

국어 문헌 자료의 처리 방법

—워드프로세서와 데이터베이스를 이용한 문헌 입력의 실제—

김 병 선

(전북대 교수·현대문학)

I. 파일에 대하여

1. 종이 파일과 디스크 파일

파일(file)이라는 말은 본시 서류철 혹은 명부를 의미하는 말이었으나, 그 의미가 확장되어 오늘날에는 컴퓨터로 작성되는 서류나 명부 및 프로그램 등도 파일이라 부른다.

보통 종이 서류철(paper file)의 외관은 약간 두툼한 표지로 감싸여진 몇 장의 종이 묶음으로 묘사될 수 있다. 대체로 그 종이 묶음에는 무엇인가 내용이 되는 글씨가 쓰여 있고, 표지에는 그 내용과 관련된 제목이 붙어 있다. 서류를 많이 다루는 사람은 그 서류의 내용이 복잡하거나 종류가 여러 가지일 때에 서류철 첫장에다가 내용 목록(contents)을 써 두고 페이지 표시를 해 두어 서류철의 내용 및 순서를 잘 알아볼 수 있게 한다. 서류철이 좀 더 두툼해지고 복잡해질 때에는 이 밖에도 목록과는 별도로 색인(index)을 붙여서 사용하기 편리하도록 만든다. 이것이 종이 서류철을 운용하는 방법이다.

디스크 파일 역시 그 기본적인 운용 개념은 이와 동일하다.¹⁾ 디스크 파일은 외관상으로는 종이 파일과는 상당히 다르지만, 또 그것보다는 훨씬

1) 디스크(disk) 혹은 디스켓(diskette)이란 컴퓨터에 입력한 정보를 기록해 두는 보조 기억 매체를 말한다. 퍼스널 컴퓨터에서는 플로피 디스크(floppy disk)와 하드 디스크(hard disk)의 두 가지 종류가 사용되며, 컴퓨터에 입력한 정보를 디스크에 모아 두었다 하여 디스크 파일이라 한다.

더 초라하고 알파하지만 이것이 컴퓨터와 맞물려 동작할 때에는 종이 화일이 도저히 따라 올 수 없는 엄청난 일을 멋지게 해낸다. 디스크 화일 역시 여느 종이 서류철과 마찬가지로 표지에 제목을 가지고 있다. 이 제목은 스티커로 써서 붙일 수 있게 되어 있기도 하지만 진짜 제목은 볼륨 라벨(volume label)이라 하여 도스(DOS)상에서 디렉토리(directory)를 볼 때 나타나게 되어 있다.²⁾

디스크 한 장에는 용량이 허용하는 한도 내에서 여러 개의 서류들을 수록할 수 있는데, 그 각각의 서류들은 사용자에 의해 이름이 붙여져 그 서류의 길이나 저장되는 날짜 및 시간과 함께 디스크상의 일정한 자리에 등록되게 되어 있다.(디스크상에서 그 화일이 차지하고 있는 번지들도 아울러서 등재된다.) 이 곳이 바로 종이 서류철의 목록 부분에 해당하는 곳이다. 디스크 화일의 내용은 '0'과 '1'이라는 두 부호의 연속으로 바뀌어 저장되어 있기에 물리적으로 관찰할 수는 없으나, 이 신호들이 컴퓨터에 의해 읽혀지면 종이 서류철은 감히 따라올 수 없는 작업 경지에 이르게 된다.

2. 랜덤 화일과 시퀀셜 화일

종이에 글을 써서 정보를 저장해 둘 때에 사람들은 대체로 다음 두 가지 방식을 사용한다. 하나는 소위 서류라는 양식을 이용하는 것이요, 다른 하나는 백지(여기에는 학생들이 사용하는 노트처럼, 행을 구분해 놓은 것 이외에는 별다른 양식을 찾아 볼 수 없는 종이도 포함된다.)를 사용하는 것이다.

(1) 랜덤 화일

가령 주소록을 작성한다고 한다면 다음 그림과 같이 선을 그어서 사용하는 것이 좋을 것이다.

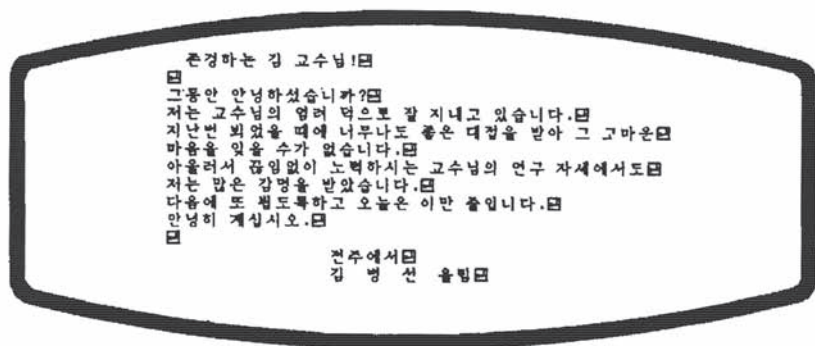
번호	이 름	우편번호	주 소
001	김 병 선	560-756	전주시 전북대학교 국어국문학과 (0652-70-3175)
002	정 영 인		부여읍 읍내리 238
003	소 강 춘	560-202	전주시 팔복동 삼화연립 A동 508호
004	김 여 름	570-101	이리시 남중 1가동 124-27
005	계갈공명		중국 상해시 선창가

2) 디렉토리란 디스크에 수록되어 있는 화일 목록을 말한다.

주소록이 이러한 양식으로 컴퓨터 디스켓에 수록된다면 그 화일은 랜덤 화일(random file)의 형태로 되어 있다고 말할 수 있다. 즉 몇 개의 정해진 항목이 있고(번호, 이름, 우편번호, 주소 등), 각각의 항목(이를 흔히 영역[field]이라 부르기도 한다.)들은 일정한 길이를 가지고 있다. 말하자면 번호는 3자리, 이름은 8자리(한글의 한 자는 영어의 두 자에 해당하기에 자릿수로서는 8자리다.), 우편번호는 7자리, 주소는 32자리를 차지하고 있는데 비록 내용이 들어가지는 않더라도 수록 인물의 수만큼의 자리는 차지하고 있는 화일인 것이다. 따라서 랜덤 화일의 크기는 대체로 한 행(이것을 레코드[record]라는 이름으로 부른다.)의 길이(여기서는 50자다.)에 행의 수(record의 수) 더하기 1한 것을 곱하면 나온다. (1을 더하는 이유는 각 항목들의 이름이 들어 있는 부분이 있기 때문이다.)

(2) 시퀀셜 화일

그런데 만일 어떤 사람이 편지를 쓴다면 편지지라고 흔히 불리는 종이에 다음과 같이 서류 양식의 구애는 받지 않고 써 나갈 것이다. (물론 기필이라든가 안부 등의 편지 내용 양식은 어느 정도 지키겠지만.)



이것이 바로 시퀀셜 화일(sequential file)의 일종이다. 이것이 디스켓에 수록될 때에는 각 행의 마지막 글자 바로 다음에 행이 바뀐다는 표시 [□가³⁾ 붙어서 저장되는데 따라서 그 화일의 총 크기는 그 화일의 전체 글자 수에 행의 수를 더하면 계산될 수 있다. 이 화일은 흔히 텍스트 화일(text file)이라고도 불리운다.

3) 여기서 □표시는 글자판(key board)의 [Enter](혹은 [Return]) 키를 누른다는 표시이다. 모든 명령은 엔터 키를 누름으로써 마감된다. 이 표시는 화면 상에는 나타나지 않는다.

3. '보석글'과 'dBASE'

현재 우리나라에서는 대다수의 컴퓨터 사용자들이 '보석글'과 'dBASE' 프로그램을 가지고 있고, 또 이를 많이 활용하고도 있다.⁴⁾ 그런데 '보석글'은 바로 시퀀셜 화일을 다루는 프로그램이고, 'dBASE'는 랜덤 화일을 다루는 프로그램인 것이다. 흔히 많이 쓰이는 말로 하자면 '보석글'은 워드프로세서(word processor : 문서 작성) 프로그램이요, 'dBASE'는 데이터베이스(data base : 자료 처리) 프로그램이다.

이 두 프로그램은 자기 장단점이 있는데 가령 '보석글'은 워드프로세서답게 입력과 수정이 손쉽고, 화면 전체를 사용하며 커서의 이동도 아주 자유로운 반면, 검색이나 정리에 있어서는 자료 처리 전문 프로그램을 따라갈 수가 없다. 한편 'dBASE'는 자료 처리에는 탁월한 면을 보이거나 원하는 문자체로 인쇄한다거나 자료의 위치를 뒤바꾼다거나 하는 데 있어서는 문서 작성용 프로그램을 따라갈 수가 없다.

위와 같이 화일을 디스크에 쓰는(write) 방식이 다르기 때문에 서로 다른 프로그램에서 다른 양식의 화일을 불러본다면 엉뚱한 결과가 나타나게 될 것이다. 그러나 이 두 프로그램(특히 'dBASE' 프로그램)은 서로 다른 양식의 화일을 사용하면서도 다행히 화일을 주고 받을 수 있는(이것을 수출[export], 수입[import]이라 부른다.) 방법이 있다. 따라서 이 두 프로그램의 장단점을 잘 살려서 화일을 주고 받는다면 아주 재미있는 결과가 나타날 수 있고, 컴퓨터의 기능을 여러 배로 확장해서 사용할 수 있게 될 것이다.

이 글에서는 특별한 프로그램용 언어(가령 베이직[BASIC]이나 포트란[FORTRAN], 코볼[COBOL], 씨[C] 등)를 사용하지 않고 '보석글' 특히 '보석글 II'와 'dBASE'의 내장 기능만을 가지고, 문헌을 입력하여 이것을 자료로 처리하는 방법에 대하여 서술하겠다.

II. 문헌의 입력

1. 어절 색인

어절 색인이란 어떤 문헌에 쓰인 전체 어절 목록을 가나다 순으로 정리

4) '보석글'은 삼보 컴퓨터의 워드프로세서 이름이다. 그리고 여기서 말하는 'dBASE'란 영문 데이터베이스 프로그램인 'dBASE II'를 삼보 컴퓨터에서 한글, 한자가 가능하도록 수정한 한글 자료 처리 프로그램 즉 'HANDB'를 지칭하고 있다. 앞으로 본문에서는 'dBASE'라는 명칭을 사용하기로 한다.

하고자 하는 색인 작업을 말한다.

이 어절 색인을 흔히 문맥에서 쓰인 중심어 색인(KWIC 색인 : key word in context)이라 부르는데 우리의 고문헌을 처리한 것으로는 일본인 福井玲에 의해 시도된 ‘月印千江之曲上 KWIC 索引’(1985)과 오종갑, 정홍교수 등에 의해 만들어진 ‘근세 국어 어휘 색인’(1986) 및 필자와 정영인 교수에 의해 만들어진 ‘철대만법 어절 색인’(1988) 등이 있다. 한편 현대시 분야로는 필자와 전정구 교수가 공동으로 만든 ‘소월시 어절 색인’이 근간될 예정으로 있다.

어절 색인을 만드는 목적은 어떤 문헌에 쓰인 모든 어절의 목록을 가나다 순으로 정리하여 어떤 특정한 어휘의 쓰임새를 검토해 본다든가, 통계적인 문체론을 시도할 때 연구의 객관성을 높인다든가, 어떤 개인이 글을 써 놓은 다음 자신의 어휘력이 어느 정도인지 무의식적으로 많이 반복해 쓰는 어휘는 어떤 것인지를 알아본다든가 문체를 교정한다든가 하는 데 있다.

따라서 어절 색인은 문헌을 다루는 많은 연구 분야에서 활용될 수 있지만 활용이 특히 더 기대되는 분야는 중세나 근세의 문헌 자료를 검토하는 국어학 분야일 것이다. 국어학 분야에서 가령 소설 문자를 검토한 논문이라든지 중세, 그리고 근대 국어에 관계된 기존의 논문들은 거의가 다 이러한 문헌 자료의 조사를 토대로 해서 얻어진 것이라 할 수 있다. 이런 점에서 문헌 자료는 좀더 신중하게, 그리고 면밀하게 검토될 필요가 있는데, 기왕의 작업들은 대체로 해당 문헌을 읽어가면서 필요한 항목만을 임의로 추출하고 있는 실정이다. 또한 자료의 추출을 위한 방법의 개선은 관심의 대상에서 제외하고 여전히 불편한 대로 수작업을 계속해서 사용해 오고 있다. 이러한 여건 속에서는 문헌의 전반적인 조사는 요구되는 시간이 너무 많아 부득이 한정된 주제에 해당되는 자료만을 선별할 수밖에 없고, 이런 식으로 선택된 자료는 좀 더 넓은 안목에서의 통찰에 제약이 가하기 때문에, 다른 주제를 위해서는 다시 얼마만큼의 시간이 동일 문헌의 조사에 필요하게 되어 궁극적으로 문헌 조사에 소요되는 시간이 효율적으로 운용되어진다고 할 수 없었다. 따라서 이러한 여건의 개선을 위해 컴퓨터를 이용하여 어절 색인집을 만들어 둔다면 국어학 연구는 그 방법면에서 진일보할 것이다.

2. 입력의 원칙

이제 '보석글'을 사용하여 문헌을 입력해 보자. 여기서 문헌을 입력하는 것은 이 문헌을 원하는 양식으로 인쇄하거나 저장해 두는 등 순수한 문서 작성상의 목적을 가지고 있는 것은 아니고, 그 문헌에 쓰인 전체 어절의 색인집을 만드는 데 있으므로, 다음 몇 가지 원칙을 지켜서 입력해야 한다.

우선 '보석글'을 부팅시킨 후 메뉴 화면(menu line)에서 [새 화일 만들(Create)]을 선택하고 'poetry'라는 화일을 만들어 준 다음 편집 상태로 들어가서 다음의 원칙대로 입력한다.⁵⁾

(1) 오른쪽 경계(RIGHT MARGIN)는 편의상 제70칼럼에 정해 둔다.

(| Esc |R)

(2) 연속 타자 모드에서 입력한다. (| Esc |W)

(3) 인쇄에 관계된 각종 양식 지정 명령들은 사용하지 않는다. 예를 들어 DOUBLE, LENGTH, FOOTNOTE 및 각종 문자체(밀줄 문자, 역상 문자, 굵은 문자, 축소 문자, 확대 문자 등)는 사용하지 않아야 한다.

(4) 하드 캐리지 리턴(HARD CARRIAGE RETURN, | Esc | Enter)은 사용하지 않는다. 문단을 구분하는 문장 재정렬(ALIGN)은 하지 않기 때문이다.

(5) 색인의 단위는 빈칸(SPACE)으로 구분된다. 행이 바뀔 때는 윗행의 끝에 빈칸이 하나 붙어 있는 것으로 간주되므로 어절을 구분한다 하여 굳이 다음 행 첫 칸을 비워 둘 필요는 없다.

(6) 문단을 구분해 주는 표시, 즉 들여쓰기(INDENT) 역시 해 줄 필요가 없다.

(7) 제목이나 이름을 입력해 줄 필요는 없다. 다만 본문만을 입력해 주는 것이 좋다.

(8) 말줄임표나 대시(—) 등으로 어절이 연결된 경우가 있으면 이를 떼어 준다. 그러한 문장 부호는 앞 어절에 붙여 준다.

(9) 사용자가 특별한 이유로 어절 색인에 넣고자 하는 접미사나 조사가 있다면 비록 띄어쓰기의 원칙에서는 벗어나더라도 그것을 자기 독립시켜 입력할 수도 있다.

위와 같은 원칙에 의하여 기타 입력을 편리하게 해 주는 보조적인 방법

5) '보석글'은 메뉴(menu) 방식과 명령(command) 방식의 두 가지로 사용할 수 있는데, 여기서는 메뉴 방식 위주로 설명하기로 한다.

령이 있는데, 그 정렬의 기준 단위는 어절(영어에서는 한 단어[WORD]이고, 우리말에서는 어절이다.)이다. (이처럼 어절 단위로 재정렬되면 저스티파이 얼라인[JUSTIFY ALIGN]일 경우 어떤 행의 경우에는 행 중간 중간의 빈 칸이 여러 칸으로 벌어지는 현상이 나타나 보기에 좋지 않기 때문에 이런 식의 재정렬법을 싫어하는 사람들도 있으나, 행의 끝이 어절 단위가 아니라 한 자 한 자 단위로 꺾여지면 보기에 좋지 모르나 검색과 치환 등의 작업이 제대로 이루어지지 않는 문제가 생긴다.)

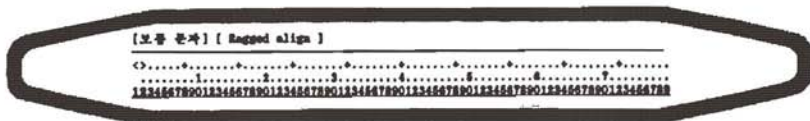
이처럼 어절 단위로 재정렬되는 점을 이용하여 어절 목록을 만들어 보자.

(1) 왼쪽 경계(LEFT MARGIN)는 제 1 칼럼에 정해 둔다. 그런데 '보석글'은 기동될 때에 이미 초기치가 1로 되어 있기 때문에 이를 변경시킨 경우가 아니라면 다시 정해 줄 필요는 없다.

(2) 오른쪽 경계(RIGHT MARGIN)는 제 2 칼럼에 정해 둔다. 문헌 입력시엔 제70칼럼에 두었으나 이를 변경해 주는 것이다. 커서를 제 2 칼럼에 두고 Esc R을 타자해 주면 된다.

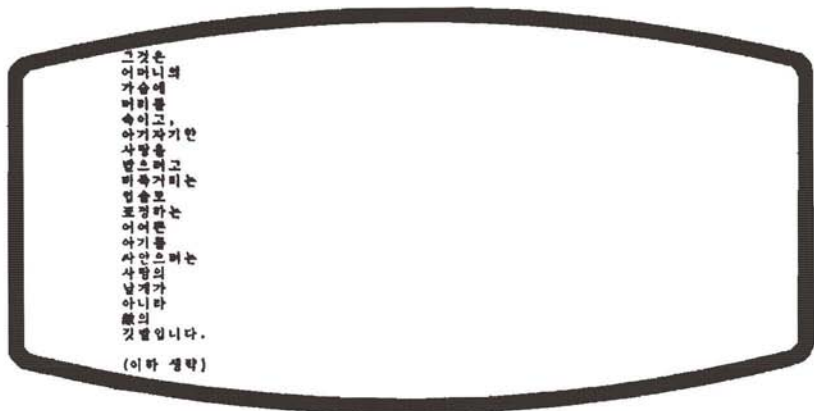
(3) 제대로 경계가 정해졌나 확인하는 방법은 화면에 눈금자(RULER)를 불러내 보는 것이다. 문서 작성 중에는 Esc G를, 편집이 끝나고 화면 아래의 명령 행에 "다음 명령은?"이라는 물음이 나왔을 때는 R 혹은 RULER를 타자해 주면 된다.

그 결과 다음 그림처럼 되어 있으면 제대로 된 것이다.



(4) 이제 [문장 재정렬] 메뉴를 선택하여 [문장 전체 정렬]을 실행해 보라. 이 경우 저스티파이 정렬(JUSTIFY ALIGN: 문서의 오른쪽이 한 줄로 잘 맞춰지는 방법이다. 따라서 행 중간의 빈칸이 원래보다 늘어나게 된다.)이나, 래그드 정렬(RAGGED ALIGN: 사용자가 입력해 놓은 빈칸을 그대로 유지한 채 정렬하기 때문에 자연스럽게 오른쪽 경계가 들쭉날쭉하게 된다.)이냐는 문제가 되지 않는다. 잠시 시간이 걸린 후 화면은 다음과 같이 바뀌어 있을 것이다. (p.117 화면)

(5) 입력한 문헌에 쓰인 어절 수의 합계는 편집 안내행(GUIDE LINE)의 총 행수에 나타나 있다. 위 화면과 같이 나열된 어절들은 텍스트의 순서 그대로이기 때문에 색인으로서는 활용할 수가 없다. 따라서 이를 다시



가나다 순으로 정리해 주어야 한다.

(6) 어절 색인을 만들기 전에 이 상태대로 디스크에 저장해 두어야 한다. 파일 이름은 사용자가 임의로 붙여도 되는데 여기서는 편의상 'poemindx.txt'라 해둔다. (좀 귀찮더라도 txt라는 확장자는 붙여 주는 것이 좋다. 이것은 나중에 'dBASE'로 파일을 수출할 때 사용되는 면장파도 같은 것이다.)

2. 단순한 어절 색인

우선 '보석글' 내부의 명령을 이용하여 정리해 보자.

(1) 글자 수가 가장 많은 어절을 찾고 그 글자 수(더 정확히는 바이트 [BYTE]수)를 알아 둔다. 이 작업은 부득이 수동식으로 할 수밖에 없다. 키보드 오른쪽의 숫자 키에 있는 [PgDn]키를 계속 눌러 가면서 확인하면 된다. 만해의 시에서 가장 긴 어절은 "정지되었습니다."이다. 이 어절의 바이트 수는 다음과 같이 계산된다. 한글 및 한자는 2바이트, 영문자, 숫자, 부호 등은 1바이트다. 따라서 이 어절의 바이트 수는 15이다.

(2) 가나다 순으로 정리를 하기 위해서는 정리 명령('보석글'에서는 이를 '배열[**SORT**]')이라 부른다.)을 사용한다. 우선 편집 상태에서 빠져 나와 ([Esc][Q], 화면 하단의 "다음 명령은?"이란 물음에 'SORT'라고 타자해 준다.

(3) 배열의 기준은 문자이므로 [문자(CHARACTER)]를 선택해 준다.

따라서 그 배경이 제시된 색인집을 만든다면 훨씬 도움이 될 것이다. 그 것이 다음에서 설명하는 고급 어절 색인이다.

3. 고급 어절 색인

고급 어절 색인은 표제 어절들만 제시하는 것이 아니라 그 문맥까지도 제시하는 색인이다. 즉 왼쪽 첫머리에는 색인 어절을 제시하고, 오른쪽에는 그 어절이 사용된 환경을 알 수 있게끔 그 어절을 중심으로 앞쪽은 3 어절, 뒤쪽은 2 어절 정도를 배열하는 것이다.

(1) 우선 단순히 어절별로만 분리해서 저장했던 화일(화일명은 poemindx.txt로 정해 두었다.)을 불러온다. (GET poemindx.txt)

(2) 이와 같은 문맥을 결합 색인을 만들려면 '보석굴'의 기능 중 끼워넣기(MERGE)를 이용하면 된다. 끼워넣기의 문법은 다음과 같다.

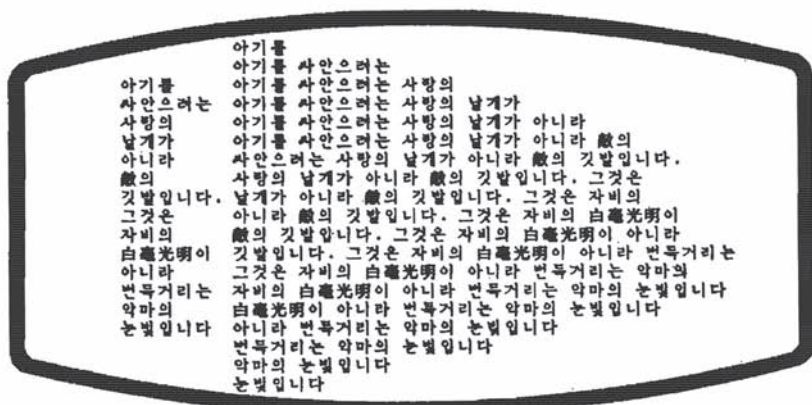
MERGE <끼워 넣고자 하는 화일명> <끼워 넣을 위치>

끼워 넣고자 하는 화일은 현재 디스크에 수록되어 있는 화일이어야 한다. 끼워 넣을 위치는 행과 열 두 가지 좌표로 지정되어야 하지만, 이 중 행의 값은 MERGE 명령을 내리기 직전 편집 상태에 있을 때 최종적으로 커서가 위치해 있던 화면의 첫 행으로 정해진다. 따라서 열만 지정해 주면 된다. 끼워 넣으려고 하는 화일은 조금 전에 불러왔던 poemindx.txt이다. 끼워 넣을 위치는 처음의 경우는 가장 길었던 어절의 바이트 수에 1을 더한 값이다. 여기서는 따라서 16이 되겠다.

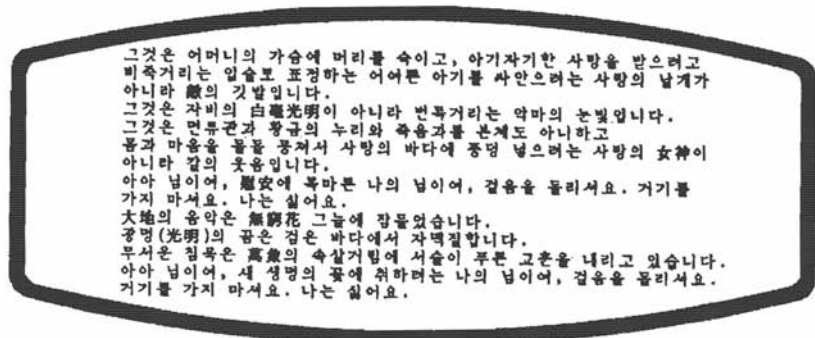
(3) 끼워 넣기 명령을 내리기 전에 화일 편집 상태로 들어가서 화일의 첫 행에 빈 행 하나를 삽입한다. [F9]



음 화면과 같이 바뀌게 된다.



(9) 이 상태에서 표제 어절이 없는 행들은 지워버리고(F10), 다시 오름차순으로 정리해 주면 우리가 원하는 어절 색인집이 나타난다.



(10) 이제 이 색인을 디스크에 저장하고, 또 프린트를 해 두면 장차의 활용에 대비할 수 있게 된다.

Ⅳ. 자료 처리

어절 색인집은 디스크 상태로나 인쇄된 문서 상태로도 그 활용 가치가 높

은 것이 되겠지만, 보다 완벽한 활용을 위해서는 전문적인 자료 처리 프로그램을 통하는 것이 좋을 것이다. 이제 여기서는 위와 같이 만들어진 어절 색인을 'dBASE' 프로그램으로 옮겨 활용하는 방법에 대해 서술하겠다.

1. 'dBASE'의 준비

'handb'를 기동시키고 새로운 데이터 파일을 만들어 주어야 하는데, 'dBASE'는 랜덤 파일을 다루기 때문에 자료의 구조를 먼저 결정해 두어야 한다. 그 구조는 앞에서 만든 고급 어절 색인을 수록할 수 있는 구조이어야 하며 아울러서 항목(FIELD)의 크기는 가급적 최소치로 결정되어야 한다.

여기서 필요한 필드는 어절과 그 문맥이다. 편의상 앞의 것의 명칭을 data라 하고, 뒤의 것의 명칭을 context라 하기로 한다. data 필드의 길이는 앞에서 여러 번 언급되었던 최대 글자 수, 즉 15이고 context 필드의 길이는 사용자가 일일이 파일을 넘겨 가면서 살펴보아 가장 긴 목록의 글자 수(바이트 수)로 정하면 된다. 만일 그것이 불편하고 귀찮다면 data 필드의 길이에 6(문맥에 사용된 어절 수)을 곱한 뒤에 적당히 줄이면 될 것이다.

(1) 이런 식으로 필드가 결정되면 그 자료 파일의 이름을 편의상 'manhae'라 하고 다음과 같이 명령을 준다.

```
.CREATE manhae6)
```

(2) 그 구조의 입력은 다음과 같이 한다.

```
FIELD NAME, TYPE, WIDTH, DECIMAL PLACES
```

```
001 data, c, 15
```

```
002 context, c, 80
```

```
003
```

(3) 'dBASE'가 데이터를 넣겠느냐고 물어 올 때에는 'N'라고 대답한다. 그리고는 방금 만들어 놓은 'manhae.dbf'를 열어 놓는다.

6) 여기서 점(.)은 데이터베이스의 프롬프트(prompt)로서, 명령을 받아들일 준비가 되어 있음을 의미한다. 대문자로 표시해 놓은 것은 'dBASE'의 내장 명령을 뜻하고, 소문자로 표시해 놓은 것은 사용자가 임의로 입력하는 내용임을 나타낸다. 그러나 실제 입력시에는 대문자 혹은 소문자 모두가 가능하다.

.USE manhae□

(4) 이제는 '보석글'로 만들어 두었던 시퀀셜 화일 'poemindx.txt'를 수입해 올 때가 되었다. 그 명령은 다음과 같이 내린다.

.APPEND FROM poemindx SDF□

여기서 SDF란 말은 Standard Data Format의 첫 자를 따서 만든 말로서 'poemindx'란 화일이 랜덤 화일이 아닌 시퀀셜 화일의 형태로 되어 있다는 것을 알리는 말이다. 드라이브는 빨간 불빛을 밝히면서 poemindx.txt 화일로부터 자료를 읽어다가 manhae.dbf 화일로 옮겨 줄 것이다. 그 작업이 끝나면 몇 개의 레코드가 수입되었는지를 다음과 같이 나타내 준다.

0000 RECORDS ADDED

이제 어절 색인의 표제어는 data 필드에, 그 문맥은 context 필드에 각각 입력이 된 것이다.

2. 자료의 검색

'dBASE'와 같은 자료 처리 프로그램은 검색(SEARCH)작업에서 탁월한 기능을 가지고 있다. 위와 같이 입력된 화일을 검색하려고 한다면 우선 정리를 해 두어야 한다. 그 정리하는 방법은 두 가지가 있는데, 하나는 SORT를 하여 새로운 데이터베이스 화일을 만드는 것이요, 다른 하나는 INDEX 화일만을 별도로 만들어 주는 방법이다. (필자는 후자를 권한다.) 여기서는 인덱스 화일을 만드는 법과 그 활용법을 서술하겠다.

(1) 인덱스 화일의 생성

여기서는 주로 data 항목을 이용할 것이므로 이 data 항목을 기준으로 인덱스 화일을 만들어 준다. 그 명령은 다음과 같다.

.INDEX ON data TO datindx□

datindx는 임의로 정한 인덱스 화일의 명칭이고, data는 정리의 기준이 되는 필드임을 나타낸다. 'dBASE'는 하나의 데이터베이스 화일에 대해 여러 개의 인덱스 화일을 운용할 수 있도록 해 준다.

(2) 인덱스 화일의 활용

위와 같은 명령으로 인덱스 화일이 만들어진 다음에는 이를 사용하기

위해서 다음 명령을 내려야 한다.

```
.SET INDEX TO datandx
```

이제 모든 레코드는 data의 순서대로 재정리된 것처럼 보이게 된다. 필자가 SORT 보다는 INDEX를 권한 이유는 검색에 있어서의 신속성 때문이다. 자료 화일의 경우에 인덱스가 되어 있으면, 사용자가 찾고자 하는 자료를 금방 찾을 수가 있다. ('dBASE' 설명서에는 단 2초만 걸릴 뿐이라고 말하고 있다.) 그런데 비해 SORT된 화일에서는 어떤 자료를 찾으라고 명령을 내리면 'dBASE' 프로그램은 첫 레코드부터 끝 레코드까지 하나씩 하나씩 검토해 나가기 때문에 그 시간이 많이 걸리는 것이다.

인덱스된 화일에서 어떤 자료를 찾으려면 SEEK나 FIND라는 명령을 사용한다. 가령 '님'이라는 어휘를 찾는다면 다음과 같이 SEEK 명령을 내린 다음 DISPLAY 하면 된다.

```
.SEEK "님"
.DISPLAY
```

만일 어절의 첫 부분이 "님"이라고 되어 있는 자료의 숫자를 알고 싶다면 다음과 같이 명령을 내린다. (이 경우는 인덱스의 여부와는 관계가 없다.) 그러면 'dBASE'는 그 수를 대답해 줄 것이다.

```
.COUNT FOR data="님"
```

그 외의 활용은 사용자들이 시도해 보기 바란다.

3. 화일의 수출

만일 'dBASE'에서 사용하던 화일을 '보석글'로 옮기고 싶다면 그 역이 가능했듯이 이것도 가능하다. 여기서는 어떤 조건을 만족하는 자료만을 '보석글'로 옮기기로 한다.

(1) 수출의 방법

'dBASE' 화일은 랜덤 화일인데 이를 시퀀셜 화일로 바꾸기 위해서는 역시 앞에서 수입할 때 사용했던 면장인 SDF란 말을 사용해 주어야 한다.

.COPY TO hanja SDF□

여기서 hanja 라는 것은 새로이 만들고자 하는 시퀀셜 화일의 이름인데 실제로 만들어지는 화일의 이름은 'hanja.txt'가 된다.

(2) 조건의 제시

편의상 한자어로 시작된 어절만을 '보석글'로 옮긴다는 조건으로 한다. 한자어로 되어 있다는 것은 그 어절이 한글의 마지막 자보다 크다(>)는 것을 뜻하므로 이를 'dBASE'의 문법에 맞게 표현한다면 다음과 같이 쓰여질 수 있다.

.COPY TO hanja SDF FOR data>"함"□

'dBASE'는 이 명령을 받은 후 어절 중에서 한자어로 시작되는 것만을 찾아 hanja.txt (txt 라는 확장자는 'dBASE'가 자동적으로 붙여준다.) 라는 새로운 화일을 만들어 준다. 사용자는 이를 '보석글'을 통해 불러 볼 수 있게 된다.

V. 전 망

여기에 서술한 것은 특별한 프로그램 언어를 통하지 않고, '보석글'이나 'dBASE'의 내장 명령만을 가지고 활용하는 방법에 대한 것이기 때문에 상당히 번거로운 것처럼 보일지도 모른다. 이 방법이 상당히 불편하다고 느끼는 사람은 프로그램 언어를 배워서 나름대로 처리하면 된다. 가령 베이직(BASIC)만 해도 랜덤 화일과 시퀀셜 화일을 다룰 수 있는 명령어가 예약되어 있으므로 얼마든지 위와 같은 작업을 할 수 있다.

그러나 스스로 프로그래밍에 자신이 없다고 느끼는 사람이나 새로운 언어를 배우는 데 시간을 바치기가 아까운 사람은 자신이 기왕에 사용하고 있는 '보석글'이나 'dBASE'와 같은 문서 작성 프로그램 및 자료처리 프로그램의 기능들을 100 퍼센트 활용하기 바란다. 그것만 가지고도 웬만한 작업은 다 해낼 수 있기 때문이다. □