

온톨로지 기반 지식검색 시스템의 시맨틱 서비스

홍광희⁰ 박제현 박주영 최중민
한양대학교 컴퓨터공학과
{khhong⁰, jhpark, parkjy, jmchoi}@cse.hanyang.ac.kr

Semantic Service in Ontology-based Knowledge Search System

Kwanghee Hong⁰ Jaehyun Park Juyoung Park Joongmin Choi
Department of Computer Science & Engineering
Hanyang University

요 약

본 논문은 시맨틱 웹의 온톨로지 기술을 이용하여 온톨로지 기반 시맨틱 검색 시스템과 서비스 시스템에 대한 구현을 위하여 문헌 정보를 대상으로 온톨로지를 OWL로 구성하였다. 여기에서 제안한 시스템은 효율적인 관리가 가능하도록 웹(Web) 환경에서 구현하였으며, 데이터의 의미와 구문 및 구조의 통일을 위해 메타데이터의 상호운용성을 고려하며 기술언어로는 XML를 사용한다. 웹 문서에서 특정정보를 추출하고 RDF 메타데이터를 생성하며, 웹 문서의 내용을 정확하게 분류하기 위해 온톨로지(Ontology)를 기반으로 한 지식검색 시스템의 시맨틱 서비스시스템을 제안한다.

1. 서론

인터넷과 정보통신 기술의 발달로 인하여 웹에서의 데이터양은 급속히 증가하였으며, 이에 따라 '정보과다'라는 새로운 문제점이 발생하게 되었다. 이러한 문제점을 해결하기 위하여 다양한 검색엔진이 개발, 연구되고 있지만, 대부분의 검색엔진이 웹 문서의 내용보다는 단어나 구문 등 단편적인 방법으로 관련성을 검사하므로, 사용자가 의미적으로 원하는 문서의 검색이 어려운 실정이다. 검색엔진에 의하여 찾아진 내용 중에는 사용자가 실제로 원하지 않는 내용들이 많이 포함되어 있으므로, 검색 만족도는 낮아지며, 검색 시간도 많이 소요되는 편이다. 따라서 이에 대한 근본적인 해결을 위해 시맨틱 웹(Semantic Web)이 등장하게 되었다. 시맨틱 웹은 웹 문서에 대해 검색어가 가진 의미가 명시적으로 표현된 새로운 웹을 말한다. [3][4] 시맨틱 웹의 목적은 컴퓨터로 하여금 웹 자원을 지능적으로 처리하고 통합할 수 있게끔 만드는 데 있다. 다시 말해서, 검색 시 의미적으로 관련 있는 웹의 검색을 가능하게 하며, 지능적 인터넷 에이전트를 개발하여 여러 이질적인 시스템간의 통신을 실현시켜 줄 수 있다.

정보 자원을 효율적으로 검색하기 위한 방법으로, 메타데이터를 추가한 웹 문서의 내용 검색 모델이 연구되고 있으며, 이러한 검색 모델을 통하여 사용자는 원하는 정보에 신속하고 정확하게 접근할 수 있다. 메타데이터는 정보에

태그를 부착하며, 텍스트 자료와 비텍스트 자료를 정확하게 검색할 뿐만 아니라 급증하는 각종 정보자원에 대한 정보관리 등의 다양한 목적을 위해 사용된다. 이러한 시점에서 대두된 RDF는 각각 다른 기술구조를 가진 다양한 메타데이터를 상호운용성의 입장에서 통합하기 위한 메타데이터 구조이다.[3] 또한, 어떤 단일 형식의 메타데이터도 모든 조건을 만족시킬 수 없으므로 메타데이터의 다양성을 인정하고 이를 포괄적으로 표현할 수 있는 구조가 RDF이다. 더블린 코어(Dublin Core)나 워릭(Warwick) 구조는 상이한 메타데이터로 기술된 웹 자원에 효과적으로 접근하기 위한 것으로, 데이터의 의미와 구문 및 구조의 통일을 위해 메타데이터의 상호운용성을 고려하며 기술언어로는 XML(Extensible Markup Language)를 사용한다.[11]

본 논문에서는 웹 문서에서 특정 정보를 추출하고 웹 문서에 대해 RDF 메타데이터를 생성하며, 웹 문서의 내용을 정확하게 분류하기 위해 온톨로지(Ontology)를 기반으로 하는 시맨틱 서비스 시스템을 제안한다.

2. 관련 연구

2.1 시맨틱 웹의 핵심기술

HTML을 이용한 웹은 네트워크 기술의 발달과 함께 급성장을 하였지만 HTML이 가지고 있는 한계점으로 인하여 더 이상의 기능적인 성장을 기대하기 어렵게 되었다. 따라서 W3C(World Wide Web Consortium)는 현재의 웹보다 더 나은 기능성과 상호 운용성을 지닌 새로운 웹인 시

맨틱 웹을 주장하게 되었다. 웹의 창시자인 팀 버너스 리는 URI(Uniform Resource Identifier), HTTP, HTML 기반으로 하는 시맨틱 웹을 고안하였고, 이는 시맨틱 웹 기술 언어들이 가지고 있는 표현력과 기능성에 따라 계층으로 나뉘어져 있다.

시맨틱 웹이 지금의 웹과 가장 큰 차이점 중의 하나는 웹을 사용하는 주체가 사람뿐만 아니라 컴퓨터도 포함된다는 사실이다, 즉 시맨틱 웹 기술 언어를 사용하여 에이전트(agent)가 능동적으로 웹을 사용할 수 있게 되고, 웹에 존재하는 지식을 컴퓨터가 이해하고 이용할 수 있다. 대표적인 시맨틱 웹 기술 언어로는 RDF, RDF Schema, DAML+OIL 및 OWL 등이 있다.

2.2 RDF/RDFS

RDF (Resource Description Framework)는 웹 상에 존재하는 자원을 표현하기 위해 만들어졌다. 기본적으로 XML을 기반으로 하며, 자원에 관한 내용은 주어(Subject), 술어(Predicate), 목적어(Object)의 조합 형태로 표현된다. RDF 명제의 목적어는 다른 자원일 수도 있고, 리터럴(Literal)이 될 수도 있다.

표 1 은 "http://www.w3.org/Home/Lassila" 라는 문서를 "http://www.w3.org/staffld/85740"이 작성했으며, 그 이름은 Ora Lassila이고, 전자우편주소는 "lassila@w3.org"라는 사실을 나타낸 것이다.

표 1 RDF 예제

```
<rdf:RDF>
  <rdf:Description
    about="http://www.w3.org/Home/Lassila">
    <s:Creator
      resource="http://www.w3.org/staffld/85740"/>
  </rdf:Description>
  <rdf:Description
    about="http://www.w3.org/staffld/85740">
    <v:name>Ora Lassila</v:name>
    <v:Email>lassila@w3.org</v:Email>
  </rdf:Description>
</rdf:RDF>
```

또한 RDF 표현은 표 2 와 같이 Subject, Predicate, Object의 형태로 나타낼 수 있으며, 시맨틱 웹은 이러한 그래프 형태로 노드와 노드 사이가 branch로 계속 연결되는 형태이다. 시맨틱 웹을 사용하는 사용자나 에이전트는 이런 그래프 형태로 모델링된 정보를 이용하게 된다.[4][11]

RDF는 정해진 표준 형식에 따라 기술함으로써 에이전트나 컴퓨터도 이해 할 수 있는 틀을 제공한다.

표 2 RDF 예제의 Triple 형태

Subject	Predicate	Object
http://www.w3.org/home/Lassila	s:Creator	http://www.w3.org/staffld/85740
http://www.w3.org/starrld/85740	v:Name	Ora Lassila
http://www.w3.org/staffld/85740	v:Email	lassila@w3.org

RDFS (RDF Schema)는 RDF에 객체지향 개념을 도입하여 웹 상의 자원을 클래스와 인스턴스로 구분할 수 있도록 한 것이다. RDF는 존재하는 자원과 자원간의 관계만 표현할 수 있지만, RDFS는 클래스와 속성 개념이 추가되어, 클래스와 클래스간의 관계를 나타낼 수 있을 뿐만 아니라 각 클래스의 인스턴스는 서로 다른 속성 값을 가질 수 있도록 구성되어 있다. [12]

표 3 RDFS 예제

```
<rdf:RDF>
  <rdfs:Class rdf:ID="Book"/>
  <rdfs:Class rdf:ID="Author"/>
  <rdfs:Property rdf:ID="hasWritten">
    <rdfs:domain rdf:ID="Author"/>
    <rdfs:range rdf:ID="Book"/>
  </rdfs:Property>
</rdf:RDF>
```

표 3은 RDFS 예제이며, "Book"과 "Author"라는 클래스를 지정하고, 이 두 클래스는 "hasWritten"이라는 속성(Property)으로 표현되어 있다.

RDF Schema는 Metadata Schema의 속성과 클래스의 정의, 클래스와 클래스간의 관계, 속성과 속성간의 관계 등을 정의하는 언어로 구성되어 있다. XML Schema 는 다양한 데이터 유형(type)의 지원과 인스턴스 조절에 큰 장점이 있으며, RDF Schema는 요소들의 의미 표현이 유리하므로 이 두 가지 기술들은 상호보완적인 것으로 이해하면 된다. 모든 메타데이터가 RDF의 Triple로 표현되면 자원에 대한 표현이 Subject, Predicate, Object를 갖춘 논리적 단위로 저장되기 때문에 거대한 Triple 지식베이스를 생성하게 되는 효과가 있다. [12][15] 이 지식베이스를 활용하여

논리적 추론이 가능한 새로운 유형의 질의에 답할 수 있고 Triple에 직접적으로 표현되어 있지 않은 지식의 추론도 가능하게 된다.

질의에 대해서 RDF 형태로 응답한다.

2.3 OWL

OWL(Ontology Web Language)은 W3C에서 시맨틱 웹을 위해 온톨로지를 표현할 수 있는 가장 최근에 제정된 표준 언어이다. OWL은 DAML+OIL보다 향상된 표현능력을 가지고 있으며, 클래스와 속성 간에 다양한 관계를 정의할 수 있다.[14] OWL에서 모든 객체는 owl:Thing 클래스의 인스턴스이며 사용자가 정의한 모든 클래스는 owl:Thing 클래스의 서브클래스가 되며, 속성에 대한 다양한 특성을 기술하는 어휘들이 제공된다. 또한 속성은 인스턴스의 속성이 가질 수 있는 값의 범위를 제한하며, 속성들 간의 논리적인 관계를 표현하기 위한 어휘와, 클래스의 논리적 관계를 표현하는 어휘 등이 있다.

2.4 RQL

RQL(RDF Query Language)은 그래프 구조의 RDF 문서에 대해 질의(Query)를 목적으로 하며, SQL과 유사한 형태로 구성되어 있다. [15]

표 4 RQL의 형식

질의명령	RQL의 형식
SELECT	Variables
FROM	Documents
WHERE	Expressions
AND	Filters
USING	Namespace Declarations

3. 시스템구조

본 논문에서는 시맨틱 웹의 핵심 기술인 온톨로지 기술을 이용하여 온톨로지 기반 시맨틱 검색 시스템과 서비스 시스템에 대한 구현을 위하여 문헌 정보를 대상으로 온톨로지를 OWL로 구성하였다. 여기에서 제안한 시스템은 효율적인 관리가 가능하도록 웹(Web) 환경에서 구현하였다. 온톨로지를 이용한 시맨틱 검색을 수행하며, 사용자 주석을 지원하여, 지능형 웹 서비스를 통한 지식의 공유와 확장이 가능하다.[6] 온톨로지 전체 구조는 그래프 구조를 가지며, 각 노드는 클래스, 속성, 객체 속성, 데이터타입 속성 등으로 표시된다. 전체 시스템 구조는 그림 1 과 같다. 이 때 모든 정보는 RDF 형태로 되어 있으며, 웹 서비스를 통해 정보를 얻고자 하는 클라이언트나 컴퓨터에게 RQL

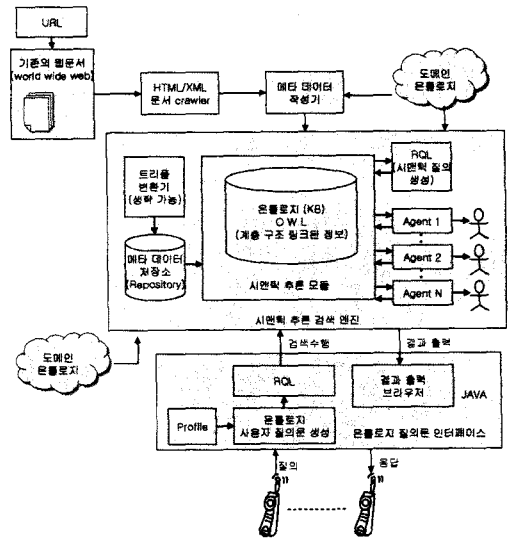


그림 1 시스템 구조

3.1 OWL로 표현된 온톨로지

본 논문에서 제안하는 시스템의 온톨로지는 OWL로 표현하였으며 OWL 언어로 작성된 온톨로지를 그대로 사용하여 추론을 할 수 없으므로 온톨로지의 변환이 필요하다. OWL 언어를 추론엔진을 통하여 일차논리형식(FOL: First Order Logic)의 추론을 하기 위해서는 온톨로지를 트리플 형식으로 변환해야 한다.

표 5 OWL로 표현된 온톨로지 예제

```

<owl:ObjectProperty
  rdf:about="http://lina.hanyang.ac.kr:8080
    /project/books.rdf#price">
  <rdfs:domain rdf:resource=
    "http://lina.hanyang.ac.kr:8080
    /project/books.rdf#Book"/>
</owl:ObjectProperty>
<project:Book
  rdf:about="http://lina.hanyang.ac.kr:8080
    /project/books.rdf#book1">
  <dc:title>Postfix:The Definitive Guide</dc:title>
  <dc:creator>Kyle Dent</dc:creator>
  <dc:publisher>O'Reilly Media</dc:publisher>
  <dc:date>December 1, 2003</dc:date>
  <dc:subject
    rdf:resource="http://lina.hanyang.ac.kr:8080
    /project/books.rdf#agents"/>
  <project:price>34.95</project:price>
</project:Book>
    
```

3.2 RQL로 표현한 질의

표 6 에서는 '기계학습'이라는 주제를 가진 책의 목록을 찾기 위해서 "Book"이라는 클래스의 인스턴스 중 "dc:subject"값으로 "machineLearning"이라는 인스턴스 값을 가지는 것을 찾을 수 있다. 표 6 은 이 내용을 SeRQL 형태로 구성한 것이다.

표 6 RQL의 Query 예제

```
SELECT DISTINCT S
FROM {S} rdf:type {O}, {S} dc:subject {SUBJ}
WHERE O=<http://lina.hanyang.ac.kr:8080
/project/books.rdf#Book>
AND SUBJ=<http://lina.hanyang.ac.kr:8080
/project/books.rdf#machineLearning>
USING NAMESPACE dc=<http://purl.org/dc/elements/1.1/>
```

3.3 시맨틱 추론검색

(단계1) 온톨로지 기반의 직접매칭 방법은 사용자의 질의 문과 메타데이터의 속성이 정확하게 매칭 되는 부분을 검색하는 방법이다.

(단계2) 온톨로지 기반의 추론검색 방법은 메타데이터에서 온톨로지의 계층구조와 공리(axiom)로 표현된 부분을 규칙기반의 추론을 통하여 사용자의 질의문을 확장시켜서 검색하는 방법이다. 제안된 두 가지 검색방법은 시간적인 효율성을 고려 시 비교적 시간이 더 걸리는 단계2의 추론 검색 방법을 매번 사용하는 것이 아니라 정확하게 매칭이 될 때는 '온톨로지 기반의 직접매칭 방법'을 사용하고, 정확하게 매칭이 안 될 때는 '온톨로지 기반의 추론검색 방법'으로 질의를 확장하여 검색함으로써 검색시스템의 효율성을 높일 수 있다.

3.4 시맨틱 웹 서비스

무선인터넷의 발달로 휴대폰이나 PDA등의 모바일 장치들의 사용이 급증하고 있으므로 웹의 다양한 정보 중 사용자가 원하는 정보만 필터링을 거쳐 모바일장치에서도 이용할 수 있도록 한다.

4. 구현 및 결과

4.1 질의문

그림 2 에서 웹 브라우저를 이용하여, 구현된 시스템으로 2.5에서 사용된 예제 질의를 전송하는 모습을 보여준다.

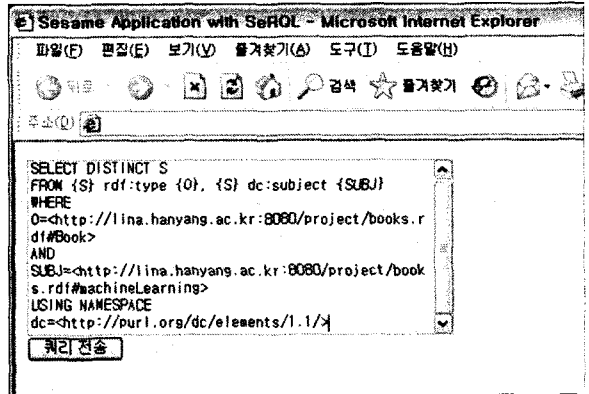
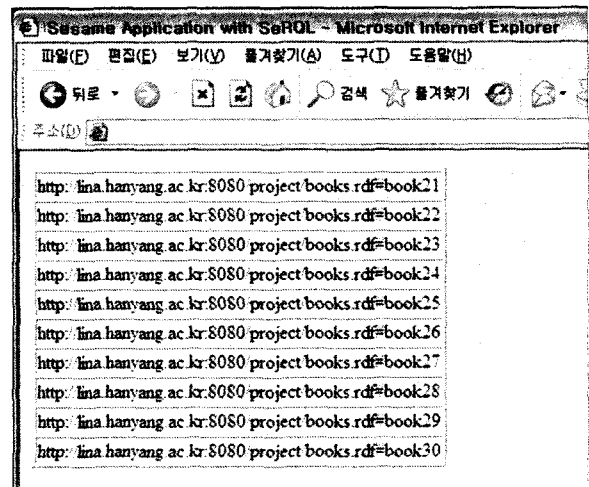


그림 2 질의 입력

4.2 질의결과

질의 결과도 또한 웹으로 확인할 수 있고, Triple의 목록 형태로 보여지며, 서로 다른 인스턴스에 대한 기술을 나타내고 있다.



5. 결론 및 향후 연구

시맨틱 웹의 핵심 기술인 온톨로지 기술을 이용하여 온톨로지 기반 시맨틱 검색 시스템과 시맨틱 웹 서비스 시스템에 대한 구현을 위하여 문헌 정보를 대상으로 온톨로지를 OWL로 구성하였다. 여기에서 제안한 시스템은 효율적인 관리가 가능하도록 웹(Web) 환경에서 구현하였다. 온톨로지를 이용한 시맨틱 검색을 수행하며, 지능형 웹 서비스를 통한 지식의 공유가 가능하다. 또한 구현된 시스템으로 예제 질의를 전송하면 질의 결과를 Triple의 목록 형태로 보여주며, 웹으로 확인할 수도 있고 서로 다른 인스턴스에 대한 기술을 나타낸다.

향후 연구에서는 현재 온톨로지의 모델링과 합병 및 변환 등의 연구가 진행될 것이며, 지능 에이전트의 기능을 향상시키고 더욱 개선된 시맨틱 추론검색시스템 구축에 연구가 계속될 것이다.

참고문헌

- [1] R. Guha, Rob McCool, Eric Miller, "Semantic Search," www2003, May 20-24, 2003.
- [2] Nicholas Gibbins, Stephen Harris, Nigel Shadbolt, "Agent-based Semantic Web Services," www2003, May 20-24, 2003.
- [3] Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," Scientific American, vol., no. 5, pp.34-43, 2001
- [4] James Hendler, "Agents and the Semantic Web," IEEE Intelligent Systems, 16(2), pp.30-37, 2001.
- [5] Jeff Heflin, James Hendler, "Semantic Interoperability on the Web," In Proceedings of Extreme
- [6] Brian McBride, "Jena: Implementing the RDF Model and Syntax Specification," In Proceedings of the Second International Workshop on the Semantic Web-SemWeb '2001, Hong Kong, China, May 1, 2001.
- [7] W3C Recommendation, "OWL Web Ontology Language Overview", <http://www.w3.org/TR/2004/REC-owl-features-20040210>
- [8] Dieter Fensel, "Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce," Springer-Verlag, 2001.
- [9] Eric Newcomer, "Understanding Web Services: XML, WSDL, SOAP, and UDDI.," Addison-Wesley, 2002.
- [10] R. Guha and R. McCool. The tap knowledge base, <http://tap.stanford.edu>
- [11] G. Karvounarakis, S. Alexaki, V. Christophides, D. Plexousakis, and M. Scholl. Rql: a declarative query language for rdf. In WWW, pages 592-603, 2002.
- [12] W3C Recommendation, "RDF Vocabulary Description Language 1.0: RDF-Schema," Dan Brickley and R. V. Guha, Editors, W3C Recommendation, 10 February 2004, <http://www.w3.org/TR/2004/REC-rdf-schema-20040210/>
- [13] Hewlett-Packard Lab., "RDQL-RDF Data Query Language," <http://www.hpl.hp.com/semweb/rdql.htm>
- [14] W3C, "Extensible Markup Language," <http://www.w3.org/XML>
- [15] W3C, "Resource Description Framework," <http://www.w3.org/RDF>