

보도자료

Super-jin의「소통광장」

국회의원 **진 성 호**
(한나라당 중랑을)

- 국회 | 서울시 영등포구 여의도동 1번지 국회의원회관 823 호
- 전화 | 02-784-3865 ● 팩스 | 02-788-3823
- 홈페이지 | <http://www.superjin.com> ● E-mail | superjin@na.go.kr



도대체 멀 한거지? 부실 의혹 ‘21세기 세종계획’

**10년간 150억원 투입하고도 당사자 외엔 아무도 모르는 사업~
몇몇 대학에 연구비만 지원, 2007년 감리보고서에선 “향후 재할용 어렵다”**

국회 문화체육관광방송통신위원회 소속 진성호 의원(한나라당, 서울 중랑乙)은 IMF 구제 금융 신청으로 온 나라가 위기에 처한 1997년 당시 8억 5천만원을 시작으로 2007년까지 150억원이 투입된 국립국어원의 ‘21세기 세종계획’에 대한 부실 의혹을 제기할 예정이다.

※ ‘21세기 세종계획’은 세종대왕께서 훈민정음을 창제한 정신을 이어받아 정보화 시대에 걸맞는 국어 정보화의 기반을 구축하기 위해 지난 1998년부터 2007년까지 10년간 국비 150억원을 투입한 대규모 국책사업으로, 우리말을 수집·정리하여 세계 수준의 고품질, 표준화된 대용량 언어자원 구축과 이를 활용하는 언어학적 전자사전을 개발하는 사업으로 기술적, 경제적으로는 국어정보화와 언어공학 발전을 위한 우리말 기간자원을 구축하는 사업이며, 사회적, 문화적으로는 실재적 쓰임새에 기초한 체계적이고 분석적인 우리말 정보·지식 자원을 구축하는 사업.

연도별 예산 및 집행 현황 (단위: 백만 원)

연도	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	합계
예산액	850	1,250	1,482	1,730	1,738	1,969	1,871	1,684	1,471	1,021	15,067

※ 세종 계획 당초 계획 대비 결과 분석 ※

(1) 기초 자료 구축

구분	계획	결과	
현대국어 말뭉치	5억 어절 (+7,000만)	9316만 어절(구어 466만 어절, 전문용어 200만 어절 포함) 9 + 7000만 어절	18.6%
현대국어 분석 말뭉치	7,900만 어절(+200만)	1,230만 어절	15.6%
북한/해외한국어 말뭉치	4,000만 어절	1,083만 어절	27%
옛문헌/방언/구비문학 말뭉치	2,600만 어절	637만 어절	24.5%
대역(한국어-외국어) 말뭉치	1,000만 어절	766만 어절	76.6%
계	6억 5500만 어절	2억 32만 어절	30%

10년 동안 약 2억 어절을 구축하였지만, 이 2억 어절 속에는 이전에 구축하여 온 각종 말뭉치, 즉 국립국어원에서 표준국어대사전 편찬을 위하여 만들어 놓은 말뭉치와 정보통신부의 'STEP 2000'이란 국가 과제로 입력해 놓은 말뭉치가 포함되어 있다. 따라서 말뭉치 구축 위주의 기초 자료 구축은 원래의 목적에 비해 약 30%만 시행되었다.

(2) 전자사전 구축

구분	목표	결과
기본 사전	80만 단어	60만 항목
전문용어 사전	35만 단어	
연어 정보 사전	170만 연어	
대역 사전	31만 단어	
개념 사전	30만 개념	

전자사전 구축은 원래의 계획에 비해 매우 축소되었다. 그 이유는 일반 국민들이나 국어 정보 처리 연구자들의 요구하는 대용량의 전자사전이 아니라, 언어학적으로 매우 정밀한 전자사전을 구축하려고 한 데에 기인한다. 따라서 전자사전 구축은 언어학적 연구로서는 성공한 셈이지만, 다양한 방면으로 활용하려는 용도로서는 그리 효율적이지 못한 편이다.

(3) 한민족 언어 정보화

구분	목표	결과
신조어·전문용어 순화	전산·전자 분야 문화·산업·경제 분야 전 분야	신조어 순화 : 없어짐 전문용어 : 전문용어 센터로 이관
남북한 언어 비교 사전	기초 연구, DB 구축 비교 사전 편찬 DB 및 사전 확충	13,000 항목 구축
한국어 교육 지원	기초 자료 지원 조직화된 DB 지원 지식 DB 및 온라인 평가	한국어 세계화 과제로 이관
언어정보 통합 네트	기반 구조 형성 서비스 체제 확충 워드 네트 구현	없어짐
남북한 및 해외동포의 정보화	개인 컴퓨터용 소프트웨어 개발 인터넷 등 통신용 및 방송망용 프로그램 개발 양적, 질적 확장 단계	없어짐
남북한 간 이질성 극복 사업	해외 한국어학 지원 (연변) 남북 교류센터 지원 남북 자판, 부호 표준 연구 활동 지원 남북한 공동 사전 남북한 언어(어휘) 비교 사전	없어짐

한민족 언어 정보화는 원래의 계획에서 크게 후퇴하였다. 뿐만 아니라 예산의 대폭 삭감과 잦은 사업 변경으로 인하여 중도에 연구책임자 및 연구자 모두가 사퇴하는 지경에까지 이르렀으나, 국립국어원에서 이를 수정하겠다는 약속을 받고 철회한 적도 있다. 그러나 그 약속은 지켜지지 않았다.

(4) 국어 정보 처리 표준화 분야

구분	목표	결과
기준 부호계	기준 부호계 틀 마련 사용자 인터페이스 마련 통신망(인터넷 등)을 통한 원격 접근 시스템 구축	<u>없어짐</u>
비표준문자 등록 및 표준화	비표준문자 수집 비표준문자 관련 정보연구(음, 훈 등) 비표준문자 등록 및 수집상황 보급 상설 운영체제 구축	비표준문자등록센터에서 수행
OCR 전용 글꼴	국내 보급 OCR시스템들의 성능 및 오류 특성 분석 OCR 전용 글꼴개발(시험안 마련) OCR 전용 글꼴보완(확정안 마련)	<u>없어짐</u>
한글 자형 표준화	한글 한자의 국내 자체의 정리. 표준화 대상 목록 작성 한자 자형 사전의 편찬. 옛 한글 자형 사전의 편찬. 기초 상용한자의 대표 자형 지정. 표준 자형과 약자체의 결정.	<u>없어짐</u>

옛 한글 및 한자코드를 정비한 것은 국민들의 국어 정보화를 원활하게 하려는 의도에 따른 것이지만, 대부분의 사업이 소멸되었다.

(5) 핵심 소프트웨어 개발 분야

구분	계획	결과
국어 처리 소프트웨어	빈도 조사 프로그램 등 문체 분석기 지능형 문서 작성기	용례 추출기(글잡이) 문장 분리기
말뭉치 구축 소프트웨어	형태소 해석기 구문 분석기 국어 생성기	지능형 형태소 분석기 구문 태그 부착기 지능형 구문 분석 도구
언어 지식 처리 소프트웨어	개념 분류 체계 분야별 시소러스 조사 결정 지식 획득 소프트웨어	<u>개발 못함</u>

언어지식 처리 소프트웨어는 관련 부처(정통부)의 미참여로 인한 기술력이 부족을 이유로 개발되지 못했다.

▶ 문제점 1. 무리한 추진으로 당초 사업목표를 달성하지 못했다!

1997년 당시 문화체육부와 과학기술처, 정보통신부 3개 부처가 공동으로 '우리말과 우리글을 바탕으로 하는 정보사회 건설'을 목표로 추진되었지만 사업 시행초 과학기술처와 정보통신부가 갑자기 빠지면서 '언어학적인 관점에서 공통적으로 필요한 언어 정보 기반 구축'이라는 연구 목적 사업으로 변질되었다.

이 두 부처가 빠지면서 각 부처에서 담당하던 과제들이 모두 추진되지 못했고, 이로 인해 결국 '국어정보화'라는 큰 목표는 절름발이 상태로 진행이 되어 사업이 종료된 지금 현재도 국어정보화 구축의 실효성에 대한 논란이 일고 있다.

▶ 문제점 2. 기존에도 같은 연구가 있었다! 중복 사업으로 국고 낭비~

1992년부터 1999년까지 국어원에서는 100억원의 예산을 들여 표준국어대사전의 자료 등 '국어정보화 사업'을 추진한 바가 있었다. 이 사업으로 한글맞춤법, 표준어, 로마자표기법, 외래어 표기관련 자료 등 상당한 양의 기초 자료가 구축되어 있었을 것으로 예상된다.

1999년 말에 구축한 국어정보보급관리시스템에 이미 한글맞춤법, 표준어, 로마자표기법, 외래어표기법 관련 일부 데이터베이스와 검색프로그램이 개발되었는데도 데이터베이스를 추가로 구축하거나 검색기능을 고급화하는 등 시스템을 보완하는 내용으로 사업을 하지 않고 21세기 세종계획이라는 별도의 어문규정 검색 시스템 구축 사업을 무리하게 추진하여 국고를 낭비하였다는 비판이 있다.

▶ 문제점 3. 국어정보보급 관리시스템 개발 허위 보고

국어정보보급관리시스템 개발사업은 국어원에서 1992년부터 1999년까지 100억원의 예산을 들여 편찬 작업을 해왔던 표준국어대사전의 자료 등 국어정보화사업을 통해 생산된 국어 관련 자료(한글맞춤법, 표준어, 로마자표기법, 외래어 표기관련 자료 등 포함)들과 문화관광부가 21세기 세종계획의 일환인 한민족언어정보화사업으로 1998년부터 1999년 8월까지 별도로 구축한 한글맞춤법, 표준어 규정 자료들을 대상으로 유니코드 파일로 변환할 수 있는 툴을 만들어 단일관리체

계의 데이터베이스로 구축하고 일반 국민들이 인터넷 웹기반에서 자료를 검색할 수 있는 기능을 제공하기 위한 시스템을 만들기 위한 것으로서 국어원에서는 1999년 7월 6일 대우정보시스템주식회사와 위 같은 과업내용으로 계약을 체결한다.

그러므로 21세기 세종계획 등에 따라 생산되는 국어 관련 연구성과물을 신속하고 체계적으로 보급하여 국민의 국어생활화에 기여하기 위해 개발한 위 국어정보보급관리시스템의 개발 효과를 높이기 위해서는 위 시스템이 과업지시내용대로 개발되도록 하여야 하고, 또한 2000년부터 2001년까지 계속 시행된 한민족언어정보화사업의 산출물들도 당초 위 국어정보보급시스템이 채택한 유니코드로 데이터베이스화하여 통합하고 인터넷 웹서비스까지 할 수 있도록 일관성있게 시행하는 것이 바람직하지만, 국어원은 1999년까지 매킨토시 컴퓨터에서 사용할 수 있는 출판용 소프트웨어 프로그램으로 작성한 표준국어대사전 포제어 자료 50만건을 대우정보시스템주식회사가 유니코드 파일로 변환시키는 중간과정에서 많은 오류가 발생되어 이를 보정하기 위한 시간 지연으로 결국 표준국어대사전 자료를 유니코드 파일로 변환시키지 못해 위 시스템에 등재하지 못하였는데도 2000년 당시 국어정보보급관리시스템 개발 사업이 과업 내용대로 완료된 것처럼 준공 처리하였다.

▶ 문제점 4. 2007년 감리보고서에서의 지적: 향후 재활용이 어렵다?

세부사업 과제별로 말뭉치 정제에 대한 오류율 목표가 존재하지만 이를 검증할 수 있는 오류에 대한 정의 및 유형, 산출 기준이 되는 모집단 즉 분모 크기 산정에 대한 정의가 수립되어 있지 않기에 오류율 산정 자체가 불가능한 상태이며, 또한 자체 검수 실적에 대한 증빙 산출물이 없기에 납품된 연구결과물의 품질상태를 확인할 수 없다고 지적하고 있으며,

기초자료 세부과제의 연구보고서 중 말뭉치 정제와 관련하여 절차의 각 항목내용에 대한 상세 내용 부재로 인하여 상세 수행 과정 파악이 안됨으로 인해 절차의 적정성 검증이 불가능할 뿐만 아니라 향후 본 과업이 되풀이 될 때 재활용이 어렵다고 감리하고 있다.

▶ 문제점 5. 10년간 특정 대학에 연구비 지원 사업으로 변질!

세종 계획이 추진되면서 정보화의 저변 인구를 확대한다는 목표 달성을 위해 1

차년도 사업의 경우에는 전국의 각 대학 소속 교수들이 연구자들로 참여하였지만, 2차 년도부터 각 중요 과제들의 연구책임자 소속 대학의 연구자들만이 참여하면서, 국어 정보화 사업이 일부 한정된 대학(서울대, 고려대, 연세대 위주)의 연구비 지원 사업으로 변질되었다.

(※ 10년 동안 연구자들이 전국적인 분포를 보이면서 과제를 시행한 분야는 '한민족 언어 정보화' 과제 뿐)

또, 세종 계획의 참여자는 원래 국어국문학을 비롯한 인문학자 70%, 그리고 정보 처리나 자연언어 처리 전문가 30%가 참여하는 것으로 계획되었으나, 실제 연구진 대부분이 국어국문학을 비롯한 어문학자들만 참여함으로써 폭넓은 국어 정보화라는 당초 목표에 부합하지 못했다.

이에 대해 진성호 의원은 “동 사업이 국고 150억원을 투입하고도 국민은 물론 관련 부처 공무원들도 모르고 있는 사업이라는 점에서 부실 의혹이 있다”고 지적하고 “현 정부가 한글 세계화 사업을 국정의 주요 정책으로 추진하려고 준비 중인 상황에서 기존 사업에 대한 면밀한 성과 분석 없이 또 중복 추진 될 우려가 있어 문화부의 정밀 감사를 요청한다”고 말했다.

※ 문의: 진성호 의원실 김용환 비서관(T. 784-3865)