

OntoFrame 기반 학술정보 분석 서비스

- 심사자 추천과 연구성과 분석 -

(The Academic Information Analysis Service using
OntoFrame - Recommendation of Reviewers and Analysis of
Researchers' Accomplishments -)

김 평 [†]

(Pyung Kim)

이 승 우 ^{††}

(Seung-Woo Lee)

강 인 수 ^{†††}

(In-Su Kang)

정 한 민 ^{††††}

(Han-Min Jung)

이 정 연 ^{†††††}

(Jung-Yeoun Lee)

성 원 경 ^{††††}

(Won-Kyung Sung)

요 약 학술정보 분석 서비스는 학술정보 온톨로지를 사용하여 연구과제의 심사자 선정과 연구자의 연구성과 분석에 필요한 정보를 제공해 주는 서비스이다. 연구과제의 심사자 추천 서비스에서는 과제의 신청 분야와 심사자의 전공 분야, 과제 신청자와 심사자의 관계, 심사자의 해당 분야에 대한 전문도를 고려하여 정확하고 공정한 심사자 추천이 이루어져야 한다. 연구성과 분석 서비스에서는 전공 분야별/기관별 연구성과물 현황, 전공 분야별 전문가 현황, 연구자 네트워크 등을 사용해서 연구자의 연구 현황 정보 제공은 물론 기관, 지역별 연구 성과 현황 정보도 제공되어야 한다. 본 연구에서는 학술정보 분석 서비스를 제공하기 위해 학술정보를 온톨로지로 구축하고, OntoFrame 기반의 추론 시스템을 적용하여 학술정보를 저장하고 지식 확장 과정을 수행한 후 심사자 추천 서비스와 연구성과 분석 서비스에 필요한 정보를 제공하였다. 본 논문에서는 학술정보 온톨로지의 구성과 OntoFrame 기반의 학술정보 시스템의 구성 및 서비스 방법을 제시하였고, 이를 통해 효과적인 학술정보 분석 서비스를 제공하였다.

키워드 : 학술정보 온톨로지, OntoFrame, 시맨틱 웹

Abstract The academic information analysis service is including automatic recommendation of reviewers and analysis of researchers' accomplishments. The service of recommendation of reviewers should be processed in a transparent, fair and accountable way. When selecting reviewers, the following information must be considered: subject of project, reviewer's major, expertness of reviewer, relationship between applicant and reviewer. The analysis service of researchers' accomplishments is providing statistic information of researcher, institution and location based on accomplishments including book, article, patent, report and work of art. In order to support these services, we designed ontology for academic information, converted legacy data to RDF triples, expanded knowledge appropriate to services using OntoFrame. OntoFrame is service framework which includes ontology,

[†] 정 회 원 : 한국과학기술정보연구원 정보기술개발단 선임연구원
pyung@kisti.re.kr

^{††} 정 회 원 : 한국과학기술정보연구원 정보기술개발단 초빙연구원
swlee@kisti.re.kr

^{†††} 비 회 원 : 경성대학교 컴퓨터정보학부 교수
dbaik@ks.ac.kr

^{††††} 정 회 원 : 한국과학기술정보연구원 정보기술개발단 책임연구원
jhm@kisti.re.kr
wksung@kisti.re.kr

^{†††††} 비 회 원 : 이화여자대학교 사회과학연구소 상임연구원
jyonlee@gmail.com

논문접수 : 2007년 11월 29일

심사완료 : 2008년 4월 25일

Copyright©2008 한국정보과학회: 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제35권 제7호(2008.7)

reasoning engine, triple store. In our study, we propose the design methodology of ontology and service system for academic information based on OntoFrame. And then we explain the components of service system, processing steps of automatic recommendation of reviewers and analysis of researchers' accomplishments.

Key words : Academic Information Ontology, OntoFrame, Semantic Web

1. 서론

연구과제를 관리하는 기관에서는 연구과제별 공정하고 전문성 있는 심사자를 선정하는 방법은 물론 연구자의 연구성과 정보를 대상으로 다양한 연구성과 통계 정보를 필요로 한다. 학술진흥재단에서는(이하 학진으로 명명함) 연구과제의 심사자 추천과정에서 공정하고 전문성 있는 심사자 선정을 위해서 다양한 상피 조건, 프로그램 관리자 제도, 동료평가제도 등을 활용하고 있다. 하지만 연구과제의 신청 분야에 적합한 분야의 심사자가 적은 경우 관련 분야의 심사자를 심사자 풀(Pool)에 포함하는 방법과 해당 분야의 심사자가 많은 경우 보다 적합한 심사자를 선별하는 방법이 요구되었다.

또한 학진에서는 연구성과 현황을 파악하기 위해 연구자의 연구성과에 기반하여 연구자별 연구 전문도, 연구성과의 연도별/분야별 연구 통계 정보, 기관 및 지역별 연구성과 통계가 필요하였다.

학술정보 분석 서비스는 심사자 추천 서비스와 연구성과 분석 서비스로 구성되어 있다. 학술정보 분석 시스템은 OntoFrame[1,2]을 기반으로 학술정보 온톨로지와 추론 시스템으로 구축되었고, 온톨로지 기반의 심사자 추천 서비스와 연구성과 분석 서비스를 제공한다. 심사자 추천 과정에서는 확률적 주제 온톨로지[3]를 이용하여 연구자의 심사 분야 확장을 통해 심사자 풀 확대, 연구자의 연구성과에 기반하여 전문도 부여, 피심사자와 심사자의 연구실적 비교를 통해 관련도 판별 등 심사자 선정에 필요한 정보를 제공한다. 연구성과 분석 과정에서는 연구자의 전공 분야를 이용해서 성과물의 분야 할당, 성과물 비교를 통해 연구자 네트워크 생성, 분야별 전문가 선정, 연도별/기관별/지역별 성과 통계를 제공한다.

본 논문에서는 학술정보 분석 서비스를 제공하기 위해 설계된 학술정보 온톨로지와 학술정보 시스템, 추론 서비스의 구현에 대해서 기술하도록 하겠다.

2. 학술정보 분석 서비스 개요

학술정보 분석 서비스는 연구과제에 적합한 심사자를 선정하기 위한 심사자 추천 서비스와 연구현황 정보를 제공하기 위한 연구성과 분석 서비스로 구성되어 있다.

2.1 심사자 추천 서비스

심사자 추천 서비스는 연구과제의 분야별 심사자 분

포와 심사자의 전문도를 고려하여 심사자를 추천하기 위한 서비스로 다음과 같은 서비스를 제공하고 있다.

- 관련 분야 제시: 확률적 주제 온톨로지를 사용해서 연구과제 분야와 관련 분야의 목록과 관련도 제시
- 분야별 심사자 수 제시: 확률적 주제 온톨로지의 가중치를 활용하여 관련 분야를 지정하고 분야별 심사자 수 제시
- 심사자와 피심사자의 관련도: 공동연구 활동, 사제관계 등 고려에 의한 배제 대상 인원 선정
- 심사자 리스트 제시: 심사자 리스트와 심사자의 해당 분야 전문도 제시

확률적 주제 온톨로지는 학문분야별 연관도를 제시한 것으로 심사를 요청한 연구과제에 적합한 분야의 심사자가 적은 경우 학문분야별 연관도에 따라 심사 가능 분야를 확장하고, 확장된 분야의 심사자를 심사자 풀에 포함시켜 심사자를 확대하는데 사용되며, 심사자 추천 과정에서의 확률적 주제 온톨로지의 유용성은 실험을 통해 검증되었다[3].

그림 1은 확률적 주제 온톨로지를 적용하였을 경우 분야별 관련도에 따라 심사자 풀의 크기 변화를 보여주는 것으로 연구과제의 지원 분야중 1순위 분야에만 적용한 경우와 3순위 분야(전 분야)까지 적용한 경우를 구분하여 보여주고 있다. 분야별 관련도는 0~1사이의 값을 가지며, 0.4 값을 전후로 심사자 풀의 크기가 커짐을 알 수 있다. 심사자 추천 서비스에서는 심사자 추천 책임자가 이 그래프를 참고하여 확률적 온톨로지의 적용 여부와 분야별 관련도를 지정할 수 있다.

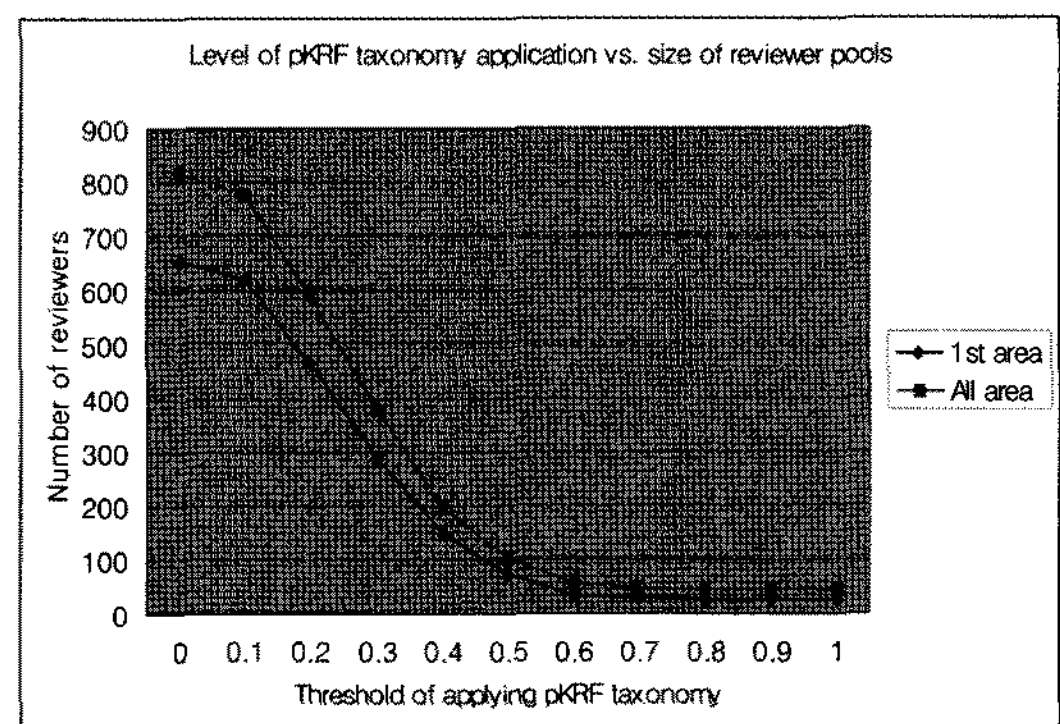


그림 1 확률적 주제 온톨로지 적용

연구성과 분석 서비스는 연구 성과물(논문, 특허, 보고서, 작품)을 통해 연구자의 연구성과 및 연구 관계를 제시하는 것은 물론 기관과 지역의 분야별, 연도별 성과 통계 정보를 제공하는 서비스로 다음과 같은 정보를 제공하고 있다.

- 분야별 연구자: 분야별 연구자 목록과 연구자별 전문도 제시
- 연구자 네트워크: 공동연구, 공저, 사제관계를 포함하는 연구자 네트워크 제시
- 분야별 관련도: 학문 분야별 관련 분야 추출 및 유사도 제시
- 연구성과 통계: 연도별, 분야별 연구성과 정보 제시 및 분야간 연구성과 비교
- 기관 통계: 대학, 연구소의 직급별, 분야별 전문가 분포 현황

온톨로지의 설계는 온톨로지를 통해 다루고자 하는 데이터와 온톨로지를 통해 제공하고자 하는 서비스를 고려되어야만 한다. 학술정보 온톨로지는 다음과 같은 내용을 고려하여 설계되었다.

- 대상 정보: 학술정보 분석 서비스를 제공하기 위해 필요한 정보 선별
 - 클래스(class): 선별된 정보의 유형별로 적합한 클래스 설계
 - 속성: 데이터타입 속성(data type property)와 객체타입 속성(object type property)로 구분하며, 학술정보 서비스에 사용되는 최소한의 속성만을 대상으로 설계
- 그림 2는 학술정보 온톨로지의 클래스와 속성을 도식화한 것으로 클래스는 원으로, 클래스의 데이터타입 속성은 해당 클래스 옆에 표기하였고, 객체타입 속성은 클래스간의 관계를 나타내는 연결선에 표기하였다.

온톨로지는 온톨로지의 스키마와 온톨로지의 인스턴스가 구분되어 관리되며, 온톨로지 스키마는 OWL을 사용하여 클래스와 클래스간 관계 및 제약 조건을 표현하였고, 온톨로지 인스턴스는 RDF 트리플[5]형태로 사실 정보를 표현하였다. 온톨로지의 인스턴스는 URI 기반의

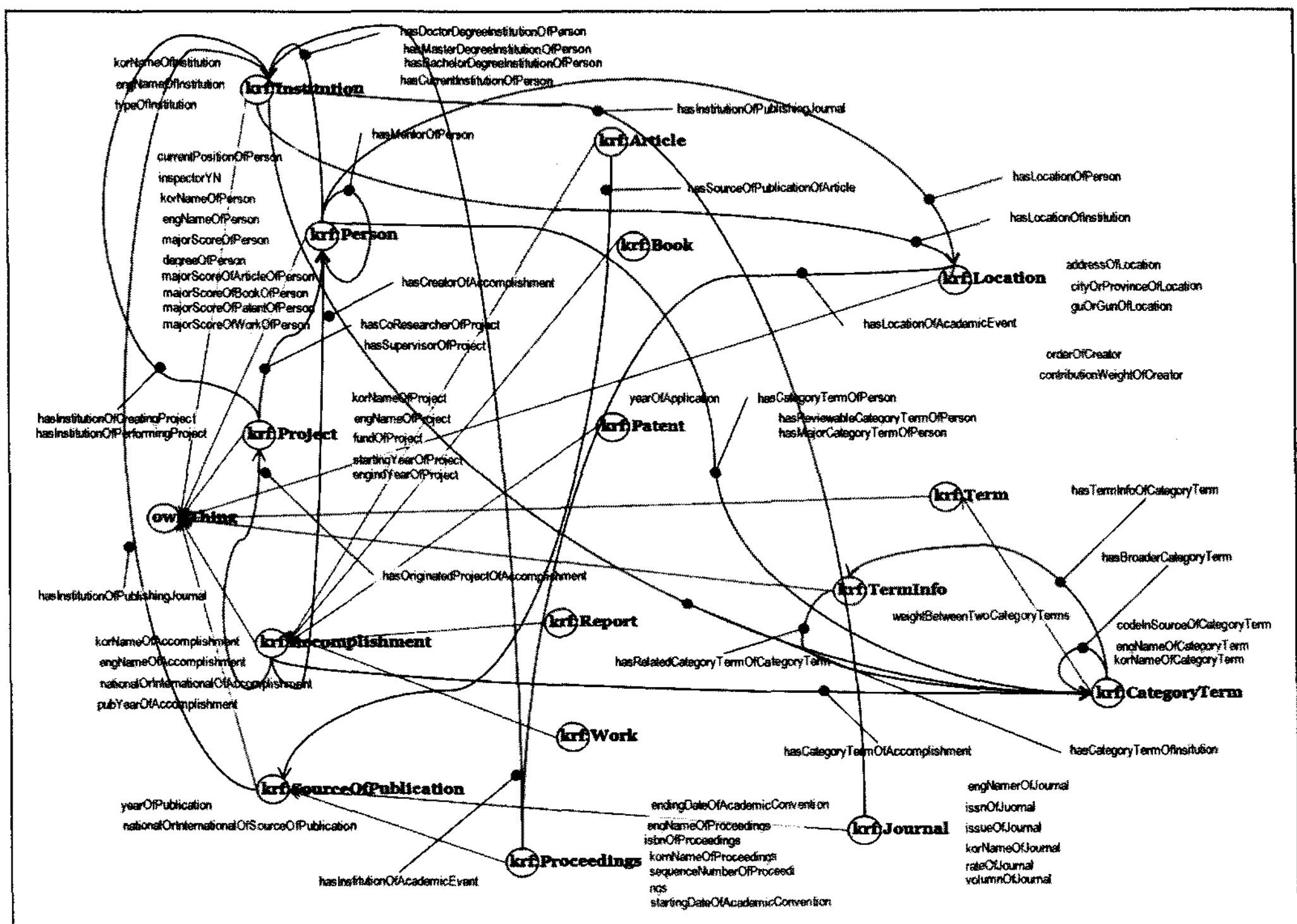


그림 2 학술정보 온톨로지

표 1 학술정보 유형별 내용

항목	내용
연구자	성명, 학력, 소속기관, 연구성과, 분야
성과물	성과물 출처, 제목, 종류 (논문, 저서, 작품, 특허, 보고서)
기관	기관명, 기관주소, 기관종류
지역	지역명
분야	분야명, 분야 체계, 주제 연관도
출처	출처명, 출판연도 (저널, 프로시딩)

정보식별 체계를 기반[6]으로, 레거시 DB의 ID를 URI 기반의 식별체계로 변환하였다.

URI는 “http://www.krf.or.kr/krf” 형태의 네임스페이스와 정보 유형을 나타내는 헤더(Prefix) “BOO_”, 식별번호를 나타내는 ID “000003_2002_1_B00386”로 구성되며 http://www.krf.or.kr/krf#PRO_000003_2002_1_B00386”와 같다. 데이터타입 속성 관계에서의 객체는 데이터 뒤에 데이터 자료형을 표기하도록 함으로써 데이터 값의 자료형에 대한 체크는 물론 데이터 자료형에 적합한 데이터 형태로 저장할 수 있다.

표 2 RDF 트리플

항목	데이터
ID	0f2cf3cd-4806-11dc-a1d6-8153d5330494
Subject	<http://www.krf.or.kr/krf#PRO_000003_2002_1_B00386>
Predicate	<http://www.krf.or.kr/krf#korNameOfProject>
Object	"스키활동에서 위험지각과 감정적 반응이 만족에 미치는 영향"^^<http://www.w3.org/2001/XMLSchema#string>
Time	2007-08-11 00:00:00.0
Object_URI	<http://www.krf.or.kr/krf#PRO_000003_2002_1_B00386>

표 2는 RDF 트리플의 구조를 보여주는데, ID는 트리플에 할당된 일련의 고유번호, SUBJECT는 관계의 주체, PREDICATE는 관계명, OBJECT는 관계의 대상, TIME은 트리플이 생성된 시점, OBJECT_URI는 해당 트리플이 유발된 객체의 URI를 나타낸다.

4. 학술정보 시스템 구성

학술정보 시스템은 다수의 레거시 DB들로부터 온톨로지 구축에 필요한 데이터를 수집하여 RDF 트리플로 변환하고, 변환된 RDF 트리플을 추론 시스템을 저장한 후 사용자의 요청에 적합한 추론 서비스를 제공한다. 그림 3은 학술정보 시스템의 구성요소와 요소별 주요 기능에 대해 보여주고 있다.

4.1 리스너

리스너는 학진의 레거시 DB를 대상으로 학술정보 온톨로지에 적합한 RDF 트리플을 생성하는 작업을 수행한다. 모든 객체는 URI를 할당받아 관리되며, 기존 DB의 식별자를 URI 식별자 형태로 변경해서 사용한다. 기존의 OntoFrame 시스템에서는 URI 서버를 사용해서 데이터 신규 등록 및 URI를 할당하는 작업을 수행하지만[7], 기존의 레거시 DB와의 연동과 기존 식별체계를 유지하기 위해서 URI 서버를 대체하는 모듈이 필요하였다. 이를 위해 레거시 DB에서 필요한 자료를 수집하고 이를 RDF 트리플 형태로 변경하는 리스너를 개발하였다.

리스너의 주요 기능을 살펴보면 다음과 같다.

- 레거시 DB 조회 : 레거시 DB를 조회하여 RDF 트리플을 생성하기 위한 정보 수집
- URI 생성 : 기존의 식별체계를 활용해서 URI 형태의 식별자를 부여
- URI 조회 : 데이터에 따라 신규 자료로 URI를 부여할 지 또는 기존에 할당된 URI를 부여할지 판단
- 전문도 계산 : 성과물의 중요도와 연구자 수에 따라 연구자별 전문도 계산
- RDF 트리플 생성 : 표 2와 같은 형태로 온톨로지에 관여하는 모든 데이터를 RDF 트리플로 생성

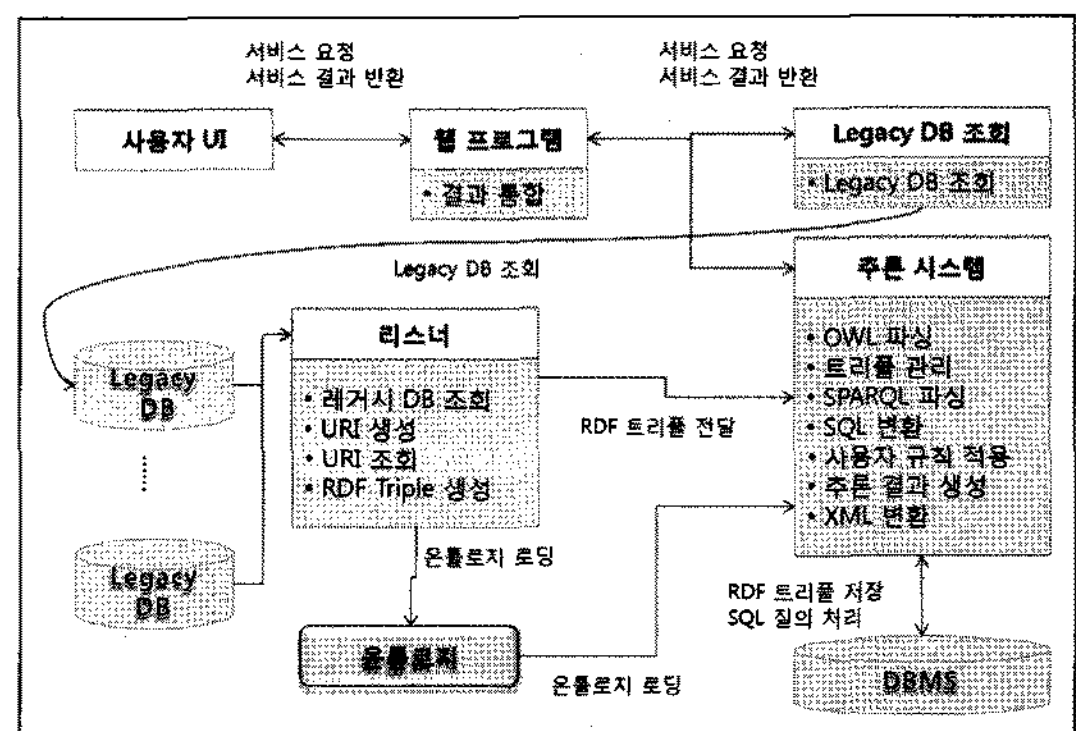


그림 3 학술정보 시스템 구성

그림 4는 학진에서 사용하는 연구자별 전문도 계산 방법으로, 최근 6년 동안의 연구 성과물을 대상으로 성과물의 유형과 공동연구를 수행한 연구자 수를 이용해서 연구자의 전문도를 계산하고 있다. 학진의 경우 연구자가 연구 성과물 입력시 공동저자의 정보를 입력하지 않고 단지 공동저자의 수만 입력하도록 되어 있다. 따라서 공동연구자를 판단하기 위한 방법으로 연구자가 입력한 성과물의 유형에 따라 항목의 비교를 통해 같은 성과물인 경우 같은 URI가 할당되도록 하였다. 즉 논문의 경우 정규화된 논문의 제목, 출처, 출판연도를 고려

■ 전문도 분석
최근 5년간 연구실적을 대학교육협의회 기준을 준용하여 수정한 항목별 가중치를 적용하여 연구자 전문도를 산출합니다.

항목		가중치
학술논문	국내 학술지(동계)	1
	국내 학술지(동계후보)	1
	국내 학술지(기타)	0.5
	국내 학술대회 발표 논문집	0.5
	국외 학술지(SCI/SSCI/A&HCI)	2
	국외 학술지(기타)	1
	국외 학술대회 발표 논문집	1
저역서	저서	국내 1.5, 국외 2
	역서	1
특허	국내	1
	국외	1
작품	미술, 음악, 연주	1
단독	-	1
공동연구	책임자	$(1 + (n-1) \times 2)$
	공동연구자	$1 + (n-1)$
	연구책임자 없는 경우	$1 + n$

그림 4 연구성과 가중치

해서 같은 성과물인지 여부를 판단하고, 같은 성과물로 판단된 경우 연구 성과물에 동일한 URI를 부여함으로써 성과물의 저자 정보를 이용해서 공동저자를 찾을 수 있도록 하였다.

4.2 추론 시스템

추론 시스템은 리스너에서 생성된 RDF 트리플을 DBMS기반 추론 시스템에 저장하고, 사용자의 요청에 적합한 추론 결과를 반환하는 역할을 수행한다. 추론 시스템은 RDF 트리플을 저장하기 위해 스키마-인지(schema-aware)[8]를 따르며, 질의에 대한 응답을 처리하기 위해 SPARQL 질의를 SQL로 변환하여 사용한다. 추론 시스템에서 수행되는 기능을 살펴보면 다음과 같다.

4.2.1 온톨로지 로딩

온톨로지는 온톨로지 스키마와 인스턴스로 구분되어 관리된다. 온톨로지 스키마는 설계가 완료되면 서비스가 추가 또는 수정되기 이전에는 재로딩될 필요가 없지만, 인스턴스의 경우에는 추가 또는 변경 작업이 빈번히 일어난다.

• 온톨로지 스키마 로딩

온톨로지 스키마 로딩과정에서는 OWL로 표현된 학술정보 온톨로지 파싱, 클래스와 속성 관계 분석 결과 저장, 클래스와 데이터타입 속성, 객체타입 속성을 저장하기 위한 RDF 트리플 저장구조를 생성한다. RDF 트리플은 클래스와 속성별로 별도의 테이블을 가지며, 테이블의 구조는 데이터타입 속성의 경우 데이터타입의 자료형을 고려해서 적합한 테이블 구조로 생성된다[6]. 온톨로지 스키마 로딩이 완료되면 네임스페이스 관리 테이블, 온톨로지 스키마 관리 테이블, RDF 트리플을 저장하기 위한 저장 테이블 생성이 완료된다.

• 인스턴스 로딩

인스턴스 로딩 과정에서는 리스너에서 생성된 RDF

트리플을 트리플 저장구조에 저장하는 작업을 수행한다. RDF 트리플에서 나타난 네임스페이스를 URI를 축약형으로 변경하여 저장하고, 데이터타입 속성인 경우 데이터타입에 적합한 테이블을 생성하고 저장함으로써 저장 공간의 낭비와 성능 저하를 방지하였다[9]

4.2.2 지식 확장

지식 확장은 온톨로지 스키마에 정의된 클래스와 속성간의 관계에 따라 지식을 확장하는 작업과 사용자 정의 규칙에 의해 지식을 확장하는 작업을 포함하고 있다. 온톨로지 스키마에 정의된 규칙이 적용되는 대표적인 예를 살펴보면 다음과 같다.

- subClassOf: 클래스간 상하위 관계를 사용해서 하위 클래스의 인스턴스를 상위 클래스의 인스턴스로 확장
- subPropertyOf: 속성간 상하위 관계를 사용해서 하위 속성의 인스턴스를 상위 속성의 인스턴스로 확장
- domain, range: 속성의 subject, object의 값에 클래스 제한이 있는 경우 속성이 존재하고 클래스의 인스턴스로 등록되지 않은 인스턴스를 클래스의 인스턴스로 등록

온톨로지 스키마에 정의되지 않았으나, 서비스를 제공하기 위해 사용자는 규칙을 적용하고, 규칙을 통해 지식을 확장할 수 있다. 사용자 규칙은 '조건부 -> 결과부'로 구분하여 규칙을 기술하며 '(?x isrl:hasCreatorInfo ?y) (?y isrl:hasCreator ?z) -> (?x isrl:createdByPerson ?z)'와 같이 표현된다. 이 규칙은 어떤 x가 저작정보 y를 가지고, 저작정보 y가 저자 z를 가지면 x는 z에 의해 생성되었다는 것을 규칙으로 기술한 것이다.

4.2.3 SPARQL-SQL 질의 변환 및 결과 생성

사용자 UI에서 전달되는 모든 질의는 추론 시스템에서 SPARQL 형태로 변환되어 처리된다. SPARQL은 추론 시스템에서 사용되는 질의 언어로서, SPARQL 질의를 DBMS 기반의 추론 시스템에서 사용하기 위해서는 SPARQL 질의를 파싱하고, 파싱된 결과를 SQL 질의로 변환하는 과정이 필요하다[10].

그림 5는 연구자 검색을 위한 SPARQL 질의의 예를 보여준다. '이정연'연구자의 이름을 가지고 연구자 URI와 전문도, 직위, 기관, 기관의 소재 지역 등을 전문도 순으로 정렬하기 위해 작성된 SPARQL이다. SPARQL 질의의 'OPTIONAL'은 속성이 존재하는 경우에 한해 값을 반환하는 구문이고, 'FILTER'는 조건을 통해 값을 필터링하는 역할을 수행한다. 이렇게 생성된 질의에 적합한 결과를 DBMS에 저장된 저장구조를 통해 얻기 위해서는 DBMS의 질의 언어인 SQL로 변환이 필요하다.

그림 6은 그림 5의 SPARQL 질의가 SQL 질의로 변환된 결과를 보여준다. SPARQL에서 나타난 Predicate는 DBMS의 테이블명으로 변경되어 처리되며, 'OP-

```

SELECT ?categoryName ?researcher ?researcherName ?position ?score ?institution ?institutionName ?locationName
WHERE {
  ?researcher krf:korNameOfPerson ?researcherName .
  ?researcher krf:currentPositionOfPerson ?position .
  OPTIONAL { ?researcher
    krf:hasMajorCategoryTermOfPerson ?category . }
  OPTIONAL { ?category
    krf:korNameOfCategoryTerm ?categoryName . }
  OPTIONAL { ?researcher
    krf:majorScoreOfPerson ?score . }
  OPTIONAL { ?researcher
    krf:hasCurrentInstitutionOfPerson ?institution . }
  OPTIONAL { ?institution
    krf:korNameOfInstitution ?institutionName . }
  OPTIONAL { ?institution
    krf:hasAreaCodeOfInstitution ?location . }
  OPTIONAL { ?location
    krf:addressOfLocation ?locationName . }
  FILTER(REGEX(?researcherName, '이정연')) .
}
ORDER BY ?score DESC

```

그림 5 연구자 검색을 위한 SPARQL 예

```

SELECT T3.obj AS categoryName, T0.subj AS researcher,
T0.obj AS researcherName, T1.obj AS position, T4.obj AS
score, T5.obj AS institution, T6.obj AS
institutionName, T8.obj AS locationName
FROM "NS0:KORNAMEOFPERSON" T0
  INNER JOIN "NS0:CURRENTPOSITIONOFPERSON" T1 ON
T0.subj=T1.subj
  LEFT JOIN "NS0:HASMajORCATEgORYTERMOFPERS" T2 ON
T1.subj=T2.subj
  LEFT JOIN "NS0:KORNAMEOFCATEgORYTERM" T3 ON
T2.obj=T3.subj
  LEFT JOIN "NS0:MAJORSOREOFPERSON" T4 ON
T1.subj=T4.subj
  LEFT JOIN "NS0:HASCURRENTINSTITUTIONOFPER" T5 ON
T1.subj=T5.subj
  LEFT JOIN "NS0:KORNAMEOFINSTITUTION" T6 ON
T5.obj=T6.subj
  LEFT JOIN "NS0:HASAREACODEOFINSTITUTION" T7 ON
T5.obj=T7.subj
  LEFT JOIN "NS0:ADDRESSOFLOCATION" T8 ON
T7.obj=T8.subj
WHERE ( T0.obj LIKE '%이정연%' )
ORDER BY T4.obj DESC

```

그림 6 연구자 검색을 위한 SQL 예

TIONAL'은 LEFT JOIN으로 처리되고, FILTER는 WHERE 절에 조건으로 반영된다. SPARQL에서 나타난 Predicate 'krf:hasMajorCategoryTermOfPerson'는 네임스페이스의 축약형과 오라클 DBMS의 테이블명 길이 제한으로 인해 'NS0:HASMajORCATEgORYTERMOFPERS'로 변경되며, 테이블명 변환을 위해 테이블명 변환을 위한 매핑 테이블이 사용된다. 변환된 SQL

질의를 사용해서 DBMS 기반 추론 시스템에서 결과를 요청하고, 요청된 결과는 후처리 모듈을 통해 사용자 UI에서 요청하는 결과를 가공된다. 최종 가공된 결과는 미리 정의된 XML 형태로 변환되어 웹 서비스를 통해 사용자 UI를 담당하는 웹 프로그램으로 전달된다.

4.2.4 ECPV 적용

RDF 트리플을 스키마 인지 방식으로 처리하게 되면 모든 테이블은 '주체-속성-객체' 형태로 지식을 단편화하고 이를 독립적인 테이블로 관리하기 때문에 사용자의 질의를 처리하는 과정에서 다수의 테이블에 대한 조인이 빈번하게 발생하게 된다. 조인에 따른 속도 저하를 방지하기 위해서 확장 클래스-속성 뷰(Expanded Class-Property View: ECPV)[11]를 사용하였다. ECPV는 기존의 트리플 형태의 지식은 그대로 유지하면서 질의처리를 보다 효율적으로 처리하기 위해서 부가적으로 생성하고 관리하는 저장 구조로서, 특정 클래스로부터 직접 또는 간접으로 관계를 가지는 속성 및 속성 값의 쌍들로 구성된다.

사람이 소속된 기관명, 기관의 이름, 기관의 위치 등이 독립적인 테이블로 저장된 경우 사람을 중심으로 사람의 이름과 기관명, 지역 정보를 얻기 위해서는 다수의 조인이 필요하다. 그림 7은 이런 경우 조인의 횟수를 줄이기 위해 사람을 중심으로 ECPV가 생성된 보여준다. 이런 ECPV의 생성 및 관리는 온톨로지 스키마를 사용해서 추론 시스템에서 자동으로 생성하고 관리하게 되며, SPARQL 질의가 SQL로 변환하는 과정에서도 ECPV 테이블을 고려해서 변환된다.

Person		Institution		KoreanNameOfPerson		nameOfInstitution		locationOfInstitution	
Subject	Object	Subject	Object	Subject	Object	Subject	Object	Subject	Object
PER_000001	INS_000001	PER_000001	홍길동	INS_000001	KISTI	INS_000001	300301	INS_000001	300301
PER_000002	INS_000002	PER_000002	이순신	INS_000002	KIST	INS_000002	300301	INS_000002	300301
PER_000003	INS_000003	PER_000003	김철수	INS_000003	ETRI	INS_000003	300301	INS_000003	300301
...

Person_ECPV				
Subject	KoreanNameOfPerson	EnglishNameOfPerson	hasInstitutionOfPerson_nameOfInstitution	locatedIn
PER_000001	홍길동	Kindong Hong	KISTI	300301
PER_000002	이순신	Sunshin Lee	KIST	300301
PER_000003	김철수	Cheonsoo Kim	ETRI	300301
...

그림 7 ECPV 적용 예제

4.3 DBMS 조회

DBMS 조회 모듈은 학술정보 온톨로지에 반영되지 않는 레거시 DB의 항목을 조회하여 사용자 UI를 구성하는 정보를 제공하는데 사용된다. 학진의 경우 학술정

보 분석 서비스와 관련된 데이터만 온톨로지에 반영하였기 때문에, 레거시 DB와의 정보 연계를 위해 학진의 ID와 정보공개 여부 정보를 온톨로지에 반영하여 추론 결과에 포함하여 제공하였다.

4.4 웹 프로그램

사용자 UI에서 전달되는 사용자의 요청을 분석하여 레거시 DB에서 조회할 내용과 온톨로지 기반 추론을 통해 얻을 수 있는 정보를 분리한다. 추론 시스템에 전달할 질의는 추론 시스템에서 제공하는 서비스 API를 웹 서비스를 통해 호출하고, 그 결과를 XML 형태로 전달받아 처리한다. 이후 웹 프로그램에서는 XML 파싱을 통해 얻은 결과와 레거시 DB 조회를 통해 얻은 결과를 가공하여 사용자 UI에 출력할 최종 데이터를 생성한다.

5. 학술정보 분석 서비스

학술정보 서비스를 제공하기 위해 구축된 RDF 트리플의 수와 시스템 환경 및 사용자 UI를 통해 심사자 추천 서비스와 연구성과 분석 서비스를 살펴보도록 하겠다.

5.1 서비스 환경

학술정보 분석 서비스를 위해 사용된 서비스 환경과 학술정보 온톨로지 구축된 RDF 트리플의 수를 살펴보면 표 3과 표 4와 같다.

표 3은 학술분석 시스템의 서비스 환경을 보여준다.

표 4는 학술정보 온톨로지를 통해 구축된 RDF 트리플의 수를 주요 클래스와 속성별로 보여준다. 트리플의 수는 온톨로지 스키마와 사용자 규칙에 의해 확장된 트리플을 모두 포함하고 있으며, Subject에 해당되는 정보 유형을 중심으로 개수를 산정하였다. 성과물은 논문, 저

표 3 서비스 환경

항목	내용
운영체제	Windows Server 2003
DBMS	Oracle 10g
WAS	Tomcat 5.5
웹서버	Apache Web server
개발언어	Java, Jsp

표 4 RDF 트리플 수

항목	인스턴스	데이터타입 속성	객체타입 속성	계
성과물	629.8 만	952.9만	344.3만	1,922.7만
과제	7.9만	15.8만	12.2만	36만
연구자	17.9만	211.1만	119.5만	348.7만
기관	1.3만	2.6만	9.9만	13.8만
분야	11.4만	11만	10.6만	33.1만
지역	0.2만	0.2만	0	0.4만
기타				912
계	668.9만	1,193.9만	496.6만	2,359.3만

서, 특허, 보고서 포함, 과제는 학진 연구과제 및 연구자별 수행과제 정보 포함, 연구자와 기관은 학진에 등록된 연구자와 기관만 포함, 기관의 연구소와 대학 및 기타로 구분된 기관들을 포함, 분야는 학진의 분야 분류체계와 확률적 주제 온톨로지 포함, 기타는 온톨로지 스키마 정보 및 네임스페이스와 테이블 매핑을 위한 정보를 포함하고 있다. 총 트리플의 수는 약 2,360만 트리플 정도이다.

5.2 추론 서비스 API

학술정보 분석 서비스를 지원하기 위해 추론 시스템에서 제공하는 추론 API는 총 14개로 구성되어 있다. 추론 API의 호출은 웹 서비스를 통해 호출되며 추론의 결과는 XML 형태로 변환되어 웹 서비스를 통해 반환된다. 다음은 심사자 추천 서비스에서 사용되는 함수의 API이다

- public String getResearcherGraph (String[] category-URIList, int areaKind, float pKRFStart, float pKRFEnd, float step);:: 심사자의 관련 분야별 분포를 보여주는 그래프 요청
- public String getResearcherList (String[] category-URIList, String explicitCond, String[] personURIList, int areaKind, float pKRFStart, float pKRFEnd));:: 특정 구간에 속한 심사자 목록 요청

5.3 심사자 추천 서비스

심사자 추천 서비스는 심사 지원 과제의 심사자를 선정하기 위한 서비스로 확률적 주제 온톨로지와 심사자의 전문도 및 심사자와 피심사자의 공동연구 관계 정보를 제공한다.

그림 8은 심사자 추천에 사용되는 사용자 UI를 보여준다. ①은 심사에 지원한 연구과제의 분야와 확률적 주제 온톨로지를 적용해서 확장된 분야 및 관련도를 보여준다. 관련도의 범위는 ②에서 0.0~1.0 사이에서 선택되

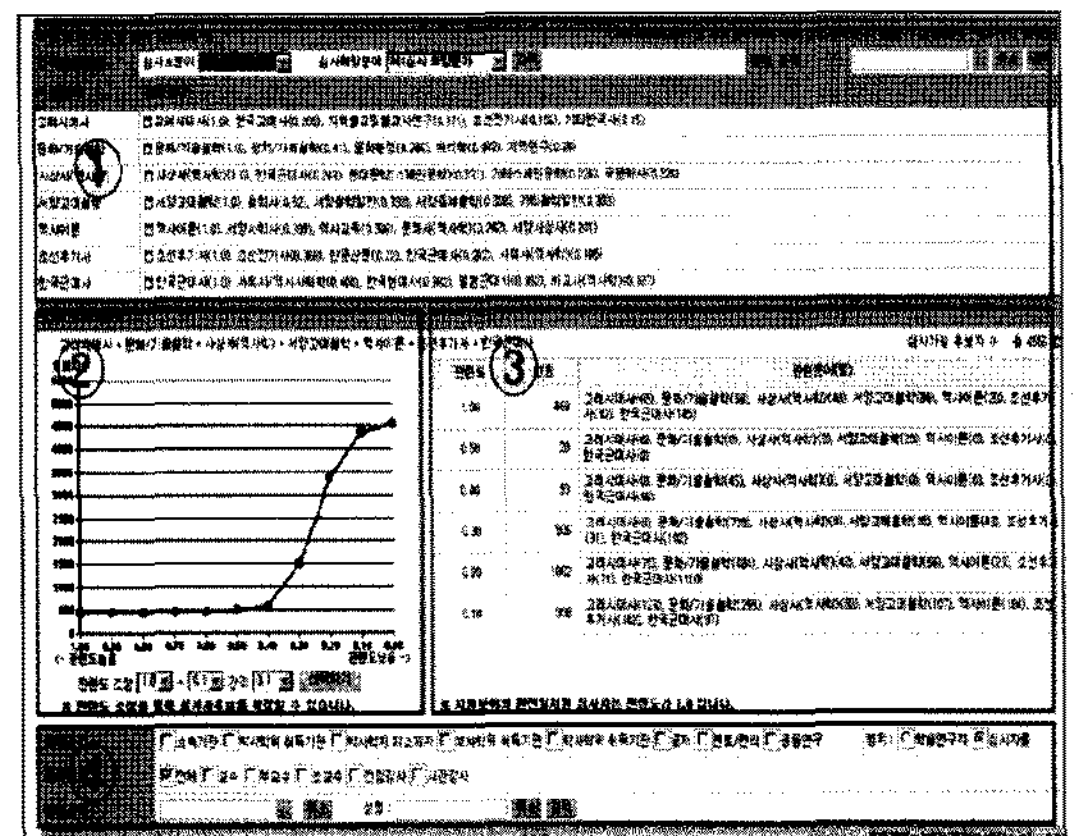


그림 8 심사자 추천 UI

번호	성명	소속	직위	전공분야 (연구분야)	기타 분야	연구도 ①	연구도 ②	연구도 ③	연구도 ④
6948	김문석	조교수	조선후기사 경희사 조선시대사, 한국사상사, 경희사	조선후기사 한국정치사상 조선전기기사 한국정치사상 조선전기기사 한국정치사상	74.5	2	0	0	
694821030110	동국대학교	서울		한국정치사상사 조선전기기사 한국정치사상 조선전기기사 한국정치사상 조선전기기사					
7789	고영섭	조교수	지역문화연구 한국문화사상사 한국문화사상사 한국문화사상사	한국문화사상사 한국문화사상사 한국문화사상사 한국문화사상사	73.7	1	0	0	
694821030110	동국대학교	서울		한국문화사상사 조선전기기사 한국문화사상사 조선전기기사 한국문화사상사 조선전기기사					
678	정재영	교수	한국고대사 삼국시대사상사	한국고대사 삼국시대사상사	66.1	1	0	0	0
694821030110	서울대학교	서울		한국고대사 삼국시대사상사 한국고대사 삼국시대사상사					
7745	박정국	조교수	서양고대사상 서양고대사상 서양고대사상 서양고대사상	서양고대사상 서양고대사상 서양고대사상 서양고대사상	64.0	1	0	0	
694821030110	한국대학교	서울		서양고대사상 서양고대사상 서양고대사상 서양고대사상					
7255	김재규	연구위원	조선전기기사 경희사상사	조선전기기사 경희사상사	62.3	1	0	0	0
694821030110	한국대학교	서울		조선전기기사 경희사상사 조선전기기사 경희사상사					

그림 9 심사자 리스트

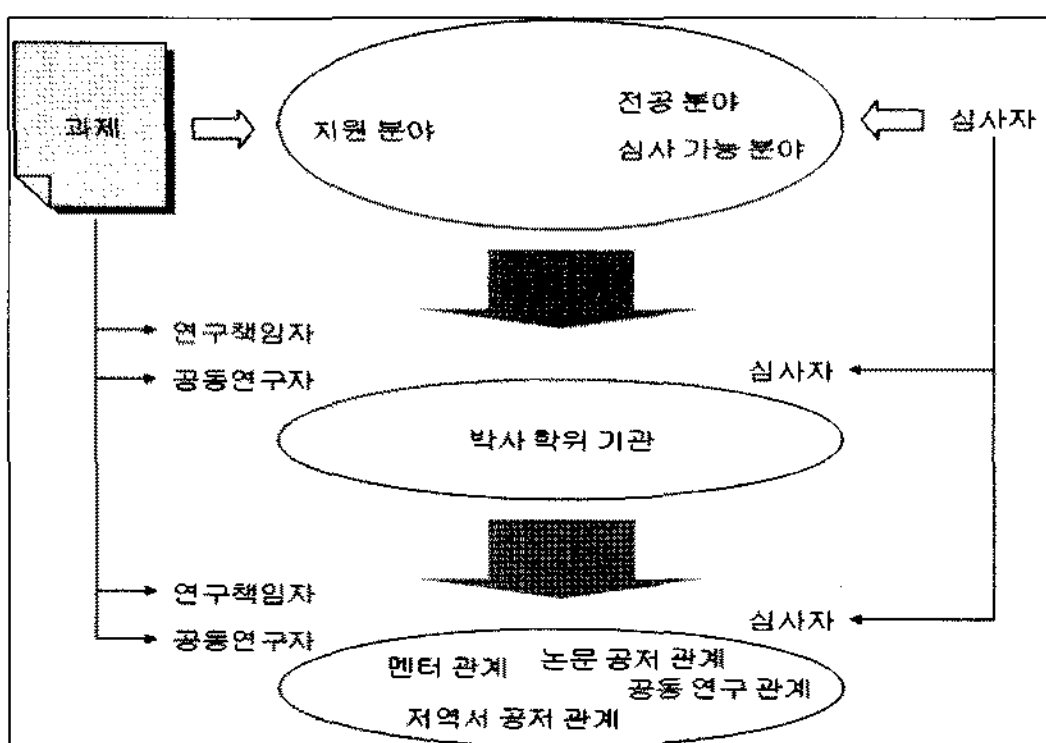


그림 10 심사자 추천 프로세스

며, 선택된 구간에 속한 심사자의 수를 그래프를 통해 관련도 구간별로 보여준다. 관련도의 구간 범위가 1.0~1.0인 경우는 확률적 주제 온톨로지를 적용하지 않은 경우이며, 관련도의 구간 범위가 0.0~1.0인 경우는 연구과제의 분야와 연관성을 가지는 모든 분야를 대상으로 심사자 풀을 확대한 경우이다. ③은 해당 관련도 구간별에 관련된 분야별로 심사자 수를 보여주며, ④는 심사자 배제 규칙과 직위에 따른 심사자 검색을 위한 화면으로 화면의 하단에는 그림 9와 같이 조건에 적합한 심사자 정보가 제시된다.

심사자 추천을 위한 과정은 그림 10과 같이 진행되며 단계별로 설명하면 다음과 같다.

- 확률적 주제 온톨로지의 관련도 구간에 따라 심사대상 연구과제의 분야를 대상으로 심사가 가능한 분야 확장
- 분야별 심사자 목록을 추출: 심사자 또는 전체 연구자 중 선택하여 심사자 목록을 구성
- 추출된 분야별 심사자 목록을 심사대상 연구과제의 분야 관련도(확률적 주제 온톨로지 적용)와 분야별로 심사자 수 산정

- 심사대상 연구과제 책임자의 연구성과를 중심으로 연구자 네트워크를 생성하고, 연구과제 책임자와 심사자 간의 공동연구관계 분석하여 배제 대상 선정

연구자 네트워크는 특정 연구자를 중심으로 공저자(성과물에 공동 참여)와 공동연구자(프로젝트에 공동 참여), 멘토 관계를 사용해서 생성되며, 심사자 추천 과정에서 심사대상 연구과제의 책임자와 심사자가 연구자 네트워크를 형성하면 심사자는 심사자 추천 대상에서 배제된다. 특정 연구자 중심의 연구자 네트워크 생성 과정은 다음과 같은 순서로 수행된다.

- 특정 연구자의 성과물과 참여한 프로젝트 목록 추출
- 추출된 성과물과 프로젝트에 참여한 연구자 목록 추출
- 특정 연구자와 멘토/멘티 관계의 연구자 목록 추출
- 특정 연구자와 공저자, 공동연구자, 멘토/멘티 관계가 포함된 연구자 네트워크 작성

5.4 연구분석 서비스

연구성과 분석 서비스는 연구자 네트워크, 학문별 관련도, 분야별/연도별 성과 통계, 기관의 성과물 통계와 인력 통계 정보를 제공한다.

그림 11은 분야, 이름, 기관, 지역 등의 조건을 통해 연구자를 검색하고, 그 결과로 연구자와 연구자별 전문도, 소속 기관 및 직위를 보여준다. 연구자 네트워크는 공동연구와 멘토/멘티 관계를 포함하는 관계와 멘토/멘티만 보여주는 관계를 구분하여 보여준다. 그림 12는 특정 연구자의 네트워크를 선택한 경우 공동논문의 건수와 공동 연구자, 멘토/멘티의 관계를 가지는 연구자 네트워크를 보여준다. 그림 13은 특정 분야를 선택한 후 확률적 주제 온톨로지를 사용하여 특정 분야와 관련 분야, 관련 분야별 관련도를 확인할 수 있다.

그림 14는 분야, 성과물의 연도, 성과물 유형의 조건에 적합한 성과물의 분야별 성과 통계정보를 제공한다.

분야별 전문가/네트워크

① 분야찾기

- ☐ 인문학
- ☐ 사회과학
- ☐ 자연과학
- ☐ 공학
- ☐ 의학학
- ☐ 농수해양
- ☐ 예술체육
- ☐ 복합학

이름

이성연

기관종류

전라기관

기관명

② 찾기

지역

서울

③ 검색

④ 다시쓰기

이름-이성연

총 15건

분야명

이름 : 전동도

기관명 :

직책 :

지역 : 네트워크

가족관계

이성연

14.23

목포대학교

교수

전남

네트워크

연도/전력

간호학사

이성연

0.0

중앙대학교

기타

서울

네트워크

연도/전력

수의내과학

이성연

3.27

충남대학교

POST-DOC

대전

네트워크

연도/전력

수학교육학

이성연

0.0

서울대학교

회사원

서울

네트워크

연도/전력

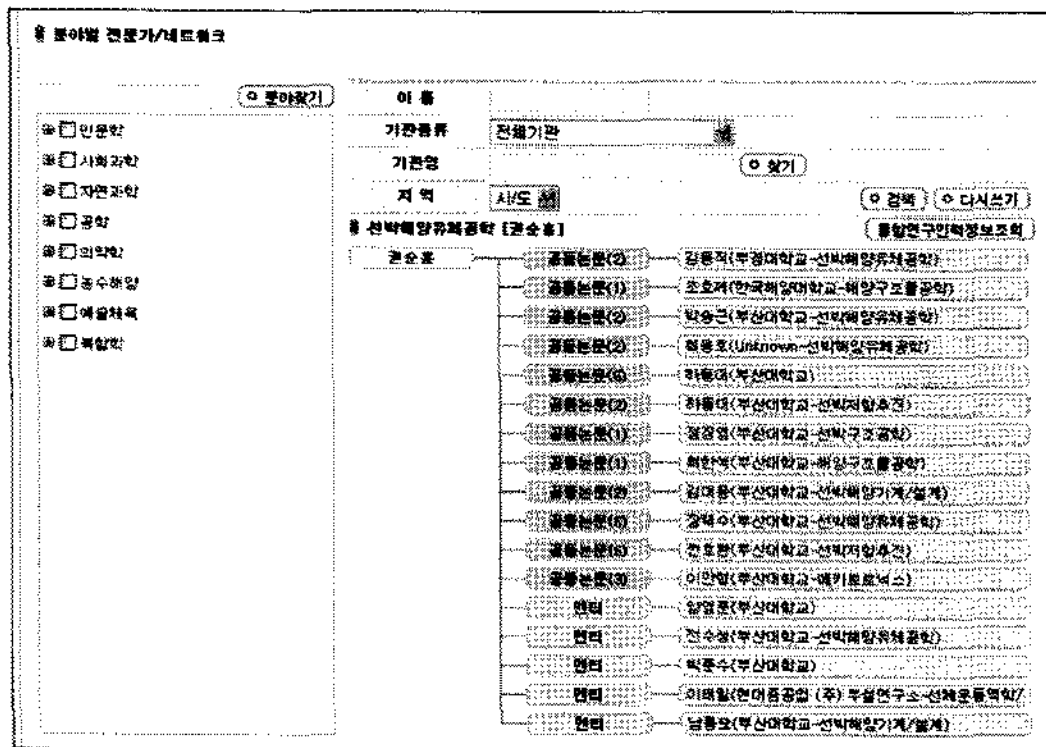


그림 12 연구자 네트워크

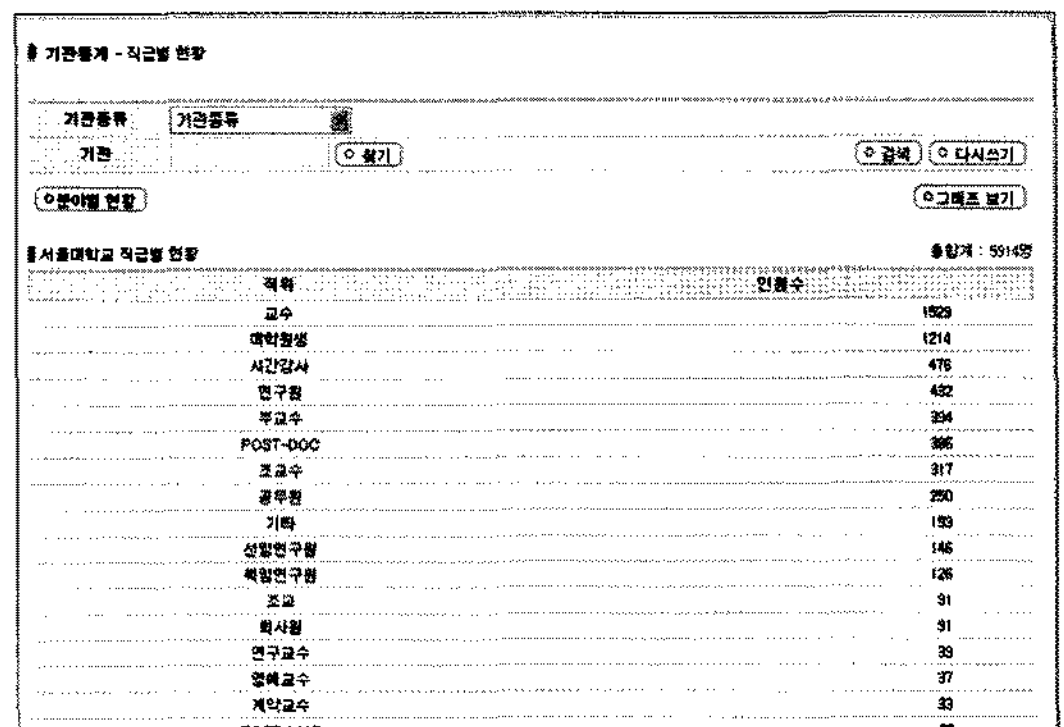


그림 15 기관의 직급별 통계

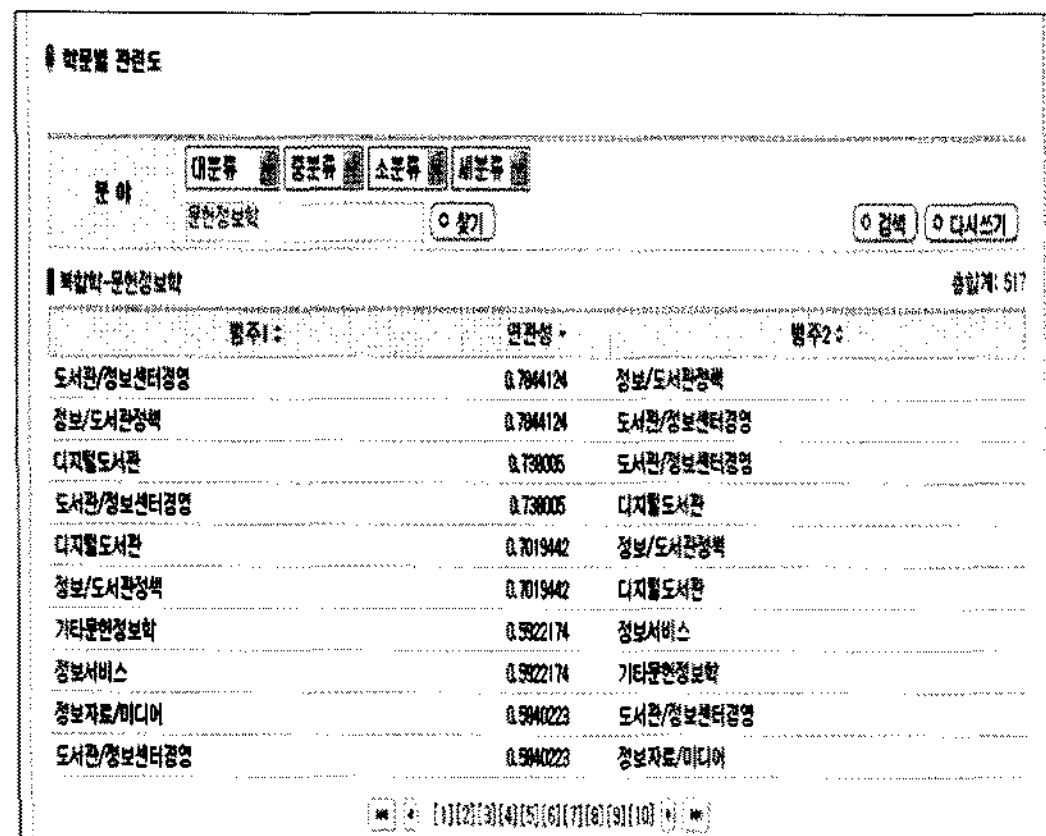


그림 13 학문별 관련도

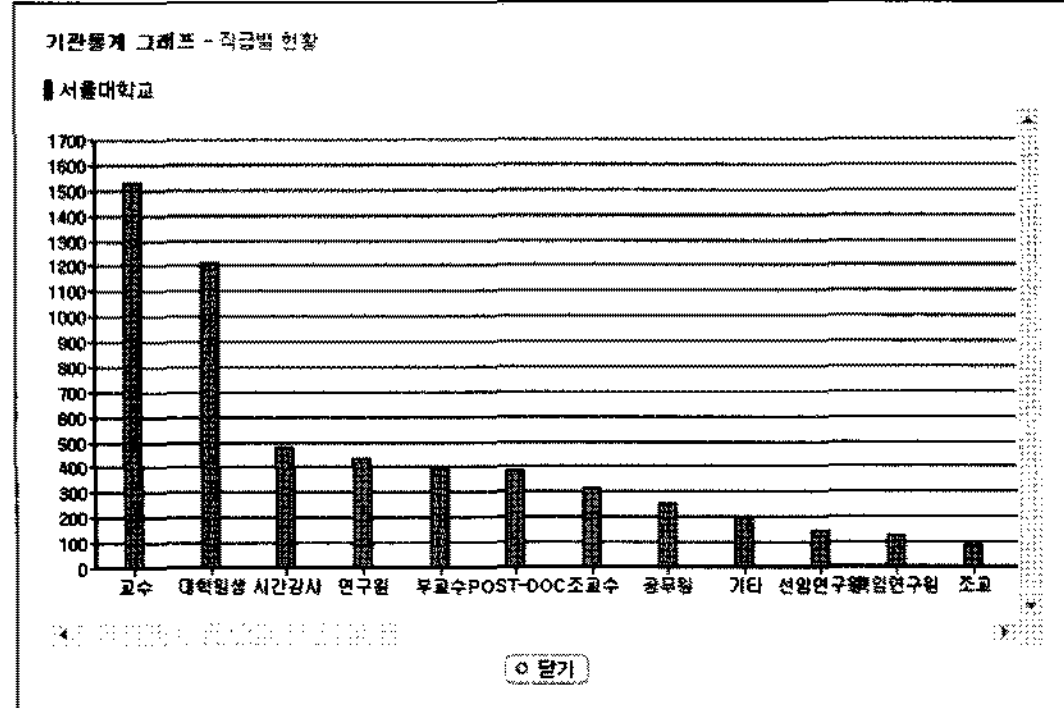


그림 16 기관통계 그래프

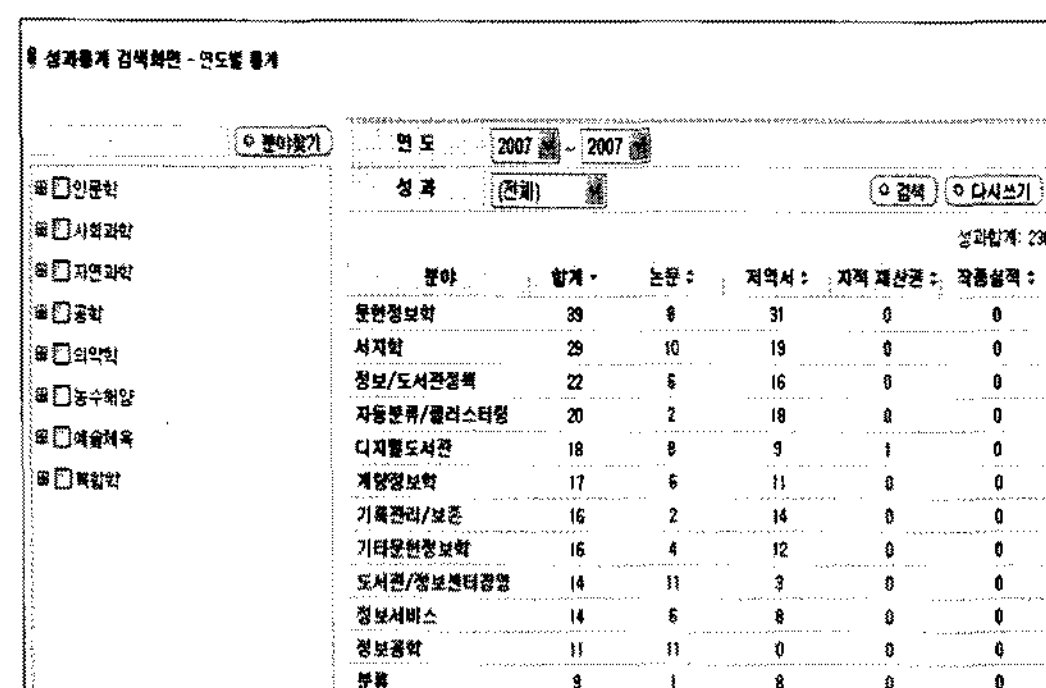


그림 14 성과물 통계

그림 15는 특정 기관을 중심으로 직급별, 분야별 현황 및 박사급 전문가 분포 현황 정보를 기관별로 제공하며, 그림 16은 기관별 직급 현황에 대한 통계 정보가 막대 그래프 형태로 표시되는 것을 보여준다.

심사자 풀의 확대는 확률적 주제 온톨로지의 결과를 그대로 이용하였고, 연구 성과통계 정보는 학진의 성과

관리 DB와의 정합성 체크를 통해 추론을 통한 결과를 검증하였다[3]. 따라서 심사자 풀 확대의 정확성과 연구 성과 정보의 정확성 실험은 별도로 수행하지 않았다.

5.5 서비스 응답 시간

학술정보 분석 서비스를 구성하는 심사자 추천 서비스와 연구성과 분석 서비스 중 심사자 추천 서비스는 결과를 생성하는데 많은 정보를 사용하고, 그 생성 방법이 복잡해서 비교적 많은 시간이 소요되고 있다. 이 문제를 해결하기 위해 심사자 추천 서비스에 ECPV를 적용하여 소요 시간을 단축하려는 실험을 하였다. 반면에 연구성과 분석 서비스는 평균 1초 미만의 응답 속도(수십 ms)를 보이기 때문에 응답 속도에 대한 추가 실험을 실시하지 않았다.

표 5 ECPV를 적용한 심사자 추천 응답 속도(단위 :ms)

구분		1회	2회	평균
without ECPV	getResearcherGraph	9969	9969	10321
	getResearcherList	13891	13421	13528
with ECPV	getResearcherGraph	3579	3438	4253
	getResearcherList	8890	8547	8903

심사자 추천 서비스의 경우 확률적 주제 온톨로지의 적용 여부와 구간별 심사자의 수, 심사자의 분야별 전문도, 연구자 네트워크를 통한 심사자 배제 등의 정보를 생성하기 때문에 서비스 응답에 많은 시간이 걸린다. 표 5는 심사자 추천 서비스에 사용되는 2개 함수의 API별 응답 속도를 ECPV를 적용한 경우와 적용하지 않은 경우로 구분하여 5회의 실험을 수행하였고, 그 중 2회의 실험 결과와 평균 값을 보여주고 있다. ECPV를 적용한 경우 40~50% 정도의 서비스 응답 시간이 단축되고 있음을 알 수 있다.

6. 결 론

학술정보 분석 서비스를 제공하기 위해서 학진에서는 OntoFrame 기반의 학술정보 시스템을 구축하였다. 학진의 연구자와 연구분야, 성과물 등의 정보가 학술정보 온톨로지로 표현되었고, 리스너를 통해 RDF 트리플을 생성하였고, 추론 시스템을 통해 추론 서비스를 제공하였다.

학술정보 분석 서비스는 심사자 추천 서비스와 연구성과 분석 서비스로 구성되며, 심사자 추천 과정에서는 확률적 주제 온톨로지를 이용하여 심사자의 심사 분야 확장을 통해 심사자 풀 확대, 연구자의 연구성과에 기반하여 전문도 부여, 피심사자와 심사자의 연구실적 비교를 통해 관련도 판별 등 심사자 추천에 필요한 정보를 제공한다. 연구성과 분석 과정에서는 연구자의 전공 분야를 이용해서 성과물의 분야 할당, 성과물 비교를 통해 연구자 네트워크 생성, 분야별 전문가 선정, 연도별 성과통계 정보를 제공한다.

학술정보 분석 서비스를 위해 구축된 RDF 트리플의 수는 약 2,360만개이며, 모든 정보 자원은 URI를 통해 식별이 된다. 학술정보 온톨로지와 RDF 트리플은 DBMS 기반의 추론 시스템을 통해 저장하였고, SPARQL 질의를 통해 추론 시스템에 저장된 트리플을 대상으로 추론 결과를 생성할 수 있었다. RDF 트리플의 저장소는 클래스와 속성들 각각이 별도의 테이블로 생성되며, 데이터타입 속성의 경우 데이터형에 적합한 테이블을 생성하여 저장하고, URI는 축약형을 사용함으로써 저장공간의 성능 저하를 방지하였다.

학술정보 분석 서비스를 위해 총 14개의 추론 서비스 API가 개발되었으며, 추론 서비스 API에서는 각각의 추론에 적합한 SPARQL 질의문을 파싱하여 SQL 변환 및 SQL 수행을 통해 결과를 생성하고, 결과를 가공한 후 XML 형태의 결과를 웹 서비스를 통해 반환하는 역할을 수행한다.

OntoFrame 기반의 학술정보 분석 서비스를 통해서 심사자 추천 과정에서는 기존의 수작업에 의존하던 심

사자 선정 과정을 연구자의 전공과 전문도, 피심사자와의 관계를 고려한 공정하고 객관적인 심사자 선정이 가능하게 되었다. 또한 연구성과 분석 서비스는 기존 시스템을 통해 획득하기 어려웠던 연구자 네트워크와 분야별 전문가 및 다양한 연구 통계 정보를 제공함으로써 연구 현황 파악을 보다 효과적으로 지원할 수 있었다.

참 고 문 헌

- [1] 성원경, 정한민, 박동인, "OntoFrame-K: 협업 연구 지원을 위한 시맨틱 웹 기반 지식정보 공유유통 플랫폼", 정보과학회지, 제24권, 제4호, pp. 65-74, 2006.
- [2] 정한민, 이미경, 성원경, 박동인, "OntoFrame-K: 연구자간 협업 지원 서비스를 위한 시맨틱 웹 기반 정보유통 플랫폼", 2006 한국컴퓨터종합학술대회 발표논문집, 2006.
- [3] 이정연, 이재운, 정한민, 강인수, 신숙경, "확률적 온톨로지와 연구자 네트워크를 이용한 심사자 자동 추천에 관한 연구", 제24권, 제3호, pp. 43-65, 2007.
- [4] OWL(Web Ontology Language), "http://www.w3.org/TR/owl-features"
- [5] RDF (Resource Description Framework), "http://www.w3.org/RDF"
- [6] 강인수, 정한민, 이승우, 김평, 성원경, "국가과학기술 R&D 기반정보 온톨로지와 추론 모델링", 2006 한국컴퓨터종합학술대회 발표논문집, 2006.
- [7] 김평, 이승우, 이미경, 구남양, 강인수, 정한민, 성원경, "OntoStore-K: URI 기반 성과 관리 시스템", 2006 한국컴퓨터종합학술대회, 2006.
- [8] Theoharis, Y., V. Christophides and G. Karvounarakis, "Benchmarking Database Representations of RDF/S Stores," in Proceedings of ISWC2005 (LNCS3729), pp. 685-701, 2005.
- [9] 김평, 이승우, 성원경, "OntoFrame의 추론 시스템 개선", 2007 한국컴퓨터종합학술대회, 2007.
- [10] 이승우, 정한민, 성원경, "R-DBMS기반 추론 서비스인 OntoThink-K에서의 SPARQL 질의 지원", 한국정보과학회 2006 추계학술대회, 2006.
- [11] 이승우, 김평, 성원경, "확장 클래스-속성 뷰를 이용한 추론시스템의 응답 속도 개선", 2007 한국콘텐츠학회 춘계종합학술대회논문집, 2007.



김 평

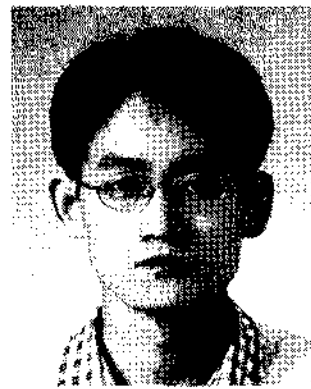
1997년 충남대학교 전산학과(학사). 1999년 충남대학교 컴퓨터학과(석사). 2000년~2003년 (주)엔퀘스트테크놀로지 개발실장. 2003년~2004년 충남대학교 외래강사. 2004년 충남대학교 컴퓨터학과(박사). 2004년~현재 한국과학기술정보연구원 선임연구원. 관심분야는 시맨틱 웹, 정보검색, 자연어 처리



이 승 우

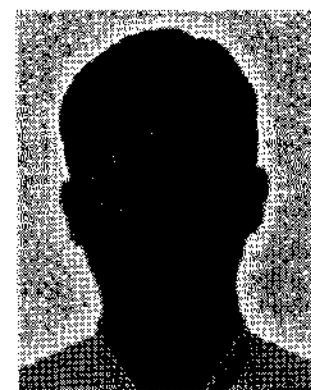
1997년 경북대학교 컴퓨터공학과(공학사). 1999년 포항공과대학교 컴퓨터공학과(공학석사). 1999년~2000년 포항공과대학교 정보통신연구소 연구원. 2005년 포항공과대학교 컴퓨터공학과(공학박사). 2005년~2006년 대구가톨릭대학교 컴퓨터교육과 강의전담교원. 2006년~현재 한국과학기술정보연구원 선임연구원. 관심분야는 시맨틱 웹, 정보검색, 자연어 처리

이스텍 연구개발본부장/상무이사. 2004년~현재 한국과학기술정보연구원 정보서비스연구팀장/책임연구원. 2004년~현재 과학기술연합대학원대학교 겸임교수. 관심분야는 시맨틱 웹, 자연어 처리



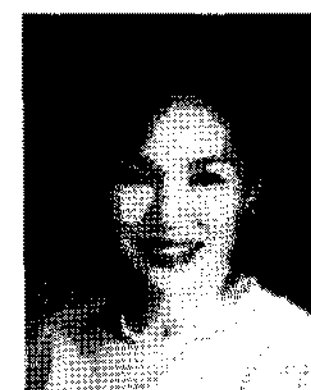
강 인 수

1995년 경북대학교 컴퓨터공학과(공학사). 1999년 포항공과대학교 컴퓨터공학과(공학석사). 2006년 포항공과대학교 컴퓨터공학과(공학박사). 1995년~1997년 (주)포스데이타. 1999년~2001년 포항공과대학교 학술정보원. 2006년~2008년 한국과학기술정보연구원. 2008년~현재 경성대학교 컴퓨터정보학부. 관심분야는 시맨틱 웹, 정보검색, 자연어 처리



정 한 민

1992년 포항공과대학교 전자계산학과(공학사). 1994년 포항공과대학교 전자계산학과(공학석사). 2003년 포항공과대학교 컴퓨터공학과(공학박사). 1994년~2000년 한국전자통신연구원 선임연구원. 2000년~2004년 (주)다이렉스트 연구소장/기술이사. 2004년~2006년 현재 한국과학기술정보연구원 책임연구원. 2004년~현재 과학기술연합대학원대학교 겸임교수. 관심분야는 시맨틱 웹, 정보검색, 정보추출, 자연어 처리



이 정 연

1993년 경기대학교 문헌정보학과(도서관학사). 1995년 이화여자대학교 문헌정보학과(문학석사). 2005년 이화여자대학교 문헌정보학과(문헌정보학박사). 1995년~1997년 대우자동차 기술연구소. 2000년~2003년 Elsevier Science. 2006년~2008년 한국학술진흥재단 지식정보센터. 2008년~현재 이화여자대학교 사회과학연구소. 관심분야는 정보서비스, 정보검색, information literacy



성 원 경

1987년 연세대학교 불어불문학과(학사) 1989년 연세대학교 불어불문학과(석사) 1996년 프랑스 파리7대학교 언어학과(박사). 1997년~1998년 한국전자통신연구원 Post-doc. 1998년~2001년 L&H Korea (주) 책임연구원. 2001년~2003년 (주)보