

URI 서버 기반 통합 검색 시스템¹

정한민⁰ 이미경 성원경
한국과학기술정보연구원 정보서비스연구팀
jhm@kisti.re.kr

Integrated Search System Based on URI Server

Hanmin Jung⁰ MiKyung Lee Won-Kyung Sung
Information Service Research Lab., KISTI

요 약

본 연구는 키워드 기반 통합 검색의 한계를 극복하고자, 시맨틱 웹 기술의 기반인 URI가 부여된 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 서비스의 동시 작업 방식으로 개체 페이지를 구성한다. 또한, 논문으로부터 자동 추출된 주제 정보를 추론을 통해 인력, 기관, 위치 등으로 전파함으로써 주제 중심의 심층적 정보 제공이 가능하다. 해외에서 실 서비스되고 있는 Citeseer, Google Scholar와의 통합 검색 결과 비교 실험을 통해 본 연구의 효용성을 실증한다.

1. 서론

매년 급격히 증가하고 있는 정보의 홍수 속에서 원하는 정보를 찾거나, 정확한 정보만을 찾거나, 또는 메타데이터 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 90여만 건의 학술 정보가 검색된다. [7]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 네비게이션(Contextual Navigation)을 통해 특정 정보 요구에 적응할 수 있는 정보 접근으로 발전할 것으로 보고 있다. 이러한 발전은 궁극적으로 본 연구가 추구하는 방향인 검색과 네비게이션의 자동화로 이어질 것이다.

정보 검색의 효율성을 제고하고자 네이버를 비롯한 일부 상용 포털에서 인력, 기관 등의 개체(Entity)를 중심으로 개체 페이지를 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. 그렇지만, 동일 직업을 가지는 경우 동명이인 구별에 한계를 가질 수 밖에 없는 키워드

¹ 관련 특허: 11/576457(미국출원), PCT/KR2006/004096(PCT출원), 10-2007-0176312(국내출원), 10-2006-0081788(국내출원)

휴리스틱이어서, 문서 내에 다른 인력 정보가 혼재되어 있는 경우 엉뚱한 인력에 대한 문서가 해당 인력 검색 결과에 포함될 수 있다는 문제점을 안고 있다(그림1 참조). Yahoo Mindset²의 경우 Research와 Shopping 주제 중 더 강조하고 싶은 쪽으로 슬라이드 바를 움직여 검색 결과를 제어할 수 있도록 하고 있지만, 이 역시 질의어에 부가적인 제약을 줄 수 있는 키워드를 주는 방식으로 보이며, 사용자가 원하는 바를 정확히 선택하기 보다는 선호도의 정도를 제공할 수 있는 수준에 그치는 한계가 있다.

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 활용한 시도가 많지 않은데 이는 기존 텍스트 기반 정보를 의미화하는데 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.



그림1. 네이버 인력 검색 결과 예('공무원 이선희' 검색 결과 내에 다른 인력 정보가 혼재되어 있음)

2005년부터 연구되고 개발된 OntoFrame은 국내 과학 기술 기반 정보를 연구자의 연구 개발 전주기에서

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

제공할 수 있도록 하는 시맨틱 웹 기술 기반 프레임워크이다[13]. 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 시스템을 설명한다.

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 설명과 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 [6], [8], [9], [10], [11], [12], [16], [18] 등이 있다.

[6]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 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[9]. 그렇지만, 웹 상에서 웹 로봇을 이용한 수집, 추출, 색인, 검색이기 때문에 식별 체계를 본격적으로 도입하는 데 어려움을 가질 수 밖에 없다. [8]은 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 [18] 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 [5], [15], [17] 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, [17]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 [11], [12]이 있다. 검색 유형을 'Navigational Search'와 'Research Search'로 구분하고 인물과 같은 특정 개체를 검색 대상으로 정의한 [11]은 사용자 의도를 반영할 수 있는 명시적 기호와 현재 네이버에서 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 수 밖에 없다. [12]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 수 밖에 없다.

3. URI 서버 기반 통합 검색 시스템

본 연구에서는 주제(Topic), 인력(Person), 출처(Source) 등 개체³에 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. 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.

[3], [13] 등의 연구를 통해 이미 2006년에 구현된 URI 서버는 온톨로지 변경과 인스턴스 생성 관리기와의 연동을 고려하지 못한 구조여서 이식성(Portability)가 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 본 연구에서 개체는 온톨로지 스키마에 정의된 클래스 중 검색 결과 페이지 구성에 이용되는 클래스로 한정하여 정의한다.

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.

그림2는 URI를 참조하여 이루어지는 통합 검색 시스템의 구성도와 흐름을 보여준다. 먼저 사용자가 입력한 질의어인 키워드는 URI 서버 내에 저장된 인스턴스를 색인한 검색 엔진(IR Engine)에 전달되어 키워드-인스턴스 매칭에 이용된다(그림2의 ①, ②). 개체 검색기(Entity Finder)는 매칭 결과에 따라 표 1과 같은 최적 개체 결정 알고리즘을 이용하여 최적 개체를 선정한다. 최적 개체를 선정하는 이유는 다중의 개체 페이지 중 어느 개체 페이지를 우선적으로 보여줄 것인가를 결정하기 위해서이다. 예를 들어, 'neural network'이라는 사용자 질의어가 입력된 경우, 주제 개체와 해당 질의어가 매칭되므로 주제 개체가 최적 개체로 선정되고, 주제 페이지가 구성된다. 표 1의 최적 개체 선정 알고리즘은 서비스 중요도와 차별화를 고려하여 서비스 설계 시 미리 정의된 것이다.

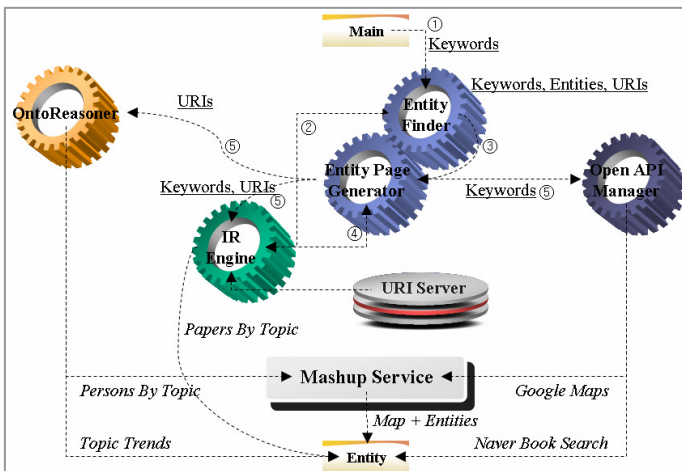


그림2. URI 기반 통합 검색 시스템 구성도

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.

표 1. 최적 개체 선정 알고리즘

검색결과 \ 개체	주제	인력	논문	출처
실패, 부분 매칭 성공			○	
주제	○			
인력		○		
논문			○	
출처				○
주제, (인력 OR 논문 OR 출처)	○			
인력, (논문 OR 출처)		○		
출처, 논문				○

개체 페이지 생성기(Entity Page Generator)는 개체 페이지를 생성하기 위해 사용자 질의어와 매칭된 URI를 각 단위 서비스 구성기에 전달한다(그림2의 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 서비스 구성기에는 추론 시스템(OntoReasoner), 검색 엔진(IR Engine), Open API 관리기(Open API Manager)가 포함된다[4]. 각 개체 페이지를 위한 단위 서비스의 예로는, 주제 페이지의 경우 추론 시스템을 호출하는 'Topic Trends', 'Persons by Topic', 검색 엔진을 호출하는 'Also Try', 'Papers by Topic', Open API 관리기를 호출하는 'Maps', 'Books' 등이 있다(그림3 참조).

```
Result = getURIInfo(keyword); // get URIs
```

```
If(Result is not NULL) { // one or more URIs
```

```
Assign each value to Topic, Person, Paper, and Event list.
```

```
getAdditionalParameters(Topic, Person, Paper, Event); // for URL generation
```

```
If(Topic is not NULL) jump to topic page;
```

```
Else If(Person is not NULL) jump to person page;
```

```

Else If(Event is not NULL) jump to event page;
Else jump to paper page;
}
Else // no URI
jump to paper page;

```

상기 코드는 표 1의 최적 개체 선정 알고리즘을 Pseudo 코드로 표현한 것으로, URI 서버 내에서 검색 키워드와 정확히 일치하는 목록(Result)을 찾고, 우선 순위에 의해 페이지 이동을 결정하는 방법을 보여준다.

4. 비교 실험

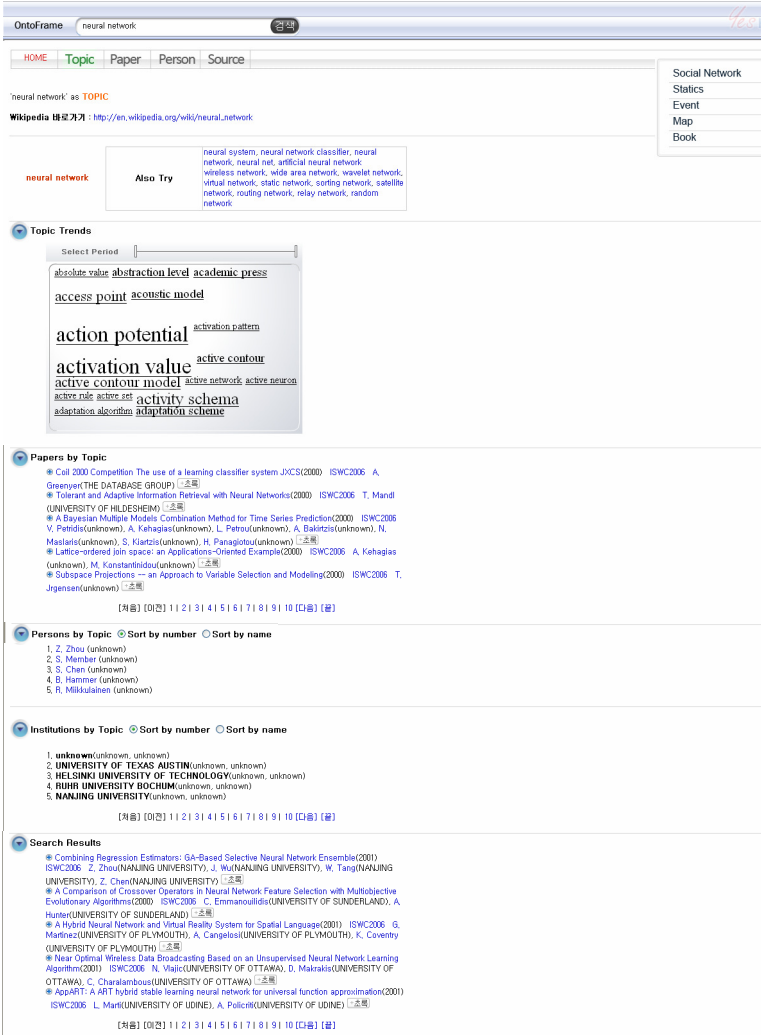


그림3. 본 시스템의 주제 검색 결과

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.

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 인력, 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 [14]에서 사용한 주제 추출 방법을 이용하여 논문에 주제를 최대 3개까지 자동으로 부여하였다. 이 주제 정보는 추론 시스템을 통해 전방 추론(Forward Chaining) 방식으로 논문 저자(인력)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 3가지를 설정하였는데⁴, 하나는 'neural network'라는 주제와 관련된 정보를 찾는 것이며, 다른 하나는 'A. J. Bernheim Brush'라는 인력과 관련된 정보를 찾는 것이며, 마지막은 'Dimensionality Reduction of Face Images Using Discrete Cosine Transforms for Recognition'이라는 논문과 관련된 정보를 찾는 것이다.

[시나리오 1: 주제 검색]

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.

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. 그림3, 그림4, 그림5는 'neural network'로 본 시스템, Citeseer, Google Scholar를 검색한 결과를 보여준다⁵.

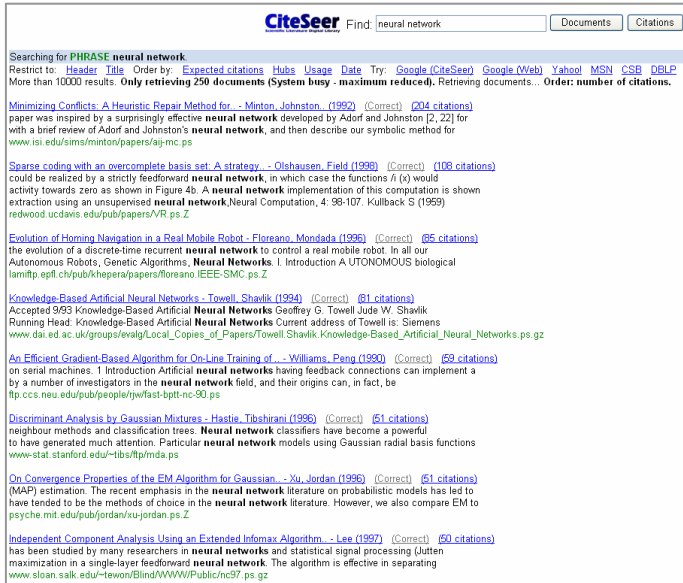


그림4. Citeseer의 주제 검색 결과

그림3과 같이 'neural network'라는 질의어로 검색 시 최적 개체 선정 알고리즘(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 그림3의 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 인력을 보여주는 'Persons by Topic', 해당 주제를 연구하는 기관을 보여주는 'Institutions by Topic', 해당 주제

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

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 바로가기 시에도 식별된 개체 정보를 정확히 제공할 수 있다.

반면에 그림4와 그림5의 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.

[시나리오 2: 인력 검색]

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

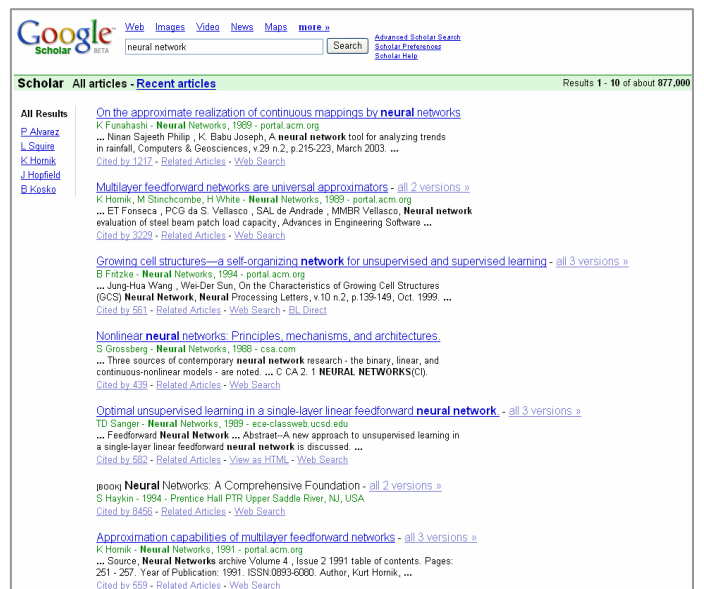


그림5. Google Scholar의 주제 검색 결과

'A. J. Bernheim Brush'는 동명이인의 대표적인 예이다⁶.

⁶ 한 명은 워싱턴 대학에 재직 중인데 도시 공학 및 시각화를

본 연구에서는 학술 정보에 포함된 저자를 공저 관계, 소속, 이메일 등 주변 정보를 이용하여 식별 해소(Identity Resolution)를 하고 있다[1]. 해당 키워드를 이용하여 검색하는 경우 URI 서버에서 3명이 검색되었으며, 그 중 정확히 매칭된 2명 중 1명으로 인력 페이지를 구성하였다(그림6 참조). 실제로 URI 'PER_000000000000000016697'을 가지는 'Bernheim Brush'는 URI 'PER_00000000000000001356'을 가지는 'A. J. Bernheim Brush'와 동일 인력이지만 현재 이 둘간에 'sameAs' 관계가 설정되어 있지 않아 별개의 인력으로 제시된다[2]. 'sameAs'는 본 통합 검색 시스템에 이미 적용되어 있으나, 본 연구의 범위를 벗어나므로 여기서는 언급하지 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 Recall을 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 이를 모두 'AG Rust'로 정규화시킴으로써 검색 결과를 풍부하게 한다.), 개체를 식별해야 하는 응용 서비스에서는 해결책이 될 수 없다. Citeseer의 경우 구두점 등 기호를 제거하고 검색하기 때문에 Google Scholar와 마찬가지로 Recall을 높일 수 있으나, 키워드 검색의 한계를 벗어나지 못한다.

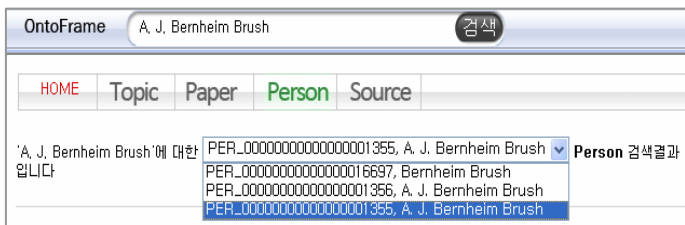


그림6. 본 시스템의 인력 검색 결과(상단 화면)

연구하고 있으며, 다른 한 명은 마이크로소프트 연구소에 재직 중인데 디지털 문서 주석을 연구하고 있다.

[시나리오 3: 논문 검색]

논문 검색은 정확한 논문 제목을 알고 있을 때 해당 논문을 바로 찾고자 하는 목적으로 이루어진다. 비교 대상 중 본 시스템과 Citeseer는 'Dimensionality Reduction of Face Images Using Discrete Cosine Transforms for Recognition'이라는 키워드에 대해 정확한 논문 정보를 제시하였으며, Google Scholar의 경우 큰따옴표("")를 사용하여야 해당 논문 정보를 상위 순위 내에서 얻을 수 있었다. Citeseer는 특정 논문에 대해 원문 정보, 인용 정보, 유사 논문 정보, 서지 정보 등 상세 정보를 제공하는데 반해 Google Scholar는 원문 정보로의 링크만을 제공한다. 본 통합 검색 시스템은 논문 페이지에서 Citeseer에서 제공하는 상세 정보 이외에 주제 정보, 인용 순위 정보를 제공한다. 또한, 유사 논문 정보의 경우에도 논문에 부여된 주제 정보간 유사도에 근거하여 제시하는 데 반해, Citeseer는 문장 수준에서의 단순 비교를 수행함으로써 대부분의 경우 동일 저자 논문이 유사 논문 정보로 제시되는 경향을 보인다. 본 시스템의 경우 URI에 기반을 두고 각 논문을 정확히 식별해내기 때문에 이러한 상세 정보의 제공이 가능하다.

이상과 같이 3가지 시나리오를 통해 본 통합 검색 시스템이 다른 키워드 검색 기반 서비스에 비해 진보된 검색 결과를 제시한다는 것을 실증하였다. 특히, 다중 개체에 대한 인스턴스를 발견한 경우에도 특정 개체 페이지를 우선적으로 구성하여 혼재된 정보 제공을 방지하고 각 개체에 대한 심층 분석 정보를 제공할 수 있다.

5. 결론

본 연구는 시맨틱 웹 기술이 실제로 정보 검색과 어떻게 결합될 수 있는가를 OntoFrame 상의 URI 서버를 중심으로 구성된 URI 서버 기반 통합 검색 시스템을 통해 설명하였다. 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] 강인수, 정한민, 이승우, 김평, 구희관, 이미경, 구남양, 성원경, 논문 원문을 이용한 동명 저자 자동 군집화, 한국콘텐츠학회 추계종합학술대회, 2006.
- [2] 강인수, 정한민, 이승우, 김평, 이미경, 성원경, 시맨틱 웹 온톨로지에서의 OWL sameAs 적용, 정보과학회논문지: 소프트웨어 및 응용 34(4), 2007.
- [3] 김평, 이승우, 이미경, 구남양, 강인수, 정한민, 성원경, OntoStore-K: URI 기반 성과 관리 시스템, 제33회 한국정보과학회 추계학술대회, 2006.
- [4] 김평, 이승우, 성원경, OntoFrame의 추론 시스템 개선, 한국컴퓨터종합학술대회(KCC), 2007.
- [5] 이재명, 김계현, 권오준, XML기반의 그리드데이터 메타데이터 통합검색시스템 개발에 관한 연구, 한국공간정보시스템학회논문지 7(2), 2006.
- [6] 최옥경, 한상용, 자동화된 통합 프레임워크를 위한 시맨틱 웹 기반의 정보 검색 시스템, 정보처리학회논문지 13-C(1), 2006.
- [7] C. Claunch, Top 10 Strategic Technologies for 2007 and Their Impact on the Data Center, In Proceedings of Gartner 25th Annual Data Center Conference, 2006.
- [8] J. Davies and R. Weeks, QuizRDF: Search Technology for the Semantic Web, In Proceedings of the 37th Hawaii International Conference on System Sciences, 2004.
- [9] 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 13th Conference on Information and Knowledge Management, 2004.
- [10] 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.
- [11] R. Guha, R. McCool, and E. Miller, Semantic Search, In Proceedings of the 12th International Conference on World Wide Web, 2003.
- [12] A. Hogan, A. Harth, J. Umbrich, and S. Decker, Towards a Scalable Search and Query Engine for the Web, In Proceedings of the 16th International Conference on World Wide Web, 2007.
- [13] 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.
- [14] H. Jung and W. Sung, Construction of Semantic Web-based Knowledge Using Text Processing, In Proceedings of The 4th International Conference on Information Technology : New Generations, 2007.
- [15] 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.
- [16] Y. Lei, V. Uren, and E. Motta, SemSearch: A Search Engine for the Semantic Web, In Proceedings of the 15th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks, 2006.
- [17] 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.
- [18] C. Rocha, D. Schwabe, and M. Aragao, A Hybrid Approach for Searching in the Semantic Web, In Proceedings of the 13th International Conference on World Wide Web, 2004.