

구조 및 의미 정보를 활용한 파스 트리 커널 기반의 온톨로지 정렬 방법

(Ontology Alignment based on Parse Tree Kernel using
Structural and Semantic Information)

손 정 우 [†] 박 성 배 ^{**}
(Jeong-Woo Son) (Seong-Bae Park)

요약 기존 온톨로지 정렬 기법은 두가지 문제점을 가지고 있다. 먼저 자질을 해당 분야 전문가가 정하기 때문에 중요한 자질들이 자질셋에 포함되지 않을 수 있다는 것이다. 다음으로는 온톨로지의 의미 정보와 구조 정보를 이용하여 유사도를 따로 계산한 후, 각각의 실험에 의해 정의된 가중치를 이용하여 전체 유사도를 계산한다. 하지만 온톨로지 상에 나타나는 의미 정보와 구조정보의 상대적인 가중치가 실험적인 방법 혹은 사용자에게 의해 결정되기 때문에 시스템이 특정 온톨로지에 한정되거나 성능이 떨어질 수 있어 문제이다. 본 논문에서는 온톨로지 정렬을 위한 파스 트리 커널을 제안한다. 온톨로지 상의 개체에 대한 유사도를 계산하기 위해 먼저 온톨로지를 트리 구조로 변환한다. 그 후, 변환된 트리 간의 유사도는 온톨로지 정렬을 위해 수정된 파스 트리 커널을 이용하여 계산한다. 이때 자질은 명시적으로 나열하지 않는다. 유사도 계산시, 파스 트리 커널에 근사 스트링 매칭 기법을 적용하여 의미 정보를 반영한다. 검증 위한 실험에서 제안한 방법은 기존의 온톨로지 정렬 기법보다 나은 성능을 보였다.

키워드 : 온톨로지 정렬, 커널메소드, 파스 트리 커널

Abstract The ontology alignment has two kinds of major problems. First, the features used for ontology alignment are usually defined by experts, but it is highly possible for some critical features to be excluded from the feature set. Second, the semantic and the structural similarities are usually computed independently, and then they are combined in an ad-hoc way where the weights are determined heuristically. This paper proposes the modified parse tree kernel (MPTK) for ontology alignment. In order to compute the similarity between entities in the ontologies, a tree is adopted as a representation of an ontology. After transforming an ontology into a set of trees, their similarity is computed using MPTK without explicit enumeration of features. In computing the similarity between trees, the approximate string matching is adopted to naturally reflect not only the structural information but also the semantic information. According to a series of experiments with a standard data set, the kernel method outperforms other structural similarities such as GMO. In addition, the proposed method shows the state-of-the-art performance in the ontology alignment.

Key words : Ontology Alignment, Kernel method, Parse tree kernel

· 본 논문은 지식경제부 및 정보통신진흥원의 정보통신 선도기반 기술개발사업(A1100-0601-0102)의 연구 결과로 수행되었습니다. 또한 본 논문은 2008년도 2단계 두뇌한국(BK)21사업에 의하여 지원되었습니다.

· 이 논문은 2008 한국컴퓨터종합학술대회에서 '구조 및 의미 정보를 활용한 파스 트리 커널 기반의 온톨로지 정렬 방법'의 제목으로 발표된 논문을 확장한 것입니다.

[†] 학생회원 : 경북대학교 컴퓨터공학과
jwson@sejong.knu.ac.kr

^{**} 종신회원 : 경북대학교 컴퓨터공학과 교수
sbpark@sejong.knu.ac.kr

논문접수 : 2008년 8월 25일

심사완료 : 2009년 2월 10일

Copyright©2009 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제36권 제4호(2009.4)

1. 서론

온톨로지(ontology)는 특정 도메인에 대한 사람의 지식을 저장하는 데이터 모델로서 도메인 개체에 대한 추론을 가능하게 한다. 시맨틱 웹 환경에서 온톨로지는 어플리케이션 간의 상호 운용을 가능하게 해준다. 하지만 같은 도메인에 여러 온톨로지가 존재하기 때문에 상호 운용을 위해서는 여러 온톨로지를 결합할 수 있는 방법이 요구된다. 온톨로지 정렬은 온톨로지 결합을 위한 하나의 방법이다.

온톨로지 정렬은 두 온톨로지의 개념(concept)과 개념, 속성(property)와 속성을 정렬하는 방법이다. 온톨로지 정렬에서 가장 중요한 요소는 각 개체간의 유사도를 계산하는 방법이다. 온톨로지 유사도는 구조적 측면과 의미적 측면의 두 가지 요소를 반영해야 한다. 최근의 연구들은 이들 두 요소를 반영함으로써 그 성능을 높이고자 했다. Hu et al.이 제안한 Falcon-AO[1]는 V-Doc[2], I-Sub[3], 그리고 GMO[4]라는 세가지 유사도 기법을 이용하였다. V-Doc와 I-Sub는 의미적 유사도를 GMO는 구조적 유사도를 계산하는데 활용되었다.

기존의 연구들이 실제 활용 가능한 성능을 보였지만, 두 가지 약점을 가지고 있다. 먼저, 대부분의 기법들이 구조적 유사도를 위해 미리 정의된 자질들을 사용했다. 이때, 온톨로지가 속한 영역의 전문가들에 의해 자질이 정의 된다. 하지만 이는 각 개체간의 상호 정보를 배제할 수 있으며, 잘 정의 된다 하더라도 중요한 정보가 자질 집합(feature set)에 포함되지 않을 수 있다. 뿐만 아니라 온톨로지의 분야가 바뀐다면 자질을 재정의 해야 한다.

두번째로 기존의 연구들은 의미적 유사도와 구조적 유사도를 각각 계산한 후, 가중치를 이용하여 결합하였다[5-7]. 이때 가중치는 실험을 통해 정의하였기 때문에 유사도 간의 상대적인 중요성을 반영하기 어렵다.

이와 같은 문제점을 해결하기 위해 본 논문에서는 MPTK(modified parse tree kernel)을 제안한다. MPTK는 온톨로지 정렬을 위해 수정된 파스 트리 커널(parse tree kernel)[8]이다. 파스 트리 커널은 복잡한 데이터 구조를 자질 나열 없이 다루기 위한 컨볼루션 커널(convolution kernel)[9]의 일종이다. 하지만 파스 트리 커널은 구조적 정보만을 고려하고 있다. MPTK에서는 근사 스트링 매칭 기법을 적용하여 구조적 정보와 의미적 정보를 자연스럽게 결합하고자 한다.

제안한 방법을 OAEI(Ontology Alignment Evaluation Initiative) 2007 데이터를 이용하여 평가하였다. 실험 결과, 제안한 방법은 기존의 방법들보다 나은 성능을 보였으며 이는 MPTK가 구조적, 의미적 정보를 잘 반영할 수 있음을 보인다.

2. 파스 트리 커널

파스 트리 커널[8]은 컨볼루션 커널[9]의 하나로 파스 트리들을 다루는데 특화된 커널이다. 파스 트리 커널에서 벡터(vector)의 자질(feature)는 각 파스 트리에 나타날 수 있는 모든 하위트리(subtree)로 이루어진다. 이때, 각 자질의 값은 하위트리의 빈도 수로 할당된다. 하지만 이러한 하위트리를 명시적으로 구한다는 것은 불가능하다. 이에 Collins와 Duffy는 아래 재귀 규칙을 두 파스 트리의 모든 노드에 대해 적용함으로써 명시적인 열거 없이 내적을 구하는 방법을 제시하였다.

규칙 1. n_1 과 n_2 가 다르면

$$C(n_1, n_2) = 0 \tag{1}$$

규칙 2. n_1 과 n_2 가 말단 노드라면

$$C(n_1, n_2) = 1 \tag{2}$$

규칙 3. 그 외

$$C(n_1, n_2) = \prod_i^{nc(n_i)} (1 + C(ch(n_1, i), ch(n_2, i))) \tag{3}$$

이 때, $ch(n_i, i)$ 는 노드 n_i 의 i 번째 자식 노드를 의미한다. 함수 $nc(n_i)$ 는 n_i 의 자식 노드의 수를 반환한다. 위의 알고리즘을 이용하여 파스 트리 T_1 과 T_2 의 내적은 아래와 같이 계산한다.

$$\langle T_1, T_2 \rangle = \sum_{n_1 \in N_1} \sum_{n_2 \in N_2} C(n_1, n_2)$$

3. 수정된 파스 트리 커널을 이용한 온톨로지 정렬

본 논문에서 제안하는 온톨로지 정렬 방법은 3단계로 이루어져 있다. 먼저 두 온톨로지를 트리 구조로 변환한 후, 개념들을 정렬한다. 이때 온톨로지 정렬을 위해 수정된 파스 트리 커널을 사용한다. 개념 정렬 시 계산된 유사도 값은 속성 정렬을 위해 저장한다. 마지막으로 속성들을 정렬한다. 그림 1은 제안한 방법의 전체 흐름을 보여준다.

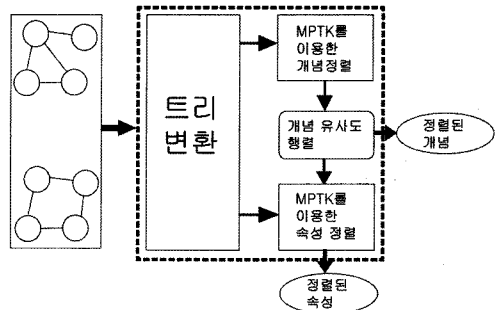


그림 1 제안한 방법의 구조

3.1 트리 변환

온톨로지는 개념, 속성, 객체(instance)로 이루어진 그래프로 볼 수 있다. 그래프 상의 노드는 개념과 객체를 나타내며, 연결선들은 속성이 된다. 개념 측면에서 온톨로지 상의 구조정보는 크게 세가지, 개념과 개념, 개념과 속성, 개념과 객체로 나눌 수 있다. 이러한 정보를 추출하여 유사도를 계산하는데 온톨로지의 그래프 구조를 그대로 이용하는 것은 높은 계산량과 연관된 구조자질을 정의의 어려움으로 인해 쉬운 일이 아니다. 본 논문에서는 온톨로지를 트리 구조로 변환함으로써 이와 같은 문제점을 해결한다. 온톨로지를 트리로 변환할 때, 트리 구조의 깊이를 제한함으로써 그래프의 순환 구조가 트리를 무한히 커지게 하는 것을 방지한다.

본 논문에서 추출하고자 하는 트리 구조는 개념의 경우, 개념-개념, 개념-속성, 개념-객체 관계를 반영하며 속성의 경우 속성-개념의 관계만을 반영한다. 하지만 속성 유사도 계산시, 이미 계산된 개념의 유사도 값을 재사용함으로써 속성-객체, 속성-속성 관계 또한 반영된다 할 수 있다. 그림 2는 본 논문에서 사용된 트리 구조들을 보여준다. 그림에서 (a)는 개념-개념 구조, (b)는 개념-속성, (c)는 개념-객체 관계를 보여준다. 마지막으로 (d)는 속성 트리이다. 예를 들어 “사람”이라는 개념

은 세가지 관계를 온톨로지에서 가진다. 먼저 “사람”과 하위 개념들인 “남자”와 “여자” 사이의 개념과 개념간의 관계, “사람”의 속성인 “나이”, “이름” 등과의 개념과 속성간의 관계, “사람”의 객체인 개별 사람들간의 개념과 객체간의 관계를 가진다. 속성의 경우, “먹다”라는 속성은 속성을 가진 “사람”(domain)과 속성의 값인 “음식”(range) 간의 관계를 가진다.

3.2 수정된 파스 트리 커널

온톨로지를 트리 구조로 변환함으로써 온톨로지의 구조적 정보를 추출하였지만 의미 정보 또한 변환된 트리에 저장되어 있다. 하지만 기존의 파스 트리 커널을 사용할 경우, 식 (1)에서의 매칭(matching) 방법으로 인해 의미 정보를 반영하기 힘들다. 예로 “Person”과 “People”는 같은 의미로 사용되지만 문자열 매칭을 통해서만 알기 힘들다.

이 문제를 해결하기 위해서 본 논문에서 제안하는 온톨로지 정렬을 위한 파스 트리 커널에서는 근사 스트링 매칭 방법을 이용한다. 본 논문에서 사용하는 근사 스트링 매칭은 레벤슈타인(levenshtein) 거리[10]와 문맥(semantic) 거리로 나눌 수 있다.

레벤슈타인 거리는 수정 거리(edit distance)의 일종으로 하나의 문자열에서 다른 문자열로 변환할 때 행해

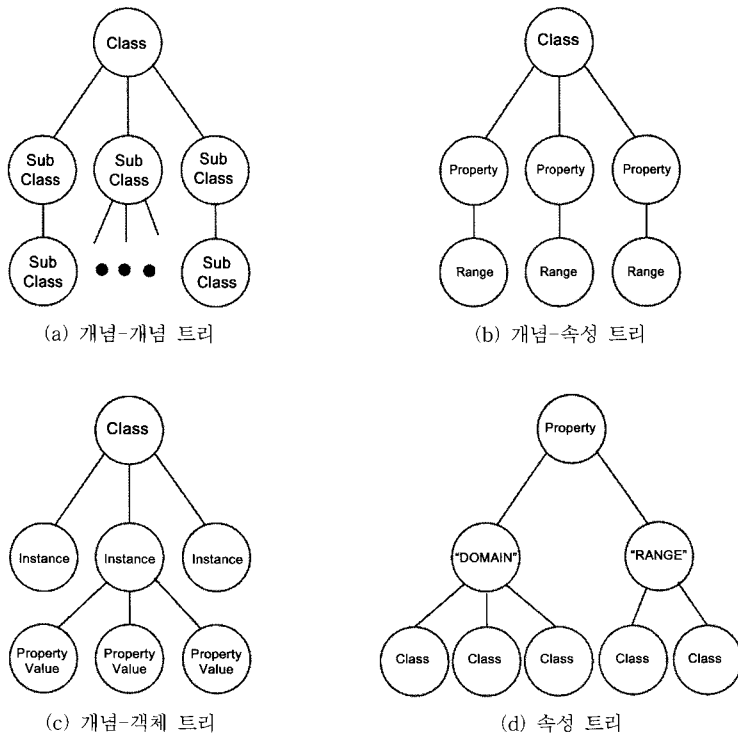


그림 2 온톨로지 변환을 위한 트리

지는 삭제, 추가, 변환의 횟수를 거리로 정의한다. 문맥 거리는 taxonomy[11]를 이용하여 비교하는 것으로 두 문자열의 의미 집합 간의 거리 중, 가장 긴 것을 두 문자열의 거리로 정의한다. 본 논문에서는 문맥 거리로 Resnik's 거리[12], Lin's 거리[13], Jiang-Conrath's 거리[14]를 사용한다.

두 스트링 사이의 거리가 구해지면 유사도는 거리의 반비례 값으로 정의 가능하다. 레벤슈타인 거리의 경우,

$$Sim(c_1, c_2) = 1 - \frac{Dist(c_1, c_2)}{\max(\text{length}(c_1), \text{length}(c_2))},$$

문맥 거리의 경우,

$$Sim(c_1, c_2) = \frac{1}{Dist(c_1, c_2)}$$

이때, c_1 과 c_2 는 비교할 두 개념을 의미한다.

이와 같은 근사 스트링 매칭 기법을 적용하기 위해 파스 트리 커널의 규칙을 아래와 같이 수정 한다.

규칙 1. n_1 과 n_2 가 다르면

$$C(n_1, n_2) = 0$$

규칙 2. n_1 과 n_2 가 말단 노드라면

$$C(n_1, n_2) = Sim(n_1, n_2)$$

규칙 3. 그 외

$$C(n_1, n_2) = \prod_i^{nc(n_1)} (1 + \max_{ch_{ch_1, 2}} C(ch(n_1, i), ch)) \quad (4)$$

위의 규칙에서 ch_{n_2} 는 n_2 의 자식 노드 집합을 의미한다. 수식 4에서 자식 노드들 간의 유사도 값 중, 가장 큰 값을 이용하여 계산하는 것은 온톨로지 정렬이 높은 유사도 값을 가지는 대상을 찾는 것이기에 가장 큰 값을 반영하도록 했다. 이외에도 전체 자식 노드와의 평균 혹은 최저 값을 반영할 수도 있다.

3.3 수정된 파스 트리 커널을 이용한 온톨로지 정렬 방법 구현

본 논문에서 제안하는 온톨로지 정렬 방법은 먼저, 개념 간의 정렬을 수행한다. 각 개념에 대해 앞 절에서 언급했듯이 세가지의 트리를 생성한 후, 각 트리들에 대해 유사도를 계산한다. 계산된 유사도들을 평균을 취해서 전체 유사도를 얻는다. 마지막으로 유사도가 가장 높은 개념들을 정렬한다. 이때, 계산된 개념 유사도는 이차원 행렬에 저장해 둔다.

온톨로지의 개념을 정렬한 후, 속성 정렬을 한다. 이때 이전 단계에서 만들어진 개념 유사도 행렬을 이용한다. 이는 속성 트리의 말단 노드들이 클래스를 나타내는 하나의 노드로 되어 있기 때문에 속성 트리만을 이용할 경우 개념의 구조적 정보가 반영되지 않기 때문이다. 이를 반영하기 위해서는 속성 트리의 말단 노드가 하나의

개념 트리가 되어야 하지만 이는 높은 계산 복잡도를 야기한다.

본 논문에서는 속성 트리를 정렬할 때, 개념의 구조적 정보를 반영하고, 계산량을 낮추기 위해 개념 유사도 행렬을 사용한다. 이를 위해 속성 정렬 시, 앞 절에서 제안한 파스 트리 커널의 두번째와 세번째 규칙 사이에 아래와 같은 규칙을 추가한다.

규칙 2.5. n_1 과 n_2 가 개념이라면

$$C(n_1, n_2) = classSim(n_1, n_2)$$

위의 규칙에서 $classSim()$ 은 개념 유사도 행렬의 n_1 , n_2 위치에서 유사도 값을 가져온다.

4. 실험

4.1 실험 데이터

제안한 방법의 검증에 위해 실험을 했다. 실험 데이터는 OAEI의 2007년도 데이터를 사용하였다. OAEI 2007 데이터는 54개의 온톨로지 로 이루어져 있으며, 이중, 101 온톨로지를 변형하여 102, 30x를 제외한 나머지 온톨로지를 생성하였다. 101 온톨로지는 33개 개념, 24개 오브젝트 속성, 40개의 데이터 타입 속성, 56개의 네임드 객체와 20개의 익명 객체로 이루어져 있다. 102 온톨로지는 완전히 다른 온톨로지이며, 30x 온톨로지는 MIT, UMBC, University of Karlsruhe에 의해 만들어진 온톨로지이다.

4.2 실험 결과

제안한 근사 스트링 매칭 방법 중 온톨로지 정렬에 대한 성능을 비교하기 위해 실험을 했다. 그림 2는 그 결과를 보여준다. 실험에서 Resnik's, Lin's, Jiang-Conrath's, 레벤슈타인 거리 중, Lin's 거리를 기반으로 한 유사도가 가장 좋은 성능을 보였다. 반면 Resnik's 거리를 기반으로 한 유사도는 가장 낮은 성능을 보였다. Resnik's 거리를 제외한 다른 유사도들이 비슷한 성능을 보였는데 이는 실험데이터가 참고문헌과 관련된 온톨로지에 기반하기 때문에 레이블에 나타난 스트링 값들이 사전에 나타나지 않기 때문이다. 제안한 방법은 사전에 나타나지 않은 단어에 대해 레벤슈타인 거리를 적용하기 때문에 비슷한 성능이 나타난 것으로 보인다. 3.2절에서 본 논문에서 제안한 커널을 계산할 때, 자식 노드의 유사도 값 중 가장 큰 값을 이용한다고 언급했다. 이에 대한 성능을 보기 위해 각 근사 스트링 매칭에 대해 두 번의 실험을 했다. 그림 3에서 MV는 평균값을 이용했을 때이며, MNV는 최대값을 이용했을 때이다. 모든 근사 매칭 방법에 대해 MNV가 더 높은 성능을 보임을 알 수 있다.

제안한 방법의 상대적인 성능을 보기 위해 세가지 방

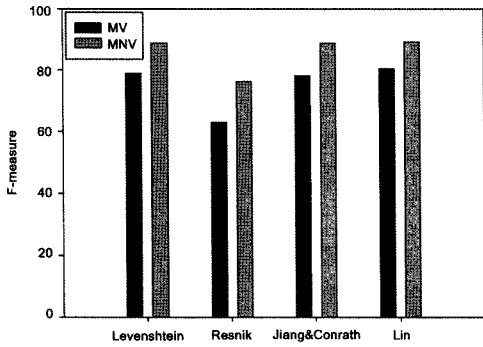


그림 3 근사 스트링 매칭 방법에 따른 성능

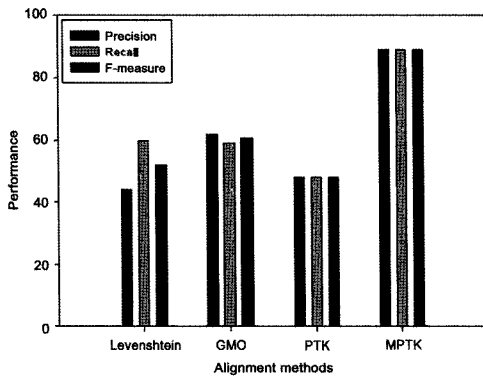


그림 4 온톨로지 정렬 방법의 성능 비교

법과 비교하였다. 레벤슈타인 거리는 가장 많은 정렬 시스템에서 문자열 매칭을 위해 사용된다. 본 실험에서는 이 방법을 기준(baseline)으로 삼았다.

다음으로 GMO는 그래프 매칭 방법의 하나로 Falcon-AO 등의 시스템에서 사용되고 있다. PTK는 온톨로지 정렬을 위해 수정되기 전의 파스 트리 커널이다. 그림 4는 실험 결과를 보여준다. 그림에서 제안한 방법은 89%의 정확율과 89%의 재현율을 보여준다. 이와 같은 결과는 비교한 다른 방법에 비해 제안한 방법이 더 나은 성능을 가지는 것을 보여준다. 특히, 제안한 방법이 파스 트리 커널에 기반하였지만 기존 파스 트리 커널이 온톨로지 정렬에 적합하지 않음이 본 실험에서 나타났다. 이

는 제안한 방법의 커널이 온톨로지 정렬에 맞게 수정되었음을 의미한다.

표 1은 제안한 방법과 OAEI 2007에 참여한 방법들의 전체적인 성능을 보여준다. 제안한 방법은 101, 103, 104, 그리고 221-247 온톨로지에 대해 100의 F-measure를 보였다. 이들 온톨로지는 101 온톨로지의 의미 정보 즉, 문자열은 유지한 채, 구조만을 바꾼 경우이기에 쉽게 정렬 가능한 것으로 보인다.

201-210 온톨로지는 구조적 정보는 유지한 채 문자열을 유의어나 외래어로 대체한 경우로 제안한 방법은 88.5의 성능을 보였다. 이때 나타난 에러의 경우 문자열이 프랑스로 대체 했을 경우로 제안한 방법은 프랑국어 사전을 사용하지 않음으로 해서 발생한 문제로 보인다. 이와 같은 결과는 제안한 방법이 의미 단계에서의 변형에 쉽게 적응함을 보여준다.

248-266 온톨로지는 의미 정보를 문자열을 임의로 정의함으로써 없앤 온톨로지이다. 특히 이들 온톨로지는 구조 정보 역시 변형되었기 때문에 가장 어려운 문제이다. 제안한 방법 역시 가장 낮은 성능인 80.2를 보였으나 다른 방법과 비교해 볼 때, 비슷하거나 높은 성능을 보인다. 마지막으로 301-304 온톨로지는 실제로 사용되는 온톨로지들로 정렬 방법의 성능을 보기 위한 가장 중요한 데이터이다. 제안한 방법은 80.1의 성능을 보였다.

제안한 방법의 전체적인 성능은 89.5로 상위 오위 안에 드는 성능이다. 제안한 방법이 사전을 제외한 어떠한 사전 정보, 실험적 방법을 사용하지 않음을 고려하면 이는 실제적인 온톨로지 정렬 방법으로서 의미있는 결과이다.

5. 결론

온톨로지 정렬은 웹 어플리케이션간의 상호 운영을 위한 가장 중요한 문제 중 하나이다. 하지만 기존 온톨로지 정렬은 두가지 중대한 문제점을 가지고 있다. 먼저, 자질들이 전문가에 의해 정해진다는 점이다. 이는 자질 정의의 어려움으로 인해 중요한 자질의 누락을 야기할 수 있다. 두번째로 구조 정보와 의미 정보의 결합이 실험적 방법에 의해 이루어진다는 점이다.

표 1 OAEI 참여 방법들과의 비교 실험 결과

	MPTK	OLA2	ASMOV	Lily	Prior+	RiMOM	SEMA	X-SOM
101	100	100	100	100	100	100	100	99
103,104	100	100	100	100	100	100	100	98.75
201-210	88.5	90.8	97.4	97.5	91.6	95.8	85.5	78.5
221-247	97	98.8	99.4	99.2	99.4	99.7	97.5	99
248-266	80.2	64	76.5	72.9	57	73.4	51.	26
301-304	80.1	68	83.5	78.8	84	74.6	73	80.5
Overall	89.5	88	92.5	92.5	87	91	82	73

본 논문에서는 파스 트리 커널에 기반한 두 개념 사이의 유사도를 측정하는 방법을 제안하였다. 온톨로지 정렬에 나타난 문제점을 해결하기 위해 기존의 파스 트리 커널을 수정하였으며 수정된 파스 트리 커널은 의미 정보와 구조정보를 자연스럽게 결합한다. 이때 구조 혹은 의미 정보를 위한 자질을 사람이 직접 결정할 필요가 없다.

일련의 실험에서 본 논문에서 제안한 방법이 비교한 다른 방법에 비해 더 나은 성능을 보였다. 이는 제안한 방법이 의미 정보뿐만 아니라 구조정보를 반영하여 온톨로지 정렬을 하기 때문이다. 특히 자질 정의를 하지 않고 유사도를 계산하기 때문에 실제 온톨로지 정렬시 큰 의미를 가질 것으로 본다.

참 고 문 헌

- [1] Hu, W.; Cheng, G.; Zhong, X.; and Qu, Y. 2007. Falcon-AO: Results for OAEI 2007. In *Proceedings of ISWC '07 Workshop on Ontology Matching*, 160-166.
- [2] Qu, Y.; Hu, W.; and Cheng, G. 2006. Constructing Virtual Documents for Ontology Matching. In *Proceedings of WWW '06*, 23-31.
- [3] Stoilos, G.; Stamou, G.; and Kollias, S. 2005. A String Metric for Ontology Alignment. In *Proceedings of ISWC '05*, 623-637.
- [4] Hu, W.; Jiang, N.; Qu, Y.; and Wang, Y. 2005. GMO: A Graph Matching for Ontologies. In *Proceedings of the K-CAP Workshop on Integrating Ontologies*, 41-48.
- [5] Mao, M., and Peng, Y. 2007. The PRIOR+: Results for OAEI campaign 2007. In *Proceedings of ISWC '07 Workshop on Ontology Matching*, 208-215.
- [6] Tan, H. and Lambrix, P. 2007. SAMBO Results for the Ontology Alignment Evaluation Initiative. In *Proceedings of ISWC'07 Workshop on Ontology Matching*, 225-232.
- [7] Tang, J.; Li, J.; Liang, B.; Huang, X.; Li, Y.; and Wang, K. 2006. Using Bayesian Decision for Ontology Alignment. *Journal of Web Semantics* 4(4): 243-262.
- [8] Collins, M., and Duffy, N. 2001. Convolution Kernels for Natural Language. In *Proceedings of NIPS 2001*, 625-632.
- [9] Haussler, D. 1999. Convolution Kernels on Discrete Structures. Technical report, UCS-CRL-99-10, UC Santa Cruz.
- [10] Levenshtein, V. 1965. Binary Codes Capable of Correcting Spurious Insertions and Deletions of ones. *Russian Problemy Peredachi Informatsii* 1, 12-25
- [11] Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.
- [12] Resnik, P. 1995. Using Information Content to Eva-

uate Semantic Similarity. In *Proceedings of IJCAI '95*, 448-453.

- [13] Lin, D. 1998. An Information-Theoretic Definition of Similarity. In *Proceedings of ICML '98*, 296-304.
- [14] Jiang, J., and Conrath, D. 1997. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In *Proceedings of ROCLING X*, 19-33.



손 정 우

2005년 경북대학교 컴퓨터공학과 졸업(학사). 2007년 경북대학교 대학원 컴퓨터공학과 졸업(석사). 2007년~현재 경북대학교 대학원 컴퓨터공학과 박사과정 관심분야는 기계학습, 자연어처리, 온톨로지



박 성 배

1994년 한국과학기술원 컴퓨터공학과 졸업(학사). 1996년 서울대학교 대학원 컴퓨터공학과 졸업(석사). 2002년 서울대학교 대학원 컴퓨터공학과 졸업(박사). 2004년~현재 경북대학교 컴퓨터공학과 교수 관심분야는 기계학습, 자연어처리, 텍스트 마이닝, 정보추출, 생명정보학

트 마이닝, 정보추출, 생명정보학