

# URI를 이용한 개체 중심적 통합 검색 시스템

## (An Entity-centric Integrated Search System Using URI)

정 한 민 <sup>†</sup>      이 미 경 <sup>\*\*</sup>      성 원 경 <sup>†</sup>  
 (Hanmin Jung)    (Mi-Kyoung Lee)    (Won-Kyung Sung)

**요 약** 본 연구는 키워드 기반 통합 검색의 한계를 극복하고자, 인스턴스를 등록하고 관리하는 URI 서버를 이용하여 개체 페이지를 구성하는 방식의 통합 검색 방안을 제안한다. 키워드로 구성된 사용자 질의어와 매칭된 URI 서버 내의 인스턴스를 분석하여 최적 개체를 선정하고 단위 서비스들을 동시에 호출하는 방식으로 개체 페이지를 구성한다. 또한, 논문으로부터 자동 추출된 주제 정보를 대상으로 추론을 수행함으로써 인물, 기관, 위치 등에 대해서도 주제 중심의 심층적 정보 제공이 가능하다. 해외에서 실 서비스되고 있는 Citeseer, Google Scholar와의 통합 검색 결과 비교 실험과 사용성 평가를 통해 본 연구의 효용성을 실증한다.

**키워드** : URI, 통합 검색, 개체, 사용성 평가

**Abstract** To overcome the limitation of keyword-based integrated search, this study shows entity-centric integrated search method using URI scheme. Our system generates entity pages in ways of analyzing user's keyword and instances matched with it, selecting optimal entity type, and calling unit services simultaneously. Topic information extracted from articles is propagated to persons, institutions, and locations by reasoning for providing topic-centric information. With comparative experiments based on search results and usability tests, we proved that this approach is superior to keyword-based integrated search served by CiteSeer and Google Scholar.

**Key words** : URI, Integrated Search, Entity, Usability Test

### 1. 서 론

매년 급격히 증가하고 있는 정보의 홍수 속에서 원하는 정보를 찾거나, 정확한 정보만을 찾거나, 또는 메타 데이터 이상의 정보를 얻고자 하는 욕구를 만족시키는 것이 갈수록 어려워지고 있다. 단적인 예로, 'neural

network'로 Google을 검색할 경우 약 6천만 건의 웹 페이지가 검색되며, Google Scholar에서도 90여 만 건의 학술 정보가 검색된다. [1]에서는 이러한 정보 검색 환경의 변화에 따라 기존의 불연속적 검색(Discrete Find)을 통한 기본적 접근 또는 단순 통합 검색이 합리적 결과를 만들 수 있도록 온톨로지 등을 이용하여 정보를 분류하고 모델링할 수 있고, 더 나아가 문맥적 내비게이션(Contextual Navigation)을 통해 특정 정보 요구에 적용할 수 있는 정보 접근으로 발전할 것으로 보고 있다. 이러한 발전은 궁극적으로 본 연구가 추구하는 방향인 검색과 내비게이션의 자동화로 이어질 것이다.

정보 검색의 효율성을 제고하고자 네이버를 비롯한 일부 상용 포털에서 인물, 기관 등의 개체(Entity)를 중심으로 개체 페이지<sup>1)</sup>를 구성하고 있으며, 비교적 효과적인 검색 결과를 제시하고 있다. 예를 들어, '이선희'의 경우 사용자 질의어에 직업을 결합함으로써 어느 정도 세분화된 개체 페이지를 제공할 수 있는 데, '가수 이선희', '골퍼 이선희', '공무원 이선희' 등이 그 예이다. 그

· 관련 특허: 11/905957(미국출원), PCT/KR2007/004887(PCT출원), 10-2007-0176312(국내출원)

· 이 논문은 제19회 한글 및 한국어 정보처리 학술대회에서 'URI를 이용한 개체 중심적 통합 검색 시스템'의 제목으로 발표된 논문을 확장한 것임

† 정 회 원 : KISTI 정보서비스연구팀 책임연구원  
 jhm@postech.ac.kr  
 wksung@kisti.re.kr

\*\* 정 회 원 : KISTI 정보서비스연구팀 연구원  
 jerryis@kisti.re.kr

논문접수 : 2008년 1월 3일

심사완료 : 2008년 5월 28일

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)

1) 단위 검색 결과를 단순히 나열한 통합 검색 결과 페이지와 달리 특정 개체와 관련된 정보들만으로 구성하는 검색 결과 페이지로 정의할 수 있다.

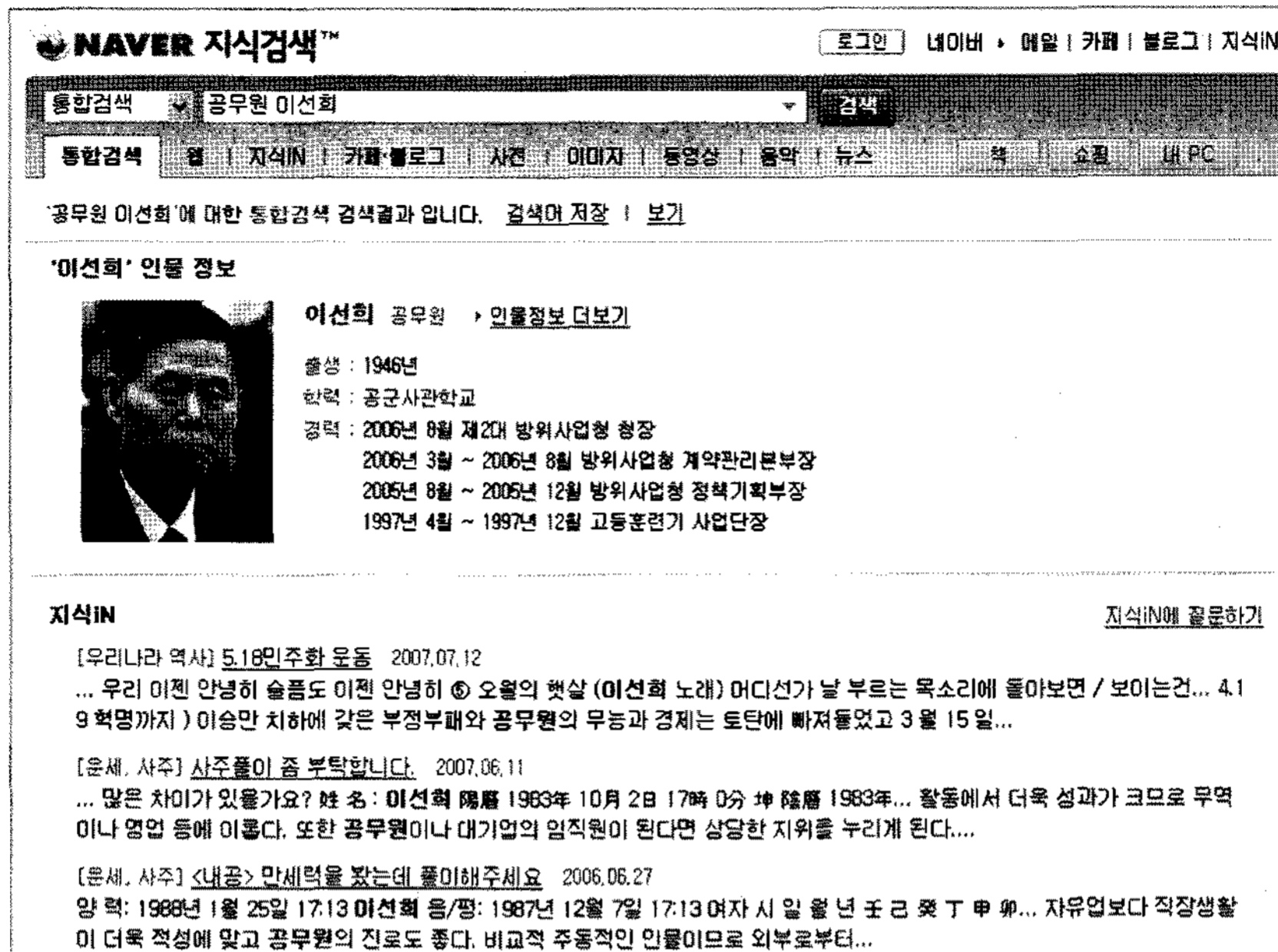


그림 1 네이버 인물 검색 결과 예 ('공무원 이선희' 검색 결과 내에 다른 인물 정보가 혼재되어 있음)

렇지만, 동일 직업을 가지는 경우 동명이인 구별에 한계를 가질 수밖에 없는 키워드 휴리스틱이어서, 문서 내에 다른 인물 정보가 혼재되어 있는 경우 엉뚱한 인물에 대한 문서가 해당 인물 검색 결과에 포함될 수 있다는 문제점을 안고 있다(그림 1 참조). Yahoo Mindset<sup>2)</sup>의 경우 Research와 Shopping 주제 중 더 강조하고 싶은 쪽으로 슬라이드 바를 움직여 검색 결과를 제어할 수 있도록 하고 있지만, 이 역시 질의어에 부가적인 제약줄 수 있는 키워드를 주는 방식으로 보이며, 사용자가 원하는 바를 정확히 선택하기 보다는 선호도의 정도를 제공할 수 있는 수준에 그치는 한계가 있다. 이러한 현실에 비추어 볼 때 식별 체계를 반영할 수 있도록 하는 통합 검색 서비스가 진화된 개체 중심적 통합 검색이 될 것으로 예측된다.

차세대 웹의 한 축인 시맨틱 웹은 URI(uniform Resource Identifier)라는 식별 체계를 그 구조의 기반으로 채택함으로써 정보간 변별 기재를 제공하고 있다. 그렇지만, 아직까지 실제 서비스에 시맨틱 웹 기술을 활용하는 시도가 많지 않은데 이는 기존 텍스트 기반 정보를 의미화하는 데 많은 노력이 필요하기 때문이다. 즉, 텍스트 마이닝, 언어 분석 등 다양한 자연어 처리 기술과 시맨틱 프로세스를 포함한 시맨틱 웹 기술 기반 프레임워크를 제대로 확보할 수 있어야 함을 기존 전제로 한다.

2005년부터 연구되고 개발된 OntoFrame은 국내 과

학 기술 기반 정보를 연구자의 연구 개발 전주기에서 제공할 수 있도록 하는 시맨틱 웹 기술 기반 프레임워크이다[2]. 본 연구는 2007년부터 Citeseer Open Access Metadata(<http://citeseer.ist.psu.edu/oai.html>)로 서비스 범위를 확장하고 시맨틱 기술에 기반을 둔 통합 검색 서비스를 제공하기 위한 시도의 하나로 개발된 URI 서버 기반 통합 검색 시스템에 대해 서술한다.

본 연구의 목적은 사용자 질의어와 URI 부여된 인스턴스간 매칭을 통해 최적 개체(Entity)를 발견하고 검색 결과로서 개체 페이지를 구성하여 사용자에게 제공함으로써 식별 체계에 기반을 둔 정교한 검색 서비스를 실현하는 데 있다. 3장과 4장을 통해 URI 서버 기반 통합 검색 시스템에 대한 서술과 해외 학술 정보 검색 서비스와의 비교 결과를 제시함으로써 본 연구의 효용성을 실증하고자 한다.

## 2. 관련 연구

최근 들어 시맨틱 기술과 시맨틱 데이터를 이용하여 통합 검색의 정확도를 높이려는 연구가 많이 진행되고 있다. 시맨틱 웹 기반 통합 검색 연구로는 [3-10] 등이 있다.

[3]은 온톨로지를 이용하여 질의어에 대한 유사어 집합을 생성하고 용어 간의 관계성으로 웹 페이지 검색 순위를 측정하는 방법을 제시하고 있다. 그렇지만, 메타 검색 방식으로 이루어지는 문자열 기반 검색이어서 키워드 기반 검색 시스템이 가지는 성능 한계를 그대로

2) <http://mindset.research.yahoo.com/>

가진다. 시맨틱 웹을 위한 검색 및 메타데이터 엔진인 Swoogle은 문자 기반의 N-Gram과 URI 표현 형식인 URIfref를 색인과 검색에 이용함으로써 질의어에 시맨틱 요소를 가미하는 방식을 제안하였다[5]. 그렇지만, 웹 상에서 웹 로봇을 이용한 수집, 추출, 색인, 검색이기 때문에 식별 체계를 본격적으로 도입하는 데 어려움을 가질 수밖에 없다. [4]는 원문에 대해 온톨로지 기반의 색인을 수행함으로써 의미 기반 검색 결과를 클래스 별로 제공하는 기법을 제시한다. 사용자 질의어와 온톨로지 상에서 매칭되는 검색 결과를 제시하기 위해 고유한 색인 구조를 제시한다는 점에서 타 연구와 구별되지만, URI가 활용되지 못하는 부분과 추론을 이용한 지식 확장 없이 웹 페이지 검색에만 연구 범위를 한정된 부분이 한계로 지적될 수 있다. 'Spread Activation'을 사용하여 관련 개념을 찾거나 검색 결과를 필터링하는 기법을 제안한 [10] 역시 온톨로지로 표현된 내부 정보를 이용하여 의미 기반 검색 결과를 제시한다. 그렇지만, 인스턴스 그래프 상의 가중치가 사용자에게 직관적으로 제시되지 못하여 TF\*IDF 검색 모델과 같이 검색 순위를 설명하지 못하는 경우가 발생할 수 있으며, 온톨로지가 데이터 증가에 따라 큰 부하로 작용할 수 있다는 문제점을 가진다.

또 다른 방향의 통합 검색 연구로 메타데이터 표준화 또는 시맨틱 스키마를 통한 데이터 통합을 제시한 [11-13] 등이 있는데, 본 연구 범위와는 다소 거리가 있으며 검색 방식 또한 키워드 기반 검색에 머무르고 있다. 특히, 다양한 포맷이나 언어 등의 이질적(Heterogeneous) 데이터 통합은 변화에 대한 지속적인 대응이나 스키마간 매핑의 어려움으로 인해 효율적으로 작동하지 못할 가능성이 크므로, 본 연구와 같이 단위 서비스 구성기에 각 분산 데이터 처리를 맡기는 방식이 대안으로 제시될 수 있을 것이다. 다만, [13]의 경우 D-Lib Magazine의 사례를 들어 제시한 질의어 선택을 위한 시각화 개념은 본 연구에서의 URI 서버 내 인스턴스 브라우징을 보조적으로 지원하기 위한 수단으로 사용될 수 있을 것이다.

본 연구가 초점을 맞추고 있는 개체 기반 검색과 관련한 흥미로운 연구로 [7,8]이 있다. 검색 유형을 'Navigational Search'와 'Research Search'로 구분하고 인물과 같은 특정 개체를 검색 대상으로 정의한 [7]은 사용자 의도를 반영할 수 있는 명시적 기호와 현재 네이머에서 사용하고 있는 휴리스틱을 제안했다는 점에서 다른 연구와 차별화된다. 그렇지만, 식별 체계 기반 시맨틱 프레임워크 상에서의 데이터 처리 방안을 제시하지 못함으로써 실제 시스템으로 발전하는 데 있어 한계를 가질 수밖에 없다. [8]은 좀더 진보된 형태의 개체

중심적(Entity-centric) 연구로서 검색 결과로 구조화된 개체 기술(Description)을 가지는 검색 모델을 제안한다. 사용자가 인스턴스, 개체, 관계(Property)를 제약해 나가는 방식으로 질의할 수 있도록 한 진보적인 특성을 가지고 있으나, 개체간 충돌에 대한 지능적 해소나 다양한 단위 서비스 활용 부재는 검색 서비스 범위를 제약시킬 수밖에 없다.

### 3. URI 서버 기반 통합 검색 시스템

본 연구에서는 주제(Topic), 인물(Person), 출처(Source) 등 개체<sup>3)</sup>에 속한 인스턴스(Instance)가 사용자 질의와 매칭되는 방식으로 통합 검색이 이루어지는데, URI 서버는 인스턴스를 등록하고 관리하는 역할을 수행한다.

#### 3.1 URI 서버

URI 서버는 온톨로지 스키마에 따라 데이터베이스를 자동 생성하고 온톨로지 인스턴스를 저장·관리하는 기능을 담당한다. URI 서버는 클래스 관리기, RDF 트리플 생성기로 구성되는데, 전자는 읽어 들인 온톨로지 스키마 내에 정의된 클래스 각각이 가지는 속성(Property)에 대한 상세 정보를 설정한다. 이 정보에는 데이터베이스 스키마 자동 생성을 위한 필드 설정 정보, URI 검증을 위한 URI 생성 정보, 인스턴스 생성 관리기를 위한 사용자 인터페이스 생성 정보, RDF 트리플 생성을 위한 트리플 생성 정보 등이 포함된다(그림 2 참조).

예를 들어 이 정보들을 서술하면 다음과 같다. Person 클래스에는 'hasInstitutionOfPerson'이라는 개체 관계 속성(Object Property)이 있는데, 필드 설정 정보는 필드명과 DBMS 데이터타입으로 '현재대표소속'과 'String'이 정의되어 있다. 'Person' URI 생성 정보는 문자 20자로 구성된 URI를 가져야 한다는 제약에 따라 'PER\_char(20)'로 정의되어 있다. 상기 개체 관계 속성에 대한 사용자 인터페이스 정보는 'Object'로 정의되는데, 이는 인스턴스 생성 관리기에서 해당 속성이 가지는 'Range'가 클래스이어야 하며 이에 따라 클래스 내 인스턴스를 선택할 수 있는 UI를 제공해야 함을 의미한다. 또한, 해당 개체 관계 속성의 트리플 생성 정보는 '? UID \$Person http://www.kisti.re.kr/isrl/ResearchRefOntology#hasInstitutionOfPerson \$hasInstitutionOfPerson'과 같이 정의되어 있어 '\$Person'과 '\$hasInstitutionOfPerson'에 해당하는 인스턴스를 획득하여 'http://www.kisti.re.kr/isrl/ResearchRefOntology#hasInstitutionOfPerson'이 포함된 트리플을 생성해야 함을 서술하고 있다.

3) 본 연구에서 개체는 온톨로지 스키마에 정의된 클래스 중 검색 결과 페이지 구성에 이용되는 클래스로 한정하여 정의한다.

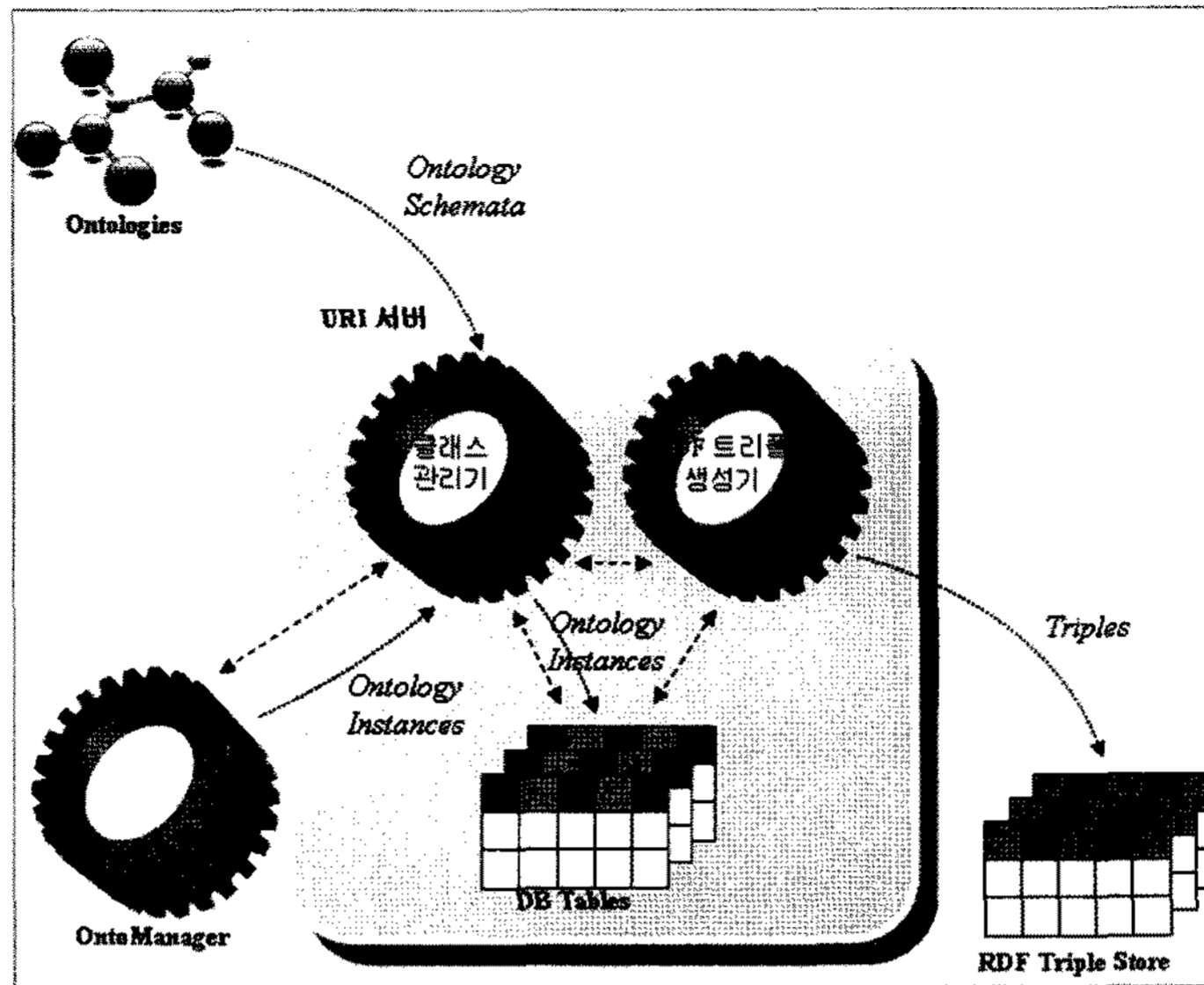


그림 2 URI 서버 구성도

클래스 관리기 설정 완료 후 자동 생성되는 데이터베이스에는 인스턴스 생성 관리기를 통해 생성된 온톨로지 인스턴스들이 저장된다. RDF 트리플 생성기는 이들을 RDF 트리플 형식으로 변환하여 추론 시스템이 직접 사용할 수 있는 지식으로서 제공한다. 온톨로지 인스턴스들과 트리플들은 모두 DBMS 기반 저장소에 저장되지만, DB 스키마에 있어서 차이가 존재한다. RDF 트리플들은 기본적으로 SPO(Subject-Predicate-Object) 형식의 단순한 DB 스키마를 가짐으로써 분산된 트리플들 간의 통합과 외부 시스템에의 지식 제공을 용이하게 해준다.

[2,14] 등의 연구를 통해 이미 2006년에 구현된 URI 서버는 온톨로지 변경과 인스턴스 생성 관리기와의 연동을 고려하지 못한 구조여서 이식성(Portability)이 다소 떨어지는 단점이 있다. 본 연구에서는 통합 검색 시스템과의 원활한 연동과 다중 온톨로지 체계에 대응하기 위해 상기에서 기술한 바와 같이 범용적인 구조로 변경하여 사용한다.

**3.2 시스템 구성**

본 연구에서는 사용자 질의어에 직접 의존하여 검색 결과를 생성하는 키워드 기반 검색 방식을 개선하고자 URI 서버 내의 인스턴스를 선행 참조하는 통합 검색 방식을 제안한다. 즉, 사용자 질의어가 입력되는 경우 먼저 URI 서버를 참조하여 해당 질의어와 일치하는 인스턴스를 획득하고 인스턴스에 부여된 URI를 이용하여 통합 검색 결과를 생성한다.

그림 3은 URI를 참조하여 이루어지는 통합 검색 시

스템의 구성도와 흐름을 보여준다. 먼저 사용자가 입력한 질의어인 키워드는 URI 서버 내에 저장된 인스턴스를 색인한 검색 엔진(IR Engine)에 전달되어 키워드-인스턴스 매칭에 이용된다(그림 3의 ①, ②). 키워드-인스턴스 매칭의 결과는 개체명, 개체 유형, 개체 URI로 구성된다. 동일 개체명을 가지더라도 개체 URI가 다를 수 있는데 동명이인 등이 이에 해당한다. 개체 검색기(Entity Finder)는 매칭 결과에 따라 표 1과 같은 최적 개체 결정 규칙을 이용하여 최적 개체를 선정한다. 예를 들어, 'neural network'이라는 사용자 질의어가 입력된 경우, 검색 엔진의 키워드-인스턴스 매칭 API를 호출함으로써 질의어와 매칭된 개체를 발견한다. 해당 API는 부분 매칭을 허용하지 않으며, 'neural network'의 경우에는 주제 개체로만 등록되어 있기 때문에 주제 개체가 최적 개체로 선정되고, 주제 페이지 구성이 결정된다. 본 통합 검색은 개체 중심으로 이루어지기 때문에 질의어 내의 키워드 순서를 그대로 유지한다. 예를 들어, 'network neural'을 질의어인 경우에는 주제 개체인 'neural network'와 매칭되지 못한다.

표 1의 최적 개체 선정 규칙은 서비스 중요도와 차별화를 고려하여 서비스 설계 시 미리 정의된 것이다. 최적 개체를 선정하는 이유는 다중의 개체 페이지 중 어느 개체 페이지를 우선적으로 보여줄 것인가를 결정하기 위해서이다. 예를 들어, 'Christian Becker'의 경우에는 6개의 서로 다른 URI를 가지는 인물 개체가 존재하며, 'Information Systems'의 경우에는 주제 개체와 출처 개체가 모두 존재한다. 전자는 동일 유형의 개체만

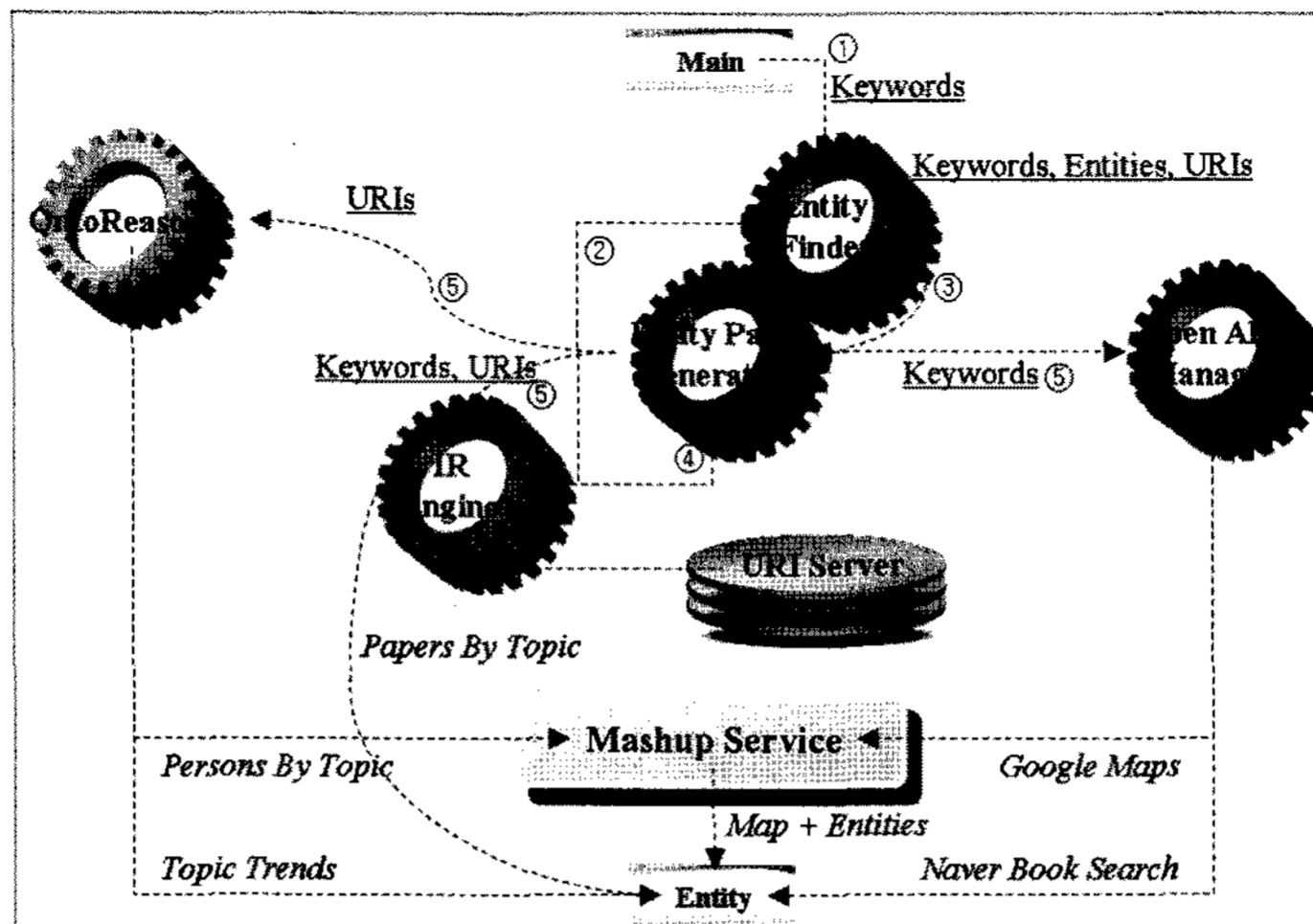


그림 3 URI 기반 통합 검색 시스템 구성도

표 1 최적 개체 선정 규칙

검색결과 \ 개체	주제	인물	논문	출처
실패, 부분 매칭 성공			0	
주제	0			
인물		0		
논문			0	
출처				0
주제, (인물 OR 논문 OR 출처)	0			
인물, (논문 OR 출처)		0		
출처, 논문				0

매칭되었으므로 URI 값이 작은 순 또는 성과가 많은 순으로 정렬하여 해당되는 개체를 먼저 보여주면 되나, 후자의 경우에는 주제와 출처 중 사용자에게 우선적으로 제공될 개체를 선정할 필요가 있다. 매칭에 실패하거나 부분 매칭만 성공한 경우에는 논문(검색 결과) 페이지가 구성되는데, 이는 Citeseer, Google Scholar 등 일반 학술 정보 검색 서비스와 유사한 결과를 생성하기 위함이다.

개체 페이지 생성기(Entity Page Generator)는 개체 페이지를 생성하기 위해 사용자 질의어와 매칭된 개체명, 해당 개체 URI를 각 단위 서비스 구성기에 전달한다(그림 3의 ⑤). 일반적인 통합 검색과 마찬가지로 각 단위 서비스는 미리 정의된 개체 페이지 내 단위 서비스를 구성하기 위해 AJAX(Asynchronous JavaScript and XML) 방식으로 호출된다. 단위 서비스는 통합 검색 결과 페이지를 구성하기 위해 호출되는 독립된 서비스 모듈로 정의할 수 있는데, 본 연구에서는 단위 서비스 구성기에는 추론 시스템(OntoReasoner), 검색 엔진(IR Engine), Open API 관리자(Open API Manager)

를 포함한다[15]. 추론 시스템은 KISTI에서 2006년 이후 개발 및 개선되고 있는 시스템으로 DBMS를 저장 구조로 채택하고 있으며 RDFS와 OWL 일부를 지원한다. 검색 엔진은 추론 시스템을 통해 지원되기에는 부담이 큰 단순 검색 기능을 특화적으로 처리하기 위해 도입되었으며, 추론 시스템과 통합 검색을 위해 필요한 역할을 분담하고 있다. 각 단위 서비스는 다수의 API를 가지며, 파라미터로 키워드-인스턴스 매칭을 통해 획득된 개체명, 개체 URI 등을 입력받아 동작한다. 단위 서비스의 API 예로는, 주제 페이지의 경우 추론 시스템을 호출하는 'Topic Trends', 'Persons by Topic', 검색 엔진을 호출하는 'Also Try', 'Papers by Topic', Open API 관리기를 호출하는 'Google Maps', 'Naver Book' 등이 있다(그림 4 참조).

매쉬업 서비스(Mashup Service)는 Open API를 통해 얻어진 결과와 다른 단위 서비스에서 얻어진 결과를 결합하는 역할을 한다. 예를 들어, 주제별 전문가('Persons by Topic')나 주제별 기관('Institutions by Topic') 정보는 추론 시스템을 호출하여 획득하며, 이들은 매쉬업 서비스에 의해 Google Maps Open API 호출을 통해 얻어진 지도 위에 출력된다. 결국 사용자에게 보여질 개체 페이지는 각 단위 서비스와 매쉬업 서비스 호출 결과들을 조합하여 구성된다.

그림 4는 주제 페이지를 구성하기 위해 호출된 검색 엔진, 추론 엔진, 외부 API들의 예를 보여준다. 18개의 API가 동시 호출되면서 주제 페이지를 구성한다. 'Papers by Topic' 결과를 얻기 위해 검색 엔진을 호출하는 `getTopicallyPapers(String searchField, String TopicURI, int displayNum, int pageNum)`나 'Persons by Topic'

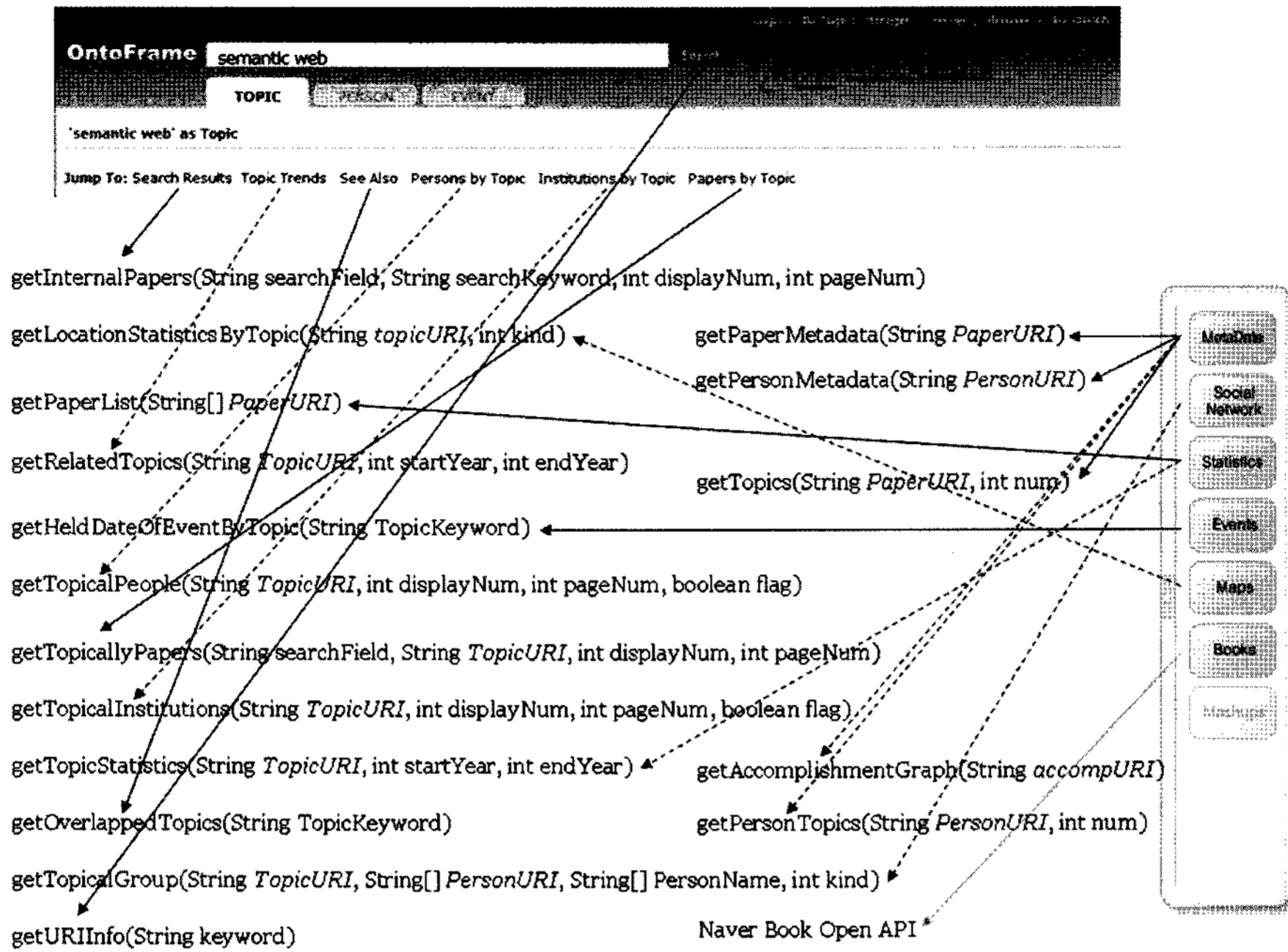


그림 4 주제 페이지 구성을 위한 검색 및 추론 API 호출도(실선: 검색 엔진 호출, 점선: 추론 엔진 호출, 회색 실선: 외부 API 호출)

결과를 얻기 위해 추론 엔진을 호출하는 getTopicalPeople (String TopicURI, int displayNum, int pageNum, boolean flag)은 URI 정보를 이용하여 해당 개체의 정확한 정보를 제공해주며, 'Also Try' 결과를 얻기 위해 검색 엔진을 호출하는 getOverlappedTopics(String TopicKeyword)나 Add-on 서비스의 'Books' 결과를 얻기 위해 호출하는 Naver Book Open API는 키워드만을 이용하여 해당 개체 관련 정보를 제공한다. 상기 API에서 'TopicURI'는 사용자 질의어와 매칭된 주제 개체가 가지는 URI이며, 'TopicKeyword'는 사용자 질의어와 매칭된 주제 개체명이다.

#### 4. 비교 실험

본 연구는 기존 키워드 수준의 통합 검색을 URI 수준의 통합 검색으로 진화시켰다는 데 의의가 있는데, 특히 URI에 기반을 둬으로써 해당 주제를 연구하는 연구자, 해당 주제에 속한 논문 등 정교한 서비스가 실현될 수 있도록 해준다. 4장에서는 현재 많은 국내·외 연구자를 대상으로 서비스 중인 CiteSeer와 Google Scholar와의 비교 실험을 통해 본 연구의 효용성을 입증하고자 한다. 이들은 2007년 4월에 실시된 사용자 인터뷰(각주 참조)를 통해 확인된 가장 많이 사용하고 있는 학술 연구 정보 서비스이며 본 시스템을 포함하여 CiteSeer 데

이타를 공통으로 포함하고 있기 때문에 본 비교 실험을 위해 사용하였다.

본 URI 서버 기반 통합 검색에 사용된 데이터는 CiteSeer Open Access Metadata로부터 수집한 2000년 이후 논문 중 중복 배제되고 주제 부여된 114,337건으로부터 추출되고 가공된 161,853명의 인물, 160,568개의 주제, 17,093개의 기관, GPS 값을 가지고 있는 730,360개의 위치 정보로 구성된다. 추론을 위해 사용된 RDF 트리플은 6,378,572개이다. 검색 엔진과 추론 시스템은 웹 서비스 방식으로 통합 검색 서비스와 연동된다.

본 실험을 위해 [16]에서 사용한 주제 추출 방법을 이용하여 논문에 주제를 최대 5개까지 자동으로 부여하였다. 이 주제 정보는 추론 시스템을 통해 전방 추론 (Forward Chaining) 방식으로 논문 저자(인물)에 전파(propagation)되고, 다시 저작 당시의 기관과 지역으로 전파되는 과정을 거쳐 여러 유형의 개체 인스턴스와 개체 관계 속성(Object Property)으로 연결된다.

비교 실험을 위한 시나리오는 2가지를 설정하였는데<sup>4)</sup>, 하나는 'neural network'라는 주제와 관련된 정보를 찾는 것이며, 다른 하나는 'A. J. Bernheim Brush'

4) 해당 시나리오는 연구자들이 학술 정보 사이트에 접근하여 입력하는 질의어 형태에 대한 사용자 인터뷰 통해 얻어진 내용을 바탕으로 구성하였다. 사용자 인터뷰는 한국과학기술정보연구원(2명)과 전자통신연구원(4명)에 근무 중인 6명을 대상으로 2007년 4월 실시하였다.

The screenshot shows the 'neural network' search results page. The main content area lists several research papers with their titles, authors, and institutions. A 'Details' sidebar on the right provides information about the top result, including its title, authors (Shuya Imajo, Masami Konishi, Tatsushi Nishi, Jun Imai), source, and a list of related topics like 'neural network (25%)', 'robot arm (25%)', 'learning (25%)', and 'trajectory generator (25%)'. A small graph is visible in the sidebar.

This screenshot displays the 'Topic Trends' and 'See Also' sections. The 'Topic Trends' section shows a list of related terms such as 'genetic algorithm', 'hidden layer', 'neural net', 'pattern recognition', 'feature vector', 'artificial neural network', 'recurrent neural network', 'decision tree', 'data point', 'output unit', 'support vector machine', 'artificial intelligence', and 'cost function'. The 'See Also' section lists related concepts like 'neural progenitor', 'neural lineage', 'neural network processor', 'neural mechanisms', 'neural filter', 'immortal neural cells', 'digital neural hardware', 'neural assemblies', 'sigmapisigma neural network', 'memorybased neural network', 'rock fracture network modeling', 'hierarchical ring network', 'generegulatory network', 'deep space network', 'data transport network', 'network cohesion', 'network timeout period', 'dualmode network', 'adhoc network', and 'ethnic trade network'. Below these are sections for 'Persons by Topic' and 'Institutions by Topic', each listing names and affiliations. A 'Papers by Topic' section is also visible at the bottom.

그림 5 본 시스템의 주제 검색 결과

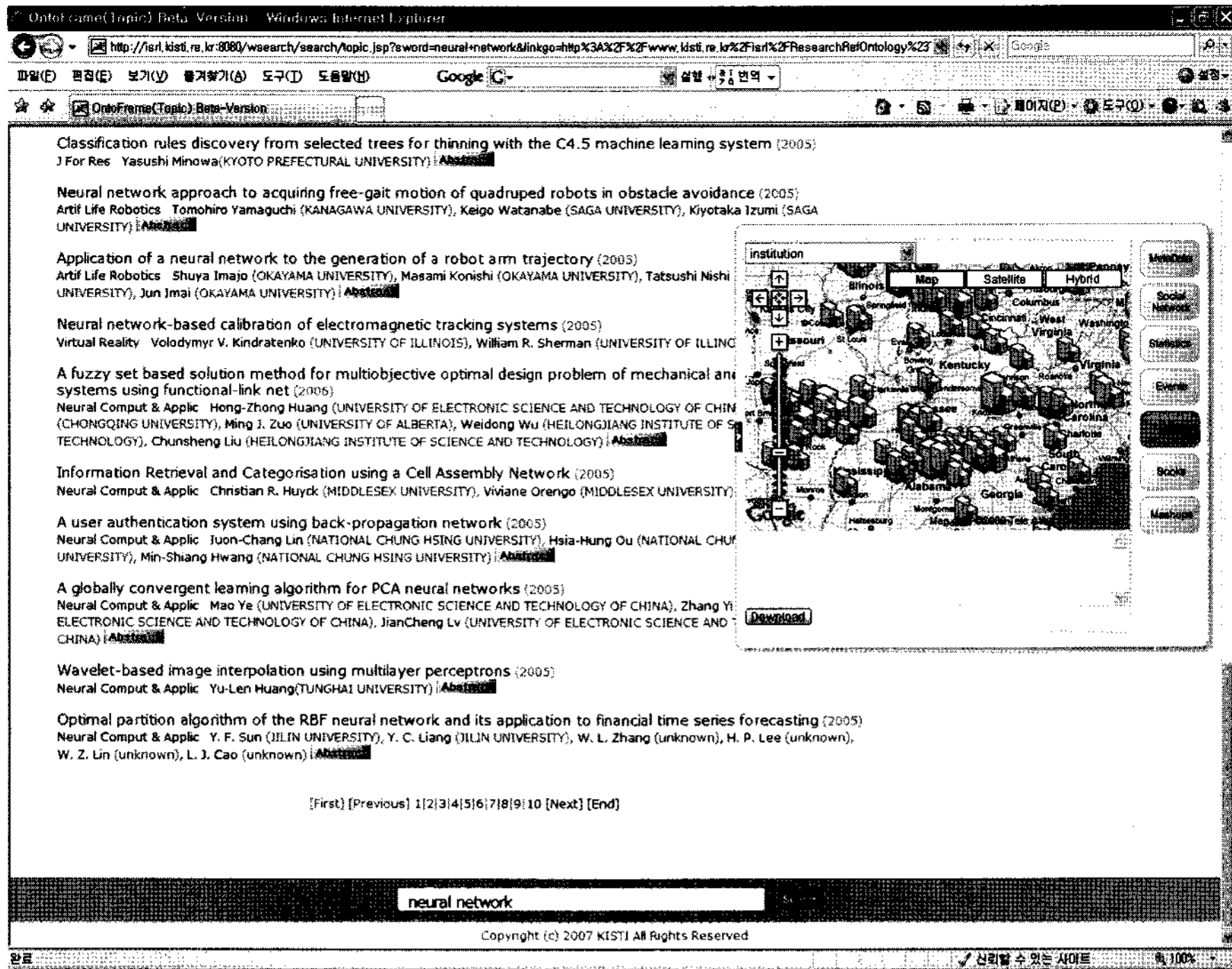


그림 5 본 시스템의 주제 검색 결과 (계속)

라는 인물과 관련된 정보를 찾는 것이다.

시나리오 1: 주제 검색

주제 검색은 일반적으로 연구 동향 파악이나 주제간 관계를 분석하기 위한 목적으로 이루어진다. 사용자는 주제를 제목으로 가지는 단순 키워드 검색 결과부터 해당 주제에 대한 상세 정보까지 여러 단계(Level)의 정보 수준을 요구할 수 있다. 그림 5, 그림 6, 그림 7은 'neural network'로 본 시스템, Citeseer, Google Scholar를 검색한 결과를 보여준다.<sup>5)</sup>

그림 5와 같이 'neural network'라는 질의어로 검색 시 최적 개체 선정 규칙(표 1 참조)에 의해 주제 페이지를 우선적으로 구성한다. 해당 질의어와 함께 'neural network classifier'와 'artificial neural network'가 'neural network'를 포함하는 주제 인스턴스로 URI 서버에 등록되어 있기 때문에 그림 8의 리스트 박스가 구성될 수 있었다. 주제 페이지에서는 해당 주제와 조어적으로 부분 매칭되는 'Also Try', 주어진 기간에 대해 연관 주제들의 추이와 빈도를 보여주는 'Topic Trends', 해당 주제로 분류되는 논문을 보여주는 'Papers by Topic', 해당 주제를 연구하는 인물을 보여주는 'Persons by Topic', 해당 주제를 연구하는 기관을 보여주

는 'Institutions by Topic', 해당 주제 키워드를 제목이나 본문 필드에 포함하고 있는 논문을 보여주는 'Search Results' 등의 정보를 제공한다.

이러한 정교한 통합 검색 서비스가 가능한 이유는 URI 서버를 이용하여 단위 서비스 입력 파라미터인 키워드와 URI를 확보할 수 있기 때문이다. 또한, 페이지 내 하이퍼링크는 URI를 내재하고 있기 때문에 바로 가기 시에도 식별된 개체 정보를 정확히 제공할 수 있다.

반면에 그림 6과 그림 7의 주제 검색 결과는 일반적인 키워드 검색 결과 이상을 제공하지 못하고 있다. 아직까지 식별 체계가 서비스 대상에 도입되지 못하여 시맨틱 서비스를 제공할 인프라가 없기 때문이다. 이러한 경우 연구 동향 분석 등 심층적 주제 분석이 불가능하다. 본 연구에서 제시하는 'Topic Trends'는 각 논문에서 추출된 상위 5개 주제들간의 공기 정보를 이용하여 관련 주제 추이를 연도별로 제공한다. 예를 들어, 'neural network'을 사용자 질의어로 입력할 때 구성되는 주제 페이지 내의 'Topic trends'는 'generic algorithm'이나 'hiddenlayer'가 2000~2001년에 'neural network'과 밀접하게 연관된 주제들인데 반해, 2005년 이후에는 'rbf architecture'나 'turning toxicity' 등이 보다 많은 연관을 가짐을 보여준다. Google Trends에서 제공하는 추이 그래프는 주제에 기반을 둔 분석 서비스가 아니라, 시간별 해당 키워드의 질의 횟수에 의존하고 있는 단순

5) 사용자 인터페이스에서는 편의상 Full Name URI 대신 Name Space를 생략한 간략화된 형태를 사용한다.



**CiteSeer** Find:  Documents Citations

---

Searching for **PHRASE neural network**  
 Restrict to: [Header](#) [Title](#) Order by: [Expected citations](#) [Hubs](#) [Usage](#) [Date](#) Try: [Google \(CiteSeer\)](#) [Google \(Web\)](#) [Yahoo!](#) [MSN](#) [CSB](#) [DBLP](#)  
 More than 10000 results. Only retrieving 250 documents (System busy - maximum reduced). Retrieving documents... Order: number of citations.

**Minimizing Conflicts: A Heuristic Repair Method for...** - Minton, Johnston, (1992) (Correct) (204 citations)  
 paper was inspired by a surprisingly effective **neural network** developed by Adorf and Johnston [2, 22] for with a brief review of Adorf and Johnston's **neural network**, and then describe our symbolic method for  
[www.isi.edu/sims/minton/papers/aij-mc.ps](http://www.isi.edu/sims/minton/papers/aij-mc.ps)

**Sparse coding with an overcomplete basis set: A strategy...** - Olshausen, Field (1998) (Correct) (108 citations)  
 could be realized by a strictly feedforward **neural network**, in which case the functions  $f_i(x)$  would activity towards zero as shown in Figure 4b. A **neural network** implementation of this computation is shown extraction using an unsupervised **neural network**, *Neural Computation*, 4: 98-107. Kullback S (1959)  
[redwood.ucdavis.edu/pub/papers/VR.ps.Z](http://redwood.ucdavis.edu/pub/papers/VR.ps.Z)

**Evolution of Homing Navigation in a Real Mobile Robot - Floreano, Mondada (1996) (Correct) (85 citations)**  
 the evolution of a discrete-time recurrent **neural network** to control a real mobile robot. In all our Autonomous Robots, Genetic Algorithms, **Neural Networks**. I. Introduction A UTONOMOUS biological  
[iamitp.epfl.ch/pub/khepera/papers/floreano.IEEE-SMC.ps.Z](http://iamitp.epfl.ch/pub/khepera/papers/floreano.IEEE-SMC.ps.Z)

**Knowledge-Based Artificial Neural Networks - Towell, Shavlik (1994) (Correct) (81 citations)**  
 Accepted 9/93 Knowledge-Based Artificial **Neural Networks** Geoffrey G. Towell Jude W. Shavlik  
 Running Head: Knowledge-Based Artificial **Neural Networks** Current address of Towell is: Siemens  
[www.dai.ed.ac.uk/groups/evalg/Local\\_Copies\\_of\\_Papers/Towell.Shavlik.Knowledge-Based\\_Artificial\\_Neural\\_Networks.ps.gz](http://www.dai.ed.ac.uk/groups/evalg/Local_Copies_of_Papers/Towell.Shavlik.Knowledge-Based_Artificial_Neural_Networks.ps.gz)

**An Efficient Gradient-Based Algorithm for On-Line Training of...** - Williams, Peng (1990) (Correct) (59 citations)  
 on serial machines. 1 Introduction Artificial **neural networks** having feedback connections can implement a by a number of investigators in the **neural network** field, and their origins can, in fact, be  
[ftp.ccs.neu.edu/pub/people/rjw/fast-bpff-nc-90.ps](http://ftp.ccs.neu.edu/pub/people/rjw/fast-bpff-nc-90.ps)

**Discriminant Analysis by Gaussian Mixtures - Hastie, Tibshirani (1996) (Correct) (51 citations)**  
 neighbour methods and classification trees. **Neural network** classifiers have become a powerful to have generated much attention. Particular **neural network** models using Gaussian radial basis functions  
[www-stat.stanford.edu/~tibs/ftp/mda.ps](http://www-stat.stanford.edu/~tibs/ftp/mda.ps)

**On Convergence Properties of the EM Algorithm for Gaussian...** - Xu, Jordan (1996) (Correct) (51 citations)  
 (MAP) estimation. The recent emphasis in the **neural network** literature on probabilistic models has led to have tended to be the methods of choice in the **neural network** literature. However, we also compare EM to  
[psyche.mit.edu/pub/jordan/xu-jordan.ps.Z](http://psyche.mit.edu/pub/jordan/xu-jordan.ps.Z)

**Independent Component Analysis Using an Extended Infomax Algorithm...** - Lee (1997) (Correct) (50 citations)  
 has been studied by many researchers in **neural networks** and statistical signal processing (Jutten maximization in a single-layer feedforward **neural network**. The algorithm is effective in separating  
[www.sloan.salk.edu/~tewon/Blind/WWW/Public/nc97.ps.gz](http://www.sloan.salk.edu/~tewon/Blind/WWW/Public/nc97.ps.gz)

그림 6 Citeseer의 주제 검색 결과

**Google Scholar** [Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)  
  [Advanced Scholar Search](#)  
[Scholar Preferences](#)  
[Scholar Help](#)

---

**Scholar** All articles - Recent articles Results 1 - 10 of about 877,000

**All Results**

**P Alvarez**  
**L Squire**  
**K Hornik**  
**J Hopfield**  
**B Kosko**

**On the approximate realization of continuous mappings by neural networks**  
 K Funahashi - *Neural Networks*, 1989 - portal.acm.org  
 ... Ninan Sajeeth Philip, K. Babu Joseph, A **neural network** tool for analyzing trends in rainfall, *Computers & Geosciences*, v.29 n.2, p.215-223, March 2003. ...  
 Cited by 1217 - [Related Articles](#) - [Web Search](#)

**Multilayer feedforward networks are universal approximators - all 2 versions »**  
 K Hornik, M Stinchcombe, H White - *Neural Networks*, 1989 - portal.acm.org  
 ... ET Fonseca, PCG da S. Vellasco, SAL de Andrade, MMBR Vellasco, **Neural network** evaluation of steel beam patch load capacity, *Advances in Engineering Software* ...  
 Cited by 3229 - [Related Articles](#) - [Web Search](#)

**Growing cell structures—a self-organizing network for unsupervised and supervised learning - all 3 versions »**  
 B Fritzke - *Neural Networks*, 1994 - portal.acm.org  
 ... Jung-Hua Wang, Wei-Der Sun, On the Characteristics of Growing Cell Structures (GCS) **Neural Network**, *Neural Processing Letters*, v.10 n.2, p.139-149, Oct. 1999. ...  
 Cited by 561 - [Related Articles](#) - [Web Search](#) - [BL Direct](#)

**Nonlinear neural networks: Principles, mechanisms, and architectures.**  
 S Grossberg - *Neural Networks*, 1988 - csa.com  
 ... Three sources of contemporary **neural network** research - the binary, linear, and continuous-nonlinear models - are noted. ... C CA 2. 1 **NEURAL NETWORKS(C)**.  
 Cited by 439 - [Related Articles](#) - [Web Search](#)

**Optimal unsupervised learning in a single-layer linear feedforward neural network - all 3 versions »**  
 TD Sanger - *Neural Networks*, 1989 - ece-classweb.ucsd.edu  
 ... Feedforward **Neural Network** ... Abstract—A new approach to unsupervised learning in a single-layer linear feedforward **neural network** is discussed. ...  
 Cited by 582 - [Related Articles](#) - [View as HTML](#) - [Web Search](#)

**book Neural Networks: A Comprehensive Foundation - all 2 versions »**  
 S Haykin - 1994 - Prentice Hall PTR Upper Saddle River, NJ, USA  
 Cited by 8456 - [Related Articles](#) - [Web Search](#)

**Approximation capabilities of multilayer feedforward networks - all 3 versions »**  
 K Hornik - *Neural Networks*, 1991 - portal.acm.org  
 ... Source, **Neural Networks** archive Volume 4, Issue 2 1991 table of contents. Pages: 251 - 257. Year of Publication: 1991. ISSN:0893-6080. Author, Kurt Hornik, ...  
 Cited by 559 - [Related Articles](#) - [Web Search](#)

그림 7 Google Scholar의 주제 검색 결과

한 통계 서비스이므로 본 연구에서 제시하는 ‘Topic Trends’에서 다루는 정보와 큰 차이가 있다. CiteSeer의 경우, 특정 논문에 대해 연도별 인용수를 그래프로 제시하고 있으나 이 역시 단순 통계 서비스로서 논문의 주제별 분류 정보를 활용하지 못하고 있다.

**시나리오 2: 인물 검색**

인물 검색은 특정 인물의 논문을 찾거나 해당 인물과 관련된 인물(공저자, 인용저자 등)을 찾기 위한 목적으로 이루어진다. 이 경우 사용자는 단순히 이름을 키워드로 입력하고 원하는 결과를 기대한다.

‘A. J. Bernheim Brush’는 동명이인의 대표적인 예이다.<sup>6)</sup> 본 연구에서는 학술 정보에 포함된 저자를 공저 관계, 소속, 이메일 등 주변 정보를 이용하여 식별 해소(Identity Resolution)를 하고 있다[17]. 해당 키워드를 이용하여 검색하는 경우 URI 서버에서 3명이 검색되었으며, 그 중 정확히 매칭된 2명 중 1명으로 인물 페이지를 구성하였다(그림 8 참조). 실제로 URI ‘PER\_000000000000000016697’을 가지는 ‘Bernheim Brush’는 URI ‘PER\_00000000000000001356’을 가지는 ‘A. J. Bernheim Brush’와 동일 인물이지만 현재 이 둘간에 ‘sameAs’ 관계가 설정되어 있지 않아 별개의 인물로 제시된다[18]. ‘sameAs’는 본 통합 검색 시스템에 이미 적용되어 있으나, 본 연구의 범위를 벗어나므로 여기서는 서술하지 않는다.

반면, Citeseer와 Google Scholar의 검색 결과를 보면, 각각 7건과 224건의 논문 검색 결과를 제시할 뿐 동명이인 문제는 처리하지 못하고 있다. Google Scholar의 경우, ‘AJB Brush’와 같이 성을 제외한 나머지 이름은 모두 1자의 문자로 약자화시키는 정규화 기법을 사용한다. 이 방식은 재현율(Recall)을 높이기 위한 하나의 해결책은 될 수 있지만(예를 들어, ‘Alistair G Rust’의 경우 Citeseer에서 ‘Alistair G Rust’와 ‘Alistair G. Rust’의 두 가지 표기 형태가 나타나는데 Google Scholar에서는 이를 모두 ‘AG Rust’로 정규화 시킴으로써 검색 결과를 풍부하게 한다(그림 7의 왼쪽 부분 참조)), 개체를 식별해야 하는 응용 서비스에서는 해결책이 될 수 없다.

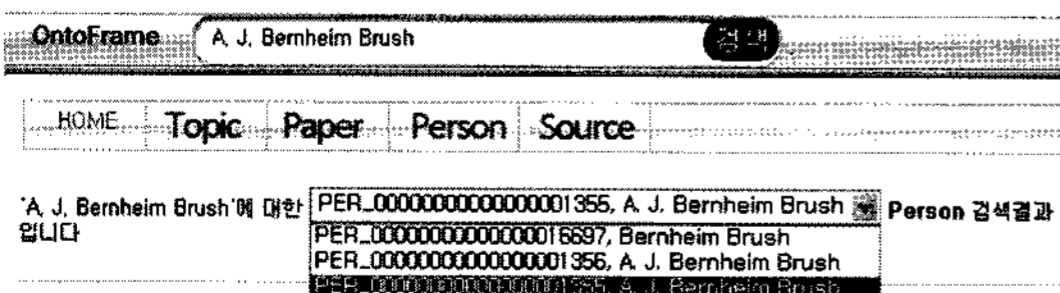


그림 8 본 시스템의 인물 검색 결과(테스트용 화면)

6) 한 명은 워싱턴 대학에 재직 중인데 도시 공학 및 시각화를 연구하고 있으며, 다른 한 명은 마이크로소프트 연구소에 재직 중인데 디지털 문서 주석을 연구하고 있다.

**5. 사용성 평가**

개체 중심적 통합 검색의 유용성을 판단하는 또 다른 방식으로 사용성 평가를 실시하였다. 사용성 평가 대상은 CiteSeer, Google Scholar, OntoFrame이며, 평가 방식으로는 7개의 작업(Task)을 사용자들에게 제시하고 각 시스템을 이용하여 정답을 찾도록 하는 작업 중심적 평가를 실시하였다. 또한, 사용성 평가 이후에 FGI(Focus Group Interview)를 통해 정성적 평가 의견을 받았다. 사용성 평가자는 KAIST 석·박사 과정 6명을 대상으로 하였는데, 이 사용자 수는 Nielsen에 의하면 약 80~90% 사이의 사용성 문제를 발견할 수 있는 규모이다(그림 9 참조).

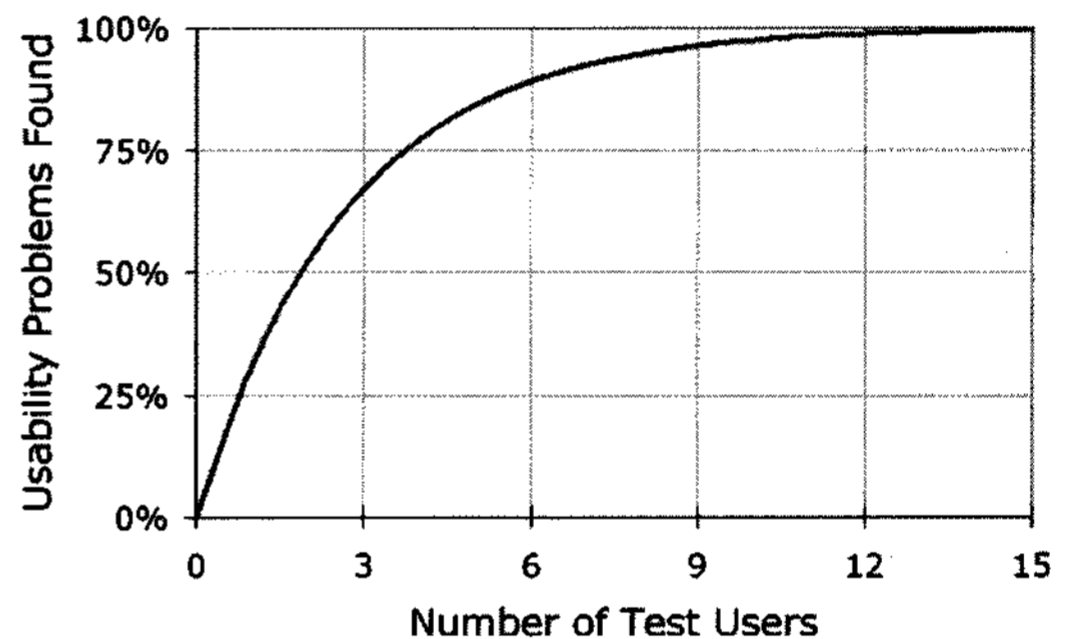


그림 9 사용자 수와 사용성 문제 발견 비율 간의 상관관계

사용성 평가 질의는 키워드 검색 등을 통해 해결할 수 있는 단순 질의(작업1~3)와 사용자의 휴리스틱, 정보 분석 등을 통해 해결할 수 있는 추론 질의(작업4~7)로 구성된다(표 2). 작업 중심적 평가 결과의 비교는 사용자들이 제시한 정답에서 발생한 편차를 이용하는데, 예를 들어, 특정 작업에 대해 6명의 사용자가 모두 동일한 정답을 10개씩 제시했다면 10/60이 될 것이며, 모두 다른 정답을 10개씩 제시했다면 60/60이 될 것이다. 이 값이 작을수록 사용자 간의 편차가 작음을 의미하며, 해당 시스템을 안정적으로 사용하고 있음을 간접적으로 증명한다.

표 3은 작업 중심적 평가 결과를 보여준다. 단순 질의를 대상으로 한 작업들(작업1~3)에서는 Google Scholar가 OntoFrame보다 약간 우수하게 나타났으나(113.58%), 연구자들이 Survey에서 필요로 하는 추론 질의를 대상으로 한 작업들(작업4~7)에서는 OntoFrame이 매우 우수하게 나타났다(Google Scholar 대비 137.63%). 전체 작업들로 살펴보더라도, OntoFrame이 Google Scholar와 비교하여 우수하였으며(113.50%), CiteSeer와 비교하여 매우 우수하였다(127.28%).

표 2 사용성 평가에 사용된 질의

작업1	제목에 "markov model"을 포함하고 있는 논문들을 찾으세요.
작업2	markov model, "markov," 그리고 "model"을 모두 제목에 가지고 있지 않으면서 "markov model"과 관련 있는 논문들을 찾으세요.
작업3	"markov model"을 사용하여 "speech recognition"을 연구하는 논문들을 찾으세요.
작업4	"markov model"과 관련된 키워드나 주제어들을 찾으세요.
작업5	"markov model"을 연구하는 전문가들을 찾으세요.
작업6	"markov model"을 연구하는 최고 전문가와 관련된 키워드나 주제어들을 찾으세요.
작업7	"markov model"을 연구하는 최고 전문가와 친근한 연구자들을 찾으세요.

표 3 작업 중심적 사용성 평가 결과

작업 번호	CiteSeer	Google Scholar	OntoFrame
1	50/60	28/60	41/60
2	51/59	57/60	47/57
3	36/44	29/60	35/54
4	46/57	55/60	35/57
5	55/59	51/60	32/50
6	40/40	46/50	30/40
7	35/43	37/43	30/50
단순 질의(작업1~3) 편차 평균	83.44% (136/163)	<b>63.33%</b> <b>(114/180)</b>	71.93% (123/171)
추론 질의(작업4~7) 편차 평균	88.94% (177/199)	88.73% (189/213)	<b>64.47%</b> <b>(127/197)</b>
전체 편차 평균	86.46% (313/362)	77.10% (303/393)	<b>67.93%</b> <b>(250/368)</b>
CiteSeer 대비 우수성	100%	112.14%	127.28%
Google Scholar 대비 우수성	89.17%	100%	113.50%

OntoFrame이 단순 키워드 검색과 부가 서비스를 제공하는 CiteSeer, Google Scholar보다 우위적인 사용성 평가 결과를 보인 데는 URI 서버를 이용하여 개체 페이지를 구성하는 방식의 통합 검색 결과가 사용자에게 보다 중점적으로 원하는 정보를 제공하기 때문이다. 이러한 결론은 FGI를 통해서도 확인할 수 있는데, 사용자 인터페이스와 단순 검색 결과에 대해서는 Google Scholar를 선호하지만, 전문적인 정보 제공 측면에서 CiteSeer나 Google Scholar가 OntoFrame에 비해 부족한 면이 많다는 지적들이 있었다. 또한, Thinking-aloud 기법[19]으로 사용자들을 관찰한 결과, CiteSeer나 Google Scholar를 이용하는 경우에 사용자 자신의 직관이나 기존 경험으로부터 얻어진 휴리스틱을 과도하게 사용하는 경향이 있었다. 이러한 경향은 사용자 간의 작업 수행에 있어서의 편차를 크게 함으로써 작업 결과

로 제시된 정답들의 편차를 크게 만드는 주요 원인이 된다.

## 6. 결론

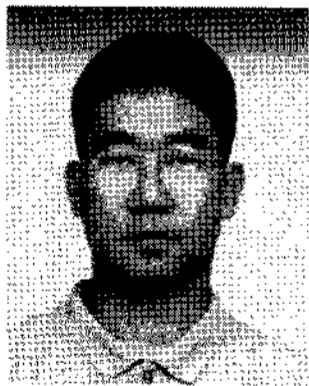
본 연구는 시맨틱 웹 기술이 실제적으로 정보 검색과 어떻게 결합될 수 있는가를 OntoFrame 상의 URI 서버를 중심으로 구성된 URI 서버 기반 통합 검색 시스템을 통해 서술하였다. 사용자 질의에 최적으로 부합하는 개체 유형을 결정하고 개체 페이지를 구성함으로써 심층적 학술 정보 서비스가 가능할 수 있도록 하였다. 향후 연구 방향은 다양한 학술 정보 서비스와의 직접 또는 간접 연계를 통한 시맨틱 웹 기술 기반 통합 검색 실제 서비스화에 초점을 맞추고 있다.

## 참고 문헌

- [1] C. Claunch, Top 10 Strategic Technologies for 2007 and Their Impact on the Data Center, In Proceedings of Gartner 25<sup>th</sup> Annual Data Center Conference, 2006.
- [2] H. Jung, M. Lee, W. Sung, and D. Park, Semantic Web-Based Services for Supporting Voluntary Collaboration among Researchers Using an Information Dissemination Platform, Data Science Journal 6(1), 2007.
- [3] 최옥경, 한상용, 자동화된 통합 프레임워크를 위한 시맨틱 웹 기반의 정보 검색 시스템, 정보처리학회논문지 13-C(1), 2006.
- [4] J. Davies and R. Weeks, QuizRDF: Search Technology for the Semantic Web, In Proceedings of the 37<sup>th</sup> Hawaii International Conference on System Sciences, 2004.
- [5] L. Ding, T. Finin, A. Joshi, R. Pan, R. Cost, Y. Peng, P. Reddivari, V. Doshi, and J. Sachs, Swoogle: A Search and Metadata Engine for the Semantic Web, In Proceedings of ACM 13<sup>th</sup> Conference on Information and Knowledge Management, 2004.
- [6] L. Ding, T. Finin, A. Joshi, Y. Peng, R. Pan, and P. Reddivari, Search on the Semantic Web, TR CS-05-09, UMBC eBiquity Publications, 2005.
- [7] R. Guha, R. McCool, and E. Miller, Semantic Search, In Proceedings of the 12<sup>th</sup> International Conference on World Wide Web, 2003.
- [8] A. Hogan, A. Harth, J. Umbrich, and S. Decker, Towards a Scalable Search and Query Engine for the Web, In Proceedings of the 16<sup>th</sup> International Conference on World Wide Web, 2007.
- [9] Y. Lei, V. Uren, and E. Motta, SemSearch: A Search Engine for the Semantic Web, In Proceedings of the 15<sup>th</sup> International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Net-

works, 2006.

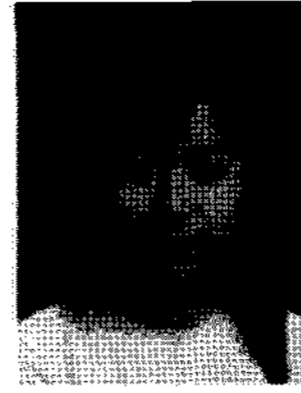
- [10] C. Rocha, D. Schwabe, and M. Aragao, A Hybrid Approach for Searching in the Semantic Web, In Proceedings of the 13<sup>th</sup> International Conference on World Wide Web, 2004.
- [11] 이재명, 김계현, 권오준, XML기반의 그리드데이터 메타데이터 통합검색시스템 개발에 관한 연구, 한국공간정보시스템학회논문지 7(2), 2006.
- [12] D. Kim, K. Jeong, H. Shin, and S. Hwang, An XML Schema-based Semantic Data Integration, In Proceedings of the 5th International Conference on Grid and Cooperative Computing, 2006.
- [13] J. Mostafa, Document Search Interface Design: Background and Introduction to Special Topic Section, In Journal of the American Society for Information Science and Technology, 55(10), 2004.
- [14] 김평, 이승우, 이미경, 구남양, 강인수, 정한민, 성원경, OntoStore-K: URI 기반 성과 관리 시스템, 제33회 한국정보과학회 추계학술대회, 2006.
- [15] 김평, 이승우, 성원경, OntoFrame의 추론 시스템 개선, 한국컴퓨터종합학술대회(KCC), 2007.
- [16] H. Jung and W. Sung, Construction of Semantic Web-based Knowledge Using Text Processing, In Proceedings of The 4<sup>th</sup> International Conference on Information Technology : New Generations, 2007.
- [17] 강인수, 정한민, 이승우, 김평, 구희관, 이미경, 구남양, 성원경, 논문 원문을 이용한 동명 저자 자동 군집화, 한국콘텐츠학회 추계종합학술대회, 2006.
- [18] 강인수, 정한민, 이승우, 김평, 이미경, 성원경, 시맨틱 웹 온톨로지에서의 OWL sameAs 적용, 정보과학회논문지: 소프트웨어 및 응용 34(4), 2007.
- [19] C. Lewis, Using the 'Thinking-aloud' Method in Cognitive Interface Design, IBM Research Report RC, 1982.



정 한 민

1992년 포항공과대학교 전자계산학과(학사). 1994년 포항공과대학교 전자계산학과(석사). 2003년 포항공과대학교 컴퓨터공학과(박사). 1994년~2000년 한국전자통신연구원 선임연구원. 2000년~2004년 (주)다이퀘스트 연구소장/CTO. 2004년~

현재 한국과학기술정보연구원 정보서비스연구팀 책임연구원 2005년~현재 과학기술연합대학원대학교 응용정보과학 겸임교수. 관심분야는 자연어처리, 시맨틱 웹, 정보 추출, 정보 검색, HCI



이 미 경

1999년 대구대학교 전자계산학과(학사) 2002년 경북대학교 컴퓨터공학과(석사) 2002년~2005년 한국전자통신연구원 연구원. 2005년~현재 한국과학기술정보연구원 연구원. 관심분야는 시맨틱 웹, 온톨로지, 추론엔진, 웹 2.0



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

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