

‘기술정보 제공을 위하여

-화학 문헌정보 서비스-

정기원

1974년 8월은 필자에게 시간적 여유가 있었던 시기였다. 그 당시 필자는 전매청 프로젝트 팀에 소속돼 있었고, 그 동안 개발한 프로그램들에 대한 테스트를 완료하고 실제 업무에 적용하는 운용단계로 넘어가는 시점이었다. 그러나 이러한 여유가 오래갈 수는 없었다. 성기수 박사님을 통해 KORSTIC(한국과학기술 정보센터)에서 어떠한 도움을 필요로 한다는 연락이 왔기 때문이다. 내용인 즉, CACS(Chemical Abstract Condensate Service)를 위한 프로그램을 개발하고 운용토록 지원해 달라는 요청이었다. 이는 화학 및 화공분야 문헌정보 서비스를 자동화하는 일이었고, 마침 그 당시 필자는 KIST 도서관의 박계홍 차장과 정덕영 씨와 함께 KIST 도서관 업무 전산화 문제를 협의하고 업무분석을 수행하고 있던 차라 KORSTIC 요청에 적극적으로 임하게 되었다.

KORSTIC 업무중에서도 핵심이 되는 업무는 문헌정보서비스라 할 수 있는데 해외에서의 문헌정보서비스 기술을 조사한 바에 의하면 이미 여러 가지 종류의 서비스가 컴퓨터로 수행되고 있음을 알았고, 그 중에서도 가장 활발한 분야가 화학 및 화공분야로서 미국화학회(American Chemical Society)는 화학문헌 초록서비스(Chemical Abstract Service)를 전산화하여 화학 및 화공분야 종사자들에게 정보를 제공하고 있었다.

특히 CACS(Chemical Abstract Condensate Service)는 최신 문헌의 서지사항들을 일주일에 한번씩 자기테이프에 수록하여 문헌정보 서비스 기관에 제공하고, 서비스 기관은 이를

처리하여 등록된 화학 및 화공분야 종사자들에게 문헌 정보를 정기적으로 알려 주는 서비스를 제공하였는데 국내에서도 이를 요청하는 이들이 다소 있었다고 한다.

서비스 수준을 보면, 문헌 수요자가 서지사항을 제시하여 해당 문헌을 찾는 저수준의 검색에서부터 키워드들을 논리적으로 합성하여 문헌 데이터베이스로부터 검색해 내는 고급수준까지 있고, 또한 문헌정보 서비스를 정기적으로 제공하여 항상 최신기술이나 동향 파악에 철저하게 대비하도록 하는 등 여러 가지 종류가 있다.

이러한 문헌정보 서비스를 계획하는 단계에서는 서비스 수요를 예측하는 일과 문헌정보를 확보하는 일, 그리고 문헌정보 검색 서비스를 수행하는 일이 가장 중요하였으므로 KORSTIC에서는 문헌정보 수요자와 문헌정보를 확보하고 KIST에서는 서비스를 위한 컴퓨터 처리를 담당기로 하였다. 그러나 KORSTIC과의 협의가 진행되는 가운데 발견한 사항은 KORSTIC이 이러한 사업에 필요한 예산 확보가 전혀 되어 있지 않았다는 사실이었다. 당시 KIST로서는 소프트웨어 개발에 필요한 인건비나 경비 및 컴퓨터 사용료에 대한 대책이 없는 사업을 수행할 수 있는 환경이 아니었으므로 사업 진행이 불가능한 상황이었다.

사업수행이 예산문제로 벽에 부딪치자 KORSTIC의 유경희 씨의 적극적인 노력과 성기수 박사님의 특별한 배려로 컴퓨터 사용료 일부만 KORSTIC에서 부담기로 하고 우선 개발업무를 수행하게 되었다.

우선 시작한 일은 화학 문헌초록 압축서비스(Cheical Abstract Condensate Service)였다. 이는 일주일에 한번씩 최신 화학분야 문헌정보가 자기테이프에 수록된 것을 미국화학회(The American Chemical Society)로부터 입수하여 문헌 수요자에게 각자의 관심 주제에 해당하는 문헌들의 서지사항을 검색하여 알려 주는 일이었다. 이를 위해 미국화학회가 제공하는 자기테이프의 표준포맷(SDF: Standard Distribution Format)을 입수하여 이를 검토하기 시작하였고 테이프 처리에 필요한 중요한 두 가지 사항을 파악하였다. 그 한 가지는 포맷 설계로서 테이프에서의 기록밀도, 문자세트와 문자표현 표준, 레코드 포맷, 응용 프로그램 인터페이스, 자료 작성시 사용한 약어, 활자, 구두점 등에 관한 표준들이었다.

특기사항은 사지사항을 기록한 모든 자료항목들은 ASCII-8 코드를 사용하였으나, 테이프 레이블들은 IBM 운영체제에 맞도록 EBCDIC 코드를 사용하였고 포맷 제어는 모두 바이너리

로 코딩되어 있었으며 파일 포맷과 레코드 설계 내용을 이해하는 데 상당한 기간이 소요되었다.

필자는 FORTRAN으로 프로그램을 작성하였고 당시 이 작업을 위한 파일 처리에 고심을 했던 기억이 생생하다. 또 FORTRAN 언어로 Bit Manipulation을 많이 행해야 하므로 프로그램이 복잡해지고 자연히 프로그램 논리가 필자 자신이 이해하기에도 어려울 정도로 복잡하게 되었다. 그 당시에는 국내 기술 수준이 구조적 프로그래밍에 익숙치 않은 시절이었고 서브루틴의 활용은 프로그램의 오버헤드를 크게 하여 컴퓨터 시간의 절약에 도움이 되지 않는다고 여겼기에 가급적 피했다.

그러나 더욱 번거로운 일은 자기 테이프 변환을 필요로 한 것이었다. 우리가 보유한 CDC 컴퓨터로 그 테이프를 읽지 못하고 SCC(서울 컴퓨터 센터)에 가서 이를 변환하고 (9track 테이프를 7track 테이프로 변환) 이를 이진 파일로 취급하여 읽은 후 다시 해독해내는 번거로운 작업을 한 것이다.

아무튼 우여곡절 끝에 제법 규모가 큰 FORTRAN 프로그램이 완성되었고 결과는 성공적이었다. 다음해(1975)에는 이 프로그램으로 검색서비스를 실시하기에 이른 것이다. 지금 회고해 보면 이러한 프로그램이 전형적인 스파게티 프로그램으로서 현재 소프트웨어 공학 전문인으로서 가장 낮은 점수를 줄 수밖에 없는, 품질수준이 낮은 프로그램이었다. 그러나 그 당시로는 잘 돌아가는 프로그램이 완성되었다는 자부심과 함께 어려운 일을 훌륭하게 해냈다는 성취감으로 가슴 뿌듯하였으니 기술수준의 격차를 새삼 느끼고 격세지감을 금할 수 없다.