

# ISO/TC 37 지식을 담은 표준 온톨로지와 언어지식자원

ISO/TC 37 부의장 겸 SC4 간사  
한국과학기술원 전산학과 교수  
국가지정언어자원은행, 시맨틱웹첨단연구센터장  
최기선  
kschoi@cs.kaist.ac.kr

## 1. 지식을 담은 표준과 역사: 서론

2004년 W3C (World-Wide Web Consortium)에서는 OWL (Web Ontology Language)의 표준 권고안을 통과시켰다. 컴퓨터 과학이 인간의 지적 능력을 기계가 하도록 하는 것이 목표의 하나라고 하겠다. 무엇을 안다고 하는 것이 중요한 이상으로, 알고 있는 것을 어떻게 활용하여 새로운 것을 알아내는 추론 능력 또한 중요하다. OWL은 인간의 알고 있는 사실들을 컴퓨터가 알 수 있도록 하는 "지식을 표현하는 방법"을 표준화한 것이다. 컴퓨터가 존재하기 전부터 지식을 어떻게 하면 기계적으로 표현할 것인가의 문제는 있어 왔고, 수학자와 철학자의 역사적 논제이기도 하다. 20세기 초에 괴델은 다음을 증명하였다: 지식을 표현하는 능력이 있는 논리적 방법 중 비교적 간단하다고 보이는 방법 (first-order predicate logic)으로 지식이 쓰여 있다고 하더라도 "기계적으로 그 추론을 완성할 수 있는지 알 수가 없다 (혹은 결정할 수가 없다)." 그 후 20세기 말에 돼서야 "기계가 결정할 수

있는" 지식표현 방법을 알아내게 되었다. 그 이름이 Description Logic (약칭 DL)이다. OWL은 DL에 힘을 입어 실용화의 단계에 들어선 것이다. 즉, 웹의 콘텐츠를 기계가 이해할 수 있게 된 것이다. 사람이 웹의 내용을 이리저리 찾아보고 결정하던 것을 앞으로 시맨틱웹 기술이 완성이 되면 기계가 사람이 하던 사이버 공간의 일을 모두 하게 되는 것이다. 이제 10년 이내에 인터넷 서핑의 시간은 획기적으로 줄어들고 조직화가 진행될 것이다. Google 등의 선진적 인터넷 검색회사는 web 2.0을 넘어서서 시맨틱웹의 세상으로 들어가기 위한 기술을 개발하여 왔고 적용하고 있다. 한국의 인터넷 검색회사보다 자동화율이 획기적으로 강한 선진 검색엔진들은 그 이면에 이미 시맨틱웹 기술을 사용하고 있다는 것이다. 향후 우리나라의 인터넷 지식검색의 기능이 완전 자동화가 되는 날이 바로 시맨틱웹 기술이 성숙하게 되는 시점이다. W3C의 기술발전예측에 의하면 그 시점이 바로 5년 후인 2010년이다.

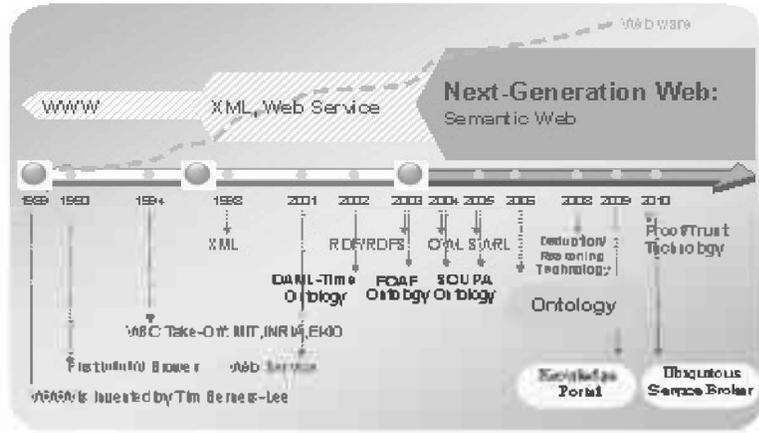
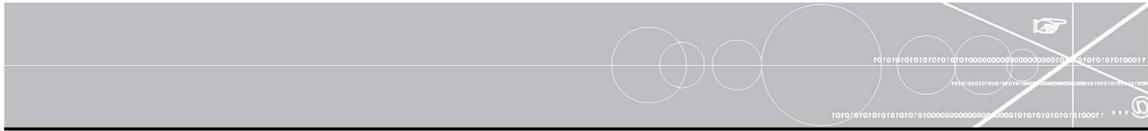


그림 1 . 웹 표준화 로드맵

HTML의 역할은 주어진 정보를 사람이 보기 좋도록 웹을 통하여 시각화하는 것이다. OWL은 사람이 아닌 기계가 보기 좋도록 웹의 정보를 이해하고 추론 가능한 형태로 하는 것이다. 그러면 그 많은 정보를 일일이 OWL로 만들어주는 것은 누가 하는가? HTML을 바탕으로 웹에 정보를 시각화하여 보여주는 일을 하는 사람은 웹 디자이너, 웹 마스터 등으로 직업화되어 있는 반면, 이제 그 직업의 내용이 보편화되어 아무나 할 수 있는 일이 되기 시작하였다. 다시 말하면 웹 디자이너나 웹 마스터의 역할이 기계화가 되었다는 뜻이다. 바로 웹의 주인공은 모든 사람 "You"이다. 그러면 OWL에 의하여 시맨틱웹으로 만들기 위하여 어느 직업군이 필요한가? 알고 있는 지식을 OWL로 써주는 시맨틱 웹 디자이너와 시맨틱 웹 마스터가 필요한 것이고, 또 자동화가 진행된다고 예측한다.

이 자동화와 기계화에 필요한 것이 바로 누구나 알아야 할 기본지식을 기계도 미리 알고 있어야 한다는 것이다. 다시 말하면, 우리 사람도 어느 정도 알고 있어야 신문도 읽고, 그 다음 단계의 교육을 받을 수 있는

이치와 같다. 우리 교육 교재가 검정 교과서로 어느 정도 표준화가 되어 있다. 이와 마찬가지로 시맨틱웹에서 기계가 꼭 알아야 할 표준화된 지식이 필요한데 이것을 "온톨로지"라고 하며 표준화의 대상이다. W3C에서는 지식을 기계가 알고 생각하게 하는 방식을 표준화하고 있다. 반면 시맨틱웹이 하는 여러 가지 일들에 필요한 기본 지식인 온톨로지는 그 쓰임새에 따라 여러 조직에서 표준화를 하고 있다.

## 2. 지식 콘텐츠 표준과 온톨로지: ISO/TC37

온톨로지는 어떻게 만들어야 하는가? 이 질문에 답을 하려는 것이 ISO/TC37의 표준화 목표가 되고 있다. 올해 2006년8월에 열린 TC37 총회에서 이와 같은 비전을 발표하여 저자는 TC37의 부의장으로 추대가 되었다.

JTC1의 여러 SC에서도 이와 비슷한 노력을 기울이고 있다. 메타데이터, 토픽맵과 같은 표준이 바로 그러한 것이다. 단, ISO/TC37의 방향은 이와 같은 지식표



현언어를 표준화하려는 것이 아니다. 지식 바로 그것, 온톨로지 만들기를 체계화하고 궁극적인 자동화를 위하여 어떠한 과정을 거쳐야 하는 것인가를 표준화하려는 것이다. ISO 9000이 경영합리화를 위한 인증표준이라면, 바로 향후 사이버 공간의 시맨틱웹의 기본이 되는 지식인 온톨로지를 만들기 위한 온톨로지 경영합리화를 위한 과정을 표준화하여야 한다는 것이다.

ISO/TC37은 1938년에 탄생하였다. 그 후 2001년도까지 "전문용어" 표준화를 위한 선도를 하여 왔다. 유명한 표준 중에 "국가 코드", "언어 코드" 표준인 ISO 639가 TC37의 산물이다. 전문용어 표준화를 위한 전

문용어 제작 과정의 표준화를 다룬 ISO 704도 역시 TC37의 표준이다. 2002년도에 저자가 제안하여 "언어자원운용" 표준을 위한 ISO/TC37의 SC4가 설립되었다. ISO/TC37에서는 한국의 지원 약속을 매우 고맙게 생각하였다.

지식은 언어에서 비롯되며, 그 운용은 "언어 자원"의 표준화에 근거한다는 자명한 사실이다. 이와 같은 맥락에서 온톨로지를 다루기 위한 ISO 639와 같은 온톨로지 코드, 온톨로지 제작 과정 표준을 위한 ISO 704의 온톨로지판, 그리고 전문용어 관리를 위한 data registry의 개념을 도입한 ISO 12620 등의 온톨로지 registry 확장판이 적용될 수 있다.

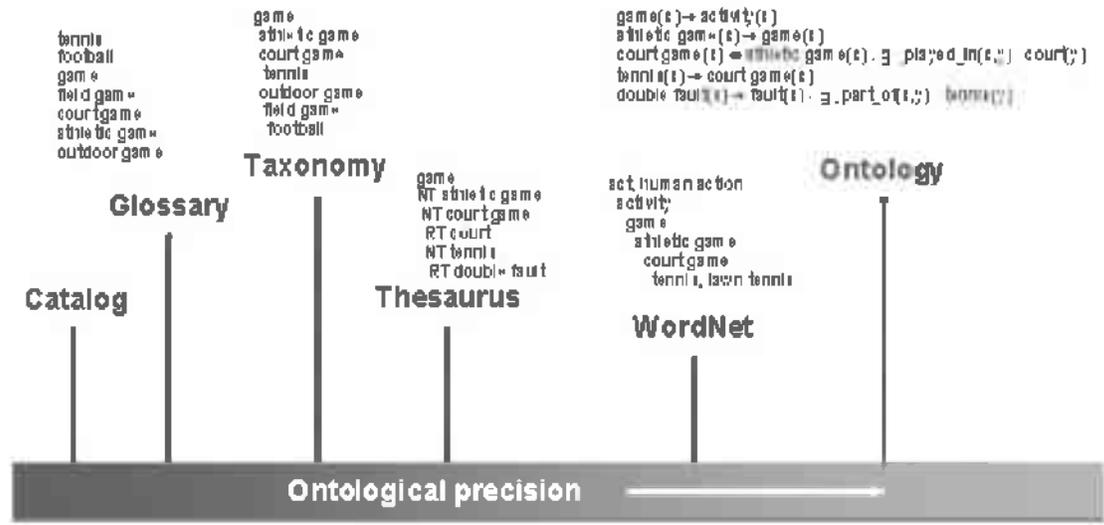


그림 . 전문용어에서 온톨로지까지: 지식표현의 정확도의 심화 과정

### 3. 기술 및 산업으로의 표준화 파급 효과

현재 웹콘텐츠는 폭발적으로 증가하고 있다. 웹페이지가 어마어마하게 늘어나서 매일 읽어야 하는 양도 많아 졌지만, 단순히 읽는 것이 아닌 웹에서 서비스를 하여 주는 것도 늘어나고 있다. 앞으로는 웹페이지 읽는 것보다 서비스를 골라서 내 목적에 맞게 활용하는 것도 폭발적으로 증가한다고 한다. 정확한 정보 습득에 대한 욕구는 지금도 마찬가지이지만 그 늘어나는 양에 대하여 정확성을 더 추구할 것이다. 바로 그 "정확성"을 준수하기 위하여, 웹을 만드는 사람이나 웹을 보는 사람이 공유화하여야 할 것이 기계의 기본 지식인 "온톨로지"이다.

사이버 공간에서 "서비스"에 의한 물건을 주고 파는 상업 활동이나 은행 업무에서부터 시작하여, 집의 홈네트워크에서 지능적인 로봇들과 가전제품들이 서로 연계하여 쓰이도록 해야 한다. 이에 필요한 것이 공통 지식으로서 "온톨로지"이며 그것이 표준화되어 서로 주고받을 수 있는 "상호운용성"이 높아야 한다. 또한, 유비쿼터스의 응용 기술 시장의 확대와 정착이 바로 온톨로지 표준화에 맞물려 있다. 다시 말하면, 유비쿼터스 환경과 지식 서비스 산업의 활성화는 물론, 차세대 인터넷 환경에서 시맨틱웹 기술의 선점을 위하여, 한국의 표준으로서 온톨로지 국제표준을 이루고자 한다. 그것을 ISO/TC37의 부의장국으로서 또한 ISO/TC37/SC4의 간사국으로서 표준을 선도할 수 있는 길이 우리에게 있다.

이것이 바로 ISO/TC37이 새롭게 "terminology, language, content resources"의 표준화를 위하여 진전하고자 하는 방향과도 일치한다.

### 4. 결론: 한국의 표준화 업무에 대한 소고

국제표준제정의 과정은 외교적 업무와 같다. "기술 외교관"의 역할을 해야 한다. 외교 업무가 총칼 없는 전쟁이라는 것은 익히 알고 있어, 이에 대한 총과 총탄 지급은 국가적으로 잘 지원이 되어야 한다는 것은 누구나 동의하고 있다. 그러나 국제표준을 위한 기술외교관으로서 간사 업무를 하는데 대하여 한국은 꽤 인식이 부족하였다고 보인다. 심지어 한 때 간사국 운영비와 같은 예산을 대폭 없애는 데에 서슴지 않았던 정부의 방침도 있었던 것 같다. 돈 문제뿐만이 아니다. 국제표준의 UN이라고 하는 ISO, IEC의 조직에 한국은 그 대표부를 잘 운영하고, 기술외교 활동을 더 제대로 벌이기 위하여 더 심각해져야 할 때이다.

#### 4.1 기술외교로서의 정보 채널 문제와 해법

지난 7월에 ISO/TC37의 의장 후보를 공고하였을 때가 있었다. 이에 본 저자를 영국 기술표준원 BSI에서 새로운 의장 후보로 추천하여 등록하였고, 새로 위원회를 구성하여 선거 활동을 시작하였다. 이에 TC37 총회가 벌어진 8월 말에 다른 나라들은 공식적인 지지 선언도 해주어서 감사를 드렸다. 그러나 정작 한국에서는 본인에 대한 의장 입후보 사실조차도 가르쳐 주지 않았고, 이에 대하여 어떻게 대처해야 하는지에 대한 정보도 없었다.

또, 3년 전부터 진행되어온 산업자원부의 동아시아 표준화 사업인 S-Dialogue에서 한중일 전문용어 표준 데이터베이스화에 대한 사업이 있었다. 이를 한국 주도로 진행하였으나 이 데이터베이스의 저작권에 대한 사항의 ISO/CS (중앙사무국)와의 협상문제가 중국에서 제기되었을 때, 일본 규격협회에서 알려주어



확인을 받은 사실도 있었다. 왜 우리나라는 이러한 정보원이나 조직에 있어서 간접적으로 움직여야 하는지 모르겠다.

이에 대한 해법으로서 ISO/CS에 현재 TC나 SC의 의장, 간사들을 6개월 내지 1년간 파견하는 것이다. 이미 중앙사무국의 임원들과 잘 알고 있고, 업무도 파악이 되고 있고, 그 필요성이 절실한 사람들을 ISO 본부에서 근무하게 하여, 우리의 인맥을 키우고, 활발한 정보 채널도 확보하는 것이다. 즉, 국제표준의 민간화와 더불어 필요성에 입각한 기술외교관을 조직적으로 키울 수 있는 밑거름이 되게 하여야 한다. 특히, 대학이나 연구소에서는 연구연가제도가 있어서, 이를 정부에서 지원을 하고 기회를 준다면 우리의 기술외교관으로서 앞으로 UN 사무총장과 같이 ISO를 지휘하는 사람들을 키우고 분위기를 만들 수 있다. 지금은 ISO/CS의 중요 인사들이 한국에 대한 인식을 올릴 필요가 있다. MPEG과 같은 대규모의 총회 유치도 중요하지만, ISO가 하려 하거나 힘이 들어 하는 일도 한국이 해줄 때, UN사무총장과 같은 지위를 ISO에서도 얻을 수 있다.

#### 4.2 기술외교와 표준지원의 조직화

표준화 업무는 걱정만 문서화에 기반을 둔다. 특히 모든 회의에는 결의문과 같은 법적 효력이 있는 문서화가 이루어져야 한다. 민간 기술외교관으로 "간사"는 이 업무를 하도록 되어 있다. 업무 매뉴얼이 있어서 이에 따르면 되겠다고 생각하지만, 즉각적으로 경우에 따라 기술표준원에는 각종 질문에 대하여 답을 잘 해줄 수 있는 체제와 사람이 필요하다. 즉, 기술표준원 수준에서 의장이나 간사 업무의 helpdesk를 두어야 한다. ISO/TC, SC의 의장이나 간사는 ISO에서 월급을 받는 사람들이 아니고, 자발적이고 사명감에 의한 노력을 하여, 그 활동비와 운영비를 국가에서 지원을 받도록 되어 있다. 국가대표 축구 선수가 비즈니스 클래스를 타고 가야 하는 이유 그 이상으로 민간 기술외교관들은 조직적인 국가적 지원과 훈련, 그리고 조직적 정보채널이 확보가 되어 좀 더 우리나라의 지식기반 표준화에 의한 산업 발전에 이바지할 수 있도록 하는 것이 더 효과적이고 예산절감이라 할 수 있겠다. **표준**