

연구 활동 지원을 위한 적응형 연구정보 지원 포털 구축에 관한 연구

A Study on Developing an Adaptive R&D Information Service Portal

최 성 필(Sung-Pil Choi)*

조 현 양(Hyun-Yang Cho)**

목 차

- | | |
|----------------------------|---------------------------|
| 1. 서 론 | 4. 지식베이스 구축 및 서비스 알고리즘 개발 |
| 2. 지식기반 전문가 시스템의 정의와 관련 연구 | 5. 연구정보 지원시스템 실행 |
| 3. 적응형 연구정보 지원서비스 모델 | 6. 한계점과 향후 연구 방향 |
| | 7. 결 론 |

초 록

본 논문은 특정 분야에서 이미 일정 수준에 있는 전문가를 활용하여 연구 초기에서 발생하는 신진 연구자들의 이러한 어려움을 좀 더 쉽게 해결해 줄 수 있는 방법을 제시하고자 하였다. 이를 위하여, 단순한 형태의 연구 계획 생성 기능(research scheduling function)을 갖춘 적응형 연구지원 포털시스템을 개발하였으며, 기존의 지식베이스 구축방법에서 탈피하여 분야 전문가가 자신의 지식을 효율적으로 표현할 수 있는 스키마 구조와 구축방법론을 제시하였다. 본 논문에서 개발된 시스템의 단순성과 확장 가능성으로 향후 다양한 발전적 모델이 될 수 있을 것이다.

ABSTRACT

This paper suggested a way to solve the problems by using domain experts who are already in the significant level of knowledge in those fields. For the purpose of achieving our goal, a very simple and efficient approach to construct the knowledge-base which can play an important role in providing researchers with essential information in need was proposed. In addition, the Adaptive R&D Information Service Portal with a new schema structure and a construction method of representing expert's knowledge efficiently was developed. With the simplicity and expandability of the proposed system it can be a good model for a similar system to be developed.

키워드: 전문가시스템, 지식베이스, 정보서비스시스템, 지식포털, 시맨틱 웹, RDF, 연구정보
Expert System, Knowledge Base, Information Service System, Knowledge Portal,
Semantic Web, RDF, R&D Information

* 한국과학기술정보연구원 정보기술개발단 정보시스템개발팀 선임연구원(spchoi@kisti.re.kr)

** 경기대학교 문헌정보학과 교수(hycho@kgu.ac.kr)

논문접수일자 2007년 11월 20일

게재확정일자 2007년 12월 5일

1. 서론

일반적으로 특정 연구 분야에서 새로운 연구를 시작하려는 많은 신진 연구자들은 그들의 첫 출발점을 찾는데 많은 어려움을 겪게 된다. 가끔은 적극적이고 헌신적으로 그들을 돕고 지원하는 선배들이나 교수들이 있어서 그 출발점 상에서의 어려움에 대한 해결책이 되기도 하지만, 대부분은 본인 스스로 해결해야 한다. 신진 연구자들이 신규 분야에서의 연구 활동을 위한 시작점은 여러 가지가 있을 수 있으나 대표적인 예를 들면, 특정 분야의 전문가가 기고한 학술 논문, 연구 개발 동향, 특정 분야에서 독보적인 연구 성과를 내거나, 왕성한 연구 활동을 수행하고 있는 전문가들에 대한 인적 사항, 그리고 기타 전문분야의 문헌 정보가 있다. 연구의 본격적인 수행을 위한 준비작업인 이러한 정보의 수집은 인터넷이 발달한 현재에도 연구자들에게 지적인 노력의 요구는 물론 매우 복잡하고 시간 소모적인 작업으로 여겨지고 있다.

신진 연구자들이 연구에 필요한 기초적인 정보를 수집하기 위하여 시간과 노력이 요구되는 반면, 이미 특정 분야에서 다양한 경험이 있는 전문가들에게는 이러한 정보는 지극히 당연한 공리(axiom)에 가깝다. 다시 말해서, 그들은 이미 자신의 관심 분야인 전문 분야의 연구에 필요한 정보를 수집하고, 활용하였으며, 종합적이고 체계적으로 정리된 정보를 가지고 있다. 해당 분야의 전문가 입장에서 볼 때 연구의 출발을 위한 기초적인 수준의 정보를 깊게 다룬다는 것은 어떻게 보면 시간 낭비에 가까울 수 있다. 어떻게 하면 신진 연구자들이 위와 같은 정보를 쉽게 획득할 수 있는가에 대한 문제는

오래 전부터 인공지능 분야에서 지식표현 및 지식베이스 구축 등과 같은 방법론으로 해결하고자 하는 많은 노력이 있었다. 이러한 노력의 결과로 정교하게 정의된 지식표현기법에 의해서 구축된 지식베이스를 활용한 분야별 전문가 시스템 개발도 많이 이루어지게 되었다. 그러나 이러한 시도에도 불구하고 대부분의 결과물들이 제대로 활용되거나 발전되지는 못하였다.

일반적으로 지식베이스를 구축하는 가장 중요한 목표 중의 하나는 바로 추론(Inference)이었다. 전통적인 인공지능 입장에서의 추론은 서술 논리(Description Logic), 1차 논리(First Order Logic) 등의 기반 로직(Logic)을 기반으로 표현된 인스턴스(Instance) 집합 즉 지식베이스에서 새로운 지식을 발견해 내는 작업이다. 이러한 복잡한 형태의 추론이 이루어지기 위해서는 대상이 되는 인스턴스 집합이 매우 정교하게 표현되어야 하며, 오류가 있어서는 안 된다. 그러나 분야별 전문가는 말 그대로 자신의 분야에 전문적인 지식과 경험을 가지고 있는 사람이다. 이러한 사람들에게 인공지능의 로직에 기반한 지식표현방법을 익히게 하고 이를 활용하여 자신의 머릿속에 있는 지식을 표현하게 하는 작업 자체가 쉽지 않았다.

이러한 상황 하에서, 본 논문은 어떻게 하면 특정 분야에서 이미 일정 수준에 올라와 있는 전문가를 활용하여 상기에 언급된 연구의 시작 단계에서 발생하는 신진 연구자들의 이러한 어려움을 어떻게 덜어줄 수 있는가에 대한 문제를 다룬다. 이를 위하여, 사용자의 현재 지식수준과 목표수준을 기반으로 한 적응형 연구지원 포털시스템과 단순하고 직관적인 지식베이스 구축 방법론을 제안한다.

본 논문은 다음과 같이 구성된다. 우선 2장에서는 전통적인 인공지능분야에서 다루었던 전문가 시스템의 다양한 응용을 비롯하여, 현재 지식표현의 한 가지 방법으로서 널리 활용되고 있는 시맨틱 웹 프레임워크와 관련 연구에 대해서 다룬다. 3장은 목표 시스템에 대한 명확한 정의와 구축되는 지식베이스의 모델링 등에 관한 내용을 포함한다. 실질적인 구현에 따른 여러 가지 고려사항 및 아이디어에 대한 설명은 4장에서 다루었다. 5장에서는 구축된 시스템에 대한 화면을 살펴보면서 실용성 및 사용자 편의성 등에 대해서 토의하고, 6장에서는 현 시스템에 대한 자가 분석을 통해 발전 방향을 모색해 본다. 마지막으로 7장에서 결론을 맺는다.

2. 지식기반 전문가 시스템의 정의와 관련 연구

전문가 시스템은 특정 분야의 전문지식을 축적된 지식베이스와 추론의 원리를 활용하여 의사결정권자들이나 기술전문가들이 필요로 하는 해답을 제공하기 위한 시스템이다. 따라서 전문가들의 깊이 있는 지식을 어떻게 그리고 얼마나 정확하게 이용 가능한 지식베이스로 변환시키느냐 하는 것이 매우 중요하다. 그러나 한 가지 분명한 것은 지식베이스에 변환되고 축적된 전문지식은 항상 부족하며, 또한 적절한 전문지식에 대한 변환이 항상 가능한 것도 아니라는 것이다. 특정 분야에 대한 깊이 있는 지식을 갖춘

이 동형 컴퓨터가 있다면 이는 결국 그 분야에 대한 수 세기에 걸친 지식을 총 망라한 하나의 전문분야지식이라고 생각할 수 있을 것이다. 이러한 시스템은 현황 평가(situation assessment)나 장기 계획(long-range planning)을 수행해야만 하는 관리자에게도 매우 유용할 것이다. 현재 특정 분야 전문지식의 아주 작은 일부만을 탑재한 수많은 소형 시스템들이 존재한다. 이러한 사실은 더 큰 시스템도 가능하리라고 생각하는 하나의 증거가 될 수 있다.

전문가 시스템의 정의에 대한 또 다른 견해는 전산학 입장에서 이를 하나의 커다란 컴퓨터 프로그램으로 보는 것이다. 다시 말해서, 전문가 시스템 혹은 지식 기반 시스템이란 다수의 전문가들이 가지고 있는 세부 분야별 지식(subject-specific knowledge)을 포함하고 있는 하나의 컴퓨터 프로그램이라고 정의하고 있다. 이런 부류의 프로그램들은 1960년대와 1970년대에 처음으로 개발되었으며, 1980년대에 와서는 상업적으로 활용되기 시작하였다. 가장 일반적인 형태의 전문가 시스템은 다음과 같은 특징을 가지고 있다. 첫째, 사용자가 해결하기 원하는 어떤 문제에 대한 상세한 수준의 정보를¹⁾ 분석하는데 필요한 규칙집합을 포함하고 있다. 둘째, 분석된 정보를 기반으로 다양한 수학적 방법론을 통한 문제 분석을 수행한다. 마지막으로, 분석된 결과를 중심으로 문제를 해결하는데 필요한, 혹은 전문가시스템의 오류를 수정하는데 필요한 액션 시나리오(user actions scenario)를 제공한다. 따라서 사용자는 전문가 시스템과의 상호 교류를 통해 자신의 최종 목

1) 이 정보는 사용자가 직접 시스템에 제공해야 하며, 만일 사용자가 의도한 바와는 다른 분석결과를 출력하면 사용자는 이를 수정할 수 있는 정보를 다시 시스템에 입력해야 한다.

표를 명확하게 할 수 있으며, 시스템의 잘못된 결과를 수정하는데 도움을 줄 수 있다. 이러한 정보교환 과정에서 전문가 시스템은 추론기능을 활용하게 된다.

위에서 기술한 바와 같이, 하나의 전문가 시스템을 만들고 이것이 실질적으로 활용 가능하게 하려면 대상 분야(target domain)에 대한 명확한 정의를 내릴 수 있어야 하며, 또한 그 분야는 충분히 표현 가능해야 한다. 다시 말해서, 대상 분야가 너무 넓으면, 수작업으로 이루어지는 지식표현단계에서 어려움을 겪게 되며, 이는 필연적으로 표현력 부족 현상을 수반한다. 이렇게 구축된 지식베이스는 그 추론능력이 매우 떨어지게 되므로, 사용자의 만족도 측면이나 활용도 측면에서 좋은 점수를 받기 힘들다.

XML과 시맨틱 웹의 출현으로 많은 사람들이 구조화된 정보 혹은 심지어 잘 조직화된 지식을 표현하는 작업에 더 이상 어렵게 생각하지 않게 되었다. 이제는 일반화된 사실로 받아들여지고 있으나, 시맨틱 웹 프레임워크의 가장 흥미로운 장점은 바로 지식을 생성하고 재사용하는 명시적인 기제를 제공한다는 것이다. 분야 전문가들은 RDF(Resource Description Framework), RDFS(RDF Schema) 혹은 OWL 표준을 기반으로 자신의 지식을 비교적 쉽게 표현할 수 있게 되었다.

전통적인 인공지능 분야에서 연구되는 논리(logic) 기반의 지식베이스 구축과 시맨틱 웹에서 설명하고 있는 직관적인 표준 프레임워크 기반(RDF, RDFS, OWL 등)의 지식표현에는 상당한 괴리가 존재한다. 또한 이러한 괴리감은 두 분야를 각각 연구하고 있는 학자들 사이에 많은 논쟁거리를 제공하고 있다. 가장 대표

적인 논쟁주제는 바로 추론(inference)이다. 사실 현재 가장 널리 활용되고 있는 자바기반 시맨틱 웹 프레임워크(Semantic Web Framework)인 Jena를 살펴봐도 이러한 논쟁거리가 사실임을 알 수 있다. 즉, Jena에서 제공하는 추론기능(inference support)은 대부분 전이관계(transitive relation) 혹은 수반 문장(entailment statement)을 다루는 것에 한정된다. 전통적인 인공지능 분야에서 다루고 있는 1차 논리나 서술 논리 기반의 범용 추론 기능은 Jena에서는 대부분 생략되어 있다. 이는 현재까지 개발된 OWL이 서술 논리를 기반으로 하고 있고, 이 서술 논리의 대부분을 포괄하는 OWL-DL에 대한 활용 시나리오나 표준 프레임워크에 대한 명확한 정의가 아직까지 확립이 되어있지 않기 때문이다.

이미 전통적인 인공지능 분야에서 주장하는 방법을 이용한 지식의 표현은 여러 가지 한계점을 드러내고 있다. 세부적으로 표현하자면, 특정 분야에 대한 전문가들의 지식을 표현하여 지식베이스를 구축하는데 있어서 인간이 소유한 지식을 기반으로 실제 행동에서 나타날 수 있는 다양한 변수에 대한 고려가 전통적인 방법에서는 부족했다. 즉, 완벽한 지식베이스를 구축하고 이에 따르는 완벽한 추론을 통한 새로운 지식의 창출이라는 하나의 커다란 프로세스가 인간의 지식습득, 표현, 변환, 발전이라는 유연하고 자연스러운 행위나 절차와는 차이가 존재한다는 사실을 깨닫게 되었다. 이러한 맥락에서 시맨틱 웹은 전통적인 지식베이스 구축에서 추구했던 목표들 중에서 실행 가능하고, 활용 가능한 것을 중심으로 선택했다고 볼 수 있다. 온톨로지 구축 프레임워크의 근간이 되

는 RDF를 보면 이러한 주장이 사실임이 입증된다. RDF에서는 세상의 모든 구체적이거나 혹은 추상적 개체가 모두 리소스로 표현될 수 있다고 가정하며, 이러한 리소스들을 표현하고, 설명하고, 상호 연관성을 결정하는 것 자체도 지식베이스 구축의 일부분이 되는 것이다. 리소스들을 가능한 한 더 정교하게 표현하기 위한 부가지원도구가 바로 RDFS이며, 이러한 정교하고 사실적인 표현을 위한 토대를 형성하는 것이 온톨로지 구축으로 확대, 발전되었고, 이러한 결과로 OWL이 탄생하게 된 것이다.

사용자에게 만족감을 주지 못하는 지식베이스나 추론 기능은 활용 가치가 없다. 본 논문의 최종 목표는 새롭게 한 분야에 진출하고자 하는 신진 연구자가 연구에 필요한 기초적인 정보를 입수하는데 있어서의 어려움과 불편함을 최소화시키기 위한 연구정보 지원 포털을 구축하는 것이다. 이를 위해서 전통적인 인공지능에서 주장하는 복잡하고 비효율적인 추론기능을 포기하고 직관적이고 구축하기 쉬운 지식베이스를 제작, 활용하는 방법을 제시하고자 한다. 또한, 분야 전문가들이라면 누구나 가지고 있는 특정 리소스²⁾에 대한 난이도 정보³⁾, 그리고 현재 연구하고 있는 분야의 세부 주제분야에 대한 연구정보는 물론 연구순서에 대한 지식 등을 직접 지식베이스에 표현하고 이를 개인화에 효율적으로 활용할 수 있는 방법을 제시하고자 한다.

본 논문은 특정 이론이나 가설 등의 실증적

증명 혹은 새로운 가설의 제안을 목표로 하지 않는다. 기존에 존재하던 전문가 시스템에서의 다양한 접근방법에 대한 고찰과 함께, 지식베이스를 구축해야 하는 분야 전문가들과 이를 활용하는 사용자들이 모두 일정 수준의 만족도를 얻을 수 있는 실질적인 시스템 개발에 그 목표를 둔다. 본 논문에서 다루지 못한 중요한 부분은 바로 구축된 시스템의 사용자⁴⁾ 기반 평가이다. 연구의 최종 목표가 사용자 중심의 서비스 시스템 개발이므로 이에 대한 평가를 위해서는 반드시 실질적인 사용자들이 관여된 만족도 조사, 분석이 수반되어야 한다. 현재 이를 위한 기반 작업을 수행 중에 있으며, 이는 향후 연구과제로 남겨둔다.

3. 적응형 연구정보 지원서비스 모델

본 연구를 통하여 개발될 시스템은 신규 연구자가 초기 연구 활동을 개시하는데 있어 요구되는 다양한 정보를 보다 쉽게 획득할 수 있도록 지원하기 위한 적응형 연구지원 포털시스템이다. 본 논문의 최종 목표시스템은 연구지원을 위한 지식베이스를 포함하고 있는 활용정보(Utilized Information)와 실지로 추론을 실행하는 실행 모듈(Executing Modules)로 구성된다. 이는 전통적인 전문가 시스템에서 채택하는 구조와 일치하는 것으로 실행 모듈의

2) 여기서 리소스는 단일 연구지원정보(예: 특정 논문, 서적, 웹사이트, 인물정보 등)를 일컫음.

3) 특정 리소스가 가지는 어려움의 정도를 이야기한다. 예를 들어, 한 분야의 서적들이라고 하더라도 그 난이도가 차이가 날 수 있고, 전문가들은 이에 대한 정보를 대부분 가지고 있다.

4) 사용자는 지식베이스를 구축하는 분야전문가와 만들어진 시스템을 활용하는 신진연구자 모두를 포괄한다.

활용 정보에 대한 독립성을 유지할 수 있다.

3.1 연구지원 지식베이스와 활용 정보

3.1.1 연구지원 지식베이스

구지원 지식베이스(Research Supporting Knowledge Base)는 신규 연구자가 특정 분야에서 연구를 시작하는데 필요한 필수 정보를 포함한다. 지식베이스를 구축하기 전에 우선적으로 고려되어야 할 사항은 본 논문에서 구축되는 지식베이스의 정의와 전체 구성요소에 대한 정형화(formalization)이다. 이와 관련하여, 광의의 지식베이스에 대한 정의와 더불어 연구지원을 위해 필요한 각종 필수 요소들을 식별하고 정의한다. 연구지원 지식베이스를 KB_R 이라고 표기하면,

$$KB_R = \{x | x \in \Sigma\} \quad (1)$$

다시 말해서, 본 논문에서의 지식베이스는 가상의 집합 Σ 에 속하는⁵⁾ 요소들의 집합이다. 여기서 Σ 은 특정 전문분야를 구성하는 자원(resource)⁶⁾ 집합으로서 다음과 같이 정의된다.

$$\begin{aligned} \Sigma = \{ \sigma | \sigma = \langle id_o, rt_o, obj_o \rangle \\ \wedge obj_o = R(id_o, rt_o) \} \end{aligned} \quad (2)$$

Σ 을 구성하는 개별 트리플(triple)들은 자원 식별자($id_o \in I$), 연관 관계($rt_o \in C$), 그리고 대상 개체($obj_o \in O$)로 구성된다.⁷⁾ 자원 식별자 집합 I 는 특정 지식베이스 내에 존재하는 모든 자원을 구별할 수 있는 단일 식별자들의 집합이다. 이 집합의 원소를 표현하는 방법은 매우 다양할 수 있다. 예를 들어, 시맨틱 웹의 핵심 프레임워크인 RDF는 이를 위한 표현 방법으로 URI(Uniform Resource Identifier)를 채택하고 있다. 연관 관계 집합 C 는 두 개의 개체를 연결시켜주는 관계정보들을 그 원소로 포함하고 있다. 마지막으로 대상 객체 집합 O 는 자원 식별자와 관계정보로 연결되는 목적 개체(objective resource)들을 그 원소로 포함하고 있는 집합으로서, 다음과 같이 표현될 수 있다.

$$O = I \cup \Omega \quad (3)$$

여기서 Ω 는 “특정 개체를 표현할 수 있는 자연어(natural language)로 구성된 모든 가능한 문자열 집합”이다. 즉, 대상 객체 집합 O 는 자원을 식별하고 지시하는데 쓰일 수 있는 식별자 집합 I 와 이 문어적 자연어 집합(literary language set, Ω)의 합집합이라고 볼 수 있다. 이는 다수의 개체가 지식베이스에 존재할 때, 특정 개체가 다른 개체와 연관관계정보로 연결되는 경우와 특정 개체에 대한 특성 지정(property description)⁸⁾

5) 추상 집합의 특성을 가지는 개체라고도 표현될 수 있다.

6) RDF(Resource Description Framework)에서 다루는 자원(resource)에 대한 개념을 차용하였다.

7) 집합 I, C, O 는 모두 포괄적인 추상 집합이다. 다시 말해서, 이 집합들의 원소들은 문자, 문자열, 숫자 등 일반적으로 사람에 의해서 표기될 수 있는 모든 개체(object)를 나타낸다.

8) 여기서 특성지정(property description)은 특정 개체에 대한 정의, 특징 등과 같은 자질들을 표현하는 것을 의미한다.

이 동시에 가능하도록 대상 객체 집합을 정의하기 위한 것이다.

식(2)에서 볼 수 있듯이, 개별 개체들 사이의 관계는 함수 $R: I \times C \rightarrow O$ 로 나타낼 수 있다. 이 함수는 자원 식별자($id_o \in I$), 연관 관계($rt_o \in C$)를 입력 받아서 그에 해당하는 대상 객체($obj_o \in O$)를 출력한다. 본 논문에서 구축되는 지식베이스의 개별 개체들은 모두 위와 같은 형태를 가진다. 이는 전통적인 인공지능 분야에서 다루는 총체적 추론의 단점인 속도의 급속한 저하나 완전성(completeness) 확보의 어려움 등을 피할 수 있음과 동시에 매우 단순한 형태의 개념망 항해(semantic lattice navigation 혹은 spreading activation)를 수행할 수 있으므로 직관적인 추론 기능을 제공할 수 있고, 지식베이스의 증분 확장(incremental expansion)을 통한 지속적인 안정성 및 포괄성 확보가 용이하다. 이러한 특성은 이미 RDF 체계에서 이미 채택하여, 웹 기반의 자원들에 대한 표준적 기술 메커니즘으로 활용되고 있으며, 본 논문도 실질적인 지식베이스는 이 RDF에 기반하여 구축되었다. 위에서 기술한 일반적인 지식베이스의 형태와 정의에 기초하여, 본 논문에서 구축하고자 하는 연구지원 지식베이스의 가장 큰 특징은 바로 함수 $R: I \times C \rightarrow O$ 의 확장 및 변형이다. 즉,

$$R: I \times C \rightarrow O \Rightarrow \widehat{R}: I \times C \times D \times G \rightarrow O \quad (4)$$

위의 식(4)에서 보는 바와 같이, 기존의 R 은 단순히 두 개의 인자(argument: 자원 식별자, 연관 관계)를 채택하였으나, 변형된 함수 $\widehat{R}$ 은 총 네 가지의 인자를 가진다. 여기서 D 는 자원 난이도(Resource Difficulty)를 나타내며, G 는 사용자 도달목표(User Goal)를 나타낸다. 자원 난이도(Resource Difficulty)란 분야 전문가들이 판단할 수 있는 현재 자원의 내용적 수준이다.⁹⁾ 예를 들어, 특정 분야에 대한 서적을 살펴보면, 같은 분야라고 하더라도 “~의 기초(Introduction to ~)”, “기본 ~ (Elementary ~)”과 같은 제목의 서적은 모두 난이도가 낮은 편에 속하는 자원이라고 볼 수 있다. 반대로 그 분야에서 최근에 만들어진 이론이나 아이디어를 집필한 서적 혹은 동일 분야에서 상위 수준으로 판단되는 서적들에 대해서는 난이도가 높다고 가정한다. 한편, 사용자 도달목표는 앞에서 언급된 자원 난이도와 동일한 개념과 특성을 가진다. 한 가지 차이점은 이것은 현재 사용자(신진 연구자)가 자신이 선택한 분야에서 도달하기 희망하는 난이도 수준으로 활용된다는 것이다. 자원 난이도와 사용자 도달목표는 다음이 표현될 수 있다.

$$\begin{aligned} D &= \{easy(1), medium(2), difficult(3)\} \\ G &= \{skin deep(1), basic(2), advanced(3)\} \end{aligned} \quad (5)$$

괄호 안의 번호는 각 구성항목을 나타내는

9) 본 논문에서는 특정 자원이 가지고 있는 내재된 난이도(innate difficulty)를 나타내지 않는다. 지식베이스를 구축하는 전문가 입장에서 이는 절대적인 기준이 없기 때문이다. 그 대신, 현재 지식베이스를 구축하는 분야 전문가의 입장에서 정의를 내리고 있다. 이는 일반적으로 지식베이스는 그것을 구축한 분야전문가의 주관성이 어느 정도 포함되어 있기 때문이고, 결국은 이 말은 지식베이스의 평가를 이론적 혹은 계산적으로 수행할 수 없음을 나타낸다. 구축된 지식베이스의 가치평가는 완전히 사용자의 몫으로 봐야 한다.

식별자이다. 피상적으로는 매우 단순하고 쉬운 정의라고 볼 수 있으나, 본 논문의 가장 큰 관심 분야 중의 하나인, 지식베이스 구축의 용이성 확보라는 측면에서는 의미가 있다. 따라서 지식베이스 구축을 담당하는 분야 전문가는 각 요소 자원마다 자신이 판단한 대로 자원 난이도를 지정한다. 그리고 구축된 지식베이스를 활용하는 사용자는 자신의 도달 목표 수준을 입력함으로써, 현재 자신이 필수적으로 참조해야 할 자원을 탐색할 수 있다.

3.1.2 세부 분야 정의

본 논문에서 다루는 특정 분야에 대한 하부 요소 분야에 대한 정의 및 활용은 다음과 같은 두 가지 목적을 가진다.

- 사용자에게 특정 분야를 구성하는 하부 분야에 대한 정보를 제공함으로써 그들로 하여금 현재 지식베이스에 대한 구조적인 접근을 가능하게 한다.
- 사용자가 인과 관계에 있는 하부 요소 분야에 대한 절차적인 접근을 할 수 있도록 도와준다.

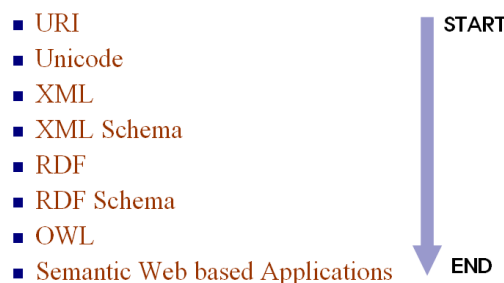
예를 들면, 아래 그림과 같이, 본 논문에서 구축되는 시맨틱 웹 분야 지식베이스는 총 8개의 세부요소 기술 분야를 포함하고 있다. 다시 말해서, 시맨틱 웹 연구의 최종 종착점으로서 “Semantic Web 기반 어플리케이션”이라는 하부 기술 분야를 지정하였고, 연구의 시작점으로서 “URI(Universal Resource Indicator)” 기술 분야를 선택하였다.

우선 사용자는 지식베이스 내에서 시맨틱 웹을 구성하는 하부 기술에 대한 전반적인 이해와 함께 하부 기술별로 전문가에 의해 분류되어 있는 자원을 수직적으로 탐색할 수 있다. 이와 동시에 특정 하부 기술을 연구하기 위해서 선행되어야 하는 기반 기술에 대한 체계적인 이해가 가능하다.

3.1.3 사용자 프로파일 정보

본 논문에서 개발되는 최종 목표 시스템이 사용자 적응형 시스템(user adaptive system)이므로, 시스템이 현재 사용자의 정보를 알기 위한 다양한 메커니즘이 필요하다. 이를 위해서는 시스템을 실질적으로 사용하는 사용자에 대

□ There are 8 sub-areas (Technologies) in the area of Semantic Web



〈그림 1〉 시맨틱 웹 분야의 세부 분야 및 연구 순서 정보

한 명확한 정의와 특징 파악이 필수적이다. 본 논문에서 구축된 지식베이스와 서비스시스템을 사용하는 사용자에게 대해서 다음과 같은 두 가지 가설을 세운다.

첫째, 본 논문에서 개발되는 서비스를 사용하는 모든 사용자는 특정 분야에서 자신의 현재 지식수준 혹은 전문성 정도를 “상(high)”, “중(middle)”, “하(low)”로 표현할 수 있다.

둘째, 본 논문에서 개발되는 서비스를 사용하는 모든 사용자는 특정 분야에서 자신의 최종 목표를 식(5)의 사용자 도달목표 집합의 특정 원소로서 정의내릴 수 있다.

위의 두 가지 가설에 근거하여 사용자 프로파일 정보는 다음의 표와 같이 정의한다. 다음의 <표 1>은 두 가지 사용자 프로파일 요소에 대한 상세 설명을 포함하고 있다.

상기의 <표 1>에서 보는 바와 같이, 사용자의 현 수준이나 최종 목표는 지식베이스를 구축할 때 전문가가 표기하는 자원 난이도(D)를

기준으로 정의된다. 그 이유는 대부분의 사람들이 가지고 있는 자신에 대한 판단기준이 워낙 다양하고 정량적으로 평가하기가 쉽지 않으므로, 지식베이스 구축 당사자인 분야 전문가의 판단기준으로서 사용자를 구분하기 위하여 사용되는 것이다.

3.2 지식베이스 실행 모듈 (Executing Modules)

본 논문에서 개발되는 서비스 시스템의 모태라고 할 수 있는 실행 모듈은 구축된 지식베이스를 바탕으로 사용자에게 세 가지의 주요 기능을 제공한다. 첫째, 시스템은 특정 분야에 대한 유용한 정보를 사용자에게 제공한다. 여기서 유용한 정보란 그 분야의 대표적인 기술 논문, 서적, 웹페이지, 그리고 인적자원 등 연구에 필요한 다양한 형태의 정보를 포함한다. 이 모든 정보는 분야 전문가 입장에서는 기본적인

<표 1> 사용자 프로파일 정보

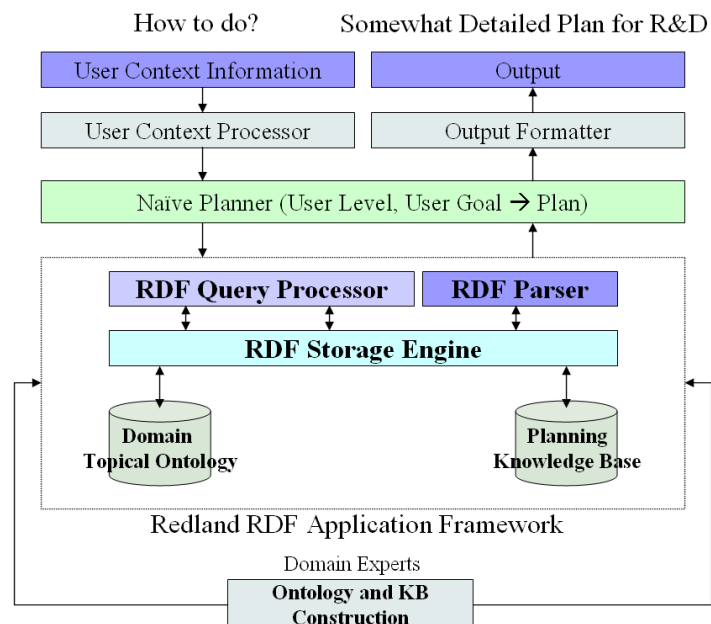
사용자 프로파일 요소	설명	
Current User Level (CUL)	사용자의 현재 지식수준 및 전문성 정도를 표현 특정 분야의 세부 요소분야에 대한 CUL에 대해서 사용자가 모두 지정해야 함 각 세부 요소분야에 대한 수준(level) 지정은 “상”, “중”, “하”로 표현	
	하 (low)	특정 분야에 대한 전문지식이 전무한 상태. 구축된 지식베이스에서 어떤 자원(개체)에 대해서도 연구를 수행한 적이 없음.
	중 (middle)	기본적인 분야지식 보유. 구축된 지식베이스에서 자원난이도(D)로써, “easy”로 구분되는 자원에 대한 지식을 보유하고 있음.
	상 (high)	중급의 분야지식 보유. 구축된 지식베이스에서 자원난이도(D)로써, “easy”, “medium”로 구분되는 자원에 대한 지식을 보유하고 있음.
Ultimate User Goal (UUG)	사용자가 생각하는 특정 분야에서의 최종 목표 내용적으로 자원난이도(D)와 같은 성격을 가지고 있음.	
	skin-deep	자원 난이도(D) 수준 “easy”에 해당하는 자원만 필요
	basic	자원 난이도(D) 수준 “easy”, “medium”에 해당하는 자원 필요
	advanced	자원 난이도(D) 수준 “easy”, “medium”, “difficult” 모두에 해당하는 자원 필요

정보이므로 지식베이스로 쉽게 구축될 수 있다. 둘째, 특정 분야의 포괄적 구조에 대한 이해를 돕기 위해서, 그 분야의 하부 요소 분야에 대한 정보를 제공한다. 마지막으로 시스템은 지식베이스 내의 각 자원을 바탕으로 사용자가 어떤 순서로 연구를 진행할 지에 대한 연구 순서 정보(Research Ordering Sequence Information)

를 제공한다. 이는 특정 분야의 세부 요소 분야의 기지정 순서 정보(pre-ordered categories)를 바탕으로 각 세부분야에 존재하는 자원들을 순차적으로 보여줌으로써 가능하다. 다음의 <표 2>는 본 연구에서 개발되는 시스템의 세 가지 주요 기능과 그에 대한 예시를 나타내고 있다.

<표 2> 실행 모듈의 3 가지 주요 기능

기능	예시
특정 분야에 대한 유용한 정보 제공	- 주요 연구 및 기술 논문 - 대표적인 교과서 혹은 참고서적 - 유용한 웹사이트 - 세부 분야별 전문가 리스트
특정 분야의 하부 요소 분야 제시	- Semantic Web의 하부 기술 분야 = OWL, RDF, RDFS, URI, Unicode, XML, XML Schema 등
사용자 프로필 기반 연구 순서 정보 제공	- URI → XML → XML Schema → RDF



<그림 2> 실행 모듈의 전체 구조

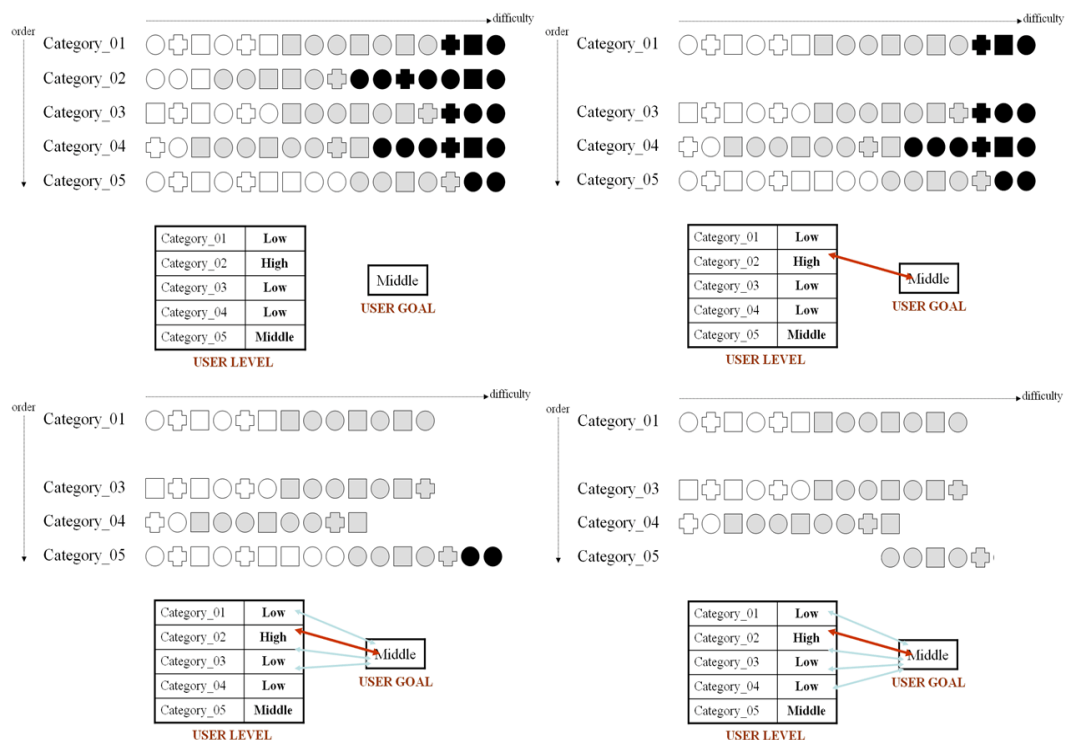
〈그림 2〉는 실행 모듈의 전체적인 구조를 보여준다. 만일 사용자가 자신의 현재 상황에 가장 적합한 정보를 얻기 원한다면, 사용자 프로파일 정보의 입력이 우선되어야 한다. 물론 사용자 프로파일을 입력하지 않은 상태에서도 시스템을 사용할 수 있지만 입력된 사용자 프로파일을 바탕으로 시스템은 필요한 정보를 제공한다.

지식베이스는 생성 및 관리의 편의성을 위하여 RDF 표준 형식으로 구축되므로, 이를 처리하기 위한 기반 엔진이 필요하다. 본 논문에서는 이를 위하여 RedLand RDF 어플리케이션 프레임워크를 채택하였다. RedLand는 매우 강력한 RDF 처리 API를 제공하며, 전체적인 구조가 유연하고 효율적이다. 기본적으로 C언

어로 구현되어 있으며, Perl, Python, Tcl 등과 같은 다양한 개발 환경을 제공한다.

3.2.1 사용자 프로파일 기반 정보제공 알고리즘

구축된 지식베이스와 입력받은 사용자 프로파일 정보를 바탕으로 시스템은 사용자에게 가장 적합한 정보를 제시한다. 기본적인 처리 알고리즘은 비교적 단순하다. 지식베이스에서 사용자 프로파일 정보에 표현된 두 가지 요소 정보 즉, 사용자의 현재 수준 정보와 사용자의 최종 목적 정보를 기반으로 필터링을 하는 개념이다. 다음 〈그림 3〉은 정보 제공 알고리즘을 예제를 통하여 도식적으로 표현하고 있다.



〈그림 3〉 사용자 프로파일 기반 정보 제공 알고리즘

각 세부 그림의 하부에 보이는 상자 안에는 현재 사용자의 프로파일 정보가 존재한다. 이 그림에서는 총 다섯 가지의 하부 요소분야가 존재하고, 각 분야마다 사용자의 현재 전문성 정도가 low, middle, high 중의 한가지로 표현되어 있다. 우측 하단에는 사용자의 최종 목표가 표기되어 있고, 이 두 가지 정보가 현재 사용자의 사용자 프로파일 정보가 된다. 각 세부 그림의 상단에는 세부 분야별로 분류된 지식베이스가 존재한다. 각 분야별로 지식베이스는 난이도에 따라 정렬되어 있다.¹⁰⁾ 상단 우측 그림에서 시스템은 “Category_02”라는 분야에서 사용자의 현재 수준이 “높고(high)”, 사용자의

최종 목표는 “중간(middle)”임을 식별한다. 따라서 “Category_02”에 속하는 모든 자원은 현재 사용자에게는 불필요한 자원이라고 판단한다. 그 이유는 본 논문에서 가정한 사용자의 전문성 수준 판정기준에서 볼 때, 이미 구축된 지식베이스 내에서는 사용자가 참고할 만한 자원이 없기 때문이다. 또한, 좌측 하단의 그림에서 “Category_04”에서의 사용자의 현 수준은 “낮음(low)”이고 사용자의 최종 도달 목표는 “중간(middle)”이므로, 현재 사용자는 난이도 기준 “어려움(difficult)” 수준의 자원은 필요가 없다. 따라서 “Category_04”의 그에 해당하는 자원은 제거된다.

```

Sub-Technologies ← Several sub-technologies related to one Research Area;
User_Level_Set ← Store each user level of all the sub-technologies (Low/Middle/High);
User_Goal ← (Low/Middle/High);
Suggested_Research_Sequence; // Resulting Sequence of Sub-Technologies
Suggested_Resources; // Suggested Resources of each sub-technology
i ← 0;
For each Sub-Technologies[i] {
  if (User_Level_Set[Sub-Technologies[i]].user_level > User_Goal)
    Discard Sub-Technologies[i];
  else {
    Append Sub-Technologies[i] into Research_Sequence;
    j ← 0;
    For each Resources[j] in Sub-Technologies[i] {
      if (Resources[j].difficult_Level <= user_goal
          AND Resources[j].difficult_Level >= User_Level_Set[Sub-Technologies[i]].user_level) {
        Append Resource[j] into Suggested_Resources;
      }
      j++;
    }
  }
  i++;
}

```

〈그림 4〉 사용자 프로파일 기반 정보제공 알고리즘 Pseudo code

10) 실질적으로는 이렇게 정렬되어 있지 않으며, 모든 자원 요소가 RDF의 리소스로 표현되어 있다. RDQL을 이용하여 검색을 수행함으로써 난이도 별 정보접근이 가능하다.

〈그림 4〉는 본 논문에서 구축되는 정보제공 알고리즘의 pseudo code를 보여준다. 이 그림에서 나타난 바와 같이, 이 알고리즘은 매우 단순하고 직관적인 형태를 가지고 있으며, 이는 알고리즘의 근간을 이루는 콘텐츠 혹은 지식베이스가 체계적으로 구축되어 있으면 그만큼 전체 시스템의 효율성도 높아진다는 것을 보여주는 좋은 예이다.

4. 지식베이스 구축 및 서비스 알고리즘 개발

4.1 RDF 기반 지식베이스 구축

서론에서 언급된 바와 같이 본 논문의 주요 목표 중의 하나는 지식베이스 구축 담당자인 분야 전문가가 보다 쉽게 구축할 수 있도록 하는 것이다. 이를 위해서, 현재 자원 표현 방법으로 가장 널리 활용되고 있는 RDF를 채택하였고, RDF 문서의 개별 필드를 정의하고 구조화하기 위한 스키마를 최대한 단순하게 설정하였다. 이 스키마를 작성하기 위해서는 RDFS를 사용하였다.

전 세계적으로, 특정 분야의 온톨로지(ontology) 및 이를 활용한 지식베이스¹¹⁾가 매우 활발히 구축되고 있다. 많은 온톨로지가 시맨틱 웹이나 전통적인 인공지능에서의 추론(inference) 기능을 가정하고 구축되고 있다. 그러나 대부분의 온톨로지 기반 어플리케이션이나 사용자 환경 등이 시간이 많이 걸리는 추론 기능을 사

용하기에는 부적합한 경우가 많다. 따라서 온톨로지 구축 당시의 목적대비 활용실적은 매우 미흡한 실정이다. 따라서 본 논문에서는 지식베이스의 구축방법이나 스키마 등은 최신 기술을 따르되, 우선적으로 사용자가 바로 활용할 수 있는 체계를 채택하였다. 그 이유는 표현력이 풍부하고 추론이 가능한 프레임워크(RDF, RDFS, OWL 등)를 기반으로 지식베이스를 구축하되, 전문가가 쉽게 자신의 핵심 노하우(know-how)를 직접적으로 표현하게 하고 이를 100% 활용하여 정보 서비스가 가능하여야 하기 때문이다.

〈표 3〉은 본 논문에서 구축되는 RDF 인스턴스 집합, 즉 지식베이스를 표현하는데 필요한 요소 항목들이다. 다양한 형태의 연구정보가 존재하지만 일단 본 논문에서는 유형별로 “사람(Person)”, “저작물(Publication)”, “기술(Technology)” 등의 기본적인 자원만을 다룬다. 표에서 나타나듯이, 총 3 개의 유형 클래스로 구성되어 있으며, 각 클래스마다 세부 유형을 나타내는 자식 클래스가 존재한다. 각 클래스는 각 클래스에 속하는 인스턴스를 나타내는 슬롯 정보가 포함되어 있다. 예를 들어, “Publication” 클래스의 자식 클래스인 “Book” 클래스는 책 제목에 해당하는 “pub_Title”, 저자에 해당하는 “pub_Author” 등과 같은 속성 집합을 가지고 있고 이들 속성은 연관 관계($rt_o \in C$) 집합의 원소라고 볼 수 있다. 이와 같이 직접적으로 보이는 연관 관계 정보 외에, “Publication”과 “Paper”는 부모-자식 관계라는 식의 내재된 연관 관계도 역시 존재한다. 슬

11) 온톨로지가 대부분 OWL 형태로 구축되므로 이를 활용한 지식베이스의 대부분도 RDF 기반의 인스턴스(instance) 집합이다.

롯 집합 중에서 “difficulty_Level” 슬롯은 현재 자원의 난이도 정보를 표시할 수 있는 슬롯이다. 지식베이스를 구축하는 분야 전문가가 직접 현재 구축 대상 자원의 난이도를 판단하여 이 슬롯에 값을 표기하게 된다.

〈그림 5〉는 위의 RDFS 스키마를 기반으로 구축된 “시맨틱 웹” 분야 지식베이스의 일부분을 보여준다. 각 자원 항목에 고유의 URI가 부

착되고, 특성 정보들도 속성 형태로 표현되고 있다. RDF에서의 모든 자원은 URI로서 식별되므로 이를 생성시켜주는 메커니즘이 필요하지만 본 논문에서는 일단 다음 그림에서와 같이 Protege에서 제공되는 URI 생성기능을 활용하였다.

이 입력 인터페이스의 가장 큰 단점은 자원끼리 연관관계로 연결할 경우 대상 자원의 URI

〈표 3〉 지식베이스를 표현하기 위한 RDF 스키마 항목

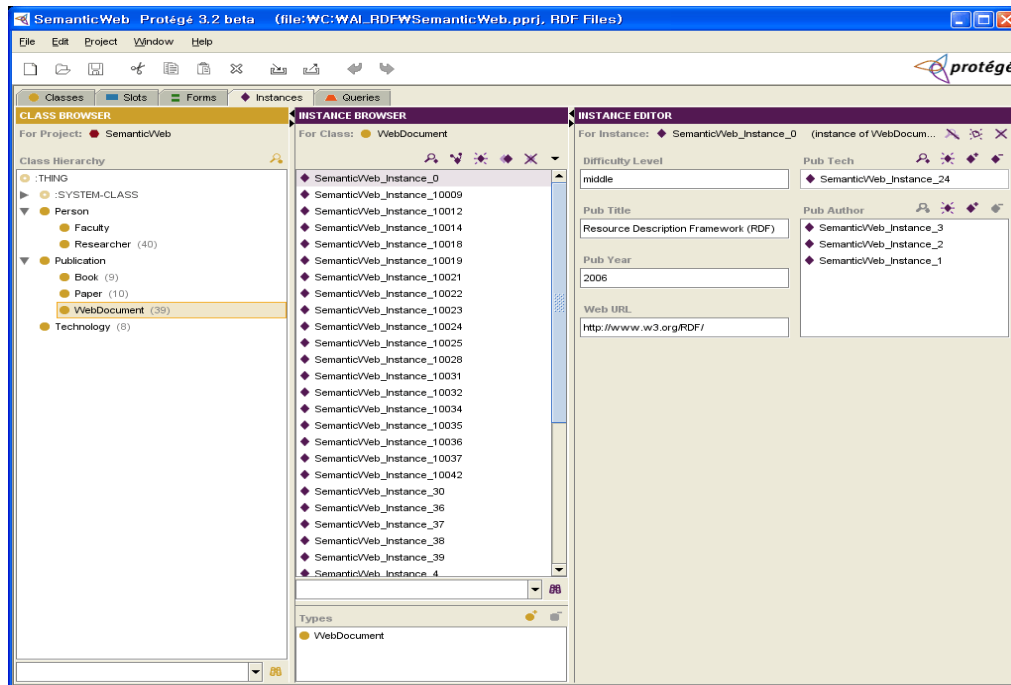
클래스	자식 클래스	슬롯(slots, properties)
Person	Researcher	person_HomePage, person_Name, tech_Area
Publication	Book	pub_Title, pub_Author, pub_Tech, pub_Year, difficulty_Level
	Paper	pub_Title, pub_Author, pub_Tech, pub_Year, difficulty_Level, download_Link
	WebDocument	pub_Title, pub_Author, pub_Tech, pub_Year, difficulty_Level, web_URL
Technology		tech_Name, tech_Order

```

<anos:WebDocument rdf:about="&anos;SemanticWeb_Instance_0"
  anos:difficulty_Level="middle" anos:pub_Title="Resource Description Framework (RDF)"
  anos:pub_Year="2006" anos:web_URL="http://www.w3.org/RDF/" rdfs:label="SemanticWeb_Instance_0">
  <anos:pub_Author rdf:resource="&anos;SemanticWeb_Instance_1"/>
  <anos:pub_Author rdf:resource="&anos;SemanticWeb_Instance_2"/>
  <anos:pub_Tech rdf:resource="&anos;SemanticWeb_Instance_24"/>
  <anos:pub_Author rdf:resource="&anos;SemanticWeb_Instance_3"/>
</anos:WebDocument>
<anos:Researcher rdf:about="&anos;SemanticWeb_Instance_1"
  anos:person_HomePage="http://www.w3.org/People/Ivan/"
  anos:person_Name="Ivan Herman"
  rdfs:label="SemanticWeb_Instance_1">
  <anos:tech_Area rdf:resource="&anos;SemanticWeb_Instance_24"/>
</anos:Researcher>
<anos:Researcher rdf:about="&anos;SemanticWeb_Instance_10008"
  anos:person_HomePage="http://www.dajobe.org/"
  anos:person_Name="Dave Beckett"
  rdfs:label="SemanticWeb_Instance_10008">
  <anos:tech_Area rdf:resource="&anos;SemanticWeb_Instance_24"/>
</anos:Researcher>

```

〈그림 5〉 구축된 지식베이스 예제



〈그림 6〉 Protege를 활용한 RDF 지식베이스 구축과정

를 알고 있어야 한다는 것이다. Protege에서는 이러한 URI 검색 및 접근 기능이 미약하기 때문에 분야 전문가가 이러한 작업을 하기에 어려운 점이 있다. 본 논문의 향후 연구 방향에서는 이런 다양한 기능을 포괄하는 사용자 중심의 지식베이스 구축 인터페이스 개발도 포함되어 있다.

〈표 4〉는 현재까지 구축된 “시맨틱 웹” 분야 RDF 지식베이스의 규모를 보여준다. 총 101개의 자원(resource)으로 구성되어 있으며, 향후 지속적으로 확장될 것이다. 본 논문에서 구축된 지식베이스는 총 2인의 분야 전문가에 의해서 구축되었으며, 연구 목적을 위한 시스템 개발로 한정하였으며, 상호 검증은 수행하지 않았다. 일단 초기 단계이므로 소규모의 지식베

이스를 통해서 시스템을 구축하여 전체 서비스에 대한 유효성 검증이 우선되어야 이 결과를 통한 증분적(incremental) 구축 및 정제가 이루어질 수 있다.

4.2 사용자 프로파일 입력 인터페이스

상기에서 언급된 두 가지의 사용자에 대한 가정을 기반으로 사용자가 직접 자신의 현재 상황을 입력할 수 있는 입력 인터페이스를 구현하였다.

〈그림 7〉은 본 논문에서 개발된 사용자 프로파일 입력 인터페이스를 보여준다. 그림의 상단 부분에는 “시맨틱 웹” 분야의 세부 요소분야마다 사용자의 현재 전문성 정도를 지정할

〈표 4〉 현재까지 RDF 지식베이스 구축 결과

클래스	자식 클래스	인스턴스(건수)
Person	Researcher	34
Publication	Book	10
	Paper	10
	WebDocument	39
Technology		8
Total		101

- Your Current Level of Knowledge about this Topic -

Topics of Semantic Web	User Level		
	Low	Medium	High
CWL	☒	☐	☐
RDF	☒	☐	☐
RDF Schema	☒	☐	☐
Semantic Web based Workflow	☒	☐	☐
URI	☒	☐	☐
Unicode	☒	☐	☐
XML	☒	☐	☐
XML Schema	☒	☐	☐

- Your Goal in this Research Topic -

Technologies of Semantic Web

skin-deep knowledge

- ☒ I want to get just Basic Information!
- ☐ I want to start my Initial Research on this topic!
- ☐ I want to get more Advanced Information about this topic!

Basic knowledge

User Level of Knowledge

Advanced knowledge

〈그림 7〉 사용자 프로필 입력 인터페이스

수 있는 선택버튼이 존재한다. 하단에는 사용자의 최종 목표를 세 가지 중에서 선택할 수 있도록 하였다. 이 인터페이스의 내용은 시스템이 지식베이스에서 자동으로 추출하여 나타낸다. 따라서 다른 분야의 지식베이스를 사용하면 자동적으로 그 분야에 맞는 사용자 프로필 입력 인터페이스가 나타나게 된다.

사용자의 최종 목표에 대해서 상세히 설명하면, “Skin-deep knowledge”는 현재 분야에서의 기초가 되는 개념이나 원리에 대해서만 알기를 원하는 사용자가 선택하는 목표이고, “Basic

knowledge”는 현재 분야를 자신의 전공이나 주요 연구 분야로 삼기를 원하는 사용자가 연구의 출발점이나 방법 등을 찾기 위한 목표라고 볼 수 있다. 최종적으로 “Advanced knowledge”는 현재 분야에서의 전문성 정도가 일정 수준에 오른 사용자가 더 심도 깊은 연구를 수행하기 위해서 선택하는 목표이다.

4.3 서비스 시나리오

위에서 기술한 모든 내용을 바탕으로 본 논

문에서 개발된 시스템은 구축된 지식베이스에 기반한 추천 서비스(suggestion service)를 수행한다. 서비스는 세 단계에 걸친 시나리오(관심 분야 선택, 사용자 프로파일 입력, 추천 정보 제공)를 통해서 이루어진다.

첫 번째 단계는 관심 분야의 선택이다. 현재 최종적으로 구축된 지식베이스는 “시맨틱 웹” 분야 한 개이나, 전체적인 시스템 구조는 복수 개의 지식베이스를 처리할 수 있는 콘텐츠 독립형 아키텍처를 지향하고 있다. 따라서 서비스 시나리오는 사용자가 자신이 원하는 관심 분야를 선택하는 과정으로부터 출발해야 한다. 우선 사용자가 자신이 원하는 전문 분야를 선택하면, 시스템은 해당 지식베이스(RDF 문서)를 분석하여 세부 요소분야 정보를 추출하게 된다.

두 번째 단계는 사용자 프로파일의 입력이다. 사용자는 자신의 현재 상태를 나타내는 사용자 프로파일 정보를 입력한다. 직접 기입하는 내용이 없으므로 직관적으로 쉽게 입력할 수 있다. 사용자 프로파일 정보에는 각 분야별로 자신의 현재 수준과 현재 분야에서의 자신의 최종 목적 정보를 선택하게 된다.

마지막 단계는 선택된 관심분야와 사용자 프로파일에 근거하여 시스템이 추천 정보 제공하는 것이다. 시스템은 입력된 사용자 프로파일 정보를 바탕으로 지식베이스에서 사용자에게 가장 적합한 정보를 제시함과 동시에, 세부 요소분야 중심 연구 순서 정보를 출력한다. 사용자는 제시된 정보를 바탕으로 자신이 살펴봐야 할 자원을 선택하여 상세 정보를 확인할 수 있다.

5. 연구정보 지원시스템 실행

본 절에서는 4.3절에서 설명한 서비스 시나리오에 따라 시스템이 구동되는 실제 모습을 화면 캡처 이미지를 통해서 설명한다. 본 논문에서 개발된 최종 시스템을 ANOS(Adaptive NOvice Supporter)로 명명하였으며, 클라이언트-서버 기반의 웹 응용시스템으로 개발되었다. 우선 아래 그림에서 최상단에 있는 도메인 선택 메뉴를 통해서 관심 분야를 선택하면, 그 분야의 세부 요소분야 정보와 함께 사용자 프로파일 정보 입력 화면이 표시된다.

입력화면을 통해서 모든 필요한 사용자 프로파일 정보를 입력한 후 “Generate Your Own Study Path!” 입력 버튼을 클릭하면 다음 그림과 같이 시스템이 생성한 정보를 보여주게 된다.

〈그림 9〉에서 보는 바와 같이, 화면 상단에는 “Suggested Research Order” 섹션이 출력된다. 이는 앞에서 설명한 추천(필터링) 알고리즘을 바탕으로 시스템이 생성한 세부 요소분야에 대한 연구 순서 정보이다. 이 화면에서는 사용자 프로파일 입력 인터페이스에서 모든 분야에 대한 사용자 전문성 정도를 “낮음”으로 입력하였으므로, “시맨틱 웹” 분야의 모든 세부 요소 분야가 출력되었다. 하단부에서는 특정 세부 요소분야에 속하는 연구정보가 출력된다. 특히 이 그림에서는 사용자의 최종 목표를 가장 낮은 수준으로 지정하였으므로, 자원 유형 “Web Document”에서 난이도(Difficulty)가 낮은(low) 자원만이 붉은 색으로 강조되었다. 결론적으로 시스템은 현재 사용자가 “Uniform Resource Identifier”라는 웹 문서만 살펴보면 충분한 것으로 판단하였다.

ANOS
Adaptive NOvice Supporter

• **Technical Domain:** Semantic Web [select]
 ○ You can select one of these listed research topics...
 ○ After choosing your interesting topic, you can see some basic information about the topic.

- Your Current Level of Knowledge about this Topic -

Topics of Semantic Web	User Level		
	Low	Medium	High
OWL	☐	☐	☐
RDF	☐	☐	☐
RDF Schema	☐	☐	☐
Semantic Web based Workflow	☐	☐	☐
URI	☐	☐	☐
Unicode	☐	☐	☐
XML	☐	☐	☐
XML Schema	☐	☐	☐

- Your Goal in this Research Topic -

- I want to get just Basic Information! ☐
- I want to start my Initial Research on this topic! ☐
- I want to get more Advanced Information about this topic! ☐

〈그림 8〉 ANOS(Adaptive NOvice Supporter) 입력화면

ANOS
Adaptive NOvice Supporter

• **Technical Domain:** Semantic Web [select]
 ○ You can select one of these listed research topics...
 ○ After choosing your interesting topic, you can see some basic information about the topic.

[Suggested Research Order]

URI ► Unicode ► XML ► XML Schema ► RDF ► RDF Schema ► OWL ► Semantic Web based Workflow ► Total Items

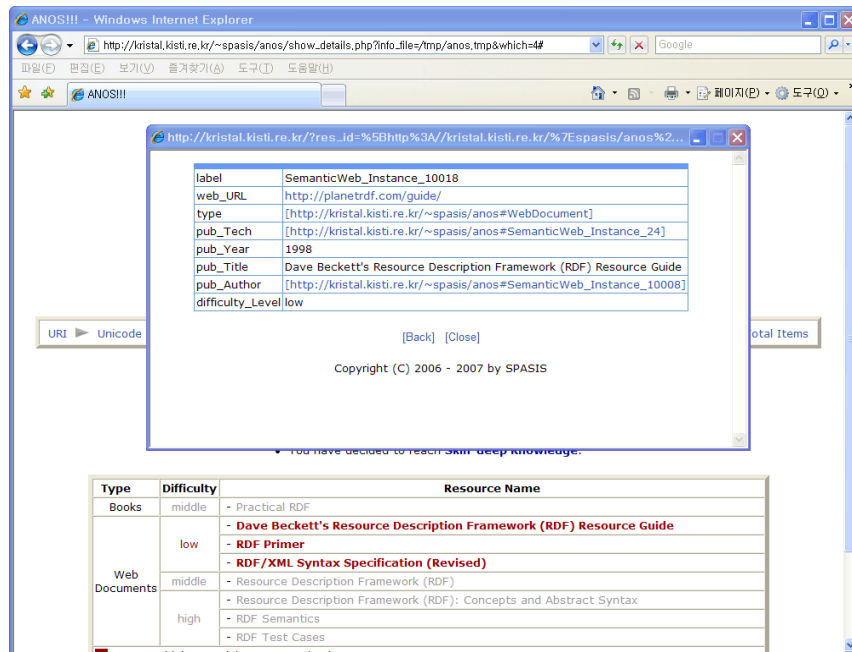
[Research Items of URI]

- Your Knowledge Level of "URI" is **LOW**.
- You have decided to reach **Skin-deep Knowledge**.

Type	Difficulty	Resource Name
Web Documents	low	- Uniform Resource Identifier
	middle	- Naming and Addressing: URIs, URLs, ...
		- Names and addresses
	high	- Uniform Resource Identifier (URI) Schemes
		- ESW Wiki - UriSchemes

■ : Items which you might want to Check Out!
 ■ : Items which you might have already seen before or which are not appropriate to your Goal!

〈그림 9〉 시스템 생성 화면 - 연구순서정보 및 분야별 상세 정보



〈그림 10〉 자원 상세 정보 출력

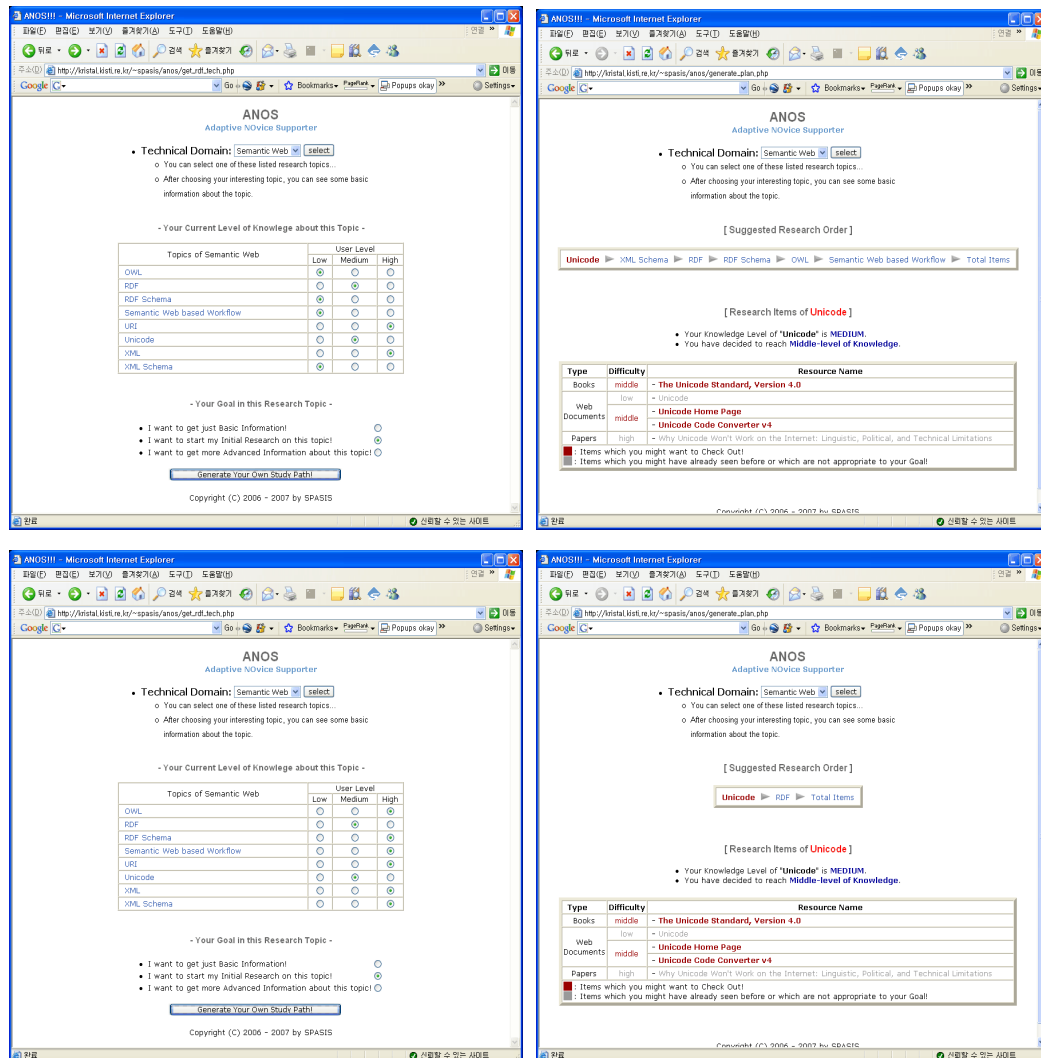
〈그림 10〉은 특정 자원에 대한 상세정보를 나타내는 팝업창을 보여준다. 앞에서 RDF 문서 구성을 위한 RDFS 기반 스키마에서 지정한 모든 정보가 출력되고, 각 필드의 데이터 특성에 따라 링크가 부여되기도 한다. 대상 자원 유형이 웹 문서이므로 사용자는 이 팝업 창을 통해서 직접 해당 URL로 이동할 수 있다.

마지막으로 사용자 프로파일의 재지정 따른 출력화면의 변화에 대해서 살펴보자. 〈그림 11〉은 두 개의 사용자 프로파일 정보가 시스템의 출력화면에 어떤 영향을 미치는지를 보여준다. 두 사용자 모두 자신의 연구 최종 목표를 “중간단계”로 지정하였고, 개별 세부 요소분야에 따른 전문성 정도에 차이를 두고 있다. 상단 그림에서는 연구 순서 정보 섹션에 거의 대부분의 세부 요소분야가 출력된 반면, 하단부에

서는 “Unicode”, “RDF” 등 두 개의 요소분야만이 출력되었다.

6. 한계점과 향후 연구 방향

본 논문에서 개발된 시스템은 기반 지식베이스 구조와 추천 알고리즘의 단순성 측면에서 많은 한계점이 있다. 우선 가장 두드러지는 한계점은 바로 연구 순서정보 출력 메커니즘이다. 직관적이고 실질적인 분야 전문가들의 지식을 최대한 풍부하게 표현하자는 것이 본 논문에서의 지식베이스 구축의 핵심이지만, 이러한 지식을 보다 쉽고 객관적으로 표현하는데 도움을 줄 수 있는 유틸리티가 필요하다. 예를 들어, 요즘 활발하게 이용되고 있는 위키피디아(Wikipedia)와



〈그림 11〉 사용자 프로파일 변경에 따른 출력화면 변화

같은 오픈된 백과사전의 정보들을 수집하여 이를 구축자에게 제시하면 매우 유용할 것이다. 시스템 서비스 측면에서 보면 분야 전문가의 명시적인 표기가 없어도 자동으로 연구 순서 정보를 생성할 수 있는 알고리즘이 필요하다. 이는 개별 세부 요소분야에 포함된 자원정보 간의 의존성 정도를 파악할 수 있는 통계적 혹

은 규칙기반 접근방법으로 가능할 수 있다.

두 번째로 향후 연구방향에서 심도 있게 다루어야 할 내용은 바로 지식베이스 구축자가 표기하는 난이도와 사용자가 명기하는 전문성 정도 수치의 불일치 가능성에 대한 식별 및 해결이다. 본 논문에서는 암묵적으로 이 두 수치가 서로 비교가능하다고 가정하고 있지만, 이에

대한 실질적인 접근 방법이 반드시 필수적이다. 다양한 방법이 제시될 수 있으나, 그 중에서 가장 단순한 방법은 바로 사용자에게 대한 우선 평가(user pre-evaluation)가 있을 수 있다. 다시 말해서, 사용자가 직접 세부 요소분야에 대한 자신의 전문성 정도를 표기하는 것이 아니라, 분야 전문가가 작성한 분야 전문성 척도용 문제 집합을 풀게 하는 것이다. 이를 기반으로 현 사용자의 전문성 정도를 분야 전문가의 관점에서 판단할 수가 있게 된다.

이러한 여러 가지 한계점에도 불구하고 본 논문에서 구축된 지식베이스와 알고리즘이 의미를 가지는 이유는 바로 단순성(simplicity)과 즉시성(immediateness)이라고 볼 수 있다. 직관적인 지식베이스 구조는 분야 전문가로 하여금 자신의 전문지식을 쉽게 표현할 수 있도록 도와주며, 사용자는 연구를 수행하는데 필요한 필수적인 정보를 보다 용이하게 획득할 수가 있다. 또한 RDF와 RDFS 기반의 지식베이스 구축방법과 콘텐츠 독립적 시스템 구조는 확장성과 발전가능성을 내포하고 있다고 할 수 있다.

7. 결 론

본 논문에서는 단순한 형태의 연구 계획 생성 기능(research scheduling function)을 갖춘 적응형 연구지원 포털시스템을 개발하였다. 또한, 기존의 지식베이스 구축방법에서 탈피하여 분야 전문가가 자신의 지식을 효율적으로 표현할 수 있는 스키마 구조와 구축방법론을 제시하였다. 이를 위하여 구축되는 지식베이스에 대한 모델링과 함께, 대상 이용자에 대한 명확한 가정 하에서 적응형 연구정보서비스 시스템인 ANOS(Adaptive NOvice Supporter)를 개발하고 이에 대한 가능성과 한계점에 대해서 제안하였다.

본 논문에서 개발된 시스템의 단순성과 확장가능성으로 향후 다양한 발전적 모델이 적용될 수 있다. 특히 현재 Web 2.0 프레임워크 하에서 중요한 비즈니스 모델로 각광받고 있는 전문가 기반의 집단 지능(Collective Intelligence) 구축, 수집, 그리고 서비스의 핵심 요소 기술로 활용될 수 있다. 또한 웹 상에 존재하는 다양한 형태의 전문 지식에 대한 자동 수집 및 분석 기술이 개발된 시스템에 첨가된다면 지식베이스의 효율적인 구축 및 심층화가 가능할 수 있을 것이다.

참 고 문 헌

Dave Beckett, Dave Beckett's Resource Description Framework(RDF) Resource Guide.
 <www.planetrdf.com/guide/>

Tom Gruber, The Pragmatics of Ontology as Language, Contract, and Content, Bio-Ontologies 2000 Workshop, 2000
 AAAI, Expert Systems.

- 〈www.aaai.org/AITopics/html/expert.html〉
- Wikipedia. Expert System.
〈http://en.wikipedia.org/wiki/Expert_system〉
- John Davies, Rudi Studer, and Paul Warren. Semantic Web Technologies: Trends and Research in Ontology-based Systems. John Wiley & Sons(July 11, 2006)
- Grigori Antoniou, Frank van Harmelen. A Semantic Web Primer(Cooperative Information Systems)(Hardcover). The MIT Press(April 1, 2004)
- Michael C. Daconta, Leo J. Obrst, Kevin Smith. The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management (Paperback). Wiley(May 30, 2003)
- David Beckett. The Design and Implementation of the Redland RDF Application Framework. Tenth International World Wide Web Conference. May 1-5, 2001
- W3C Semantic Web Activity. Resource Description Framework(RDF)/W3C Semantic Web Activity.
〈www.w3.org/RDF/〉
- XML.com. What Is RDF.
〈www.xml.com/pub/a/2001/01/24/rdf.html〉
- Planet RDF. 〈www.planetrdf.com/〉
- Redland RDF Application Framework.
〈<http://librdf.org/>〉
- RDQL - A Query Language for RDF.
〈www.w3.org/Submission/RDQL/〉
- Jena Tutorial - RDQL. jena.sourceforge.net/tutorial/RDQL/
- Russell, Stuart J. & Norvig, Peter(2003). Artificial Intelligence: A Modern Approach(2nd ed.). Upper Saddle River, NJ: Prentice Hall. ISBN 0-13-790395-2