

한글 DBpedia 온톨로지 스키마 구축

강민서^o, 김재성[†], 김선동[‡], 이재길[§]

^{o§}한국과학기술원 지식서비스공학과

[†]한국과학기술원 전산학과

[‡]한국과학기술원 산업 및 시스템공학과

e-mail : {minseo, kaka7537, sundong.kim, jaegil}@kaist.ac.kr

Building a Schema of the Korean DBpedia Ontology

Min-Seo Kang^o, Jae-Sung Kim[†], Sun-Dong Kim[‡], Jae-Gil Lee[§]

^{o§}Dept. of Knowledge Service Engineering, KAIST

[†]Dept. of Computer Science, KAIST

[‡]Dept. of Industrial & Systems Engineering, KAIST

요 약

시맨틱웹의 구현 도구로써 온톨로지가 있다. 온톨로지는 지식개념의 의미적 연결을 하는데 사용된다. 영어 위키피디아를 토대로 한 영어 DBpedia 온톨로지는 스키마(owl파일 형태)와 인스턴스 모두 잘 구축이 되어있다. 그리고 영어 DBpedia의 각 Class에 한글은 레이블의 형태로 달려있다. 하지만 한글 레이블을 가지고 있지 않은 영어 DBpedia의 Class들이 절반이 넘기 때문에 한글 Class들만으로 된 스키마 구축은 의미가 있다. 한글 Class들로 만들어진 스키마가 있다면 두 한글 온톨로지 사이의 클래스 매칭 알고리즘을 위한 실험이나 한글 온톨로지 자동 증강 알고리즘의 연구 등에 유용하게 쓰일 수 있을 것이다. 본 논문에서 구축한 한글 DBpedia 온톨로지 스키마는 영문 DBpedia 온톨로지의 계층구조와 한글 클래스와 영문 클래스 사이의 매핑정보를 바탕으로 구축되었다. 그리고 기존에 제공되는 한글 DBpedia 온톨로지 클래스의 영어매핑 정보가 있는 한글 프로퍼티와 영어매핑 정보가 없는 한글 프로퍼티를 모두 한글 클래스의 프로퍼티로 입력해주었다.

주제어: Ontology, DBpedia, Schema, Korean

1. 서론

팀 버너스리에 의해 제시된 웹 기술인 시맨틱웹(Semantic Web)은 기존의 웹을 확장해서 컴퓨터가 이해할 수 있는 잘 정의된 의미 기반의 웹을 만드는 것이 목표이다. 이는 다양한 정보자원의 처리를 자동화 하고 데이터의 통합과 재사용 등을 컴퓨터가 인간의 도움 없이도 수행할 수 있게 하여 인간과 컴퓨터 모두 잘 이해할 수 있는 웹을 만드는 것을 포함한다[1].

시맨틱웹의 구현 도구로써 온톨로지(Ontology)가 있다. 온톨로지는 지식개념의 의미적 연결을 하는데 사용된다. 시맨틱웹은 기존의 웹보다 더 정교한 검색을 요구했고, 정보 검색 시스템의 지능화를 촉진하게 되었다. 이 계기를 통해서 웹 자원을 효과적으로 관리하기 위한 새로운 정보 검색의 도구가 필요해졌고 온톨로지가 이러한 도구로써 각광받게 되었다. 온톨로지는 자연어의 기계 번역과 인공지능 분야에서도 사용되며 최근 시맨틱웹과 시맨틱웹 서비스에서 중요한 요소로써 사용되고 있다 [1].

영어 위키피디아[2]를 토대로 한 영어 DBpedia[3] 온톨로지는 스키마(owl파일 형태)와 인스턴스 모두 잘 구축이 되어있다. 그리고 영어 DBpedia의 각 Class에 한글은 레이블의 형태로 달려있다. 하지만 한글 레이블을 가지고 있지 않은 영어 DBpedia의 Class들이 절반이 넘기

때문에 한글 Class들만으로 된 스키마 구축은 의미가 있다. 한글 Class들로 만들어진 스키마가 있다면 두 한글 온톨로지 사이의 클래스 매칭 알고리즘을 위한 실험이나 한글 온톨로지 자동 증강 알고리즘의 연구 등에 유용하게 쓰일 수 있을 것이다.

본 논문에서는 한글 DBpedia 온톨로지 스키마 구축 사례를 바탕으로 그 스키마를 어떻게 구축했는지에 대한 방법과 구축된 DBpedia 스키마로 어떤 활용을 하였는지 그 사례들을 살펴 볼 것이다.

본 논문의 구성은 다음과 같다. 제 2장에서는 온톨로지 스키마 구축과 관련된 연구는 어떤 것이 있는가를 살펴본다. 제 3장에서는 한글 DBpedia 온톨로지 스키마를 어떻게 구축했는지 설명한다. 그리고 제 4장에서는 한글 DBpedia 온톨로지 스키마의 활용 사례들은 어떤 것이 있는지 살펴보고 마지막으로 제 5장에서 결론을 내린다.

2. 관련 연구

인공지능의 한 분야로써 지식개념을 표현하기 위해 RDFS나 OWL과 같은 여러 가지 모델이 등장하였다. 이 모델들은 Entity, Entity간의 관계(Relationship), 각 Entity들이 어떤 클래스에 속하는지 등을 나타내는 방법으로 Subject-Predicate-Object의 형태인 트리플을 이용하여 표현하였다. 그리고 이러한 표현 모델들을 바탕으

로 DBpedia, YAGO[4], freebase.com 등 여러 영어 온톨로지들이 만들어졌다.

Question Answering이나 Information Retrieval과 같은 많은 어플리케이션들은 Knowledge를 활용한다. 아키텍처도, 대부분의 어플리케이션들은 오직 하나의 Knowledge Source만을 활용하고 있었다. 여러 Knowledge Source들을 이용하지 못하고 하나의 Knowledge Source만을 이용하는 것은 이들의 성능을 저하시키는 큰 요소 중에 하나였고, 이들이 여러 Knowledge Source들을 융합한 하나의 거대한 온톨로지를 활용할 수 있다면 그들의 성능을 훨씬 향상시킬 수 있었다. YAGO 온톨로지는 이러한 필요성을 충족시키기 위해 등장한 온톨로지로서, Wikipedia와 WordNet을 융합하였고, 높은 Coverage와 Quality를 제공한다. YAGO 온톨로지는 100만 개 이상의 Entity를 표현하고 있으며(사람, 조직, 도시 등), 각 Entity를 표현하는 500만 개 이상의 Facts와 Is-A 계층 구조, 그리고 Entity간의 관계를 포함하고 있다. 현재는 YAGO 온톨로지의 확장인 YAGO2s[5]가 등장하였는데, 이 온톨로지는 YAGO에 GeoNames를 추가로 융합하여 기존에 불충분했던 시·공간상의 정보를 보완하여 제공하는데 초점을 맞추었다.

YAGO와 YAGO2s 뿐만 아니라 DBpedia, freebase.com 등 풍부한 영어 온톨로지들은 자연어의 기계 번역과 인공지능 분야 등에서 활용되고 있으며 더 정교하고, 지능화된 서비스를 제공할 수 있게 하는 핵심요소가 되었다.

3. 한글 DBpedia 온톨로지 스키마 구축 방법

온톨로지의 스키마를 구축하기 위해서는 우선 클래스들 간에 계층구조가 잘 정의되어야 한다. 기존에 한글 DBpedia와 영어 DBpedia 사이의 클래스 매핑정보(그림 1)와 영어 DBpedia 스키마(그림 2)는 제공이 되어있으므로 이 정보들을 활용해서 한글 DBpedia의 클래스(총 116개)의 계층구조를 만들어주었다(그림 3). 그리고 Activity, Agent, Athlete 등 한글 매핑이 없는 영어 DBpedia 클래스 중에서 계층구조상 중요한 역할을 하는 영어 DBpedia의 클래스들은 영어 클래스를 그대로 넣어서 계층구조를 만들어주었다.

한글 DBpedia 클래스의 한글 프로퍼티는 영문 매핑이 있는 한글 프로퍼티와 영문 매핑이 없는 한글 프로퍼티 두 가지 종류가 있다. 이러한 한글 프로퍼티들은 기존에 구현이 되어 있는 정보이다. 본 한글 DBpedia 온톨로지 스키마에서는 두 종류의 한글 프로퍼티를 모두 한글 클래스의 프로퍼티로 입력해주었다. 그리고 계층구조상 중요한 역할을 해서 넣어주었던 영어 DBpedia 클래스의 영어 프로퍼티들도 넣어주었다.

Mapping ko:NBA 팀 정보

This is the mapping for the Wikipedia template [NBA 팀 정보](#). Find usages of this Wikipedia template [here](#).

Test this mapping (or in namespace [File](#) or [Creator](#)) with some example Wikipedia pages. Check which properties are not mapped yet.

[Read more about mapping Wikipedia templates.](#)

Template Mapping (help)	
map to class	BasketballTeam

그림 1 한글 DBpedia 클래스 'NBA 팀 정보'와 영어 DBpedia 클래스 'BasketballTeam' 사이의 매핑

Ontology Classes

- [owl:Thing](#)
 - [Activity](#) (edit)
 - [Game](#) (edit)
 - [BoardGame](#) (edit)
 - [Sport](#) (edit)
 - [Athletics](#) (edit)
 - [Boxing](#) (edit)
 - [BoxingCategory](#) (edit)
 - [BoxingStyle](#) (edit)
 - [HorseRiding](#) (edit)
 - [Agent](#) (edit)
 - [Deity](#) (edit)
 - [Family](#) (edit)
 - [NobleFamily](#) (edit)
 - [Organisation](#) (edit)
 - [Band](#) (edit)
 - [Broadcaster](#) (edit)
 - [BroadcastNetwork](#) (edit)
 - [RadioStation](#) (edit)
 - [TelevisionStation](#) (edit)
 - [ClericalOrder](#) (edit)
 - [ComedyGroup](#) (edit)
 - [Company](#) (edit)
 - [Airline](#) (edit)
 - [Brewery](#) (edit)
 - [BusCompany](#) (edit)
 - [LawFirm](#) (edit)
 - [Publisher](#) (edit)
 - [RecordLabel](#) (edit)
 - [Winery](#) (edit)
 - [EducationalInstitution](#) (edit)
 - [College](#) (edit)
 - [Library](#) (edit)
 - [School](#) (edit)
 - [University](#) (edit)
 - [GeopoliticalOrganisation](#) (edit)
 - [GovernmentAgency](#) (edit)
 - [Group](#) (edit)
 - [Legislature](#) (edit)

그림 2 영어 DBpedia 스키마



그림 3 한글 DBpedia 클래스의 계층구조

4. 한글 DBpedia 온톨로지 스키마 활용 사례 4.1 한글 DBpedia의 자동 스키마 진화

위키피디아를 토대로 만든 온톨로지인 DBpedia가 있다. 이러한 온톨로지의 자동증강 방법은 웹의 구조화에 큰 기여를 할 수 있다. 2013년에 시작된 엑소브레인 프로젝트¹⁾에는 한글 DBpedia를 기반으로 새로운 트리플을 입력 받아서 기존의 온톨로지를 자동 증강시키는 시스템을 개발하는 것이 있다(그림 4). 이 연구에서 온톨로지의 스키마를 자동 증강하는 두 가지 알고리즘을 제안하는데 두 알고리즘 모두 인스턴스가 지닌 프로퍼티, 즉 rdf-triple 단위에서 진행된다. 이 두 알고리즘들을 사용한 결과는 다음과 같다. 첫째, 단계별로 입력받은 인스턴스와 하위 클래스의 프로퍼티를 이용하여 확률적 격상 알고리즘을 적용한 결과 상위 클래스의 스키마가 정교해졌다. 둘째, 클래스에 속해있지 않았던 인스턴스들이 자동 증강 알고리즘을 통해 가장 적합한 클래스에 속하게 되었다. 온톨로지가 정교해지면서 다시 분류된 인

스턴스들과 새로 들어온 트리플 셋들을 바탕으로 이 두 알고리즘들은 반복적으로 작동한다. 그리고 이를 통해서 한글 DBpedia 온톨로지를 자동증강 하였다[6].

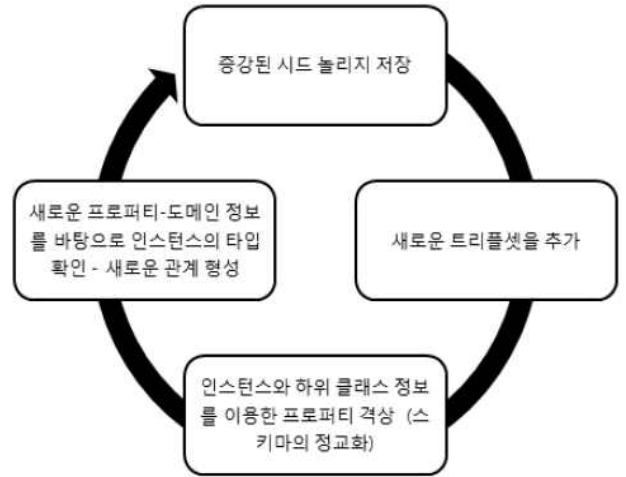


그림 4 한글 DBpedia 지식베이스 자동증강 원리[6]

본 논문에서 구축한 한글 DBpedia 온톨로지 스키마는 앞서 설명한 알고리즘의 실험에 사용되었다. 지식베이스의 자동증강 알고리즘의 실험을 위해서는 DBpedia 같이 다양한 클래스, 프로퍼티, 인스턴스를 포함하는 온톨로지가 필요했고 본 논문에서 구축한 온톨로지는 실험 데이터셋의 역할에 적합하였다.

4-2 서로 다른 두 온톨로지 사이의 클래스 매핑

서로 다른 두 온톨로지 사이의 클래스 매핑은 온톨로지의 중요한 연구분야 중 하나이다. 두 온톨로지 사이의 클래스 매핑을 하게 되면 여러 온톨로지에서 제공하는 정보를 모두 이용하여 Query에 답을 해줄 수 있다. 또한 다른 온톨로지를 활용하여 특정 온톨로지를 개선 할 수도 있다. 서로 다른 두 온톨로지 사이의 클래스 매핑 알고리즘을 개발하고 성능을 테스트하기 위해서는 다양한 클래스, 프로퍼티, 인스턴스를 가진 규모가 큰 온톨로지가 필요하다. 본 논문에서 구축한 한글 DBpedia 온톨로지 스키마는 이러한 알고리즘의 성능을 측정하기 위한 데이터셋으로 적합하다.

5. 결론

현재 한글로 된 잘 갖춰진 온톨로지 스키마가 없는 상황에서 본 논문에서 구축한 한글 DBpedia 온톨로지 스키마는 향후 한글 온톨로지 연구에서 중요한 역할을 할 수 있을 것이다. 이 온톨로지 스키마는 지식베이스 자동 증강 알고리즘의 성능을 테스트하기 위한 데이터셋으로 사용 가능하고 서로 다른 두 온톨로지 사이의 클래스 매핑 알고리즘을 개발할 때에도 사용 가능하다. 본 온톨로지 스키마는 영문 DBpedia 온톨로지의 계층구조와 한글 클래스와 영문 클래스 사이의 매핑정보를 바탕으로 구축되

1) <http://exobrain.kr/twodintro>

었다. 그리고 기존에 제공되는 한글 DBpedia 온톨로지 클래스의 영어매핑 정보가 있는 한글 프로퍼티와 없는 한글 프로퍼티를 모두 한글 클래스의 프로퍼티로 입력해 주었다.

6. 사사

본 연구는 미래창조과학부 및 정보통신기술연구진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음. [10044494, WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발]

참고문헌

- [1] 강민서, 김재성, 김선동, 이재길, "증강된 온톨로지를 위한 시각화 도구", 한국정보과학회 2014 한국컴퓨터종합학술대회 논문집, pp.285-287, 2014.
- [2] Wikipedia. <http://www.wikipedia.org>, 2014.9.12.
- [3] DBpedia. <http://www.dbpedia.org>, 2014.9.12.
- [4] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum, "YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia", In Proceedings of the 16th International Conference on World Wide Web. ACM, pp.697-706, 2007.
- [5] Johannes Hoffart, et al, "YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia", Artificial Intelligence Journal, 194, pp.28-61, 2013.
- [6] 김선동, 강민서, 이재길, "한국어 디비피디아의 자동 스키마 진화를 위한 방법", 한국정보처리학회 2014 춘계학술발표대회 논문집, pp.741-744, 2014.