



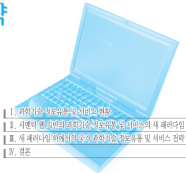
과학기술정보통신부 (정보유통 및 서비스)

시맨틱 웹 기반의 차세대 과학기술 정보유통 및 서비스 전략



김 성 혁 숙명여자대학교 문헌정보학과 교수

ksh@sookmyung.ac.kr

- 
- I. 과학기술 정보유통 및 서비스 현황
 - II. 시맨틱 웹 기반의 과학기술 정보유통 및 서비스의 새 패러다임
 - III. 새 패러다임 하에서의 국가 과학기술 정보유통 및 서비스 전략
 - IV. 결론

본 글은 웹과 디지털도서관이라는 환경 하에서 향후 정보유통 및 서비스의 전략, 특히 최근 확산되고 있는 시맨틱 웹(The Semantic Web)과 온톨로지(ontology)를 기반으로 하는 정보서비스의 새 패러다임을 제시하였다. 특히 국가의 과학기술 정보의 유통과 서비스를 책임지고 있는 KISTI 측면에서 현대의 정보유통 및 서비스의 문제점을 찾아보고, 이를 해결하기 위하여 시맨틱 웹과 온톨로지를 이용한 차세대 정보유통 및 서비스를 왜, 무엇을, 어떻게 그리고 누가 할 것인가에 대한 전략을 펼쳐가 지금까지 KISTI 평가와 경영진단에 직·간접으로 참여했던 경험과 연구를 중심으로 제시하였다.

1. 과학기술 정보유통 및 서비스 현황

미국 캘리포니아 대학 버클리의 켈퍼스에 간행된 최근 보고서에서 레이먼 등은(Peter Lyman and Hal Varian) 테라 바이트 전 세계적으로 1~2 엑사바이트(exabytes)의 정보가 생산된다고 하였다. 이 정보의 대부분은 텍스트, 이미지, 소리, 수치데이터 등인데, 이 중에서 인쇄된 문서는 단지 0.003% 만을 차지한다고 하였다. 이렇게 생산되는 정보의 많은 부분이 디지털 형태로 생산, 저장 및 검색될 수 있으며, 실제로 이들 정보의 90%는 오늘날 디지털 형태로 저장되고 있다. 그러나 이들 정보의 극소수만이 정보센터나 디지털도서관의 장서(디지털과 인쇄형태로 된 정보 자료의 집합, 즉 하이브리드 장서)를 통해 서비스되

고 있다.

정보센터나 디지털도서관의 장서에 장애 없이 보편적이고 동등하게 접근을 제공하는 것은 모든 국민의 활동, 예를 들어 교육, 직장, 오락, 문화, 연구 등에 지대한 영향을 미치게 될 것이다. 기사의 장벽을 없애고, 시기적절한 자원의 공유 및 콘텐츠 전달을 지원하는 것은 모든 국민의 삶의 질과 작업의 생산성을 크게 향상시킬 수 있다. 그러나 앞서 지적한 방대한 양의 정보를 어떻게 정보센터나 디지털도서관의 장서로 포함시켜 국민들에게 정보서비스를 제공할 것인가는 매우 중요한 과제이다.

1990년대 초반 웹과 디지털도서관이 등장하면서 전 세계의 정보유통 및 서비스는 양적 및 질적인 측면에서 전대미문의 변화를 가져왔다. 특히 디지털 정



보의 양적인 증가는 정보의 품질과 저작권 문제를 야기시키면서 정보센터, 디지털도서관, 시스템사업자 및 이용자에게 항상 정보의 접근성(accessibility), 획득성(availability), 확장성(scalability)에 관한 관심을 유발시키고 있다.

디지털도서관 연구 및 관련 기술은 이러한 문제를 해결하기 위한 노력이라고 판단된다. 그러나 유감스럽게도 디지털도서관 연구는 국내에 유입되면서 디지털도서관 기술이 미래의 정보유통 및 서비스에 끼칠 거대한 비전을 보지 못하고 단지 인쇄형태의 문헌을 디지털화 하여 정보유통 및 서비스를 온라인으로 수행하는 시스템 정도로 인식되었다고 판단된다. 즉, 디지털도서관 연구를 도서관이라는 협의적인 물리적 건물로 여겼기 때문에 디지털도서관이 리드하는 기술의 중요성을 간과하였던 것이라고 본다. 오늘날 미국, 유럽연합, 호주 등에서는 디지털도서관을 차세대 정보유통 및 서비스의 플랫폼(platform) 내지는 미들웨어로 확장해 가고 있다. 하나의 예로, 최근 그리드(Grid)는 미들웨어를 통한 이 기종 컴퓨터이션 및 데이터 자원을 통합하여 데이터 마이닝과 같은 고차원의 지식을 추론하려는 응용기술을 개발하고 있다. 유럽연합의 보고서에 따르면 이러한 응용기술은 디지털도서관에서 제공하는 기능 중의 하나이기 때문에 그리드는 디지털도서관의 특별한 사례로 볼 수 있다고 하였다.

본 글은 웹과 디지털도서관이라는 환경 하에서 향후 정보유통 및 서비스의 전략, 특히 최근 확산되고 있는 시맨틱 웹(The Semantic Web)과 온톨로지(ontology)를 기반으로 하는 정보서비스의 새 패러다임을 제시하였다. 특히, 국가의 과학기술 정보의 유통과 서비스를 책임지고 있는 KISTI 측면에서 현재의 정보유통 및 서비스의 문제점을 찾아보고, 이를 해결하기 위하여 시맨틱 웹과 온톨로지를 이용한 차세대 정보유통 및 서비스를 왜, 무엇을, 어떻게 그리고 누가 할 것인가에 대한 전략을 필자가 지금까지 KISTI 평가와 경영진단에 직간접으로 참여하였던 경험과 연구를 중심으로 제시하였다. 따라서 본 글은 KISTI의 차세대 정보유통 및 서비스의 해결책이라기보다는 해결 해 나가야 할 과제에 이해하는 것이 바람직하다고 할 수 있다.

II. 시맨틱 웹 기반의 과학기술정보 유통 및 서비스의 새 패러다임

지금까지 우리는 정보유통 및 서비스를 위한 다양한 시스템을 개발하고 운영하여 왔다. 그 결과 우리는 지금 수많은 검색시스템, 유통 및 서비스시스템, 검색연진, 포털시스템 등 서로 이질적인 시스템 속에서 원하는 정보를 찾고, 공유 및 교환하기 위한 사무를 벌이고 있지만 아직도 뚜렷한 해결책을 찾지 못하고 있다. 잉딩(Ying Ding) 등에 의하면 아직도 정보 탐색, 정보추출, 정보원 유지 및 자동문서생산 등에서 다음과 같은 문제점을 갖고 있다고 지적하였다.

첫째로 정보탐색의 경우, 기존의 키워드 기반 검색 시스템에서는 사용하는 용어의 모호성으로 인해 부적합한 문헌이 검색되기도 한다. 또한 유사한 자료가 거의 다른 용어로 표현 및 저장되어 있기 때문에 적합한 정보를 검색하지 못하는 문제가 있다.

둘째로 정보추출의 경우, 오늘날 대부분의 사용자는 직접 원하는 정보를 추출하기 위하여 검색 결과에 대한 브라우징과 자세한 읽기를 통해서 만이 가능하다. 즉, 소프트웨어 에이전트가 이러한 서비스를 수행하기 위해서는 아직 그 기술이 부족하다. 나아가 서로 이질적인 정보원을 통합하는 문제도 해결되지 못하고 있다.

셋째로 서로 이질적이고 미 구조화된 대규모의 정보원을 구축하고 유지하는 것은 많은 비용과 시간이 소요되는 일이고, 마지막으로 이용자의 프로파일, 또는 특정의 요구에 따라 정보원이 적용하고 동적으로 변화되어 이질적인 정보원을 통합한 새로운 문서를 자동생산하기 위한 기술도 아직은 부족하다.

이러한 문제점을 해결하기 위해서는 시스템적 접근이 아닌 환경적 접근을 통해서 만이 가능하다. 즉, 이질적이고 다양한 시스템들이 공존하는 환경이 유지되기 위해서는 시스템들 간, 프로그램들 간의 정보나 데이터의 의미적인 소통이 될 때 가능하게 되는 것이다. 지금까지 정보는 기계가독형(machine-readable) 형태로 축적되고 이용되어 왔지만 미래에는 기계가 이용 가능한(machine-usable) 또는 기계 처리가 가능한(machine-processable) 또는 기계가



이해 가능한 (machine-understandable) 형태로 정보가 축적되어야만 한다.

시맨틱 웹은 1990년대 초 웹을 처음 개발한 베르너 리(Tim Berner-Lee) 등에 의해 'Scientific American' 2001년 5월 호에 처음 소개되면서 태동된 기술이다. 지금까지 웹을 지탱하고 발전시킨 모토가 있다면 그것은 '단순화(simple)'이었을 것이다. 이러한 웹의 단순화는 웹 발전의 장애물이 되었다. 즉, 자동화되고 지능적인 웹이 되기 위해서는 컴퓨터 프로그래밍과 프로그램, 또는 에이전트 간(소프트웨어와 사람) 주고받는 데이터나 정보의 의미를 서로 공유할 수 있어야 가능하다는 것이다. 이렇게 시작된 시맨틱 웹의 근간은 온톨로지를 통해 구현될 수 있다고 한다. 온톨로지에 대한 정의는 매우 다양하지만 정보기술 측면에서 본다면 그루버(T. R. Gruber)가 내린 정의가 가장 적합하다고 볼 수 있다. 그는 온톨로지란 '공유되는 개념화의 공식적이고 명시적인 규격'(a formal, explicit specification of a shared conceptualization)으로 정의하면서 '개념화'는 실세계 현상의 추상화 모델로, '명시적'은 사용하는 개념들의 유형과 개념들의 사용에 따른 제한 등이 분명하게 정의되어야 함을, '공식적'은 온톨로지가 기계가독형 형태로 되어 있어야 한다는 것을 의미한다.

시맨틱 웹을 구현하기 위한 온톨로지 구축 표준과 기술은 W3C를 중심으로 하는 RDF와 OWL 그리고 ISO/IEC JTC1/SC34를 중심으로 하는 토픽 맵 관련 기술로 요약될 수 있다. 이들에 대한 자세한 설명은 지면 제약 상 다루지 않았다. 다만 아래 (그림 1)을 통해 그 관계를 파악할 수 있을 것으로 판단된다. 최근에는 W3C를 중심으로 두 표준간 호환을 위한 연구들이 진행되고 있다.

TMCL	표현법	OWL RDF Schema
Topic maps		RDF
XTM, HyTM, LTM	문법	RDF/XML, N3

그림 1 두 표준간의 비교

지금까지 과학기술 분야의 다양한 포털로 운영되어 온 정보유통 및 서비스는 시스템적으로 해결하기 위한 노력이었다면 새로운 정보유통 및 서비스는 시맨틱 웹과 온톨로지를 이용한 새로운 환경 구축으로 전환되어야 할 것이다. 정보센터 측면에서는 자동화된 정보유통 및 서비스를, 이용자 측면에서는 지능적이고 이용자 중심적인 서비스를 시스템 측면에서 사람이 개입되지 않는 시스템 운용 등 특정 시스템 의존적인 정보 유통에서 이질적이고 분산적인 환경, 즉 디지털도서관 플랫폼에 시맨틱 웹과 온톨로지가 지원되는 새로운 패러다임을 창조하고 구축하여야 한다.

III. 새 패러다임 하에서의 국가 과학기술정보 유통 및 서비스 전략

KIST는 과학기술 정보유통 및 서비스와 슈퍼컴퓨팅서비스를 앞대 축으로 운영되고 있는 국가출연 정보센터이다. 본 글에서는 정보유통 및 서비스 측면에서만 다루고 슈퍼컴퓨팅 부분은 제외하였다. 그러나 필자의 개인적인 견해로는 슈퍼컴퓨팅서비스도 디지털도서관이라는 범주에 속하기 때문에 정보유통 및 서비스와 매우 밀접한 관계가 있다고 판단된다.

디지털도서관과 시맨틱 웹 기반의 새로운 패러다임은 분산환경, 이질적인 정보, 협업체계, 정보의 디지털화 등을 전제로 하고 있다. 특히, 과학기술 정보의 대부분을 외국 정보에 의존하고 있는 국내의 상황을 고려할 때, 새 패러다임에 적합한 KIST의 정보유통 및 서비스 체계 구축은 매우 중요하다고 할 수 있다. 따라서 본 글에서는 다음과 같은 전략 제안을 통해 새 패러다임의 방향을 설정하였다.

첫째로 과학기술 1차 정보원의 확보가 절대적으로 필요하다. 국가적 차원에서 KIST는 과학기술도서관의 역할을 담당하여야 한다. 1차 정보원의 확보 없는 정보유통 및 서비스는 명분에 불과하다. 이미 참여정부에서 이에 관한 구체적인 논의가 진행되고 있다는 것은 매우 고무적이라고 할 수 있다.

둘째로 국내 및 전 세계에서 생산되는 모든 과학기술 정보를 KIST가 수집하여 서비스하기에는 불가능



하다. 따라서 KIST는 국내에서 유통되는 과학기술 정보의 국가 메타 데이터 레지스트리(National Metadata Registry of Science and Technology Information) 역할을 수행할 수 있는 체계를 구축하여야 한다. 이와 병행하여 수집된 1차 정보원 및 특정 분야의 레포지토리(repository) 기능도 필요하며, 동시에 국내외 다른 정보센터와의 협업체계 구축이 필요하다.

셋째로 정보유통 및 서비스와 관련된 표준화와 연구 기능이 필요하다. KIST는 국내의 정보유통 및 서비스를 리드하는 역할을 담당하여야 한다. 현재의 체계로는 이를 가능하게 하기에는 역부족이라 판단된다. 표준화와 연구는 전문 인력의 확보로 통해 가능하다. 남의 것을 흉내 내거나 모방하는 표준화와 연구가 아닌 미래 지향적이고, 창의적인 아이디어가 필요한 시점이다. 이는 표준화와 연구에 국한되는 것이 아니라 정보유통 및 서비스에 모두 필요하다.

넷째로 새로운 환경과 새로운 기술에 적용할 수 있는 정보유통 및 서비스 체계의 구축이 필요하다. 정보기술의 발전 속도는 우리가 예측한 것 보다 한발 앞서 나가고 있다. 따라서 항상 변화하는 환경에 적용할 수 있어야 한다. 이는 이 부분에 종사하는 전문가에게도 그대로 적용된다고 볼 수 있다.

다섯째로 정보유통 및 서비스에 종사하는 KISTI 전문가들의 태도에 대한 변화가 필요하다. KIST에 종사하는 모든 전문가는 정보센터나 도서관이 서비스기관이라는 것을 명심하여야 한다. 서비스기관은 고객이 필요하다. 고객이 없는 서비스기관은 존재할 가치가 없는 것이다.

IV. 결론

웹 서비스와 관련된 표준 및 기술의 등장, 전자상거래와 관련된 표준 및 기술, 그리고 온톨로지 기반의 시맨틱 웹 등장 등으로 그 어느 때보다 정보유통 및 서비스는 변화의 소용돌이 속에 있다. 다양한 표준과 기술들의 병합에 의해 새로운 표준과 기술들이 등장하고 이를 정보유통 및 서비스에 활용하기 위한 다양한 연구가 진행되고 있고, 일부는 이미 현장에서 활

용되고 있다.

지금까지의 정보유통 및 서비스는 시스템 의존적이었지만 디지털도서관 플랫폼과 온톨로지 기반의 시맨틱 웹을 적용한 새 패러다임 하에서의 정보유통 및 서비스는 다양하고 이질적인 시스템이 공존하는 환경 구축이 필요하다. 이러한 환경이 구축되기 위해서는 무엇보다도 정보유통 및 서비스와 관련된 표준화가 선행되어야 한다. 온톨로지도 표준화의 결과로 볼 수 있다.

결론적으로 본 글에서 제시한 차세대 정보유통 및 서비스 전략이 실현되기 위해서는 무엇보다도 왜 새로운 방향과 전략이 필요한지에 대한 KISTI 내의 공감대가 필요하다. 이미 본 글의 초안에서 이에 대한 설명을 하였다. 그 다음 무엇을 할 것인지에 대한 구체적인 아젠다가 등장하여야 한다. 새 패러다임으로 가기 위해서 미래의 비전을 제시하고, 그에 적합한 정보유통 및 서비스 프레임워크를 만들고 여기에 들어갈 콘텐츠 관리, 이용 및 영향력 등을 구체적으로 연구하여야 한다. 나아가 누가, 어떻게, 이를 수행할 것인가에 대한 수행 주체와 방법론에 대한 연구가 필요하다.

그러나 표준화와 기술의 개발 및 발전은 궁극적으로 인간에게 달려있다. 즉, 이러한 변화의 핵심은 인간이고, 인간만이 담당할 수 있기 때문에 KISTI 종사자들의 패러다임 변화가 무엇보다도 중요하다. 이미 21세기 시맨틱 웹의 시대에 들어있는데도 아직까지 19세기 혹은 20세기적인 패러다임 머물러 있다면 정보유통 및 서비스는 발전할 수 없다. 정보유통 및 서비스가 어려운 이유는 정보검색에는 정답이 없기 때문이다. 그러나 정답에 가까운 해결책을 찾기 위한 노력은 인류의 역사가 계속되는 한 멈추지 않을 것이다. **KISTI**

참고 문헌

- [1] Brainstorming report, 2001. DIGITAL LIBRARIES - Future Directions for a European Research Programme, June 13-15, 2001.
- [2] Ying Ding, Dieter Fensel, Hans-Georg Stok, The Semantic Web: from Concept to Percept, OGI, 2003.
- [3] Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001.
- [4] T. R. Grube: A Translation Approach to Possible Ontology Specifications, Knowledge Acquisition, 5: 199-220, 1993.
- [5] James Hendler: Agents and the Semantic Web, the IEEE Intelligent Systems Journal, march/April 2001.