

지식공유 기반의 XMDR을 이용한 검색 시스템 설계

황치곤^o 이민노 박유신 정계동 최영근

광운대학교 정보통신대학원 정보통신학과

{duck1052^o, oceanil, gdchung, love0840, ygchoi}@kw.ac.kr

Design of Retrieval system using XMDR based knowledge sharing

Chigon Hwang^o, Minnoh Yi, Yooshin Park, Gyedong Jung, Youngkeun Choi

Dept. of information communication, Kwangwoon Univ. graduate school of information communication

요 약

최근 대부분 기업들 환경에서의 정보 시스템들은 지역적으로 분산되어 있으며 다양한 형태로 구성되어 있으므로, 사용자 의사 결정을 지원하는데 필요한 통합된 정보를 얻는 것은 어려운 일이다. 따라서 이러한 문제를 효율적으로 정보 검색에 적용하기 위해 사용자에게 단일 인터페이스를 제공하고, 이기종 시스템들 간에 구축된 데이터베이스 시스템들은 각각 독립성을 유지하면서 하나의 인터페이스처럼 투명성을 제공할 필요성이 있다. 이를 위해 ISO/IEC 11179에서 연구 중인 XMDR의 개념을 이용하여 정보검색에서 발생하는 "의미적 상호운용성(semantic interoperability)"이라는 문제점을 해결하고, 이 XMDR에 지식 인스턴스 계층을 통한 지식공유를 가능하게 함으로써 단순 검색의 한계점을 극복할 수 있도록 한다. 본 논문에서는 데이터 표현에 사용되는 명칭, 속성, 관계성에 대한 이질적인 문제를 해결하기 위한 표준 온톨로지, 각 레거시 시스템을 연결하는 중간자(mediation)역할을 수행하는 로케이션 온톨로지, 지식공유가 가능하도록 하는 지식 인스턴스 계층으로 구성하는 방법을 제안한다. 또한 지식 인스턴스 계층은 협업적인 검색 환경 하에서 각각의 정보시스템에서 다양한 형태의 지식을 공유 및 통합에 있어 구조화 되지 않은 지식들을 어떻게 공유할 것인가에 대한 개념적인 모델을 제시한다.

1. 서 론

최근 대부분 기업들 환경에서의 정보 시스템들은 지역적으로 분산되어 있으며 다양한 이기종의 소스들로 구성되어 있으므로, 사용자 의사 결정을 지원하는데 필요한 통합된 정보를 얻는 것은 어려운 일이다. [1] 따라서 이러한 문제를 효율적으로 정보검색을 하기위하여 사용자에게는 단일 인터페이스를 제공하고 이기종 시스템들 간에 구축된 데이터베이스 시스템들은 각각 독립성을 유지하면서 하나의 인터페이스처럼 투명성을 제공할 수 있다. 그러나 정보 검색에 있어서 "의미적 상호운용성(semantic interoperability)"이라는 문제가 대두되고 있으며, 상호운용성을 위한 의미를 정의하는데 있어 기본적인 접근은 메타데이터를 이용하는 것이다. 온톨로지는 용어의 사용이 다르고 개념들 간의 관계적 논리구조가 다르다고 할지라도 정보의 처리방식이 단순한 패턴매칭이 아닌 내용 중심적이기 때문에 정보의 통합 공유가 가능하다. 즉 온톨로지가 제공하는 의미정보와 관계성을 표현함으로써 웹상에서 좀 더 나은 의미적 상호운용성을 확보할 수 있다. [2]

더불어 지식공유를 위해 지식 관리 시스템에서 사례 기반 추론(Case-Based Reasoning: CBR)이 널리 이용되고 있다. [3] CBR은 사용자가 입력한 적은 특성에 의존한 추론으로 적절한 추천을 하지 못하고, 사례 기술에 제약을 받으며, 사례 검색에서 사용자가 요구하는 결과를 의미에 맞게 검색하지 못한다. 또한 용어의 특성을 기술 할 수 없고, 용어와 용어 사이의 관계를 기술 할 수 없기 때문에, 사용자의 유연성이 없고, 의미상으로 정확한 검색 결과를 제공하지 못한다는 단점이 있다. [4][5]

2005년 1월부터 ISO/IEC 11179에서는 XMDR(eXtended Meta-Data Registry)에 대한 프로젝트가 수행되고 있다. 이 XMDR은 데이터 요소들, 용어법(Terminology), 그리고 MDR을 등록하고 검색하는 것을 더욱 증가시키고, 메타데이터의 다양한 타입들과 의미론적 명세와 질의를 수행할 수 있도록 확대된 능력을 지원하기 위해 MDR 표준인 ISO/IEC 11179 계열의 확장을 목적으로 한다. 이것은 구조화된 메타데이터에 카테고리 분류, 온톨로지 그리고 시소러스 등의 개념을 결합한 것이다. [5]

본 논문에서는 기존 기업이나 업체가 구성한 레거시 시스템을 있는 그대로 유지하면서 통합 검색할 수 있는 시스템을 ISO/IEC 11179에서 논의되고 있는 XMDR의 개념을 도입하여 제안하고자 한다. 이 시스템은 업무용 구분하여 공통된 용어를 정의한 카테고리 계층, 각 업무에 따른 상품 사이의 관계 또는 상품을 구입과 관련된 지역 사이의 관계

를 표현한 지식 인스턴스 계층, 각 레거시시스템의 데이터베이스에 표현된 데이터의 이질성을 해결하기 위한 표준을 제공하는 표준 온톨로지 및 레거시 시스템들에 접근하기 위한 로케이션 온톨로지로 구성된 XMDR 계층으로 구성된 검색시스템을 제안한다.

또한 협업시스템의 검색 환경 하의 각 정보시스템은 다양한 형태로 지식을 관리하고 있다. 이러한 지식을 공유하고, 통합하기 위한 방안으로 개념적인 도메인 모델을 이용한 XMDR 모델을 제시한다. 따라서 협업시스템의 각 데이터베이스와 데이터웨어하우스 간의 통합을 위한 XMDR을 도입함으로써 사용자에게 데이터의 위치 투명성을 제공함으로써 정보 검색의 개선된 방안을 제시하고자 한다.

2. XMDR

본 논문에서는 XMDR을 바탕으로 협업시스템에서 지식공유를 위한 XMDR 모델을 제시한다. 제시된 모델은 그림 1에서와 같이 5개의 계층으로 구성된다.

2.1. 카테고리 계층(Category Layer)

카테고리 계층은 업무 구분에 따른 용어들에 대한 정의들로 사용자가 필요한 상품과 서비스들에 대한 용어 및 의미가 정의되는 계층으로, 공유되는 용어들의 집합이다. 이 계층에서 제시된 용어는 협업시스템에서 사용되는 용어들에 대한 표준으로 각 협업시스템에서 공유되고 사용자에게 제시되는 용어들의 집합이다.

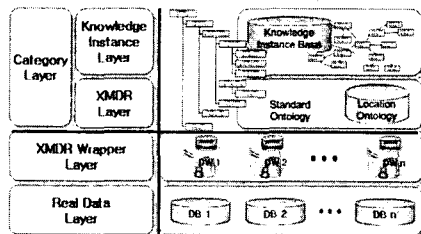


그림 1 XMDR 모델

2.2. 지식 인스턴스 계층(Knowledge Instance Layer)

이 계층은 협업시스템에서 상품 간의 관계 또는 서비스가 제공되는

지역 간의 관계성을 정의한 계층이다. 즉, 실제 데이터 사이의 관계성을 정의한 것이다. 검색하고자 하는 상품이나 서비스에 대한 요구를 만족시키지 못할 경우 관계성을 분석함으로써 사용자가 요구하는 상품과 유사하거나 대체 가능한 상품 또는 서비스가 존재하는 지역에 관한 정보를 제공한다. 이 계층의 생성은 레거시 시스템의 DW(DataWarehouse)에서 관리하는 데이터를 수집하고, 수집된 데이터의 분석을 통해 생성하므로 실제 사용되는 데이터들의 관계성을 파악하여 이를 사용자의 검색에 제공한다.

정의된 연관관계

- 동등관계(Equal-Bridge) - 표현은 다르지만 의미상 동등한 의미를 가지는 관계를 표현(제품명, 상품명, 품명 등 같다)
- 상속관계(Is-Bridge) - 분류 계층상 직접 연결되어 상속된 하위 개념을 표현(오리와 거위는 물새라는 개념을 상속받는다.)
- 부분관계(Part-Bridge) - 관계상 하위개념에 상위개념의 일부가 되는 관계를 표현(시, 군, 구는 도에 포함되는 관계)
- 대체관계(Substitute-Bridge) - 대체되어 사용될 수 있는 관계를 표현(상품정보에서 볼펜과 연필은 대체관계)

지식 인스턴스 계층에서 정의된 관계와 카테고리 계층을 여행 정보를 예제로 그림 2와 같이 표현할 수 있다.

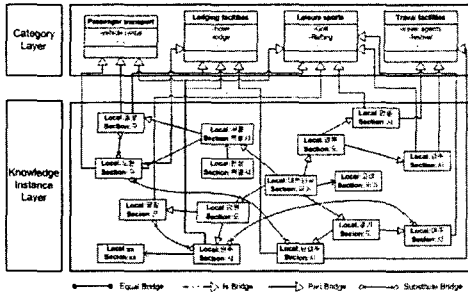


그림 2 카테고리 계층과 지식 인스턴스 계층의 관계 표현

그림 2은 사용자의 요구에 해당하는 상품 또는 서비스 정보가 카테고리 계층에 존재하지 않을 경우 검색키워드의 관계성을 파악한다. 파악된 관계를 사용자에게 제공함으로써 원하는 정보를 검색할 수 있도록 지원한다. 이를 통해 여행 정보 검색 서비스를 제공하는 시스템에서는 원하는 지역에 서비스가 존재 하지 않을 경우 그 지역과 관련된 인근 지역에 있는 시설까지도 검색할 수 있고, 상품 정보를 제공하는 시스템에서는 상품간의 유사성이나 대체할 수 있는 상품을 검색할 수 있도록 상품의 관계성을 구성한 계층이다.

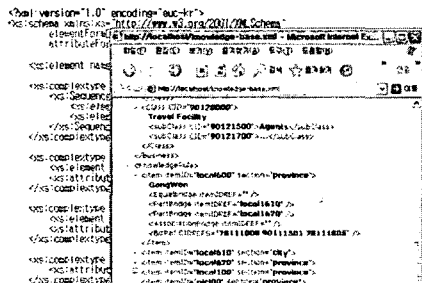


그림 3 XML로 표현된 지식 베이스

이 계층을 XML로 표현하기 위해서 그림 2과 같이 XML Schema 문서에 의해서 유효성을 보장하는 XML로 표현한다.

2.3. XMDR 계층(XMDR Layer)

XMDR은 가입된 레거시 시스템의 데이터베이스 정보를 수집하여 온톨로지를 구성하고, 서비스나 상품에 대한 관계성을 표현한 지식 인스턴스 베이스를 포함하여 구성한다.

온톨로지의 구성은 그림 4에서 표현된 것과 같이 정의된 속성에 의해서 각 레거시 시스템의 MDR정보를 바탕으로 한 분석을 통해 생성된다. 상품 분류 기준인 카테고리에 의거한 각 상품의 데이터 표현은 표준 온톨로지 영역, 쇼핑물의 접근 정보에 관한 것은 로케이션 온톨로지 영역 그리고 상품검색시스템에서는 유사성이나 대체할 수 있는 상품의 관계성들을 지식 인스턴스 베이스(Knowledge Instance Base)를 추가함으로써 XMDR 생성된다.

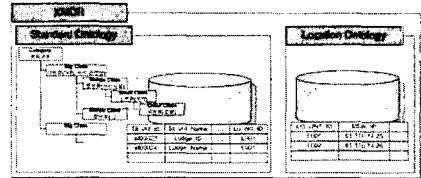


그림 4 XMDR의 상세 구성도

XMDR의 구성은 그림 5와 같이 표준 온톨로지와 로케이션 온톨로지를 바탕으로 참여한 레거시 시스템의 데이터에 대한 요소를 정의하며, 데이터와 시스템의 이질적인 문제를 극복할 수 있도록 표준이 필요하다. 따라서 XMDR은 문서 구조 정의를 통해 유효성 제약을 지니게 한다. 온톨로지의 설계 정보에 따른 XML Schema 문서를 정의하여 XMDR의 구조를 표준화 하고 각 레거시 시스템의 스키마 변경 등에 의한 XMDR의 임의 변경이 일어나지 않도록 하여 각 레거시 시스템의 데이터교환에 신뢰성을 확보하도록 한다.

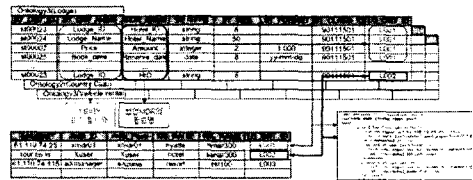


그림 5 표준 온톨로지와 로케이션 온톨로지 간의 매핑

이 XMDR에서 추가로 참여하는 경우, 매핑정보에 해당하는 XMDR Wrapper를 레거시 시스템에 설치시켜 XML_QUERY가 전송될 때 검색 에이전트에 의해 해당 레거시 시스템에 적합한 질의로 변환되어 데이터 이질성 문제를 해결할 수 있도록 한다. 이를 XMDR로 표현하기 위한 과정으로 그림 2과 같이 표준 온톨로지와 로케이션 온톨로지를 매핑시킨다. 이렇게 매핑된 것은 표 1과 같이 XML 문서로 표현된다.

표 1 XMDR의 XML 표현

```
<?XML version="1.0" encoding="euc-kr" ?>
<KnowledgeBase xmlns:xs="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="xmdr.xsd">
  <XMDR>
    <Standard_Location id="SL0001">
      <Location type="ip">61.110.74.25</Location>
      <Database auth_id="xmdr01" auth_pass="xmdr01">hyatt</Database>
      <table>hr300</table>
      <Standard_st_ont_id="st00023">
        <std_name type="string" size="8" format="no" classification="xxxx">
          Lodge_ID</std_name>
        <mdr_name>Hotel_ID</mdr_name>
      </Standard>
    </Standard_Location>
    <Standard_Location id="SL0002">
      <Location type="url">http://www.co.kr</Location>
      <Database auth_id="User" auth_pass="User">Hotel</Database>
      <table>sahar300</table>
      <Standard_st_ont_id="st00023">
        <std_name type="string" size="8" format="no" classification="xxxx">
          Lodge_ID</std_name>
        <mdr_name>HID</mdr_name>
      </Standard>
    </Standard_Location>
  </XMDR>
</KnowledgeBase>
```

2.4. 데이터웨어하우스 계층(Data Warehouse Layer:DW)

데이터웨어하우스 계층은 협업에 참여하는 기업의 레거시 시스템에서 구축된 데이터웨어하우스로 구성되는 계층으로, 레거시 시스템과

XMDR의 매핑을 수행하는 XMDR 래퍼에 의해 표준을 데이터웨어하우스에 인식시키는 계층이다.

이 계층에서는 레거시 시스템 자체의 변화는 없이 서버에서 제공되는 XMDR 래퍼를 이용하여 각 레거시시스템의 관리자가 XMDR과 데이터웨어하우스의 매핑을 수행하고 매핑된 정보를 XMDR 래퍼에서 관리함으로써 표준 질의를 레거시시스템에 적합하게 변환 후 질의를 수행하고, 수행된 결과는 다시 표준으로 변환하게 한다. 이렇게 함으로써 사용자에게는 요구한 정보를 제공하고, 참여한 기업의 데이터웨어하우스 정보는 유출하지 않고 사용자에게 제공할 정보만 공개하도록 한다. 그림 6과 같이 서버에서 제공되는 XMDR 표준함록과 레거시시스템의 데이터웨어하우스에 XMDR 래퍼로 연결한다.

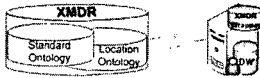


그림 6 XMDR과 DW

3. 시스템 개요 및 구성

본 논문에서 구축하고자 하는 시스템은 그림 7과 같이 4부분으로 구성된다. 첫째, 인터페이스이다. 인터페이스는 고객과 관리자에게 제공할 화면을 구성하고, 구성에 필요한 요소를 웹서버(Web Server)의 검색을 통해 인터페이스를 생성하는 역할을 한다. 둘째, 웹서버이다. 웹서버는 데이터의 관계성을 밝혀 연관된 지식을 검색할 수 있도록 제공하는 지식 인스턴스 베이스(Knowledge Instance Base)와 각 레거시시스템의 DW에 구성된 데이터베이스의 구조나 의미에 대한 관계성을 정의하고, 이를 유지, 관리하는 XMDR로 구성된다. 셋째, 실질적인 데이터의 이동에 관한 부분을 관리하기 위한 에이전트로 구성되는 네비게이트 관리자(Navigate Manager) 그리고 넷째, 레거시시스템에 있는 데이터웨어하우스(DW : Data Warehouse)로 구성된다.

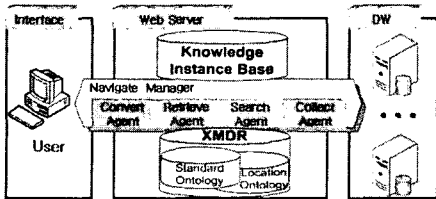


그림 7 시스템 구성도

3.1. Interface

인터페이스는 지식 인스턴스 베이스와 표준 온톨로지에 의해서 생성된다. 지식 인스턴스 베이스에 저장된 정보를 통해 검색 키워드의 연관관계 정보를 제공받고, 표준 온톨로지에서는 XMDR에서 구현된 XML 문서를 검색하여 사용자에게 제공된 표준 항목을 생성한다.

3.2. Web Server

참여하는 레거시시스템의 DW에 Wrapper에 의해서 표준으로 변환하게 되므로 웹서버는 각 레거시에 제공하기 위한 표준인 XMDR과 연관관계 분석을 통한 데이터의 연관성을 제공하기 위한 지식 인스턴스 베이스를 유지관리하고, 데이터의 이동과 변환을 담당하는 에이전트로 구성된다.

4. 구현 및 비교분석

4.1. 구현

본 시스템에서 적합한 검색 시스템의 인터페이스로 여행 정보 검색에 대해 XMDR를 적용한다. 그림 8은 사용자 인터페이스로 사용자의 요구에 의한 지식 인스턴스 베이스의 적용과 레거시시스템을 접근하기 위한 표준 온톨로지가 적용되는 부분이다. 지식 인스턴스 베이스에 의한 연관성 분석을 여부에 대한 지점과 표준 온톨로지의 상품 카테고리에 의한 상품선택을 할 수 있도록 지원하고, 선택된 상품에 대한 표준 항목을 지원하도록 표현한다.

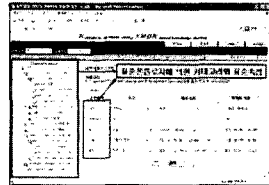


그림 8 적용1-검색 인터페이스

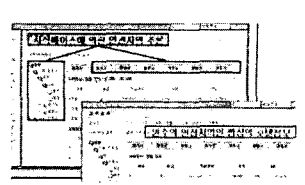


그림 9 적용2 - 검색결과

그림 9에서는 표현된 표준 항목에 의해 선택된 검색조건으로 검색한 결과는 사용자가 원하는 여행 지역의 정보를 우선적으로 제공하고, 해당 지역에 대한 연관관계를 분석한 결과에 해당하는 부분을 탭으로 제공하고, 사용자의 요구 지역인 전주와 인접한 지역인 평창에 관한 정보를 보여 주고 있다. 이와 같이 연관관계를 분석함으로써 사용자가 원하는 정보와 확장된 정보까지 제공할 수 있도록 한다.

4.2. 타 시스템과 비교

데이터의 상호운용성을 확보하기 위한 기반구조를 구성하기 위한 연구가 진행되고 있다. 특히 Microsoft를 비롯한 여러 단체에서 이용하고 있는 BizTalk, UN/CEFACT와 OASIS가 공동으로 개발한 ebXML, 전세계 IT에 종사하는 기업들의 비영리 컨소시엄이 정의하는 e-Business 표준 프레임워크인 RosettaNet 등의 프레임워크가 있다. 표 3에서와 같이 위의 3개 프레임워크를 대상으로 다음 항목으로 비교분석하였다.

표 3 데이터 통합을 위한 프레임워크들과의 비교

	BizTalk Framework	ebXML Framework	RosettaNet Framework	본 시스템
XMDR 지원	부분지원	부분지원	부분지원	지원
데이터교환 자동화	지원	부분지원	부분지원	지원
범용 명세 스키마	지원	없음	지원	지원
적용 범위	모든 분야	모든 분야	IT/EC 분야	모든 분야
저장소 구조	중앙 집중	분산	분산	집중과 분산
데이터 관계성	부분지원	부분지원	부분지원	지원

5. 결론 및 향후과제

본 시스템에서는 분산 데이터의 상호운용을 위한 어질성 극복을 위해 데이터 교환과 공유를 위한 표준 명세인 MDR을 바탕으로 XMDR 설계 및 구축하였고, 분산 시스템 프레임워크의 어질성 극복을 위해 XML기반의 메시지 교환 시스템 구축하였다. 그리고 협업 시스템에서 지식 공유를 위해 지식 인스턴스 베이스를 구축하고, 협업을 위한 레거시 시스템의 DW에서 관리하는 데이터를 수집하고, 수집된 데이터의 분석을 통해 생성하므로 실제 사용되는 데이터들의 관계성을 파악하여 이를 사용자의 검색에 제공한다.

협업 시스템에서 기업 간의 정보공유를 위한 정보 통합을 위해서는 XML 기반의 XMDR 기반이 되어야 한다. 따라서 본 논문 XMDR은 웹 서비스 및 기업의 DW에 적용이 가능할 것으로 기대된다.

이후는 P2P환경에서 Grid개념을 도입하여 데이터의 검색뿐만 아니라 데이터 교환과 수정이 가능한 시스템으로 확장할 필요가 있다.

참고문헌

[1]Hea-Sook Park, Jong-Hwan Kim, Doo-Kwon Baik, "Component-based Modeling and Verification of an Information Retrieval System using EPDM", Proceedings of The 4th ICACT, pp821-826, Feb, 2002.
 [2]이현경, 이응봉, "분산된 웹 정보자원의 공유 및 재이용을 위한 온톨로지 이용에 관한 연구", 『충남대학교 대학원 논문정보학과』, 2002.
 [3]김은경, "신장 질환 진단을 위한 규칙 기반 추론과 사례 기반 추론의 통합", 정보과학회논문지, 제24권 제10호, pp.1093-1100, 1997
 [4]고은성, 김여정, 김운, 강지훈, "지역 온톨로지를 이용한 지능형 여행정보 제공 시스템", 한국정보과학회 2004년 춘계학술대회, VOL.31 NO.01 pp.0610-0612 2004, 04
 [5]A. Adamodt, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and system Approaches", AI Communications ISO Press, Vol7:1, p.39-59, 1994
 [6]http://xmdr.org