국가 선진화를 위한 '시맨틱 기술'의 필요성

글 _ 옥 철 영 교수 · 울산대학교 컴퓨터정보통신공학부 · okcy@ulsan.ac.kr

'시맨틱 테크놀로지(Semantic Technology; 이하 시맨틱 기술)'를 우리말로 어떻게 표현할 수 있을까? '의미 기술', '의미 처리 기술', '의미적인 기술', '의미 있는 기술' 등으로 표현할 수 있지 않을까? '의미(意味)'라는 단어를 국어사전에서는 『①말이나 글의 뜻, ②행위나 현상이 지닌 뜻, ③사물이나 현상의 가치』로 풀이하고 있다. 이는 영어로 표현하면 'meaning; significance, import, meaningful' 등으로 번역할 수 있을 것이다. 그렇다면 시맨틱 기술을 '의미 분석(semantic analysis)과 의미 처리(semantic processing) 등의 정보처리 기술과 사용자가 만족하는 의미 있는(meaningful) 서비스 기술이 융합된 총체적인 기술'이라고 정의내릴 수 있지 않을까 한다. 나아가 한국과학기술정보연구원(KISTI)의 입장에서의 시맨틱 기술은 '과학기술 지식정보에 대한 의미적 분석, 관리, 저장, 검색 등의 일련의 의미 처리를 통해, 과학기술 지식정보를 이용하는 연구자에게 의미 있는 서비스가 가능하도록 하는 기술'로 정의할 수 있을 것이다.

현재 시맨틱 기술은 시맨틱 웹의 등장으로 인해 시맨틱 웹 관련 기술이 대부분 시맨틱 기술이라는 오해를 불러오고 있다. 데이터를 가공하고 정제하여 서비스 환경에 맞게 정보처리를 하는 일련의 과정 역시 정보서비스 측면에서는 시맨틱 기술이라고 정의내릴 수 있기 때문에 정확하게 어떤 것이 시맨틱 기술이라고 정의내리기가 힘들다. 하지만 최근의 동향이나 연구적 관점을 통해 이 시맨틱 기술에 대한 대략적인 연구개발 성향을 분석할 수 있을 것이다. 소프트웨어 관점에서의 시맨틱 기술은 일반적으로 데이터와 문서 파일에서 의미를 분석하거나 애플리케이션 코드에서 의미를 분석하는 의미 중심적(meaning-centered) 기술을 말하는 것으로, 토픽(topic)과 개념(concept)을 자동 인식(autorecognition)하거나, 특정 데이터에서 정보나 의미를 추출하고, 의미적으로 정보들을 범주화(categorization)하는 등의 일련의 의미 처리 과정을 일컫는다.

그리고 시맨틱 웹은 웹상에 존재하는 정보를 사람뿐만 아니라 기계가 의미를 파악하고, 사용자의 요구에 적합한 결과만을 찾아주는 의미 기반 검색이나 사람과 기계, 또는 기계와 기계 상호간에 협업을 원활히 수행함으로써 사람을 대신한 자동적인 서비스가 가능하도록 하는 차세대 웹으로서, 시맨틱 웹 기술은 이러한 차세대 웹을 만들기 위한 웹 온톨로지 기술, 추론 기술, 웹 서비스 기술 등 일련의 기술들을 의미한다.

덧붙여 언어 처리에서의 시맨틱 기술은 형태소 분석, 구문 분석, 의미 분석, 화용 분석 등의 일련의 자연언어 분석 과정이 언어의 의미를 파악하는 일련의 시맨틱 기술이라 할 수 있다. 시맨틱 웹에서 다루는 연구 주제를 살펴보면 인공지능, 언어처리, 웹 프로그래밍, 정보저장 · 관리 등 다양한 분야의 기술들이 시맨틱 웹을 만들기 위해서 필요하다는 것을 알 수 있다. 이러한 측면에서 시맨틱 기술은 시맨틱 웹 구현에 필요한 일련의 기술을 포함하여 지식정보의 의미적인 확장 · 공유 · 인식을 가능하게 하는 일련의 기술이라 할 수 있을 것이다.

현재 국내에서 시맨틱 기술 기반으로 연구 개발된 사례는 그리 많지 않다. 한국과학기술정보연구원의 OntoFrame®과 Semantic KRISTAL(현 KISTI-STA®), 조달청의 온톨로지 기반 상품 검색 시스템, 한국문화콘텐츠진흥원(KOCCA)의 문화콘텐츠 유통정보 포털 시스템, 한국교육학술정보원(KERIS)의 의미기반 검색 시스템, 한국정보문화진흥원(KADO)의 국가지식포털, KT의 시맨틱 웹 기반 한글 검색엔진, KAIST의 국가 IT Core 온톨로지 구축 및 활용 등 최근 들어 시맨틱 기술 기반의 연구 개발이 조금씩 그 결과물을 선보이고 있다.

특히 한국과학기술정보연구원은 2004년부터 시맨틱 웹 기술 연구 개발을 시작하여, 세밀하게 준비된 기술과 시나리오를 바탕으로 2006년에는 연구자들의 연구 개발 전주기를 지원하는 시맨틱 웹 기반 정보서비스 프레임워크 OntoFrame®을 발표한 바가 있다. 이 프레임워크는 URI 서버 기반 온톨로지 개체 관리 기법, 과학기술 시소러스, 국가과학기술 기반정보 온톨로지 등을 비롯하여 온톨로지 기반 추론 서비스, 성과정보 맵 서비스 등 시맨틱 웹 요소 기술과 응용 기술을 결합하여, 시맨틱 웹 특화 서비스 기술 분야를 창출하였다. 또한 오랫동안 자체 개발한 정보검색관리시스템인 KRISTAL과 의미적 언어자원을 결합한 의미기반 정보검색 서비스인 Semantic KRISTAL을 2006년에 개발하여 과학기술 시맨틱 정보서비스의 또 다른 모습을 선보이기도 했다.

Category	Semantic Technology	Information Technology
Represent meaning	▪ Semantic web standards ▪ Language-neutral ▪ Machine interpretable ▪ Semantic metadata ▪ External to system (sharable) ▪ Knowable at run time	▪ Data format; program language ▪ Language-based ▪ Human knowledge required ▪ Syntax/structure metadata ▪ Internal to system ▪ Predefined at design time
Discover, acquire, create, access	▪ Meaning-centered. ▪ Auto-recognition, information extraction, and categorization ▪ Topic and concept-based search with enhanced recall and precision (relevance)	▪ Message, file, page, and document-centered ▪ Manual, full-text & statistical categorization ▪ Data, word, & document search with limited recall and precision
Organize, integrate, interoperate	▪ Taxonomy, ontology, knowledge base ▪ Knowledgebase easily modified for new concepts, relationships, properties, constraints and instances. ▪ Integrate data, content, applications, and processes via shared ontology	▪ RDBMS, OODBMS, flat file. ▪ Database structure difficult to modify to change /add new relationships (e.g., requires coding, reorganization) ▪ Integrate data, content, applications, and processes via point-to-point interfaces
Reason, interpret, infer, answer	▪ Application reasons via logic constraints, rules, axioms separate from cod ▪ Knowledge answers questions such as how, why, and what if, (combining theory & data) ▪ Machine can learn (infer new knowledge), simulate, test and adapt based on experience.	▪ Application reasons via fixed algorithm that is embedded in the application code. ▪ Information processing provides situation awareness, e.g.,: what, where, when, and how much ▪ Preprogrammed responses are like instinct. Logic updated off-line (new version) Program does not learn or adapt
Provision, present, communicate, act	▪ Auto-discover, provision, semantic web services ▪ Ontology-linked composite applications give 360-view of relevant data in context. ▪ Auto-generate text, documents, graphics, drawings & dialogs ▪ Auto-personalize, customize, version (e.g. languages) ▪ Autonomics: systems with self-knowledge can self-configure, self-optimize, self-protect, self-heal, and self-manage. ▪ Services & products that know, learn, & reason as humans do	▪ Manually discover and implement data and application connections & interfaces ▪ Human search to find data & information and then put it into useful context for decisions ▪ Computers as "electronic pencils" for humans to author and develop content, visuals, and media formats

Source: TopQuadrant

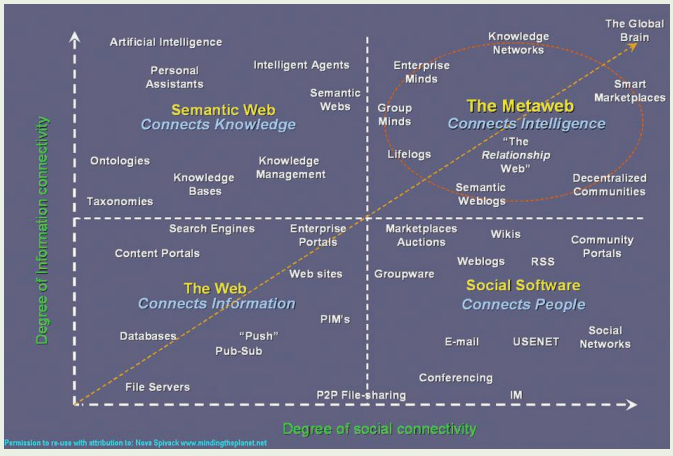
〈그림 1〉 Semantic Technology & Information Technology

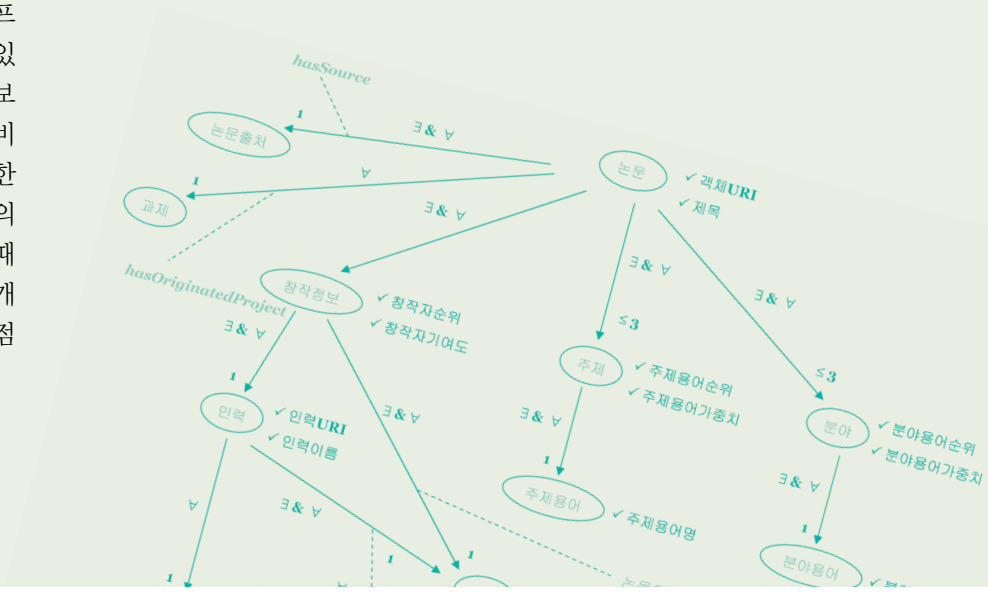
국외의 기술선진국에서도 이미 Esperonto, Ontoweb, Empolis K42, Mondeca TM, Swoogle, ODESew 등과 같은 다양한 시맨틱 기술(엄밀하게는 시맨틱 웹 요소 기술) 기반의 서비스 및 포털 시스템을 개발하고 있다. 또한 시맨틱 기술과 관련된 일련의 연구 개발 성과물 역시 상당히 높은 수준에까지 이르렀다고 할 수 있다. 이런 점과 비교하여 국내에서는 아직까지도 인포메이션 기술과 시맨틱 기술의 연계성 연구 및 실험적 개발 수준에서 벗어나지 못하고 있는 실정이다. 그러나 앞에서 언급되었던 국내의 여러 기관의 시맨틱 기술 기반 실용화 기술 연구 개발 노력은 우리나라의 시

맨틱 기술 연구의 실용성을 앞당기는 그 시작이라 할 수 있을 것이다.

시맨틱 기술 개발은 어떠한 기술을 보유하고 있으며, 어떠한 서비스를 하고 있는가에 따라 기술 수준의 우수성을 평가할 수 있다. 즉 시맨틱 기술 연구 개발은 먼저 특화된 서비스에서 기술적 우수성을 검증받은 뒤 점차적으로 서비스 분야를 확대해 나가면서 평가해야 한다. 이는 시맨틱 기술이라는 것이 면밀한 지식정보 데이터에 대한 분석에서부터 서비스 시나리오에 이르기까지 기술적 · 서비스적 환경에 의존적인 기술이기 때문이다. 이런 측면에서 일반 포털 서비스에서 탈피하여, 과학기술 연구자 나아가 전문가를 위한 시맨틱 기술 기반의 특화된 시맨틱 포털 개발이 높은 수준의 기술과 서비스를 질적 · 양적으로 성장시킬 수 있는 국가적 차원의 지속적 연구 개발로 이어진다면, 과학기술 연구자 및 분야별 전문가의 연구 개발 환경 변화의 중심에 설 수 있을 것으로 예상된다.

최근 정보통신부에서 수립한 Post IT839 전략 품목 중 3대 인프라에 소프트웨어 인프라, 8대 서비스에 지식기반 IT 서비스가 새롭게 채택됨으로써, 지식기반 IT 서비스의 일환으로 시맨틱 웹에 대한 관심이 급증하고 있으며, 특히 지식기반 서비스에 필요한 온톨로지, 시맨틱 기술 개발은 미래 지식기반 경제의 인프라를 구축하는 데 필요한 요소로 주목받고 있다. 즉 지식기반 IT 서비스는 산재된 지식정보 자원의 체계화와 공유 가능한 지식기반 서비스를 활성화하겠다는 의미로 해석된다. 또한 국내외적으로 〈그림 2〉와 같이 기술 마케팅의 급변화가 실제 기술 개발보다 앞서가는 이때에 선도적이고 미래예측적인 시맨틱 기술 개발은 그만큼 선도적 입지를 굳힐 수 있는 시점이라 할 수 있다.



〈그림 2〉 앞서 가는 Web 환경 : Metaweb 소개

이러한 일련의 국가 정책 방향과 국내외 기술 동향에 맞추어, 한국과학기술정보연구원은 고성능의 과학기술 지식정보 인프라를 제공하는 중추적인 역할을 담당하는 국가기관으로서, 현재 IT의 혁신적인 시맨틱 기술을 적극 활용하여 자동화되고 지능화된 지식정보를 제공하여 보다 강력한 과학기술 지식정보 인프라를 제공하는 데 앞장서야 할 것이다. 또한 한국과학기술정보연구원이 과학기술 연구자를 위한 시맨틱 세상의 기술적·서비스적 중심이 되기를 바란다.