

정보를 표현하는 기법으로서의 RDF와 토픽맵(Topic maps)과의 비교

A Comparison on RDF and Topic Maps,
as the Standards for Representing Information

이혜원, 연세대학교 문헌정보학과, hwlee@yonsei.ac.kr

Lee, Hye-Won, Dept of Library & Information Science, Yonsei University

< 초 록 >

효율적이고 체계적인 정보관리를 위해 최근 연구들은 시맨틱웹(semantic web), 지식관리, 메타데이터의 통합 등에 많은 관심을 두고 있다. 그러한 연구들은 자원(resources)의 기술을 어떻게 표현할 것인가에 대한 기술구조와 그 구조를 표현하기 위한 기계 언어 등을 다루고 있다. 특히 자원의 기술을 어떻게 표현할 것인가에 대한 기술적인 구조로 가장 널리 사용되는 것은 RDF와 토픽맵(Topic Maps)을 들 수 있다. 정보조직이나 시맨틱웹 등의 연구에서 자주 등장하는 위의 개념들을 정확하게 이해하고 무엇보다 그 개념들 간의 관계를 알아보는 것이 중요할 것이다. 본 연구에서는 RDF와 토픽맵에서, 정보 즉 표현하고자 하는 대상을 표현하는 방법을 살펴보고, 두 기법간의 상호운용성에 대한 선행연구로 RDF와 토픽맵의 유사점과 차이점을 비교하고자 한다.

1. 서론

시간의 흐름에 따라 정보의 규모가 양적으로 질적으로 증가하는 정보사회에서는 그 정보자체를 보다 체계적으로 수집·축적하여, 효율적인 관리와 사용을 유도하는 활용성의 측면에 관심을 두고 있다. 그러한 효율성이나 체계적인 정보관리를 위해 최근 연구들은 시맨틱웹(semantic web), 지식관리, 메타데이터의 통합 등의 커다란 연구주제를 가지고 그 안에서 다양한 기법을 소개하고 있다. 다양한 기법 중에는 자원의 기술을 어떻게 표현할 것인가에 대한 기술 구조와 그 구조를 표현하기 위한 기술

적인 언어 등이 있다. 특히 자원의 기술을 어떻게 표현할 것인가에 대한 기술적인 구조로 가장 널리 사용되는 것은 자원기술구조(Resource Description Framework: 이하 RDF라 함)과 토픽맵(Topic Maps)을 들 수 있다. 그리고 자원을 표현하는 언어로는 DAML(DARPA Agent Markup Language), OIL(Ontology Inference Layer)등이 있으며, 최근 월드 와이드 웹(World Wide Web: 이하 W3C라 함)에서는 이 두 언어를 합쳐서 DAML+OIL이라는 최고 수준의 시맨틱웹 구축 언어를 개발하고 있다.

정보조직이나 시맨틱웹 등의 연구에서 자주

등장하는 위의 개념들을 정확하게 이해하고, 무엇보다 그 개념들 간의 관계를 알아보는 것이 중요할 것이다. 특히 RDF와 토픽맵을 이용하여 효율적인 정보조직체계를 갖추기 위한 이론적인 연구(Garshol 2005, Manola & Miller 2003)가 진행되고 있으며, 더 나아가 이 두 기법간의 상호운용성(interoperability)에 대한 연구가 국내외적으로 관심을 모으고 있다. 하지만 두 기법의 특성 및 표현하고자 하는 대상을 기술하는 방법 등을 비교하거나, 그 기법들 간의 상호운용성에 대한 국내연구가 미진한 상태이다.

본 연구에서는 RDF와 토픽맵에서, 정보 즉 표현하고자 하는 대상을 표현하는 방법을 살펴보고, 두 기법간의 상호운용성에 대한 선행연구로 RDF와 토픽맵의 유사점과 차이점을 비교하고자 한다.

2. 정보를 표현하는 기법으로서의 RDF와 토픽맵

2.1 RDF

RDF는 W3C에서 자원에 관한 정보를 표현하기 위한 기법으로 제안된 것이다. RDF는 특히 웹 페이지의 표제와 저자, 갱신일자를 포함하여 웹 문서의 저작권과 계약정보, 또는 일부 공유자원에 대한 접근가능일자 등 웹 자원에 관한 메타데이터를 표현하기 위한 것이다. 그

러나 “웹 자원”이라는 개념을 일반화함으로써, 비록 웹에서 직접 검색되지는 않더라도 웹에서 식별되는 사물에 관한 정보를 표현하기 위해서 RDF를 사용할 수 있다. RDF는 이러한 정보를 사람들에게 단지 제시하기 위한 것이 아니라 어플리케이션으로 처리하기 위한 것이다. RDF는 의미의 손상 없이 이러한 정보를 어플리케이션 상호간에 교환할 수 있도록 표현하기 위한 공통의 틀을 제공한다.

RDF는 웹 식별기호(Uniform Resource Identifiers: 이하 URI라고 함)를 사용하여 사물을 식별하고, 아울러 간단한 속성(properties)과 속성 값(value)으로 자원을 기술한다는 생각에 토대를 두고 있다. 이를 통해서 RDF에서는 자원에 관한 간단한 선언문을, 그 자원을 표현한 노드와 호(아크), 그리고 이들 자원의 속성과 값으로 구성된 그래프로 표현할 수 있다 (Manola & Miller 2003).

- 주어 노드
- 목적어 노드
- 주어 노드에서 목적어 노드로 향한 술어 호(아크)

RDF는 자원에 관한 단순한 선언문을 표현한다는 생각에 기초하고 있으며, 각 선언문은 주어(subject; 자원 resource)와 술어(predicate; 속성 property), 목적어(object; Literal)로 구성되어 있다.

[그림 1]은 영어로 작성된 문장을 RDF 선언문으로 표현해 보면 [그림 1]과 같다.

`http://www.example.org/index.html` has a creator whose value is John Smith
 문장을 RDF 선언문으로 표현하면:
 주어 `http://www.example.org/index.html`
 술어 `http://purl.org/dc/elements/1.1/creator`
 목적어 `http://www.example.org/staffid/85740`

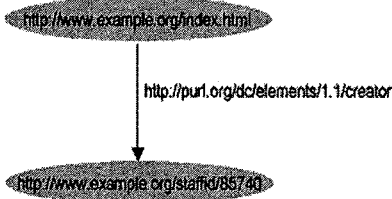
[그림 1] 단순한 RDF 선언문

`http://www.example.org/index.html` has a `creation-date` whose value is `August 16, 1999`
`http://www.example.org/index.html` has a `language` whose value is `English`

[그림 3] 추가된 구문들

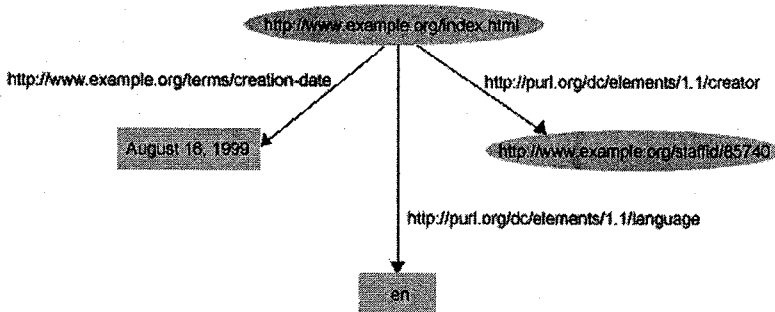
"creator"와 "John Smith"라는 단어를 사용하는 대신, 처음 선언문의 주어뿐만 아니라 술어와 목적어를 식별하기 위하여 각각 URI를 사용하였다.

위의 RDF 선언문을 노드와 호(아크)로 모델화하면 [그림 2]과 같다.



[그림 2] 단순한 RDF 그래프

[그림 2]처럼 모든 선언문은 상응되는 노드와 호로 표현된다. 따라서 [그림 3]과 같은 구문들을 RDF 그래프로 추가로 표현하는 경우에는 [그림 4]의 그래프를 사용할 수 있다(적절한 URI를 사용하여 "creation-date"와 "language" 속성을 지정할 수 있다).



[그림 4] 동일 자원에 대한 복수의 선언문

아울러 RDF에서는 이러한 그래프를 기록하고 교환하기 위하여 XML에 기반한 구문도 제공하고 있다.

2.2 토픽맵(Topic Maps)의 TAO

토픽맵은 분산 환경 하에서 지식 구조를 정의하고 정의된 구조와 지식 자원을 연계하는데 쓰이는 기술 표준이라 할 수 있다. 토픽맵의 표준으로는 SGML(Standard Generalized Markup Language)에 기반을 둔 ISO/IEC 13250:2000이 있으며, 웹 환경에서의 사용을 위해 XML 구문체계를 이용한 XML Topic Maps 1.0(2001)이 있다.

2.2.1 토픽(Topic)의 T

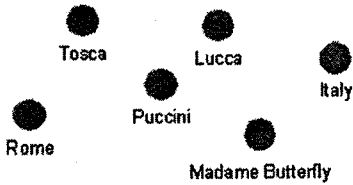
토픽맵에서 토픽은 가장 일반적이면서도 포괄적으로 정의되어 있다. 즉 토픽맵에서 지원하는 토픽은 사람, 엔티티(entity), 개념 등이며 이들의 존재 유무에 상관이 없다. 또한 그 토픽들에 대한 특징도 존재 유무에도 상관이 없으며 특징에 대한 어떠한 주장도 가능하다 결

국 어떠한 것도 토픽으로의 전환이 가능하다는 것이다.

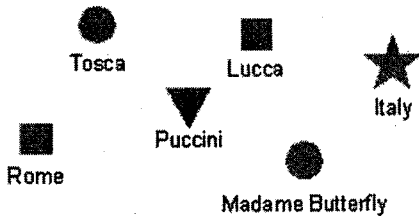
토픽맵은 어떤 단어(word)를 가지고 특정한 주제(subject)를 어떻게 정의하는 것에 대한 기준이다. 그 정의 또한 단어를 가지고 표현하므로, 실제 사물(thing)을 표현하는데 쓰는 용어(term)를 가지고 토픽을 결정하게 된다(Pepper 2002).

토픽의 종류(Topic kinds)는 나타나지 않을 수도 있으며, 하나 이상의 다양한 토픽이 제공될 수 있다. 토픽맵에서는 도서의 색인처럼 이름이나 작품, 장소 등 기준으로 그 토픽의 종류를 구분할 수 있다.

Dictionary of Opera에서는 "Tosca", "Madame Butterfly", "Rome", "Italy" 등의 주제를 토픽으로 표현하고 있을 뿐만 아니라, 작곡가인 "Giacomo Puccini" 그리고 그의 출생지 "Lucca" 또한 사전의 엔트리로 사용되고 있다.



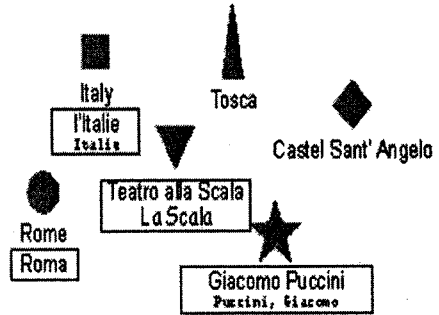
위의 여러 가지 토픽들을 그 종류에 따라서 작곡자(Giacomo Puccini), 오페라(Tosca, Madame Butterfly, Rome, Italy), 나라(Italy), 도시(Rome, Lucca) 등으로 나누어 볼 수 있다.



Topic kinds

그리고 토픽을 표현하는 이름(Topic name)

에도 하나의 토픽에 대하여 사용하는 이용 환경에 따라 여러 가지 방식으로 제공될 수 있다.

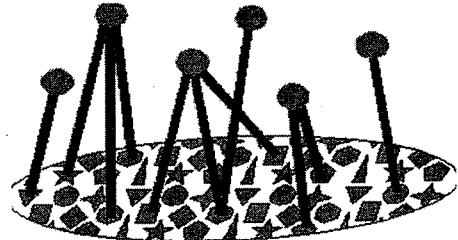


Topic names

2.2.2 Occurrence(어커런스)의 O

토픽맵에서는 하나의 토픽을 설명하기 위해 하나 또는 하나 이상의 정보자원을 연결함으로써 의미적으로 관련된 것들을 모아준다. 그러한 관련된 정보원들을 그 토픽에 대한 어커런스라 한다(Pepper 2002).

그러므로 정보자원과 토픽 간에는 1:M의 관계가 성립할 수 있지만, N:M의 관계는 성립되지 않는다.



Occurrences

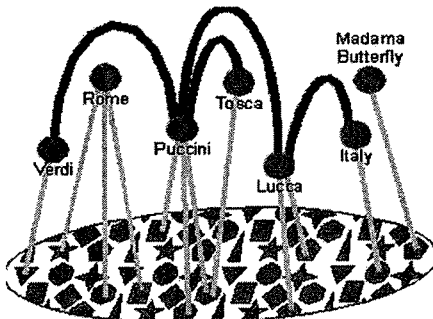
2.2.3 Association(관계)의 A

관계는 토픽맵 안에 정의된 토픽들 간의 관

계를 설정하는데 쓰인다(Pepper 2002). 토픽 간의 관계를 정의할 수 있다는 것은 특정한 토픽맵을 구성하는 토픽 각각의 문맥(context) 정보를 제공하게 되는 것이다. 그러므로 토픽간의 관계를 기술한다는 것은 표현하고자 하는 대상으로 보다 정확하게 정의하며, 효율적인 지식 관리가 이루어질 수 있는 환경을 제공하게 된다.

Dictionary of Opera에서 제시한 토픽들을 정리해 보면 아래와 같은 문장들로 표현되어질 수 있다.

- "Tosca was written by Puccini"
- "Tosca takes place in Rome"
- "Puccini was born in Lucca"
- "Lucca is in Italy"
- "Puccini was influenced by Verdi"



Topic associations

3. RDF와 토픽맵과의 비교

정보를 표현하는 기법으로서의 RDF와 토픽맵과의 비교 연구에서는 표현하고자 하는 대상 즉 사물(thing), 그 사물의 성질(properties), 타입(type), 이름(names), 종류(kinds) 등을 어떻게 표현했는가를 먼저 살펴보아야 한다(Garshol 2001).

3.1 표현하고자 하는 대상 즉 사물(things)

RDF나 토픽맵의 중심적 개념은 사물 그 자체이다. 그러므로 그 사물을 어떻게 표현하고 정의할 것인가에 초점을 맞추고 있는 것에 대해서는 공통점이라고 할 수 있겠다. 그 사물에 대해서 RDF는 자원(resources)라는 말로, 토픽맵에서는 토픽(topics)이라고 부르는 것이다.

RDF는 각각의 자원에 대해 상징적인 구분자인 URI로 표현하였으며, 토픽맵에서의 토픽은 표현하고자 하는 대상 즉 사물을 실제 정보환경에서 어떤 주제로 정의되는지에 대해 중점을 두고 있다(Garshol2001).

Garshol(2001)가 쓴 *Topic Map, RDF, DAML, OIL: a comparison*

<
<http://www.ontopia.net/topicmaps/materials/tmrdfoildaml.html>>이라는 논문을 예로, 사물에 대해 그 표현을 정리해 보면 [표 1]과 같다.

3.2 관계(Relationship)

RDF나 토픽맵에서 표현되는 관계에 대한 차이점은 크게 세 가지로 나뉠 수 있다(Garshol 2001).

첫 번째로, RDF는 하나의 사물과 다른 사물과의 관계를 설명하지만, 토픽맵에서는 여러 개의 사물들을 여러 번 연결시킴으로써 각 관계에 대해 좀더 정확한 정의를 보여주고 있다. 토픽맵의 관계를 RDF와 유사해 보이기도 하지만 좀더 많은 시간과 노력이 요구되어 진다. 각 기법에서 표현되어 지는 관계를 [그림 5.6]에서 보여준다.

두 번째로는 토픽맵의 관계는 양방향으로 향하게 된다(two-way). 이는 Garshol가 Ontology에 소속되어 있다고는 할 수 있지만 그 동시에 Ontology가 Garshol를 채용하고 있

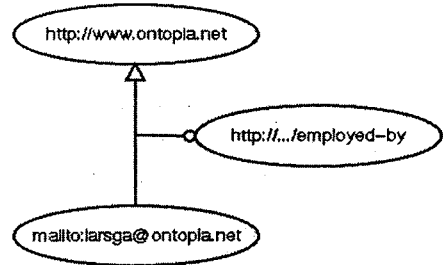
다는 사실을 알려줄 수는 없다. 반면 RDF는

Thing	RDF	Topic Maps
Ontopia	http://www.ontopia.net	<i>subject indicator</i> = http://www.ontopia.net
Lars Marius Garshol	mailto:larsga@ontopia.net	<i>subject indicator</i> = mailto:larsga@ontopia.net
This paper	http://www.ontopia.net/ topicmaps/materials/ tmrfdamlail.html	<i>subject address</i> = http://www.ontopia.net/ topicmaps/materials/ tmrfdamlail.html

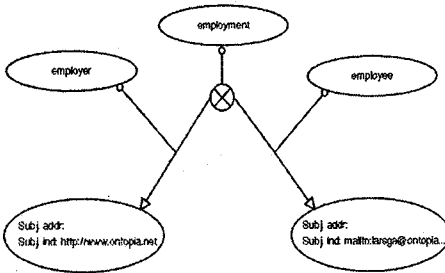
[표 1] Three things

사물들의 관계에 있어서 아크의 방향을 역으로 추적하여 반대의 속성을 확인시킬 수 있다[그림 5,6]에서 확인할 수 있음).

세 번째는 아주 미묘한 차이라고 할 수 있다. 만약 RDF 선언문이 추상적인 사물들의 관계를 정의하고 있다면 그 차이를 정확하게 찾아낼 수 없다. 사물의 속성들을 할당할 때 정확한 의미 전달을 위해서 어떠한 경우에는 URI가 아닌 그 객체를 설명하는 실제 표현(literals)이 필요하기 때문이다.



[그림 6] RDF의 관계 표현

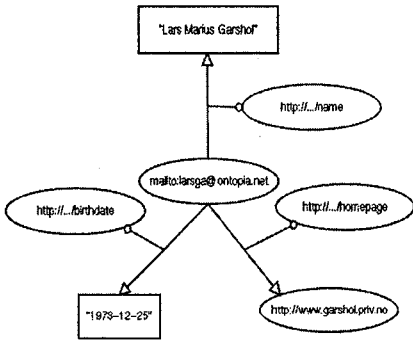


[그림 5] 토픽맵의 관계 표현

3.3 속성(Attributes)

속성은 사물을 표현하기 위한 정보의 단편들을 말하는 것이며, 그 사물에 대한 속성들로 인해 사물의 정의가 좀더 정확하게 제시될 수 있다.

Garshol라는 사물에 대해 그 속성들을 좀 더 정확하게 표현하면 [그림 7]과 같다. [그림 7]에서는 Garshol의 이름, 출생일자와 홈페이지에 대한 정보를 한꺼번에 제공해 주고 있다.



[그림 7] RDF로 표현한 세가지 속성

하지만 토픽맵에서는 관계에 대한 표현이 아주 다르게 이루어지고 있다. 토픽맵에서는 하나의 토픽과 관련된 정보의 단편들을 어커런스라는 개념을 사용하였다. 어커런스는 구현하고자 하는 토픽맵의 외부에 있는 자원일수도 있고, 또는 그 토픽맵 내부의 자원일 수도 있으며, 마지막으로 그 자원의 URI에 의해 표현될 수도 있다. 어커런스는 다양한 종류에 의해 구분되어지고 그 종류들이 바로 토픽이 되는 것이다.

RDF의 속성을 설명하는 예를 기준으로 토픽맵과의 차이점을 알아보면 다음과 같다. Garshol라는 사물에 대해 그 속성들 중 출생일자와 홈페이지에 관한 정보는 토픽맵에서도 어커런스를 통해 표현되어질 수 있지만, Garshol의 이름은 어커런스가 아닌 이름(names; 2.2.1에서 언급)을 통해 제시되어져야 한다.

속성에 대한 RDF와 토픽맵과의 차이점을 세 가지로 찾아볼 수 있다(Garshol 2001).

첫 번째로는, 이름(names)은 RDF와 토픽맵 두 기법 다 표현할 수 있다. 하지만 만약 표현하고자 하는 속성의 이름을 알아낼 수 없는 스키마를 가진 소프트웨어에서는 토픽맵 만을 사용할 수 있을 것이다.

두 번째로는, 간단한 속성(simple properties)

은 토픽맵과 RDF에서 유사하게 사용되어진다. 문자열(string)은 특정한 사물을 표현하는 것이고 다른 사물을 통해 문자열과 그 사물과의 관계가 무엇인지 알 수 있다.

마지막으로, RDF의 관계나 선언문에서는 사물과 주제적으로 관련된 자원(resources relevant to a thing)들을 구별할 수가 없다. 하지만 토픽맵에서는 어커런스의 종류를 이용하여 정보를 의미적으로 좀 더 명확하게 구별한다.

4. 결론

RDF와 토픽맵이라는 기법에 대해 각각 살펴보고, 그 차이점을 살펴보았다. RDF와 토픽맵은 표현하고자 하는 대상 즉 사물(things)에 대해 중점을 두는 개념이라는 출발점은 같다. 하지만 그 사물을 표현하는 방법은 달랐다.

그 사물에 대해서 RDF는 자원(resources)라는 말로, 토픽맵에서는 토픽(topics)이라고 지칭하였다. 또한 RDF는 자원을 표현하는 속성들의 관계가 토픽맵에 비해 간단하면서 아크의 방향의 유연성을 강조하였다. 하지만 토픽맵에서는 좀더 실제 정보환경에 맞는 표현(literals)을 제시하므로 RDF의 URI로 표현할 수 없는 자원들 간의 관계를 좀 더 정확하게 정의할 수 있었다.

마지막으로 사물에 대한 속성들에 대한 비교에서는 RDF에서는 이름을 하나의 속성으로 사용하지만 토픽맵에서는 속성이 아닌 하나의 주된 사물 즉 토픽으로 접근하고 있다. 또한 RDF는 주제적으로 적합한 사물들을 연결하는 데는 부족하지만 토픽맵에서는 자원에 대한 다양한 어커런스의 종류를 이용하여 좀 더 명확한 의미를 파악할 수 있다.

앞으로의 후속 연구로, 두 기법 간의 상호운용성에 대한 구체적인 방안을 살펴봐야 할 것

이다. W3C에서도 그러한 연구가 활발하게 진행되고 있는데, RDF를 기준으로 할 것인가, 아니면 토픽맵을 기준으로 할 것인가에 대한 다양한 접근이 이루어지고 있다.

6.20]<<http://www.ontopia.net/topicmaps/materials/tao.html>>

참고문헌

한국전자거래진흥원. 2003. *Topic Maps 응용 표준 및 활용 가이드라인 개발*.

Garshol, L. M. 2001. *Topic Map, RDF, DAML, OIL. a comparison*. Ontopia. [cited 2005. 6.22] <<http://www.ontopia.net/topicmaps/materials/tmrdfoidaml.html>>

Garshol, L. M. 2005. *The Linear Topic Map Notation: definition and introduction, version 1.3*. Ontopia. [cited 2005. 6.22] <<http://www.ontopia.net/download/ltn.html>>

Manola, F., Miller, E. 2003. *RDF Primer* [cited 2005. 6.22]<<http://www.w3.org/TR/2003/PR-rdf-primer-20031215>>

Pepper, S. 2002. *Ten Theses on Topic Maps and RDF*. Ontopia. [cited 2005. 6.20] <<http://www.ontopia.net/topicmaps/materials/rdf.html>>

Pepper, S. 2002. *The TAO of Topic Maps: findin the way in the age of infoglut*. Ontopia. [cited 2005.