

농촌어메니티자원 검색을 위한 온톨로지 활용방안

Ontology-based Information Retrieval Algorithm in Rural Amenity Resources

이지민*·박미정·이정재 (서울대)

Jimin Lee · Meejeong Park · Jeong-Jae Lee

Abstract

Effective information "query and retrieval" process is one of the fundamental problems in the field communication and information science and has become especially important due to dramatic increase in magnitude of information to be processed in modern era. Of particular importance at information exchange process, our study focuses on compositions of proper queries and retrieval of rural amenity resources. This particular task has been difficult because the rural amenity resources does not necessarily carry measurable traits and also contains huge amount of data. In this Letter, we propose an alternative approach to the architecture of the resource information system by use of a noble retrieval algorithm based on ontology. Test of efficiency and applicability of this new scheme was conducted, and it showed that this has possibility to be effective information retrieval process of rural amenity resources.

I. 서론

인터넷이 일상생활에 보편화되면서 정보의 획득과 배포가 용이해졌다. 또한 정보의 집중을 막고 여러 사람들이 정보를 공유하기 위해 각종 정보와 서비스가 인터넷을 기반으로 제공되고 있다. 전자정부 및 각각의 공공기관에서는 토양, 저수지, 수자원 등에 관한 정보가 웹으로 제공되고 있다. 그러나 이러한 정보의 제공은 양방향으로 정보가 유통하는 인터넷이라는 장점을 살리지 못한 채 정해진 정보만 일방적(단방향)으로 제공되고 있다.

또한 웹상에 유통되는 정보의 양이 기하급수적으로 증가하고 있으며 이에 따라 각종 정보검색시스템과 포털사이트를 통하여 웹정보를 검색함에 한계가 나타나고 있다. 검색어를 입력하면 수천개 혹은 수만개의 웹사이트 및 웹문서가 검색되는데, 이는 정보의 양이 매우 많아 단순한 키워드 검색만으로는 원하는 정보를 찾기 어렵기 때문이다. 이러한 한계로 인하여 '지식검색'서비스가 제공되는 예도 있으나 이 또한 그 양이 증가하면서 원하는 정보를 찾기 어려운 기존의 웹문서 검색의 한계를 갖고 있다. 이러한 한계에도 사용자는 원하는 정보에 대해 정확한 검색결과를 얻고자 하는 요구는 점차 증대되고 있다.

증가하는 일반 사용자의 요구를 충족하기 위해 새로운 기술이 주목받고 있다. 이 중 차세대 인터넷 기술인 '의미를 기계가 처리할 수 있는 시멘틱 웹'과 컴퓨터가 정보의 의미를 보다 정확하게 이해하고 조작할 수 있도록 메타정보를 정의하는 '온톨로지'가 인공지능, 컴퓨터공학, 자연어처리 분야에서 활발히 연구되고 있다. 특히 온톨로지는 정보의 재사용성, 정보의 공유 등을 위한 해결책으로 제시되고 있으며, 지식의 표현과 처리를 위한 기본적인 요소로 간주되고 있다.

공공기관의 토양, 저수지, 수자원 등의 정보가 제공되는 바와 같이 농촌어메니티(농촌공간의 쾌적성)를 형성하는 농촌공간에 존재하는 유·무형자원정보를 조사구축하여 자원을 체계적으로 보존·관리하며 잠재적 가치를 발굴하여 부가가치를 창출하고자 하는 노력이 진행되고 있다. 농촌어메니티자원은 농촌공간에 존재하는 다양한 자원을 모두 포함하여 그 종류가 매우 다양하며, 전국적으로 자원정보가 조사된다면 그 양 또한 매우 방대할 것이다. 따라서 농촌어메니티자원정보에서도 인터넷공간에서의 정보검색

과 같은 문제가 발생할 것으로 예상된다.

따라서 본 연구에서는 우리나라에 분포하는 다양한 농촌어메니티자원정보를 효율적으로 활용하고, 사용자가 쉽게 원하는 정보를 찾을 수 있으며, 널리 공유될 수 있도록 온톨로지 활용방안을 제시하고자 한다.

II. 이론 및 선행연구

1. 온톨로지란

온톨로지는 철학분야에서 존재론을 일컫는 단어로 '존재하는 것과 그 범주를 연구하는 학문'이라 할 수 있으며, 이러한 개념은 컴퓨터공학과 인공지능 분야에서 재정의되어 사용되고 있다. 온톨로지는 '공유된 개념화에 대한 형식적 명시적인 구체화' (Gruber, 1995) 의 정의가 가장 널리 사용되고 있으며, 단순한 사전(dictionary)이나 분류(taxonomy)의 개념에 대상이 되는 문제영역에서 사용되는 어휘들의 관계까지 정의하는 것이다(장태우, 2004). 같은 전산분야에서도 온톨로지는 서로 다른 의미로 사용되기도 하여 웰티(C. Welty)는 온톨로지 분류되는 것들이 매우 넓은 스펙트럼 내에 병존하고 있다고 지적하였다(한국전산원, 2004).

본 연구에서는 기존의 온톨로지 정의에서 공통적으로 나타나는 '개념, 관계, 속성' 을 중심으로 '특정 분야에 대한 개념과 개념들간의 관계(분류체계), 그리고 개념들의 속성을 나타내는 것'으로 온톨로지를 정의하였다

2. 온톨로지 활용사례

온톨로지를 설명하거나 활용된 예를 보여주고자 할 때 주로 사용되는 것이 'Wine Ontology'이다. 와인 온톨로지는 스탠포드대학에서 와인정보가 제공되는 웹사이트 자료를 바탕으로 와인의 종류, 지역, 맛, 성분, 어울리는 음식들의 정보를 정의하고 그 정보들간의 관계를 정의한 것으로, 어떤 음식에 어울리는 와인을 자동으로 검색하여 추천해주는 agent를 통해 활용될 수 있다. 이처럼 knowledge base를 구축하고자 하거나 knowledge base를 통해 원하는 정보를 제공하기 위해 가장 기본적으로 구축되어야 하는 것이 온톨로지이다(wine agent site).

해외에서는 FAO에서 농업, 산림, 어업 및 환경을 주제로 "AGROVOC"이라는 다언어의 온톨로지를 구축하고 있으며, W3C를 중심으로 RDF, OWL등의 온톨로지를 표현하는 언어에 대한 연구 등 온톨로지 구축, 온톨로지 활용 에이전트 및 시스템에 대한 다양한 연구가 진행되고 있다. 국내에서도 여러 분야에서 온톨로지와 관련된 연구가 진행되고 있으며 대표적인 활용사례로는 조달청의 제품정보 온톨로지와 백과사전 인물 온톨로지 기반의 검색시스템인 ETRI의 AnyQ를 들 수 있다.

조달청의 제품정보를 온톨로지로 구축한 사례는 50만개 품목의 상품정보를 어휘·속성·속성값·측정단위·규칙 등 9개 사전으로 구축하여 상품 검색 기능을 대폭 확장한 시스템으로 사용자가 원하는 상품을 정확히 찾아낼 수 있도록 온톨로지가 지원하고 있다. ETRI의 백과사전 인물정보를 온톨로지로 구축한 사례는 백과사전 데이터를 knowledge base로 구축하여 사용자 질문에 유동적으로 정답을 제시하는 시스템을 구축한 바 있다.

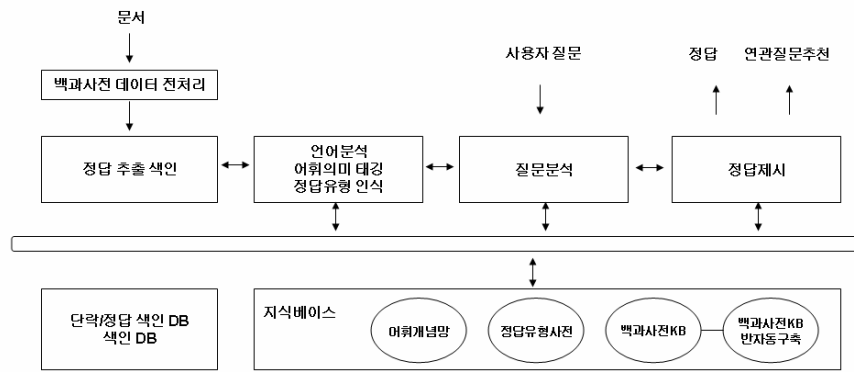


그림 1 백과사전 인물정보검색 AnyQ 시스템(anyq.etri.re.kr)

III. 농촌어메니티자원과 농촌어메니티자원 정보 제공

1. 농촌어메니티자원종류 및 특징

농촌어메니티자원은 여러 연구에 의해 다양한 분류기준이 존재하며, 그 중 박창석 외(2002)는 기존의 어메니티, 농촌어메니티 및 자원의 분류체계를 바탕으로 농촌어메니티자원을 자연적자원, 문화적자원, 사회적자원으로 분류하였다. 이 연구에서는 대·중·소분류의 자원분류체계를 제시하고 설문조사를 통하여 자원의 중요도를 도출하였다. 이 연구결과의 자원분류체계가 수정되어 농촌진흥청에서 ‘농촌어메니티 자원조사사업’에 이용되고 있다.

박의 연구에서 제시한 농촌어메니티자원의 분류체계는 표1과 같으며, 이처럼 각기 다른 종류의 인문사회적, 자연적, 문화역사적 자원들로 구성되어 있음을 알 수 있으며, 자원의 근원이 되는 것이 사람, 자연자원, 인공적 시설물 등으로 서로 다른 특징을 보인다. 김정섭은 농촌어메니티를 생성의 기원에 따라 자연적, 인위적, 자연과 인간상호적 자원으로 분류하기도 하였는데, 이처럼 농촌어메니티자원은 종류 뿐 아니라 각 자원이 가지는 속성이 서로 상이하게 다른 특성을 가진다.

표 1. 농촌어메니티자원의 분류체계 (박창석 외, 2002) (농촌자원개발연구소, 2005)

대분류	중분류	자원(박)	조사자원(농자연)
자연적 자원	환경자원	대기질, 수질, 소음이 없는 환경	환경오염자원
	자연자원	비옥한 토양, 미기후, 지형, 동물, 수자원, 식생, 습지 혹은 생물서식지	수자원 지형자원 식물자원 동물자원
문화적 자원	역사자원	문화재 사적 등 지정 전통건조물, 비지정 전통 건조물, 신앙공간, 전통주택, 마을의 전통적인 요소, 마을상징물, 유명인물, 풍수지리나 전설	전통자원
	경관자원	농업경관, 하천경관, 산림경관, 주거지경관	경관자원
사회적 자원	시설자원	공동생활시설, 기반시설, 공공편의시설, 환경관리 시설, 정보기반시설, 농업시설	시설물자원
	경제활동	도농교류시설, 특산물생산, 특용작물생산	특산자원
	공동체 활동	생활공동체활동, 농업공동체활동, 씨족행사, 마을문화활동, 마을놀이, 마을관리 및 홍보활동	공동체자원

2. 농촌어메니티자원정보 제공 시스템

농촌어메니티자원은 앞에서 살펴본 것과 같이 상이하게 다른 자원들의 집합으로 구성되어 있어 기존 하나의 자원종류를 위한 정보시스템(수자원정보지원, 저수지정보지원 등)의 기능만으로는 자원정보를 제

공, 활용하기 어렵다. 기존의 자원정보시스템은 하나의 자원종류만을 대상으로 하여 제공하는 정보의 수가 제한적이며 단순히 정보를 보여주는 기능만으로 대부분의 정보를 제공할 수 있었다. 또한 다양한 자원정보를 갖고 있더라도 정해진 제한된 자원정보를 일방적으로 제공하여 사용자의 요구와는 무관하게 서비스되고 있다.

농촌어메니티자원정보시스템은 전국에 분포하는 다양한 자원의 정보가 통합된 시스템으로 그 데이터량은 매우 방대해질 것임을 예상할 수 있다. 또한 다양한 자원의 속성에 의해 종합적으로 형성되는 농촌어메니티는 단순히 조사된 농촌어메니티자원정보 뿐 아니라 각 자원과 관련된 다른 데이터베이스와 연계되어야 제대로 표현될 수 있다. 다른 데이터베이스와 연계될 경우, 그 데이터량은 기하급수적으로 증대될 것이다. 데이터베이스에는 무수히 많은 정보가 저장되어 있지만 일반적인 검색으로는 검색어에 따라 검색결과가 없거나 무수히 많은 결과를 얻게 될 것이다. 원하는 정보를 정확히 검색해 내기 위해서 다음과 같이 온톨로지를 활용한 농촌어메니티자원정보시스템의 구성을 다음과 같이 제안하고자 한다.

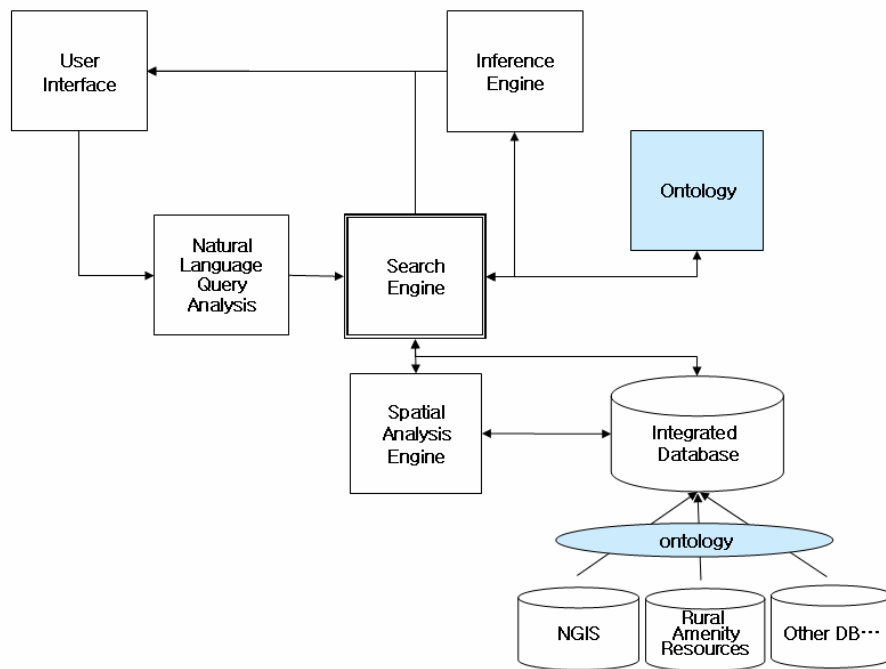


그림 2 농촌어메니티자원정보시스템 구성안

농촌어메니티자원정보시스템은 사용자인터페이스, 질의분석기, 검색엔진, 추론엔진, 공간분석엔진, 온톨로지와 통합데이터베이스로 구성된다. 통합데이터베이스는 NGIS, 조사된 농촌어메니티자원자료와 일반사용자 및 다른 기관에서 제공하는 관련 데이터베이스가 온톨로지를 통해 통합되며, 이렇게 통합된 자료에서 사용자가 원하는 정보를 온톨로지와 검색엔진, 지리적 공간분석엔진, 추론엔진을 통하여 제공하게 된다. 사용자가 질의를 하게 되면 질의분석을 통해 검색어가 검색엔진에 전달되며, 이러한 검색어는 온톨로지와 연계되어 통합DB를 검색하게 된다. 온톨로지는 시스템 내에서 개념을 정의하고 개념간의 관계를 정의해두고 있으므로, 이를 통하여 사용자가 질의에 대한 원하는 정보를 얻을 수 있도록 한다. 또한 공간분석이 필요한 질의는 공간분석엔진에 통합DB의 NGIS자료가 전달되어 이뤄지며, 추론엔진에서는 일정규칙을 통해 관련 DB의 자료에서 원하는 정보를 검색하여, 질의에 대한 결과는 질의에 대한 자원정보, 공간정보, 추론된 정보 등을 갖게 된다.

IV. 농촌어메니티자원 온톨로지 활용방안

1. 농촌어메니티 자원 온톨로지 예시와 온톨로지 활용한 검색 예시

농촌어메니티자원 중 수자원에 해당하는 온톨로지를 U-win(울산대)의 자료를 기반으로 구성하였다. U-win의 '물'과 관련된 체계를 검색하고, 그 자료를 대상으로 고유명사를 제외하며, 농촌어메니티자원의 조사대상을 고려하여 수정하였다. 그 결과 각 단어는 정의와, 동의어, 상하관계를 갖는 특징을 가진다.

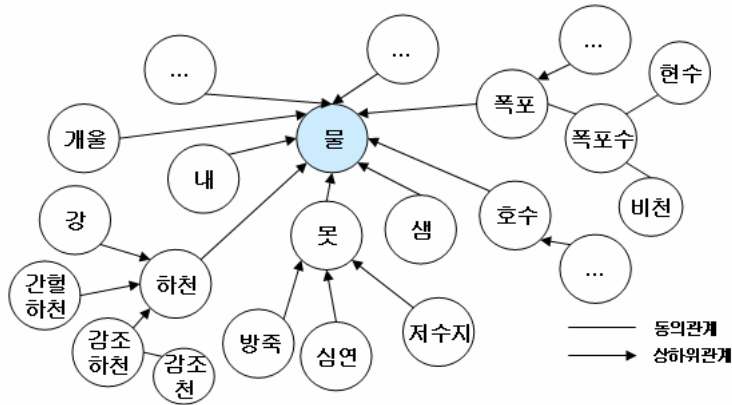


그림 3 수자원 온톨로지 일부

온톨로지를 활용하여 검색할 경우, 단순한 단어매칭이 아닌 개념의 상하위 체계를 통해 사용자가 질의하는 것에 대한 답을 줄 수 있다. 이를 구현하기 위해 본 연구에서는 다음 그림과 같이 확장검색프로세스를 정의하고 임의의 자원조사DB를 대상으로 수자원온톨로지를 활용하여 "못"을 검색어로 테스트해 보았다. 사용자가 '못'을 질의할 경우, 자원조사 DB에서 검색한 후 결과가 없을 시 확장검색이 실시된다. 여기서 동위어는 같은 상위어를 가지는 단어를 뜻한다. 온톨로지는 단순한 구조가 아닌 이론적 기반을 토대로 구축된 개념들 간의 관계이므로 이러한 확장검색의 결과는 의미를 가진다.

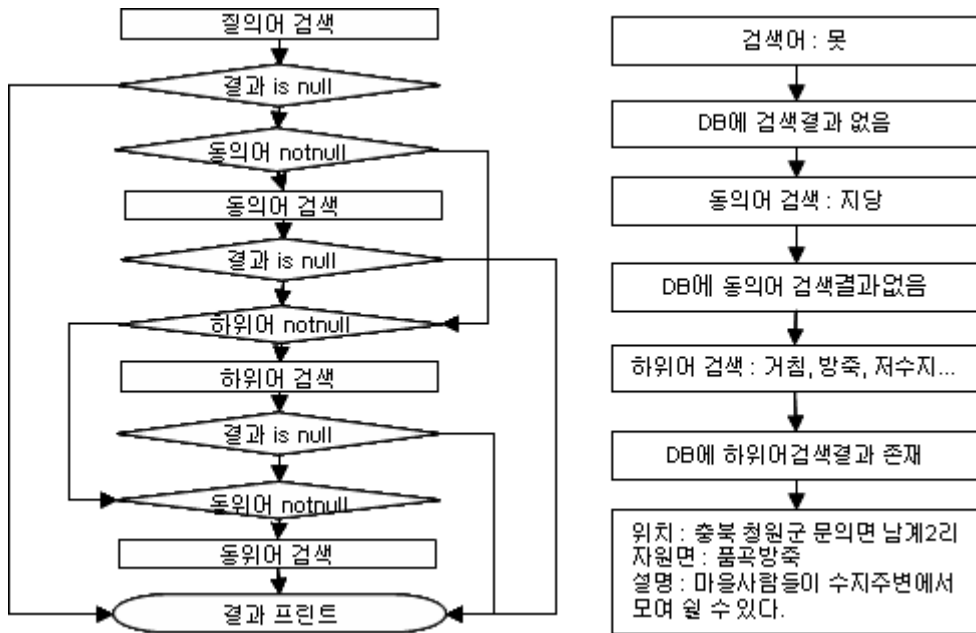


그림 4 온톨로지 활용한 검색 프로세스

기존의 단순 검색은 일치하는 자료가 없을 경우에는 답변을 하지 못하지만, 온톨로지를 통하여 동의어, 하위어, 동위어 검색이 가능함을 알 수 있다. 이러한 기능 외에도 온톨로지가 더욱 확충되고 각 개념간의 관계가 정립된다면, 보다 효율적인 정보검색이 이루어질 수 있을 것이다.

V. 결론

인터넷이 일상생활화되면서 누구나 손쉽게 정보를 얻을 수 있게 되었으며, 또한 정보의 양의 증대로 정보의 공유 및 활용에서의 문제점이 지적되고 있다. 단순히 정보의 일반적인 전달은 그 양이 많아 사용자에게 오히려 혼란을 가중하게 된다.

농촌어메니티자원은 자원종류가 각기 다르며 서로 상이한 속성을 갖고 있기 때문에, 농촌어메니티정보를 전부나 혹은 일부만을 보여줄 경우 농촌어메니티를 제대로 표현할 수 없을 뿐만 아니라 사용자가 원하는 정보를 제공할 수 없게 된다. 따라서 본 연구에서는 온톨로지를 활용하여 사용자가 원하는 정보를 보다 정확하게 검색할 수 있도록 농촌어메니티자원 정보시스템을 설계하였으며 일부 온톨로지를 구축하고 예제를 통하여 그 효율성을 검증해 보았다.

예제를 통하여 일반검색과는 달리 온톨로지를 활용하여 확장검색이 가능함을 살펴볼 수 있었으며, 온톨로지와 데이터베이스가 보다 확충된다면 다양한 자원간의 관계를 통하여 사용자가 원하는 정보제공에 효율적인 정보시스템이 가능할 것으로 예상된다.

본 연구는 농촌진흥청 농촌자원개발연구소에서 지원받은 "농촌어메니티 자원 조사사업"에 의한 결과 중 일부입니다.

참고문헌

박창석, 전영옥, 조영국, 2002, 농촌어메니티에 기초한 농촌자원 중요도 평가 및 순위적 관계분석, 대한국토도시계획학회지 37권 6호

김상범, 이상영, 정남수, 이지민, 조순재, 이정재, 2004, 자료정규화를 통한 농촌어메니티자원 조사표의 표준화, 농촌계획 10권 4호

한국전산원, 2004, 웹 온톨로지 개발 지침 연구, 한국전산원 연구보고서

장태우, 2004, 전자상거래 통합을 위한 온톨로지 구축, 서울대학교 박사학위논문

<http://www.fao.org/agrovoc>

<http://www.ksl.stanford.edu/projects/wine/explanation.html>