

東洋學資料 電算化의 構想과 課題

金 炫

목 차

- | | |
|------------------------|------------------------|
| I. 왜 전산화인가? | 확장의 문제 |
| II. 東洋學 데이터 베이스의 구축 방안 | IV. 컴퓨터를 이용한 古典 자료의 정리 |
| III. 漢籍 자료의 전산화를 위한 코드 | V. 정보화 사회에 있어서의 東洋學 |

I. 왜 전산화인가

산업분야의 종사자들에게 있어 생산성의 향상은 곧 그들이 추구하는 모든 사업의 성패를 좌우하는 첩경으로 간주된다. 같은 노력을 들이고서도 더욱 나은 상품을 만들어 낼 수 있게 하는 것은 바로 기술의 발전에서 얻어지는 소중한 결과이며, 그러한 기술을 앞서 터득한 자만이 치열한 경쟁의 와중에서 우위를 점하게 되는 것이 제반 산업분야의 속성인 만큼, 이 생산성 향상이라는 문제는 그들의 일차적인 관심거리가 될 수 밖에 없는 것이다.

물론, 연구자들에 의한 학문 연구를 산업분야의 기술 문제와 직접적으로 연관지을 수는 없는 일이다. 인문학, 그중에서도 東洋學은 엄청난 양의 문헌자료들과의 고통스러운 씨름이 이른바 '연구'라고 하는 것의 대부분을 차지한다. 옛 문헌 혹은 다른 연구자들의 논저를 한 자 한 자씩 검토하며 의미 맥락을 찾아가는 그러한 노력은 그 자체로 창의적인 연구의 중요한 일부분으로서 결코 생략될 수 없는 것임에 분명하다. 그

것이 아무리 짜증스럽고 귀찮은 일이라 할지라도 그것을 떠나 달리 東洋學의 연구를 모색할 길이 없는 이상, 문헌자료와의 싸움을 방기할 수는 없는 것이다. 그러나, 東洋學의 문헌연구에 있어 우리가 기울이고 있는 현재의 모든 노력이 과연 최선의 방법인가에 대해서는 再考의 여지가 없지 않다.

東洋學 연구를 과제로 삼고 있는 우리의 앞에는 아직도 정리되지 않은 史料가 산재해 있으며, 양적으로 적지 않은 현재까지의 연구성과들도 효율적으로 관리되지 못하고 있다. 또한 ‘東洋學’ 또는 ‘國學’이라고 테두리 지을 수 있는 여러 분과에서 심도 깊은 연구가 지속적으로 진행되고 있고, 그에 따라 가치 있는 연구 결과들이 끊임없이 배출되고 있으나 그 가운데 적지 않은 부분이 해당 분야에서 보다 발전된 연구의 토대가 되지 못한 채 休眠하고 있는 것은 과연 어디에서 그 이유를 찾아야 할까?

우리의 관심분야가, 美·日에 의해 주도되고 있는 과학 기술분야라면, 또는 독일·프랑스와 같은 외국의 문화라면 ‘문헌자료의 정리’라는 문제를 그렇게 심각하게 생각하지 않아도 좋을지 모른다. 이러한 문제에 있어서는 누군가, 우리보다 먼저 그 일을 해 줄 사람이 있을 가능성이 보다 높기 때문이다. 그러나, 東洋學이란 과거부터 우리의 것이었던 것, 오늘날 우리에게 의해서 계승되고 발전되어야 하는 것에 대한 연구이다. 이 분야에 관련된 사료의 정리와 연구 결과의 관리 또는 일차적으로 우리의 책임일 수밖에 없다.

물론, 東洋學 자료의 정리에 반드시 컴퓨터를 동원해야만 하는 것은 아니다. 우리가 컴퓨터 연구 그 자체를 목적으로 삼고 있지 않는 한, 컴퓨터는 어디까지나 도구에 불과하다. 도구란 일의 성격과 이용자의 기호에 따라 사용할 수도 사용하지 않을 수도 있는 것이다. 단, 우리 앞에 놓인 과제가 엄청나다든 사실과 그 과제를 해결해야 할 이 분야의 연구자가 저극히 한정된 수에 불과하다는 사실을 고려할 때, 우리는 주

어진 여건 내에서 최선의 결과를 얻을 수 있는 효율적인 방법을 강구하지 않을 수 없다. 東洋學 분야의 연구 생산성을 제고할 수 있는 방법의 하나로 필자가 구상해 온 東洋學 자료 전산화의 추진 방향을 이 글을 통해 밝히고자 한다.

II. 東洋學 데이터베이스의 구축 방안

컴퓨터를 통해 학문 연구의 효율을 증진시키는 방법은 여러 가지로 생각할 수 있겠지만, 그중에서도 학술정보 데이터베이스의 구축은 東洋學 뿐 아니라, 여타의 인문사회과학 분야에서도 그 필요성이 절실히 대두되고 있는, 가장 넓은 범위의 수요를 가지고 있는 과제이다.

데이터베이스란 서로 유관한 관계를 맺고 있는 자료들을 수집하여 그 가운데에서 필요한 정보를 선별하여 조회할 수 있도록 가공해 놓은 것을 말한다. 일반적으로 컴퓨터하면 의례 연상되는 기능의 하나가 자기가 알고 싶어하는 어떠한 사실에 대한 구체적인 정보를 그 즉시 제공해 준다는 것인데, 컴퓨터로 하여금 이러한 일을 가능케 해 주는 것이 바로 데이터베이스의 역할이라고 할 수 있다. 더군다나 오늘날에 있어 이러한 데이터베이스의 효용을 더욱 높여 주고 있는 것은 컴퓨터와 통신장비를 결합시킨 데이터 커뮤니케이션의 기술이다. 학술정보를 저장해 놓은 특정 컴퓨터가 있는 곳을 수고스럽게 찾아가지 않아도, 먼 곳에서 장거리 전화를 걸어 친척의 안부를 확인하듯 원거리 단말기를 조작함으로써 세계 어느 곳의 데이터베이스도 현재 우리가 있는 곳에서 이용할 수 있다는 이야기이다. 이와 같은 데이터 커뮤니케이션의 기술은 이미 우리 나라의 어느 곳에서도 구미의 학술 데이터베이스를 조회

할 수 있도록 해 주고 있다. 한국 데이터통신에서 중계하고 있는 미국 록히드사의 'Dialog' 데이터베이스는 그 한 예가 될 것이다.¹⁾

1) 다이알로그(Dialog) 데이터베이스 중 인문·사회과학 분야의 파일들

파일번호	데이터베이스명	제작기관명	주제분야	수록기간	축적정보건수	갱신주기 별수가건	갱신주기
38	AMERICA HISTORICAL AND LIFE	ABC-Clio, Information Services	미국·캐나다의 역사문학	1964~	194,000	11,500	3/Y
56	ARTBIBLIOGRAPHIES MODERN	"	미술전반	1974~	75,400	3,500	SA
64	CHILD ABUSE AND NEGLECT	Dept of Health, and Human Services	아동 학대사범	1965~	13,8000	—	"
39	HISTORICAL ABSTRACTS	ABC-Clio, Information Services	역사학	1973	178,300	—	Q
202	INFORMATION SCIENCE ABSTRACTS	IFI/Plenum Data Company	정보과학	1966~	74,000	—	BM
36	LANGUAGE AND LANGUAGE BEHAVIOR ABS	Sociological Abstracts, Inc.	언어·언어학용	1973~	66,600	1,500	Q
61	LISA(LIB AND INF. SCI. ABSTRACTS)	Library Association Publishing	도서관학·정보과학	1969~	64,000	500	M
249	MIDEAST FILE	Learned Information Ltd.	중동지방정보	1979~	30,900	800	"
71	MLA BIBLIOGRAPHY	Modern Language Association	현대어·언어학, 문학	1970~	538,500	—	A
49	PAIS INTERNATIONAL	PAIS, Inc.	사회·공공문제	1976~	221,500	25,000	M
57	PHILOSOPHER'S INDEX	Philosophy Documentation Center	철학	1940~	109,500	1,000	Q
91	POPULATION BIBLIOGRAPHY	University of North Carolina	인구문제	1966~	66,700	1,000	BM
11	PSYCINFO	American Psychological Association	심리학	1967~	456,500	2,700	M
190	RELIGION INDEX	American Theological Library Association	종교, 인문, 사회과학	1975~	—	—	BA
97	RILM ABSTRACTS	City University of New York	음악학·음악이론	1971~	47,700	—	
7	SOCIAL SCISEARCH	Institute for Scientific Information	사회과학전반	1972~	1,364,000	10,000	M
37	SOCIOLOGICAL ABSTRACTS	Sociological Abstracts, Inc.	사회학	1963~	143,000	1,700	S/Y
93	U. S. POLITICAL SCIENCE DOCUMENTS	University of Pittsburgh	정치·행정	1975~	24,500	900	Q

「Dialog 분야별 일람표」 p.5, 한국데이터통신주식회사

그러나, 외국의 학술정보를 이렇듯 손쉽게 얻을 수 있는 길이 현실화 되어 있는 데 반해, 우리의 문화, 우리의 학문에 대한 정보를 효율적으로 관리할 수 있는 장치는 아직 미비한 상태에 있다. 미국의 어느 학회에서 발표된 최근의 논문에 대한 정보는 위성통신의 중계에 의해 우리의 컴퓨터 단말기 화면에 언제든지 모습을 드러낼 수 있는 상황인데 반해 수년 전에 발표된 국내의 저명한 학자의 연구 논문 한 편을 찾기 위해서는 고속버스를 타고 서울에 올라와 장서 수십만 권 이상의 대학 도서관을 찾아야만 하는 것이 우리의 슬직한 현실이다. 우리에게 컴퓨터가 없는 것이 아니며, 데이터베이스를 구축할 기술이 축적되어 있지 않은 것도 아니다. 은행의 온라인 예금이 얼마만큼 편리한 것인지를 경험해보지 못한 사람은 거의 없다. 기차 간의 좌석이 텅텅 비어 있는데도 입석표밖에는 구할 수 없는 우스꽝스런 일도 이제는 거의 일어나지 않는다. 그런데도 우리의 학문 연구분야에서만은 그와 같은 편의를 입지 못하는 이유는 일차적으로 우리의 무관심에 있다고밖에 할 수 없다.

아무리 뛰어난 기능을 가진 컴퓨터라고 해도 그것은 어디까지나 기계적인 도구에 불과하며, 경이로운 검색 능력을 가진 데이터베이스 관리 프로그램을 개발했다고 해도 거기에 우리가 필요로 한 데이터들이 입력되어 있지 않은 이상 그것은 아무것도 쓰여 있지 않은 백지를 호화로운 장정으로 제본한, 의미없는 책자에 불과하다. 우리의 학문 연구에 직접적인 도움을 줄 수 있는 소중한 자료들을 수집 분류 정리해 나아가는 우리 자신의 노력이 있음으로써만 컴퓨터를 이용한 획기적인 연구 생산성의 향상이 이룩될 수 있는 것이다.

우리 나라의 東洋學 연구자들이 이러한 문제에 그렇게 깊은 관심을 기울이지 못하는 이유는 컴퓨터 엔지니어링은 전적으로 이공분야의 일로서 우리와는 무관한 것이라는 의식이 고정관념으로 자리잡고 있기 때문이기도 하겠지만, 또 하나의 이유는 컴퓨터가 제공하는 데이터베이스

의 효능에 대한 충분한 이해가 없기 때문이기도 하다. 현재 우리 나라에서도 비약적으로 발전하고 있는 컴퓨터 엔지니어링의 기술과 東洋學 분야의 전문 연구자들의 학적인 노력이 결집되어 國學 분야 제분과의 연구결과를 종합적으로 수록한 國學 관계 자료 데이터베이스가 구축되었다고 가정할 경우, 우리가 거기에서 얻을 수 있는 이점은 과연 어떠한 것들일까?

가장 먼저 생각할 수 있는 것은 정보 수집의 용이성이다. 자신의 연구하고자 하는 어떤 문헌 자료의 내용을 검색하는 데 있어서, 또는 자신이 관심을 가지고 있는 분야에 대한 선배 학자들의 연구 성과들을 종합하는 과정에서 우리가 현재 소모하고 있는 시간과 노력을 현저하게 절감시켜 줄 수 있는 한 가지 방법이 바로 학술 데이터 베이스의 구축과 이용이다. 다이알로그 데이터베이스의 경우 거기에 수록된 5천 5백 만건의 각종 학술자료는 분류주제, 제목, 저자명, 내용 앵스트랙트, 게재 학술지명 등 어느 항목의 어느 단어로도 검색이 가능하다.²⁾ 그와 같은 방식으로 우리의 학술정보를 관리해 줄 수 있는 東洋學 데이터베이스가 만들어질 경우, 거기에서 얻어지는 연구 능력의 향상은 분명 하찮은 것으로 간주될 수만은 없을 것이다.

두 번째로 생각할 수 있는 것은 學際間 연구의 가능성을 증진시킬 수 있다는 점이다. 오늘날 우리나라의 東洋學 연구는 철학, 사학, 사회학, 등 분과별 연구에만 머물고 있는 것이 아니라, 각 분과의 연구성과들을

2) 다이알로그(Dialog) 데이터베이스의 자료 수록 예 및 검색 항목

```

SAMPLE RECORD
Title Of Accession Number
-----> 052017
HEN TO APPLY THE CATEGORICAL IMPERATIVE. -----> /TI
AU: -----> POTTER JR., NELSON
JN: -----> PHILOSOPHIA (ISRAEL), 5, 306-416, 0 75.
LA: -----> Languages: ENGLISH
DT: -----> Doc Type: JOURNAL ARTICLE
JA: -----> Journal Announcement: 105
A PROCEDURE FOR SUCCESSFULLY (I.E., WITHOUT THE USE OF SUBSIDIARY MORAL
PRELUSIONS) APPLYING THE "ON PERSONAL LAW" FORMULATION OF KANT'S CATEGORICAL
IMPERATIVE IS SET FORTH, USING AS A MODEL APPLICATION KANT'S ARGUMENT THAT IT IS
WRONG TO MAKE A LYING PROMISE. ALSO INCLUDED IS AN ACCOUNT OF "MAXIMS" AND OF
THE "CONTRADICTION" RESULTING WHEN WE ATTEMPT TO UNIVERSALIZE AN IMMORAL MAXIM.
EVALUATION (A CAUSAL CRITERION OF ACTION-DESCRIPTION IS PROPOSED).
-----> /AB
Descriptors: FINIS: CATEGORICAL IMPERATIVE; END; DESCRIPTION: MAXIM
Named People: KANT -----> /DE
-----> /JA
  
```

종합함으로써 우리 문화의 전체적인 윤곽을 파악하고자 하는 이른바 學際間 연구라고 하는 종합적 연구 방법에 대한 관심과 수요가 새롭게 일고 있다. 이러한 연구 방법을 실현시키기 위해서는 무엇보다도 먼저 분과별 연구 성과에 대한 자유롭고 편리한 정보교환이 가능해야 한다. 엄청난 양의 학술정보들이 한꺼번에 쏟아지는 오늘날의 학문세계에서 어느 한 분야를 주전공으로 삼고 있는 사람에게 다른 분과에서 나온 업적들을 빠짐없이 살펴보도록 요구하는 것은 좀 무리한 일이다. 하지만, 그 여러 분과에서 쏟아지는 논저의 목록이나 앵스트랙트 정도만이라도 용이하게 볼 수 있는 장치가 마련되어 자신이 필요로 하는 자료를 취택할 수 있다면, 연구자들은 그로부터 상당한 도움을 얻을 수 있을 것이다.

세 번째로는 연구 결과의 公示性 증진에 기인한 학문 연구의 질적 수준향상을 들 수 있다. 긍정적이든 부정적이든, 오늘날 대학 사회에 몸담고 있는 연구자들은 적어도 1년에 한 편 정도의 논문을 써 내야 한다는 숙제를 받고 있으며, 이들의 논문을 활자화하기 위한 수많은 종류의 학술지들이 전국의 곳곳에서 쏟아져 나오고 있다. 그리고, 그러한 학술지들의 상당 부분은 (그것이 전국 규모의 조직을 갖고 있는 유명 학회이거나 간행 연수가 오래인 명문대학의 학회지가 아닌 이상) 그러한 학술지가 있는지, 거기에 어떠한 논문들이 실려 있는지조차 알려지지 않은 채 묵은 자료가 돼 버리기도 한다. 대학 교수들이 그들의 전공 분야를 막론하고 1년에 한 편 이상의 논문을 발표해야 한다는 것이 과연 바람직한 제도인지의 여부는 차치하더라도, 그 많은 양의 논문들이 쓰여져 나와 결국은 필자 주위의 몇몇 사람들에게만 읽히고 마는 경우도 적지 않다는 사실에 대해서 한 번쯤 생각해 볼 필요가 있다. 다수의 사람들에게 읽혀지고 평가받을 논문을 쓸 때의 자세와 아무도 읽어 볼 사람이 없을 것이라는 확신(?) 속에서 숙제 때우기를 하는 자세는 결

코 동일할 수 없다. 연구과정에서는 분명 적지 않은 시간과 노력을 투여하고서도 정의 없는 끝맺음으로 연구의 가치를 감소시키는 비능률을 방지할 수 있는 방법은 찾아질 수 없을까?

우리가 학술 데이터베이스의 구축을 생각한다면, 전국 규모의 학술지에 게재된 논문이건 어느 지방대학의 학과 학회지에 실린 글이건, 그것은 모두 학술 정보로서 동일한 가치를 갖게 될 것임을 예견할 수 있다. 인도양의 고도에 은거하는 어느 작가로 하여금 도시 사회에는 한 발자국도 들여 놓지 않은 가운데 세계적인 베스트셀러를 출판하게 만든 것이 바로 컴퓨터가 발휘한 데이터통신의 기능이였다. 어느 시골 구석에서 쓰여진 논문이라도 그것이 전국적인 통신망을 가지고 있는 데이터베이스의 자료로 등록되기만 하면, 그것은 그 순간부터 전국의 모든 연구자들의 공유 자료가 되는 것이며, 거기에 혹시 내제되어 있을지도 모를 학적인 불철저성은 다수에 의한 가차없는 비판의 대상이 될 것이다.

현재까지의 기초적인 연구 성과에 힘입어 東洋學 분야에서 우리가 연구해야 할 범위는 더욱 넓어졌으며, 그 자료 또한 더욱 방대해졌다. 산업분야의 상품 생산과 학계의 학문 연구를 동차원에 놓고 비교할 수는 없으나 산업 부문의 생산성이 급속하게 향상되어 가고 있는 현실에 비추어 볼 때, 인문학 분야, 특히 東洋學 분야의 연구 생산성은 학자 개인의 성실한 노력에도 불구하고 담보의 상태에 있다는 느낌을 떨쳐버릴 수 없다. 東洋學 관계의 학술 정보를 효율적으로 관리하고 전달할 수 있는 東洋學 학술 데이터 베이스의 구축은 東洋學이 미래의 학문으로까지 지속적인 생명력을 갖기 위해서는 필수적으로 해결되어야 할 과제이다.

그렇다면, 그 과제는 언제, 누구에 의해서 추진되어질 것인가? 東洋學 관련 제분과의 학술 자료를 광범위하게 수집해서 정보화한 대형 데

이타베이스의 구축은 현실적으로 (현시점에서) 가능한 일인가?

여기에는 몇 가지 중대한 장애요인이 있음을 솔직히 인정하지 않을 수 없다. 그 첫번째는 시스템 구축의 경제적인 문제이다. 다수의 이용자들에게 일정 수준 이상의 공통된 편의를 제공할 수 있는 데이터베이스가 구축되기까지는 그 엄청난 양의 자료 수집과 입력에 수반되는 적지 않은 노력이 투여되어야 한다는 전제가 있다. 마치, 누구나 편리하게 사용할 수 있는 한 권의 사전이 나오기 위해서는 그 속의 모든 단어들을 수집하고 그것의 의미를 정의하는 편찬자들이 엄청난 노력이 전제되어야만 하듯이... 그리고, 그 같은 전제 조건은 이 사업에 관심이 있는 소수의 회생적 노력에 의해서 해결될 문제만도 아니다. 조직과 설비 그리고 인력이 뒷받침되어야 하고, 그것들을 유지할 장기적인 재원이 있어야 하는 것이다. 구미의 경우와 같이 정보의 판매가 가장 유망한 방법이기도 하나 '정보도 일종의 상품'이라는 의식이 아직 일반화되어 있지 않은 우리 사회에서 정보 판매에 가능성을 두고 이 사업을 시작하기는 너무도 큰 경제적 위험성이 수반된다고 할 수 있다. 그밖에 한글, 漢字 정보처리의 불완전성 및 코드 표준화의 문제 등의 기술적 난제들도 없지 않은 것은 아니나, 더욱 어려운 것은 東洋學의 학문 지식을 컴퓨터 정보로 가공할 수 있는 전문가가 우리 학계에 거의 전무하다는 사실과 이 사업에 대한 학계의 일반적인 무관심이라고 지적해야 할 것이다.

이상과 같은 장애 요인으로 인해 東洋學 자료 데이터베이스를 당장 엮어 내는 것은 매우 힘든 일임에 분명하다. 그렇지만 중요한 사실은 위와 같은 장애 요인이 있다고 해서 막연히 팔짱끼고 기다리고만 있는 한, 東洋學 데이터베이스의 구축은 미래의 어느 시점에서든 이루어지지 못하리라는 점이다. 東洋學 자료의 전산화는 오직 이 문제에 관심을 갖고 있는 東洋學 연구자들에 의해서만 추진될 수 있다. 컴퓨터 엔지니어니

들은 단지 그들이 갖고 있는 기술 수단을 제공해 줄 수 있을 뿐이다. 東洋學 정보시스템의 구축이 우리의 東洋學 연구와는 영원히 무관한 의미없는 일이라면 모를까, 그것이 언젠가 우리의 손으로 구축되어야 할 필연적 과제라면, 현시점에서 가능한 작은 일부터 단계적으로 추진해 나아가는 차선의 방책을 취함으로써 미래의 東洋學 정보 시스템 구축에 대비해 나아가는 것이 현명한 태도일 것이다.

필자는, 점진적인 東洋學 자료 전산화 추진의 일례를 高麗大學校 民族文化研究所의 國學 자료 전산화 사업에서 찾을 수 있다고 생각한다.

민족문화연구소에서는 소내의 행정적인 업무를 자동화하기 위한 목적에서 마이크로 컴퓨터 시스템을 도입했으나, 학술 부문의 컴퓨터 응용에 관심을 가져온 몇몇 연구진들은 이 시설을 행정사무뿐 아니라 國學 분야의 자료 정리에 활용하여 그 결과를 民研과 관계를 맺고 있는 국내외의 연구진들에게 서비스할 계획을 수립하고 소프트웨어 개발 및 연구원 교육, 자료 수집 등의 사업에 착수하였다.

민족문화연구소의 전산시스템과 전산화 사업 규모는 기업이나 과학기술 분야 연구소의 전산 시스템 규모에 비교할 때 매우 보잘것없는 것임에 분명하다. 그러나, 불모의 상태와 다름없는 현시점의 東洋學 전산화 분야에서 발전적인 첫걸음을 내디딘 것이라는 점에서 民研의 전산시스템에 큰 의의를 부여하지 않을 수 없다.

民研의 학술 자료 전산화 사업이 시사해 주는 바를 볼 때, 東洋學 자료 데이터베이스 구축을 위한 점진적인 과정의 1단계는 연구소, 학회, 대학의 학과 등 東洋學 관련연구 조직에서 (비교적 경제적 부담이 적은) 마이크로 컴퓨터를 구입하는 것이라고 할 수 있다. 연구자료의 정리를 위해 컴퓨터를 구입하는 것은 연구 기능 활성화를 위해 충분히 명분이라는 일이나, 그 사실만으로는 컴퓨터에 대한 이해가 부족한 톱 매니저(혹은 다른 동료 연구자)의 동의를 얻기 어려울지도 모른다. 이때, 이

들을 설득할 수 있는 또 하나의 명분은 사무자동화의 도구로서의 컴퓨터의 기능이다.

요즈음은 인문학계에서도 공동연구 내지 연구 인력의 조직화에 대한 필요성이 증대해 감에 따라 개인별 연구의 심화 못지않게 연구 조직(학회 또는 연구소)의 기능이 강화되어 가고 있다. 조직을 유지하기 위해서는 곧 사무가 있기 마련이고, 조직이 커가면 커갈수록 이 사무의 양도 급속도로 팽창하게 된다. 퍼스널 컴퓨터와 같은 작은 시스템도 이 사무자동화의 측면에서 획기적인 기여를 한다는 것은 이제 누구에게나 상식화된 이야기이다.

제반 문서의 워드 프로세싱, 회원들의 주소록 관리 및 뉴스레터 등의 인쇄물 발송관리를 위해 도입된 연구소나 학회의 퍼스널 컴퓨터는 그 학회의 연구 분야에 관련된 문헌정보를 관리해 주는 소규모의 전문 데이터베이스를 구축할 수 있는 기능을 함께 보지하고 있다. 그 조직에 관계하고 있는 다수의 연구자들로부터 손쉽게 얻어지는 각종 문헌 정보들이 자유로운 데이터 편집과 검색 기능을 갖춘 컴퓨터 시스템 속에 종합적으로 정리될 경우, 그 각각의 정보들은 개별 정보로서 존재할 때보다 몇 배의 높은 가치를 지니게 될 것이다. 이 데이터베이스에 수록된 정보를 다수의 관련 연구자들에게 서비스하는 방법은 1차적으로 인쇄물에 의한 공급일 것이다. 즉, 화일에 수록된 내용을 몇 가지 항목별로 소트(순차정렬)하여 프린터로 출력한 후, 그 복사물을 배부하거나 학회지에 게재하는 방법인데, 사용자 임으로 정보를 검색할 수 있는 툴(tool)을 제공하지는 못하나 지속적인 데이터의 관리(추가 및 수정)가 용이하다는 점에서 기존의 수작업에 의한 정보 서비스보다 유리한 점을 가지고 있다. 그 후 과학기술처나 체신부의 구상대로 국민보급형 퍼스널 컴퓨터가 일반화되면 학회나 연구소에서 수집한 학술정보는 디스켓의 형태로 보급될 수도 있을 것이다. 이상과 같은 점진적인 東洋學 정보

시스템 구축의 제 2 단계 사업에는 다음과 같은 2가지 주요 기능이 내포되어 있다. 그 첫번째는 정보 관리가 연구생산성의 향상에 큰 보탬이 된다는 의식의 계몽이며, 또 한 가지는 장차 구축되어질 종합적인 데이터베이스의 구성 요소가 될 전문 분야별 화일의 구축이 이 과정에서 단계적으로 이루어지게 된다는 점이다.

이렇게 해서 '정보'로서 가공된 학술 데이터의 보급이 그 초보적인 면모를 갖추게 되면 보다 편리한 정보 송수신 방법(데이터 통신망의 이용)에 대한 수요가 있게 될 것이며, 또한 이 시점에서는 데이터 통신망의 이용이 사회 각 분야에서 활발하게 실용화될 전망이다. 이미 체신부에서는 '87년 2월부터 일반 전화선에 컴퓨터를 연결하여 데이터 통신에 사용하는 것을 허용했으며, 전국의 우체국을 정보센터화 하기 위한 전초 작업으로서 '87년 12월까지 국내 4개 지역(여의도, 영동, 충무, 고흥)에 컴퓨터 우체국을 시험 가동할 예정이다.³⁾

이상과 같은 환경이 조성되면 제 3 단계 사업인 東洋學 정보 센터의 설립이 가능해진다. 東洋學 정보 센터의 주요 기능은 첫째, 다량의 학술정보를 효율적으로 관리할 수 있는 데이터 베이스 관리 시스템을 구축하는 일, 둘째, 각 연구소에서 수집한 자료의 종합 정리 및 전문분야별 연구기관에서 취급되지 않은 연구 자료를 수집, 정리하는 일, 셋째, 외국의 연구기관과의 東洋學 관련 정보 교류 및 외국의 학술정보 수집, 넷째, 데이터 분배업자와 계약을 맺어 일반 연구자들에게 정보를 서비스하는 일 등으로 요약할 수 있다.

물론, 東洋學 정보 센터의 설립과 같은 규모가 큰 사업은 東洋學 분야 연구자들의 공동의 노력 뿐 아니라, 정부 차원에서의 지원까지도 필

3) 이밖에도 과학기술처에서는 1990년대까지 1가구 1 단말기 보급 계획을 추진하고 있으며, 1997년~2001년에는 생활 정보 서비스 분야의 국내 시장 규모가 2조 5천억 원에 이를 것으로 전망하고 있다.

(「2000년대를 향한 과학기술발전 장기 계획」, 1986. 12. 과학기술처)

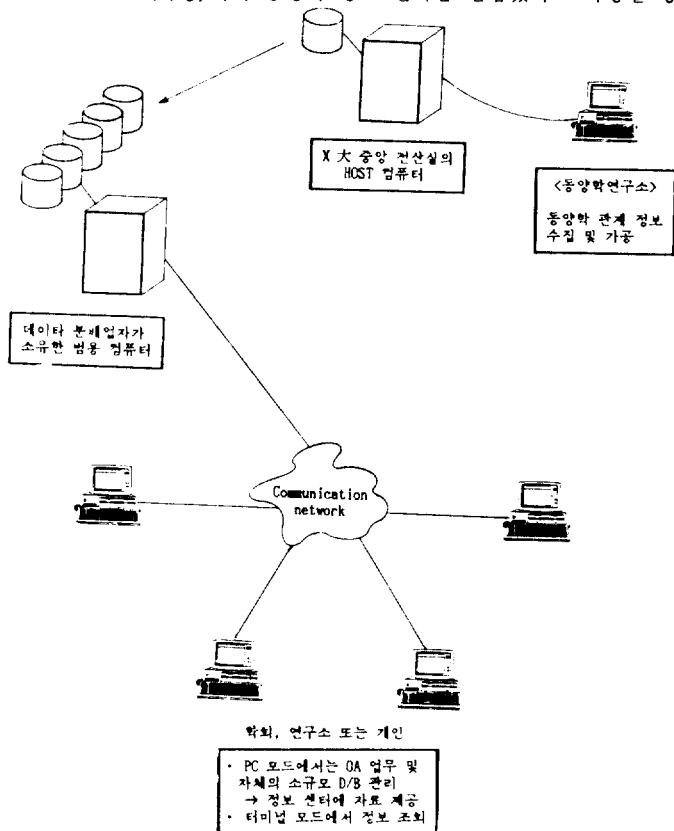
요한 일이므로 현시점에서 그것의 조속한 실현을 장담할 수는 없는 일이다. 그러나 현재의 우리들의 관심과 노력으로 할 수 있고, 또 해야 하는 당면 과제는 앞에서 언급한 1 단계 및 2 단계의 사업을 점진적으로 추진해 나아가는 것이며, 아울러, 미래의 종합적인 東洋學 정보 센터의 구축에 대비해 장기적인 생명력을 갖는 '정보 가공 방식'의 표준안을 마련하는 것이다. 학술정보 데이터베이스 구축을 위한 정보 가공 방식의 표준안에 대해서는 자료 분류방식의 표준안 제정⁴⁾, 자료 입력 양식의 표준안 제정⁵⁾ 및 표준화된 한글, 漢字 코드 사용⁶⁾ 등을 생각할 수 있다.

Ⅲ. 漢籍資料의 전산화를 위한 코드 확장의 문제

앞 장에서 언급한 東洋學 자료 데이터베이스에서 궁극적으로 다루어 질 자료는 1 차 자료(原史料)와 이에 대한 연구 결과로서의 2 차 자료(연구자의 논문, 저술)를 모두 포괄해야 한다. 그러나 본고의 제 2 장에서는 주로 2 차 자료를 중심으로 한 데이터베이스 구축에 대해 설명하

- 4) 도서관의 도서 분류 기준(KDCP, DDC 등)과 같은 주제 분류 기준을 보다 전문적인 입장에서 학계 내의 공동 약속으로 정하여 자료 분류 및 검색의 편의를 도모하는 것. 이 경우, 앞으로는 비교 연구, 학제간 연구가 보다 활성화될 것이라는 점도 고려해야 한다.
- 5) 데이터베이스에 수록될 자료 카드의 작성 양식에 대한 약속을 의미한다. 기존의 자료에 대해서는 데이터베이스 제작자들이 초록을 뽑아야 하겠지만, 새로 출간되는 자료의 경우에는 저자 자신이 이 약속에 따라 전산 입력용 카드를 작성하여 첨부하는 것이 바람직하다. (석·박사 학위 논문에 외국어 초록을 첨부토록 하는 것과 유사한 규정이 마련될 수 있다면 좋을 것이다. 또한 이러한 작업은 훗날의 편의를 위해 자료의 전산 처리가 시행되고 있지 않은 현시점에서부터 시작할 필요가 있다.)
- 6) 표준화된 코드를 사용하여 데이터를 입력하였을 경우 언제 누구에 의해서 수집된 데이터이건 훗날 재입력의 절차를 거치지 않고도 정보 센터의 대형 데이터베이스에 등록될 수 있다. (코드 표준화에 대한 세부 사항은 3장에서 언급)

〈그림 2-1〉 동양학 정보 센터의 시스템 구성도(X大 부설 동양 학연구소(가칭)에서 동양학 정보 센터를 설립했다고 가정할 경우)



었다. 그것은, 1차 자료의 전산화를 위해서는 필수적으로 선행되어야 할 몇 가지 기술적 과제들이 아직 미해결의 상태로 남아있기 때문이다.

東洋學 분야의 학술 자료 전산화를 가로막고 있는 가장 큰 장애 요인은 현재까지 개발된 컴퓨터 시스템의 漢字 처리 능력의 한계이다. 5,000자에 못 미치는 상용 漢字만으로는 漢籍資料의 전산화가 용이하지 못하기 때문이다. 따라서, 현재까지 東洋學 분야에서 추진된 漢籍資料의 전

〈표 2-1〉 국내 데이터베이스 수요에 대한 예측

	87—91	92—96	97—01
인 구 (만 명)	4,416	4,689	4,939
G N P (조 원)	121	325	343
D/B 서비스 매출액 (조 원)	0.5	1.4	2.5
가 입 자 (만 명)	15	100	200
D/B 수 (개)	100	600	1,250
D/B 분 배 기 관	20	60	100

「2000년대를 향한 과학기술발전 장기 계획」(1986. 12, 과학기술처)

산처리하는 漢字 한 글자 한 글자에 대응하는 숫자 코드를 약정하여 그것으로 입출력 작업한 시행한 것이었다. 그러나, 오늘날 우리 나라 정보산업계의 기술 수준은 이 문제에 대한 보다 근본적인 해결책을 제시해 줄 만큼 성장했다고 할 수 있다. 漢字 데이터의 직접적인 입출력을 가능케 하는 학술용 오퍼레이팅 시스템의 개발을 도모할 수 있게 된 것이다. 이 작업의 구체적인 추진은 학술용 漢字에 대한 코드 영역 부여에서부터 출발해야 한다.

컴퓨터에 어떠한 데이터를 입력하고 그 입력 내용을 조회할 수 있기 위해서는 그 데이터를 구성하고 있는 모든 문자가 컴퓨터에 내장(롬 칩 혹은 폰트 화일 속에)되어 있어야 하고, 그 내장 방법 또한 일정한 논리적 규칙에 따라야 한다. 아울러, 그 문자 내장의 논리적 규칙이 컴퓨터마다 일정하다는 전제에서만 각 컴퓨터 간의 데이터 호환성이 보장되는 것이다. 컴퓨터 문자의 코드화 및 코드 표준화가 이루어져야 하는 이유가 바로 여기에 있고, 한글, 漢字의 컴퓨터 코드 표준안이 제정된 것도 이러한 이유에서이다.

1986년 12월 한국표준연구소에서 제정한 컴퓨터용 한글 漢字 코드 표

준 시안의 내용 중 2 바이트 코드의 자종 수를 살펴 보면 다음과 같다.

* 코드 체계 :

1. 기본 코드—2 바이트 완성형
2. 보조 코드—16 비트 조합형 및 7 비트 한글 자모

* 기본 코드의 구성 :

1. 특수 문자(기호, 로마문자, 그리스 문자, 페션 조작 등) 및 숫자……
1128자
2. 한글……2,350자
3. 한글 사용자 정의 영역……94자
4. 漢字……4,888자
5. 漢字 사용자 정의 영역……94자⁷⁾

이상에서 한국표준연구소의 컴퓨터용 코드 표준 시안에서는 4,888자에 달하는 漢字 코드 영역을 부여하고 있음을 알 수 있다. 이 4,888자는 1982년 한국과학기술원에서 제정한 표준안⁸⁾의 1,692자에 비하면 3배 가까이 증가된 것이며, 국내의 사무용 퍼스널 컴퓨터들에 채택된 漢字 폰트의 자수에 비해서도 크게 뒤떨어지는 것은 아니나, 東洋學 관련 漢籍문헌의 전산화를 생각하는 우리가 보기에는 크게 미흡한 자수임이 분명하다.

東洋學 자료, 좀 더 범위를 좁혀서 말한다면 國學 자료의 전산화를 위해서는 기본적으로 몇 자 정도의 漢字가 지원되어야 하는지에 대해서는 아직 구체적인 통계 조사가 이루어져 있지 못하다. 다만, 이러한 문제에 대해서 선구적인 작업을 시행한 계명대 사학과와 김기협 교수의 조선왕조실록 자료에 의한 샘플 조사에 의하면, 국내 漢籍 사료의 전산화를 위해서는 기본적으로 약 1만 자의 漢字가 필요하다고 한다.⁹⁾

7) 「정보교환용 부호 시안 및 해설」, 1986.12, 한국표준연구소

8) KSC 5619-1982 정보교환용 漢字 부호

9) 김기협, 「漢文資料의 컴퓨터 처리방법 연구」, 『中國學誌』 제 1집, (啓明大學校 中國學研究所), 1983.

이 문제는 어떠한 방식으로 해결되어야 할까?

오직 漢籍 자료의 전산화만을 위해 漢字 코드 영역을 새롭게 제정하는 작업은 비교적 용이한 일이라고 할 수 있다. 우리의 편의대로 코드 영역을 정의하고 그에 따라 폰트 화일을 만들어 내면 그만이다. 그러나 앞에서도 언급하였듯이 표준화되지 않은 코드의 사용은 데이터 호환성의 결여라는 심각한 문제를 야기시킨다. 특정 인물, 특정 연구소에서만 사용할 시스템이라면 몰라도 다수의 연구자들이 저마다의 단말기를 가지고 공동의 자료를 이용할 수 있는 시스템을 설계하기 위해서는 코드

<표 3-1> 표준 코드를 ASCII 코드의 8단위 부호계에 적용시켰을 경우의 코드 테이블

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P	'	p								
1			!	1	A	Q	a	q								
2			"	2	B	R	b	r								
3	문		#	3	C	S	c	s								
4	프		\$	4	D	T	d	t								
5	름		%	5	E	U	e	u								
6	관		&	6	F	V	f	v								
7	드		'	7	G	W	g	w								
8	요		(8	H	X	h	x								
9	요)	9	I	Y	i	y								
A	요		*	:	J	Z	j	z								
B			+	;	K	[k	[
C			,	<	L	\	l									
D			-	=	M]	m	}								
E			.	>	N	^	n	-								
F			/	?	O	_	o	DEL								

2 바이트 문자의
제1 바이트 및
제2 바이트 영역

의 표준화라는 문제를 의면할 수는 없는 것이다.

이번에 제정된 표준시안도 국내 컴퓨터계의 가장 심각한 문제인 한글 漢字 코드의 불일치 문제를 어느 정도 해소할 수 있을지 의문이지만, 현재로서는 이 시안만이 다수의 약속이라고 할 수 있는 유일한 '표준'인 만큼, 東洋學 자료 전산화를 위한 漢字 코드 확장도 기본적으로 이 표준안의 테두리를 크게 벗어나서는 안 된다는 전제를 세우는 것이 타당할 것이다.

漢字 코드의 확장 방법을 강구하기 위해 먼저 표준안에서 시행한 2 바이트 코드 제정 방법을 살펴 보기로 한다.

표 3-1에서 보이는 바와 같이 표준 코드 시안에서 제정한 2 바이트 코드는 그것의 제 1 바이트 영역 및 제 2 바이트 영역을 모두 AIH에서 FEH로 제한함으로써 8,836자만을 사용할 수 있도록 한 것이다. (94자 (AIH-FEH)×94자) 표준 코드가 이렇게 좁은 영역에 한정된 이유는 7 단위 부호계 및 8 단위 보호계에 공통적으로 적용될 수 있도록 하기 위해서였지만, 사무자동화 정도에서 쓰이는 현대문의 사용 정도만을 고려한다면 이와 같은 2 바이트 코드 영역 제한은 사용자의 측면에서도 매우 편리한 것이다. 한글 漢字 영역과 영문자 영역이 앞, 뒤 바이트에서 모두 구분되므로 데이터 입력시 구분문자를 삽입할 필요가 없을 뿐 아니라, 데이터의 내용을 체크하는 프로그램(비정형 데이터 검색 프로그램과 같은)을 개발하는 일도 매우 용이해진다. (8단위 부호계를 사용하는 경우)

그러나, 최소한 1만 자 이상의 2 바이트 코드 사용이 필수적인 우리의 입장에서는 이와 같은 2 바이트 코드 영역의 제한은 매우 심각한 장애 요인이 아닐수 없다. 漢字 코드의 확장을 위해서는 부득이 표준안의 기본 골격만을 준수하는 선에서 제한의 일부를 수정하는 수밖에 없다.

필자는 다음과 같은 방법으로 표준 漢字 코드의 2 바이트 코드 영역

을 8천여 자 정도 더 확장하는 것이 가능하다고 생각한다.

제 1 바이트 : 기존의 코드 체계에서와 마찬가지로 AIH-FEH 범위 내의 것만을 사용—데이터 입력시 구분 코드(SO/SI)의 사용 없이 1 바이트 문자와 2 바이트 문자를 혼용하여 쓸 수 있도록 ⇒ 이 경우 제 1 바이트 문자는 94자가 된다.

제 2 바이트 : 기존 코드 체계에서 사용하지 않았던 OOH-9FH의 코드 영역 중 ASCII 코드의 콘트롤 코드 영역(OOH-1FH, 7FH) 및 특수문자(! " # % \$ & ' () _ = + * 등)와 숫자(1234567890) 영역(20H-3FH)을 제외한 나머지 영역(40H-9FH, 7FH는 제외)을 추가 ⇒ 이 경우 제 2 바이트 문자는 기존의 94자 외에 95자가 추가될 수 있다.

추가된 문자 수 : $94 \times 95 = 8,930$ 자……도합 $8,838 + 8,930 = 17,768$ 자

이와 같은 방식으로 2 바이트 코드 영역을 확장할 경우, 표준 코드 안에 맞도록 개발된 시스템 프로그램을 수정하는 노력을 최소화할 수 있으며, 이 수정 프로그램으로는 확장된 코드의 데이터뿐 아니라 표준 코드만으로 작성된 데이터도 무리없이 수용할 수 있다는 장점이 있다. 단, 이러한 코드 확장 방식의 제한은 8 단위 부호계에만 적용될 수 있다는 점과, 13,813자 이상의 漢字를 사용할 수 없다는 점이다. ASCII 코드는 원래 7 단위 부호계로 개발되었으나 요즘에는 대부분 8 단위로 확장되어 사용되므로 굳이 7 단위 부호계의 적용을 고집할 필요는 없다고 생각되나, 漢字 사용자수가 1만 3천여 자로 제한되는 것은 또 하나의 장애 요인이 아닐 수 없다. 궁극적으로 우리가 접할 수 있는 모든 漢文 자료를 전산화하고자 한다면, 보다 항구적인 코드 확장 방법을 강구해야 할 것이나, 대체로 1만 자 정도의 漢字만 마련되면 國學자료의 99%는 전산화될 수 있다는 예상과 상업적으로 매우 제한된 수요부에는 없는 東洋學資料의 전산화 문제에 무한정의 노력을 투여할 수단은 없다는 현실적인 문제를 고려할 때, 漢文 자료 처리를 위한 코드 확장

의 노력은 일단 위와 같은 선에서 1차적인 목표를 두는 것이 바람직하리라고 생각한다.

이상에서 언급한 바와 같은 코드 확장의 문제는 이 분야의 전문가들의 광범위한 조력을 얻어 해결해 나아가야 하며, 또한 코드 영역의 확장에 뒤이어 제기되는 학술용 漢字 폰트의 개발, 이 폰트를 액세스할 수 있도록 시스템 프로그램을 수정하는 문제 등도 전문 기술진에게 그 개발을 의뢰해야 할 것이다. 그러나 보다 중요하고 시급히 해결해야 할 과제는 표준 漢字 이외에 도대체 어떠한 漢字들이 國學 자료의 전산 처리를 위해 필요한가 하는 물음에 대한 해답을 구하는 일이다. 이 과제는 오직 東洋學 연구자들의 전문 지식에 의해서만 해결될 수 있다. 古典 자료의 전산 처리를 위해 東洋學 연구자들은 우선 표준 漢字 이외에 古典 자료 처리를 위해 필요한 기본 漢字를 조사하여 수준별 폰트 화일¹⁰⁾을 만들 수 있는 근거 자료를 확보해야 할 것이다.

10) 수준별 폰트 화일 제작의 구상(案)

1 수준: 표준 漢字(4,888자)

2 수준: 國學 연구에 있어서 특히 비중이 높은 문헌에 수록된 漢字 중 표준 漢字로 지정되지 않은 것(四書와 같은 기본 經書, 『三國史記』, 『三國遺事』, 『高麗史』, 『經國大典』, 『文獻備考』, 『東國輿地勝覽』 등의 문헌), 현대어에 쓰이지 않는 한글 古字

3 수준: 기타 고전 문헌의 전산화 작업 가운데에서 발견되는 희귀 漢字들의 도입을 폰트 화일화하여 기존의 폰트 화일에 병합할 수 있도록 코드 영역만 부여

제 3 수준의 폰트 화일은 문헌에 따라 여러가지가 만들어질 수 있다.

- 문자 등록: 특정 문헌의 전산화 작업 과정에서 문자를 디자인하여 등록.
- 사용: 문헌 입력·검색시 해당 폰트를 병합 사용토록 한다.

EX) 『東醫寶鑑』 D/B를 검색할 경우

제 1 수준 FNT+제 2 수준 FNT+『東醫寶鑑』 검색용 FNT 사용

文科榜目 D/B를 검색할 경우

제 1 수준 FNT+제 2 수준 FNT+『文科榜目』 검색용 FNT 사용

※ 이러한 작업이 가능하기 위해서는 문자 등록 유틸리티 및 폰트 병합 유틸리티 등의 소프트웨어가 지원되어야 한다.

Ⅳ. 컴퓨터를 이용한 古典 자료의 정리

코드 확장 및 폰트 개발이 이루어져 漢籍 자료의 전산 처리를 위한 환경이 조성되면 그 다음의 과제는 漢文 문헌을 컴퓨터에 입력하여 정보화하는 것이다. 이 경우, 古典 자료를 어떻게 정보화(데이터베이스화)할 것이냐에 대한 해답은 정보화할 자료의 성격에 따라 달라진다. 十三經과 같은 經學 연구의 기본 텍스트는 그 속의 모든 내용에 대한 낱자 검색 및 단어 검색이 가능하도록 비정형 자료 검색 시스템을 구축¹¹⁾해야 할 것이며, 『文科榜目』, 『東國輿地勝覽』과 같이 정형화할 수 있는 자료는 항목별로 조건 검색이 가능하도록 정형 데이터 검색 시스템을 구축하는 것이 효율적이라고 생각한다. 단, 자료의 성격에 맞는 데이터베이스 관리 프로그램의 개발은 전문 프로그래머에게 의뢰해야 하나, 자료의 성격 규정, 입력 조회 항목의 선정 등은 어디까지나 東洋學 연구자들의 전문지식에 의해서만 가능한 일인 만큼, 東洋學 연구자들에게는 컴퓨터 시스템의 활용에 대비한 자료 분석의 훈련이 요청된다.

古典 자료를 컴퓨터에 입력하는 것은 해당자료의 데이터베이스를 구축하여 필요한 정보를 손쉽게 얻고자 하는 목적에서 이루어질 수도 있으나 현실적으로 보다 긴박한 과제는 고문헌 자료의 校閱 및 그에 따른 텍스트의 再刊行이라고도 할 수 있다. 우리 나라의 國學 연구자들이 사용하는 古典 텍스트는 대부분 영인본으로서, 수백년 전의 우리 조상들이 읽던 그대로의 모습을 지닌 자료이다. 동일한 모습의 텍스트를 통해 古人과 무언의 대화를 나누는 것은 나름대로의 심원한 의미를 지니는

11) 비정형 漢文 문헌 자료 검색시스템의 설계에 대해서는 <부록>을 참고하기 바람

일이지는 하나, 반면 이러한 연구방법의 일반화는 우리 학계에서 문헌 분석의 축적이 거의 이루어져 있지 못하다는 사실을 반증하고 있는 것이기도 하다. 다시 말해, 개개의 연구자들에 의해 이루어진 상세한 문헌 연구가 텍스트의 校閱로 이어지지 못하고 있는 실정이므로 어떤 문헌을 독해한 선학의 노력은 그 문헌을 다시 연구하고자 하는 후학의 노력을 크게 절감시키지 못하고 있는 것이다.

물론, 東洋學 연구자들이 고문헌 연구를 위해 수준 높은 漢文 독해력을 지녀야 함은 필수적인 요건이다. 그러나, 아무런 문장 부호와 주기가 없는 원문을 독파해 내기 위해서는 전문 연구가라 할지라도 적지 않은 시간과 노력이 필요하다. 더군다나 미래의 학문 발전을 위해서는 東洋學 분야의 연구자뿐 아니라, 타 분야(외국의 학문 및 현대사회과학 등)의 연구자들에게까지 우리의 古典자료가 개방되어야 할 필요가 있다 이를 위해서는 궁극적으로 주요 古典에 대한 ‘완전 번역’이 이루어져야 하겠지만 그 중간 과정으로서 고문 校閱과 고문헌의 再刊行도 중요한 의미를 갖는다.

그러나, 문장부호와 주기를 포함한 고문헌의 再刊行이 우리 학계에서 거의 이루어지고 있지 못하는 이유는 校閱, 조판상의 오류에 의한 改惡의 가능성이 높기 때문이다. 아무리 훌륭한 한학자들이 校閱을 하고, 숙련된 편집자들이 교정을 본다고 해도 단 한 번의 작업으로 완벽한 텍스트의 校閱이 이루어지기는 불가능하다. 이 문제에 대한 유일한 해결 방법은 오류가 발견될 때마다 끊임없는 수정 작업을 되풀이하는 것이지만, 한 번 활자화된 것을 수정하는 것도 결코 쉬운 일은 아니다. 원고의 수정 및 추가에 대해 무한한 융통성을 부여하는 컴퓨터 조판에 우리가 관심을 갖게 되는 이유는 바로 여기에 있다.

國學 研究者라면 누구나 한 질씩은 가지고 있는 『興猶堂全書』의 경우를 들어 보기로 하자. 茶山學 연구의 주 텍스트인 신조선사간 『興猶堂

『全書』는 지금으로부터 50여 년 전 정인보 선생을 비롯한 志士들이 민족운동의 일환으로써 다산 정약용의 필사본문집을 활자본으로 再刊行한 것으로서 오늘날 우리가 가지고 있는 어떠한 漢籍 문집보다도 체계적이고 정확한 내용을 담고 있는 것으로 정평이 있다. 그러나, 거기에도 수많은 탈자, 오자 및 편차의 임의 설정, 원고 내용의 임의 삭제 등의 심각한 문제가 내재되어 있으며, 그러한 이유에서 여유당 전서의 再刊行이 조속히 이루어져야 한다는 견해가 수년 전부터 연구자들 사이에서 적지 않게 운위되어 왔다. 하지만, 이와 같은 작업은 엄청난 재정과 인력의 동원이 요구된다는 점 이외에도 새로운 간행본이 기존 『興猶堂全書』의 문제점을 어느 정도의 수준까지 개선할 수 있겠느냐는 의구심에서 아무도 그 같은 작업의 실시를 주창하고 나서지 못하는 실정이다.

컴퓨터 조판에 의한 漢籍 문헌의 再刊行은 1회의 작업에 의한 완성을 요구하지 않는 데에 가장 큰 매력이다. 비록 최초의 입력과정에서 수많은 오류가 발생했다고 해도, 그 오류가 전문가들의 눈에 발견되는 한 그에 대한 수정은 언제나 용이하게 보장될 수가 있는 것이다. 더군다나 컴퓨터 조판은 텍스트의 校閱 작업 이외에 인명, 지명, 서명, 주제어 등의 주요 단어 인덱스까지도 자동 생성해 낼 수 있는 부수적 기능을 가지고 있다는 점도 고려해 둘 필요가 있다. 다음 세대의 보다 발전된 학문연구를 위한 기반을 조성한다는 의미에서 컴퓨터 시스템에 의한 古典 텍스트의 再刊行을 보다 주의깊게 강구해 볼 필요가 있을 것이다.

컴퓨터를 이용한 古典 자료의 정리 문제에 있어서 또 하나 언급해야 할 사항은 漢籍 자료의 효율적 입력 방법에 대한 연구의 필요성이다.

漢文 문헌을 비롯한 모든 종류의 데이터는 일단 컴퓨터에 입력되고 나면 정보 활용의 경이로운 능률 증진을 보장하지만, 그것이 컴퓨터 속으로 들어가는 과정은 인간의 노동력에 의존하는 매우 원시적인 방법이 그 주종을 이루고 있는 것이 현실이다. 키 보드를 두드려 자료를 입력

하는 방법은 손으로 글자를 쓰거나 활자를 조합하는 방법에 비하면 비교적 신속한 것이라고도 하겠지만, 『여유당전서』나 『조선왕조실록』과 같은 방대한 漢文 자료의 입력이라는 문제를 생각하면 보다 능률적인 입력 방법의 개발이라는 문제가 대두되지 않을 수 없는 것이다.

현재 우리에게 일반화되어 있는 漢字 입력 방법은 먼저 한글로 해당 漢字의 음을 입력하여 동일한 음을 가진 漢字들은 모두 호출한 후 그 가운데에서 자신이 쓰고자 하는 글자를 번호(또는 커서키)로 선택하는 방법이다. 여기에 더하여, 단어 단위로 漢字 단어를 호출하는 방법, 단어 연상에 의한 입력 방법 등도 개발되어 있으나 이러한 방법은 상용 漢字 및 단어를 이루고 있는 漢字에 대해서만 적용시킬 수 있으므로 순수한 漢文 데이터의 입력에는 그다지 효율적인 방법이 되지 못한다고 할 수 있다. 전문적인 漢籍 학술 자료의 입력에 적용시킬 수 있는 방법은 부수의 조합에 의한 입력 방법일 것이다. 국내에서는 아직 개발되어 있지 않으나, 우리보다 많은 漢字를 상용자로 쓰고 있는 중국에서는 발음(주음부호, 병음자모)에 의한 입력방법과 함께 부수에 의한 입력방법을 채택하고 있다.

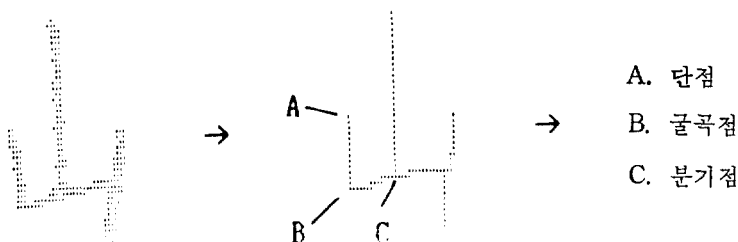
그외에 수 년 이내에 가장 효율적인 데이터 입력 방법으로 각광받을 것으로 예견되는 이미지 프로세싱에 의한 데이터 입력법을 생각할 수 있다. 이것은 복사기와 비슷한 구조의 이미지 스캐너로 문서 한 장을 촬영하면, 거기에 쓰여진 모든 문자 데이터들이 컴퓨터 코드로 변환되어 입력되는 방법인데, 타이핑된 영문자의 경우에는 이미 퍼스널 컴퓨터의 수준에서도 실용화되어 있고, 漢字의 경우는 1986년 IBM JAPAN에서 2,400자의 손으로 쓴 漢字를 식별할 수 있는 시스템을 개발했다고 발표한 바 있다.

문자 패턴 인식에 사용되는 대표적인 방법은 문자의 선을 細線化하여 특징점을 추출하고 그 특징점의 갯수 및 상대적인 거리 등에 따라 문자

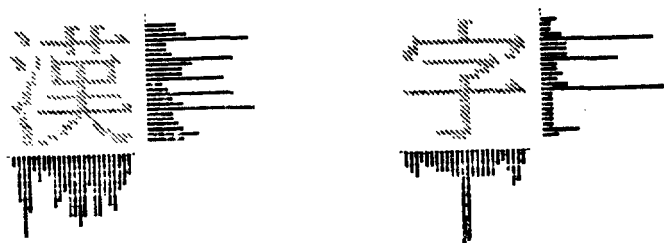
를 인식하는 방법¹²⁾, 글자를 구성하는 점의 집합이 나타나는 스펙트럼의 형상에 따라 인식하는 방법¹³⁾, 문자 윤곽의 凹凸에 따라 특정 문자를 식별하는 방법 등이 있다.

우리의 古典 자료 입력에 이와 같은 이미지 프로세싱에 의한 문자 패턴 인식법을 도입하는 것도 충분히 고려해 볼 가치가 있는 일이다. 단, 손으로 쓴 漢字의 자동 입력은 너무 부담스러운 과제이므로 활자화 된 문헌에 대한 처리를 우선적으로 고려하는 것이 바람직할 것이다. 그 구체적인 추진 방법은 첫째, 우리 나라의 古活字 중 많은 문헌의 인쇄에 사용된 활자로서 자형이 유사한 활자체를 선정하고¹⁴⁾ 그중에서도 사용 빈도가 높은 글자를 2,000자 내외의 규모로 선정한 후, 선정 문자의 패턴을 분석하여 개개의 문자를 식별해 낼 수 있는 룰(rule)을 추출하고,

- 12) 특징점의 종류: 단점(-), 굴목점(ㄣ), 분기점(⊥), 교차점(+) 등
(글자) (細線化) (특징점 추출)



- 13) 가로 세로 점의 집합에 의한 스펙트럼: 동일한 글자면 자체가 조금씩 다르더라도 대체로 유사한 형태의 스펙트럼이 발생된다.



- 14) 예컨대, 甲寅字體

그에 따라 문자 패턴 자동 인식 소프트웨어를 개발하는 것이다.

V. 정보화 사회에 있어서의 東洋學

지금까지 필자는 東洋學의 자료 전산화라는 주제와 관련하여 그것의 실현을 위한 세부 과제들을 몇 가지 항목에 걸쳐 논의하였다. 이장에서는 좀 더 포괄적인 주제로서 컴퓨터 시스템에 대한 의존도가 나날이 높아가는 오늘날의 정보화 사회에서 새롭게 모색되어야 할 東洋學 연구자들의 사회적인 기능에 대해서 언급하고자 한다.

요즈음의 언론 매체에서 빈번하게 이야기 하는 ‘정보화 사회’의 개념을 명확하게 규정짓기는 힘들다. 다만, 그에 대한 다양한 설명들 가운데 가장 포괄적이라고 할 만한 것을 추린다면, “물질, 에너지에 더하여 정보에 대한 의존도가 높아진 사회”, 또는 “정보의 생산 및 관리가 사회 기능의 중심이 되는 사회”라는 말을 들 수 있을 것이다.

여기에서 말하는 ‘정보’란 간단히 우리의 오관을 통해 두뇌 속으로 전달되는 ‘앎’을 의미한다고 할 수 있다. 단, 이 정보의 가장 중요한 특징은 개개인의 처지와 문제의식에 따라 그것의 가치가 각기 달리 부여된다는 점이다. 정보라는 말 자체는 정보화 사회 이전의 산업사회에서도 중요한 의미를 갖는 단어로 간주되었으며, 매스 미디어를 통한 정보의 전달이 사회의 중요한 기능으로 인식되어졌다. 그러나, 그와 같이 매스 미디어를 통해 광범위하게 전파되었던 산업사회의 정보와 정보화 사회라고 이름하는 오늘날, 또는 미래의 사회에서 컴퓨터 단말기를 통해 제공되는 정보의 사이에는 어느 정도의 의미상의 상이점이 발견되어진다. 전자의 경우 그것은 대중이 공통적으로 관심을 갖는 일반적인 정

보였다면, 후자의 정보는 고도로 전문화된 분업화 사회에서 각 전문 분야의 사람들이 필요로 하는 전문 정보를 의미하고 있다.

바꿔 말해, “다종 다양한 전문 정보에 대한 수요가 급증하는 사회”라고도 할 수 있는 ‘정보화 사회’의 인력 구조에서 나타나는 두드러진 특징은 정보처리 인력에 대한 수요가 급격하게 증가하고 있다는 점이다. 그리고, 여기에서 말하는 정보처리 인력이란 오늘날의 시스템 엔지니어나 프로그래머와는 다른 개념이라는 점에 주목할 필요가 있다. 현재까지는 컴퓨터를 다루는 것 자체가 고도의 전문 기술을 요하는 것이었으므로 컴퓨터에 대한 지식을 우선적으로 습득한 사람만이 다른 전문 분야의 정보를 처리할 수 있었지만, 컴퓨터로 처리해야 할 정보의 성격이 점차 고도의 전문성을 띠게 되는 반면, 누구든 자신의 목적에 맞는 방법으로 컴퓨터를 조작할 수 있도록 해주는 엔드 유저(end user) 중심의 패키지(package)가 광범위하게 개발되는 앞으로의 정보화 사회에서는 분야별 전문지식을 습득한 사람들이 정보 산업의 제 1 선에서 활동하게 될 것이라는 예측이다. 이처럼 특정 분야의 전문지식을 가지고 있으면서, 그 지식을 전산가독형 정보로 가공할 수 있는 사람을 지식 공학자(knowledge engineer)라고 부른다.

이 점에 있어 필자가 이야기하고자 하는 것은 우리 東洋學(보다 넓게는 인문과학) 분야에서도 정보화 사회를 리드해 나아갈 수 있는 정보처리 능력의 배양이 필요하다는 점이다. 우리의 정보, 다시 말해 우리의 전문지식은 우리들에 의해서만 가공될 수 있으며, 그와 같은 정보의 가공은 東洋學의 연구 수준 提高를 위한 필수적인 요건이 되기 때문이다.

교육정책을 주관하는 문교 당국에서는 정보화 사회에 대비한 광범위한 정보처리 인력 배양의 필요성을 인지하고 초·중·고등학교의 교과과정 개편을 추진하고 있다. 그러나, 정보화 사회에서 보다 중요시 되는

정보는 전문화된 정보인 만큼 전문 지식을 다루는 대학에서의 교육이 가장 중요한 의미를 갖는다.

정보화 사회를 리드해 나아갈 수 있는 자질은 ‘전산개론’을 수강하고 프로그래밍 언어를 배우는 데서보다 자신의 전문 분야와 관련하여 필요한 정보를 수집하고 그것을 목적에 맞게 가공하는 훈련을 쌓는 데서 성숙하게 된다. 이제는 우리의 대학 사회에서도 전문 지식의 교육과 아울러, 이와 같은 정보 가공 훈련의 기회를 학생들에게 제공해 줄 수 있어야 할 것이다. 이 점에 있어서는 대학의 경제적인 여건 등 장애 요인이 없는 것이 아니지만, 실현 가능한 일에서부터 방법을 찾아 본다면, 도서관 목록실에 터미널 몇 대를 설치해서 전 도서 자료의 전산화는 불가능하다 하더라도 전문 학술지의 목차만이라도 입력하여 학생들 스스로 조회해 볼 수 있게 하는 방법, 학과에서 퍼스널 컴퓨터를 구입하여 학생들로 하여금 전공 교과목의 참고 도서 리스트와 내용 앵스트랙트를 공동으로 작성케 하고, 차기 수강생들은 이를 이용하면서 그 내용을 계속 보완해 나아갈 수 있게 하는 것 등의 방법을 찾을 수 있을 것이다.

정보의 가공은 단지 컴퓨터에 의한 정보 처리만을 의미하지는 않는다. 그것이 갖는 본래의 목적은 우리에게 전달된 갖가지 지식을 진정 쓸모 있는 것으로 만드는 데 있으며 컴퓨터는 그것을 위한 하나의 도구에 불과하다. 다만, 컴퓨터 시스템에 의한 정보 가공의 훈련은 우리 사회의 거의 모든 분야에서 필요로 하는 효율적인 정보 관리의 능력을 배양시켜 준다는 점에서 장래의 학문 연구자들 뿐 아니라, 정보화 사회가 필요로 하는 모든 고급 인력들에게 필요한 일이라고 할 수 있다.

東洋學을 연구하는 우리의 과제는 일종의 자연 상태의 데이타라고도 할 수 있는 수 많은 古典 자료에서부터 우리의 과거를 알게 하고 그 결과 미래를 조망할 수 있게 하는 유용한 정보를 추출해 내고 그것을 오

늘날의 유의미한 언어로서 재조직하여 우리와 동일한 전통의 맥락에서 오늘을 살아가는 우리 사회의 동시대인들에게 파급시키는 일이다. 이러한 과제를 안고 있는 우리들에게 있어 '제 3의 물결'이라고도 일컫는 사회상의 변화(정보 사회화)는 낭만적인 기대감으로만 대할 수 있는 성질의 것이 아님을 분명하게 인식해야 한다.

산업사회화 과정에서 갑작스럽게 몰아닥친 외래문화의 홍수 속에서 잃어서는 안될 우리의 것을 너무나 많이 잃어 버린 우리의 입장에서 볼 때 미래의 정보사회화는 더 심각한 문제의 소지를 안고 있다. 정보 통신 기기의 광범위한 보급으로 인해 정보의 '수입 자유화'가 점차 가속화되어 가는데 반해 그러한 뉴 미디어를 우리의 고유한 문화 정보 창출에 이용하려는 노력을 기울이지 않을 경우, 비데오텍스의 단말기를 통해 우리의 안방까지 거침없이 파고 들어올 외래 정보의 홍수 속에서 우리는 또다시 浮遊하게 될지도 모른다. 東洋學 데이터 베이스를 구축하고 漢籍 자료를 전산화하는 일련의 작업은 東洋學의 연구 생산성을 향상시켜 이로 인해 절감되는 시간과 노력을 보다 창의적인 연구 활동에 투여할 수 있게 하기 위함이다. 그리고, 여기에서 말하는 '창의적인 연구'의 범주 가운데에는 생활 수준의 향상에 따라 우리의 전통문화에 관심을 기울일 여유를 얻게 된 일반인들에게 '가치 있는 우리의 것'을 전달해 줄 수 있는 문화 정보의 창출도 중요한 부분으로 포함되어야 한다.

<부 록>

非定型 漢文 文獻資料 검색 시스템의 화일 구조

非定型 漢文 文獻資料 검색 시스템의 구조를 알아 보기 위한 일례로 四書의 본문 전체에 대해 낱자 검색 및 단어 검색을 수행할 수 있는 데이터베이스의 화일 구조를 살펴 보기로 한다. 편의상 다음의 몇 구절이 四書의 본문 전체라고 가정한다.

1. 大學

제 1 장 大學之道，在明明德，在新民，在止於至善.

제 2 장 湯之盤銘曰：“苟日新，日日新，又日新.”

제 3 장 詩云：“邦畿千里，惟民所止.”

2. 中庸

제 1 장 天命之謂性，率性之謂道，修道之謂教.

제 2 장 子曰：“中庸其至矣乎！民鮮能久矣”

제 3 장 子曰：“道之不行也，我知之矣。知者過之，愚者不及也.”

3. 論語

學而-1 子曰：“學而時習之，不亦說乎！有朋自遠方來，不亦樂乎！”

學而-2 有子曰：“其爲人也孝弟而好犯上者，鮮矣.”

爲政 子曰：“爲政以德，譬如北辰居其所而衆星共之.”

4. 孟子

梁上 孟子見梁惠王，王曰：“叟不遠千里而來，亦將有以利吾國乎？”

孟子對曰：“王何必曰利，亦有仁義而已矣.”

梁下 莊暴見孟子曰：“暴見於王，王於暴以好樂，暴未有以對也.”

曰：“好樂何如？”

< data file 의 구성 >

經	章	節	LN	TEXT
1	1			善。止於新。至日新。之謂教。有者朋而而而。過自鮮來來已好。之遠矣星。矣。暴者不及亦也。乎！ 之。邦性。講。中道。爲。惡。子。何。如？ 明。荀。里。其。不。時。爲。以。王。何。如？ 德。新。惟。請。我。不。弟。如。利。於。王。何。如？ 在。日。所。民。道。久。上。其。千。仁。語。 新。日。止。能。矣。乎。犯。居。遠。有。王。主。 在。日。所。民。道。久。上。其。千。仁。語。 民。新。道。久。上。其。千。仁。語。 在。日。所。民。道。久。上。其。千。仁。語。 新。日。止。能。矣。乎。犯。居。遠。有。王。主。 在。日。所。民。道。久。上。其。千。仁。語。 新。日。止。能。矣。乎。犯。居。遠。有。王。主。 在。日。所。民。道。久。上。其。千。仁。語。 新。日。止。能。矣。乎。犯。居。遠。有。王。主。 在。日。所。民。道。久。上。其。千。仁。語。 新。日。止。能。矣。乎。犯。居。遠。有。王。主。
1	2		1	
1	3		2	
2	1			
2	2			
2	3			
3	1			
3	2			
3	3			
4	1			
4	2			
4	3			

↓

< keyword file >

< posting file >

經	章	節	(LN)
大學之道	1	1	01
在明明德	1	1	02
在明明德	1	1	03
在明明德	1	1	04
在明明德	1	1	06
在明明德	1	1	07
在明明德	1	1	08
在明明德	1	1	09
在明明德	1	1	11
在明明德	1	1	12
在明明德	1	1	13
在明明德	1	1	15
在明明德	1	1	16
在明明德	1	1	17
在明明德	1	1	18
在明明德	1	1	19
學之云命曰子	4	2	21
學之云命曰子	4	2	22
學之云命曰子	4	2	23
學之云命曰子	4	2	24
學之云命曰子	4	2	25
學之云命曰子	4	2	26
學之云命曰子	4	2	29
學之云命曰子	4	2	32
學之云命曰子	4	2	33
學之云命曰子	4	2	34
學之云命曰子	4	2	35

→

經	POSTER POINT	全
居	001	01
居	002	03
居	003	01
居	004	01
居	005	01
居	006	01
居	007	01
居	008	01
居	009	01
居	010	03
居	011	01
居	012	01
弟	083	01
弟	084	01
弟	085	01
弟	086	10
弟	087	02
平	100	04
平	101	03
平	102	01

&

經	POSTER POINT	全
居	001	01
居	002	01
居	002	02
居	002	03
居	003	01
居	004	01
居	005	01
居	006	01
居	007	01
居	008	01
居	009	01
居	010	01
居	010	02
居	010	03
居	011	01
居	012	01
弟	083	01
弟	084	01
弟	085	01
弟	086	01
弟	086	02
弟	086	03
弟	086	04
弟	086	05
弟	086	06
弟	086	07
弟	086	08
弟	086	09
弟	086	10
弟	087	01
弟	087	02
平	100	01
平	100	02
平	100	03
平	100	04
平	101	01
平	101	02
平	101	03
平	102	01

위와 같이 데이터 화일 및 색인 화일(keyword file 및 posting file)이 구성되었으면 입력된 데이터의 모든 글자 및 단어에 대한 검색이 가능하다.

〈A. 낱자 검색〉

- 四書에 쓰인 “見” 자의 용례를 찾을 경우(사용자가 “search 見” 등의 명령을 입력하면 검색 시스템은 다음과 같은 일련의 작업을 수행한다.)

1. keyword file에서 “見”을 찾는다.

⇒ |見|002|03|…“見”字는 四書 전체에서 3번 쓰였다고 답하고, posting file의 002-01, 002-2, 002-03에서 “見”字의 위치가 지정됨을 확인한다.

2. posting file에서 002-01, 002-02, 002-03을 찾아 四書에서 “見”字가 쓰인 위치를 찾는다.

⇒ |002|01|4|1| |03|…孟子 梁惠王上的 3번째 글자
|002|02|4|2| |03| 孟子 梁惠王下的 3번째 글자
|002|03|4|2| |10| 孟子 梁惠王下的 10번째 글자

3. data file의 4-1-0과 4-2-0를 찾아 수록된 text의 내용을 보여준다.

|4|1| |…孟子見梁惠王，王曰：“叟不遠千里而來，亦將有以利吾國乎？”

孟子對曰：“王何必曰利，亦有仁義而已矣。”

|4|2| |…莊暴見孟子曰：“暴見於王，王語暴以好樂，暴未有以對也。”

曰：“好樂何如？”

〈B. 단어 검색〉

- 四書에서 ‘何如’의 용례를 찾을 경우(사용자가 ‘search 何如’ 등의 명령을 입력하면 검색 시스템은 다음과 같은 일련의 작업을 수행한다.)

1. 위의 낱자 검색의 경우와 동일한 과정을 거쳐 ‘何’ 字의 위치와 ‘如’ 字의 위치를 찾는다.

|何|096|02| ⇨ |096|01|4|1| |37|…孟子 梁惠王上 37번째 字
 |096|02|4|2| |34|…孟子 梁惠王下 34번째 字
 |如|053|02| ⇨ |053|01|3|2| |11|…論語 爲政 11번째 字
 |053|02|4|2| |35|…孟子 梁惠王下 35번째 字

2. 선택된 데이터 중 經, 章, 節이 같고, 글자 위치가 이웃해 있는 것을 찾는다.

⇨ |4|2| |34| & |4|2| |35|

3. data file 의 4-2-0를 찾아 수록된 text 의 내용을 보여준다.

|4|2| |…莊暴見孟子曰：“暴見於王，王語暴以好樂，暴未有以對也。”
 曰：“好樂何如？”

※ 四書 전체가 아닌 論語, 孟子 중에서, 혹은 孟子 梁惠王上篇 중에서 특정 단어를 검색하고자 할 때에는 검색 명령에 經, 章, 節의 범위를 제한하는 명령을 추가하면 된다. 그러나, 四書 전체에 대한 검색이든 특정 章, 節에 대한 검색이든 검색작업은 인덱스를 통해 수행되므로 회답 시간에 차이는 없다.