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

손정우○ 박성배
경북대학교 컴퓨터 공학과
jwson@sejong.knu.ac.kr, sbpark@sejong.knu.ac.kr

# An Ontology Alignment Based on Parse Tree Kernel for Combining Structural and Semantic Information Without Explicit Enumeration of Features

Jeong-Woo Son○ Seong-Bae Park
Kyungpook National University
School of Electrical Engineering and Computer Science
Machine Learning Lab.

## 1. 서론

온톨로지는 특정 도메인에 대한 사람의 지식을 저장하는 데이터 모델로써 도메인의 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. 본 논문에서는 온톨로지 정렬을 위해 구조 및 의미 정보를 활용한 파스 트리 커널 기반의 방법을 제안한다. 제안한 방법은 구조 정보를 활용하여 유사도를 구하는 파스 트리 커널[2]에 의미 정보를 결합하여 구조 정보와 의미 정보를 자연스럽게 결합한다.

## 2. 파스 트리 커널 기반의 온톨로지 정렬

### 2.1 트리 변환

본 논문에서 제안하는 방법은 먼저 온톨로지를 트리 구조로 변환한다. 온톨로지는 컨셉, 프로퍼티, 인스턴스로 이루어진 그래프로 볼 수 있다. 그래프 상의 노드는 컨셉과 인스턴스로 나타내며, 연결선은 프로퍼티가 된다. 이와 같은 온톨로지의 구조적 정보를 트리로 변환하기 위해 트리 템플릿을 사용한다. 그림1은 사용한 트리 템플릿을 보여준다. 먼저 컨셉의 구조 정보를 추출하기 위해 세가지 트리 템플릿을 사용한다. 이들은 컨셉-컨셉, 컨셉-프로퍼티, 컨셉-인스턴스의 정보를 가진다. 다음으로 프로퍼티를 추출하기위해 한가지 트리를 사용한다.

### 2.2 온톨로지 정렬을 위한 파스 트리 커널

파스 트리 커널은 컨볼루션 커널[1]의 하나로 파스 트리들을 다루는데 특화된 커널이다. 파스 트리 커널에서 벡터의 자질는 각 파스 트리에 나타날 수 있는 모든 subtree로 이루어진다. 이때, 각 자질의 값은 subtree의 빈도 수로 할당된다. 두 파스 트리 간의 유사도는 이들 파스 트리 벡터 간의 내적을 이용하여 계산할 수 있다. 하지만 파스 트리 벡터를 생성하기 위해 모든 자질을 나열하는 것은 불가능하다. 이를 명시적으로 나열하지 않고 계산하기 위해 다이나믹 알고리듬을 이용한다.

온톨로지를 트리형태로 변환하여 커널을 이용하여 비교하는 것은 온톨로지의 구조에 중점을 두고 있다. 온톨로지가 가지는 의미 정보는 이러한 구조에 반영되고 있지만 구조를 비교하는 것만으로 의미정보를 충분히 반영할 수는 없다. 이와 같은 문제를 해결하기 위해 본 논문에서는 파스 트리 커널을 이용하여 두 노드를 비교할 때, 정확한 스트링 비교 대신 근사 스트링 비교(approximate string matching)을 이용한다. 스트링의 비교는 일반적으로 두 스트링 사이의 거리를 정의함으로써 구현되어진다. 본 논문에서는 이들 알고리듬 중, Resnik's distance[3], Lin's distance[4], Jiang-Conrath's distance[5]를 사용한다.

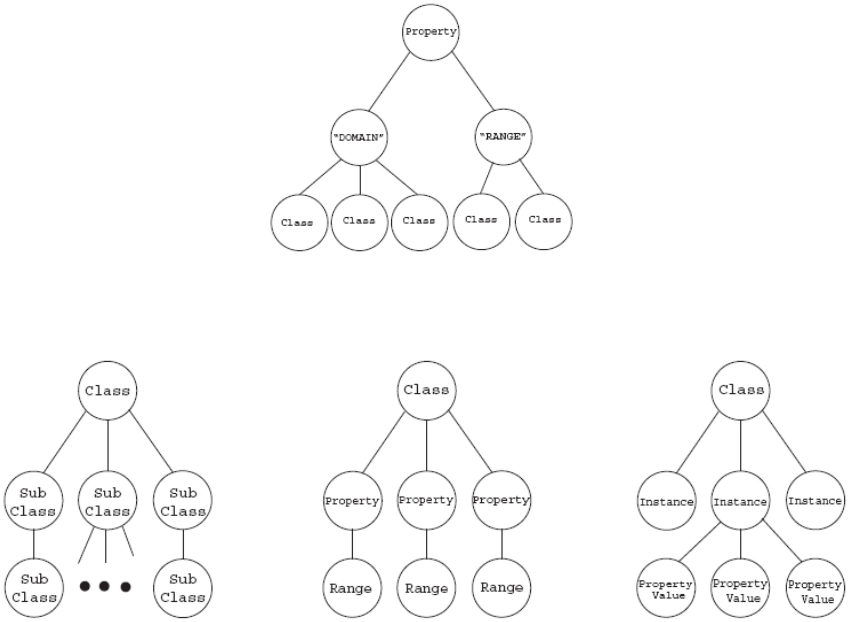


그림 1. 온톨로지 변환을 위한 트리 템플릿

### 3. 실험

제안한 방법을 검증하기 위해 Ontology Alignment Evaluation Initiative 2007 (이하 OAEI 2007)의 데이터를 사용하여 실험을 하였다. OAEI 2007 데이터는 54개의 온톨로지로 이루어져 있으며 이 중, 101 온톨로지를 변형하여 나머지 온톨로지를 만들었다. 표 1은 실험 결과를 보여 준다.

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 |

실험에서 제안한 방법은 비교한 다른 방법들과 비슷한 성능을 보였지만 제안한 방법은 자질을 정의하기 위한 외부 지식을 전혀 사용하지 않기 때문에 실제 온톨로지 정렬시 큰 의미를 가질 것으로 본다.

### 4. 결론

온톨로지 정렬은 웹 어플리케이션간의 상호 운영을 위한 가장 중요한 문제 중 하나이다. 본 논문에서는 파스 트리 커널에 기반한 두 컨셉 사이의 유사도를 측정하는 방법을 제안하여 온톨로지 정렬에 나타난 문제점인 자질 정의의 어려움을 해결하고자 했다. 검증을 위한 실험에서 제안한 방법은 기존의 방법과 비슷한 성능을 보였다. 제안한 방법은 자질을 정의하기 위한 외부 지식이 필요하지 않기 때문에 이와 같은 결과는 실제 온톨로지 정렬시 의미를 가질 것으로 본다.

### Acknowledgement

이 논문은 2008년도 2단계 두뇌한국(BK) 21사업에 의하여 지원 되었음

### 참고문헌

[1] D. Haussler, Convolution Kernels on discrete structures. Technical report, UCS-CRL-99-10, UC Santa Cruz, 1999.

[2] M. Collins and N. Duffy, Convolution Kernels for natural language. In *Procedings of the 15th Neural Information Processing Systems*, pages 625-632, 2001.

[3] P. Resnik, Using information content to evaluate semantic similarity. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence,* pages 448-453, 1995.

[4] D. Lin, An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning,* pages 296-304, 1998.

[5] J. Jiang and D. Conrath, Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of the 10th International Conference on Research in Computational Linguistics.* pages 19-33, 1997.