



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



**저작자표시.** 귀하는 원저작자를 표시하여야 합니다.



**비영리.** 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



**변경금지.** 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

慶北大學校 工學碩士學位論文

# 온톨로지에 기반한 인스턴스들 간의 의미 관계 측정

大學院 컴퓨터工學科

韓 溶 振

2008年 6月

慶北大學校 大學院

# 목 차

1. 서 론 .....	1
2. 관련 연구 .....	3
2.1 온톨로지에 기반한 인스턴스 간의 의미적 관계.....	3
2.2 두 인스턴스 사이의 의미적 관계에 대한 서열.....	4
2.3 의미적 관계를 활용한 응용.....	5
3. 특정한 컨텍스트에 대한 인스턴스들 간의 관계.....	8
3.1 인스턴스들에 대한 서열화.....	8
3.2 인스턴스 간의 관계 정도 측정.....	9
3.3 인스턴스 간의 의미적 관계 정의.....	13
4. 인스턴스 간의 의미적 관계 정도 측정 방법.....	16
4.1 의미적 관계의 세가지 경우.....	16
4.2 두 인스턴스 간의 의미적 관계 정도 측정.....	18
4.2.1 목표 클래스를 공유하는 경우 .....	21
4.2.2 목표 인스턴스와 관계하는 의미적 관계 정도.....	25
5. 실험 및 평가 .....	28
5.1 이벤트 온톨로지를 이용한 실험.....	28
5.2 의미적 관계 정도 측정에 대한 평가.....	32
5.3 인스턴스 서열화에 대한 평가.....	35
6. 결론 및 향후 과제 .....	41
참 고 문 헌 .....	43
영 문 초 록 .....	47
부 록 .....	48

# 1. 서론

온톨로지는 기계가 접근할 수 있고, 사람이 이해할 수 있는 개념화에 대한 명시적인 명세(explicit specification of conceptualization) [8]를 제공한다. 온톨로지에서 사용하는 개념화는 실세계에 존재하는 것들 혹은 관념적인 것들 사이의 관계로써 표현된다. 따라서, 온톨로지에 기반한 시맨틱 웹 기술은 정보 간의 관계로부터 새로운 정보를 추론하거나 정보에 대한 재구성을 가능하게 한다 [22, 23, 25]. 어떤 대상들 간의 관계 정도를 측정하는 것은 이러한 시맨틱 웹 기술의 핵심 요소이다 [24].

웹 자원을 온톨로지로 표현하기 위한 모델인 OWL의 경우, 관심 대상이 되는 유일한 객체들을 인스턴스(instance)로 표현하고, 이들 간의 관계는 오브젝트 프라퍼티(object property)를 통해 표현한다. 웹상의 텍스트나 테이블의 정보 혹은 DB 등의 정보로부터 어떤 객체와 객체 간의 직접적인 관계 정보를 찾고 이것을 OWL에 정의된 개념 간의 관계 정보에 맞게 인스턴스화하는 연구가 진행되고 있다 [5, 6, 7, 8, 15, 16, 14].

이렇게 웹상의 정보를 온톨로지를 통해 의미적 관계로 표현함으로써 어떤 대상들 간에 명시적으로 들어나지 않았던 복잡하고 새로운 의미적 관계(semantic association)를 형성한다. 두 인스턴스 사이에 존재할 수 있는 의미적 관계는 그 자체로 새로운 정보이다. Boanerges [6]는 두 인스턴스 사이에 중요한 의미적 관계를 찾기 위해 의미적 관계 정도(degree of semantic association)를 측정하는 방법을 제안했다. 이들은 의미적 관계를 인스턴스 간의 패스에서 찾고, 각 패스에 대해 주로 컨텍스트 정보에 기반하여 관계 정도를 측정한다. 여기서 컨텍스트는 관계에 대한 의미를 명시적으로 표현해주는 클래스와 프라퍼티 간의 관계를 말한다. 어떤 컨텍스트를 만족하는 인스턴스 간의 패스는 여러개 있을 수 있다. 두 인스턴스 간의 관계는 인스턴스와 프라퍼티를 통해 연결되는 패스

로 나타나거나 온톨로지의 스키마를 통해 연결되는 패스로 나타나기도 한다. 본 논문은 이러한 정보로부터 특정 컨텍스트로 표현되는 두 인스턴스 사이의 관계 정도를 측정한다.

기존의 연구들은 주로 대상들 간에 직접적으로 연결된 정보만을 주요하게 다뤘다 [10, 24]. 하지만, 온톨로지에서는 인스턴스 간의 직접적인 연결은 없지만 스키마를 통해 연결되는 의미있는 관계가 있을 수 있다. 본 논문에서 제안하는 함수는 인스턴스 간에 직접적인 연결이 있는 관계는 물론, 스키마를 통해 연결되는 관계에 기반하여 관계 정도를 측정한다.

실험에서 책과 영화에 대한 인물들의 경험을 표현한 인벤트 온톨로지를 대상으로 특정 장르에 대한 인물들 간의 관계 정도를 측정한다. 제안한 방법의 성능 평가를 위해 직접적으로 연결된 관계만을 고려한 함수와 비교한다. 대부분의 경우 제안한 방법이 사람의 직관에 의한 결과와 비교적 높은 상관관계를 보였다.

본 논문의 구성은 다음과 같다. 먼저, 2장에서 관련 연구에 대해 설명하고, 3장에서는 본 논문에서 특정 컨텍스트가 주어졌을 때 두 인스턴스 간의 관계 정도를 찾는 문제를 논의한다. 이어서 4장에서 제안하는 두 인스턴스 간의 의미적 관계 정도에 대한 측정 방법을 기술한다. 5장에서는 실험을 통해 제안된 방법의 유용성을 입증해 보인다. 마지막으로 6장에서 결론을 맺는다.

## 2. 관련 연구

온톨로지는 프라퍼티를 통해 연결되는 인스턴스들 사이의 관계 정보를 표현한다. 웹의 자원을 온톨로지화할 때, 추출한 객체들은 인스턴스나 데이터 값에 해당하고, 관계 정보는 온톨로지의 프라퍼티에 대응한다. 두 인스턴스들 사이의 관계 정보가 구축되면, 인스턴스와 프로퍼티로 연결되는 패스로부터 어떤 두 인스턴스들 사이의 간접적인 의미적 관계를 관찰할 수 있다. 두 인스턴스들 사이의 직접적인 관계는 이미 알고 있는 정보이지만, 다른 인스턴스를 매개로해서 연결되는 간접적인 관계는 그 자체로 새로운 정보이다. 다음 절에서부터 온톨로지의 인스턴스들 간의 의미적 관계를 찾고, 두 인스턴스들 간의 관계 정도를 측정하는 연구에 대해 논의한다. 이어서 이러한 의미적 관계를 활용한 응용에 대해 설명한다.

### 2.1 온톨로지에 기반한 인스턴스 간의 의미적 관계

Kemafor [11]는 RDF 데이터 모델(RDF Data Model)을 대상으로 객체들 간의 복잡한 관계를 일반화하고 이것을 기반으로 세 가지 유형의 의미적 관계를 정의했다. 의미적 관계의 대상은 모든 인스턴스와 데이터값이다. RDF 데이터 모델의 오브젝트 프라퍼티와 데이터 타입 프라퍼티는 두 인스턴스 혹은 인스턴스와 데이터값 사이의 관계를 질의하는데 이용된다. 다시말하면, 프라퍼티의 연속(property sequence)으로 표현되는 관계 정보를 통해 찾고자 하는 관계 유형을 제약한다. 질의 결과는 객체와 프라퍼티의 패스로 표현되고, 질의 대상이 되는 의미적 관계는 다음의 세가지 형태로 정의한다.

- 1) 두 객체(인스턴스나 데이터 값) 사이에 연결이 있는 패스: 두 객체 사이에 간접적인 연결 정보를 제공한다.
- 2) 두 객체에서 시작하는 각각의 패스가 같은 프라퍼티의 연속으로 표

현되는 관계: 두 패스는 유사하다.

- 3) 두 객체에서 시작하는 각각의 패스가 유사하면서 같은 객체를 공유하는 경우

Kemafor의 경우, 자체 프로그램 알고리즘을 통해 의미적 관계에 대한 질의를 처리했다. 온톨로지로부터 찾은 의미적인 관계는 그 자체로 새로운 정보를 제공해 준다. 따라서, 의미적 관계를 찾기 위한 연구도 활발하게 진행되고 있다 [2, 3, 4, 12, 13]. 예를 들어, 기존의 SPARQL의 문법을 활용하거나 새로운 문법을 추가하여 패스에 대한 질의 처리 방법이 제안되었다 [12, 13].

만약, 두 인스턴스 사이의 의미적인 관계 정도를 측정할 수 있다면 온톨로지의 인스턴스들 사이의 관계에 대해 보다 구체적이고 다양한 분석이 가능할 것이다. 다음 절에서는 이러한 관점에서 두 인스턴스 사이에 존재할 수 있는 의미적 관계를 서열화하는 연구에 대해 소개한다.

## 2.2 두 인스턴스 사이의 의미적 관계에 대한 서열

온톨로지의 두 인스턴스는 스키마 상의 정의에 따라 다양한 의미적 관계를 가질 수 있다. Boanerges [6]은 인스턴스 층에서의 두 인스턴스 사이의 여러 가지 패스에 대해 의미적 관계 정도를 측정하는 여섯 가지 방법을 제안했다. 이들 여섯 가지 방법은 다음과 같다.

- 1) Context: 의미적 관계에 해당하는 패스에 대해 특정 클래스를 지나가는 경우 가중치를 준다.
- 2) Subsumption: 의미적 관계에 해당하는 패스에 대해 클래스 계층구조에서 하위에 있는 클래스를 지나가는 경우 가중치를 준다.
- 3) Trust: 의미적 관계를 구성하는 인스턴스들에 대한 신뢰정도를 부여하고 가중치를 결정한다.

- 4) Rarity: 의미적 관계를 구성하는 인스턴스와 프라퍼티(property)의 타입(Type)에 대해 해당 타입을 가지는 인스턴스들이 희소하게 나타날수록 높은 가중치를 준다.
- 5) Popularity: 의미적 관계를 구성하는 요소들 중 들어오는 관계(Incoming relationships)와 나가는 관계(outgoing relationships) 빈도수를 근거로 가중치를 준다.
- 6) Length: 의미적 관계가 가지는 패스의 길이를 가중치로 활용한다.

각각의 측정 방법은 사용자가 어떤 응용을 목적으로 어떤 질의를 하느냐에 따라 그 중요성이 달라질 수 있다. 따라서, Boanerges는 각 방법에 대해 가중치를 부여함으로써 도메인에 따라 유연하게 대처하도록 했다.

### 2.3 의미적 관계를 활용한 응용

온톨로지에 기반한 인스턴스들 간의 의미적 관계를 찾고, 관계 정도를 측정함으로써 새로운 정보를 추론할 수 있다. Amit [1]은 테러와 관련된 정보를 구축하고 있는 온톨로지로부터 테러와 관련된 인물을 찾기 위해 인스턴스들 간의 의미적 관계를 활용하고 있다. 2.1절에서 소개한 context를 활용하면 테러와 관련이 큰 클래스를 지나는 패스에 대해 높은 가중치를 줌으로써 테러와 관련된 인물이나 조직을 찾게 된다. 돈세탁 혐의를 찾기 위해 어떤 계좌에서 다른 계좌로 이체 거래가 많은 경우를 색출할 수 있을 것이다. 계좌를 클래스로 하고 이체하는 관계를 표현한 온톨로지가 있다면, 계좌간의 이체 관계에서 Length가 큰 패스에 대해 높은 가중치를 주게 된다.

추천 시스템에서는 개인화된 온톨로지를 구축하고 인스턴스 간의 의미적 관계로부터 개인 정보의 유사한 정도를 얻는데 활용한다. Yolanda [18]는 개인화된 TV프로그램 온톨로지를 기반으로한 TV프로그램 추천 방법을 제안했다. TV프로그램 온톨로지는 프로그램에 대한 계층적인 구



조를 표현하고 프로그램의 속성을 표현하기 위한 프라퍼티를 정의하고 있다. 사용자에게 어떤 프로그램을 추천하기 위해 사용자가 본 프로그램 간의 유사도를 측정하고, 두 사람에 대한 TV온톨로지의 인스턴스들 사이에 유사성을 측정하여 프로그램을 추천하는데 활용한다.

프로그램 간의 유사성을 측정하기 위해 2.1절에서 소개한 subsumption과 유사한 방법을 활용하였다. 즉, 두 프로그램이 공통적으로 속하는 클래스가 하위에 있을수록 유사성은 큰 값을 가진다. 이러한 방법은 WordNet을 이용한 단어 간의 유사성을 측정 [17]하거나 유전자 온톨로지(Geo-ontology)에서 인스턴스들 간의 유사성 [9]을 밝히는데 활용된 바 있다.

두 사람에 대한 TV 온톨로지의 인스턴스 정보에 대한 유사성은 온톨로지의 스키마 및 인스턴스를 벡터 표현하고 두 벡터 간의 상관관계로부터 유사성을 측정한다. 이 때, 인스턴스 간의 속성값이 같은 경우에 대해 가중치를 부여한다.

이상에서 살펴 본 것과는 다른 관점에서 온톨로지의 스키마 상에서 클래스와 클래스 사이의 관계 정도를 측정하는 연구도 있었다 [19, 20]. 이들은 인스턴스로부터 얻은 확실적인 정보를 이용해 프라퍼티로 연결되는 클래스와 클래스 사이의 관계 정도를 측정한다. 이것은 클래스의 계층 구조로부터 관계를 측정하는 것과는 구별된다.

지금까지 온톨로지의 인스턴스 혹은 객체 간의 관계를 찾고 관계 정도를 측정하는 방법, 그리고 이것들을 활용한 응용에 대해 살펴보았다. 이러한 연구들의 공통점은 의미적 관계에 대해 인스턴스 간에 연결이 있는 패스를 주요 대상으로 한다는 것이다. 하지만, 인스턴스 간에 직접적인 연결은 없지만, 클래스를 통해 간접적으로 연결되는 관계 또한 유용한 정보로 활용될 수 있다. 이러한 정보를 두 인스턴스간의 관계 정도를 측정하는데 사용하는 것은 새로운 접근 방법이다.

본 논문에서는 프라퍼티의 연속으로 찾은 기존 연구 [11]보다 제한적인 방법으로 의미적 관계를 찾는다. 즉, 클래스와 프라퍼티의 연속으로 표현된 특정한 컨텍스트 측면에서 두 인스턴스 간의 의미적 관계를 찾는다. 그리고 앞서 논의한 클래스를 통한 간접적인 관계와 인스턴스들 간의 직접적인 관계를 활용해서 의미적 관계 정도를 측정한다. 다음 장에서 특정한 컨텍스트에 대한 인스턴스들 간의 관계를 찾는 문제를 기존 연구와 비교해서 자세하게 논의한다.

### 3. 특정한 컨텍스트에 대한 인스턴스들 간의 관계

본 장에서는 특정한 컨텍스트에 대한 인스턴스들 간의 관계에 대해 논의한다. 먼저 어떤 인스턴스와 관계하는 다른 인스턴스들에 대한 서열화와 인스턴스 간의 관계 정도를 측정하는 문제에 대해 설명한다. 마지막으로 앞서 논의한 문제를 풀기위해 유의미하게 바라볼 의미적 관계를 정의한다.

#### 3.1 인스턴스들에 대한 서열화

본 절에서는 2.2절에서 소개한 두 인스턴스 사이의 의미적 관계[6]에 대한 서열화 문제와 비교해서 본 논문에서 목표로 하는 인스턴스들에 대한 서열화에 대해 설명한다. 2.2절에서 소개한 두 인스턴스 사이의 의미적 관계에 대한 서열화 문제[6]는 그림 1과 같이 표현할 수 있다.

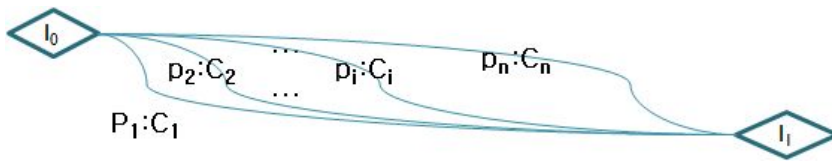


그림 1. 두 인스턴스 간의 의미적 관계에 대한 서열화.

인스턴스  $I_0$ 와  $I_1$  사이에는 서로 다른 컨텍스트를 갖는 의미적 관계들이 있을 수 있다. 그림 1에서 두 인스턴스 사이에 있을 수 있는 패스  $p_i$ 는  $C_i$ 라는 컨텍스트를 갖는다. 이렇게 서로 다른 컨텍스트를 갖는 개별적인 의미적 관계들에 대해 어느 것이 두 인스턴스  $I_0$ 와  $I_1$ 에 큰 의미가 있는지 혹은 흥미로운 관계인지를 찾는다.

반면, 본 논문에서는 특정한 컨텍스트에 대해 두 인스턴스가 갖는 의미적 관계 정도를 측정하고 어떤 인스턴스에 대해 관계하는 다른 여러 인스턴스들을 서열화한다. 특정한 컨텍스트  $C_k$ 에 대해 인스턴스  $I_0$ 가 다른

인스턴스들과 관계를 가질 때 이들에 대해 서열화하는 문제는 그림 2와 같이 표현된다.

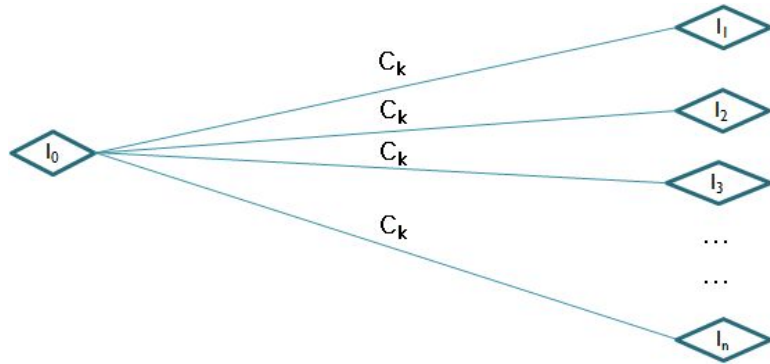


그림 2. 어떤 인스턴스에 관계하는 인스턴스들에 대한 서열화.

특정한 컨텍스트에서 인스턴스 간의 관계를 본다는 것은 하나의 관점에서 인스턴스 간의 관계 정도를 측정하는 것이다.

본 논문은 그림 2에서  $C_k$ 라는 동일한 컨텍스트 상에서 인스턴스  $I_0$ 가 다른 인스턴스들과 관계를 가질 때, 관계 정도에 따라 이들 인스턴스들을 서열화하는 것이 목표이다. 서열화된 결과는 추천 시스템에 응용되거나 인스턴스 간의 관계에 대한 분석에 활용될 수 있다. 다음 절에서는 주어진 하나의 컨텍스트에서 두 인스턴스 사이의 관계 정도를 측정하는 문제를 설명한다.

### 3.2 인스턴스 간의 의미적 관계 정도 측정

온톨로지의 인스턴스들은 스키마의 정의에 따라 프라퍼티와 인스턴스들의 패스로 표현되는 관계를 가진다. 본 논문은 스키마에서 표현된 컨텍스트에 대해 어떤 두 인스턴스들이 여러 개의 패스를 가질 때, 두 인스턴스들 간의 관계 정도를 측정한다.

실세계의 사람이나 어떤 대상은 온톨로지의 인스턴스로 표현될 수 있

다. 인스턴스는 URI를 통해 유일성을 보장받는다. OWL모델의 경우 인스턴스들 간의 관계는 온톨로지의 오브젝트 프라퍼티를 통해 표현된다. 어떤 두 인스턴스의 관계는 여러 개의 오브젝트 프라퍼티와 인스턴스들의 연속으로 나타날 수도 있다. 이것은 온톨로지 제작자에 의해 정의된 스키마를 통해 가능한 유형이 결정된다. 그림 3은 스키마에 정의된 클래스와 오브젝트 프라퍼티의 관계를 통해 인스턴스에서 나타날 수 있는 패스를 표현하고 있다.

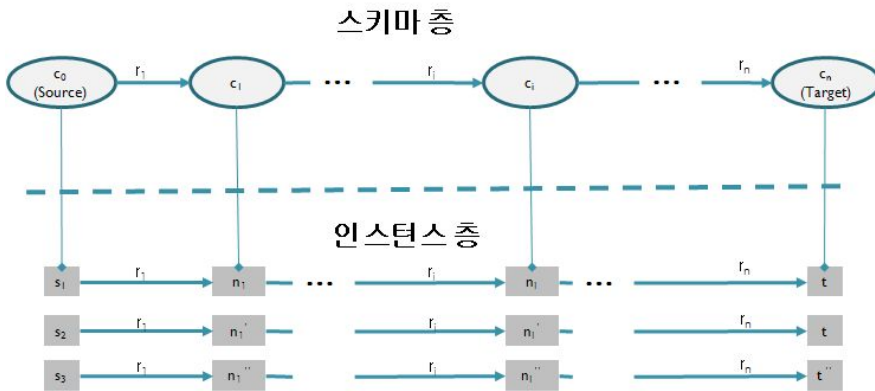


그림 3. 어떤 컨텍스트에 해당하는 인스턴스 층의 패스.

인스턴스  $s$ 는 클래스  $C_0$ 에 속하고, 스키마에 정의된  $C_0$ 에서  $C_n$ 까지 클래스와 오브젝트 프라퍼티의 관계에 따라 인스턴스  $t'$ ,  $t''$ ,  $t'''$ 로 가는 패스를 가진다. 이와 같이 스키마 레벨에서 정의된 클래스와 오브젝트 프라퍼티의 관계에 따라 인스턴스 레벨에서 어떤 특정 인스턴스에서 시작하는 여러 개의 패스가 나타날 수 있다. 본 논문은 이렇게 어떤 두 인스턴스  $s$ 와  $s'$ 가 인스턴스 층에서 동일한 컨텍스트의 패스를 여러 개 가질 때 이들 간의 관계 정도를 측정한다. 그림 4는 사람과 책 사이에 buy라는 property로 관계를 갖는 간단한 온톨로지 스키마와 스키마를 만족하는 인스턴스들 간의 패스 예를 이다.

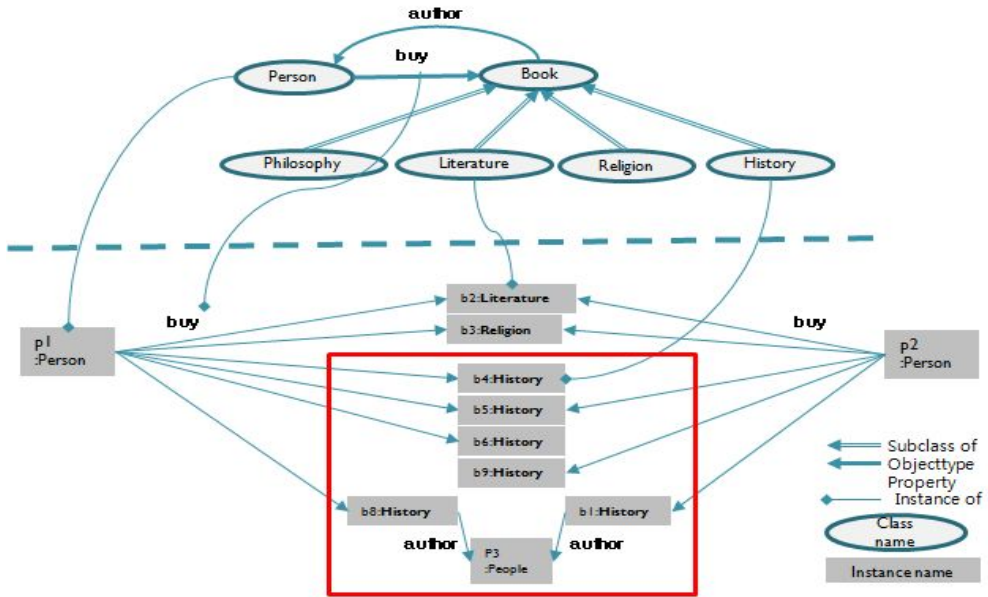


그림 4. 어떤 인스턴스에 관계하는 인스턴스들에 대한 서열화.

클래스 Person은 프라퍼티 buy를 통해 클래스 Book과 관계를 가진다. 따라서, Person-buy-Book이라는 클래스와 프라퍼티의 관계로부터 Person의 어떤 인스턴스와 Book의 어떤 인스턴스 사이에 buy로 연결되는 관계가 있을 수 있다. 또한, Person은 Book의 서브클래스(Subclass)인 History와 관계에서, Person-buy-History라는 클래스와 프라퍼티로 표현되는 관계를 만들어 낸다. Person-buy-Book이라는 스키마 상의 관계에 대해 인스턴스 P<sub>1</sub>은 총 7개의 패스를 가진다. Book의 subclass인 History와의 관계를 보면 총 5개의 패스를 가진다. 반면 P<sub>2</sub>의 경우 Book으로 가는 총 5개의 path 중 3개가 History에 관계한다. 주어진 예에서 본 논문에서 풀고자 하는 문제는 다음의 물음으로 요약된다.

“Person-buy-History라는 관계로부터 두 인스턴스 P<sub>1</sub>과 P<sub>2</sub>는 얼마나 의미적인 관계를 가질 것인가?”

2.2절에서 Boanerges[6]가 제안한 인스턴스 층에서 두 인스턴스들 간

의 관계 정도를 측정하는 방법 여섯가지를 소개했다. 이 방법들은 두 인스턴스 사이의 특정 패스에 대해 의미적인 관계 정도를 측정한다. 예를 들어, 그림 2에서  $P_1$ 으로부터 History의 인스턴스로 가는 각 패스에 대해 관계 정도를 측정한다. 그런데, 이렇게 동일한 컨텍스트를 가지는 각 패스에 대해 관계 정도를 측정할 경우, Context, Association Length, Subsumption, Rarity 이렇게 네 가지 방법은 모든 패스에 대해 동일한 값을 준다. Trust는 인스턴스나 패스에 대해 사람이 주관적인 값을 주는 경우로, 일반적인 온톨로지에서는 얻기 어려운 값이다. Popularity의 경우, 어떤 패스가 얼마나 많이 쓰이는 인스턴스로 구성되었는지를 측정하기 때문에 클래스와 프라퍼티가 제한된 관계에 대해서도 패스를 구성하는 인스턴스에 따라 다른 결과를 제시할 수 있다. 따라서, Popularity를 이용해 인스턴스 레벨에서의 각 패스에 대한 의미적인 관계를 측정하고 이를 바탕으로 스키마 레벨에서의 의미적인 관계를 측정하는 방법을 고려할 수 있다. 하지만, 두 인스턴스  $P_1$ 과  $P_2$ 가 같은 인스턴스를 공유하지 않을 경우에, 두 인스턴스가 얼마나 의미적인 관계를 가질 것인지에 대한 측정 방법은 제안된 바가 없다.

Harith[10]는 온톨로지에서 정형화된(formal) 관계에서 비정형(informal) 관계를 추론하는 문제를 제시했다. 그림 2에서 buy라는 관계로 연결되는 인스턴스들은 정형화된 관계를 갖는다. 이러한 관계로부터  $P_1$ 이나  $P_2$ 가 역사(History) 책에 대한 관심정도를 구하는 것은 비정형 관계를 추론하는 것이다. 이들은 온톨로지의 인스턴스들이 프라퍼티를 통해 연결되는 관계에 대해서 비정형 관계를 추론했다. 하지만, Harith가 제안한 방법은 Person-buy-History라는 관계로부터 두 인스턴스  $P_1$ 과  $P_2$ 가 인스턴스 레벨에서 연결이 있는 경우만 고려할 수 있다. 두 인스턴스 사이에 특정한 인스턴스를 공유하지 않더라도 패스가 유사한[12] 경우,  $P_1$ 과  $P_2$ 가 얼마나 의미적인 관계를 가질 것인지에 대한 단서가 될 수 있다.

즉 그림 4의 예에서  $P_1$ 과  $P_2$ 가 같은 역사책을 사지 않았더라도 역사책을 산 경우가 많을 때, 둘 사이에 역사에 관해 공통적으로 흥미를 가지는 정도는 그렇지 않은 경우에 비해 상대적으로 높을 것이다.

다음 절에서는 어떤 패스를 의미있는 것으로 볼 것인지를 설명하고 '의미적 관계'를 정의한다.

### 3.3 인스턴스 간의 의미적 관계 정의

2.1 절에서 Boanerges[12]가 정의한 온톨로지에서의 세 가지 의미적 관계를 소개했다. Boanerges는 의미적인 관계를 프라퍼티의 연속으로 제약해서 질의하는데 이용한다. 이 경우, 프라퍼티의 도메인이나 레인지(range)에 오는 정보에 따라 스키마 층의 관계에 대한 의미가 다르게 해석될 수도 있다. 예를 들어, 그림 2의 예에서 buy의 레인지로 올 수 있는 클래스로 Cloth(옷)나 Footwear(신발)이 있을 경우, 인스턴스 레벨에서 표현된 관계는 의미적으로 다른 관계로 해석될 수 있다. 즉, 'Person-buy-Book' 과 'Person-buy-Cloth'는 서로 다른 의미적 관계를 표현한다. 특정 의미적인 관계에 대해 어떤 인스턴스와 관계하는 다른 인스턴스들을 서열화하기 위해서는 의미적 관계를 보다 제한적으로 정의할 필요가 있다.

의미적 관계를 정의하기 위해 먼저 스키마 레벨에서 ConceptPath(CP)를 정의하고 인스턴스 층에서는 InstancePathSet을 정의한다. ConceptPath는 온톨로지에 정의된 클래스와 클래스 간의 관계 중 사용자가 관심을 가지는 컨텍스트에 해당한다. 즉, ConceptPath는 유한 개의 클래스와 오브젝트 브라벨티의 연속으로 정의된다. 앞으로 ConceptPath의 처음과 끝에 위치한 클래스를 각각 시작 클래스, 목표 클래스라고 부른다. ConceptPath를 수식으로 정의하면 다음과 같다.





InstancePathSet은 어떤 인스턴스  $s$ 에서 시작하는 인스턴스 층의 패스 중, 스키마 층의 패스 유형이 ConceptPath와 같은 패스들에 대한 집합이다. 수식으로 정의하면 다음과 같다.



$\text{InstancePathSet}(s_1, CP)$ 와  $\text{InstancePathSet}(s_2, CP)$ 의 각 패스들은 같은 CP를 가진다. 이러한 관계로부터 특정 컨텍스트를 통한 두 인스턴스 사이의 의미적 관계(Semantic Association)를 정의하면 다음과 같다.

특정 컨텍스트를 통한 두 인스턴스 사이의 의미적 관계:

두 인스턴스  $s_1$  과  $s_2$ 로부터 시작하는 어떤 두 패스가 같은 CP를 가지면 각 패스는 같은 목표 클래스를 공유한다. 이러한 경우,  $s_1$  과  $s_2$ 는 서로 의미적인 관계(Semantic Association)를 가진다.

다음 장에서는 의미적 관계를 근거로 해서 어떤 두 인스턴스 사이에 의미적인 관계 정도(Degree of Semantic Association)를 측정하는 방법에 대해 논의한다.

## 4. 인스턴스 간의 의미적 관계 정도 측정 방법

두 인스턴스 사이의 의미적 관계는 이들 두 인스턴스가 연결되는 형태에 따라 크게 세가지 경우로 나눌 수 있다. 즉, 목표 클래스만 공유하는 경우, 목표 인스턴스를 공유하는 경우, 그리고 목표 인스턴스는 다르지만, 속성을 공유하는 경우가 있다. 각 경우는 두 인스턴스로부터 시작하는 각 패스들의 집합으로 표현할 수 있다. 다음 절에서부터 이러한 각 집합을 정의하고, 이것을 기반으로한 의미적 관계 정도를 측정하는 방법을 논의한다.

### 4.1 의미적 관계의 세가지 경우

의미적 관계에 대한 정의로부터 의미적인 관계 정도는 구체적으로 세가지 경우로 나눌 수 있다. 목표 클래스(Target Class)만을 공유하는 경우, 목표 인스턴스(Target Instance)를 공유하는 경우, 그리고 목표 인스턴스를 공유하지는 않지만, 목표 인스턴스들의 속성이 같은 경우가 있다. 각 경우는 두 인스턴스가 연결되는 형태와 관계 정도에 따라 구분된다. 다음 그림 5는 두 인스턴스  $s_1$ 과  $s_2$  사이에 목표 클래스만을 공유하는 상황을 표현하고 있다. 점 선으로 표현된 원은 ConceptPath에서 정의된 클래스들의 연속을 의미한다.  $C_n$ 은 목표 클래스에 해당한다.

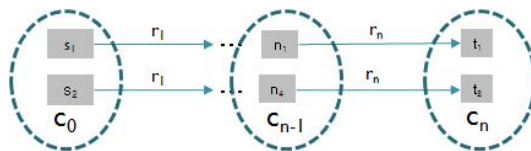


그림 5. 목표 클래스만 공유하는 경우.

목표 인스턴스를 공유하는 경우는 어떤 두 인스턴스가 같은 목표 인스턴

스를 매개로 의미적인 관계를 가진다. 그림 6은 목표 인스턴스를 공유하는 경우를 표현하고 있다.

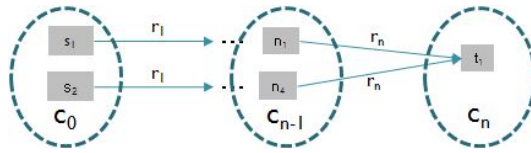


그림 6. 목표 인스턴스를 공유.

그림 6와 같이 목표 인스턴스를 공유하는 경우는 인스턴스  $s_1$ 과  $s_2$  사이에 여러 개가 있을 수 있다. 앞서 정의한 의미적 관계에 따르면 모든 의미적 관계는 목표 클래스를 공유한다. 목표 인스턴스를 공유하는 경우 또한 목표 클래스를 공유한다. 따라서, 목표 인스턴스를 공유하는 관계들의 의미적 관계 정도는 목표 클래스만을 공유하는 경우에 비해 더 크다고 가정할 수 있다.

목표 인스턴스들 간의 속성값이 일치하는 것은 두 인스턴스 사이에는 연결되는 패스가 존재하는 경우이기도 하다. 그림 7은 두 목표 인스턴스는 다르지만 속성값이 같은 경우를 나타내고 있다.  $C_{n+1}$ 은 목표 인스턴스의 속성  $r_{n+1}$ 에 대한 값으로 올 수 있는 인스턴스들에 대한 클래스이다.  $t_1$ 과  $t_2$ 는  $r_{n+1}$ 에 대한 속성값  $p$ 를 공유한다.

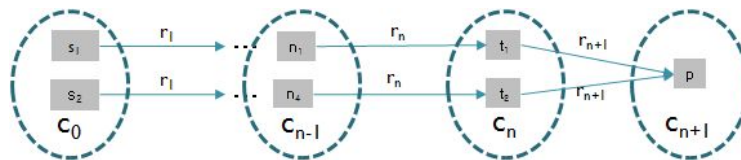


그림 7. 목표 인스턴스의 속성값 일치.

그림 7과 같이 두 인스턴스  $s_1$ 과  $s_2$ 에 대한 각 목표 인스턴스는 여러 개의 속성을 가질 수 있다. 그리고 이러한 속성값의 일치는 여러 개의 홑

(hop)을 거쳐서 나타날 수도 있다. 예를 들면,  $C_{n+1}$ 에 해당하는 인스턴스들 사이에 속성값이 일치하는 것은 두 개 홉을 거친 경우에 해당한다.

한편, 목표 인스턴스의 속성값을 공유하는 경우도 목표 클래스를 공유한다. 따라서, 의미적 관계 정도는 목표 인스턴스를 공유하는 경우에 비해 작지만, 목표 클래스만을 공유하는 경우에 비해서는 크다고 볼 수 있다.

이러한 세가지 의미적 관계의 경우를 인스턴스와 속성으로 구성되는 패스들에 대한 집합으로 표현하면 다음과 같다.



$P^{sc}$ ,  $P^{si}$ ,  $P^{sv}$ 는 서로 교집합을 가지지 않는다. 이들에 대한 합집합은 어떤 컨텍스트에 대한 두 인스턴스 사이의 의미적 관계로 표현되는 모든 패스들의 집합과 같다.

다음으로 이러한 의미적 관계의 세가지 경우에 대한 의미적 관계 정도를 측정하는 방법을 논의한다.

## 4.2 두 인스턴스 간의 의미적 관계 정도 측정

4.1 절에서는 온톨로지에서 나타날 수 있는 인스턴스 간의 관계 정도를 세 가지 경우로 정의했다. 두 인스턴스 사이의 의미적인 관계 정도는 목표 클래스나 목표 인스턴스를 공유하는 형태에 따라 서로 다른 값을 가질 수 있다. 인스턴스 간의 관계 정도 측정 방법을 설명하기에 앞서 3.2절에

서 클래스 Person과 클래스 Book 사이에 buy라는 관계로 표현된 온톨로지의 예를 다시 살펴보자.

그림 1의 예에서 사람은 같이 읽었던 책의 종류에 따라 각기 다른 의미를 부여할 수도 있다. 하지만, 그러한 정보는 주관적이기 때문에 온톨로지에 명시적으로 표현하는 것은 어렵다. 또한, 사람이 아닌 객체 간의 의미적인 관계를 측정한다면 공유하는 인스턴스에 따라 관계 정도를 구분하는 것은 어려운 일이다. 예를 들어, Company-produce-Product라는 관계가 있다고 할 때, 기업 간에 생산하는 품목이 같다면 두 회사 간에는 경쟁 관계에 있거나 각 회사에서 공통된 품목에 대해 관심을 가질 확률이 높다고 볼 수 있다. 하지만, 공통적으로 생산하는 품목이 여러 가지가 있을 때 각각에 대해 의미를 두는 정도를 구분하기는 어렵다. 다만, 두 회사 간에 공통으로 생산하는 품목이 있다면 관계 정도가 높다고 볼 수 있다. 따라서, 온톨로지의 인스턴스들 간에 공유하는 인스턴스가 있다면 공유하는 개수에 비례해서 관계 정도가 높다고 가정한다. 또한, 같은 인스턴스는 아니지만 인스턴스 간의 속성이 같다면 어느 정도 관계 정도가 높다고 할 수 있다. 두 인스턴스가 같은 CP를 가진다면 공통된 클래스를 갖게 되고, CP가 같다는 것은 각 인스턴스가 목표 클래스에 대해 관계를 갖는 의미가 동일하다. 따라서, 인스턴스를 공유하지 않더라도 두 인스턴스 사이에는 어느 정도 의미적인 관계가 있다고 가정한다. 만약, 인스턴스 a가 각 인스턴스b, c와의 관계에서 공유하는 인스턴스가 같다면 b, c가 인스턴스 a에 관계하는 정도는 서로 다른 목표 인스턴스들에 대해 공통된 속성을 갖는 개수와 목표 클래스로 향하는 패스가 얼마나 많은지에 따라 결정될 것이다.

이러한 측면에서 본 논문에서는 어떤 컨텍스트 상에서 두 인스턴스 사이의 의미적인 관계 정도를 다음의 세 가지 요소로 구성한다. P는 의미적 관계를 가지는 패스들에 대한 집합을 의미한다.

$v_p^c$  : 목표 클래스를 공유하는 것에 대한 의미적 관계 정도

$v_p^i$  : 목표 인스턴스를 공유하는 것에 대한 의미적 관계 정도

$v_p^p$  : 서로 다른 두 목표 인스턴스가 하나 이상의 홉을 통해 인스턴스를 공유하는 것에 대한 의미적 관계 정도

두 인스턴스 사이의 의미적 관계를 가지는 패스들의 집합에 대한 의미적 관계 정도는 이러한 세 가지 요소에 선형적으로 의존한다고 가정한다. 이것을 식으로 표현하면 다음과 같다.



이러한 의미적 관계 정도로부터 두 인스턴스 사이에 나타나는 모든 의미적 관계에 대한 패스들의 집합 P에 대한 의미적 관계 정도는 다음 식과 같이 표현된다.



즉, 목표 클래스를 공유하는 경우에 대한 의미적 관계 정도는 목표 클래스를 공유하는 것에 대한 의미적 관계( $v_p^c$ )에 의해서만 결정된다. 반면, 목표 인스턴스를 공유하는 경우에 대한 의미적 관계 정도는  $v_p^c$  와 목표 인스턴스를 공유하는 것에 대한 의미적 관계 정도( $v_p^i$ )에 의해 결정된다. 마지막으로 목표 인스턴스의 속성이 같은 경우에 대한 의미적 관계 정도는  $v_p^c$  와  $v_p^p$  에 의해 결정된다.

다음 절에서부터 목표 클래스를 공유하는 것에 대한 의미적 관계 정도와 목표 인스턴스와 관계하는 의미적 관계 정도를 측정하는 방법에 대해 논의한다.

#### 4.2.1 목표 클래스를 공유하는 경우

목표 클래스를 공유하는 경우에 대해 두 인스턴스 사이의 관계 정도를 구하는 것은 본 논문에서 새롭게 제안하는 방법이다. 이것은 두 인스턴스가 목표 클래스에 얼마나 공통적으로 의미를 부여하는지에 근거해서 의미적 관계 정도를 측정한다.

먼저 어떤 인스턴스가 목표 클래스에 의미를 부여하는 정도에 대해 살펴보자. 온톨로지가 실세계의 객체들 간의 의미적인 관계를 표현한다고 할 때, 온톨로지의 인스턴스는 오브젝트 프라퍼티를 통해 관계 정보를 표현한다. 예를 들어, 3.2절의 그림 4에서 어떤 사람이 어떤 종류의 책을 샀는지에 대한 정보를 반영한다. 어떤 사람이 역사에 관심이 많다면 철학이나 소설 책 보다 역사 책을 더 많이 살 것이다. 따라서, 그 사람의 역사책에 대한 관심 정도는 책을 사는 사건들 중에서 역사 책을 산 사건의 상대적인 크기로 이해할 수 있다.

이러한 관점에서 온톨로지의 인스턴스가 어떤 목표 클래스에 의미를 부여하는 정도는 그 오브젝트 프라퍼티를 통해 갈 수 있는 패스의 경우의 수에 대해 해당 클래스의 인스턴스로 가는 패스의 수에 상대적으로 비례한다고 가정한다. 어떤 인스턴스  $s$ 가 클래스  $C$ 에 의미를 부여하는 상황은 그림 8과 같이 표현할 수 있다.

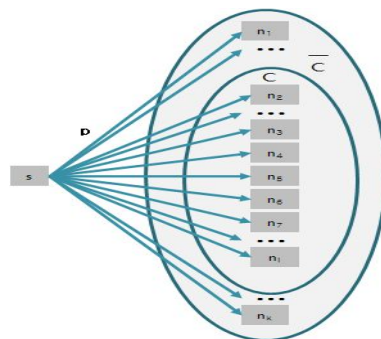




그림 8. 어떤 인스턴스  $s$ 가 클래스  $C$ 에 의미를 부여하는 정도.

CP가 한 개의 홉(hop)으로 정의된다면 어떤 인스턴스  $s$ 가 오브젝트 프라퍼티  $p$ 를 통해 목표 클래스  $C$ 에 의미를 부여하는 정도는 가정에 의해 식 3과 같이 표현된다.  $w_s^p$ 는 인스턴스 층에서의 한 개의 패스에 대해 인스턴스  $s$ 가 목표 클래스에 의미를 부여하는 정도이다. 이것을 식으로 표현하면 다음과 같다.

$w$ 를 계산할 때 분자와 분모에 1을 더한 것은 패스가 전혀 없는 경우에 대해 0이되는 것을 보정하기 위해서이다.

CP가 여러 개의 홉으로 구성된 경우 각 홉에 대해 식 3의 방법을 적용한다. CP에 대해 시작 인스턴스가 목표 클래스에 의미를 부여하는 정도를 구하기 위해 그림 8과 같은 상황을 가정할 수 있다.

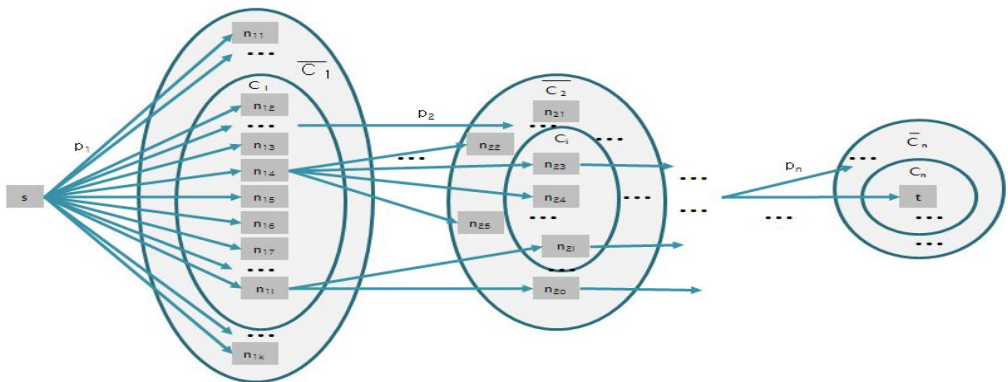


그림 8. CP의 홑이 2 이상일 때.

그림 8에서 시작 인스턴스  $s$ 에서 목표 인스턴스  $t$ 까지 갈 때,  $n$  개의 프라퍼티 연속과 클래스 연속을 거치게 된다. 시작 인스턴스  $s$ 에서 목표 인스턴스  $t$ 까지 가는 패스가 있을 때,  $s$ 가 목표 클래스  $C_n$ 에 의미를 부여하는 정도는 각 홑에서 식 3의 방법을 적용하여 한 개의 패스에 대해  $s$ 가 목표 클래스에 의미를 부여하는 정도를 표현한다. 예를 들어, 그림 8에서 시작 인스턴스  $s$ 에서 목표 인스턴스  $t$ 까지 가는 패스가 있고, 첫 번째 홑을 거쳤을 때 식 3을 적용하면 시작 인스턴스는  $s$ 가 되고, 목표 클래스는  $C_1$ 이 된다. 두 번째 홑을 보면 시작 인스턴스를  $n_{14}$ 로 하고 목표 클래스를  $C_2$ 로 해서 식 3을 적용할 수 있다. 이와 같은 방법으로 CP를 만족하는 한 개의 패스에 대해 인스턴스  $s$ 가 목표 클래스에 의미를 부여하는 정도를 벡터로 표현한다. 이러한 벡터를 InterestVector라 부르고 다음과 같이 정의한다.

InterestVector의 요소  $w$ 는 각 홑에 대해 이전 홑의 인스턴스가 해당 홑의 클래스에 의미를 부여하는 정도를 의미한다. 이러한 InterestVector는 시작 인스턴스  $s$ 로부터 CP를 만족하는 모든 패스의 개수만큼 얻을 수 있다.

다음으로 CP를 만족하는 모든 패스로부터 시작 인스턴스  $s$ 가 목표 클래스  $C_n$ 에 의미를 부여하는 정도인 InterestVector<sup>sum</sup>구한다. 이것은 모든 InterestVector에 대해 각 요소들을 더해서 구한 것이다. 이것을 식으로 표현하면 다음과 같다.

다음으로 두 인스턴스로부터 얻은 각각의 InterestVector<sup>sum</sup>을 이용해 두 인스턴스가 목표 클래스에 공통적으로 의미를 부여하는 정도를 구한다.

먼저, InterestVector IV로부터 각 인스턴스가 목표 인스턴스에 의미를 부여하는 정도를 DOI(Degree of Interest)로 정의한다. 다음은 DOI(IV)에 대한 정의와 구하는 방법을 표현하고 있다.

본 절의 목표는 두 인스턴스가 공통적으로 목표 클래스에 의미를 부여하는 정도를 찾는 것이다. 먼저, 앞서 정의한 각 인스턴스에 대한 InterestVector<sup>sum</sup>을 구한다. 다음으로 두 인스턴스로부터 시작하는 패스의 각 요소들이 다음 클래스에 얼마나 공통적으로 관심을 가지는지를 표현한 InterestVector<sup>common</sup>을 구한다. 다음은 InterestVector<sup>common</sup>을 정의한 것이다.

InterestVector<sup>common</sup>의 각 요소는 두 InterestVector<sup>sum</sup>의 대응하는 요소에 대해 조화평균을 구한 값으로 한다. 조화평균은 합이 같은 임의의 두 수에 대해 두 수의 차가 작을수록 높은 평균값을 부여한다. 따라서 의미를 부여한 정도의 합이 같더라도 두 값의 크기가 비슷할 때 더 많은 값을 부여할 수 있다.

이렇게 구한 InterestVector<sup>common</sup>에 대한 DOI를 구함으로써 최종적으로  $v^c$ 을 얻을 수 있다. 이 과정을 식으로 표현하면 다음과 같다.

본 절의 내용은 다음과 같이 요약된다. 특정 컨텍스트에 대한 두 인스턴스 사이의 의미적 관계 정도를 측정하기 위해 먼저 각 인스턴스가 목표 클래스에 대해 의미를 두는 정도를 구한다. 다음으로 두 인스턴스 사이에 공통된 의미 관계 정도를 측정한다. 목표 클래스를 공유한다는 점에서 공통된 관계 정도를 조화 평균을 활용해서 구했다.

#### 4.2 목표 인스턴스와 관계하는 의미적 관계 정도

앞에서 논의한 바와 같이 목표 인스턴스와 관계하는 의미적 관계 정도는 두 가지, 목표인스턴스를 공유하는 것에 대한 의미적 관계 정도( $v^j$ )와 목표 인스턴스의 속성을 공유하는 것에 대한 의미적 관계 정도( $v^p$ )가 있다. 이들 각각은  $P^{si}$ 와  $P^{sp}$ 에 관계한다.  $P^{sc}$ 의 경우, 목표 클래스만을 공유하기 때문에 목표 인스턴스와 관계하는 의미적 관계 정도는 가지지 않는다.

먼저,  $v^p$ 를 구하기 위해 속성의 대상을 정의한다. OWL모델에서 어떤 인스턴스는 오브젝트 프라퍼티(Object property)와 데이터타입 프라퍼티

(Datatype property)를 가진다. 같은 클래스에 속하는 어떤 서로 다른 두 인스턴스들은 모든 프라퍼티의 값이 같을 수도 있다. 따라서, 각 인스턴스의 속성으로 인스턴스 아이디를 추가한다. 두 개의 목표 인스턴스가 속성을 공유할 경우, 전체 속성의 개수에 대해 공유하는 속성의 개수만큼 그 둘은 유사하다고 볼 수 있다. 이러한 관점에서 목표 인스턴스의 속성을 공유하는 경우에 대한 의미적 관계 정도를 구하면 다음과 같은 식으로 표현된다.

위 식에서 두 인스턴스는 각 홑에서 일치하는 속성의 개수가 많을수록 높은 의미적 관계를 가진다. 홑의 길이가 길어지면 그 만큼 의미적 관계 정도는 상쇄되어 측정된다.

목표 인스턴스를 공유하는 경우는 모든 속성값이 같고, 동일한 목표 인스턴스를 갖는 경우로 이해할 수 있다. 다시 말하면, 인스턴스의 아이디까지 같다면 결국 목표 인스턴스를 공유하는 경우와 같다. 따라서, 목표 인스턴스를 공유하는 것에 대한 의미적 관계 정도는 다음과 같이 풀이 된다.

위 식에서 목표 인스턴스를 공유하는 경우에 대한 의미적 관계 정도는 목표 인스턴스를 공유하는 패스에 대한 집합의 원소의 개수와 같다.

다음 장에서는 event-ontology를 이용한 실험 방법을 설명하고 제안한 의미적 관계 정도 측정 방법에 대한 성능을 평가한다.

## 5. 시험 및 평가

본 장에서는 크게 두 가지 시험을 한다. 먼저, 어떤 질의와 관련하여 두 인스턴스 사이의 의미적 관계 정도에 대한 측정 방법의 성능을 평가한다. 다음으로 의미적 관계 정도를 이용한 활용으로 어떤 인스턴스와 관계하는 다른 인스턴스들에 대한 서열화 문제를 풀고, 그 결과에 대해 평가한다. 다음 절에서는 먼저 이러한 시험에 사용할 이벤트 온톨로지와 시험 방법에 대해 설명한다.

### 5.1 이벤트 온톨로지를 이용한 시험

이벤트 온톨로지는 인물들의 책이나 영화에 대한 경험 정보를 표현한다. 이벤트 온톨로지의 스키마와 스키마를 따르는 인스턴스 예를 그림 9에서 표현하고 있다.

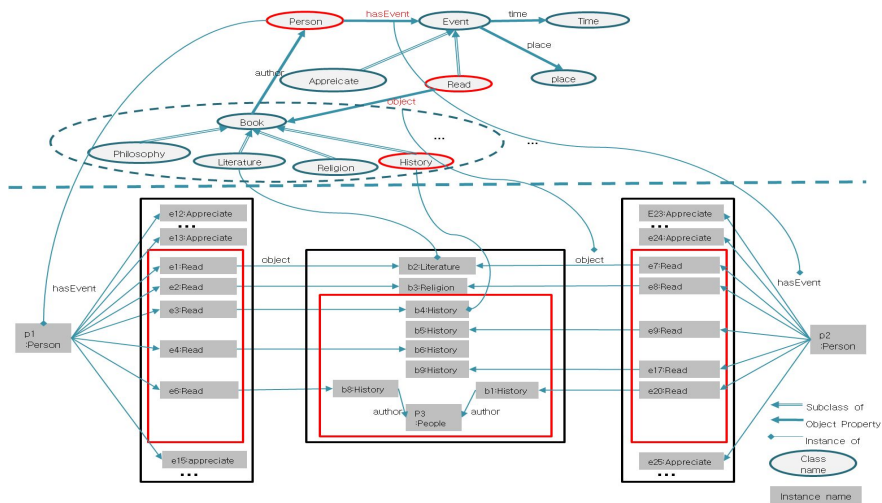


그림 9. 이벤트 온톨로지와 인스턴스 예

이벤트 온톨로지의 클래스 Event는 Appreciate과 Read를 서브클래스

로 가진다. 각 서브 클래스는 프라퍼티인 object의 값으로 Movie와 Book을 갖는다. 그림 9에서는 Read에 대한 object의 값인 클래스 Book만 표현하고 있다. 클래스 Book은 Philosophy, Literature 등의 책에 대한 분류 정보들을 서브클래스로 가진다. 클래스 Person의 경우, hasEvent를 통해 클래스 Event와 관계를 가진다. Person의 인스턴스는 실제 영화를 보고 책을 읽었던 인물들에 해당한다.

이러한 스키마 표현에 따라 인스턴스 층에서 나타날 수 있는 인스턴스들의 예를 그림 9의 아래 쪽에서 표현하고 있다. Person의 인스턴스 p1과 p2는 프라퍼티 hasEvent를 통해 여러 개의 Event 인스턴스를 가진다. Event 인스턴스들은 하위 클래스인 Read 혹은 Appreciate으로 구분되고, 각 인스턴스들은 프라퍼티 object을 통해 Book 혹은 Movie의 인스턴스들과 관계를 가진다. 그림 9에서는 Read의 인스턴스들이 Book의 인스턴스와 관계를 가지는 경우만을 표현하고 있다.

사용자는 이러한 온톨로지 정보로부터 역사를 주제로 자신과 논의할 만한 사람을 찾을 수 있다. 스키마 층에서 다른 색으로 표시되어 있는 [Person, hasEvent, Read, object, History]는 인물이 역사 책과 관계를 가지는 것을 표현한 ConceptPath에 해당한다. 인스턴스 층에서는 이러한 ConceptPath를 따르는 InstancePathSet을 표현하고 있다. 프라퍼티 hasEvent와 관계하는 큰 사각형은 클래스 Event로 가는 인스턴스 집합을 의미하고 그 속에 작은 사각형은 Event로 가는 인스턴스 중 서브 클래스인 Read로 가는 인스턴스 집합을 의미한다. 각 Read의 인스턴스에서 Book으로 가는 인스턴스와 History로 가는 인스턴스 가는 인스턴스 같은 방법으로 표현하고 있다. History는 목표 클래스에 해당하고 그림 9에서 목표 클래스로 가는 목표 인스턴스들 간의 관계를 확인할 수 있다.

이러한 이벤트 온톨로지에 적합한 실제 데이터를 인스턴스화 하기위해 11명의 인물을 대상으로 설문조사를 실시하였다. 설문조사 과정은 다음과



같다. 1) 인물들의 경험 대상이 되는 책과 영화의 리스트를 수집한다. 2) 실험에 참가한 11 명의 인물들에게 수집된 책과 영화 리스트를 주고 흥미가 가거나 인상 깊게 봤던 책, 영화 리스트를 선택 받는다. 3) 실험에 참가한 사람들에게 질의를 주면, 참가자들은 질의와 관련해 흥미 있게 이야기할 만한 책 혹은 영화에 대해 0에서 3 사이의 정수값으로 점수를 준다.

설문 조사 과정 1)과 2)를 통해 온톨로지를 구축하기 위한 정보를 얻는다. 총 221 권의 책과 913 편의 영화 리스트를 수집하였고, 각 인물들로부터 책의 경우 40권에서 최고 49권의 리스트를 받았고 영화는 27편에서 281편의 리스트를 받았다.

설문 조사 과정 3)은 인물들이 어떤 질의와 관련하여 흥미를 공유하는 정도를 결정하고 이것을 근거로 어떤 인물과 관계하는 다른 인물들을 서열화한 결과를 얻기 위한 것이다. 먼저, 어떤 인물이 질의와 관련하여 책 혹은 영화 리스트에 점수를 부여하고 인물들이 체크한 책 혹은 영화리스트에 대해 부여된 점수에 따라 합을 구함으로써 의미적 관계 정도를 결정한다. 이렇게 구한 의미적 관계 정도에 따라 어떤 질의에 대해 어떤 인물과 관계하는 다른 인물들을 서열화한다. 이러한 결과를 의미적 관계 정도 측정에 대한 성능을 평가하고 서열화 결과를 평가하는데 활용한다.

과정 3)에 이용할 질의는 다음과 같다.

- 1) 심리철학에 관한 주제를 가지고 이야기를 할 때 호감이 가는 사람을 서열화 한다면?
- 2) 경제경영에 대해 논의할 때 호감이 가는 사람을 서열화 한다면?
- 3) 액션물 주제로 이야기할 때 호감이 가는 사람을 서열화 한다면?
- 4) 애니메이션물 주제로 이야기할 때 호감이 가는 사람을 서열화 한다면?

질의는 이벤트 온톨로지에서 두 인스턴스 간의 의미적 관계 정도를 측

정하고자 하는 컨텍스트에 해당한다. 즉, 컨텍스트는 각 인물들이 다른 인물들과 책이나 영화에 대해 흥미를 가지는 관계이다. 같은 대상을 경험했거나 비슷한 경험을 많이 했을 때 흥미도는 높은 값을 가질 것이라고 가정한다. 이러한 가정은 본 논문에서 두 개의 시작 인스턴스가 같은 목표 인스턴스를 공유하거나 목표 인스턴스들의 속성값이 같은 경우, 혹은 목표 클래스를 공유하는 경우로 이해할 수 있다. 이들 컨텍스트로부터 의미적 관계를 얻기 위한 ConceptPath를 표 1에서 정리하고 있다.

표 1. 컨텍스트에 대한 ConceptPath.

컨텍스트(질의)	ConceptPath
1	[Person, hasEvent, Read, Psychology_Philosophy]
2	[Person, hasEvent, Read, Economics_Business]
3	[Person, hasEvent, Appreciate, Action]
4	[Person, hasEvent, Appreciate, Animation]

표 1의 ConceptPath로부터 각 인물 인스턴스들이 인스턴스층에서 가지는 InstancePathSet을 얻을 수 있다. 이러한 정보로부터 제안한 방법을 적용하여 질의와 관련해 어떤 인물과 관계하는 다른 인물들을 서열화할 수 있다.

본 논문에서 인스턴스 층에서는 연결이 없더라도 스키마 층에서 보면 관계를 가지는 경우를 고려하여 두 인스턴스 사이에 관계 정도를 측정하는 방법을 제안했다. 따라서, 두 인스턴스 사이에 인스턴스 층에서 연결이 있는 경우만 고려한 방법과 비교하여 제안한 방법의 성능을 비교 분석한다. 인스턴스 층에서 연결이 있는 경우는 본 논문에서 고려하는 의미적 관계 중 목표 인스턴스를 공유하는 경우에 해당한다.

질의와 관련하여 두 인스턴스 사이에 의미적 관계 정도를 측정한 성능은 사람에 의한 결과와의 상관관계를 통해 평가한다. 서열화 결과는 사람에 의한 결과와의 순위상관관계를 비교해서 평가한다. 서열화 결과에 대한

관련 정도를 측정하는 방법으로 스피어만의 순위 상관계수(Spearman Rank Correlation Coefficient)[21]가 있다. 스피어만의 순위 상관계수는 두 개의 순서형 자료에 대해 순서가 완전 일치하는 경우 1을, 완전 반대인 경우 -1을 반환한다. 스피어만의 순위 상관계수는 다음과 같이 구한다.

## 5.2 의미적 관계 정도 측정에 대한 평가

먼저, 11 명을 대상으로 한 설문조사를 통해 각 인물들이 관심 있게 본 영화와 책의 리스트를 받았고, 6 명의 사람들에게 질의와 관련된 다른 인물과 흥미를 공유할 만한 정도에 대한 측정값을 얻었다.

첫 번째 질의로부터 심리철학에 관해 이야기할 때 인물들에 대한 호감 정도를 사람이 결정한 값은 표 2와 같다.

표 2. 심리 철학에 관해 이야기할 때 인물들에 대한 호감정도.

실험대상	p1	p2	p3	p4	p5	p6	
학습	p1	-	19	13	13	17	9
	p2	18	-	24	17	13	13
	p3	18	11	-	22	19	5
	p4	17	31	28	-	23	19
	p5	32	30	24	24	-	16
	p6	25	36	32	19	9	-
테스트	p7	12	7	7	14	7	1
	p8	7	5	2	4	1	0
	p9	14	31	19	20	14	18
	p10	14	8	5	13	5	6
	p11	14	29	15	17	6	14

표 2는 인물 p1, p2, p3, p4, p5, p6 이렇게 여섯 명의 인물에게 첫 번째 질의를 주고, 본인을 제외한 나머지 인물들에 대한 호감 정도를 결정한 값이다. 의미적 관계에 대한 가중치를 결정하기 위해 이들 여섯 명의 그룹을 학습 데이터로 활용하고 각 여섯 명이 나머지 다섯 명에 대해 관계 정도를 결정한 값을 테스트 데이터로 활용한다.

가중치는 각 인물이 결정한 다른 인물들에 대한 호감 정도와 제안한 방법에 의한 의미적 관계 정도 사이의 상관관계를 구하고, 이것들에 대한 평균값이 최대가 되도록 하는 값으로 결정한다. 이렇게 하기 위해 목표 인스턴스를 공유하는 것에 대한 가중치  $\beta$ 와 목표 인스턴스의 속성이 같은 것에 대한 가중치  $\gamma$ 를 0에서 1사이에 0.01을 단위로 변화 시켜서 제안한 방법에 의한 의미적 관계 정도를 구하고, 사람이 정한 값 사이의 상관관계를 구한다. 목표 클래스를 공유하는 것에 대한 가중치  $\alpha$ 는 1에서  $\beta$ 와  $\gamma$ 를 뺀으로써 얻을 수 있다.  $\beta$ 와  $\gamma$  값의 변화에 따라 학습 데이터의 여섯 인물과 비교한 상관관계에 대한 평균값을 그래프로 표현하면 그림 10과 같다. 첫 번째 질의와 관련하여 평균 상관관계 값이 가장 높게 나온 지점은  $\alpha=0.02$ ,  $\beta=0.87$ ,  $\gamma=0.11$  인 경우이다.

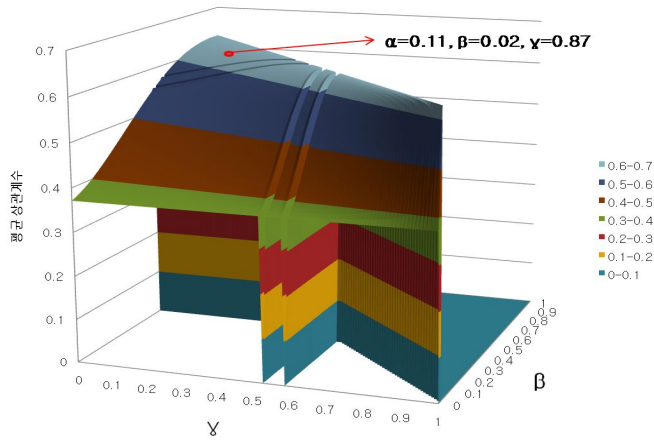


그림 10. 가중치 변화에 따른 평균 상관관계수의 변화와 가중치 결정

이렇게 구한 가중치를 테스트 그룹에서 인물 간의 의미적 관계를 측정하는데 이용하고 그 결과와 테스트 데이터 사이의 상관관계를 구함으로써 성능을 평가한다. 비교를 위해 목표 인스턴스를 공유하는 것에 대한 의미적 관계 정도에 대한 결과를 사용한다. 목표 인스턴스를 공유하는 것에 대한 의미적 관계 정도는  $\beta$ 가 1이고 나머지 가중치가 모두 0인 경우이다. 나머지 세 개의 질의에 대해서도 동일한 방법으로 가중치를 결정하고, 테스트 데이터를 이용해 성능을 평가한다. 각 질의 별 학습 데이터로부터 결정한 가중치 값과 그에 따르는 의미적 관계 정도와 사람이 정한 호감정도 값 사이의 상관관계를 그림 11에서 표현하고 있다. 각 그래프의 x축은 질의 대상 인물에 대한 구분이다. Base는 목표 인스턴스를 공유하는 경우에 대한 상관관계 결과를 의미하고, DSA는 제안한 방법에 의한 결과를 나타낸다. DSA에 대한 가중치는 그래프의 아래에 있는 괄호 안에 표시하고 있다.

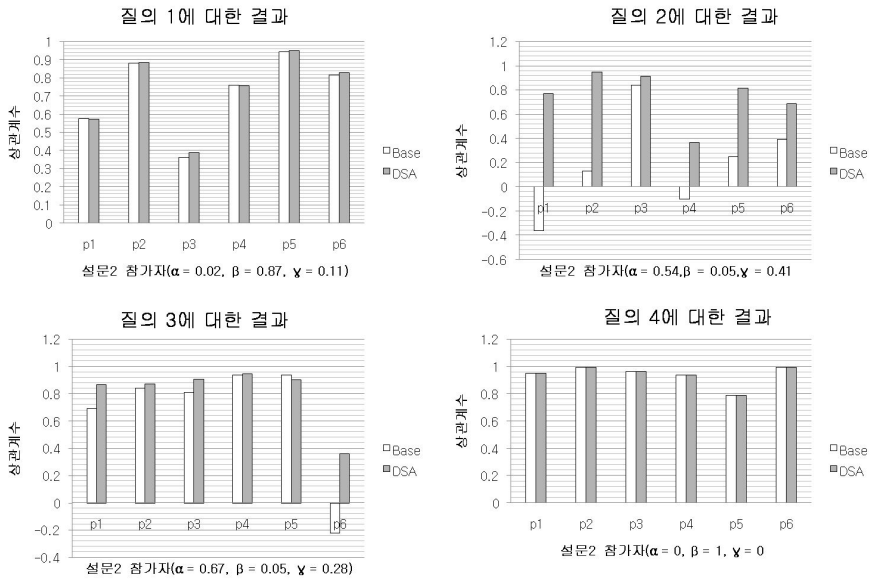


그림 11. 질의 별 Base와 DSA에 대한 상관계수

질의 1에 대해 Base와 DSA는 거의 비슷한 결과를 보이고 있다. p1에 대해서는 Base에 의한 상관계수가 다소 높게 나왔지만, p2, p3, p5, p6에서 모두 DSA에 의한 상관계수가 모두 다소 높게 나왔다. 질의 2에서는 DSA의 결과가 현저하게 높은 상관계수를 보이고 있다. 이때,  $\alpha$ 의 값은 0.54로 질의 1의 0.02에 비해 월등히 높은 비중을 차지하는 것을 알 수 있다. 질의 3의 결과에서도 DSA에 의한 상관계수의 값이 대체로 높게 나타나고 있으며  $\alpha$ 의 값도 0.67로 높은 편이다. 질의 4의 경우 학습 데이터로부터 얻은 가중치의 값이  $\beta$ 가 1인 경우로 DSA가 Base와 동일한 형태로 결정되었다. 따라서, 상관계수의 결과는 같다.

질의 2와 3의 경우 목표 클래스를 공유하는 것에 대한 의미적 관계 정도가 높은 가중치를 가진다. 따라서, 목표 인스턴스만을 고려한 의미적 관계 정도가 낮은 상관계수를 가지는 것으로 나타났다. 질의 1과 4의 경우 목표 클래스를 공유하는 것에 대한 가중치가 낮게 결정되었고 이 경

우, DSA와 Base의 결과는 거의 비슷하거나 DSA가 다소 높은 상관계수를 나타냈다. 질의 4에서는 목표 클래스를 공유하는 것에 대한 가중치가 0인데, 이것은 두 인스턴스들 사이에 목표 인스턴스를 공유하는 개수의 차이가 크게 나타나기 때문으로 풀이 된다. 즉, 의미적 관계 정도는 목표 인스턴스를 공유하는 값에 크게 의존한다. 하지만, 공유하는 목표 인스턴스의 개수가 비슷할 경우, 의미적 관계 정도는 목표 클래스를 공유하는 것에 대해 많이 의존하는 것이다.

다음 절에서는 질의에 대해 어떤 인물과 관계하는 다른 인물들을 의미적 관계 정도에 따라 서열화한 결과를 평가한다.

### 5.3 인스턴스 서열화에 대한 평가

인스턴스에 대한 서열화 결과를 평가하는 것은 의미적 관계 정도에 대한 평가 시험과 유사하다. 먼저, 학습 데이터와 제안한 방법에 의한 서열화 결과 사이의 순위상관계수를 구하고, 각 인물들로부터 얻은 순위상관계수의 평균을 구해서 가장 큰 평균을 가지는 경우에 대해 가중치를 결정한다. 결정된 가중치로부터 테스트 그룹의 인물들에 대한 의미적 관계 정도를 구하고, 서열화 결과를 얻는다. 이러한 결과와 테스트 데이터에서 얻은 서열화 결과 사이의 순위상관계수를 비교함으로써 제안한 방법에 의한 인스턴스에 대한 서열화 결과를 평가한다.

표 2의 결과로부터 질의에 대해 어떤 인물과 관계하는 다른 인물들을 호감도 정도에 따라 서열화한 결과를 얻는다. 호감도 정도가 같은 경우, 질의 대상 인물들의 판단에 따라 서열화하였다. 이렇게 얻은 테스트 데이터를 활용하여 평균 순위상관계수를 최대로 하는 가중치를 결정한다. 그림 12는 질의1에 대해 가중치 변화에 따른 평균 순위상관계수의 변화를 표현하고 있다. 평균 순위상관계수가 최대가 되는 것으로 가중치를  $\alpha = 0.37$ ,  $\beta = 0.63$ ,  $\gamma = 0$ 인 값으로 결정했다.

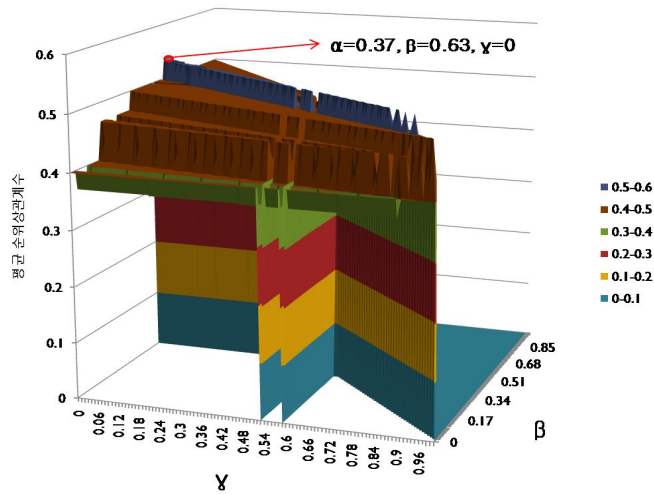


그림 12. 가중치 변화에 따른 평균 순위상관계수의 변화와 가중치 결정

이와 같은 방법으로 질의 2, 3, 4에 대해서도 가중치를 결정한다. 결정한 가중치를 이용해 테스트 데이터의 서열화 결과와 제안한 방법에 의한 서열화 결과 사이의 순위상관계수를 구했다. 이것을 그래프로 표현하면 그림 13과 같다. 각 그래프의 x축은 질의 대상에 대한 구분이고, y축은 순위상관계수를 나타낸다. DSA에 대한 가중치는 그래프의 아래에 있는 괄호 안에 표시하고 있다.



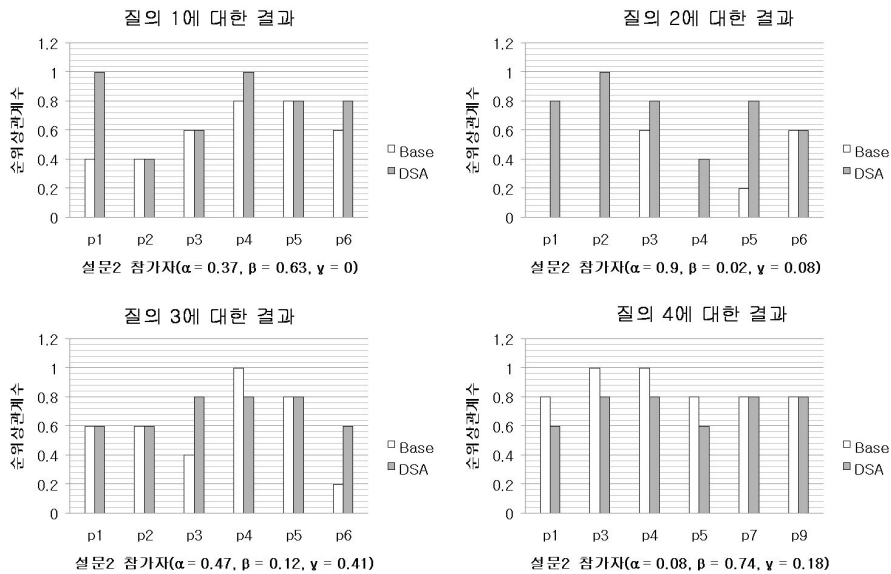


그림 12. 질의 별 Base와 DSA에 대한 순위상관계수

질의 1, 2, 3에 대해 대체로 DSA에 의한 순위상관계수가 높은 것으로 나타났다. 특히, 질의 2의 경우 의미적 관계 정도 측정에 대한 평가에서와 같이 DSA에 의한 서열화 결과가 비교적 높은 순위상관계수를 보이고 있다. 질의 4의 경우 Base의 결과 높게 나왔는데, 이것은 학습데이터 중에 한 인물에 대한 순위상관관계 결과와 다른 인물에 대한 결과와 상당한 차이를 보임으로써 결정된 가중치 값이 테스트 데이터의 분포를 반영하지 못한 결과로 분석된다. 테스트 데이터의 경우  $\beta$  값이 1이 될 때, 가장 높은 평균 순위상관계수를 나타냈다. 질의 4에서 DSA의 결과가 Base의 결과보다 낮게 나타났지만, 전체적으로 볼 때 Base에 의한 결과보다 안정적이고 높은 순위상관계수를 보이고 있다.

다음으로 상위에 서열화된 결과로부터 Base와 DSA에 의한 결과를 비교한다. 상위에 서열화된 결과가 다르더라도 하위에 서열화된 결과가 비슷한 경우 순위상관계수가 높은 값을 가질 수도 있고, 상위에 서열화된

결과가 일치하더라도 하위에 서열화된 결과가 다를 경우, 낮은 값을 가질 수도 있다. 따라서, 상위 서열에 대해 각 관계 정도 측정 방법이 사람이 서열화한 결과와 얼마나 일치하는지를 알아 볼 필요가 있다.

각 질의에 대해 사람이 결정한 서열의 4위까지 정보에 대해 해당 순위 안에 있는 요소를 Base와 DSA에 의한 서열의 결과가 포함하는 경우에 대한 통계 정보를 분석하면 그림 17와 같다. 그래프의 x축은 인물에 대한 구분이고 괄호 안의 b는 Base를 d는 DSA를 의미한다. 두 가지 방법에 의한 각 인물과 관련하는 다른 인물을 서열화한 결과에서 사람이 서열화한 상위 4위의 결과를 포함하는 개수를 그래프 아래에 표로 표시하고 있다. 5위에 대한 결과는 서열화 대상이 5명이기 때문에 모든 경우 통계 값이 5이기 때문에 생략했다.

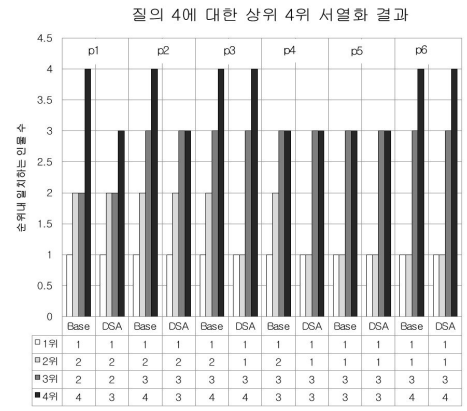
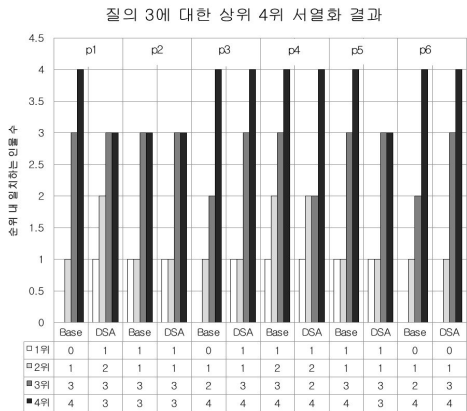
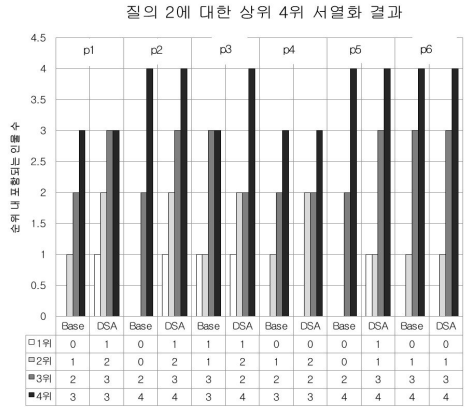
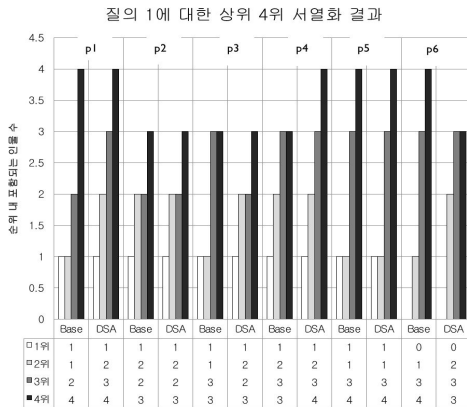


그림 17. Base와 DSA에 대한 상위 4위의 서열화 결과 비교

질의 1의 경우 1위에서의 사람이 서열화한 결과에 해당하는 인물이 Base와 DSA에 의한 결과 내에 포함되는 개수는 동일하게 나타났다. 하지만, 2위에서의 결과를 보면 p1, p3, p6에서 DSA가 Base보다 우수한 것을 알 수 있다. 질의 2의 경우, 1위에 대해 사람이 서열화한 결과를 포함하는 인물의 수가 p1, p2, p5에서 DSA가 더 많은 것을 알 수 있다. 질의 3에서도 1위에 대해 p1, p3에서 DSA가 사람이 서열화한 인물을 더 많이 포함한다. 질의 4의 경우, 순위상관계수의 결과는 Base가 비교적 높게 나타났지만, 1위에 대한 결과는 모두 동일한 것을 알 수 있다. 2위 이후에 Base의 결과가 비교적 사람이 서열화한 인물을 많이 포함하는 것으로

로 나타나고 있다. 이것은 앞서 분석한 바와 같이 학습 데이터와 테스트 데이터 간의 분포 차이에 따라, 목표 클래스를 공유하는 경우가 테스트 데이터에 맞지 않게 다소 높은 가중치가 적용된 결과로 풀이된다. 하지만, 질의 4에서 Base와 비교해서 사람이 서열화한 인물 수를 적게 포함 하기는 하지만, 다른 질의 결과와 비교해 볼 때 대체로 사람이 서열화한 인물 수를 많이 포함하고 있다. 그리고 Base의 결과와 비교할 때, 1위 서열화한 인물이 사람의 서열화 결과와 일치하는 경우가 더 많다.

본 시험에서는 인스턴스 간의 관계 정도에 대한 측정 방법을 책과 영화를 본 이력을 표현한 이벤트 온톨로지에 적용해 봤다. 시험 결과 질의에 따라 의미적 관계에 대한 가중치에는 다소 차이가 있었지만, 목표 클래스를 공유하는 정보를 반영할 때 최고 순위상관계수를 얻을 수 있었다. 마지막으로 상위에 서열화 결과에 대한 분석 결과, 목표 인스턴스를 공유하는 것은 관련성이 가장 높은 인스턴스를 찾는 데 큰 비중을 차지하지만, 더 정확한 서열화 결화를 얻기 위해서는 목표 클래스를 공유하는 정보가 유용하게 쓰일 수 있음을 보였다.

## 6. 결론 및 향후 과제

본 논문에서는 특정 컨텍스트를 통해 두 인스턴스가 가지는 의미적인 관계 정도를 측정하는 방법을 제안했다. 컨텍스트는 온톨로지에 정의된 클래스와 프라퍼티 간의 관계로 표현한다. 이러한 컨텍스트를 만족하는 인스턴스 간의 패스 정보로부터 두 인스턴스 간의 의미적 관계 정도를 측정하였다. 실험 결과 목표 클래스만 공유하는 경우도 두 인스턴스 간에 의미적 관계를 측정하는데 유의미한 정보를 제공한다는 것을 알 수 있었다. 또한, 어떤 인스턴스와 관계하는 다른 인스턴스들을 관계 정도에 따라 서열화했을 때, 사람의 직관에 의한 서열화 결과와 제안한 방법의 결과가 비교적 높은 순위상관계수를 보였다.

따라서, 본 논문의 공헌은 다음의 두 가지로 요약된다. 우선, 특정 컨텍스트를 통한 두 인스턴스 간의 관계를 일반화하여 정의하였다. 다음으로 두 인스턴스 간에 관계 정도를 측정하는 방법을 제안하였고, 실험 결과 기존에 고려되지 않았던 스키마를 통한 인스턴스 간의 관계가 유의미하다는 것을 보였다.

실제 두 인스턴스 사이에 의미적인 관계 정도를 얻는 것은 어려운 문제이다. 사람의 경우, 온톨로지에 데이터로 존재하는 정보로부터 의미적인 정보를 활용할 수도 있고, 온톨로지에 표현되지 않은 직관적인 지식을 활용해서 객체들에 대한 관계 정도를 판단한다. 또한, 엄밀한 의미에서 컨텍스트는 단순히 하나의 ConceptPath만을 의미하지 않을 수도 있다. 예를 들어, “역사와 관련된 주제로 논할 만한 사람은?” 이라는 질의에서 역사 책을 읽었던 사건 정보가 주요한 근거가 될 수도 있지만, 역사에 관련된 영화를 본 경우도 근거로 이용될 수 있다.

향후 과제로 여러 개의 컨텍스트를 활용한 의미적 관계를 찾고 이것에

대한 관계 정도를 측정하는 방법을 고려해 볼 수 있다. 그리고 제안된 방법은 사용자에게 기반한 추천 시스템에도 활용할 수 있다. 자신과 관계 정도가 큰 인물이 경험한 자료 중 본인이 경험하지 못한 것을 추천받을 수 있을 것이다. 또한 잘 모르는 분야에 대해 관심 정도가 큰 인물로부터 필요한 정보를 얻을 수도 있다.

## 참 고 문 헌

- [1] Amit, S., Boanerges, A., Budak, I., Chris H., Cartic, R., Clemens, B., Yashodhan, W., David, A., Sena F., Kemafor, A., Krys, K., "Semantic Association Identification and Knowledge Discovery for National Security Applications," In *Journal of dataBase Management on DataBase Technology for Enhancing national Security*, 16(1), 2005.
- [2] Alexaki S., Christophides V. , Karvounarakis G., Plexousakis D., and Tolle K., "The ICS-FORTH RDFSuite: Managing Voluminous RDF Description Bases," In *SemWeb*, 2001.
- [3] Alkhateeb, F., Baget JF., Euzenat, J. "Complex path queries for RDF," Poster paper in *International Semantic Web Confoerence 2005*, Galway, Ireland, 2005.
- [4] Barton, S, "Designing Indexing Structure for Discovering Relationships in RDF Graphs," In *Proc. of Databses, Texts, Specifications, and Objects*, Galway, Ireland, pp. 7-17, 2004.
- [5] Bernardo, M., Emanuele P., Octavian P., and Manuela S. "Ontology Population from Textual Mentions: Task Definition and Benchmark," In *Proc. of The 2nd Workshop on Ontology Learning and Population at COLING/ACL2006*, Sydney, Australia, pp. 26-32, 2006.
- [6] Boanerges, A., Christian, H., Budak, A., Cartic, R., and Amit, P., "Ranking Complex Relationships on the Semantic Web," In *IEEE Internet computing*, 9(4), pp. 37-44, 2005.
- [7] Celjuska, D., and Vargas-Vera M. "Ontosophie, A Semi-Automatic System for Ontology Population from Text," In

*Proc. of The International Conference on Natural Language Processing ICON 04*, Hyderabad, India, 2004.

- [8] Gruver, T., <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- [9] Haiying, W., Francisco, A., "Gene Expression Correlation and Gene Ontology-Based Similarity: An Assessment of Quantitative Relationships," In *Proc. of Computational Intelligence in Bioinformatics and Bioengineering*, pp. 25- 31, 2004
- [10] Harith, A., Kieron, O., Nigel, S., "ONTOCOPI: Methods and Tools for Identifying Communities of Practice," In *Proc. of the IFIP 17th World Computer Congress - TC12 Stream on Intelligent Information Processing*, Montreal, Canada, 2002
- [11] Kemafor, A., Amit, S., "p-Queries: Enabling Querying for Semantic Associations on the Semantic Web," In *Proc. www2003.*, Budapest, Hungary, 2003.
- [12] Kemafor, A., Angela, M., Amit, S.,, "SPARQLer: Extended Sparql for Semantic Association Discovery," In *LNCS The Semantic Web: Research and Application*, pp. 145-159, 2007.
- [13] Krysz, J., Maciej, J., "SPARQ2R: Towards Support for Subgraph Extraction Queries in RDF Databases," In *Proc. www2007*, Banff, Alberta, Canada, 2007.
- [14] Mary, E. C., and Raymond J. M. "Relational Learning of Pattern-Match Rules for Information Extraction," In *Proc. of the Sixteenth National Conference on Artificial Intelligence*, Menlo Park, California, pp. 328-334, 1999.
- [15] Miller, S., Fox, H., Ramshaw L., and Weischedel R. "A novel



- use of statistical parsing to extract information from text," In *Proc. of The Association for Computational Linguistics North American*, Seattle, USA, pp. 226-233, 2000.
- [16] Muslea, I., Minton, S., Knoblock C. A., "STALKER: Learning extraction rules for semistructured Web-Based information sources," In *Proc. of the Association for the Advancement of Artificial Intelligence*, Madison, Wisconsin, pp. 74-81, 1998.
- [17] Philip R. "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," In *Proc. International Joint Conferences on Artificial Intelligence*, pp. 448-453, 1995.
- [18] Yolanda, B., Jose, J., Pazos, A., Martin, L., Alberto, Gil., "AVATAR: An Improved Solution for Personalized TV Based on Semantic Inference," In *Consumer Electronics, IEEE Transactions*, 52(1), pp. 223 - 231, 2006.
- [19] Xuan T., HaiHua L., Xiaoyong D., "Measuring Semantic Association in Domain Ontology," In *Proc. of International Conference on Semantics, Knowledge and Grid*, pp. 515-518, 2007
- [20] Xuan T., Xiaoyong D., Haihua L., "Computing Degree of Association Based on Different Semantic Relationships," In *18th International Workshop on DataBase and Expert Systems Applications*, Regensburg, Germany, pp. 372-376, 2007
- [21] Spearman C. , "The proof and measurement of association between two things," In *The American Journal of Psychology*, 100(3/4), pp. 72-101, 1904
- [22] Tim B., James H. and Ora L., "The Semantic Web, Feature

Article," In *Scientific American*, 5(1), 2001.

- [23] Yong H., Se P., Seong P., Young L., and Kweon K., "Time Variant Event Ontology for Temporal People Information," In *Fuzzy Logic and Interlligent Systems*, 7(4), pp. 301-306, 2007.
- [24] Zhdanova, A., Predoiu L., Pellegrini T., Fensel D., "A Social Networking Model of a Web Community," In *Proceedings of the 10th International Symposium on Social Communication*, 2007.
- [25] 한용진, 박세영, 이영화, 김권양, "이벤트 온톨로지 기반의 의미 정보 검색," *정보과학회 논문지:컴퓨팅의 실제 및 레터*, 14(1) pp. 96-100, 2008.

# **A Measurement of Semantic Association between Instances Based on an Ontology\***

**Yong-Jin Han**

*Department of Computer Engineering  
Graduate School, Kyungpook National University  
Daegu, Korea  
(Supervised by Professor, Se Young Park)*

(Abstract)

This paper proposes a measurement for a relationship between two instances of an ontology through a single context. The context is the meaning of the relationship. Thus, the single context is just one restricted context. Any other two instances could have the different degree of relationship through a single context. The ability to measuring for the relationships is fundamental to tasks in identifying community or constructing social network. Most of researches for such tasks are based on connected path between entities. However, there could be no connected, but meaningful relationships between instances in an ontology. The proposed method is not only based on connected paths, but also relationships through the schema. We focus to measure for a relationship between two instances through a single context. The experiment using an event-ontology showed that the relationships through the schema are meaningful to measure for relationships between instances.

---

\* A thesis submitted to the Council of Graduate School of Kyungpook National University in partial fulfillment of the requirements for the degree of Master of Engineering in June 2008.

## 부 록

### 1. 실험용 영화 리스트

연번	영화 제목	감독
1	007 네버세이 네버어게인	어빈커쉬너
2	007 어나더데이	리타마호리주연
3	28일후	DannyBoyle
4	34번가의 기적	레이메이필드
5	48시간	월터힐
6	48시간 2	월터힐
7	51번째주	RonnyYu
8	7인의 신부	스탠리도넨
9	가을날의 동화	장완정
10	강아지똥	권오성
11	게티스버그	로날드F_맥스웰
12	갯어웨이	로저도날드슨
13	고스트 앤 다크니스	스티븐훅킨즈
14	고양이를 부탁해	정재은
15	골든 차일드	마이클리치
16	국화꽃 향기	이정욱
17	그림속 나의마을	YoichiHigashi
18	긴급명령	필립노이스
19	닥터모로 D.N.A	존프랑켄하이머
20	달마야 놀자	박철관
21	데몰리션 맨	MarcoBrambilla
22	더티댄싱	에밀아돌리노
23	동갑내기 과외하기	김경형
24	동승	주경중
25	드라이빙 미스 데이지	브루스베레스포드
26	디 아더스	AlejandroAmenabar
27	디어헌터	마이클치미노
28	라스트맨 스탠딩	월터힐
29	라이드 위드 데블	앙리
30	라이언 일병 구하기	StevenSpielberg
31	라이온 킹	민국롭
32	라파누이	케빈레이놀즈
33	러브 & 워	리차드아텐보로
34	러브 어페어	글렌고든캐런
35	레인 맨	배리레빈슨
36	로닌	존프랑켄하이머
37	로드무비	김인식
38	로맨싱 스톤	로버트저메키스
39	로맨싱커플	캐빈W스미스
40	(로이 샤이더의)째즈클럽	BobFosse

## 2. 실험용 책 리스트

	<b>서명</b> 지식 ⑥ : 가슴으로 읽는 우리시대의 지식(智識) . 1-2
	<b>저자명</b> EBS, 지식채널E제작팀
	<b>청구기호</b> 020.13 이49ㅈ
	<b>이 책은</b> <b>갈등</b> 은 '말'이 되어 '성찰'로 이어져야 한다! 우리 시대의 지식 『지식 ⑥』, 2005년 9월EBS에서는 과학사회의간교육문학 등 여러 분야의 '지식'을 키워드로 구성한 '지식채널 ⑥'를 방송하기 시작했다. 일주일에 3~4편이 하루 3~4번씩 5분간 방송되었을 뿐이지만 '지식채널 ⑥'의 영상과 메시지는 우리의 눈과 귀 그리고 마음과 머리를 사로잡았다.
	<b>서명</b> (조선의 집) 동굴에 들다 : 창덕궁과 창경궁으로 떠나는 우리 역사 기행
	<b>저자명</b> 한영우
	<b>청구기호</b> 611 한64ㄷ
	<b>이 책은</b> 이 책은 창덕궁의 정문인 돈화문에서 출발해 발길이 닿는 순서를 따라 각 전당을 소개하며 그곳에서 일어난 주요 역사적 사건들을 함께 살펴본다. 창덕궁과 창경궁 그리고 최근 일반에 공개된 후원을 누비며 일반인이 쉽게 알 수 없는 공간까지 카메라로 담아내었다. 또한 사라진 건물은 '동굴도'를 통해 되살았으며 '동굴도' 전경을 잇달아 옛 지도를 들고 직접 궁궐을 찾을 수 있게 했다.
	<b>서명</b> 그림으로 본 음식의 문화사
	<b>저자명</b> 밴디너케네스
	<b>청구기호</b> 654.9594 변229ㄴK-
	<b>이 책은</b> 이 책은 15세기 르네상스 시대부터 포스트모더니즘 시대까지 등장하는 회화 작품들을 대상으로 음식과 식생활의 역사를 검토한다. 저자는 그림 속 작은 음식 하나에도 그 시대의 모습이 담겨 있다고 말하면서 유명 작품들은 물론 덜 알려진 많은 화가들의 작품들을 새로운 시각에서 분석하고 있다. 인문과 예술의 세계를 자유롭게 넘나들며 음식 문화와 관념의 변화를 미시사적 관점에서 짚어낸다.
	<b>서명</b> 인간 없는 세상
	<b>저자명</b> 와이즈먼앨런
	<b>청구기호</b> 472.5 와이69ㄹKㄹ
	<b>이 책은</b> 인간이 없는 세상이 어떻게 변화하는 지 보여주는 『인간 없는 세상』. 이 책은 어느 날 갑자기 인류가 사라지면 어떤 일이 생기는지에 대한 질문과 그에 대한 답을 찾아가는 여정을 담은 것으로 인류가 지구상에 남겨져 할 유산에 대한 고찰과 인류가 사라진 후 세상을 상상하게 그려낸다.
	<b>서명</b> 재왕의 책
	<b>저자명</b> 윤희진
	<b>청구기호</b> 911.05 윤98ㅈ
	<b>이 책은</b> <재왕의 책>은 '재왕'과 '책'이라는 특별한 두 존재의 만남을 소재로 우리 역사의 역동적인 현상들을 살펴보는 책이다. 재왕과 책이 만나 빚어낸 상황을 그리고 현상 위에 숨은 다양한 변화의 동력들을 예리하게 파헤치고 있다. 또한 이 고전들이 태어날 당시의 시대적 배경오랜 시간을 넘어 현재까지 전해지는 메시지 등을 함께 들려준다.
	<b>서명</b> 한국 2030
	<b>저자명</b> 전상인
	<b>청구기호</b> 331.54 전51ㄹ
	<b>이 책은</b> 『한국 2030』은 학계와 경제관계법조계산업계언론계문화예술계 등 각 분야의 지성인들이 대한민국의 현재를 진단하고 미래를 구상하기 위해 결성한 '한국미래학회'가 지난 2005년 6월 11일 서울 프레스센터에서 개최한 확대발표회의 세미나 결과물로 '2030년 한국의 미래를 예측한 보고서'이다.
	<b>서명</b> 세계도시기행
	<b>저자명</b> 김현
	<b>청구기호</b> 980.24 김94ㅈ
	<b>이 책은</b> 전 세계 150여 개국 250여 개의 도시를 유럽과 아시아오세아니아아프리카아메리카 등 권역별로 나눠 그 곳에 맞는 부부만의 맞춤 여행법과 그곳의 역사와 풍경을 들려준다. 또한 부부가 함께 배낭여행을 하면서 지켜야 할 부부배낭여행 10계정도 소개한다.