </tr> <tr> <td>513005 영광방송기풀비</td> <td>5대(100미리)</td> <td>250,000</td> </tr> <tr> <td>513006 영광방송기풀비</td> <td>한대(500미리)</td> <td>350,000</td> </tr> <tr> <td>513105 영광방송기풀비</td> <td>2대(100미리)</td> <td>200,000</td> </tr> <tr> <td>513204 영광방송기풀비</td> <td>3대(100미리)</td> <td>250,000</td> </tr> <tr> <td>513207 영광방송기풀비</td> <td>마든공비 족대(400)</td> <td>260,000</td> </tr> </tbody> </table>				상품명	내용	가격	545003 광양청매살(간)	1.2kg	220,000	550101 대수피문어	3.8kg(100미리)	180,000	513005 영광방송기풀비	5대(100미리)	250,000	513006 영광방송기풀비	한대(500미리)	350,000	513105 영광방송기풀비	2대(100미리)	200,000	513204 영광방송기풀비	3대(100미리)	250,000	513207 영광방송기풀비	마든공비 족대(400)	260,000
상품명	내용	가격																									
545003 광양청매살(간)	1.2kg	220,000																									
550101 대수피문어	3.8kg(100미리)	180,000																									
513005 영광방송기풀비	5대(100미리)	250,000																									
513006 영광방송기풀비	한대(500미리)	350,000																									
513105 영광방송기풀비	2대(100미리)	200,000																									
513204 영광방송기풀비	3대(100미리)	250,000																									
513207 영광방송기풀비	마든공비 족대(400)	260,000																									

(b) 지식기반 검색 결과 출력

그림 4. 검색된 데이터 출력
Fig. 4 Output of retrieval data

데이터베이스 접속 시스템			
파일(F) 작업(W) 검색(S) 레코드(R) 보기(V) 도움말(H)			
세부설명			
항 목 : 음식점 품 명 : 광양청매살(간) 내 용 : 1.2kg 가 격 : 220,000 담당자 : 광양청매살(간)영원정동호회집 연락처 : 광양다민투체국-(0687)772-01			
이미지는 상세정보를 보여주는 곳입니다. 이전보기 상품검색(S)			

그림 5. 데이터 정보와 이미지 출력

Fig. 5 Output of information and image data

그림 4의 결과에서 보여주는 바와 같이 일반검색의 경우 단순한 SQL 질의에 의한 검색이므로 이용자의 질의에 대해 막연한 데이터나 검색 조건에 일치하지 않는 데이터를 제공하지만, 지식기반 데이터베이스 검색의 경우에는 이용자의 질에 대해 만족할 만한 데이터 제시하고 검색조건에 일치하는 데이터가 없을 경우에는 가장 근접한 데이터를 출력한다.

그림 4.5는 검색 질의의 상자에 의해 검색된 데이터 중에서 하나의 데이터를 선택할 경우, 해당 데이터의 상세 정보를 보여주는 경우이다. 선택된 데이터의 텍스트 정보와 해당 이미지를 보여준다.

5. 결 론

기존 데이터베이스 검색시스템은 방대한 데이터로부터

터 사용자가 원하는 요긴한 소량의 데이터를 검색할 수 없고 사용자의 막연한 검색 질문에도 대응할 수 없다는 문제점이 있다. 이러한 문제점을 해결하기 위한 연구의 일환으로 본 논문에서는 FCM을 이용하여 지식기반 데이터베이스를 구축하고 이를 언어적으로 표현하여 사용자의 막연한 질문에도 대응할 수 있는 데이터베이스 검색 시스템의 구축 기법을 제안하였다. 구체적인 연구내용으로는, FCM의 클러스터 증가 및 재초기화 알고리즘을 제안하고, 이미지 데이터와 연동하는 기법을 제안하였다. 제안된 기법을 우체국에서 발매하는 우편주문안내책자를 이용한 선물 고르기 데이터베이스 검색 시스템에 적용하여 그 실효성을 확인하였다.

앞으로 보다 실용적인 검색시스템의 구축을 위해서는 입력 데이터의 편중성과 시스템 관련 문제점의 해결, 그리고 클러스터링을 위한 정성적인 속성의 효율적인 정량화 기법 및 시스템과 이용자간의 보다 원활한 인터페이스 방법, 사물에 대한 인간의 개념적 사고와 같은 개념적인 속성을 갖는 데이터를 클러스터링 할 수 있는 개념 클러스터링에 관한 연구가 필요하다.

참 고 문 헌

- [1] T. Gaasterland, P. Godfrey, J. Minker, "An Overview of Cooperative Answering", Journal of Intelligent Information System, 1, pp.123-157, 1992.
- [2] S. Miyamoto, "Fuzzy Sets in Information Retrieval and Cluster Analysis", Theory and Decision Library, Series D. Kluwer Academic Publisher, 1990.
- [3] Jun Ozawa and Koichi Yamada, "Generating a fuzzy model from a database and using it to find alternative data", proc. of First Australian and New Zealand Conference on Intelligent Information Systems, ANZIIS-93, pp.560-564, 1993.
- [4] I. Jung, G. K. Park, W. Hwang, "Intelligent Retrieval System using FCM", proc. of Korea Fuzzy Logic and Intelligent Systems Society Fall Conference '95, Vol. 5, No. 2, pp.40-44, 1995.
- [5] 박계각, 서기열, 임정빈, "지식기반 데이터베이스 검색 시스템의 구축", 한국해양정보통신학회 '99 추계 종합학술대회지, pp.450-453, 1999.
- [6] 주혜종, 박상원, 이상필, "데이터베이스 총론", 정일, pp.391-454, 1999.
- [7] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithm", Plenum Press, New York, 1981.
- [8] '98우편주문판매상품안내, 우체국, 1998.
- [9] 김성원, "한글 액세스97", 혜지원, 1997.
- [10] 김용성, "Visual C++ 6.0 완벽가이드", 영진출판사, 1998.

저 자 소 개

박 계 각 (Gyei-Kark Park)



1982년 한국해양대학교 항해학과 졸업
1986년 동 대학원 해사수송과학 석사
1993년 일본 동경공업대학교 박사
1995년 ~ 현재 목포해양대학교 교수
관심분야 : 퍼지제어, 지능제어, DB
지식처리

Phone : 061-240-7128

Fax : 061-240-7281

E-mail : gkpark@mail.mmu.ac.kr

서 기 열 (Ki-Yeol Seo)



1995년 동신대 전자공학과 졸업
1998년 동 대학원 제어계측 전공 석사
2000~현재 목포해양대학교 해사정보계
측공학 전공 박사과정
관심분야 : 퍼지응용, 지능제어, 음성인식

Phone : 061-240-7128

Fax : 061-240-7281

E-mail : vito@hanmir.com

양 원 재 (Won-Jea Yang)



1990년 한국해양대학교 항해학과 졸업
1999년 동 대학원 해사정보공학 석사
1999년 ~ 현재 동 대학원 박사과정
관심분야 : 퍼지응용, 의사결정, 평가

Phone : 061-240-7128

Fax : 061-240-7281

E-mail : ywj07@hanmail.net

천 대 일 (Dea-II Cheon)



1997년 목포해양대학교 해상운송시스템
학부 졸업
2000년 ~ 현재 동 대학원 해사정보계측공
학 전공 석사과정
관심분야 : 퍼지응용, 지능제어, 음성인식

Phone : 061-240-7128

Fax : 061-240-7281

E-mail : chiefoff@dreamwiz.com