

RDF

시맨틱 웹이 연결하고자 하는 객체에는 월드와이드웹의 디지털 문서뿐 아니라 그 내용의 문맥을 이루는 요소-인물, 장소, 사건, 문헌, 개념 등-까지도 포함된다. 그러한 객체 사이의 관계성을 표시해 주는 방법은 여러 가지가 있을 수 있겠으나, 월드와이드웹의 세계에서는 RDF(Resource Description Framework)라고 불리는 형식이 표준으로서의 지위를 얻고 있다.¹⁾ 웹 주소와 같은 명명법으로 객체에 식별자를 부여하고 두 객체 사이의 관계성을 약속된 서술어로 기술하는 방식이다. ‘A’라는 텍스트와 ‘B’라는 인물 사이에 ‘~의 저자는 ~이다’라는 관계성을 서술어로 부여하는 RDF 문의 모양새는 다음과 같다.



RDF를 사용하여 월드와이드웹 자원 사이의 의미적 연관 관계를 기술하는 방법을 다음의 예시를 통해 알아보기로 하자.

‘구글’이나 ‘네이버’와 같은 월드와이드웹 포털에서 ‘인문정보학’이라는 단어를 검색하면, ‘한국학중앙연구원 인문정보학 교수 김현’(a), ‘인문정보학의 모색’이라는 책(b) 등에 관한 웹 문서가 검색 결과로 제공될 것이다. 이 결과를 가지고 (b)의 저자가 (a)라고 하는 사실을 파악하는 것은 검색 결과를 받아 본 이용자의 몫이다. 시맨틱 웹은 이용자가 이 문서를 읽기 전에 그 관계를 먼저 파악하여 더욱 친절하고 효용성 있는 안내를 해 주고자 한다. 하지만 그것은 컴퓨터가 스스로 알아서 할 수 있는 일이 아니고, 처음에 그 문서를 만든 생산자가 그와 같은 관계 정보를 ‘컴퓨터가 파악할 수 있는 형태로’ 문서 안에 포함시킬 때 비로소 가능해질 수 있는 일이다.

RDF 형식으로 웹 자원 상호간의 관계를 기술하기 위해서는 두 가지 조건이 선결되어야 한다. 첫째는 관계성을 부여하고자 하는 객체가 월드와이드웹 상에

1) RDF(the Resource Description Framework)는 월드와이드웹 자원의 메타데이터를 기술하는 형식이다. 2004년 W3C(World Wide Web Consortium)의 권고안이 제시되었으며, 시맨틱 웹 활동의 일환으로 운용되고 있다. RDF가 기술하는 웹 자원의 속성은 인간이 아닌 컴퓨터가 읽고 해석하기 위한 것이며, 인간을 대신하여 지능적인 정보 처리를 하게 하기 위한 것이다.
<http://www.w3.org/RDF/>

서 유일하게 식별될 수 있는 이름을 가져야 하는 것, 두 번째는 객체와 객체 사이의 관계를 설명하는 관계 서술어의 표준적인 형태를 약속해야 한다는 것이다. 이 예시에서 보이도록 하는 (a)와 (b)의 관계 기술을 위해 이 두 가지 객체의 고유한 이름을 아래와 같이 정하기로 한다.

(a) ‘김현’이라는 사람의 식별자:

<http://digerati.aks.ac.kr/terms/person#김현>

(b) 『인문정보학의 모색』이라는 책의 식별자:

<http://digerati.aks.ac.kr/terms/bibliography#인문정보학의모색>

이 두 객체의 식별자는 월드와이드웹 문서의 주소와 같은 형식으로 부여되었지만 웹상에 이에 대응하는 문서가 존재하는 것은 아니다. 세상의 모든 사물, 심지어는 추상적인 개념까지도 디지털 세계에서 객체로 식별될 필요가 있을 때에는 웹 주소 형태의 식별자(URI 또는 IRI)²⁾를 부여할 수 있다.

그 다음 필요한 것은 이 객체들 사이의 관계를 표현하는 서술어이다. “~는 ~의 저작이다”라는 관계 서술을 다음과 같이 하기로 한다.

[p] 서술어: ~는 ~의 저작이다

<http://purl.org/dc/terms/creator>

RDF 문의 주어와 목적어로 쓰인 두 객체의 식별자와 마찬가지로 서술어도 웹 주소 형태로 기술하였다. 이것 역시 서술어가 의미하는 바가 모호해지지 않도록 하기 위한 장치이다. ‘<http://purl.org/dc/terms/>’는 웹 자원 메타데이터 기술 용어의 표준화를 위한 더블린 코어 스키마(Dublin Core Schema)를 의미한다. 여기서 ‘<http://purl.org/dc/terms/creator>’는 웹상에 물리적으로 존재하는 문서의 주소가 아니라 ‘~는 ~의 저작이다’라는 뜻을 명시적으로 알리는 서술어임을 이해하도록 하자.

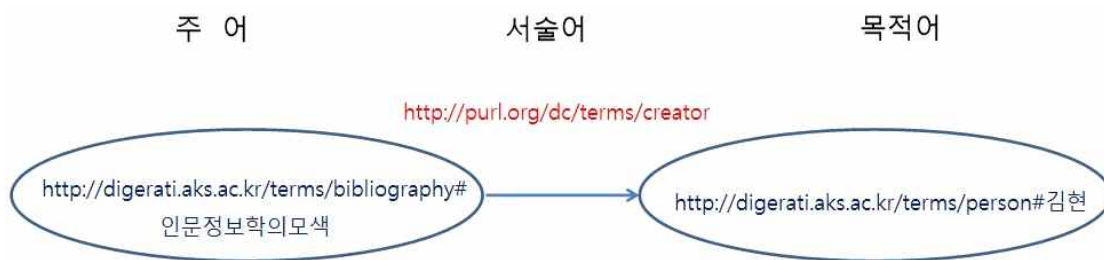
이제 주어, 목적어, 서술어로 쓸 디지털 언어의 어휘들이 정해졌기 때문에 그 어휘들을 영어의 3형식 문장과 같은 RDF 구문 형식에 집어넣으면 RDF 문이 완성된다.

[예시] RDF 문의 구문 형식

2) 이 장의 주 5) 참조.

주어	서술어	목적어
http://digerati.aks.ac.kr/terms/bibliography#인문정보학의모색	http://purl.org/dc/terms/creator	http://digerati.aks.ac.kr/terms/person#김현

[예시] RDF Graph



[예시] RDF/XML³⁾

```

<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/terms/">
  <rdf:Description rdf:about="http://digerati.aks.ac.kr/terms/bibliography#인문정보학의모색">
    <dc:creator rdf:resource="http://digerati.aks.ac.kr/terms/person#김현"/>
  </rdf:Description>
</rdf:RDF>
  
```

월드와이드웹 문서 또는 그 속에 담긴 문맥 요소의 상호 연관 관계를 RDF 문으로 만들고 나면, 그 다음에 어떠한 일이 어떻게 이루어지는지 보기로 하자. 아래에 예시한 3편의 HTML 문서는 웹 브라우저를 통해 볼 때 일반적인 HTML 문서와 다른없는 형태로 표현된다. 그렇지만 이 문서의 내부에는 유관한 것들을 연결시켜 주는 정보가 포함되어 있다.⁴⁾

3) RDF/XML: RDF 문을 XML로 표현하기 위한 W3C(World Wide Web Consortium)의 제안 형식.
<https://www.w3.org/TR/rdf-syntax-grammar/>

4) RDF 문을 HTML로 쓰인 웹 문서 안에 담아내는 형식을 RDFa(Resource Description Framework in Attributes)라고 한다. 이는 다른 시맨틱 웹 관련 기술표준과 마찬가지로 W3C(World Wide Web Consortium)가 제안하는 규약이다.
<https://www.w3.org/TR/xhtml-rdfa-primer/>

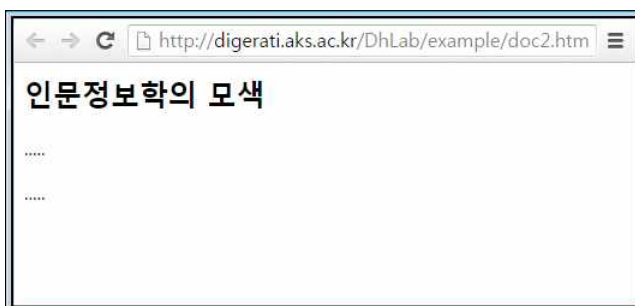
[예시] 문서1: doc1.htm

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:dc="http://purl.org/dc/terms/"
      xmlns:foaf="http://xmlns.com/foaf/0.1/"
      version="XHTML+RDFa 1.0" >
  <head>
    <title>이력서: 한국학중앙연구원 인문정보학 교수 김현</title>
    <link rel="dc:relation" href="http://digerati.aks.ac.kr/terms/person#김현"/>
  </head>
  <body about="http://digerati.aks.ac.kr/terms/person#김현">
    이름: <span property="foaf:name">김현</span><br/>
    홈페이지: <span property="foaf:homepage">www.xuanflute.com</span><br/>
    .....
  </body>
</html>
```



[예시] 문서2: doc2.htm

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:dc="http://purl.org/dc/terms/"
      version="XHTML+RDFa 1.0" >
  <head>
    <title>서평: 인문정보학의 모색</title>
    <link rel="dc:relation" href="http://digerati.aks.ac.kr/terms/bibliography#인문정보학
의모색"/>
  </head>
  <body about="http://digerati.aks.ac.kr/terms/bibliography#인문정보학의모색">
    <h2 property="dc:title">인문정보학의 모색</h2>
    <p property="dc:description"> ..... </p>
    .....
  </body>
</html>
```



[예시] 문서3: doc3.htm

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:dc="http://purl.org/dc/terms/"
      xmlns:foaf="http://xmlns.com/foaf/0.1/"
      version="XHTML+RDFa 1.0">
  <head>
    <title>디지털 인문학 도서 목록</title>
  </head>
  <body>
    .....
    <div about="http://digerati.aks.ac.kr/terms/bibliograhya#인문정보학의모색">
      <span property="dc:creator"
resource="http://digerati.aks.ac.kr/terms/person#김현">김현</span>,
      <span property="dc:title"><cite>인문정보학의 모색</cite></span>,
      <span property="dc:created">2012</span>.
    </div>
    .....
  </body>
</html>
```



이와 같은 문서가 월드와이드웹 상에 개별적으로 존재할 때에는 기존의 웹 문서와 별다른 차이를 보이지 않는다. 하지만 이 문서 속의 RDF 정보가 웹 포털에 의해서 탐색·추출되면 현재와는 다른 데이터 서비스가 가능해진다. 아래의 세 가지 예시는 위의 세 문서 속에 포함된 RDF 문을 기계적으로 추출하여 RDF 그래프 형태로 보인 것이다.

[예시] 문서1에 포함된 관계 정보

관계 정보	문서1 -[관계가 있다]→ #김현 -[이 사람의 속성]→ 이름, 홈페이지 주소
RDF GRAPH	<pre> graph LR A([http://digerati.aks.ac.kr/dhlab/example/doc1.htm]) -- http://purl.org/dc/terms/relation --> B([http://digerati.aks.ac.kr/terms/person#김현]) B -- http://xmlns.com/foaf/0.1/name --> C[김현] B -- http://xmlns.com/foaf/0.1/homepage --> D[www.xuanflute.com] </pre>
RDF/XML	<pre> <?xml version="1.0" encoding="utf-8" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/terms/" xmlns:foaf="http://xmlns.com/foaf/0.1/"> <rdf:Description rdf:about="http://digerati.aks.ac.kr/dhlab/example/doc1.htm"> <dc:relation> <rdf:Description rdf:about="http://digerati.aks.ac.kr/terms/person#김현"> <foaf:name>김현</foaf:name> <foaf:homepage>www.xuanflute.com</foaf:homepage> </rdf:Description> </dc:relation> </rdf:Description> </rdf:RDF> </pre>

[예시] 문서2에 포함된 관계 정보

관계 정보	문서2 -[관계가 있다]→ #인문정보학의모색 -[이 책의 속성]→ 저자 이름, 해제
RDF GRAPH	<pre> graph LR A([http://digerati.aks.ac.kr/dhlab/example/doc2.htm]) -- http://purl.org/dc/terms/relation --> B([http://digerati.aks.ac.kr/terms/bibliography#인문정보학의모색]) B -- http://purl.org/dc/terms/title --> C[김현] B -- http://purl.org/dc/terms/description --> D[.....] </pre>
RDF/XML	<pre> <?xml version="1.0" encoding="utf-8" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/terms/"> <rdf:Description rdf:about="http://digerati.aks.ac.kr/dhlab/example/doc2.htm"> <dc:relation> <rdf:Description rdf:about="http://digerati.aks.ac.kr/terms/bibliography#인문정보학의모색"> <dc:title>인문정보학의 모색</dc:title> <dc:description> </dc:description> </rdf:Description> </dc:relation> </rdf:Description> </rdf:RDF> </pre>

[예시] 문서3에 포함된 관계 정보

관계 정보	#인문정보학의모색 -[~는 ~의 저작이다]→ #김현
RDF GRAPH	<pre> graph LR A([http://digerati.aks.ac.kr/terms/bibliography#인문정보학의모색]) -- "http://purl.org/dc/terms/creator" --> B([http://digerati.aks.ac.kr/terms/person#김현]) A -- "http://purl.org/dc/terms/title" --> C[인문정보학의 모색] A -- "http://purl.org/dc/terms/created" --> D[2012] </pre>
RDF/XML	<pre> <?xml version="1.0" encoding="utf-8" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/terms/"> <rdf:Description rdf:about="http://digerati.aks.ac.kr/terms/bibliography#인문정보학의모색"> <dc:creator rdf:resource="http://digerati.aks.ac.kr/terms/person#김현"> <dc:title>인문정보학의 모색</dc:title> <dc:created>2012</dc:created> </rdf:Description> </rdf:RDF> </pre>

이러한 정보는 각각 별개의 웹 문서에서부터 추출된 것이지만, 그것들은 서로 연결될 수 있는 명확한 근거를 가지고 있다. ‘#김현’이라는 인물과 ‘#인문정보학의모색’이라는 책이 고유한 식별자로 표시되었고, 개체와 개체 사이의 관계, 개체와 속성의 관계가 누구나(컴퓨터를 포함하여) 인식할 수 있는 표준적인 서술어로 기술되었기 때문이다. 더 많은 사람, 더 많은 책, 그리고 장소와 사건, 개념 용어까지도 이러한 방식으로 연결되면 월드와이드웹은 ‘단어를 찾아 주는’ 수준에서 더 진화하여 ‘지식을 엮어 주는’ 역할까지 할 수 있게 되리라는 것이 시맨틱 웹의 이상이다.

[예시] 식별자를 매개로 연결된 문서1, 문서2, 문서3의 정보

