

# 온톨로지 구축 방법과 사례<sup>†</sup>

울산대학교 최호섭 · 임지희 · 배영준 · 최수일 · 옥철영

## 1. 서 론

방대한 지식 정보를 체계화하기 위해서는 제학문적 기술이 요구된다. 즉 각 분야별 지식 정보에 대한 명확한 학문적·이론적 지식을 비롯하여 기술적 지식, 응용적 지식 등 연구 개발에 활용하고자 하는 지식 정보에 따라 상당한 지식이 요구된다. 이러한 요구 사항은 지식 정보 관리 체계 확립에서는 반드시 필요한 이론적·실제적 기반이며, 실질적인 응용 시스템 개발에서 일반 사용자, 전문 연구자, 관리자 측면에서도 효율성을 기대할 수 있는 상당한 중요성을 가지는 요소이다. 그리하여 분야별 전문 지식 정보를 다루는 응용 시스템에서는 분야별 전문 지식을 비롯하여, 응용 시스템을 통한 서비스 유통 과정 등을 충분히 고려하지 않으면 안 된다.

이러한 지식 정보와 관련된 최고의 핵심 기술로 부각되고 있는 것이 바로 시맨틱 웹으로서, 시맨틱 웹의 등장은 몇몇 분야에서 이론적·실험적 수준으로 연구되어 오던 시소러스, 온톨로지, 의미망 등과 같은 지식베이스, 지식 정보 처리, 웹 지식 정보의 공유 기술 등에 대한 연구 개발 관심을 증대시켰다고 할 수 있다.

시맨틱 웹뿐만 아니라 제학문적으로 관심이 되고 있는 온톨로지는 광의적으로는 '실세계에 존재하는 모든 개념(concept)과 그 개념들의 속성(attribute)은 무엇이며, 그리고 개념들이 상호간 의미적으로 어떻게 연결되어 있으며(relation), 개념들의 실질적인 사례는 무엇인가(instance)'에 대한 정보를 가지고 있는 의미적인 개념 집합체로 정의할 수 있다. 나아가 온톨로지는 협의적으로는 특정 영역에서의 정보교환용으로 합의된 어휘를 만들기 위해 사물, 사건 및 관계 등과 같은 개념들의 집합으로 정의할 수 있다. 이러한 온톨로지에 대한 이전 연구들은 단순히 개념이나 어휘들의 집합 또는 그들 간의

연결에 초점이 맞춰진 것이라면, 최근에는 전산학적으로 온톨로지 구축을 위한 표준화 작업, 온톨로지 구축 언어 개발, 분야별 온톨로지의 대규모 구축 방안, 재사용과 상호운용성을 고려한 온톨로지 구축 등과 같은 실질적이고 실용적인 온톨로지를 구축하기 위한 연구가 활발하다.

그러나 온톨로지 관련 연구에서 국내외적으로 공통적으로 지적될 수 있는 문제는 온톨로지에 대한 개념적 정리와 실질적인 구축 방법이다. 이러한 문제는 온톨로지에 대한 다양한 학문적 의견 차이에서 비롯되었다고 할 수 있다. 즉 철학적 온톨로지, 언어학적 온톨로지, 문헌정보학적 온톨로지, 공학적 온톨로지 등 다양한 연구 분야에서 언급되고 있는 온톨로지에 대한 인식 차이를 비롯하여, 시소러스·어휘망·어휘분류·의미망 등과 같이 온톨로지와 비슷한 구조를 가진 기존의 연구 개발 결과들과의 형태적·의미적 유사성 때문에 발생한 문제라고 할 수 있다.

본고에서는 최근 제학문적으로 연구 중인 온톨로지에 대한 다양한 시각적 차이를 알아보고, 본 연구팀이 연구 개발한 몇 가지 온톨로지 사례를 통해 온톨로지를 구축하는 원리를 간략하게 소개한 뒤, 이러한 온톨로지 구축 사례에서 발견되는 공통점과 차이점은 어떤 것이 있는지를 살펴보도록 한다.

## 2. 온톨로지에 대한 인식과 구축 방법

### 2.1 온톨로지에 대한 인식

언어처리적 입장에서 시소러스, 어휘망, 온톨로지, 의미망 등은 하나의 언어적 지식을 총망라한 지식베이스(knowledge base)로 인식하는 경우가 많다. 즉 지식베이스는 언어로 표현된 지식을 체계적이고 의미적으로 관리할 수 있는 의미적 지식 보관소 역할을 담당한다는 것을 의미한다. 이러한 인식은 문헌정보학에서의 시소러스(thesaurus)나 어휘통제집(controlled vocabulary), 언어학에서의 의미장(semantic field or

<sup>†</sup> 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 육성지원사업의 연구결과로 수행되었으며, 일부 구축 사례는 2004, 2005년 한국과학기술정보연구원의 위탁과제로 수행되었음.

lexical field)이나 의미 부류(semantic class), 인공지능에서의 의미망(semantic network), 특정기관별 표준분류체계 (standard classification), 시맨틱 웹에서의 온톨로지 등에서도 유사하다고 할 수 있다.

특히 웹이라는 방대한 지식 정보 공간과 다양한 영역의 지식 정보를 대상으로 하는 시맨틱 웹에서의 온톨로지 관련 기술에 대한 집중적인 연구 개발은 온톨로지와 유사한 구조를 가진 관련 연구 결과물의 연구까지 확대되었다고 할 수 있다. 이는 온톨로지의 내부 구조 속에 포함되는 개념성, 관계성(특히 계층성), 속성 등의 몇몇 특성이 시소러스나 어휘망 등과 일치하는 부분들이 많았기 때문이다. 그리하여 새로운 온톨로지를 구축하는 연구 이상으로, 기존의 WordNet[1,2,3,4], Cyc[5,6,7], UMLS[8,9], MeSH[10,11] 등과 같은 연구 결과물을 기반으로 상위 수준의 온톨로지(upper level ontology)나 특정 영역의 온톨로지(domain ontology) 등으로 활용하는 연구가 국내외적으로 진행되고 있는 실정이다.

이렇듯 기존의 시소러스, 의미망 등이 온톨로지로 활용되는 경우는 '개념'과 '관계'라는 두 가지 특성을 통해 그 연관성이 분석될 수 있다[12]. 먼저 '개념'은 여러 관념 속에서 공통된 요소를 추출하고 분석하여 얻은 하나의 공통 관념이라 정의할 수 있다. 그리고 이러한 '개념'은 일반적으로 언어로 표현되는데, 이를 '개념어'라고도 지칭한다. 바로 개념이 언어(엄밀히는 문장, 구절, 어휘)로 표현되기 때문에, 언어로 표현된 시소러스나 의미망 등의 형태를 갖춘 것을 온톨로지라 판단할 수 있는 것이다. 다음으로 '관계'는 어휘의미론(lexical semantics)에서의 의미관계(semantic relation)와 개념관계(conceptual relation)로 기술될 수 있는데, 시소러스나 WordNet에서 일반적으로 사용되고 있는 상하·동의·유의·부분전체 관계 등의 의미관계 유형과 온톨로지에서도 사용되는 넓은 의미의 의미관계인 구성원 관계(has\_member), 위치관계(has\_location) 등의 개념관계 유형이 있다. 이러한 '관계'는 계층적 구조를 비롯하여 의미적 네트워크 구조를 형성하는 것의 핵심으로써 시소러스, 어휘망, 어휘 분류 등을 온톨로지와 비슷하게 판단할 수 있는 요소 중 하나이다[12,13,4,15].

앞서 언급한 온톨로지의 내부 구조 및 특성, 시소러스, 어휘망 등과 같은 유사 구조 체계, 그리고 국내외의 온톨로지 관련 연구 내용을 종합해 볼 때 그림 1과 같은 온톨로지에 대한 인식 범위를 설정할 수 있다. 즉 온톨로지는 어휘, 용어, 어휘목록, 사전, 전문분야사전 등과 같은 어휘 집합을 기반으로 하며, 통제어휘집, 분류,

시소러스, 어휘망 등을 포함하는 개념, 관계, 속성 등이 내부적으로 형성된 상의의 지식 구조 체계라 할 수 있는 것이다.

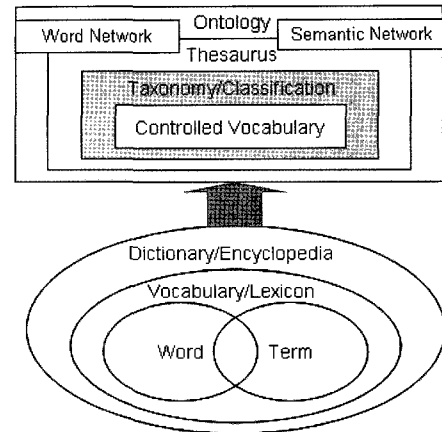


그림 1 온톨로지에 대한 인식 범위

## 2.2 온톨로지 구축 방법

“온톨로지를 개발하는 것은 쉬운 일이 아니다. 일반적인 데이터 모델링과 같이, 온톨로지 모델링도 풍부한 경험과 폭넓은 지식을 요구한다. 어떤 사람은 온톨로지 모델링의 어려움을 표현하기 위해 이것이 기술(technology)이라기보다는 예술(art)적 차원이라고 한다.” [16]라고 기술한 바와 같이 온톨로지 구축은 상당한 지식과 노력이 요구되는 연구 개발 작업이다.

몇몇 특정한 도구를 사용한 온톨로지 구축 방법 [17,18]이나 기존의 온톨로지 구축 방법론을 정리한 연구 [16,19] 등이 있으나, 엄밀하게는 국내외적으로 온톨로지를 실질적으로 구축하는 이론적 체계와 원리는 아직까지 마련되지 않았다. 이는 최근의 온톨로지 관련 연구들이, 온톨로지의 실질적인 구축보다도 기존의 시소러스, 의미망, 분류 체계 등을 이용한 온톨로지 구축이나 기구축된 온톨로지를 이용한 애플리케이션 개발이 대부분을 차지하는 것을 통해서도 알 수 있다. 이를 해결하기 위해서는 WordNet, UMLS, Cyc 등과 같이 오랜 시간 동안 축적된 연구 결과를 종합적으로 검토할 필요가 있으며, 온톨로지와 관련된 이론 습득을 비롯하여 다양한 기반 기술 및 요소 기술을 연구해 가야 할 것이다[12].

그림 2는 특정 영역(domain)의 온톨로지를 구축할 때의 일반적인 단계를 나타낸 것이다. 본고에서는 본 연구팀에서 개발한 포괄적인 의미에서의 온톨로지라 판단할 수 있는 지능형 한국어 어휘망, 면역학 온톨로지, 국가과학기술 R&D 기반정보 온톨로지를 중심으로 온톨로지 구축 방법과 사례를 간략하게 살펴보기로 한다.

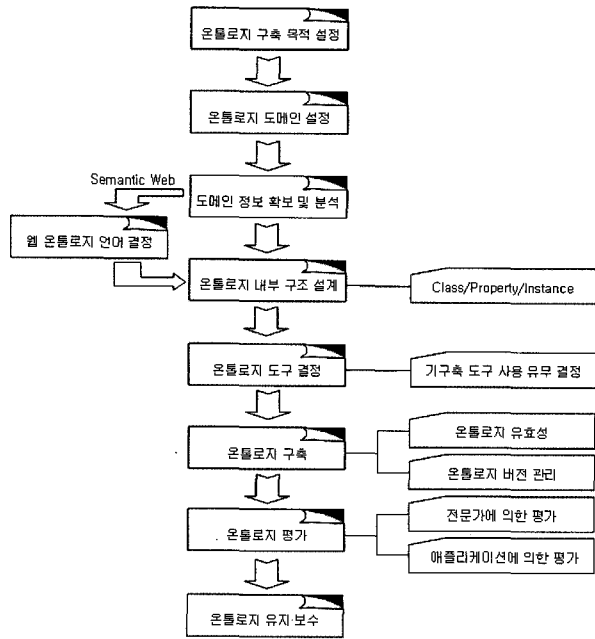


그림 2 온톨로지의 일반적인 구축 단계

### 3. 지능형 한국어 어휘망

자연 언어의 어휘적 의미, 구문적 의미, 담화적 의미를 바탕으로 행위나 현상, 상태 등에 담긴 의미론적·개념론적 특성을 포함하고 있는 의미적 언어 자원 구축에 대한 연구는 다양하게 이루어지고 있는 실정이다. 국외에서는 WordNet, EuroWordNet, Cyc, HowNet, Lexical FreeNet, EDR 등이 대표적이며, 국내에서는 카이스트의 CoreNet, ETRI의 어휘개념망, 부산대의 KorLex 등이 대표적이라 할 수 있다.

본 연구팀이 2002년부터 개발 중인 한국어 사용자 어휘 지능망(User Word Intelligent Network, 이하 UWIN)은 한국어의 공통적이고 개별적인 속성을 바탕으로 한국인의 보편적인 인지 체계와 개념 관계를 파악하여 이를 어휘의 의미적·개념적 네트워크로 형성한 온톨로지적 어휘망이라 할 수 있다.

UWIN은 한국어정보처리를 비롯하여 정보검색, 기계번역, 시맨틱 웹 등 다양한 분야에 이용될 수 있는 대규모 어휘 지식 베이스를 목표로 하고 있다. 현재 온톨로지 기반 의미적 주석(ontology-based semantic annotation)과 유사한 단어 중의성 해소(word sense disambiguation)와 의미 태깅(semantic tagging) 기술에 활용되고 있으며, 이 외에도 복합명사 자동 생성, 전문분야별 개념체계 자동 생성, 정보검색에서의 질의 확장, 어휘 학습 시스템 등 다양한 기술에서 활용되고 있다. 나아가 UWIN 영어 버전을 구축 중에 있어 조만간 한국어를 중심으로 한 한영 대역 UWIN이 개발

될 계획이며 WordNet과의 사상 구조(mapping structure)도 기대할 수 있게 되었다.

UWIN은 현재 25만 여 어휘가 구축된 상태이다. UWIN의 구축 대상은 한국어 어휘 전체(모든 품사 및 언어 단위)로서, 핵심적 대상은 명사, 동사, 형용사이며, 부수적 대상은 부사, 관형사, 대명사, 감탄사, 조사, 수사, 의존명사 등이며, 북한어, 방언, 옛말, 전문용어, 고유명사, 어근, 어미 등 한국어 어휘 전체를 대상으로 연구 개발 중이다<sup>1)</sup>.

#### 3.1 UWIN의 내부 구성 체계

UWIN의 내부 구성 체계는 의미론적 측면의 의미관계/개념관계/의미자질, 통사론적 측면의 격관계/구문패턴/하위범주화, 형태론적 측면의 품사 정보/형태 정보, 언어처리적 측면의 기계가독사전/시소러스/의미망 등 어휘가 가지고 있는 최대한의 정보를 체계화시키고자 하였다.

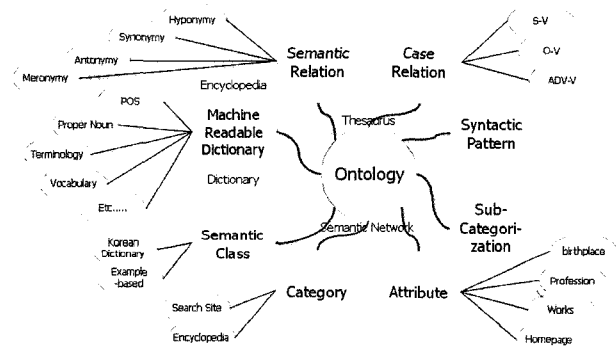


그림 3 UWIN의 대략적인 내부 구조

또한 UWIN의 확장성 및 활용성을 고려하여, 어휘 정보 통합 관리 체계를 형성할 수 있도록 설계하여 단순 어휘 네트워크, 시소러스, 온톨로지, 표준분류체계, 사전/백과사전/전문용어사전, 말뭉치 등이 유기적으로 연결된 구조 체계를 가지고도록 설계되었다.

#### 3.2 UWIN의 구축 원리

##### 3.2.1 기초 자원 및 기본 구축 지침

UWIN의 기초 자원은 국어사전, 백과사전, 전문용어사전, 기타 자원 등을 약 100만 여 어휘를 확보하여 실질적인 구축 대상을 쉽게 추출하고 어휘적 정보를 효율적으로 획득할 수 있도록 하였다.

기본 구축 지침은 기초 어휘 선정, 어휘 표기 방법, 동형이의어/다의어 처리 방법 등과 같이 일반적으로 많은 연구자들이 어휘망을 구축할 때 고려하지 않는 사항들을 지침화하였다. 그리하여 먼저 기초 어휘는 상위

1) 본고에서는 지면 관계상 중요한 몇 가지 사항만 기술한다.

층위 설정에 밑바탕이 될 수 있는, 언어생활에서 빈도수가 높고 분포가 넓으며 합성어나 파생어와 같은 이차조어의 근간이 되는 최소한의 필수 어휘를 선정하였다. 다음으로 어휘 표기 방법은 한글맞춤법 및 외래어표기에 의거한 단어 표기, 사용자의 표기 오류 처리 방법, 순화어 이용, 한자어/외래어 등에 대한 원어 정보 표기, 복합어의 형태 구성 정보 등을 설정하여 지침을 마련하였다. 동형이의어 처리는 국어사전의 어휘 어깨번호를 이용하고 이를 의미태그로 활용하도록 함과 동시에, 여러 사전의 어깨번호가 다를 경우에는 사상 구조(mapping structure)를 형성하고 대표 사전의 동형이의어 표기 방식을 핵심 표기 방식으로 설정하였다. 마지막으로 한 어휘가 가지는 다의적 특성으로 인한 의미적 부담감 해소 및 의미적 계층 구조의 세밀성, 의미관계 설정의 용이성 등을 고려하여 어휘와 뜻풀이의 관계를 1:N의 관계가 아닌 1:1의 관계로 설정하여 다의어를 관리하도록 하였다.

### 3.2.2 중심 구축 지침

WordNet을 비롯한 국내외 어휘망 구조를 가진 언어 자원들은 어휘들 간의 계층적 구조 형성 원리와 의미/개념 관계에 대한 설정이 어휘망 구축의 가장 중요한 이론적·실제적 기반이 된다고 할 수 있다. 그리하여 기존에 구축된 국내외 어휘망, 시소러스, 온톨로지 등을 분석해 보면, 분류적 구조와 계층적 구조가 혼용된 체계를 포함하거나, 어휘의 실생활 사용 패턴을 중심으로 한 의미관계를 사용하거나, 개별 언어의 특수성을 반영하지 않은 구조 체계가 많이 나타난다.

그리하여 UWIN에서는 최대한 계층적 구조를 기준으로 층위를 형성하도록 하였다. 여기서의 계층적 구조

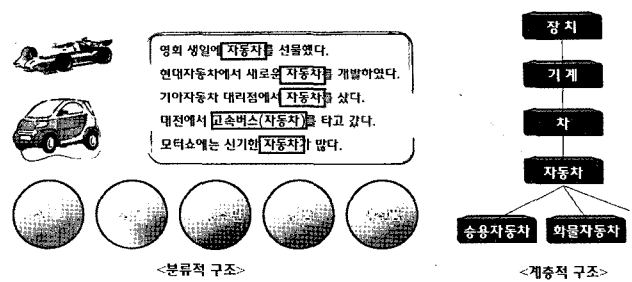


그림 4 분류적 구조와 계층적 구조

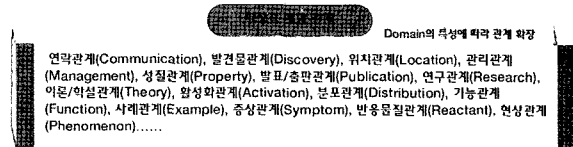
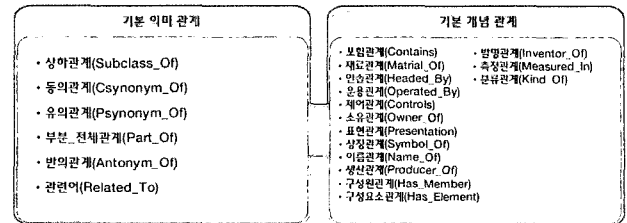


그림 5 UWIN의 의미관계와 개념관계

란 의미론적 상하관계를 중심으로 층위(level)를 이룬 집합체로서, 상하위 층위(또는 노드)가 의미적으로 밀접한 연관성을 가짐과 동시에, 엄밀한 의미에서의 상하관계가 형성되고 보편적인 개념화(또는 추상화) 과정으로 인식되는 구조를 말한다. 이러한 개념적 계층 체계는 WordNet과도 유사하다.

다음으로 UWIN에서의 어휘들간의 의미관계와 개념관계 설정은 그림 5와 같이 설정한다. UWIN의 의미관계는 어휘의미론에서의 다루고 있는 보편적인 상하관

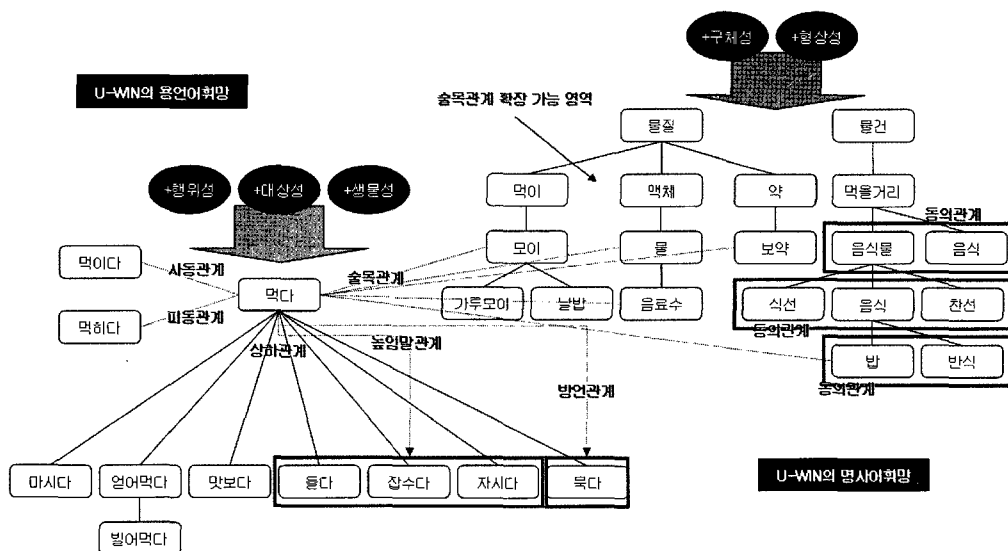


그림 6 UWIN의 구축 사례 일부

계, 동의관계, 유의관계, 부분전체관계 등을 사용하는 데, 문장의 표면 구조(surface structure)를 중심으로 하여 어휘의 계열(paradigmatic)/통합(syntagmatic) 관계를 이용하여 어휘의 의미관계를 설정한다. 또한 개념관계는 의미관계의 세부적이지 확장적인 의미적 속성을 이용한 관계를 말한다. 이는 기본 개념관계와 확장적 개념관계로 분리하여 일반적인 개념관계와 특정 영역 중심적인 개념관계로 설정할 수 있다.

마지막으로 각종 사전과 말뭉치에서 추출한 의미 속성(또는 의미 자질)에 대한 상속 체계는 어휘망 구축의 필수 정보이다. UWIN은 어휘의 속성 상속 체계와 더불어 속성 제약(constraint) 정보를 어휘망 구축에 포함한다. 덧붙여 UWIN은 의미 속성 정보의 활용과 동시에 기존의 분류 체계와의 연결 구조를 통한 어휘 정보 통합 관리 체계의 기본틀을 제공하는데, 기구축된 분류 체계나 시소러스 등과의 유기적 연결 구조를 통해 어휘망의 개념적 확장성을 확보할 수 있다.

### 3.3 UWIN의 구축 사례

그림 6은 UWIN의 구축 사례 일부를 보이고 있다.

## 4. 번역학 온톨로지

의학(특히 의료정보학), 생물학 분야와 관련된 온톨로지는 다른 분야의 온톨로지보다 상당히 많은 연구 개발이 진행되고 있으며, UMLS, MeSH, GALEN[27], SNOMED[28], Gene Ontology[29] 등이 대표적이라 할 수 있다.

본 연구팀에서 개발한 번역학 온톨로지(Immunology Ontology)는 번역학 관련 전문가들이 사용하는 지식관리시스템(KMS)에서의 시소러스 및 온톨로지 검색이 가능하도록 함과 동시에, 한영 대역 온톨로지를 개발하여 지식관리시스템에서의 전문가의 활용성을 극대화시키기 위한 온톨로지이다.

### 4.1 시소러스 구축 원리의 활용

비전문가에 의해 특정 분야의 온톨로지를 구축하는 것은 쉬운 일이 아니다. 특히 의학이나 생물학 분야의 특정한 온톨로지를 구축하는 것은 전문가 정확한 검증과 평가 없이는 힘든 일이다.

이러한 점을 고려하여 번역학 관련 사전과 관련된 학술 문서 등을 UWIN 구축 원리와 유사하게 분석하고 정제하여 번역학 시소러스를 우선적으로 반자동으로 구축한 다음, 온톨로지로 확장 적용하고자 하였다. 그리하여 시소러스에서 사용되는 계층관계, 연관관계, 대등관계 등을 각각 UWIN의 상하관계, 관련어, 동의/유의

관계로 대치하여 개념화 구조를 취하는 온톨로지 계층 구조가 형성되도록 하였다.

## 4.2 번역학 시소러스 구축 방법

### 4.2.1 번역학 전문용어 데이터베이스

번역학 온톨로지를 구축하기 위해서 선행되어야 하는 것이 바로 번역학 전문용어의 확보이며, 나아가 번역학과 직접적으로 관련되어 있는 인접 학문에 대한 전문용어까지 고려해야 한다. 그리하여 번역학 전문용어 데이터베이스는 번역학 사전을 비롯하여 생화학·미생물학·의학 관련 사전, 그리고 대규모 국어사전이나 백과사전 등을 적절히 활용하여 구축하였다. 이는 엄밀한 의미에서 특정 분야에서만 사용되는 전문용어라고 판단하기 어려운 부분이 발생할 수 있으므로, 번역학 사전 및 번역학 관련 문서 등을 분석하여 인접 학문의 전문용어까지 확보하기 위한 확장적 특정 분야 전문용어 데이터베이스 구축 방법이라 할 수 있다. 그리하여 번역학과 관련된 어휘를 32,000여 개를 구축하여, 식별자/한글표제어/뜻풀이/의미태그/영어표제어/한자정보/약어정보/관련어/상용어/전문용어태그/출처정보 등으로 구분하여 번역학 전문용어 데이터베이스를 확보하였다.

### 4.2.2 상하관계 설정

번역학 시소러스 및 온톨로지 구축에서 상하관계 설정은 비전문가로서는 가장 어려운 작업 중 하나이지만 반드시 필요한 핵심적인 작업이다. 그리하여 번역학 시소러스의 상하관계 설정은 분류적 구조가 아닌 UWIN의 엄밀한 의미에서의 상하관계를 이용한 계층적 구조 원리를 이용하였다. 이는 아래의 예를 통해 쉽게 이해할 수 있다.

#### ▶ 엄격한 상하관계

X is Y {wife, doctor, scientist, servant...} →  
X is Y' {person}

#### ▶ 유사(문화적) 상하관계

X is Y {knife; clock; snake; dog...} → X is Y' {weapon; present; poisonous animal; pet...}

이러한 번역학 시소러스의 계층적 구조의 원칙을 MeSH 기반 한영 의학용어 시소러스와의 비교, 그리고 Ogden&Richards의 의미삼각도를 이용하여 그림 7과 같이 설명할 수 있다. MeSH 기반의 한영 의학 용어 시소러스는 '세포'의 포함 여부에 따라 상위어 '인체구조'에 포함시킴과 동시에, '세포막'과 '세포질' 또한 '세포'와의 관련성 때문에 '세포'의 하위어로 설정하고 있다. 이와는 달리 번역학 시소러스에서는 용어 자체가 가지는 의미와 지시물(referent), 사고(thought)에 대한 종합

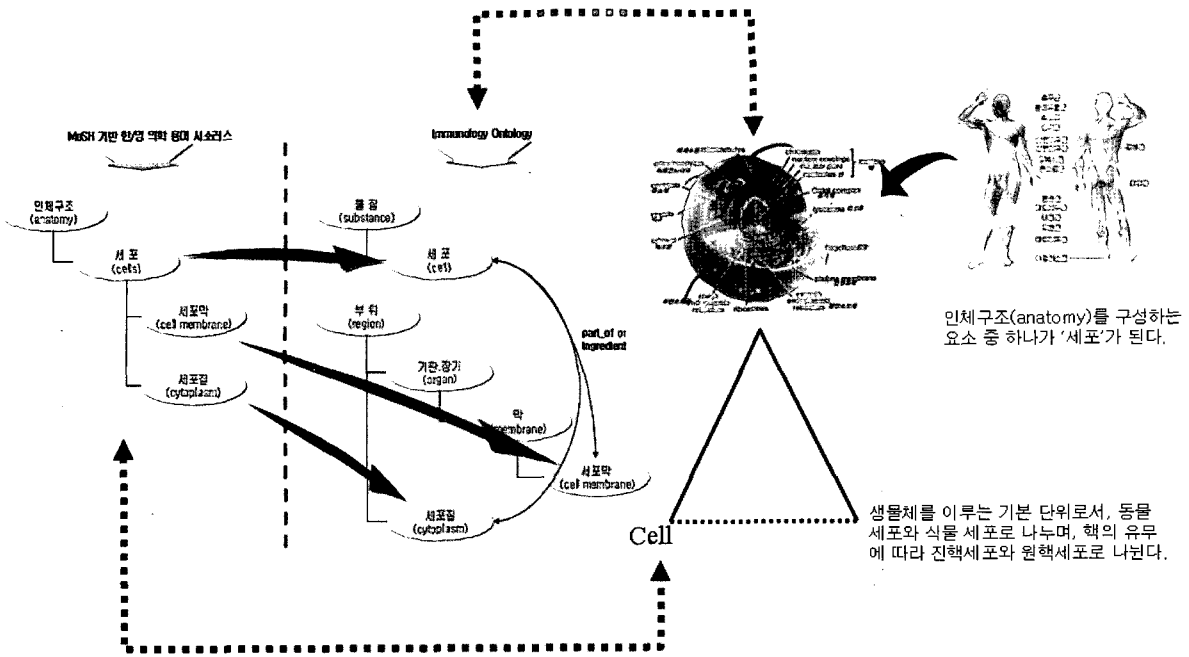


그림 7 의미상각도를 이용한 MeSH와의 비교

적인 고찰을 통해 상하관계를 설정하여, '세포'의 경우 '물질'의 하위어로 설정하고, '세포막'과 '세포질'은 '부위'의 하위어에 위치하도록 구축한다. 그리고 온톨로지 측면에서 '세포'의 부분전체관계 및 구성성분관계로서 '세포막'과 '세포질'을 설정함으로써 다양한 개념적 관계가 성립되도록 구축한다.

이러한 상하관계의 기본 원칙을 기반으로, 번역학 전문용어에 대한 상하관계를 반자동으로 구축한다. 번역학 시소러스에서 사용된 세 가지 상하관계 반자동 구축 원리는 다음과 같다.

첫째로 표제어 정보를 이용하는 것이다. 한글, 영어 표제어의 형태적 정보를 충분히 활용하는 것으로, 먼저 한글 및 영어 표제어의 특정 단어에 대한 대역 상태를 살펴 다음, 한자어일 경우에는 한자 정보의 동일성 여부를 판별하여 상하관계를 설정한다. 이는 복합명사류일 경우에 적절히 이용될 수 있는 방법이다. 또한 복합명사의 경우에는 주로 오른쪽 성분이 핵심적인 의미 정보를 가지고 있는 경우가 많기 때문에 대부분의 복합명사는 오른쪽 성분을 이용하여 상하관계를 설정할 수 있다. 그리고 단일어일 경우에는 한자 정보와 영어 표제어의 파생 여부를 판별하여 그 원형의 의미를 밝혀 상하관계를 연결한다. 그러나 단일어일 경우에는 뜻풀이나 관련 정보를 참조하지 않으면 안 된다. 이외에도 한자어의 구성 상태, 한자 접미사의 사용 여부, 영어의 파생 접사 사용 여부, 표제어의 동의어 형태 정보 등을 고려해야 한다.

둘째로 뜻풀이 정보를 이용하는 것이다. 용어의 뜻풀

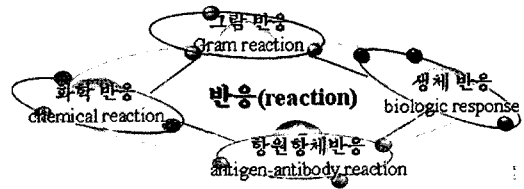


그림 8 표제어 정보를 이용한 상하관계 설정

이가 상위어 정보를 갖는 경우, 즉 유개념어를 가지는 경우는 이를 상위어로 설정할 수 있다. 또한 표제어의 형태적 정보와 다른 뜻풀이가 명시되어 있을 경우에는 뜻풀이를 우선시 할 수 있다. 다만 그 쓰임에 대한 명확한 조사가 뒤따라야 한다. 예를 들어 '전사인자(transcription factor)'는 "유전자를 활성화시키거나 억제하여 형질 발현을 조절하는 단백질"로서, '전사인자'의 상위어로 '인자(factor)'로 설정할 수 있으나, '전사인자' 자체가 하나의 단백질을 설명하는 용어로 굳어진 상태이므로 이 용어의 상위어는 뜻풀이 정보에 의해 '단백질 (protein)'로 설정할 수 있다.

마지막으로 기타 정보를 이용하는 것이다. 같은 종류의 표제어 형태 정보를 포함하고 있으나 뜻풀이가 다른 경우에는 유개념어를 통해 상하관계를 확인한 후 동일

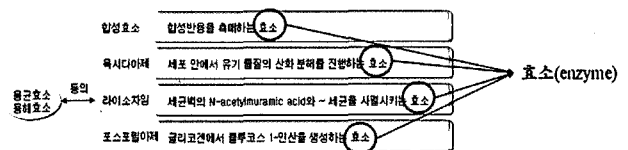


그림 9 '효소(enzyme)'에 대한 상하관계 설정

한 상하관계로 연결시키거나, 참조사전을 통해 보편적인 개념을 찾아 상하관계를 연결한다. 그리고 뜻풀이의 유개념어가 명확하지 못한 경우나 두 가지 이상의 유개념어가 등장하는 경우에는 작업자의 주관적 판단으로 상하관계를 연결시킬 수 있으나, 반드시 시소러스 구축 작업시 문서화해야 한다. 나아가 번역학에서 일반적으로 사용하는 분류 체계가 정해진 용어들의 경우에는 전체적인 시소러스의 계층 구조의 틀이 변하지 않는 한 번역학의 용어 분류 방식을 그대로 채택하도록 한다.

#### 4.2.3 동의관계와 관련어 설정

동의관계와 관련어는 번역학 사전을 비롯하여 각종 사전이 제공하는 동의어 집합과 관련어 집합을 그대로 이용하고자 하였다. 동의관계의 경우, 동의어를 설정하는 명확한 방법이 제시되지 못하고 전문가에 의한 판단과 사용 실태를 분석해야 하는 어려움이 따르므로 전문 용어사전이나 기타 사전에서 제공하는 동의어 정보를 활용하는 것이 현재까지는 최선의 방법이라 할 수 있다. 그러나 전문 분야의 문서나 사전의 뜻풀이 정보를 상호 비교하여 유사한 의미의 용어를 추출하는 연구는 진행될 수 있을 것이다.

### 4.3 번역학 온톨로지 구축 방법

번역학 온톨로지는 번역학 시소러스에서 사용되었던 번역학 전문용어 데이터베이스를 이용함과 동시에 계층적 구조를 그대로 이용하였다. 다음으로 번역학 온톨로지에 표현되는 개념에 대한 정의는 전문용어 데이터베이스의 뜻풀이와 관련 정보를 통하여 획득하였다. 그리고 개념의 조직화를 위해, 상위 개념을 중심으로 개념이 가지는 개별적인 속성과 공통적인 속성을 분석하고, 개념들 간의 관계를 시소러스의 기본 의미 관계와 UWIN에서의 기본 개념 관계, 도메인 특성 개념 관계 등 세 가지 관계로 나누어 온톨로지를 표현하고자 하였다.

#### 4.3.1 개념 표현

번역학 온톨로지의 구축 대상 개념은 번역학 전문용어를 비롯한 관련 분야 용어, 일반 용어가 된다. 개념들 간의 다양한 관계가 성립되기 때문에, 번역학 시소러스에서 표현된 용어보다 좀더 많은 용어 집합이 필요하다. 즉 번역학 시소러스보다 번역학 온톨로지가 좀더 많은 개념 집합을 바탕으로 구축된다는 것을 알 수 있다.

#### 4.3.2 관계 표현

번역학 온톨로지에서의 관계는 기본 의미관계, 기본 개념 관계, 특정 분야 개념 관계로 구분하였다. 기본 의미 관계와 기본 개념 관계는 시소러스 및 UWIN의 관계 체계를 그대로 반영하고, 특정 분야 개념 관계는 특정 분야의 특수성과 표현성에 따라 생성될 수 있는 개

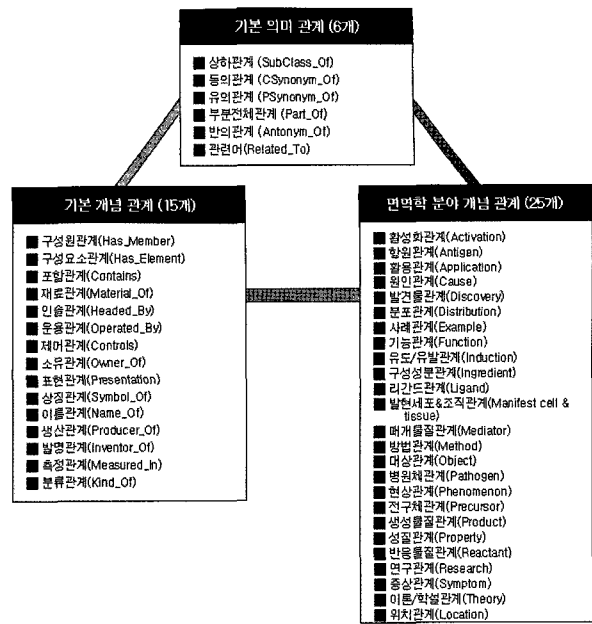


그림 10 번역학 온톨로지에서 사용되는 관계

념 관계이다. 그리하여 번역학 온톨로지에서도 사용되는 관계들을 정리하면 그림 10과 같다.

번역학 온톨로지에서도 사용되는 번역학 분야 개념 관계는 번역학 사전과 관련 사전의 뜻풀이, 그리고 문서 표현, 표, 그림 등을 이용하여 추출하였다. 예를 들면 '질병/질환(disease)'과 관련된 개념 관계를 추출할 경우에는, 뜻풀이나 문서, 표에서 나타나는 증상, 발병대상, 발병부위 등을 고려하여 추출 단서를 통해 번역학 분야 개념 관계를 설정할 수 있다. 그림 11은 질병과 관련된 의미 및 개념 관계를 추출하는 방법을 간략하게 나타낸 것이다.

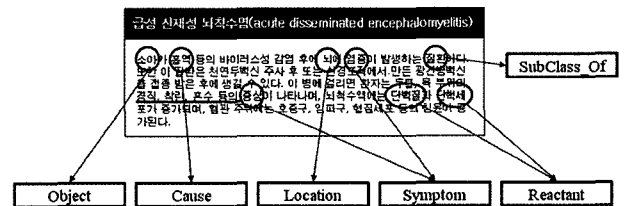


그림 11 질병(급성 산재성 뇌척수염)에서의 의미 및 개념 관계 추출

#### 4.3.3 속성 체계

번역학 온톨로지의 개념의 속성(attribute)은 한 개념의 특징적인 성질들을 나타내는 것으로, 개념들은 속성 체계를 상위 계층으로부터 상속을 받거나, 유사 속성으로 변형되어 속성의 상속성을 유지한다. 번역학 온톨로지에서도 이러한 속성의 상속 체계를 이용하여 계층 구조 형성이나 관계 설정에 적절하게 이용하였다.

특히 번역학 온톨로지에서는 특정한 물질들이 가지는

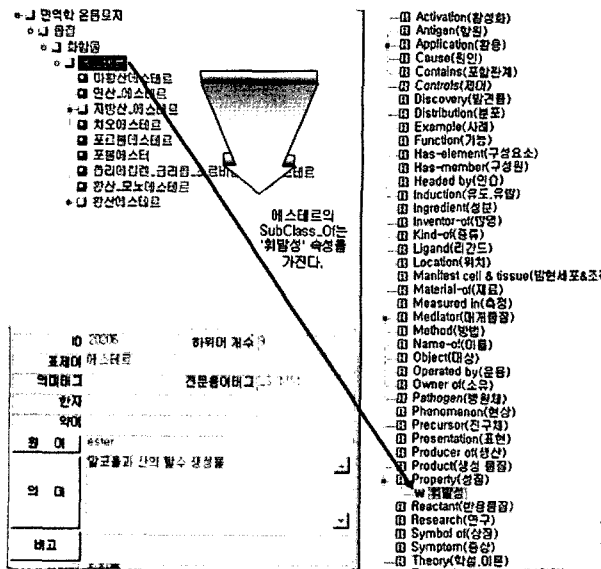


그림 12 에스테르(ester)의 속성 상속 체계

성질이 중요한 역할을 담당한다. 예를 들면 그림 12처럼 '에스테르(ester)'의 경우 휘발성(volatility)을 가지며, '글로불린(globulin)'의 경우 불용성(insolubility)이지만, subClassOf인 '가성글로불린(pseudogloblin)'은 가용성(solubility)이다. 즉 '글로불린'의 subClassOf인 '가성글로불린'은 가용성 속성을 가지고 있으므로 '글로불린'의 제약 조건(constraint condition)이라 할 수 있다. 이렇듯 면역학 온톨로지에서는 물질에 대한 속성 정리를 체계적으로 관리하는 것이 중요하다.

#### 4.4 면역학 온톨로지의 구축 사례

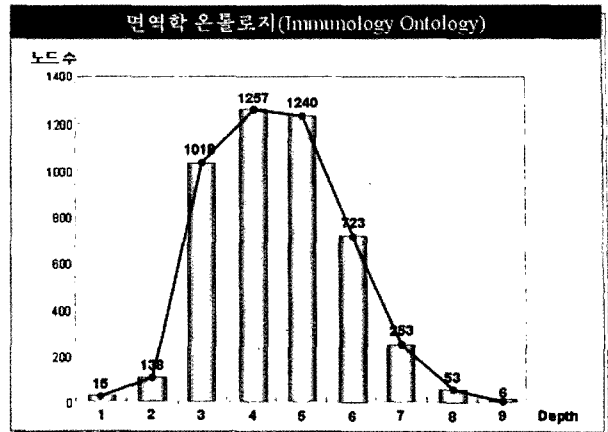


그림 13 면역학 온톨로지의 계층 분포

#### 5. 국가과학기술 R&D 기반정보 온톨로지

국가지식정보자원 및 전문지식정보를 체계적으로 관리함과 동시에 효율적인 지식정보 서비스가 가능하도록 하기 위하여, 이를 효율적으로 체계화시키고 공유화시키기 위한 온톨로지 구축 연구가 진행 중이다(22,23, 24,25,26).

본 연구팀에서 개발한 국가과학기술 R&D 기반정보 온톨로지(Science and Technology R&D Reference Ontology; 이하 STRDR 온톨로지)는 국가과학기술 R&D 기반정보를 분석·활용하여 한국의 과학기술

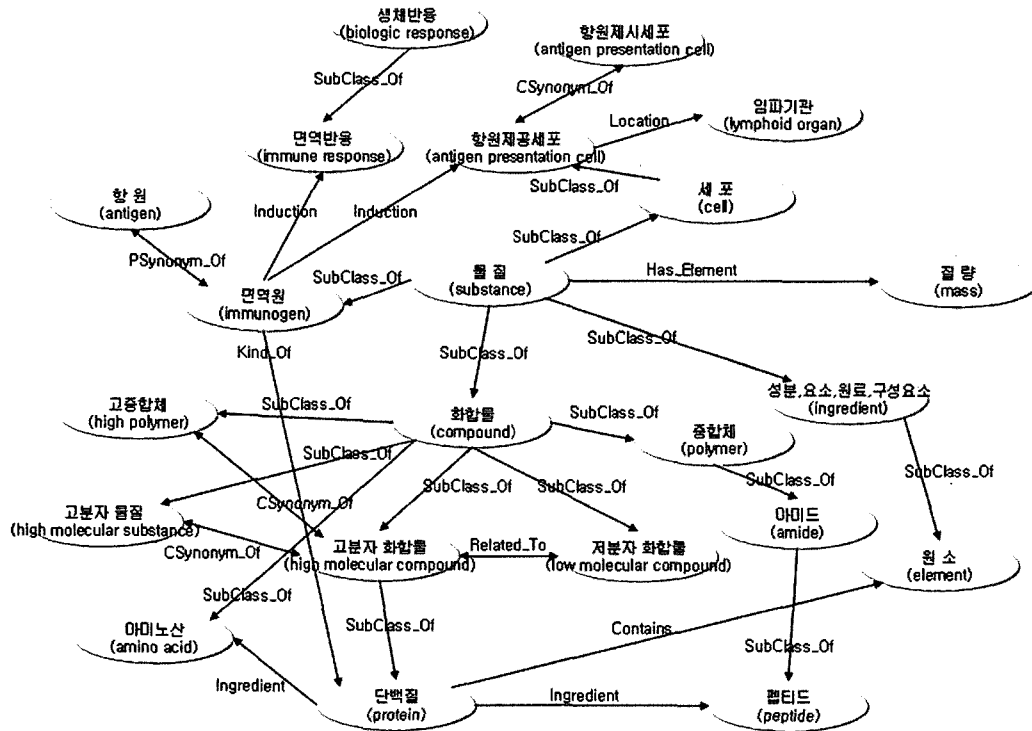


그림 14 면역학 온톨로지의 일부



R&D 지식 정보를 과학기술 전문가를 비롯한 관리자, 일반 사용자 등에게 효율적으로 제공할 수 있도록 체계화시킨 OWL 기반의 온톨로지이다. 이 온톨로지는 시맨틱 웹 환경에서의 과학기술 지식 정보에 대한 자료 저장 구조 및 관리 체계를 반영하고 있으며, 과학기술 통합 정보유통서비스의 효율성을 고려하여 구축하고 있는 온톨로지로서, 시맨틱 웹 기술 기반의 정보유통서비스 시스템에 활용되는 온톨로지이다.

## 5.1 특정 영역 결정 및 분석

### 5.1.1 특정 영역 결정 및 분석

STRDR 온톨로지는 크기는 국내 과학기술 R&D 기반정보에 대한 전반적인 관리 분야가 될 수 있으며, 세부적으로는 프로젝트, 학술정보, 과학기술 인력 분야 등으로 나눌 수 있다.

이러한 특정 영역의 정보 분석을 위한 기초 자료는 다음과 같다.

- KISTI 내부성과물관리 데이터베이스
- 과학기술 통합 검색 YesKISTI
- 한국과학재단/한국학술진흥재단 통합연구인력정보
- 한국과학재단 연구관리시스템/연구정보시스템
- 국가과학기술인력 종합관리시스템
- 연구장비/인력 종합검색시스템
- 국가연구개발사업 통합공고 포털사이트

STRDR 온톨로지 구축을 위한 특정 영역 정보 분석은 사업, 프로젝트, 학술정보(출판물, 지적소유권 등), 기관, 과학기술인력 등의 중점 대상 R&D 정보를 중심으로 이루어진다. 그리하여 기초 자료를 바탕으로 공통적으로 쓰이는 개념이나 특성, 개별적으로 쓰이는 개념이나 특성을 파악하고, 또한 각각의 세부 영역 정보가 다른 세부 영역 정보와 어떠한 관계를 맺고 있는지를 파악한다.

### 5.1.2 Class/Property/Individual 생성 요소 분석

국가과학기술 R&D 기반정보에 대한 전체적인 정보 흐름 구조를 파악한 뒤에는, 온톨로지로 표현 가능한 내부 구조 요소를 파악·분석해야 한다. 즉 Class, Property, Individual 등으로 생성 가능한 요소를 분석·추출하여 온톨로지로서의 반영 여부를 판단해야 한다. 이러한 온톨로지 생성 요소 분석을 위하여 몇 가지 전제를 사용한다.

- 체계적인 특정 영역 정보를 담고 있는 자료 중 데이터베이스의 테이블명, 각 테이블의 필드명, 표의 제목, 표의 각 열과 행의 제목은 온톨로지의 Class 가 될 수 있다. 예) 과제, 지적재산권, 소속기관...

- 데이터베이스나 표의 레코드나 셀 등에 포함된 정보는 Individual이나 특정한 값(value 또는 Datatype) 이 될 수 있다. 예) 한국과학기술정보연구원, 부산대학교, 2005-01-01, 051-501-2391...
- 데이터베이스 테이블 필드 구조와 표의 하부 구성을 통해 Property나 Restriction 정보를 획득할 수 있다. 예) 연구책임자⇒이름(한/영), 소속기관, 소속부서, 직위, 전화번호, 팩스번호...

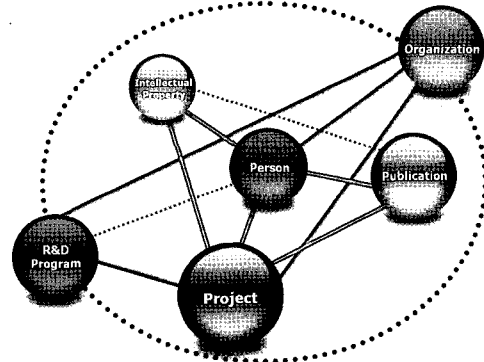


그림 15 국가과학기술 R&D 기반정보의 연관 정보

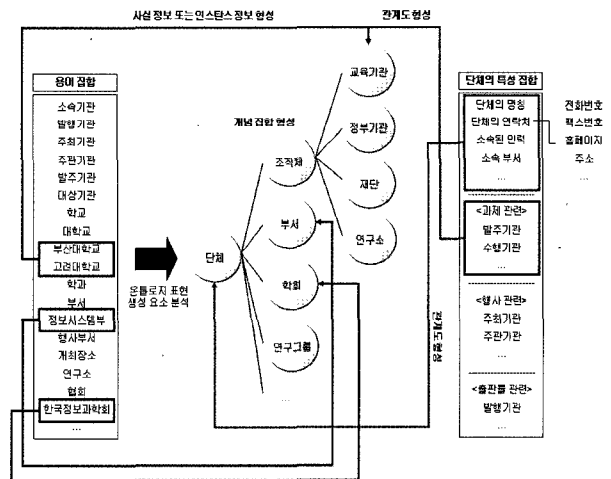


그림 16 '단체'와 관련된 국가과학기술 R&D 기반정보 분석 사례

이러한 전제를 바탕으로 국가과학기술 R&D 기반정보 분석을 통한 사용되는 용어의 집합과 공통 개념 집합, 개념들 간의 관계도를 온톨로지를 구축하기 전 설계할 수 있다. 이러한 정보 분석에서는 언어학적 분석 방법이나 언어처리적 분석 방법 등 적절한 분석 기법이 사용될 수 있다.

## 5.2 웹 온톨로지 언어 결정

### 5.2.1 OWL

STRDR 온톨로지는 시맨틱 웹 기반 기술에서 활용

가능한 구조로 구축하기 위하여 OWL 기반의 온톨로지 언어를 사용한다. OWL은 개념의 일관성을 확보하여 Class와 Property의 개념 및 그들 사이의 관계를 보다 명료하게 정의하도록 하는 웹 온톨로지 언어로서, RDF와 RDFS, DAML+OIL의 모델링 요소를 확장 강화하기 위하여 설계되었다고 할 수 있으며, 기존 언어보다 폭넓은 의미 표현 수단을 제공한다. 특히 STRDR 온톨로지에서는 OWL의 세 가지 하위 언어 중, OWL Lite와 OWL DL을 모두 포함하고 최대의 의미표현력과 RDF의 유연한 문법을 모두 활용할 수 있는 OWL Full 언어를 사용한다.

## 5.2.2 OWL 기반 온톨로지의 필수 요소

### 5.2.2.1 Classes

Class는 Individuals로 구성된 하나의 집합으로 생각할 수 있다. 즉 Class는 동일한 속성을 지니고 있어 하나의 부류로 모아지는 Individuals의 집합을 정의한다. 나아가 임의의 Class를 세분화하여 subClassOf를 이용한 계층 구조를 통해 Individuals의 집합을 정의할 수도 있다. 예를 들어 Conference, Workshop, Seminar 등은 Class:Event의 subClassOf로 구조화될 수 있으며, '한국정보과학회 춘계학술대회' 등은 각종 발표논문, 주최, 주관 등을 가지는 학술대회이므로 Class:Conference의 Individuals에 속한다.

Class 생성에서 중요한 점은 특정 영역에 필요한 Class만을 생성해야 한다는 점이다. 이것은 Class 생성을 위한 특정 영역에 대한 분석이 이루어진 후 결정되어야 할 요소이며, Properties 및 Individuals에 대한 분석이 병행되어야 한다는 점을 의미한다.

### 5.2.2.2 Properties

Property는 Class(또는 Individual) 사이의 관계 및 Class(또는 Individual)와 데이터값 사이의 관계를 표현한다. 또한 Property는 Individuals에 대한 binary relations로 생각할 수 있으며, Individual과 데이터값 사이의 binary relations로 생각할 수도 있다. 즉 특정한 Class와 Class 사이의 관계를 표현하는 것과 마찬가지로 때문이다.

Property는 ObjectProperty와 DatatypeProperty로 구분되는데, ObjectProperty는 Class 요소들간의 관계를 표현할 때 사용되고 DatatypeProperty는 Class 요소가 취해야 하는 데이터의 형식과 값을 표현할 때 사용된다. 예를 들면, Class:ConferencePaper는 hasAuthor 등과 같은 ObjectProperty를 통해 Class:Person과 관계를 가지며, Class:ConferencePaper는 hasYear,

hasFirstPage, hasLastPage 등과 같은 데이터값과 관계를 가지는 것이다.

더 나아가 Property 설정에서는 rdf:domain과 rdf:range를 통해 Property 범위를 제한시켜야 한다. 즉 Property의 적용 도메인(domain: 엄밀히 Class) 집합을 rdf:domain으로 제한 표현하며, Property의 값이 될 수 있는 도메인(엄밀히 Class 또는 Data-type) 집합을 rdf:range로 제한 표현해야 한다.

Property 제약 사항(constraints) 관리를 위해서는, Property Restriction(owl:restriction 구문)의 allValueFrom, someValueFrom 제약 조건을 통해 Property의 값을 제한할 수 있으며, Cardinality(minCardinality/maxCardinality/Cardinality) 제약 조건을 통해 Property의 값의 개수를 제한할 수도 있다.

### 5.2.2.3 Individuals

Individual은 Class의 인스턴스이자 특정 영역의 표현된 실제 객체들이며, Property를 통해 상호 관계를 맺는다. Individual은 특정 영역 온톨로지의 Class 생성 및 Property 설정에 중요한 역할을 담당한다. 특정 Class의 Individual은 Property를 통해 다른 Class의 Individual 또는 Individual Datatype과 관계를 가진다. 이 Individual은 직접적 인스턴스(direct instance)와 Property를 통한 인스턴스(property instance)로 구분될 수도 있다.

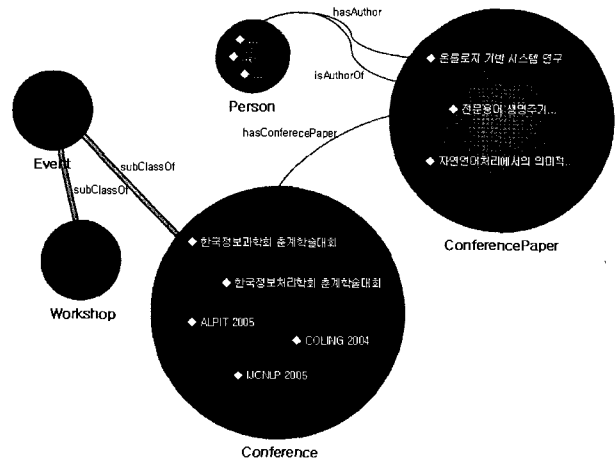


그림 17 Classes/Properties/Individuals의 간략한 표현

## 5.3 온톨로지 내부 구조 설계

온톨로지 내부 구조 설계는 Class, Property/Restriction, Individual에 대한 표기 방식을 비롯하여 이러한 필수 구성 요소의 생성 원칙이 기술되어야 한다.

STRDR 온톨로지의 내부 구조에서 사용된 표현 언어는 Class-영어, Property/Restriction-영어, Individual-

한국어로 표기한다. 그리고 Class Name 설정 방법, Property Name 설정 방법, Class/Property 생성 원칙, Restriction 표현 방법, Individual 생성 원칙 등 세밀한 온톨로지 내부 구조를 설계가 이루어져야 한다.

### 5.4 온톨로지 편집 도구 결정 및 구축

STRDR 온톨로지 구축을 위하여 OWL 지원 여부와 추론 엔진과의 연동성 부분, 온톨로지 구축의 효율성, 온톨로지 구축을 위한 인터페이스 환경 등을 고려하여, 스탠포드 대학에서 지속적으로 개발하고 있는 Protégé 온톨로지 편집 도구를 활용한다. OWL Plug-In을 탑재한 Protégé (v3.1.1)를 사용하여, OWL로 표현하고자 하는 온톨로지를 편리하게 구축할 수 있으며, Jena와의 연동서비스도 일부 제공하고 있어 추론서비스 설계에도 유용성이 있다고 판단된다. 다만 아직까지는 대규모 온톨로지를 구축하기에는 문제가 있는 듯하다.

이러한 일련의 준비 과정을 거친 뒤, Protégé를 이용하여 STRDR 온톨로지를 구축한다. 온톨로지 구축에서도 유효성 검사와 버전 관리 등을 고려하여야 한다. 그리하여 온톨로지 유효성 검사는 구축 지침 반영 여부, 구축된 일부 데이터(sample data)를 이용한 추론 서비스 시스템에서의 수시 활용 검사, Protégé에서 지원하는 consistency 체크를 이용한 내부 문법 구조 검증 등을 시행한다. 다음으로 온톨로지 버전 관리는 일정한 합의를 통해 내부 문법 구조에 반영하거나 저장 형식의 구별을 통해 관리하도록 한다.

### 5.5 온톨로지 평가 및 유지 보수

국내외적으로 온톨로지에 대한 평가 지침이 마련되어 있지는 않다. 그리하여 보편적으로는 전문가에 의한 평가, 애플리케이션에서의 적용을 통한 평가 등으로 온톨로지 평가를 수행할 수 있다. 또한 온톨로지 내부 문법 구조에 대한 자동 문법 오류 검사를 통한 평가, 추론 엔진을 통한 추론 규칙 및 결과 평가, 온톨로지 내부 구조의 정밀 평가 등이 추가될 수도 있을 듯하다.

다음으로 구축된 온톨로지에 대한 유지 보수는 평가에 의한 합의로 온톨로지 유지 보수가 가능해야 한다. 이는 온톨로지 버전 관리 문제와도 연관성을 고려해야 한다.

## 6. 결 론

이상과 같이 온톨로지에 대한 일반적인 인식 설명과 본 연구팀이 개발한 몇 가지 온톨로지를 통해, 온톨로지의 다양한 개발 방법을 살펴보았다. 하지만 아직까지 온톨로지와 관련된 세밀한 연구가 더 진행되어야 할 것

```

<rdf:comment rdf:datatype="&xs:string">연구자 클래스</rdf:comment>
<rdf:subClassOf rdf:resource="#Person"/>
<rdf:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#hasDepartments"/>
    <owl:someValuesFrom rdf:resource="#Department"/>
  </owl:Restriction>
</rdf:subClassOf>
<rdf:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="#hasHeadOf"/>
    <owl:someValuesFrom rdf:resource="#Project"/>
  </owl:Restriction>
</rdf:subClassOf>
</owl:Class>

<owl:Class rdf:about="#Seminar">
  <rdf:comment rdf:datatype="&xs:string">세미나</rdf:comment>
  <rdf:subClassOf rdf:resource="#Event"/>
  <rdf:subClassOf>
    <owl:Restriction>
      <owl:allValuesFrom rdf:resource="#SeminarArticle"/>
      <owl:onProperty rdf:resource="#hasDocuments"/>
    </owl:Restriction>
  </rdf:subClassOf>
</owl:Class>

<owl:Class rdf:about="#SeminarArticle">
  <rdf:comment rdf:datatype="&xs:string">세미나 논문 클래스</rdf:comment>
  <rdf:subClassOf rdf:resource="#Document"/>
  <rdf:subClassOf>
    <owl:Restriction>
      <owl:allValuesFrom rdf:resource="#Seminar"/>

```

그림 18 STRDR.OWL의 일부

```

<!-- ConferencePaper Class Instance (sample) -->
<BaseInformation:ConferencePaper rdf:about="#지능형로봇 환경에서의 질의처리">
  <BaseInformation:hasAbstract rdf:datatype="&xs:string">지능형로봇 (URC: Ubiquitous Robotic Comparison) 개발은 정보통신부의 IT 신성장동력 9대 사업의 하나로서 URC를 위한 인프라시스템 개발, URC를 위한 내장형 컴포넌트 기술개발 및 표준화, URC용 차세대 웹 기반 서비스 플랫폼 개발, 생체인식 표준화 및 시각인식 기술개발로 구성되어 있다. 본 논문은 이 중 URC용 차세대 웹 기반 서비스 플랫폼 개발에서 핵심적으로 요구되는 사용자 질의처리를 위해 정보추출을 도입한다. 정보추출은 정보나 행동을 요구하는 목적지향성 질의를 처리하기 위한 적합한 방법론으로, 본 연구에서 사용하고 있는 Lexico-Semantic Pattern 기반 언어처리는 이미 개방형 질의응답 시스템, 이메일과 같은 비정형 텍스트, 네비게이션 도메인에서의 질의분석에서 이미 그 효율성을 입증한 바 있다. 우리는 정보추출을 지능형로봇 서비스를 위한 날씨, 교통, 그리고 홈오트메이션 도메인 상에서의 질의로부터 질의의미타입인 Predicate과 의미내포적인 Argument를 추출하고 로봇이 이해할 수 있는 질의구조를 생성하는데 이용한다. 본 연구에서는 다양한 애매성과 형태변화가 심한 어휘들을 처리하기 위기 위해 추출 규칙들을 비문법기관과 문맥</BaseInformation:hasAbstract>
  <BaseInformation:hasAuthor rdf:resource="#박동민"/>
  <BaseInformation:hasAuthor rdf:resource="#전충녕"/>
  <BaseInformation:hasAuthor rdf:resource="#성원경"/>
  <BaseInformation:hasAuthor rdf:resource="#손주환"/>
  <BaseInformation:hasAuthor rdf:resource="#정현민"/>
  <BaseInformation:hasDate rdf:datatype="&xs:date">2005-02-01</BaseInformation:hasDate>
  <BaseInformation:hasEnglishKeyword rdf:datatype="&xs:string">Query Processing, Ubiquitous Robotic Comparison, Information Extraction</BaseInformation:hasEnglishKeyword>
  <BaseInformation:hasEvents rdf:resource="#HCI_학술대회"/>
  <BaseInformation:hasFirstPage rdf:datatype="&xs:integer">102</BaseInformation:hasFirstPage>
  <BaseInformation:hasKoreanKeyword rdf:datatype="&xs:string">질의 처리, URC, 정보 추출</BaseInformation:hasKoreanKeyword>
  <BaseInformation:hasLastPage rdf:datatype="&xs:integer">107</BaseInformation:hasLastPage>

```

그림 19 Class:ConferencePaper의 Individual Sample

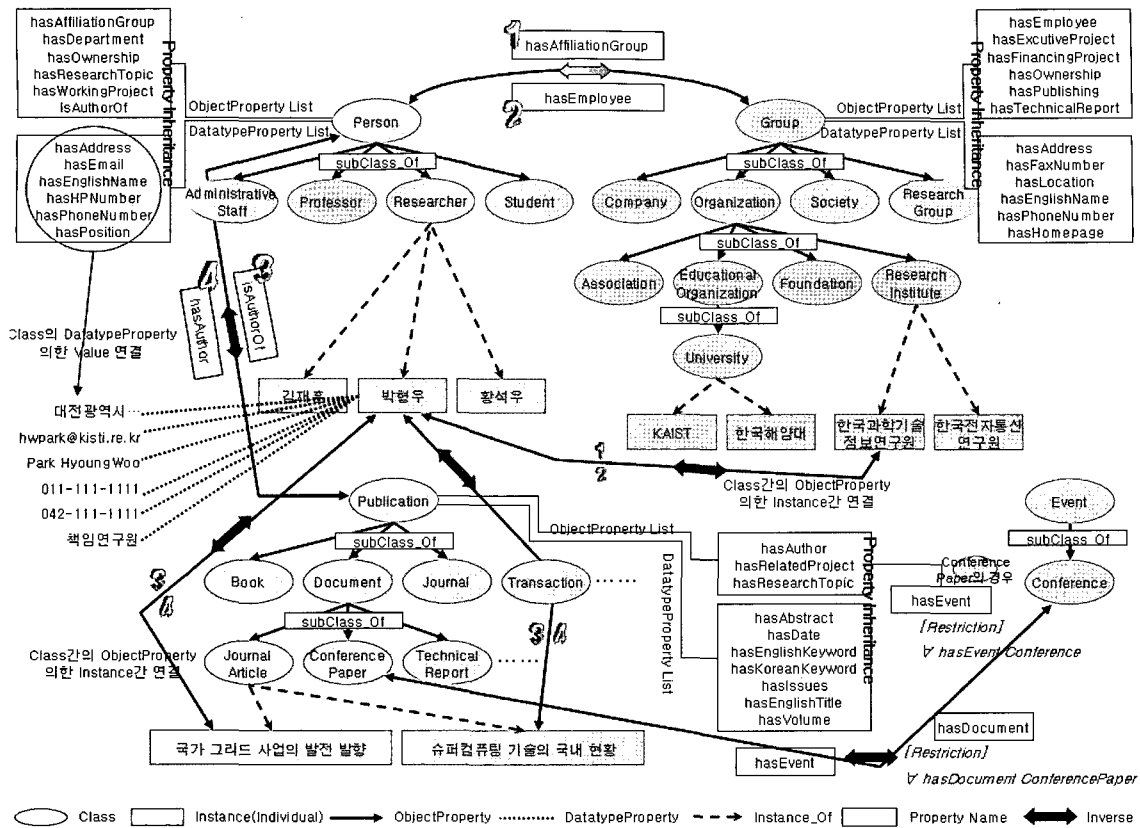


그림 20 국가과학기술 R&D 기반정보 온톨로지 도식적 표현

으로 판단된다.

최근 온톨로지 기술에 대한 연구 개발은 차세대 인공지능 기술 개발이라고 할 만큼 많은 연구자와 기술자들에게 관심의 대상이 되고 있다. 그러나 국제경쟁력을 가진 실질적인 온톨로지에 대한 구축 작업이 미흡한 상태라 할 수 있다. 지식정보의 기하급수적인 증가에 따라, 지식정보를 체계화하고 효율적인 관리 방법에 대한 관심이 증가하고 있지만, 아직까지 국내는 이론적 기반 문제, 기술적 문제, 산업적 문제, 장기간의 연구 개발 지원 미비 등의 연구 개발 구축 환경의 어려움으로 인해 실험적 수준으로 온톨로지와 같은 지식정보 관리 체계 구축 작업이 진행되고 있는 실정이다.

국내의 온톨로지 및 시맨틱 웹 연구 상황은 IWSC 2007의 한국 유치로 통해서 한 단계 더 성장할 수 있는 계기가 마련되었다고 할 수 있다. 이러한 시기에 빠른 시일 내에 온톨로지에 대한 제학문적인 연구 기반 확보를 비롯하여, 정부 차원의 장기간의 지원이 절실히 필요하다고 할 수 있다.

### 참고문헌

[1] Fellbaum, C., WordNet : An Electronic

Lexical Database, The MIT Press, 1998.

[2] Vossen, P., EuroWordNet : A Multilingual Database with Lexical Semantic Networks, The Kluwer Academic Publishers, 1998.

[3] Snasel, V., Moravec, P., Pokorny, J., "WordNet Ontology Based Model for Web Retrieval," International Workshop on Challenges in Web Information Retrieval and Integration(WIRI) 2005, pp. 220-225, 2005.

[4] Niles, I., Pease, A. "Mapping WordNet to the Suggested Upper Merged Ontology," In Proceedings of the 2003 International Conference on Information and Knowledge Engineering, 2003.

[5] <http://www.opencyc.org>

[6] Lenat, D.B. Guha, R.V., Building large knowledge-based system : Representation and inference in the CYC project, Addison Wesley, 1990.

- [7] Kiryakov, A.K., Simov, K.I., "Mapping of EuroWordnet Top Ontology into Upper Cyc Ontology," KAW 2000, 2000.
- [8] Schulze-Kremer, S., Smith, B., Kumar, A., Revising the UMLS Semantic Network, MedInfo2004, 2004.
- [9] Kashyap, V., Borgida, A., "Representing the UMLS Semantic Network Using OWL," ISWC 2003, pp. 1-16, 2003.
- [10] Abasolo, J.M., Gómez, M., "MELISA. An Ontology-based agent for information retrieval in medicine," ECDL 2000 Workshop on the Semantic Web(SemWeb2000), pp. 73-82, 2000.
- [11] <http://www.nlm.nih.gov/mesh/>
- [12] 최호섭, 옥철영, "정보검색시스템과 온톨로지," 정보과학회지, 제22권, 제4호, pp.28-35, 2004.
- [13] Wielinga, B.J., Schreiber, A.Th., Wielemaker, J., Scandberg, J.A.C., "From Thesaurus to Ontology," Proceedings of the international conference on Knowledge capture, pp. 194-201, 2001.
- [14] Gruber, T., "Toward Principles for the design of ontologies used for knowledge sharing," International Journal of Human-Computer Studies, vol.43, no.5/6, pp. 907-928, 1995.
- [15] Uschold, M., Gruning, M., "ONTOLOGIES: Principles, Methods, and Applications," AIAI-TR-191, Artificial Intelligence Applications Institute (AIAI), the University of Edinburgh, 1996.
- [16] 류광택, 웹 온톨로지 개발 지침 연구, 한국전산원 연구보고서, 2004.
- [17] Horridge, M., Knublauch, H., Rector, A., Stevens R., Wroe, C., "A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools - Edition 1.0," The University of Manchester, 2004.
- [18] <http://protege.stanford.edu>
- [19] Kavi, M. Ontology Development for Machine Translation : Ideology and Methodology, Technical Report MCCS-96-292, Computing Research Lab, New Mexico State University, 1996.
- [20] Uschold, M., King, M., Moralee, S., Zorgios, Y., "The Enterprise Ontology," AIAI-TR-195, Artificial Intelligence Applications Institute (AIAI), the University of Edinburgh, 1997.
- [21] Beeferman, D., "Lexical discovery with an enriched semantic network," In Proceeding for the Workshop on Applications of WordNet in Natural Language Processing System, ACL/COLING 1998, 1998.
- [22] 오삼균, "국가지식정보자원관리를 위한 시맨틱 웹 설계 및 정책 방향에 관한 연구", 한국비블리아, 제15권 제1호, pp. 43-67, 2004.
- [23] AKT Reference Ontology, <http://www.aktors.org/publication/ontology/>
- [24] Maedche, A., Motik, B., Stojanovic, L., Studer, R., Volz, R., "Ontology for Enterprise Knowledge Management," IEEE Intelligent Systems, 18(2), pp. 26-33, 2003.
- [25] <http://suo.ieee.org>
- [26] Niles, I., Pease, A., "Toward a Standard Upper Ontology," In Proceedings of the 2nd International Conference on Formal Ontology in Information Systems, 2001.
- [27] <http://www.opengalen.org>
- [28] <http://www.snomed.org/snomedct/>
- [29] <http://www.geneontology.org>
- [30] Davies, J., Fensel, D., Harmelen, F.V., Towards The Semantic Web, JOHN WILEY & SON Ltd, 2003.
- [31] Fensel, D., Hendler, J., Lieberman, H., Wahlster, W., Spinning the Semantic Web, The MIT Press, 2003.
- [32] Daconta, M.C., Obrst, L.J., Smith, K.T., The Semantic Web, Wiley Publishing Inc. 2003.
- [33] Guarino, N., Giaretta, Pierdaniele., "Ontology and Knowledge Bases-Toward a Terminological clarification," In N. Mars (ed.), Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing, pp. 25-32, IOS Press, 1995.

최 호 섭



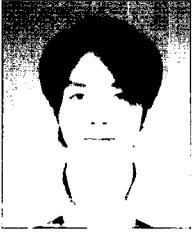
1998 경남대학교 국어국문학과(문학사)  
2000 경남대학교 국어국문학과(문학석사)  
2000~2001 한국전자통신연구원  
지식정보검색연구팀 파견연구원  
2002~현재 (주)시소러스 선임연구원, 감사  
2002~현재 울산대학교 컴퓨터정보통신  
공학과 박사수료  
관심분야: 한국어정보처리, 온톨로지, 지식  
베이스, 언어이해, 지식처리  
E-mail : hoseop@ulsan.ac.kr

임 지 희



2003 울산대학교 컴퓨터정보통신공학부  
(공학사)  
2005 울산대학교 컴퓨터정보통신공학과  
(공학석사)  
2005~현재 울산대학교 컴퓨터정보통신  
공학부 박사과정  
관심분야: 한국어정보처리, 온톨로지, 정보  
검색  
E-mail : jihlim@mail.ulsan.ac.kr

배 영 준



2004 울산대학교 컴퓨터정보통신공학부  
(공학사)  
2006 울산대학교 컴퓨터정보통신공학과  
(공학석사)  
2006~현재 울산대학교 컴퓨터정보통신  
공학과 박사과정  
관심분야: 한국어정보처리, 전문용어, 정보  
검색  
E-mail : young4862@mail.ulsan.ac.kr

최 수 일



2005 울산대학교 컴퓨터정보통신공학부  
(공학사)  
2005~현재 울산대학교 컴퓨터정보통신  
공학과 석사과정  
관심분야: 한국어정보처리, 어휘학습, 기계  
학습  
E-mail : our1223@mail.ulsan.ac.kr

옥 철 영



1982 서울대학교 컴퓨터공학과(학사)  
1984 서울대학교 컴퓨터공학과(석사)  
1993 서울대학교 컴퓨터공학과(박사)  
1994 러시아 TOMSK 공과대학 교환교수  
1996 영국 GLASGOW 대학교 객원교수  
1984~현재 울산대학교 컴퓨터정보통신  
공학부 교수  
관심분야: 한국어정보처리, 의미분별, 온톨  
로지, 지식베이스, 기계학습, 문  
서분류  
E-mail : okcy@ulsan.ac.kr