

# 분산 환경에서의 쿼리 변환을 위한 온톨로지 매핑 결합

정재은  
영남대학교 전자정보공학부  
(j2jung@intelligent.pe.kr)

온톨로지 기반 분산 정보시스템 환경에서는 시스템들 간의 자동화된 정보 공유를 지원하기 위해서 온톨로지 간의 의미적 이질성(semantic heterogeneity)을 해결해야 한다. 일반적으로 전문가들에게 미리 온톨로지들 간의 명시적 매핑(explicit mapping between ontologies)을 요청하고 있다. 하지만, 온톨로지 매핑의 고비용성 문제 때문에 모든 정보 시스템들의 온톨로지 간의 일대일 매핑이 이루어지기 힘들다. 따라서, 본 논문에서는 분산된 온톨로지들 간의 매핑 정보를 수집하고 재사용(reuse) 하고자 한다. 즉, 분산 환경에서 활용 가능한 온톨로지 매핑 정보의 결합(composition)을 위한 방법론을 제안한다. 이를 통해 주어진 두 온톨로지 간의 매핑 정보를 간접적으로 예측할 수 있게 된다. 특히, 본 연구에서 제안하는 온톨로지 매핑 결합 기법의 성능 평가를 위하여, 자동화된 쿼리 전송(propagation) 및 변환(transformation) 시스템에 적용하였다.

논문접수일 : 2008년 07월    논문수정일 : 2008년 10월    게재확정일 : 2008년 10월    교신저자 : 정재은

## 1. 서론

다양한 분야의 정보시스템들은 각각의 도메인 온톨로지(domain ontology)의 구축을 통해 각 시스템 내의 정보 및 자원의 효과적인 관리와 유통 프로세스들을 실현하였다(Brandt, S. C., et al., 2006; Chau, K. W., 2007; Tsou, J. C., 2008; Chou, T. H., 2008; Lu, T., Guan, F. and Gu, N., et al., 2008). 일반적으로 각각의 정보시스템의 온톨로지들은 해당 분야의 전문가들에 의해 생성된다. 그럼으로써 분산 환경에서는 시스템들 간의 자동화된 정보 공유를 지원하여야 한다. 이와 같은 시스템

상호운용성(interoperability)을 지원하기 위해서는 각각의 도메인 지식을 자동으로 이해하기 위한 별도의 방법론을 필요로 한다. 이와 같은 시스템 간의 의미적 이질성(semantic heterogeneity) 문제를 해결하기 위해서는 각 분야 전문가들에 의해 해당 온톨로지들 간의 의미적 연관 관계, 즉, 온톨로지 매핑(mapping)<sup>1)</sup> 정보를 요구한다.

일반적으로 온톨로지 간의 매핑 과정은 단순히 해당 분야의 전문성 부족뿐만 아니라, 온톨로지의 복잡한 내부 구조(방대한 클래스의 수, 클래스들 간의 관계와 같은) 인식의 어려움들 때문에 매우 고비용의 작업임에 틀림없다.

\* 이 논문은 2008년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(NO. R01-2008-000-10196-0).

도메인 전문가의 개입에 의한 온톨로지 매핑 고비용성 문제를 해결하기 위하여, 주어진 임의의 두 온톨로지 간에 자동화된 매핑 알고리즘에 대한 몇몇 연구들이 진행되었다(Shvaiko, P. and Euzenat, J., 2005).

하지만, 이와 같은 매핑 알고리즘들은 일반적으로 확장성(scalability)에 문제점을 보인다( Euzenat, J., Shavaiko, P., 2007) 온톨로지들 간의 매핑 정보를 명시적이고 직접적으로(explicit and direct mapping) 제시할 수 있다는 점에 반해, 일반적인 분산 환경에서와 같이 온톨로지 개수가 증가함에 따라 온톨로지 매핑 프로세스의 복잡도는 지수적으로 증가하게 된다.

이와 같은 문제점을 해결하기 위해서 본 논문에서는 기존에 이미 존재하는 온톨로지 매핑 정보를 재사용함으로써 정보시스템들 간의 보다 효율적인 시스템 상호운용성을 지원하고자 한다. 다시 말해, 정보시스템  $S_i$ 와  $S_j$ 의 두 온톨로지  $O_i, O_j$  간의 직접적인 매핑  $M(O_i, O_j)$ 을 계산하는 대신에, 기존에 존재하는 매핑  $M(O_i, O_k), M(O_k, O_j)$ 의 적절한 결합(Composition)을 통해  $S_i$ 와  $S_j$ 간의 정보공유를 가능하게 하고자 한다.

이를 위하여, 온톨로지와 온톨로지 매핑 정보의 형식화를 위한 몇 가지 정의들을 언급해야 하며, 특히, 우리는 매핑의 신뢰성을 정량화하기 위해 새로운 척도(measurement)를 소개하고자 한다.

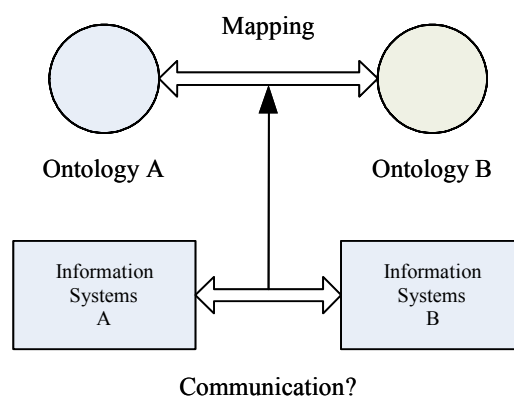
본 연구에서는 이와 같은 매핑 결합 알고리즘의 성능을 평가하기 위해서, 분산 환경에서의 멀티 에이전트 프레임워크를 통해 서로 간의 매핑 정보를 공유할 수 있도록 하였다. 즉, 이질적인 (heterogeneous) 에이전트 시스템들 간의 쿼리 전송을 통해 얼마나 정확한 수준의 정보 검색 효율을 유지

하는지를 측정하고자 한다.

다음의 제 2장에서는 온톨로지 기반의 분산 정보시스템을 구성하는 요소들을 정의하고 의미 유사도(semantic similarity) 기반의 온톨로지 매핑 알고리즘을 소개한다. 또한, 제 3장에서는 온톨로지 매핑의 결합에 대해 설명하고자 하며, 이와 더불어 제 4장에서는 몇 가지 매핑 결합에서의 고려되어야 할 이슈들을 다룬다. 마지막으로 제 5장에서는 본 논문의 결론을 도출하고, 향후 연구 방향을 소개한다.

## 2. 온톨로지 기반의 분산 정보시스템

본 논문에서 다루고 있는 온톨로지 기반의 분산 정보시스템은 정보시스템들 간의 정보 및 지식을 자동화된 모듈(machine-processible)에 의해 서로 공유하기 위해 개발되었다. 즉, 이와 같은 정보시스템은 *i*) 온톨로지와 *ii*) 지식 공유를 위한 정보시스템의 온톨로지와의 매핑 정보로 모델링 된다.



<그림 1> 온톨로지 매핑과 정보시스템의 정보 공유

즉, <그림 1>과 같이, 두 개의 정보시스템 A와 B간에 자동화된 통신을 위해서는 각각 정보시스템

1) 본 논문에서 말하는 온톨로지 매핑은 Alignment와 Matching과 동일한 의미로 간주될 수 있다.

들의 온톨로지 간의 매핑 정보를 활용하여야 한다. 온톨로지는 다양한 형태로 표현될 수 있지만, 본 연구에서의 온톨로지는 다음과 같이 정의된다.

- **Definition 1 (온톨로지)** 온톨로지  $O$ 는 Class들의 집합  $C = \{c_1, \dots, c_N\}$ 와 Class간의 관계를 표현하기 위한 Relation들의 집합  $R = \{r_1, \dots, r_M\}$ 로 이루어진다.

$$O = \langle C, R, T \rangle \quad (1)$$

여기서,  $T = |C| \times |C| \times r$ 는 Class간의 관계를 나타낸다. 간단히, 예를 들어보면 Class들간의  $r$ 이 "subclass"인 경우만 고려한다면, Taxonomy와 같은 형태로 표현된다. 각각의 요소들은 특정 단어(term)로 레이블(labeling) 되어 있다. 온톨로지 매핑을 위한 가장 기본적인 정보이다.

이와 같은 온톨로지들은 임의의 매핑 알고리즘에 의해 온톨로지 요소들 간의 특정 관계가 있음을 알 수 있으며, 이와 같은 요소들 간의 관계를 Correspondences라 부른다.

- **Definition 2 (Correspondence)** 두 온톨로지 간의 Correspondences는  $\langle e_1, e_2, R \rangle$ 와 같은 Triple 형식으로 표현되는데,  $e_1$ 와  $e_2$ 은 각각 온톨로지  $O_1$ 과  $O_2$ 의 요소들이다. 특히,  $R$ 은 두 요소들간의 관계를 표현하는 것으로서, equivalence( $\equiv$ ), subsumption( $\supseteq, \subseteq$ ), disjunction( $\perp$ ) 등이 가능하다.

이와 더불어, 각 Correspondence는 해당 온톨로지 요소들 간의 매핑이 얼마나 신뢰할 수 있는지

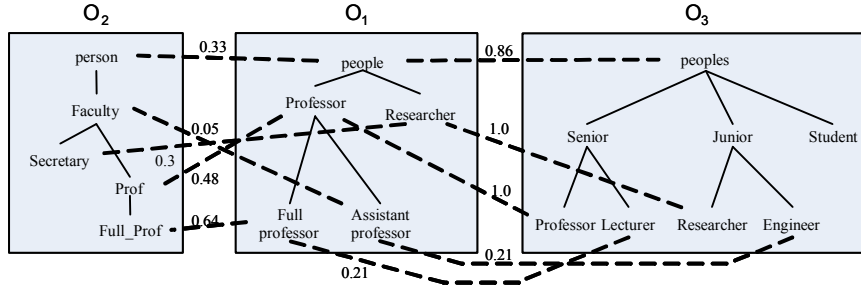
를 정량화하기 위해 새로운 Measurement를 추가하게 된다.

- **Definition 3 (Confidence)** 각각의 Correspondence에는 Confidence 값을 할당하여 신뢰(정확성)의 정도를 알 수 있도록 한다. 따라서  $\langle e_1, e_2, R \rangle$ 는  $\langle e_1, e_2, R, CF \rangle$ 로 확장하여 표현한다.

두 온톨로지  $O_1, O_2$  요소들 간의 Correspondences 집합을 본 논문에서는 온톨로지 매핑(ontology mapping)이라 부르고,  $M(O_1, O_2)$ 로 표현한다. 특히, 본 논문에서와 같이 Alignment API (Euzenat, J., 2004)를 이용한 의미 유사도(semantic similarity) 기반의 온톨로지 매핑 알고리즘을 사용하고 있다고 한다면, Confidence 값 역시 자동으로 구할 수 있다. 온톨로지 매핑 알고리즘은 기본적으로 온톨로지를 이루고 있는 모든 요소들(즉,  $H = \{Class, Relation, Instance\}$ )간의 Lexical similarity의 총합을 최대화 하는 상태를 찾는다 (Euzenat, J., 2004; Jung, J. J., 2007). 다시 말해, 아래 수식의 값을 최대화 하는 mapping 상태를 찾고자 한다.

$$\begin{aligned} Sim_H(O_i, O_j) \\ = \sum_{\Psi \in N(H)} \pi_{\Psi}^H MSim_G(\Psi(O_i), \Psi(O_j)) \end{aligned} \quad (2)$$

여기서,  $N(H)$ 는 온톨로지 내의 모든 구성 요소들의 집합을 가리킨다. 예를 들어,  $H = Class$  라고 한다면  $N(C) = \{Superclass, Subclass, Sibling-class\}$ 라 할 수 있다( $MSim_G$ 는 Equ. (3) 참조). 본 연구에서는 Class와 Relation 만을 고려하므로( $G = \{C, R\}$ ), 위 식은 아래와 같이 표현될 수 있다.



<그림 2> 의미 유사도를 이용한 온톨로지 매핑(Jung, J. J., 2007)과 분산 정보 시스템 환경

$$Sim_H(O_i, O_j) = \sum_C \pi_C^H MSim_C(C(O_i), C(O_j)) + \sum_R \pi_R^H MSim_R(R(O_i), R(O_j))$$

$$MSim_C(S, S') = \frac{\max \left( \sum_{\langle g_x, g_y \rangle \in Pairing(s, s')} Sim_C(g_x, g_y) \right)}{\max(|S|, |S'|)} \quad (3)$$

각 구성 요소의 가중치를 위한  $\pi_{\psi}^H$ 의 총합은 다음과 같이 정규화된다.

$$\sum_{\psi \in N(H)} \pi_{\psi}^H = 1$$

예를 들어, class의 경우, 하나의 클래스와 관련을 갖고 있는 주위의 클래스들(e.g., Subclass, Superclass, Sibling class)을 고려한다면, 위의 수식은 다음과 같이 전개된다.

$$Sim_C(c, c') = \pi_L^C sim_L(L(c), L(c')) + \pi_{SupC}^C MSim_C(SupC(c), SupC(c')) + \pi_{SubC}^C MSim_C(SubC(c), SubC(c')) + \pi_{SibC}^C MSim_C(SibC(c), SibC(c'))$$

여기에서,  $MSim_C$ 은 두 집합 요소들간의 모든 조합들 중에서 유사도의 합을 최대로 하는 상태를 가리키는데, 주어진 두 집합  $S, S'$ 이 주어졌을 때, 최적의 상태는 아래 수식을 이용해서 찾을 수 있다.

여기서, 함수 Pairing은 단순히 두 집합의 구성 요소들간의 모든 가능한 조합을 순차적으로 생성하기 위한 것이다.

간단한 예로서 <그림 2>에서와 같이 두 온톨로지  $O_1$ 와  $O_2$  간의 매핑을 통해 다음 표와 같은 매핑 정보를 얻을 수 있다.

<표 1> 두 온톨로지간의 Correspondences

$\langle e_2, \rangle$	$e_1, \rangle$	$R, \rangle$	$CF \rangle$
$\langle Full\_Prof, \rangle$	Full professor,	$\equiv, \rangle$	.64 >
$\langle Prof, \rangle$	Professor,	$\equiv, \rangle$	.48 >
$\langle person, \rangle$	people,	$\equiv, \rangle$	.33 >
$\langle Secretary, \rangle$	Researcher,	$\equiv, \rangle$	.3 >

이와 같은 정의들을 바탕으로, 본 연구에서 다루고자 하는 온톨로지 기반의 분산 정보시스템을 아래와 같이 표현하고자 한다.

- **Definition 4 (분산 정보시스템)** 온톨로지 기반의 분산 정보시스템  $G$ 에 참여하고 있는 각 정보시스템  $S_i$ 은 로컬 온톨로지  $O_i$ 를 가

고 있으며, 정보공유를 필요로 하는 정보시스템의 온톨로지와 매핑 정보  $M(O_i, O_j)$ 를 저장하고 있다. 따라서,  $G$ 는 다음과 같다.

$$G = \{M(O_i, O_j) | L(S_i, S_j)\} \quad (4)$$

여기서  $L$ 은 정보시스템들 간의 매핑 정보 유무를 나타내기 위한 행렬을 가리킨다.

- **Example 1** : 예를 들어, 분산 정보시스템  $G$ 에 참여한 정보시스템이 <그림 2>와 같이 구성되어 있다고 가정하자. 각 정보시스템들간의 매핑 정보는 i)  $\langle S_1, S_2 \rangle$ 와 ii)  $\langle S_1, S_3 \rangle$ 만이 존재하므로, 아래와 같이 표현된다.

$$G = \left\{ M(O_1, O_2), M(O_1, O_3) \mid L_G = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right\}$$

정보시스템들간에 쿼리 기반의 정보 공유가 이루어진다고 한다면,  $S_1$ 에서 운용 가능한 쿼리  $q = \text{"Researcher"}$ 가  $S_2$ 에서 사용되기 위해서는  $M(O_1, O_2)$ 을 참조함으로써  $q' = \text{"Secretary"}$ 와 같이 Transformation되어 사용될 수 있다.

하지만, 이와 같은 분산 정보시스템에서의 문제는 모든 온톨로지 간의 일대일 매핑 정보를 구하기 불가능하여 매핑 정보가 존재하지 않는 시스템들간의 상호운용성이 힘들다는 것이다. 예를 들어, <그림 2>의  $S_2$ 로부터의 쿼리는  $S_3$  내부의 정보 검색은 불가능하다.

### 3. 쿼리 변환을 위한 매핑 결합

본 연구에서는 정보 시스템들 간의 매핑 정보의

재사용(결합)을 통해 존재하지 않은 매핑 정보를 간접적으로 생성하고자 한다. <그림 2>에서  $O_2, O_3$  간의 매핑이 존재하지 않기 때문에,  $S_2$ 와  $S_3$  간의 통신을 위해서 다음과 같은 간접적인 온톨로지 매핑을 예측하여 사용한다.

$$\widetilde{M}(O_2, O_3) = Compose(M(O_2, O_1), M(O_1, O_3)) \quad (5)$$

이와 같은 매핑 결합을 통해, 정보시스템  $S_2$ 의 에이전트는 간접적으로 생성된 쿼리를 변환하여 시스템  $S_3$ 에 저장된 정보를 검색할 수 있다.

#### 3.1 쿼리 변환

일반적으로 분산 정보 시스템들 간의 통신을 위한 쿼리는 다음과 같이 표현된다.

$$q ::= c | \neg q | q \wedge q' | q \vee q'$$

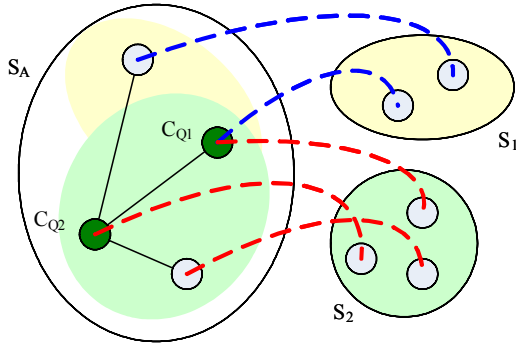
이와 같은 쿼리에 포함되어 있는 class의 추출을 통해 Query-activated Class  $C_Q$ 를 구성한다.

$$C_Q = \{c \mid c \in q, c \in O_{Src}\}$$

쿼리 송신자의 온톨로지( $O_{Src}$ )로부터 얻을 수 있다. 예를 들어, <그림 3>의 정보시스템  $S_A$ 가 쿼리를 생성하였다면,  $C_{Q1}$ 과  $C_{Q2}$ 와 같은 온톨로지  $O_A$ 에 매칭되는 클래스를 찾을 수 있다.

또한, 분산 정보시스템들 간의 통신을 위한 쿼리 언어로써 다음과 같이 SparQL<sup>2)</sup>을 이용할 수 있다. <그림 2>에서 정보시스템  $S_1$ 으로부터 쿼리  $Q_1$ 는 다음과 같이 표현되며 "Full professor"와 "Researcher"에 대한 정보를 찾게 된다.

2) SPARQL, <http://www.w3.org/TR/rdf-sparql-query/>.



<그림 3> 쿼리에 대한 매핑 결합의 선택

$C_{Q2} = \{ \text{"Full\_Prof", "Secretary"} \}$
<Full_Prof, Full professor, 0.64>
<Secretary, Researcher, 0.3>
<Full professor, Lecturer, 0.21>
<Researcher, Researcher, 1.0>
$C_{Q3} = \{ \text{"Lecturer", "Researcher"} \}$

하지만 이와 같은 쿼리 변환을 위한 매핑 결합에 있어서 심각한 문제는 결합을 위한 경로가 하나 이상일 수 있다는 것이다.

```

SELECT ?x, ?y
WHERE {
    ?x ?p "Full professor."
    ?y ?p "Researcher."
}
    
```

여기서, 우리는 Query-activated Class  $C_{Q1}$ 를 다음과 같이 구할 수 있다.

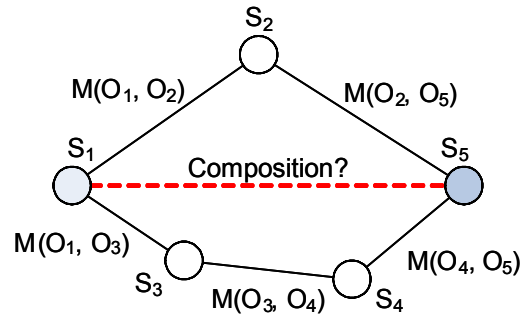
$$C_{Q1} = \{ \text{"Full professor", "Researcher"} \}$$

이 쿼리가 정보시스템  $S_2$ 로 전송된다면, 매핑 정보  $M(O_1, O_2)$ 에 의해서, 해당 쿼리가 아래와 같이  $Q_2$ 로 변환된다.

$$C_{Q2} = \{ \text{"Full_Prof", "Secretary"} \}$$

결과적으로,  $S_2$ 에서 변환된 쿼리의 클래스 "Full\_Prof", "Secretary"가 효과적으로 운용될 수 있다.

이와 더불어, 앞에서 언급한 식 (5)와 같은 매핑 정보의 결합 기법을 통해, 시스템  $S_2$ 으로부터  $S_3$ 로의 쿼리 변환이 가능하다.



<그림 4> 온톨로지 매핑 결합

<그림 4>의 점선과 같이 S1과 S5간의 간접 통신을 위해서는

- $Compose(M(O_1, O_2), M(O_2, O_5))$  또는
- $Compose(M(O_1, O_3), M(O_3, O_4), M(O_4, O_5))$

와 같은 매핑 결합이 가능하다. 따라서, 쿼리 변환에 있어 보다 적합한 매핑 결합을 선택할 수 있어야 한다. 이를 위해서 본 논문에서는 주어진 쿼리의 분석을 통한 휴리스틱 접근법을 제안한다.

### 3.2 Semantic Coverage Ratio

주어진  $C_Q$ 로부터 본 논문에서는 매핑 결합을

위한 경로를 설정하게 되는데, 이 때 Semantic coverage ratio( $\tau_Q$ )를 다음과 같이 두 개의 휴리스틱(H1, H2)을 이용하여 계산하고 비교한다.

- H1 : 쿼리  $Q$ 에 의한 Query-activated Class  $C_Q$ 가 보다 많은 Correspondence 클래스와 일치할수록,  $\tau_Q$ 가 증가한다.

$$\tau_Q^{H1}(S_{Src}, S_{Dest}) = \frac{|\{c \mid c = e_{Src} M(O_{Src}, O_{Dest})\}|}{|C_Q|}$$

- H2 : 쿼리  $C_Q$ 와 일치하는 Correspondence의 CF값이 높을 수록,  $\tau_Q$ 가 증가한다.

$$\tau_Q^{H2}(S_{Src}, S_{Dest}) = \sum_{e_k \in C_Q} CF_k$$

예를 들어, <그림 3>에서 쿼리  $Q$ 에 의해 활성화(activated)된 클래스가 녹색 원들이라고 하자. 휴리스틱 H1에 의하면,  $\tau_Q^{H1}(S_1, S_1) = \frac{1}{2} = 0.5$  와  $\tau_Q^{H1}(S_1, S_2) = \frac{2}{3} = 0.67$ 임을 확인할 수 있으므로, 주어진 쿼리의 변환은 우선  $S_2$ 로 전달될 가능성이 크다.

### 3.3 Transformation Path Selection

최종적으로 우리가 찾아야 하는 정보는 쿼리의 소스(source)로부터 Destination까지의 가능한 매핑 결합 경로들 중에서 최적의 경로이다. 그러므로, 앞에서 소개한 Semantic coverage ratio( $\tau_Q$ )의 Serial Aggregation을 수행한다.

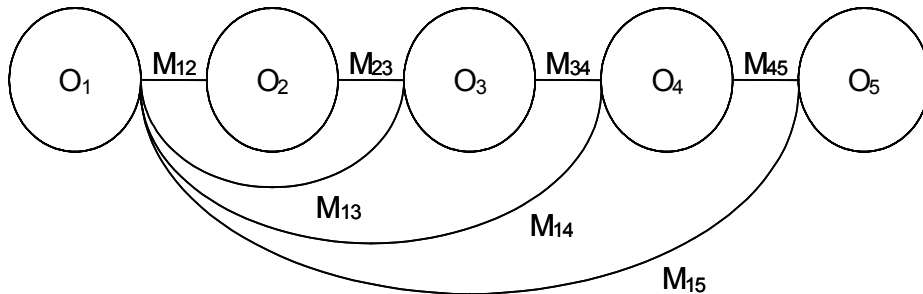
$$\tau_Q(S_{Src}, S_{Dest}) = \max(\tau_Q(S_{Src}, S_i) \cdot \tau_Q(S_i, S_{i+1}) \cdot \dots \cdot \tau_Q(S_j, S_{Dest}))$$

다음의 수식을 최대로 하는 매핑 경로를 통해 쿼리가 변환이 이루어져야 한다.

## 4. 실험 및 토론

본 논문에서 제안하는 온톨로지 매핑 결합의 성능을 실험하기 위해 5개의 정보시스템으로 이루어진 분산 정보시스템을 구축하였으며, 각각의 정보시스템의 자원을 어노테이션(annotation)함으로서 온톨로지  $\{O_1, O_2, O_3, O_4, O_5\}$ 들이 생성되었다 (Jung J.J., Euzenat, J., 2007). 또한 전문가에 의해 모든 온톨로지들간의 매핑 정보를 구하였다.

매핑 정보의 결합에 대한 성능 평가를 위해 <그림 5>과 같이 모든 결합 조합을 비교하였다.



<그림 5> 온톨로지 매핑 결합 평가

### 4.1 Mapping Composition 평가

주어진 온톨로지 간의 매핑이 자바 API인 OLA 알고리즘(Euzenat, J., 2004; Shvaiko, P., Euzenat, J., 2005)에 의해 자동으로 구해졌으며, <표 2>는 온톨로지 매핑 결과를 보여주되 각각 Direct mapping과 비교를 통해 Precision과 Recall를 계산한 결과를 보여준다.

<표 2> 매핑 정보의 결합에 대한 성능 평가

Direct Mapping	Composed Mapping	Recall	Precision
M <sub>12</sub>	M <sub>13</sub> * M <sub>23</sub>	0.76	0.65
	M <sub>14</sub> * M <sub>34</sub> * M <sub>23</sub>	0.73	0.62
	M <sub>15</sub> * M <sub>45</sub> * M <sub>34</sub> * M <sub>23</sub>	0.66	0.6
M <sub>23</sub>	M <sub>24</sub> * M <sub>34</sub>	<b>0.86</b>	<b>0.74</b>
	M <sub>25</sub> * M <sub>45</sub> * M <sub>34</sub>	<b>0.74</b>	<b>0.72</b>
	M <sub>12</sub> * M <sub>15</sub> * M <sub>45</sub> * M <sub>34</sub>	<b>0.72</b>	<b>0.65</b>
M <sub>34</sub>	M <sub>35</sub> * M <sub>45</sub>	0.73	0.66
	M <sub>31</sub> * M <sub>15</sub> * M <sub>45</sub>	0.67	0.63
	M <sub>23</sub> * M <sub>12</sub> * M <sub>15</sub> * M <sub>45</sub>	0.66	0.56
M <sub>45</sub>	M <sub>41</sub> * M <sub>51</sub>	0.64	0.72
	M <sub>42</sub> * M <sub>12</sub> * M <sub>51</sub>	0.63	0.73
	M <sub>43</sub> * M <sub>23</sub> * M <sub>12</sub> * M <sub>51</sub>	0.58	0.63

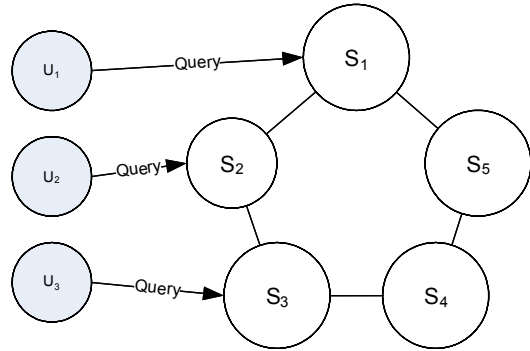
M<sub>23</sub>의 경우 가장 높은 성능을 보였다(Recall = 0.77, Precision = 0.7).

$$P = \frac{|DirectMapping \cap CompositeMapping|}{|CompositeMapping|}$$

$$R = \frac{|DirectMapping \cap CompositeMapping|}{|DirectMapping|}$$

특히, 모든 경우에 있어서 Composition을 계속해서 거듭할수록 Recall과 Precision이 각각 특정 비율로 감소함을 볼 수 있었다.

다음은, 쿼리 전송을 위한 Path 선택에 있어서 중요한 역할을 하게 되는 Semantic coverage ratio의 계산을 위해서 제 3.2장에서 소개되었던 휴리스



<그림 6> 실험을 위한 분산 정보시스템 테스트베드

틱 H1과 H2를 각각 평가해보고자 한다.

실험을 위해 <그림 6>과 같이 세 명의 사용자(U<sub>1</sub>, U<sub>2</sub>, U<sub>3</sub>)들로부터 SPARQL 문법에 따라 생성된 10개의 쿼리를 오로지 S<sub>1</sub>, S<sub>2</sub>, S<sub>3</sub>를 통해서만 시맨틱 어노테이션이 이루어져 있는 실험 데이터(Jung, J. J., 2007)을 검색하도록 하였다.

<표 3> Query path selection 평가

User	IS	Recall	Precision
U <sub>1</sub>	S <sub>1</sub>	0.73	0.55
	S <sub>2</sub>	0.76	0.32
	S <sub>3</sub>	0.65	0.61
	S <sub>4</sub>	0.62	0.74
	S <sub>5</sub>	0.76	0.43
U <sub>2</sub>	S <sub>1</sub>	0.78	0.74
	S <sub>2</sub>	0.89	0.72
	S <sub>3</sub>	0.82	0.65
	S <sub>4</sub>	0.715	0.51
	S <sub>5</sub>	-	-
U <sub>3</sub>	S <sub>1</sub>	0.68	0.56
	S <sub>2</sub>	0.67	0.68
	S <sub>3</sub>	0.82	0.75
	S <sub>4</sub>	0.86	0.34
	S <sub>5</sub>	0.65	0.68

각각의 사용자들은 검색된 결과를 비교함으로써 <표 3>과 같이 Recall과 Precision값을 계산하



였다. 이 결과로부터, Path의 길이가 길어질수록 Recall 값이 감소함을 알 수 있었으며, Precision은 비슷한 수준을 유지하였다.

## 5. 결론 및 향후 연구 방향

Zimmermann and Euzenat(Zimmermann, A., Euzenat, J., 2006)는 분산 정보시스템 환경에서 존재하는 온톨로지 매핑들의 결합(composition)을 생성하는데 있어서 다음과 같은 세 가지 의미가 있음을 정리하였다.

- Simple semantic composition
- Integrated semantic composition
- Contextualized semantic composition

본 연구에서 제안하는 매핑 결합 방법론은 Simple semantic composition과 Integrated semantic composition에 한하여 적용하고자 한다. 하지만, 사용자들로부터 요청된 쿼리의 문맥(context)을 온톨로지 매핑에 반영할 수 있는 방법론을 고려하여야 한다.

기본적으로 분산 환경에서의 소프트웨어 에이전트와 같은 자동화된 시스템들은 필요한 지식 및 정보의 검색을 위해 협업적 통신 기능(또는 서비스)을 필요로 한다. 이를 위해서는 시스템 간의 이질성 해결을 위한 메시지 변환이 효과적이다. 하지만, 변환을 거듭할수록, 초기에 생성된 의도가 상실되는 소위 Whispering problem가 발생한다. 이와 같은 문제를 해결하기 위해서, 본 연구에서는 이형적 분산 환경에서 효과적인 정보 검색을 위한 온톨로지 기반의 쿼리 변환 프레임워크를 제안하였다. 온톨로지 매핑은 온톨로지 구성요소 간의 의미적 유사도를 최대화함으로써 온톨로지 간의 의미적 연관관계를 발견하였으며, 주어진 쿼리를 변

환 및 전달함으로써 정보 손실을 최소화하고자 하였다. 특히, 정보 손실의 최소화를 위해, 최적화된 쿼리 전달 경로상의 매핑 정보의 결합을 통한 쿼리 변화 기법을 제안하였다. 실험을 통해, 본 논문에서 제안한 매핑 결합 방법론이 Recall과 Precision 측면에서 각각 77%, 70%의 성능을 보였는데, 이는 명시적 매핑의 성능과 유사한 수준의 결과를 보였다.

향후 연구로서, 우선 가장 시급한 연구로는 분산 환경에서의 온톨로지 매핑 알고리즘의 복잡도 분석이 필요하다. 또한, 제안된 휴리스틱들의 평가를 위한 시스템의 구현이다. 또한, 이와 같은 온톨로지 매핑 결합을 이용하여, Semantic P2P 환경(Jung J. J., Euzenat, J., 2007)에서의 지식 공유 시스템을 구현할 수 있다. 게다가, 우리는 블로그의 트랙백(trackback)이나 RSS 피드와 같은 기능을 통해 정보 공유 서비스를 제공하고 있다. 하지만, 각 블로그나 사용자의 문맥(context)을 고려하지 않고 있으므로, 제공되는 정보가 대부분 무의미하다. 또한, 블로그 수의 증가뿐만 아니라 Network Isolation 문제에 의해 효과적인 정보 공유가 이루어지기 어렵다. 이 문제점들을 해결하기 위해서, 본 논문에서 소개한 매핑 결합 기법을 이용하여 블로그 시스템의 BlogRoll과 같은 연결 정보를 활용하여 정보공유 서비스를 보다 향상시키고자 한다. 특히, 사회망 분석법을 적용하기 위해, Blog Overlay Network(BON) 플랫폼(정재은, 구철모, 2007)을 설계하여, 문맥(context)에 따른 Community identification을 수행하고자 한다. 이와 더불어, 본 논문에서 제안한 방법론이 블로거들의 Mental 모델에 의해 발생하는 태그(tag) 정보들간의 매칭이 협업을 위한 해당 블로그에 저장되어 있는 정보들 간의 공유를 지원하고 있음을 보이고자 한다.

## 참고문헌

- 정재은, 구철모, “블로그 환경에서의 정보 공유를 위한 상황 비교 기반의 블로그 오버레이 네트워크”, *Telecommunication Review*, 17권 4호 (2007).
- Brandt, S. C., et al., “Ontology-based Information Management in Design Processes”, In : Proc. 9th International Symposium on Process Systems Engineering, 2006.
- Chau, K. W., “An ontology-based knowledge management system for flow and water quality modeling”, *Advanced Engineering Software*, Vol.38, No.3(2007), 172~181.
- Zimmermann, A. and Euzenat, J., “Three Semantics for Distributed Systems and their Relations with Alignment Composition”, In : Proc. 5th International Semantic Web Conference(ISWC'06), 2006, 16~29.
- Euzenat, J., “An API for Ontology Alignment”, In : Proc. Third International Semantic Web Conference(ISWC'04), 2004, 698~712.
- Shvaiko, P. and Euzenat, J., “A Survey of Schema-Based Matching Approaches”, *Journal of Data Semantics IV, LNCS*, Vol.3730(2005), 146~171.
- Jung, J. J., “Ontological Framework based on Contextual Mediation for Collaborative Information Retrieval”, *Information Retrieval*, Vol.10, No.1(2007), 85~109.
- Jung J. J. and Euzenat, J., “Towards Semantic Social Networks”, In : Proc. 4th European Semantic Web Conference(ESWC'07), 2007, 267~280.
- Lanzola, G., Gatti, L., Falasconi, S. and Stefanelli, M., “A framework for building cooperative software agents in medical applications”, *Artificial Intelligence in Medicine*, Vol.16, No.3(1999), 223~249.
- Tsou, J. C., “The ontology of a supply chain model based on IDEF5 and Ontolingua”, *PRODUCTION PLANNING and CONTROL*, Vol.19, No.3(2008), 265~274.
- Chou, T. H., Vassar, J. A. and Lin, B. S., “Knowledge management via ontology development in accounting”, *KYBERNETES*, Vol. 37, No.1-2(2008), 36~48.
- Lu, T., Guan, F. and Gu, N., et al., “Semantic classification and query of engineering drawings in the shipbuilding industry”, *INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH*, Vol.46, No.9(2008), 2471~2483.
- Jung, J. J., “Exploiting semantic annotation to supporting user browsing on the web”, *Knowledge-Based Systems*, Vol.20, No.4(2007), 373~381.
- Euzenat, J. and Shavaiko, P., *Ontology matching*, Springer-Verlag, 2007.

Abstract

## Ontology Mapping Composition for Query Transformation on Distributed Environments

Jason J. Jung\*

Semantic heterogeneity should be overcome to support automated information sharing process between information systems in ontology-based distributed environments. To do so, traditional approaches have been based on explicit mapping between ontologies from human experts of the domain. However, the manual tasks are very expensive, so that it is difficult to obtain ontology mappings between all possible pairs of information systems. Thereby, in this paper, we propose a system to make the existing mapping information sharable and exchangeable. It means that the proposed system can collect the existing mapping information and aggregate them. Consequently, we can estimate the ontology mappings in an indirect manner. In particular, this paper focuses on query propagation on the distributed networks. Once we have the indirect mapping between systems, the queries can be efficiently transformed to automatically exchange knowledge between heterogeneous information systems.

**Key Words** : Semantic P2P Networks, Ontology Mapping, Social Network

---

\* Department of Computer Engineering Yeungnam University

## 저자 소개



정재은

저자는 현재 영남대학교 컴퓨터공학과에 재직 중이다. 연구분야는 Description logic과 Knowledge engineering이다.