[UCI]I804:24011-200000234083

2007年　2月
碩士學位論文

시맨틱 웹 환경에서 온톨로지 기반의 지능형 상품 검색 시스템 설계 및 구현

朝鮮大學校　大學院

電算統計學科

姜　　來　　求

시맨틱 웹 환경에서 온톨로지 기반의 지능형 상품 검색 시스템 설계 및 구현

Design and Implementation of Ontology-Based Intelligent Product Search System in Semantic Web Environment

2007年 2月 23日

朝鮮大學校 大學院

電算統計學科

姜 來 求

시맨틱 웹 환경에서 온톨로지 기반의 지능형 상품 검색 시스템 설계 및 구현

指導敎授　丁　采　榮

이 論文을 理學 碩士學位申請 論文으로 提出함.

2006年　10月

朝鮮大學校　大學院

電算統計學科

姜　來　求

姜來求의 碩士學位論文을 認准함

委 員 長　朝鮮大學校 敎授 _____ 印

委　　員　朝鮮大學校 敎授 _____ 印

委　　員　朝鮮大學校 敎授 _____ 印

2006年　11月

朝鮮大學校　大學院

목 차

목 차 ····· i
표 목 차 ····· iii
그림목차 ····· iv
Abstract ····· vi

Ⅰ. 서 론 ····· 1
A. 연구 배경 ····· 1
B. 연구 목적 및 방법 ····· 2

Ⅱ. 관련연구 ····· 4
A. 시맨틱 웹(Semantic Web) ····· 4
B. 온톨로지(Ontology) ····· 6
1. 온톨로지의 구성 요소 ····· 7
2. 온톨로지의 기능 ····· 8
3. 시맨틱 웹에서의 온톨로지의 역할 ····· 11
C. 온톨로지 언어의 개요 ····· 12
1. XML(eXtensible Markup Language) ····· 13
2. RDF(Resource Description Framework) ····· 14
3. OWL(Web Ontology Language) ····· 17

Ⅲ. 온톨로지 설계 ····· 21
A. EOE 방법론 ····· 21
1. 요구사항 분석 ····· 23
2. 온톨로지의 개념 설계 ····· 24

3. 온톨로지의 구현 ····· 26
B. 자연어 처리기능 ····· 29
C. 유의어 처리기능 ····· 31
D. 추론기능 ····· 33

Ⅳ. 시스템 구현 및 평가 ····· 35
A. 상품 검색 시스템 구현 환경 ····· 36
B. 분류별 검색 실험 및 평가 ····· 39
C. 자연어 검색 실험 및 평가 ····· 44
D. 추론기능을 이용한 상품의 확장 검색 ····· 53

Ⅴ. 결론 및 향후 과제 ····· 58

참고문헌 ····· 60

부록 *1* - 상품 도메인 온톨로지의 클래스 수목도 ····· 63
부록 *2* - 불용어 리스트 ····· 67

표 목 차

【표 2-1】 용어학과 시소러스, 온톨로지의 비교 ····· 11
【표 2-2】 RDF의 요소 ····· 16
【표 3-1】 온톨로지의 클래스 사전 일부 ····· 26
【표 3-2】 불용어 리스트 일부 ····· 29
【표 3-3】 온톨로지에 정의된 유의어 목록의 일부 ····· 31
【표 4-1】 제조사 관련 유의어 목록 일부 ····· 42
【표 4-2】 유제품 관련 하위 클래스 ····· 48
【표 4-3】 과일 관련 하위 클래스 ····· 52
【표 4-4】 제조사 도메인 온톨로지 클래스 일부 ····· 56

그 림 목 차

【그림 2-1】 시맨틱 웹의 계층 구조 ····· 4
【그림 2-2】 온톨로지 기술의 응용분야 ····· 12
【그림 2-3】 RDF, XML, 메타데이터의 관계 ····· 14
【그림 2-4】 문장의 예 ····· 16
【그림 2-5】 트리플 표현 ····· 16
【그림 2-6】 이용한 자원 기술의 예 ····· 16
【그림 2-7】 웹 온톨로지 언어의 발전 단계 ····· 18
【그림 2-8】 OWL 언어의 종류 ····· 19
【그림 3-1】 EOE 방법론을 이용한 온톨로지 구축 과정 ····· 23
【그림 3-2】 온톨로지 클래스의 개념과 관계 일부 ····· 27
【그림 3-3】 OWL로 표현한 상품 도메인 온톨로지 ····· 28
【그림 3-4】 불용어 제거를 통한 상품 검색 과정 ····· 30
【그림 3-5】 상품 도메인 온톨로지 ····· 32
【그림 3-6】 추론기능을 이용한 검색 ····· 33
【그림 4-1】 온톨로지 기반의 상품 검색 시스템 구조 ····· 35
【그림 4-2】 사용자 인터페이스(UI : User Interface) ····· 37
【그림 4-3】 사용자 인터페이스 소스 코드 일부(PowerBuilder로 개발한 부분)···· 38
【그림 4-4】 분류별 검색 방법 ····· 39
【그림 4-5】 분류별 검색 결과 1 ····· 40
【그림 4-6】 분류별 검색 결과 2 ····· 41
【그림 4-7】 분류별 검색 결과 3 ····· 43
【그림 4-8】 자연어 검색 방법 ····· 45
【그림 4-9】 자연어 검색 결과 1 ····· 46
【그림 4-10】 자연어 검색 결과 2 ····· 47
【그림 4-11】 자연어 검색 결과 3 ····· 49
【그림 4-12】 국내 온라인 할인점에서 검색한 결과 1 ····· 50

【그림 4-13】 자연어 검색 결과 4 ····· 51
【그림 4-14】 국내 온라인 할인점에서 검색한 결과 2 ····· 52
【그림 4-15】 온톨로지를 이용한 검색 결과 ····· 53
【그림 4-16】 추론기능을 이용한 유사상품 확장 검색 결과 ····· 54
【그림 4-17】 추론기능을 이용한 제조사 및 금액 확장 검색 결과 ····· 55
【그림 4-18】 XML문서로 제공된 상품의 상세 정보 ····· 57

ABSTRACT

Design and Implementation of Ontology-Based Intelligent Product Search System in Semantic Web Environment

Kang Rae-Goo
Advisor : Prof. Jung Chai-yeoung Ph.D.
Department of Computer Science & Statistics,
Graduate School of Chosun University

With development of information technologies, advances in Web technologies produced a large amount of information, made network such as Internet a critical element inevitable for daily life and brought improvement in quality of information like free information exchange over Internet.

Conventional Web has put emphasis on convenience to facilitate use of HTML so that everyone can easily access Hypertext information, resulting in a drastic increase in the volume of information.

As a result, users are now confronted with challenges in searching desired information.

Typical approaches that have been used till now to search information are usually based on keywords using frequency of words or word information. Therefore, since many documents irrelevant to user's query are included in the searched results, user has to search the represented results further to obtain a desired result.

To resolve this problem, interests have been focused on a semantic Web in which a search engine itself searches the information required for user through conceptualization, and ontology is at the center of it.

Ontology can be regarded as a kind of dictionary that defines relations between words. It explicitly expresses the conceptual architecture in a

specific domain to enable modeling of concepts closer to a real world around domain as a platform for building a search system.

To build an ontology, we collected product related documents of each manufacturer, analyzed the documents and built a product domain ontology in OWL based on classification and features frequently found in those documents. In addition, synonym list, deduction and natural language processing functions were added to the ontology to allow expansion and deduction in searching.

A database was built with the product data of about 40,000 items of Mart A operating on/off-line discount stores nationwide, and the user interface was developed using JSP and PowerBuilder9.0.

We applied two techniques for ontology-based product search. One is searching by class in which users directly select a class of desired product from the ontology classes. The other is the natural language search method in which the keyword entered by user is searched through natural language processing.

From the experiments on the two ontology-based searching methods, we could see that, unlike the conventional keyword search methods, it is possible to implement precise searching by giving meaning to the keyword itself using the concepts of classes and attributes, synonym list and deduction function defined in the ontology. Moreover, it is also feasible to search products with similar characters or those unexpected by user by semantic expansion of the searched products.

By building an ontology-based intelligent product search system with this search architecture, we could obtain searching results that is impossible with the existing keyword-based searching methods and implement more precise searching.

Ⅰ. 서 론

A. 연구 배경

1989년 Tim Berners-Lee에 의해 처음 제안된 월드와이드웹(World Wide Web)은 HTML(Hypertext Markup Language)을 이용하여 누구나 쉽게 하이퍼텍스트(Hypertext) 정보에 접근할 수 있게 편리성에 중점을 둠으로써 결과적으로 폭발적인 정보의 증가를 가져왔다.

사용자들은 이러한 정보를 이용하기 위해 브라우저에 URL(Uniform Resource Locator)주소를 직접 입력하거나 하이퍼링크(Hyperlink)를 통해 정보를 찾기도 하고 때로는 검색엔진의 도움을 받기도 한다. 이러한 단순함과 편리함이 웹의 성장을 가져온 중요한 열쇠가 되었다.

이처럼 사용하기 편리하다는 점이 웹의 주요 성장 요인이기도 하지만 반면에 많은 양의 정보가 폭발적으로 생성되는 현 상황에서 정보의 저장과 검색에 있어서 많은 문제점이 생겨나고 있다.

현재의 웹 사이트에서는 데이터베이스를 이용하여 자료를 저장하고 스키마(Schema) 구조를 통해 자료를 찾아 사용자들에게 단순하게 제공하고 있다. 이러한 시스템은 사용자가 자주 사용하는 용어와는 관계없이 웹 사이트 개발자나 데이터베이스 설계자가 일방적으로 정해 놓은 키워드를 통해서만 정보를 검색할 수 있다는 문제점을 가지고 있다. 정보를 검색하는 데 있어서 지금까지의 웹은 정적이었다고 볼 수 있다. 검색어에 의미를 부여하여 사용자가 원하는 정보를 자동으로 추출하고 그 정보의 속성을 나타낼 수 있는 지능적인 방법이 없었기 때문이다.

이러한 한계를 극복하기 위해서 1999년 W3C(World Wide Web Consortium)에서 시맨틱 웹(Semantic Web)을 제안하였다.

시맨틱 웹이란 “컴퓨터가 정보의 의미를 파악하고 조작할 수 있는 웹”으로서 사람이 원하는 정보를 컴퓨터가 자동으로 추출하도록 하는 기술을 말하며 온톨로지(Ontology)를 이용하여 지식을 공유하게 된다.

온톨로지는 넓은 의미로는 데이터베이스라고 할 수 있지만 데이터보다 복잡한 형태의 지식과 관련되어 있다는 의미에서 지식베이스(Knowledge base)라고 부르기

도 한다. 하지만 엄격히 말해 온톨로지는 지식 내용과 절차적 추론 과정을 포함하는 포괄적 의미의 지식 보다는 용어 사이의 개념적 관계에 국한되어 있기 때문에 지식베이스와 구별되는 또 다른 형태의 데이터베이스라고 할 수 있다.[1]

온톨로지에 관한 연구는 인공지능 분야의 시작과 함께 지식 표현 분야의 핵심으로 활발히 연구가 이루어져 왔다. 최근에는 시맨틱 웹의 출현과 더불어 온톨로지의 중요성이 인식되면서 이 분야는 새로운 연구 분야가 아닌 응용분야로서 연구되고 있다.

웹상의 정보가 증가함에 따라 지금까지의 검색엔진은 주로 단어의 사용 빈도수나 어휘정보를 이용하여 문서의 유사도를 측정하고 순위를 부여하기 때문에 사용자가 찾고자 하는 정보와 관계가 없는 동형이의어나 동음이의어와 같은 단어가 포함된 문서들까지도 동일한 결과로 나타냈기 때문에 이로 인해 사용자는 제시된 결과를 한번 더 검색을 해야 하는 부담을 안게 되었다. 하지만 온톨로지를 이용하여 정보를 검색할 경우엔 사용자가 찾고자 하는 정보를 보다 빠르게 찾을 수 있고 또한 검색된 결과의 정확도를 향상시킬 수 있다는 장점이 있다.[2]

B. 연구 목적 및 방법

본 논문에서는 상품 도메인 온톨로지를 구축하여 웹 검색 시스템에 활용함으로써 검색 성능을 향상시키고 원하는 상품을 보다 빠르고 정확하고 검색하는 데 목적이 있다.

온톨로지를 구축하기 위해 각 제조사의 상품 관련 문서를 수집하고 해당 문서를 분석하여 빈번하게 발생하는 상품의 특징과 분류체계를 기반으로 상품 도메인 온톨로지를 구축하였다.

예를 들어 사용자가 “서울우유에서 생산한 500원짜리 우유 중에서 과일 맛이 나는 우유는 무엇인가?”라는 검색어를 입력했을 때 기존의 검색엔진에서는 사용자가 입력한 검색 문장 중 단어의 빈도수나 어휘 정보를 키워드로 하여 검색하게 된다. 즉 위의 예에서 입력한 검색어에서 분류해 낼 수 있는 키워드로는 “서울우유”, “500원”, “과일 맛”, “우유” 정도로 분류가 가능하며 결과적으로 서울우유에서 생산한 500원짜리 우유들 중에서 “과일 맛 우유”만을 검색하게 될 것이다. 하지만 사용자가 원하는 검색 결과는 “과일 맛 우유”라는 단일 상품이 아닌 딸기, 바나나, 복숭

아 등 과일 맛이 나는 모든 우유에 대한 검색 결과를 원하고 있을 것이다.

이러한 문제점을 해결하기 위해 본 논문에서는 상품 도메인 온톨로지를 구축하여 의미기반 검색을 통해 사용자가 원하는 정보를 보다 빠르고 정확하고 검색해 낼 수 있도록 하였다.

실험에 사용한 상품 데이터는 국내 A할인점에서 사용 중에 있는 상품 데이터를 이용해 구성하였으며 본 논문에서는 온톨로지 연구와 관련된 내용을 기술하였기 때문에 데이터베이스 구축 내용은 다루지 않았다.

논문의 구성은 다음과 같다. 2장에서는 시맨틱 웹과 온톨로지에 관련된 이론 및 현재의 동향을 살펴보고, 3장에서는 온톨로지의 구축 방법을 기술하고 이를 바탕으로 직접 온톨로지를 구축하였다. 4장에서는 구축된 온톨로지를 이용한 상품 검색 시스템을 구현하고 그 결과를 평가하였다. 마지막으로 5장에서는 결론 및 향후 연구 과제를 제시하였다.

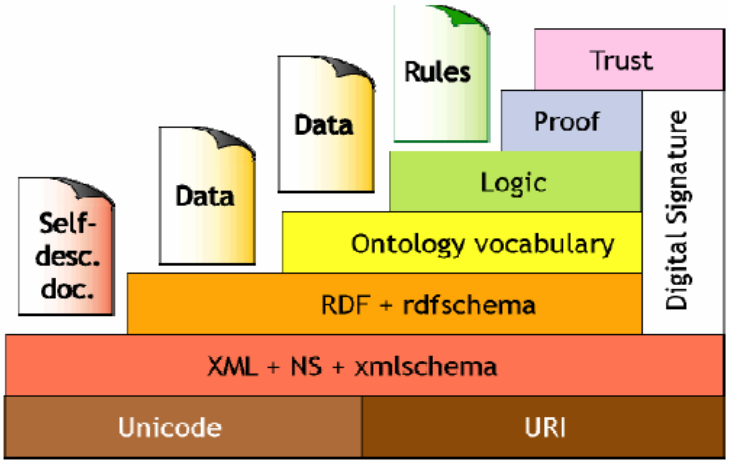
Ⅱ. 관련 연구

A. 시맨틱 웹(Semantic Web)

현재의 웹은 HTML문서를 기반으로 하여 사람들이 쉽게 읽고 이해하도록 정보의 외적인 모습만을 강조한 웹으로서 사용자가 원하는 정보를 신속하게 파싱하여 문서를 구성하고 하이퍼링크를 통해 문서와 문서를 연결해주는 매우 제한적인 기능만을 수행해 왔다.

이러한 현재 웹의 문제점을 보안하기 위해 1990년대 말에 시맨틱 웹이 Tim Berners-Lee에 의해 제안되었다. 그는 시맨틱 웹이 기존의 웹과 완전히 구별되는 새로운 웹의 개념이 아닌 현재의 웹을 확장하여 웹에 올라오는 정보에 잘 정의된 의미를 부여하고 이를 통해 컴퓨터와 사람이 협동적으로 작업을 수행할 수 있도록 하는 패러다임이라고 정의하였다.[3][4]

현재의 웹을 개선하기 위한 시맨틱 웹의 계층 구조는【그림 2-1】과 같다.



【그림 2-1】 시맨틱 웹의 계층 구조

이 계층구조에서 보면 가장 하위 레벨은 웹 프로토콜(Web Protocol)에서 자원을

지칭하기 위한 주소지정(Addressing) 방법인 URI(Uniform Resource Identifier)와 유니코드(Unicode)가 밑받침되고 이를 기반으로 XML(eXtensible Markup Language)과 Namespace, RDF(Resource Description Framework)와 RDF Schema, 그리고 온톨로지의 순서로 위치한다. 그 다음 계층인 Logic에 대해서는 인공지능의 추론연구를 바탕으로 일부 연구가 시작되었다. 또한 Proof와 Trust는 시맨틱 웹 정보의 신뢰성과 보안에 관한 내용으로서 아직 개념 정도만 대두되고 있을 뿐 차후 연구 과제로 남아있다.[5][6]

① URI(Uniform Resource Identifier) : 웹상에 존재하는 리소스들을 구분하기 위한 리소스 ID, 또는 고유이름이다. 이것은 시맨틱 웹에서 데이터를 연결해주고 정보를 추출해내는 데 있어서 필수적인 요소이다. 즉 정보라는 것은 리소스를 기술해 주는 가장 정확한 메타데이터라고 볼 수 있다.

② XML(eXtensible Markup Language) : 사용자가 임의로 태그를 정의할 수 있기 때문에 어떠한 내용을 기술하고 있는지 쉽게 알 수 있을 뿐만 아니라 자료를 주고받을 때 웹상에서 보는 정보를 수작업을 통해 데이터베이스화 시킬 필요 없이 문서를 그대로 받아서 컴퓨터가 직접 정보를 처리할 수 있다.

③ RDF(Resource Description Framework) : 웹의 리소스에 관한 메타데이터를 표현할 수 있게 해 주는 언어이다. 웹상의 모든 문서와 리소스에 관한 자료들을 컴퓨터가 처리할 수 있는 일관된 형식으로 나타낼 수 있고 이로 인해서 시맨틱 웹이 가능해진다.

④ 온톨로지(Ontology) : 시맨틱 웹의 핵심 기술들 중 활발한 연구가 수행되는 분야로서 개념을 표준 용어로 정의하고 용어들 간의 계층구조 및 연관관계를 정의함으로써 지식을 표현한다. 시맨틱 웹에서는 지식 도메인별로 다수의 분산된 표준 온톨로지를 구축하고 상호 참조에 의해 지식 기반 검색 및 서비스가 가능하도록 한다.

⑤ 논리(Logic) : 시맨틱 웹은 추론의 기능이 반드시 필요하다. 추론이란 웹상에

이미 존재하는 자료들을 논리적 법칙을 통해 새로운 결론과 지식을 유도해 내는 시맨틱 웹의 논리계층에 속하는 일종의 프로그램으로서 스스로 추론하고 결론을 내릴 줄 알아야 하며, 정보의 홍수 속에서 인간에게 가장 좋은 선택을 가능하게 해주어야 한다.

⑥ 증명(Proof) : 컴퓨터가 의외의 결론을 도출했을 때나 추론을 통해 옳은 결론을 내고 있는지 알고 싶을 때 증명이라는 기능은 필수적이다. 현재까지 논리계층과 마찬가지로 증명계층은 여러 단체에서 표준을 제정하기 위해 연구 중이다. 이 단계를 거침으로써 "정보를 스스로 이해하고 처리하는" 시맨틱 웹을 완성하게 된다.

⑦ 신뢰의 웹(Trust) : 시맨틱 웹에서는 정보의 진위 여부를 확인하기 위해 전자서명을 사용한다. 전자서명이란 일상생활에서 한 사람의 서명이 그 사람의 진위성을 대변하는 것과 같이 문서에 코드를 추가해 어떤 사람이 작성하였는지를 입증하는 기법이다. 사람과 그의 서명을 맞춰보아 진위성을 결정하는 과정은 신뢰엔진을 통해 할 수 있다. 신뢰의 웹은 사실상 시맨틱 웹이 완성되어 가면서 가장 중요한 요소 중 하나가 될 것이기 때문에 신용등급에 관한 연구가 활발히 진행되고 있다.[7]

B. 온톨로지*(Ontology)*

지식베이스 구축을 위한 하나의 방안으로 활발히 연구되고 있는 온톨로지는 정보검색은 물론 추론을 지원함으로써 지식을 수집하고 표현하는 유용한 수단으로 이용되고 있다.

온톨로지는 원래 철학 용어이나 지금은 철학 이외의 여러 다른 분야에서도 널리 사용되고 있다. 철학에서 온톨로지는 "존재에 대한 체계적 이론으로서 세상의 것들을 추상화의 방식으로 계통을 세워서 설명하는 것"으로 정의된다. 한편 인공지능 분야에서의 온톨로지는 세상의 어떤 부분을 모델링하여 표현할 수 있는 용어의 집합으로 정의되고 있다.[8]

여기서 표현할 수 있는 용어의 집합이란 용어를 클래스, 관계, 함수 등의 개체로

해석하고 그 해석을 공리로 제안하며, 제안된 용어의 의미를 인간이 읽을 수 있는 텍스트와 연결하는 것을 뜻한다. 온톨로지는 특정분야의 용어 그리고 용어의 관계 정의는 물론 용어의 조합규칙과 용어의 확장에 대한 관계도 정의한다.[9]

Rudi Studer 등은 온톨로지를 공유된 개념화의 형식적, 명시적 명세로 정의하고 있다. 여기에서 공유된다는 뜻은 서로 간의 합의된 지식을 말하며, 개념화란 세상의 어떤 현상에 대한 추상적 모델을 형식적이란 기계가 해독할 수 있음을 의미하고, 명시적 명세란 개념, 속성, 관계, 함수, 제한, 공리가 형식적으로 정의되는 것을 말한다.

온톨로지는 단순히 개념 클래스의 의미를 정의하는 것이 아니라, 개념 클래스와 관련된 관계, 개념 클래스 간의 연산 함수 그리고 지식 추론을 위한 공리규칙과 개념 인스턴스를 총체적으로 정의하여 영역의 지식체계를 형식화하는 데 목적이 있다. 즉, 온톨로지는 영역 지식 표현을 위한 형식 체계인 것이다.[10]

1. 온톨로지의 구성 요소

온톨로지에서 지식은 주로 클래스, 관계, 함수, 공리, 인스턴스의 다섯 가지 요소를 이용하여 형식화 되며 각 요소의 정의는 다음과 같다.[11]

① 클래스(Class) : 클래스는 추출된 개념에 해당되는 것으로서, 보통 분류된 개념의 조직단위를 말한다.

② 관계(Relation) : 관계는 개념들 사이에 상호관계의 유형을 표현한다. 관계 R은 n개의 집합의 곱에서 부분집합으로 정의된다.

③ 함수(Functions) : 함수는 관계의 특별한 경우로 관계가 값을 가질 때 성립한다.

④ 공리(Axom) : 공리는 논리의 전개나 추론의 기저로 무조건 '참'이라고 인정되는 문장이다. 공리는 온톨로지 요소의 의미정의, 속성 값의 제한규정, 관계의 논한 등 여러 목적으로 활용되며, 온톨로지에서 정의한 정보의 정확성을 검증

하거나 새로운 정보를 추론하는 데 이용된다.

⑤ 인스턴스(Instance) : 인스턴스는 특정 요소의 표현, 즉 요소의 실례에 해당한다.

2. 온톨로지의 기능

온톨로지는 보통 재사용을 목적으로 만들어지며, 이러한 온톨로지를 효과적으로 활용하고 응용하기 위해서는 먼저 이에 대한 체계적이고 기능적인 고찰이 필요하다. 온톨로지의 구체적인 활용법은 프레임웍으로 구조화하여 사용하는 것이고 활용을 위한 기능적인 측면의 역할은 다음과 같다.

a. 커뮤니케이션

온톨로지는 개념적으로나 용어적으로 혼돈을 줄 수 있는 부분에 개념모델링의 방법을 이용하여 단일화된 프레임웍을 제공함으로써 요구와 견해가 다른 사람들 사이에 공유된 이해와 커뮤니케이션을 가능하게 한다.[11]

b. 상호 운용성

지금까지 서로 다른 소프트웨어를 통합하는 방법으로 API와 같은 부분적인 해결책은 있었지만 상호운용이 가능한 도구는 없었다. 온톨로지는 시스템 통합에 있어서 상호 운용성을 제공하는 도구가 될 수 있다. 상호운용성은 차원에 따라 내적 상호운용성과 외적 상호운용성 그리고 특정 영역의 개념을 통합하는 온톨로지와 도구들을 통합하는 온톨로지가 있다.[11]

c. 표준명세

명세 표시로서 온톨로지는 대상 분야를 모델링하고 목적에 따른 요구사항을 규격화하는 것이다. 온톨로지는 명세적인 의미 표현이 풍부할 수록 요구 사항을 만드는 데 있어서 모호성을 줄여준다. 온톨로지는 시스템에 대한 선언적인 명세와 추론을 제공하기 때문에 지식베이스 시스템 구축에 광범위하게 이용된다. 명세로서 온톨로지의 장점은 문서화, 유지, 신뢰성, 지식의 재사용에 있다.[12]

d. 지식관리

지식관리의 목적은 생산성을 향상시키고 새로운 가치를 창출하며 경쟁력의 향상을 위해 조직의 지적 재산을 개발하는 데 있다. 지식관리의 구성은 동일한 분야나 여러 분야에 종사하는 전문가들이 협력과 커뮤니케이션을 통하여 새로운 지식을 창조하고 통합하며 보급하는 과정으로 이루어진다.

지식관리에 대한 연구와 시스템 개발은 1990년 이후 지식공학의 발전과 함께 상당한 진전을 이루었다. 이전의 지식관리는 데이터베이스와 인터넷을 통해 이루어졌고 지식이 저장되는 곳은 간단한 테이블과 같이 전형적인 데이터베이스 형태이거나 반구조화된 텍스트였다. 따라서 정교한 지식표현이 어려웠고 지식을 포착하기 위한 체계적인 절차도 갖추어져 있지 않았다.

그러나 최근에는 지식공학의 발달로 지식유도와 모델링, 지식표현 그리고 자동추론을 구현하는 기술들을 적용하여 보다 유연하게 지식을 관리할 수 있게 되었다. 특히 온톨로지의 개념구조는 의미 웹에서 기계처리가 가능한 데이터에 핵심적 역할을 수행하여 지식베이스의 일관성을 향상시킬 수 있는 하나의 방안으로 이용되고 있다. 지식의 창조, 통합 및 보급 과정에서 온톨로지의 역할은 다음과 같이 요약할 수 있다.[13]

① 온톨로지를 이용한 의미적 내용은 지식의 식별과 수집 및 개발에 있어서 전문가는 물론 일반인에게도 지식의 포착과 창조의 수단이 될 수 있다.

② 온톨로지는 지식 통합에 있어서 개념적 일반화와 표현적 형식화를 제공한다. 여기서 개념적 일반화란 정보를 동일한 의미적 맥락에서 다른 정보와 연결시키는 것이고, 표현적 형식화란 정보를 접근하고 해석할 수 있는 컴퓨터 메커니즘에 접속시키는 것을 의미한다.

③ 온톨로지를 이용하여 개념화한 형식적 구조는 지식 보급에 있어서 표준화되고 집중적인 접근 체계를 갖출 수 있어 문제 해결을 위한 실시간의 학습 환경을 제공할 수 있다.

e. 정보검색

이용자 요구정보의 식별과 정보검색을 효율적으로 수행하기 위해서는 정보시스템의 주제가 되는 용어를 명확하게 개념적으로 구조화하여, 용어의 차이에 관계없이 분산된 환경에서도 정확한 정보에 접근할 수 있어야 한다. 그러나 기존의 키워드 기반 정보검색시스템은 다음과 같은 문제점을 가지고 있다.[14]

① 동형이의어나 동음이의어 같은 경우 키워드 기반의 검색에서는 용어의 의미 차이로 인해 정확한 정보 검색이 어렵다.

② 정확한 정보를 추출하기 위해서는 인간의 개입이 요구되고 있고, 시스템에 공통의 개념이 공유되지 않아 분산되어 있는 자원에서 정보를 통합하기가 어렵다.

③ 구조화되지 않은 텍스트 자원은 지속적인 유지와 관리가 어렵고, 자원의 규모가 커질수록 정확한 정보를 찾는 데 있어서 시간적 소모가 많이 발생한다.

④ 다양하고 광범위한 웹상의 정보에 동적으로 대처할 수 있는 의미적 기계표현이 요구된다.

현재 웹에서 사용되고 있는 대부분의 검색 방법이 키워드를 기반으로 하는 검색시스템이고 사용되는 시소러스 또한 그 구조가 유연하지 않으며 언어적 변화를 수용하는데 있어서 많은 어려움이 있다.[15]

반면 온톨로지는 언어의 의미적 개념체계를 가지는 용어학과 결합하여 인간과 컴퓨터의 의사소통에 보다 효과적이고 유연한 자원을 제공하며 해당분야에서 시소러스보다 심층적인 지식체계를 구현할 수 있다.

용어학과 시소러스, 온톨로지의 차이점은【표 2-1】과 같다.[16]

	용어학	시소러스	온톨로지
정의의 형식성	* 자연어로 개념어의 의미를 정의 * 이용자 중심의 의도된 의미 해석 * 속성 정의 부재	* 비형식적 용어 정의 * 속성정의 부재	* 형식언어로 개념적 지식 규정 * 개념의 명확한 해석 * 형식적 속성 정의
컴퓨터 적용성	* 기계적 의미 표현 빈약	* 기계적 의미 표현 빈약	* 데이터의 유지와 지식 표현의 기계적 의미 표현이 풍부
대 상	* 번역자, 주제전문가 등의 이용자	* 색인과 검색을 위한 자	* 지식공유 목적의 인간과 컴퓨터
기 본 엔트리	* 개념	* 용어	* 개념
관 계	* 개념 용어의 다양한 어의적 관계	* 용어의 계층, 대등, 연관관계	* is-a, part-of 및 기계와 인간 사이에 공유가 가능한 명확한 개념 관계

【표 2-1】용어학과 시소러스, 온톨로지의 비교

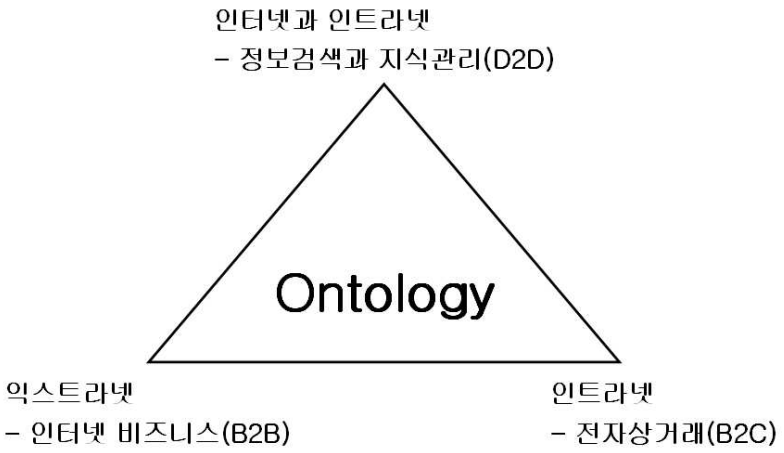
【표 2-1】에서 보는 바와 같이 시소러스는 일반적으로 용어의 관련구조를 형성하는 반면 온톨로지는 개념구조 형성을 목적으로 한다. 또한 용어학과 시소러스는 온톨로지에 비해 정의의 형식성과 기계의 적합성이 떨어지기 때문에 그 모호한 의미 구조로 인하여 검색 시스템에서도 효율성을 기대하기 어렵다.[17]

3. 시맨틱 웹에서의 온톨로지의 역할

온톨로지에 대한 연구는 인공지능 지식표현 기술과 관련된 중요한 분야로서 시맨틱 웹과는 독립적으로 발전해 왔다. 이와는 대조적으로 시맨틱 웹은 하나의 독립된 기술이라기보다는 RDF와 같은 메타데이터 언어를 비롯해 여러 분야의 기술들이 함께 발전한 분야로서 최근에는 온톨로지 기술이 시맨틱 웹을 구현하기 위한

중요한 요소가 되었다. 따라서 시맨틱 웹을 기반으로 한 온톨로지에 대한 연구는 더욱 활기를 띠게 되었으며 응용 범위도 갈 수록 확산되고 있다. 온톨로지를 넓은 의미에서는 데이터베이스로 볼 수 있지만 절차적 지식에 대한 표현보다는 어떤 영역에서 사용되는 어휘들의 집합을 개념적으로 표현하는 방법으로 봐야한다.

시맨틱 웹의 요소 기술로서 온톨로지가 응용되는 분야는【그림 2-2】에서 보여주듯이 정보검색과 지식관리(D2D), 전자상거래(B2C), 인터넷 비즈니스(B2B)등 세 분야로 크게 구분할 수 있다.



【그림 2-2】 온톨로지 기술의 응용분야

시맨틱 웹의 관점에서 볼 때 사용자가 원하지 않는 데이터를 미리 걸러주며, 관련성이 높은 정보의 발견 가능성을 높여주기 위해 온톨로지를 이용하여 검색 단어에 의미를 부여할 필요가 있다. 하지만 현실적으로 모든 분야에 사용되는 개념들을 수집하여 이들 간의 관계를 표현한다는 것은 사실상 불가능하다. 따라서 모든 분야가 아닌 특정 분야나 특정 영역만을 위한 개별적인 온톨로지 개발이 추진되고 있다.[18]

C. 온톨로지 언어의 개요

시맨틱 웹에서는 서로 다른 사전 체계를 가진 문서들 간에도 변환이 쉽게 이루

어지도록 개념 간의 계층관계나 개념 정의 간의 정합성 등을 자동으로 계산할 수 있는 온톨로지 언어를 제공하고 있다.

시맨틱 웹 환경에서 효과적으로 지식을 표현하기 위해 W3C에서는 다양한 언어를 개발하여 사용하고 있으며, 2004년 2월에는 시맨틱 웹을 기술하기 위한 중요한 언어들 중 하나인 개정된 자원 기술 프레임워크(RDF)와 웹 온톨로지 언어(OWL)를 시맨틱 웹 기술 언어로 인정한다고 발표했다.

1. XML(eXtensible Markup Language)

XML(eXtensible Markup Language)은 SGML의 하위 집합으로 구성된 마크업 언어로서 시맨틱 웹과는 별개로 HTML의 비구조성을 극복하기 위해 제안된 표준이다. HTML에 비해 XML은 구조화된 문서를 작성할 수 있는 방법을 제공하고 있다. 시맨틱 웹과 관련되어 XML문법을 사용하는 이유는 구조화된 문서의 생성을 유도한다는 점과 태그 이름을 사용자가 임의로 정의할 수 있기 때문에 의미정보를 손쉽게 태그에 적용할 수 있기 때문이다. 단, 시작 태그와 종료 태그가 쌍으로 이루어져야 한다는 것과 중첩구조가 지켜져야 한다는 제약 조건이 반드시 만족되어야만 한다.[19][20]

이러한 구조화된 문서 생성과 손쉬운 태그 정의로 인해 시맨틱 웹은 XML을 기반으로 하고 있는 것이다. 즉 다양한 정보를 일정한 형식으로 정의하고, 표준화된 방식으로 상호 교환할 수 있는 기술 기반을 제공하여 컴퓨터가 스스로 처리할 수 있도록 함으로써 시맨틱 웹의 구현에서 있어서 대단히 중요한 기반이 되고 있다.

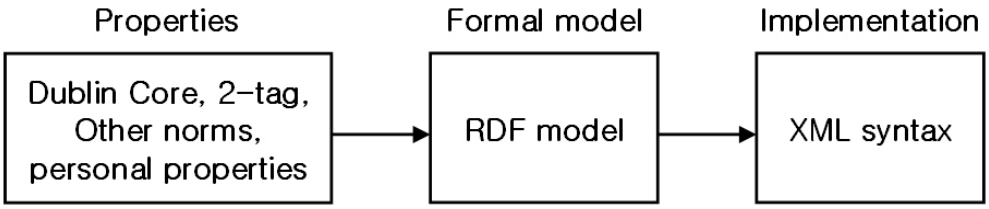
그러나 아직까지 XML은 시맨틱 웹에서 요구하는 수준의 데이터와 의미, 객체 형태의 데이터 구조, 데이터 구조 사이의 관계, 그리고 추론을 위한 규칙을 표현할 수 있는 직접적인 기술 방법 등이 없는 상태이다. 태그의 사용에 대한 규칙은 DTD나 XML Schema에 의해 정의될 수 있으나 이들 사이의 관계를 컴퓨터에게 알려줄 방법이 없다. 이는 XML이 문서 구조를 지정할 수는 있으나 문서 구조가 갖는 의미에 대해서는 아무 것도 정의하지 못하기 때문이다. 이러한 단점을 극복하기 위해 XML을 기반으로 하여 객체 형태의 데이터 구조 정의 및 데이터 구조간의 관계 표현이 가능한 새로운 언어인 RDF가 W3C에서 발표되었다.[21]

2. RDF(Resource Description Framework)

RDF(Resource Description Framework)는 XML의 문제점을 해결하고 의미에 초점을 맞추기 위해 제시된 기반구조로서 W3C에서 발표한 메타데이터를 기술하는 포맷으로 웹의 리소스에 관한 메타데이터의 기술과 교환을 위한 표준으로 사용되고 있다. RDF는 다양한 메타데이터간의 상호운용을 위해 의미(Semantic), 구조(Structure) 및 구문(Syntax)에 대한 공통적인 규칙을 제공한다. RDF의 출현은 사람들에게 많은 의문을 던져주고 있다.

RDF가 왜 필요하고, 어떻게 사용될 수 있을까? 이에 대한 답은 상호운용이란 말로부터 실마리를 풀 수 있다. 굳이 시맨틱 웹이나 온톨로지 분야에 국한시키지 않아도 이 말 자체만으로도 매우 폭넓게 사용되고 있다. 상호운용은 서로 다른 객체 사이에서 특별한 노력 없이도 동작할 수 있는 것을 말한다. IT환경에서 상호운용에 대한 관심의 원인은 월드와이드웹에서 찾을 수 있다. 월드와이드웹의 폭발적인 성장은 데이터의 양적 증가도 함께 불러 일으켰다. 이러한 환경에서 사람들은 끊임없이 데이터를 공유하기 위해 힘써 왔고 그 노력의 일환이 메타데이터이다. 메타데이터는 자원의 내용을 일정한 규칙에 따라 기술할 수 있는 틀이 될 수 있다. 그러나 목적에 따라 다른 메타데이터 구문과 스키마 정의 언어가 달라 이 또한 효과를 장담할 수 없었다. RDF는 이러한 문제를 해결하는 데 주된 목적을 갖고 있다. RDF는 데이터 모형을 이용하여 메타데이터 스키마를 일관된 방법으로 표현한다. 다시 말해, 다양한 특성을 갖는 메타데이터 스키마가 있을 때 RDF는 이것들을 표현할 수 있는 추상적 모형을 제공하고, 컴퓨터가 스스로 처리 가능하도록 XML을 이용하여 구현한다.[22][23][24]

【그림 2-3】은 RDF, XML 그리고 메타데이터와의 관계를 나타내고 있다.



【그림 2-3】RDF, XML, 메타데이터의 관계

RDF는 자원을 기술하기 위한 모형을 제공하며 RDF 데이터 모형은 기술하는 자원의 속성과 자원 간의 관계를 표현한다. 데이터 모형에 기술되는 내용들은 문법적으로는 서로 상이할 수 있지만, RDF 파서(parser)에 의해 동일한 의미로 처리되기 때문에 문법에 중립적이다. RDF 데이터 모형은 크게 자원(Resource), 속성타입(Property type), 진술문(Statement)으로 구성된다.[23]

① 자원(Resource) : URI(Universal Resource Identifier)로 기술되는 모든 객체를 말한다. URI는 어떤 객체를 인식하기 위한 수단으로 그 대상은 웹 자원에만 한정되지 않는다. URI는 특정한 웹 페이지가 될 수도 있지만, 오프라인에 있는 책상이 될 수도 있다. 우리가 알고 있는 URL(Uniform Resource Location)은 URI의 부분집합이라 할 수 있다. 다시 말해, 자원의 범위는 웹 자원(URL로 표현되는)에만 한정되지 않는다. 실세계에 존재하는 모든 객체를 자원으로 표현할 수 있고 하나의 자원은 여러 개의 속성 유형과 속성 값을 포함할 수도 있다. 예컨대 “http://www.ontology.or.kr/default.html” 은 URL로 표현된 자원이라 할 수 있다.

② 속성타입(Property type) : 자원의 속성명을 의미한다. 속성명은 “저자명” 또는 “파일형식” 등과 같이 특정 자원의 속성에 대해 적절한 이름을 부여하는 것이다. RDF에서의 기술은 동일한 자원을 참조하는 속성들의 집합을 말한다. 속성 값은 다른 자원 혹은 문자열이나 숫자와 같은 형태일 수도 있다.

③ 진술문(Statement) : 특정 자원을 기술하기 위해 지정된 속성과 그 속성 값을 의미한다. 이 진술문은 주어(Subject), 서술어(Predicate), 목적어(Object)로 구성된다. 이 때 진술문의 목적어는 어떤 자원이나 리터럴이 되기도 한다.

RDF 데이터 모형은 트리플(triples), 그래프, XML을 이용하여 표현할 수 있다. 이들 표현 방식은 서로 대등한 의미를 갖는다.【그림 2-4】는 RDF를 소개할 때 사용되는 가장 기본적인 문장이다.[24]

Ora Lassila는 자원 http://www.w3.org/Home/Lassila의 작성자이다.

【그림 2-4】 문장의 예

위 진술문은 【표 2-2】 와 같은 요소로 구성이 된다.

주어(Subject)	http://www.w3.org/Home/Lassila
서술어(Predicate)	작성자
목적어(Object)	"Ora Lassila"

【표 2-2】 RDF의 요소

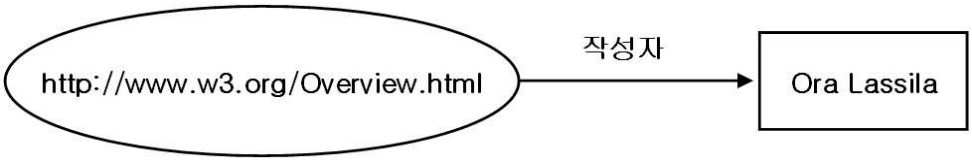
또한 트리플을 이용하여 【그림 2-5】 와 같이 표현할 수 있다.

{ 작성자, [http://www.w3.org/Home/Lassila], “Ora Lassila” }

【그림 2-5】 트리플 표현

트리플 표현은 RDF나 OWL언어가 XML 형식으로 읽힐 수 있도록 변환시키는 작업을 말하는 것으로서 여기서 기호 [/]는 URI에 의해 식별되는 자원을 의미하고, 인용부호는 리터럴을 의미한다.

RDF는 【그림 2-6】 과 같이 방향 그래프를 기반으로 자원을 기술할 수 있다.



【그림 2-6】 이용한 자원 기술의 예

그래프는 두 개의 유한 집합 즉, 정점(vertex)들의 집합과 간선(edge)들의 집합으로 이루어진 구조를 말한다. 예컨대, 간선 e_i는 정점 v_j로부터 정점 v_k로의 간선이

된다.

$$e_i=(v_j,\ v_k)$$

이 때, v_j를 간선의 시작점(initial vertex), 정점 v_k를 간선의 종착점(terminal vertex)이라고 부른다. 이와 같은 구조는 간선에 방향을 지정하기 때문에 이를 방향 그래프라고 한다. 또 그래프를 구성하는 요소인 정점과 간선에 라벨(label)을 붙일 수 있는데, 이러한 그래프를 라벨 그래프라 한다. 그래프는 다이어그램으로 편리하게 도식화 할 수 있다. 여기서 정점은 원으로, 리터럴은 사각형으로, 그리고 간선은 두 정점들을 잇는 화살표로 표현된다.

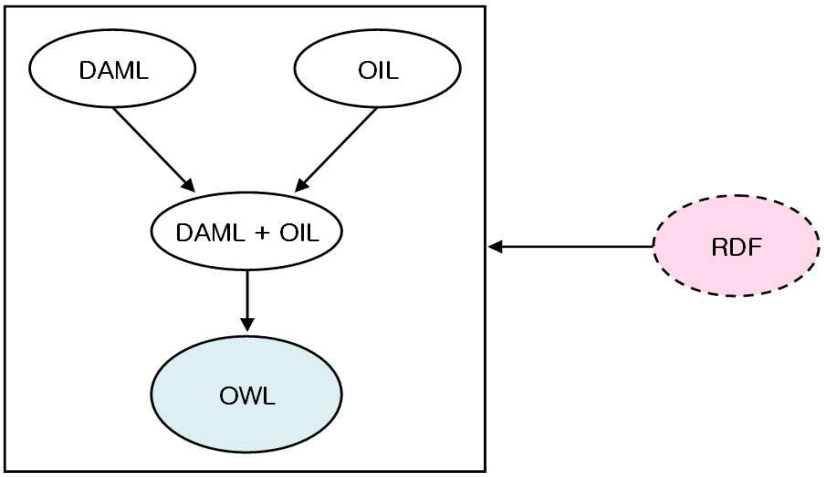
주의할 점은 화살표의 방향이다. 그래프에서 화살표의 방향은 항상 주어에서 시작하여 목적어를 지시해야 한다. 따라서 【그림 2-6】은 다음과 같이 해석할 수 있다.[25][26]

◉ Ora Lassila는 자원 http://www.w3.org/Home/Lassila의 값이다.

◉ 자원 http://www.w3.org/Home/Lassila는 Ora Lassila와 같은 값을 가진 속성 작성자를 갖고 있다.

3. OWL(Web Ontology Language)

OWL은 문서에 포함된 정보를 어플리케이션을 이용하여 자동으로 처리하고자 할 때 사용하는 언어이다. OWL을 이용하면 임의의 어휘를 구성하는 용어의 의미와 용어들 간의 관계를 명시적으로 표현할 수 있다. 이처럼 용어와 용어들 간의 관계를 표현한 것을 온톨로지(Ontology)라 한다. OWL은 XML, RDF, RDF Schema보다 더 많은 의미 표현 수단을 제공하므로, 웹상에서 컴퓨터가 스스로 해석할 수 있는 컨텐츠를 작성하는 데 있어서 다른 어떤 언어들보다 뛰어난 언어로서, DAML+OIL로부터 파생된 언어이다.[27]



【그림 2-7】 웹 온톨로지 언어의 발전 단계

OWL은 표현력이 서로 다른 세 개의 하위 언어로 구성된다.

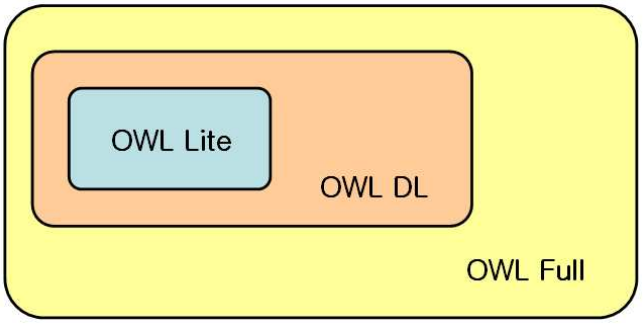
① OWL Lite는 클래스의 계층과 간단한 제약 사항만을 제공하고 있다. 따라서 다른 하위 언어보다 표현력의 제한이 많다. 예컨대 관계차수 제약이 가능하지만, 차수의 값은 0 또는 1만을 사용할 수 있다. OWL Lite는 이론적 복잡도가 낮기 때문에 다른 분류체계나 시소러스를 빠르게 OWL로 변환하는 데 적합하다.

② OWL DL은 기술논리(Description Logic)를 기반으로 한 언어이다. 기술논리는 OWL의 형식적 기반이 된 논리학의 한 분야로서 계산적 완전성과 결정 가능성을 유지하면서 최대의 표현력을 제공한다. 계산적 완전성은 모든 결론이 계속 될 수 있다는 것이고 결정 가능성은 모든 계산이 유한한 시간 내에 끝난다는 것을 말한다. OWL DL은 OWL에서 정의한 모든 어휘를 포함하고 있다. 그러나 어휘를 사용할 때는 정해진 제약 사항에 따라야 한다. 예컨대 클래스는 다른 클래스의 하위 클래스로 선언할 수 있지만, 다른 클래스의 인스턴스로 선언할 수는 없다.

③ OWL Full은 RDF의 모든 문법을 사용할 수 있으며 최대의 표현력을 제공한

다. 단, 계산적 보장은 하지 않는다. OWL Full은 DL과 Lite의 모든 기능을 포함하는 관계이며 유효성과 호환성은 하위 언어 중 가장 완벽하다. 그러나 OWL Full을 지원하는 추론엔진을 개발하는 것은 쉽지 않다.

OWL 하위 언어들은【그림 2-8】과 같은 포함 관계를 갖는다. 즉, OWL DL은 OWL Lite의 확장이고, OWL Full은 OWL DL의 확장 언어이다.



【그림 2-8】OWL 언어의 종류

각 하위 언어 사이에는 다음과 같은 관계가 성립한다. 반면 그 역은 성립하지 않는다.

◉ 모두 올바른 OWL Lite의 온톨로지는 올바른 OWL DL의 온톨로지이다.
◉ 모두 올바른 OWL DL의 온톨로지는 올바른 OWL Full의 온톨로지이다.
◉ 모두 타당한 OWL Lite의 결론은 타당한 OWL DL의 결론이다.
◉ 모두 타당한 OWL DL의 결론은 타당한 OWL Full의 결론이다.

OWL을 이용하려면 먼저 어떤 OWL 하위 언어가 주어진 요구 사항에 최적인지 결정해야 한다.

OWL Lite와 OWL DL 중 어떤 언어를 사용할지 판단하는 기준은 OWL Lite보다 풍부한 OWL DL의 표현력을 필요로 하는가의 여부이다.

OWL DL과 OWL Full 사이의 판단 기준은 RDF Schema가 제공하는 메타 모델 구축 기능을 필요로 하는가의 여부이다.

OWL Full을 이용하면 추론 기능 지원을 보장 받을 수 없다. 왜냐하면 아직까지는 OWL Full을 완벽하게 지원하는 추론 시스템은 구현된 바가 없기 때문이다.[25]

Ⅲ. 온톨로지 설계

지금까지의 웹은 폭발적인 정보량의 증가로 인해 원하는 문서를 찾기 위해서는 검색 결과에 대한 여러 차례 검증 절차를 거쳐야만 하는 불편함을 사용자에게 가져다주었다. 대부분의 인터넷 쇼핑몰은 상품 검색시 주로 키워드를 기반으로 검색을 함으로써 사용자가 입력한 검색어가 데이터베이스의 상품 목록과 일치해야만 정확한 검색이 이루어지며 또한 유사한 상품을 검색함에 있어서 키워드가 서로 다를 경우엔 전혀 검색할 수 없는 단점이 있었다. 이와 같은 단순한 일차원적인 검색은 고객이 찾고자 하는 상품의 대한 정확한 검색은 물론이며 구매 조건에 맞는 상품을 추천하거나 안내하는 데도 한계가 있다. 또한 찾고자 하는 상품에 대한 정보를 고객 스스로 판단하여 키워드로 사용해야 하는 어려움도 있다. 이는 기존 웹에서 주로 사용되고 있는 HTML이 의미 정보를 제공하지 못하기 때문이다.[28]

본 논문에서는 웹상에서 보다 정확한 상품 검색을 위해 시맨틱 웹 환경에서 온톨로지를 이용하여 사용자가 입력한 질의를 의미적인 해석을 통해 원하는 결과를 보다 쉽고 정확하게 검색할 수 있도록 하였다.

온톨로지 구축을 위해 현재 온·오프라인 쇼핑몰에서 유통 중인 생필품(공산품) 데이터의 상품명과 분류, 각 상품의 특징 등을 분석하여 구축하였다. 검색 방법으로는 사용자가 상품에 대한 속성과 특징을 직접 선택하여 원하는 상품을 검색할 수 있는 분류별 검색 방법과 찾고자 하는 상품에 대한 기본적인 정보가 없이도 원하는 상품을 검색할 수 있도록 자연어 처리 기능을 추가한 자연어 검색 방법이 있다.

A. EOE 방법론

온톨로지를 개발하는 데 있어서 특별히 정해진 규칙이 있는 것은 아니다. 온톨로지의 개발은 도메인 온톨로지의 모델링과 그 온톨로지를 이용하여 검색에 활용하게 될 어플리케이션의 확장성에 달려 있다.

먼저, 도메인의 특성과 온톨로지의 사용 용도를 고려하여 온톨로지의 구성영역을 정하고 온톨로지가 응답해야 할 질의들의 대하여 온톨로지 내에 포함되어야 할 세

부사항들을 정리한다.

온톨로지는 어떤 특정 도메인에서 사용되는 정보들과 그 정보들 간의 관계를 정의해 놓은 것으로서 구축하고자 하는 분야의 전문가와 상의를 통해 개념과 관계의 구조를 정한 뒤 이들을 기반으로 구축하게 된다. 실제의 응용시스템에서는 도메인마다 특징적인 지식을 포함하는 온톨로지가 필요하다. 왜냐하면 유통, 경제, 의학, 공학 등과 같이 각 분야마다 사용되는 개념이 각각 다르기 때문이다.

구축할 온톨로지의 초기 안을 작성한 후에는 반복적인 과정을 거쳐 평가하고 그 분야의 전문가와 상의를 통해 세부적인 사항들을 세워 가며 지속적으로 수정하고 보완해가는 것이다. 그 과정 중에 온톨로지 설계자 및 개발자는 구축하려는 온톨로지에 대한 장단점과 구축 중에 발생할 수 있는 문제점에 대한 대비책을 강구하고 고려하여 모델링을 해나가야 한다.[29]

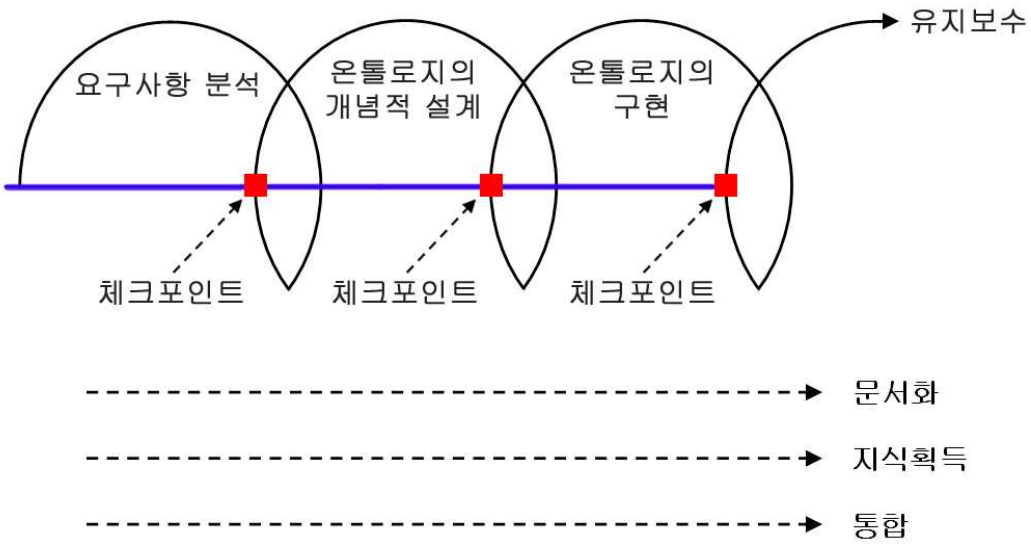
본 논문에서 구축 할 도메인 온톨로지는 국내 온・오프라인 할인점에서 유통 중인 공산품 및 생필품 약 40,000여 종을 대상으로 하였다. 즉, 국내에서 유통 중인 공산품 및 생필품 관련 도메인 온톨로지를 구축함으로써 소비자로 하여금 상품 검색 시 상품에 대한 사전 지식이 없더라도 보다 정확한 상품 검색이 가능하도록 하기 위한 목적이 있다.

본 논문에서는 온톨로지를 구축하기 위한 여러 가지 방법론 중 EOE(Evolving Ontology Engineering) 방법론을 사용하였다. EOE 방법론은 진화형 프로토타입을 바탕으로 하고 있다.

진화형 프로토타입은 시간 경과에 따라 요구사항이 분명해지면서 온톨로지를 점진적으로 구축할 수 있는 견고한 아키텍처를 제공하며 일회성 프로토타입 보다 구축하는 데 있어서는 더 많은 시간이 소요되지만, 신뢰성 있고 쉽게 확장할 수 있기 때문에 온톨로지를 구축하는 데 많이 사용되는 방법 중 하나이다. 또한 요구사항의 불확실성을 제거하기 위해 문서 프로토타입을 사용한다. 이것은 모든 요구사항을 문서화함으로써 사용자와 온톨로지 개발자가 요구 사항에 대해 공감하고 있는지를 테스트할 수 있어 최종 산출물(온톨로지)의 위험을 낮출 수 있다. 문서 프로토타입은 상용화 도구를 이용하여 전자 프로토타입으로도 만들 수 있다.[25]

즉, EOE 방법론은 프로토타입 기법을 기반으로 하여 온톨로지를 구축하기 때문에 온톨로지 개발자와 사용자 사이의 상호 이해와 지식 교환을 보다 수월하게 할 수 있다.

【그림 3-1】은 EOE 방법론을 이용해 본 논문에서 구축하고자 하는 온톨로지의 구축 과정을 나타내고 있다.



【그림 3-1】EOE 방법론을 이용한 온톨로지 구축 과정

설계자는 온톨로지를 구축하고자 하는 분야의 전문가와의 상의를 통해 온톨로지 실행 계획을 먼저 작성한다. 요구사항 분석, 온톨로지의 개념적 설계, 온톨로지의 구현 단계는 온톨로지를 실제 개발하기 위한 핵심적인 절차이다. 온톨로지를 구축하는 전 생명주기 동안 문서화와 지식 획득 작업은 지속적으로 수행해야 하며 이때 EOE 방법론은 각 단계마다 순환적인 프로세스를 진행할 뿐만 아니라 각 단계가 마무리되기 전에 점검을 통해 미흡한 부분에 대해 수정을 해야 한다.[25]

1. 요구사항 분석

요구사항 분석 단계는 온톨로지의 개발 목적이 무엇인가를 구체화하는 단계이다. 또한 온톨로지가 어떻게 구성이 되어야 하는지, 온톨로지의 특징은 무엇인지를 명확하게 정의해야 한다. 이 단계는 온톨로지 개발에 대한 전체적인 계획을 수립하고 정의하기 때문에 정확성이 무엇보다 강조된다.

a. 요구사항 유도

이 단계에서는 의사소통이 주된 도구로 활용이 되기 때문에 에러 발생이 특히 높다. 따라서 요구 사항을 정확히 도출하기 위해 고객, 도메인 전문가, 개발자 사이에 긴밀한 협력 관계가 이루어져야 한다.

b. 요구사항 분석

요구사항 분석의 목적은 현실적으로 수행 가능한 프로젝트인지 예측하고, 이를 통해 설계, 구성, 테스트 작업을 진행할 수 있도록 하는 데 있다.

c. 요구사항 명세

분석된 요구사항은 체계적이고 쉽게 사용, 또는 검토할 수 있는 방식으로 문서화가 이루어져야 한다.

d. 요구사항 검증

검증단계는 요구사항 설명이 적절한지를 검토하고 바람직한 품질 특징을 알려주는 단계로서 고객의 요구를 충족시키기 위해 반드시 필요하다.

2. 온톨로지의 개념 설계

개념적 설계는 구축하고자 하는 온톨로지에 필요한 클래스를 결정하는 것으로 시작한다. 하지만 클래스를 결정하고 설계한다는 것은 매우 어렵고 불명확하다. 이 과정은 온톨로지 개발자와 관련 분야 전문가의 경험과 직관에 강하게 의존하게 된다.[25]

a. 용어 사전 작성

본 논문에서는 클래스를 작성하기 위해 명사 분석 기법을 사용했다. 이 분석 방법은 온톨로지 개발에 사용된 모든 문서에서 명사 혹은 명사절을 수집한 후 적절한 클래스 집합을 만들어내는 것이다. 이 분석 방법에는 특별한 도구가 필요하지는 않고 간단한 스프레드시트를 이상적인 도구로 사용할 수 있다. 명사 분석 과정은 다음과 같은 절차에 의해 수행 된다.

◉ 온톨리지 구축에 사용된 모든 문서를 수집한다.
◉ 도메인에 사용되는 용어들을 수집한다.
◉ 수집된 문서에서 클래스로 이용 가능한 명사/명사구를 추출하여 스프레드시트에 입력한다.
◉ 명사/명사구 추출 시 중복이나 동의어를 고려하지 않고 작업한다.
◉ 스프레드시트에 입력된 명사/명사구를 정렬하여 중복된 내용을 삭제한다.
◉ 동의어는 사용자가 상품 검색 시 추론 기능을 이용할 수 있도록 별도로 보관한다.

b. 클래스 이름 정의 및 클래스 계층 구조 정의

클래스의 이름을 정의하는 것은 온톨로지를 구축하는 데 있어서 매우 중요하다. 만약 어떠한 클래스의 이름이 정확하지 않다면 그에 따른 속성이나 다른 클래스와의 관계를 명확하게 정의하는 것은 불가능할 것이다. 일반적으로 클래스의 이름을 정의하고 계층 구조를 정의하는 데 있어서 다음과 같이 몇 가지 고려할 규칙이 있다.

◉ 클래스 이름을 명확하게 정의한다.
◉ 클래스 이름에 대한 다국어 표현을 고려한다.
◉ 유사어는 통합하여 하나의 클래스로 정의한다.
◉ 단수형을 복수형의 IS-A 관계로 정의하지 않는다.
◉ 형제 관계는 같은 단계로 정의한다.
◉ 하나의 클래스에 서브 클래스가 12개 이상일 경우엔 새로운 부모 클래스를 만든다.
◉ 다중 상속을 고려한다.
◉ 복잡한 관계에 있는 클래스들은 새로운 클래스로 표현하고, 이 클래스와 기존 클래스의 관계를 재정의 한다.

클래스의 이름을 정의하는 데 있어서는 어떠한 집합이 클래스인지 아닌지를 판단할 수 있어야 한다. 예를 들어, “A 자동차”를 클래스로 정의했다고 가정했을 때 이 클래스의 속성은 자동차 명, 배기량, 차량번호 등 두 개 이상을 속성으로 정의할 수 있지만 개체는 하나 뿐이기 때문에 “A 자동차” 하나만으로는 클래스로 정의

할 수 없다. 클래스로 정의되기 위해서는 두 개 이상의 속성과 두 개 이상의 개체가 필요하다.

이러한 기준을 적용하여 상품 도메인 온톨로지의 클래스를 구축해야 한다. 예를 들어, 음료는 제조업체, 주원료, 용량 등 두 가지 이상의 속성을 가지고 있고 탄산음료, 과즙음료와 같은 하위 개체를 가지고 있기 때문에 클래스로 정의할 수 있다.

클래스 명	상위 클래스	하위 클래스
유음료	Daily 식품	우유, 두유, 유산균음료
디저트	Daily 식품	호상, 젤리, 푸딩, 무스
유제품	Daily 식품	버터, 마아가린, 치즈
탄산음료	음료	콜라, 사이다, 기타 탄산음료
과즙음료	음료	천연과즙, 기타 과즙음료
국산주류	주류	소주, 맥주, 청주, 전통주, 탁주
수입주류	주류	위스키, 브랜디, 와인
· · ·	· · ·	· · ·

【표 3-1】 온톨로지의 클래스 사전 일부

【표 3-1】은 본 논문에서 구축한 온톨로지 클래스 사전의 일부를 나타내고 있다. 온톨로지를 구출할 때 클래스는 다음과 같이 세 단계 트리구조로 분류할 수 있다.

◉ 최상위 클래스 : 부모를 갖지 않는 가장 상위에 존재하는 클래스
◉ 핵심 클래스 : 최상위 클래스 속성을 물려받은 자식 클래스
◉ 행위 클래스 : 실세계에서 지속적으로 데이터가 생성되는 내용을 갖는 클래스

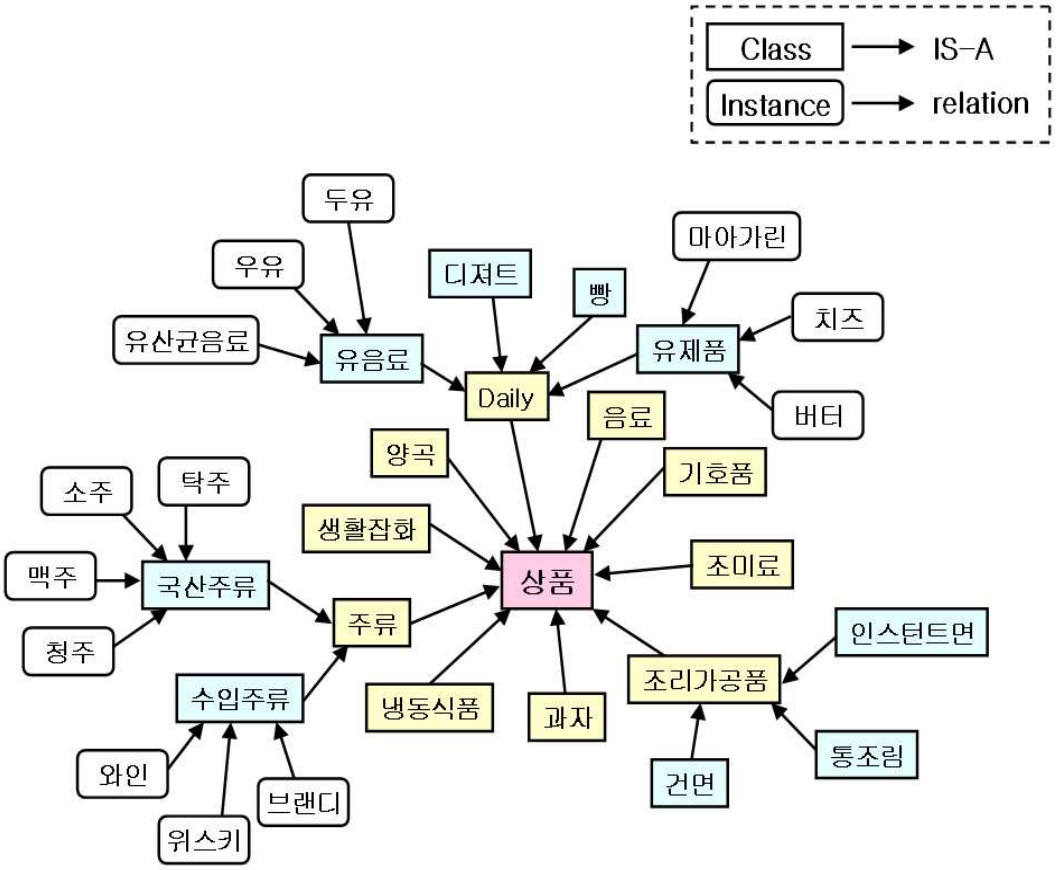
3. 온톨로지의 구현

온톨로지를 구축함에 있어서 최상위 클래스는 부모를 갖고 있지 않기 때문에 제일 상위에 위치하게 된다. 핵심 클래스란 부모를 가지고 있는 클래스로서 각 분류의 조상 클래스가 된다. 핵심 클래스는 관점에 따라 조금씩 달리 설계를 하기도 한

다. 하지만 보편적으로 온톨로지를 구축함에 있어서 핵심 클래스는 2단계 혹은 3단계 수준에 위치하게 된다.

마지막으로 행위 클래스는 트리구조에서 가장 하위에 위치해 있고 자손으로 비유된다. 최상위 클래스와 핵심 클래스를 제외한 모든 것을 행위 클래스로 간주한다.

클래스의 계층 구조를 결정하는 방법에는 보통 상향식 방식, 하향식 방식, 조합 방식이 있다. 본 논문에서는 하향식 방식을 사용하여 클래스 계층 구조를 결정하였다. 하향식 방식은 우선 도메인의 일반적인 클래스를 정의한 후 클래스를 세부적으로 명세한다.[30] 예를 들면 “상품” 클래스에 대한 정의를 시작한 후 “상품” 클래스의 하위 클래스인 “주류”, “Daily”, “음료”등을 생성하고 기술한다. 그런 다음 “주류”에 대한 하위 클래스를 다시 생성한다. 이러한 과정을 통해 클래스의 계층 구조를 결정하게 된다.



【그림 3-2】 온톨로지 클래스의 개념과 관계 일부

【그림 3-2】는 하향식 방식으로 본 논문에서 구축한 상품 도메인 온톨로지 클래스의 개념과 관계를 축약하여 보여주고 있다. 각 클래스 간의 가장 기본적인 관계는 포함관계로서 IS-A관계라고도 표현한다.

이와 같은 개념과 관계를 이용하여 상품 도메인 온톨로지를 구축하였다. 구축할 온톨로지에 존재하는 개념과 그들의 관계는 Protégé 3.1을 이용하여 OWL로 표현하였다.[31]

Protégé 3.1은 스탠퍼드 대학의 의학정보그룹인 SMI(Stanford Medical Informatics)에서 이용자 그룹과 공동으로 개발해 나가는 온톨로지 편집환경으로서 지식 기반 모델의 설계와 구축을 위한 통합개발환경을 제공하고 있다.

```
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns="http://www.owl-ontologies.com/unnamed.owl#"
  xml:base="http://www.owl-ontologies.com/unnamed.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class rdf:ID="유음료">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >마실수 있는 유제품</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="Daily"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="무스">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="디저트"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="축육가공품">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >가공된 육류</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Class rdf:about="#Daily"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="맥주">
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >수입, 국내산 맥주</rdfs:comment>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="국산술"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="수입주류">
```

```
    </owl:ObjectProperty>
    <owl:Thing rdf:ID="상품"/>
  </rdf:RDF>

  <!-- Created with Protege (with OWL Plugin 2.1, Build 284)  http://protege.stanford.edu -->
```

【그림 3-3】OWL로 표현한 상품 도메인 온톨로지

【그림 3-3】은 상품 도메인 온톨로지를 Protégé 3.1을 이용하여 OWL로 표현한 내용의 일부를 보여주고 있다. 여기서 한글로 표현된 클래스명과 속성 등은 이해를 돕기 위해 재구성한 것으로서 실제 OWL로 표현함에 있어서는 모두 영문으로 작성하였다. 대부분의 온톨로지에는 하나의 온톨로지 헤더가 존재한다.【그림 3-3】에서는 최상단 부분에 정의되어 있다. 헤더에는 네임스페이스 및 주석, 버전 정보 등이 선언되어 있다. <rdfs:subClassOf>는 해당 클래스가 "하위 클래스"임을 나타내며 <rdfs:subClassOf>과 </rdfs:subClassOf>사이에 "상위 클래스"명이 위치한다.

위와 같이 구축된 클래스만 가지고 사용자가 원하는 검색 조건에 대한 충분한 답을 내리기에는 무리가 있다. 따라서 각 클래스에 대한 세부적인 설명을 클래스의 속성으로 정의해야 한다. 이렇게 정의된 속성을 통해 보다 정확한 검색이 이루어질 수 있다.

B. 자연어 처리기능

사용자가 원하는 상품을 정확히 검색하기 위해서는 무엇보다 사용자가 입력한 검색어를 명확하게 가공해야 할 필요가 있다. 일반적으로 사용자가 검색어로 입력한 내용은 키워드 위주의 검색어가 아닌 자연어 위주의 검색어일 것이다.

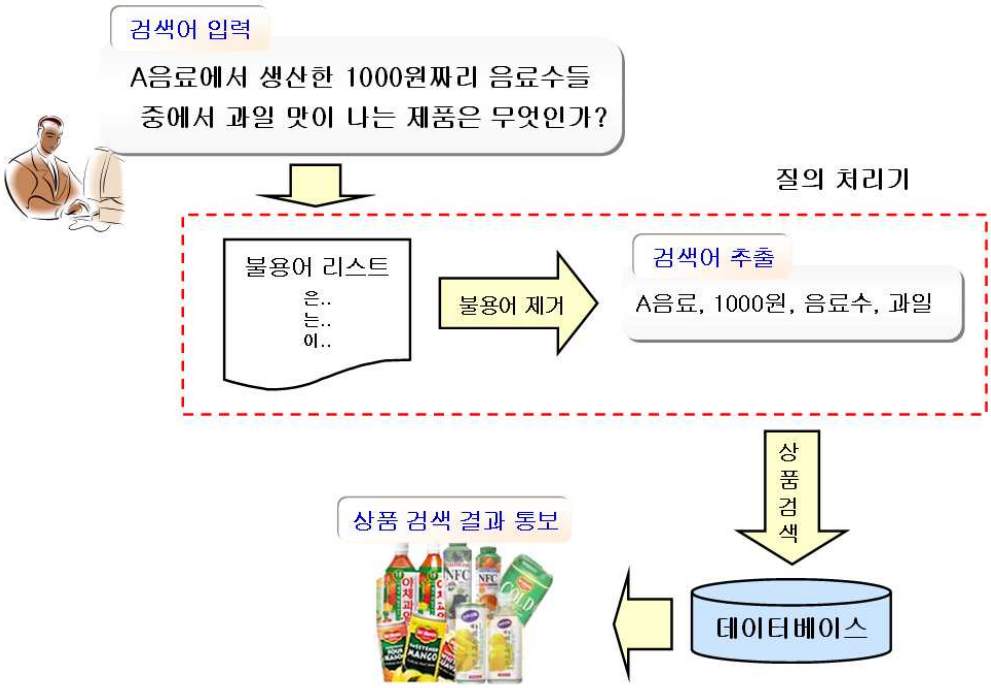
이러한 자연어 위주의 검색어에서 불용어를 제거함으로써 검색어를 보다 명확하게 하여 정확한 검색이 가능하도록 하는 것이다.

불용어 리스트		
~은	~것	~만든
~는	~들	~짜리
~이	~의	~에서
~가	~에	~생산
~중	~인	~에는
~맛	~무엇	~있는가

【표 3-2】불용어 리스트 일부

【표 3-2】는 검색어에서 불필요한 단어를 제거하기 위해 본 논문에서 사용한 불용어 리스트의 일부를 보여주고 있다.

불용어 리스트는 자주 사용되는 조사와 동사를 기본적으로 포함하고, 온·오프라인 쇼핑몰 전문가의 자문을 통해 사용자들이 상품을 검색할 때 자주 사용하는 단어를 파악하여 작성하였다.



【그림 3-4】불용어 제거를 통한 상품 검색 과정

【그림 3-4】는 질의 처리기에서 불용어 리스트를 통해 불용어를 제거하여 상품을 검색하는 과정을 설명하고 있다.

위 그림에서 보는 바와 같이 사용자가 “A음료에서 생산한 1000원짜리 음료수들 중에서 과일맛이 나는 제품은 무엇인가?” 라는 검색어를 입력했을 때 불용어 리스트를 통해 불용어를 제거하게 되면 “A음료”, “1000원”, “음료수”, “과일” 이라는 특정 단어만 남게 된다. 이렇게 재구성된 단어를 검색어로 사용하여 온톨로지를 통해 의미를 부여함으로써 데이터베이스에 저장된 상품들 중 키워드와 일치하는 상품은

물론이며 비록 키워드가 일치하지 않더라도 검색어의 의미적인 해석을 통해 사용자가 원하는 상품을 정확하게 찾아 낼 수 있다.

C. 유의어 처리기능

“가격” 이란 단어와 “판매 금액” 이라는 단어는 서로 다른 단어이지만 하나의 뜻으로 쓰이는 유의어이다. 이러한 유의어를 사용자가 검색 조건으로 입력했을 때 같은 의미로 판단할 수 있도록 온톨로지에 유의어 목록이 정의되어 있어야 한다.

대 표 어	유 의 어
가 격	금액, 판매가, 판매단가, 판매금액, 소비자가, ...
수 량	개수, 입, 개, ...
제조사	제조업체, 생산업체, 거래처, ...
색 상	색깔, 색, 컬러, ...
짠 맛	짭짤, 소금 맛, ...
유제품	우유, 두유, 유산균음료, ...
· · ·	· · ·

【표 3-3】 온톨로지에 정의된 유의어 목록의 일부

【표 3-3】 은 상품 검색 온톨로지에 정의해 놓은 유의어 목록의 일부를 나타내고 있다.

사용자의 단어 선택 성향에 따라 같은 의미의 단어일지라도 서로 다른 단어를 선택하여 검색하게 되는 경우가 많이 있다. 이 경우 검색어가 서로 다르다 할지라도 검색된 결과는 항상 같은 결과가 나와야 한다.

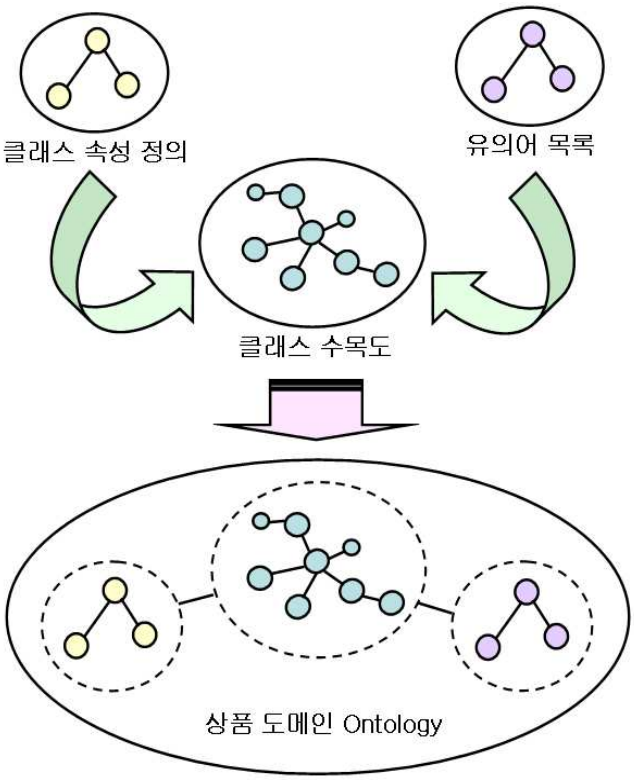
예를 들어 사용자들은 개인의 성향에 따라 다음과 같은 검색어를 입력할 수 있다.

“가격이 500원인 A제조사에서 생산한 커피”

“판매가가 500원인 A제조사에서 생산한 커피”

위의 두 가지 검색어의 결과는 항상 동일한 결과가 나와야 한다. 그러나 기존의 키워드 검색을 이용하여 검색할 경우 “가격”과 “판매가”를 동일한 의미로 파악하지 못하는 경우가 대부분이다. 즉, 데이터베이스에 저장된 상품 데이터 중 키워드가 “가격”이라고 저장된 상품이 있는 반면 “판매가”라는 키워드로 저장된 상품이 존재하기 마련이다. 이러한 경우 키워드 검색에서는 “가격”과 “판매가”를 동일한 의미로 파악하지 못하기 때문에 검색 결과, 서로 다른 검색 결과를 갖게 된다.

하지만 온톨로지를 이용하여 검색할 경우에는 구축해 놓은 유의어 목록을 참조하여 “가격”과 “판매가”를 동일한 의미로 파악하게 된다. 따라서 “가격” 으로 검색한 결과와 “판매가”로 검색한 결과는 항상 동일하게 나타나며, 데이터베이스에 저장 된 상품 데이터 중 “가격”이라는 키워드와 “판매가”라는 키워드를 가진 모든 상품들을 동일하게 검색하게 된다.

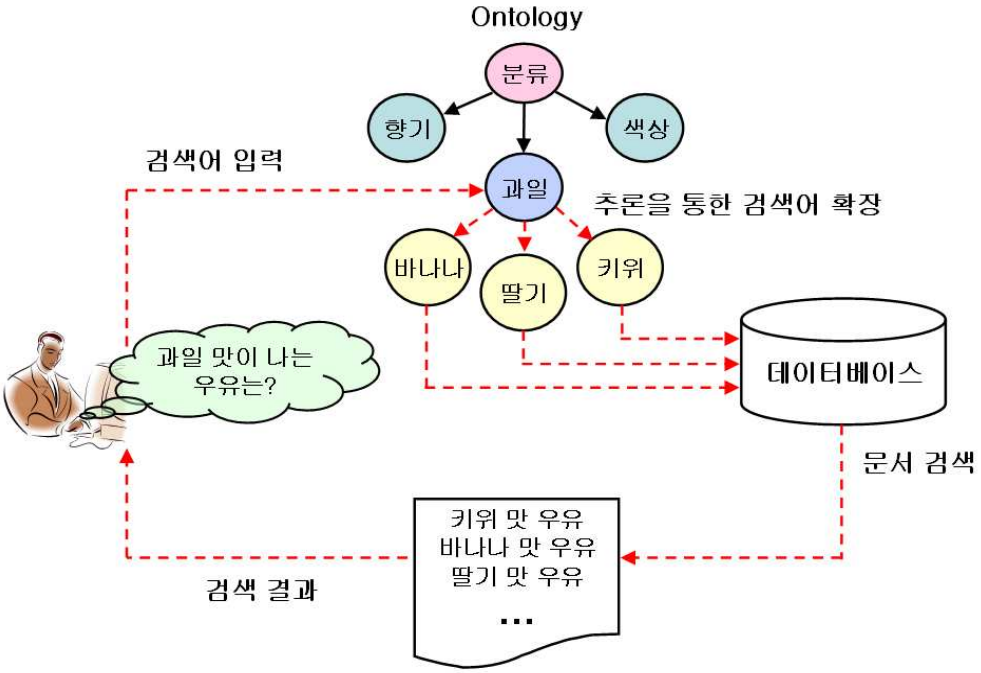


【그림 3-5】 상품 도메인 온톨로지

【그림 3-5】는 클래스 수목도에 클래스 속성과 유의어 목록을 추가한 온톨로지를 나타내고 있다.

이처럼 유의어 목록을 정의해 놓음으로써 사용자의 검색어 선택 성향에 따라 검색 결과가 달라지게 되는 것을 미연에 방지하고 항상 동일한 검색 결과를 유도해 낼 수 있다.

D. 추론기능



【그림 3-6】추론기능을 이용한 검색

【그림 3-6】은 추론기능을 이용한 상품 검색 과정을 나타내고 있다. 예를 들어 사용자가 "과일 맛이 나는 우유는?"이라는 검색 조건을 입력 했을 때, 기존의 검색 방법에서는 "과일"과 "우유"를 검색어로 이용하여 검색을 진행하였다. 하지만 이 경우 데이터베이스에 저장된 우유 상품들 중에서 "과일"이라는 키워드를 가진 상품 데이터가 있을 경우에만 검색이 이루어졌다. 키워드 검색 방법에서는 사용자가 입

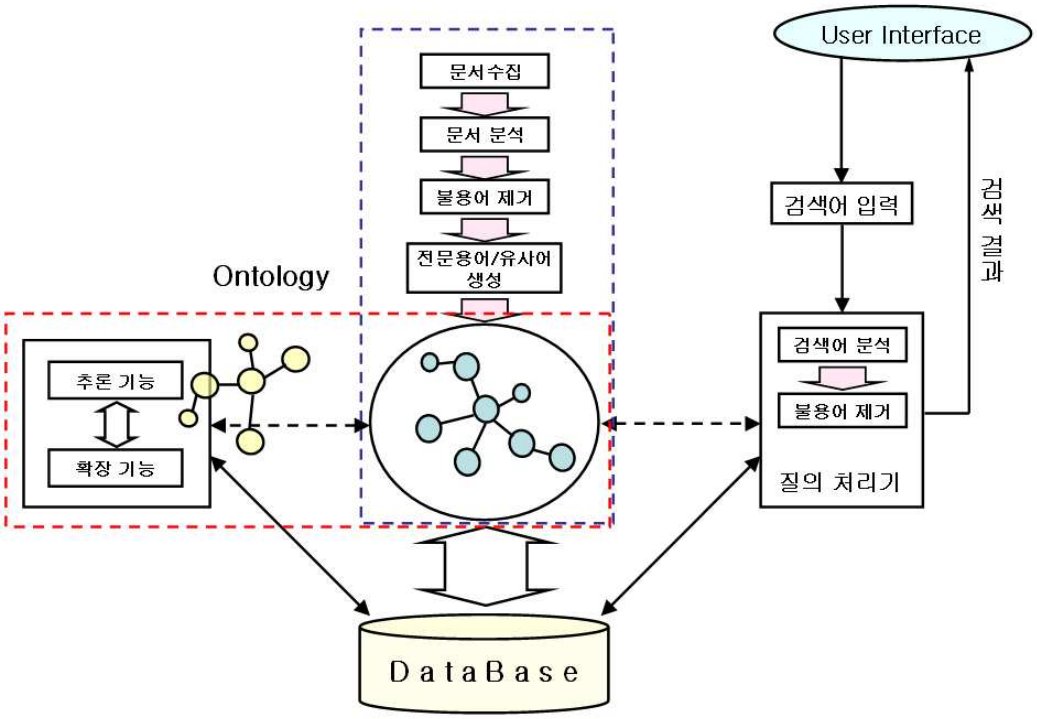
력한 검색어와 데이터베이스에 저장된 상품 정보에 관한 키워드가 정확히 일치했을 경우에만 검색이 가능했기 때문이다.

그러나 온톨로지의 추론기능을 이용하여 검색할 경우 “과일” 클래스에 미리 정의해 놓은 “바나나”, “딸기”, “키위” 등과 같은 하위 클래스를 참조해 “과일”뿐만 아니라 그 하위 클래스에 존재하는 모든 상품 목록까지 검색하게 된다. 그 결과 데이터베이스에 저장되어 있는 우유 상품들 중 “과일” 키워드를 포함하는 상품이 없다 하더라도 그 하위 클래스 목록까지 검색을 진행함으로써 “과일 맛이 나는 우유는?”에 해당하는 결과로 “키위 맛 우유”, “바나나 맛 우유”, “ 딸기 맛 우유”라는 과일 맛과 관련된 모든 상품 데이터를 검색하게 된다.

Ⅳ. 시스템 구현 및 평가

지금까지의 웹에서 검색 방법으로는 주로 키워드 검색이 이용되어 왔다. 하지만 차세대 웹으로서 시맨틱 웹이 대두되기 시작하면서 온톨로지를 활용한 의미적인 검색이 가능하게 되었다.

이 장에서는 본 논문에서 구축한 온톨로지를 이용하여 상품 검색을 했을 경우 얼마나 정확하게 원하는 결과를 검색해 내는지 실험하였다.



【그림 4-1】 온톨로지 기반의 상품 검색 시스템 구조

본 논문에서 구축하고자 하는 온톨로지 기반의 상품 검색 시스템의 구조는 【그림 4-1】과 같다.

사용자 인터페이스(UI : User Interface)는 일반적으로 사용하고 있는 GUI 환경에서 사용자가 원하는 검색어를 입력하고 검색 결과를 통보 받을 수 있도록 시스

템을 설계하였다.

상품 검색 절차는 먼저 사용자 인터페이스를 통해 사용자가 입력한 검색어를 질의 처리기에 전송한다. 질의 처리기에 입력된 검색어는 불용어 제거를 통해 검색어를 보다 명확하게 재구성한다. 재구성된 검색어에 온톨로지를 통해 의미적인 내용을 부여함으로써 상품 검색시 사용자가 원하는 상품을 정확하게 데이터베이스에서 검색할 수 있도록 하였다.

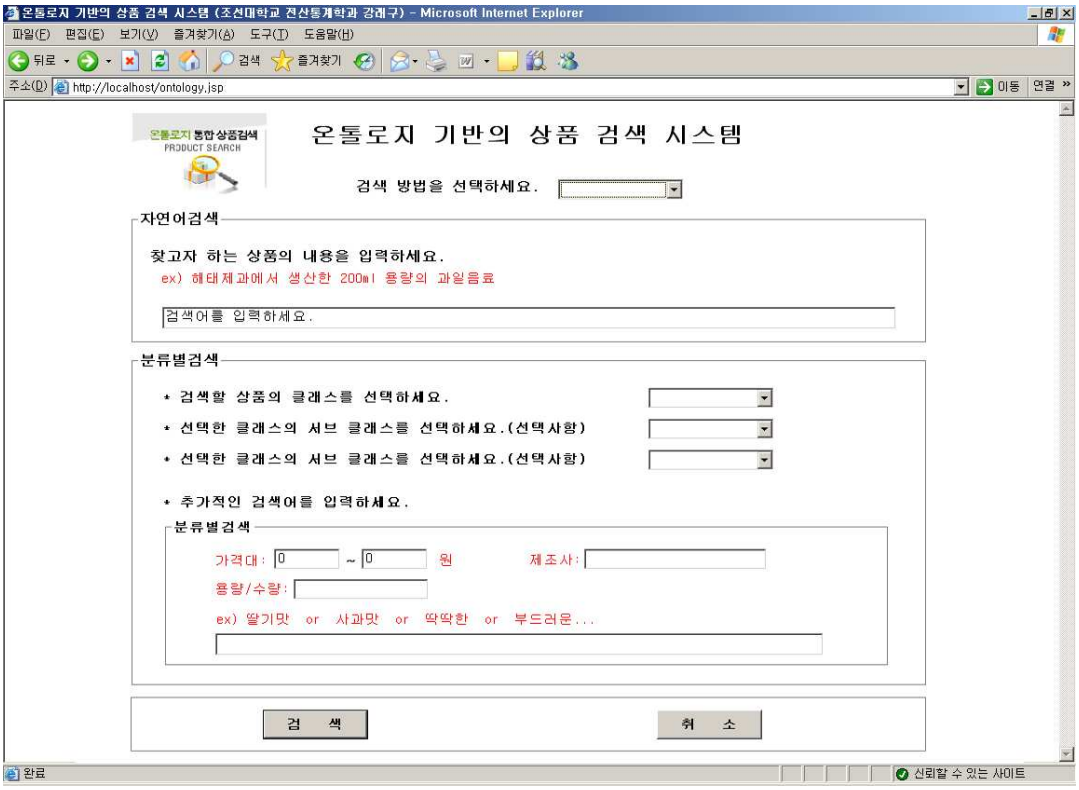
또한 추론 및 확장 기능을 통해 유사 상품이나 키워드가 정확히 일치하지 않은 상품이다 하더라도 검색 조건에 적합한 상품이라면 모두 검색이 가능하게 된다.

A. 상품 검색 시스템 구현 환경

본 논문에서 구축한 온톨로지 기반의 상품 검색 시스템의 구현 환경은 다음과 같다.

- ◉ OS - Windows XP Professional
- ◉ User Interface - JSP, PowerBuilder9.0
- ◉ Web Server - Apache
- ◉ Java Container - Tomcat 4.1
- ◉ DBMS - Oracle 9i
- ◉ Ontology Tool - Protégé 3.1
- ◉ Ontology 언어 - OWL

실험을 위해 사용할 상품 데이터는 현재 전국적으로 온·오프라인 할인점을 운영 중에 있는 A할인점의 상품 데이터 약 40,000여개의 품목을 Oracle 9i로 구축하여 사용하였다.



【그림 4-2】 사용자 인터페이스(UI : User Interface)

【그림 4-2】는 본 논문에서 상품 검색을 위해 사용하게 될 사용자 인터페이스 화면을 보여주고 있다. 사용자 인터페이스는 JSP와 PowerBuilder9.0을 이용하여 웹 상에서 사용 가능하도록 직접 구현하였다. 상품을 검색하는 방법은 두 가지로 구성하였다. 하나는 사용자가 상품에 대한 기본적인 지식을 가지고 있을 때 사용할 수 있는 검색 방법으로서 검색하고자 하는 상품의 대·중·소분류를 선택하여 선택된 분류를 검색어로 사용하는 분류별 검색 방법이다. 다른 하나는 입력된 검색어를 불용어 리스트를 통해 불필요한 단어를 삭제하여 검색어를 보다 명확하게 처리함으로써 자연어 위주로 입력된 검색어에 사용하기 적합한 자연어로 검색하는 방법이다.

두 가지 검색 방법을 이용해 만들어 진 검색 조건을 온톨로지를 이용해 의미적인 내용을 부여함으로써 원하는 상품을 데이터베이스에서 정확하게 검색할 수 있으며 검색 조건의 추론 및 확장을 통해 유사 상품이나 비슷한 분류의 상품까지도

검색할 수 있다. 또한 검색된 상품들 중 특정 상품에 대한 자세한 정보가 필요할 경우 해당 상품을 선택하게 되면 상품 정보를 XML문서로 제공하도록 구성하였다.

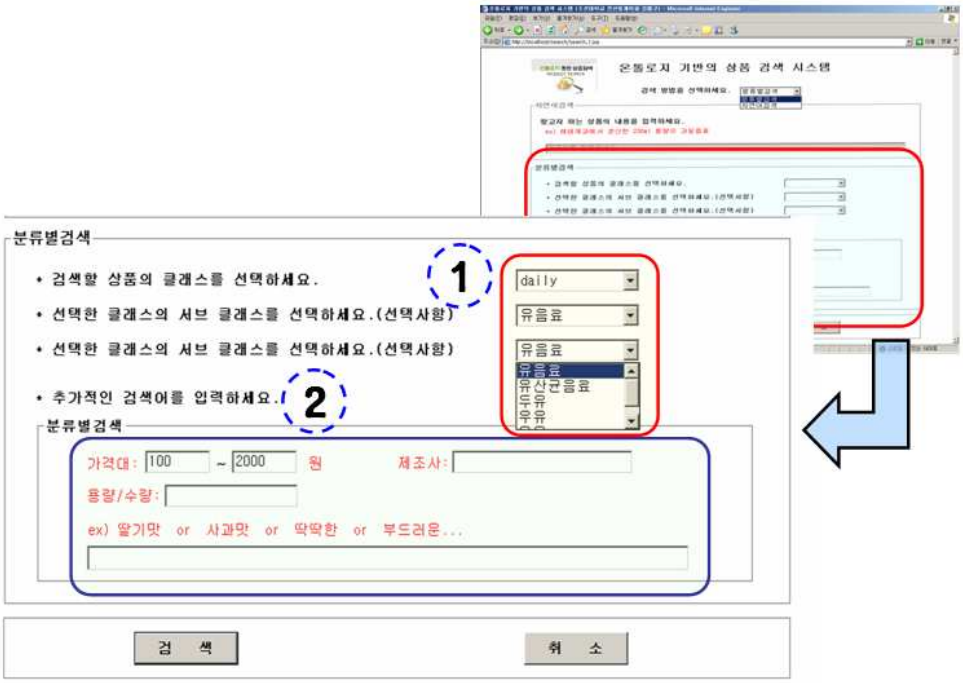
```
If i_index = 1 Then          //분류별 검색
select grp_name
  into :s_grp_name
  from mmgrpm
 where grp_code = :grp_code;

s_dpt_name = drop_2.text
s_line_name = drop_3.text
s_sle_2 = sle_2.text

 select dpt_code
   into :dpt_code
   from mmdptm
  where dpt_name = :s_dpt_name;

 select line_code
   into :line_code
   from mmlinm
  where line_name = :s_line_name;

s_tot = '찾고자 하는 상품이 ' + s_grp_name + '에 ' + s_dpt_name + '에 ' + s_line_name + '~r~n' + s_sle

i_mess = messagebox("알림",s_tot,Question!,YesNo!,1)

If i_mess = 2 Then
   messagebox("알림", "검색 내용을 다시 입력하세요.")
   dw_1.setfocus()
else
   l_sale_f = st_13.text
   l_sale_t = st_14.text

   If l_sale_t = '0' Then
      l_sale_t = '1000000'
   End If
```

```
        OPEN cur_12; //커서 시작
           open(win_result)
        DO while sqlca.sqlcode = 0

        FETCH cur_12 INTO :s_name, :l_up_salex, :s_ven_name, :s_itm_codex, :s_itm_spec;

        l_up_salex = l_up_salex

        do while len(s_name) < 30
           s_name = s_name + ' '
        loop

        do while len(l_up_salex) < 10
           l_up_salex =  ' ' + l_up_salex
        loop

        do while len(s_ven_name) < 20
           s_ven_name = s_ven_name + ' '
        loop

        do while len(s_itm_codex) < 5
           s_itm_codex = s_itm_codex + ' '
        loop

        do while len(s_itm_spec) < 10
           s_itm_spec = ' ' + s_itm_spec
        loop

        win_result.lb_1.AddItem(s_name + '   ' + s_itm_codex + '   ' + string(l_up_salex) + '원' + '
        LOOP  // 0 이나오면 ITEM에 계속 입력하고 그렇지 않으면 Close cur_dept; // (반복한다)
     Close cur_12; //커서 close
  End if

 End If
```

【그림 4-3】 사용자 인터페이스 소스 코드 일부(PowerBuilder로 개발한 부분)

B. 분류별 검색 실험 및 평가

찾고자 하는 상품에 대한 기본적인 지식이 있을 경우 사용하는 검색 방법으로서 상품이 해당하는 분류를 선택하고 추가적인 검색 내용을 입력하면 입력된 값들이 파라미터로 전달되고 전달된 값을 가지고 데이터베이스에서 상품을 찾게 된다.



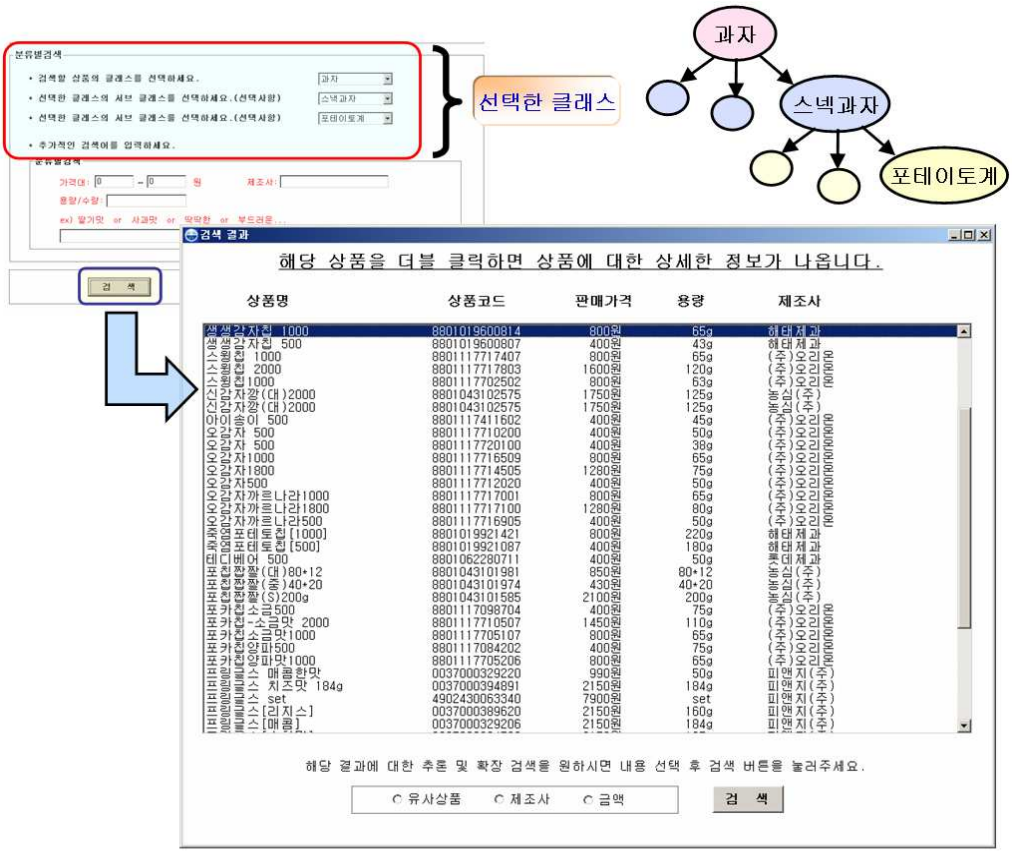
【그림 4-4】 분류별 검색 방법

【그림 4-4】는 분류별 검색 방법을 나타내고 있다. ①과 ②에 해당 하는 부분이 온톨로지로 구축된 부분으로서 사용자가 쉽게 사용할 수 있도록 재구성하여 사용자 인터페이스 화면에 디자인 하였다.

①의 경우 각각의 상품 정보를 분석하여 구축한 온톨로지를 사용자로 하여금 검색하고자 하는 상품의 대·중·소분류로 활용할 수 있도록 클래스를 도식화 하였다. 즉, 검색을 원하는 상품에 대한 기본적인 지식을 알고 있을 경우 찾고자 하는 상품의 대·중·소분류를 메뉴에서 선택하게 되면 선택된 각각의 분류와 구축된

온톨로지의 상·하위 클래스를 1:1로 매칭을 시켜 내부적으로는 사용자가 클래스를 직접 선택하여 상품을 검색하는 것과 같은 개념이 된다.

②의 경우는 좀 더 구체적이고 세분화 된 검색을 하고자 할 때 추가적으로 입력이 가능한 부분으로서 주로 유의어 목록과 추론기능을 이용하게 된다. 가격, 제조사, 제품 용량, 수량뿐 만 아니라 각 상품의 특징을 추가적으로 입력함으로써 상품 검색을 보다 상세하게 할 수 있도록 해 준다. 또한 검색어와 상품 데이터의 키워드가 일치하지 않더라도 온톨로지에 구축된 유의어 목록과 추론기능을 이용하여 검색 조건에 해당되는 상품이라면 모두 검색이 가능하다.

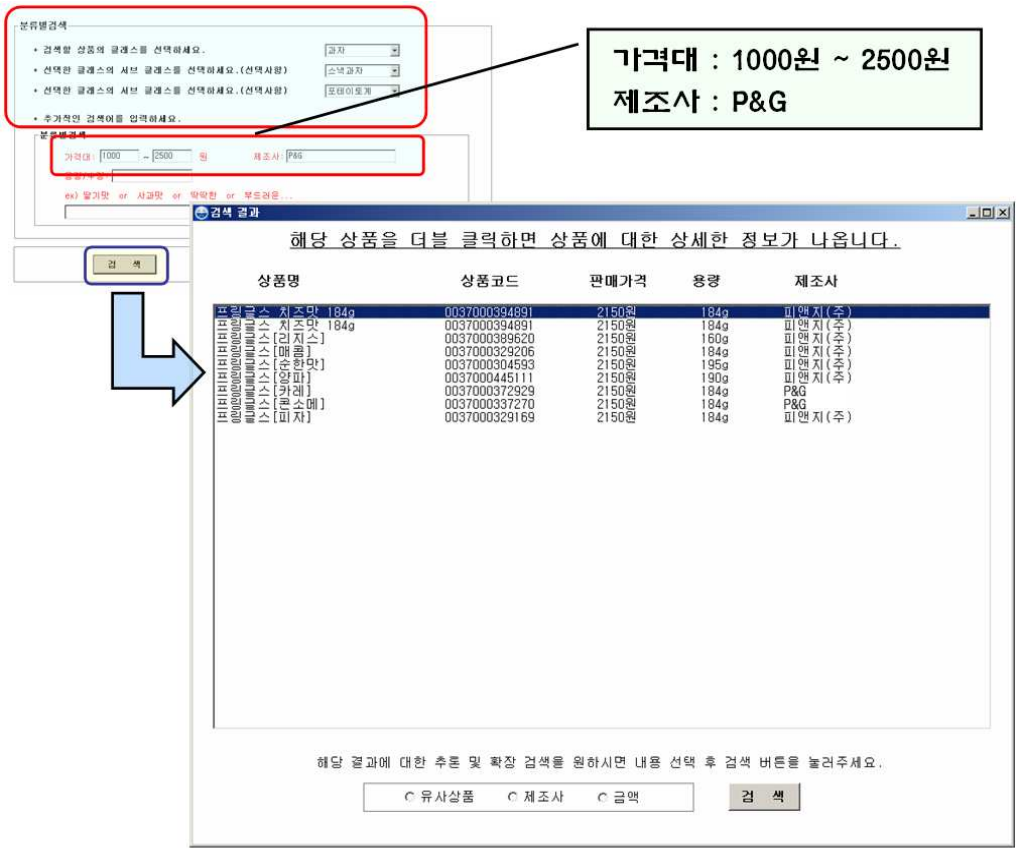


【그림 4-5】 분류별 검색 결과 1

【그림 4-5】는 분류별 검색 결과를 보여주고 있다. 그림에서 보면 사용자는 검

색하고자 하는 상품의 대·중·소분류를 “과자” -> “스넥과자” -> “포테이토계” 순의로 선택하였고, 이렇게 선택된 분류를 가지고 상품 도메인 온톨로지를 이용하여 상품을 검색하게 된다. 이 때 선택된 대·중·소분류와 온톨로지의 상·하위 클래스는 1:1로 서로 매칭되어 검색이 진행된다. 이러한 과정을 통해 검색된 상품들은 모두 “과자” 클래스의 하위 클래스로서 그 중 “스넥과자” 클래스에 속하고 “스넥과자” 클래스의 하위 클래스인 “포테이토계” 클래스에 속하는 상품들이다.

이처럼 내부적으로 구축되어 있는 온톨로지의 상·하위 클래스 계층을 사용자가 인터페이스 화면에서 대·중·소분류를 이용하여 직접 선택할 수 있도록 디자인함으로써 사용자로 하여금 찾고자 하는 상품이 속해 있는 온톨로지 클래스를 직접 선택할 수 있는 효과를 주는 것이다.



【그림 4-6】 분류별 검색 결과 2

【그림 4-6】은 【그림 4-5】의 검색 결과에 추가적인 검색어 입력을 통해 좀 더 세부적인 상품 검색을 실행한 결과이다. 검색 조건에 나타나 있듯이 가격대와 제조사를 추가 검색어로 입력하여【그림 4-5】에 나타난 결과 중 가격대가 “1000원~2500원” 사이에 해당하는 상품들 중 제조사가 “P&G”에 해당하는 상품들만을 검색하도록 하였다.

이 때 기존의 키워드 검색에서는 추가적으로 입력된 검색어 중 가격과 관련된 검색어는 특별한 문제가 없지만 제조사와 관련된 검색어에서는 문제가 발생하게 된다.

키워드 검색에서는 검색어로 입력한 제조사의 영문 명칭인 “P&G”와 한글 명칭인 “피앤지(주)”를 같은 의미로 판단하지 못한다. 그렇기 때문에 “P&G”와 “피앤지(주)”가 동일한 제조사임에도 불구하고 데이터베이스에 저장된 상품의 제조사 키워드가 어떠한 명칭으로 저장되어 있느냐에 따라 서로 다른 검색 결과를 나타내게 된다.

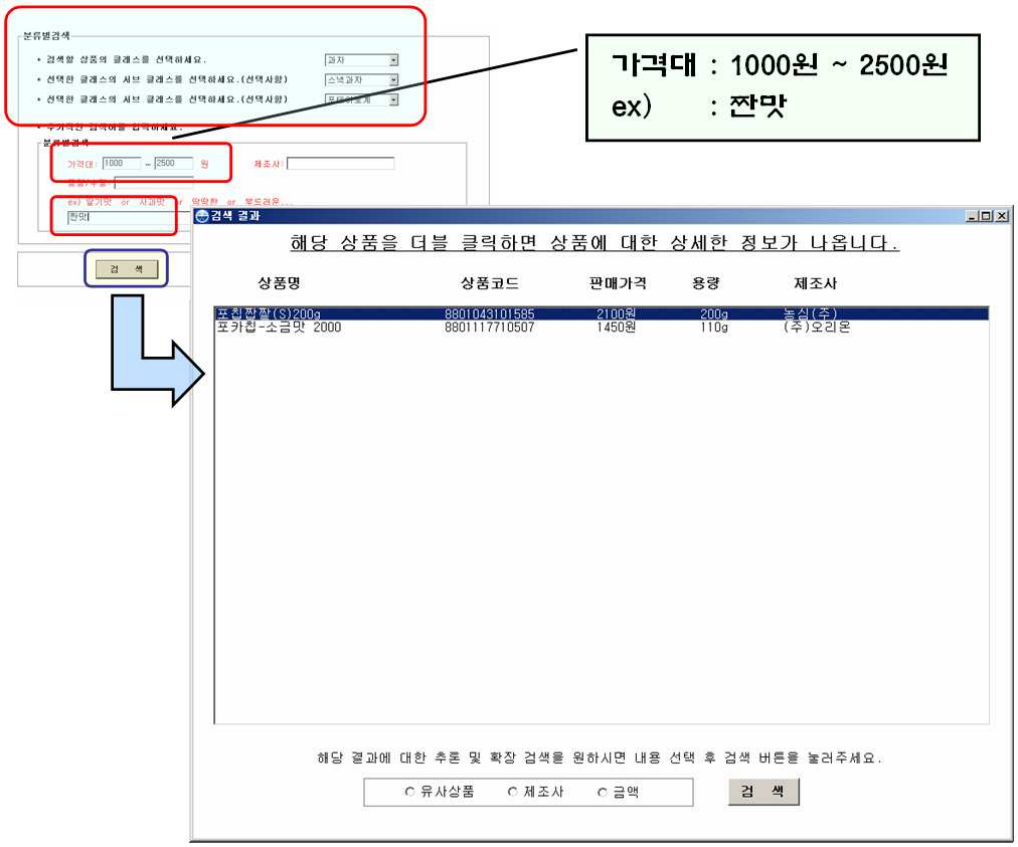
하지만【그림 4-6】에서처럼 온톨로지를 이용한 상품 검색에서는【표 4-1】과 같이 온톨로지에 구축되어 있는 제조사 관련 유의어 목록을 통해 “P&G“와 “피앤지(주)“가 같은 의미로 사용되고 있음을 추론하여 검색하기 때문에 둘 중 어떠한 명칭을 사용하더라도 항상 동일한 검색 결과를 나타내게 된다.

대 표 어	유 의 어
P&G	피앤지, 피앤지(주), 피엔지, 피엔지(주)
Unilever	유니레버, 유니레버코리아, 유니래버, 유니래버코리아
오리온	동양제과, (주)오리온
남양유업	남양, 남양우유
매일유업	매일, 매일우유
· · ·	· · ·

【표 4-1】제조사 관련 유의어 목록 일부

이는 이미 각 상품들의 특징과 자주 쓰이는 단어 및 유의어 등을 분석하여 구축해놓은 온톨로지를 통해 검색어의 의미적인 부분까지 파악하여 검색하기 때문에

가능한 일이다.



【그림 4-7】 분류별 검색 결과 3

【그림 4-7】은 【그림 4-5】의 검색 결과에 가격대와 ex) 부분을 검색어로 추가하여 검색한 결과이다. 가격대가 “1000원~2500원”사이에 해당하는 상품들과 “짠맛”이 나는 상품들을 추가적인 검색어로 사용하여 검색하였다.

【그림 4-7】의 경우도 키워드 검색에서는 정확한 검색 결과를 얻을 수 없다. 키워드 검색에서는 “짠맛”이라는 검색어가 상품 키워드와 정확히 일치하는 상품에 대해서만 검색이 가능하기 때문이다.

다시 말해 키워드 검색에서는 검색어와 키워드가 정확히 일치해야만 검색이 가능하며 그와 유사한 의미의 검색어에 대해서는 서로 다른 상품으로 인식을 한다.

그러나 온톨로지를 이용한 검색에서는 온톨로지를 구축시 정의해 놓은 유의어 목록을 통해 “짠맛”, “짭짤”, “소금맛”과 같이 유사한 의미의 단어들을 모두 같은 의미로 인식하여 검색을 진행하기 때문에【그림 4-7】과 같이 “짠맛”이라는 검색어로 검색을 하더라도 결과는 “짭짤”, “소금맛”과 같이 유사한 의미의 키워드를 가지고 있는 상품들까지 검색 결과로 얻을 수 있는 것이다. (“【표 3-3】온톨로지에 정의된 유의어 목록의 일부” 참조)

따라서 서로 다른 검색어라 할지라도 그 의미만 유사하다면 모두 같은 검색어로 인식하게 됨으로써 항상 같은 검색 결과를 얻게 된다.

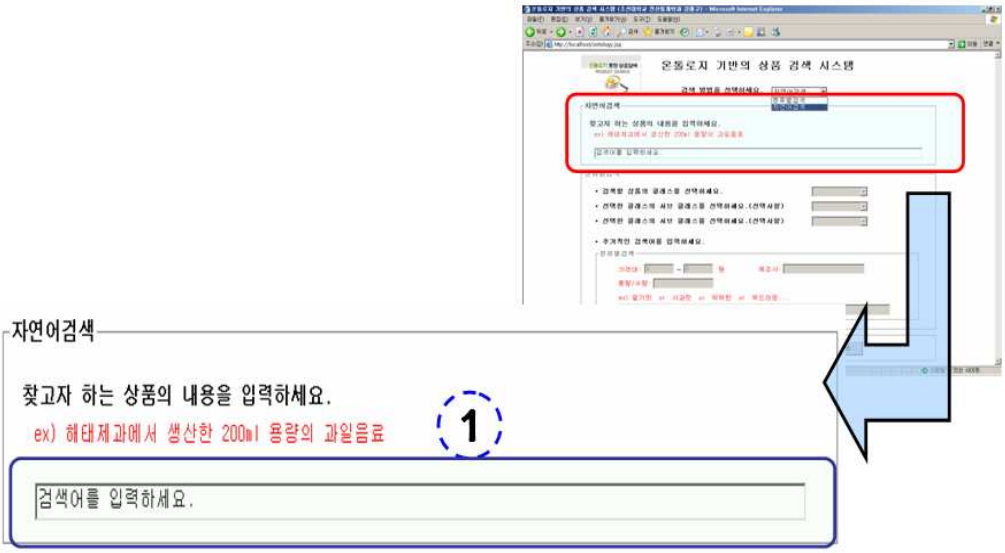
C. 자연어 검색 실험 및 평가

분류별 검색보다는 이번 장에서 설명할 자연어 검색이 온톨로지의 장점을 보다 적극적으로 활용할 수 있는 검색 방법이다. 대부분의 사용자들은 찾고자 하는 상품에 대한 기본적인 지식이 없는 경우가 많다. 이러한 경우 사용자들은 자신이 찾고자 하는 상품을 검색할 때 어떤 키워드로 상품을 검색해야 하는지 알지 못하기 때문에 정확한 검색을 하기 어려웠다.

기존의 키워드 검색 방법에서는 위와 같은 문제점을 특별히 해결할 수 있는 방법이 없었다. 사용자 스스로 검색된 결과에서 몇 번의 검색을 다시 진행하고 나서야만 원하는 결과를 얻을 수 있었다. 하지만 이렇게 검색된 결과도 100% 신뢰할 수 있는 결과는 아니다.

왜냐하면 검색할 때 정확한 키워드를 입력하지 않았거나 상품의 특징, 동형이의어, 동음이의어 등과 같이 사용자가 모르는 상품 정보나 의미를 파악해야 하는 검색어에 대해서는 정확한 판단을 하지 못하기 때문이다.

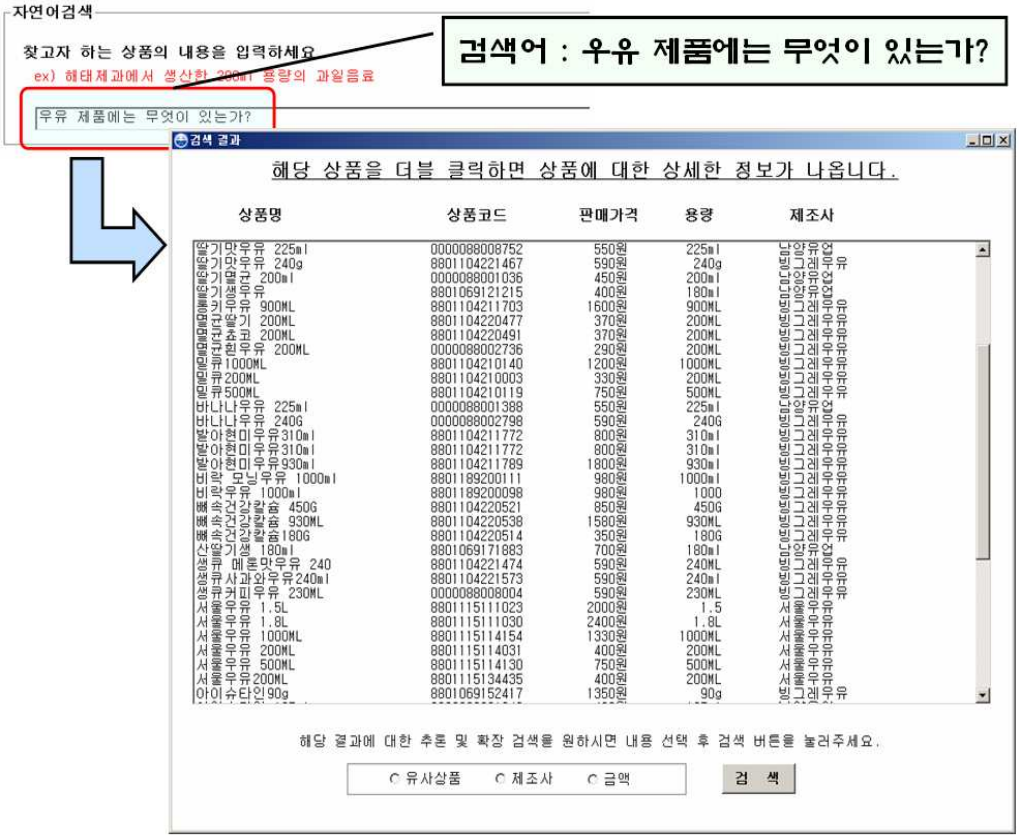
반면 온톨로지를 이용한 자연어 검색을 활용하면 이와 같은 문제점을 쉽게 해결할 수 있다.



【그림 4-8】 자연어 검색 방법

【그림 4-8】은 자연어 검색 방법을 보여주고 있다. ①에 찾고자 하는 상품에 관한 검색 문장을 입력한다. 입력한 검색 문장을 불용어 리스트를 통해 검색시 불필요한 단어를 제거하여 검색어를 재구성한다. 이렇게 재구성된 검색어를 가지고 상품을 검색하게 된다.

이 때 재구성된 검색어와 데이터베이스에 저장된 각 상품 데이터의 키워드가 서로 일치하지 않는다 하더라도 온톨로지에 구축된 클래스와 추론 및 유의어 목록을 통해 검색어의 의미관계를 파악함으로써 검색 조건에 적합한 상품이면 검색이 가능하게 된다.



【그림 4-9】 자연어 검색 결과 1

【그림 4-9】는 기본적인 자연어 검색 결과를 보여주고 있다.

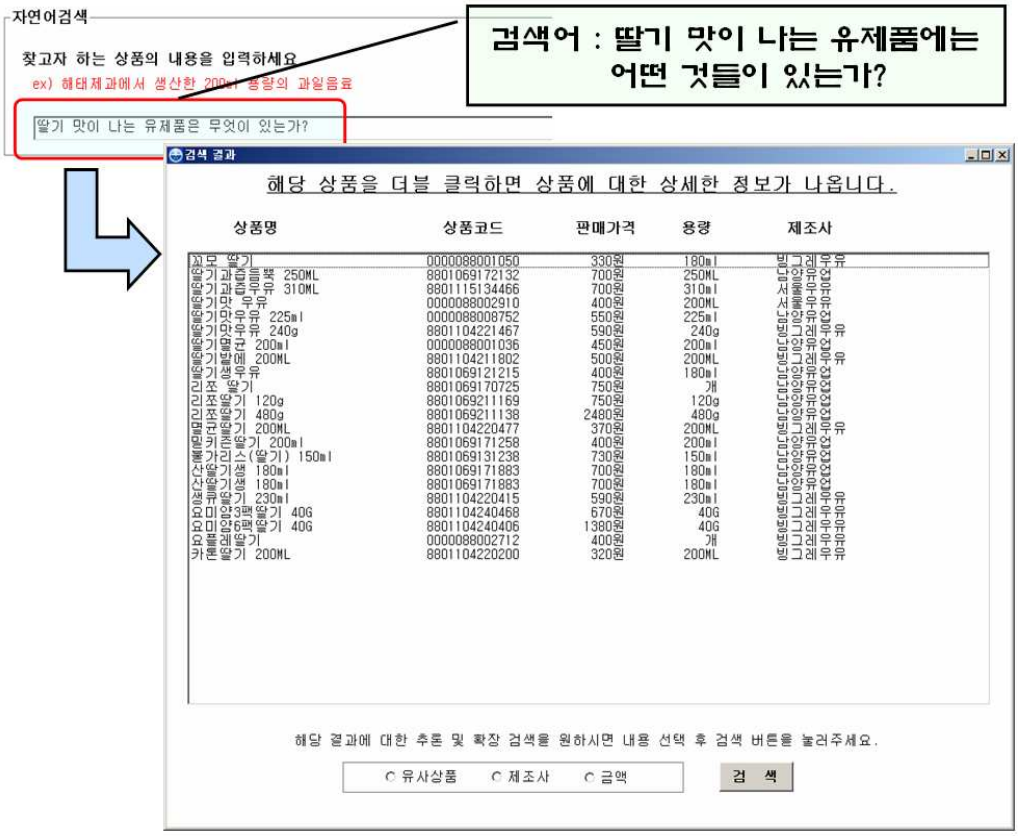
“우유 제품에는 무엇이 있는가?”라는 문장을 사용자가 검색 문장으로 입력했을 때 입력된 검색 문장을 불용어 리스트를 이용해서 검색시 불필요한 단어를 제거하게 되면 “우유”와 “제품”이라는 두 개의 단어만 남게 된다.

기존의 키워드 검색 방법에서는 “우유”와 “제품”을 각각 다른 검색어로 판단을 하고 “우유”라는 키워드와 “제품”이라는 키워드가 있는 상품을 데이터베이스에서 검색할 것이며 결과적으로 “우유”와 “제품” 키워드가 있는 상품은 모두 검색 될 것이다. 이러한 경우 사용자는 “우유”라는 검색어로 다시 한번 검색을 해야만 정확한 결과를 얻을 수 있는 번거로움이 있었다.

그러나 온톨로지를 이용해서 검색할 경우 “우유”와 “제품” 두 단어를 별개의 단

어가 아닌 하나로 연결된 의미의 단어로 추론을 하여 “우유 제품”이라는 검색어로 상품을 검색하게 된다.【그림 4-9】처럼 간단한 검색 문장으로 상품을 검색할 경우에는 기존의 키워드 검색 방법을 사용하더라도 크게 불편함을 느끼지는 못했을 것이다. 하지만 검색 문장이 좀 더 복잡해지고 다양한 키워드를 포함하고 있을 경우에는 정확한 검색을 진행하기에 많은 어려움이 있었다.

【그림 4-10】은 좀 더 복잡한 검색 문장을 온톨로지를 이용하여 검색한 결과를 보여주고 있다.



【그림 4-10】자연어 검색 결과 2

【그림 4-10】에서 입력된 검색 문장은【그림 4-9】의 검색 문장보다 복잡한 구조로 되어 있다. 입력된 검색 문장인 “딸기 맛이 나는 유제품은 어떤 것들이 있는가?”에서 불용어 리스트를 통해 검색어를 재구성하면 “딸기 맛”과 “유제품”으로 재구성된다.

이 중 “딸기 맛”이라는 단어는 상품을 검색하는 데 있어서 그 단어 자체를 검색어로 그대로 사용하더라도 문제가 없다. 반면 “유제품”이라는 단어는 두 가지 의미로 해석할 수 있다. 하나는 “유제품”이라는 단일 상품으로 해석할 수 있고 다른 하나는 우유나 유산균음료, 두유 등과 같은 상품들을 포괄적으로 가리키는 분류에 의미로 해석할 수 있다. 이러한 두 가지 의미 중【그림 4-10】에서 입력한 검색 문장의 “유제품”은 분류를 가리키는 “유제품”이라고 볼 수 있으며 검색 결과에서도 딸기 맛과 관련된 모든 유제품들이 검색된 걸 확인할 수 있다.

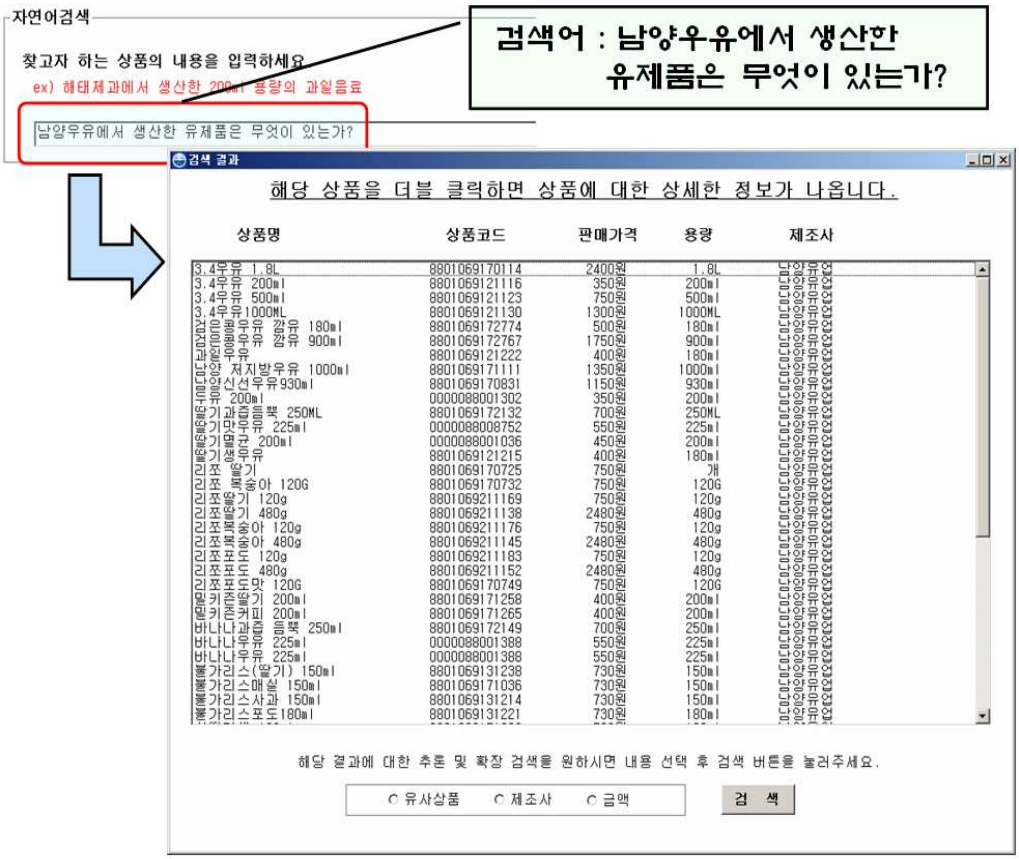
그렇다면 “유제품”이라는 검색어만으로 어떻게【그림 4-10】과 같이 유제품과 관련된 모든 상품들을 검색 결과로 얻을 수 있었을까?

바로 온톨로지를 이용하여 검색하기 때문에 가능한 일이다. 사용자가 입력한 검색어에 의미를 부여함으로써 “유제품”이 단일 상품이 아닌 분류로서 사용되고 있음을 판단하여 온톨로지를 구축시 정의해 놓은 클래스 정보를 바탕으로 “유제품”을 상위 클래스로 가지는 하위 클래스에 속해 있는 모든 상품들까지도 검색에 활용하게 된다.

상위 클래스	하위 클래스
유 제 품	우 유
	발 효 유
	두 유
	유 음 료
	유산균음료
	치 즈
	버 터

【표 4-2】유제품 관련 하위 클래스

【표 4-2】는 본 논문에서 구축한 온톨로지에 정의되어 있는 유제품을 상위 클래스로 갖는 모든 하위 클래스를 보여주고 있다. 유제품 클래스에 속해 있는 하위 클래스의 상품들까지도 검색을 함으로써 유제품과 관련이 있는 모든 상품들을 찾아내게 된다. 이러한 절차를 통해 상품을 검색한 결과가【그림 4-10】에 나타나 있다.



【그림 4-11】 자연어 검색 결과 3

【그림 4-11】도 【그림 4-10】과 같은 방법으로 검색을 진행하여 얻은 결과 이다.【그림 4-10】과는 다르게 제조사를 검색어로 사용하여 검색하였다.

그림에서 보듯이 “남양우유에서 생산한 유제품은 무엇이 있는가?”라는 검색 문장을 사용하여 검색하였다. 검색 문장에서는 “남양우유”라는 제조사 명칭을 사용하여 검색을 했지만 결과는 “남양유업”이라는 제조사로 검색이 되었다.

사실 “남양우유”라는 제조사는 존재하지 않는다. 정확한 제조사 명은 “남양유업”이 맞다. 그런데 어떻게 “남양우유”라는 제조사 명에서 “남양유업”이라는 제조사 명을 유추해냈을까? 이것은 온톨로지의 추론기능과 유의어 목록을 통해 가능하다. (“【표 4-1】 제조사 관련 유의어 목록 일부” 참조)

제조사와 관련해서 기술해 놓은 유의어 목록을 통해 “냠양우유”와 “남양유업”을

같은 제조사로 추론하여 검색을 진행하였기 때문에 가능한 결과이다.

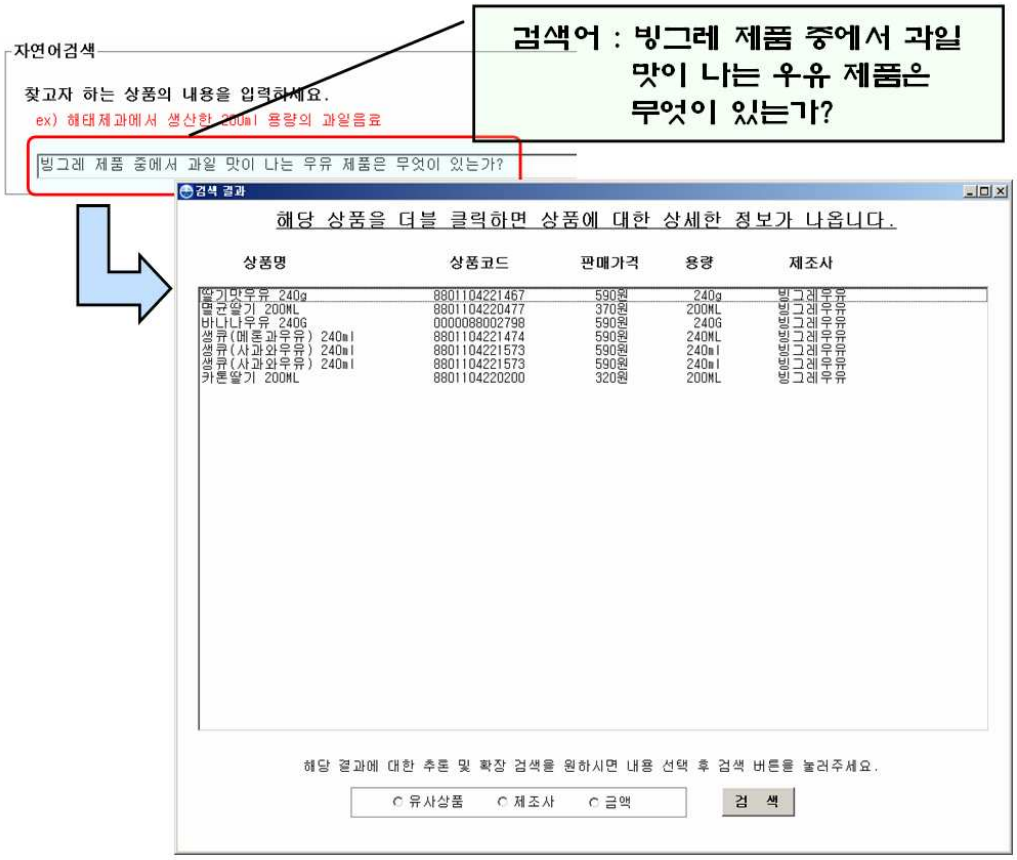
이렇듯 유의어 목록과 추론을 통해 만들어진 제조사 명칭을 검색어로 사용하여 검색을 함으로써 【그림 4-11】과 같은 검색 결과를 얻을 수 있는 것이다.

【그림 4-12】는 국내 최대 온·오프라인 할인점인 L할인점(위쪽 그림)과 E할인점의 온라인 쇼핑몰에서 【그림 4-11】에서 사용한 검색 문장으로 상품을 검색한 결과를 보여주고 있다.

【그림 4-12】국내 온라인 할인점에서 검색한 결과 1

검색 결과 L할인점에서는 유제품과는 전혀 관련이 없는 상품들만 검색되었다. 검색된 상품들을 살펴보면 “남양우유에서 생산한 유제품은 무엇이 있는가?”라는 검색 문장에서 “한”과 “생산”이라는 엉뚱한 단어를 검색어로 사용하고 있는 걸 확인할 수 있다. 그리고 E할인점에서는 단 하나의 상품도 검색하지 못했다.

이처럼 키워드를 기반으로 하는 검색에서는 검색 문장이 복잡해질 경우 정확한 검색을 하는 데 많은 어려움이 있지만 온톨로지를 이용하여 검색할 경우에는 문장을 의미적으로 해석함으로써 보다 정확한 검색이 가능하다.



【그림 4-13】 자연어 검색 결과 4

【그림 4-13】은 좀 더 복잡한 형태의 검색 문장을 온톨로지를 이용해 검색한 결과를 보여주고 있다. 그림에서 보면 “빙그레 제품 중에서 과일 맛이 나는 우유 제품은 무엇이 있는가?”라는 검색 문장으로 상품을 검색하였다.

위 검색 문장을 불용어 리스트를 통해 불필요한 단어를 삭제하게 되면 “빙그레 제품”, “과일 맛”, “우유 제품”으로 재구성이 되며 이렇게 재구성된 세 개의 단어를 검색어로 사용하게 된다. 이 중 “빙그레 제품”과 “우유 제품”은 【그림 4-9】에서 기술한 바와 같이 하나의 단어로 인식하여 검색하게 된다. 그리고 “과일 맛”은 온톨로지에 정의해 놓은 과일 관련 클래스 정보를 토대로 검색을 하게 된다.

상위 클래스	하위 클래스	
과 일	사 과	복숭아
	배	자 두
	딸 기	참 외
	포 도	바나나
	키 위	· · ·

【표 4-3】 과일 관련 하위 클래스

【표 4-3】은 온톨로지에 구축된 클래스 중 과일을 상위 클래스로 갖는 하위 클래스를 나타내고 있다. “과일 맛”이라는 키워드를 갖는 별개의 상품도 존재할 수 있지만 하나의 상품이라기보다는 과일을 총칭하는 뜻의 “과일 맛”이라고 보는 것이 보다 정확한 판단일 것이다. 결국 “과일”을 상위 클래스로 갖는 모든 하위 클래스에 속해 있는 상품들을 검색 대상으로 하고 재구성된 검색어인 “빙그레 제품”과 “우유 제품”을 공통적으로 포함하고 있는 상품들을 검색하게 된다. 그 결과 【그림 4-13】과 같은 검색 결과를 얻게 되었다.

【그림 4-14】 국내 온라인 할인점에서 검색한 결과 2

【그림 4-14】는 “빙그레 제품 중에서 과일 맛이 나는 우유 제품은 무엇이 있는가?”라는 검색 문장을 온라인 쇼핑몰에서 검색 한 결과이다. 【그림 4-12】의 결과와 마찬가지로 L할인점에서는 검색 조건과 전혀 관련이 없는 상품들을 검색하였다. 이 때 사용된 검색어로는 “제품은”, “제품”, “맛”, “맛이”와 같이 검색 조건과 무관한 단어를 검색어로 사용하였다. 또한 E할인점 역시 【그림 4-12】의 결과와 마찬가지로 검색 조건에 맞는 상품을 단 한 가지도 검색해 내지 못했다.

두 곳의 국내 할인점의 검색 결과에서처럼 키워드를 기반으로 하는 검색에서는 자연어 위주로된 검색 문장에서 정확한 검색어를 찾아낸다는 게 사실상 힘들다는 것을 알 수 있다. 반면 온톨로지를 이용할 경우 검색 문장을 의미적으로 해석하여 정확한 검색어를 찾아낼 수 있고 또한 클래스, 유의어, 추론 등 온톨로지의 여러 기능을 이용하여 보다 정확한 상품을 검색할 수 있는 장점이 있다.

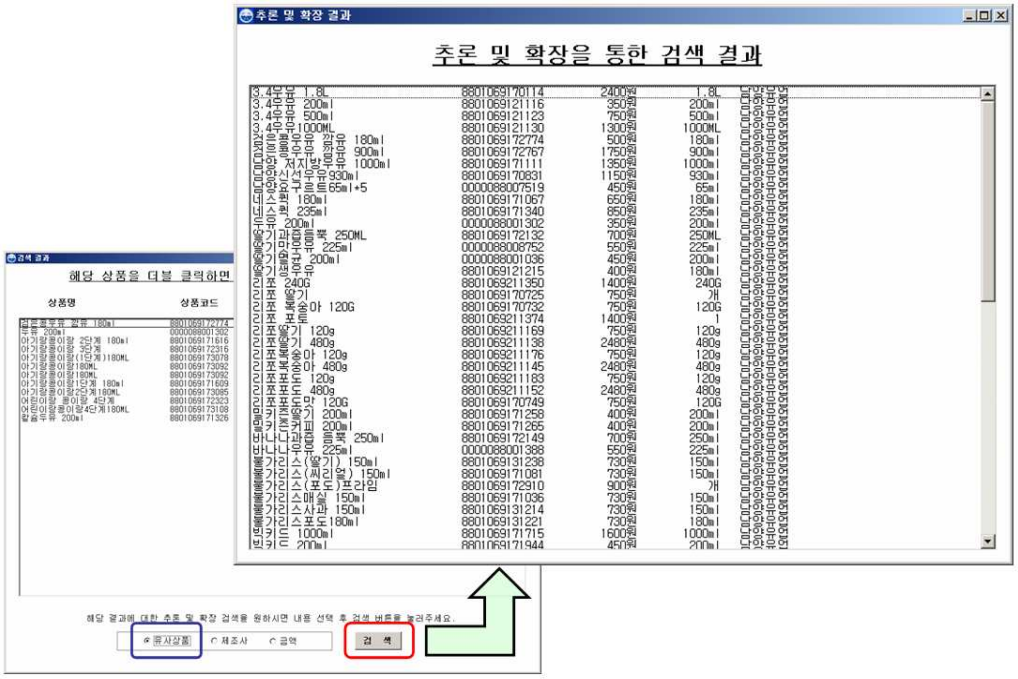
D. 추론기능을 이용한 상품의 확장 검색

【그림 4-15】 온톨로지를 이용한 검색 결과

온톨로지를 이용한 분류별 검색과 자연어 검색을 통해 사용자는 원하는 상품을 보다 정확하게 검색할 수 있으며 또한 추론기능을 이용하여 검색된 결과와 관련이 있는 유사한 상품들을 확장시켜 추가적으로 검색할 수도 있다.

【그림 4-15】는 “남양에서 생산한 1000원 이하의 두유 제품은 무엇이 있는가?” 라는 문장으로 검색한 결과이며 해당 검색 결과에서 온톨로지의 추론기능을 이용하여 유사상품들을 추가적으로 검색한 결과를【그림 4-16】에서 보여주고 있다.

확장 검색 결과를 보면 남양에서 생산한 제품들 중에서 두유 제품과 유사하다고 판단할 수 있는 우유, 요구르트, 유산균음료 등 모든 유제품들로 확장하여 검색한 것을 확인할 수 있다.



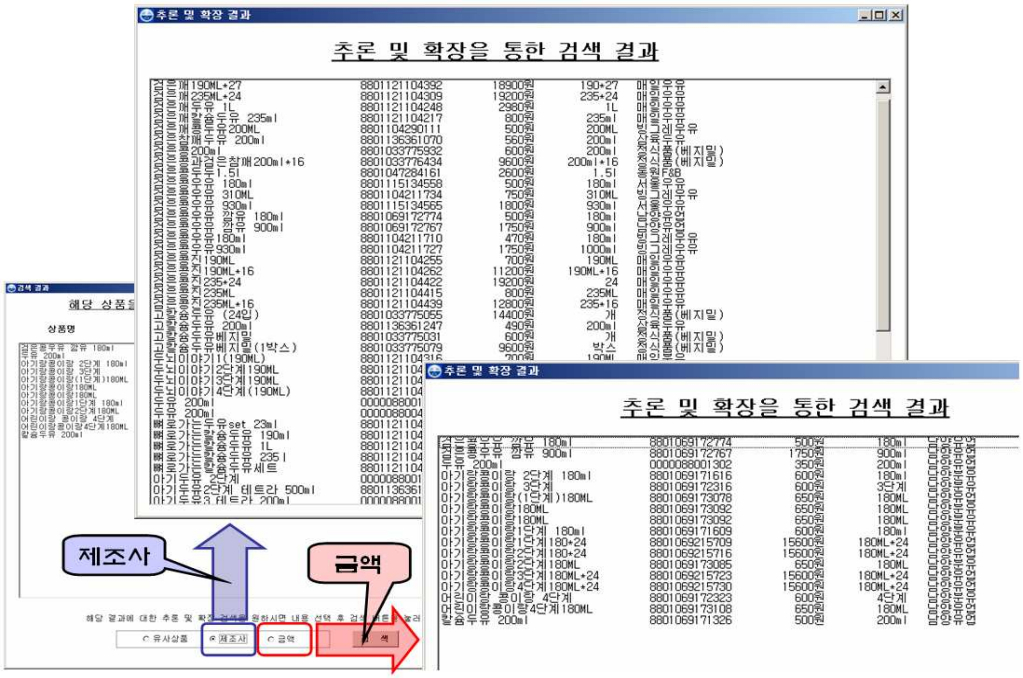
【그림 4-16】추론기능을 이용한 유사상품 확장 검색 결과

추론기능을 통한 검색 방법은 온톨로지에 구축된 클래스 구조를 이용하여 해당 결과에 대한 클래스를 추론함으로써 유사상품들을 확장 검색하게 된다.

“두유 제품”을 포함하고 있는 클래스의 상위 클래스를 찾아내고, 그 상위 클래스에 속해 있는 하위 클래스들을 추론한다. 이렇게 추론된 클래스에 속해 있는 모든

상품들을 대상으로 남양에서 생산한 상품들만을 검색함으로써【그림 4-16】과 같이 남양에서 생산한 제품 중에서 두유 제품과 관련이 있는 제품 즉, 모든 유제품들로 확장하여 검색하게 되는 것이다.(“【표 4-2】유제품 관련 하위 클래스” 참조)

또한 추론기능을 이용하면 유사상품 뿐만 아니라 제조사나 금액과 관련된 상품들도 확장 검색이 가능하다.



【그림 4-17】추론기능을 이용한 제조사 및 금액 확장 검색 결과

【그림 4-17】은 추론기능을 이용하여 제조사 및 금액과 관련된 확장 검색을 진행한 결과를 나타내고 있다. 이 때 제조사 확장 검색은 제조사 도메인 온톨로지의 클래스를 이용하여 추론하게 된다.

최상위 클래스	상위 클래스	하위 클래스	
제 조 사	유제품 제조사	남양유업	서울우유
		해태유업	정식품
		매일우유	삼육두유
		롯데우유	동원F&B
		빙그레	대상
		CJ	· · ·
	스넥 제조사	해태제과	롯데제과
		오리온	· · ·
	· · ·	· · ·	· · ·

【표 4-4】 제조사 도메인 온톨로지 클래스 일부

【표 4-4】는 제조사 도메인 온톨로지의 클래스 일부를 나타내고 있다.

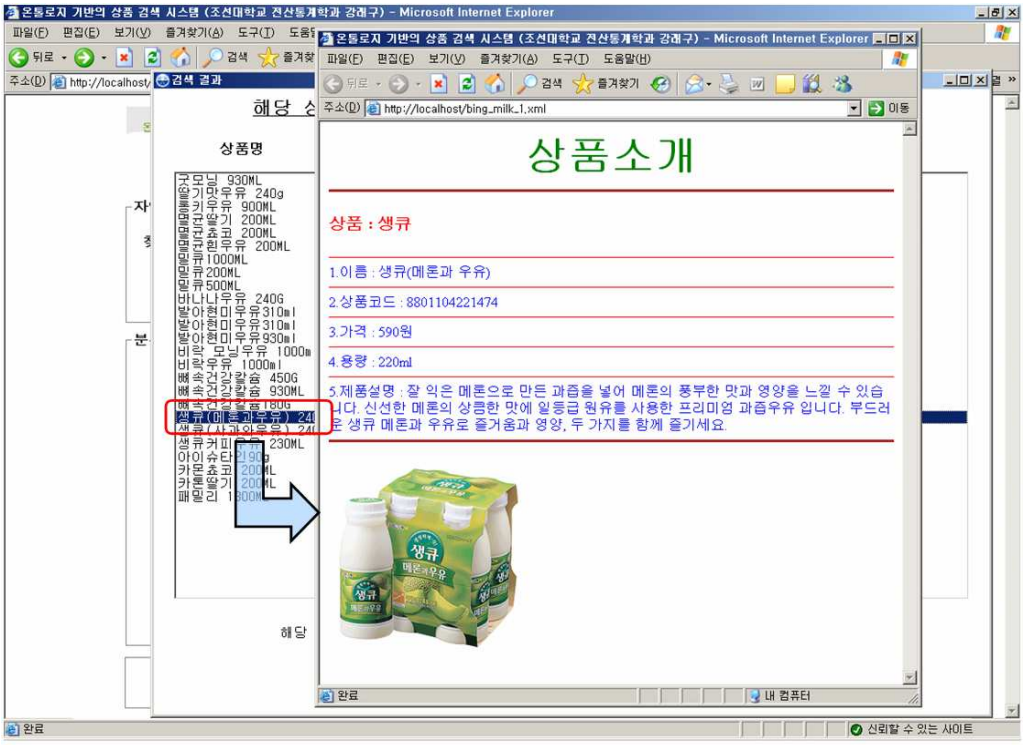
"남양에서 생산한 1000원 이하의 두유 제품은 무엇이 있는가?"라는 검색 문장에 대한 제조사 확장 검색을 진행하기 위해서는 제조사 도메인 온톨로지를 이용하여 검색해야 한다. 먼저 "남양"을 포함하고 있는 상위 클래스를 찾아낸다.

【표 4-4】에서처럼 "남양"은 "유제품 제조사"라는 상위 클래스에 속해 있고 "유제품 제조사"는 국내에서 유통 중인 유제품을 생산하는 모든 제조사들을 하위 클래스로 포함하고 있으며, "유제품 제조사"에 속해 있는 하위 클래스의 모든 제품들을 대상으로 "두유"와 관련된 상품들만 확장하여 검색하게 된다.

금액관련 확장 검색은 "1000원 이하"라고 한정해 놓은 금액 관련 검색어를 모든 금액으로 확장하여 검색하였다.

이와 같이 추론기능을 이용하여 거래처 및 금액을 확장 검색한 결과가【그림 4-17】이다.

지금까지 살펴본 분류별 검색과 자연어 검색을 통해 검색된 결과 중에서 원하는 상품에 대한 상세한 정보를 얻고자 할 때 해당 상품을 더블클릭하면【그림 4-18】에서처럼 해당 상품에 대한 상세한 정보를 XML문서로 제공하고 또한 해당 제조사 사이트로 연결이 가능하도록 추가적으로 설계하였다.



【그림 4-18】XML문서로 제공된 상품의 상세 정보

V. 결론 및 향후 과제

Tim Berners-Lee에 의해 처음 제안된 월드와이드웹은 HTML을 이용하여 쉽고 간편하게 사용할 수 있도록 편리성을 추구하였고, 그 결과 정보량의 폭발적인 증가를 가져왔다. 이처럼 사용하기 편리하다는 장점이 있는 반면 많은 양의 정보가 폭발적으로 증가하는 현 상황에서 사용자가 원하는 정보를 정확하게 검색하지 못한다는 문제점 또한 갖고 있다.

지금까지의 검색 방법으로는 찾고자 하는 정보를 검색할 때 주로 단어의 빈도수나 어휘 정보를 이용하는 키워드 기반의 검색이 주로 쓰이고 있다. 즉 사용자가 검색어로 사용하고자 하는 용어와는 관계없이 웹 사이트 개발자나 데이터베이스 설계자가 일방적으로 정해 놓은 키워드로만 검색할 수 있다는 문제점이 있었다. 이 때문에 사용자가 원하는 정보를 찾기 위해서는 찾고자 하는 정보에 대한 정확한 키워드를 알아야만 하는 문제점이 있고, 정확한 키워드를 통해 검색을 진행했다 하더라도 키워드만 일치하면 동형이의어나 동음이의어처럼 같은 단어로 쓰이지만 실질적인 뜻이 다른 문서들까지도 동일한 결과로 나타내는 경우가 많이 있다.

이로 인해 사용자는 제시된 결과를 한번 더 수동으로 검색을 해야 하는 부담을 안게 되었다.

HTML 기반의 웹은 데이터에 의미를 부여하여 사용자가 원하는 정보를 자동으로 추출하고 정보와 그 정보의 속성을 나타낼 수 있는 방법이 없기 때문에 검색시 위와 같은 문제점이 발생하는 것이다.

이러한 문제점을 해결하기 위해 1999년 W3C에서는 시맨틱 웹을 제안하였고 그 중심에 온톨로지가 있다.

온톨로지는 흔히들 지식베이스라고 부르기도 한다. 하지만 엄격히 말해 온톨로지는 용어 사이의 개념적 의미를 정의하기 때문에 지식베이스와 구별이 되는 또 다른 형태의 데이터베이스라고 할 수 있다.

온톨로지는 단순히 개념의 의미만을 정의하는 것은 아니다. 개념 클래스와 관련된 관계, 연산 함수, 지식의 추론 등 개념 인스턴스를 총체적으로 정의하여, 영역의 지식체계를 형식화하는 데 그 목적이 있다. 즉, 온톨로지는 지식 표현을 위한 형식적 체계인 것이다.

본 논문에서는 온톨로지를 상품 검색에 활용하기 위해 상품 도메인 온톨로지를 직접 구축하여 자연어 및 분류별 검색을 통해 얼마나 정확한 검색을 하게 되는지 실험하였다.

이 실험을 위해 Protégé 3.1을 이용하여 OWL로 온톨로지를 구축하였고, JSP와 PowerBuilder9.0을 이용하여 사용자 인터페이스를 만들어 자연어 및 분류별 상품 검색을 하였다. 그 결과, 온톨로지에 구축된 클래스 개념과 속성, 유의어, 추론 등 다양한 온톨로지 기능을 통해 사용자가 원하는 상품을 정확히 찾아내는 걸 확인하였고, 사용자가 예상치 못한 상품에 대해서도 추론을 통해 상품 범위를 확장시켜 검색을 하였다. 또한 온톨로지 기반의 상품 검색 시스템과 키워드 검색을 사용하고 있는 국내 최대 온・오프 할인점인 L할인점과 E할인점의 온라인 쇼핑몰에서 동일한 검색 조건으로 상품 검색을 진행한 결과 본 논문에서 구현한 상품 검색 시스템이 키워드를 기반으로 하는 검색 시스템보다 훨씬 정확한 검색 결과를 나타내고 있는 것을 확인하였다.

이처럼 온톨로지를 이용하여 검색할 경우 개념과 추론 등을 통해 스스로 검색어 자체에 의미를 부여함으로써 사용자가 원하는 상품을 보다 정확하게 검색할 수 있고, 사용자가 예상하지 못한 상품에 대해서도 검색이 가능하다는 장점이 있다.

매년 새로운 신상품이 다량으로 출시되고 있는 현 시점에서 초기에 구축해 놓은 온톨로지를 영구적으로 이용한다는 것은 불가능한 일이다. 정확한 상품 검색을 처리하기 위해서는 신상품과 새로운 분류의 상품들에 맞게 지속적으로 온톨로지를 수정 및 확장해 나가야 할 것이다. 또한 상품 데이터베이스 설계에 있어서도 상품들의 특징을 정확하게 분석하여 간단하면서도 명확하게 정의를 해 놓아야만 온톨로지를 통한 검색시 보다 빠르고 정확한 검색이 가능할 것이다.

따라서 온톨로지를 구축하고 효과적으로 이용하기 위해서는 지속적인 유지보수를 고려해야 할 뿐만 아니라, 상품 데이터베이스 설계에 있어서도 보다 간단하고 명확한 설계가 이루어져야 할 것이다.

참 고 문 헌

[1] 김홍기, 김학래, 이강찬, 정지훈, 이재호 외, “월드와이드 웹에서 시맨틱 웹으로,” 『마이크로소프트웨어』, pp.242-301, 2002.

[2] Guarino, N., “Formal Ontology and Information Systems,” Proceedings of FOIS'98, pp.3-15, 1998.

[3] Berners-Lee T., Hendler J. and Lassila, O., 2001, The Semantic Web, Scientific American.

[4] McIlraith.S.A, son T.C, and Honglei Zeng, Semantic web services, IEEE Intelligent Systems, Vol.16, No.2, pp.46-53, 2001.

[5] World Wide Web Consortium, http://www.w3.org.

[6] 최중민, “시맨틱 웹의 개요와 연구동향”,『정보과학회지』, 제21권 제3호, pp.4-10, 2003.

[7] 이은주, “시맨틱 웹을 위한 문서검색시스템”, 계명대학교 석사학위논문, 2003.

[8] Kuhanandha Mahalingam and Michael N. Huhns “An Ontology Tool for Query Formulation in an Agent-Based Context”, 1997.

[9] Robert Neches et. al. “Enabling Technology for Knowledge Sharing” AI Magazine, Winter, p.42, 1991.

[10] Rudi Studer, V. Richard Benjamins, and Dieter Fensel. “Knowledge Engineering : Principles and Methods.” Data & Knowledge Engineering. Vol.25, No.1-2, pp.184-185, 1998.

[11] Oscar Corcho, Mariano Fernández-López, and Asunción Gómez Pérez. OntoWeb Technical Roadmap v1.0. IST Programme of the Commission of the European Communities as Project No. IST-2000-29243, 2000.

[12] Robert Jasper and Mike Uschold. A Framework for Understanding and Classifying Ontology Applications. Seattle : Boeing math and Computing Technology, pp.9-11, 2000.

[13] Gerhard Fischer and Jonathan Ostwald. “Knowledge Management : Problems, Promises, Realities, and Challenges.” IEEE Intelligent System. January/February,

pp.60-62, 2000.

[14] Dieter Fensel et al. “On-To-Knowledge : Ontology-Based Tool for Knowledge Management.”, pp.1-2, 2000.

[15] Winfried Schmitz-Esser, “New Approaches in thesaurus Application.” International Classification, Vol.18, No.3, pp.143-146, 1991.

[16] Johann Gamper, Wolfgang Nejdl, and Martin Wolpers. “Combining Ontologies and Terminologies in Information Systems”, pp.12-13, 1999.

[17] 이현실, “합성 온톨로지 기반의 한의학 처방 지식관리 시스템”, 『한국학술정보』, 2006.

[18] 김홍기, 김학래, “생명이 없는 공간과 의사소통을 : 온톨로지”, 『(주)마이크로소프트웨어』, pp.268-277, 2002.

[19] Robin Cover, “XML and Semantic Transparency”, Cover pages, http://oasis-open.org/cover/xmlAndSemantics.html, November 1998.

[20] Jeff Helfin and James Hendler, “Semantic Interoperability on the Web”, Proceedings of Extreme Markup Languages 2000. Graphic Communications Association, Alexandria, VA, pp.111-120, 2000.

[21] “Connecting XML, RDF and Web Technologies for Representing Knowledge on the Semantic Web”, XML Europe 2002, Barcelona, May 2002.

[22] Stefan Decker, Dan Brickley, Janne Saarela, and Jurgen Angele, A Query and Inference Service for RDF, The Query Languages Workshop(QL'98), W3C, 1998.

[23] RDF Model and Syntax Specification, W3C Recommendation 22 February http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/

[24] http://www.w3.org/TR/REC-rdf-syntax/

[25] 한국전산원, “웹 온톨로지 개발 지침 연구”, 2004.

[26] WD-RDF-Schema Website, http://www.w3.org/TR/2002/WD-rdf-schema-20021112/

[27] WD-RDF-Schema Website, http://www.w3.org/TR/2004/REC-owl-features-20040210/

[28] 김영민, 이상춘, “시맨틱을 이용한 연구논문 검색 시스템”, 『인터넷 정보학회 논문지』 Vol.4 No.3, pp.15-22, 2003.

[29] 임수연, “전문용어의 의미관계 정보를 이용한 도메인 온톨로지의 구축”, 경북

대학교 박사학위 논문, 2004.
[30] Natalya F. Noy, Deborah L. McGuineww. “Ontology Development 101 : A Guide to Creating Your First Ontology”,http://protege.stanford.edu/publications/Ontology_development/Ontology101-noy-mcguinness.html.
[31] http://protege.stanford.edu.

부록 *1* - 상품 도메인 온톨로지의 클래스 수목도

Daily	유음료	유음료
		유산균음료
		두유
		우유
		버터
		마아가린
		치즈
	간식	호상
		젤리
		푸딩
		무스
	빵	브레드
		롤
	축육가공품	햄
		소시지
		베이컨
냉동식품	소재냉식	농산냉식
		수산냉식
	조리냉식	한식조리냉식
		양식조리냉식
		중식조리냉식
	아이스크림	소프트
		아이스크림
		아이스밀크
		프로즌요구르트
		빙과
		얼음
조미료	기초조미료	소금
		설탕
		향신료
		화학조미료
		풍미조미료
		참기름
	주요조미료	식용유
		된장
		식초
		맛술
		액젓
		식용유세트
		고추장
		쌈장
		물엿
		참깨
		고춧가루
	가공조미료	드레싱
		마요네즈
		케첩
		소스
		불고기양념
		카레
		짜장
조리가공품	건면	소면
		우동
		냉면
		평면
		당면
		중면

조리가공품	인스턴트면	봉지면
		컵면
	파스타	스파게티
		마카로니
	인스턴트스프	인스턴트스프
		인스턴스라이스
		인스턴트죽
	통조림	과일통조림
		농산물통조림
		수산물통조림
		축산물통조림
		참치캔
	이유식	과일
		야채
		혼합
		라이스
		스프
		분유
		베이비쥬스
	분말류	밀가루
		조리용가루
		베이킹가루
		쌀가루
		보리가루
		잡곡가루
		전분
		콩가루
		건미역
		김
		건포도

음료	탄산음료	콜라
		사이다
		기타탄산
		스포츠음료
		건강음료
		미네랄워터
	과즙음료	천연과즙
		기타과즙
	기타음료	스포츠음료
		미네랄음료
과자	주머니과자	캔디
		초코릿
		츄잉껌
		카라멜
		젤
	양식건과자	비스킷
		쿠카
		크래커
		웨하스
		종합세트
		파이
	스넥과자	콘계
		포테이토계
		밀가루계
		씨리얼
		후레이크
		비스켓
		군과자
		튀긴과자

과자	한식건과자	누른과자
		엿
	한식반생과자	양갱
		모나카
양곡	곡류	쌀
		보리
		잡곡
		콩
		찹쌀
		현미
		깨
일용품	세제	세탁세제
		주거세제
		부엌세제
		표백세제
		고체세제
		다림세제
	세면목욕제	화장비누
		샴푸린스
		트리트먼트
		바디용품
		치약
		칫솔
		유아샴푸
		헤어케어
		염색약
		면도제품
	방충방취제	방충제
		살충제
		방향제
일용품	방충방취제	방취제
		탈취제
		방습제
	위생용품	생리용품
		종이기저귀
		위생용품
	가정지	화장지
		티슈
		물티슈
		키친타올
	키친소모품	랩
		호일
		알미늄박성형
		취사장갑
		행주
		고무장갑
		일회용품
생활잡화	실내잡화	실내화
		정리대
		플라스틱류
		청소용품
		매트
		우산
	욕실잡화	락스
		트래밍
		타올
		세면용품
		샤워기
	주방잡화	컵

생활잡화	주방잡화	조리기구
		수세미
		플라스틱제품
		씽크용품
		스테인레스제품
	자동차용품	방향제
		오일
		왁스
		미러
		핸들커버
		시트
		핸즈프리
		와이퍼
		악세사리
		워셔액
		겨울용품
		코팅제
		세정제
	유아용품	젖병
		꼭지
		노리개
		유아칫솔
주류	국산주류	소주
		맥주
		청주
		전통주
		탁주
	수입주류	위스키
		브랜디
		와인

부록 2 - 불용어 리스트

1	가	28	들의
2	개	29	듯
3	거나	30	따라
4	건	31	때
5	걸	32	때문
6	것	33	라
7	게	34	라는
8	곳	35	로
9	과	36	로서
10	과의	37	로써
11	그	38	를
12	그외	39	마다
13	그이	40	만
14	그전	41	만큼
15	그쪽	42	맛
16	까지	43	며
17	나는	44	면서
18	네	45	몇
19	는	46	무엇
20	되	47	번
21	되는	48	번째
22	되다	49	부터
23	되도록	50	뿐
24	두	51	생산
25	둘	52	서
26	들	53	서로
27	들도	54	세

55 수
56 에
57 에게
58 에서
59 여럿
60 옆
61 와
62 원
63 으로
64 으로서
65 은
66 을
67 의
68 이
69 이것
70 이때
71 이라는
72 인가
73 있는가
74 잘
75 짜리
76 정도
77 제품
78 쪽
79 처럼
80 하
81 하나
82 하는
83 하다
84 하도록
85 한