

컨텍스트에 기반한 두 인스턴스 사이의 의미 관계 정도 측정 A Measurement for the Degree of Semantic Relationship Between Two Instances Based on Context

한용진* · 박세영* · 박성배* · 김권양**

Yong-Jin Han*, Se Young Park*, Seong-Bae Park* and Kweon-Yang Kim**

* 경북대학교 컴퓨터공학과

** 경일대학교 컴퓨터공학과

요 약

실세계의 객체들은 서로 직접적인 관계를 맺고 있고, 이러한 직접적인 관계를 통해 연결되는 새로운 간접적인 관계를 가진다. 온톨로지는 객체들 사이의 관계에 대한 의미를 명시적으로 표현한다. 따라서, 온톨로지를 이용함으로써 객체 간의 새로운 관계를 발견할 수 있다. 새롭게 발견된 객체들 간의 관계는 커뮤니티를 찾거나 소셜네트워크를 구축하는데 활용된다. 이러한 응용에서 두 객체 간에 관계 정도를 측정하는 것은 중요한 문제이다. 본 논문은 온톨로지에 기반하여 객체들 간의 관계 정도를 측정하는 방법을 제안한다. 기존 연구에서는 주로 객체들 간의 연결되는 패스를 주요하게 다루어 왔지만 두 객체 사이에 연결된 패스는 없더라도, 온톨로지의 스키마를 통해 의미를 가지는 관계가 있다. 본 논문에서 제안한 방법은 객체들 간에 패스로 연결되는 관계는 물론 객체가 속하는 스키마를 통해 관계하는 정보들도 활용해서 두 객체 간의 관계 정도를 측정한다. 실험 결과 두 객체 사이에 스키마를 통한 관계에서 많은 의미 관계가 있음을 보였다.

키워드 : 온톨로지, 의미관계, 컨텍스트

Abstract

Entities in reality have direct relationships between each other. They also have new and indirect relationships through such direct relationships. An ontology gives explicit meaning of such relationships. Thus, we can discover new relationships between entities based on an ontology. Such new relationships are applied in identifying new communities or constructing social networks. Measuring for the degree of relationship is an important problem in such domains. This paper proposes a measurement for the degree of relationship between entities based on an ontology. Most of researches are based on connected paths between entities. However, there are meaningful relationships between two entities through the schema in an ontology even though there are no connected paths between the entities. The proposed method measures for the degree of relationships between two entities not only based on connected paths, but also relationships through the schema. The experiment result shows that the relationships through the schema are meaningful to measure the degree of relationship between entities.

Key Words : Ontology, Semantic relationship, Context

1. 서 론

온톨로지는 기계에 의해 처리될 수 있는 개념화에 대한 명시적인 명세를 제공한다[1]. 온톨로지에서 개념화는 실세계에 존재하는 객체와 객체들 간의 관계를 표현함으로써 실현된다[2]. 온톨로지의 대표적인 표현 언어인 OWL(Web Ontology Language)의 경우, 유일한 객체들을 인스턴스로 표현하고, 이들 간의 관계를 오브젝트 프라퍼티로 표현한다[3]. 인스턴스들은 오브젝트 프라퍼티를 통해 직접적인 관계를 맺기도 하지만, 이들 인스턴스 사이에 인스턴스와 오브젝트 프라퍼티로 구성된 패스에 의해 새로운 관계를 맺기도 한다. 전자의 경우는 인스턴스 사이의 관계를 쉽게 알 수 있지만, 후자의 경우는 이들 사이에 연결되는 인스턴스와

오브젝트 프라퍼티로 구성된 패스를 통하여 새로운 관계를 알 수 있다. 이러한 관계들이 가지는 의미는 온톨로지 스키마에 정의된 클래스와 오브젝트 프라퍼티에 의해 명시적으로 표현된다. 본 논문에서는 이러한 관계를 단일 컨텍스트라고 정의한다. 주어진 단일 컨텍스트에서 인스턴스 간의 관계 정도 분석은 새로운 커뮤니티를 찾거나 소셜네트워크를 구성하는데 응용될 수 있다[4, 5].

여러 장르의 책에 대한 블로그 사용자들의 관심 정보를 온톨로지로 구축하게 되면, 특정 장르에 대한 사용자들 간의 공통적인 관심 정도를 측정할 수 있다. 책에 대한 장르 중 역사에 관해 두 사용자의 공통 관심 정도를 측정할 때, 역사 장르에 속하는 같은 책을 읽었거나 혹은 같은 저자의 책을 읽었다면 두 사용자 사이에 공통적인 많은 요소를 가진다. 또한, 두 사용자가 책 제목은 다르지만 같은 장르에 속하는 책을 많이 읽었다면, 특정 장르에 대해서 공통된 흥미를 가질 가능성이 높다. 여기서 역사와 같이 특정 장르에 속하는 책에 대한 경험이 관계의 의미가 되는 컨텍스트에

접수일자 : 2008년 1월 15일

완료일자 : 2008년 9월 30일

** 교신 저자

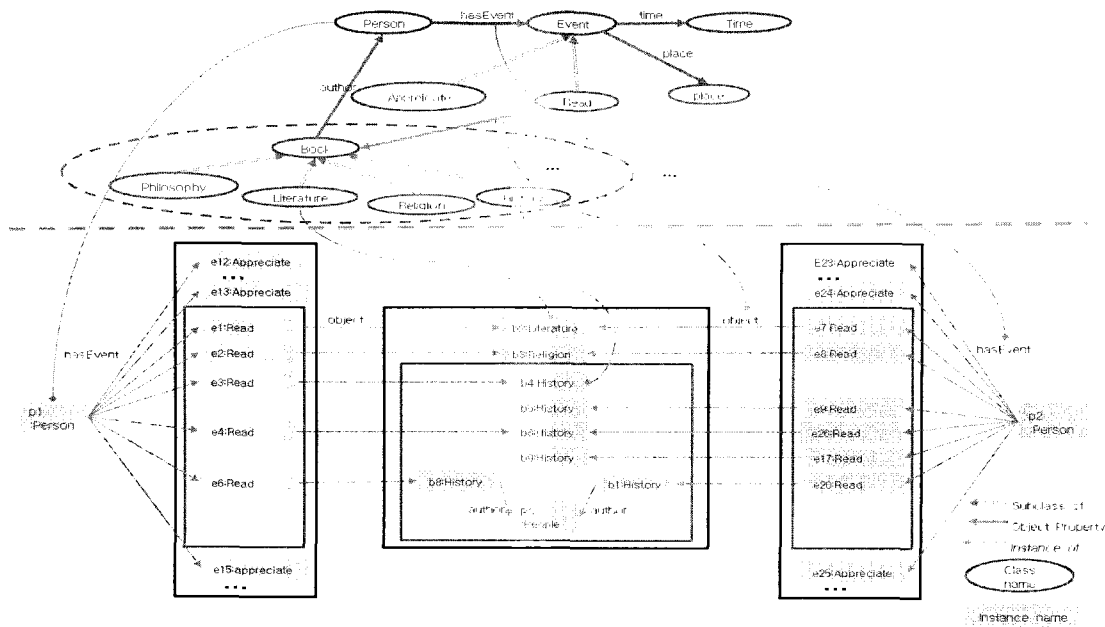


그림 1. 이벤트 온톨로지의 스키마와 인스턴스 예
Fig. 1. The schema of the event-ontology and its example instances

강한다. 주어진 컨텍스트에 대해 여러 인스턴스 간의 관계 정도를 측정하고 이들 관계를 연결하면 그 자체로 가치가 있는 하나의 네트워크를 구성하게 된다. 관계 정도에 한 한계치를 설정함으로써 유의미한 커뮤니티를 발견하거나 소셜네트워크로 활용할 수 있다[5].

객체 간의 관계 분석에 대한 기존 연구에서는[4, 5, 6] 주 두 객체 사이에 연결된 패스 정보를 주요하게 다루었다. 앞에서 든 예에서 책 제목이 같은 책을 읽은 경우 그 책 매개로 두 사용자를 연결하는 패스에 의해 이들 간의 관계를 표현할 수 있다. 하지만, 두 사용자에 해당하는 인스턴스 사이에는 명시적인 연결은 없지만, 스키마를 통해 관계 가지는 의미관계가 있을 수 있다. 기존 연구들은 이러한 경우에 대해 두 인스턴스 사이에 관계를 가지는 정도를 고차지 않고 있다.

본 논문에서 제안한 방법은 두 인스턴스가 간의 관계 정도를 측정하기 위해 인스턴스 층에서 연결된 패스에 의해 현되는 관계 외에 스키마 층을 통해 연결되는 관계들도 용한다. 제안한 방법의 유효성을 보이기 위해 책과 영화 대한 도메인 온톨로지와 이들 도메인에 대한 사용자의 읽 정보를 표현한 이벤트 온톨로지를 구축하였으며 이들 온톨로지 상에서 특정 장르에 대한 인물들 간의 관계 정도를 측정한다. 인스턴스 간의 연결된 패스만을 이용한 결과와 비교한 결과, 대부분의 경우, 제안한 방법의 결과가 읽의 직관에 의한 결과와 비교적 높은 상관관계가 있음을 얻는다.

2장에서 컨텍스트와 두 인스턴스 사이의 관계에 대해 정하고, 3장에서는 주어진 컨텍스트에 있는 두 인스턴스 사이의 관계 정도 측정 방법을 소개한다. 4장에서 실험 및 결과에 대해 설명하고, 5장에서는 관련 연구에 대해 소개한다. 2로 6장에서 결론을 맺는다

2. 단일 컨텍스트에 대한 관계

프라퍼티 연속(property sequence)[7]은 인스턴스 사이에 간접적으로 연결된 새로운 관계를 표현한다. 이러한 관계의 의미는 인스턴스가 속한 클래스를 통해 명확해지기 때문에 [8] 인스턴스 사이의 관계에 대한 의미를 클래스와 프라퍼티의 연속으로 표현한다. 이것을 단일 컨텍스트라(Single Context) 부르고 다음과 같이 정의한다.

정의 1. 단일 컨텍스트 : C_0, C_1, \dots, C_n 을 온톨로지의 클래스라 하고, r_1, r_2, \dots, r_n 을 같은 온톨로지의 오브젝트 프라퍼티라 할 때 C_{i-1} 과 C_i 는 r_i 에 의해 관계를 가진다. ($1 \leq i \leq n$). 이 때 C_0 에서 C_n 에 이르는 관계를 단일 컨텍스트라고 정의하고, $[C_0, r_1, C_1, r_2, C_2, r_3, \dots, r_n, C_n]$ 으로 표기한다. C_0 를 시작 클래스라 부르고, C_n 을 목표 클래스라 부른다.

본 연구의 목표는 “어떤 단일 컨텍스트에 대해 두 인스턴스가 얼마나 관계가 있는가?” 에 대해 답을 얻는 것이다. 예를 들면, 본 논문의 실험에서 평가 초점은 책이나 영화 같은 특정 장르에 대한 인물 간의 공통적인 관심 정도를 측정하는 것이다. 그림 1은 실험에 사용된 이벤트 온톨로지를 나타낸다.

그림 1에서 제시한 이벤트 온톨로지는 책이나 영화에 대한 블로그 참가자들의 경험을 온톨로지로서 표현한 것으로 Person 클래스는 hasEvent 프로퍼티를 통해 Event 클래스와 관계를 가진다. Event 클래스는 어떤 사건을 묘사하는 기본적인 요소들로 정의된다. 즉, 사건의 대상과 사건이 일어난 장소, 시간이 Event 클래스의 프로퍼티가 된다. Event 클래스는 두 개의 서브 클래스, Appreciate와 Read 클래스를 가진다. Read 클래스는 object 프로퍼티를 통해 Book 클래스와 관계한다.

어떤 블로그 참가자가 “사람들이 역사라는 장르에 대해서 얼마나 공통 흥미를 가지는가?” 에 관심이 있다고 가

정하자. 그러면 단일 컨텍스트는 [Person, hasEvent, Read, object, History]가 된다. 이러한 단일 컨텍스트에 대한 인스턴스 간의 관계를 정의하면 다음과 같다.

정의 2. 단일 컨텍스트에 대한 관계 : i_1 과 i_2 인스턴스 각각 단일 컨텍스트에 속하는 패스들을 가진다고 할 때, i_1 과 i_2 는 단일 컨텍스트에 대한 관계를 가진다고 정의한다. 각 패스의 첫 번째 인스턴스를 시작 인스턴스라 하고, 마지막 인스턴스를 목표 인스턴스라고 부른다.

예를 들면, 그림 1에서 p_1 과 p_2 인스턴스는 단일 컨텍스트 [Person, hasEvent, Read, object, History]를 만족하는 패스를 가지기 때문에 두 인스턴스는 단일 컨텍스트에 대해 관계를 가진다. 이러한 단일 컨텍스트에 대한 두 인스턴스 사이의 관계 정도(Degree of Relationships for a single Context:DRC)는 다음과 같이 정의한다.

정의 3. 단일 컨텍스트에 대한 관계 정도 : 두 인스턴스가 어떤 단일한 컨텍스트에 대해 관계를 가질 때, 이러한 관계에 대한 가중치로 DRC를 정의한다.

온톨로지는 개체들 간의 직접적인 연결 관계에 기반해서 만들어지며 일반적으로 직접적인 관계를 통해 유추할 수 있는 간접적인 관계를 온톨로지 구축 과정에 고려하지는 않는다. 따라서, 의미있는 컨텍스트를 따라 간접적이지만 의미를 가지는 연결 관계를 찾음으로써 이들 개체간의 복잡한 관계를 규명하는 연구 주제가 활발하다. Aleman-Meza 등은[9] 두 인스턴스 사이에 존재하는 의미 관계를 찾기 위해 관계 정도를 측정하는 방법을 제안했다. 이러한 관계 정도의 측정은 관심의 대상이 되는 컨텍스트를 찾는데 활용할 수 있다. 본 논문에서는 Aleman-Meza의 연구에서와 같이 사용자에게 의해 흥미있는 컨텍스트가 주어진다고 가정한다. 따라서, 주어진 단일한 컨텍스트에 대해 두 인스턴스 사이에 관계 정도를 측정하는데 초점을 맞춘다.

3. 단일 컨텍스트에 대한 관계 정도 측정

본 논문에서 제안한 DRC 측정은 두 블로그 참가자 인스턴스가 단일 컨텍스트에 대해 목표 클래스에 얼마나 관심을 가지는지에 기반한다. 예를 들어, 두 개의 시작 인스턴스가 단일 컨텍스트 [Person, hasEvent, Read, object, Novel]에 대해 다른 인스턴스들에 비해 공통적인 패스를 많이 가진다면 이들 두 시작 인스턴스는 Novel이라는 장르에 대한 공통된 관심정도가 크다고 볼 수 있다. 두 인스턴스가 Novel 클래스에 속하는 같은 목표 인스턴스를 공유하는 경우에 둘 사이에 관계 정도는 그렇지 않은 경우보다 크다. 또한, 목표 인스턴스는 비록 다르지만, 목표 인스턴스의 프로퍼티인 저자나 출판사가 같다면 두 인스턴스는 의미있는 관계를 가진다고 할 수 있다. 이러한 관계는 Anyanwu 등이[7] 정의한 의미 관계에 기반하여 다음과 같은 세 가지 유형의 의미적 관계로 나눌 수 있다.

- 1) 단일 컨텍스트만을 공유하는 경우
- 2) 동일한 목표 인스턴스를 공유하는 경우
- 3) 목표 인스턴스의 프로퍼티가 같은 경우

1)의 경우는 두 인스턴스가 각각 동일한 단일 컨텍스트를 만족하는 패스를 가지지만, 둘 사이에 연결된 패스는 없는 경우이다. 예를 들어, 그림 1에서 p_1 과 p_2 에서 시작하는 각 패스 $p_1-e_3-b_4$ 와 $p_2-e_9-b_5$ 는 서로 동일한 단일 컨텍스트를 가지지만 p_1 과 p_2 사이를 연결하지는 않는다. 이 경우, 두 인스턴스는 단지 단일 컨텍스트인 [Person, hasEvent, Read, object, History]를 공유하고 있으며, 이들 두 블로그 참가자가 역사에 관한 책을 읽었다는 점에서 공통의 관계를 가지는 경우이다.

2)의 경우는 두 인스턴스가 단일 컨텍스트에 대해 동일한 목표 인스턴스를 가지는 경우로, 그림 1에서 패스 $p_1-e_4-b_6-e_{26}-p_2$ 가 그 예이다. 여기서 참가자 p_1 과 p_2 는 b_6 라는 동일한 역사책을 읽었다는 점에서 서로 관계를 가진다.

마지막으로 3)의 경우는 두 인스턴스가 가지는 각 패스에 대해 목표 인스턴스가 가지는 프로퍼티가 같은 경우이다. 그림1에서 패스 $p_1-e_5-b_8-p_3-b_1-e_{20}-p_2$ 은 각 목표 인스턴스 b_8 과 b_1 의 프로퍼티인 *author*에 대한 값이 p_3 으로 동일하다. 이 경우 두 참가자는 저자가 같은 책을 읽었다는 점에서 서로 관계를 가진다. 즉, 두 인스턴스에서 시작하는 각 패스의 목표 인스턴스는 다르지만, 그 프로퍼티 값이 같다.

직관적으로 볼 때, 각 경우에 대해 두 개의 패스만 주어진다면, 관계 정도의 대소 관계는 2) > 3) > 1) 가 된다. 의미적 관계에 대한 세 가지 연관 관계를 기준으로DRC를 다음의 세 가지 요소로 나눈다: 1) 단일 컨텍스트를 공유하는 것에 대한 DRC, 2) 목표 인스턴스를 공유하는 것에 대한 DRC, 3) 목표 인스턴스의 프로퍼티가 같은 것에 대한 DRC. 전체 DRC는 이러한 세 가지 요소에 선형적으로 의존한다고 가정한다. 이것을 수식으로 표현하면 다음과 같다.

$$DRC(i_1, i_2) = \alpha v^c + \beta v^j + \gamma v^p, \quad \alpha, \beta, \gamma: \text{가중치}, \alpha + \beta + \gamma = 1$$

v^c , v^j , v^p 는 각각 단일 컨텍스트만을 공유하는 것에 대한 DRC, 목표 인스턴스를 공유하는 것에 대한 DRC, 목표 인스턴스의 속성이 같은 것에 대한 DRC 값을 나타낸다. v^c 는 세 가지 의미적 관계 1), 2), 3) 모두에 적용된 DRC의 요소이다. v^j 는 목표 인스턴스를 공유하는 경우인2)에만 해당한다. 마지막으로 v^p 는 서로 다른 두 목표 인스턴스의 프로퍼티가 같은 경우 3)에만 해당한다.

3.1 단일 컨텍스트를 공유하는 것에 대한 DRC

단일 컨텍스트에 대해 두 인스턴스가 목표 클래스에 얼마나 공통된 관심을 가지는지를 측정하기 위해 먼저 각 인스턴스가 목표 클래스와 관련된 정도를 구하고, 다음으로 두 인스턴스가 목표 클래스에 공통으로 관심을 가지는 정도를 구한다.

그림 1에서 시작 인스턴스 p_1 과 p_2 는 단일 컨텍스트 [Person, hasEvent, Read, object, History]를 만족하는 여러 개의 패스를 가진다. 안쪽 네모에 속하는 인스턴스들은 단일 컨텍스트에 속하는 것들이다. 바깥쪽 네모는 오브젝트 프라퍼티를 통해 이전 인스턴스와 관계하는 모든 인스턴스들을 포함한다. 예를 들어, hasEvent를 통해 p_1 과 관계하는 Event의 인스턴스 8개 중 Read에 속하는 것은 5 개이다. 따라서, p_1 은 Read에 5/8만큼 흥미를 가진다고 볼 수 있다. 즉, 주어진 단일 컨텍스트 $[C_0, r_1, C_1, \dots, r_i, C_i, \dots, r_n, C_n]$ 에 대해 C_i 의 어떤 인스턴스 s 가 프라퍼티 r_{i+1} 을 통해 C_{i+1} 에 흥미를

가지는 정도는 s 와 관계하는 인스턴스들 중에서 C_{i+1} 에 속하는 것의 개수에 비례하고, r_{i+1} 을 통해 관계하는 모든 인스턴스의 수에 반비례한다. 한편, C_i 에 속하는 인스턴스는 여러 개 있을 수 있다. 각 인스턴스마다 r_{i+1} 을 통해 C_{i+1} 에 흥미를 가지는 정도를 구하고, 이들을 합해서 C_i 의 인스턴스들이 프라퍼티 r_{i+1} 을 통해 C_{i+1} 에 흥미를 가지는 정도를 구한다. 이것을 $w_{C_i}^{r_{i+1}}$ 라 하고, 수식으로 표현하면 다음과 같다.

$$w_{C_i}^{r_{i+1}} = \sum_{s \in C_i} \left(\frac{|C_{i+1}|_s}{|C_{i+1} \cup \bar{C}_{i+1}|_s} \right) + \frac{1}{|C_{i+1} \cup \bar{C}_{i+1}|_s},$$

$$|C_{i+1} \cup \bar{C}_{i+1}|_s \neq 0, (0 \leq i \leq n-1)$$

C_i 에 속하는 인스턴스 $s(s \in C_i)$ 와 관계하는 인스턴스들은 단일 컨텍스트의 클래스인 C_{i+1} 에 속하는 것과 r_{i+1} 과 관계하지만, C_{i+1} 이 아닌 다른 클래스에 속하는 경우(C_{i+1})로 나눌 수 있다. $C_{i+1} \cup C_{i+1}$ 은 r_{i+1} 을 통해 C_i 와 관계하는 모든 클래스이다. $|C_{i+1}|_s$ 는 인스턴스 s 에서 C_{i+1} 로 가는 패스의 개수를 의미한다. $w_{C_i}^{r_{i+1}}$ 가 0이 되는 경우를 보정하기 위해 C_i 가 r_{i+1} 을 통해 관계하는 모든 클래스의 인스턴스 개수($|C_i \cup C_i|$)의 역수를 더해준다.

패스의 각 홉(hop)에 있는 인스턴스들 다음에 연결된 클래스에 대해 이와 같은 흥미 정도를 가지며 다음과 같이 V (Interest Vector)라 정의한다.

정의 4. IV_s : 단일 컨텍스트의 각 클래스에 있는 인스턴스들이 다음에 연결된 클래스에 대한 흥미 정도를 나타낸 벡터

$$IV_s = (w_{c_1}^{r_1}, w_{c_2}^{r_2}, \dots, w_{c_1}^{r_1}, \dots, w_{c_n}^{r_n}), 1 \leq i \leq n$$

IV_s 벡터의 각 요소 값이 클수록 목표 클래스로 가는 패스의 개수가 클 가능성이 높다. 즉, 인스턴스 s 가 목표 클래스에 흥미를 가지는 정도는 IV_s 의 각 요소의 값에 비례한다. 이것을 DOI(degree of interest)라 하고 다음 식과 같이 정의한다.

$$DOI(IV_s) = \prod_{i=1}^n w_{c_i}^{r_i}$$

다음으로 두 인스턴스가 목표 클래스에 대해 공통적으로 가지는 관심의 정도를 구하기 위해 두 인스턴스가 단일 컨텍스트의 각 클래스에 대한 공통 흥미 정도를 표현한 CIV(Common Interest Vector)를 정의한다.

정의 5. CIV(s_1, s_2) : 두 인스턴스 s_1 과 s_2 가 단일 컨텍스트에 대해 관계를 가질 때, 단일 컨텍스트의 각 클래스에 공통적으로 가지는 흥미 정도를 표현한 벡터

두 인스턴스 s_1 과 s_2 가 어떤 단일 컨텍스트에 대해 관계를 가지면, 이들에 대한 IV_{s_1} 과 IV_{s_2} 를 각각 구할 수 있다. 이들 두 벡터의 대응하는 각 요소에 대한 조화 평균을 구함으로써 두 인스턴스가 단일 컨텍스트의 각 클래스에 대해 공통적으로 흥미를 가지는 정도를 구한다. 수식으로 표현하면 다음과 같다.

$$IV_{s_1} = (w_1, w_2, \dots, w_i, \dots, w_n)$$

$$IV_{s_2} = (w'_1, w'_2, \dots, w'_i, \dots, w'_n), n: \text{단일 컨텍스트의 길이}$$

$$CIV(IV_{s_1}^{sum}, IV_{s_2}^{sum}) = (w_1^{com}, w_2^{com}, \dots, w_i^{com}, \dots, w_n^{com}), w_i^{com} = \frac{2w_i w'_i}{w_i + w'_i}$$

조화평균은 두 수의 합이 같을 때, 두 수의 차가 적을수록 큰 값을 반환한다. 즉, 어떤 클래스에서 다음 클래스로 가는 패스의 수를 합한 값이 같더라도 각 패스의 수가 비슷할수록 더 큰 조화평균값을 가진다.

마지막으로 두 인스턴스가 목표 클래스에 공통적으로 흥미를 가지는 정도는 CIV에 대한 DOI를 구함으로써 얻을 수 있다. 이 값이 곧 v^c 가 된다. 수식으로 표현하면 다음과 같다.

$$v^c = \frac{1}{f^c} DOI(CIV(IV_{s_1}, IV_{s_2})) = \frac{1}{f^c} \prod_{i=1}^n w_i^{com},$$

$$f^c = \max(DOI(IV_{s_1}, IV_{s_2}))$$

f^c 는 정규화를 위해 나누어 준다.

3.2 동일 목표 인스턴스를 공유하는 것에 대한 DRC

두 인스턴스가 단일 컨텍스트에 대해 같은 목표 인스턴스를 공유하는 경우, 공유된 인스턴스의 개수에 비례해서 흥미를 공유하는 정도가 커진다. 즉, 블로그의 두 참가자가 같은 책을 많이 읽었을수록 이야기할 수 있는 공통 주제가 많다.

$P_{\langle x, y \rangle}$ 를 두 인스턴스 x, y 에 대해 목표 인스턴스를 공유하는 패스의 집합이라고 할 때, 목표 인스턴스를 공유하는 것에 대한 DRC는 다음 식과 같이 정의한다.

$$v^j = \frac{|P_{\langle x, y \rangle}^j|}{f^j}, f^j = \max(|P_{\langle x, y \rangle}^j|)$$

즉, v^j 는 공유하는 목표 인스턴스의 수에 비례한다. 이 식에서 f^j 는 정규화를 위해 나누어 준다.

3.3 목표 인스턴스가 유사한 것에 대한 DRC

시작 인스턴스 x, y 가 단일 컨텍스트에 대해 서로 다른 두 목표 인스턴스의 프로퍼티가 같은 경우, 두 패스에 대한 관계 정도는 일치하는 프로퍼티 값의 개수에 비례한다. 예를 들면, 두 사람이 같은 책은 아니지만, 저자가 같고, 같은 출판사의 책을 읽었다면, 두 사람은 공통된 관심을 가질 가능성이 높다.

다음은 목표 인스턴스가 유사한 것에 대한 DRC를 구하는 방법이다.

$$v^p = \frac{1}{f^p} \sum_{p \in P_{\langle x, y \rangle}^p} \frac{m(p)}{k}, f^p = \max\left(\sum_{p \in P_{\langle x, y \rangle}^p} \frac{m(p)}{k}\right)$$

위 식에서 $P_{\langle x, y \rangle}^{sp}$ 는 시작 인스턴스 x, y 의 서로 다른 두 목표 인스턴스의 프로퍼티 값이 같은 경우에 대한 두 패스들의 집합을 나타낸다. 함수 $m(p)$ 는 어떤 두 패스 p 에 대해서 목표 인스턴스들이 같은 값을 가지는 프로퍼티 개수를 반환한다. k 는 목표 인스턴스가 가지는 프로퍼티 개수로 컨텍스트에 따라 목표 인스턴스가 가질 수 있는 프로퍼티도 다르다. 목표 인스턴스들의 프로퍼티가 많이 일치할수록 또한 일치하는 목표 인스턴스들이 많을수록 관계 정도는 커진다. f^p 는 정규화를 위해 나누어 준다.

4. 실험 및 평가

본 논문에서 제안한 단일 컨텍스트에 대한 관계 정도 측정 방법의 유효성을 보이기 위해 먼저 책과 영화에 대한 도

메인 온톨로지와 이들 도메인에 대한 사용자의 경험 정보를 표현한 이벤트 온톨로지를 구축하였다. 온톨로지 구축을 위해 실험 참가자 11명을 대상으로 흥미있게 본 책과 영화에 대한 이벤트 인스턴스 정보를 수집하였고, 수집된 인스턴스를 이용해 이벤트 온톨로지를 구축하였다. 실험 참가자를 대상으로 수집한 이벤트 온톨로지의 인스턴스는 221 권의 책과 913 편의 영화로 구성되며, 참가자 별로 책은 40권에서 49권까지 영화는 27편에서 281편에 대한 관심 관계를 가진다.

실험 참가자에게 질의가 주어지면 각자 해당 질의에 대해 다른 참가자들과 흥미를 공유하는 정도를 판단하게 한다. 흥미에 대한 공유 정도는 상대가 읽었던 각 책에 대한 흥미 정도를 0에서 3 사이의 점수를 매겨서 정량적으로 구한다. 다음은 실험에서 사용한 네 가지 질의이다.

- 질의 1 : 철학에 대해 논의할 때 상대에게 호감이 가는 정도는?
- 질의 2 : 경제, 경영에 대해 논의할 때 상대에게 호감이 가는 정도는?
- 질의 3 : 액션에 대해 이야기할 때 상대에게 호감이 가는 정도는?
- 질의 4 : 애니메이션에 대해 이야기할 때 상대에게 호감이 가는 정도는?

표 1은 질의 1에 대해서 참가자로부터 직접 입력받은 참가자 간의 공통 흥미 정도를 보여준다.

표 1. 참가자 간의 공통 흥미 정도

Table 1. The degree of common interests between candidates

		P1	P2	P3	P4	P5	P6
학 습	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	-
테 스 트	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

DRC 추정 방법의 가중치를 결정하기 위해 6 명의 참가자(P1, P2, P3, P4, P5, P6)에 대해 수집한 공통 흥미 정도를 학습자료로 사용하고, 나머지 5 명의 참가자(P7, P8, P9, P10, P11)에 대한 자료를 성능 평가를 위한 테스트 데이터로 사용한다. 두 인물 간의 관계 정도를 측정하기 위해서는 먼저, 질의에 대응하는 단일 컨텍스트를 설정하여야 한다. 표 2는 네 가지 질의에 대응하는 단일 컨텍스트를 보여준다.

표 2. 질의에 대응하는 단일 컨텍스트

Table 2. Single context w.r.t. each queries

질의	단일 컨텍스트
1	[Person, hasEvent, Read, object, Philosophy]
2	[Person, hasEvent, Read, object, Economy]
3	[Person, hasEvent, Appreciate, object, Action]
4	[Person, hasEvent, Appreciate, object, Animation]

다음으로 Protege API를 이용해 어떤 두 인스턴스에 대해 단일 컨텍스트를 만족하는 패스를 각각 구하고, v^c , v^i , v^p 를 결정한다. DRC 추정에서 의미 관계에 따른 가중치를 결정하기 위해 먼저 임의값을 대입했을 때 관계 정도값과 표 1의 학습 데이터의 각 열 데이터 사이의 상관관계를 구한다. 여섯 개의 상관관계에 대한 평균값을 최대로 하는 것을 가중치로 결정한다. 결정한 가중치를 이용해 테스트 데이터 상에서 인물 간의 관계 정도를 측정하고, 표 1의 테스트 데이터와의 상관관계를 비교함으로써 성능을 평가한다. 비교 분석을 위해 인스턴스 층에서 목표 인스턴스를 공유하는 요소만을 고려한 방법(Base)과 비교하였다. 그림 2는 4개 질의에 대해 Base와 본 논문에서 제안한 방법(DRC)을 비교한 결과를 보여준다.

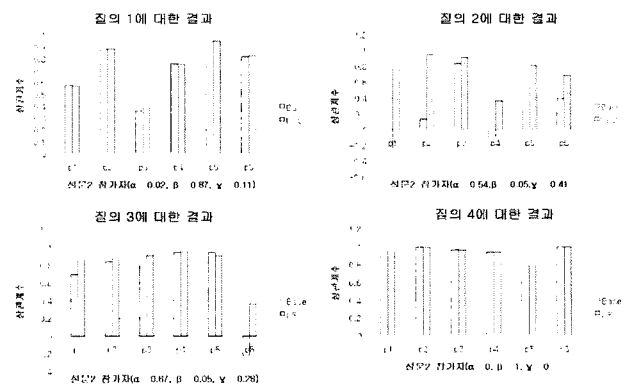


그림 2. 테스트 데이터와 DRC, Base에 대한 상관관계
Fig. 2. Correlation between the test data and the results by DRC or Base

그림 2의 각 그래프에서 x축은 설문조사의 대상인 참가자를 나타내고, y축은 표 1의 테스트 데이터들과 상관관계를 측정한 값이다. 각 그래프의 하단에는 학습 데이터 상에서 결정된 가중치 값을 명시하고 있다. 상관관계가 높을수록 측정값과 테스트 데이터의 값이 유사함을 의미한다.

그림 2에서 대부분의 경우 DRC가 Base와 비슷하거나 더 높은 상관관계를 보임을 알 수 있다. 특히, 질의 2의 경우, Base는 음의 상관관계를 보이기도 하지만, DRC의 경우 P4를 제외한 나머지의 경우 0.7이상의 높은 상관관계를 보이고 있다. 질의 3에 대해서도 P6에 대해 Base는 음의 상관관계를 보이지만, DRC는 다른 경우에 비해 낮지만, 양의 상관관계를 보임을 알 수 있다. 질의 4의 경우 목표 인스턴스를 공유하는 것에 대해서만 가중치값을 가지므로 DRC, Base는 같은 결과를 보이고 있다. 이러한 결과는 목표 인스턴스의 공유에 따른 가중치 반영이 다른 관계에 비해 상대적으로 커다는 사실을 보여준다. 따라서 대부분의 경우, DRC의 결과가 Base에 비해 안정적이고, 테스트 데이터와 더 높은 상관관계를 보이고 있음을 알 수 있다.

5. 관련 연구

Anyanwu 등은 RDF 데이터 모델에서 두 인스턴스 사이의 유의미한 관계를 “의미 관계(semantic association)”라고 정의하고, 의미 관계를 찾는 방법을 제안했다[7]. 이 연구에서 두 인스턴스 사이의 의미 관계는 프라퍼티 연속으로

정의한다. 두 인스턴스는 직접적인 연결 관계뿐만 아니라 다른 인스턴스들을 매개로 연결되는 패스를 통해 간접적인 관계를 가진다. 따라서, 인스턴스 간의 패스를 찾는 방법에 대한 연구가 있어왔다[10, 11, 12]. 최근에는 SPARQL 문법에 기반하여 의미적 관계를 찾는 연구가 있었다[13, 14].

이들 연구를 통해 찾은 두 인스턴스 간의 관계는 그 자체로 새로운 정보를 제공할 뿐 아니라 두 인스턴스 사이의 관계 정도를 측정함으로써 인스턴스 간의 관계에 대한 구체적인 분석이 가능하게 한다. Aleman-Meza 등은[8] Anyanwu의 정의[7]로부터 의미적 관계를 찾을 때, 인스턴스들이 속한 클래스를 구체화함으로써 사용자의 관심 사항을 정확하게 반영할 수 있음을 보였다. 예를 들면, 돈세탁 혐의를 찾고자 할 때, Bank 클래스에 속하는 인스턴스를 포함하는 의미적 관계는 유의미할 수 있다. 이들은 또한 두 인스턴스 사이에 의미적 관계 정도를 측정하는 여섯 가지 방법을 제안했다. 여섯 가지 방법 중 네 가지는 클래스 정보에 기반하여 관계 정도를 측정한다. 따라서, 같은 클래스에 속하는 인스턴스들로 이루어진 의미적 관계들은 이들 네 가지 방법에 대해서 동일한 관계 정도를 나타낸다. 본 논문에서는 이러한 점에 착안하여 클래스와 오브젝트 프라퍼티로 표현되는 단일 컨텍스트를 정의하였다. 하지만 본 논문에서 제시한 단일 컨텍스트는 프라퍼티 연속보다 관계에 대한 의미를 명확하게 제약할 수 있다는 점에서 차별성을 가진다.

두 클래스 사이의 관계 정도를 측정하는 최근 연구에서는[15, 16] 두 클래스 사이의 관계에 대한 인스턴스들의 통계 정보를 이용했다. 두 클래스 사이의 관계를 측정한다는 점에서는 본 논문에서 제안한 단일 컨텍스트와 유사한 점을 가지지만 본 논문에서는 클래스로 표현되는 관계에 대한 두 인스턴스 사이의 관계 정도를 측정한다는 점에서 차별성을 가진다.

두 인스턴스 사이에 관계 정도를 측정함으로써 시맨틱 웹과 관련한 다양한 응용 분야에 관계 정도를 적용할 수 있다. Sheth 등은[6] 테러리스트나 돈세탁 혐의를 찾기 위해 온톨로지에 기반한 의미 관계를 이용하고 있다. 이들은 두 인스턴스 사이에 의미를 가지는 패스를 찾는 것에 연구의 초점을 맞추고 있다. Zhdanova 등은[5] 태그에 대한 네트워크를 모델링하고 이러한 모델에 기반하여 소셜네트워크를 구축하기 위해 커뮤니티를 찾는 알고리즘을 제안했다. 그들은 개념과 인스턴스로 이루어진 일반적인 온톨로지 모델에 주체(actor)를 추가함으로써 확장된 형태의 온톨로지 모델을 제안했다. 그리고 [actor1-link-object-link actor2]와 같은 양방향 링크(bi-directional link)에 기반하여 커뮤니티를 찾는다.

이들 연구에서는 두 객체 사이에 연결된 패스를 의미있게 본다는 공통점을 가진다. 그러나 두 인스턴스 사이에 연결된 패스는 없지만, 스키마를 통해 관계를 가지는 의미있는 관계가 있으며 본 연구에서 이러한 관계를 중요한 관계로 정의하고 있다.

온톨로지에 기반한 추천 시스템에서 Fernandez 등은[17] TV 프로그램 추천을 위해 온톨로지를 개인 프로파일로 활용한다. 두 사용자의 프로그램 시청에 대한 유사도를 측정할 때, 단지 두 사용자 간의 전체적인 유사도만을 측정하며 따라서 본 논문에서처럼 특정 클래스에 대한 공통된 흥미 정도를 측정하지는 않는다.

6. 결 론

본 논문에서는 단일 컨텍스트에 대한 두 인스턴스 사이의 관계에 대해 정의하고 관계 정도를 측정하는 방법을 제안했다. 제안한 방법은 두 인스턴스 사이에 연결된 패스뿐만 아니라, 스키마를 통해 관계하는 정보를 활용했다. 실험 결과, 이러한 스키마를 통한 두 인스턴스 사이의 관계 또한 유의미함을 보였다.

본 논문의 주된 기여는 다음과 같이 요약된다. 먼저, 1) 단일 컨텍스트에 대해 두 인스턴스 사이에 관계 정도를 측정할 문제를 제시하였고, 2) 두 인스턴스 사이의 연결된 패스와 스키마를 통한 관계에 기반하여 관계 정도를 측정하는 방법을 제시하였으며, 3) 실험을 통해 네 가지 질의에 대해 제안한 방법이 사람의 직관에 의한 결과와 높은 상관관계를 가짐을 보였다.

향후 계획으로 제안한 방법을 특정 응용에 적용하여 평가할 것이다. 관계 정도에 대한 유의미한 한계치를 설정함으로써 소셜네트워크를 구성하거나 커뮤니티를 찾는 데 활용할 수 있다.

참 고 문 헌

- [1] Gruber, T., <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
- [2] Tim, B., James, H. and Ora, L., "The Semantic Web, Feature Article", *In Scientific American*, 5(1), 2001.
- [3] McGuinness, D., and Harmelen, F., "OWL Web Ontology Language Overview", *W3C Recommendation*, 2004.
- [4] Li, D., Tim, F., Anupam, J., "Analyzing Social Networks on the Semantic Web", *IEEE Intelligent Systems*, 8(6), 2004.
- [5] Zhdanova, A., Predoiu, L., Pellegrini, T., and Fensel, D., "A Social Networking Model of a Web Community", *In Proc. of the 10th International Symposium on Social Communication*, pp. 22-26, 2007.
- [6] Sheth, A., Aleman-Meza, B., and Arpinar, I., "Semantic Association Identification and Knowledge Discovery for National Security Applications", *In Journal of database Management on Database Technology for Enhancing national Security*, 16(1), pp. 33-53, 2005.
- [7] Anyanwu, K. and Sheth, A., "p-Queries: Enabling Querying for Semantic Associations on the Semantic Web", *In Proc. WWW2003*, pp. 690-699, 2003.
- [8] Aleman-Meza, B., Halaschek-Wiener, C. Arpinar, I. and Sheth, A., "Context-Aware Semantic Association Ranking", *In First International Workshop on Semantic Web and Databases*, pp. 33-50, 2003.
- [9] Aleman-Meza, B., Halaschek-Wiener, C., Arpinar, I., Ramakrishnan, C., and Sheth, A.,

"Ranking Complex Relationships on the Semantic Web", *In IEEE Internet Computing*, 9(4), pp. 37-44, 2005.

- [10] Alexaki S., Christophides V., Karvounarakis G., Plexousakis D., and Tolle K., "The ICS-FORTH RDFSuite: Managing Voluminous RDF Description Bases", *In SemWeb*, pp. 1-13, 2001.
- [11] Alkhateeb, F., Baget JF., Euzenat, J. "Complex path queries for RDF", *Poster paper in International Semantic Web Conference 2005*, pp. 52-53, 2005.
- [12] Barton, S, "Designing Indexing Structure for Discovering Relationships in RDF Graphs", *In Proc. of Databases, Texts, Specifications, and Objects*, pp. 7 - 17, 2004.
- [13] Anyanwu, K., Maduko, A., and Sheth, A., "SPARQ2L: Towards Support for Subgraph Extraction Queries in RDF Databases", *In Proc. www2007*, pp. 797-806, 2007.
- [14] Kochut, K., and Janik, M., "SPARQLeR: Extended Sparql for Semantic Association Discovery", *In LNCS The Semantic Web: Research and Application*, pp. 145-159, 2007.
- [15] Tian, X., Li, H., and Du, X., "Measuring Semantic Association in Domain Ontology", *In Proc. of International Conference on Semantics, Knowledge and Grid*, pp. 515-518, 2007.
- [16] Tian, X., Li, Du, X., and Li, H., "Computing Degree of Association Based on Different Semantic Relationships", *In Proc. 18th International Workshop on Database and Expert Systems Applications*, pp. 372-376, 2007.
- [17] Fernandez, Y., Arias, J., Nores, M., Solla, Q., and Cabrer, M., "AVATAR: An Improved Solution for Personalized TV based on Semantic Inference," *In Consumer Electronics, IEEE Transactions*, 52(1), pp. 223-231, 2006.



박세영 (Se Young Park)
 1980년 : 경북대 전자공학과(학사)
 1992년 : KAIST(석사)
 1999년 : 프랑스 파리 7대학(박사)
 1982~2000년 : ETRI 책임연구원
 2003년~2005년 : 정보통신연구진흥원
 전문위원
 2005년~현재 : 경북대학교 컴퓨터공학부 교수

관심분야 : 시멘틱웹, 자연어 처리, 정보 검색, 소셜네트워크 분석

Phone : 053-950-6551
 Fax : 053-950-2122
 E-mail : seyong@knu.ac.kr



박성배 (Seong-Bae Park)
 1994년 : 한국과학기술원 전산학과 (학사)
 1996년 : 서울대 컴퓨터공학과 (석사)
 2002년 : 서울대 컴퓨터공학과 (박사)
 2004년~현재 : 경북대학교 컴퓨터공학과 교수

관심분야 : 자연어처리, 기계학습, 정보검색, 바이오인포매틱스

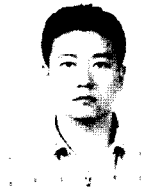
Phone : 053-950-7574
 Fax : 053-957-4846
 E-mail : seongbae@knu.ac.kr



김권양 (Kweon Yang Kim)
 1983년 : 경북대학교 전자공학과(학사)
 1990년 : 경북대학교 전자공학과(석사)
 1998년 : 경북대학교 컴퓨터공학과(박사)
 1983~1988년 : ETRI 연구원
 1999년~2000년 : University of Central Florida 방문교수
 1991년~현재 : 경일대학교 컴퓨터공학부 교수

관심분야 : 시멘틱웹, 한글공학
 Phone : 053-850-7287
 Fax : 053-850-7609
 E-mail : kykim@kiu.ac.kr

저 자 소 개



한용진 (Yong-Jin Han)
 2006년 : 경북대 컴퓨터공학과 (학사)
 2008년 : 경북대 컴퓨터공학과(석사)

관심분야 : 시멘틱웹, 기계학습, 소셜네트워크 분석
 E-mail : yjhan@sejong.knu.ac.kr