

농촌어메니티자원 정보서비스를 위한 Ontology를 활용한 외부정보 통합방안

이지민 · 이정재*

서울대학교 대학원 · *서울대학교 조경/지역시스템공학부

Integration with External Information Using Ontology for Rural Amenity Resources Information Service

Lee, Ji Min · Lee, Jeong Jae*

Graduate School, Seoul Nat'l Univ. · *Dept. of Rural Systems Eng., Seoul Nat'l Univ.

ABSTRACT : Rural amenity resources consist of natural, cultural and social resources. Some of resources have been surveyed in other institutes and organizations and information system about these resources has been constructed. Integrating with information of existing system or database is helpful to produce useful rural amenity resources information to users. This integration of distributed information is based on semantic integration. In this paper, we used ontology for semantic integration and introduced the method of external information integration using ontology. We designed ontology table to represent ontology in database and integrated distributed information using ontology. To examine the improvement of information service efficiency and the applicability, rural amenity resources information and tourist attraction information database were constructed and integrated. Also, the information service efficiency with information integration was evaluated using recall ratio and compared to the information service efficiency without information integration.

Key words : External information, Integration, Ontology, Rural amenity resources

1. 서론

컴퓨터기술의 발달에 따라 대부분의 자료가 전산화되고 많은 양의 데이터들이 지속적으로 발생되고 있어 그동안 축적해온 많은 정보를 효율적으로 활용하기 위해 데이터베이스로 구축하려는 노력이 꾸준히 진행되어 왔다. 농업농촌분야에서는 농촌진흥청의 농업기술종합정보서비스(ATINS), 농림부의 한국농림수산정보센터(AFFIS), 농촌공사의 농어촌종합정보포털, 물관리종합정보 등과 같이 농업·농촌과 관련된 정보를 데이터베이스화 해왔다. 그러나 이러한 데이터베이스 구축은 각 기관에 의해 개별적으로 추진되어 특정 목적에 제한적으로 이용되는 특성을 가진다. 즉, 이러한 데이터베이스들은 각기 다른 목적에 의해 서로 다른 체계와 구성을 가지게 되어, 여러 기관에서 다른 목적으로 재사용되기 어렵고, 자료가 분산되어 있어 자료통합을 통한 범용적인 사용목적에 제약을 갖고 있

다. 따라서 분산된 정보는 필요에 의해 공유하기 어려우므로 분산된 정보를 활용하기 위해 효과적인 통합방안이 요구된다.

농촌어메니티 자원은 농촌공간에 존재하는 공동체와 공간을 형성하는 자원들로 농촌공간에서의 총체적인 삶의 질을 결정하는 자원으로 인식할 수 있다. 이러한 농촌어메니티 자원은 박창석 외(2002)의 연구에서 37종의 자원으로 분류된 바 있으며, 이는 정남수(2005)의 연구에서 10종의 분류로 재분류된 바 있다. 이러한 농촌어메니티 자원분류체계를 살펴보면 크게 자연자원, 문화자원, 사회자원으로 구성되어 농촌공간에 존재하는 자연환경 및 경관, 문화자원 및 시설, 공동체 및 행사 등이 포함되어 매우 다양한 분야와 연관됨을 알 수 있다. 농촌어메니티 자원은 '농촌어메니티 자원조사사업'을 통해 조사되고 있으나, 이러한 다양한 종류의 자원은 관련된 분야의 구축된 기존의 데이터베이스를 활용함이 유용할 것이다.

이질적이고 분산된 정보는 정보의 재사용성을 저하시키며 정보처리에 어려움을 발생시키므로 정보공유에 대한 중요성이 높아지고 있다. 이러한 정보공유 중요성의 증대

Corresponding author : Lee, Ji Min

Tel : 02-880-4595

E-mail : habi@skypond.snu.ac.kr

로 최근 정보의 의미론적인 통합을 위한 온톨로지 구축에 대한 필요성이 제기되고 있다(장태우, 2004). 온톨로지는 단순한 사전이나 분류의 개념에 더해 대상이 되는 문제영역에 사용되는 어휘들의 관계까지 정의하는 것으로 ‘개념화에 대한 명시적인 구체화’라고 정의된다. 이러한 온톨로지는 Tim Berners-Lee에 의해 주창된 기계가 이해할 수 있는 차세대 웹인 ‘시멘틱 웹’의 주요 요소로 인식되고 있으며, 웹상의 콘텐츠뿐만 아니라 이질적이고 분산된 정보들을 통합하는데 에도 의미영역의 공유를 담당하는 온톨로지가 중요한 역할을 할 수 있다고 지적된 바 있다(장태우, 2004). 온톨로지에서 용어관계를 정의하는데 BT(broader term : 상위개념어), NT(narrower term : 하위개념어), RT(related term : 관련개념어) 등이 사용되거나(한유석 외, 2004) 이러한 관계를 나타내는 is-a, is kind of 등이 사용된다.

본 연구에서는 이러한 온톨로지를 이용하여 이종의 정보를 통합하는 방안을 개발하고 이를 통하여 농촌어메니티 자원데이터베이스와 외부정보와 통합하여 농촌어메니티 자원정보의 활용성을 증대하고자 한다.

II. 기본이론

1. 농촌어메니티 자원정보

농촌어메니티 자원조사는 농촌공간이 가지는 잠재적 가

치인 농촌어메니티에 대한 관심이 고조되면서 이를 발굴하여 농촌어메니티 자원을 체계적으로 관리하고 보전하기 위해 농촌진흥청 농촌자원개발연구소에서 ‘농촌어메니티 종합기술 구축사업’의 일환으로 2005년부터 전국 1,230개 읍면을 대상으로 수행되고 있다. 이러한 자원조사는 앞서 제시한 박창석(2002)의 연구와 정남수(2005)의 연구에서 제시한 자원분류체계를 따르고 있으며 각 자원항목에 대한 정보를 기 구축한 관련기관을 조사한 결과는 다음 표 1과 같다.

표 1에서 보는 바와 같이 37종(정(2005)의 연구에서 10종)의 농촌어메니티 자원은 환경부, 문화재청, 농촌공사 등의 기관에서 기 구축한 정보와 관련이 있음을 알 수 있다. 이러한 기 구축된 정보는 농촌어메니티자원 정보와 통합된다면 정보의 공유 및 재사용성 증대 측면에서 높은 시너지 효과를 나타낼 것으로 판단된다.

2. 분산 데이터의 통합방안

데이터베이스 통합에 관한 연구는 멀티데이터베이스를 기반으로 한 데이터베이스 스키마통합에 관한 연구가 주를 이루고 있다(Mahendar, 1996, 김기중, 1998, 이상태, 1999). 즉, 멀티데이터베이스는 이기종 데이터베이스를 단일 전역스키마를 통해 액세스하기 때문에 전역스키마의 생성이 필수적이므로, 스키마 통합과 관련한 연구가 다양하게 이

표 1. 농촌어메니티 분류별 관련기관

분류(정)	분류(박)	자원항목(박)	관련기관
환경오염자원	환경자원	대기질, 수질	환경부
		소음	
지형자원	자연자원	토양	한국문화관광정책연구원(관광지식정보시스템)
		미기후	
특이지형			
수자원		수자원공사, 농촌공사(물관리정보시스템)	
동물자원		동물, 식생, 습지 혹은 생물서식지	
식물자원			
전통자원	역사자원	문화재	문화재청(국가문화유산종합정보서비스)
		비지정전통건조물, 전통주택	
		신앙공간, 마을상징물	
		유명인물, 풍수지리나 전설	
경관자원	경관자원	농업경관, 하천경관, 산림경관, 주거지경관	
시설자원	시설자원	공동생활시설, 기반시설, 공공편익시설, 환경관리시설, 정보기반시설, 농업시설	농촌공사(농촌지형정보)
특산자원	경제활동자원	특산물, 특용작물	
		도농교류(민박 등)	
공동체 자원	공동체활동자원	생활공동체, 농업공동체	마을홈페이지 (전통테마마을 홈페이지 등)
		씨족행사, 마을문화활동, 마을놀이, 마을관리 및 홍보	

워져 왔다. 최근 메타데이터관리기술을 활용한 전자상거래 B2B연구(이진수 외, 2002)와 XML을 이용한 분산된 데이터베이스 스키마를 통합하는 연구들(고광은, 2002, 탁우현 외, 1999)이 진행된 바 있다.

Y. Breitbart(1990)는 기존 멀티데이터베이스에 대한 문헌연구를 통해 멀티데이터베이스가 가지는 주요 문제점들로 데이터불일치, 의미론적 다양성, 트랜잭션 관리 및 질의 최적화를 지적하였다. 그 중 데이터불일치는 이름의 불일치, 데이터 타입의 불일치, 데이터 구조, 단위나 데이터 누락에 의해 발생한다고 지적하였으며, 의미론적 다양성은 각 데이터베이스들이 갖고 있는 정보에 대해 의미적으로 동등하거나 유사한 데이터 항목을 식별해야 한다고 지적하였다.

장태우(2004)는 데이터들을 통합하는데 있어 의미론적인 문제를 해결하기 위해서는 분산된 데이터에 대한 구조적이고 의미론적인 통합이 우선적으로 처리되어야 하므로, 이러한 멀티데이터베이스의 데이터 불일치 및 의미론적인 문제는 온톨로지의 부재로 지적하고 전자상거래 통합을 위한 온톨로지를 구축하였다. 또한 Noriaki Izumi(2002)는 e-비즈니스 핸드북(e-business process Handbook)과 Wordnet을 이용하여 비즈니스 요소 온톨로지와 비즈니스 과정 온톨로지를 구축하여 전자상거래에서의 정보통합을 위한 프로토타입 시스템을 구축한 바 있다.

온톨로지를 활용한 데이터베이스 통합 노력으로는 RDF를 이용하여 메타데이터를 표현하고 유사성을 고려하여 스키마 통합을 통한 데이터베이스 통합에 관한 연구가 진행된 바 있다(오상경, 2002). L. Kerschberg et. al(2004)은 여러 이질적인 정보원에서 정보 자원의 통합은 매우 중요한 문제로 인식하고 George Mason대학에서 진행되는 knowledge sifter라는 프로젝트를 소개하여 온톨로지를 이용하여 검색과 정보 공유를 쉽게 제공할 수 있는 에이전트 구조를 제시한 바 있다. Denjing Dou(2006)는 데이터베이스 스키마를 온톨로지로 변환하고 개념간의 의미론적 관계를 정의하는 방법을 사용하여 Ontogate를 사용하여 관계형 데이터베

이스를 통합한 바 있다. 그러나 이러한 방법은 Ontogate와 OntoEngine이라는 특정 소프트웨어를 사용하여야 하며 DBMS와 별도로 구축되어야 하는 한계를 가진다.

이러한 기존의 온톨로지를 활용한 이질적인 데이터베이스를 통합하려는 노력은 멀티데이터베이스 통합방법의 특징과 같이 유사한 데이터베이스일 때 각 데이터베이스의 모든 정보체계를 포함하는 온톨로지를 구축하는 개념이었다. 그러나 유사한 정보를 다루지 않고 정보체계가 매우 상이할 경우 새롭게 온톨로지를 구축하고 활용하는 매우 어려운 작업이 될 것이다.

III. Ontology를 활용한 분산정보 통합방안 개발

1. Ontology database 구조설계

온톨로지는 일정한 체계에 의한 용어 및 용어간의 관계를 정의하여 정보공유와 재사용을 위한 것으로 정의될 수 있으며, 온톨로지를 구축하는 방법으로는 온톨로지 언어 사용방법과 관계형 DBMS상의 구축방법이 존재한다. 이러한 각각의 방법은 다음 표 2와 같은 특징을 가진다(심준호, 2005).

시멘틱 웹의 등장으로 웹 온톨로지 구축을 위한 OWL(Web Ontology Language) 등이 소개되어 이를 통한 온톨로지 구축의 노력이 여러 분야에서 진행되고 있다. 그러나 심준호(2005)의 연구에서 지적된 바와 같이 현재까지 이론중심의 실험적 단계에 머무른 상태이므로 이를 바로 지금까지 운용해 온 시스템에 바로 적용하기 어려운 특성을 가진다. 이에 반해 온톨로지를 관계형 데이터베이스 상에 구축한다면 온톨로지 정보를 표현하거나 추론기능을 사용함에 한계를 가지지만 기존의 구축된 데이터베이스와 쉽게 연계될 수 있으며 실제 시스템에 적용되어 현실화될 수 있는 장점을 가지므로 본 연구에서는 온톨로지 데이터베이스를 사용하였다.

표 2. 온톨로지 구축방법의 비교(심준호, 2005)

구축방법	온톨로지 언어 기반	관계형 데이터베이스상의 온톨로지 데이터베이스
특징	<ul style="list-style-type: none"> o Semantic Web 기반 o OWL, DAML+OIL, Topic Maps, Description Logics, FOL, Logics, ... 	<ul style="list-style-type: none"> o Relational DBMS 기반 o EER, Table, SQL, Relational Algebra & Calculus, ...
장점	<ul style="list-style-type: none"> o 이론 중심, 이상적 - 다양한 온톨로지 정보의 표현 	<ul style="list-style-type: none"> o 동작 중심, 현실적 - 상용 DBMS & 응용 프로그래밍 활용 - 상용화 가능
단점	<ul style="list-style-type: none"> o 표현된 knowledgebase를 운용할 reasoning tool의 부재 - 실험적 연구실 수준 - Robust 한 시스템 부재 	<ul style="list-style-type: none"> o 온톨로지 정보의 표현과 기능의 시스템적 한계 o 온톨로지 추론의 지원의 한계

온톨로지 데이터베이스 구축하기 위해 RES(Radial Edge Structure)를 활용하여 온톨로지 테이블 구조를 설계하였다. RES는 Weiler에 의해 제안된 3차원 비다양체 형상모델링 방법으로 위상요소는 Model, Region, Shell, Face, Edge, Vertex로 구성되며 계층구조를 가진다. 신일섭(2002)의 연구에서 위상학적 형상화 방법인 RES를 관계형 데이터베이스로 구현하고자 한 연구가 있었으며, 본 연구에서는 그 연구결과를 바탕으로 온톨로지 데이터베이스 테이블 구조를 설계하고자 하였다.

온톨로지의 상하위 계층구조를 표현함을 목표로 웹 온톨로지 언어인 OWL의 데이터 구조인 RDF(Resource Definition Framework)을 고려하여 RES 요소 중 Shell, Edge, Vertex를 이용하여 온톨로지를 표현하기 위한 데이터베이스 구조를 설계하였다. RDF는 자원(Resource)을 기술하기 위한 모형을 제공하는데, 기술되는 자원은 고유한 속성(Property)을 가지고 있다. 자원을 표현하는 속성은 속성유형(Property Type)으로 구분된다. 속성유형은 '저자'나 '서명' 등과 같이 자원의 속성을 적절한 이름으로 표현한 것이다. 또한 속성 유형에 상응하는 값은 속성값(Value)이라 하며, 문자열이나 숫자 등과 같이 자연어로 상세하게 기술되며 다음 그림 1과 같다(한승희, 1998).

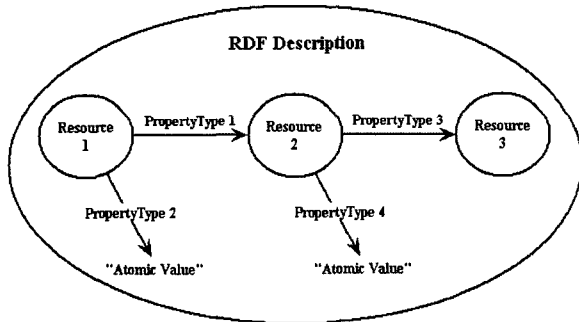


그림 1. RDF 구조.

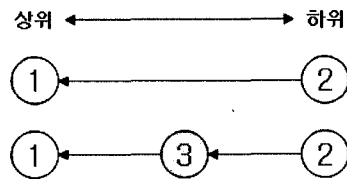


그림 2. 용어와 관계 분리구조.

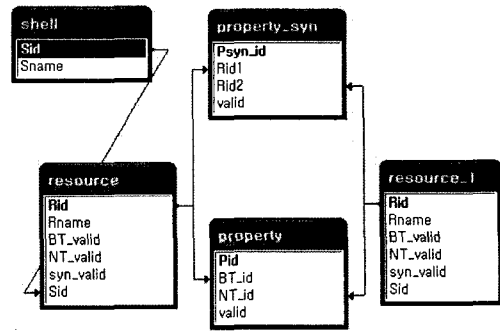


그림 3. 온톨로지 데이터베이스 테이블 구조.

이와 같이 자원(용어)과 관계를 분리한 구조는 아래 그림과 같이 용어의 계층이 관계에 의해 결정지어지므로 새로운 용어가 추가되더라도 관계만 수정되어질 뿐 전체 온톨로지의 수정은 필요로 하지 않는 장점을 가진다.

온톨로지를 데이터베이스 상에 표현하기 위해 사용한 Shell, Edge, Vertex는 각각 서브도메인(sub domain), 관계(relation), 용어(resource or term)로 정의하였다. 설계된 온톨로지 데이터베이스 구조는 다음 그림 3과 같다. 관계를 나타내는 edge는 상하위관계를 나타내는 테이블과 동의 관계를 나타내는 테이블로 property와 property_syn 2개의 테이블로 구성하였으며, 전체 도메인을 구성하는 부분들을 구분하기 위해 서브도메인으로 표현하였다.

2. Ontology를 활용한 분산정보 통합

이질적인 두 개의 데이터베이스를 통합하기 위해 본 연구에서는 각각의 데이터베이스에서 통합하고자 하는 정보부분만 추출하여 그 정보체계를 온톨로지로 구축하고 각각의 온톨로지를 유지하면서 동시에 통합하는 방법을 개발하고자 한다. 각각의 데이터베이스에서의 정보체계는 각 분야에서 의미 있는 체계를 따르고 있을 것이므로 이를 유지하는 것이 정보의 획득과 이해를 높일 수 있다.

용어와 관계 분리하지 않을 경우

용어	계층레벨
1	1
2	2

용어	계층레벨
1	1
2	3
3	2

용어와 관계 분리한 경우

용어	관계 (상)	관계 (하)
1	1	2
2		

용어	관계 (상)	관계 (하)
1	1	3
2	3	2
3		

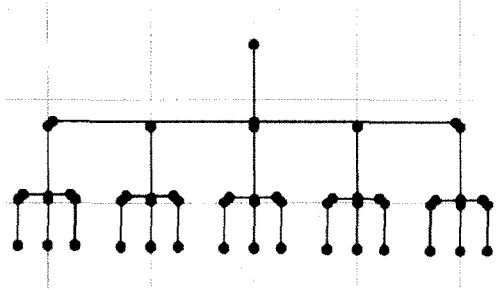


그림 4. 파일시스템 구조(Clay Shirtky, 2005).

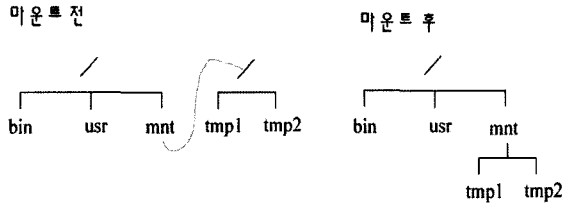


그림 5. 파일시스템의 마운트.

먼저 온톨로지의 계층구조는 그림 4와 같이 파일시스템의 계층구조와 비슷함을 알 수 있다. 온톨로지는 계층구조(hierarchy)와 동일한 것은 아니지만 이러한 구조적 특징을 통해 온톨로지를 통합하는 방법을 개발하였다. 온톨로지의 구조가 파일시스템과 유사한 구조를 가지는 점을 착안하여 파일시스템에서 마운트(mount)와 링크(links)에 대해 고찰해 보았다.

마운트의 경우 그림 5와 같이 어느 한 파일시스템을 마운트하기 위해서는 마운트 포인트(mount point)를 잡고 마운트 시키면 그 포인트를 루트로 하는 파일시스템구조로 형성된다. 온톨로지로 생각해 볼 경우 두 개의 온톨로지가 존재하며 같은 항목이 존재할 경우 하나의 온톨로지가 좀더 세분화된 용어(개체)들로 구성될 경우 파일시스템의 마운트 개념으로 통합이 가능할 것이다. 그러나 본 연구에서 고려하는 것은 같은 개념이 존재하지만 이질적인 온톨로지의 경우를 대상으로 하므로 마운트의 개념을 활용할 수 없었다.

파일시스템의 링크(links)는 여기서는 심볼릭링크(symbolic links)의 경우를 뜻하며, 윈도우체제에서는 바로가기(short-cut)이라고 이해해도 좋다(Clay Shirtky, 2005). 파일시스템 상에서 링크는 하나의 파일이 서로 다른 이름으로 서로 다른 위치(디렉토리)에 존재할 수 있도록 하는 것이다. 그러나 링크를 생성하는 것은 원하는 정보가 위치한 디스크 상의 장소를 가리키는 포인터를 생성하는 것으로 데이터를 갖고 있지 않는다(Marshall K. M. et. al, 1984). 온톨로지 측면에서 살펴보면, 저수지의 경우 수자원의 분류 아래 존재하기도 하면서 동시에 수리구조물 분류 아래 존

재할 수도 있다. 이를 파일시스템 상에서 표현하면 그림 6과 같이 연결된 링크로 존재할 수 있다. 또한 파일시스템에서 파일의 고유의 정보를 갖고 있는 inode는 웹온톨로지에서 각 개체가 갖는 고유한 URI와 같은 개념으로 생각할 수 있다.

이러한 파일시스템의 링크개념을 도입하여 온톨로지를 이용하여 분산된 정보를 통합하는 방법으로 다음과 같은 온톨로지 링크(ontology links)방법을 제안하였다. 분산된 정보인 각각의 데이터베이스에서 메타데이터를 통해 온톨로지를 구축한 다음 이 두 개의 온톨로지에서 같은 용어(개체)들을 연결하는 방법이다. 이러한 방법은 전역적 스키마를 생성하지 않는 데이터베이스 스키마 통합방법과 유사하나, 의미론적인 관계를 가지므로 이로 인한 오류를 해소할 수 있다는 면에서 장점을 가진다. 각 온톨로지의 용어(개체)들을 연결(link)할 때 동의어 사전을 활용하여 용어의 형태는 다르지만 의미론적으로 동일한 의미를 가진 용어들을 연결할 수 있다.

각각의 데이터베이스와 온톨로지 테이블을 가진 상태에서 온톨로지 연결정보를 별도의 테이블에 저장하여 분산된 정보를 통합하여 활용한다. 이를 통해 각각의 데이터베이스를 변환시키지 않고 연계되는 정보를 공유할 수 있는 장점을 가진다.

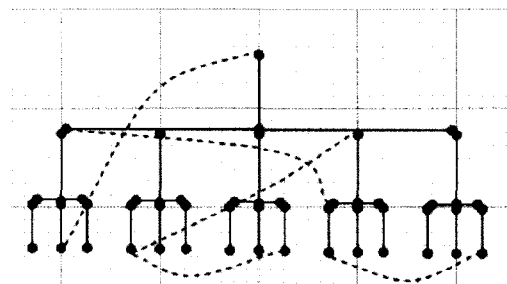


그림 6. 파일시스템의 링크(Clay Shirtky, 2005).

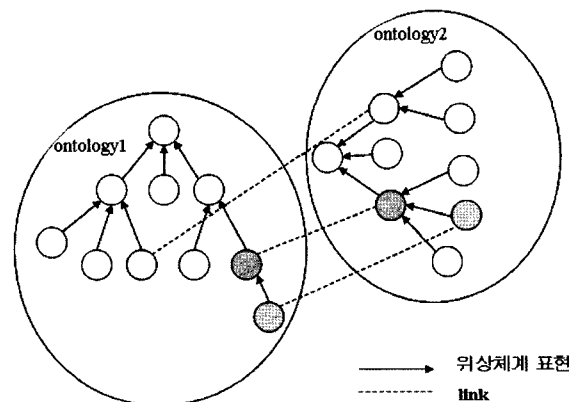


그림 7. 온톨로지 링크 통합방법.

표 3. 농촌어메니티자원 분류체계
(농촌어메니티자원조사사업 10종 분류체계)

Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	
농촌어메니티 자원	수자원	하천	식물자원	환경오염	마을숲	
		저수지			비보숲	
		호수			군락	
		약수터			보호수	
		공동우물			노거수	
		댐			냄새	
	동물자원	포유류	전통자원	경관자원	시설물자원	소음
		조류				토양오염
		어류				전통건축물
		곤충				신앙공간
		양서파충류				전통주택
		무척추동물				전통적인 마을안길
	특산자원	수공예품	정관자원	시설물자원		마을 상징물
		도자기				유명인물
		친환경 농법				풍수지리나 전설
		식품				주거지경관
		특용작물				농업경관
		섬유				산림경관
	공공단체자원	농업공동체활동	시설물자원			하천경관
		생활공동체활동				해안경관
		도농교류활동				공동생활시설
		마을 문화 활동				기반시설
		마을놀이				공공편익시설
		마을관리및홍보활동				환경관리시설
	지형자원	습지				정보기반시설
		생물서식지				농업시설
		바위				
		절벽				
		봉				
		소				
담						
폭포						
특이토양						
미기후발생지역						

표 4. 관광지식정보시스템의 자연 및 생태환경 분류체계

Level 1	Level 2	Level 3
자연 및 생태환경	동식물	희귀 보호식물
		희귀 보호동물
		식물자생지
		조류서식지
		어류서식지
		천연보호구역
	보호구역	생태계보전지역
		습지보호지역
		강
	수변 및 해양자원	천
		섬
		폭포
		연못
		호수, 댐
		늪지
		갯벌
	산악 및 평지자원	산
		봉
		고개(언덕)
		약수터
		계곡
		동굴 및 광물
		숲, 수림
		절벽
화석지		
오솔길, 동산, 산책로		
분화구, 오름		
경승지	대	
	전망대 및 조망시설	
	5경, 8경, 10경, 12경	
	일출, 낙조	
	경관도로	

출처 : 한국문화관광정책연구원
관광지식정보시스템
(http://www.tour.go.kr)

IV. 적 용

앞에서 제시한 온톨로지를 통한 분산정보 통합방법을 그 적용성을 살펴보기 위해 농촌어메니티 자원정보와 외부정보인 관광지식정보 통합에 적용해 보았다. 먼저 각각에 해당하는 온톨로지를 구축하기 위해 농촌어메니티자원 분류체계와 관광지식정보의 자연 및 생태환경 분류체계를 이용하였다. 각각의 자원분류체계를 살펴보면 표 3과 표

4에서 보는 바와 같이 3단계의 계층구조를 가지고 있다. 이러한 자원분류체계를 살펴보면 표 3과 표 4에 음영으로 표시한 바와 같이 서로 다른 분류체계이지만 같은 자원들이 서로 다른 위치에 존재함을 알 수 있다. 이러한 자원들을 온톨로지를 통해 연결하면 농촌어메니티자원정보와 외부정보를 통합하여 정보를 제공할 수 있는 장점을 가진다. 연결될 수 있는 자원리스트들은 다음 표 5와 같다.

표 5. 연결되는 자원리스트

농촌어메니타자원	관광지식정보
하천	천
저수지	호수, 댐
호수	호수, 댐
약수터	약수터
댐	호수, 댐
절벽	절벽
봉	봉
소	늪지
폭포	폭포

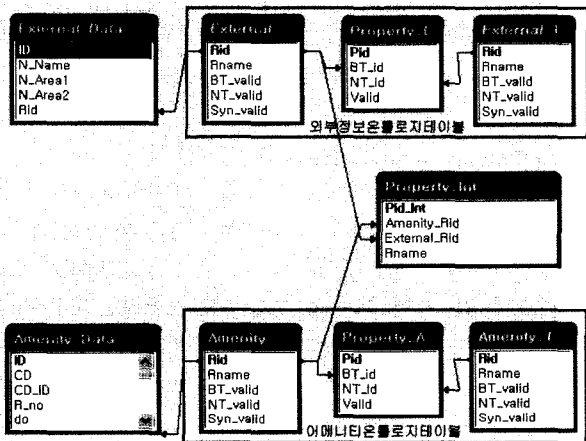


그림 8. 외부정보 통합 테스트 데이터베이스.

데이터베이스 상에서 농촌어메니타 자원정보와 외부정보의 온톨로지 테이블을 구축하고 이를 서로 연결한 정보를 테이블에 저장한다. 각 테이블간의 관계는 그림 8과 같으며, External테이블과 Amenity테이블이 온톨로지의 Resource테이블 역할을 하며 각각의 Rid가 실제 데이터가 저장된 테이블과 연결되며, Rid는 외부정보와 통합하기 위한 Property_Int테이블에서 표 5에서 제시한 자원리스트들이 각각 온톨로지의 Rid를 통해 연결정보를 가진다.

개발된 통합방안의 적용성과 효율성을 판단하기 위해 먼저 2005년 조사된 지역 중 전라북도 지역의 농촌어메니타 자원정보를 Amenity_Data 테이블로 구축하고 농촌어메니타자원 분류체계를 온톨로지 테이블로 구축하였으며, 관광지식정보의 자연 및 생태환경 정보를 외부정보로 구축하였다.

통합된 데이터베이스를 기반으로 정보를 제공할 경우 효율성을 판단하기 위해 농촌어메니타 자원데이터 테이블(Amenity_Data)과 외부정보인 관광지식정보 데이터 테이블(External_Data)만 존재할 경우와 온톨로지와 이를 연결한 링크테이블(Property_Int)이 존재할 경우에 대하여 자원정보를 얻기 위한 검색을 통해 비교해 보았다.

표 6. '하천'자원 검색결과 비교

	농촌어메니타자원		관광지식정보	
	검색결과 수	재현율	검색결과 수	재현율
온톨로지 비이용 (명칭검색)	97	0.78	10	0.03
온톨로지 이용	123	1.0	320	1.0
총 데이터 개수	2,990		46,302	

온톨로지 테이블이 없을 경우 데이터는 자료분류체계인 위상구조정보를 갖고 있지 않기 때문에 직접 데이터를 검색해야 한다. 이러한 경우 같은 자원 종류로 분류되었으나 검색되지 않는 자원이 존재한다. 예를 들어 하천의 경우 외부정보 상에 '천' 자원분류로 구분된 자원은 46,302개의 자원정보 중 320개 존재하였으나 단순히 자원명을 '하천'으로 검색할 경우 10개의 자원이 검색되었다. 검색효과를 측정하는 방법으로는 1966년 Curil Cleverdon에 의해 만들어진 재현율(recall)과 정확률(precision)이 있으나 본 연구에서는 관련 없는 것이 검색되지 않음을 가정하여 검색된 결과 중에 적합한 결과의 비율을 나타낸 재현율을 이용하여 비교하였다.

$$\text{재현율 (recall ratio)} = \frac{\text{적합한 검색된 결과}}{\text{적합한 검색대상}} = \frac{w}{w+x}$$

	검색됨	검색되지 않음
관련 있음	w	x
관련 없음	y	z

온톨로지 링크테이블의 효율성을 판단하기 위해 각각의 온톨로지 테이블 만 구성한 경우와 온톨로지 링크테이블을 통해 연결된 경우를 비교해 보기 위해 '저수지' 검색을 이용하였다. 각각의 온톨로지에서 '저수지'의 rid를 검색하여 이에 해당하는 자원들을 검색하였다. 관광지식정보 온톨로지에는 '저수지'가 존재하지 않는다. 따라서 단순히 온톨로지를 통해 검색할 경우 검색결과는 0개이다. 그러나 온톨로지 링크테이블에 '저수지'와 관광지식정보의 온톨로지 상의 '호수·댐'을 연결하여 저장하였기 때문에 온톨로지 링크를 이용한 검색결과는 197개가 나타났다. 이러한 검색결과들의 효율성을 비교하면 다음 표 7과 같으며, 온톨로지를 이용하지 않고 각각의 데이터베이스에서 명칭으로 검색하는 것보다 온톨로지링크를 통해 통합할 경우 검색효율이 증가함을 알 수 있다.

또한 온톨로지 링크 테이블을 통해 연결된 자원정보를 미리 선언해 두기 때문에 외부정보에서 보완이 되는 정보인지 아닌지 미리 알고 있어 연결정보가 있을 경우에만 검색을 하게 된다. 예를 들어 '공동우물'과 같은 경우 외부정보 상에 우물과 관련된 정보가 존재하지 않으므로

표 7. '저수지'자원 검색결과 비교

	농촌어메니티자원		관광지식정보		통합정보	
	검색결과 수	재현율	검색결과 수	재현율	검색결과 수	재현율
온톨로지 비이용 (명칭검색)	32	0.62	151	0.76	183	0.74
온톨로지 이용	51	1.0	0	0	51	0.2
온톨로지 링크	51	1.0	197	1.0	248	1.0
총 데이터 개수	2,990		46,302		49,292	

외부정보를 검색하지 않고 농촌어메니티 자원정보만 제공하도록 하여 검색결과뿐만 아니라 불필요한 검색을 수행하지 않는 장점을 가진다.

이렇게 통합된 데이터베이스를 기반으로 정보를 제공할 경우를 순창군 지역의 약수터를 검색할 경우를 일례로 살펴본다. 어메니티 자원정보에서 '약수터'에 해당하는 자료를 제공하며 이와 동시에 '약수터'는 관광지식정보와 연결된 자원이므로 관광지식정보에서 갖고 있는 '강천약수터' 정보를 함께 제공하게 된다. 이 경우 해당 정보를 찾기 위해 외부정보 전체를 검색하는 것이 아니라 연결정보를 통해 해당정보의 유무를 파악하여 갖고 온다. 관광지식정보데이터베이스에서 갖고 있는 약수터 정보의 웹주소를 통하여 이를 함께 제공한다면 다음 표 8과 같다.

표 8에서 보는 바와 같이 어떤 지역의 특정 자원종류에 대한 정보를 얻고자 할 경우, 농촌어메니티 자원정보만으로 정보를 제공할 수 있지만, 같은 지역 내에 존재하는 자원이 조사되지 않았을 경우 외부정보를 통해 정보의 보완이 이루어 질 수 있다. 또한 외부정보인 관광지식정보 측면에서도 기존에 조사되어 제공되지 않고 있던 자원정

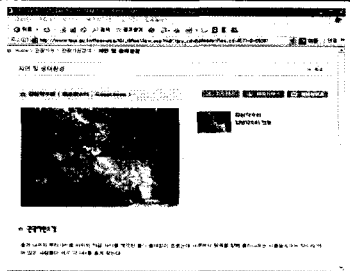

보를 농촌어메니티 자원정보를 통해 추가되거나 같은 자원일 경우 제공되지 않던 좌표정보 등의 정보를 추가적으로 제공할 수 있어 정보서비스의 질과 양을 증대시킬 수 있을 것으로 판단된다.

V. 결론

각기 다른 목적에 의해 서로 다른 체계와 구성을 가지도록 구축된 데이터베이스 및 정보들은 여러 기관에서나 다른 목적으로 재사용되기 어렵고, 자료가 분산되어 있어 자료통합을 통한 범용적인 사용목적에 제약을 갖고 있다. 따라서 정보를 효율적으로 관리하고 공유하기 위한 효과적인 통합방법을 필요로 한다. 이러한 데이터베이스의 통합을 위해서는 의미론적인 통합이 우선되어야 하며, 이 질적이고 분산된 정보들을 통합하는데 의미영역의 공유를 담당하는 온톨로지가 중요한 역할을 할 수 있다.

다양한 자원종류를 포함하는 농촌어메니티 자원은 기존에 구축된 다양한 정보와 관련되며 이러한 외부정보를 통하여 조사된 자원정보를 보완하고 정보를 통합하여 제공할 필요가 있다. 따라서 본 연구에서는 온톨로지를 데이터베이스 상에 표현하기 위한 테이블 구조를 설계하고 이를 이용하여 외부정보와 농촌어메니티자원정보를 통합하는 방안을 개발하였다. 이 통합방안의 적용성을 판단하기 위해 2005년도 조사지역 중 일부지역의 농촌어메니티자원정보와 관광지식정보 중 자연 및 생태환경 정보를 구축하였으며, 정보제공 시 효율성을 살펴보기 위해 온톨로지를 이용하지 않을 경우와 통합하지 않은 경우에 대해 검색효율을 각각 비교하였다.

표 8. 통합된 정보 제공 (예)

농촌어메니티 자원조사 정보		관광지식정보	
자원속성	자원속성정보	자원속성	자원속성정보
위치	전라북도 순창군 팔덕면 산동리 (팔왕리)	위치	전라북도 순창군 팔덕면 청계리
좌표	35°22'51" , 127°05'26"	자원명	강천약수터
자원명	약수터	웹주소	http://www.tour.go.kr/Resource/101_GResView.asp?hidCtgr_y_cd=B&hiddenRes_cd=4577-B-05037
설명	장재마을 약수터	사진	
이용현황	식수		
사진		웹페이지	

본 연구결과는 온톨로지를 이용하여 의미론적 통합이 이루어지는 장점을 가지며 이로 인해 검색의 효율성이 증가하고 어메니티 자원정보나 외부정보에 수정을 가하지 않고 온톨로지와 연결되는 정보를 통해 분산된 정보통합이 가능하였다. 또한 이러한 외부정보통합을 통해 농촌어메니티 자원정보 서비스의 일례를 살펴본바 외부정보와 농촌어메니티 자원정보가 서로 보완되어 제공될 수 있음을 확인하였다.

이 연구는 농촌진흥청 농촌자원개발연구소의 용역 연구(농촌어메니티자원도 관리 및 활용기술 연구)지원에 의하여 수행되었음

참고문헌

1. 강신재, 2004, 의미 분석을 위한 말뭉치 기반의 온톨로지 학습, 한국산업정보학회 논문지 9(1) : 17-23
2. 고광은, 2002, XML을 이용한 분산 관계형 데이터베이스 스키마의 통합, 강릉대학교 대학원 컴퓨터공학과 석사학위논문
3. 김기중, 1998, 통합 데이터베이스를 위한 스키마 통합 방법, 컴퓨터정보통신연구 6(2) : 1-12
4. 박창석, 전영욱, 조영국, 2002, 농촌어메니티에 기초한 농촌자원 중요도 평가 및 순위적 관계 분석, 대한국토 도시계획학회지 국토계획 37(6) : 21-35
5. 신일섭, 2002, Radial Edge Data Structure와 RDBMS를 이용한 3-D 모델링, 서울대학교 석사학위논문
6. 심준호, 2005, 현실적인 상품정보 온톨로지 데이터베이스 구축, <http://blog.naver.com/iwithin?Redirect=Log&logNo=80016760410>
7. 오상경, 2002, 이중 관계형 데이터베이스의 추출을 통한 포괄적 RDF 스키마 생성과 데이터베이스 관리, 서울대학교 석사학위논문
8. 이경하, 이상, 이강찬, 이규철, 1999, 룰 기반 미디어이터의 설계 및 구현, 한국정보과학회 봄학술발표논문집 : 185-187
9. 이상태, 주경수, 1998, 객체-관계형 데이터 모델을 토대로 한 다중 데이터베이스 시스템의 스키마 통합에 관한 연구. 1998년 한국멀티미디어학회 춘계학술발표논문집, 256-261
10. 이진수, 노희영, 2002, B2B 전자상거래에서 메타데이터를 이용한 데이터베이스의 통합, 강원대학교기초과학연구소 기초과학연구 제13집 : 45-56
11. 장태우, 2004, 전자상거래 통합을 위한 온톨로지 구축, 서울대학교 박사학위논문
12. 정남수, 2005, 농촌어메니티 자원의 평가를 위한 조사표와 항목개발, 한국농촌계획학회지 농촌계획 11(4) : 9-15
13. 탁우현, 류영호, 이종환, 김경석, 1999, 멀티데이터베이스에서 XML을 이용한 스키마 통합, 1999년 한국정보과학회 가을 학술발표논문집 26(2) : 78-80
14. 한승희, 2002, 메타데이터간의 상호호환성을 위한 RDF 구조연구, 연세대학교 대학원 문헌정보학과 메타데이터 연구보고서 YS-MTR-1998-8
15. 한유석, 설근수, 2004, 한국어 시소러스 연구, 한국문화사
16. Clay Shirky, 2005, Ontology Is Overrated, O'Reilly ETech conference http://www.shirky.com/writings/ontology_overrated.html
17. Dejing Dou, Paea LePendu, 2006, Ontology-based Integration for Relational Databases, Proceedings of the 2006 ACM symposium on Applied computing, 461-166
18. Larry Kerschberg, Mizan Chowdhury, Alberto Damiano, Hanjo Jeong, Scott Mitchell, Jingwei Si, and Stephen Smith, 2004, Knowledge Sifter: Ontology-Driven Search over Heterogeneous Database, Proceedings of the 16th International Conference on Scientific and Statistical Database Management, 431-432
19. Mahendar Madhavaram, Dia L. Ali, Ming Zhou, 1996, Integrating Heterogeneous Distributed Database System, Computers Ind. Engng. 31(1/2) : 315-318
20. Marshall Kirt McKusick, Willam N. Joy, Samuel J. Leffler, Robert S. Fabry, 1984, A Fast File System for UNIX, ACM Transactions on Computer Systems (TOCS) Vol. 2, Issue 3, 181-197
21. Noriaki Izumi, Takahira Yamaguchi, 2002, Integration of heterogeneous repositories based on ontologies for EC applications development, Electronic commerce Research and Applications 1(1) : 77-91
22. Y. Breitbart, 1990, Multidatabase interoperability, ACM SIGNOD Record 19(3) : 53-60

* 접수일 : 2006년 10월 21일