

RDF 메타데이터를 이용한 응용프로그램의 검색

심현진, 김기태, 조선문, 고훈준*, 유원희
인하대학교 컴퓨터 정보공학과
경인여자대학 컴퓨터정보기술학부
e-mail:g2022035@inhavision.inha.ac.kr

Searching Application Program by Using RDF Metadata

Hyun-Jin Sim, Ki-Tae Kim, Sun-Moon Jo,
Hoon-Joon Kouh*, Weon-Hee Yoo
Dept. of Computer Science & Engineering, Inha University
*School of Computer Information Technology,
Kyungin Women's College

요 약

인터넷에 있는 많은 응용프로그램들의 용도와 부가적인 정보를 아는 것은 매우 어렵다. 웹에는 수 많은 정보들이 존재하고 이것들을 필요에 따라 정확하게 검색하는 것이 점점 어려워지고 있다. 이러한 어려움을 해결하기 위해 현재 메타데이터를 사용하여 웹 자원을 검색하는 연구가 활발하게 진행되고 있다. W3C는 메타데이터의 상호운용성과 통합을 위한 환경으로 RDF(Resource Description Framework)를 표준으로 정했다. 본 논문에서는 응용프로그램이나 프로그램 개발에 필요한 작은 단위의 프로그램에 대해 RDF를 이용하여 메타데이터로 만들고 이것을 대상으로 검색 가능하도록 하여 사용자가 원하는 정보를 보다 정확한 검색을 수행할 수 있도록 하였다.

1. 서론

웹은 간단한 HTML을 이용하여 편리성을 추구한 덕분에 일반 사용자 누구나 쉽게 정보를 접근하거나 게시할 수 있게 되었고, 결과적으로 폭발적인 정보의 증가를 가져왔다. 웹에는 현재 수많은 기관과 단체, 개인들이 서로 다른 목적으로 생성한 많은 문서와 정보가 포함되어 있다. 사용자들은 이러한 정보를 이용하기 위해 브라우저에 URL 주소를 직접 입력하거나 하이퍼링크를 따라가기도 하고, 검색 엔진의 도움을 받기도 한다. 이러한 단순성이 현재 웹의 성장을 가져온 중요한 작용을 했지만 이런 단순성이 웹의 정보가 감당할 수 없을 정도로 방대한 현 상황에서는 문제점으로 작용하기도 한다.

현재의 웹 검색엔진은 사용자의 질의와는 관계없는 많은 문서를 결과로 가져올 수 있고 이로 인해

사용자는 필요한 정보를 찾아내는데 많은 어려움이 있다. 또한 HTML로 여러 관련 문서를 확장하거나 통합, 공유하는 것은 매우 어렵다. 시맨틱 웹은 메타데이터의 개념을 통하여 웹 문서에 의미 정보를 더하여 소프트웨어 에이전트가 이 의미 정보를 추출할 수 있도록 하는 것이다. 또한 의미 정보의 추출뿐 아니라 정보의 확장이나 공유 등도 가능하게 된다. 시맨틱 웹은 기존의 웹과 완전히 구별되는 새로운 개념이 아니라 현재 웹을 확장하여 웹에 있는 정보에 잘 정의된 의미를 부여하고 이를 통해 컴퓨터와 사람이 협동적으로 작업을 수행할 수 있도록 하는 것이다. 따라서 기존의 웹 데이터에 의미를 부여할 수 있는 표현 방법이 필요하다[1].

본 논문에서는 프로그램 개발자들이 제작하여 인터넷에 게시된 응용프로그램에 메타데이터를 추가해

서 응용프로그램을 찾는 다른 사람들에게 각각의 요구에 더 적합한 프로그램을 검색하는데 효율적인 작업이 되도록 RDF 메타데이터를 사용할 것을 제안한다. 응용프로그램에 대한 요약정보를 RDF로 기술한 이 메타데이터를 프로그램과 같이 저장하고 메타데이터를 검색하여 필요한 응용프로그램을 찾는 사용자에게 직관적인 형태로 결과를 보여주는 시스템을 설계하였다.

2. 관련 연구

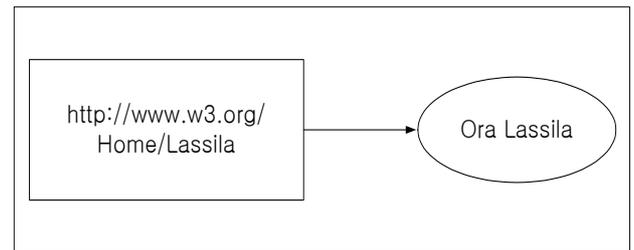
시맨틱 웹의 궁극적인 목적은 웹에 있는 정보를 컴퓨터가 좀 더 이해할 수 있도록 도와주는 표준과 기술을 개발하여 시맨틱 검색, 데이터 통합, 작업의 자동화 등을 지원하는 것이다. 세부적으로는 정보를 검색할 때 더욱 정확한 결과를 가져오고 서로 다른 이질적인 정보를 통합하고 비교한다. 어떤 리소스에 대해서도 의미적이고 기술적인 정보를 연관시킨다. 웹 서비스의 자동화를 위해 데이터에 정보를 추가할 수 있다.

메타데이터(Metadata)는 데이터에 관한 데이터이다. 전달하려고 하는 1차 자료에 접근하기 위한 2차 자료로써 도서관의 도서목록이나 지도의 범례 등도 일종의 메타데이터라고 할 수 있다. 메타데이터를 이용해서 웹에 있는 방대한 자원들을 종류와 특성에 따라 간단하게 기술하며 원하는 자원에 쉽게 접근할 수 있고 정확한 검색에 유용하게 쓰일 수 있다. 메타데이터의 핵심은 간결성이다. 현재 대표적인 메타데이터로는 더블린 코어(Dublin Core)가 있다[2]. 네트워크에 있는 자원의 기술에 필요한 요소를 정하고 쉽고 정확한 검색을 목적으로 1995년에 합의된 메타데이터이다. 더블린 코어는 15개의 핵심 기술요소로 이루어져 있다. 이들 기술 요소들은 자원의 내용, 다른 자원과의 관계, 자원의 속성, 형식, 시간을 나타낸다. 더블린코어를 기초로 다른 분야에서도 확장하여 사용하는 것이 가능하다. 메타데이터의 목적은 정보검색의 처리과정을 줄여주고 원하지 않는 데이터를 걸러주어 관련성이 있는 정보의 발견 가능성을 증가시키고 검색능력을 향상시키는 데에 있다.

RDF는 메타데이터를 처리하기 위한 기초로써 현재 W3C의 표준으로 정해져 있다[3]. 기계가 이해할 수 있는 정보를 교환하는 응용프로그램 간에 상호운용성을 제공하며 자원의 자동적인 처리를 하는데 편리하다. 자원검색에 있어서는 개선된 검색능력을 제공하고 목록분야에서는 콘텐츠나 콘텐츠 간의

관계를 기술할 수 있다. RDF는 자원의 속성과 값을 XML의 방식으로 표현하고 데이터에 의미를 부여한다[4].

기본적인 데이터 모델은 자원, 속성, 문으로 구성되어 있으며 객체중심이 아닌 속성중심의 구조를 가지고 있다. 자원은 속성을 가지며 웹에서 URI(Uniform Resource Identifiers)로 식별되는 모든 객체를 가리킨다. 웹 페이지의 전체나 일부가 될 수도 있고 웹 페이지들의 집합을 나타낼 수도 있다. 속성은 자원의 기술을 위해 사용된 특징, 관계를 말한다. 문은 자원과 특성과 값을 말하며 각각 주부, 술부, 객체라고 한다. 객체는 다른 자원일 수도 있고 문자열을 가리키기도 한다.



[그림 1] 간단한 RDF 데이터 모델

[그림 1]은 간단한 RDF 데이터 모델이다. “Ora Lassila는 자원 <http://www.w3.org/Home/Lassila>의 작성자이다.”라는 의미이다.

[표 1] RDF 데이터 모델의 주부, 술부 객체

Subject	http://www.w3.org/Home/Lassila
Predicate	Creator
Object	"Ora Lassila"

[표 1]는 [그림 1]의 RDF 데이터 모델을 주부, 술부, 객체로 나눈 것이며 이것은 각각 자원, 속성, 문에 대응된다.

```
<rdf:RDF>
<rdf:Description
  about="http://www.w3.org/Home/Lassila">
<s:Creator>Ora Lassila</s:Creator>
</rdf:Description>
</rdf:RDF>
```

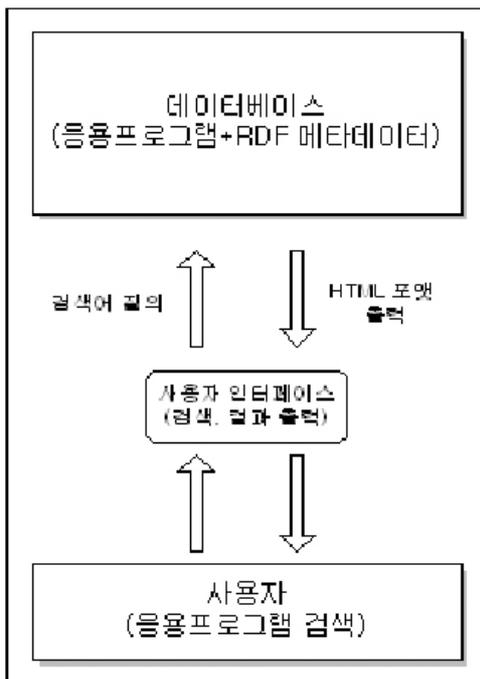
[그림 2] RDF/XML 표현

[그림 2]은 위의 데이터 모델을 RDF/XML 표현으로 나타낸 것이고 이 표현 방법으로 본 논문에서 제안할 시스템에 사용하였다. 웹에 존재하는 응용프로그램들에 대한 각각의 요약정보를 RDF/XML 표

현으로 나타내어 실제 응용프로그램과 함께 저장하여 검색에 이용한다.

3. RDF를 이용한 응용프로그램의 요약

본 논문에서는 응용프로그램의 개발자들이 만든 각각의 프로그램에 대한 요약 정보를 RDF를 이용하여 기술할 것을 제안한다. 어떤 응용프로그램을 검색하려는 사용자는 검색을 통해 발견된 프로그램이 자신이 원하던 프로그램이 맞는지 확인하는 것이 쉽지 않다. 응용프로그램의 이름으로 그 프로그램의 목적이나 기능을 유추해보거나 설치를 하여 확인해보는 방법은 검색을 하는 사용자에게 불편함과 시간의 낭비를 가져온다. 그러나 응용프로그램의 특징들을 간략히 표현한 메타데이터가 있고 그 메타데이터를 검색을 통한 결과로써 적절한 형태로 보여준다면 검색을 하는 사용자는 프로그램의 역할과 기능을 더욱 쉽고 정확하게 알 수 있을 것이다.



[그림 3] 전체 시스템의 개략적인 구조와 동작

[그림 3]는 제안하는 시스템의 개략적인 구조이다. 시스템은 크게 응용프로그램과 메타데이터를 저장한 데이터베이스와 검색 인터페이스로 이루어져 있다. 동작은 메타데이터의 생성, 응용프로그램과 메타데이터의 등록, 검색, 결과 화면 출력의 순서로 진행된다. 프로그램의 개발자는 프로그램의 요약 정보를 RDF로 기술한다. 그리고 RDF로 기술한 이 메타데이터를 데이터베이스에 등록해야 한다. 이 메타데

이터의 내용으로는 개발자, 프로그램의 이름, 프로그램의 수정 또는 갱신에 관한 정보, 관련된 프로그램에 대한 정보를 포함할 수 있다. 이렇게 데이터베이스에 응용프로그램과 RDF 메타데이터가 등록이 되면 필요한 프로그램을 검색하고자 하는 사용자는 검색 인터페이스를 통해 검색작업을 수행할 수 있다. 적절한 검색어를 입력하면 RDF로 표현된 프로그램들의 요약 정보가 검색이 되고 요약 정보에는 실제 프로그램이 저장된 위치와 함께 개발자가 작성한 메타데이터를 웹 브라우저를 통한 HTML 형태로 보여주게 된다. 검색작업을 한 사용자는 이 결과화면을 보고 자신의 요구에 맞는 필요한 프로그램을 더욱 정확히 찾을 수 있다. [그림 4]는 응용프로그램의 요약정보를 나타낸 RDF 메타데이터의 일부분이다.

```
<?XML version="1.0">
<?namespace href="http://www.w3c.org/TR/WD-rdf-syntax#"AS="RDF"?>
<?namespace href="http://www.inhapds.com/"AS="PDS"?>
<RDF:RDF>
  <RDF:Description RDF:Href="ftp://ftp.inhapds.com/pub/notepad/mynotepad.exe">
    <PDS:Name>My Notepad</PDS:Name>
    <PDS:Version>0.90</PDS:Version>
    <PDS:Release>1</PDS:Release>
    <PDS:Vendor>Inha Univ.</PDS:Vendor>
    <PDS:Size>13244</PDS:Size>
    <PDS:URL>http://mynotepad.com</PDS:URL>
    <PDS:BuildDate>Sun Mar 29 19:44:53 EST 1998
  </PDS:BuildDate>
  <PDS:Summary>
    .....
```

[그림 4] 프로그램의 요약정보를 파일의 일부

이렇게 응용프로그램과 그에 대한 메타데이터를 저장한 데이터베이스는 특정영역에 대해 집중시키는 것이 중요하다. 또는 데이터베이스 내에서 일정한 기준으로 분류화를 하는 것도 가능하다. 공통된 환경에서 실행 가능한 특징이나 공통된 언어로 구현된 프로그램 단위들을 한 분류로 만드는 것이 검색성을 높일 수 있을 것이다.

웹에 존재하는 많은 응용프로그램들과 그것들을 필요로 하는 사용자의 요구는 다양해졌다. 응용프로그램들은 각각의 특성도 있기 때문에 그 것을 설명한 메타데이터의 필요는 중요하다.

4. 결론 및 향후 연구

본 논문에서는 개발자들에 의해 만들어진 많은 응용프로그램을 효율적으로 검색하는 방법으로 RDF 메타데이터를 사용하는 방법을 제안하였다. 프로그램에 대한 설명 또는 부가적인 정보들을 실제로 검색의 대상이 될 수 있는 메타데이터로써 만들어 검색하는 사용자가 메타데이터에 나와 있는 정보로부터 자신에게 필요한 프로그램을 찾을 수 있도록 하는 시스템이다.

향후 연구를 통해 개선되어야 할 점으로는 개발자들이 RDF로 표현된 메타데이터들을 작성할 수 있는 생성기를 통해 가능한 한 메타데이터들이 일정한 형식을 갖출 수 있도록 하는 것이다. 또는 메타데이터들을 자동적으로 생성할 수 있는 방법에 대한 연구이다. 또한 응용프로그램과 메타데이터가 물리적으로 여러 장소에 분산될 경우의 관리에 대한 연구이다. 예를 들어 버전 정보가 다른 응용프로그램과 메타데이터가 여러 곳에 분산될 경우에 검색을 하는 사용자에게 혼란을 줄 수 있으므로 이러한 문제를 해결하는 연구도 필요할 것이다.

참고문헌

- [1] Sheila A. McIlraith, Tran Cao Son, Hongglei Zeng, Semantic Web Services, IEEE INTELLIGENT SYSTEMS, 2001
- [2] Diane Hillmann, Using Dublin Core, <http://dublincore.org/documents/usageguide/>, 2003
- [3] Ora Lassila, Ralph R. Swick, Resource Description Framework (RDF) Model and Syntax Specification, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>, 1999
- [4] Michel Klein, XML, RDF, and Relatives, IEEE INTELLIGENT SYSTEMS, 2001