

**사례
발표**

**시맨틱 웹 기반 플랫폼상에서의
웹 2.0 활용 서비스**

목 차

- 1. 서 론
- 2. OntoFrame 소개
- 3. 서비스 개선
- 4. 결 론

정안민 · 이미경 · 성원경
(한국과학기술정보연구원)

1. 서 론

OntoFrame 사용성 평가의 일환으로 2007년 초에 실시된 사용자 인터뷰 결과에 따르면, 모든 인터뷰 대상자들이 연구개발을 위한 사전 조사 단계에서 Google(<http://www.google.com/>)과 Google Scholar(<http://scholar.google.com/>)를 이용한다고 답변하였다¹⁾. 경쟁 정보 서비스인 Citeseer(<http://citeseer.ist.psu.edu/>), YesKiSTi(<http://www.yeskisti.net/>), NDSL(<http://www.ndsl.or.kr/eng/newindex.html>) 등에 비해 사용 빈도 면에서 월등히 앞선 이유로 정교한 검색 결과와 빠르게 갱신되는 풍부한 정보를 꼽았다. 이런 현실에서 연구개발 지원 서비스를 독자적으로 개발하여 성공할 수 있을까라는 의구심이 들기도 하지만, 사용자 인터뷰 결과의 다른 부분에서 그 방향성을 찾을 수 있다고 본다. 예를 들어, Citeseer의 경우 몇 년 전까지만 해도 대표적인 IT 분야의 연구 정보 서비스였지만, 지금은 Google에 밀린 상태이다. 그렇지만, 과반수의 사용자들이 아직까지 Citeseer를 이용하고 있다고 답했는데, 그 이유는 해당 서비스가 인용 논문 정보, 해당 저자

의 다른 논문 정보, 연관 논문 정보 등 보다 심층적인 정보를 제공하고 있기 때문이었다.

OntoFrame의 연구개발 전주기 지원 서비스는 논문 검색에만 국한되는 것이 아니라, 연구 동향 분석, 전문가 발굴, 연구자 네트워크 분석 등 다양한 연계 정보를 연구개발 전주기에 걸쳐 제공함으로써 종합적인 연구개발 프로세스의 효율성을 추구하는 것을 목표로 하고 있다. 즉, 단순 문자열 기반 논문 검색 서비스라면 Google, Citeseer, NDSL 등을 앞서기 힘들지만, 검색 수준에서는 획득할 수 없는 2차 연구 정보를 서비스하여 연구자의 연구개발 노력을 대폭 감소시키는 방향에서 그 효율성을 찾고 있다. 이러한 목표 달성을 위해서는 시맨틱 웹 기술 기반의 정보 관리 프레임워크 구축이 선결되어야 한다. URI(Uniform Resource Identifier), 온톨로지(Ontology), 추론(Inference, Reasoning)으로 대표되는 시맨틱 웹 기술은 종적 정보의 횡적 구성화, 정보 식별화를 통해 단편적 정보 조각을 연

1) 본 인터뷰를 위해 한국과학기술정보연구원 연구원(Full-time:2명)과 한국전자통신연구원 연구원(Full-time:1명, Part-time 대학원생: 3명)을 선정하였다.

계·통합하기 위한 필수 요소이다.

한국과학기술정보연구원(KISTI)는 2005년부터 OntoFrame이라는 이름으로 연구개발 전주기 지원 시스템을 개발해오고 있다[2][3]. 2005년의 ‘연구자 그룹 및 전문가 추천’, ‘연구자 추적’, ‘연구 맵’ 서비스를 시작으로, 2006년에는 ‘연구자 네트워크’, ‘연구자 정보’, ‘연구 성과 맵’, ‘통계 정보’, ‘성과 정보’, ‘기관 정보’ 서비스를 개발하였다. 국내·외 연구 소개를 통해 이미 그 성과를 인정받고 있지만, 연구자에게 좀 더 실질적인 도움이 되는 서비스를 제공하기 위해 지속적으로 개선하고 있다.

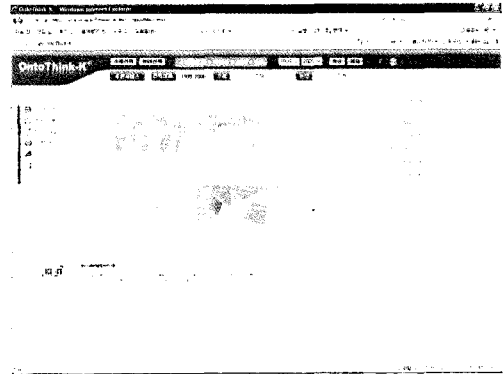
시맨틱 웹 기술이 정보 조직화 측면에서 필요한 기반 기술이라면, 웹 2.0 기술은 서비스 유연성 측면에서 필수불가결한 기술이다. 특히, Open API, AJAX(Asynchronous JavaScript and XML), Mashup 서비스, Social Networking 등은 서비스를 풍부하고 효과적으로 구성하는데 도움을 준다. OntoFrame은 차세대 웹 기술로서 두 축을 이루고 있는 시맨틱 웹 기술과 웹 2.0 기술의 결합을 통해, 연구개발 전주기 지원 서비스를 원활히 제공하는 것을 목표를 삼고 있다.

2장에서는 시맨틱 웹 기술에 기반을 두고 개발된 OntoFrame을 소개하고, 3장에서는 2007년에 개발 중인 시맨틱 웹 기술과 웹 2.0 기술의 결합 서비스를 설명한다.

2. OntoFrame 소개

OntoFrame은 성과, 인력, 기관 등을 포함하는 국가 과학기술 R&D 기반 정보를 등록하고 관리하는 서비스인 OntoStore와 추론 시스템과의 상호작용을 통해 시각화 기반 정보를 제공하는 추론 서비스인 OntoThink로 구성된다[4][6]. 본 장에서는 OntoFrame의 주요 특징인 URI 기반 개체 관리 기술과 추론 기술을 중심으로 연구개발 전주기 지원 서비스를 설명한다²⁾.

2.1 서비스 구성



(그림 1) OntoThink 인터페이스 예

추론 서비스 인터페이스는 상단의 공통 제약 설정과 하단의 6개 세부 추론 서비스(‘연구자 네트워크’, ‘연구자 정보’, ‘연구 성과 맵’, ‘통계 정보’, ‘성과 정보’, ‘기관 정보’)로 구성된다. 공통 제약 설정을 통해 특정 주제나 분야, 관심 있는 성과물 유형(논문, 특허, 보고서), 관찰 기간을 지정하면 모든 세부 추론 서비스에 적용된다. ‘연구자 네트워크’는 상위 20인의 전문가, 공저자 관계 기반 연구자 그룹 및 네트워크, 인용 관계 기반 연구자 그룹 및 네트워크를 보여준다. ‘연구자 정보’는 연구자 검색을 통해 URI 기반으로 연구자를 찾고 해당 연구자의 연구 주제·분야 정보를 제시한다. ‘연구 성과 맵’은 지역별 연구 활성도를 보여주는데 광역 행정 단위와 시·군·구 단위로 살펴볼 수 있다. ‘통계 정보’는 최대 3인까지의 연구자 성과물을 연도별로 비교할 수 있도록 통계 그래프를 제공한다. ‘성과 정보’는 성과물을 검색하여 유형별로 그 결과를 제시하고, ‘기관 정보’는 연구자 정보와 같은 방식으로 URI 기반 기관 정보와 해당 기관의 주요 연구 주제·분야 정보를 제시한다.

2006년도 서비스 대상은, KISTI 내부 성과 정

2) 관련 특허: 10-0659370, 10-0644317, 10-0626817

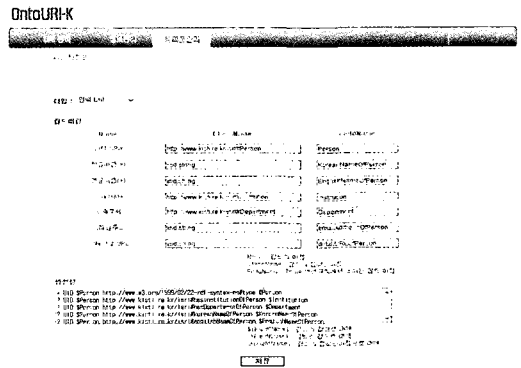
보의 경우 학술대회 논문 1,172건(2002년 ~ 2006년), 학술지 논문 279건(2002년 ~ 2005년), 연구 보고서 734건(1999년 ~ 2006년), 특허 64건(2002년 ~ 2005년)을, 외부 성과 정보의 경우 원문 포함 국내 학술대회 논문 11,406건을 포함한다. 추론 시스템 내의 RDF 트리플 수는 1,459,047개(2006.9 기준)이다. 이는 지식 확장(전방 추론; Forward-Chaining Inference)에 의해 획득된 확장 트리플을 포함하는 개수이다. 확장 트리플 생성 비율이 낮은 편인데, 그 이유는 첫째, 스키마 기반 확장 규칙에서 다루는 13개 RDF(S) 합의 규칙이 본 추론 서비스에서 별다른 역할을 하지 않아 1개를 제외하고 사용하지 않았기 때문이며, 둘째, 정보 등록 인터페이스와 URI 서버를 통해 개체 정합성을 검사함으로써 별도로 OWL 제약(Restriction)을 다루고 있지 않기 때문이다[1][7].

2.2 URI 기반 개체 관리

관리하고자 하는 정보가 대용량화되면서 온톨로지 상에서의 개체 관리 문제가 시맨틱 웹 영역에서 주요 이슈로 대두되고 있다. 개체 관리 시스템으로서의 URI 서버는 모델링 대상이 되는 실세계(Real World)와 온톨로지 사이에 위치하여, 온톨로지에 등록될 실세계 개체들의 정합성을 검사하고, 온톨로지 인스턴스를 저장하며, 더 나아가 등록된 특정 개체와 관련된 RDF 트리플 생성까지를 담당한다[4][6]. URI 서버는 온톨로지 개체의 정합성 검사를 위해 관계형 데이터베이스(RDBMS) 기술을 이용한다. 즉, 온톨로지 스키마 상의 클래스, 속성, 클래스 간 관계를 E-R 모델에서의 Entity, Attribute, Relationship으로 대응시키고, 온톨로지서에서 결여된 클래스 내 개체의 고유성 보장을 위해 Entity 모델링 시 Key Attribute들을 지정한다. 또한, 개체 속성값의 도메인 검사를 위해 RDBMS의 Trigger 기술을 사용함으로써 개체 등록이 레코드 삽입

(Record Insert)과 관련된 Stored Procedure에서 수행되도록 한다.

URI 서버의 또 다른 기능인 RDF 트리플 생성은 추론을 지원하기 위해 URI 서버에서 관리되는 개체들을 트리플 형태로 변환하는 역할을 한다. 이때 이용하는 정보로는 URI 서버 내 저장소 테이블과 온톨로지 상의 클래스를 매핑시켜주는 DB-클래스 매핑 정보, XML 네임스페이스(Namespace)를 포함하는 Full-named URI와 Prefix를 매핑시켜주는 NS-Prefix 매핑 정보, 개체별 필드로부터 인스턴스 트리플을 생성할 수 있도록 규칙을 정의한 DB-트리플 변환 규칙이 있다(그림 2참조)[6].



(그림 2) DB-클래스 매핑 정보 및 DB-트리플 변환 규칙 편집 인터페이스 예(인력 URI)

2.3 추론

RDBMS 응용 서비스의 문제로 지적되어 온 유지보수 및 서비스 확장의 경직성을 시맨틱 웹 기술 기반 응용 서비스에서는 추론 기술을 이용하여 해결한다. OntoFrame은 온톨로지를 이용하여 새로운 지식을 내재된 지식을 통해 창출하는 추론을 수행하기 위해 자체적으로 개발한 추론 시스템을 사용한다[6]. 전방 추론에 의한 지식 확장 과정에서 스키마 기반 확장 규칙과 사용자

- Article
 - x hasPublication y, y yearOfPublication z -> x hasPublicationYear z
- Thesis
 - x yearOfPublishingThesis y -> x hasPublicationYear z
- Patent
 - x yearOfApplication y -> x hasPublicationYear y
- Report
 - x hasOriginatedProject y, y endingYearOfProject z -> x hasPublicationYear z
- Person
 - x hasCreationInformation y, y hasCreator z -> x createdByPerson z
 - x hasInstitutionOfPerson y, y hasLocationOfInstitution z -> x locatedIn z
 - x createdByPerson y -> y creatorOf x
- Institution
 - x hasCreationInformation y, y hasInstitutionOfCreator z -> x createdByInstitution z
- Location
 - x createdByInstitution y, y hasLocationOfInstitution z -> x createdByLocation z

(그림 3) 사용자 정의 확장 규칙 예

정의 확장 규칙을 참조한다. 그림 3은 사용자 정의 확장 규칙의 예를 보여주는 데, 예를 들어 'Person'과 'Location'에서의 'createdByPerson', 'createdByInstitution'은 온톨로지에 기정의된 객체 속성 관계(Object Property)가 아니라 사용자 정의 확장 규칙에 의해 신규 생성된 유도 속성 관계(Derived Property)로서, 추가적으로 'creatorOf'와 'createdByLocation'을 생성하는데에도 이용된다. 새로운 관계를 추가로 정의하기 위해 RDBMS 응용 서비스의 경우 스크립트를 갱신하고, 필드를 추가로 정의하고, 뷰(View) 테이블을 생성하는 등의 복잡한 유지보수 작업을 필요로 하지만, 추론의 경우 규칙 추가를 통해 간단히 이 문제를 해결할 수 있다.

2.4 서비스 분석 및 개선 방향

OntoFrame은 세계적으로도 찾아보기 힘든 수준의 URI 관리 기술을 연구개발 전주기 지원 서비스에 적용한 시맨틱 웹 기술 기반 시스템이다. 추론 서비스를 위해 자체 개발한 추론 시스템을 이용하였으며, RDBMS를 RDF 트리플 저장소로 활용하여 안정성과 성능을 높였다. 그렇지만 이

러한 특징에도 불구하고, 연구자의 연구개발에 실질적으로 도움을 주기에 다소 부족한 요소들이 있다.

2.4.1 연구 정보의 다양성과 풍부성

2000년 이후 대표적인 국내 IT 관련 연구 정보³⁾를 대상으로 서비스하고 있지만, 연구자들이 해외 논문을 상대적으로 많이 참조하고 있는 현실에서 국내·외를 망라하여 좀 더 다양하고 풍부한 연구 정보를 제공할 필요가 있다. 서론에서 언급한 바와 같이 Google이 연구 정보 제공처로서 비중 있는 역할을 하는 이유도 풍부하고 적시에 갱신되는 정보가 있기 때문이다. 해외 논문 정보로 서비스 대상을 확대하고, Google, Google Scholar, Citeseer, Rexa.info(<http://rexa.info/>), OntoWorld(http://ontoworld.org/wiki/Main_Page) 등의 연구 정보 제공 사이트와의 연계를 시도할 필요가 있다.

2.4.2 개체별 정보 서비스

서비스 구성이 '연구자 네트워크', '연구 성과

3) 한국정보과학회, 대한전자공학회, 한국정보처리학회, 한국통신학회 등에서 주관한 국내 학술대회 논문집

맵', '통계 정보' 등 특정 서비스별로 구성되어, 개체 간 정보 연계성이 떨어진다. 이를 위해 논문, 인력, 주제, 출처 등 개체별 정보 제공과 상호 연계를 통해 관심 있는 개체를 집중적으로 살펴볼 수 있는 방안을 제공할 필요가 있다.

2.4.3 Open 플랫폼 활용

자체적인 정보와 인터페이스만을 이용하여 서비스를 구성하는 것은 정보 제공 범위를 한정시키고 플랫폼의 상호운용성(Interoperability)을 떨어뜨린다. 이러한 문제는 이질적 정보 간 결합이 가능하도록 하는 Open API, Mashup 서비스 등 웹 2.0 기술을 활용하여 해결할 필요가 있다.

2.4.4 개인화 서비스

모든 연구자에게千篇일률적인 정보만을 제공한다면, 효율적인 방식으로 원하는 정보를 획득하는데 장애가 될 수밖에 없다. 각 연구자가 관심을 가지고 있는 정보를 선별적으로 제공하고 원하는 방식으로 맞춤 서비스를 구성할 수 있게 한다면 여러 단계를 거치지 않고도 연구자 의도에 맞는 서비스를 제공할 수 있다.

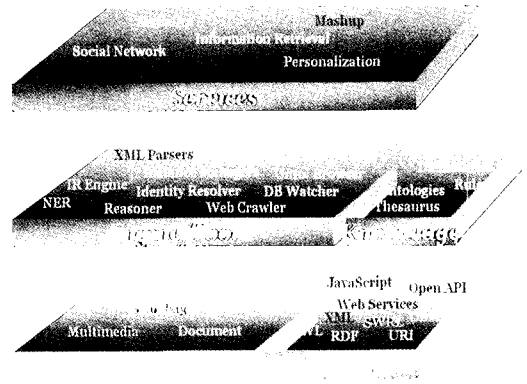
3. 서비스 개선

차세대 웹 기술로서 각광받고 있는 시맨틱 웹 기술과 웹 2.0 기술은 상호 보완적인 적용이 가능한데, 예를 들어 시맨틱 웹 기술은 정교한 개체 표현과 용이한 지식 확장을 가능하도록 하며, 웹 2.0 기술은 오픈 플랫폼 상에서의 유연한 서비스 접목을 도와주기 때문이다. 결국, 웹 2.0은 Open API와 개방형 구조를 통해 데이터 중심의 서비스가 가능하도록 한다는 점에서 개체 기반으로 2차 연구 정보를 제공하고자 하는 OntoFrame의 서비스 방향에 부합한다. 또한, 웹 서비스들의 조합을 통한 Mashup 서비스와 클라이언트 확장 기술을 이용한 편리하고 호환성 있는 사용자 인터페이스 제공은 시맨틱 웹 기술 기반 프레임워크와 응용 서비스 간의 원활한 연동을 가능하게

한다. 본 장에서는 2.4절에서 제시한 서비스 개선 방향 중 웹 2.0 기술이 어떻게 OntoFrame 상에서 실현될 수 있는지를 보여준다.

3.1 서비스 구성

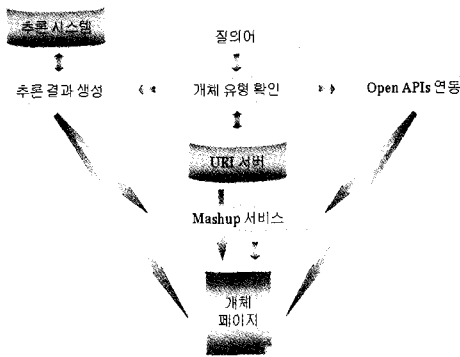
웹 2.0 기술이 현상을 통해 생겨난 비이론적인 개념이지만, [5]를 통해 기술적 특성들이 어느 정도 정리가 되어 있다. 특히, 데이터 중심의 서비스 개념과 개방형 구조상에서의 Mashup 서비스가 OntoFrame 서비스에 도입될 수 있다.



(그림 4) OntoFrame 구조

(그림 4)는 3개 계층으로 이루어진 OntoFrame 구조를 보여주는데, Web Services, Open API, Mashup 서비스 등 웹 2.0 기술이 추가되어 있다. Open API 중에서도 Google Maps API (www.google.com/apis/maps/), Google AJAX Search API ([code.google.com/apis /ajaxsearch/](http://code.google.com/apis/ajaxsearch/)), Naver Book Search API (<http://openapi.naver.com/page.nhn?PageId =05>) 등이 서비스에 이용된다.

(그림 5)는 사용자 질의부터 서비스 응답까지의 흐름을 보여준다. 개체 유형 확인을 위한 URI 서버 검색에서는 입력된 문자열 기반 질의어가 URI 서버에 이미 등록된 개체에 해당하는지를 확인한다. 입력 이름, 기관 이름, 주제명 등



(그림 5) OntoFrame 서비스 흐름

이 검색 대상이며 매칭되는 경우에 해당 개체 페이지를, 매칭되지 않는 경우에 성과 개체 페이지를 우선적으로 구성할 수 있도록 함으로써, 단순한 문자열 기반의 통합 검색 결과 페이지 이상의 개체 기반 서비스를 제공할 수 있다. 예를 들어, 질의어로 “시맨틱 웹”을 입력한 경우, 주제(주제어 개체) 페이지를 통해 관련 주제어, 상·하위 주제어, 연도별 주제 트렌드, 해당 주제로 분류된 논문, 해당 주제 연구자, 해당 주제 연구 기관, 해당 주제 연구자 그룹, 해당 주제 관련 연도별 논문 통계, 해당 주제 관련 학술대회, 해당 주제 관련 지역별 분포(Google Maps API와 연동), 해당 주제 관련 도서(Naver Book Search API와 연동) 등을 보여준다. 이전의 OntoFrame에 비해 개체별로 심층적인 서비스가 가능하도록 함으로써 OntoFrame을 기존 논문 검색 사이트들과 차별화시킬 수 있다. 개체 유형 확인 후 해당 주제 관련 지역별 분포와 같이 Open API와 연동하여 Mashup 서비스로 구성하는 경우나 해당 주제 연구자와 같이 메타 정보를 통해 직접 획득할 수 없는 정보를 생성해야 하는 경우, 추론 시스템을 통한 추론 결과 생성 과정이 필요하다. Google AJAX Search API, Naver Book Search API 등은 질의어를 통해 직접 결과를 획득할 수 있도록 해주므로 통합 개체 페이지의 일부로서 바로 삽입할 수 있다. 이와 같이 다양한 경로를

통해 획득된 정보는 개체 페이지에 통합되어 사용자에게 서비스됨으로써, 웹 2.0이 추구하는 데이터 중심의 서비스가 실현될 수 있다. 또한, AJAX, FLEX 등 클라이언트 확장 기술을 활용하여 편리하고 호환성 있는 UI를 제공하여, 시맨틱 웹 기술의 가치를 한층 효과적으로 보여줄 것이라 기대한다.

5. 결론

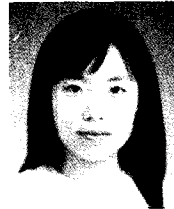
시맨틱 웹 기술에 기반을 두고 개발된 OntoFrame은 대내·외 서비스들을 위해 상용화 단계를 거치고 있다. 웹 2.0 기술이 매력적으로 다가오는 이유는 시맨틱 정보만으로 연구개발에 필요한 모든 정보를 적시에 그리고 효과적으로 제공하기 어렵기 때문이다. 다만, 웹 2.0 기술 자체만으로 정교한 서비스를 개발하기도 힘들다. [8]에서 언급한 바와 같이 Flickr(<http://www.flickr.com/>) 등 대표적인 Open API를 이용한 서비스가 태그(Tag)나 문자열에만 의존하는 방식이라 검색 품질이 떨어지는 문제가 발생할 수 있기 때문이다. 시맨틱 웹 기술, 특히 URI 기반 개체 확인 기술이 필요한 이유가 여기에 있다. 본 연구는 시맨틱 웹 기술과 웹 2.0 기술이 OntoFrame의 연구개발 전주기 지원 서비스에서 어떻게 결합될 수 있는가를 보여줌으로써, 두 차세대 웹 기술의 진정한 Mashup 가능성을 타진하였다. 현재 개선이 진행 중인 OntoFrame의 성공적인 서비스가 국내·외 시맨틱 웹 기술 발전에 기여하기를 기대한다.

참고문헌

- [1] S. Harris and N. Gibbins, 3store: Efficient Bulk RDF Storage, In Proceedings of the 1st International Workshop on Practical and Scalable Semantic Systems, 2003.

- [2] H. Jung, OntoFrame-K: Semantic Web-based Information Dissemination Platform, The 9th International Forum on Metadata Registry, 2006.
- [3] 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, Vol. 6, 2007.
- [4] 김평 외, OntoStore-K: URI 기반 성과 관리 시스템, 제33회 한국정보과학회 추계학술대회, 2006.
- [5] 전중홍, 이승윤, 웹 2.0 기술 현황 및 전망, 전자통신동향분석, Vol. 21, No. 5, 2006.
- [6] 정한민, 강인수, 이미경, 이승우, 성원경, OntoThink-K@: DBMS 기반 추론 서비스, 제33회 한국정보과학회 추계학술대회, 2006.
- [7] 정한민 외, URI 서버에 기반한 국가 R&D 기반정보 온톨로지 설계 및 구현, 정보관리연구, Vol. 37, No.2, 2006.
- [8] 정한민, 이미경, 성원경, Open API 기술 동향, 주간기술동향, Vol. 1296, 2007.

2000년~2004년 ㈜다이렉스트 연구소장/기술이사
 2004년~현재 한국과학기술정보연구원 책임연구원
 2005년~현재 과학기술연합대학원대학교 겸임교수
 관심분야 : 자연어처리, 시맨틱 웹, 정보 추출, 정보 검색, HCI
 이 메 일 : jhm@kisti.re.kr



이 미 경

1999년 대구대학교 전자계산학과(학사)
 2002년 경북대학교 컴퓨터공학과(석사)
 2002년~2005년 한국전자통신연구원 연구원
 2005년~현재 한국과학기술정보연구원 연구원
 관심분야 : 시맨틱 웹, 온톨로지, 추론엔진, 웹 2.0
 이 메 일 : jerryis@kisti.re.kr



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)보이스텍 연구개발본부장/상무이사
 2004년~현재 한국과학기술정보연구원
 정보서비스연구팀장/책임연구원
 2004년~현재 과학기술연합대학원대학교 겸임교수
 관심분야 : 자연어처리, 시맨틱 웹
 이 메 일 : wksung@kisti.re.kr

저자약력



정 한 민

1992년 포항공과대학교 전자계산학과(학사)
 1994년 포항공과대학교 전자계산학과(석사)
 2003년 포항공과대학교 컴퓨터공학과(박사)
 1994년~2000년 한국전자통신연구원 선임연구원