

시맨틱웹과 온톨로지: 가능성과 한계

부산대학교 권혁철

1. 들어가며

원시시대에는 귀동냥으로 정보를 획득했지만, 시대가 흐름에 따라 문자를 통한 정보획득이 가능해지면서 축적된 정보의 활용이 가능해졌다. 인쇄술 발달은 정보의 대중화를 이루었으며, 매스컴은 문자를 넘어서서 음성과 멀티미디어 정보를 실시간에 저렴한 비용으로 배포할 수 있게 했다.

그러나 이런 모든 정보유통의 발전도 지난 10여 년간 이루어낸 웹 중심의 인터넷이 이루어낸 정보유통의 변화와 비교할 수 없다. 인터넷은 정보소비자와 정보생산자의 구별을 없애고 정보획득 장벽을 제거함으로써 정보유통의 민주화와 대중화를 가져왔다. 멀티미디어 정보를 비롯한 모든 종류의 정보를 거의 공짜에 가까운 비용으로 배포하고 획득할 수 있게 했다.

이런 정보유통의 혁명은 사회민주화와 세계화를 이루어내면서, 동시에 정보획득 불균형의 확대라는 새로운 문제를 야기했다. 더구나 인터넷에 의해 무차별로 배포되는 정보에서 알짜 정보를 찾고, 거짓 정보, 쓰레기 정보, 불필요한 정보를 제거하는 것은 현재 사용하는 단순링크(링크에 정보가 없이 지시만 하는)에 기반을 둔 인터넷에서는 현실적으로 불가능하다.

정보를 이용하려면, 먼저 다양한 방법으로 정보를 획득하고, 이를 자신이 원하는 목적에 따라 분석하고 가공하여 정보를 통합하고, 이를 바탕으로 정보를 활용하거나 또는 통합한 정보를 다시 배포하게 된다. 그런데 단순링크 기반 인터넷에서는 색인에 의한 홈페이지 검색시스템이나 주제에 따라 홈페이지를 분류한 검색시스템에서 제공하는 정보를 바탕으로 필요한 정보를 획득해야 한다. 그런데 정보검색 시스템이 제시하는 정보는 지나치게 많으며, 분류시스템의 분류방법은 대부분 사용자가 원하는 형태로 되어있지 않다. 따라서 정보획득이 우연에 의해 이루어질 수밖에 없다.

이 문제를 가장 쉽게 해결하는 방법은 기계가 언어나 멀티미디어 정보를 인간처럼 이해하여 처리할 수 있는

기술을 개발하는 것이다. 그러나 이런 기술이 조만간 개발될 가능성은 없다.

한편, 오랫동안 특정 분야에서 정보를 획득한 경험자는 자기만의 경험적 지식을 바탕으로 필요한 정보를 획득한다. 그리고 일부 전문가는 정보획득에 필요한 정보(일종의 메타정보)를 모은 사이트를 만들기도 한다. 그런데 같은 분야의 전문가라도 각자 정보획득에 사용하는 방법 또는 표현 방법에 차이가 있다.

그런데 이런 전문가가 정보획득에 도움을 주고자 만든 사이트 자체가 역시 단순링크 기반으로 정보가 표현됨으로써 한계가 있다. 더구나 다양한 전문가가 각자 다른 관점에서 정보를 찾는 방법을 표현했다면 모든 사이트를 방문하여 분석하고, 그 결과를 바탕으로 필요한 정보를 획득해야 하므로 어려움이 크다.

정보이용자는 자신이 원하는 정보만을 자신이 원하는 형태로 받기를 원한다. 지금처럼 거의 무한으로 확장되고 있으면서 거짓정보와 진짜정보, 필요한 정보와 불필요한 정보가 뒤섞인 인터넷 환경에서 정보유통의 효율화를 이루면서, 정보이용의 격차를 극복하고, 정보이용을 극대화하려면 사용자가 원하는 정보만을 획득할 수 있는 방법론이 제공되어야 한다.

이 목적으로 Tim Berners-Lee가 제시한 새로운 인터넷 환경이 시맨틱웹이다. 시맨틱웹이 정말 이런 문제를 해결할 수 있을지는 아무도 모른다. 그러나 지금까지 제시된 방법론 중에서는 가장 설득력이 있다. 이 글에서는 시맨틱웹의 장점, 한계, 극복해야 할 점 등을 중심으로 시맨틱웹의 가능성과 한계를 살펴보겠다. 또 시맨틱웹 관련 연구에서 가장 큰 논란이 되는 '온톨로지'에 대해 나름으로 정의해보겠다.

2. 시맨틱웹은 무엇이 다른가?

인간은 의미 또는 개념에 따라 정보를 획득하고 가공한다. 그러나 기존기술로는 기계가 언어를 이해하기는 불가능하다. 따라서 Tim Berners-Lee는 의미를 인

간이 직접 메타정보로 제공하는 방법을 생각했다. 즉, 인터넷에서 접근할 수 있는 또는 세상에 존재하는 객체(object, URI로 표현)¹⁾에 대해 의미에 기초하여 정보의 내용과 개념 따위를 표현하는 체계를 만든 것이 시맨틱웹이다. 정보 제공자나 이용자가 경험에서 얻은 정보 획득에 필요한 메타정보 또는 자기가 접근한 사이트나 페이지의 내용에 대한 정보 따위를 의미에 기초하여 기록하여 인간뿐 아니라 기계도 활용하게 하는 것이다.

그런데 '의미(뜻, semantic)' 그 자체를 정의하거나 표현하기는 매우 어렵다. '커피'라는 말은 그 자체로 뜻이 있지만, 응용분야에 따라 의미를 더 세분화해야 할 수도 있다. 즉, 문맥에 따라 '커피나무', '커피열매', '음식' 또는 '음료수' 따위로 다르게 해석할 수 있다(홍재성(서울대) 교수의 발표 중). 따라서 어떤 차원에서 의미를 표현하느냐에 따라 달라질 수 있다. 따라서 우리가 쓰는 단어를 그대로 의미표현에 쓸 수는 없다.

시맨틱웹에서 뜻하는 '의미표현'은 인간뿐 아니라 기계(프로그램)도 이해할 수 있는 '의미'를 뜻한다. 기계가 이해한다는 말은 매우 중요하다. 즉, 인간이 메타정보를 주면, 그것을 이해하여 프로그램이 의미에 따라 정보를 통합하여 제공할 수 있어야 한다는 전제조건을 가정한 것이다. 따라서 응용분야에서 기계가 하나로만 해석할 수 있는 '의미' 단위를 가정해야 한다.

그런데 이런 의미를 모두가 공감하게 표현하는 방법은 없다. 이에 따라서 일종의 name space에서 이야기하는 의미 표현방법을 시맨틱웹이 도입하고 있다. 즉, 나는 "여기서 이 단어를 이런 의미로 사용한다."라고 선언하고, 그에 따라 의미를 표현하는 접근을 취한다.

name space는 어떤 용어(terminology)를 어떤 의미(gloss, 주석으로 구체화)로 사용하겠다고 정의한 것이다. 예를 들어, '역학'을 쓸 때, 물리학에서 이야기하는 '역학(力學)'으로 쓰겠다거나, 주역에서 말하는 '역학(易學)'으로, '역학(疫學, 질병 관련)' 또는 '역학(曆學)'으로 쓰겠다는 것을 의미에 따라 구별하여 쓸 수 있게 하는 것이다.

용어뿐 아니라 개념을 표현할 때도 의미해석에 중의성이 없어야 한다. Dublin Core에 의해 낱자를 표현하면 해석이 언제나 일정한 것이 예가 된다. 예를 들어, "URI가 가리키는 객체(사이트)에는 어떤 정보가 있는데 신뢰도는 어느 정도다"라는 것을 개념화하여 기계가

이해할 수 있는 형태로 표현할 수 있어야 한다. 이렇게 함으로써 기계가 뜻을 이해하고, 정보를 정확히 추출하며, 또한 다양한 정보를 통합하여 제공할 수 있다.

특정한 분야나 목적에서 사용하는 용어나 개념을 정의한 것을 온톨로지라 한다. 하지만 '온톨로지'가 무엇인지에 대해서도 학자마다, 사람마다 해석이 다르다. 이에 대해서는 다음 장에서 더 구체적으로 설명하겠다.

그런데 특정분야의 온톨로지를 기계가 해석하려면 온톨로지 자체를 표현하기 위한 방법이 표준화되어야 한다. Tim Berners-Lee가 중심이 되어 W3C에서 온톨로지(정보)를 개념화하여 표현하기 위해 표준화한 언어가 RDF(Resource Description Framework)이며, RDF에 형(type)을 제공하여 의미적 완결성을 보장하기 위한 방법이 RDF Schema이다. RDF가 3개 요소(URI, Property, Value)로 지식을 표현하는 것은 이것이 가장 단순하면서 보편적인 지식표현 구조이기 때문이다. 그리고 웹 온톨로지를 기계가 해석할 수 있게 새로운 응용영역 용어(vocabulary)를 정의하고, 표현의 무결성을 검증하고 추론할 수 있게 온톨로지를 개념화하는 언어가 OWL(Web Ontology Language)다.

그런데 인간이 의미에 따라, 중의성이 제거된 개념화를 거쳐 메타데이터로 정보를 표현하기는 매우 어렵다. 더구나 같은 내용을 사람에 따라 다른 방법으로 표현할 수도 있다. 따라서 시맨틱웹은 분야별로 개념표현 방법을 표준화하고, 이를 바탕으로 메타데이터를 표현한다. 그러나 인간이 정보를 표준화한 방법으로 제공해야 한다는 것은 시맨틱웹의 한계가 될 수 있다.

Tim Berners-Lee는 이 문제가 개방환경인 인터넷의 특성에 의해 많은 사람(다수의 힘)이 다양한 형태로 메타데이터를 제공할 것이므로 해결할 수 있다고 본다.

그런데 현재 상용화된 시맨틱웹 기반 응용 시스템은 많아야 100개 내외의 개념만 활용하고 있다. 이는 언어처리나 지식처리 분야에서 사용하는 개념의 수에 비하면 매우 적다. 그 이유는 응용분야를 제약함으로써 그만큼 필요한 지식을 표현하는 데 쓰이는 개념의 수가 적은 데 기인할 수도 있고, 개념의 수가 너무 다양해지면 그만큼 인간의 노력이 더 들기 때문일 수도 있다. 기존 시맨틱웹 응용 시스템은 인간이 직접 표현한 메타정보와 데이터베이스의 구조에서 얻은 메타정보 및 기계가 자동처리하여 얻은 메타정보(정보검색시스템이 제공한 순위정보, 다른 시맨틱웹이 제공하는 정보)를 바탕으로 정보를 통합하여 제공한다.

최근 시맨틱웹의 응용영역은 크게 확장되고 있다. 인터넷에서 프로그램과 프로그램이 의미에 따라 정보를 교환하기 위한 시맨틱웹에 기반을 둔 웹서비스를 구상하기

1) 여기서 'URI(Universal Resource Identifier)'로 표현한 것은, 지금까지 사용한 URL처럼 장소를 중심으로 접근하는 것이 아니라, 객체 자체로 접근함을 뜻한다. 객체는 우리가 정보를 줄 수 있는 모든 것에 해당하며 심지어 이 글의 각주인 이 부분도 객체로서 URI를 가질 수 있다. URI는 객체에 부여한 이름(주민등록번호를 생각해 볼 것)이다.

도 하고(시맨틱웹서비스), 업무 흐름(business process)을 시맨틱웹에 기초하여 정의하고 이를 인터넷으로 통합하려고 노력하고도 있다. 또 제품의 사양, 하드웨어/소프트웨어 특성, 개인의 성향, 개인이 느끼는 가계나 제품에 대한 인상 등을 시맨틱웹에 의한 메타데이터로 표현하려고 하고도 있다.

특히 물류처럼 상품의 생산에서 소비까지 모든 과정을 통괄하면서, 각 과정에 참여하는 기관이나 기업이 각자의 온톨로지를 보유하고, 또 정보유통과정이 어떻게 이루어질지 알 수 없는 환경에서 기계가 정보를 해석하고, 통합하고, 생성하여 제공하려면 기계가 의미를 해석하여 정보를 유통해야 하며, 현재 이런 목적으로 웹을 변화시키려는 시도가 시맨틱웹밖에 없는 만큼 시맨틱웹은 미래 인터넷환경의 발전에 중심이 될 것으로 보고 있다.

3. 시맨틱웹, 개념화 그리고 온톨로지

'Semantic Web Challenge 2003'에서 시맨틱웹 응용시스템의 최소한 요구조건을 다음과 같이 정의했다.

- (1) 응용시스템이 활용하는 정보가
 - 지리적으로 분산되고
 - 소유권이 다양하여 제어가 어려우며
 - 구조적으로나 의미상으로 이질적이며
 - 실세계 자료를 바탕으로 해야 하고,
- (2) 정보가 계속 변하는 개방형 환경에서 작동하는 응용시스템이어야 하며,
- (3) 응용시스템이 사용하는 자료가 형식론적 표현(formal description)을 따라야 한다.

위에서 말하는 '지리적'이란 정의는 인터넷이 가진 분산환경을 뜻하지만, 실체는 '개념적으로 분산되어 있다.'라는 말이 더 적합하다.

시맨틱웹에서 개념화는 매우 중요하다. 그런데 응용 목적, 응용분야, 개념화하는 사람에 따라 사용하는 용어나 개념화에 차이가 나게 마련이다. 그러나 인터넷의 개방적 특성으로 볼 때, 이 모든 차이를 인정하면 제어가 어려워질 가능성이 크다. 따라서 시맨틱웹은 공통의 온톨로지를 이용한 개념화를 가정한다. 시맨틱웹에서 뜻하는 온톨로지는 특정영역에서 사용하는 어휘를 뜻한다. 이는 name space로 구현된다.

온톨로지에 쓰인 어휘들은 여러 가지 방법으로 체계화할 수 있는데, 가장 손쉬운 방법이 'is-a' 관계에 의한 시소러스 형태의 체계화이다. 이 외에도 'part-of'나 여러 다른 방법으로 어휘 간 관계(RT: Relational Term)를 정의할 수 있다. 그런데 온톨로지는 사용하는 언어가

다르거나, 온톨로지를 정의한 집단(전문분야)이 다르면 달라진다.

한편, 같은 내용을 같은 온톨로지로 표현하더라도 개념화가 다르면 결과는 다르다. 개념화할 때 사용하는 온톨로지가 다르면 만들어진 메타데이터도 달라진다. 같은 내용을 다른 목적에서 다른 온톨로지로 개념화하여 표현하면 달라진다.

그런데 시맨틱웹 응용시스템은 필요한 정보가 다른 온톨로지에 기초하여 다른 개념화에 의해 표현되었더라도, 온톨로지를 번역하고 개념을 추론하여 해당 응용시스템이 요구하는 정보로 가공하여 응용시스템의 메타데이터 형태로 변환할 수 있어야 한다고 주장한다. 이를 위해서는 메타데이터의 의미를 표현하는 언어가 형식론적으로 표현되어야 한다. 이렇게 되어야 추론을 통하여 지식을 변환할 수 있다.

위에서 말한 (1)은 다른 온톨로지를 사용함에 따라서 메타데이터에 쓰인 용어의 차이, 개념화를 다르게 함에 따른 표현구조나 의미표현 방법의 차이, 메타데이터를 만든 목적의 차이에 따른 개념화의 차이 등 개방적 구조에 따른 문제점을 극복하고, 정보를 획득하여 가공하고 통합하여 제공할 수 있는 응용시스템을 시맨틱웹 응용시스템으로 정의한다.

그러면 시맨틱웹에서 주장하는 온톨로지와 자연언어 처리나 지식표현에서 말하는 온톨로지에 차이는 있을까? 기본적으로 RDF는 모든 지식을 표현할 수 있다. 따라서 제한 없이 자유롭게 쓴다면 차이가 없다. 더구나 언어처리나 지식표현에서 의미를 표현하기 위해 의미요소를 정의하는, 예를 들어 몬테규가 사용하는 의미표현은 name space와 유사한 면이 많다.

그러나 시맨틱웹은 일반상식을 표현하기 위한 의미체계를 가정하지 않는다. 그러므로 언어나 응용과 무관한 의미체계(universal semantics)를 추구하지 않는다. 따라서 언어처리나 지식표현에서 요구하는 수준보다는 매우 낮은 수준의 의미표현을 가정한다. 즉, 응용영역을 제한하고, 그 영역에서 문제를 해결하는데 필요한 수준에서 의미를 표현하려고 한다. 따라서 지금까지 구현된 시맨틱웹 응용시스템은 100개 내외의 개념화만 사용하고 있다.

의미표현에서 더 중요하고 더 큰 차이는 언어처리나 인공지능에서는 문제에 내재하는 중의성(ambiguity, 전산학에서는 모호성으로 번역)을 기계가 제거하는 데 온톨로지를 사용하지만, 시맨틱웹은 처음부터 중의성이 제거된 자료를 대상으로 하며, 추론과정에서도 중의성이 발생하지 않는다고 본다는 면이다. 사람이 처음부터 메타데이터를 만들 때, 또는 다른 프로그램에서 정보를

가져올 때 이미 자료에 중의성이 없다는 가정은 시맨틱 웹이 의미기반의 메타데이터를 쓰되 언어처리나 인공지능에서 극복하지 못한 기술적 장벽을 피하려는 시도로 보인다.

OWL은 표현력보다는 기계가 온톨로지를 처리할 수 있느냐에 중심을 두었다. OWL은 표현력과 기계처리가능성에 따라 OWL Lite, OWL DL과 OWL Full로 나뉜다. OWL Lite는 표현력이 가장 약하며 단순한 의미적 지원만 하며, OWL DL(Description Logics)은 계산할 수 있는 한에서 최대의 표현력을 제공한다. OWL Full은 RDF가 제공하는 최대한의 표현력은 제공하지만, 대신 계산에 대해 보장하지 않으므로²⁾ 사용자(즉, 응용 프로그램)가 알아서 처리해야 한다.

언어처리나 인공지능에서 요구하는 표현력은 일차술어논리를 넘어서지만 일차술어논리 자체가 계산이 불가능하므로 논리에 의한 이 분야 문제해결은 한계에 봉착했다. OWL Full은 이론적으로 볼 때 일차술어논리의 표현력을 가졌다고 판단한다. 따라서 언어처리나 인공지능이 요구하는 표현력은 있지만, 시맨틱웹 구현을 위해 사용될 가능성은 크지 않다.

따라서 언어처리나 인공지능을 위한 지식표현방법을 시맨틱웹 응용에 바로 적용하는 것은 매우 위험하다. 더구나 인공지능에서 말하는 시맨틱 네트워크를 시맨틱웹과 동일시하는 것은 잘못되었다. 언어처리나 인공지능을 위한 지식표현은 일반지식 표현을 대상으로 하므로 온톨로지의 번역이니 통합이라는 개념을 쓰지 않지만, 시맨틱웹이 이 개념을 쓰는 이유도 영역에 따라 또는 응용분야에 따라 세분화해 다르게 표현된 정보를 필요에 따라 기계가 활용하고, 통합하려는 시도에서 나왔다. 따라서 OWL DL 수준에서 응용영역별, 응용목적별로 다른 온톨로지가 만들어질 것이다.

그런데 아주 제한된 수준에서 만들어진 온톨로지가 아니고, 실제 문제에 활용된 온톨로지를 자동으로 통합하고, 이를 통해 필요한 정보를 의미에 따라 번역하여 활용할 수 있을까? 그렇게 희망적이지는 않다. 한 가지 해결방법은 워드넷과 같은 일반적 의미체계를 기반으로 영역별 온톨로지를 만드는 방법이 있다. 이렇게 하면 어느 정도 자동으로 온톨로지를 통합할 수 있다. 이에 따라 최근에는 온톨로지를 통합하면서 인간이 어느 정도 개입하는 방법을 주로 쓰고 있다. 그렇더라도 무작위로 만든 온톨로지를 통합하기는 쉽지 않을 것이다. 이에 대한 연구가 잘 되느냐가 시맨틱웹의 성공에 큰

영향을 줄 것이다.

온톨로지가 통합되면 개념화하여 표현된 지식의 변환이 가능할까? 이 또한 부정적인 면이 많다. 물론 '사람과 직업, 주소', '자동차를 만든 회사' 따위의 간단한 지식은 쉽게 변환할 수 있다. 그러나 선호도를 '상, 중, 하'로 분류한 사이트와 '1-10'의 단위로 분류한 사이트의 개념을 통합하려면 어려움이 상당하다. 더구나 어떤 사람의 건강에 대한 평가를 수치로 표현한 정보를 얻더라도, 그 수치가 구체적으로 무엇을 뜻하는지를 정보를 제공한 사이트에서 정확히 얻지 못한다면 수치는 의미가 없다. 그런데 시맨틱웹은 개방성을 가정하므로 이런 문제는 쉽게 해결하지 못할 것이다. 더구나 수치에 대한 판단이 통계를 바탕으로 지속하여 바뀐다면 그 내용을 메타데이터로 받지 않는 한 사용할 수 없다.

지식(메타데이터) 또는 정보의 재활용은 시맨틱웹의 성공에 중요한 요소다. 온톨로지의 개발은 어렵고 비용이 많이 들므로 대부분 기존에 개발된 온톨로지를 쓰거나 약간 바꾸어 쓸 것이라는 면은 시맨틱웹의 성공에 긍정적인 면이다. 또 어떤 전문분야 사람이 자신이 가진 정보획득 지식이나 경험을 표현하는 방법은 비슷하므로, 한번 개발한 온톨로지나 개념화 방법은 비슷한 영역에서 같이 사용될 것이라는 것도 같다. 또 전자상거래와 같이 자신이 제공한 정보를 필요한 사람이 최대한 쉽게 획득할 수 있게 하려는 의도가 있는 사이트는 표준화만 한다면 이 온톨로지를 따를 것이므로 역시 시맨틱웹이 큰 힘을 발휘할 것이다.

여러 기관이 정보를 공유해야 하는 분야도 시맨틱웹 개념이 잘 활용될 수 있다. 또 웹서비스를 활용한 ASP(Application Service Provider), 분산정보처리 등도 시맨틱웹이 잘 활용될 수 있는 분야로 본다.

4. 온톨로지와 시소러스

시맨틱웹과 관련하여 항상 문제가 되는 것은 "도대체 '온톨로지'가 무엇인가?"에 대한 정의다. 일부 사람은 이를 시소러스를 확장한, 즉 관계정의(RT)가 추가된 시소러스로 보기도 한다. '온톨로지'는 '응용영역 지식을 개념화하여 표현하는 언어'라는 측면에서 보면 시소러스와 관련성이 있다.

그러나 실제성(Reality)이 전제된다는 면에서 보면 차이가 있다. '시소러스'는 특정 분야에서 사용하는 용어 간의 관계를 정의한다. 그리고 일반적으로 검색에 중심을 둔 응용을 가정한다. '온톨로지'는 응용(실제성)의 관점에서 개념화를 하며, 그 응용이 검색이라면 시소러스와 비슷해질 수 있다.

'색깔'에 대해서 살펴보자. 색깔은 우리가 사용하는

2) 일차술어논리(First-order Predicate Logic)는 부분계산가능 문제(partially solvable problem)이다. 따라서 술어논리로 표현되었다고 해서 이를 연산할 수 있지는 않다.

언어로는 '빨강', '노랑', '하양' 등의 용어(단어)로 개념화한다. 시소러스는 각 전문분야에서 사용하는 용어와 이 용어 간 관계를 정의한다. 그러나 색깔을 전혀 다른 개념화 방법인 'RGB(Red, Green, Blue)' 값으로 개념화할 수도 있다. 또한, 표준색상표에 따라 '빨강'과는 어떤 특정 'RGB' 값이 서로 치환(번역)될 수 있다. 우리가 언어를 사용할 때는 '빨강' 등 단어를 쓰지만, 모니터에 어떤 응용 프로그램이 '빨강'을 표현하려면 RGB 값으로 바꾸어야 한다. 그런데 RGB 값으로 표현할 수 있는 모든 영역에 대응하는 색깔을 표현하는 단어가 있지는 않다. 따라서 RGB 값을 단어로 번역할 때는 유사한 색깔 이름으로 대체할 수밖에 없고, 이에 따라 정보 손실이 발생할 수 있다.

이는 개념화 방법에 따라 같은 색깔에 대한 온톨로지의 표현 방법이 어떻게 달라지는지를 보여준다. 또는 온톨로지 간에 왜 번역이 필요하며, 번역 과정에서 정보를 잃어버릴 수도 있음을 보여 준다. 그러나 다른 온톨로지를 잘 이해하고 이에 따라 번역이 잘만 이루어질 수 있다면 색깔에 대해 다른 개념화를 한 두 프로그램이 서로 정보를 인간의 개입 없이 정보를 주고받고, 이용할 수 있다.

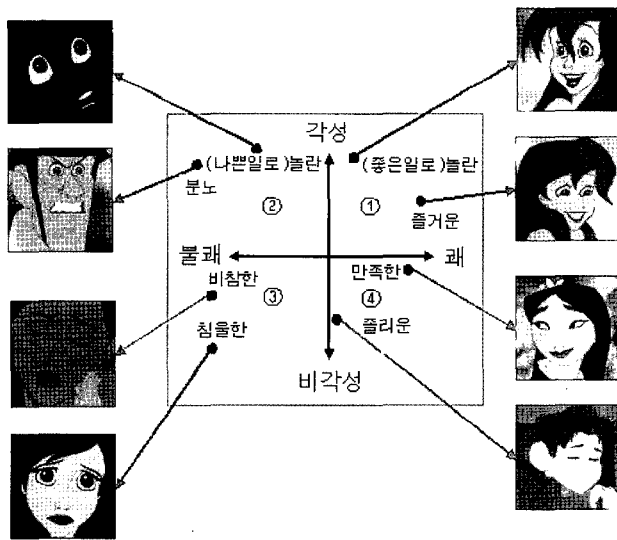


그림 1 감정의 수치화

그림 1은 인간의 감정을 '쾌-불쾌'와 '각성-비각성'으로 개념화하고, 이를 영역별로 나누어 우리가 사용하는 용어로 정의한 후, 이 결과를 바탕으로 만화영화의 주인공의 표정을 예로 보여준다. 이 그림에서 '입 모양', '눈 모양', '눈썹 모양' 등은 감정값에 차이가 난다. 만화영화 캐릭터의 감정표현은 인간의 감정 표현보다 과장을 하므로 명확히 얼굴에 표현되는 만큼 이와 같은 분석과 표현은 만화영화나 게임을 만드는 데 큰 도움을 준다.

그런가 하면 전혀 다른 방법으로 감정을 개념화한 사람도 있다. 즉 인간의 감정을 수십 또는 그 이상으로 나누고 이를 개념화한 연구자도 있다. 그러나 궁극적으로 인간의 감정을 개념화한 만큼 다른 감정 개념화 방법과 그림 1의 방법은 서로 온톨로지 측면에서 변환할 수 있으리라 본다.

같은 내용을 응용에 따라, 기관에 따라 또는 개인에 따라 다르게 개념화한 예는 얼마든지 있다. 날짜 표현 방법, 이름 표현방법, 제품 사양 표현방법, 가격 표현방법 등은 나라나 기관, 기업 등에 따라 달라진다.

지금까지는 이와 같은 온톨로지가 프로그램 또는 자료 속에 내재되어 있어 다른 온톨로지를 사용하는 프로그램의 정보를 이용할 수 없었으며, 또한 프로그램이 서로 통신을 하려면 온톨로지를 프로그래머가 맞추어 줘야 했다. 시맨틱웹에서는 온톨로지를 프로그램에서 분리하여 정의하고, 이에 따라 정보를 표현할 수 있게 함으로써 정보를 온톨로지를 기반으로 해석하여 사용할 수 있게 하여 임의의 프로그램(에이전트) 간에 정보를 공유하고, 다른 목적에서 개발된 정보를 활용할 수 있게 한다. 시맨틱웹에서 온톨로지를 표준화된 방법으로 정의할 수 있는 체계 또는 언어가 OWL이다.

5. 맺는말

기존 인터넷은 정보과부하에 따라 한계에 도달했으며, 어떤 형태로든 이 문제를 해결해야 한다. 이 해결방법은 의미에 의한 정보접근일 수밖에 없다. 또 인터넷에 있는 정보의 양으로 볼 때 기계가 능동적으로 의미에 기초하여 사용자가 원하는 형태로 정보를 통합하고 가공하여 제공해야 한다. 그러나 이런 문제를 해결할 만큼의 지능성이 있는 기계의 개발은 쉽지 않다. 따라서 이 문제의 극복 방법으로 현재 가장 실현가능성이 큰 것이 시맨틱웹이다.

자연언어처리나 인공지능이 의미적 중의성 문제를 해결하는 기술을 가정하고 있는 데 비해, 시맨틱웹은 의미적 중의성을 인간이나 다른 정보제공 소스가 제거한 의미적 중의성이 없는 정보소스만을 대상으로 하므로 자연언어처리나 인공지능의 장벽인 의미처리 문제는 발생하지 않을 것이다.

그러나 온톨로지 개발과 표준화, 온톨로지 통합과 개념 및 정보의 변환, 의미부착의 어려움과 비용 등으로 아직 결정적 시맨틱웹 응용시스템은 개발되지 않았다. 그러나 최근에 시맨틱웹의 가능성을 보여주는 여러 연구가 이루어지고 시스템도 만들어졌다. 하지만, 남은 문제가 훨씬 많다. 특히 독립하여 개발한 온톨로지에 바탕하여 개념화한 다양한 정보소스에서 필요한 정보를

찾아서 의미에 따라 통합하는 기술은 아직 아주 제한된 영역을 제외하면 이루어지지 않았다.

시맨틱웹은 현재 인터넷이 단순링크에 기반을 둬으로써 발생하는 문제를 해결하는 대안으로 응용영역을 끝없이 넓혀가겠지만, 동시에 우리가 지금 시맨틱웹에서 바라는 기대치에는 도달할 수 없을 것이다. 따라서 시맨틱웹의 활용을 통한 정보과부하 문제의 해결과 더불어, 자연언어처리와 인공지능 기술도 함께 발전시키면서 시맨틱웹의 한계와 문제를 해결하는 방법만이 미래 정보환경에 대응하는 방법으로 본다.

참고문헌

- [1] Semantic Web Challenge, IEEE Intelligent Systems, May/June, 2004.
- [2] J. Davis et al, Towards the Semantic Web: Ontology-driven Knowledge Management, John Wiley & Sons, LTD, 2002.
- [3] J. Hjelm, Creating the Semantic Web with RDF, John Wiley & Sons, Inc., 2001.
- [4] W3C Recommendation, OWL Web Ontology Language, Feb., 2004.
- [5] <http://www.w3.org/2001/sw/>

권 혁 철



1982 서울대학교 공과대학 전산학(학사)
1984 서울대학교 공과대학 전산학(석사)
1987 서울대학교 공과대학 전산학(박사)
1988~현재 부산대학교 전자전기정보
컴퓨터공학부 교수
1988~현재 한국정보과학회 프로그래밍
언어 연구회 운영위원
1990~현재 한국정보과학회 한국어정보
처리 연구회 운영위원

1992~1993년 미국 Stanford 대학 CSLI연구소 연구원
1992~1993년 Xerox Palo Alto Research Center 자문위원
2003~현재 BK21 산업자동화 및 정보통신분야 인력양성사업단
단장
2004~현재 한국정보과학회 이사
2006~현재 한국인지과학회 이사
관심분야 : 한국어정보처리, 정보검색, 프로그래밍언어, 인공지능,
시맨틱웹
E-mail : hckwon@pusan.ac.kr
