

『온톨로지 알고리즘 I』  
(齊藤孝 저, 최석두·한상길 역, 한울  
아카데미, 2008)

이 미 영\*

‘온톨로지’ 그리고 ‘알고리즘’.

쉽지 않은 두 단어의 조합만으로도 이 책에서 다룰 내용이 전자적 환경위에서 그에 부응해야 하는 지금 시대를 이야기할 것임을 짐작할 수 있을 것이다. 최근 기록학의 논문주제가 매우 다양해진 한편, 기록관리시스템 설계, 온라인 서비스 및 콘텐츠 개발, 폭소노미 도입을 통한 검색기능 개선 등 전자적 환경을 전제한 논문이 많은 비중을 차지하고 있는 게 사실이다. 이런 시점에서 『온톨로지 알고리즘 I : 기록·정보·지식의 세계』의 출판은 반가운 소식이다.

저자는 지적 체계를 체계화하려는 욕망의 결정이 분류이며, 나누기 위해 고심한 결과가 개인의 암목지가 되고 결국 주지의 형식지로 응축되며, 온톨로지 알고리즘이란 이 지적 아픔의 결

---

\* 명지대학교 기록정보과학전문대학원 박사과정 수료. 인문학박물관 연구원·사서

정이라고 이 책을 시작한다. 즉 온톨로지 알고리즘은 인간의 지성을 밝히는 지식 모형인 것이다. 그리고 저자는 역사를 알고 날마다 철학하는 일을 바로 온톨로지 알고리즘의 원점으로 보고 이 책을 통해 기록정보학에서 필요불가결한 온톨로지 알고리즘을 재구축하고자 하였다. 따라서 이 책은 철학, 논리학, 언어학 분야의 기초원리 뿐만 아니라 정보과학, 인지과학, 소프트웨어 공학 등 시대발전이 낳은 학문분야도 담고 있다.

이 책에서 온톨로지는 세 가지로 구분한다. ‘존재론’과 ‘인식론’으로부터 그 계보를 갖는 철학적 온톨로지가 첫 번째이며, 자연언어처리의 개념 사전이나 시맨틱웹의 온톨로지 기능을 담고 있는 사전적 온톨로지가 두 번째, 그리고 정보학적 온톨로지가 세 번째 이다. 기록정보학과 전자도서관, 사회정보학이 전공 분야라는 저자는 정보학과 시스템측면에만 치우치지 않고 철학적 바탕의 온톨로지 개념과 매우 전자적인 시스템을 넘나들며 매우 이론적이면서 시스템적인 이야기를 아우르고 있다.

역자서문에서도 밝히고 있듯이 이 책의 원서명은 『記録・情報・知識の世界：オントロジ・アルゴリズムの研究』(기록·정보·지식의 세계: 온톨로지 알고리즘의 연구)로, 동저자가 후속편으로 출판한 『意味論からの情報システム：ユビキタス・オントロジ・セマンティックス』(정보시스템의 의미론: 유비쿼터스·온톨로지·시맨틱스)와 전·후편으로 이어져있고 주제는 같으나 관점이 다르다는 것을 나타내기 위하여 서명을 일치시키고 번호와 부서명을 붙여 I, II로 구분하였다. 그리고 이 서평은 전편에 대한 것이다.

총 8장으로 구성되어 있는 이 책은 한 장 한 장 읽어다가만 만만치 않다.

제1장은 ‘전체상’이란 제목의 도입부분으로 기록, 정보, 지식

의 정의와 각각의 사이클과 알고리즘에 대해 설명하고 있다. 이 책에서 보는 기록은 기록하는 ‘행위’이자 기록화의 결과인 ‘기록물’을 포함하는 것으로 기록정보의 대상은 기록물로서의 형태를 갖춘 것이다. 구체적으로 도서·문헌, 신문, 영화, 음악, 멀티미디어 출판물, 전자 출판물 등이 있으며, 그 형태라고 하는 것은 물리적인 것뿐만 아니라 논리적인 것도 포함하고 있기 때문에 프로그램의 성과물인 소프트웨어나 인터넷상에 떠다니는 웹도 ‘기록물’로 보고 있다. 좀 더 독특한 정의는 기록이라고 하는 행위에서는 ‘주제’를 명확하게 설정하고 그것을 중심으로 편집이나 출판이라는 기록처리(documentation)가 이루어진다는 것으로, ‘주제’가 정해지지 않아 모호한 것은 기록이 아니라고 보고 있다. 따라서 주제는 기록정보의 핵심이 된다.

제2장 ‘기록’에서는 종래의 문자를 이용하여 편집된 인쇄물의 기록 알고리즘은 전형적인 텍스트로 보고, 하이퍼텍스트는 문자·음성·영상 등의 개별 매체를 통합한 멀티미디어를 실현가능한 기록 알고리즘으로써 다차원적이고 비선형구조라고 정리하고 있다. 이어서 노드와 링크로 구성된 하이퍼텍스트의 지식구조와 시스템을 설명하면서 기록의 대상이 되는 주제를 명확하게 하는 주제의 생성단계에서 아이디어 생성이라는 온톨로지 알고리즘이 필요한데, 이 구조는 하이퍼텍스트적이며 아이디어가 외재화되어 비선형 아이디어 네트워크가 형성되어 가고 이것이 정련, 편집되어 간다는 것이다. 인터넷을 통해 개인은 자유롭게 기록을 하고 유사 전자기록을 발신할 수 있지만 이는 온톨로지 알고리즘에 기반한 진정한 기록이라고 말할 수 없다는 것이 저자의 주장이다.

제3장 ‘정보’에서는 기록은 주제 분석(정보 알고리즘)에 의하여 정보화된다고 전제한 뒤, 주제 분석의 목적과 실례, 범주, 순

서, 연구과제 등을 다루었다. 기록정보학에서 주제분석이란 기록물로부터 내용을 파악하고 주제라고 하는 개념을 명확하게 하는 것으로 구체적으로는 분류나 색인을 이용하여 정보를 조직화하는 것이며 이 과정에서 대상 기록물에 대한 정보 맥락에서 보는 관점이 중요하다고 지적한다.

제4장 ‘지식의 지식’에서는 사물과 다름없는 기록과 정보의 존재론, 즉 온톨로지를 다양한 철학적 개념(철학, 논리학, 형이상학, 수학, 기호론과 의미론, 언어학 등)에서부터 살펴보고 있다. 온톨로지는 존재를 의미하는 것과 그 속성으로서의 개념, 그리고 그 체계를 밝히는 것이라는 저자는 다양한 방법으로 기록이 담고 있는 의미의 형성과 작용, 처리에 대한 설명을 이론적으로 정리하였다.

제5장 ‘분류학과 분류시스템’에서는 ‘안다, 나눈다’라는 온톨로지의 기본 개념을 이해하고 분류시스템과 시소러스를 설명하고 있다. 그런데 이 장에서 설명하고 있는 분류시스템은 NDC, DDC, UDC, 패킷분류 등 도서분류시스템이며 시소러스는 기본적인 개념과, 원리, 표현방식만을 다루는 한계가 있다.

제6장 ‘지식 모형’에서는 기록정보학의 정보 모형을 시작으로 인공지능의 지식 모형, 지식 공학의 지식 모형, 인지과학의 지식 모형, 소프트웨어 공학의 지식 모형, 지식관리의 지식 모형에 이르기까지 각 분야의 정보 모형에 대한 접근을 하고 있다. 기록정보학의 지식은 분류하는 과정에서 암묵지로 생겨나는데, 즉 정보를 일정한 구분원리로 분할하고 그것을 어떤 단위의 이용가능한 형태로 구조화한 결과로 얻을 수 있으며 이때 지식의 표현 단위는 그 지식을 이용하는 목적의 관점에서 구조화해야 할 필요가 있다. 원래 지식 모형이란 외부표현만으로 구축할 수 없으며 내부표현을 포함해야 하기 때문에 인지과학적 접근방법

도 중요한 것이다.

제7장 ‘지식 지도’에서 다루는 지식 지도란 온톨로지 알고리즘을 그림으로 설명한 것이며, 지식 지도를 대규모로 만든 것이 분류시스템이다. 이 장에서도 역시 다양한 분야의 지식지도와 그 특성에 접근하고 있는데, 지식이라는 것은 사실 또는 개념 그 자체뿐 아니라 이들 상호간의 관계성에 의해 표현되기 때문에 다양하게 도해되기도 하고 연결되기도 한다는 것을 그림으로 보여주고 있다.

마지막 제8장 ‘온톨로지 알고리즘 구현’에서는 철학의 언어인 온톨로지가 정보처리의 세계에서 갑자기 각광을 받게 된 것은 시맨틱웹(Semantic Web)이라 불리는 인터넷 애플리케이션에서 사용되었기 때문이란 내용으로 시작하고 있다. 온톨로지 알고리즘은 다양한 모형화에 의하여 시스템으로 설계되고 실제 컴퓨터에서 가동시킬 수 있는 소프트웨어로 구현되어, 대규모로써는 디지털 도서관 또는 시맨틱웹 메타데이터 개념사전으로 개발된다. 이후 시맨틱웹에 관한 설명에서 시맨틱웹의 온톨로지 기능은 공유되는 개념화의 형식화(대상으로 하는 현상의 추상적인 모형으로, 흥미를 갖는 개념과 그 개념간의 관계를 밝히고 메타데이터 등으로 적용하는 것)와 명시적(개념의 타입과 개념간의 정의가 명시적)이어야 한다는 조건을 만족시켜야 한다. 이외에도 서지 암묵지를 획득하기 위한 지식 모형인 K-Agent, K-Map을 중점 내용으로 다루고 있다.

지금까지의 내용으로 구성된 이 책은 정보학을 전공하지 않은 이들에겐 다소 이해하기 어려운 것이라 짐작된다. 더군다나 현대기록학의 아버지라는 쉘렌버그(Schellenberg)의 『현대기록학개론』을 통해 기록의 개념과 가치, 기록과 도서의 차이, 기록관리기관과 도서관의 차이, 그리고 그 각 기관에서의 자료관

리방법론의 차이부터 공부해 온 사람들에게 이 책의 기록, 정보, 지식, 도서의 개념은 혼란스럽거나 받아들이기 어려운 부분도 있을 거라 생각된다. 하지만 그 모호한 경계야말로 우리가 더욱 고민하고 연구해야 할 새로운 지점일 수 있다.

기록이 다양한 정보의 컨테이너라는 점에는 이견이 없다. 그러나 도서, 문헌을 기록의 형식의 하나로 보고, 기록을 있는 그대로의 사실을 정확하게 담고 있는 객체이기 보다는 고쳐 쓰고, 편집한 것이라고 전제하는, 즉 기록을 의도된 결과로서만 바라보는 듯한 저자의 시각은 많은 논의점을 담고 있다. 좀 더 이 책의 아쉬운 점을 얘기하자면 먼저 몇몇 중요 용어에 대한 개념정의 부분이다. 이 정의는 모든 학문분야를 아우를 만큼 너무나 일반적이어서 특정 분야내에서 재정의되어 특징적, 제한적으로 사용됨으로써 해당 분야 연구자들의 이해와 합의를 얻기 쉬울 수 있다는 점을 간과하고 있으며, 사례는 지나치게 도서, 도서관 중심이어서 기록학분야에서 실질적인 도움을 얻기에는 다소 부족하다.

또한 몇몇 용어간의 개념차이도 눈에 들어온다. 앞서 언급했던 기록도 그렇고, 이 책에서 온톨로지 알고리즘을 포함한 기록 정보화와 그 관련 기술을 의미하는 매우 전자적인 개념으로 사용되고 있는 'Documentation'도 그러하다. 도큐멘테이션을 기록화로 해석하거나 전략이라는 단어와 연결해서 기록의 생산 및 수집전략의 개념으로 이해하는데 익숙해있던 연구자들에게 이는 생소하기만 하다. 아마도 Documentation이 컴퓨터 분야나 인쇄·출판 분야에서 빈번하게 사용되는 용어이고, 일본에서는 설계의 문서화나 매뉴얼 작성 기술을 의미하는 용어로 정착되었다는 저자의 설명에서 비롯된 듯하다<sup>1)</sup>.

---

1) 저자는 Documentation이 주로 19세기 말 유럽에서 탄생한 것으로 조사하였고,

마지막 장에서 언급되어 있는 것처럼, 저자는 기록의 온톨로지 알고리즘과 정보의 알고리즘 각각을 구분하는 것이 아니라 ‘기록, 정보, 지식’의 세계를 가로지르는 온톨로지 알고리즘을 얘기하고 싶었을 것이다. 이 시도야 말로 이 책의 중요한 가치이지만 그 어려운 과정에서 앞서 말한 몇 가지 아쉬운 점을 가질 수밖에 없었을 거라 이해되기도 한다. 다양한 학문분야를 넘나들며 자신의 전공분야에 대한 이론적 기술적 전개를 풍부히 해나가는 저자의 역량은 존경스럽다. 기록정보학, 전자도서관, 사회정보학이 전공분야라는 저자는 경험을 바탕으로 ‘기록 정보 지식의 세계’를 폭넓게 다루고자 노력했다. 기록학분야의 책이 너무나 드물고 귀하게 출판되는 현실에서 ‘기록’이라는 단어가 들어간 것만으로도 이 책은 너무나 반가웠다. 또한 역자가 메타데이터 설계, 시소러스 개발, 색인, 검색 및 인터페이스 설계 등 다양한 정보학분야의 연구를 해 왔음을 알기에 더 이상 적절한 번역자는 없을 거라 믿어 의심치 않았다. 문헌정보학에서 통용되는 일본식 용어가 사용되었지만 document, record 등의 단어 하나도 조심스럽게 번역한 느낌이 들었다. 하지만 내용으로 봤을 때 결론적으로 기록학관점에서 관련자들이 가진 어떤 갈등을 시원하게 풀어주진 못하고 있다는 점은 이 책의 아쉬운 점으로 남을 수밖에 없다.

아마 이 책의 효용성도 기록관리분야보다는 도서관분야에서 빛을 발할 것이다. 이는 이 책이 ‘주제’에 따른 분류와 알고리즘 설계를 기반으로 하기 때문이다. 공공 기관의 기록관이나 모 기관으로부터 기록물을 이관받는 기록관은 기록과 관련된 행위

---

문헌이라는 단위를 세밀하고 자세하게 다루는 새로운 서지학이 documentation이며, 이를 위한 분류법으로 고안된 것이 UDC(Universal Decimal Classification)라 한다. 그리고 주제분석은 Documentation의 중심 주제이다.

와 기록을 생산한 기관의 조직적 구조를 먼저 고려하고 업무분석이나 조직분류체계를 바탕으로 기록정보를 분류한다. 따라서 이런 기관들보다는 도서관 및 주제기록관이나 수집형 기록관(메뉴스크립트 보존소)에서 관리방법론에 대한 아이디어를 더 많이 얻을 수 있을 것으로 보인다.

우리나라에서도 이론적, 기술적 개념을 하나로 꿰뚫는 기록학 분야의 도서가 빠른 시일내 나올 수 있기를 기대해 본다.