

분산 이기종 환경에서 XML기반 메타데이터를 이용하는 통합 정보 검색 시스템

박성준^o 강지훈 김영국
충남대학교 컴퓨터학과
{sjpark, jhkang, ykim}@cs.cnu.ac.kr

An Integrated Information Retrieval System using Metadata based on XML in Distributed Heterogeneous Environments

Sung-Joon Park^o Ji-Hoon Kang Young-Kuk Kim
Dept. of Computer Science, ChungNam National University

요 약

인터넷 쇼핑몰 허브 사이트는 분산 이기종 시스템들로 구성되는 다수의 독립된 인터넷 상점들을 연합하여 공동 포인트 적립, 공동 상품 검색 등의 통합된 서비스를 제공하는 가상 공동체이다. 본 논문에서는 인터넷 쇼핑몰 허브 사이트에서 다수의 이기종 데이터베이스에 대한 검색 질의를 생성하고, 검색 결과를 통합하여 고객에게 제공하는데 필요한 메타데이터를 효율적으로 관리하기 위해 XML을 이용하는 분산 이기종 통합 정보 검색 시스템을 제안한다.

1. 서론

최근 인터넷의 보급이 폭발적으로 늘어남에 따라 인터넷 사용자들을 대상으로 온라인 상에서 상품과 정보, 그리고 기타 서비스를 팔려고 하는 인터넷 비즈니스 사이트들이 기하급수적으로 늘어나고 있다[1, 2]. 또한, 소규모 전문 인터넷 상점들과 서비스 제공 사이트들을 연합하여 각 사이트에서 확보한 회원들을 공유하여, 공동 마케팅, 공동 포인트 적립 등의 통합된 서비스를 제공하는 허브(Hub) 사이트가 등장하고 있다[4]. 이러한 허브사이트의 특징은 이에 참여하는 각 가맹점들이 지역적으로 분산되어 있으며, 서로 다른 플랫폼 위에서 각 가맹점 고유의 시스템을 독립적으로 구축 운영함으로써 자율성을 제공한다. 이와 같은 분산 이기종 환경에서 통합된 상품 검색 기능을 제공하기 위해서는 각 가맹점의 인터넷 주소, 데이터베이스 명, 테이블 명, 그리고 테이블을 구성하고 있는 각 필드들에 대한 메타 정보를 허브 사이트에 유지하면서, 고객의 상품 검색 요청 시 각 가맹점에 동시에 검색 질의를 할 수 있어야 한다.

본 논문에서는 인터넷 쇼핑몰 허브 사이트에서 다수의 이기종 데이터베이스에 대한 검색 질의를 생성하고, 검색 결과를 통합하여 고객에게 제공하는데 필요한 메타데이터를 효율적으로 관리하기 위해 XML을 이용하는 분산 이기종 통합 검색 시스템을 제시한다.

본 논문의 구성은 제2장에서 관련연구에 관하여 기술하고, 3장에서는 시스템 구성도 및 통합 검색 시나리오를 기술한다. 4장에서는 시스템 설계 및 구현 결과에 대하여 기술하고, 마지막으로 5장에서 결론 및 향후 계획에 관하여 기술한다.

2. 관련연구

분산 이기종 통합 검색 방법은 클라이언트/서버 구조를 기반으로 한다. 클라이언트/서버 구조는 2단계 클라이언트/서버 구조와 3단계 클라이언트/서버 구조로 분류한다. 2 단계 클라이언트/서버 구조는 클라이언트에서 사용자와 인터페이스 하기 위한 화면 운용(표현 규칙의 운용 : Presentation Rule), 업무 규칙(Business Rule)의 구현, 그리고 자료의 접근 규칙(Data Access Rule)의 운용에 대한 작업을 맡고, 서버는 주로 데이터를 관리하는 작업을 맡는 방식이다. 이 방식은 클라이언트에서 서버 접근 정보 및 검색 질의를 위한 모든 정보를 관리하여야 한다. 3단계 클라이언트/서버 구조는 사용자와 인터페이스 하기 위한 화면 운용을 클라이언트 시스템이 담당하고, 업무 규칙의 구현과 자료의 접근 규칙의 운용을 서버 시스템에서 담당하는 방법으로 서버와 클라이언트가 작업을 분담하는 형태로 시스템을 운용하는 방식이다.

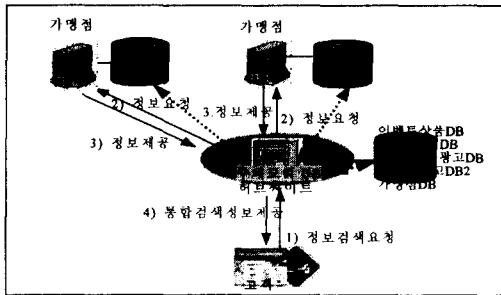
본 논문에서 제안하고 있는 프로토타입 시스템은 새로운 가맹점이 허브사이트에 추가될 때, 가맹점이 가지고 있는 기존 머천트 시스템을 수정하지 않고 그대로 이용할 수 있어야 한다는 요구 조건을 만족시키기 위해 2단계 클라이언트/서버 구조를 따른다.

따라서, 고객의 검색 요청을 분석하고 각 가맹점 사이트에 존재하는 상품정보 데이터베이스에 접근하여 검색을 수행한 후, 고객에게 통합된 검색 결과를 제공하기 위해서는 각 가맹점에 대한 자료 접근 규칙을 관리하고 연결을 유지하기 위한 메타데이터를 클라이언트 사이트에 유지할 필요가 있다.

3. 분산 통합 검색시스템 구성도

[그림 1]은 본 논문에서 제안하고 있는 인터넷 쇼핑몰 허브 사이트를 위한 통합 검색 서비스를 제공하기 위한 시스템 구성도

이다. 물리적 구성 요소로는 서로 다른 지역에 분산되어 있으며, 이기종 데이터베이스 환경을 가지는 여러 종류의 다양한 소규모 전문 인터넷 상점인 가맹점 사이트, 허브 사이트, 그리고 각 가맹점 또는 허브 사이트를 방문하여 광고를 선택하거나, 상품 정보를 검색 또는 구매 행위를 하는 고객들로 구성된다. 허브 사이트는 각 가맹점 페이지의 일정 영역에 고객의 특성에 따라 개인화 맞춤 서비스를 제공할 수 있으며, 고객으로부터 상품명, 회사명, 카테고리, 가격 등과 같은 키워드 검색을 요청 받으면, 분산 이기종 환경의 다수 가맹점 사이트로부터 통합된 검색 정보를 실시간으로 제공한다.



[그림1] 분산 통합 검색 시스템 구성도

정보 검색 기능을 제공하기 위한 동작 시나리오는 다음과 같다.

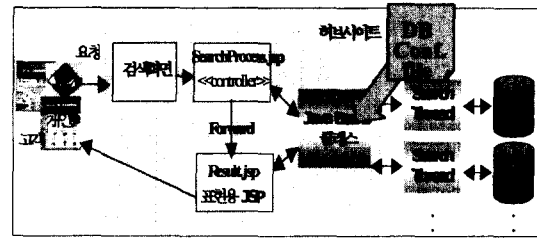
- 1) 허브 사이트에서는 고객의 검색 요청을 받아 각 가맹점에 동시에 검색을 요청하기 위해 쓰레드를 이용하며, 검색 요청에 따른 각 가맹점에 보유하고 있는 다수의 이기종 데이터베이스에 대한 질의를 생성하기 위해 메타데이터를 유지한다.
- 2) 각 가맹점 사이트에서는 허브 사이트로부터 질의 요청을 받아 질의를 수행한다.
- 3) 각 가맹점으로부터 수행된 검색 결과를 허브 사이트에 전달한다.
- 4) 허브 사이트는 각 가맹점으로부터 제공받은 검색 결과를 통합하여 고객에게 보여준다.

본 논문에서는 상기 동작 시나리오 중 첫 번째에서 메타데이터를 유지하기 위한 방법으로 기존의 텍스트 기반 통합 검색 시스템의 문제점을 보완하고자 메타데이터를 실시간으로 관리하기 위한 기법으로 XML 기반 통합 검색 시스템을 설계하고 구현하였다. 텍스트 기반 통합 검색 시스템은 허브 사이트에 가맹점이 새로 추가되거나, 각 가맹점 DB 정보가 변경되는 경우, 허브 사이트에 존재하는 구성(Configuration) 파일에 실시간으로 반영하기 위해서는 가맹점 관리자가 허브사이트 관리자에게 연락을 취해야만 하는 문제점을 가지고 있다. 반면 제안하고 있는 XML 기반 통합 검색 시스템은 DB 정보 변경 시 각 가맹점 관리자가 허브사이트에 접근하여 직접 변경 사항을 XML 파일에 적용함으로써 실시간으로 메타데이터 정보를 반영할 수 있으며, 각 가맹점 사이트로 하여금 정보 누출에 대한 염려를 해소할 수 있다.

4. 시스템 설계 및 구현 결과

4.1 Text 기반 통합 검색

[그림 2]는 고객이 허브 사이트를 방문하여 상품 검색을 하는 경우 서비스를 제공하기 위한 구조도이다.

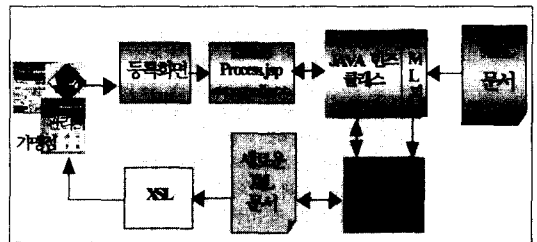


[그림 2] 허브 사이트 상품 통합 검색 제공 구조

허브 사이트에서 상품 검색을 제공하는 단계는 사용자의 입력을 받아들이는 허브 사이트측의 JSP Search 프로세스 부분, 실제 비즈니스 로직이 있는 JSP Java Bean, JSP Bean에 포함되어 검색 기능을 하는 Thread class, 각 가맹점의 데이터베이스로 5단계로 나눌 수 있다. 사용자가 검색항목과 키워드를 입력하고 검색 버튼을 누르면, 이 내용을 SearchProcess JSP 프로세스로 넘어가고, JSP 프로세스에서는 입력된 검색항목과 키워드를 파라미터로 하여 SearchHandler JSP Java Bean에 전달한다. SearchHandler는 Text 기반으로 서버에 기록되어 있는 각 가맹점의 데이터베이스에 대한 정보를 읽어 온 후, JSP로부터 전달된 검색항목과 키워드에 대한 정보를 바탕으로 SearchThread들을 생성하여, 각 SearchThread에 검색 명령을 내리고, SearchThread는 SearchHandler에게서 받은 정보를 바탕으로 각 nShop의 데이터베이스에 연결하고 지정된 질의를 이용하여 데이터베이스로부터 데이터를 얻는다. SearchHandler는 검색 명령을 내린 후 SearchThread들로부터 얻은 결과들을 조인하여 검색 결과 페이지를 생성한 후 이 문서의 결과를 사용자에게 제공한다. 이 때 JSP Bean은 검색 결과를 저장하고 있다가, 사용자가 정렬 명령을 내리면 결과들을 정렬하여 페이지를 새로 생성한 후 결과를 제공한다.

4.2 XML 기반 통합 검색

[그림 3]은 메타데이터에 대한 정보를 저장하기 위한 방법으로 DOM API를 이용하여 데이터베이스에 대한 정보를 기존 XML 문서에 추가, 변경, 삭제하기 위한 구조도이다.



[그림3] XML 문서 접근, 추가, 변경, 삭제를 위한 구조도

각 가맹점으로부터 상품 검색에 필요한 메타데이터 정보 변경 또는 새로운 가맹점 추가 시, 각 가맹점 관리자로부터 직접 상품 검색에 필요한 관련 정보를 허브 사이트의 등록 화면에 입력하게 되면 DbInfoProcess JSP 프로세스로 넘어가고, JSP 프로세스에서는 입력된 정보를 파라미터로 자바 빈즈 클래스로 전달한다. 자바 빈즈 클래스에서는 기존의 XML 문서가 있는 경우 XML 파서를 이용하여 DOM 객체를 메모리에 생성한 후, 추가,

변경, 삭제 요구를 반영한 새로운 XML 문서를 생성한다. 또한 새로 생성된 XML 문서에 대한 정보를 가맹점 관리자가 확인할 수 있도록 XML 스타일 시트를 이용하여 새로운 XML 문서의 내용을 브라우저에 제공한다. 일단 XML 문서에 메타데이터 관련 정보가 반영되면 고객의 상품 검색 요구 시 통합 검색을 제공하기 위한 구조는 Text 기반 통합 검색 구조와 유사하다. 다음은 XML 문서의 내용을 보여주고 있다.

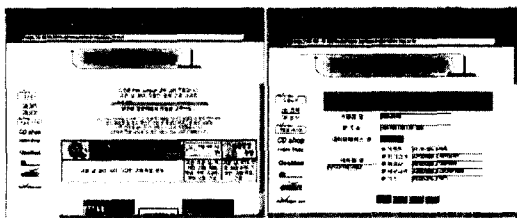
```
<?xml version="1.0" encoding="euc-kr"?>
<?xml:stylesheet type="text/xsl" href="dbinf.xsl"?>
<DBCONF>
  <NSHOP>
    <SHOPNAME>CDShop</SHOPNAME>
    <IPADDRESS>168.188.128.128</IPADDRESS>
    <DBNAME>oracle</DBNAME>
    <TABLE>
      <TABLENAME>CDGOODSTABLE</TABLENAME>
      <GOODSNAME>CDGOODS_NAME</GOODSNAME>
      <CATEGORY>CDGOODS_DESCRIPTION</CATEGORY>
      <COMPANY>CDGOODS_COMPANY</COMPANY>
      <DESCRIPTION>CDGOODS_DESCRIPTION</DESCRIPTION>
      <PRICE>CDGOODS_PRICE</PRICE>
    </TABLE>
  </NSHOP>
</DBCONF>
```

다음은 DOM API를 통해서 XML 문서의 SHOPNAME 엘리먼트를 접근하기 위한 방법과 <NSHOP> 엘리먼트 뒤에 <SHOPNAME>이라는 엘리먼트를 만들고, AMAZON 이라는 Text 노드를 추가하는 예이다.

```
/*SHOPNAME이라고 이름 붙여진 엘리먼트와 그 자식노드들을
node_list라고 선언하고 하고 있다.*/
NodeList node_list = rootnode.getElementsByTagName("SHOPNAME");
/*텍스트노드를 화면에 뿌려준다.*/
String shopname = node_list.item(0).getFirstChild().getNodeValue();
System.out.println("가맹점 : " + shopname);
NodeList add_node_list = doc.getElementsByTagName("DBCONF");
add_node_list.item(add_node_list.getLength()-1).appendChild(addEle);
Element addEle = doc.createElement("SHOPNAME");
Text addText = doc.createTextNode("AMAZON");
addEle.appendChild(addText);
add_node_list = doc.getElementsByTagName("NSHOP");
add_node_list.item(add_node_list.getLength()-1).appendChild(addEle);
```

4.3 구현 결과

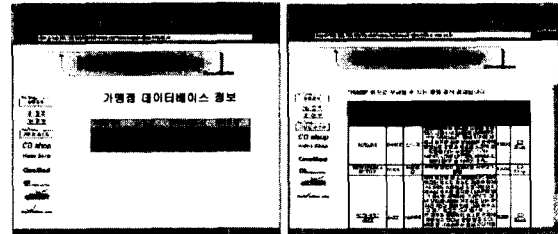
본 논문에서 제안한 시스템을 구현하기 위한 구현 환경은 운영체제로 Linux, Windows NT 기반에 데이터베이스로 Oracle과 MySql을 이용하였다. 구현 언어로는 Java와 JSP를 이용하였으며, XML DOM 파서로 IBM4J를 이용하였다.



[그림 4] 허브 사이트 개인화 및 가맹점 DB 정보 등록 화면

[그림 4]는 허브 사이트 회원이 허브 사이트를 방문한 경우

고객의 특성에 따른 개인화 맞춤 광고를 제공하는 초기 화면과 각 가맹점에서 데이터베이스 관련 정보 변경 시 XML 파일에 이를 반영하기 위한 등록 화면이다. 가맹점 관리자는 등록 화면에서 가맹점 명, IP 주소, 데이터베이스 명, 그리고 상품 관련 데이터 명 및 관련 필드 명을 입력하고 추가 버튼을 선택하면 새로운 XML 파일에 반영된다.



[그림 5] XML 파일 정보 및 상품 정보 검색 결과 화면

[그림 5]는 데이터베이스 등록 화면을 통해 새로운 XML 파일을 생성 한 후, 결과를 확인하기 원하는 경우 새로운 XML 파일의 내용 정보를 제공하는 화면과 고객이 상품 검색을 요청한 경우 XML 파일을 기반으로 통합 검색한 결과이다.

5. 결론 및 향후 계획

이제까지 본산 이기중 환경의 인터넷 쇼핑몰 허브 사이트에서 각 가맹점 사이트에 존재하는 상품들에 대한 통합된 검색 정보를 제공하기 위한 방법으로 XML을 기반으로 하는 시스템의 설계 및 구현 결과를 살펴보았다.

본 논문에서 제시한 XML 기반 통합 검색 시스템은 Text 기반 통합 검색 시스템에 비하여 각 가맹점 데이터베이스 구성 정보에 대한 추가 변경 시 실시간으로 반영할 수 있으며, 허브 사이트에서 검색에 필요로 하는 정보를 각 가맹점에서 직접 입력하도록 함으로써 정보 노출에 대한 불안감을 해소할 수 있다.

향후 연구 과제로는 각 가맹점 요구 사항에 맞는 개인화 페이지를 제공하기 위한 연구와 동시에 여러 고객이 검색을 요구하는 경우 성능 향상을 위한 연구가 필요하다. 또한 본 시스템은 프로토타입 시스템으로 상업용 사이트에 적용하기 위한 연구가 필요하다.

참고문헌

- [1] Ee-Peng Lim, Wee-Keong Ng, "An Overview of the Agent-Based Electronic Commerce System(ABECOS) Project", Bulletin of the Technical Committee on Data Engineering, Vol. 23, No. 1, Mar. 2000
- [2] G.-H. Yan, W.-K. Ng, and E.-P. Lim, "Toolkits for a Distributed, Agent-Based Web Commerce System", In International IFIP Working Conference on Trends in Distributed Systems for Electronic Commerce(TrEC '98), Hamburg, Germany, June 1998.
- [3] Allen Cliff, Kania Deborah, and Yaeckel Beth, Internet World Guide to One-to-One Web Marketing, John Wiley & Sons, Inc., 1998
- [4] Wholesale Hub, <http://www.wholesalehub.com>
- [5] Richard Anderson et al., "Professional XML", Wrox Press Ltd., 2000
- [6] W3C, "DOM Level 1 Specification", <http://www.w3.org/TR/1998/REC-DOM-Level-19981001>