

공간정보사업의 중복사업 검색을 위한 의미기반검색 시스템의 설계*

박상언** · 임재익*** · 강주영****

Design of Semantic Search System for the Search of Duplicated Geospatial Projects*

Sangun Park** · Jay Ick Lim*** · Juyoung Kang****

■ Abstract ■

Geospatial information, which is one of social overhead capital, is predicted as a core growing industry for the future. The production of geospatial information requires a huge budget, so it is very important objective of the policy for geospatial information to prevent the duplication of geospatial projects. In this paper, we proposed a semantic search system which extracts possible duplication of geospatial projects by using ontology for geospatial project administration. In order to achieve our goal, we suggested how to construct and utilize geospatial project ontology, and designed the architecture and process of the semantic search. Moreover, we showed how the suggested semantic search works with a duplicated projects search scenario. The suggested system enables a nonprofessional can easily search for duplicated projects, therefore we expect that our research contributes to effective and efficient duplication review process for geospatial projects.

Keyword : Semantic Web, Semantic Search, Geospatial Project, PC Web Page, Mobile Web Page, Transformation Technique

논문투고일 : 2013년 07월 26일 논문수정완료일 : 2013년 09월 08일 논문게재확정일 : 2013년 09월 12일

* 이 논문은 2012년, 2013년 국토연구원 국가공간정보정책 총괄 및 조정사업의 지원을 받아 연구되었음.

** 경기대학교 경영정보학과

*** 아주대학교 e-비즈니스 학과, 교신저자

**** 아주대학교 e-비즈니스 학과

1. 서론

공간정보는 도로, 철도, 항만과 같은 사회간접자본으로 공장이나 백화점 혹은 도로 등을 지을 때 중요한 역할을 하는 정보이며 이로 인해 다양한 시설물과 관련하여 생산, 유통, 소비 등 경제활동의 기반이 된다[1]. 따라서 막대한 자본이 투자되고 투입자본의 회수에 장기간이 소요되므로 정부를 중심으로 투자가 이루어진다. 선진국의 국가공간정보기반을 보면 공간정보가 사회간접자본이라는 인식이 잘 나타나며, 공간정보의 생산, 유통, 활용을 촉진하기 위한 많은 정책을 추진하고 있다. 또한 공간정보는 다양한 기관에서 미래의 핵심적인 성장산업으로 예측되고 있으며, 모든 행정정보의 80% 이상이 공간정보와 관련이 되어 있기 때문에[9], 전자정부의 핵심기반이기도 한다. 이로 인해 공간정보는 중요한 국가자산으로 인식되고 있으며 철저히 관리되고 있다.

공간정보정책의 목적은 공간정보의 중복구축을 막고 공간정보의 공유 및 연계로 융복합 시너지효과를 창출하는데 있다[1]. 특히 공간정보의 구축에는 막대한 예산이 소요되기 때문에 중복구축을 막는 것은 매우 중요한 작업이다. 그럼에도 불구하고 공간정보의 중복구축을 방지하는 것이 쉽지 않은 이유는 기존에 구축된 공간정보사업들과 다양한 지자체 및 중앙기관에서 올라오는 사업계획들간의 중복성을 검토하는 것이 매우 어려운 작업이기 때문이다. 구체적으로 그 원인을 살펴보면, 첫째 중복성을 검토할 수 있는 기준이 정립되어 있지 않으며, 둘째 중복성 검토 대상이 되는 사업이나 산출물들이 매우 광범위하고, 셋째 중복성 검토가 전문가의 지식 및 경험에 의존하고 있기 때문이다. 공간정보사업의 중복성을 검토하는 과정과 관련지식을 검토하여 이를 지식화 및 체계화하면 이전 시스템에 비해 쉽게 중복사업을 검색하는 것이 가능하다.

의미기반검색은 시맨틱 웹 기술과 온톨로지를 활용함으로써 사용자가 원하는 바를 보다 정확하

게 검색할 수 있도록 하기 위한 도구이다[20]. 공간정보사업의 중복성 검토와 관련된 지식을 온톨로지로 정형화하고 기존의 중복성 검토 과정을 의미기반검색으로 구현한다면 중복성 검토 작업을 간편하고 정확하게 수행하는 것이 가능해진다. 이렇게 함으로써 공간정보사업을 기획하는 단계부터 누구나 중복성 검토를 쉽게 하도록 할 수 있다면 공간정보의 중복구축을 막는 작업을 현재의 시스템보다 효과적, 효율적으로 수행할 수 있을 것으로 기대된다.

따라서 본 논문에서는 먼저 의미기반검색 프레임워크에 대해 기존연구를 살펴본 후에 이 프레임워크에 따라 공간정보사업에 대한 중복사업 검색을 위한 의미기반검색 시스템을 설계하고자 한다. 설계를 위해서 먼저 중복사업 검색을 위해 필요한 온톨로지와 온톨로지의 활용방안을 제시하고 나서 이를 기반으로 시스템 구조와 프로세스를 제안하고자 한다. 그리고 중복사업 검색 시나리오를 만들고 시나리오에 따라 시스템에서 어떻게 의미기반검색이 수행되는지 살펴보고자 한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 의미기반검색 프레임워크와 의미기반검색의 활용현황 그리고 공간정보 분야의 시맨틱 웹 표준화현황을 살펴본다. 제 3장에서는 의미기반검색을 이용한 공간정보 중복사업의 검색 방안에 대해 살펴보고, 제 4장에서는 본 논문에서 제안하는 의미기반검색 시스템의 구조와 프로세스를 설명한다. 제 5장에서는 중복사업 검색에 의미기반검색을 적용하는 시나리오를 설명하고, 제 6장에서는 프로토타입의 구현내용과 효과에 대해 서술하고 마지막으로 제 7장에서 결론을 서술한다.

2. 관련 연구

2.1 시맨틱 웹과 의미기반검색 프레임워크

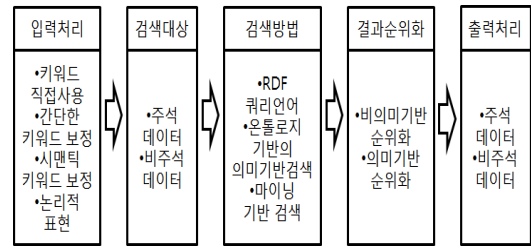
시맨틱 웹은 일반적으로 소프트웨어 혹은 기계가 웹에 존재하는 정보를 이해하고 이를 처리할

수 있는 기술로 이해된다. 팀 버너스 리는[5] 웹에 존재하는 자원들에 대한 의미적 관계를 명시적으로 표현하기 위한 방안으로 시맨틱 웹(Semantic Web)을 정의하였다. 온톨로지를 이용하면 이러한 자원들을 개념화하고 이들 간의 관계를 의미적으로 연결함으로써 보다 지능화 된 웹을 위한 기반을 마련할 수 있다.

최근 다양한 의미적 연결을 가진 데이터가 확산됨에 따라 시맨틱 웹의 활용성이 높아지고 있다. 그 예 중 하나인 링크드 데이터(Linked Data)[6]는 웹 환경에서 문서 중심이 아닌 데이터 중심의 웹 환경(Web of Data)을 구축하고자 하는 노력이다.

시맨틱 웹 기술의 발달에 따라 의미기반검색, 정보의 통합 및 분석, 그리고 데이터 마이닝 등 다양한 분야에서 시맨틱 웹의 활용이 가능해졌다[12]. 특히 의미기반검색에서는 소프트웨어가 웹 상의 자원 혹은 정보의 의미를 이해함으로써 사용자의 의도를 보다 정확히 반영하고 이에 일치하는 정확한 결과를 가져오는 검색이 가능해졌다[11]. 또한, 링크로 연결되어 있는 자원 혹은 정보들을 탐색함으로써 구조 속에 내포되어 있는 다양한 의미를 찾아낼 수 있는 마이닝도 가능해졌다[11]. 이러한 의미기반검색 및 마이닝을 활용함으로써 키워드 기반의 막연한 검색에 비해 정확한 검색을 수행하고 데이터 속에 숨어 있는 지식들을 발견하는 것이 가능해지고 있다[11, 16].

의미기반검색에 대한 공통적인 정의를 찾기는 어려우나, 최근 연구에서 의미기반검색을 “사용자의 명시적인 의도뿐만 아니라 묵시적인 의도를 이해하기 위해 온톨로지와 기타 방법들을 이용하여 사용자가 제시한 키워드 검색어에 대하여 원하는 결과들을 검색하는 과정”으로 정의하였다[11, 20]. 의미기반검색을 보다 명확히 이해하기 위해서는 의미기반검색의 프레임워크를 살펴 볼 필요가 있다. [그림 1]에서와 같이, 의미기반검색 프레임워크는 입력처리, 검색대상, 검색방법, 결과 순위화, 출력데이터 유형의 다섯 구성요소로 이루어져 있다[20].



[그림 1] 의미기반검색 프레임워크[20]

입력처리(Input Processing)는 사용자로부터의 키워드를 처리하는 과정이다. 일반적인 키워드 검색이 만족스럽지 않은 이유는 입력 키워드의 의미적인 모호성 때문이므로 사용자가 명시적 혹은 묵시적으로 의도하는 바를 알기 위해서는 입력 키워드의 정확한 의미를 알아내는 과정이 요구된다[20].

검색대상(Target Source)은 검색엔진이 검색을 수행해야 하는 대상으로서, 일반적인 웹 환경에 해당하는 주석이 없는 데이터(Un-annotated Data)와 RDF/S[7, 17], OWL[4] 등의 시맨틱 마크업 언어로 주석이 되어 있는 데이터로 나뉜다[20].

검색방법(Search Methodology)은 검색엔진이 검색을 수행하는 방법으로, RDF 쿼리언어(RDF Query Language)는 SPARQL[13]과 같은 RDF 쿼리 언어를 이용하여 검색하는 것을 말하며, 사용자가 입력한 검색어를 구조적인 쿼리 언어로 변환한 후에 이를 실행함으로써 원하는 결과를 도출한다. 이 외에 온톨로지 기반 의미기반검색(Ontology-based Semantic Search)과 마이닝 기반 검색(Mining-based Search)이 있다[20].

결과 순위화(Result Ranking)는 검색결과를 주어진 기준에 따라 순위화하여 출력하는 단계를 말한다. 출력처리(Output Data Type)는 검색결과 형태를 지정하며, 타 시스템과의 연동이 필요한 경우에는 검색결과를 주석이 있는 데이터(Annotated Data)로 시맨틱 주석을 붙여서 출력하고, 그냥 사용자에게 보여주기 위해서는 주석이 없는 데이터(Un-annotated Data)로 소스에 대한 주석처리 없이 원본을 출력한다[20].

2.2 의미기반검색 적용 사례

의미기반검색을 적용하고자 한 연구로 시맨틱 웹 기술을 호텔 검색에 활용한 연구가 있다[24]. 이 논문에서는 여행객들의 다양한 요구를 미리 정형화된 틀로 입력하도록 한 후, 이를 SPARQL 기반의 질의로 변환하여 원하는 결과를 가져오도록 하였다. “깨끗한 호텔”, “안전한 호텔” 등의 모호한 질의를 SWRL[14]을 이용한 규칙으로 표현하고 이러한 규칙에 따라 질의를 수행했다는 점이 특징이다[14]. 또 다른 연구로 부동산 지식 정보시스템 설계에 의미기반검색을 활용한 연구가 있다[8]. 이 연구 역시 매우 유사한 접근방법을 이용하였으며, 호텔 대신 부동산을 대상으로 하여 이용자의 다양한 요구를 표현할 수 있는 온톨로지를 설계하고 주관적인 인지용어에 대해 SWRL[14]로 표현하여 처리하였다. 역시 웹 기반의 정형화된 인터페이스를 통하여 이용자의 요구를 표현하도록 하고, 이를 SPARQL[13]로 변환하여 검색을 수행하였다.

논문이 아닌 실제 검색 서비스에서 의미기반검색을 활용한 첫째 예로 구글의 지식그래프(Knowledge Graph)를 들 수 있다[10]. 지식그래프는 구글 검색엔진의 검색결과를 향상시키고자 사용된 지식베이스로서, 2012년 구글 홈페이지를 통해 개념이 발표되었으며 영어 검색에서 서비스를 시작하여 점차 확대하고 있다. 다양한 소스로부터 수집된 특정 토픽에 대한 상세정보들을 구조적으로 연결함으로써 검색의 결과를 향상시키는 것이 목표이며, 상호 연관성이 있는 데이터들을 속성에 따라 연결하고 사용자가 이러한 연결 관계를 따라가도록 함으로써 기존과는 달리 정확하고 연관성 있는 정보들을 얻을 수 있도록 하였다. 최근 국내에서 개발된 다른 서비스 사례로는 국토포털 인문지리 시맨틱 검색이 있다[21]. 이는 인문지리정보의 검색의도를 해석하여 정보를 제공하는 시맨틱 포털 서비스로서, 사용자의 검색의도를 분석하여 결과를 분류별로 제공하며 연관정보를 함께 제공

함으로써 검색 결과의 정확성을 높이고자 하였다. 3차원 지도 기반의 위치정보와 결합하여 지리정보를 동시에 제공하고 있으며 또한 시기별로 중요 내용을 담고 있는 자료에 대해서는 타임슬라이드를 이용하여 시계열 정보를 제공하고 있다.

2.3 공간정보 분야 시맨틱 웹 표준화 및 적용 현황

공간정보 분야의 시맨틱 웹 표준화 현황과 관련하여 먼저 지리정보 국제표준화기구에서는 지리정보에 대한 시맨틱 웹 표준을 추진하고 있다. 여기서 추진되는 ISO 19150 프로젝트는 온톨로지를 지리정보의 상호운용성 향상에 활용하기 위한 방안을 개발하는 것을 목표로 하고 있다[15]. 최근 ISO 19150-3에서는 지리정보 온톨로지에서의 시맨틱 연산에 대한 내용을 논의하고 있으며, ISO 19150-4에서는 지리정보 분야의 서비스 온톨로지에 대해 논의하고 있는 중이다[15].

W3C는 2007년에 W3C 인큐베이터 그룹 보고서 발간을 통해 공간정보 온톨로지의 필요성과 중요성을 보이고 온톨로지를 활용하는 업무 예제들을 정리하였다[23]. 이 보고서에서 지식의 지리적인 속성들을 기계(프로그램)에서 표현하고 발견하며 실행할 수 있도록 하고 다양한 조직 및 기관에서 이를 활용할 수 있도록 하는 지리정보 시맨틱 웹(Geospatial Semantic Web)의 개념을 제시하였다[23].

OGC(Open Geospatial Consortium)는 개방형 공간정보 컨소시엄으로서, GeoSemantics 워킹그룹을 통해 꾸준히 공간정보 분야에서의 시맨틱 웹 활용을 추진하고 있다[19]. 2009년 시맨틱 웹 환경에서 공간정보를 검색하기 위한 GeoSPARQL 표준안을 제안하고 표준화 워킹그룹을 통해 지속적으로 내용을 보완하고 있다[19]. 2010년부터는 Geo-Semantic 표준을 위한 참조모델을 개발하기 시작하였으며, 2012년까지 정기적인 모임을 통해 지속적으로 표준화 방안을 논의하고 있다[19].

SOCoP(Spatial Ontology Community of Practice)은 공간정보 온톨로지 표준 개발을 위한 온라인 포럼으로[22], OGC와 연계하여 공간정보 분야에서의 시맨틱 웹 활용과 온톨로지 표준 구축을 위한 활동을 수행하고 있으며 활동내용과 개발내역 등을 온라인 포럼을 통해 게시하고 있다.

공간정보 분야에서 온톨로지를 활용하는 방안에 대해서도 활발하게 연구가 진행되고 있다. 건축에서 기본적인 건축공간의 구조와 상호관계를 온톨로지로 표현하고자 한 연구에서는 건물공간을 개념화하고 OWL을 이용하여 건물의 공간과 벽체 등을 표현하고 공간의 기능을 추론하도록 하였다[2]. 공간정보 중에서 지하공간정보를 대상으로 하여 온톨로지를 개발하고 검증한 연구에서는 상위 온톨로지, 측정온톨로지, 공간온톨로지로 레이어를 설계하고 MDA를 이용하여 시맨틱 공유를 목적으로 온톨로지를 개발하였다[3].

이상과 같이 공간정보에 대한 시맨틱 웹 표준화 및 활용에 대한 노력이 다양하게 진행되고 있으나, 본 연구에서 공간정보사업의 산출물 유형을 표현해줄 수 있는 표준은 찾을 수 없었다. 이는 이상의 표준들이 공간정보 자체에 대한 표준인 것이 비해 본 연구는 공간정보사업과 산출물을 대상으로 한다는 것이 가장 큰 이유라고 할 수 있다. 따라서 본 연구에서는 자체적으로 필요한 온톨로지를 설계하고 구축하여 이용하고자 한다.

3. 의미기반검색을 이용한 공간정보 중복사업 검색방안

3.1 공간정보사업 중복사업 검색의 문제점 및 공간정보사업 온톨로지의 필요성

공간정보사업의 예산 규모가 크기 때문에 중복사업을 방지하는 것이 매우 중요한 일임에도 불구하고 지금까지 공간정보사업에 대한 중복성을 검토하는 것이 쉽지 않았던 데에는 다음과 같은 이유가 있다.

첫째, 중복성을 검토할 수 있는 구체적 기준이 정립되어 있지 않기 때문에 공간정보사업의 담당자가 스스로 공간정보사업의 중복성을 검토하기 어렵고 검토자 역시 이를 명확하게 판단하기 어렵다는 문제가 있다. 제안된 공간정보사업에 대한 사전검토를 위해 검토지표 및 검토대상이 제공되거나 이에 대한 설명이나 가이드라인이 없기 때문에 검토자의 주관에 따라 평가될 수밖에 없다는 문제점이 존재한다. 또한 이로 인해 중복성 검토에 대한 기준이나 주제가 검토자의 관점에 따라 달라질 수 있다.

둘째, 중복성 검토 대상이 되는 사업이나 산출물들이 매우 광범위 하다. 다양한 중앙기관 및 자치단체에서 공간정보사업을 수행할 뿐만 아니라, 기존에 수행된 공간정보사업 및 산출물에 대한 자료가 다양한 기관에 분산되어 있기 때문에 중복성 검토를 위해서는 이러한 광범위한 대상 및 소스를 모두 직접 접촉하여 필요한 자료를 확보해야 한다. 그러나, 현재로서는 이러한 검토 대상 및 소스들을 효과적으로 검색할 수 있는 시스템적인 지원이 없다. 자료들을 구할 수 있다 하더라도 중복성 여부를 확인시켜줄 수 있는 시스템적인 지원이 역시 부족하다.

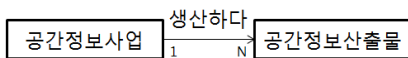
셋째, 중복성 검토가 전문가의 지식 및 경험에 의존하는 경향이 있다. 실제로 담당자의 지식 및 경험이 중복성 검토에 결정적 영향을 미치고 있으며 이로 인해 담당자의 기억에 의존하여 중복성 검토가 이루어지는 상황이 많다. 예를 들어 유사 혹은 중복이 의심되는 두 사업의 사업계획서를 직접 비교 검토한다 해도 전문적 지식이 없이는 중복성을 판단하기 어렵다. 이는 사업계획서 내용에 중복성을 판단할 수 있는 자료가 부족하기 때문이다. 따라서 중복성 검토 기준의 정립으로부터 시작하여 중복성 검토를 위한 자료의 준비까지 체계적이고 일관성 있는 작업이 반드시 선행되어야 한다.

이상과 같은 문제점들을 해결하기 위해서는 유사 및 중복사업의 기준 및 판단의 근거를 제시하는 공간정보사업 온톨로지가 요구된다.

즉, 중복사업의 검토를 위해서는 어떤 유형의 공간정보사업이 중복성 검토 대상이 되는 지를 제시해주는 기준이 필요한데, 이를 위해서는 중복성 검토 관련 내용의 지식화가 선결과제이며 중복성 검토를 위한 기준의 정형화가 필요하다. 예를 들어 유사 및 중복 관점의 분류체계가 온톨로지에서 제시된다면 두 사업이 동일한 분류체계에 속하는 경우, 이 두 사업은 유사 혹은 중복의 가능성이 있음을 파악할 수 있다. 또한 중복성 검토 대상 및 내용에 대한 전문가의 지식 및 경험을 정형화함으로써 유사 혹은 중복성의 가능성을 비전문가도 알 수 있도록 하고, 시스템에 이러한 지식을 체계화함으로써 시스템적인 추천이 가능하도록 하여야 한다.

3.2 공간정보사업 온톨로지 예시 및 활용방안

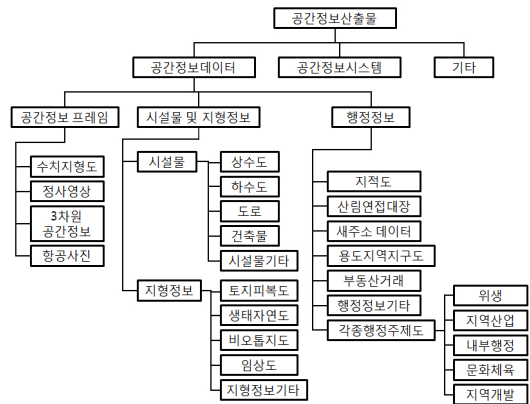
먼저 정확한 검색을 위해 공간정보사업과 산출물의 분리할 필요가 있다. 하나의 공간정보사업은 다양한 형태의 여러 산출물을 생산할 수 있음에도 불구하고 지금까지 공간정보사업 단위로 데이터베이스를 구축하였기 때문에 사업에서 생산되는 산출물들을 전부 표현하지 못하는 단점이 있었다. 그러나, 공간정보사업의 유사 혹은 중복을 검토하기 위해서는 산출물을 모두 분석하여야 하며, 산출물 중 하나라도 유사 혹은 중복의 가능성이 있으면 이를 검색할 수 있어야 한다. 따라서 공간정보사업과 산출물을 분리하여 둘 간의 관계를 표현하고, 사업 단위로 표현하던 부정확한 속성들을 산출물 단위로 표현함으로써 검색의 정확성 향상이 가능하다. 그리고 이에 맞추어 사업 단위의 분류체계 보다는 산출물 단위의 분류체계를 사용하는 것이 바람직하다. [그림 2]는 하나의 사업이 여러 산출물을 생산하는 관계를 도식화한 것으로 사업과 산출물을 연결함으로써 산출물 단위의 유사 및 중복성 검토를 통해 사업에 대한 검사도 가능하도록 하였다.



[그림 2] 공간정보사업과 산출물의 관계

둘째로 앞서 설명한 바와 같이 공간정보 산출물에는 유사 및 중복을 표현할 수 있는 분류체계가 요구된다. 이 때 공간정보 산출물은 하나의 관점에서만 분류를 정의하기 어렵고, 복합적인 관점에서 분류되어야 한다. 예를 들어 하나의 공간정보 산출물 중에서 공간정보데이터와 관련된 산출물은 공간정보 프레임 관점, 시설물 및 지형정보 관점, 그리고 행정정보 관점 등에서 분류를 정의할 수 있다. [그림 2]는 이러한 세 가지 관점의 분류체계를 반영한 공간정보 데이터의 분류체계 예시를 보여준다.

공간정보 프레임 관점에서는 수치지형도, 정사영상, 3차원 공간정보 등의 분류체계를 가질 수 있으며, 시설물 및 지형정보 관점에서는 시설물과 관련하여 상수도, 하수도, 도로, 건축물 등의 분류체계를 가지고, 지형정보와 관련하여 토지피복도, 생태자연도, 비오톱지도, 임상도 등의 분류체계를 가질 수 있다. 마지막으로 행정정보와 관련해서는 지적도, 산림연접대장, 용도지역지구도, 부동산거래 등의 분류체계를 가질 수 있다.

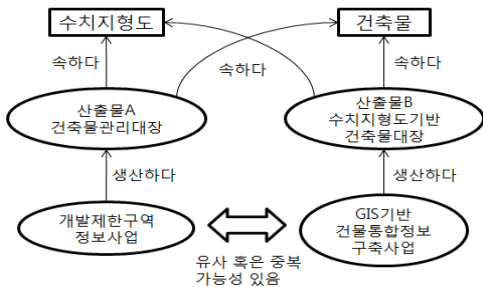


[그림 3] 공간정보산출물 분류체계 예시

이러한 공간정보 산출물 분류체계를 통해 유사 혹은 중복의 가능성이 있는 사업을 파악하는 것이 가능하다. [그림 3]과 같이 공간정보 분류체계는 상/하위 관계를 갖고 있으며 하위의 분류는 상위의 분류의 일부에 속하는 체계로서, 공간정보 산출물들을 체계적으로 분류할 수 있는 도구로 활용이 가능하다. 공간

정보 산출물들을 이 분류체계에 따라 분류하게 될 경우, 동일한 분류에 있으면 유사 및 중복 가능성이 있는 산출물로 파악이 가능하다. 이 때, 하나의 산출물은 각각 세 관점에 있는 분류체계에 모두 속할 수 있다. 예를 들어 특정 산출물은 공간정보 프레임 관점에서 수치지형도에 속하고, 시설물 및 지형정보 관점에서는 건축물에 속하고, 행정정보 관점에서는 용도지역지구도에 속할 수 있다. 이러한 상황에서 산출물이 속해 있는 분류체계 중 어느 하나라도 동시에 속해 있는 다른 산출물이 있다면 이 산출물과 유사 혹은 중복의 가능성이 있다고 판단할 수 있다.

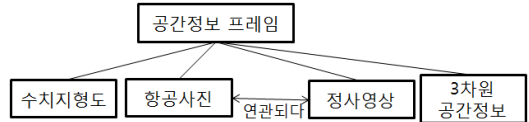
[그림 4]는 분류체계를 이용한 유사 혹은 중복 가능성 검출의 예를 보여준다. 개발제한구역정보사업과 GIS 건물통합 정보사업은 각기 건축물관리대장과 수치지형도기반 건축물대장을 산출물로 생산하는데 두 산출물은 동시에 수치지형도와 건축물 분류체계에 속하므로 유사 및 중복 가능성이 있다. 이를 기반으로 두 사업 간의 유사 및 중복 가능성을 검출하는 것이 가능하다.



[그림 4] 분류체계를 이용한 유사 및 중복사업 검출 예시

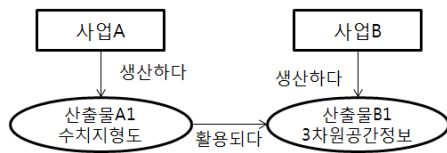
공간정보사업의 산출물 간 유사도를 평가할 수 있는 가장 기본적인 방안으로 분류체계를 사용하는 것 외에, 공간정보사업의 특성을 살려 온톨로지의 추가정보를 구축하고 이를 활용하는 것이 가능하다. 그러한 첫째 예로 공간정보 산출물의 세부 분류 간 연관성 표현을 들 수 있다. 공간정보 산출물이 서로 다른 분류에 속해 있더라도 두 분류 간에 상호 연관성이 있는 경우에는 유사 및 중복의 가능성을 검토할 필요가 있다.

[그림 5]는 분류간의 연관성을 표시한 예를 보여 준다. 공간정보 산출물 분류체계에서 각 분류들은 ‘연관되다’라는 관계를 이용하여 상호 연관성을 표시할 수 있다. [그림 4]에서 “공간정보 프레임”의 하부 분류체계인 항공사진과 정사영상은 비록 별도의 분류체계로 관리되지만 상호 연관성이 강하기 때문에 양쪽에 속한 산출물들은 그 유사성을 반드시 검토해야만 한다. 따라서 서로 다른 분류체계에 속하더라도 상호 연관성이 있는 분류체계인 경우에는 유사 및 중복의 가능성이 있다고 보고 검토대상으로 추천할 수 있다.



[그림 5] 공간정보 산출물의 세부 분류 간 연관성 표현 예시

세부분류간 연관성 외에 공간정보 산출물 간의 생태계를 표현함으로써 이를 활용하는 것이 가능하다. 일반적으로 하나의 공간정보 산출물은 다른 산출물에서 기본자료로 사용이 되는 경우가 많다. 이러한 공유 및 연계에 대한 정보를 표현하고 유지할 경우, 산출물 혹은 사업 간의 연관성을 파악하기가 쉽고 이를 이용하여 보다 효율적인 공간정보의 활용이 가능하다. [그림 6]은 공간정보 산출물 간의 상호 활용현황을 표현한 예를 보여 준다. 산출물들은 ‘활용되다’ 관계를 이용하여 상호 연결될 수 있다. 예를 들어 그림의 사업 B가 생산하는 산출물 B1 ‘3차원공간정보’는 사업 A가 생산하는 산출물 A1 ‘수치지형도’를 활용하여 만들어졌음을 그림과 같이 표현하는 것이 가능하다.



[그림 6] 공간정보 산출물 간 활용현황 표현 예시

이상과 같은 방법으로 두 산출물이 속해 있는 분류체계와 분류체계 간의 관계를 이용하여 유사 및 중복 가능성이 있는 사업의 검출이 가능하나, 두 사업에 대해 보다 세부적으로 유사 및 중복 가능성을 검토하기 위해서는 산출물의 속성을 파악할 필요가 있다. 예를 들어 두 산출물이 동일한 수치지형도에 속해 있다고 하더라도 축척이 각기 1/500과 1/1500으로 다르다면 중복의 가능성은 낮아진다고 할 수 있다. 산출물의 속성은 분류체계에 따라 달라지므로 각 분류체계마다 고유한 속성들을 정의하여야 한다.

4. 중복사업 검색을 위한 의미기반 검색 시스템 구조 및 프로세스

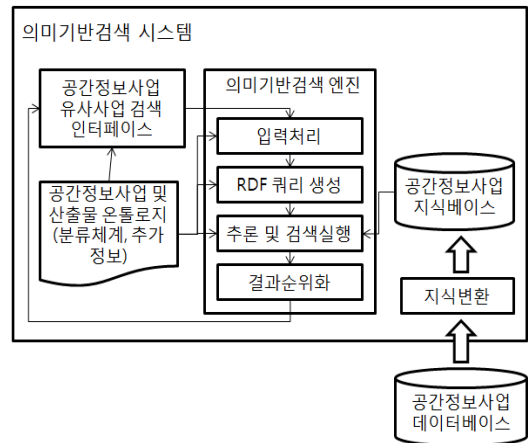
4.1 의미기반검색 시스템 구조

본 논문에서 제안하는 공간정보사업에 대한 중복사업 검색을 위한 의미기반검색 시스템은 인터페이스, 온톨로지(분류체계와 추가정보), 의미기반 검색엔진, 그리고 지식베이스로 구성된다. [그림 6]은 의미기반검색 시스템의 구조를 보여 준다. 그림에서 인터페이스는 사용자로부터 검색에 필요한 정보를 받아 이를 검색엔진에 전달하고, 검색이 종료된 후 검색결과를 사용자에게 보여주는 역할을 수행한다.

온톨로지는 의미기반검색을 위해 필요한 사업 및 산출물에 대한 메타정보들 즉 앞서 설명한 분류체계와 추가정보들을 담고 있다. 이 정보는 기본적으로 검색엔진에서 활용되며, 이 틀에 따라 사용자로부터 필요한 정보를 입력받을 수도 있으므로 인터페이스에서도 사용될 수 있다. 의미기반검색엔진은 인터페이스로부터 검색에 필요한 정보를 받아 이를 처리하는 역할을 한다. 의미기반검색 프레임워크에서 설명된 바와 같이 입력처리, 쿼리 생성 후 쿼리를 통한 검색의 실행, 그리고 결과를 순위화하는 역할을 수행한다.

공간정보사업 지식베이스는 온톨로지에서 정의

된 형태에 따라 실제 사업들의 정보를 주석이 있는 데이터 형태로 저장하는 곳이다. 이 정보들은 기존의 공간정보사업 데이터베이스로부터 읽어 들여 변환할 수 있다. 향후에는 데이터베이스로부터 실시간으로 변환하여 사용하는 것도 가능하다.



[그림 7] 의미기반검색 시스템 구조

[그림 7]에서 검색 인터페이스는 주어진 상황에 맞는 다양하게 제공될 필요가 있다. 예를 들어 사업담당자 혹은 사업검토자가 정해진 틀에 따라 사업의 개요를 입력하고 유사 및 중복 사업을 검색하고자 하는 경우에는 입력된 내용으로부터 구조적 표현을 생성하고 이를 SPARQL[13] 기반의 쿼리로 변환함으로써 검색을 수행하여 유사 및 중복 가능성이 있는 사업들을 리스트로 추천할 수 있다. 반면에 사업담당자 혹은 사업검토자가 임의의 키워드 조합으로 자유롭게 연관성이 있는 사업들을 검색하고자 하는 경우에는 시맨틱 키워드 보정을 통해 키워드 조합으로부터 사업에 대한 구조적 표현을 생성한 후에 검색을 수행하는 것도 지원할 수 있다. 마지막으로 사업담당자 혹은 사업검토자가 사업과 산출물 간의 연관관계를 따라가면서 자유롭게 사업들을 검색하고자 하는 경우에는 자유로운 검색이 가능하도록 탐색(navigation) 방식도 지원이 가능하다. 그러나, 정해진 틀에 따라 사업의 개요를 입력하면 가장 정확한 검색이

가능하므로, 본 논문에서는 입력된 내용을 기반으로 SPARQL 기반의 쿼리를 생성하여 검색을 수행하는 프로토타입을 구현하였다.

4.2 의미기반검색 시스템 프로세스

본 논문에서 제안하는 공간정보사업에 대한 중복사업을 검색하기 위한 의미기반 시스템의 검색 프로세스는 다음과 같다. 우선 본격적인 검색 프로세스에 앞서 선행단계로 의미기반검색을 위한 검토대상 사업의 구조적 표현이 요구된다. 검토의 대상이 된 사업을 사업과 산출물이 분리된 형태로 표현하고, 사업과 산출물 간의 연관성을 표시한 후에, 각 산출물에 대해서는 그 산출물이 속해 있는 분류체계들을 연결하여야 한다. 그리고 산출물이 속한 분류체계에 따라 각 산출물의 속성을 표시하고, 사업에 대해서도 필요한 속성들을 표시하여야 한다. 그리고 기존 사업들도 이러한 구조적 표현에 따라 표현되어 저장되어 있다는 가정 아래, 검토대상 사업과 기존 사업들의 구조를 비교함으로써 유사사업에 대한 의미기반검색이 가능해진다.

검색의 첫 단계는 구축지역에 따른 중복가능 사업의 검색이다. 구축지역은 사업의 속성으로 표현되어 있으며, 기존의 사업들 중에서 사업의 구축지역이 동일한 사업들만 유사사업 검토의 대상이 된다. 이 때, 구축지역에 대한 행정정보 역시 온톨로지를 이용하여 구축함으로써 구축지역에 대한 확장된 비교를 가능하도록 하였다. 예를 들어 현재 검토대상 사업의 구축지역이 수원시인 경우에 구축지역이 동일하게 수원시인 기존 사업뿐만 아니라 수원시의 상위(예 : 전국, 경기도)와 하위(예 : 팔달구, 영통구) 분류체계에 속하는 모든 지역에 대해서도 유사사업 검토대상이 되어야 한다.

검색의 둘째 단계는 사업의 산출물 분류체계에 따라 유사사업을 검색하는 단계이다. 검색 첫 단계에서 대상이 된 모든 사업들을 대상으로 현재 사업의 모든 산출물에 대하여 동일분류체계에 있는 산출물들을 검색함으로써 유사사업을 검색한다. 이

때 [그림 4]의 예와 같이 동일 분류체계에 속하는 산출물을 생산하는 사업이 유사 및 중복 검토대상이 될 뿐만 아니라, [그림 5]와 같이 연관성이 있는 분류체계에 속하는 산출물도 유사 및 중복성을 검토하기 위해 대상으로 선정된다.

셋째 단계는 사업 및 산출물의 속성을 비교하고 유사도를 결정하여 유사사업을 순위화하는 단계이다. 산출물은 동시에 여러 분류체계에 속할 수 있으므로 두 산출물은 하나의 분류체계뿐만 아니라 여러 개의 분류체계에 동시에 속하는 것이 가능하다. 따라서 동시에 속한 분류체계의 수가 많을수록 높은 순위를 갖게 된다.

또한 동일 분류체계에 속하는 경우는 연관된 분류체계에 속하는 경우보다 순위를 높게 책정되며, 분류체계가 일치하더라도 두 사업 간에 시간 차이가 크면 유사 및 중복성이 떨어지므로 두 사업 간의 시간이 가까울수록 높은 순위를 갖게 된다. 마지막으로 분류체계에 따른 속성 즉 축척이나 해상도와 같이 그 분류체계의 고유한 속성에 대한 속성값이 가까울수록 높은 순위를 갖게 된다. 이상의 순위기준에 따라 유사사업을 순위화하여 유사도가 가장 높은 사업부터 출력하면 사업 검토자 혹은 담당자는 주어진 순위와 세부 속성들에 따라 유사 및 중복성을 판단하게 된다.

5. 의미기반검색 적용 시나리오

공간정보사업은 크게 중앙정부 사업과 지자체 사업으로 분류될 수 있는데, 여기서는 2012년도의 지자체 사업 중 하나를 대상으로 유사사업을 검토하는 과정을 대상으로 함으로써, 기존 검색의 한계를 보이고 의미기반검색의 효과를 보이고자 한다. 대상 사업은 “고정밀 멀티 공간정보 시범구축 사업”이며 사업의 개요와 중복 검토 대상이 되는 사업들을 나열하고, 기존 검색에 의해 중복 사업이 도출될 수 없음을 보인 후에, 의미기반검색에 의해 실제로 중복가능성이 있는 사업을 찾아가는 과정을 보이고자 한다.

5.1 중복사업 검색 시나리오 개요

시나리오에서 유사 및 중복의 검토대상이 되는 사업의 이름은 “고정밀 멀티공간정보 시범구축 사업”이며 시행년도는 “2012년”, 대상지역은 “수원시”이고 사업내용은 “정사영상 제작, 1/500 수치지형도, 3차원 공간정보”이다.

시나리오에서 사용할 비교대상 사업 리스트는 <표 1>과 같다. <표 1>은 현재 검토 대상인 사업과 관련이 있는 사업들의 리스트를 보여 준다. 이 중에서 상위 두 건은 구축지역이 달라 실질적인 유사 및 중복 가능성이 있는 사업이 될 수 없다.

<표 1>에서 셋째 줄에 있는 국립지리원의 “1대 1000 수치지형도 수정제작 수원지구” 사업은 수치지형도를 생산하기 때문에 검토대상 사업과 중복 가능성이 있다. 또한 마지막 줄에 있는 “2009년 디지털 항공사진 촬영 및 판독용역”은 항공사진을 산출물로 생산하고 있기 때문에 직접적인 연관성은 없어 보이나, 항공사진과 정사영상은 상호 간에 밀접한 연관성이 있기 때문에 유사 및 중복의 가능성이 있다.

5.2 기존 검색의 한계

기존 검색에서는 전문지식이 없는 경우 검토 자체에 많은 어려움이 있었다. 우선 검토대상 사업은 제목만으로는 산출물에 대해 알기가 어렵고, 내용을 검토해야만 정사영상, 수치지형도, 3차원 공간정보 등을 생산하는 사업임을 알 수 있다. 특히 이 사업은 최종적으로 3차원 공간정보를 생산하는 사업이지만 그 과정에서 정사영상과 수치지형도가

생산되므로 부가적으로 생산되는 두 산출물에 대해서도 유사 및 중복 검토를 해야 하는데, 이는 전문지식이 없이는 파악하기 어렵다. 따라서 유사 및 중복사업 검토를 위해서는 수원시에서 집행된 사업 중에 정사영상, 수치지형도 혹은 3차원 공간정보를 생산하는 모든 사업들을 대상으로 검토를 수행해야 한다.

기존 검색에 있어 둘째 문제는 전문지식이 있다 하더라도 현재 검색 시스템으로는 지원이 어렵다는 점이다. 위에서 언급한 전문지식을 알고 있더라도 현재의 검색 시스템으로는 검색이 원활하지 않다. 먼저 구축지역이 동일하게 수원인 사업들 중에서 정사영상, 수치지형도, 3차원 공간정보를 생산하는 사업을 찾아야 하는데, 현재의 “공간조회” 시스템에서는 시 단위로는 검색이 안 되고 도 단위로만 검색이 되기 때문에 정확히 수원시를 대상으로 하는 사업을 찾는 것은 불가능하다. 지역에 대한 검색을 포기하고 산출물을 중심으로 사업을 개별 검색하는 경우에 현재 시스템에서는 제목으로만 검색이 가능하므로 3차원 공간정보, 정사영상, 수치지형도를 사업 제목과 매칭시켜 검색해야 한다. 그러나, 예를 들어 3차원 공간정보를 생산하는 사업을 검색하고자 하는 경우, <표 1>에서 둘째 줄과 같이 “디지털 입체영상” 등 다른 이름으로 사용되기 때문에 사업명으로서의 검색이 모든 결과를 가져온다고 보장할 수 없다. 수치지형도로 검색한 경우, 현재는 103개의 사업이 검색되는데 각 지역에서 생산된 이들 사업의 내용을 전부 확인해야 하는 번거로움이 있다.

<표 1> 시나리오 비교대상 사업 리스트

사업기관	사업명	산출물	대상지역	속성
지자체/광양시	광양시 3차원 공간정보 DB 고도화	디지털 정사영상	광양시 전역	축척 : 1/1000
지자체/목포시	2012년 광주시 디지털 입체영상 데이터베이스 구축사업	디지털 항공정사영상	목포시 전역	GSD : 25cm, 51cm
국립지리원	1대 1000 수치지형도 수정제작 수원지구	수치지형도 제작	수원시 전역	축척 : 1/1000
지자체/수원시	2009년 디지털 항공사진 촬영 및 판독용역	디지털 항공사진 촬영	수원시 전역	GSD : 10cm~15cm

5.3 의미기반검색을 위한 구조적 표현

검토대상 사업의 중복사업에 대한 의미기반검색을 위해서는 사업에 대한 정보를 온톨로지 기반의 구조적 표현으로 변환하여야 한다. [그림 8]은 현재 검토대상 사업인 “고정밀 멀티 공간정보 시범구축사업”에 대한 구조적 표현을 N3 형태로 보여준다. 사업(p7)의 구축지역(builtIn)이 “수원”이고, 세 개의 산출물(d72, d73, d71)을 생산하고 있으며, 시행년도(execYear)는 “2012년”임이 표현되어 있다. 산출물은 각각 “정사영상”(Orthoimagery), “3차원 공간정보”(3DSpace), “수치지형도”(DigitalMap)에 속해 있음을 나타낸다. “1/500 수치지형도”(d73)의 경우에는 속성으로 축척(scale)의 값이 “1/500”임을 나타낸다. 구조적인 표현에서는 사업수준의 정보와 산출물수준의 정보 그리고 산출물 분류체계를 서로 연결함으로써 산출물 분류체계를 통해 사업수준에서의 중복사업 검색이 가능하도록 하고 있다.

이 표현과 저장되어 있는 다른 사업들의 구조적 표현을 비교함으로써 유사 혹은 중복가능성이 있는 사업을 검색하는 것이 가능하다. 따라서 비교

```
@prefix default: <http://www.krihs.re.kr/SpatialProject.owl#>
default:p7
  a default:GeoProject ;
  rdfs:label "고정밀 멀티 시범구축사업"^^xsd:string ;
  default:builtIn default:수원 ;
  default:execYear "2012년"^^xsd:string ;
  default:produce default:d72, default:d73, default:d71 .
default:d72
  a default:Orthoimagery ;
  rdfs:comment "정사영상제작"^^xsd:string ;
  default:producedBy default:p7 .
default:d73
  a default:DigitalMap ;
  rdfs:comment "1/500 수치지형도"^^xsd:string ;
  default:producedBy default:p7 ;
  default:scale "1/500"^^xsd:string .
default:d71
  a default:3DSpace ;
  rdfs:comment "3차원 공간정보"^^xsd:string ;
  default:producedBy default:p7.
```

[그림 8] 검토대상 사업의 온톨로지 표현

대상 사업 리스트에 있는 사업들에 대해서도 구조적 표현이 요구된다. <표 1>에 있는 각 사업들은 위에서와 동일한 방식으로 표현되어 온톨로지에 저장된다. [그림 9]는 그 중 하나인 “1대 1000수치지형도 수정제작 수원지구” 사업에 대한 N3 표현을 보여 준다.

```
default:p5
  a default:GeoProject ;
  rdfs:label "1대1000 수치지형도 수정제작 수원지구"^^xsd:string ;
  default:builtIn default:수원 ;
  default:execYear "2010년"^^xsd:string ;
  default:organization "국립지리원"^^xsd:string ;
  default:produce default:d5 .
default:d5
  a default:DigitalMap ;
  rdfs:comment "수치지형도 제작"^^xsd:string ;
  rdfs:label "d5"^^xsd:string ;
  default:producedBy default:p5 ;
  default:scale "1/1000"^^xsd:string.
```

[그림 9] 비교대상 사업의 온톨로지 표현

5.5 의미기반검색의 수행

의미기반 검색은 제 4.2절에서 설명된 단계에 따라 수행된다. 그러나, 실제 구현에서는 각 단계들이 순차적으로 수행되기보다는 사용자로부터의 검색 요구를 각 단계의 내용이 모두 포함된 SPARQL 질의로 변환하여 원하는 결과를 가져오게 된다. [그림 10]은 현재의 시나리오에서 동일한 분류체계에 속하는 사업들을 가져오기 위해 사용하는 SPARQL 질의를 보여준다.

SPARQL 질의문에서 SELECT 문은 사업정보를 보여주기 위해 온톨로지로부터 가져와야 하는 항목들을 표현한다. WHERE 절에서 첫 6줄은 검색의 첫 단계인 구축지역에 따른 검색을 수행한다. 기본적으로 구축지역이 “수원”인 사업들이 대상이 되나, 제 4.2절에서 설명한 바와 같이 구축지역이 “수원”을 포함하고 있거나, “수원”의 하부지역인 경우에도 대상이 되어야 한다. 셋째 줄과 넷째 줄

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX base: <http://www.krihs.re.kr/SpatialProject.owl#>
SELECT ?pj ?pname ?year ?region ?products ?pdname
?pname ?gsd ?scale
WHERE {
  {{?pj base:builtIn ?region
    FILTER(?region = base:수원)}
  UNION {?pj base:builtIn ?region.
    base:수원 base:partOf ?region}
  UNION {?pj base:builtIn ?region.
    ?region base:partOf base:수원}}.
  ?pj rdfs:label ?pname.
  ?pj base:execYear ?year.
  ?pj base:produce ?products.
  ?products rdfs:comment ?pdname.
  ?products rdf:type ?sort.
  ?sort rdfs:label ?pname.
  OPTIONAL {?products base:GSD ?gsd}.
  OPTIONAL {?products base:scale ?scale}.
  {{?sort rdfs:label "정사영상"}
  UNION {?sort rdfs:label "3차원공간정보"}
  UNION {?sort rdfs:label "수치지형도"} }. }
ORDER BY DESC(?year) ?pj

```

[그림 10] 중복사업 검색 SPARQL 질의

은 수원의 상위지역을 나타내며, 다섯째 줄과 여섯째 줄은 하위지역을 나타낸다.

WHERE 절의 마지막 세 줄은 검색의 둘째 단계인 동일분류체계의 결과를 가져오기 위해 사용된다. 현재 시나리오에서는 정사영상, 3차원 공간 정보, 수치지형도가 해당 분류이므로 이 중 하나라도(UNION) 산출물이 속하는 경우 대상이 된다.

유사분류체계에 속하는 산출물을 갖고 있는 사업들을 찾기 위해서는 위 WHERE 절의 열째 줄과 열한 번째 줄 대신 아래의 조건을 사용한다. 즉, 유사분류체계(similarTo)로 연결된 분류에 속한 산출물(products)들에 대해서 검색을 수행한다.

```

?products rdf:type ?sort1.
?sort1 base:similarTo ?sort.
?sort1 rdfs:label ?pname.

```

이상과 같이 SPARQL을 이용하여 검색을 하면 <표 1>의 “1대 1000 수치지형도 수정제작 수원지

구”가 수치지형도에 해당하는 산출물을 갖고 있으므로 동일분류체계로 검색이 되며, “2009년 디지털 항공사진 촬영 및 판독용역”은 산출물인 항공 사진이 정사영상과 연관되어 있으므로 유사분류체계로 인해 검색이 된다.

검색의 마지막 단계는 속성(분류체계의 동일/연관 여부, 구축년도)에 따라 중복가능성이 있는 사업들을 순위화하는 단계이다. 동일분류체계와 유사분류체계는 다른 SPARQL 질의어를 사용하므로 자연스럽게 순위화가 되며, 시행년도는 위 SPARQL 질의어인 ORDER BY DESC(?year)를 이용하여 순위화가 이루어진다. 산출물의 속성은 SPARQL이 아닌 프로그램 내에서 처리하였다.

시나리오에서 검색의 최종결과는 <표 1>의 “1대 1000 수치지형도 수정제작 수원지구”가 직접적인 분류체계 연관성으로 인해 1순위, “2009년 디지털 항공사진 촬영 및 판독용역”이 연관 분류체계로 인해 2순위로 출력된다.

6. 의미기반검색 프로토타입 시스템의 구현 및 효과

6.1 시스템 구현

본 논문에서 제시한 공간정보사업의 중복사업 검색을 위한 의미기반검색 시스템의 프로토타입 개발을 위한 첫 단계는 온톨로지의 구축이다. 온톨로지는 프로티지(<http://protege.stanford.edu/>)를 이용하여 제 4장에서 제시한 분류체계와 부가정보 및 행정구역에 대한 스키마를 OWL 문서로 작성하였다. 제 5장에서 제시한 구조적인 형태로 기존의 공간정보사업들을 표현하기 위하여 먼저 기존의 데이터베이스에 있는 데이터들을 엑셀로 변환한 후에 엑셀로부터 OWL로 변환하였으며 이를 위하여 매핑마스터[18]를 사용하였다.

프로토타입 구현의 둘째 단계는 의미기반검색 시스템의 개발로, 온톨로지를 다루기 위해 Jena 라이브러리(<http://jena.apache.org/>)를 사용하고 웹 기



[그림 11] 프로토타입 의미기반검색 실행결과

반으로 시스템을 구축하기 위해 톱캣 7.0을 설치하였으며, Eclipse를 이용하여 JSP 기반의 웹 프로젝트로 개발하였다.

구현된 검색은 두 가지 방법을 지원한다. 첫째는 논문에서 제시된 바와 같이 미리 입력된 사업 정보를 가져와서 그 사업에 대한 중복사업을 검색하는 것이고, 둘째는 검색을 위한 속성을 입력함으로써 속성에 해당하는 중복사업을 검색하는 것이다.

[그림 11]은 대상 사업의 속성을 직접 입력하여 중복사업을 검색하는 화면이다. 시나리오의 “고정밀 멀티 시범구축 사업”에 대한 중복사업을 찾기 위해 구축지역과 산출물의 분류체계를 그림 상부와 같이 입력하면 그림 하부와 같은 결과를 볼 수 있다. 프로토타입에는 보다 많은 사업정보가 포함

되었기 때문에 제 5장의 시나리오에 비해 더욱 많은 중복사업이 추출되었다. 그림에서 중복사업 결과는 산출물이 동일분류체계로 직접 연결된 경우이며, 연관사업결과는 분류체계 간의 연관성에 의해 검색된 결과를 보여준다. 각 사업들에 대하여 화면 우측의 유사사업검색 버튼을 사용하면 검색된 결과에 대해 반복적으로 중복사업 검색을 수행할 수 있다.

6.2 의미기반검색 시스템의 효과

공간정보사업에 대한 중복사업의 검색 시스템에서 본 논문에서 제시한 의미기반검색을 활용하는 경우 다음과 같은 효과를 기대할 수 있다.

첫째, 기존의 시스템에서는 검색되지 않는 중복

사업들을 쉽게 검색할 수 있다. 과거에는 공간정보 사업에 대한 지식이 있더라도 시스템을 통해 중복가능성이 있는 사업들을 검색하는 것이 불가능했으나 제안된 시스템에서는 산출물에 대한 간단한 입력만으로 검색이 가능하다.

둘째, 전문적인 지식이 없더라도 정해진 속성을 입력함으로써 중복가능사업에 대한 검색이 가능하다. 현재까지는 공간정보사업에 대해 전반적인 지식이 없는 경우, 중복사업에 대한 검토 자체가 불가능했기 때문에 사업 담당자나 중간 검토자가 중복 여부를 판단하기 매우 어려웠다. 그러나, 제안된 시스템을 도입할 경우, 새로운 사업을 제안하는 담당자나 중간 검토자가 직접 중복가능성이 있는 사업들을 검색하고 내용을 검토함으로써 사업 제안단계에서 중복사업의 가능성을 낮추는 것이 가능하다.

셋째, 현재 공간사업에 관련된 정보들은 자치제자 중앙단체 등 다양한 기관에 분산되어 있다. 현재의 시스템은 시맨틱 웹을 기반으로 되어 있기 때문에, 향후에 각 기관들이 제안된 표준을 따르는 경우 시맨틱 웹에서 제공하는 의미적인 상호운용성으로 인해 각 기관에 분산된 사업정보들에 대해서도 검색이 가능해진다. 즉 지금까지 수작업에 의존하여 다양한 사업정보의 원천들을 검색하던 것을 현재 시스템보다 효율적 효과적으로 수행할 수 있을 것으로 기대된다.

7. 결 론

공간정보는 사회간접자본의 하나로서 미래의 핵심적인 성장산업으로 예측되고 있으며, 전자정부의 핵심기반이기도 하다. 공간정보의 구축에는 막대한 예산이 소요되기 때문에 중복구축을 막는 것이 공간정보정책의 중요한 목적 중 하나이다. 본 논문에서는 의미기반검색을 이용하여 공간정보사업의 중복사업을 검색할 수 있도록 함으로써 중복사업 검토 작업을 보다 효과적, 효율적으로 수행할 수 있도록 지원하고자 하였다. 이를 위하여 의

미기반검색과 온톨로지의 구축 및 활용방안을 제시하였으며 검색시나리오를 제시하여 의미기반검색이 어떻게 수행되는지를 보였다. 본 논문에서 제안된 시스템을 이용함으로써 기존의 시스템에서는 제공되지 않는 중복사업 검색이 가능하도록 하였으며, 이를 이용하여 전문가가 아니더라도 쉽게 중복성 검토를 함으로써 향후 공간정보사업의 중복사업 검색이 보다 효과적, 효율적으로 이루어질 수 있을 것으로 기대된다.

향후 연구로 먼저 현재 구축된 프로토타입의 온톨로지를 보다 정교하게 만드는 작업이 필요하다. 논문에서 제시된 분류체계와 세분류 간의 연관성 및 생태계 활용 외에도 다양한 속성들을 개발하여 추가함으로써 보다 다양한 기준에 따른 검색이 가능하도록 지원하고자 한다. 시스템과 관련하여 사업의 중복성을 수치화할 수 있는 모델을 만들고자 한다. 또한 연관성을 통해 간접적으로 연결된 경우 등에 대해 연관성을 수치화함으로써 연관성 파악이 구체적으로 제시될 수 있도록 만들고자 한다.

참 고 문 헌

- [1] 국토해양부, “2011년도 국가공간정보정책에 관한 연차보고서”, 2011.
- [2] 김성아, “OWL을 이용한 공간 온톨로지 구현 방법론의 기초적 연구”, 『대한건축학회논문집-계획계』, 제21권, 제6호(2005), pp.51-58.
- [3] 이상훈, 장병욱, “시맨틱 공유를 위한 MDA 기반 지하공간정보 온톨로지 모델 개발”, 『지형공간정보』, 제17권, 제1호(2009), pp.121-129.
- [4] Antoniou, G. and F. van Harmelen, “Web Ontology Language : OWL”, In : S. Staab and R. Studer(eds.), Handbook on Ontologies, (2004), pp.67-92.
- [5] Berners-Lee, T., J. Hendler, and O. Lassila, “The Semantic Web”, Scientific American.
- [6] Bizer, C., T. Heath, and T. Berners-Lee, “Linked Data-The Story So Far”, *Inter-*

- national Journal on Semantic Web and Information Systems*, Vol.5, No.3(2009).
- [7] Brickley, D. and R. V. Guha, "RDF Vocabulary Description Language 1.0 : RDF Schema", W3C Recommendation 10 February 2004. <http://www.w3.org/TR/rdf-schema/>.
- [8] Cho, J. H. and M. H. Kang, "Design of a Real Estate Knowledge Information system Based on Semantic Search", *Journal of the Korea Industrial Information System Society*, Vol.16, No.2(2011), pp.111-124.
- [9] Folger, P., "Geospatial Information and Geographic Information Systems (GIS) : An Overview for Congress", *CRS Report for Congress*, 2011.
- [10] Google, "Knoweldge Graph", 2012, <http://www.google.com/insidesearch/features/search/knowledge.html>.
- [11] Grimes, S., "Breakthrough Analysis : Two +Nine Types of Semantic Search", *InformationWeek*, 2010.
- [12] Guha, R., R. McCool, and E. Miller, "Semantic Search", *WWW2003*, 2003.
- [13] Harris, S. and A. Seaborne, "SPARQL 1.1 Query Language", W3C 2013, <http://www.w3.org/TR/rdf-sparql-query/>.
- [14] Horrocks, I., P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosf, and M. Dean, "SWRL : A Semantic Web Rule Language Combining OWL and RuleML", W3C Member Submission, <http://www.w3.org/Submission/SWRL/2004>.
- [15] ISO, "ISO/TC 211 Programme Maintenance Group(PMG)", 2012, <http://www.isotc211.org/pmg/programme.htm>.
- [16] John, T., "What is Semantic Search?", *Techulator*, 2012.
- [17] Klyne, G. and J. Carroll, "Resource Description Framework(RDF) : Concepts and Abstract Syntax", W3C Recommendation 10 February 2004, <http://www.w3.org/TR/rdf-concepts/>.
- [18] O'onnor, M. J., C. Halaschek-Wiener, and M. A. Musen, "Mapping Master : A Flexible Approach for Mapping Spreadsheets to OWL", 9th International Semantic Web Conference (ISWC), Shanghai, China, 2010.
- [19] OGC Network, "2010 OGCGeosemantics Summit", 2010, <http://www.ogcnetwork.net/geosemantics>.
- [20] Park, J., N. Kim, M. Choi, Z. Jin, and Y. Choi, "Semantic Search-A Survey", *Journal of Intelligence and Information Systems*, Vol. 17, No.4(2011), pp.19-36.
- [21] Protal, L., "Academic Geographic Semantic Search", National Geographic Information Institute, 2012, <http://www.land.go.kr/portal/main.do>.
- [22] Semantic Community, "Spatial Ontology Community of Practice", Vol.284, No.5(2001), pp.34-43.
- [23] W3C, "W3C Geospatial Ontologies", 2007, <http://www.w3.org/2005/Incubator/geo/XGR-geo-ont-20071023/>.
- [24] Yoo, D. H. and Y. M. Suh, "An Ontology-based Hotel Search System Using Semantic Web Technologies", *The Journal of Society for e-Business Studies*, Vol.13, No.4(2008), pp.71-92.

◆ 저 자 소 개 ◆

**박 상 언 (supark@kgu.ac.kr)**

현재 경기대학교 경상대학 경영정보학과 부교수로 재직중이며, 한국과학기술원 전산학과에서 학사, 한국과학기술원 경영공학과에서 석사 및 공학박사 학위를 취득하였다. 주요 관심분야는 웹기반 지능정보시스템, Semantic Web Mining, Semantic Web portal, 온톨로지 매핑, 지능형 웹서비스, 지능형 전자상거래 등이다. IEEE Transactions on Knowledge and Data Engineering, International Journal of Human Computer Studies, International Journal of Electronic Commerce, Journal of Intelligent Information Systems 등의 국제 저널 외에 국내외 학회 및 학술지에 논문을 발표하였다. 수상경력으로는 한국 지능정보시스템학회 우수논문상 수상 (2005년, 2006년, 2007년)과 지식정보산업 연합학회 우수논문상(2008년)이 있다.

**임 재 익 (limjay@ajou.ac.kr)**

현재 아주대학교 경영대학 e비즈니스학과 부교수, 아주대학교 국제대학원 원장으로 재직 중이며, 서울대학교 경제학과에서 학사, 미국 University of Iowa에서 MIS 전공으로 경영학 박사학위를 취득하였다. 주요 관심분야는 정보통신, CRM, MIS 이며, 관련 분야에서 해외 저명 학술지인 Operations Research를 포함 국내외 저명 학술지에 다수의 논문을 발표 및 게재하였다.

**강 주 영 (jykang@ajou.ac.kr)**

현재 아주대학교 경영대학 e비즈니스학과 부교수로 재직중이며, 포항공과대학교 컴퓨터공학과에서 학사, 서울대학교 컴퓨터공학과에서 석사, 한국과학기술원 경영공학전공에서 공학박사학위를 취득하였다. 주요 관심분야는 시맨틱 웹, 지능형 전자상거래, 클라우드 컴퓨팅, 빅데이터 등이다. 관련 분야에서 몇편의 저서를 기술하고, 국내외 학회 및 해외 저명 학술지 등에 40여건 이상의 논문을 발표 및 게재하였다. 수상경력으로는 한국지능정보시스템학회 우수논문상 수상, 한국경영정보학회 최우수 사례상 수상 등이 있다.