

데이터베이스 이야기

유경희

(사)한국복지정보통신협의회 부회장

독일 뮌헨에서 필자가 국제 표준화회의의 참석을 위하여 독일의 고도 뮌헨을 방문할 기회를 가졌다. 오늘이 사흘째인데 내일은 귀국행 비행기를 탈 작정이다. 지금 이글은 호텔의 방안에서 쓰고있다. 이것을 다 쓴 다음 세수하고 조반을 마치고 회의장에 들어가야 한다. 회의 도중에 잠시 나와서 귀국행 비행기의 탑승 확인전화를 하고 나서 노트북 컴퓨터를 들고 아래층으로 내려가서 특수 플러그를 빌려서 하이텔에 연결하여 이 글을 전송할 작정이다.

왜냐하면 원고 마감일이 오 늘이기 때문이다. 어제와 그제, 필자가 하고 있는 하이텔 원로 방의 게시판에 뮌헨기행 1, 2, 3, 4 그리고 한 개의 메일을 전송하였다. 이상하게도 01410은 연결이 안되었고 747-7071로 연결을 하였다. 전송하고 있는 모습을 본 미국 사람들이 놀라워하는 것을 등뒤로 느꼈다.

이번 회의의 요점은 “국제전자출판표준”을 개정하자는 내용의 회의였으며, 미국의 Graphic Communication Association (GCA)가 주관해서 개최하는 SGML Europe 96이라는 5일간의 행사의 진행과 함께 개최하는 것이었다. SGML이란 “Standard Generalized Markup Language”라는 말의 약자이며 이미 원안은 국제표준으로 채택이 되었으며 한국에서도 KS로 제정한바가 있다.

이와 바로 연결된 “전자출판표준”(Electronic Manuscript Preparation and Markup)이라는 표준이 1994년도에 공표되었으며 한국에서는 아직 표준화되지 않았다. 쉽게 말하면 전자출판표준이라고 말할 수가 있다. 국제표준화기구(ISO)의 제46기술위원회(TC46) 제4분과위원회(SC4) 제6작업반(WG6)에서 담당하고 있다. 전자출판표준이 생겨나게 되어서

수많은 저술가, 편집가, 학자들의 관심을 불러일으켜서 붐이 일어나고 있다는 사실을 작년의 보스톤회의에서 보고온바가 있으며, 그때는 이미 “워드프로세서”의 시대는 지났다.

이제는 SGML의 시대가 왔다고 할만큼 키보드를 두들기는 사람들의 관심을 사고 있다. 흔히들 어느 작가가 책을 출판하려고 하면 200자 원고지에 글을 써서 무조건 출판사에 맡겨야 하며, 출판사에서는 그 복잡한 편집, 교정이라는 과정을 거쳐서 타자 식자를 해야 한다. 그리고 원판을 제작해서 인쇄출판하는 과정이 상식으로 되어 있다. 그러나 SGML 시스템으로서는 글쓰기 자체부터 SGML로 하고 모든 과정이 모두 자동적으로 이루어지게 된다. 뿐만 아니다. 이걸로 바로 “데이터베이스”를 만든다. 즉 데이터베이스 제작 도구로서 활용된다는 것이 획기적이다. 또한 이것으로서

CD-ROM을 바로 제작할 수도 있다. 음성발생장치를 연결하면 소리로 발생하기도 한다. 그렇다면 작곡가가 화면에서 작곡하면 금방 음악으로 연주해 보여주기도 한다. 점자기를 연결하면 점자책도 발간하여 준다. 여기에 멀티미디어를 수용할 수도 있다. 화면에서 화상이나 영상을 보면서 글을 쓸 수가 있다. 회의에서 들은 이야기이지만 인터넷의 WWW에 수록된 글의 90%의 글은 non-SGML문건이라고 한다. 이것을 SGML로 변환하는 프로그램도 수록되어 있다고 한다.

지금 이 “전자출판시스템”에서는 복잡한 수학 공식의 수용 방안을 강구하기 위한 개정안을 심의 중에 있다. 앞으로는 모든 멀티미디어를 수용할 수 있도록 계속 개정해 나갈 방침으로 있다고 한다. 분야별 요구로서는 화학 공식의 표현 서지정보의 표현 등이 요구되고 있다고 한다. 이미 모든 한글과 모든 한자를 포함한 유니코드가 국제표준화가 되었으며 또한 국내에서도 국가표준화가 되었다. 그렇다면 한국에서는 무엇보다도 SGML을 이 유니코드체계와 접목을 시켜서 한국판 “전자출판시스템”을 완성하여야 할 것이다. 이것은 데이터베이스를 만드는 기능도 가지고 있으므로 한국판 데이터베이스의 제작 시스템으로 봐도 될 것이다.

지금 시각은 독일 시간으로 아침 6시40분, 서울시간으로 오후 1시40분. 이 호텔 방에서는 국제 연결이 안되어서 할 수 없이 아래층으로 내려가야 한다. 아마도 호텔의 전화교환기에 문제가 있는 모양이다.

그래서 직통전화를 빌려서 쓰는데 문제의 전화 플러그가 표준화되어 있지않아서 들고 간 플러그가 맞지 않는다. 결국 그걸 빌려야 하는데 그것을 빌리는데 줄을 서야 할 정도이다. 시간을 봐서 오늘 중으로 보내도록 하겠다. 이제는 데이터베이스의 개념이 달라졌다. 영숫자만의 데이터베이스 개념으로 정책을 수립하거나 직접 상업용으로 만들거나 해서는 안된다고 본다. 한가지 유의할 점은 한글·한자 문장 처리 시스템이 없이 데이터베이스를 만든다는 것은 문제가 있다. 영숫자 일변도의 사고방식으로서 안된다는 것이다. 혹자는 한글·한자는 표현만 가능하면 된다고 하지만 역시 문장 처리 기능을 컴퓨터에 어느 정도 부여하여야 우리글의 데이터베이스화가 이루어질 것이다.

다시 말하지만 표현만 가지고는 안된다는 것이다. 멀티미디어란 우선 글, 소리(말), 그림, 영상(동영상) 등을 데이터베이스에 담아야 한다. 우리글의 문장 처리를 개발하지 않고 맹목적인 멀티미디어로 간다면 결

국은 중국적으로 우리의 말이 없어져 버리고 말게 될 것이다. 가령 공공데이터베이스의 개발 기준에 반드시 각종의 표준화를 기하는데 더욱 신중하여야 한다. 표준화에 대하여 무신경한 사람이 개발하는 데이터베이스는 결코 바람직하지 못하다. 초고속이나 멀티미디어에만 가중치를 더 줌으로서 우리글자의 처리 능력을 고려하지 않는다면 그것도 바람직하지 못하다. 한가지의 꿈이지만 데이터의 표준화와 함께 한글·한자 등을 처리할 수 있는 시스템으로서 SGML을 기반으로 한 “한국 SGML센터”와 같은 기관이 있어서 여기서 수많은 데이터베이스, CD-ROM, 책자 등을 제작할 수 있었으면 좋겠다. 1986년에 공포된 국제 표준 SGML, 이어서 이에 관련된 표준 TR 9573, HTML, DSSSL 등에 이어서 1994년에 12083 (전자원고 작성 및 마크업)이 공포됨으로서 SGML산업이라고 불릴 만큼 수 백개 기업이 관련 소프트웨어 개발에 투자하고 있다. 불행히도 한국의 기업은 찾아볼 수가 없었다. 일본에서도 이번 유럽 회의에 10여명이나 참석하였으며, 이미 일본어 SGML을 개발해서 보급하고 있는 실정이다. 