

## 학술 연구 분야에서의 추론 기반 연구자네트워크 생성 Construction of Researcher Network in the Academic Research Area based on Inference

이승우, 김평, 정한민, 구희관, 성원경  
한국과학기술정보연구원

Lee Seung-Woo, Kim Pyung, Jung Han-Min,  
Koo Hee-Kwan, Sung Won-Kyung  
KISTI

### 요약

인간관계를 파악하기 위한 사회망(Social Network)에 대한 연구는 사회과학 분야에서의 중요성으로 인해 꾸준히 진행되어 왔다. 학술 연구 분야에서도 연구자들 사이의 관계를 파악하는 것은 마찬가지로 중요성을 갖는다. 특히, 연구자들 사이의 공동 연구나 인용 형태에 대한 네트워크를 파악하는 것은 학술 연구에 대한 정책 입안이나 사업의 평가 등에 유용하게 활용될 수 있다. 본 연구에서는 국가과학기술 R&D 기반정보 온톨로지를 활용하여 공저자 관계와 인용 관계를 보여주는 두 가지 연구자 네트워크를 생성하는 방법에 대해 기술한다. RDF 트리플로 저장된 연구 성과물과 참여 연구 인력들에 대한 정보로부터 SPARQL 질의를 통하여 공저자 및 인용 관계의 연구자 쌍을 추론하고 후처리 과정을 통해 공저자 관계 및 인용 관계로 연결된 연구자 네트워크를 생성하는 과정을 소개한다.

### Abstract

The research about social network for analyzing human relationship has been steadily worked due to the importance in the social science field. It is also important that analyzing the relationship between researchers in the academic and research fields. Especially, the network by joint research or citation between researchers is useful to evaluating projects or making policy on academic and research fields. This paper describes a method that generates two kinds of researcher networks showing co-authorship and citation relationship between researchers based on national R&D reference information ontology. We infer pair of researchers in co-authorship or citation relationship by SPARQL query from the ontology which is composed of research outcomes and their participating researchers in RDF triples. By postprocessing, we construct researcher network which links researchers in co-authorship and citation relationship.

## I. 서론

인간관계를 파악하기 위한 사회망(Social Network)에 대한 연구는 사회과학 분야에서의 중요성으로 인해 꾸준히 진행되어 왔다[1]. 학술 연구 분야에서도 연구자들 사이의 인맥관계를 파악하는 것은 마찬가지로 중요성을 갖는다. 인용 네트워크와 협업 네트워크를 학술 연구 분야에서의 사회망의 예로 들 수 있다. 연구자들 사이의 공동 연구나 인용 형태에 대한 네트워크를 파악하는 것은 학술 연구에 대한 정책 입안이나 사업의 평가 등에 유용하게 활용될 수 있다.

공저자 관계 네트워크는 협업 네트워크의 하나로 공동으로 연구 성과를 발표한 연구자들의 관계를 표현하는 네트워크를 말한다. 이 네트워크를 분석함으로써 연구자들 사이의 학술 연구적 인맥관계를 파악하는 것이 가능하다. 특히, 연구 인맥이 넓은 전문가를 추천하거나, 반대로 연구 인맥관계가 없는 심사 평가자를 추천하는 데에 유용하게 활용될 수 있다.

인용 관계 네트워크는 연구 성과물 사이의 인용 관계를 바

탕으로 연구자들 사이의 인용-피인용 관계를 표현하는 네트워크를 말한다. 이 네트워크를 분석함으로써 공저자 관계 네트워크에서와 마찬가지로 비슷한 분야를 연구하는 연구자들을 그룹핑할 수 있을 뿐만 아니라, 그 분야에서의 연구 전문가를 파악하는 데에도 도움이 된다. 이는 일반적으로 피인용 회수가 클수록 우수한 연구자로 평가하는 데에 근거한다.

본 논문에서는 국가과학기술 R&D 기반정보 온톨로지[2]를 활용하여 추론을 통해 공저자 관계와 인용 관계를 보여주는 두 가지 연구자네트워크를 생성하는 방법에 대해 기술한다. 연구 성과물과 연구 인력에 대한 정보를 포함하는 이 온톨로지는 RDF 트리플로 저장되며 SPARQL 질의를 통하여 공저자 및 인용 관계의 연구자 쌍을 추론하고 후처리 과정을 통해 공저자 관계 및 인용 관계로 연결된 연구자 네트워크를 생성하는 과정을 소개한다. 온톨로지를 기반으로 하기 때문에 연구 성과물의 주제 혹은 분야, 성과물의 유형, 출판 연도 등을 달리하여 연구자 네트워크를 동적으로 생성할 수 있다.

## II. 관련 연구

학술 연구 분야에서 대부분의 사회망에 대한 연구는 정보 자원으로부터 사회망을 생성하고 이를 분석하며 시각화하는 내용이 주류를 이루고 있다.

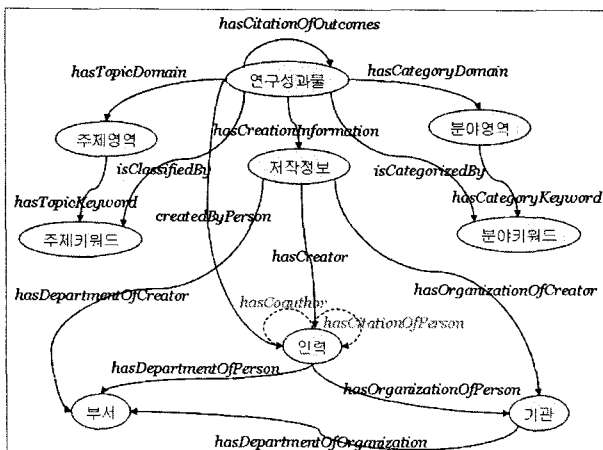
[3]은 10년간의 JCDL 학술대회 자료로부터 공저자 관계 네트워크를 생성하고 분석하였다. 이 네트워크를 표현하기 위해 이진 무방향 네트워크 모델과 가중치 부여된 방향 네트워크 모델의 두 가지를 도입하였고 네트워크를 분석하기 위한 여러 가지 수단을 제시하였다. 특히, 네트워크에서의 각 저자의 영향력을 가리키는 *AuthorRank*를 정의하였다.

[4]는 3년간의 JSAI 학술대회 자료를 바탕으로 연구자들의 협업 네트워크를 생성하고 년도별로 네트워크를 분석하여 연구자 네트워크의 진화과정을 보였다.

그러나 이러한 연구들은 정적인 네트워크를 다루는 데에 반해 본 논문에서는 온톨로지를 기반으로 추론 기능을 이용함으로써 동적인 네트워크를 생성할 수 있다. 온톨로지를 기반으로 하는 네트워크 분석에 대한 연구로 [5]가 있다. [5]는 기관 내의 멤버들 사이의 연결의 강도를 분석하여 잠재적인 COPs(Communités of Practice)를 찾아내는 것을 지원한다.

## III. 온톨로지

본 논문의 연구자 네트워크 추론은 국가과학기술 R&D 기반정보 온톨로지[8]를 바탕으로 한다. 그림 1은 이 온톨로지의 일부를 보인 것으로 온톨로지서 정의된 클래스들의 관계를 보여준다. 실선으로 표시된 속성은 RDF 트리플 형식으로 생성된 지식 혹은 규칙을 통해 확장된 지식을 의미하며 점선으로 표시된 속성 (*hasCoauthor*와 *hasCitationOfPerson*)은 두 가지의 연구자 네트워크를 생성하기 위해 추론을 통해 얻어지



▶▶ 그림 1. 온톨로지의 클래스와 속성

는 지식을 의미한다. '저작정보'는 연구 성과물에 대한 저자와 함께 저자 순위와 저자 가중치를 설정하기 위한 추가된 클래스이다.

그림 1에서 볼 수 있듯이, 각 연구 성과물에는 주제와 분야가 할당된다. 또한, 그림에서는 생략되었지만, 연구 성과물은 보고서, 논문, 특허 등의 유형으로 구분되며, 출판된 연도를 속성값으로 갖는다. 따라서, 본 온톨로지 기반의 연구자 네트워크는 연구 성과물의 주제 혹은 분야, 성과물의 유형, 출판 연도 등을 달리하여 동적으로 생성될 수 있다.

## IV. R-DBMS기반의 추론

본 논문에서 설명하는 추론을 통한 연구자 네트워크 생성은 R-DBMS 기반으로 하는 추론 서비스인 *OntoThink-K*<sup>®</sup>[6]의 한 부분을 구성한다. 이 장에서는 *OntoThink-K*<sup>®</sup>에 대해 간략히만 기술하며 보다 자세한 내용은 [6]과 [7]을 참고하기 바란다.

R-DBMS 기반의 *OntoThink-K*<sup>®</sup>는 크게 지식 생성, 확장, 추론의 세 부분으로 구성된다.

지식 생성에서는 온톨로지[8]와 URI 서버[2]로부터 자동으로 RDF 트리플 형식의 지식을 생성한다. 지식 확장에서는 RDF(S)에 기반한 추론 규칙을 이용하여 전방향 추론을 수행하여 기존의 지식을 확장한다. 생성 및 확장된 트리플 형식의 지식은 R-DBMS에 저장 및 관리된다. 지식 추론에서는 SPARQL (SPARQL Protocol And RDF Query Language) 형식의 질의를 R-DBMS에 저장된 RDF 트리플 상에서 처리하기 위해 SQL로 변환하고 실행하여 그 결과를 가공하여 사용자에게 제공한다.

## V. 연구자 네트워크 생성

본 논문에서는 두 가지 종류의 연구자 네트워크를 다룬다. 공저자 관계 네트워크는 협업 네트워크의 하나로 공동으로 연구 성과를 발표한 연구자들의 관계를 표현하는 네트워크이며 인용 관계 네트워크는 연구 성과물 사이의 인용 관계를 바탕으로 연구자들 사이의 인용-피인용 관계를 표현하는 네트워크이다.

연구자 네트워크 생성은 두 단계로 구성된다. 첫 번째는 SPARQL 질의를 통해 주어진 제약만 만족하는 연구자 쌍을 얻는 추론 과정이며, 두 번째는 연구자 쌍 목록으로부터 네트워크를 생성하는 후처리 과정이다.

연구자 네트워크는 연구 성과물의 주제 혹은 분야, 성과물의 유형, 출판 연도 등을 제약으로 설정하여 생성된다. 다시 말해,

특정 주제 혹은 분야에 속하면서 특정 유형에 해당되며 특정 기간에 출판된 연구 성과물들을 기준으로 공저자 관계 네트워크와 인용 관계 네트워크를 생성한다. 이러한 4가지 공통계약은 SPARQL 질의에서 다음과 같이 WHERE 절로 표현될 수 있다. 여기서 T는 특정 유형, S1과 S2는 특정 주제, C1과 C2는 특정 분야, Y1과 Y2는 특정 연도기간을 가리킨다.

```
WHERE {
  ?outcome rdf:type T .
  ?outcome isrl:isClassifiedBy ?S . FILTER (?S=S1 || ?S=S2) .
  ?outcome isrl:isCategorizedBy ?C . FILTER (?C=C1 || ?C=C2) .
  ?outcome isrl:hasPublicationYear ?Y .
  FILTER (?Y>=Y1 && ?Y<=Y2) .
}
```

▶▶ 그림 2. 4가지 공통계약에 대한 SPARQL 표현

### 1. 공저자 관계 네트워크

#### 1.1 공저자 관계 연구자 쌍 추론

공저자 관계 네트워크를 생성하기 위해 첫 단계로 RDF 트릴플 형태에 저장된 온톨로지로부터 SPARQL 질의를 통해 공저자 관계에 있는 연구자 쌍을 추론한다. 먼저 4가지 공통 계약을 만족하는 연구 성과물을 찾고 이 연구 성과물을 공동으로 저작한 연구자 쌍을 찾는 것이다. 온톨로지 상에는 연구 성과물과 이것의 저자 사이의 관계가 있으므로 이로부터 동일 연구 성과물에 대한 저자들의 모든 쌍을 얻을 수 있으며, SPARQL 질의에서는 그림 3의 WHERE 절과 같이 표현될 수 있다.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX isrl: <http://www.kisti.re.kr/isrl#>
SELECT ?creator1 ?cName1 ?creator2 ?cName2
WHERE {
  ?outcome rdf:type isrl:Article .
  ?outcome isrl:isClassifiedBy ?S .
  FILTER (?S=isrl:TOP_GE0006956) .
  ?outcome isrl:hasCreationInformation ?CI1 .
  ?CI1 isrl:hasCreator ?creator1 .
  ?CI1 isrl:orderOfCreator ?order1 .
  ?outcome isrl:hasCreationInformation ?CI2 .
  ?CI2 isrl:hasCreator ?creator2 .
  ?CI2 isrl:orderOfCreator ?order2 .
  FILTER (?order1<?order2) .
  FILTER (?CI1!=?CI2) .
  ?creator1 isrl:nameOfPerson ?cName1 .
  ?creator2 isrl:nameOfPerson ?cName2 .
  FILTER(?creator1!= 'http://www.kisti.re.kr/isrl#PER_0000000000') .
  FILTER(?creator2!= 'http://www.kisti.re.kr/isrl#PER_0000000000') .
}
```

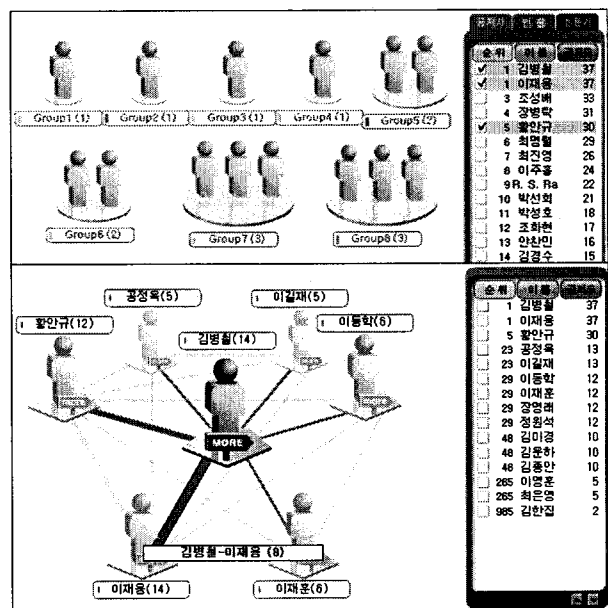
▶▶ 그림 3. 공저자 관계 연구자 쌍 추론을 위한 SPARQL질의 의 (주제, 유형 제약)

여기서, WHERE 절의 처음 두 줄은 특정 주제와 유형으로 연구 성과물을 제한하기 위한 것이다. “FILTER (?order1<?order2)”는 저자 쌍을 저자 순위에 따른 순서쌍으로 제한함으로써 저자 쌍이 순서만 바뀌어 중복으로 나타나는 것을 막는다. 또한, “FILTER (?CI1!=?CI2)”는 동일 저자가 공저자 쌍으로 잘못 추론되는 것을 막는다. 마지막 두 줄의 FILTER 구문은 특정 URI로 등록되지 않은 연구자 (즉, unknown 연구자)는 제외하기 위한 것이다.

#### 1.2 공저자 관계 네트워크 생성

연구 성과물을 기준으로 얻어진 모든 공저자 쌍으로부터 공저자 관계 네트워크를 만든다. 이 네트워크는 본질적으로 방향성이 없고 가중치를 갖는 그래프로 모델링될 수 있다.

연구자 사이의 연결 (즉, 공저자 관계)의 가중치는 공저자로 작성된 성과물의 수를 사용한다. 즉, 공동으로 더 많은 연구 성과물을 저작할수록 연결의 강도가 커지는 것이다. 또한 각 연구자에 대한 모든 연결의 가중치를 합산하여 각 연구자의 공저자 관계 점수를 계산함으로써 네트워크상에서 각 연구자들의 순위를 정한다. 다음으로 공저자 관계로 연결된 연구자들을 그룹핑한다. 그림 4는 이렇게 생성된 공저자 관계 그룹(위)과 Group7에 대한 네트워크(아래)의 예이다. 공저자 그룹은 공저자 관계 점수 순으로 상위 20명의 연구자들에 대한 그룹만을 보여주며 공저자 네트워크는 중심 연구자에 연결된 상위 6명의 연구자들에 대한 네트워크를 보여 준다. 이 네트워크에서 연결선의 굵기는 그 연결의 가중치를 나타내며 정확한 수치는 툴팁으로 표시된다.



▶▶ 그림 4. 공저자 관계 그룹(위) 및 네트워크(아래)

## 2. 인용 관계 네트워크

### 2.1 인용 관계 연구자 쌍 추론

공저자 관계 네트워크와 마찬가지로, 인용 관계 네트워크를 생성하기 위해 첫 단계로 RDF 트리플 형태로 저장된 온톨로지에서부터 SPARQL 질의를 통해 인용 관계에 있는 연구자 쌍을 추론한다. 먼저 4가지 공통 제약 조건을 만족하는 연구 성과물을 찾고 이 연구 성과물들 사이에 인용 관계가 있을 때 그 연구 성과물의 저자 쌍을 찾는 것이다. 이를 SPARQL 질의로 표현하면 그림 5의 WHERE 절과 같이 표현될 수 있다.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX isrl: <http://www.kisti.re.kr/isrl#>
SELECT ?citingOutcome ?citingCreator ?citingName ?citedCreator
?citedName
WHERE {
?citingOutcome isrl:isCategorizedBy ?citingC .
FILTER (?citingC=isrl:CAT_010100) .
?citingOutcome isrl:hasPublicationYear ?citingY .
FILTER (?citingY>=2004 && ?citingY<=2006) .
?citedOutcome isrl:isCategorizedBy ?citedC .
FILTER (?citedC=isrl:CAT_010100) .
?citedOutcome isrl:hasPublicationYear ?citedY .
FILTER (?citingY>=2004 && ?citingY<=2006) .
?citingOutcome isrl:hasCitationOfOutcomes ?citedOutcome .
?citingOutcome isrl:createdByPerson ?citingCreator .
?citedOutcome isrl:createdByPerson ?citedCreator .
?citingCreator isrl:nameOfPerson ?citingName .
?citedCreator isrl:nameOfPerson ?citedName .
FILTER(?citingCreator!= 'http://www.kisti.re.kr/isrl#PER_0000000000') .
FILTER(?citedCreator!= 'http://www.kisti.re.kr/isrl#PER_0000000000') .
}
ORDER BY ?citedCreator ?citingOutcome
    
```

▶▶ 그림 5. 인용 관계 연구자 쌍 추론을 위한 SPARQL질의 예 (분야, 연도 제약)

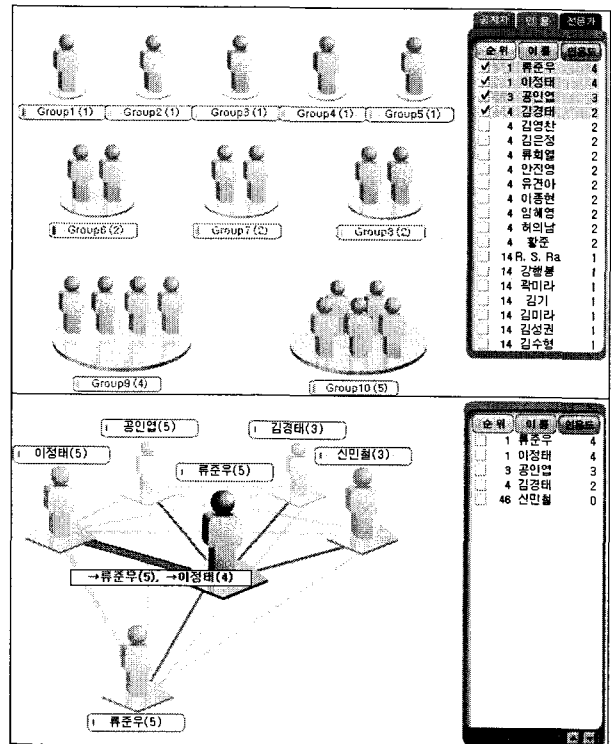
여기서, WHERE 절의 처음 8줄은 인용관계에 있는 두 성과물을 특정 분야와 연도로 제한하기 위한 것이다. SELECT 절에서 인용하는 성과물(citingOutcome)을 함께 얻는 것은 피인용 저자의 점수를 인용하는 성과물의 수로 계산하기 위함이다.

### 2.2 인용 관계 네트워크 생성

연구 성과물 사이의 인용관계를 기준으로 얻어진 모든 인용 관계 저자 쌍으로부터 인용 관계 네트워크를 만든다. 이 네트워크는 본질적으로 방향성이 있으며 가중치를 갖는 그래프로 모델링될 수 있다.

연구자 사이의 연결(즉, 인용 관계)의 가중치는 연구자 사이의 인용 관계 수로 정의한다. 그러나 인용 관계 점수를 계산할 때에는 인용하는 성과물의 저자 수로 나눈 가중치를 합산한다. 인용 관계 점수는 인용 저자의 수가 아닌 인용 성과물의 수를 따르는 것이 합리적이기 때문이다. 이 인용 관계 점수를

기준으로 네트워크상에서 각 연구자들의 순위를 정할 수 있다. 다음으로 인용 관계로 연결된 연구자들을 그룹핑한다. 그림 6은 이렇게 생성된 인용 관계 그룹(위)과 Group9에 대한 네트워크(아래)의 예이다. 인용 그룹은 인용 관계 점수 순으로 상위 20명의 연구자들에 대한 그룹만을 보여주며 인용 네트워크는 중심 연구자에 연결된 상위 6명의 연구자들에 대한 네트워크를 보여 준다. 이 네트워크에서 연결선의 굵기는 그 연결의 가중치를 나타내며 정확한 수치와 방향성은 툴팁으로 표시된다.



▶▶ 그림 6. 인용 관계 그룹(위) 및 네트워크(아래)

## VI. 결론

본 논문에서는 국가과학기술 R&D 기반정보 온톨로지를 활용하여 공저자 관계와 인용 관계를 보여주는 두 가지 연구자네트워크를 생성하는 방법에 대해 기술하였다. RDF 트리플로 저장된 연구 성과물과 참여 연구 인력들에 대한 정보로부터 SPARQL 질의를 통하여 공저자 및 인용 관계의 연구자 쌍을 추론하고 후처리 과정을 통해 공저자 관계 및 인용 관계로 연결된 연구자 네트워크를 동적으로 생성하는 과정을 소개하였다.

공저자 관계 네트워크는 방향성이 없는 가중치 그래프로 모델링하고 인용 관계 네트워크는 방향성 있는 가중치 그래프로 모델링하였다. 공저자 관계와 인용관계의 빈도를 가중치로 계산하고 이로부터 각 연구자의 공저자 관계 점수와 인용 관계 점수를 계산하여 네트워크상의 연구자들의 순위를 정하였다.

이러한 연구자 네트워크를 파악하는 것은 학술 연구에 대한 정책 입안이나 사업의 평가 등에 유용하게 활용될 수 있다.

#### ■ 참고 문헌 ■

- [1] S. Wasserman and K. Faust, "Social Network Analysis", Cambridge Press, 1994.
- [2] 정한민, 강인수, 구희관, 이승우, 성원경, "URI서버에 기반한 국가 R&D 기반정보 온톨로지 설계 및 구현", 정보관리연구 37(2), 2006.
- [3] X. Liu, J. Bollen, M. L. Nelson and H. Van de Sompel, "All in the family? a co-authorship analysis of JCDL conference(1994-2003)", [http://public.lanl.gov/liu\\_x/trend.pdf](http://public.lanl.gov/liu_x/trend.pdf).
- [4] Y. Matsuo and Y. Yasuda, "An Analysis of Researcher Network Evolution on the Web", ANDI05, 2005.
- [5] K. O'Hara, H. Alani and N. Shadbolt, "Identifying Communities of Practice: Analysing Ontologies as Networks to Support Community Recognition", IFIP02, 2002.
- [6] 정한민, 강인수, 이미경, 이승우, 성원경, "OntoThink-K<sup>®</sup>: DBMS 기반 추론 서비스", 한국정보과학회 춘계학술대회, 2006.
- [7] 이승우, 정한민, 성원경, "R-DBMS기반 추론 서비스인 OntoThink-K<sup>®</sup>에서의 SPARQL 질의 지원", 한국정보과학회 춘계학술대회, 2006.
- [8] 강인수, 정한민, 이승우, 김평, 성원경, "국가과학기술 R&D 기반정보 온톨로지", 한국콘텐츠학회 춘계종합학술대회, 2006.