



웹! 그 흐름에 대한 이해  
시맨틱 웹

---

<b>웹! 그 흐름에 대한 이해</b>	<b>1</b>
<b>웹의 역사</b>	<b>3</b>
인터넷 = 웹?	3
최초의 웹 주소 info.cern.ch	3
웹의 기본적인 개념	7
<b>시맨틱(뜻살린) 웹</b>	<b>9</b>
공개와 공유는 웹의 기본 철학이다	9
웹의 미래상	12
뜻 살린 웹	13
RSS와 숨은자료metadata	15
<b>시맨틱 웹의 진행과정</b>	<b>18</b>
XML을 기반으로 한 시맨틱 웹의 구조	18
온톨로지(존재론)로 인공지능의 기초를 다진다	18
숨은자료metadata의 양과 정보의 질은 비례 관계이다	20
시맨틱웹의 기반 기술과 현재 진행상황	20

## 웹의 역사

### 인터넷 = 웹?

인터넷과 웹은 구별된다. 인터넷은 컴퓨터들을 이용해서 여러 가지 정보들을 주고 받는 통신 수단이다. 반면에 웹은 그런 인터넷을 사용하는 방법과 같은 것이다. 가령 그림을 그린다면 캔버스는 인터넷에 해당되고 웹은 그 위에 그려진 그림의 양식이나 내용을 말한다. 웹은 메일서비스, 원격접속서비스 등등의 인터넷을 통해서 제공되는 많은 서비스 중에 하나일 뿐이다. 하지만 그 서비스 중에서 가장 강력한 영향력을 가지고 있다. 사람들이 “인터넷=웹”이라고 착각한다는 사실은 웹이 얼마나 영향력이 있는지를 시사한다. 그럼 지금부터 그 웹에 대해서 알아보자.



### 최초의 웹 주소 [info.cern.ch](http://info.cern.ch)

우리가 보통 말하는 “웹”은 World Wide Web을 줄여 말하는 것이다. 흔히 WWW 나 W3로 부르기도 한다. WWW는 유럽입자물리연구소(CERN: European Organization for Nuclear Research)에서 처음 시작되었다. 매일 얻어지는 엄청난 양의 연구결과 및 자료의 효율적인 공유를 목적으로 팀 버너스 리의 제안으로 연구가 시작되어 개발되었다.



CERN에서 근무하던 팀 버너스 리는 ‘정보의 연결’에 관심이 많았다. 연구시간에 짬을 내어서 그 머리 속의 작업을 실천에 옮기게 되었고 1989년 3월 오늘날의 웹의 설계도가 담긴 보고서를 제출하게 된다. 내용은 하이퍼 텍스트를 통한 정보의 효율적인 관리였다. 바로 이것이 웹의 시초이다. 처음 올린 보고서는 채택되지 않았다. 그는 웹을 개발하기 위해 로버트 카일리아우(Robert Cailliau)라는 네트워크 전문가의 도움을 받아 1990년에 네트워크 보고서를 다시 올린다. 역시 처음에는 채택되지 않았으나 마침 넥스트에서 만든 컴퓨터인 'NEXT Cube'<sup>1</sup>가 CERN에 들어오면서 네트워크 실험에 대한 허락이 떨어진다.



그림 1. 웹의 창시자, 팀 버너스 리

<sup>1</sup> NeXT Cube 란 NeXT사에서 만든 컴퓨터를 말한다. 위 그림에서 보이는 컴퓨터가 바로 NeXT Cube 이다. 이 컴퓨터에서 World Wide Web이 개발되었다.

Vague but exciting ...

CERN DD/OC  
Information Management: A Proposal

Tim Berners-Lee, CERN/DD  
March 1989

## Information Management: A Proposal

### Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control

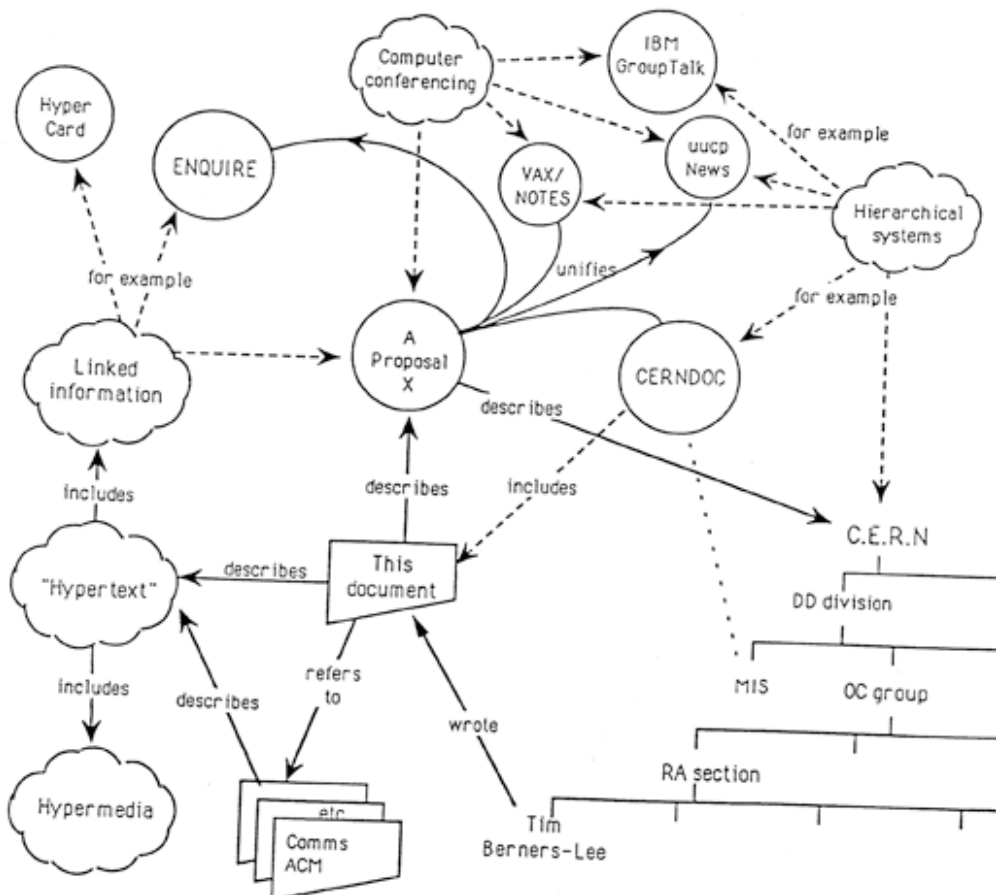


그림 2. 팀 버너스 리가 제출한 보고서

그는 그것의 이름을 WWW(World Wide Web)으로 결정하고 1990년 10월부터 웹 코딩을 시작해 HTML 문서와 HTTP 프로토콜, URI 등을 만들었다. 1990년 11월에는 클라이언트 프로그램(웹 브라우저)을 만들고 12월에는 브라우저로 HTML 문서를 연결했다. 역시 같은 달인 12월에 최초의 웹주소인 'info.cern.ch'를 만들고 여러 시스템에서 'info.cern.ch'의 접속에 성공함으로써 마침내 웹 시대를 열게 되었다. 1990년 웹의 탄생이 세상의 삶의 패러다임을 바꾸어 놓았다. 지금은 'info.cern.ch'에 접속하면 웹의 탄생에 대한 자세한 설명이 나온다.



그림 3. 최초의 웹사이트인 info.cern.ch 는 오늘날에도 접속이 가능하다.

팀 버너스 리는 그 이후에도 계속 웹을 발전시켰다. CERN에서 사용할 수 있는 웹을 개발하는 한편 학술대회에 웹을 공개하면서 세상에 알리기 시작했다. 1994년 10월 MIT에서 W3C<sup>2</sup>를 출범하면서 본격적인 웹의 시대를 시작하였다. 그 후 웹이 급속도로 발전하여 오늘날에 이르러 전세계를 묶어주는 공간으로 발전하였다

<sup>2</sup> World Wide Web Consortium 의 약자이다. WWW의 각종 표준을 정의하고 웹에 대한 새로운 연구를 하는 등 다양한 활동을 하는 기관이다.

---

결론적으로, WWW 세계 규모의 거대한 정보 망, 공간을 뜻한다. 팀 버너스 리가 제안한 웹은 이전의 인터넷 서비스와 확연한 다른 점이 있었고, 이것이 바로 웹이 전세계인을 사로잡은 이유이다. 가장 확연히 다른 점은 하이퍼텍스트가 있다는 점인데, 이것을 통해서 연결된 다른 정보에 대해서 쉽게 접근할 수 있었다. 단순히 빗빗하게 글만 있는 것과는 달리 하이퍼링크를 통해서 전세계의 모든 정보가 하나로 묶여 상호작용을 할 수가 있었던 것이다. 웹의 주요 특징에 대해서는 다음 장에서 자세히 알아보자.

---

## 웹의 기본적인 개념

웹은 기본적으로 4가지 기본적인 아이디어의 조합이다:

**하이퍼텍스트:** 정보의 형식중 하나로써, 하나의 문서에서 다른 문서로 내부적인 연결(하이퍼링크라 불린다)을 통해서 이동할 수 있는 것.

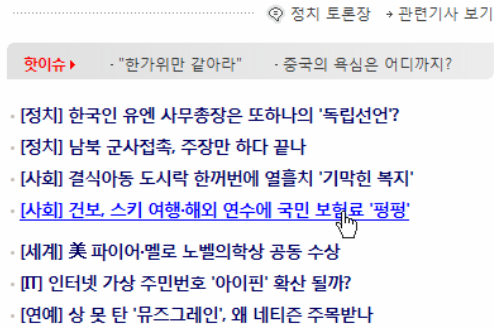


그림 4. 우리는 하이퍼링크를 통해서 다른 페이지로 이동한다.

**Universal Resource Identifiers(URI)**<sup>3</sup>: 네트워크에 존재하는 특정 자원(문서, 파일, 디스크 등의 자원)을 찾을 수 있도록 하는 고유한 식별자.

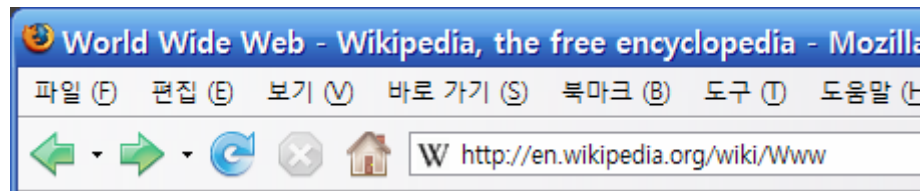


그림 5. 위키백과 wiki 안 Www 라는 문서를 접근할 때의 URI. 우리는 이를 사용해서 다양한 자료를 접근할 수 있다.

---

<sup>3</sup> 널리 쓰이는 용어로는 URL(Uniform Resource Locator)가 있다. 하지만 URL은 URI에 포함되는 개념이다.

**마크업 언어:** HTML 등과 같이 문서의 의미적인 정보를 저장하는 형태의 언어.



그림 6. 우리는 이러한 복잡한 문법 덕분에 보기 편한 웹 페이지를 얻을 수 있다.<sup>4</sup>

**서버-클라이언트 구조<sup>5</sup>:** 서비스 받는 사람과 서비스를 제공하는 사람이 구분되어 있다. 예를 들어서 우리는 네이버 지식인 서비스를 이용하는 클라이언트이고, 네이버는 서버가 된다.

<sup>4</sup> 이러한 문법을 HTML(Hyper Text Markup Language) 이라고 한다. 하이퍼텍스트를 사용해 꾸미는 언어라는 뜻이다. 여기서 꾸민다는 뜻은 사용자에게 보이는 정보 외에 사용자에게 보이지는 않지만 컴퓨터에 의해서 이해되는 다양한 정보를 담는 것을 뜻한다.

<sup>5</sup> 서버-클라이언트 구조에 대한 자세한 내용은 다음 자료를 참고하자:

[http://talent.kaist.ac.kr/bbs/zboard.php?id=board\\_study&no=142](http://talent.kaist.ac.kr/bbs/zboard.php?id=board_study&no=142)

[http://talent.kaist.ac.kr/bbs/zboard.php?id=board\\_study&no=154](http://talent.kaist.ac.kr/bbs/zboard.php?id=board_study&no=154)

[http://talent.kaist.ac.kr/bbs/zboard.php?id=board\\_study&no=202](http://talent.kaist.ac.kr/bbs/zboard.php?id=board_study&no=202)

---

## 시맨틱(뜻살린) 웹

### 공개와 공유는 웹의 기본 철학이다

우리는 인터넷을 하면서 수많은 정보를 받아들이고, 생산하면서 생활하고 있다. 웹의 창시자인 팀 버너스 리가 웹을 만들면서 꿈꿔왔던 세상은 ‘정보의 공개와 공유를 통한 인간의 행복을 추구하는 세상’ 이라고 할 수 있다. 그가 추구한 웹의 기본적인 원칙은 언제 어디서나 정보에 접근할 수 있는 완벽한 정보의 연결 이었다. 이러한 철학은 웹의 근본적인 4가지 아이디어에 잘 나타나 있다.

웹의 이러한 철학을 잘 따르는 곳 중에서 위키백과Wikipedia란 곳이 있다. 이곳은 기본적으로 백과사전을 모토로 하는 사이트이다. 즉, 인터넷 판 백과사전 이라고 할 수 있는데, 일반적인 백과사전이 출판사와 수많은 전문가들이 모여 작성한 것이라면 위키백과는 인터넷 접속이 가능한 수만 명의 사람들이 모여 작성한 것이다. 위키백과는 참여하는 사람들에 특별한 제한을 두고 있지 않기 때문에 수많은 사람들이 참여하고, 한번 만들고 출판되는 백과사전과는 달리 항상 최신 정보로 업데이트 될 수 있다는 장점을 가지고 있다. 위키백과 프로젝트는 불과 시작한지 5년 만에 150만 페이지에 달하는 방대한 분량의 정보가 모아졌고, 현재에도 수많은 사람들이 페이지를 생성하고 수정하고 있다.

위키백과를 보면 페이지를 찾기 위한 URI 가 체계적으로 정리 돼있고, 각 페이지마다 다른 페이지로 연결되는 주요 단어들이 링크로 연결되어 있고, 필요에 따라 강조 해야 할 단어들은 굵게 표시가 되어있는 등 HTML을 적극적으로 잘 사용하고 있는 것을 알 수 있다. 이러한 이유 말고도 위키백과가 웹의 철학을 잘 반영했다고 평가 받는 주 이유중 하나는 바로 사용자 참여가 가능한 읽기/쓰기 웹Read/Write Web<sup>6</sup>이라는 것이다.

---

<sup>6</sup> Read/Write Web 이란 웹의 근본적인 철학 중 하나이다. <http://www.readwriteweb.com/> 참고



그림 7. WWW의 취지에 잘 부합되는 위키백과

이러한 웹의 철학을 이용한 것 중에서 검색엔진이란 것이 있다. 우리는 이미 네이버나 구글과 같은 검색엔진을 매일 사용하고 있다. 검색엔진은 인터넷의 정보를 쉽게 찾아갈 수 있도록 해주는 유용한 도구이다. 검색엔진에서 우리는 원하는 정보를 키워드로 입력하면 검색된 결과를 링크를 통해서 찾아갈 수 있다. 링크를 위해 필요한 두 가지 전제 조건은 공개와 공유다. 공개되지 않은 정보에 연결할 방법은 누구라도 가지고 있지 않다. 그래서 공개 정신과 공유 정신은 웹의 기본 정식이자 웹 사용자에게 요구하는 기본 조건이다. 그러나 한국에서는 남의 사이트에 있는 문서와 DB를 긁어와 보여주는 것으로 돈을 번 검색 포털들이 정작 자신들의 지식 DB를 검색하려는 검색엔진은 거부하는 상황이 벌어지고 있다.

국내 기업의 인터넷 정신 부재를 보여주는 가장 단적인 예로 2005년 중순에 발생한 엠파스의 열린검색과 네이버의 지식인 서비스 논쟁을 들 수 있다. 엠파스가 열린검색 서비스를 시작하면서 네이버의 지식인 게시물을 검색 가능하게 하자 네이버는 엠파스의 열린 검색을 비난했다. 네이버 자신은 남의 사이트에 있는 질문답변 게시물을 검색해 보여주는 것으로 성장했으면서 엠파스가 네이버의 질문답변 게시판인 지식인의 내용을 검색하는 것을 비난하는 사태가 벌어진 것이다.

엠파스가 지속적으로 열린검색 서비스에서 네이버의 지식인 게시물을 포함시키자, 네이버는 시간 별로 지식인 게시물의 링크를 변화시키며 링크가 불가능 하도록 대응했다. 이로 인해서 네이버 지식인에 링크를 걸었던 과거 웹 페이지들은 모두 쓸모 없게 되어버리고 말았다. 웹 정신에 충실하고 웹 자원을 아끼기 위해 걸어진 링크가 모조리 무용지물이 된 것이다.

링크가 안 되는 웹 사이트 게시물을 일부러 만든 곳이 국내 1위 포털이니 부끄러운 일이다.<sup>7</sup>

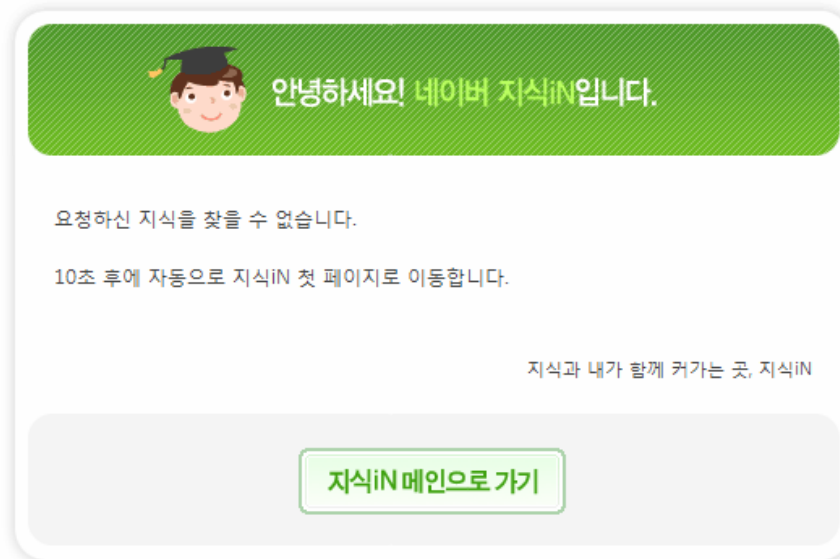


그림 8. 기존 링크가 모두 깨져 볼 수 없는 게시물. 링크를 걸 수 없는 사태를 만들었다.

정보 관련 업체는 공개와 공유 혜택을 가장 많이 받으며 성장한다. 인터넷 성공을 이야기 할 때 가장 먼저 이야기하는 기업 하면 야후, 아마존, 이베이다. 그런데 이들 기업은 정보를 공개, 공유하는 기술을 개발함으로써 성공했다. 공개와 공유 혜택을 받아 성장한 기업은 다시 공개와 공유를 개발해야 할 것이다. 위 사례처럼 공개와 공유를 막고 이기적으로 자기 자신만 정보를 간직하려 한다면 그것은 인터넷의 철학에 반하는 것이고, 바람직하지 못하다고 할 수 있다.

<sup>7</sup> 김중태(2006), 『웹 2.0 시대의 기회, 시맨틱 웹』, 디지털미디어리서치, 45~55면

## 웹의 미래상

몇 십 년 뒤의 세상을 상상해 보자. 이미 그 먼 미래에는 가벼운 컴퓨터가 우리와 상시 결합되어 있다. 웹에 연결되어 있는 것은 물론이다. 친구와 영화를 보고 싶어 컴퓨터에 음성인식 기능으로 컴퓨터에게 최근 재미있는 영화들을 보여달라고 한다. 컴퓨터가 여러 영화관의 웹 서버에 접속해 정보를 받아온다. 데이터들은 모두 컴퓨터의 인공지능이 이해할 수 있는 간결한 형태로 제공된다. 이를 이해한 컴퓨터는 자기의 주인이 그 동안 선호했던 영화들을 토대로 하나를 추천한다. 주인은 그 영화 추천에 흡족하고 친구의 일정과 자신의 일정을 각각 알아본 뒤 빈 시간을 서로 잡아 약속을 한다. 컴퓨터는 영화 예매 사이트를 접속해 이 시간대의 영화 표를 자동 예매한다.



그림 9. 미래엔 영화 표 하나를 예매하기 위해 이런 광고로 가득한 웹 페이지는 볼 일이 없을 것이다. 우리는 영화를 예매하려는 것이지 광고를 보러 온 것이 아니지 않는가? 미래에는 이러한 무식한 광고가 사라지고 자신이 즐겨보는 영화들을 바탕으로 새로운 영화를 추천 받는 세상이 올 것이다.

영화의 한 장면 같은 이런 미래상은 그리 멀지 않았다. 이미 몇 십 년 전부터 이러한 세상을 만들기 위해 웹을 향상 시키려는 연구가 계속되었다. 그 연구의 중심이 바로 ‘시맨틱 웹’이다. 시맨틱의 뜻을 사전에서 찾아보자.

**se-man-tic** [simæntik] a.

### 1 의미의

▶ semantic change 의미 변화

### 2 의미론의

▶ a semantic approach 의미론적인 접근

시맨틱 웹에서의 시맨틱의 뜻은 바로 의미론적인 접근<sup>8</sup>을 뜻한다. 즉, 좀 더 쉬운 말로 하면 뜻을 살린 웹을 뜻한다. 뜻 살린 웹이란 무엇일까? 이것에 대해서 좀 더 알아보자.

## 뜻 살린 웹

블로그를 예를 들어 시맨틱 웹이 무엇인지 알아보자.

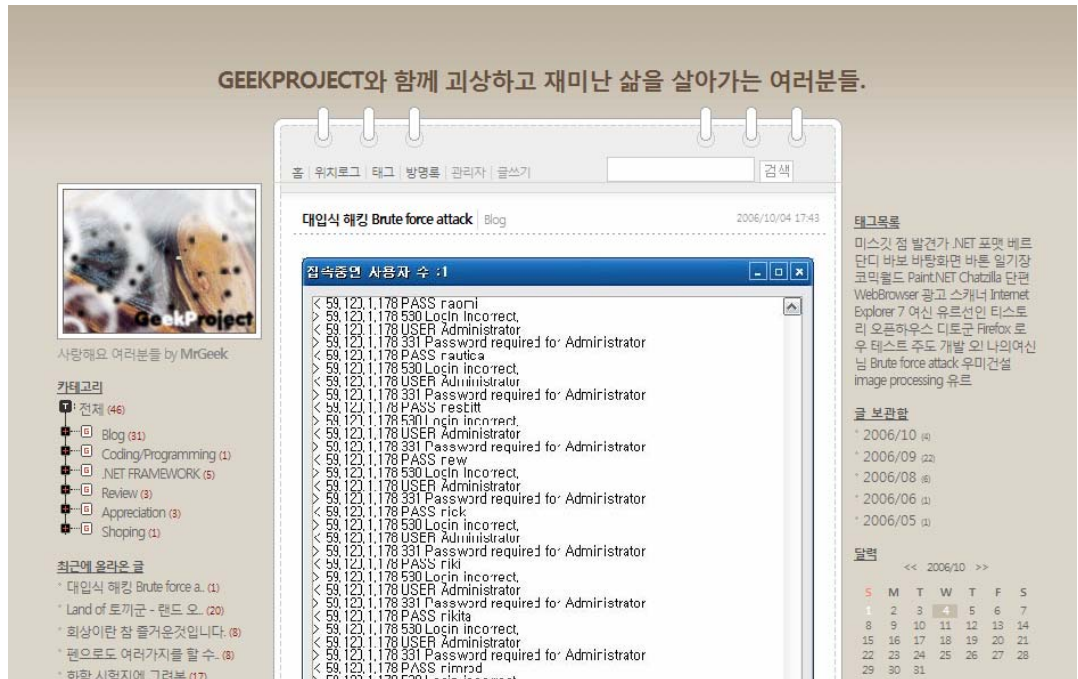


그림 10. 우리는 본문이 어디 있는지를 이해 하지만, 컴퓨터는 하지 못한다.

위와 같은 블로그<sup>9</sup>에서 우리는 본문이 본문인지를 어떻게 알까? 장식용으로 제작된 이미지와 실제 블로그 내용에 포함되는 이미지는 어떻게 구분이 될까? 왼쪽에 있는 카테고리를 컴퓨터에서 인식할 수 있을까? 현재의 웹에서 그러한 웹 페이지에 포함된 문서의 진정한 뜻을 이해할 수 있는 컴퓨터는 없다. 왜냐하면 근본적으로 위 웹 블로그의 내부 문서인 HTML에서 그러한 설명을 전혀 하지 않기 때문이다.

위 블로그의 소스<sup>10</sup>를 분석해 보자.

<sup>8</sup> 보다 정확한 뜻의 이해는 다음 사이트를 참고해 보자: <http://ko.wikipedia.org/wiki/의미론>

<sup>9</sup> 블로그란 Web과 Log의 합성어로 인터넷에 공개된 개인 페이지, 혹은 공개 일기장을 뜻한다

<sup>10</sup> 여기서 소스란 실제 이 페이지의 내부 구조를 뜻한다. 즉, 컴퓨터가 이해하는 문법을 소스라고 한다. 웹 페이지에서 소스는 앞서 설명한 HTML로 이루어져 있다.

```
소스:http://geekproject.net/ - Mozilla Firefox
파일(F) 편집(E) 보기(V) 도움말(H)
<!-- 공지사항 -->

<!-- 보호글 -->

<!-- 본문글 -->
<div class="entry">
  <!-- 글 제목, 카테고리, 글쓴 날짜 -->
  <div class="titleinfo">
    <h2><a href="/entry/대입식-해킹-Brute-force-attack">대입식 해킹 Brute force attack</a>
    <div class="name"><a href="/category/Blog">Blog</a></div>
    <div class="date">2006/10/04 17:43</div>
  </div>
  <!-- 본문 -->
  <div class="article"><div class="imageblock center" style="text-align: center; clear: both;
FTP서버에 저런식으로<br />
거들<br />
ack"을 걸어왔다.<br />
이다.<br />
려고 하다가<br />
올린다.<br />
force attack]]<br />
wikipedia.org/wiki/Brute_force_attack">http://en.wikipedia.org/wiki/Brute_force_attack</a><br />
스워드같은 평어문이라면 몰라도 "할리우드 보안체계" 라면.</font></div>
  <!-- 본문 관련 태그 -->
  <div class="tag">
     <a href="/tag/Brute%
```

그림 11. 컴퓨터가 이해하는 페이지 소스에서 본문을 따로 구별할만한 정보가 없다.

위 소스에서 본문은 무엇일까? 상식적으로 생각해 본다면 <!--본문--> 이라고 쓰여져 있는 부분에 본문이 있을 것이라고 생각할 수 있다. 하지만 다른 블로그도 이와 같은 구조를 가지고 있을까? 그렇지 않을 것이다. 블로그들은 종류가 수없이 많고 HTML 소스의 형태도 다른 구조로 되어 있을 것이다. 컴퓨터가 알아서 이러한 내용들을 인식하고 처리하는 것은 불가능에 가까울 것이다.



---

그 안에 본문이 예쁘게 저장되어 있다. 블로그의 내용을 꾸미기 위한 소스는 전혀 없다. 오직 내용만 잘 정리된 형태로 제공되고 있는 것이다. 만약에 이러한 형태를 모든 블로그들이 지키고 있다면 컴퓨터가 쉽게 처리할 수 있지 않을까?

위에서 title 태그 안에 있는 내용이 제목이라고 우리는 짐작할 수 있다. 이렇듯 실제 블로그 글의 타이틀인 ‘대입식 해킹.....’ 을 title 태그라는 숨겨진 정보metadata로 감싸 표현 함으로써 컴퓨터는 이 문장이 블로그 글의 제목이라는 것을 알 수 있게 되는 것이다.

위와 같은 내용만을 담아 컴퓨터가 쉽게 블로그의 내용을 읽을 수 있도록 하는 표준을 RSS<sup>11</sup> 라고 한다. RSS 표준은 사이트의 정보를 표현하기 위해서 숨겨진정보metadata를 어떠한 키워드(title, description 따위의 이름)으로 표현할 것인지 표준으로 약속한 것을 말한다. RSS는 사이트의 정보를 출판(Syndicate) 하는 역할을 하는 표준을 정한 것이다. 이러한 표준은 이미 블로거들에 의해 대중화가 되어 많이 쓰이고 있고, 이제는 각종 뉴스 웹 사이트나 야후, 구글등에서 적극적으로 자신들의 정보를 공개하기 위해서 적극적으로 도입 중이다. 더불어 이러한 공개된 RSS 소스들을 읽는 프로그램들<sup>12</sup>도 수없이 등장했다.

---

<sup>11</sup> RSS는 Really Simple Syndication, RDF Site Summary, Rich Site Summary 등의 약자이다. RSS 가 이와 같이 많은 의미를 가지는 이유는 다양한 그룹에서 RSS 라는 이름으로 자기 자신들만의 표준을 만들었기 때문이다. 자세한 것은 위키백과의 글을 읽어보자:

[http://en.wikipedia.org/wiki/RSS\\_\(protocol\)](http://en.wikipedia.org/wiki/RSS_(protocol))

<sup>12</sup> RSS Aggregator, RSS 리더라고 불린다. 대표적으로 Google Reader 나 HanRSS 등이 있다.

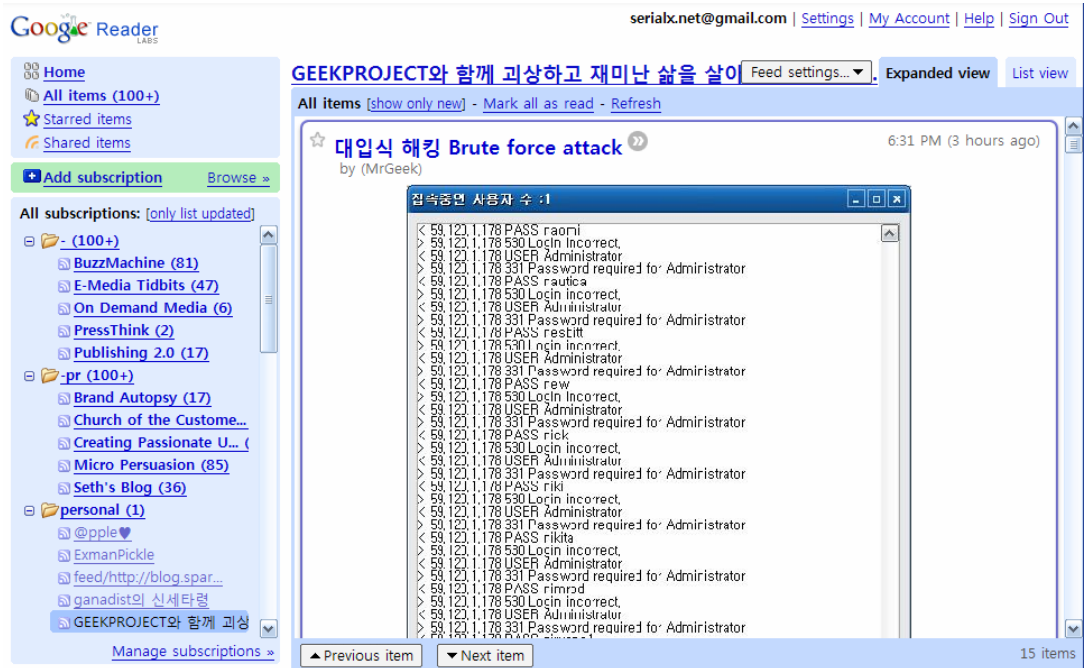


그림 14. 대표적인 웹 RSS 리더 Google Reader. 우리는 표준화된 RSS 와 이를 이용한 RSS 리더를 통해서 실제로 그 사이트에 접속하지 않더라도 편하게 새로 올라온 글을 볼 수 있다.

위 화면을 보면 웹 브라우저로 접속했던 블로그의 내용만 따로 Google Reader 에서 볼 수 있음을 알 수 있다. 또한 블로그에 글이 새로 올라오면 사용자에게 알려주는 등 편한 기능들이 있다.

---

## 시맨틱 웹의 진행과정

### XML을 기반으로 한 시맨틱 웹의 구조

시맨틱 웹은 기본적으로 어떤 형태로 이루어질까? RSS는 무엇으로 이루어 졌는지 부터 알아보면 되겠다. RSS는 XML<sup>13</sup>로 이루어져 있다. XML 은 다양한 데이터를 표현하기 위한 좋은 수단중 하나로 꼽힌다. 그 이유는 XML 의 견고함과 확장성 때문이다. XML로 우리 세상에 있는 다양한 것들을 표현할 수 있고, 컴퓨터가 처리하기 매우 좋은 구조를 가지고 있기 때문에 시맨틱 웹의 표현 수단으로써 선호되고 있다.

RSS를 보기만 하더라도 인간이 이해할 수 있는 형태로 명쾌한 구조로 되어 있으면서도(인간이 보기 좋다는 뜻은 아니다) 컴퓨터도 처리하기 좋은 형태로 존재하고 있음을 알 수 있다. XML을 다루는 것 자체는 이 글에서 다루는 부분이 아니므로 관심이 있다면 관련 XML 자료를 보는 것이 좋겠다.

### 온톨로지(존재론)로 인공지능의 기초를 다진다<sup>14</sup>

앞서 말했던 영화 예약에 관한 이야기는 멀지 않은 미래의 모습이다. RSS 나 XML 같은 내용들은 이미 수 년 전부터 연구되고 표준화 된 것들이고 현재도 시맨틱 웹을 위한 연구와 표준화는 지속적으로 이루어지고 있다. 그 중에서 중요한 것 중 하나가 바로 온톨로지 이다.

컴퓨터에서 온톨로지는 존재론으로 번역 되는데, 컴퓨터에서 사용하는 각종 개념에 대한 설명 또는 정의라고 볼 수 있다. 철학에서 사용하는 존재론을 컴퓨터에서 사용하는 이유는 무엇일까? 뇌가 우주의 축소판이고, 컴퓨터가 뇌의 축소판이기 때문이다.

일반 분야의 온톨로지와 컴퓨터 분야에서 사용하는 온톨로지가 다른 점이 있다면 기계가 이해할 수 있느냐 없느냐 하는 차이이다. 인공지능 분야에서 오래 전부터 논의되던 온톨로지가 시맨틱웹 분야에서 새삼 중요한 부분으로 떠오른 이유는 기계들이 낱말의 뜻을 이해할 수 있도록 하는 것이 시맨틱웹이기 때문이다. 사람은 어떤 개념을 쉽게 이해할 수 있지만 기계가 이를 이해하기란 쉽지 않다. 따라서 기계가 이해할 수 있도록 매우 구체적으로 개념을 설명하고 관계를 표현하는 기술이 필요한데, 이것이 온톨로지 기술에 의해 이루어지는 것이다.

온톨로지가 일종의 사전이고 기계가 이 사전의 의미를 파악하는 것이 핵심이므로 온톨로지

---

<sup>13</sup> XML은 Extensible Markup Language의 약자이다. HTML을 포함하는 개념으로, 웹 페이지를 표시하는 것뿐만이 아니라 다양한 데이터를 표현할 수 있는 언어이다.

<sup>14</sup> 김중태(2006), 『웹 2.0 시대의 기회, 시맨틱 웹』, 디지털미디어리서치, 90~100면

는 필연적으로 인공지능 기술에 의지할 수밖에 없다. 현재 컴퓨터 분야의 인공지능과 인공지능 프로그램은 ‘근대적 의미론의 창시자’로 부르는 독일 철학자인 고트롭 프레게가 개발한 술어 논리predicate logic 체계에 크게 의한다. 이후 알란 튜링과 존 맥카시에 의해 계승된 인공지능은 지속적인 연구를 통해 시맨틱웹 개발의 핵심으로 떠올랐다.

시맨틱웹에서 사용하는 온톨로지의 정의는 어떤 낱말에 대한 뜻(개념)과 각 낱말 사이의 관계를 잘 설명한 것을 뜻한다. 우리가 ‘사과’라고 말했는데 어떤 사람은 이를 먹는 사과로 이해하고, 어떤 사람은 ‘미안하다’라는 뜻으로 이해할 경우 혼란이 온다. 그래서 낱말의 정의는 매우 중요하다.

온톨로지는 낱말에 대한 정의와 개념을 명확하게 해주는 학문이다. 예로 ‘아버지’라는 낱말을 설명할 때 ‘남자, 기혼자, 결혼해 아이를 낳은 남자’라는 뜻 설명이 필요할 것이고, 이때 사용하는 ‘남자’ ‘결혼’ ‘아이’와 ‘아버지’라는 낱말의 구조와 상호 관계를 설명해둔 내용이 온톨로지인 것이다.

만약 아버지를 ‘아이가 있는 남자’로만 설명하고 아이와의 관계만 설명해준다면 컴퓨터는 아버지에서 ‘기혼’을 추론하지 못할 것이고 기혼자를 위한 추천 상품을 내놓지 못할 것이다. 이렇듯 관계를 잘 설명해 놓는다면 인공지능 프로그램은 이를 토대로 유용한 일들을 사람에게 해줄 수 있을 것이다.

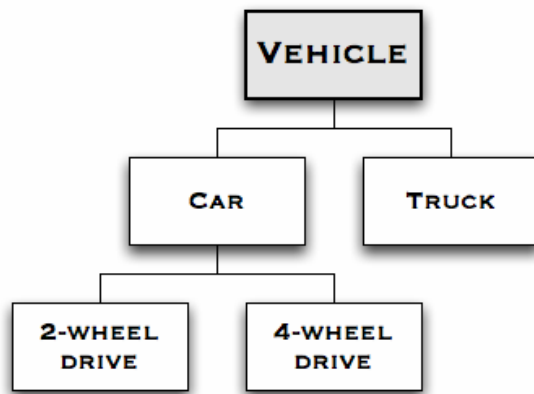


그림 15. 2륜 구동 차와 4륜 구동 차는 차에 속하고 차와 트럭은 탈것에 속한다. 이러한 정의를 하는 것이 온톨로지이다.

온톨로지에서는 개념을 얼마나 잘 정의하느냐도 중요하지만 이렇게 설명한 내용이 다른 시스템에서도 똑같이 이해되도록 하는 것이 중요하다. A 시스템에서 구축한 온톨로지가 B 시

---

시스템에서 다른 의미로 받아들여진다면 A 와 B 시스템 사이의 자동화는 물 건너 가는 셈이다. 따라서 전 세계적으로 사용이 가능한 온톨로지 구축과 이를 통한 인공지능 처리 기술에 시맨틱웹 관계자들이 주력을 삼고 노력하고 있다.

#### 숨은자료metadata의 양과 정보의 질은 비례 관계이다

시맨틱 웹은 숨은자료metadata의 양을 늘리고, 또한 질적인 향상(온톨로지와 함께)을 추구하는 것이다. 온톨로지에서 개념을 정의하는 것 자체도 숨은자료라고 말할 수 있기 때문에 숨은자료의 존재는 필수적이라고 할 수 있다.

숨은자료가 많다는 것은 그 자료에 대해서 자동화가 많이 이루어질 수 있다는 뜻이기도 하다. 예를 들어서 간단한 HTML 문서를 보면, 우리는 어떤 글을 강조하고 싶어 <strong> 태그<sup>15</sup>를 삽입한다. 그렇게 되면 웹 브라우저는 이 태그를 인식해 <strong> 태그에 둘러 싸여진 글을 굵게 표시하게 된다. 사용자에게는 보이지 않지만 컴퓨터가 이해하는 정보가 바로 숨은자료이다.

#### 시맨틱웹의 기반 기술과 현재 진행상황

현재 온톨로지 구축은 매우 더디게 진행되고 있다. 이는 온톨로지용 언어인 OWL이 2004년에 겨우 W3C 에서 제정된 것으로 알 수 있다. 실질적으로 온톨로지를 응용한 시맨틱웹 구현은 이제 겨우 시작 단계에 불과하며 갈 길이 멀다. 현재 온톨로지의 활용 분야로 지식관리Knowledge Management 와 웹 상거래Web Commerce가 주요 분야로 지목되고 있으며, 실제로 지식관리와 전자상거래 분야에서 온톨로지의 활용을 활발하게 연구하고 있다.

현재 시맨틱웹 분야에서 진행되고 있는 것은 RDF를 이용한 RSS 기술이나 꼬리표, XML을 이용한 응용 프로그램, Ajax 기술 정도로 매우 저차원적인 응용에 그치고 있다. 물론 그 저차원적인 응용마저도 혁명적인 변화를 가져오고 있다는 사실을 볼 때 이와 같은 온톨로지의 개념이 지금 전 세계적으로 퍼져있는 태그와 같이 보편화 된다면 그 위력은 더욱 클 것임을 쉽게 이해할 수 있다.

---

<sup>15</sup> HTML에서 Strong태그는 표면적으로 B태그와 동일한 기능을 수행한다. 하지만 B 는 단순히 이 문장을 굵게 표시하라는 의미만을 가지고 있지만 Strong태그는 이 안에 있는 것을 강조하라는 의미를 가지고 있으므로 의미를 더 포함한다고 할 수 있겠다.

---

## 참고문헌

- [1] 김중태(2006), 『웹 2.0 시대의 기회, 시맨틱 웹』, 디지털미디어리서치
- [2] 김태우(2006), 「태우's Completely Incomplete Guide to Web 2.0」  
[http://twlog.net/web\\_2.0\\_guide.html](http://twlog.net/web_2.0_guide.html)
- [3] 웹의 역사, 2006.10.25  
<http://info.cern.ch>
- [4] 위키백과(<http://wikipedia.org>) 문서, 2006.10.29  
<http://en.wikipedia.org/wiki/Www>  
<http://en.wikipedia.org/wiki/Html>  
<http://en.wikipedia.org/wiki/Semantics>  
[http://en.wikipedia.org/wiki/Semantic\\_web](http://en.wikipedia.org/wiki/Semantic_web)
- [5] 주영재, “주영재 블로그”, 2006.10.30.  
<http://geekproject.net/>
- [6] “Web 2.0 Hub”, 2006.10.30, <http://web2hub.com/>
- [7] “O'Reilly Radar”, 2006.9.24, <http://radar.oreilly.com/>
- [8] “Read/Write Web”, 2006.10.12, <http://www.readwriteweb.com/>
- [9] “PRAK's Blog: Versioning Up the Web!”, 2006.10.31, <http://www.fortytwo.co.kr/tt/>