

OntoFrame-K[®]: 협업 연구 지원을 위한 시멘틱 웹 기반 지식정보 공유·유통 플랫폼

한국과학기술정보연구원 성원경·정한민·박동인

1. 서 론

오늘날에는, 국가 간의 기술 경쟁이 더욱 치열해지면서, 국가 경쟁력 확보를 위하여 최단 기간 내에 최고의 연구성과를 얻을 수 있도록 연구·개발자들의 협업 연구를 가상공간상에서 지원하기 위한 국가 차원의 “소프트 인프라”에 대한 필요성이 강조되고 있다. 이러한 필요성에 따라, 각국은 그리드 및 e-Science에 대한 연구를 통하여 연구·개발 자원 및 장비의 공동 활용을 위한 여러 방안들을 마련하고 있다. 그러나 실질적인 연구·개발자들 간의 협업 연구를 위해서는, 장비 및 실험 데이터의 공동 활용 이외에도, 지식정보의 공동 활용, 특히 연구·개발자들이 개인적으로 소장하고 있는 지식정보의 공동 활용 역시 필요하다. 왜냐하면 연구·개발자들이 소장하고 있는 지식정보는 연구·개발자 본인이 직접 만들었거나 또는 손수 중요한 정보로 선별한 것이므로, 개인 소장 지식정보의 효율적인 공유·유통은 연구·개발자들의 연구를 가속시킴으로써 해당 분야를 더욱 빠르게 발전시키고 이는 다시 연구·개발자들의 협업 연구를 더욱 촉진하여 궁극적으로는 국가 과학기술 경쟁력을 향상시킬 수 있기 때문이다.

과학기술 경쟁력 강화를 위해서는 연구·개발 활동의 생산성 향상도 필요하다. 90년대 중반에 실시된 미국 과학재단의 조사에서는, 연구·개발자들은 전체 연구·개발 시간의 30% 이상을 필요한 정보를 수집, 분류 그리고 선별하는 데 사용하고 있으며 이런 상황은 정보자원의 폭증으로 인해 더욱 심화될 것이므로 연구·개발 활동의 생산성 향상을 위해서는 무엇보다도 정보획득 과정의 개선이 필요함이 지적되었다. 이에 대한 해결 방안으로서, W3C나 ISO와 같은 국제 표준화 기관들과 선진 연구기관들에서는 정보자원을 내용 또는 의미를 기준으로 관리·유통하고 기존 기술의 한계를 극복하기 위한 대안으로서 시멘틱 웹(semantic web) 기술과 온톨로지(ontology)에 대한 연구를 90년대 후반부터 수행하여 오고 있으며, 최근에는 그간의 성과의 실

용화를 위한 다양한 응용 연구도 활발히 수행 중이다.

이상과 같은 기술환경 변화에 적극 대응하기 위하여, 한국과학기술정보연구원(KISTI)은 2004년부터 “지식 기반 정보유통 플랫폼 개발 연구” 사업을 수행하고 있다. “지식 기반 정보유통 플랫폼(OntoFrame-K[®](이하 “OFK”라 약칭함.))”은 정보 공유·유통 기술과 시멘틱 웹 기술이 융합·적용된 새로운 개념의 정보유통 플랫폼으로서, 이미 검증된 전문성과 높은 부가가치를 지닌 개인 소장 과학기술 지식정보를 공유·유통시킬 수 있는 “자발적 가상 협업 연구 커뮤니티(self-organizing virtual research community)”의 구현을 지원하기 위한 시스템이다. 시멘틱 웹 기술은 OFK의 유통 대상 정보인 지식정보의 모델링에서뿐만 아니라 다양한 추론 기반 응용 서비스들에도 적용되어 데이터들 간의 연관 관계를 보다 효율적으로 관리·제시함으로써 OFK에 기반한 협업 연구 커뮤니티를 통한 정보획득 과정의 효율화에 활용되고 있다. 특히 OFK에서 시멘틱 웹 기반 서비스의 핵심 기반 지식자원으로 활용되는 온톨로지 와 시소러스는 각각 언어 독립적인 지식과 언어 종속적 지식의 모델링과 추론 과정에 활용되고 있으며, 2005년도 현재까지, 전체 과학기술 분야를 대상으로 한 “국가 과학기술 R&D 기반정보 온톨로지(National Science & Technology R&D Reference Ontology)”와 “범용 과학기술 분야 시소러스”가 구축되었다. 2006년도부터는 IT, BT, NT 및 기초 과학기술 분야 등과 같은 과학기술 각 분야를 위한 분야별 기반 지식자원들을 지속적으로 구축·적용함으로써, 개별 분야 지식정보 처리를 위한 OFK 전문성을 확장해 나갈 예정이다.

한국과학기술정보연구원의 “지식 기반 정보유통 플랫폼 개발 연구” 사업은 2010년까지 수행될 예정이며, 2008년까지의 1단계 사업과 2010년까지의 2단계 사업으로 나뉘어 추진되고 있다. 1단계는 개별 분야 가상 협업 연구 커뮤니티들의 구현 및 이를 위한 기반 지식 자원, 시멘틱 웹 기반 기술, 그리고 이에 기초한 응용 서비스들의 개발을 주요 목표로 하며, 2단계는 시멘틱 웹

기술과 그리드 기술이 접목된 시맨틱 그리드(semantic grid) 기술을 개발하고 이를 기반으로 개별 가상 협업 연구 커뮤니티들을 연계하여 국가적 차원의 지식정보 공동 활용 체제를 마련하는 것을 목표로 한다.

이상과 같이 OFK에 대한 연구·개발은 앞으로도 계속 진행될 예정이나, 본 고에서는 2005년도까지 구현된 사항들을 중심으로 하여 현재까지의 OFK에 대하여 소개하고자 한다.

2. 관련 연구

OFK는 과학기술 분야에 종사하는 연구·개발자들 간의 자발적인 협업 연구 지원을 목표로 정보유통 기술과 시맨틱 웹 기술의 결합에 기초하여 개발되고 있다. 이에 본 장에서는, 협업 연구 지원을 위해서, 기존의 정보유통 기술 관련 연구들과 시맨틱 웹 기술 관련 연구들에서 개선되어야 하는 사항들을 살펴보고 OFK에서는 현재 어떠한 해결 방안이 제시되고 있으며 향후 OFK가 추구해야 할 방향을 가늠해 보고자 한다.

협업 연구 지원을 위해서는 유통되고 있는 지식정보의 식별(identification)과 유통 과정 중 발생될 수 있는 변경 및 수정에 따른 정보의 이력(version 또는 history) 관리가 필요하다. 이와 관련된 정보유통 관련 연구들을 살펴보면, P2P(Peer-To-Peer)를 포함한 많은 연구들이 식별자(identifier) 검사 등과 같은 방법을 사용한 정보에 대한 검증에 관심을 가지고 있음을 알 수 있다. 예를 들어, [2]와 [5]의 연구에서는 식별자를 미디어 문서 파일의 헤더에 부착하는 방식을 사용하여, 동일 문서는 동일 식별자를 가질 수 있도록 하는 방안을 제시하였다. 그러나 [2]와 [5]에는 사용자가 의도적으로 문서의 헤더 정보를 바꾸거나 또는 특정 문서에 대한 이력을 추적하고자 하는 경우에 대응하기 위한 적절한 해결 방안이 제시되지 않았다. 문서 파일에 추가된 부가정보(또는 메타데이터)의 신뢰성 또한 연구자들 간의 협업 연구 지원에 매우 중요하다. 부가정보는, 정보유통의 신뢰도 확보를 위하여, 등록된 문서 파일에 대한 검증 및 특정 문서 파일의 이력 추적에 활용될 수 있기 때문에 부가정보의 신뢰도는 결과적으로 정보유통 과정 자체의 신뢰성을 좌우할 수 있기 때문이다. 그러나 앞의 연구들뿐만 아니라 대부분의 관련 기존 연구들에서는 부가정보 생성 과정에 대한 효율적인 통제 방안이 제시되지 않음으로 인해 생성된 부가정보의 신뢰성 및 체계성을 확보하기 어려울 뿐만 아니라 그로 인해 이전 버전 문서 파일들 간의 올바른 연결 관계를 제시할 수 없다는 문제점이 존재한다[3,4,11].

이러한 문제들을 해결하기 위해서, OFK에는 “문서

컨테이너(document container)”라는 개념이 도입되었다. 문서 컨테이너는 문서 파일 자체에 관련 메타데이터와 관리 정보들이 인코딩되어 추가·통합된 파일 형태로, 유통을 위해 문서 파일을 등록할 경우, 자동적으로 해당 메타데이터 역시 동시에 등록되게 함으로써 온톨로지 기반의 추론 과정을 효율적으로 지원할 수 있도록 한다. 또한, 문서 컨테이너는 자체적으로 이력을 추적할 수 있는 관련 문서 정보와 메타데이터를 내포하고 있으므로, 문서 컨테이너가 등록될 때 문서 파일 자체에 대한 자동적인 검증이 가능하며, 사용자가 원하는 정보를 내려 받는 경우에도 문서 컨테이너 형식으로 내려 받은 후 OFK의 클라이언트로 문서 컨테이너에 포함된 메타데이터 및 문서 파일을 볼 수 있게 함으로써 신뢰도가 높은 지식정보 유통이 가능하도록 한다¹⁾.

시맨틱 웹과 관련된 기존 연구들로는, 영국 EPSRC(Engineering & Physical Science Research Council)에서 주관하는 AKT(Advanced Knowledge Technologies) 사업에서 개발된 CAS(CS AKTive Space), 독일 Karlsruhe 대학의 SEAL(SEmantic portAL)과 같은 비교적 성공적인 시맨틱 웹 응용 시스템들이 해외에서 발표된 바가 있다. 반면 국내의 경우, 아직까지는 실세계(real world) 정보를 대상으로 시맨틱 웹 기술을 적용한 사례는 발견하기 어려우며, 대개는 가상 정보나 실험실 규모의 간단한 정보를 다루는 수준에 머물고 있는 실정이다. 국내외의 시맨틱 웹 기반 서비스들의 공통적인 문제점으로는 주로 정적인 지식 서비스를 제공하기 위해 설계되고 구현되어 계속적으로 추가되는 정보를 반영하기 위한 동적 지식화 과정을 가지고 있지 않다는 것을 지적할 수 있다[7,9,10,12]. OFK에서는 등록된 지식정보에 해당하는 정보를 DB-to-OWL 변환²⁾을 통하여 지식정보의 등록 시점에 맞추어 새로운 시맨틱 웹 정보를 즉각적으로 반영할 수 있도록 하였으며 변환 결과 생성된 데이터 인스턴스들은 지식의 일관성을 유지하고 일괄 처리가 가능하도록 “국가 과학기술 R&D 기반 정보 온톨로지”에도 삼입·추가되도록 하였다.

끝으로, 다음 그림 1은 정보유통 기술과 시맨틱 웹 기술 관련 논문 맵으로서 양 기술 분야의 연구동향을 전체적으로 조망·분석하기 위한 것이다³⁾.

1) 문서 컨테이너에 포함되는 부가정보의 생성 과정에 대한 통제는 3.3에서 소개한다.

2) DB-to-OWL 변환에 대해서는 3.3에서 소개한다.

3) 그림 1의 논문 맵은 2000년부터 2004년까지의 정보유통 기술 관련 국내 논문 23편과 시맨틱 웹 기술 관련 국내 논문 52편을, 해외 논문에 대해서는 정보유통 기술 관련 논문 51편과 시맨틱 웹 기술 관련 논문 154편을 분석한 결과이다.

(출처: <http://ndsl.or.kr/eng/newindex.html>)

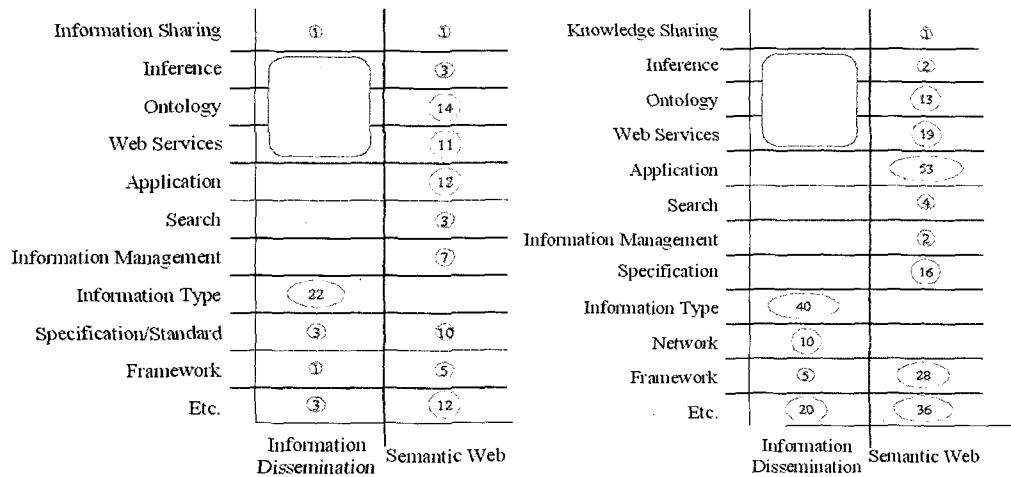


그림 1 정보유통과 시멘틱 웹 관련 논문 맵(좌측: 국내 논문, 우측: 해외 논문)

그림 1에 의하면, 정보유통의 경우 추론, 온톨로지, 웹 서비스에 관한 연구가 전무하며, 시멘틱 웹 또한 정보유통, 정보공유에 관한 연구가 빈약한 실정이다. 대다수의 정보유통 관련 연구들은 정보의 형식과 명세에만 초점을 맞추고 있어서 신뢰성 있는 정보유통 방안을 제시하지 못하고 있으며, 시멘틱 웹 관련 연구들은 응용 분야와 프레임워크에 많은 관심을 가지고 있으나, 체계화된 정보의 확산 방안 마련을 보완할 필요가 있는 것으로 분석될 수 있다. OFK는 이러한 양자 간의 독립적 연구 방향을 통합하기 위해 시멘틱 웹 기술을 이용하여 정보를 체계화하고 연구·개발자들 사이에서 자발적으로 유통될 수 있도록 하기 위한 것이다. OFK의 이러한 시도는 의미를 가진 정보의 확산을 위한 기반을 마련하기 위함이다.

3. OFK의 구성

3.1 정보유통 서비스와 지식 서비스

연구·개발자들 간의 자발적인 협업 지원을 위한 정보유통 플랫폼으로서의 OFK는 중앙 지식서버를 통해 신뢰성 있는 정보만을 교환할 수 있도록 클라이언트-서버 모델로 구성된다. 각 연구·개발자인 클라이언트는 중앙 지식서버를 통해 지식정보를 등록하기도 하고 다른 연구·개발자가 등록한 지식정보를 내려 받기도 할 수 있다. 중앙 지식서버는 등록된 지식정보를 검증하고 해당 정보의 이력을 추적하며, 온톨로지 기반의 추론 서비스를 통해 연구·개발자들에게 국가 과학기술 R&D 기반정보를 포함하는 지식 서비스를 제공한다. 지식 서비스는 온톨로지를 사용하는 지식화와 추론 서비스를 포함하며 시멘틱 웹 기반 프레임워크상에서 동작한다. 현재 제공하고 있는 추론 서비스는 “연구자 네

트워크(Communities of Practice: CoP)”, “연구자 추적”, “연구 맵”으로 구성된다.

하나의 OFK는 특정한 연구 주제나 학문 분야를 가지는 하나의 가상 협업 연구 커뮤니티에 해당하며, 중앙 지식서버를 통해 다른 커뮤니티와 연결될 수 있다⁴⁾.

3.2 정보유통 서비스

OFK의 정보유통 서비스는 그림 2에서와 같이 클라이언트-서버 모델로 구성된다. 개별 연구·개발자들이 생성한 정보의 등록 및 교환은 중앙 지식서버를 통해 이루어지며, 연구·개발자들에 의해 갱신된 정보 역시 중앙 지식서버에 다시 등록된다. 이러한 과정의 순환적인 흐름은 연구·개발자들 간의 자발적인 정보유통 서비스를 보장하기 위한 것이다.

다음에 제시된 그림 3은 OFK의 정보유통 서비스의 흐름을 예시한 것으로서, 서비스 흐름의 각 단계에 대한 설명은 다음과 같다.

- ① 문서 생성(Document Creation): 연구·개발자들은 공유하고자 하는 논문, 과제, 지적재산권과 같은 성과정보에 해당하는 문서들을 생성한다.
- ② 문서 등록(Document Upload): 연구·개발자들은 생성한 문서나 문서 컨테이너(⑤ 참조)를 중앙 지식서버에 등록한다. 등록 과정에서 생성일, 파일명, 파일 타입, 생성자(creator)나 성과정보와 같이 미리 정의된 메타데이터들이 추가된다. 성과정보 관련 메타데이터들은 추론을 위해 자동적으로 지식화 과정을 거친다(3.3 참조).
- ③ 문서 URI 할당(URI Assignment for Document): URI 서버는 문서 식별자로서 등록된 각 문서 객

4) 다중 커뮤니티 모델은 아직 구현되지 않았으나, 2008년 이후에 시멘틱 웹 서비스와 시멘틱 그리드를 이용하여 구현할 예정이다.

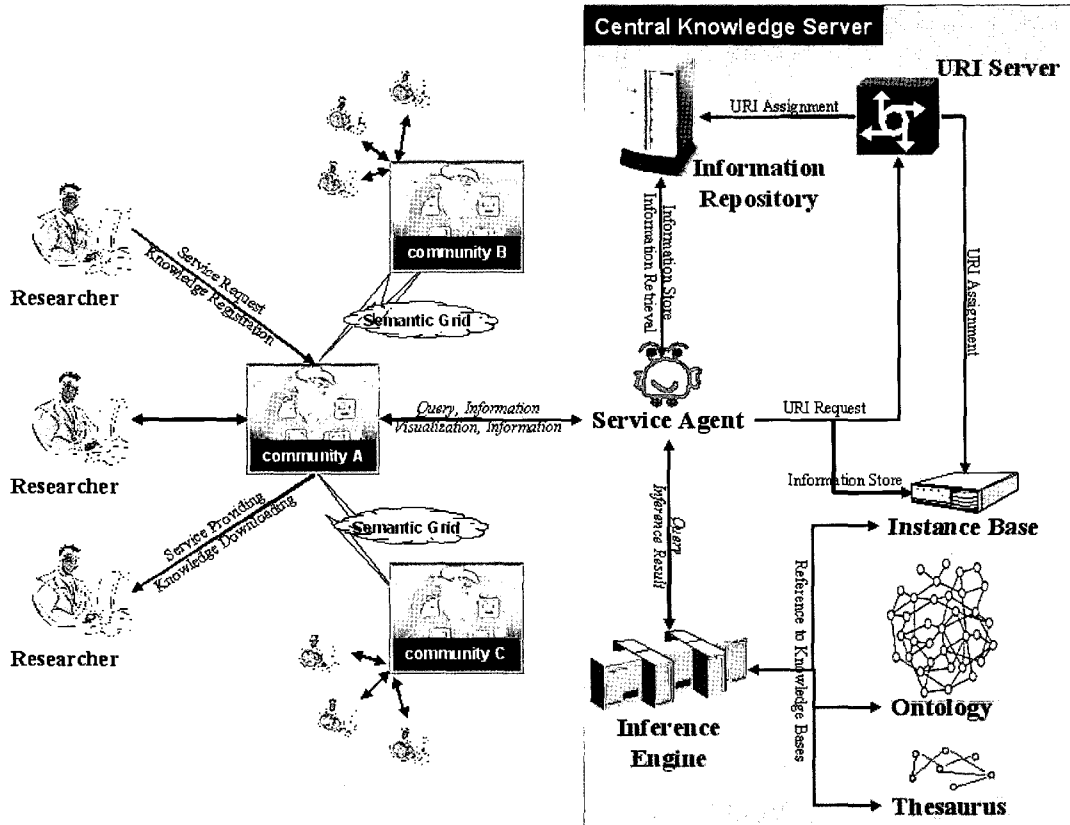


그림 2 정보유통 플랫폼 상에서의 정보유통 서비스와 지식 서비스의 연관도(커뮤니티 A는 가상 협업 연구 커뮤니티이며, 다른 커뮤니티들과 중앙 지식서버를 통해 연결될 수 있다.)

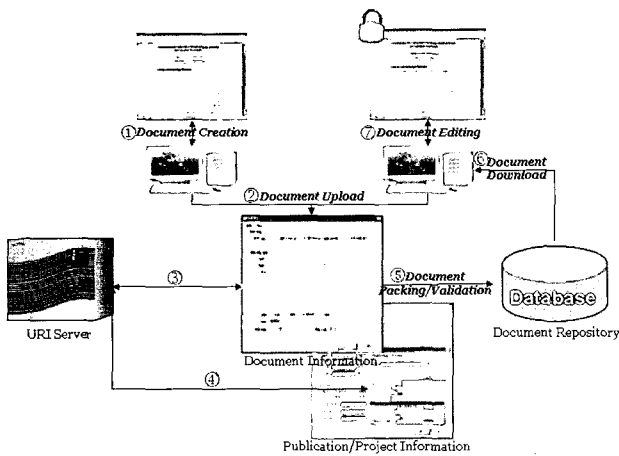


그림 3 정보유통 서비스 흐름도

체에 URI를 할당한다. 중앙 지식서버는 식별자를 이용하여 해당 문서 파일의 이력을 추적한다.

- ④ 메타데이터 URI 할당(URI Assignment for Metadata): 온톨로지에 클래스로 정의된 필드들에 대해 URI들을 할당한다. 이 과정은 3.3의 ①과 같다.
- ⑤ 문서 패키징/검증(Document Packing/Validation): 앞에서 기술한 바와 같이 기존 P2P 시스템들은

문서 파일에 대한 검증 방식의 약점들로 인해 정보유통 과정의 신뢰성 확보에 한계를 가진다. 연구·개발자들 간의 원활한 협력이 이루어지기 위해서, 정보유통 서비스는 검증되지 않은 정보를 차단하고 이력 추적과 관리가 가능해야 한다. OFK에서는 이러한 요구사항들을 만족시키기 위해 전술한 바와 같은 문서 컨테이너 개념을 도입하였다. 공유 대상으로 등록된 모든 문서들은 검증과 패키징 단계를 거치며 문서 패키징의 결과로서 생성되는 문서 컨테이너는 등록자가 설정해 놓은 접근 권한에 따라 다른 연구·개발자들의 접근이 허용된다⁵⁾. 연구·개발자가 클라이언트에서 문서 컨테이너에 대한 편집을 완료한 후에 문서 컨테이너는 자동적으로 다시 패키징이 되며, 문서 컨테이너가 등록되는 경우에 문서 저장소(document repository) 내의 관련 정보들과 비교하는 방식으로 해당 문서 컨테이너를 검증한다.

- ⑥ 문서 다운로드(Document Download): 연구·개

5) 현재 OFK의 정보유통 서비스는 다음의 세 가지 접근 권한의 설정을 허용한다: Read-Only, Read-Write (metadata + text), Read-Write (only metadata)

발자들은 정보검색이나 추론의 결과로 얻어진 정보들을 내려 받을 수 있다. 선택된 정보는 문서 컨테이너 형식으로 제공된다.

⑦ 문서 편집(Document Editing): 문서 컨테이너는 내려 받은 연구자에게 허용된 권한에 따라 편집과 관련된 메뉴들을 활성화시킨다.

3.3 지식화

지식화는 메타데이터 형태의 국가 과학기술 R&D 기반정보를 온톨로지 참조를 통해 재구성하는 작업으로 정의할 수 있다. 추론 서비스는 이러한 지식화의 결과로서 생성되는 RDF triple 데이터들을 사용한다.

- ① 메타데이터 URI 할당(URI Assignment for Metadata): 3.2의 ④와 같다.
- ② 메타데이터 저장(Metadata Storing): 등록된 메타데이터(URI 또는 literal)는 RDBMS 기반의 메타데이터 저장소에 저장된다.
- ③ DB-to-OWL 변환(DB-to-OWL Conversion): 이 변환은 RDF 인스턴스 정보를 담고 있는 OWL 문서 생성을 위한 것으로 서비스 시나리오에 의존적인 부분이다. 추가된 정보를 즉각적으로 반영하기 위해 필요한 부분으로 온톨로지 로딩 시간을 단축하기 위해 추가된 인스턴스들만을 RDF triple store에 반영한다.

표 1 한국과학기술정보연구원의 성과정보로부터 얻어진 국가 과학기술 R&D 기반정보의 예

Subject	Predicate	Object
nlp:신기술아이디어_사업화타당성_평가사업_20036)	nlp:hasProjectMember	nlp:서성호
nlp:Electrospinning_Nanofiber_양산기술평가_및_시장분석	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	nlp:SuperintendentProject
nlp:Electrospinning_Nanofiber_양산기술평가_및_시장분석	nlp:hasBeginningDate	"2003-06-10"^^http://www.w3.org/2001/XMLSchema#string
genid:ARP13440017)	http://www.w3.org/1999/02/22-rdf-syntax-ns#first	nlp:Workshop

6) "nlp:"는 "http://www.kisti.re.kr/OntoFrameK/ReferenceInformation.owl#"의 prefix이다.
7) Jena에 의해 생성된 Bnode (blank node)

- ④ RDF 파싱(RDF Parsing): Jena⁸⁾의 RDF/XML 파서(ARP: Another RDF Parser)는 기반정보용 OWL 문서와 인스턴스용 OWL 문서를 파싱하여 영속 모델(Persistent Model)을 생성한다.
- ⑤ 영속 모델 생성(Persistent Model Creation): OWL 기반의 온톨로지는 표 1에서와 같은 RDF triple 데이터들로 변환되며 빠른 추론을 위해 Jena가 저장소로 사용하는 RDBMS에 저장된다.
- ⑥ 인스턴스 추가(Instance Appending to OWL-based Ontology): DB-to-OWL 변환으로부터 얻어진 인스턴스들은 국가 과학기술 R&D 기반정보 온톨로지에 추가된다. 이는 시스템이 초기화되는 경우나 최적화 작업을 수행하는 등의 배치 처리가 필요한 경우를 대비하기 위해서이다.

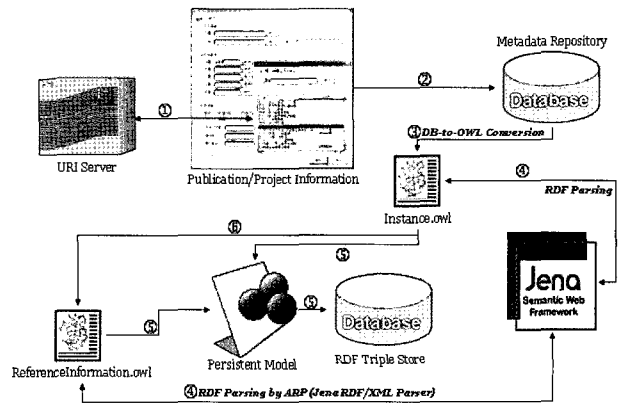


그림 4 지식화 흐름도

3.4 추론 서비스

정보유통 서비스는 직관적이고 명확한 정보를 제공하는 반면에 추론은 fact, 규칙 및 후처리⁹⁾ 등을 통해 명시적으로 드러나 있지 않은 간접적인 정보를 발굴하고 제공한다.

- ① RDQL 생성(RDQL Creation): OFK에서는 연구자 네트워크 추적, 연구자 연구동향 추적, 연구맵 추적의 세 가지 추론 서비스를 제공한다. 연구자 네트워크 추적 서비스는 특정한 연구 주제와 관련된 연구·개발자들의 네트워크를 찾아내고 공저자 관계에 대한 분석 및 발표 논문 수 또는 수행한 사업의 규모를 기준으로 하여 연구자 네트워크에 포함된 연구·개발자들을 순위화하여 제시한다. 연구자 연구동향 추적과 연구맵 추적은 특정 개인으로서의 연구·개발자나 특정 기관의 연구 동

8) <http://jena.sourceforge.net/>

9) Communities of Practice의 전문가 순위 결정 모듈 등이 여기에 해당한다.

향 정보를 추론을 통하여 얻고 연구 주제별로 관련 통계 정보와 함께 제시한다. OFK는 이와 같은 추론 서비스를 위해 미리 정의된 RDQL(RDF Data Query Language) 템플릿을 사용한다. 연구·개발자가 입력 인자들(연구 주제를 위한 literal, 연구자 URI 등)을 입력하면 입력된 내용들은 미리 정의된 템플릿에 삽입된다. 다음은 연구 주제로서 “시스템”이 선택된 경우에 과제명(?project), 참여 연구원(?member), 그리고 과제 금액(?fund)을 요청하는 RDQL의 예이다.

```
SELECT ?project, ?member, ?fund
WHERE (?project rdf:type nlp:Project),
      (nlp:시스템 nlp:isResearchTopicOf
       ?project),
      (?project nlp:hasProjectMember
       ?member), (?project nlp:hasFund ?fund)
```

② 시소러스 탑재(Thesaurus Loading): OFK는 사용자 질의와 등록된 지식 간의 어휘 불일치를 해소하기 위해 시소러스를 사용한다[8]. 현재는 연구 주제에 한해 적용하고 있는데, 사용자가 선택한 특정 연구 주제가 시소러스의 개념어와 매칭되는 경우에 시스템은 자동적으로 매칭된 개념어의 하위 개념어들로 확장한다. 추론 서비스는 이러한 질의 확장을 이용하여 사용자의 질의에 대한 재현율을 높인다. 다음은 “시스템” 개념어에 대한 하위 개념어들의 예를 보여준다.

- ```
시스템
├─ 정보 시스템
│ ├── 지리 정보 시스템
│ ├── 종합 정보 시스템
│ └─ ...
├─ 관리 시스템
│ ├── 지식 관리 시스템
│ ├── 콘텐츠 관리 시스템
│ └─ ...
└─ ...
```

③ 추론(Inference): Jena는 RDF triple 데이터들과 추론 규칙들을 이용하여 추론의 결과로서 인스턴스들이나 triple들을 생성한다. OFK에서는 “rdfs:subClassOf” property를 위한 5개, 최상위 루트 클래스들(“Class: IntellectualProperty”, “Class: Project”, “Class: ResearchTopic”)을 위한 5개의 규칙들을 생성하였다. 다음은 연구 주제 확장을 위한 규칙의 예를 보여준다.

```
(?x nlp:subTopicOf ?y) (?y nlp:subTopicOf ?z) →
(?x nlp:subTopicOf ?z)
```

- ④ 추론 결과 생성(Inference Result Generation): 이것은 추론 결과 시각화의 입력 데이터 생성을 위한 서비스 의존적인 후처리를 수행하는 과정으로 전문가 순위화 모듈이 여기에 해당한다. 또한, 연도별 성과 정보 수와 같은 통계 정보도 생성한다.
- ⑤ 추론 결과 시각화(Inference Result Visualization): 연구·개발자들에게 정보를 효율적으로 제공하기 위해 플래시 기반의 시각화 모듈을 사용하여 추론 서비스를 제공한다(부록 B 참조).

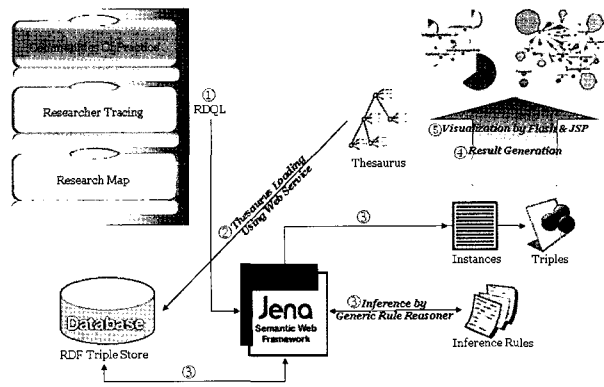


그림 5 추론 서비스 흐름도

### 3.5 구현 현황

한국과학기술정보연구원의 “지식 기반 정보유통 플랫폼 개발 연구” 사업은 시멘틱 웹 기반 지식정보 공유·유통 플랫폼인 OFK의 개발과 OFK의 구동에 필요한 기반 지식자원인 시소러스, 온톨로지 및 데이터 인스턴스들의 구축을 목표로 한다.

OFK가 지원하는 정보유통 서비스(부록 A)는 웹 기반 인터페이스를 가진다. 웹 상에서 연구·개발자는 중앙 지식서버에 접근하여 문서나 문서 컨테이너를 등록하고(부록 A의 (f)), 공유된 문서 검색(부록 A의 (b))을 통해 원하는 정보를 문서 컨테이너 형식으로 내려받을 수 있다.

기반 지식자원을 이용하여 추론을 통해 제공되는 추론 서비스(부록 B)는 연구자 네트워크 추적(부록B의 (b), (c), (d)), 연구자 연구동향 추적(부록B의 (e)), 연구 맵 추적(부록B의 (f))으로 구성된다. 2005년에는 한국과학기술정보연구원의 내부 성과정보 데이터베이스를 서비스에 활용하였으며, 2006년부터는 외부 성과정보를 추가하고 한국과학기술정보연구원의 KOI(Knowledge Object Identifier) 등과 같은 공인된 외부 식별자를 URI로서 활용할 계획이다[1].

OFK의 국가 과학기술 R&D 기반정보 온톨로지는 Event, Group, Intellectual Property, Person, Project, Publication, 그리고 Research Topic의 7개 최상위 클래스들로 구성된다. 전체 클래스의 수는 74개이며, 75개의 Property들과 someValuesFrom, allValuesFrom, 그리고 minCardinality를 포함하는 제약들을 포함하고 있다. OFK의 운영을 위해서, 2005년도에는, 앞서 언급한 바와 같이, 한국과학기술정보연구원으로부터 수집한 2,300여 건의 내부 성과정보로부터 인스턴스들을 구축·활용하였다. 이들 인스턴스 데이터들을 기초로 하여, Jena를 통해 최종적으로 37,656 RDF triple 데이터들을 생성하였다.

OFK의 또 다른 핵심 기반 지식자원으로 2005년도에 구축된 범용 과학기술 시소러스는 약 15,000여 개념어들과 50,000여 개의 개념/관계 facet 정보를 포함하고 있으며 현재 OFK에서는 질의 확장에 사용되고 있다[6,8]. 범용 과학기술 시소러스에 포함된 개념어들은, 한국과학기술정보연구원에서 제안한 “용어 생명주기(life-cycle)” 개념 및 그 측정 기법을 활용하여 약 3,000만 어절의 과학기술 분야 말뭉치로부터 추출된 약 50,000여 용어들로부터 선정된 것들이다[6].

#### 4. 결 론

지금까지 본 고에서는 2004년에 시작하여 2010년까지 수행될 예정인 한국과학기술정보연구원의 “지식 기반 정보유통 플랫폼 개발 연구” 사업에서 개발된 OFK(OntoFrame-K<sup>®</sup>)에 대하여 2005년도까지 구현된 사항을 중심으로 소개하였다.

OFK는 개인이 소장하고 있는 과학기술 지식정보를 자발적으로 공유·활용할 수 있게 함으로써 과학기술 분야의 산·학·연 연구·개발자들 간의 실질적인 협업 연구를 지원하기 위한 소프트 인프라로서의 “자발적 가상 협업 연구 커뮤니티(self-organizing virtual research community)”의 구현을 목표로 한다. 커뮤니티를 통한 지식정보의 공유·유통 및 관리의 효율화를 위해서, OFK는 정보 공유·유통 기술과 시멘틱 웹 기술의 융합에 기초하여 구현되고 있으며, 특히 OFK의 핵심 지식자원인 온톨로지와 시소러스는 범용 과학기술 분야를 필두로 하여 개별 과학기술 분야를 대상으로 연차적으로 구축·적용됨으로써 전문 연구·개발자들로 구성된 개별 전문 분야 커뮤니티에서 유통되는 지식정보 처리의 전문성 확보를 위해서 활용되고 있다.

OFK가 추구하는 자발적 가상 협업 연구 커뮤니티는 그림 6에 제시된 Gartner의 예측 자료에 의하면 향후 2~5년 이내에 상용화가 가능할 것으로 예상되고 있는

corporate semantic web의 전형적인 예이다[10].

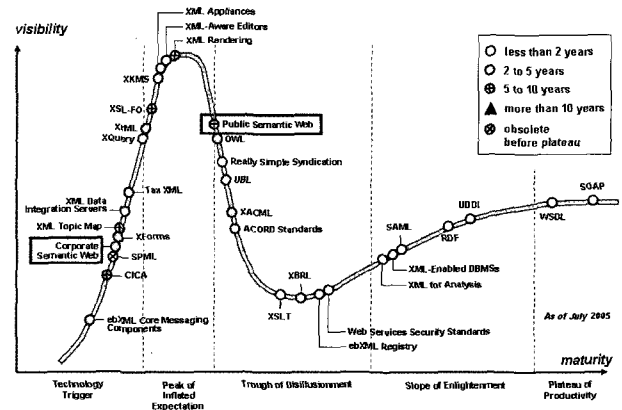


그림 6 Hype Cycle for XML Technologies 2005 (Gartner, 2005)

그러나 Gartner의 예측대로 2~5년 이내에 corporate semantic web이 성공적으로 적용되기 위해서는 무엇보다도 대용량 데이터에 대한 처리 및 추론이 가능하여야 할 것이다. 사실 현재까지 수행되고 있는 대다수의 시멘틱 웹 기술 관련 연구들은 실험실 수준의 소규모 데이터들만을 대상으로 하고 있으나 corporate semantic web의 실질적 활용이 가능하기 위해서는 현재의 EIP(Enterprise Information Portal)에서 처리하고 있는 정도의 양에 해당하는 데이터 처리가 가능해야 할 것이다. OFK는 현재 한국과학기술정보연구원의 연구성과 데이터베이스 전체를 활용하고 있으므로 중규모 이상의 데이터 처리가 가능한 상태다. 그러나 연구자들의 협업 연구를 실질적으로 지원하기 위하여 대용량 데이터 처리 기술을 개발 중에 있다.

시멘틱 웹 기술의 성공을 위해서는 목적하는 응용 서비스가 명확히 정의될 필요 또한 있다. 이는 어떤 응용 또는 서비스를 목적으로 하느냐에 따라, 핵심 자원인 온톨로지 및 시소러스와 같은 부가적 지식자원의 구축 방향이 결정되기 때문이다. 2005년도까지의 OFK는 3장에서 언급된 바와 같은, 연구자 네트워크(CoP) 추적, 연구자 연구동향 추적 및 연구 맵 추적 등과 같은 서비스를 제공하고 있다. 2006년도에는 보다 효율적으로 연구·개발자들의 연구를 지원할 수 있는 다양한 서비스를 제공하기 위한 연구·개발을 수행하여 나갈 것이다.

10) Gartner에 의하면 그림 6의 public semantic web은 웹 포털과 같은 일반적인 웹 환경에 시멘틱 웹 기술을 적용한 것으로서, 그간 많은 연구가 있었음에도 불구하고 향후 5~10년 이후에나 상용화가 가능할 것으로 예측되고 있으며 반면 제한된 영역 혹은 분야를 대상으로 하는 corporate semantic web은 가까운 시일 내에 상용화가 가능할 정도로 기술이 안정될 것으로 예측되고 있다.

## 참고문헌

- [1] 신동구, 김재수, 윤정모, 권이남, 전성진, 정택영: 식별체계 간 연계시스템 구축에 관한 연구. 한국정보과학회 추계학술대회 (2005)
- [2] 이원주, 김승연, 전창호: P2P 환경에 유효한 허위 데이터 감축 알고리즘. 한국정보과학회 한국컴퓨터종합학술대회 (2005)
- [3] 이해주, 최범석, 홍진우, 석종원: 디지털 방송 콘텐츠 보호 유통 시스템 설계 및 구현. 한국정보처리학회 논문지. Vol. 11-C. No. 6 (2004)
- [4] 최성원, 이상환, 이상기, 조성남, 석중호: 식별체계와 공개키 기반의 디지털 콘텐츠 유통모델 연구. 한국정보과학회 추계학술대회 (2004)
- [5] Clarke, I., Sandberg, O., Wiley, B., and Hong, T.: Freenet: A Distributed Anonymous Information Storage and Retrieval System. Proceedings of the International Workshop on Designing Privacy Enhancing Technologies. Springer-verlag (2000)
- [6] Jung, H., Koo, H., Lee, B., and Sung, W.: Toward Managing the Life Cycle of Terms Using Term Dominance Trend. Proceedings of the Pacific Association for Computational Linguistics (2005)
- [7] Jung, H., Seon, C., Kim, J., Son, J., Sung, W., and Park, D.: Information Extraction for User's Utterance Processing on Ubiquitous Robot Companion. LNCS 0302-9743. Springer-verlag (2005)
- [8] Jung, H., Sung, W., and Park, D.: Korean Science & Technology Thesaurus Construction Based on Concept/Relation Facets. Proceedings of the 3rd Global WordNet Conference (2006)
- [9] Seon, C., Jung, H., Kim, J., and Son, C.: Formal Description Generation from Users Query on Web-based Service Platform. Proceedings of the 1st International Workshop on Intelligent and Ubiquitous Robotics (2004)
- [10] CAS: <http://www.triplestore.aktors.org/demo/AKTiveSpace/>
- [11] DVB CPCM Functional Model Definitions and Applications. DVB-CPT (2003)
- [12] SEAL: [http://www.aifb.uni-karlsruhe.de/Projekte/viewProjektenglish?id\\_db=58](http://www.aifb.uni-karlsruhe.de/Projekte/viewProjektenglish?id_db=58)

---

### 성 원 경



1983~1987 연세대학교 불어불문학과(학사)  
 1987~1989 연세대학교 불어불문학과(석사)  
 1990~1996 프랑스 파리7대학교 언어학과(박사)  
 1997~1998 한국전자통신연구원 post-doc  
 1998~1999 Lernout & Hauspie Inc. 선임연구원  
 1999~2001 L&H Korea(주) 책임연구원  
 2001~2002 ㈜보이스텍 전문위원

2002~2003 ㈜보이스텍 연구개발본부장/ 상무이사  
 2004~현재 한국과학기술정보연구원 정보시스템연구팀  
 2004~현재 과학기술연합대학원대학교(UST) 겸임교수  
 관심분야: 자연어처리, 시맨틱 웹  
 E-mail: wksung@kisti.re.kr

### 정 한 민



1988~1992 포항공과대학교 전자계산학과(학사)  
 1992~1994 포항공과대학교 전자계산학과(석사)  
 2000~2003 포항공과대학교 컴퓨터공학과(박사)  
 1994~2000 한국전자통신연구원 선임연구원  
 2000~2001 다이렉트닷컴 선임연구원  
 2002~2004 ㈜다이렉트 연구소장/ 기술이사

2004~현재 한국과학기술정보연구원 선임연구원  
 2005~현재 과학기술연합대학원대학교(UST) 겸임교수  
 관심분야: 자연어처리, 시맨틱 웹, 정보 추출, 정보 검색  
 E-mail: jhm@kisti.re.kr

### 박 동 인



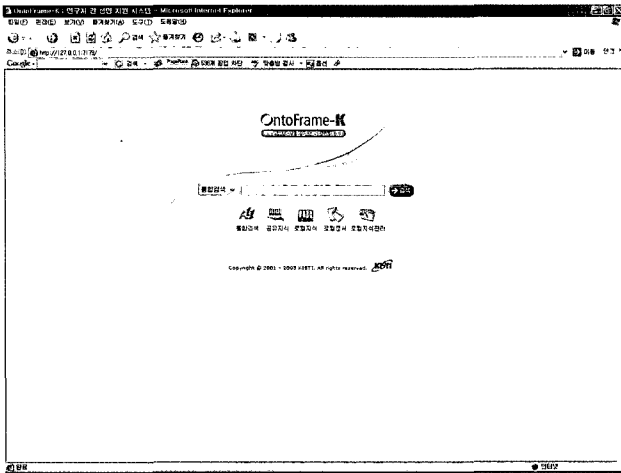
1979 서강대학교 전자공학과(학사)  
 1979~1994 시스템공학연구소 실장  
 1994~1996 시스템공학연구소 국어공학센터장  
 1997~1998 시스템공학연구소 자연어정보처리연구부장  
 1999 (주)SnL 대표이사  
 2000 L&H Korea(주) 음성언어기술 연구소장

2001~2002 나눔정보테크(주) 기술이사  
 2002~2003 리눅스원(주) 대표이사  
 2003~현재 한국과학기술정보연구원 NTIS 사업단장  
 관심분야: 자연어처리, 시맨틱 웹, 정보 검색  
 E-mail: dipark@kisti.re.kr

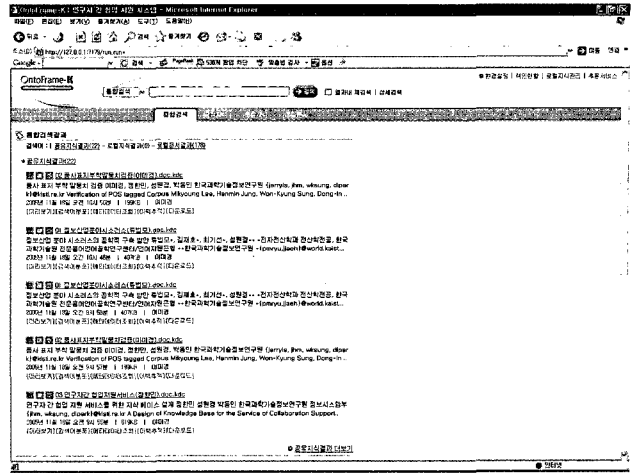
---



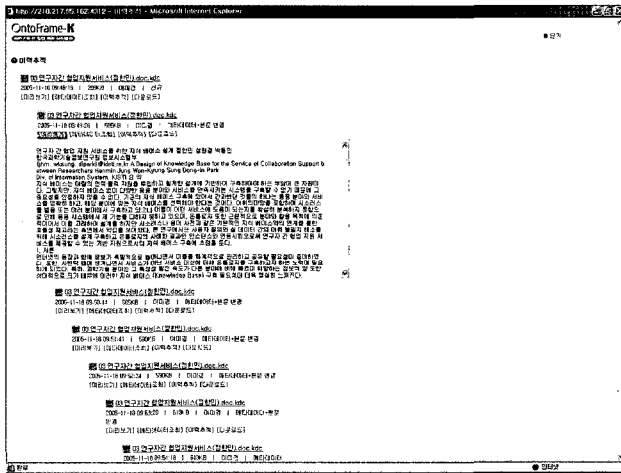
# 부록 A: OFK의 정보유통 서비스 화면



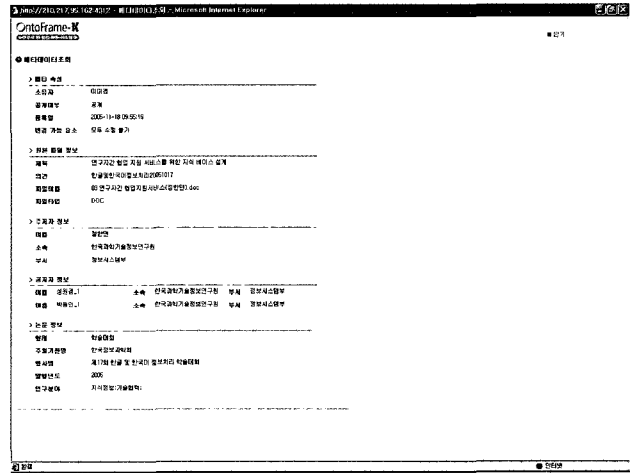
(a) OntoFrame-K 초기 화면



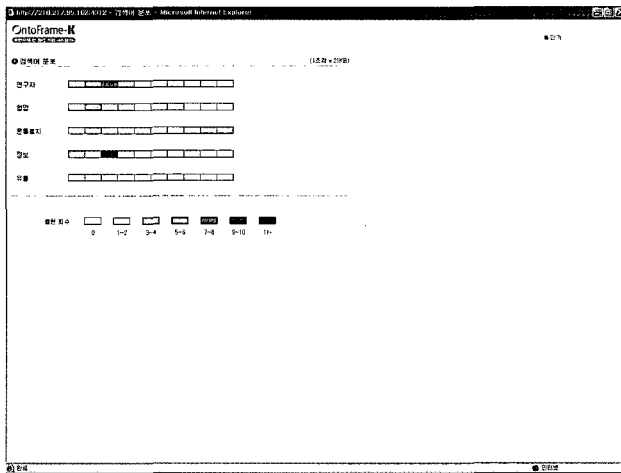
(b) 공유문서 검색



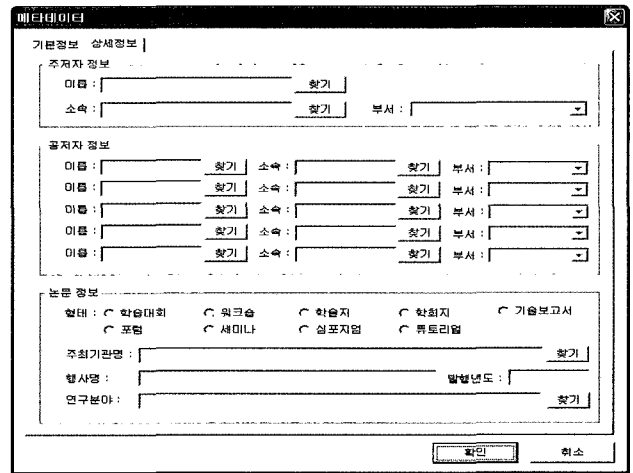
(c) 이력 추적



(d) 메타데이터

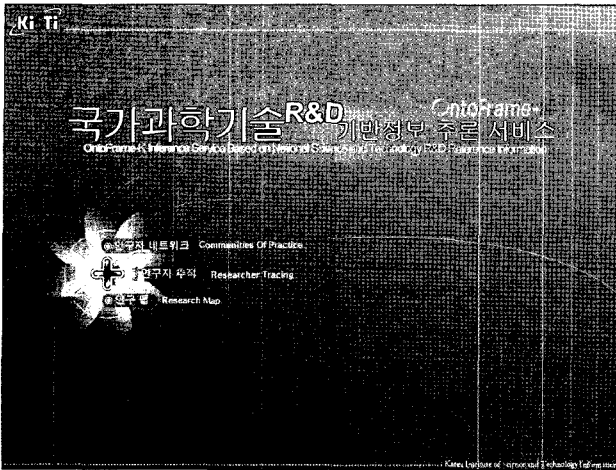


(e) 검색어 분포

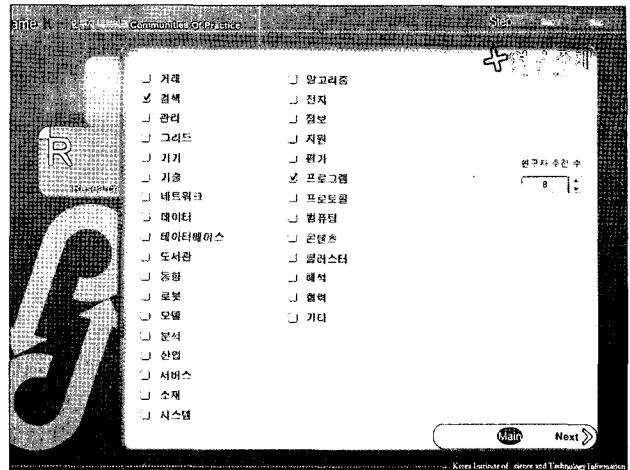


(f) 메타데이터 등록 화면

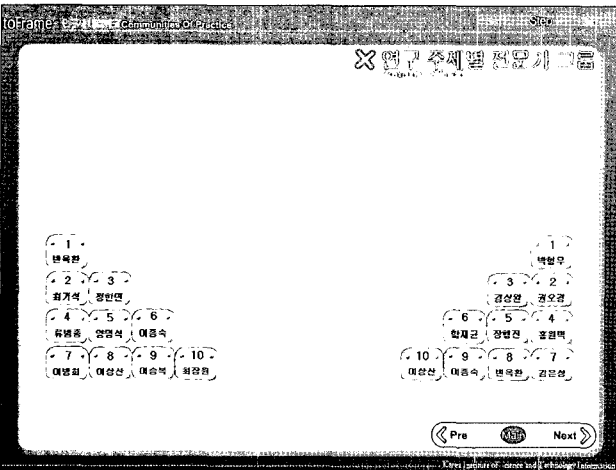
## 부록 B: OFK의 지식 서비스 화면



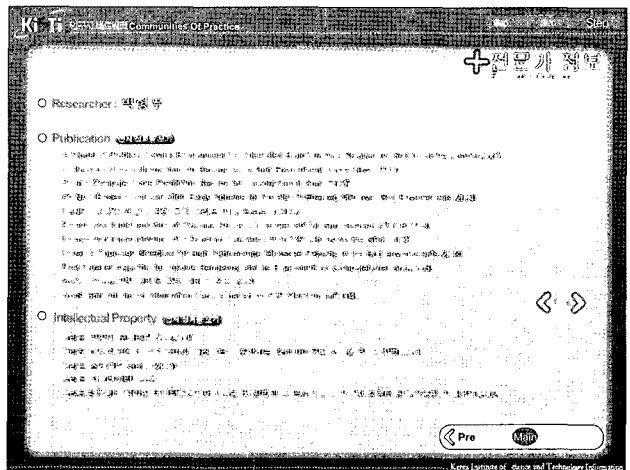
(a) 지식 서비스 초기 화면



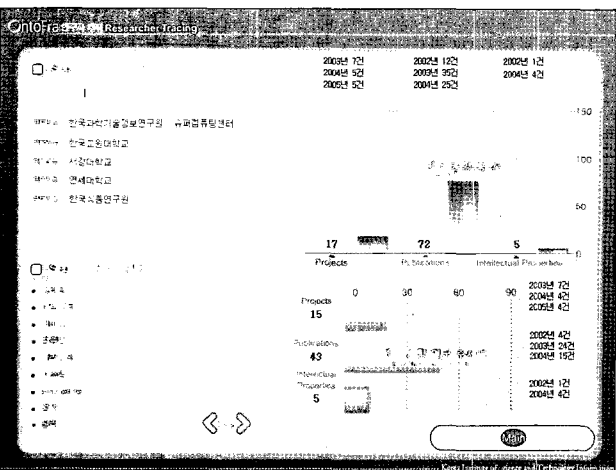
(b) 연구자 네트워크: 연구 주제 선택



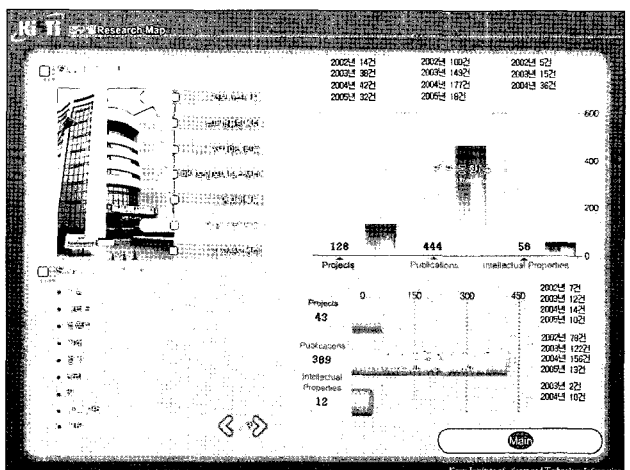
(c) 연구자 네트워크: 전문가 목록



(d) 연구자 네트워크: 성과 정보



(e) 연구자 연구동향 추적



(f) 연구 맵