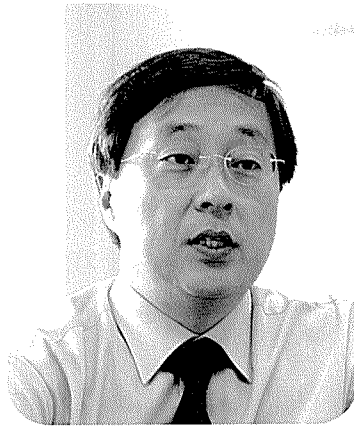


신택틱 (syntactic) 웹에서 '시맨틱 (semantic) 웹' 으로

글 | 이정우 연세대학교 정보대학원 교수 jlee@yonsei.ac.kr

● 언어기호화에서 정보를 분류할 때에 보통 세 가지로 구분한다. '신택스, 시맨틱, 그리고 프래그매틱스'. 구문론이라고 번역할 수 있는 신택스는 정확한 형태와 형식에 관한 이야기이고 시맨틱에서는 형식을 넘어서 의미의 전달에 관한 이야기를 하고 있다. 화용론이라고 번역되는 프래그매틱스에서는 언어 사용의 사회적인 행태를 연구한다. 화용론에서 다루는 개념들은 함축, 가변성, 타협성, 적응성, 보편성, 대화성 등으로서 언어의 구조적 의미를 넘어서 상황과 과업의 차이 그리고 개인적인 차이까지 포함하여 전달되는 사회적 관습과 문화에 관한 부분까지 포함한다. 데이터베이스로 말하자면 데이터 타입을 맞추는 것이 신택스이고 정합성을 유지하는 것이 시맨틱, 그리고 사용자가 필요한 때에 필요한 정보를 주는지를 분석하는 것이 프래그매틱스라고 쉽게 해석해 볼 수 있다.



李政雨
 서강대학교 경영학 석사(국제경영)
 미국 조지아주립대학교 경영학 박사(컴퓨터 정보시스템)
 미국 조지아주립대학교
 컴퓨터 정보시스템학과 전임강사
 미국 네바다대학교 경영정보학과 조교수
 현재 연세대학교 정보대학원 교수

XML (Extensible Markup Language)과 더불어 새롭게 떠오르는 시맨틱웹에서는 초점이 애플리케이션에서 데이터로 움직여 가고 있다. Daconta는 데이터가 스마트해지는 네 단계를 <그림>과 같이 그리고 있다.

2001년 5월 Tim Berners-Lee, James Hendler, Ora Lassila는 Scientific American의 기사를 통해 인터넷의 미래 비전을 발표했다. 이 비전은 시맨틱웹에 관한 것으로서 인터넷이 지금의 단순한 데이터의 원천이라는 위상에서 벗어나 지식과 이해의 베이스가 될 것이라는 전망이다. 시맨틱웹의 내용이 컴퓨터와 에이전트, 그리고 프로그램들에 의해 사용되고 최종 사용자에게는 실제적으로 나타나지 않

기 때문에 그동안 우리가 보아왔던 킬러 애플리케이션들과는 많이 다르다. 가트너의 보고서 'By 2005'에서는 시맨틱웹이 근간이 될 가벼운 온톨로지 구축 프로젝트가 시스템 통합 프로젝트의 75% 이상에서 포함되어 시행될 것이라고 예측하고 있다.

최고의 정보를 가지고, 그것을 어디에서 찾을지 알고 있으며 가장 빨리 이용할 수 있는 조직이 승리한다. 지금은 정보의 과잉 속에 압도되고 있고, 제대로 된 정보를 통해 결정을 내리는 것이 가능하도록 올바른 정보를 빠르게 찾을 수 있어야 한다. 또한 원천 데이터의 소유도 중요하지만 이를 지식으로 활용하는 게 더욱 중요하다.

너무 많은 정보, 키워드 검색의 비효율성, 신뢰할 수 있는 정보의 부족, 자연어 처리를 할 수 있는 컴퓨터 시스템의 부족 등이 나타나고 있으며 이를 해결하기 위해서는 정보의 구조화가 필요하다. 지식을 얻기 위해서 정보를 파일과 데이터베이스에 퍼붓는 것

시맨틱(semantic)웹의 개념

인터넷은 정보의 바다다. 수많은 정보들이 매일 쏟아진다. 하지만 아직 정보의 바다를 헤엄치는 데에 있어 우리가 컴퓨터로부터 받는 도움은 신택스 수준이다. 그렇다면 컴퓨터가 우리를 도와서 시맨틱을 프로세싱 할 수 있는 웹은 어떻게 하면 만들 수 있을까? 우선 데이터에 대한 생각의 전환이 필요하다. 역사적으로 데이터는 애플리케이션의 구성물로 한정되어 왔다. 즉 데이터는 데이터 프로세스 안의 부수적인 의미로만 인식되었다.

만으로는 부족하고 기계가 이해할 수 있는 마크업 언어를 우리가 가진 정보에 삽입해야 하고, 어떠한 정보를 신뢰할 수 있는 것인지 식별할 수 있어야 한다. 새로운 정보를 발견했을 때 그 정보가 신뢰할 수 있는 것임을 입증할 근거를 가져야 하고, 이미 가지고 있는 정보를 새로운 정보와 연관시킬 수 있어야 한다. 마지막으로 새로운 지식을 활용할 도구를 가져야 한다. 이것이 '시맨틱웹'의 기본적인 개념이다.

새로운 정보를 발견했을 때 그 정보가 신뢰할 수 있는 것임을 입증할 근거를 가져야 하고, 이미 가지고 있는 정보를 새로운 정보와 연관시킬 수 있어야 한다. 마지막으로 새로운 지식을 활용할 도구를 가져야 한다. 이것이 '시맨틱웹'의 기본적인 개념이다.

지어 말하고 있다. 그는 2002년 6월 Meet the Press에서의 인터뷰에서 "만일 FBI에 CIA로 연결된 컴퓨터가 있어서 비행학교를 검색할 수 있어, 지난 10년간 비행학교와 연관된 모든 보고서들이 검색되었다면 좋았을 것이다. 특히 인공지능을 이용해 쿼리를 할 필요 없이 보고서 내에서 그러한 형식의 것들을 볼 수 있는 것이다."라고 말했다.

Mueller가 묘사하고 있는 것은 사용자뿐 아니라 소프트웨어 에이전트도 이미 정부가 가지고 있는 데이터베이스의 데이터들 간의 숨은 관계를 찾을 수 있도록 해

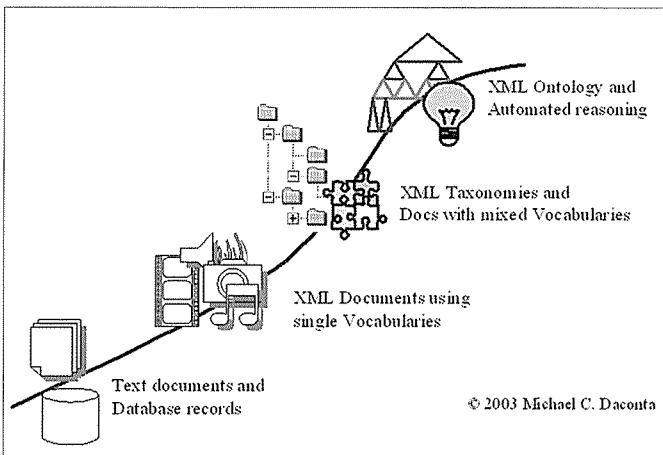
시맨틱웹의 효과

조직이 아직 활용되지 않은 수많은 정보를 가지고 있다는 것은 기회이다. 만약 조직이 거대하다면, 업무가 중복되는 프로젝트들을 가지고 있을 수 있다. 프로젝트를 통해 배운 교훈을 공유할 수 있고, 경쟁력 있는 지적 정보를 공급할 수 있고, 이에 많은 시간과 업무를 줄여 줄 수 있는 프로젝트를 가지고 있을 것이다. 만약 소프트웨어 에이전트를 통해 검색과 분석이 가능한 Knowledge Base를 보유하고 있다면, 많은 시간과 비용을 절감해주는 웹 기반 애플리케이션을 가질 수 있다. 다음 섹션들은 이들의 예를 제공한다.

의사결정지원

단지 데이터가 아닌 지식을 바로 이용할 수 있도록 함으로써, 보다 나은 의사결정을 내리는 것이 가능하다. 지난 10년간 정보기관들이 가지고 있었던 정보관리의 딜레마에 대해 생각해보자. FBI의 지휘관인 Robert Mueller는 이 문제를 9.11테러와 연관

(그림) The Smart Data Continuum



주는 시맨틱웹이다. 그의 말 속에는 상호운용성과 데이터 공유도 언급되고 있다. 대개 상이한 조직들은 상이한 데이터베이스와 서버를 가지고 있고, 그 조직만의 솔루션에 묶여있기 때문이다. 시스템 통합을 하는 사람들은 각기 다른 시스템들이 '대화할 수 있도록' 하기 위해 노력해 왔다. 웹서비스의 출현은 이러한 장벽을 없앨 수 있도록 해주고 있다. 많은 양의 정보를 가지고 있다면, 데이터 안에 내포되고 숨은 관계가 있을 것이다. 시맨틱웹을 사용함으로써 그것들을 찾아내는 데 도움을 얻을 수 있다.

기업들도 연방정부와 같은 정보관리의 딜레마를 가지고 있다. 기업들은 하부조직, 부서, 그룹, 프로젝트 등의 정보를 발생시키는 다양한 부분들을 가지고 있다. 이들 그룹의 능력을 열어주기 위해서는 그룹의 정보를 통합하고 그룹들 간의 관계를 이해해야만 한다. 가장 간단한 예는 상황보고 프로세스이다. 각 직원들은 상황보고서를 쓴다. 매니저는 이들 보고서를 취합하여 프로젝트 상황보고서를 작성한다. 프로젝트 매니저가 속한 부서의 매니저는 보고서를 받아 부서 상황보고서를 작성한다. 마지막으로 부서 매니저의 상급자는 부서 상황보고서를 모으고 요약서를 작성하여 기업의 총수에게 제출한다. 이 과정에서 정보는 여과되어 최종 보고서는 의사결정을 하는 데 사용되는 이해 가능한 보고서가 될 수 있도록 한다. 하지만 불행히도 중요한 정보들은 대부분 걸러내어진다(특히 개인 수준의 프로젝트에서 원수되는 업무들 간의 관계와 같은 것들).

최종사용자가 충분한 정보를 가지고 의사결정을 내리는 것이 가능하도록 의사결정에서 소프트웨어 에이전트의 분석(소프트웨어 에이전트가 데이터·정보를 분석한다는 의미), 최종사용자와 시스템 간의 상호작용에 중점을 둔 의미해석이 가능한 의사결정 지원시스템(DSS)을 만드는 작업이 완료되어 가고 있다. 의사결정 지원시스템이 아니더라도, 소프트웨어 에이전트는 지식 베

스를 모니터링하고 경고를 해주는 역할을 할 수 있다.

Enigmatic의 기술담당최고책임자인 Duncan Johnson-Watt는 2001년 Information Week의 논설에서 다른 예를 제시하고 있다. 그는 미증권거래위원회의 등록서류가 Semantic Tag를 포함하고 있다면, 규제하는 사람들이나 투자하는 사람들이 내부 주식 거래와 같은 경우에 자동으로 경고 메시지를 전해주는 프로그램을 만들 수 있을 것이라고 제안하고 있다. 우수한 의사결정을 내리기 위해서는 우수한 지식을 가지고 있어야 한다. 시맨틱웹은 그것을 달성할 수 있도록 해준다.

제안서 입찰도 기업 시맨틱웹의 중요한 활용분야이다. 잠재고객, 입찰되고 있는 업무, 요구되는 Skill Set 등에 대한 지식을 더 많이 가지고 있다면, 입찰에 이길 수 있는 더 나은 기회를 갖는 것이다.

비즈니스 개발

조직의 구성원들이 사업에서 성공하도록 도움을 줄 수 있는 최신 정보를 갖는 것이 중요하다. 대부분의 경우 조직들은 판매직원들과 함께 전문위원회의 구성원들을 파견할 수 있을 만큼의(금전적) 능력을 가지고 있지 못하다. 판매직원이 잠재적 고객과 만나고 있는 상황을 생각해보자. 얘기를 진행하면서, 판매직원은 고객이 어떠한 주제에 매우 관심있다는 것을 발견한다. 잠재적 고객은 “우리는 생체인식을 사용하는 온라인 전자상거래 시스템을 구축해 줄 기업을 찾고 있습니다.”라고 말한다. 만약 판매직원이 기업의 지식 베이스에 신속하게 접속할 수 있다면, 그 기회를 활용할 수 있는 중요한 정보를 찾을 수 있다. 기업의 지식 베이스를 신속하게 이용함으로써, 판매직원은 “우리는 어제 그 주제에 관하여 백서를 작성하였고, 지난 달에는 우리의 엔지니어들이 내부용 생체인식 솔루션의 초기 모델을 개발했습니다. 실제로 시스템을 살펴볼 수 있도록 일정을 잡아볼까요?”라고 신속하게 대응할 수 있을 것이다. 조직에서 운용되고 있는 시맨틱웹을 통해 새로운 비즈니스 기회를 잡을 수 있다.

제안서 입찰도 기업 시맨틱웹의 중요한 활용분야이다. 잠재고객, 입찰되고 있는 업무, 요구되는 Skill Set 등에 대한 지식을 더 많이 가지고 있다면, 입찰에 이길 수 있는 더 나은 기회를 갖는 것이다. 과거의 상황보고서, 제안서, 습득한 교훈, 경쟁력 있는 정보가 상호 연관되어 있는 지식 베이스를 가지고 있다면, 제안서를

위한 귀중한 정보를 활용할 가능성이 있다. 제안서를 작성하는 팀이 지식 베이스에 정보를 입력할 수 있고, 기업 내에 그 정보를 분석할 수 있는 소프트웨어 에이전트가 보유했던 상태라면, 에이전트는 가지고 있지만 그것을 깨닫고 있지 못한 정보 위에 ‘점들을 연결해 줄’ 수 있다.

전자상거래 산업의 전문가들은 시맨틱웹이 사용자들을 연결해주는 역할(matchmaking)을 할 수 있을 것이라고 믿는다. Matchmaking은 기업들이 잠재적 비즈니스 파트너 혹은 고객들과 접촉할 수 있도록 하는 프로세스이다. 전통적으로 이 프로세스는 고용된 브로커에 의해 처리되었으며, 많은 이들이 서비스의 홍보를 처리하고 홍보되는 서비스들에 대한 쿼리를 할 수 있는 Matchmaking 서비스를 개발하는 것을 제안해왔다. 전문가들은 시맨틱웹 기술이 이러한 요구사항들을 충족시킬 수 있다고 주장하고 있으며, 시맨틱웹이 Matchmaking과 협상을 자동화할 수 있을 것이라고 믿고 있다. 시맨틱웹을 통해 비즈니스 기회를 극대화하기 위한 기회는 무한하다.

정보공유와 지식 발견

정보의 공유와 의사소통은 어느 조직에서나 매우 중요한 일이지만 대부분의 조직들이 성장하고 더 많은 정보를 모으게 되면서, 이는 주요 고심거리가 된다. 우리는 모두 바퀴를 또 다시 창안해 내지 않는 것의 중요성을 이해하고 있지만, 얼마나 많은 노력의(의도하지 않은) 중복이 일어나는가? 조직이 거대해져 가면서, 의사소통의 격차는 불가피해진다. 약간만 노력한다면 기업의 지식 베이스에 적어도 프로젝트에 관한 설명과 각 팀이 하고 있는 작업의 레지스트리를 포함할 수 있다. 직원들이 얼마나 쉽게 관련된 정보를 찾을 수 있게 될 것인지 생각해보라. 시맨틱웹이 접합된 웹서비스를 사용함으로써 그러한 레지스트리를 만들 수 있다.

운영과 자동화

지금까지 우리는 조직 내에서의 지식 공유에 관한 다소 명백한 예시들에 대하여 논의했다. 그러한 지식 베이스를 갖추는 것의 부수적인 효과는 운영 업무를 자동화 하는 소프트웨어 프로그램의 능력이다. 예를 들어 여행 예약은 시맨틱웹과 웹서비스가 고통스러운 업무를 쉽게 해주는 예시의 하나이다. 여행객 배치를 하는 것은 운영상의 악몽과 같다. 각각의 사람들이 선호하는 여행 스케줄을 가지고 있고, 다음과 같은 항목들을 고려해야만 한다.

- 선호하는 교통수단(자동차, 기차, 버스, 비행기)
- 선호하는 호텔과 호텔에 관련된 보상

- 선호하는 항공회사와 자주 사용하는 할인제도
- 합류점과 호텔의 근접도
- 선호하는 호텔방(금연, 큰 사이즈, 바, 로비의 무선 네트워크)
- 차량대여 옵션과 보상
- 가격(여행사에 대한 1일 기준 숙박과 교통비)

여행객 배치의 결정을 내리는 순서도를 만드는 것은 복잡한 프로세스가 될 수 있다. 예를 들어 여행거리가 100마일 미만이라면 자동차를 대여할 것이다. 여행거리가 100 ~ 300마일 사이라면 기차나 버스를 탈 것이다. 여행거리가 300마일 이상이라면 비행기를 탈 것이다. 비행기를 타게 된다면, 자주 사용하는 할인제도를 써서 일등석을 예약하지 않는 한, 가장 저렴한 비행기표를 찾을 것이다. 비행기표를 예약한다면, 채식주의 식단을 원할 것이다. 합류점과 호텔 간의 거리에 따라 호텔의 가격을 달리하기를 원할 것이고, 선호하는 방이 있을 것이고 등등... 간단한 여행예약의 논리를 맵핑하기 시작하면서부터, 당신은 이것이 수시간은 걸릴 복잡한 작업이라는 것을 깨달을 것이다.

직접 손으로 배치를 마무리 짓는 것은 많은 시간을 소요할 것이다. 다행히 시맨틱웹과 웹서비스의 통합으로 작업의 대부분은 자동화될 수 있다. 모든 법칙과 개인적으로 선호하는 여행방법을 기업의 지식 베이스에 가지고 있다면, 갈등 해소를 위한 원칙으로 설정된 기계가 읽을 수 있는 형태의 법칙들을 사용하여 여행 애플리케이션이 여행 배치를 선택해 줄 것이다. 온라인 여행·호텔 서비스 상의 (관계가 정의되어 있는) Semantic Tag에 접속함으로써, 여행 애플리케이션은 옵션들을 비교, 대조, 평가하고 가장 적합한 것들의 리스트를 출력할 것이다. 즉 시맨틱웹이 접합된 웹서비스는 조직 내의 지루하고 복잡한 업무를 자동화할 수 있는 잠재력을 가지고 있다.

시맨틱웹을 위한 기술을 이미 보유하고 있는가?

인터넷에 시맨틱웹을 적용하는 것은 아직 비전이지만, 시맨틱웹을 위한 기술적인 구성요소들은 소규모이지만 국지적으로 개발되어 구축되고 있다. 비전을 현실로 만들 조각들이 조금씩 구성되어 가고 있는 것이다.

시맨틱웹이 인터넷 상에서 데이터의 통합을 지향하고 있는 바 개방표준이 절실히 필요한 데 지난 5년간에 걸쳐서 이러한 개방표준으로의 전환이 이루어져왔다.

W3C, Internet Engineering Task Force(IETF), Organization for the Advancement of Structured

Information Standards(OASIS)는 기업과 학술기관으로부터 상호운용성을 위한 폭넓은 지원을 받아왔다. XML의 지원은 SOAP기반 웹서비스(인터넷 상에서 상호운용이 가능한 인터페이스를 제공)와 같은 XML기반 기술의 지원을 낳았다.

RDF는 정보 요소들간의 관계를 실제적으로 모델링할 수 있는 방법을 제공한다. XML을 선택으로 사용하면서 RDF는 시맨틱웹의 온톨로지의 기초가 된다. XML Topic Map(XTM)은 데이터를 분류하기 위한 정보의 Taxonomy를 표현하기 위한 수단이며 웹서비스는 이러한 데이터들을 활용하는 소프트웨어 프로그램들이 서로 의사소통을 하기 위한 수단이 된다.

웹서비스에 있어 산업계의 최근 초점은 상이한 조직의 웹 서비스로부터 교환되는 정보를 활용하여 상호 통합과 조정을 이룰 수 있는 시맨틱을 이르는 웹서비스이다. 기업들은 시맨틱웹과 관련된 기술에 많은 비용을 투자하고 있다.

온톨로지 언어(OWL, DAML+OIL)는 이제 활용될 만큼 표준화 최적화가 되었으며 많은 기업들이 지식 베이스에 시맨틱을 집어 넣어 확장하기 위해서 이들 언어들을 사용하고 있다.

웹서비스에 있어 산업계의 최근 초점은 상이한 조직의 웹 서비스로부터 교환되는 정보를 활용하여 상호 통합과 조정을 이룰 수 있는 시맨틱을 이르는 웹서비스이다. 기업들은 시맨틱웹과 관련된 기술에 많은 비용을 투자하고 있다. Adobe는 RDF를 중심으로 소프트웨어 메타데이터를 재구성하고 있고, 문서관리를 위한 웹 온톨로지를 사용하고 있다.

이 변화로 다른 소프트웨어가 PDF 파일에 있는 정보를 이해하는 것이 가능해졌다. IBM은 최근 the Institute of Search and Text Analysis in California를 설립하고, 시맨틱웹 연구에 크게 투자하고 있다. 독일의 Ontoprise와 같은 기업은 온톨로지를 사업화하여 지식 모델링, 지식 검색, 지식 통합 도구를 만들어내고 있다.

인터넷이 우리 눈앞에 다가 오는 데 몇 년 안 걸렸듯이 시맨틱웹이 우리 눈앞에 다가와 있음을 느낄 날이 멀지 않았다. **Users**