

쟁점과 토론

정보양식의 변모에 따른 전자텍스트의 새로운 구조 연구

장 노 현*

I. 들어가는 말	IV. 전자텍스트의 초구조
II. 지식정보와 전자텍스트의 개념 변화	V. 끝맺기
III. 전자텍스트의 거시구조: 단위텍스트	

I. 들어가는 말

디지털 정보통신 기술의 발전으로 우리 삶의 여러 부문이 큰 변화에 직면하고 있다. 지식의 생산과 축적 그리고 유통과 활용의 방식도 이러한 변화의 물결에서 예외가 아니다. 푸코는 에피스테메(épistémè)¹⁾, 즉 각 시대는 인식론적 구조물들, 언설적 실천들을 묶어줄 수 있는 독특한 관계들의 집합이 존재한다고 생각했다. 이는, 각 시대마다 사유를 지배하는 독특한 사유문법(思惟文法)이 있으며, 이에 따라 지식을 획득, 정리, 유포되는 방식도 서로 달라질 수 있음을 의미한다. 디지털 문화도, 인쇄문화가 보편화되어 있던 근대 사회에서 지식의 정리 및 유포방식과는 다른 새로운 에피스테메를 가지게 된다. 15세기 구텐베르크의 인쇄 혁명이 그랬던 것처럼²⁾, 디지털 매체와 정보통신 기술의 발전은 지식을 새로운 방식으로 재구성

* 한국정신문화연구원 전문위원 서사학·인문정보학 전공

1) 미셸 푸코(이정우 역), 『지식의 고고학』(민음사 1992), 266~267 쪽 참고

2) 피에르 레비는 “16세기의 위대한 인쇄업자들은 학식 있는 사람들이었으며 휴머니스트였고 동시에

하고 유형화시키도록 요구하고 있다. 매체의 형식이 매체의 내용을 유도하고 있는 셈이다.

지식정보를 둘러싼 사회문화적 상황이 이렇게 변화를 거듭하고 있지만, 현재 지식의 획득, 정리, 유포 방식은 아직도 기존의 방식을 답습하고 있는 경향이 강하다. 디지털문화만이 가질 수 있는 독특한 지식의 구성방식은 물론이거니와 유통방식에서도 근본적인 변화를 찾아보기 쉽지 않다. 인쇄매체에 맞게 구성된 지식체계가 단순히 디지털의 옷을 입고 나타나는 정도에 그치고 있는 형편이다. 우리가 광범위하게 사용하고 있는 워드프로세서들, 예컨대 한글이나 MS 워드 등은 인쇄문화 시기의 사유체계를 단지 디지털 파일 형태로 변환시키기에 적합한 형식만을 제공하고 있을 따름이다. 정부도 디지털 형식의 지식정보자원의 효율적인 관리를 통해 지식정보의 활용성 제고에 노력하고 있다. 하지만 현재 구축 완료되었거나 구축 중에 있는 대부분의 디지털 지식정보자원은 재활용성이나 확장가능성이라는 측면에서 높은 점수를 받기는 어려운 형편이다. 지식의 확장과 재창출 등의 학문적 활용도가 낮고, 디지털 지식정보자원은 양적인 측면에서 절대적으로 부족하다. 이는 디지털 기반에서 지식을 생산하고, 처리하고, 나아가 잘 활용하기 위한 체계가 갖추어져 있지 않은 데서 기인한다고 볼 수 있다.

그러나 이러한 흐름 속에서도, 큰 틀에서 변화의 조짐이 나타나고 있다. 워드프로세서들이 다양한 애플리케이션 프로그램들과 혼합되면서 진정한 의미의 디지털 문화를 가시화시키는 쪽으로 변화하고 있다. 이러한 변화들이 지속적으로 일어난다면, 앞으로 생성될 모든 지식은 사유의 단계에서부터 생성, 축적, 유통, 활용에 이르기까지 제반 과정에서 디지털문화만의 독특한 구조와 체계를 갖게 될 것이다. 그러면 ‘디지털화 될 수 없는 지식은 더 이상 지식이 아니다’라는 말도 가능해지리라 본다.

이제, 디지털 지식정보의 생산, 처리, 유통에 관련된 수많은 요소와 학문분야의 학제적인 접근을 통해 돌파구를 찾아야 할 시점에 이르렀다. 현재 언어·문화와 관련된 지적 자원에 대한 대규모 데이터베이스 구축에만 열을 올리고 있는 상황을

기술자들이었으며, 지식의 조직과 지적 교유의 새로운 양식을 탐색하는 사람들이었다. 새로운 지적 테크놀로지에 관해 우리는 르네상스 시기와 비견될 수 있는 시기를 맞이하고 있다고 상상할 필요가 있다” 고 했다. 『지능의 테크놀로지』(철학과현실사 2000), 164 쪽

돌아보고, 보다 근본적인 차원에서 정보화시대에 걸 맞는 지식의 체계화와 구조화 방법을 찾아내야 한다. 보다 구체적으로는 전자문서 즉 전자텍스트의 구조와 이의 가공처리에 대한 깊이 있는 통찰을 통해, 생산과 유통 등의 제반과정에 보다 효율적으로 작동할 수 있는 전자텍스트 구조를 탐색하는 일이 시급하게 되었다.

전자텍스트는 효율적인 생성 및 처리 유통을 위해서 세 개의 층위로 구성될 수 있다. 이는 전자텍스트 처리에 대한 필자의 오랜 경험을 바탕으로 한다. 세 개의 층위는 각각 미시구조, 거시구조, 초구조의 이름으로 부를 수 있다. 이는 텍스트학의 선구자 중 한 사람인 반 데이크의 용어를 차용한 것이다. 본 논문은 전자텍스트를 구성하는 이들 세 층위에 대한 본격적인 논의의 출발점이 될 것이다. 따라서 논의의 여러 부분이 아직 미 확정적이고 유동적일 수 있다. 그런 만큼 이번 연구에서는 다루는 대상을 제한하여, 그에 대한 논의를 좀 더 구체화시킬 예정이다.

따라서 본고에서는 전자텍스트의 세 층위 중에서 우선 거시구조와 초구조만을 다루고, 미시구조는 다음의 논문에서 다루겠다. 내용구조의 층위인 거시구조에 대한 논의에서는 단위텍스트를 중심으로, 전자텍스트의 생산과 이를 통한 재생산의 측면을 다룰 것이다. 형식구조의 층위인 초구조 논의에서는 개념과 필요성 및 초구조에서 정의되어야 할 다양한 기술(description) 요소들을 살펴보겠다. 이는 전자텍스트의 관리와 활용의 측면에 관련되는 부분이다.

II. 지식정보와 전자텍스트의 개념 변화

1. 지식정보의 개념과 특성의 변화

데이터 혹은 자료는 외부 대상과 관련된 사실들을 나타내기 위해 사용하는 구체적인 숫자나 상징물을 말한다. 이에 비해 정보는 이들이 유용하게 조직되거나 선택적인 연관성 하에 재조직된 상태를 가리키는 것으로, 작업을 하거나 찾아볼 때 의미 있고 유용하도록 처리된 데이터이다. 데이터는 수동적 혹은 자동적인 처리에 의해 바뀌기 전까지는 같은 상태를 유지한다는 측면에서 정적이다. 그러나 데이터와 비교할 때 정보는 끊임없이 변한다는 점에서, 그리고 무한대의 방법으로

처리되고 표현될 수 있다는 점에서 동적이다. 결국 데이터는 저장되는 것이고, 정보는 추출되는 것이다.³⁾

그리고 정보를 평가하거나 이해하는 데 필요한 체계화의 원리 혹은 구조화의 원리를 갖추게 되면, 정보는 지식의 단계로 확장된다. 이처럼 체계성 혹은 구조화의 정도에 따라 데이터, 정보, 지식으로 나누어 볼 수 있지만, 사실 정보와 지식을 뚜렷하게 구분해 내기란 어려운 일이다. 따라서 이 글에서는 정보와 지식의 개념을 두루 포함하는 지식정보라는 용어를 사용하기로 하겠다. 즉 지식정보는 학습, 지각, 판단, 추리 등 인간의 지적·정신적 활동의 재료이자 결과물이라고 할 수 있다.

정보, 특히 지식정보는 시대에 따라, 매체의 변화에 따라 체계화의 양식이나 구조적 유형이 변하게 된다. 따라서 매체를 달리하는 각 시대는 지식정보의 양식과 유형에 대해 느끼는 친숙성은 서로 다르다. 반면 동일한 매체를 사용하는 사람들은 동일한 양식과 유형을 공유하게 된다. 그런데 어떤 시대가 변화의 상태에 놓이게 되면 친숙함은 붕괴되고 낯설음이 그 자리를 대신해 간다. 이러한 변화의 과정이 시작되었을 때, 사람들은 대부분 친숙한 양식에 대해서는 우호적인 반면 낯선 양식에 대해서는 강한 거부감을 갖는다. 낯선 양식을 강하게 비판하기도 하고, 심지어는 무시해 버리려는 강한 충동과 유혹을 느끼기도 한다. 혹은 새로운 유형의 지식정보가 전통적인 양식이나 기준, 정의 등에 부합하지 못하거나 상응하지 않기 때문에 그것을 지각하지 못할 수도 있다.

최근 들어, 이런 현상은 디지털 매체가 근대적 인쇄 매체를 대체해 나가는 변화의 과정에서도 목격되고 있다. 디지털 매체의 급속한 확산에 따라, 많은 사람들이 친숙함과 낯설음 사이에서 방황하고 있다. 인쇄 매체에 동기화되어 있던 사람들은 새로운 디지털 매체 즉 전자텍스트가 갖추어야 할 바람직한 지식정보의 모델을 찾으려는 노력을 백안시 하려는 경향이 있다. 생성과 공유, 그리고 활용을 위해 바람직한 디지털 지식정보의 구조적 모델에 대한 모색작업이 거의 진전이 없는 상황이다. 인쇄문화가 추구했던 지식 구조화의 모델을 답습한 즉 인쇄문화의 사유와 글 쓰기 체계를 바탕으로 새로운 디지털 지식정보를 만들어 간다는 것은 지극히 위험

3) Michael J. Hernandez, *Database Design for Mere Mortals* (SciTech, 2003), pp. 36-38.

한 발상이다. 그것은 생성 공유 활용이라는 측면에서 효율성이 떨어지기 때문이다. 따라서 디지털 문화에 맞는 지식정보의 구조에 대한 탐색이 시급한 과제로 등장하게 된다. 그러면 인쇄문화에 동기화되었던 지식정보와는 다른, 새로운 시대의 지식정보는 어떤 특성을 갖는가? 그것이 지향해 나갈 바람직한 방향은 무엇인가?

인쇄 매체에 적합하게 조직되었던 전통적인 지식정보는 정당성이 확보된 참인 믿음으로 간주되었다. 자연히 정당성의 확보 과정에서 비판적이고 객관적인 자질이 중시되었다. 피에르 레비는 문자로 인해 사람들의 지식과 자기정체성이 부분적으로 분리되면서 비판적 진리에 대한 요구가 생겨났고, 지식의 활자화된 전달에 적합한 탈문맥적 의사소통 상황으로 인해 객관적 진리에 대한 요구가 강해졌다고 말한다.⁴⁾ 특히 활자매체는 지식의 생산과 수용에 관여하는 문맥적 상황과는 가능한 한 무관한 텍스트를 만들도록 자극하였다.

또한 기존의 지식정보는 지속적으로 동조할 수 있는 유일성에 기반하고 있었다. 어떤 문제가 제기되었을 때, 그것에 대한 답은 한 가지라야 했다. 문제 해결을 위해 다양한 방법이 동원될 수는 있지만 어떤 경우든 도달하게 되는 지점은 동일하며, 그것은 가장 정확한 한 가지 답이 되었다. 근대적 지식 환경에서는 진실이 여럿일 수 없다는 가정은 확고하였다. 나아가 한번 확정된 지식은 새로운 것에 의해 완전히 대체될 수 있음이 증명되기까지 도그마의 지위가 보장되었다. 유동성은 허락되지 않고 고정불변성과 정확성이라는 환상이 일반화 되었다. 그리고 그것은 인쇄매체의 특성에 의해 강화되고 지속되었다.

그런데 전자텍스트가 등장하면서 지식정보에 대한 전통적인 개념은 크게 흔들리고 있다. 비판적이고 객관적인 진리를 지식의 규범으로 만들어주는 조건들이 급속하게 변하고 있는 것이다. 우리는 웹 검색을 통해 얻게 되는 웹문서의 정확성을 크게 신뢰하지 않는다. 웹문서는 불확실성과 변화가능성을 속성으로 한다는 판단이 개입되어 있기 때문이다. 다시 말해서 웹을 통해 접근 가능한 전자텍스트는 고정성과 정확성이라는 지식의 도그마적 기준에서 벗어난다.

웹 속의 전자텍스트들은 고정된 정보가 아닌 변화가능성을 언제나 내포하고 있다. 그것은 상황과 문맥에 따라 다른 의미로 읽히도록 유도될 수 있다. 또한 정확

4) 피에르 레비(강형식, 임기대 옮김), 『지능의 테크놀로지』 (철학과현실사, 2000), 181 쪽

한 한 가지 대답보다는 취사선택이나 시뮬레이션이 가능한 지식정보에 대한 선호도가 높아졌다. 수많은 정보 중에서 내가 원하는 답을 취사선택하는 이데올로기적 자유로움을 지향한다. 다시 말해서 지식이 유동적이라는 것은 어느 시대에나 사실이었으나 이제는 과거의 어느 시기보다도 유동적 속성을 더욱 적극적으로 택하게 되고, 대신 정확성을 어느 정도는 포기한다⁵⁾ 우리는 다소 적절하다고 판단되는 것들의 시뮬레이션을 통해 정확성을 대체해 나가고 있다. 전자텍스트 환경에서의 지식정보의 개념과 특성의 변화를 피에르 레비는 다음과 같이 언급한 바 있다.

정보화된 지식은, 최초 구전의 경우처럼 변함없이 살거나 혹은 스스로 변함없기를 바라는 사회를 동일하게 보존하는 것을 지향하지 않는다. 또한 이론 혹은 해석학인 문자에서 탄생된 규범적 양식 대신에 진리를 추구하지도 않는다. 정보화된 지식은 실행의 속도와 적절성을 추구하고, 더욱이 신속성과 작업적 변화의 적절성을 추구한다.⁶⁾

결국 이러한 지식정보는 유일한 진리를 추구하는 데는 관심이 별로 없다. 고전 문헌처럼 단순히 읽히거나 해석되기 위한 것이 아니며, 우리는 더 이상 비판하지 않고 오류수정(디버깅)할 뿐이다.

모든 지식정보의 성격은 그것이 지니는 구조적 특성에 반영되게 된다. 디지털 지식정보 역시 그것만의 구조적 특성을 지닌다. 전자텍스트는 기존의 인쇄텍스트와는 다른 구조적 특성을 지니고 있다. 우선 전자텍스트는 파편화된 정보, 분절성이 강한 정보이다. 인쇄된 정보가 논리적 연속성과 균질화의 원리에 큰 비중을 두는 텍스트라면, 전자텍스트는 기본적으로 분자화를 지향한다. 전자텍스트를 생산하는 과정에서 우리가 맨 처음 부딪치는 문제는, 어떻게 긴 문서를 쪼개서 독립된 부분들로 만드느냐 하는 점이다. 모든 문서는 더 이상 나눌 수 없는 작은 부분들로 나누어질 때, 즉 분자화 되었을 때 가장 적절한 전자텍스트로 생성될 수 있다.⁷⁾

5) 김보영, 「지식정보의 의미에 관한 단상 인터넷 문화사를 쓰기 위하여」 정보트러스트 운동 홈페이지 (www.infotrust.or.kr).

6) 피에르 레비(강형식, 임기대 옮김), 앞의 책, 180쪽.

7) 이에 대해서는 장노현, 「하이퍼텍스트 서사에 관한 연구」(한국정신문화연구원 박사학위논문, 2002)

또한 전자텍스트는 원래의 맥락을 자주 벗어나는 운명을 타고 난 텍스트이다. 맥락을 벗어나는 속성은 분절성에서 연유한다고 볼 수 있다. 분절된 전자텍스트는 하이퍼링크나 검색 기능으로 인해 어떤 연계적 맥락을 벗어난 상태에서도 호출될 수 있다. 따라서 텍스트가 생성될 때의 맥락을 벗어나더라도, 정보 자체가 변형되거나 불충분한 상태로 읽히게 되면 안 된다. 만약 어떤 정보가 원래의 맥락을 벗어나면 어떤 일이 생기는가? 가령 누군가가 “강여는 기분이 좋아졌다” 라고 말했다고 가정해 보자. 이 발화는 그것이 발화되는 순간의 특정한 맥락 속에서는 참이 된다. 그러나 이 발화는 하루가 지난 다음날에도 계속 참일 수는 없다. 발화의 맥락이 달라졌기 때문에 정보 자체의 진위에 문제가 발생할 수 있는 것이다. 즉 시간적 맥락을 벗어나는 순간, 이 문장의 의미는 다른 해석의 가능성을 갖게 된다. 전자텍스트는 구어에서처럼 상황과 맥락을 자주 벗어난다. 원래의 시간적, 혹은 구조적 맥락에서 벗어난 상태로 이용자에게 전달될 가능성이 높은 텍스트이다.

따라서 전자텍스트는 자립성이 강한 텍스트로 구성되어야 한다. 여기서 말하는 자립성이란 컨텍스트 즉 전후 맥락에 의지하지 않음을 뜻한다. 전자텍스트는 인쇄된 책과는 달리, 앞 맥락과 뒤 맥락이 항상 고정불변하지 않는다 따라서 전자텍스트는 개별단위마다 최소한의 자립성을 유지하고 있어야 한다.

또한, 전자텍스트는 지배가치 중심의 획일화에서 벗어나는 구조를 갖는다. 전통적인 지식체계에서는 지배가치를 중심으로 모든 지식은 서열화되고 획일화된다. 그러나 인터넷의 등장으로 이런 근대적 지식체계가 도전을 받고 있다. 이제 지식체계는 다양화되고, 같은 대상을 두고도 서로 다른 담론들이 얼마든지 존중받을 수 있는 체계로 변화되었다. 이러한 상황을 들뢰즈는 ‘고원’의 개념으로 표현하였다. 새롭게 구상되는 전자텍스트의 구조는 이런 패러다임의 변화를 어떻게 수용할 것인가에 대한 고민에서부터 시작되어야 한다.

2. 새로운 전자텍스트에 대한 요구

기존의 웹문서는 지식정보 자원의 효율적인 생산과 처리, 유통에는 불합리한 측

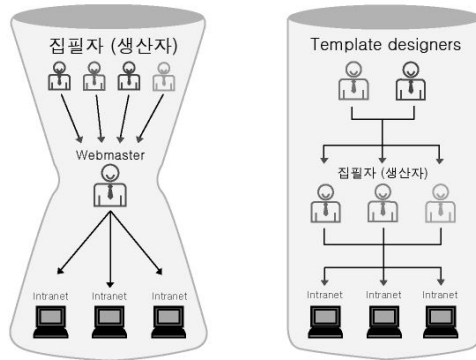
에서 자세히 논의했음.

면이 많다. 이들은 작고 정적인 웹사이트 내에 갇힌 폐쇄적인 형태로 생산되고 유통되었다. 이에 따라 다른 사이트의 텍스트와는 어떤 관련이나 상호성을 띠지 못하였다. 전자텍스트의 디자인도 개별 사이트를 설계한 디자이너의 취향이 그대로 반영될 뿐이며 해당문서 자체의 특성이 제대로 고려되는 경우가 드물었다. 디자이너들은 해당문서의 내용이나 구조를 제대로 파악하지 못하고, 내용구조와는 동떨어진 채로 단순히 시각적인 효과만을 고려하여 디자인에 임하는 경우가 많았다. 그리고 한번 디자인이 완성되면 문제점이 발견되어도 수정하기 어렵고, 처음처럼 다시 디자인을 해야 하는 번거로움을 피할 수 없었다.

하지만 이런 디자인 관련 문제는 오히려 주변적이고 사소한 측면에 해당한다. 기존의 전자텍스트는 필자가 원고를 완성하여 웹사이트 개발자들에게 넘기면 그것으로 필자의 역할은 종료된다. 스스로 HTML에 대한 이해가 있는 필자가 아니라면 그는 더 이상 자신의 원고가 어떤 가공 처리 과정을 거치는지 알지 못한다. 또 이 과정에서 어떤 의견을 개진할 수도 없다. 원고의 구조화는 웹콘텐츠 기획자의 손에 맡겨지고, 여기에 디자이너가 시각적인 디자인 요소를 입힌다. 그 후 다시 웹마스터에게 넘어가 서버에 등록되면 웹을 통해 이용 가능한 상태가 된다. 이러한 일련의 과정 어디에서도 필자는 자신의 원고에 대한 가공 처리 및 이용에 대해서 어떤 의견도 제출할 수 없다. 이처럼 필자의 역할이라는 측면에서 볼 때 인쇄문화 시대의 필자와 하등 달라진 것이 없다. 혹시 필자가 원고의 잘못된 부분이나 보충하고자 하는 부분을 있을 때에는, 인쇄문화 시대의 필자가 그랬던 것처럼 번거로운 과정을 거쳐야 한다. 자신의 의견을 반영하여 수정이 이루어지기 위해서는 많은 시간과 노력을 필요하였다. 또한 필자라고 해도 자신이 집필한 지식정보를 업데이트하는 데는 큰 어려움이 따랐던 것이다.

전자텍스트가 웹사이트라는 형식으로 완성되면, 그 텍스트는 유통의 과정에 들어간다. 유통(관리)의 과정에서도 여러 가지 어려움이 따른다. 이 과정에서의 모든 권한은 서버관리자나 웹마스터에게 있으며, 이 과정에서도 필자는 배제될 수 밖에 없었다. 특히 서버관리자나 웹마스터가 전자텍스트의 내용과는 무관하게 기술적인 측면만을 관리하는 경우가 많은 상황 속에서 지식정보에 대한 효율적인 이용이나 가치의 확대재생산 등 활용성을 최대화 할 수 있는 가능성은 애초부터 막혀있다고 해도 과언이 아니다. 그림에서 보는 것처럼, 지식정보는 웹마스터에게서 병목현상

을 일으킨다. 정보의 생성·유통 과정에서 발생하는 이런 병목현상은 진정한 정보의 공유와 활용에 많은 제약으로 작용한다. 그림은 전자텍스트의 생성과 유통 과정의 변화 양상을 도식적으로 보여주고 있다.8)



<전자텍스트의 생성·유통 과정의 변화>

그런데 최근 들어 웹문서에 대한 요구가 변화하고 있다. 우선 텍스트 생성과 유통에 관여하는 주체의 변화를 꼽을 수 있다. 전자텍스트의 생산과 가공처리 과정이, 프로그래머나 웹마스터 중심의 체계(모래시계형 모델)로부터 다양한 필자를 중심으로 하는 체계(원통형 모델)로 변화되고 있다 또한 유통 즉 검색이나 웹서핑을 이용한 전자텍스트의 이용 측면에서도 생산자 중심으로부터 이용자 중심으로 변해가면서 이용자의 역할이 점차 확대되고 있다. 이에 따라 지식정보는 확대재생산이 가능한 구조로 설계되고 제작될 것이 요구되고 있다.

좀 더 자세히 살펴보자. 하나의 전자텍스트가 접근 경로가 다른 이용자들에게도 유용한 정보가 되기 위해서는 폐쇄적인 형태나 구조를 가져서는 안 된다. 즉 생산 처리, 유통이 보다 원활하게 이루어질 수 있도록 개별 웹사이트의 범주를 벗어나는 범용의 전자텍스트 구조를 필요로 하게 되었다. 범용성을 가진 전자텍스트 구

8) 이 그림은 Microsoft Content Management Server 2002를 위한 CMS Seminar 발표문 발표자 이 미경, 한국글로벌닐리지 강사)에서 뽑아 필자가 약간 변형시켰다. 왼쪽의 모래시계형 모델은 기존의 지식정보의 생산과 유통체계를, 오른쪽의 원통형 모델은 향후 새롭게 전개될 지식정보의 생산과 유통체계를 표현한 것이다.

조에 대한 기술(description) 원칙과 스키마가 개발되고, 이를 손쉽게 처리할 수 있는 애플리케이션 시스템이 개발된다면, 전자텍스트의 생산과 처리 유통 과정은 보다 편해질 것이다. 이것은 전자텍스트의 생산을 비롯한 전 과정에서 필자의 역할이 증대함을 의미한다. 필자가 직접 자신의 원고를 처리할 수 있는 시스템에서는 전자텍스트의 개발과 배포가 빨라지고, 이에 따라 업데이트도 자주 이루어질 수 있게 된다. 전자텍스트의 생산 유통에 대한 필자의 직접적인 개입은, 중간 과정에 관여하던 사이트 개발자들의 역할을 최소화시켜 전자텍스트를 개발하고 관리하는 비용을 크게 줄여줄 것이다. 체계의 재편은 디지털정보를 단순정보에서 본격적인 지식정보로 한 차원 높이는 계기가 될 것이다.

한편 이용자의 측면에서도 고정된 실체를 가진 파일 단위의 전자텍스트보다는 좀더 유연한 형식의 전자텍스트를 원하고 있다. 디지털의 장점은 변형과 재활용이 자유롭다는 점이다. 전자텍스트는 이런 장점을 최대한 살린 구조여야 한다. 이용자들은 자신들의 경험에 적합하고 자신들이 원하는 구조로 재편집과 재활용이 가능한 전자텍스트를 원한다. 그런데 기존의 전자텍스트들은 모두가 웹콘텐츠 기획자나 개발자들이 의도하고 디자인한 형태로 볼 수 있었을 뿐이다. 이는 인쇄문화 시대의 방식을 그대로 답습하고 있는 것에 지나지 않는다. 진정한 디지털 지식정보는 이용자들이 정보를 자신만의 독특한 요구대로 해체하고 통합할 수 있는 형태로 제공되어야 할 것이고, 또한 최근 들어 이러한 가시적 변화의 조짐이 보이고 있다.⁹⁾

III. 전자텍스트의 거시구조 : 단위텍스트

최근의 변화 요구에 부응하는, 새로운 개념과 구조를 지니는 전자텍스트를 모색하는 것이 이번 절의 과제이다. 전자텍스트가 어떤 구조를 갖게 될 때 가장 전자텍스트다운 전자텍스트, 웹문서다운 웹문서가 만들어지게 될까?

9) 이러한 이상적인 시스템을 필자는 ‘순환형 지식정보시스템’으로 이름 붙였다. 한국정신문화연구원에서 추진하고 있는 <한국향토문화전자대전> 사업에서는 이 시스템을 구현하는 것을 최종적인 목표로 하고 있다.

앞에서 이야기했듯이, 여기서는 거시구조와 초구조만을 논의하게 된다. 거시구조는 내용구조의 일종이다. 거시구조는 법조문, 계약서, 학위논문 등에서 쉽게 확인할 수 있다. 즉 이런 텍스트들은 제도적 혹은 관습적으로 내용구조가 확정되어 있는 텍스트들이다. 현대의 대학원 교육을 통해 양산되는 학위논문을 예로 들어보자. 학위논문은 크게 제출지, 인준지 등의 형식적 표지부분¹⁰⁾과 서론, 본론, 결론 등의 내용부분, 그리고 참고문헌, 초록 등의 부가부분으로 이루어져 있다. 내용 부분을 좀 더 자세히 들여다보면, 서론에서는 연구목적이나 문제제기, 연구사 정리, 적용할 방법론 등이 제시된다. 본론은 일반적으로 3장5장으로 더욱 세분되어 중심논의가 펼쳐지고, 결론에선 연구결과를 정리하고 남은 과제 및 전망을 제시한다. 이와 같이 학위논문은 형식적 관습에 따라 정해진 구조대로 쓰여지기 마련이다.

그러나 모든 텍스트가 그와 같이 합의된 구조를 갖고 있는 것은 아니다. 오히려 거시구조가 크게 유동적인 텍스트들이 더욱 많다고 할 수 있다. 따라서 앞으로 다양한 텍스트들의 구조에 대한 기초연구가 학문적 차원에서 방대하게 이루어져야 할 것이다. 여기서는 거시구조를 ‘단위텍스트’라는 개념에서 논의할 것이다. 단위텍스트는 거시구조의 가장 작은 단위에 해당하는 개념이다.

1. 단위텍스트의 개념과 필요성

필자는 이미 다른 글¹¹⁾에서 ‘단위텍스트’라는 용어를 만들어 사용했다. 그 글에서 단위텍스트는 하이퍼텍스트 시스템에서의 문자화된 텍스트라고 규정했다. 하이퍼텍스트는 통상 하이퍼링크(hyperlink)라고 불리는 기능에 의해 작동하기 시작하는데, 이러한 하이퍼링크는 단위텍스트(바르트는 이를 ‘렉시야’라고 불렀고, 랜도우는 ‘텍스트 블록’이라 부른다)들 사이를 자유롭게 이동할 수 있게 하는, 문과 같은 기능을 한다고 했다. 필자는 그 글을 쓸 때만 해도, 단위텍스트라는 것은 하이퍼링크를 통해 순간적으로 이동할 때마다 만나게 되는 ‘텍스트 덩어리’라는 두루

10) 학위논문의 형식적 표지들은 4장에서 다룰 초구조에 속하는 요소들이다

11) 장노현, 앞의 글. 필자는 이 글에서 하이퍼텍스트의 두 가지 구성요건으로 단위텍스트와 하이퍼링크를 들었고, 주로 링크의 기능성에 주목하였다. 이제 본 논문에서는 단위텍스트 자체에 초점을 맞춰 논의함으로써 하이퍼텍스트의 두 가지 구성요건을 모두 다루게 되는 것이다.

몽술한 생각을 하고 있었다. 그러나 빠르게 변화하는 웹 환경 하에서 텍스트의 생산·처리·활용(검색) 및 확대재생산의 효율성을 높이기 위해서는 전자텍스트의 새로운 구조 모색이 절실하게 필요해졌다. 이러한 구조 모색 과정에서 단위텍스트에 대한 보다 정밀한 논의가 요구된다.

현재 웹상에서는 대개의 정보가 파일 단위로 생성되고 저장되며, 또한 유통된다. 하나의 파일로 저장된 정보는, 그것이 어떤 확장자를 가지더라도, 이용자의 화면에서는 처음 저장되던 형태 그대로 나타난다. 저장의 순간, 파일의 생성 순간에 갖추어진 정보의 구조와 형태와 내용이 동일하게 유지되는 것이다.¹²⁾ 우리는 통상 디지털 정보가 무형성이나 유동성 등의 속성을 지니고 있다고 말한다. 그러나 앞의 언급은, 즉 정보가 저장 당시의 구조와 형태와 내용을 지속적으로 동일하게 유지하고 있다는 말은, 그것이 사실은 어떤 고정된 실체를 지니고 있음을 의미한다. 고정된 실체를 가진 디지털 정보, 이것은 엄밀한 의미에서 볼 때 디지털 매체의 속성을 온전하게 갖추지 못한 정보에 해당한다. 다시 말해서 웹문서로 통칭할 수 있는 기존의 전자텍스트들은 변화무쌍하며 유동적 특성을 지닌 온전한 디지털 텍스트라고 볼 수 없다.

단위텍스트의 개념은, 그동안 파일단위로 관습화되어 있던 웹문서에 대한 새로운 인식에서 출발한다. 단위텍스트는 기본적으로 기존의 파일이 지칭하는 정보단위보다는 훨씬 작게 세분화된 정보단위이다. 그런데, 지식정보를 구조화하는 데 있어서, 아주 작게 세분된, 새로운 정보단위가 필요한 이유는 무엇일까? 우선 이용자의 측면에서 대답할 수 있다. 현재의 상황에서 이용자가 원하는 정보를 검색하게 되면 불필요한 잉여정보가 함께 나타난다. 한마디로 검색의 정확도에 문제가 있는 것이다. 이는 검색엔진 자체의 문제이기도 하지만 텍스트 구조의 문제이기도 하다. 텍스트가 훨씬 작게 세분화되어 단위텍스트로 잘 구조화 되어 있다면 검색의 정확도가 훨씬 높아질 것이다. 단위텍스트로 잘 짜여진 지식정보에서는 이용자가 검색을 통해 원하는 정보에 보다 정확하고 빠르게 접근할 수 있게 된다.

또한, 단위텍스트는 텍스트의 재활용성, 즉 확대재생산의 가능성을 높일 수 있

12) 물론 한 파일의 범위를 벗어나 파일 간 연결이 이루어질 경우에는 현재의 전자텍스트(웹문서)에서도 처음의 주어진 문맥을 벗어날 수 있다. 이 문제에 대해서는 그간 하이퍼텍스트의 다선형성 혹은 비선형성의 개념으로 필자를 비롯한 많은 논자들이 논의를 펼친 바 있다.

다. 반 데이크는 텍스트의 재생산, 재구성과 관련해서 “재생산규칙은 바로 현재 있는 정보를 세분화하고 확장하고 납득할 만하게 세목화해야 한다”¹³⁾고 했다. 즉 단위텍스트로 세분화되었을 때 텍스트의 재활용성이 높아지게 된다. 단위텍스트는 언제든지 저장된 원래의 파일에서 빠져나와 다른 문맥 속에서 다른 단위텍스트와 결합할 수 있게 된다. 이를 통해 텍스트의 재활용성은 크게 증가할 수 있다.

이러한 이유에서 볼 때, 향후 생산되는 모든 디지털정보는 단위텍스트를 중심으로 세분화되어 구조화되어야 한다. 그러나 논리적 연속성을 금과옥조로 여기는 인쇄문화적 글쓰기에 익숙한 필자들이, 것처럼 세분화된 단위텍스트 형태로 글쓰기를 하는 것은 결코 쉬운 일이 아닐 것이다. 모든 글쓰기는 오랜 시간에 걸쳐 관습화되어 왔다. 그것은 글을 구성하는 방식이나 글의 길이 등에 암묵적인 영향력을 행사한다. 따라서 이런 습관을 한순간에 바꿔 단위텍스트를 기반으로 하는 글쓰기를 강요할 수는 없다. 단위텍스트를 기반으로 하는 새로운 글쓰기에 적합한 습관이 형성될 때까지는, 기존의 글쓰기 습관과 원칙은 존중될 필요가 있다. 현 단계에서 텍스트 생산자들은 처음부터 단위텍스트의 형태로 글을 쓸 필요는 없다. 하나의 텍스트는 현재 한글이나 MS 워드에서 글을 쓰는 일반적인 글쓰기 원리에 따라 글을 쓰면 된다. 다만, 그것이 온라인상의 시스템에 저장되는 과정에서 어떤 처리 절차를 거치면서 단위텍스트로 나누어져 저장되면 된다. 이러한 저장 과정은 시스템에 구현되어 있는 자체 기능의 도움을 받게 될 것이다.

2. 단위텍스트의 자질

인쇄문화적 글쓰기 습관에 따라 쓰여진 전통적인 텍스트가 단위텍스트가 되기 위해서는, 어떤 원리에 따라 혹은 어떤 조건 하에서 나누어지고 세분화되어야 하는가? 즉 하나의 단위텍스트를 형성하는 기준은 무엇인가 이를 해결하기 위해서는 진지한 학문적인 논의의 장이 필요할 것이라고 본다. 예컨대 텍스트학의 한 분야로 텍스트정보학¹⁴⁾이라는 분야를 설정하고, 여기서 다양한 실험과 검증 과정을

13) 반 데이크(정시호 역), 『텍스트학』(아르케 2001), 318 쪽

14) 미국에서는 Humanities Computing(인문전산학)이라는 용어가 널리 보급되어 있다. 하지만 지식 정보의 생산, 가공, 유통, 활용 등과 관련하여 텍스트 자체를 연구하는 학제적 학문분야의 필요성

통해 해결방안을 모색해 가야 한다.

그렇지만 한 가지 분명한 사실은, 단위텍스트는 이용의 측면에서 가장 효율적인 크기와 형태를 지녀야 한다는 점이다. 검색 시에 불필요한 부가정보가 배제된 정확한 검색결과를 가져올 수 있는 형태와 크기를 지녀야 하며, 단위텍스트를 재 활용하여 새로운 정보를 창출하는 데 도움이 되는 구조여야만 할 것이다. 이를 위해서 단위텍스트는 책이나 전통적인 파일 단위의 웹문서와는 다른 자질을 가져야 한다. 그것을 크기의 자질과 구조의 자질로 나누어 설명해 볼 수 있다.

단위텍스트의 크기는 전통적인 웹문서로서의 파일이 어떤 크기를 가져야 할 것인가를 결정하는 것만큼이나 애매하고 해결하기 어려운 면이 있다. 그것은 다분히 자의적인 성격이 강할 수 있다. 그러나 본 연구에서는 단위텍스트를 문단(paragraph)을 기준으로 나누는 시험적인 방안을 제안한다. 통상 문단은 5~6개의 문장으로 이루어진, 하나의 의미를 드러내는 글의 덩어리라고 한다. 그러므로 단위텍스트 역시 문장 5~6개를 가진, 통일된 의미를 표현하고 있는 단위정보라고 일단 정의해 볼 수 있다. 이러한 단위정보는 고정된 실체를 가진다. 현재의 웹문서는 파일단위에서 고정된 실체를 갖지만, 새로운 개념의 전자텍스트에서는 단위텍스트가 고정된 실체성으로 다가오게 되는 것이다.

한편, 동영상이나 소리파일의 경우는 일단 제작될 당시의 파일을 하나의 단위텍스트로 취급할 수 있다. 그 후 부분만을 절단하여 새로운 파일을 만들었을 경우 그것은 또 다른 단위텍스트가 된다. 즉 동영상과 소리파일에서는 파일단위와 단위텍스트의 크기가 같다.

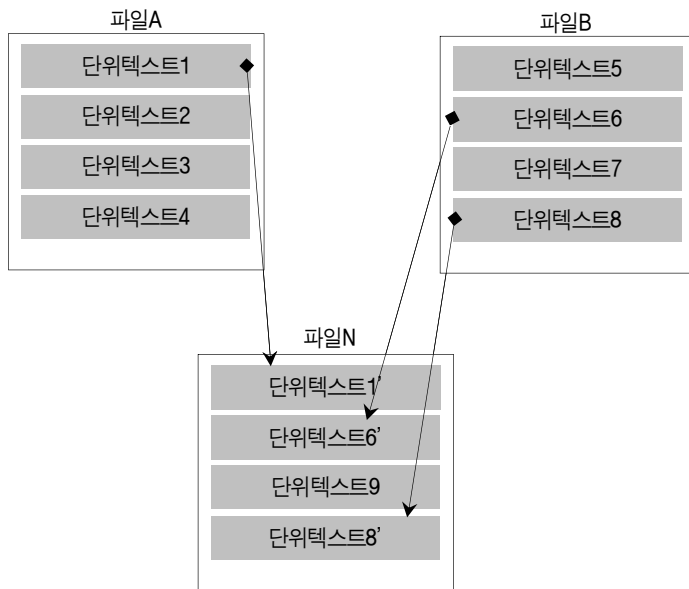
단위텍스트는 뒤에서 논의할 초구조의 정보를 상속할 수 있는 자질을 지녀야 한다. 초구조의 정보는 단위텍스트들 간의 관여가치를 크게 해주는 역할을 한다. 관여가치가 커질수록 그 단위텍스트는 이용자에 의해 재발견될 가능성이 더욱더 크다고 가정할 수 있다. 하나의 단위텍스트는 어떤 문맥에서도 다른 단위텍스트와 잘 융합될 수 있는 구조적 자질을 가졌을 때 가장 이상적이다. 이를 위해서 단위텍스트는 ‘그러나’, ‘그리고’, ‘또한’ 등의 문두 접속부사의 사용을 배제해야 한다. 또한 지시대명사의 사용에도 주의해야 한다. 전통적인 글쓰기에서는 문단이 바뀌

에 대한 인식에서 ‘텍스트정보학’이라는 용어(김병선 교수가 제안함)를 사용하였다. 이는 텍스트학의 하위 분과라는 의미를 명확하게 하기 위해서 ‘전자텍스트학’이라고 불러도 좋을 듯하다

있을 때에도 지시대명사의 사용에 지장을 받을 필요가 없었다. 그러나 단위텍스트 체계 하에서는 다른 단위텍스트로 넘어가면 지시대명사를 사용을 금할 필요가 있다.

3. 단위텍스트의 해체와 조립

현재의 웹문서가 지닌 구조 하에서는 파일 단위 정보에서 의미가 발생한다. 이 파일 단위 정보는, 앞에서 살펴보았듯이, 저자의 사유가 반영된 저자의 텍스트이다. 이용자들은 파일단위로 저장된 텍스트를 열람함으로써 지식정보를 이용하게 된다. 아래 그림을 통해 살펴보자. 이용자들은 파일A, 파일B를 최종적인 형태의 정보로 이용하게 된다. 그러나 새로운 개념의 전자텍스트 이용자는, 원래 저자의 사유체계가 반영된 완성된 텍스트를 해체하고, 이를 재구성할 수 있다. 저자의 사유체계보다는 자신의 사유체계에 맞는 형태로 텍스트를 재구조화함으로써 보다 섬세한 접근과 활용이 가능해진다. 이러한 섬세한 접근과 활용은 단위텍스트로 인해 가능해지는 것이다.



개별 단위텍스트는 단독적으로도 최종 정보가 될 수 있다. 즉 그림에서 볼 수 있는 파일A 속의 네 개의 단위텍스트들은 저마다 최종적인 정보가 될 수 있다. 이 사용자들의 화면에는 파일A 전부가 디스플레이되지 않고 이용자의 요구에 부응하는 각각의 단위텍스트만 디스플레이 된다. 이것은 파일단위의 정보가 단위텍스트를 중심으로 해체됨으로써 이용자의 요구에 가장 부합되는 정보로 재구성되는 과정을 나타낸다.

그러나 단위텍스트들은 해체되는 것으로 끝나지 않는다. 그것은 다시 새로운 조립을 지향한다. 2 개 이상의 단위텍스트들이 모여져서 새로운 정보단위를 형성할 수 있다. 그림에서 파일A에 속해있는 단위텍스트6과 파일B에 속해있는 단위텍스트6, 단위텍스트8이 조합되면서 새로운 정보단위인 파일N을 형성하고 있는 것을 볼 수 있다. 파일의 경계를 넘어 진행되는 단위텍스트의 자유로운 해체/조립의 반복 과정은, 어떤 것(something) 내로의 전이와 어떤 것 밖으로의 전이를 통해 자신의 영토를 재구축하는 과정이라도 표현할 수 있다. 새로운 파일N을 만드는 것은 개별 단위텍스트들의 인식적 거리를 가깝게 조정하기 위한 것이다. 경험적 인식의 거리가 가까워졌을 때, 각각의 단위텍스트들은 하나의 통합된 인식의 체계 속에 구조화될 수 있다. 단위텍스트의 조합은 기존 지식정보를 확대재생산하는 과정이다. 개별 단위텍스트들은 어떤 단위텍스트와 조합되느냐에 따라 정보의 가치와 의미가 달라질 수 있다. 이러한 과정을 통해 지식정보의 양은 확대되게 된다. 뿐만 아니라, 이용자의 요구에 가장 부합되는 새로운 텍스트로 재탄생하게 된다.

단위텍스트의 조합을 통해 생성된 파일N은 한 이용자의 요구와 목적에 충실한 정보일 뿐만 아니라, 추후의 다른 이용자들에게도 유용한 정보가치를 지닐 수도 있다. 이러한 경우 파일N은 시스템 내에 저장되게 된다.¹⁵⁾ 시스템이 요구하는 저장요건을 충족시키기 위해서, 파일N을 생성해 낸 이용자는 문장이나 접속어 등을 손질하여 둘 필요가 있다. 파일N에서 단위텍스트1', 단위텍스트6', 단위텍스트8'로 표기된 것은 최소한의 문법적 자질이 변화되었음을 표시한다. 경우에 따라서는 단

15) 파일N을 시스템에 저장할 것인지 아닌지는 우선 파일N을 생성해 낸 이용자가 결정하게 된다. 그 이용자가 저장 결정을 하면 시스템의 저장 인터페이스를 이용하여 저장 요청을 하게 되고, 시스템은 이를 받아들일 것인지 대기목록에 이를 등록해 둔다. 그 후 시스템과 자료를 관리하는 전문가의 판단에 의해 최종적인 저장 과정을 거칠 수도 있고, 아니면 저장이 거부되어 폐기될 수도 있다.

위텍스트⁹처럼 새로운 내용의 단위텍스트가 집결되어 파일N의 완성도를 높일 수도 있다.¹⁶⁾

그런데, 파일A와 파일B 속의 단위텍스트들이 자유롭게 해체되거나 조립되기 위해서는 요구되는 것은 무엇인가? 무엇보다도 각 단위텍스트들은 전세계 어떤 시스템 속의 단위텍스트와도 구별되는 자신만의 식별자를 가져야 한다. 각 단위텍스트들은 고유한 네트워크 주소를 식별자로 쓸 수 있을 것이다. 이는 단위텍스트들이 자유롭게 해체되거나 통합되기 위한 최소한의 자질이 된다. 각각의 고유한 네트워크 주소를 가진 단위텍스트들은 이 주소를 통해서 검색시스템과 상호작용을 하게 된다. 검색시스템에 포착된 단위텍스트들은 해체와 조립 과정에서 언제나 자신의 고유한 네트워크 주소를 갖고 이동하게 된다. 단위텍스트이 파일N의 단위텍스트 1'로 재활용되더라도 그것은 원래의 네트워크 주소를 보유하게 된다는 말이다. 그렇기 때문에 단위텍스트가 다른 저자에 의해 재활용되더라도 원래의 저자 정보가 그곳에 남아있기 때문에 원저자의 저작권은 침해받지 않을 수 있다.

IV. 전자텍스트의 초구조

1. 초구조의 개념과 필요성

모든 지식정보는 어떠한 형태로 생산되어 어떤 방식으로 관리, 이용되며, 나아가 어떻게 지속적으로 더 많은 유용한 정보를 축적해 나갈 것인가에 대한 전망을 필요로 한다. 그 전망은 지식정보가 보다 효율적으로 관리될 수 있는 방향으로, 보다 쉽게 찾을 수 있는 방향으로, 더 많은 사람이 공유할 수 있는 방향으로, 축적과 재생산이 보다 수월하게 이루어질 수 있는 방향으로 모색되어야 한다. 근대적 인쇄문화 시대에는 국가도서관이나 정부기록보존소 등의 기록관리 기관들에서 주로 그러한 업무를 담당해 왔다. 각 기관들은 일정한 분류체계에 의거해서 각종 문서나 기록을 보관하고 관리해 왔다. 그러나 기관들은 상호간에 큰 연관성을 맺지 않

16) 이런 재활용을 통해 보다 많은 이용자들이 필자로 등장할 수 있다. 포스트모더니즘에서 말하는 독자의 저자화라는 생각이 새로운 전자텍스트에서 실현되고 있는 것이다.

고 독자적인 운영 상태라도 좋았다.

그런데 디지털 매체가 널리 보급되면서 상황이 많이 바뀌었다. 이제 디지털 문서나 기록들은 기록관리 기관(관리자)들의 상호 밀접한 연계 속에서 관리되어야 할 뿐만 아니라, 기록의 생산자(일반적으로 저자라고 일컬어지는 사람들), 그리고 기록의 이용자 등과도 상호 밀접한 연관을 갖지 않으면 안 되게 되었다. 기록생산자, 기록관리자, 기록이용자 모두가 상호작용하면서 전자텍스트에 관여하는 상황 속에 놓이게 되었다. 이제 개별 기관이나 개인을 넘어서서, 인류가 생산한 모든 지식정보가 거대한 네트워크에 연결된 정보시스템 하에서 생산, 관리, 활용될 수 있는 환경으로 옮겨 가야 한다. 이러한 범매체적, 범주제적 거대 네트워크가 효율적으로 작동하기 위해서는 그 안에 담긴 모든 전자텍스트가 일정한 표준에 따라 구조화된 상태로 작성되어야 한다. 한 마디로 전자텍스트 구조에 대한 새로운 합의가 필요한 상황이다. 이를 위하여 필자는 전자텍스트의 ‘초구조’를 제안하고자 한다.

앞에서도 말했던 것처럼, ‘초구조(Superstructure)’는 반 데이크가 자신의 저서인 『텍스트학』에서 사용했던 용어이다. 그가 생각하는 초구조는, 텍스트의 유형을 표시해 주는 총괄적 구조로서, 내용과 무관하게 존재하는 일종의 텍스트 형식을 말한다.¹⁷⁾ 그러나 이 용어는, 반 데이크가 스스로 밝히고 있듯이, 아직은 ‘구속력 있는 일반 개념으로 정립되지 않았다’. 본고에서는 대체적인 의미에서 초구조라는 용어를 차용하되, 구체적인 의미와 쓰임에서는 용어에 대한 새로운 개념규정이 필요할 것으로 생각된다. 우선 본고에서 초구조라는 용어는 전자텍스트의 최상위 층위를 지칭하는 의미가 될 것이다. 이는 정보의 보존과 공유, 검색과 활용 등을 위한 메타데이터(혹은 메타텍스트)가 위치하는 층위이다 따라서 전자텍스트의 메타

17) 반 데이크, 앞의 책, 213~222쪽 참고. 반 데이크는 자신이 말한 ‘초구조’ 개념이 이야기에서 가장 쉽게 설명된다고 말한다. 우리는 어떤 이야기를 듣거나 읽고 난 후에 그것이 광고나 강의 혹은 경찰조서가 아닌 어떤 이야기였다는 사실을 알게 된다. 즉 가택 침입에 대한 전형적인 이야기 구조는 경찰보고서, 신문조서, 보험신청 피해보고서 등의 텍스트와는 다른 것이다. 이런 텍스트 종류는 모두 상이한 커뮤니케이션 기능과 사회적 기능 뿐만 아니라 서로 다른 구조를 갖고 있다는 점으로도 구분된다. 반 데이크는 텍스트의 유형을 표시해 주는 이러한 총괄적인 구조를 초구조라고 부른다. 이러한 초구조는 텍스트의 내용에 일정한 제한을 부과한다 하더라도 이야기의 내용(거시구조)와는 무관한 것이다.

데이터(혹은 메타텍스트)를 전자텍스트의 초구조라고 볼 수도 있다.

초구조는 지식정보의 여러 측면에 대한 기술적(descriptive)¹⁸⁾ 요소들로 구성된다. 이 기술 요소들은 디지털 지식정보의 관리, 검색, 활용 등이 보다 신속하고 효율적으로 이루어지 위한 전제조건에 해당한다. 또한 검색의 효율성이라는 측면에서도 지식정보에 대한 표준화된 기술 요소로서의 초구조의 중요성이 더욱 부각되어야 한다.

그런데 아직까지도 대부분의 전자텍스트는 체계화된 기술정보를 갖고 있지 않다. 기술정보를 기초로 하는 텍스트 관리와 공유, 그리고 검색법 등에 대한 적극적인 모색이 필요하지만, 그렇지 못한 상황이다. 현재는 기술정보 즉 일종의 메타텍스트를 활용하는 대신, 관리자라는 인적 지원을 매개로 정보이용자에게 봉사하는 측면이 강하다. 모든 단계에서 정보는 동적으로 유지되어야 하며 더 나아간 지식정보를 제공하기 위해서는 정보 자체가 끝없이 수정되고 갱신되고 추가되어야 한다. 뿐만 아니라 정보 자체의 양적 크기가 급속하게 증가하고 있다. 이러한 상황에서 인적 지원에만 의존하는 정보 관리 및 봉사는, 얼마 지나지 않아 한계점에 도달하고 말 것이 자명한 일이다. 지금 우리 시대에서 잘 설계된 기술표준 즉 지식정보의 초구조를 정의해 두지 않으면 장차 전자텍스트의 관리와 활용에서 큰 어려움이 발생하게 될 것이다.

전자텍스트의 초구조를 구체화함에 있어서, 정보의 원활한 공유와 활용에 초점을 맞출 것인가, 아니면 정보 자체의 특성을 최대한 살리는 데 초점을 맞출 것인가 하는 문제도 고려해 보아야 한다. 하지만 정보 자체의 특성은 앞에서 다룬 정보의 거시구조와 더 밀접한 관련이 있는 것으로 본다면, 초구조는 전자에 초점을 맞추어서 설계되어야 할 것으로 보인다.

초구조 설계의 성패는, 전자텍스트에 초구조가 필요하다는 인식과 합의에 도달하는 데에 우선적으로 달려 있다. 그 후 초구조에 대한 정확한 개념과 필요한 요소들을 도출해야 한다. 이런 작업이 어느 정도 진척된다면 이를 적극 수용해서 실험하고 정착시키는 단계가 필요하다. 이 단계에서는 너무 원칙만을 고집하기 보다는 탄력적인 적용이 이루어지도록 고려하고, 나아가 제도적 지원을 통해 빨리 정

18) 이 글에서 ‘기술’ 혹은 ‘기술적’이라는 용어는, 특별한 지시가 없을 경우에 모두 Description 혹은 Descriptive 의 의미로 사용된다.

착될 수 있도록 도와야 한다.

2. 초구조의 기술(description) 요소

전자텍스트의 초구조는 텍스트의 생산에서 관리, 활용에 이르는 모든 단계에서 필요한 기술요소를 포괄해야 한다. 초구조는 텍스트의 화용론적, 의미론적 요소들을 두루 포괄하는 넓은 영역을 거느린다. 그런 만큼 초구조의 전체적인 면모는 새로운 연구영역으로 남겨둘 수 밖에 없다. 그러나 초구조에 속할 수 있는 몇 가지 요소들의 예를 보여주는 것마저 불가능한 것은 아니다.

우선, 화용론적 중요성을 가진 초구조 요소로, 저자/화자, 장소, 날짜 등을 생각해 볼 수 있다. 구어에서와는 달리, 글로 쓰여진 텍스트의 경우는 화자가 누구인지, 언제 어디서 텍스트가 생산되었는지 화용적 맥락으로는 즉시 분명해지지 않을 경우가 있다. 이러한 경우에 초구조 요소로서의 저자/화자, 장소, 날짜 등은 중요한 의미를 갖게 된다. 텍스트가 갖는 진실성이나 타당성도 텍스트 생산의 시간과 장소에 의해 결정되는 경우가 많음을 상기할 때, 이러한 초구조 요소가 얼마나 중요한지 인식할 수 있게 된다.¹⁹⁾ 전통적인 인쇄문화에서는 종종 이러한 요소들이 해당 텍스트와 같은 공간에 표시되기도 했다. 신문보도처럼 텍스트가 생산되는 날짜를 신문의 상단에 따로 표시해 두는 경우가 그런 경우에 해당한다. 서간문의 경우에도 텍스트가 생산되는 날짜와 장소가 끝부분에 표시되었다. 물론 서적의 경우에도 따로 표기되는 출판 일에 대한 정보를 통해 텍스트가 생산되는 시기를 추정할 수 있도록 한다. 이처럼 텍스트 전달체에 따라 다양한 형태를 유지하던 화용론적 요소들이 전자텍스트에 와서는 초구조 속에서 일괄적으로 통합될 수 있게 된다. 이렇게 될 경우 다양한 텍스트 전달체의 정보를 통합적으로 이용하기 쉬워지게 된다.

또한, 서적의 경우에는 서문, 후기(혹은 발문) 등이 있다 이들은 독자에게 맥락에 대한 정보를 제공하는 특별한 화용적 과제를 갖는다. 즉 텍스트를 쓰게 된 동기과 원인이나 목적, 텍스트의 내용과 기능에 관계되는 취지, 텍스트 제작시 있었

19) 반 테이크, 앞의 책, 250쪽 참조

던 특별한 문제들의 소개, 텍스트가 노리는 독자층 등이 그것이다²⁰⁾ 서문에 자주 쓰이게 되는 이러한 요소들도 전자텍스트의 초구조를 이루는 기술요소들이 될 수 있다. 그리고 책의 후기들에서 간혹 접하게 되는 텍스트를 둘러싸고 있는 사회문화적 환경의 변화와 이에 따른 텍스트 해석의 변화가능성에 대한 주석 등도 초구조의 한 요소라고 볼 수 있다. 반 데이크가 ‘동반적인 텍스트’라고 불렀던 서문이나 후기 등에는 전자텍스트의 초구조를 이룰 수 있는 여러 요소들에 대한 예시가 내포되어 있는 셈이다.

그렇다면, 전자텍스트의 초구조를 설계하는 데 우리가 참고할 만한 전례는 없는가? 앞에서 초구조는 메타데이터가 위치하는 층위라고 했다 따라서 메타데이터 기술원칙에 대한 국제적인 표준으로 만들어진 Dublin Core와 ISAD(G)을 살펴보면 도움을 얻을 수 있을 것이다. 먼저, 더블린 코어(Dubin Core)는 1995년 미국 오하이오 주의 더블린 회의와 1996년 Warwick University 회의를 통해 만들어졌다. 이는 전자 데이터베이스에 담긴 서지정보를 이용자가 찾을 수 있게 하는 기본적인 데이터요소를 확인하는 것이 목적이었으므로, 이를 통해 확인된 15개의 데이터 요소는 다음과 같다. 표제(title), 제작자(creator), 주제(subject), 설명(description), 발행처(date), 기타 제작자(contributor), 날짜(date), 자료유형(type), 형식(format), 식별자(identifier), 출처(source), 언어(language), 관련자료(relation), 내용범위(coverage), 이용조건(rights). 이들 각 요소는 선택적이며 반복적으로 사용할 수 있다. 더블린 코어의 기술요소들은 가능한 한 단순한 형태를 유지하여 비전문가들도 정보 자원에 대한 간단한 기술레코드를 손쉽게 작성하거나 이해할 수 있도록 하는 데 주안점이 있다. 하지만 이러한 단순성에 대한 고려가 텍스트의 적절한 관리나 정확한 탐색에 대한 요구와 알맞은 균형을 유지했는지 의심해 보지 않을 수 없다. 지식정보사회가 본격화되면, 지식정보의 정확한 탐색이 더욱 중요해질 것으로 보인다. 결국 더블린 코어의 기술요소와 기술원칙을 더욱 확장하고 상세화시켜, 우리가 목표로 하는 전자텍스트의 초구조 속에 포함시켜야 할 것으로 보인다.

한편, ISAD(G)²¹⁾는 일관성과 완결성을 가진 기술요소의 국제적인 표준화 노력의 일환으로 마련되었다. 서문, 서론, 용어집 등의 앞부분, 본문 3장, 그리고 부록

20) 반 데이크, 앞의 책, 252쪽.

21) General International Standard Archival Description

으로 구성되어 있다. 본문 3장에서는 26개의 기술요소들이 7개영역으로 나뉘어 조직되어 있다. 7개의 영역에 포함되는 기술요소들은 다음 표와 같다.

영역	요소	비고
식별 영역	참조코드 표제 일자 기술계층 기술단위의 규모와 매체	필수영역
배경 영역	생산자명 행정연혁/개인이력 기록물 이력 수집/이전의 직접적 출처	필수영역
내용과 구조 영역	범위와 내용 평가, 폐기, 처리일정 정보 추가 정리체계	필수영역
열람과 이용조건 영역	열람조건 재생산조건 언어와 자체 물리적 특성과 기술적 요구사항 검색도구	필수영역
연관자료 영역	원본의 존재와 위치 사본의 존재와 위치 관련기술단위 출판주기	필수영역
주기 영역	주기	필수영역
기술통제 영역	아키비스트 주기 규칙과 협약 기술일자	

이들 각 기술요소는 요소명, 요소의 목적, 사용을 위한 규칙, 지시와 규칙의 이행, 예시 등의 체계로 설명된다. 예컨대 식별영역에 속하는 표제부분은 이렇게 설명되어 있다.

3.1.2 Title*Purpose:*

To name the unit of description.

Rules:

Provide either a formal title or a concise supplied title in accordance with the rules of multi level description and national conventions.

If appropriate, abridge a long formal title, but only if this can be done without loss of essential information.

For supplied titles, at the higher level, include the name of the creator of the records. At lower levels one may include, for example, the name of the author of the document and a term indicating the form of the material comprising the unit of description and, where appropriate, a phrase reflecting function, activity, Subject, location, or theme.

Distinguish between formal and supplied titles according to national or language conventions.

Examples:

Helen Lucas fonds (Fonds)

The Christmas Birthday Story production records (Series)

The Christmas Birthday Story (Item)

Canada, York University Archives

St. Anthony Turnverein organizational records (Fonds)

U.S., Minnesota Historical Society

Papers of J. Lawton Collins (Fonds)

Appointment Books, 1948-1955 (Series)

U.S. National Archives & Records Administration

그러나 ISAD(G)는 모든 기록물을 대상으로 한 기술 원칙을 제정하는 데 목적이 있었기 때문에, 전자텍스트를 위한 기술요소로서 그대로 받아들이는 데는 문제가 있다. 따라서 그것이 디지털 정보에 어떻게 적용될 수 있는지 심각하고 구체적으로 고려해 볼 필요가 있다. 즉 이 표준이 제안하는 기술요소들을 어떻게 하나의 검색도구 안에 통합하여 접근점을 제공할 것인가에 대한 명확한 모델을 구성해내고, 또 각 데이터요소가 사용되는 구체적인 방식에 대한 지침 부족의 문제, 연관자료 등의 항목에서 정보요소가 중복되거나 결여되는 문제 등도 해결해야 한다.²²⁾ 이러한 문제들을 하나씩 해결해 나가면서 바람직한 전자텍스트의 초구조에 접근해 나가는 것이 필요하다.

22) 한국국가기록연구원, 「ISAD(G)를 적용한 한국기록물기술규칙 개발에 관한 연구」(한국국가기록연구원, 2002)에서 ISAD(G)에 대한 기본적인 문제들이 검토된 바 있음

이상의 논의를 종합해 볼 때, 초구조에 포함될 기술요소들이 간단히 정의될 수 있는 것이 아님을 알 수 있다. 초구조의 필수요소로 누구나 인정할 수 있는 참조 코드, 표제, 생산자, 생산장소, 날짜 등은 물론이고, 텍스트 생산의 동기·원인·목적, 텍스트의 내용·구조, 연관자료, 독자층과 열람 및 이용조건, 텍스트 생산의 사회문화적 환경 등 다양한 요소들이 초구조 속에서 잘 정의될 수 있어야 한다. 이를 위해서는 다양한 형태의 텍스트 전달체에 대한 구조분석 작업이 선행되어야 한다.

예컨대, 낱장이나 낱장의 묶음으로 된 텍스트, 단행본 텍스트, 편집류 텍스트들이 갖게 될 각 기술구조와 이들 상호간의 연관성을 고려한 계층적 기술구조에 대한 고려도 필요하게 된다. 또한 원본텍스트, 수정텍스트, 주석텍스트의 경우에도, 각각에 대한 기술요소들은 물론이고 상호 관련성에 기초한 기술요소들도 정의해야 한다. 수집을 통해 집적된 텍스트는 다양한 유래를 갖는 것들이 혼재하기 때문에, 이에 대한 기술원칙을 정하는 일은 좀 더 복잡한 양상을 보인다. 그리고 최종적으로는, 각각의 기술요소들을 누가, 언제 기술할 것인지 즉 기술의 주체와 기술의 시점 등에 대한 원칙도 마련해야 한다. 이처럼 전자텍스트의 초구조를 정의하는 데는 복잡다기한 문제들이 관련되지만, 텍스트들 간의 상호관련성에 대한 고려는 특히 중요한 문제이다. 텍스트는 하나 하나가 독립적인 정보가치를 갖기도 하지만, 여러 개의 텍스트가 일련의 내용적인 계층성과 상관관계 아래서 그 가치가 더욱 확장된다. 텍스트의 이러한 성격을 초구조에 적절히 반영해야만, 생산자나 이용자가 효율적으로 생산하고 활용할 수 있는 전자텍스트가 완성될 것이다. 아직 갈 길이 멀다.

V. 끝맺기

본고는 문제를 해결하기 보다는 새로운 문제를 제기하는 데 초점이 맞추어졌다. 지식정보를 생산하고 관리하고 유통하는 체계는 지식정보 자체의 표현형식을 결정하는 중요한 요소가 된다. 지식정보의 표현 형식은 단지 내용을 담는 그릇이나 매개가 아니라 내용을 ‘선규정’하는 표상의 장치이며 또한 그것은 지식의 경계를 결

정하고 사유를 전개하는 영역이 된다. 이에 따라 인쇄매체와는 다른, 디지털매체에 맞는 새로운 시대의 지식정보 표현 형식이 개발되어야 한다. 즉 지식정보를 담는 그릇으로서의 전자텍스트에 대한 새로운 개념과 구조를 정의할 필요가 생겨나게 되었다. 이는 전자텍스트를 보다 효율적으로 생산하고, 가공하고, 관리하고, 유통하고, 활용하기 위한 문제와 직결된 중요한 사안이다.

전자텍스트의 구조에 대한 연구는, 텍스트정보학이라는 학문분과를 신설하고 본격적으로 이루어져야 한다. 전자텍스트는 크게 미시구조, 거시구조, 초구조로 나눌 수 있는데, 이들 각각은 텍스트정보학의 중요한 연구과제들이 될 수 있다. 이러한 산적한 연구과제들 중에서 본고는 거시구조와 초구조에 대한 문제들을 부분적으로 다루었다. 먼저 거시구조와 관련해서는 단위텍스트의 문제를 다루었다. 단위텍스트는 기존의 파일이라는 정보단위보다는 훨씬 작게 세분화된 정보단위이다. 단위텍스트들은 이용자의 요구에 부합되게 해체되거나 새롭게 조합될 수 있는데, 이런 특성으로 인해 텍스트의 활용성과 확대재생산성은 극대화될 수 있게 된다.

초구조는 정보의 보존과 공유, 검색과 활용에 쓰일 메타데이터가 위치하는 전자텍스트의 최상위 층위를 말한다. 이는 지식정보의 여러 측면에 대한 기술적 (descriptive) 요소들로 구성된다. 초구조의 설계는 전문가들의 합의를 바탕으로 이루어져야 할 문제이면서, 동시에 정보텍스트학과 같은 학제적 학문분과의 도움을 받아 다양한 실험과 검증을 거쳐야 한다. 따라서 선부른 접근은 불가능하다. 하지만 본고에서는 초구조의 다양한 기술요소들을 살펴봄으로서 장차 초구조 연구의 기초를 마련하고자 하였다.

디지털문화는 표준화를 지향하는 속성을 지닌다. 그리고 이를 통한 개방과 공유의 정신은 초창기 디지털문화가 지니고 있는 힘이기도 했다. 하지만 표준화를 지향하는 움직임이 늘상 긍정적인 힘으로 작용하는 것만은 아니다. 그것은 다른 말로 하면 획일화와 단순화를 의미하기도 한다. 따라서 전자텍스트를 크게 미시구조, 거시구조, 초구조로 나누어 각각의 표준적인 구조를 정의하려는 이 글은 문화적 다양성을 거세하는 쪽에 힘을 보태고 있는 듯이 보일 수도 있다. 하지만 이 글에서 제안한 단위텍스트는 도그마적인 위치를 점하는 텍스트를 견제하는 다양한 변종의 텍스트를 보다 쉽게 생산할 수 있도록 하기 위해 만들어진 개념이다. 이는 사유의 다양성과 텍스트의 다양성을 최대한 수용하기 위한 전자텍스트의 가장 기

본적인 개념이 될 것이다. 또한 초구조를 선부르게 규정하지 않고, 그 기술요소의 확정을 유예시킨 것도 텍스트의 다양성을 충분히 고려하여 바람직한 초구조를 정의해 내기 위한 글쓰기 전략이다.

이제 전자텍스트 연구의 많은 부분이 텍스트정보학의 학제적 연구과제로 남게 되었다. 이러한 과제들을 해결하는 과정에서도, 전자텍스트의 생산과 활용의 효율성은 최대한 고려되어야 한다. 하지만 지식정보의 구조화 및 표준화 작업은 언제나 ‘문화적 다양성을 어떻게 지켜내고, 확대시켜 나갈 것인가?’ 하는 또 다른 문제도 염두에 두어야 한다.

참고문헌

- 손동현 외, 『정보사회의 철학적 진단』, 서울: 철학과현실사, 1999.
- 장노현, 「하이퍼텍스트 서사에 관한 연구」, 한국정신문화연구원 박사학위논문, 2002.
- 천정환, 『근대의 책읽기 독자의 탄생과 한국 근대문학』, 서울: 푸른역사, 2003.
- 한국국가기록연구원, 「ISAD(G)를 적용한 한국기록물기술규칙 개발에 관한 연구」, 한국국가기록연구원, 2002.
- Michael J. Hernandez, *Database Design for Mere Mortals*, SciTech, 2003.
- Michel Foucault (이정우 역), 『지식의 고고학』, 서울: 민음사, 1992.
- Pierre Levy (강형식, 임기대 옮김), 『지능의 테크놀로지』, 서울: 철학과현실사, 2000.
- Richard Light (채규혁 역), 『차세대 웹의 혁명, XML』, 서울: 도서출판 대림, 1998.
- Teun A. van Dijk (정시호 역), 『텍스트학』, 서울: 아르케, 2001.

- | | |
|---|-----------------------|
| ● 투고일: 2003. 12. 13 | ● 심사완료일: 2003. 12. 28 |
| ● 주제어(key word): 전자텍스트(electronic text), 지식정보(knowledge-information), 단위텍스트(Unit Text), 웹문서(Web document), 초구조(Superstructure), ISAD(G) | |