

지능적 비즈니스를 위한 XML 기반 실시간 의미정보지향 서비스 시스템 설계

홍성용*, 진혜진**

*사바나주립대학교 컴퓨터정보시스템학과

**고려대학교 산업시스템정보공학과

*e-mail : hongs@savstate.edu,

** e-mail : hjjin2003@korea.ac.kr

XML based Real-Time Semantic Information Oriented Service System Design for Intelligent Business

Seong-Yong Hong*, Hye-Jin Jin**

Dept. of Computer Information System, Savannah State University

Dept. of Industrial Systems and Information Engineering, Korea University

요 약

최근 e-비즈니스나 인터넷 쇼핑몰 사이트에서는 많은 양의 상품 이미지 정보와 콘텐츠를 취급하고 있으며, 이로 인하여 효과적인 실시간 의미정보 검색의 필요성이 대두되고 있다. 본 논문에서는 XML 기술과 퍼지기술을 이용하여 웹상의 상품 정보를 의미적으로 검색할 수 있는 시스템에 대해 기술한다. 본 논문에서 제시한 시스템은 상품 이미지 XML 문서를 이용하여 메타데이터를 추출하고, 추출된 메타데이터를 의미정보로 자동 변환하여 실시간 서비스할 수 있는 방법을 제안한다. 따라서, 인터넷 비즈니스를 좀 더 지능화 할 수 있는 시맨틱 웹 환경이나 Web2.0 을 구현하는데 기여할 수 있을 것이라 기대한다. 또한 사용자의 입장에서는 실시간 의미정보를 서비스 받을 수 있는 장점을 가지게 될 것이다.

1. 서론

최근 인터넷의 사용이 급격히 증가함에 따라 새로운 기술 용어와 웹 서비스(Web Service) 개념이 여러 분야에서 활발히 연구되고 있다. 특히, XML(eXtensible Markup Language) 기술이 보급된 이후로, XML 을 기반으로 하는 웹 서비스 기술 개발이 활발하게 이루어져 가고 있으며, 이에 따른 사용자의 참여도 많이 증가하고 있다[1]. 또한 이러한 시대적 변화는 새로운 인터넷 관련 용어를 탄생시키고 사용하고 있다. 대표적인 예로 UCC(User-Created Content)라는 용어는 사용자가 직접 콘텐츠를 제작, 편집하여 웹을 통해 서비스하는 개념을 말한다. 이뿐만 아니라, 최근에 웹을 웹 2.0 이라는 이름으로 새로운 웹 문화가 빠르게 변화되어 가

고 있으며, 시맨틱 웹과 같이 정보의 표현을 좀 더 의미적으로 표현하고 서비스할 수 있는 개념을 많이 연구하고 있다[2]. 즉, 이러한 웹 문화의 변화는 단순한 정보를 제공하기 위한 비즈니스환경을 위해 사용하던 인터넷 환경을 일반 사용자의 입장에서 쉽고 빠르게 이해하면서 사용 할 수 있도록 변화가 일어난 것이다. 쉽게 표현하면, “찾아가는 정보가 아닌 찾아오는 정보” 시대를 만들어 나가고 있는 것이라고 할 수 있다. 예로, 인터넷 블로그(blog) 사용이나 RSS, Ajax 와 같은 기술은 사용자의 적극적인 인터넷 사용을 도모하고 있다. 하지만, 이러한 모든 기술에 공통된 문제점 중에 하나는 사용자가 정보를 직접 제작하거나 표현해야 한다는 것이다. 또한 제작된 정보 표현은 의미적으로 제공되지 못하고 있다. 본 논문에서 말하는 의미적 정보라는 것은 메타데이터에 대하여 사용자가 필요로 하는 의미적 정보를 지닌 표현을 말한다. 예를

* This research was supported by Savannah State University of Georgia.

들어 P 상품의 가격이 수치적으로 표현되어 있을 때, 그 수치의 의미가 사용자에게 의미적으로 전달될 수 있도록 변환되는 것이다.

따라서, 본 논문에서 실시간 의미정보 지향 서비스를 하기위해 상품 카탈로그 이미지를 표현한 XML 문서를 사용하여, 상품의 가격에 해당하는 메타데이터를 먼저 추출하고 해당 상품이 비싼 가격대의 상품인지, 싼 가격대의 상품인지를 자동으로 분석하여 실시간 상품 이미지에 의미정보를 서비스 혹은 검색 가능하게 하는 것이다.

본 논문에서는 의미정보를 자동으로 생성하기 위해 퍼지집합 이론을 적용할 수 있는 방법을 제시하고, 웹상의 실제 상품 카탈로그 이미지 정보에 반영하여 의미성을 해석할 수 있는 알고리즘을 연구하고 적용하였다. 또한, 상품 이미지 메타데이터의 갱신에 따른 트리거 이벤트(Trigger Event)를 설계하고, 적용 사례를 살펴본다.

본 논문의 구성은 다음과 같다. 2 장에서는 관련연구에 대하여 소개하고, 3 장에서는 상품이미지 메타데이터에 의한 퍼지 데이터를 자동 생성하기 위한 알고리즘과 의미정보 생성기법을 설명한다. 4 장에서는 실험 평가에 대하여 살펴본다. 마지막으로 결론과 향후 연구 과제에 대하여 기술한다.

2. 관련 연구

상품 이미지와 같은 멀티미디어 데이터는 다양하고 방대한 양의 정보를 포함하고 있어 효율적인 검색을 하기 위해서는 데이터를 기술하는 방법이 구조적이고 체계화된 형태의 메타데이터가 요구된다. 그래서 XML 로 이미지 메타데이터를 표현하는 연구가 활발히 진행되고 있으며, 이러한 연구들은 이미지 데이터가 내포하고 있는 다양하고 복잡한 정보를 표현하고 있다. 그러나 이와 같은 이미지 표현 방법들은 이미지 자체의 데이터만을 고려하거나 단순한 메타데이터만을 표현하므로 실제 인간이 표현하는 의미정보 표현의 어려움이 있다. 또한, 사용자로부터 의미성을 포함한 효과적인 검색을 지원할 수 없고, 표현의 애매성을 고려하지 못하고 있다. 인간적 표현의 의미성이란 데이터의 수치적 표현이 변화됨에 따라 인간이 표현할 수 있는 의미가 달라지는 것을 말한다. 예를들어, 2004 년에 50,000 원인 핸드폰이 “매우 비싸다”고 생각했다면, 2005 년에는 “싸다”고 표현할 수 있으며, 현재 2007 년에는 “매우 싸다”라고 표현할 수 있다. 즉, 같은 상품의 수치적 데이터가 시간적으로 변화됨에 따라 인간의 의미성 부여는 변화하게 된다. 본 장에서는 XML 관련 기술과 수치적 데이터에 의미성을 부여하기 위한 퍼지 기술에 대해서 살펴본다.

2.1 XML 기술

XML 은 최근 시맨틱 웹 환경을 만들어 나가기 위한 가장 기본적인 마크업 언어로 사용되고 있으며, RDF, RSS 와 같은 Web2.0 기술에 토대가 되는 웹 표준 언어이다[2]. 다양한 정보 형태를 가진 웹 문서의

효과적인 관리를 위해 XML 관련 연구들이 현재 국내외에서 활발히 진행되고 있다. 특히 멀티미디어 정보를 표현하기 위해 XML 을 사용하고 다양한 검색 방법을 제공하기 위해 XML 질의어를 사용하고 있다. 관련 질의어는 XPath, XML-QL, XQL, Quilt, XQuery 등이 있으며, 이들은 XML 문서의 구조적 특성을 반영한 구조-내용기반 검색을 지원하고 있다[3]. 이러한 XML 질의어를 활용하면 내용 기반 검색, 구조 기반 검색 등과 같은 다양한 형태의 XML 문서 검색 질의를 표현할 수 있다. 그림 1 은 상품이미지에 대한 XML 표현 예를 보이고 있다.

```
<?xml version="1.0" encoding="UTF-8" ?>
<images>
<image>
<id>i00001</id>
<ref> image01.jpg</ref>
<subject> cellular phone </subject>
<category> phone </category>
<object> <size unit="cm"> 5 *6 </size>
<made> <company> samsung </company>
<nation> korea </nation>
<brand> Anycall </brand>
<date> Oct 15 2006</date>
</made>
<price unit="won">50000</price>
</object>
</image>
</images>
```

(그림 1) 상품 이미지 XML 표현 예

2.2 퍼지 기술

최근 인간과 비슷하게 생각하고, 일하는 컴퓨터를 만들고자 하는 연구가 활발하게 진행되고 있다. 컴퓨터가 인공지능을 가지고 인간이 원하는 바를 제대로 수행하기 위해서는 인간이 사용하는 숫자는 물론이고 애매한 표현을 처리할 수 있어야 한다. 이러한 애매한 표현을 처리할 수 있는 이론적인 바탕을 제공하는 것이 바로 퍼지 이론(fuzzy theory)이다[4]. 예를 들어, 인간이 일반적으로 사용하는 표현으로 “매우비싼(very expensive)” 혹은 “비싼(expensive)”과 같은 표현은 정형적인 형태의 데이터로 표현하기가 어렵다. 또한, 이미지 색상에서도 “약간 붉은 (light red)” 혹은 “거의 붉은(dark red)”과 같은 정보는 애매한 표현으로 나타내게 된다. 퍼지 집합(fuzzy set)은 애매한 개념을 다루는 집합 개념이다. 의미적으로 접근되어 애매성이 있으며 구별하기 힘든 데이터를 효과적으로 관리하기 위한 방안의 하나로 퍼지 집합을 많이 사용한다.

전체 상품이미지 집합 I 를 I={i1, i2, i3,.....,in }로 정의한다면, 이 경우 퍼지 집합 A 는 공식(1)과 같이 정의한다.

$$A = \mu_A(i_1)/i_1 + \mu_A(i_2)/i_2 + \dots + \mu_A(i_n)/i_n$$

$$= \sum_{k=1}^n \mu_A(i_k)/i_k \quad \dots\dots(1)$$

또한, 연속적인 표현에서는 소속 함수를 사용하여 공식(2)와 같이 정의한다.

$$\int \mu_A(i)/i \quad \dots\dots\dots(2)$$

두 퍼지 집합 A, B 에 대한 연산은 공식(3)과 같이 정의한다.

$$A \cap B = \int_i \mu_A(i) / i \cap \int_i \mu_B(i) / i = \int_i (\mu_A(i) \wedge \mu_B(i)) / i$$

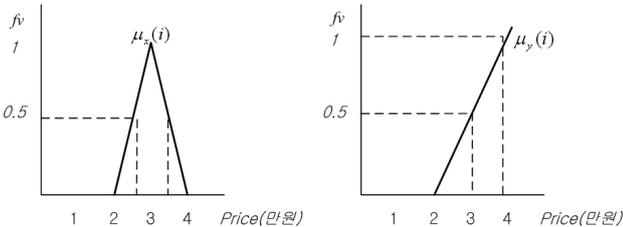
$$A \cup B = \int_i \mu_A(i) / i \cup \int_i \mu_B(i) / i = \int_i (\mu_A(i) \vee \mu_B(i)) / i$$

$$A^c = \mu_{A^c}(i) = 1 - \mu_A(i) = \int_i (1 - \mu_A(i)) \dots\dots\dots (3)$$

3. 의미정보 생성 기법

본 논문에서는 상품이미지에 대한 XML문서로 부터 가격정보를 추출하여, 추출된 수치적 데이터를 의미적 정보로 자동 생성할 수 있는 기법을 소개한다.

그림 3은 상품 카탈로그 이미지의 메타데이터 속성 중 가격을 퍼지 집합 표현에 적용한 예를 보이고 있다. 그림3의 (a)는 가격의 범위를 2~4만원으로 한정된 소속 함수로 표현된 경우를 나타내며, (b)는 2만원 이상으로 연속적 소속 함수로 나타낸다.



(a) 2~4만원 소속함수 (b) 2만원 이상 소속함수
(그림 3) 가격 메타데이터 퍼지 집합표현

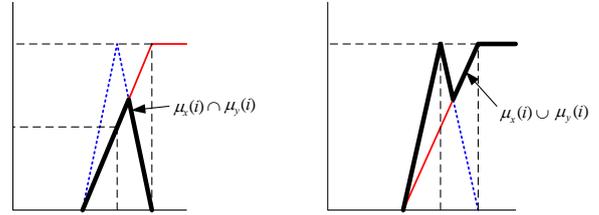
전체 상품이미지 집합 I 에 관한 퍼지 집합 A 인 경우 $\mu_A : I \rightarrow [0,1]$ 되는 소속함수 $u_A(i)$ 에 의해 특성 지워진 집합이고, 소속함수 $u_A(i)$ 는 A 에 관한 I 의 소속도를 나타낸다. 이 경우 $u_A(i)$ 의 값이 1 에 가까우면 i 의 A 에 속하는 정도가 크고, 반대로 0 에 가까우면 i 의 A 에 속하는 정도가 적은 것을 나타내고 있다. 퍼지 부분집합 A 는 원소 I 와 소속도 $u_A(i)$ 쌍의 집합으로 식(4)와 같이 표현한다.

$$A = (i, \mu_A(i)) \mid i \in I \dots\dots\dots (4)$$

그림 3 의 (a)의 경우 가격이 30,000 원정도인 핸드폰 이미지에 대한 퍼지 집합은 $i_x = \{ \text{가격이 30,000 원 정도인 핸드폰 상품} \}$ 라고 정의할 수 있고, 쌍의 집합으로 나타내면 $\{ \dots(25,000,0.5), (30,000,1.0), (35,000,0.5) \dots \}$ 와 같은 수치적 집합을 가지게 된다. 그림 3 의 (b)는 가격이 비싼 핸드폰 이미지에 대한 퍼지 집합을 나타내고 있다. 만약, 가장 가격이 낮은 핸드폰이 20,000 이고, 가장 비싼 핸드폰 가격이 40,000 이라고 한다면 $i_y = \{ \text{가격이 비싼 핸드폰 상품} \}$ 이라고 정의할 수 있고, 쌍의 집합으로 나타내면 $\{ (20,000,0.0), (30,000,0.5), (40,000,1.0) \}$ 이다.

그림 4 의 (a)는 그림 3 에서 정의된 두 경우를 교집합으로 표현 함으로서 “가격이 3 만원 정도이면서 비싼 핸드폰 상품” 에 대한 퍼지 집합을 표현하고 있다. 또한 (b)는 “가격이 3 만원 정도이거나 비싼 핸드폰 상품” 에 대한 퍼지 집합을 표현하고 있다.

상품 이미지 메타데이터 중에서 가격정보에 대한 퍼지 데이터를 생성하기 위해서 그림 5 와 같은 알고리즘을 제안한다. 제안된 알고리즘의 기본연산 방식은 식 (5)와 같다.



(a) 퍼지 교집합 표현 (b) 퍼지 합집합 표현
(그림 4) 퍼지 집합 연산 표현

$$\left(\left[x - \min_{x \in X} x \right] / \left[\max_{x \in X} x - \min_{x \in X} x \right] \right) \dots\dots\dots (5)$$

여기서, x 는 각 상품이미지 메타데이터중에 실제 가격 데이터를 의미한다. 만약 상품이미지 i_x 의 가격이 35,000 원이라면, *real_value* 는 35000 가 된다. *min_value* 는 전체 가격데이터중에 가장 최소값을 의미한다. 만약, 전체 상품이미지가 10,000 개일 경우에 그 중에서 가장 적은 가격 정보를 가진 i_x 의 가격이 18,000 원이라면, *min_value* 는 18000 이 된다. *max_value* 는 전체 가격 데이터중에 최대값을 의미한다. 만약, i_x 의 가격이 49,000 원으로 전체 상품이미지 가격중에 가장 큰 가격 정보를 가지고 있다면, *max_valu* 는 49000 이 된다. 그러기 때문에 우선 이미지 메타데이터로부터 각 최소값과 최대값의 데이터를 추출하고, 각 튜플(tuple)에 가격데이터를 추출하여 연산한다. 연산된 퍼지 값은 *fuzzy_value* 에 저장되고, 데이터베이스의 퍼지 사전 테이블에 저장한다.

```

Algorithm Generate_fuzzy_data (int num, int row_count, float
min_value, float max_value, float real_value, float fuzzy_value)
begin
num ← 1;
row_count ← select_from_database(i_n); /* 전체 record 의 수 */
min_value ← select_from_database(min, i_x); /* 최소값 */
max_value ← select_from_database(max, i_x); /* 최대값 */
while num < (row_count+1)
begin
real_value = select_from_table_where(num);
front_value ← real_value - min_value;
rear_value ← max_value - min_value;
fuzzy_value ← (front_value / rear_value);
if fuzzy_value ≠ null do
begin
update_to_table_where(num);
end
num ← num+1;
end
end
    
```

(그림 5) 가격정보 퍼지데이터 생성 알고리즘

퍼지 데이터 생성 알고리즘은 저장 프로시저를 생성하여 데이터베이스에 저장한다. 저장된 프로시저는 이미지 메타데이터를 기반으로 연산을 하여 퍼지 데이터를 생성해 낸다. 생성된 퍼지데이터는 의미정보를 생성하기 위해 그림 6 의 알고리즘을 통해 의미정보를

자동생성하여 저장하게 된다. 만약 새로운 이미지 메타데이터가 삽입되거나 수정되면, 트리거(Trigger) 동작에 의해 저장 프로시저는 자동으로 재 수행을 하고, 새로운 퍼지 데이터를 생성해 실시간 저장하게 된다.

```

Algorithm Generate_semantic_information (int get_price, float fuzzy_data)
begin
  fuzzy_data ← select_from_table(get_price)
  If fuzzy_data < low_weight then
    { If SI_data == null
      insert_into_table_value(Ti) /* Ti: Term */
    Else
      update_to_table_value(Ti) }
  Else If fuzzy_data < middle_weight then
    { If SI_data == null
      insert_into_table_value(Ti)
    Else
      update_to_table_value(Ti) }
  Else If fuzzy_data < high_weight then
    { If SI_data == null
      insert_into_table_value(Ti)
    Else
      update_to_table_value(Ti) }
  Else
    { If SI_data == null
      insert_into_table_value(Ti)
    Else
      update_to_table_value(Ti) }
end
    
```

(그림 6) 트리거 기반 의미정보 자동생성 알고리즘

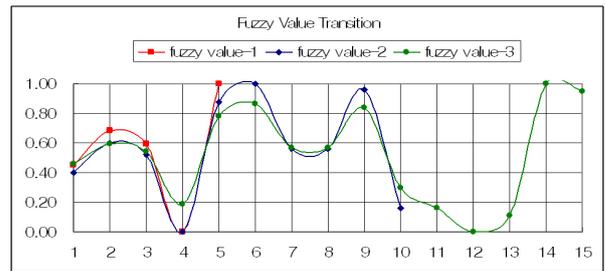
그림 6 알고리즘에서 *low_weight*, *middle_weight*, *high_weight* 는 각각 퍼지 데이터의 가중치를 의미한다. 매우 싼 가격대의 이미지는 $\{0.25 \geq weight \geq 0.0\}$ 의 가중치를 부여할 수 있으며, 보통 싼 가격대는 $\{0.5 \geq weight \geq 0.25\}$ 가중치를 부여할 수 있을 것이다. 또한 비싼 가격대의 가중치는 $\{0.75 \geq weight \geq 0.5\}$ 가중치를 부여하고, 아주 비싼 가격대의 가중치는 $\{1.0 \geq weight \geq 0.75\}$ 로 부여 한다. 본 논문에서는 가중치에 대한 적용은 특정 측정치를 사용하지 않고, 개념적인 비율(1/n)로 적용하여 *low_weight* = 0.25, *middle_weight*=0.5, *high_weight*=0.75 로 적용하여 실험하였다. 즉, T_i 는 'very cheap', 'cheap', 'expensive', 'very expensive'와 같이 4 분류로 적용하였으므로 $0.25(1/4=0.25)$ 단위로 가중치를 적용한 것이다. 표 1 은 의미정보 생성 알고리즘에 의해 생성된 결과 테이블을 보이고 있다.

<표 1> 의미 정보 생성 테이블 결과

iid	price	fuzzy value	Semantic Information
i00001	25,000	0.4000000000000000	cheap
i00002	30,000	0.5999999999999999	expensive
i00003	28,000	0.5200000000000000	expensive
i00004	15,000	0.0000000000000000	very cheap
i00005	37,000	0.8800000000000000	very expensive
i00006	40,000	1.0000000000000000	very expensive
i00007	29,000	0.5600000000000000	expensive
i00008	29,000	0.5600000000000000	expensive
i00009	39,000	0.9599999999999999	very expensive
i00010	19,000	0.1600000000000000	very cheap

4. 실험 평가

본 장에서는 논문에서 제안한 시스템 설계를 기반으로 실시간 의미정보지향 서비스 시스템을 구현하고, 의미 정보의 변화에 대하여 설명한다. 제안된 시스템은 웹 환경에서 상품 이미지를 위한 XML 문서를 사용하며, 관계형 데이터베이스에 매핑되어 XML 데이터를 저장관리 할 수 있다. 시스템 구현을 위해 MS-SQL Server 시스템을 사용하였다. 그림 7 은 초기 5 개의 가격데이터를 분석한 후 10 개 그리고 15 개의 가격정보로 변화 되었을 때, 실시간적으로 사용자에게 제공되는 상품 가격에 대한 실시간 의미정보 서비스 형태를 그래프로 표현한 예를 보이고 있다.



(그림 7) 실시간 의미정보 변화 그래프

그림 7 의 결과에서 알 수 있듯이 2 번 상품이미지의 가격정보에 대한 의미는 초기에 “very expensive”라는 의미로 서비스 되었지만, 새로운 가격정보에 의해 “expensive”라는 의미로 변화된 것을 알 수 있다.

5. 결론 및 향후과제

본 논문에서는 XML 과 퍼지기술을 적용한 웹상의 상품이미지를 실시간 의미정보 지향 서비스를 할 수 있는 시스템을 제안하였다. 실시간 의미정보 지향서비스를 가능하게 하기 위하여 퍼지 데이터 적용방안을 연구 하였으며, 상품 이미지 가격 메타데이터로부터 퍼지데이터를 자동 생성할 수 있는 알고리즘을 제안하였다. 따라서 사용자에게 좀 더 친숙하고 편리한 웹 서비스를 가능하게 할 수 있으며, 의미성을 함축한 의미 기반 검색을 가능하게 하였다. 이러한 서비스는 사용자의 의사결정을 좀더 빠르고 편리하게 할 수 있을 뿐만 아니라, 효율적인 상품 이미지 정보 검색에도 활용될 수 있다. 또한, 웹상의 멀티미디어 데이터에 대한 의미 해석을 자동화할 수 있는 시맨틱 웹의 발전에 기여할 수 있을 것이다.

참고문헌

- [1] World Wide Web Consortium, “Extensible Markup language(XML),” 1998, <http://www.w3.org/XML>.
- [2] World Wide Web Consortium, “Semantic Web,” 2001, <http://www.w3.org/2001/sw>.
- [3] Daniela Florescu and Donald Kossmann., “Storing and Querying XML Data using an RDBMS,” Data Engineering Bulletin,22(3), September 1999.
- [4] Buckley, James J. Eslami, Esfandiar “An Introduction to Fuzzy Logic and Fuzzy Sets”, MIT Press publishers, 1998.