

□ 기획연재 □

컴퓨터 과학 산책(9)

### 정보폭발은 문화파괴로 갈 것인가?

- 소프트웨어 산업이 나가야 할 방향 -

한국과학기술원 최기선\*

#### 0. 언어는 목표인가? 뜻의 작은 구멍은 있어도 뜻은 무너지지 않는다?

임진왜란, 일제시대는 왜 있었나? 조선은 일본의 중국 침략을 위한 교두보 역할을 한 것일까? 아니면 그 목표의 전부이었나? 언어정보소프트웨어의 대표적 문제인 윈도우95 코드 문제와 이에 뒤이은 워드프로세서의 국가별 코드화와 지역화는 그것으로서 끝나는 목표의 전부인가 아니면 그 다음의 다른 소프트웨어로 이어지는 교두보인가. 해답은 언어정보소프트웨어는 응용시스템이 아니라, 기본 시스템소프트웨어로서 향후 소프트웨어 산업의 교두보 역할을 한다는 것이다.

인터넷 정보검색, 워드프로세서, 시스템소프트웨어에 언어와 정보가 결부되지 않은 것이 없기 때문이다. 한국어정보처리는 그 길 위에 놓여져 있는 것이다. 뜻의 작은 구멍이 나지 않도록 뜻을 쌓는 방법을 강구하여야 하는 절실한 시기이다. 대란을 준비하는 방법을 생각한다.

#### 1. 정보대란의 대두

가격파괴, 정보홍수와 같은 도발적 어휘가 비도발적으로 들린다. 곧 이어 정보파괴, 정보폭발과 같은 비보편적 용어가 보편화할지도 모른다. 정보라는 어휘가 언제부터 일상용어화 하였는지 알 수 없지만, 십 년전만 하더라도 정보는 중앙정보부나 안전기획부의 전문용어로

서 보통 사람이 마음대로 쓸 수 있는 말이 아닌 비보편적 용어이었다. 십 년도 안되어, 말의 해방이 이루어진 것이다. 디지털 혁명, 정보혁명의 결과로서 얻어진 결과이다. 도발적 어휘의 폭발과 감수성의 비도발적 약화가 무감각적, 비전공적, 비체계적, 비생산적인 결과가 되고 있다.

#### 2. 공동적 공감각적 노력 집약이 필요한 항목

이에 대한 체계적·과학적 접근이 필요한바, 제2의 코드혼란과 국가적 코드 대란이 일어나서는 안된다고 생각한다. 이와 같은 폭발과 파괴에 따른 대란은 준비의 부족에 있다 하겠다. 제2차 대란의 대상은 어디인가? 그것은 용어에 있다. 코드로서 표현되는 것은 글자이며, 의미있는 글자는 용어이다. 전문적 용어의 비표준적 사용은 제2차 정보대란의 원인이다. 정보는 용어로서 생산, 축적되고 색인, 검색이 이루어진다. 이에 따라, 다음과 같은 구체적 항목에 대한 공동적 노력이 필요하다고 하겠다. 즉, 개념의 언어 자원의 확립이 범국가적으로 이루어져야 하는 것은 당연하다:

- 사전정보로서 일반용어의 계산언어학적 집대성 및 보급,
- 전문용어의 집대성 및 보급,
- 또, 이에 대한 가공 및 관리기술로서 다음의 항목이 필요하다고 하겠다. 즉,
- 정보가공, 활용기술
- 정보개발체계
- 정보관리체계

\*중신회원

### 3. 비파괴적, 문화적 준비

마이크로소프트의 한국어 코드문제-이것은 코드대란으로 표현할 수 있다. 이 코드 대란에 의한 정보파괴는 한국 문화 파괴에 해당한다는 대중 및 정부의 태도가 있었다는 것을 상기한다. 다음에 이어질 파괴의 대상은 정보 일반에 관한 것인 바, 다음 대란의 대상이 되는 것은 정보의 근간이 용어인 것이다. 일상용어, 전문용어의 한국어정보처리적 집대성, 이를 위한 정보개발, 관리체계의 표준화 및 실용화 기술의 선결이 필수사항이다.

### 4. 차세대 운영체제와 통신은 언어정보 처리 도구이다

더 심각한 것은 앞으로의 운영체제와 통신환경은 고도의 언어정보처리 도구일 수 밖에 없다. 시스템 소프트웨어의 발전 역사는 기계와 컴퓨터의 차이를 줄이기 위한 줄기차고 끈질긴 노력의 거울이다. 지금까지의 접근방법은, 영어의 비교우위론 위에서 만들어진 소프트웨어를 지역별 코드입력이 가능한 소프트웨어로 전환하는 방법이다. 다른 하나는, 유니코드와 같은 소프트웨어 변형이 필요하지 않은 방법을 쓰는 것이다.

언어정보소프트웨어의 세계화는 양방향성이 있다. 외국의 유명 언어정보소프트웨어와 국내의 산물이 같은 시장을 동시에 공급할 수 있는 방법론이 생긴 것이다. 그러나, 과연 그러한가? 같은 유니코드를 쓰더라도 각 언어에 따른 문화적 특성을 고려하는 문화적 방법이 우선하여야 한다. 외국의 유명소프트웨어 회사는 국내의 문화적 준비가 없을 때, 비문화적 언어정보 소프트웨어를 준비할 수 밖에 없을 것이다. 우리는 한국어정보처리를 바탕으로 문화적 언어정보소프트웨어의 기반과 이에 따른 차세대 운영체제와 통신방법에 대비하여야 한다.

### 5. 정보검색은 다음의 코드대란의 문화적 침략 대상이다

정보검색시스템에 들어가는 소프트웨어 기술

은 종래의 데이터베이스 관리시스템과 운영체제와 같은 시스템소프트웨어적 부분과 언어정보에 해당하는 언어정보시스템의 둘로 나누어 생각할 수 있다. 정보검색 소프트웨어는 운영체제와는 차원이 다른 면이 있다. 첫째는 그 보급의 정도가 아직은 운영체제와 같은 필수불가결한 것이 아니되, 곧 운영체제의 일부분이 될 것이라는 시간적 전환점에 와 있다.

둘째, 정보검색 소프트웨어는 데이터베이스 관리시스템과는 다른 문화적 특성을 갖는다는 점이다. 데이터베이스 관리시스템은 외국의 유명 브랜드를 쓰더라도 문화적 침략성이 없지만, 언어정보의 핵심인 정보검색 시스템은 문화적 특성을 갖는 특수성이 있다. 이스라엘의 Enigma와 같이 이스라엘 히브리어 정보검색 소프트웨어에서 출발하여 국제화된 소프트웨어로 커질 수 있는 예가 있다. 거꾸로 외국의 정보검색소프트웨어가 한국어, 그리고 한국 문화의 결정체인 용어에 대한 심각한 고려없이 들어올 때, 그 한국어의 사용자들은 시장을 빼앗기면서도 제대로 서비스를 받지도 못하는 상황이 된다.

### 6. 응용이 응용이 아닌 때가 되는 시대

15 년전에 한국어정보처리를 연구한다고 하면, 과연 먹고 살 수 있을까 하던 말을 들었고, 십 년 전에는 응용의 응용이라고 하였고, 지금은 “응용” 소프트웨어라고 한다. 지금 마이크로소프트 윈도95에 네트워크 소프트웨어가 같이 들어가 있고, 컴퓨터 사용자가 가장 시간을 많이 보내는 소프트웨어 중의 하나인 넷스케이프가 과연 “응용” 소프트웨어로 남아있을 것인가는 “아니다”라는 것이다. 곧 사용자에게 보편화될 시스템소프트웨어는 컴파일러가 아니라 정보검색시스템이나 워드프로세서와 같은 언어정보 소프트웨어일 것이기 때문이다. 이제 언어정보처리라는 “응용”이 아니라 “기본” 소프트웨어인 것이다. 왜 그러한가?

사용자 개인의 정보량도 많아질 것이다. 문서 탐색도 정보검색이고, 인터넷 정보탐색도 정보검색이고, 기계번역도 일종의 정보검색일 수 밖에 없다. 이제 응용이 응용이 아닌 시대

에 들어선 것이다.

### 7. 과연 한국의 소프트웨어 산업은 애니메이션 제작만 남는가

한국의 소프트웨어 산업은 줄기찬 노력을 하였지만, 변두리적 노력만 하여왔고, 세계의 중심에 서질 못하였다. 정보검색, 기계번역, 워드프로세서와 같은 언어문화적 소프트웨어의 육성은 향후 한국 소프트웨어 산업의 발전 방향을 위하여 후퇴할 수 없는 징검다리이며 교두보이다. 이제 여기서 머뭇거리면 그 다음에 다가올 소프트웨어산업은 다시 외국 소프트웨어의 수입과 지배에 들어갈 것이다. 다만 남는 것은 외국 소프트웨어에 의한 애니메이션 하청업이 영상사업이라는 포장으로 남을 것이기 때문이다.

### 8. 한국어정보 소프트웨어가 한국의 문화적 생존이다

우리는 무엇을 필요로 하는가? 우리 국민은 무엇이 필요한가? 일생을 두고 배우는 영어, 한자교육에 대한 열의, 자기 생활 관리, ... 일생을 두고 괴롭히는 언어장벽의 극복, 이에 대한 개인적 콤플렉스, 국제적으로 성장 못하는 문제, ... 국가적으로는 무엇이 필요한가? 국가적 정보, 외국 정보의 집대성 미흡에 의한 국제적 열등감 ... 과연 고도의 문화적, 과학적 접근이 필요한 언어정보소프트웨어를 외국이 판매할 것인가? 또, 그러한 소프트웨어를 외국이 가지고 있다고 하더라도 과연 우리에게 팔 것인가? 아무 준비도 없는 상대방에게 좋은 도구를 팔 이유가 없다. 그 상대방이 그 가치를 모르기 때문이다. 좋은 도구는 개발자의 전유물인 것이다. 따라서, 개인의 생존의 혈로에 서있는 불로초를 우리가 가질 수 있는가? 이에 외국의 정보검색 소프트웨어를 지금 쉽고 싸다고 하여 우리 기술을 포기한다면 단순히 MS-DOS의 한글화 수준의 문제가 아니라, 이제는 문화적 열등감 문제와 소프트웨어 산업 전반적 열등감으로 넘어간다. 과연 우리는 변두리에서 북만 칠 것인가?

### 9. 기본기술은 옷을 갈아 입으며 발전한다.

인터넷, CALS, 에이전트, ... 이들에 대한 기본 기술은 정보 처리 기술이며, 정보의 대부분은 언어로 표현된다. 선진국은 왜 언어정보소프트웨어에 대한 기반적 노력을 기울여 왔는가? 일본의 전자사전주식회사(EDR; 통신부 지원), 미국의 정보검색 (TREC; 국방부 지원), 유럽공동체의 다양한 언어공학적 지원 (Language Research Plan), 전세계의 초고속통신망과 디지털라이브러리에 대한 노력이 이에 속한다. 언어정보처리는 이제 “응용”의 “응용”도 아니요, “응용”도 아니다. “기본” 시스템소프트웨어이며 컴퓨터와 사람의 차이를 줄일 수 있는 “기본” 기술인 것이다.

### 10. 국제표준화의 주도적 역할 자임

국제적 표준의 주역은 그 기술 개발자 만의 특권이다. 개발되지 않은 기술에 대한 표준은 없다. 한국의 국제적 표준작업에서 항상 뒤떨어진 역할을 생각할 때, 언어정보처리의 표준화 작업에 대한 적극적 태도에 의한 문화권 제 국가의 입장 공감대 형성 및 지원에 의한 국제적 입장 강화가 이루어져야 할 절호의 기회라는 사실을 생각하여야 할 때이다.

### 11. 한국인을 위한 언어정보시스템은 세계인을 위한 것이다

수입대체품이 아닌, 적극적 수출지향적 언어정보시스템의 개발을 지금 이야기하고 있다. 가장 훌륭한 방어는 공격에 있다고 한다. 동반자적 국제인, 국제국가로서 우리의 심장이식수술을 체형에 맞지 않는 외국인의 심장을 쓸 것인가?

### 12. 결 론

한국어 정보처리를 위한 전방위적 노력, 정부의 지원이 필요한 때이다. 이에 대하여, 교육에서의 혁신, 언어자원에서 혁신, 산학연 협

동에서의 혁신이 필요하다.

### 12.1 전산학 교육의 언어정보처리의 조기 교육화

전산학의 교과과정에서 언어정보처리는 학부에서부터 조기 교육이 필요하다. 이 분야의 필요수요에 비하여 공급자 입장인 교육기관의 교과과정은 미국의 교육체계가 좋다고 하여 그것만을 따라가고 있다. 우리는 우리실정에 맞는 교과과정이 필요하다. 이에 대하여 언어정보처리는 과학철학적 입장에서 정규과학화하여야 하며, 패러다임을 설정할 수 있는 언어공학적 기반을 만들어야 한다. 종래의 인공지능(AI)은 자원보다는 그 가공기술에 초점을 두었다. 이제 이 언어정보처리의 문제는 가공보다는 자원에 있다. 이에 따라, 하나의 패러다임으로서 정보아키텍처(IA<sup>1)</sup>)와 같은 우리의 학문적 체계를 두어야 하겠다.

1. IA는 AI에 반하여, 자원에 초점을 두는 패러다임이다.  
IA=AI-1

### 12.2 언어자원 개발과 활용기술의 집대성

언어자원의 사유화는 국가적 노력의 낭비이다. 국가적 표준화된 방법에 의한 전자사전, 코덱스, 음성데이터 등에 대한 집대성과 활용기술의 축적이 필요하다. 이 기술과 자원은 차세대 운영체계의 핵심이다. 일차 자원을 그대로 파는 것보다는 가공 자원화, 혹은 가공 기술에 의한 부가가치가 있는 상품으로 유도되어야 한다.

### 12.3 學研產官 협동에 의한 실용화 조기 정착

이 기술은 전산학, 언어학, 문헌정보학 등의 학제적 연구의 산물이다. 전담 연구부서에 의한 끈질긴 집대성과 추진력이 필요하다. 學研에 의한 표준화와 집대성에 대한 노력은 官에 의한 표준화와 지속적 지원이 대두된다. 이에 따라, 산업은 이를 이용하여 상품화에 노력한다. 우리 언어정보처리의 고도화에 의한 선진 소프트웨어화의 지속적 노력이 산업에 반영되어야 한다.

## ● '정보문화의 달' 기념 강연회 및 전시회 ●

- 일 자 : 1996년 6월 20일(목)
- 장 소 : 경북대학교 전자계산소 4층
- 내 용 : 강연회, 패널토의, 시연회 등
- 주 최 : 영남지부
- 문 의 처 : 경북대학교 컴퓨터공학과 유기영 교수  
T. 053-950-5553