

IV. 인문지식 시맨틱 웹

월드와이드웹 상에서 유통되는 디지털 텍스트의 양이 폭발적으로 증가하면서 그 가운데 서로 유관한 데이터가 서로에 대해 의미를 갖고 모여질 수 있도록 하는 기술이 탐구되기 시작했다. 이른바 시맨틱 웹(Semantic Web)이라고 하는 것이다.¹⁾ 시맨틱 웹은 인터넷상에 존재하는 여러 가지 자원들이 각각 별개의 것으로 보이기보다, 서로에 대해 어떠한 의미적 관계를 갖는지를 알게 하는 것을 지향한다.²⁾

IV-1. 시맨틱 웹이란?

‘시맨틱 웹’의 이상은 월드와이드웹 상에서 ‘지식의 조각’이라고 할 수 있는 개체들(individuals)이 서로서로 어떤 의미로 관계를 맺는지 명시적으로 드러내고, 그 바탕 위에서 개체와 개체 사이의 관계를 추론하여 새로운 사실을 발견하거나 종합적인 지식을 얻으려는 것이다.³⁾ 이 이상적인 목표를 이루기 위한 실천적 노력의 첫걸음은 한 편의 월드와이드웹 문서를 작성할 때 그것이 다른 무엇과 어떠한 관계를 갖는지 표시해 주는 것이다. 서로에 대해 ‘관계가 있음’을 알리는 데이터의 집합을 ‘링크드 데이터’(Linked Data)라고 하며, 그 링크드 데이터가 인터넷상에서 공개되어, 누구라도 그 연결을 확장해 갈 수 있는 것을 ‘링크드 오픈 데이터’(Linked Open Data, LOD)라고 한다.⁴⁾

1) ‘시맨틱 웹’은 어떤 기술이나 표준을 특정하는 말이기보다, 미래의 ‘월드와이드웹’(World Wide Web)에 대한 기대와 이상을 표현하는 용어이다. 하지만 그 이상을 실현하기 위해서는 기술을 개발하고 표준을 정하는 노력이 있어야 한다. 시맨틱 웹은 월드와이드웹의 창시자 팀 버너즈리가 1998년 제안한 이후, 월드와이드웹 컨소시엄(World Wide Web Consortium: W3C)을 중심으로 그것의 구현을 위한 기술의 확장과 표준화 작업이 진행중이다.

2) 시맨틱 웹이란 컴퓨터가 인식할 수 있도록 명확한 의미를 부가한 데이터의 웹이다. 컴퓨터가 인식할 수 있는 데이터 기술 규칙을 만드는 것은 정보기술의 영역에서 추구할 과제이지만, 그 규칙을 응용하고 확장하여 미래의 사이버 공간에서 인문지식이 소통될 수 있도록 만드는 것은 디지털 인문학자들의 과제이다.

3) 시맨틱 웹에 대해 팀 버너즈리는 “현재의 웹에 명확한 의미의 정보를 부가하여 사람과 컴퓨터가 협업할 수 있도록 하는 것”이라고 하였다. (Berners-Lee et al. The Semantic Web, Scientific American (May 2001))

4) 시맨틱 웹(Semantic Web)과 LOD(Linked Open Data): 팀 버너즈리는 시맨틱 웹에 다가가기 위한 실천 전략으로 ‘Linked Data’와 ‘Linked Open Data’의 개념을 소개하였다.

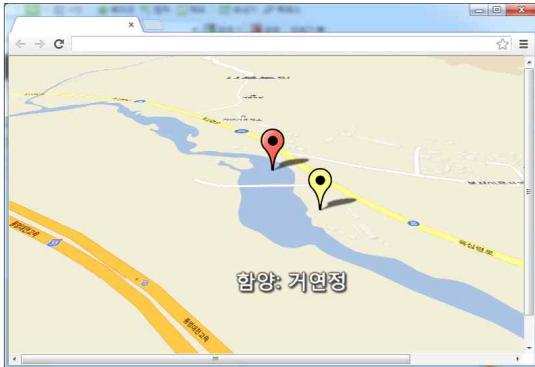
☞ Tim Berners-Lee, “Linked Data”, W3C Design Issues (2006, 2009),

<https://www.w3.org/DesignIssues/LinkedData.html>

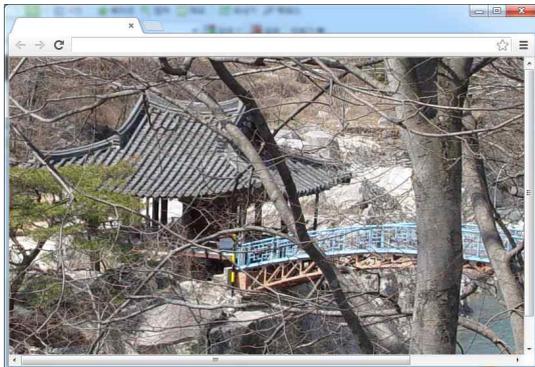
‘시맨틱 웹’이 데이터의 연결망으로서의 미래의 웹에 대한 포괄적인 구상이라고 한다면, LOD는 HTTP URI(Uniform Resource Identifier)와 RDF(Resource Describing Framework) 등의 데이터

시맨틱 웹을 통해 무엇을 얻을 수 있는지 간단한 예시를 통해 알아보기로 하자.

[예시] 내용적으로 관련한 월드와이드웹 자원



[1-1]
함양 거연정 위치
Google Maps



[1-2]
함양 거연정 사진
문화재청 문화재 DB



[1-3]
함양 거연정 기사
한국학중앙연구원 한국민족문화대백과사전

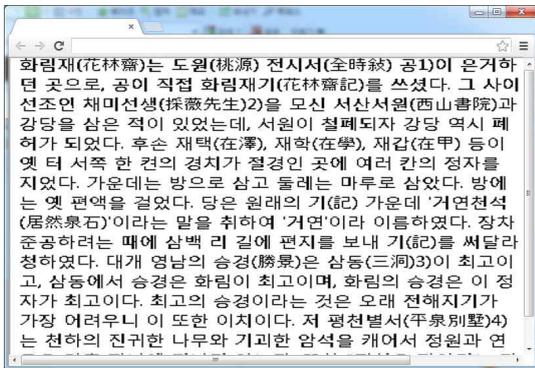
기술 형식에 의존하여, 일정한 영역에서 데이터의 연결을 구현하는 구체적인 방법의 제안이라고 할 수 있다.



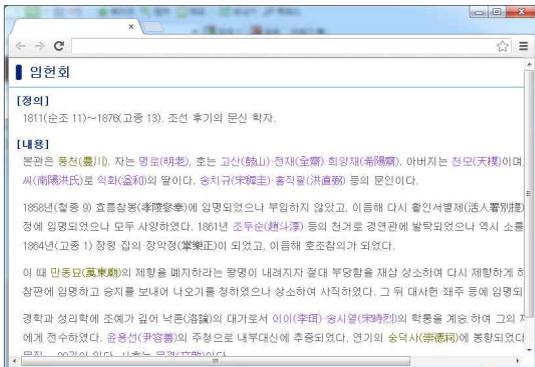
[1-4]
거연정기 편액 사진
개인 웹 페이지



[1-5]
거연정기 한문 원문
고전번역원 문집총간 DB



[1-6]
거연정기 번역문
개인 웹 페이지



[1-7]
임헌회(거연정기 저자) 기사
한국학중앙연구원 역대인물 DB

위에 예시로 든 7개의 디지털 자료는 모두 경상남도 함양군 남강천에 있는

유서 깊은 누정 ‘거연정’과 관련이 있는 것이다. 그것들은 내용면에서 관련성이 있지만 각각 다른 곳에 별개의 문서로 존재한다. 이러한 자료를 고립된 자료원으로 두지 않고, 의미의 연결 고리를 좇아 문맥을 이루게 하려면 어떠한 노력이 더해져야 할까?

현재의 월드와이드웹 환경에서는 그 문서들에 공통으로 포함되어 있는 어휘-이름테면 ‘거연정’-를 이용해서 서로 관계가 있을 수 있는 후보 문서를 모두 탐색한 후에, 이용자의 눈으로 그 결과 하나하나를 확인하면서 필요한 것을 취하는 방법이 가능할 뿐이다.

시맨틱 웹은 데이터가 생산될 때 유관한 자료의 의미적 연관 관계를 약속된 방식으로 명시함으로써 보다 지능적인 데이터 연계가 이루어지게 할 것을 제안한다. 위에 예시한 월드와이드웹 상의 데이터에 대해 그 관계성을 정의하는 데이터를 다음과 같이 만들었다고 가정해 보자.

- 개체의 고유한 이름(IRI: Internationalized Resource Identifier) 정의⁵⁾

N1: 월드와이드웹 상에서 ‘함양 거연정’을 유일하게 지목하는 식별자

N2: 월드와이드웹 상에서 ‘거연정기’를 유일하게 지목하는 식별자

N3: 월드와이드웹 상에서 ‘임헌회’를 유일하게 지목하는 식별자

D1: [1-1] 문서의 월드와이드웹 주소

D2: [1-2] 문서의 월드와이드웹 주소

D3: [1-3] 문서의 월드와이드웹 주소

D4: [1-4] 문서의 월드와이드웹 주소

D5: [1-5] 문서의 월드와이드웹 주소

D6: [1-6] 문서의 월드와이드웹 주소

D7: [1-7] 문서의 월드와이드웹 주소

- 개체간의 관계성 정의

N1의 누정기는 N2이다.

N2의 저자는 N3이다.

D1은 N1의 지도이다.

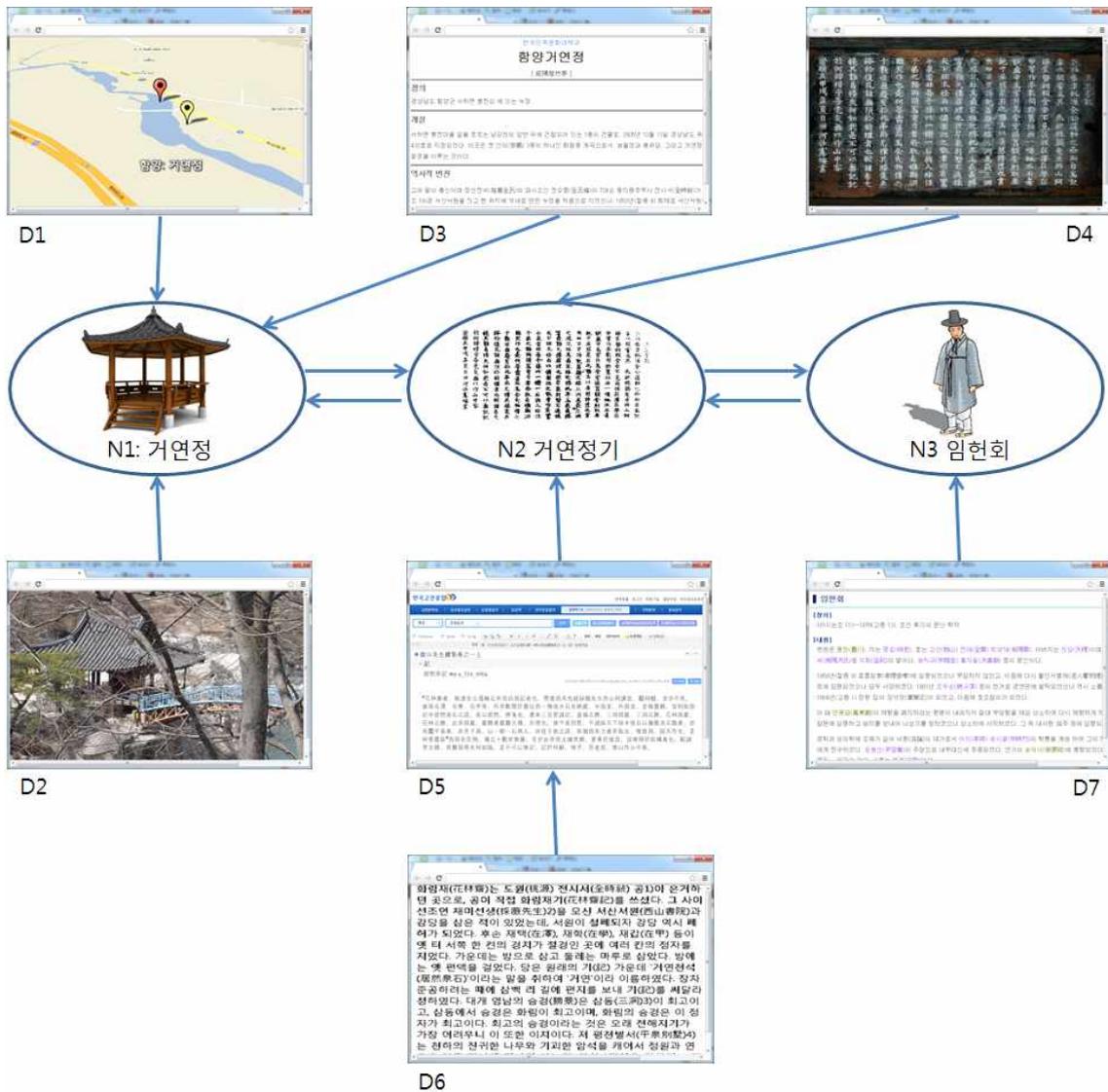
D2는 N1의 사진이다.

D3는 N1에 대한 설명이다.

5) IRI(Internationalized Resource Identifier): 인터넷상의 특정 자원을 유일하게 식별할 수 있게 하는 이름. 종래의 URI(Uniform Resource Identifier)는 ASCII 문자(영문 알파벳과 기호)를 사용했던 데 반해 IRI는 한글, 한자를 포함하는 세계문자부호계(Unicode/ISO 10646)의 모든 문자를 사용할 수 있도록 하였다. IETF(Internet Engineering Task Force)에서 2005년에 제안.
<http://tools.ietf.org/html/rfc3987>

- D4는 N2의 사진이다.
- D5는 N2의 한문 원문 텍스트이다.
- D6는 N2의 번역문이다.
- D7은 N3에 대한 설명이다.

이러한 방식으로 관계성을 기술한 정보가 [예시 3]에서 보인 개별 문서와 함께 인터넷상에 존재할 때, 우리가 현재의 월드와이드웹에서는 얻을 수 없는 어떤 것을 더 얻게 될 것인지는 쉽게 이해할 수 있을 것이다. 그 이전까지는 웹상에서 고립된 형태로 존재하던 지식의 조각이 다른 조각들과 함께 모여 커다란 의미체를 이루게 된다.



IV-2. 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.⁶⁾ 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

6) 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/>

선결되어야 한다. 첫째는 관계성을 부여하고자 하는 객체가 월드와이드웹 상에서 유일하게 식별될 수 있는 이름을 가져야 하는 것, 두 번째는 객체와 객체 사이의 관계를 설명하는 관계 서술어의 표준적인 형태를 약속해야 한다는 것이다. 이 예시에서 보이하고자 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)⁷⁾를 부여할 수 있다.

그 다음 필요한 것은 이 객체들 사이의 관계를 표현하는 서술어이다. “~는 ~의 저자이다”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.

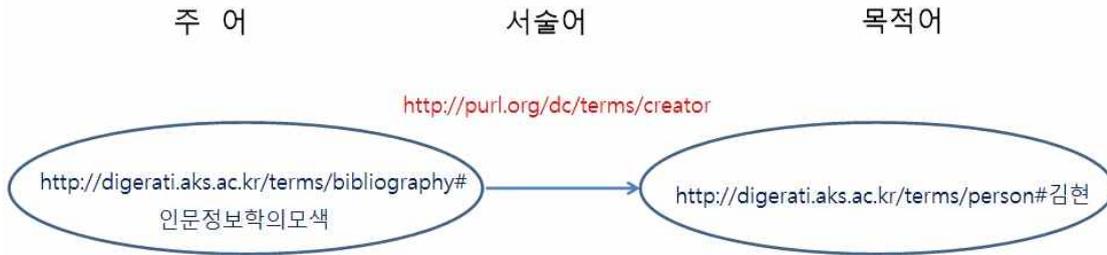
이제 주어, 목적어, 서술어로 쓸 디지털 언어의 어휘들이 정해졌기 때문에 그 어휘들을 영어의 3형식 문장과 같은 RDF 구문 형식에 집어넣으면 RDF 문이 완성된다.

[예시] RDF 문의 구문 형식

7) 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/XML⁸⁾

```
<?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.⁹⁾

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

9) 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/bibliography#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

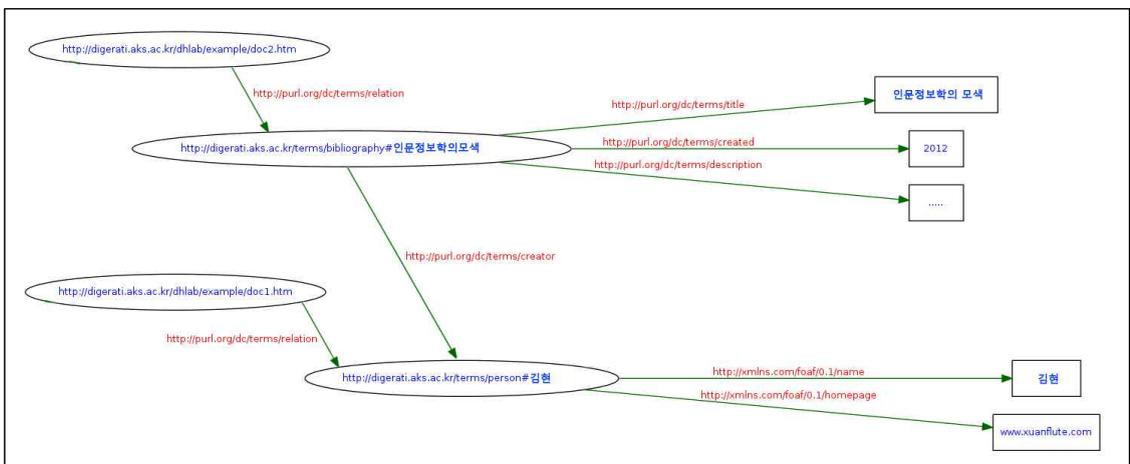
관계 정보	문서2 -[관계가 있다]→ #인문정보학의모색 -[이 책의 속성]→ 저자 이름, 해제
RDF GRAPH	<pre> graph LR A([http://digerati.aks.ac.kr/dhlab/example/doc2.htm]) -- http://purl.org/dc/terms/relation --> B([http://digerati.aks.ac.kr/terms/bibliography#인문정보학의 모색]) B -- http://purl.org/dc/terms/title --> C[김현] B -- http://purl.org/dc/terms/description --> 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	<p>The graph shows a central node for the document '인문정보학의 모색' (http://digerati.aks.ac.kr/terms/bibliography#인문정보학의 모색). It is connected to a person node '김현' (http://digerati.aks.ac.kr/terms/person#김현) via the property 'http://purl.org/dc/terms/creator'. The document node is also connected to a title node '인문정보학의 모색' (http://purl.org/dc/terms/title) and a date node '2012' (http://purl.org/dc/terms/created).</p>
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



IV-3. 온톨로지

앞에서 살펴보았듯이 시맨틱 웹을 구현하는 핵심 요소는 데이터와 데이터를 명시적으로 연결해 주는 개체 식별자와 관계성 서술 어휘들이다. 이러한 어휘는 누구든지 자기 필요에 따라 만들어낼 수도 있지만, 그렇게 되면 시맨틱 웹이 목적하는 ‘소통’의 효과를 거두기 어렵다. 같은 언어를 쓰는 사람끼리 의사소통이 가능하듯이, 의미가 통하는 웹이 만들어지기 위해서는 RDF와 같은 데이터 기술문의 어휘에 관해 일정한 약속을 만들고 이를 지켜야 한다. 정보기술의 세계에서는 그와 같은 성격의 약속을 ‘온톨로지’(ontology)라는 이름으로 부른다.

1) 온톨로지 설계

‘온톨로지’란 정보화의 대상이 되는 세계를 전자적으로 표현할 수 있도록 구성한 데이터 기술 체계이다.¹⁰⁾ 원래 온톨로지라는 말은 철학에서 ‘존재론’이라고 번역되는 용어로서 ‘존재에 대한 이해를 추구하는 학문’의 의미를 갖는 말이었다. 그러한 용어가 정보과학 분야에서 중요한 개념으로 등장하게 된 것은 인간이 세계를 이해하는 틀과 컴퓨터가 정보화 대상(콘텐츠)을 이해하는 틀 사이에 유사성이 있다고 보았기 때문이다. 그 틀은 바로 대상을 구성하는 요소들에 대응하는 개념들과 그 개념들 간의 연관 관계이다.¹¹⁾

넓은 의미에서는 모든 정보화의 틀이 다 온톨로지일 수 있겠지만, 대상 자원을 ‘클래스’(class)로 범주화하고, 각각의 클래스에 속하는 개체(individuals)들이 공통의 ‘속성’(attribute)을 갖도록 하고, 그 개체들이 다른 개체들과 맺는 ‘관계’(relation)를 명시적으로 기술하는 것이 가장 일반적인 온톨로지 설계 방법이라고 할 수 있다.

다음의 예시를 통해 인문지식을 정보화하기 위한 온톨로지는 어떻게 구상하

10) 정보기술 분야에서 말하는 ‘온톨로지(ontology)’에 대한 가장 일반적인 정의는 그루버(Gruber, Thomas, 1959~)가 말한 ‘명시적 명세화의 방법에 의한 개념화’(explicit specification of a conceptualization)이다. (Gruber, ‘A Translation Approach to Portable Ontology Specifications’, *Knowledge Systems Laboratory Technical Report KSL 92-71*, Stanford University, 1992) 여기서 ‘개념화’(conceptualization)라는 것은 정보화하고자 하는 대상 세계를 일정한 체계 속에서 파악하는 것, 예를 들면 그 세계에 무엇이 있고, 그것은 어떤 속성을 품고 있으며, 그것들 사이의 관계는 무엇인가 하는 일정한 질문의 틀 속에서 대상 세계를 이해하는 방식이라고 할 수 있다. ‘명세화’(specification)란 대상 세계에 존재하는 개체, 속성, 관계 등을 일목요연한 목록으로 정리하는 것, 그리고 ‘명시적’(explicit)이라는 그 정리된 목록을 사람뿐 아니라 ‘컴퓨터가 읽을 수 있도록’(machine readable) 한다는 것이다.

11) 김현, 「한국 고전적 전산화의 발전 방향 - 고전 문집 지식 정보 시스템 개발 전략 -」, 『민족문화』 28 (2005)

는 것을 좋을지 생각해 보기로 하자.

[예시] 명승 40호 ‘담양 소쇄원’ 안내판의 문화유산 설명문

담양 소쇄원(潭陽 瀟灑園)

명승 제40호
전라남도 담양군 남면 지곡리

소쇄원(瀟灑園)은 자연과 인공을 조화시킨 조선 중기의 대표적인 원림(園林)으로 우리나라 선비의 고고한 품성과 절의가 풍기는 아름다움이 있다.

양산보(梁山甫, 1503~1577)가 조성한 것으로 스승인 조광조(趙光祖)가 유배를 당하여 죽게 되자 출세의 뜻을 버리고 이곳에서 자연과 더불어 살았다. 소쇄원이라 한 것은 양산보의 호(號)인 소쇄옹(瀟灑翁)에서 비롯되었으며, 맑고 깨끗하다는 뜻이 담겨 있다.

오곡문(五曲門) 담장 밑으로 흐르는 맑은 계곡 물은 폭포가 되어 연못에 떨어지고, 계곡 가까이에는 제월당(霽月堂: 비개인 하늘의 상쾌한 달이라는 뜻의 주인집)과 광풍각(光風閣: 비온 뒤에 해가 뜨며 부는 청량한 바람이란 뜻의 사랑방)이 들어서 있다.

소쇄원에는 영조 31년(1775) 당시 모습을 목판에 새긴 「소쇄원도(瀟灑園圖)」가 남아 있어 원형을 추정할 수 있다. 이곳은 많은 학자들이 모여들어 학문을 토론하고, 창작 활동을 벌인 선비정신의 산실이기도 하다. 지금의 소쇄원은 양산보의 5대손 양택지에 의해 보수된 모습이다.



위 예시의 글은 전라남도 담양군에 위치한 조선시대의 대표적인 정원인 소쇄원(瀟灑園)의 문화재 안내판 해설문이다. 이 문화유산과 관련된 다양한 지식의 실마리를 위의 해설문에서 찾고, 이것을 유형별로 정리하면 다음과 같은 키워드 목록을 얻을 수 있을 것이다.

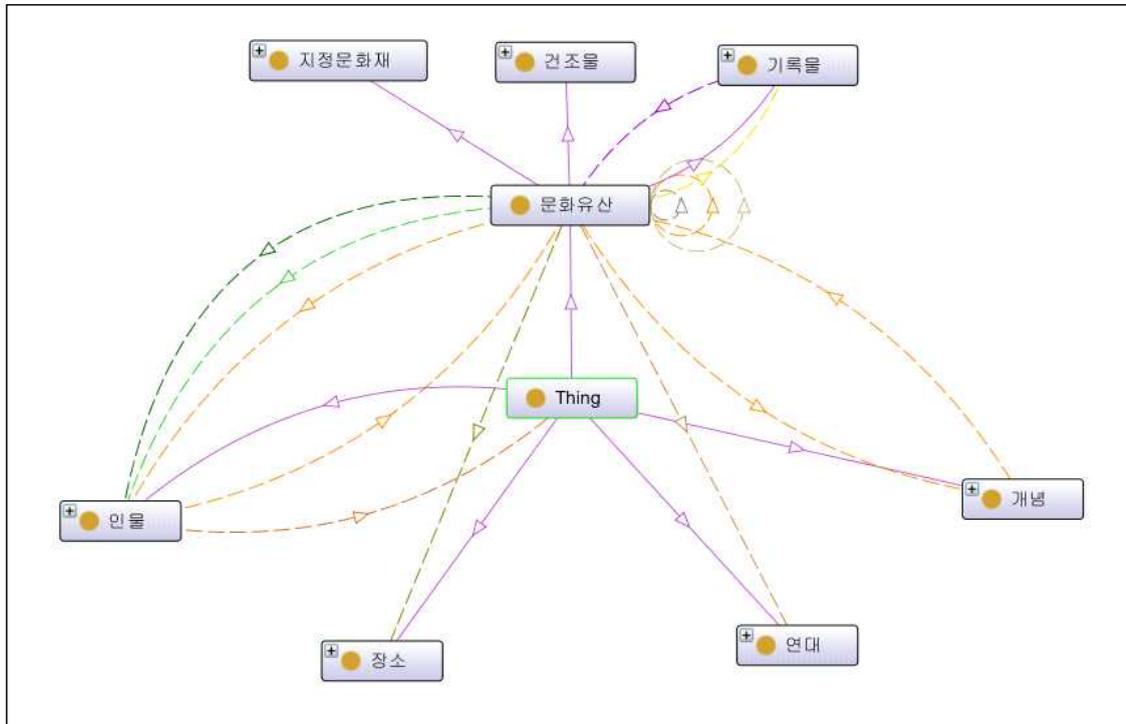
유형 범주		키워드
문화 유산	지정문화재	담양 소쇄원
	건조물	제월당, 광풍각, 오곡문
	기록물	소쇄원도
인물		양산보, 조광조, 양택지
장소(공간)		전라남도 담양군 남면 지곡리
연대(시간)		영조 31년(1775)
개념		원림(園林), 절의(節義)

위의 표에 정리한 어휘들은 담양 소쇄원 문화재 안내판 해설문에 적힌 내용의 뼈대를 이루는 문맥 요소(Contextual Element)들이다. 이 문맥 요소들은 문장 속의 단어로만 존재하는 것이 아니라, 소쇄원에 얽힌 역사, 그것의 문화적 의미에 관한 지식을 대표하는 용어라고 할 수 있다. 이것을 근거로, 담양 소쇄원 관련 지식의 관계도를 만들어 보기로 한다.

① 개체(individual) 탐색

온톨로지 설계를 위해 제일 먼저 할 일은 정보화 하고자 하는 지식 세계에 어떠한 지식 요소들이 있는지 탐색하고, 그 성격을 분석하는 것이다. 위의 예시에서 목록화한 개개의 요소들을 온톨로지 용어로는 ‘개체’라고 한다. 이것은 정보화하고자 하는 지식의 단위 요소이자, 관련 지식의 네트워크에서 관계의 접점이 될 노드(node)이다.

Class:



이 그림의 가운데에 그려진 'Thing'은 대상 세계의 모든 존재를 포괄하는 가상의 최상위 범주이다. 그 아래에 '문화유산', '인물', '장소', '연대', '개념' 등 5가지 차상위 클래스를 두기로 하고, 다시 '문화유산' 클래스 밑에 '지정문화재', '건조물', '기록물' 등의 하위 클래스를 두었다. 문화유산 클래스를 이렇게 중층으로 구성한 것은 예시로 든 문화유산 사례에서 보듯이 '담양 소쇄원'이라는 지정문화재(명승 제40호) 안에 '광풍각', '제월당'과 같은 부속 건조물도 있고, '소쇄원도'와 같은 기록물도 있기 때문에 디지털 정보세계가 그와 같은 체계를 반영할 수 있도록 하기 위해서였다. 문화유산의 범위를 넓히게 되면 또 다른 하위 요소도 추가될 수 있을 것이다. 유적지 경내에 있는 노거수(老巨樹), 묘정비(廟庭碑) 등도 고려 대상이 될 수 있다.

③ 속성(attribute) 설계

클래스 설계가 이루어졌으면, 각각의 클래스에 속하는 개체(individual)들이 어떤 속성을 갖는지를 살피고, 이 속성을 담을 수 있는 틀을 만들어야 한다. 예를 들어 '인물' 클래스에 속하는 개체들에 대해서는 이름, 이름의 한자 표기,

별명(자, 호, 봉작호, 시호 등), 생몰년, 관직 등의 정보를 부여하는 것이다. 클래스 설계와 마찬가지로 속성 설계도 ‘이렇게 해야 한다’는 절대적인 기준은 없다. 대상 개체를 어떤 관점에서 보고 있으며, 그 정보를 어떠한 목적으로 사용할 것인가를 기준으로 삼으면 된다. 간단한 예시를 위해 여기서는 다음과 같이 최소한의 속성만을 부여하기로 한다.

Attributes:

Class	한자	유형	생년	몰년	위도	경도	서력	설명
지정문화재	○	○						○
건조물	○	○			○	○		○
기록물	○	○						○
인물	○	○	○	○				○
장소	○	○			○	○		○
연대	○	○					○	○
개념	○	○						○

④ 관계성(relation) 설계

각각의 클래스에 속하는 개체들이 서로 어떠한 의미적 연관 관계를 맺고 있는지 분석하여, 그 관계성을 표현할 수 있는 서술어를 정한다. 이 관계성 서술어는 주어와 목적어를 수반하여 RDF(Resource Description Framework) 문을 완성하게 된다. 이 때 주어 역할을 할 수 있는 개체들의 클래스를 ‘정의역’(定義域, domain)이라고 하고, 목적어 역할을 하는 개체들의 클래스를 ‘치역’(值域, range)라고 부른다. 담양 소재원과 관계가 있는 개체들을 기준으로 아래와 같은 17개의 관계 서술어를 정하였다.¹²⁾

12) 온톨로지의 어휘들은 컴퓨터 시스템이 인식할 수 있도록 일관성 있게 사용하면 되는 것이기 때문에 원천 어휘를 한국어로 하느냐 영어로 하느냐는 기술적으로 문제되지 않는다. 하지만 온톨로지 공유를 통해 소통하고자 하는 범위가 국제적이라면, 세계 여러 나라에서 이미 사용하고 있는 동의어의 사용이 권장된다. 이 책에서의 온톨로지 설계는 교육의 목적으로 하는 것이기 때문에 학생들이 이해하기 쉽도록 한국어 어휘를 위주로 하였다. 다만 관계성 서술어는 한국어 어휘를 쓸 경우 주어와 목적어의 식별이 모호해질 수 있기 때문에 편의상 영어식의 조어(造語)를 사용하기로 한다.

Relations:

Relation	Domain (A)	Range (B)	설명
created	문화유산	연대	A는 B에 만들어졌다
creator	문화유산	인물	A는 B가 만들었다
renovator	문화유산	인물	A는 B가 중건(중수)했다
depicts	기록물	문화유산	A는 B를 묘사한다
isDepictedIn	문화유산	기록물	A는 B에 묘사되었다
hasDescendant	인물	인물	A는 B를 후손으로 두었다
isDescendantOf	인물	인물	A는 B의 후손이다
hasDisciple	인물	인물	A는 B를 문인으로 두었다
isDiscipleOf	인물	인물	A는 B의 문인이다
hasPart	문화유산	문화유산	A는 B를 포함한다
isPartOf	문화유산	문화유산	A는 B의 일부이다
hasSon	인물	인물	A는 B의 아버지이다
isSonOf	인물	인물	A는 B의 아들이다
isEnshrinedIn	인물	문화유산	A는 B에서 제향된다
isFriendOf	인물	인물	A와 B는 친구이다
isLocatedIn	문화유산	장소	A는 B에 위치한다
isRelatedTo	개념, 인물, 문화유산	개념, 인물, 문화유산	A와 B는 관계가 있다

⑤ 개체(individual)에 대한 속성, 관계성 부여

지금까지 수행한 네 가지 과정이 온톨로지 설계에 해당하는 일이라고 한다면 이 다섯 번째 작업은 설계된 온톨로지에 따라 데이터베이스를 구축하는 일이다.¹³⁾ 문맥 요소로 발굴된 모든 개체를 해당 클래스에 귀속시킨 후, 각각의 개체에 고유한 속성을 부여하고, 개체와 개체 사이의 관계성을 지정한다.

§ Individual ‘담양 소쇄원’

- * Class: 문화유산 > 지정문화재
- * Attribute
 - 한자: 潭陽 瀟灑園
 - 유형: 명승
- * Relation

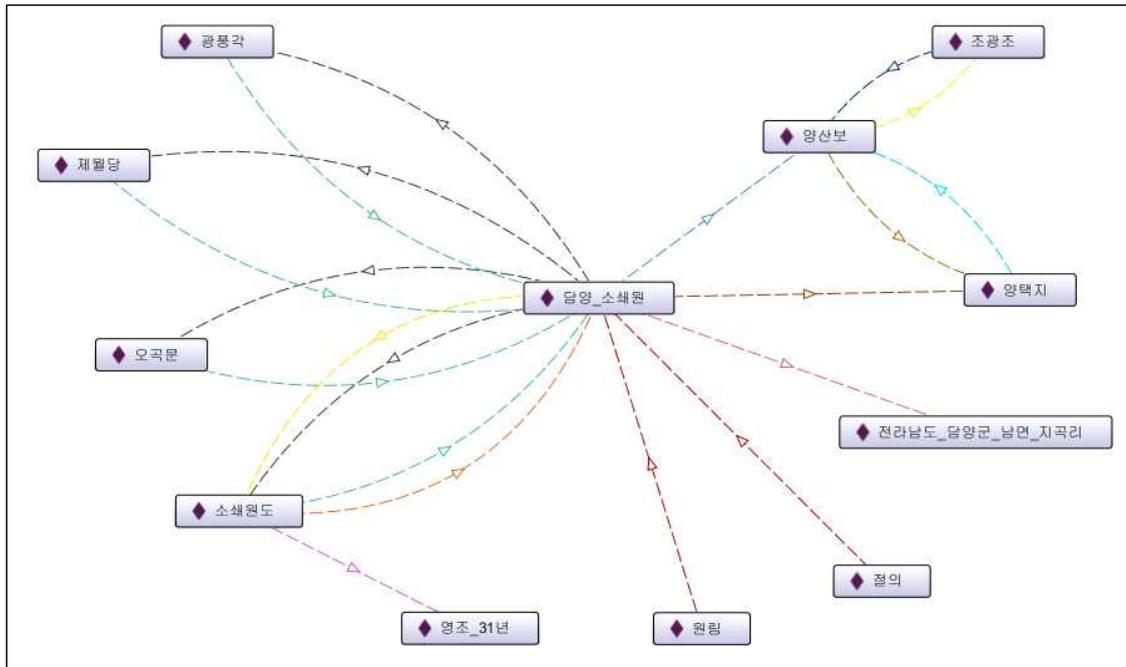
13) 온톨로지 설계에 의해 만들어진 데이터베이스를 일반적인 관계형 데이터베이스(Relational Database, RDB)와 구분하여 ‘온톨로지 기반 데이터베이스’(Ontology Based Database, ODB)라고 부르기도 한다. 실험적인 ODB는 온톨로지 설계 도구를 이용하여 구현할 수 있다.

hasPart: 광풍각
 hasPart: 소쇄원도
 hasPart: 오곡문
 isDepictedIn: 소쇄원도
 creator: 양산보
 renovator: 양택지

§ Individual: '양산보'

- * Class: 인물
- * Attribute:
 - 생년: 1503
 - 몰년: 1557
 - 한자: 梁山甫
- * Relation
 - hasDescendant: 양택지
 - isDiscipleOf: 조광조

12개 개체(individuals)에 개체에 대해 모두 이러한 작업을 수행하면 아래 그림과 같이 관계의 맥락을 명시적으로 드러내게 된다.

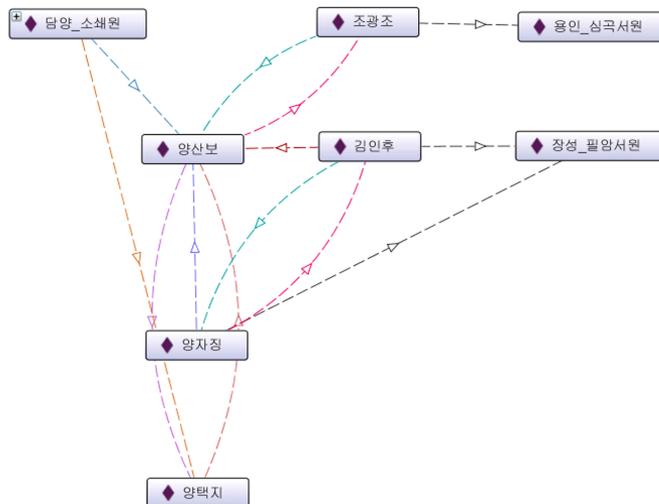


위의 그래프에 담겨 있는 정보는 다음과 같은 것들이다.

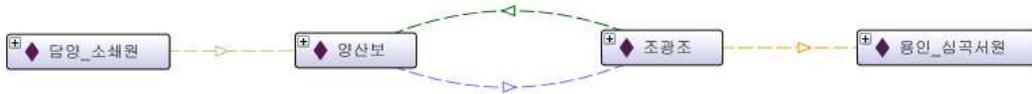
[문화유산]‘담양 소쇄원’은 [장소]‘전라남도 담양군 남면 지곡리’에 위치한다.
 [문화유산]‘담양 소쇄원’은 [인물]‘양산보’가 세웠다.
 [인물]‘양산보’는 [인물]‘조광조’의 제자이다.
 [문화유산]‘담양 소쇄원’은 [개념]‘원림’/‘절의’와 관련이 있다.
 [문화유산]‘담양 소쇄원’ 안에는 [건조물]‘광풍각’/‘제월당’/‘오곡문’이 있다.
 [문화유산]‘담양 소쇄원’ 안에는 [기록물]‘소쇄원도’가 있다.
 [기록물]‘소쇄원도’는 [문화유산]‘담양 소쇄원’을 묘사했다.
 [기록물]‘소쇄원도’는 [연대]‘영조 31년’에 제작되었다.
 [인물]‘양택지’는 [인물]‘양산보’의 후손이다.
 [인물]‘양택지’는 [문화유산]‘담양 소쇄원’을 중수했다.

이 데이터는 이것의 원시 데이터라고 할 수 있는 안내판 해설문 속의 정보를 컴퓨터가 읽고 해석할 수 있는 형태로 가공한 것이라고 할 수 있다. 이와 같은 기계가독형(Machine-Readable) 데이터를 만들어서 얻는 이점은 무엇일까? ‘담양 소쇄원’이라는 하나의 문화유산 정보만 디지털화한 상황에서는 온톨로지 기반 데이터베이스의 장점을 드러내기 힘들다. 하지만 이러한 정보가 수천 건, 수만 건 있다고 가정해 보면, 다음과 같은 효과를 쉽게 기대할 수 있다.

여러 가지 기대 효과 가운데 가장 우선시되는 것은 별개의 문화유산 해설문 속에 담겨 있는 정보들이 서로 관련성을 좋아 광대한 의미의 연결망을 만들 수 있을 것이라는 점이다. ‘용인 심곡서원’에는 양산보의 스승 조광조가 제향되어 있다. ‘장성 필암서원’에 제향된 김인후는 양산보의 친구이고, 김인후와 함께 제향된 그의 제자 양자징은 양산보의 아들이다. 이 세 곳의 문화유산들은 공통의 관련 인물을 매개로 다음 그림과 같은 관계망을 형성한다.



이와 같이 서로서로 의미의 연결고리를 맺고 있는 지식의 네트워크 속에서는 그 관계를 추론하여 새로운 지식을 얻는 것도 가능하다. 우리가 이 예시를 위해 만든 데이터 안에서도 ‘소쇄원을 만든 사람의 스승은 누구인가?’, ‘그를 제향한 서원은 어디인가?’라는 질문을 던지면, 컴퓨터는 그 질문에 대해 정확한 답변을 제공한다. 이 온톨로지를 공유하는 더 많은 데이터가 만들어졌을 경우, 컴퓨터는 보다 유용한 지식 탐구의 동반자 역할을 할 수 있을 것이다.



2) 온톨로지 기술 언어

어떤 대상을 디지털 공간에 재현하기 위해 만든 온톨로지는 같은 분야의 지식을 다루는 여러 사람들이 지식의 소통을 위한 공동의 약속으로 사용할 때 큰 효용성을 발휘한다. 예를 들자면, 월드와이드웹의 세계 안에는 수많은 사람들이 개인적으로 작성한 ‘문화유산 답사기’가 존재하는데, 그 글 속에 담긴 객관적 사실 정보들이 앞에서 예시한 바와 같은 ‘문화유산 온톨로지’에 의해 기술될 수 있다면, 그 결과는 월드와이드웹 속에 ‘한국의 문화유산에 관한 시맨틱 웹’이 만들어지는 것이고, 우리는 그것을 통해 더욱 풍성한 문화유산 관련 지식을 얻을 수 있게 될 것이다.

특정 분야의 지식에 관한 온톨로지가 여러 사람들, 또는 다수의 컴퓨터 시스템 사이에서 공유될 수 있기 위해서는 그 설계의 내용이 명시적이고 표준적인 언어로 기술되어야 한다. 앞에서 살펴보았던 RDF(Resource Description Framework)는 가장 기본적인 형태의 온톨로지 기술 언어에 속한다고 할 수 있다. RDF는 단순히 주어와 목적어 사이의 관계를 서술하는 것을 위주로 한다. 그런데 그 관계성이 어떤 대상들 사이에서 유효한지, 그리고 다양한 관계 서술어 사이에는 또 어떤 상관관계가 존재하는지 등도 점검할 필요성이 제기됨에 따라 RDF보다 좀 더 확장된 기능의 온톨로지 기술 언어들이 등장하였다.

학계, 산업계에서 실험적으로, 또는 특정 영역에서 실용적으로 쓰이는 온톨로지 기술 언어는 매우 다양하지만, 월드와이드웹 상에서 운용되는 시맨틱 웹의 구현을 목적으로 하는 온톨로지 기술 언어로는 RDF를 확장시켜 나아간

RDFS(RDF Schema)와 OWL(Web Ontology Language)이 표준적인 약속으로 권장되고 있다.¹⁴⁾ RDF, RDFS, OWL은 별개의 언어가 아니라, 먼저 정해 놓은 약속을 기반으로 하면서 더 유용한 수단으로 발전시켜 간 것이라고 할 수 있다.

```

xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"

```

OWL 문서에서 기본적으로 참조하는 이름 공간(Name Space)

OWL은 얼마나 복잡하고 정교한 온톨로지를 설계하느냐에 따라 사용하는 어휘의 종류가 달라질 수 있지만, 핵심적인 것은 클래스, 속성, 관계성, 개체를 선언하고, 각각의 개체에 속성을 지정하거나 개체 사이의 관계성을 지정하는데 필요한 어휘들이다.

OWL 어휘	의미
owl:Ontology	온톨로지 선언
owl:Class	클래스 선언
owl:ObjectProperty	관계성 선언
owl:DatatypeProperty	속성 선언
owl:NamedIndividual	개체 선언
rdfs:subClassOf	서브 클래스 선언
rdfs:domain	정의역 지정
rdfs:range	치역 지정
rdf:datatype	데이터 타입 지정

OWL(Web Ontology Language)의 주요 어휘

14) RDF Schema: Resource Description Framework Schema. RDFS, RDF(S), RDF-S, RDF/S 등 다양한 형태의 약어로 지칭되고 있다. RDFS는 RDF 형식으로 표현하고자 하는 대상의 온톨로지를 기술하는 어휘(vocabulary)를 제공한다. RDF Schema는 1998년 W3C에 의해 처음 제안되었으며, 현재 통용되는 권고안은 2004년에 발표되었다.
<https://www.w3.org/TR/rdf-schema/>
 OWL(Web Ontology Language)는 RDFS보다 확장된 온톨로지 기술 언어이다. W3C에 의해 2004년에 제안되었으며, 이것을 부분적으로 수정·확장한 OWL2는 2009년에 발표되었다.
<https://www.w3.org/TR/owl2-overview/>

온톨로지 기술 언어를 사용하면 우리가 앞에서 예시로 다루었던 ‘담양 소쇄원’ 등의 문화유산 관련 정보를 월드와이드웹 상에서 통용되는 시맨틱 데이터로 기술할 수 있다.

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns="http://digerati.aks.ac.kr/terms/CHoKorea#"
  xml:base="http://digerati.aks.ac.kr/terms/CHoKorea"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:CHoKorea="http://digerati.aks.ac.kr/terms/CHoKorea#">

  <!-- 온톨로지 선언-->

  <owl:Ontology rdf:about="http://digerati.aks.ac.kr/terms/CHoKorea"/>

  <!-- 클래스 선언 예시: ‘문화유산’ 클래스를 선언-->

  <owl:Class rdf:about="http://digerati.aks.ac.kr/terms/CHoKorea#문화유산"/>

  <!-- 속성 선언 예시: ‘인물’ 클래스의 개체는 ‘생년’ 속성을 갖는다-->

  <owl:DatatypeProperty rdf:about="http://digerati.aks.ac.kr/terms/CHoKorea#생년">
    <rdfs:domain rdf:resource="http://digerati.aks.ac.kr/terms/CHoKorea#인물"/>
  </owl:DatatypeProperty>

  <!-- 관계성 선언 예시:
  ‘isLocatedIn’이라는 관계성은 ‘문화유산’과 ‘장소’ 클래스 개체 사이에 성립할 수 있다 -->

  <owl:ObjectProperty rdf:about="http://digerati.aks.ac.kr/terms/CHoKorea#isLocatedIn">
    <rdfs:domain rdf:resource="http://digerati.aks.ac.kr/terms/CHoKorea#문화유산"/>
    <rdfs:range rdf:resource="http://digerati.aks.ac.kr/terms/CHoKorea#장소"/>
  </owl:ObjectProperty>

  <!-- 개체 선언 및 속성/관계성 지정 예시:
  개체 ‘양산보’는 .....의 속성을 가지며 개체 ‘양택지’ 등과 .....의 관계성을 갖는다-->

  <owl:NamedIndividual rdf:about="http://digerati.aks.ac.kr/terms/CHoKorea#양산보">
    <rdfs:type rdf:resource="http://digerati.aks.ac.kr/terms/CHoKorea#인물"/>
    <생년 rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1503</생년>
    <몰년 rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1557</몰년>
    <한자 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">梁山甫</한자>
    <hasDescendant rdf:resource="http://digerati.aks.ac.kr/terms/CHoKorea#양택지"/>
    <isDiscipleOf rdf:resource="http://digerati.aks.ac.kr/terms/CHoKorea#조광조"/>
  </owl:NamedIndividual>

</rdf:RDF>

```

OWL 어휘를 XML/RDF 문법으로 표현한 웹 온톨로지 기술 문서

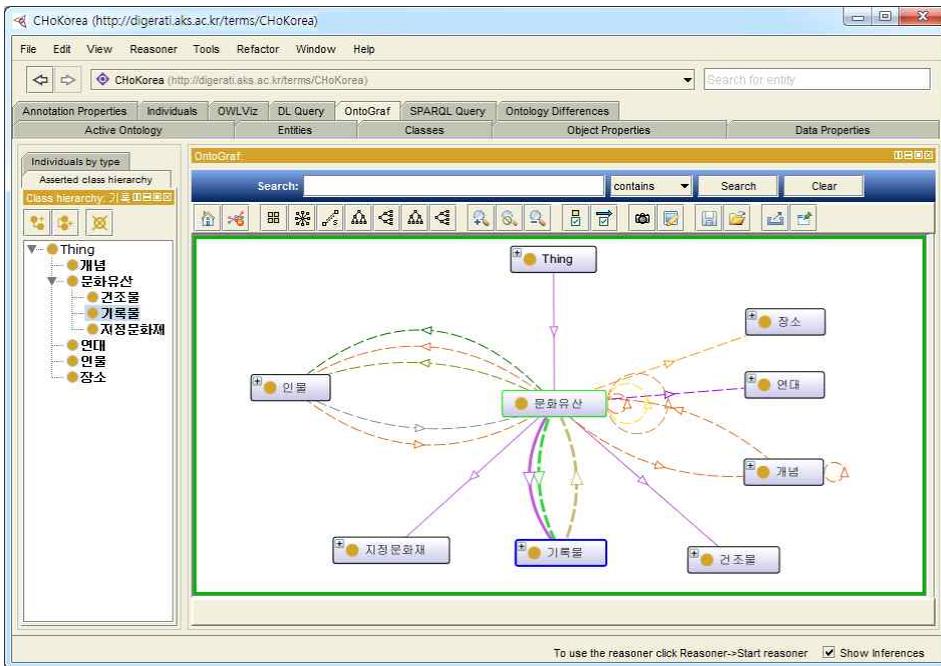
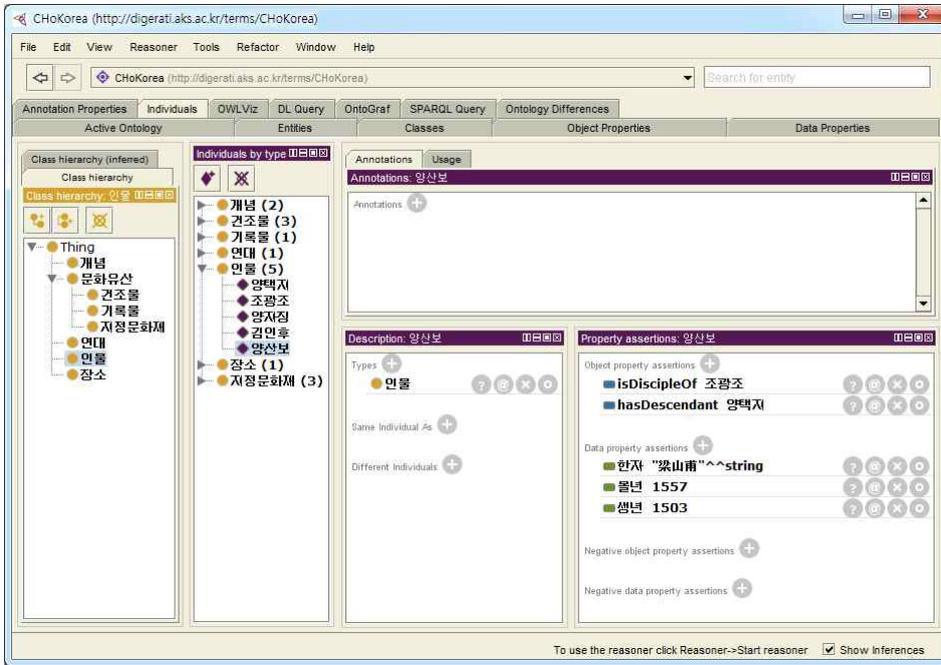
이와 같은 웹 온톨로지 기술 문서에서 우리가 의미 있게 보아야 할 것은 대

상 세계가 존재하는 모습을 표현하는 모든 데이터가 유일하게 식별될 수 있는 이름을 갖도록 했다는 점이다.

이 온톨로지 전체가 '<http://digerati.aks.ac.kr/terms/CHoKorea>'라는 고유한 이름을 부여받았으며, RDF 문에서 동사처럼 쓰이는 관계성, 속성 서술어, 그리고 주어와 목적어가 되는 모든 개체가 온톨로지 이름을 접두어로 삼는 고유한 식별자를 갖는다. 실제 세계에서 쓰이는 이름과 용어가 있는데도, 이와 같이 작위적인 유일 식별자를 또 다시 부여하는 이유는 무엇인가? 어떤 이름이 어떤 대상을 가리키는지 분명하게 알 수 있게 하기 위해서이다. 여기서 '분명하게 안다'는 것은 사람의 능력이 아니라 컴퓨터 시스템의 기능을 두고 하는 말이다.

3) 온톨로지 편집기

온톨로지의 설계는 현재 우리의 관심 분야인 어느 영역의 대상 세계를 특별한 목적을 지닌 관점에서 바라보고 해석하는 일이기 때문에, 기술적인 작업 이전에 대상 세계를 적정하게 분석하는 과정이 중요하다. 대상 세계를 어떻게 정보화 할 것인지 구체적인 구상이 이루어진 후 그것을 온톨로지 기술 언어로 표현하는 것은 '온톨로지 편집기'(Ontology Editor)의 도움을 받아서 수행하면 편리하다. 온톨로지 편집기는 '온톨로지'의 설계와 검증, 지속적인 확장 및 수정, 보완 기능을 제공하는 소프트웨어이다. 이용자는 이 소프트웨어상에서 클래스, 속성, 관계성, 개체 등 온톨로지 핵심 요소들을 디자인할 수 있고, 데이터의 입력을 통해 어느 정도 규모의 실험적인 데이터베이스를 구축할 수도 있다. 온톨로지 편집기상에서 데이터 클래스의 구조, 개체 상호간의 관계성 등을 시각적인 그래프로 확인하는 것도 가능하다.



온톨로지 편집기 Protégé™¹⁵⁾

15) Protégé™: 온톨로지 설계 및 시각화 기능을 제공하는 소프트웨어. 미국 스탠포드 대학의 의료정보학센터(Stanford Center for Biomedical Informatics Research)에서 개발 연구를 이끌고 있는 오픈 소스 소프트웨어. 다양한 온톨로지 에디터 가운데 교육 및 연구 분야에서 가장 많이 쓰이고 있는 제품이다. <http://protege.stanford.edu/>

4) SPARQL

온톨로지를 설계하고 그에 따라 데이터를 입력하면 ‘온톨로지 기반 데이터베이스’(Ontology Based Database, ODB)를 구축할 수 있다. ‘스파클’(SPARQL, Simple Protocol and RDF Query Language)과 같은 RDF 질의어는 그와 같은 데이터베이스상에서 필요한 정보를 찾는 방법을 제공한다.¹⁶⁾ 다음은 스파클 질의어의 간단한 사용 예시이다.

SPARQL	질의 내용																
<pre> PREFIX chk: <http://digerati.aks.ac.kr/terms/CHoKorea#> PREFIX dc: <http://purl.org/dc/terms/> SELECT ?picture WHERE { chk:조광조 chk:hasDisciple ?X. ?Y dc:creator ?X. ?Y chk:isDepictedIn ?picture. } </pre> <p>☞ 소재원도</p>	<p>조광조의 제자 X가 만든 Y가 묘사된 그림은?</p> <p>☞ 소재원도</p>																
<pre> PREFIX chk: <http://digerati.aks.ac.kr/terms/CHoKorea#> PREFIX dc: <http://purl.org/dc/terms/> SELECT ?site WHERE { chk:담양_소재원 dc:creator ?X. ?Y chk:isFriendOf ?X. ?Y chk:isEnshrinedIn ?site. } </pre> <p>☞ 장성_필암서원</p>	<p>담양 소재원을 만든 X의 친구 Y가 제향된 곳은?</p> <p>☞ 장성_필암서원</p>																
<pre> PREFIX chk: <http://digerati.aks.ac.kr/terms/CHoKorea#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> SELECT ?site ?part WHERE { {?site rdf:type chk:지정문화재} UNION {?site chk:hasPart ?part} } </pre> <p>☞</p> <table border="1" style="margin-left: 20px;"> <thead> <tr> <th>site</th> <th>part</th> </tr> </thead> <tbody> <tr> <td>담양_소재원</td> <td></td> </tr> <tr> <td>장성_필암서원</td> <td></td> </tr> <tr> <td>용인_심곡서원</td> <td></td> </tr> <tr> <td>담양_소재원</td> <td>광풍각</td> </tr> <tr> <td>담양_소재원</td> <td>제월당</td> </tr> <tr> <td>담양_소재원</td> <td>오곡문</td> </tr> <tr> <td>담양_소재원</td> <td>소재원도</td> </tr> </tbody> </table>	site	part	담양_소재원		장성_필암서원		용인_심곡서원		담양_소재원	광풍각	담양_소재원	제월당	담양_소재원	오곡문	담양_소재원	소재원도	<p>지정문화재에 속하는 문화유산과 그것에 부속된 문화유산은?</p>
site	part																
담양_소재원																	
장성_필암서원																	
용인_심곡서원																	
담양_소재원	광풍각																
담양_소재원	제월당																
담양_소재원	오곡문																
담양_소재원	소재원도																

16) SPARQL(Simple Protocol and RDF Query Language): RDF 형식의 데이터를 탐색할 수 있는 데이터베이스 질의어. W3C(World Wide Web Consortium)에서 2008년에 처음 제안하였다. 2013년 이후 SPARQL 1.1이 W3C의 공식 권장안이다. <https://www.w3.org/TR/sparql11-query/>

스파클은 하나의 독립적인 데이터베이스 안에서 정보를 찾기보다 온톨로지를 통해 정보의 연결고리를 맺고 있는 시맨틱 웹의 넓은 세계에서 그 관계성을 좇아 의미 있는 정보를 탐색하는 것을 목적으로 한다. 위의 예시에서 보듯이, 'PREFIX' 지정을 통해 이 질의문에서 어떤 온톨로지를 어떤 접두사(prefix)로 참조할지를 정하고, 이하의 질의문에서 사용하는 모든 어휘(개체, 클래스, 관계성, 속성의 이름)는 앞에서 정의된 접두사를 첨부함으로써 그 대상과 의미를 명시적으로 식별할 수 있게 한다. 이러한 방법을 통해 월드와이드 웹 상에 존재하는 다양한 지식 요소들 사이의 관계를 기계적으로 추론할 수 있도록 하는 것이다.

IV-3. 시맨틱 웹과 인문지식

명시적인 기계가독형 언어로 기술된 웹 온톨로지가 월드와이드웹의 세계에서 공유되고, 특정 분야의 지식과 정보를 생산하는 사람들이 이것을 참조하여 상호 소통할 수 있는 데이터를 생산한다면, 그 분야 지식의 활용과 교육, 연구의 양상이 크게 바뀌게 될 것임은 자명하다. 문제는 누가 어떠한 방법으로 그 일을 시작할 것이며, 또 누가 어떠한 역할로 그 일을 발전시켜 가겠는가 하는 것이다.

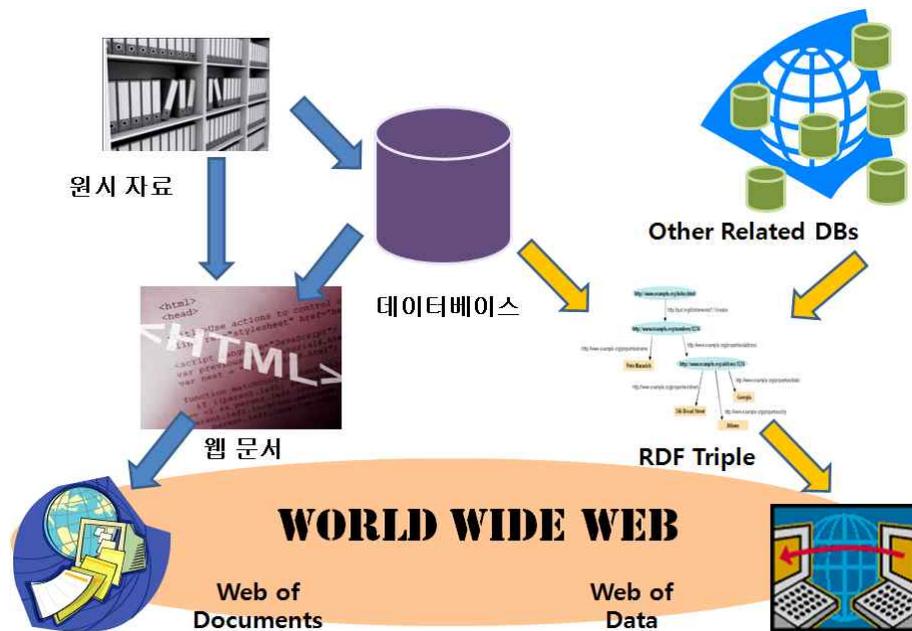
정보화의 대상 세계를 기계적 식별과 분석이 가능하게끔 명시적으로 명세화 하자는 것은 시맨틱 웹의 고유한 주장이 아니며, 모든 정보처리 기술의 기본 전제라고 해도 무방하다. 사실상 회사, 공공기관, 특정 목적의 커뮤니티가 운영하는 데이터베이스 안에서는 이미 오래 전부터 시맨틱 웹이 추구하는 것과 다름없는 일들이 일어나고 있었다. 대학에서 운영하고 있는 학사관리 시스템을 예로 들어 보자. 학생들은 누구 하나 예외 없이 '학번'이라고 하는 고유번호를 갖고 있으며, 강의를 담당하는 교수들도 모두 교원번호를 가지고 있다. 모든 개설 교과목은 강좌코드로 식별되고, 강의실과 강의시간도 다른 장소, 다른 시간대와 구별되는 분명한 이름을 가지고 있다. 그 틀 안에서 어느 학생이 어느 과목을 수강하는지, 그 과목은 어느 교수가 담당하고 언제 어디에서 강의하는지 분명하게 알 수 있게 되고, 컴퓨터는 기계적으로 개설 강의 목록과 과목별 출석부, 학생들의 수강 이력과 성적표를 산출한다. 누구든 이 데이터베이스에 접근할 수만 있다면 간단한 질의어 한 문장을 데이터베이스 관리시스템에 입력함으로써 지난 3년간 인문정보학 강의 수강생 가운데 역사, 문학, 철학, 예

술 전공 학생의 비율을 조회할 수도 있을 것이다.

시맨틱 웹은 어느 한 기관, 조직, 업무의 폐쇄된 영역이 아니라 월드와이드 웹의 무한히 개방된 영역에서 이와 유사한 데이터의 명시적 연계를 이루어내 고자 하는 구상이다. 그것은 과연 가능한 일일까?

학교나 회사 내의 사무적인 일이 요구하는 정보와 달리 월드와이드웹의 세 계에서는 형식과 내용을 한정할 수 없는 다양한 정보가 무한히 넘쳐나고 있다. 이러한 것을 빠짐없이 취급하는 데이터베이스를 만드는 것은 가능하지 않을 뿐 아니라 필요하지도 않다. 시맨틱 웹에 대해 우리가 오해하지 말아야 할 것 은 그것이 전 세계의 모든 데이터를 망라적으로 수용하려는 것은 아니라고 하 는 점이다.

시맨틱 웹은 디지털 세계에 존재하는 수많은 층위의 다양한 정보 가운데 그 것의 문맥 요소를 명시적으로 식별하고, 그 관계성을 정밀하게 탐색할 필요성 이 있는 것만을 대상으로 삼는다. 그리고 이 '필요성'이라는 것에 정해진 기준 이 있는 것은 아니기 때문에, 어떤 영역에서든 지식의 소통과 이를 위한 협업 이 필요한 곳에서는 그 분야의 필요성에 부응하는 온톨로지를 설계하고 이를 공유하는 방법으로 기존의 웹의 한 부분을 시맨틱 웹으로 만들어 갈 수 있다.



미래의 월드와이드웹: 기존의 웹과 시맨틱 웹의 공존

특정 조직 내의 데이터베이스 시스템과 달리 시맨틱 웹은 월드와이드웹이라 는 개방된 세계에서 만들어지는 것이며, 그 개방의 장점을 최대한 이끌어내는 의 도를 갖고 있기 때문에 어떠한 주제의 시맨틱 웹이든 그것을 이용하거나 데이

터의 생산에 참여할 수 있는 범위는 항상 열려 있다. 하지만 시맨틱 웹이 일종의 ‘약속’이라는 점을 주목하면, 그 약속을 만들고, 그 약속에 따라 데이터를 만들고, 그렇게 만들어진 데이터를 이용하는 사람들이 존재할 때만 그 세계가 의미를 갖게 된다는 점 또한 부인할 수 없다. ‘시맨틱 웹’이라는 개념으로 설명하는 ‘개방적인 데이터베이스’는 특정 기관이나 조직의 소유물은 아니되, 관심을 공유하는 사람들의 조직적인 노력 없이는 만들어질 수 없는 것이다.

인문학 분야의 학술적인 일에 종사하는 사람들이 ‘인문지식 시맨틱 웹’을 만들고자 한다면, 그 필요성에 공감하는 사람들이 공동의 약속을 정하고, 그들이 이제부터 만들어낼 ‘인문지식 데이터’ 속에 그 약속에 의한 메타데이터를 첨가하는 노력을 기울여 가야 한다.

그 노력은 한꺼번에 전체를 만들고자 하는 노력이 아니다. 관심이 모여지는 곳에서부터 부분적인 것을 만들되, 그 범위와 경계를 한정하지 않음으로써 향후에 더 확장될 수 있게 하고, 경우에 따라서는 전혀 다른 영역에서 만들어진 시맨틱 웹 데이터와도 소통할 수 있게 하는 것이다. 예를 들어, 우리나라 여러 지방의 마을 또는 가정에서 전승되고 있는 민간 의학에 관한 정보를 시맨틱 웹 데이터로 생산하는 일을 생각해 보자. 치료하고자 하는 병증, 약재로 쓰이는 식물, 약재를 가공하는 방법, 치료 효과에 대한 주장 등을 일정하게 기술할 수 있는 방법을 정하고 이를 사람들에게 알리면, 대학이나 공공기관의 연구팀이 아닌 개인 블로거들도 자신의 웹 페이지에 쓴 자기 마을의 민간 한방요법 이야기가 시맨틱 웹의 한 노드로 기능하게 할 수 있다. 그리고 시맨틱 웹 상의 다른 정보에 의해 그 블로그에 지방 사투리로 쓰인 약용식물의 이름에 상응하는 학명이 무엇인지 알려지면, 식물학, 화학, 제약학 영역에서 제공되는 정보를 블로거의 글과 연결해서 살펴보고, 그 신빙성을 검증하는 일도 가능해진다.

이러한 일들이 가능성에 대한 담론에 머물지 않고, 이루어질 수 있는 방향으로 나아가기 위해서는 어떠한 실천적인 노력이 필요할까? 나의 웹 문서에 쓰인 정보를 시맨틱 데이터로 기술하는 데 필요한 ‘이름’을 정하는 일이 우선되어야 한다. 바꿔 말해, 특정 주제의 시맨틱 웹을 위한 온톨로지를 설계하고 그 안에서 주어, 서술어, 목적어로 쓰일 수 있는 어휘들의 표준적인 이름을 제시함으로써 누구든지 그 어휘를 이용하여 ‘소통할 수 있는’ 데이터를 생산하게 한다는 것이다.

이 장의 첫 번째 절에서 살펴보았듯이, ‘함양 거연정’에 대해 이야기하는 여러 가지 웹 문서들은 ‘함양 거연정’이라는 건조물, ‘거연정기’라는 기록물, ‘임헌회’라는 인물 등 세 가지 개체의 고유한 이름을 매개로 서로 연결될 수 있었다. 두 번째 절에서는 ‘김현’이라는 인물과 ‘인문정보학의 모색’이라는 저서가

같은 방식으로 연결되는 구체적인 사례를 보였다. 그리고 세 번째 절에서는 이러한 종류의 개체에 더하여 장소, 연대, 개념어 등에도 고유한 이름을 부여함으로써 문화유산 안내문을 기계가독적인 텍스트로 전환해 보았다. 이 과정에서 다루어 본 몇 가지 범주와 그것에 속하는 개체들, 그리고 그 개체들 사이의 상관관계는 인문학 분야의 지식, 그 중에서도 한국의 전통문화 관련 지식의 정보화를 위한 데이터 모델의 핵심적인 구성 요소이다. 다시 말해, 그와 같은 범주에 속하는 중요 개체들이 표준적인 이름을 갖도록 하는 ‘한국 문화유산 온톨로지’가 누군가에 의해 만들어지고 많은 사람들이 이를 공유한다면, 우리나라의 디지털 정보세계 전체에서 통용되는 ‘문화유산 시맨틱 웹’이 만들어질 수 있다는 것이다. 그 일은 누가 해야 할까?

인문지식 시맨틱 웹의 구현이 교육이나 실험의 차원이 아니라 실효성 있는 지식 소통의 기능을 하기 위해서는 이 분야의 학술 진흥에 관한 공익적 책임을 지고 있는 기관들이 그 역할을 담당해야 한다. 시맨틱 웹의 노드와 링크가 될 요소들을 찾아내고 그것에 고유한 이름을 부여하는 일은 그 분야의 대상 지식을 분석할 수 있는 전문가라면 누구나 할 수 있다. 하지만, 그 일을 연구자들이 저마다 개인적인 차원에서 하려 한다면, 표준을 만드는 일은 점점 더 어려워질 것이다. 우리나라의 인문 분야 학술 연구기관들은 이미 인문지식 시맨틱 웹의 노드로 활용될 수 있는 문맥 요소 데이터를 작지 않은 규모의 데이터베이스로 구축해 놓고 있다.¹⁷⁾ 이 데이터를 기반으로 기본적인 인문지식 시맨틱 웹 온톨로지를 구축하는 것은 그 데이터를 축적해 온 공공기관들이 담당해야 할 과업이다.

한국사, 문화유산, 고전문학 등 주제 영역별로 통용될 수 있는 온톨로지가 만들어지고 누구나 이를 참조할 수 있게 되면 개인 연구자의 연구 활동이나 대학의 인문학 교육 과정에서 생산되는 다양한 데이터가 시맨틱 웹의 세계에 진입할 수 있는 길이 열리게 된다. 연구자들이 논문을 발표하거나, 교사가 수업 자료를 만들 때, 학생들이 리포트를 작성할 때, 심지어는 학계의 일원이 아

17) 예를 들어 한국학 분야의 정부 출연 연구기관인 한국학중앙연구원이 운영하고 있는 다음과 같은 데이터베이스로부터 ‘한국학 시맨틱 웹’의 기초자원이 될 수 있는 데이터(고유한 식별자를 가진 개체 목록 및 개체 사이의 관계를 기술하는 RDF 데이터)를 생산할 수 있다.

보유 자원: 데이터베이스	산출 가능한 시맨틱 웹 자원
역대인물정보	우리나라 전근대 인물의 식별자 / 혈연적·사회적 관계 정보
한국사연표	한국사의 주요 사건 식별자 / 사건의 전개와 관련된 시간 정보
한국향토문화전자대전	인물, 작품, 문화유산, 역사적 사건과 지리적 공간 사이의 관련성 정보
한국민족문화대백과사전	한국역사상의 주요 인물, 문화유산, 기록유산, 개념·용어 식별자 / 역사적 인물과 역사적 시공간의 관련성 정보

닌 누군가가 문화유적지를 돌아보며 기행문을 남길 때에도 자신이 지금 웹에 남기는 흔적이 #어떤 인물, #어떤 사건, #어느 장소, #어느 문헌, #어떤 개념과 관계가 있는지, 온톨로지 참조를 통해 명시적으로 밝힐 수 있게 되는 것이다.

이렇게 만들어진 시맨틱 웹 데이터는 월드와이드웹 상에서 어떻게 활용될까? 앞의 스파클(SPARQL) 사용 예시에서 보았듯이, 웹상에서 단어를 검색하는 데 머무는 것이 아니라 사실과 사실간의 관계를 추론하여 새로운 지식을 얻는 일이 가능해질 것이다.

시맨틱 웹 데이터를 집적하여 사실의 추론이 가능한 데이터 검색 서비스를 제공하는 역할은 미래의 '시맨틱 웹 포털'이 담당할 역할이다. 시맨틱 웹 포털은 국가가 운영하는 공공 데이터 포털¹⁸⁾이나 민간 상업 포털이 기능 확장을 통해 구현할 수도 있고, 특정 지식 영역의 커뮤니티에서 온톨로지 기반 데이터베이스에 관심 주제의 시맨틱 데이터를 집적하는 방법으로 구현할 수도 있다.

한국 문화 연구자: 문화유산 지식 온톨로지를 활용하여 역사, 지리, 교육문화, 예술 등 분야별 지식 자원의 연계 데이터(Linked Data)를 생산

정보 서비스 포털: 한국 문화 연구자들이 생산한 데이터가 실제적인 시맨틱 웹 서비스로 이어지도록 기술적 환경 제공



공공기관(문화재청, 한국학중앙연구원, 국립중앙박물관, 국립중앙도서관, 고전번역원 등): '한국 문화유산 지식 네트워크'의 '노드'(Node)와 '링크'(Link)를 만드는 데 필요한 어휘 자원의 데이터베이스(온톨로지)를 구축하고 활용 모델 제시

한국 문화유산 시맨틱 웹의 구현을 위한 협력 구도

시맨틱 웹 포털의 기술적인 환경을 구축하는 일은 정보기술 전문가들이 주도적인 역할을 담당할 것이다. 이 영역의 일은 기술 표준이 확립되어 가고 있고, 또 성공적인 선행 모델이 있기 때문에 언제든지 필요하면 만들어질 수 있는 것이라고 해도 무방하다. 반면, 인문학적 내용을 담은 시맨틱 웹 데이터의

18) <http://www.data.go.kr>

생산은 인문학을 중심에 두는 디지털 인문학의 영역에서 만들어져야 한다. 디지털 세계에서 다양한 인문지식 자원들이 의미의 맥락을 좇아 무한히 이어지는, 그 문맥 속에서 새로운 사실을 발견하기도 하고, 알려진 사실의 신뢰성을 검증할 수도 있는 인문지식 시맨틱 웹의 구현은 그 네트워크의 노드와 링크 하나하나를 정밀하게 선택하고 다듬는 노력 위에서 만들어진다. 디지털 인문학의 역할은 인문지식 데이터에 대해 바로 그와 같은 실천적 노력을 투입하는 것이다.