

# 우리나라 온라인 인물데이터베이스의 비교 평가 연구

## Evaluating Online Biographical Databases in Korea: A Comparative Study

장 혜 란(Chang, Hye Rhan)\*

### <목 차>

- |                  |            |
|------------------|------------|
| 1. 서론            | 3.1 최신성 분석 |
| 2. 연구내용과 방법      | 3.2 포괄성 분석 |
| 2.1 평가대상 데이터베이스  | 3.3 완전성 분석 |
| 2.2 평가요소의 설정     | 3.4 다양성 분석 |
| 3. 인물데이터베이스 비교분석 | 4. 결론 및 제언 |

### 초 록

온라인 상용 데이터베이스에 대한 이용자 불만족 요인을 기초로 하여, 우리나라 인물 데이터베이스를 분석 평가 하고 외국의 인물데이터베이스 현황과 비교하였다. 구체적으로 최신성, 포괄성, 완전성, 그리고 다양성을 평가하였으며 문제점과 발전방향을 도출하였다.

주제어 : 인물데이터베이스, 현황분석, 최신성, 포괄성, 완전성, 다양성, 발전방향

### Abstract

Current states of Korean biographical databases are evaluated based on the dissatisfaction factors identified by a user survey. currency, coverage, completeness, and variety of the Korean biographical databases are evaluated and compared with overseas biographical databases. The results of the analysis are given with problems and suggestions for the future improvements.

Key Words : biographical databases, comparative evaluation, currency, coverage, completeness, variety

\* 상명대학교 문헌정보학과 교수(chrhan@sangmyung.ac.kr).

· 접수일 : 2001. 11. 15 · 최초심사일 : 2001. 11. 30 · 최종심사일 : 2001. 12. 7

## 1. 서 론

컴퓨터와 통신기술의 발전으로 데이터베이스 생산과 유통이 크게 증가하였으며, 급속도로 확장되고 있는 인터넷 정보, 특히 WWW의 성장은 데이터베이스 탐색을 도서관 및 정보센터는 물론 일반 대중에게도 확산시켰다. 특히 우리 나라의 데이터베이스 산업은 정부의 정보화 정책에 힘입어 1990년대를 통하여 지속적으로 발전하였으며(연평균 42.2%, 1992-1998), 개인용 컴퓨터의 보급도 일반화되었고, PC통신과 인터넷을 이용한 정보의 탐색 또한 정착되어 가고 있다.

그러나 현재 서비스되고 있는 데이터베이스 중에는, 이용자 요구와 탐색에 대한 체계적인 연구가 선행되지 않고, 기존의 인쇄본을 그대로 혹은 일부 증보하여 기계가독형으로 변환시키거나 새로운 데이터베이스가 생산됨으로써, 데이터베이스의 장점을 살리기 어려운 것들이 상당한 부분을 차지하고 있다. 실제로 국내 데이터베이스 이용자를 대상으로 상용 온라인 데이터베이스의 만족도를 조사한 결과를 보면, 이용자들은 수록 정보의 내용과 검색 결과에 대한 높은 수준의 불만을 나타내고 있다. 데이터베이스 산업에 대한 사회 전반의 요구는 양적인 발전을 탈피하여 질적인 향상을 기대하고 있다.

유용한 데이터베이스를 구축하고 질적 향상을 도모하기 위하여는 평가 연구가 필수적이다. 선진 외국에서도 초창기에는 기존의 자료를 기계가독형으로 변환시킴으로써 데이터베이스를 구축하기 시작하였으나, 그 나라들은 일찍이 서지와 참고자료가 발전되어 있었다. 그리고 데이터베이스 산업의 성장과 더불어 유사 분야나 주제를 다루는 데이터베이스에 대한 비교 평가 연구가 실무자들을 중심으로 지속적으로 수행되고 있다. 우리 나라의 경우에는 유사한 분야나 주제에 대한 데이터베이스가 생산되고 있음에도 불구하고 체계적으로 분석한 비교연구는 극히 드문 편이다. 다만 인터넷 WWW 전자신문을 비교 분석한 연구(한상길, 1996)와, 웹 기반 데이터베이스의 평가기준을 설정하고 4종의 인물데이터베이스를 분석 평가한 최근의 연구가(박준식 박경호, 2001)있다.

본 연구는 이와같은 상황을 염두에 두고, 우리나라에서 생산되는 온라인 인물데이터베이스를 이용자 불만족 요인을 기초로 하여 분석 평가하고, 가능한 경우에 외국의 인물데이터베이스 현황과 비교 분석함으로써, 문제점을 식별하고 보다 나은 정보를 제공할 수 있는 방안을 모색하기 위한 것이다. 이와 같은 연구는 데이터베이스 제작자에게 품질을 개선할 수 있는 지침을 제공하고 이용자들의 만족도를 높일 수 있는 방안을 마련함으로써 데이터베이스 산업과 정보이용 활성화에 기여할 수 있을 것으로 보인다.

## 2. 연구내용과 방법

### 2.1 평가대상 데이터베이스

본 연구는 우리나라와 외국의 인물정보 데이터베이스를 대상으로 한다. 개인의 이력에 대한 정보를 포함하는 인물 데이터베이스는 학생들의 과제해결, 일상생활에서 생기는 지적 호기심의 충족, 그리고 특정 업무나 연구를 위한 사전 지식을 제공하기 위하여 사용되고 있다. 따라서 인물데이터베이스는 각종 도서관의 참고봉사에서 일관성 있게 사용도가 높은 정보원이다. 최근의 조사 결과에 의하면 우리나라 인물데이터베이스에 대한 온라인 이용자들의 요구가 높게 나타나 있다. 상용 온라인 데이터베이스 이용 조사의 소분류 분석 결과를 보면, 인물/기관에 대한 정보가 이용 사례 전체의(N=482) 약 25%로 상위 6위를 차지하고 있는 것으로 나타나 있다(한국데이터베이스 진흥센터, 2001, pp.112-113).

데이터베이스가 구축되고 활용됨에 따라 인물 정보원의 탐색 양상이 크게 변화되었다. 전통적으로 인물 정보원의 사용은 이름 등을 알고 있는 개인에 대한 탐색이 주류를 이루었으나(일종의 known item search), 데이터베이스를 통하여 알지 못하는 불특정 다수를 대상으로 레코드를 구성하는 필드를 조합 탐색하는 것이(일종의 unknown item search) 가능해졌다. 또한 공공기록이 데이터베이스화 되어 공개됨으로써, 저명인사가 아닌 일반인에 대한 탐색이 인물 탐색의 한 부분을 차지하게 되었다.

한국데이터베이스진흥센터에서 발행하는 “알기쉬운 한국의 데이터베이스 목록(1999)”을 보면, 총 124종의 데이터베이스가 인물/기관 정보로 분류되어 있다. 그 중에서 기관 정보 데이터베이스, 즉 기관의 상품과 서비스, 기관홍보, 기관소속인물 등을 수록하고 있는 데이터베이스를 제외하고, 전적으로 인물 정보만을 다루는 데이터베이스는 20종이다. 20종의 인물 데이터베이스 중에서 지역위주의 소규모 데이터베이스(예, 화순군청 인물 등), 서비스 중단된 데이터베이스, 문학작품과 더불어 문학인을 소개하는 데이터베이스 등을 제외하면 15종이 남는다. 여기에 포함이 안되어 있는 조선일보 인물정보, 연합뉴스 인물정보, 여성인명사전, KOLIS 법조인명록, 4종을 포함하여 총 19종이 분석대상 우리나라 인물데이터베이스이다. 19종의 데이터베이스는 모두 국내 PC통신과 인터넷을 통하여 접속 가능하다. 분석대상 데이터베이스명칭이 <부록 1>에 열거되어 있다.

비교 평가를 위하여 사용된 외국의 데이터베이스는 대표적인 권위있는 디렉토리인 “Gale Directory of Databases”(v.1, online databases, March 2000)에 수록되어 있는 인물데이터베이스로 설정하였다. 이 디렉토리에는 총 73개의 엔트리가 인물 분야로 분류되어 있다. 이 중에서 기관과 함께 혹은 문학작품과 함께 인물을 다룬 데이터베이스를 제외하고 전적으로 인

#### 4 한국도서관·정보학회지 (제 32권 제 4호)

물에 대한 사항만 수록하고 있는 데이터베이스는 총 46종이다. 분석대상 46종의 외국 인물데이터베이스 명이 <부록 2>에 열거되어 있다.

## 2.2 평가요소의 설정

정보시스템은 크게 축적된 정보와 정보에 대한 접근을 제공하기 위한 액세스메카니즘으로 나누어 볼 수 있다. 텍스트기반 데이터베이스의 탐색 기능은 어느 정도 표준화되어 있고, 인터페이스의 개선은 하드웨어나 소프트웨어의 발전과 밀접한 관계가 있다. 데이터베이스 품질을 결정하는 요인 중에서 일차적으로 고찰하여야 할 중요한 사항은 그 기초가 되는 내용이다.

최근에 하이텔 이용자 2,047명을 표집하여 수행된 국내 데이터베이스 이용 실태 조사에 의하면, 우리나라 상용 온라인 데이터베이스의 수록 정보와 검색 결과에 대한 불만사항이 나타나 있는데 그 순위와 정도는 다음과 같다(최명규, 1998).

- 1) 쓸만한 정보가 많지 않다(34.9%),
- 2) 최신정보가 부족하다(16.9%),
- 3) 정보의 수록 건수가 적다(15.8%),
- 4) 정보가 구체적이지 못하다(13.1%),
- 5) 정보가 다양하지 못하다(11.9%),
- 6) 정보가 부정확한 경우가 있다(7.4%).

본 연구에서는 위와 같은 조사결과에 나타난 이용자 불만 요인에 기초하여 평가요소를 설정하고자 한다.

첫번째 제시된 불만요인은 데이터베이스의 존재 여부와 수록정보 내용 전반의 질을 의미하는 것인데, 우리나라에는 이미 다수의 인물데이터베이스가 존재하며, 유용성 전반에 대하여는 다음에 기술하는 평가요소로 대신할 수 있다.

두번째 제시된 불만요인인 최신 정보의 부족은 온라인 데이터베이스의 장점인 신속성의 결핍을 지적하고 있다. 최신성은 정보가 신속하게 데이터베이스에 적재되는 정도를 말한다. 최신성을 평가하기 위한 방법은 데이터베이스의 유형에 따라 다르다. 서지나 전문 데이터베이스의 경우에는 원정보의 생산일자와 레코드 입력일자와의 시차를 측정요소로 삼지만, 디렉토리 데이터베이스의 경우에는 갱신빈도의 적절성을 측정요소로 간주할 수 있다. 우리나라 인물데이터베이스는 디렉토리 유형이 주류를 이루므로, 본 연구에서는 최신성 분석을 위하여

갱신빈도를 측정하기로 결정하였다.

세번째 제시된 불만요인, 정보의 수록건 수가 적다는 것은 포괄성과 관계가 있다. 포괄성은 단순히 수록된 레코드의 수를 측정하는 것이 아니라, 기본적으로 해당 데이터베이스에 포함되어야 할 정보가 실제로 어느 정도 포함되어 있는가를 나타내는 것으로 레코드를 단위로 비율(%)로 표시한다. 추가적으로 수록 레코드 수의 적절성, 소급정도 등이 측정요소로 간주될 수 있다. 우리나라 데이터베이스는 비교적 최근에 구축되어 서비스되기 시작하였으며, 인물데이터베이스도 예외가 아니다. 따라서 소급정도는 대단히 낮다. 본 연구에서는 포괄성 분석을 위하여 수록 레코드 수의 적절성과포괄정도(%)를 측정요소로 삼았다. 모든 데이터베이스의 포괄성을 측정하기는 불가하므로, 수록 범위가 알려져 있는 특정 데이터베이스군을 선택하여 포괄성을 측정하고 이와 더불어 중복성을 고찰해 보고자 한다.

네번째 제시된 불만요인, 정보가 구체적이지 못하다는 것은 내용의 완전성과 관계가 있다. 완전성은 보통 레코드 구성내용의 적절성, 탐색가능 필드의 적절성, 수록레코드 전체중에서 필드값이 모두 기록된 레코드의 비율 등을 측정하여 평가한다. 인물데이터베이스는 대체로 피전자에게 설문을 보내어 자발적으로 제공한 정보를 기초로 하며, 피전자가 원치 않는 부분은 기술하지 않아도 된다. 즉 해당항목의 포함/불포함 여부는 피전자 자신이 결정하므로, 필드값이 모두 기록된 레코드의 비율은 측정요소로 적절치 못하다. 따라서 완전성에 관한 사항은 레코드 구성 내용의 적절성과 탐색가능 필드의 적절성을 측정 요소로 삼았다.

다섯 번째로 제시된 불만요인, 정보가 다양하지 못하다는 점은 위에서 언급한 레코드 구성 내용의 적절성과도 관계가 있지만 주로 데이터베이스의 유형과 관계가 있다. 인물에 관한 정보를 가지고 있는 레코드는 전화번호부와 같은 단순한 형태로부터 정기간행물 기사, 자서전과 같은 도서에 이르기까지 다양하다. 인물정보는 탐색 목적과 필요로 하는 정보의 양에 따라 사용될 데이터베이스의 유형이 결정된다. 텍스트 데이터베이스의 유형은 사전, 디렉토리, 서지, 전문 등 다양하다. 따라서 다양성에 대한 측정 요소는 데이터베이스의 유형으로 설정하였다.

마지막으로 제시된 불만족 요인은 정보의 부정확성이다. 인물데이터베이스의 경우에 부정확 요인은 피전자가 잘못 기록한 정보와 입력작업중에 생기는 에러의 두 가지로 나눌 수 있다. 지난 20여년 동안 에러탐지를 위한 기법이 개발적용됨으로써 입력에러에 대한 문제는 어느 정도 성공적으로 해결되고 있다. 따라서 이 요인은 본 연구에서 평가요소로 설정하지 않았다.

이상과 같이 설정된 평가요소는 최신성, 포괄성, 완전성, 다양성이며, 분석 결과는 다음과 같다.

### 3. 인물데이터베이스 비교 분석

#### 3.1 최신성 분석

다음은 분석 대상 우리나라 인물데이터베이스 총 19종에 대한 최신성 분석이다. 15종에 대한 갱신빈도는 “알기쉬운 한국의 데이터베이스 목록”에 수록된 엔트리를 기초로 산출되었고, 추가된 4종에 대하여는 연구자가 직접 조사하였다.

<표 1> 우리나라 인물데이터베이스의 최신성

갱신빈도	데이터베이스 수
수시	2
일간	3
주간	1
2년	1
불규칙	12
계	19

<표 1>에 나타난 바와 같이 인물데이터베이스의 상당 부분이 (12종, 63%) 불규칙한 갱신 빈도를 가지고 있다. 피전자가 직접 등록, 수정하는 경우에는 불규칙적인 갱신이 불가피하지만, 별도로 제작자에 의한 레코드 전체에 대한 정기적 갱신과 저명인사의 신상변화에 따른 보완은 필수적인 것이다. 이와 같은 비계획적인 레코드 유지 방법은 정보에 대한 신뢰성을 크게 저하시킨다.

<표 2>는 외국의 인물데이터베이스 46종에 대한 최신성 분석이다. 갱신빈도는 “Gale Directory of Databases”에 수록된 엔트리 내용을 기초로 산출하였다.

<표 2> 외국 인물데이터베이스의 최신성

갱신빈도	데이터베이스 수
수시	4
일간	5
주간	3
격주간	1
월간	5
계간	2
반년간	5
연간	5
2-3년	5
5년	1
불규칙	2
미상	8
계	46

전체의 약 78%인 36종이 정기적으로 갱신되고 있는 것으로 나타났다. 데이터베이스 유형에 관계없이, 최신성이 필요한 분야, 즉 정치 법률 경제 관련 인물데이터베이스의 경우는 수시~일간으로 갱신되는 반면, 전형적인 디렉토리인 Who's Who 유형은 연간이 지배적이다. 월간이나 계간이 주를 이루는 서지 데이터베이스보다는 전문 데이터베이스가(AP, 일간; Baseline Celebrity, NYT Biographical File, Washington Post Bio Stories, 주간) 빠르게 갱신되고 있다. 최근에 생산되기 시작한 몇 가지 데이터베이스의 경우에 갱신빈도가 명시되어 있지 않았다(예, *Dictionary of Literary Biography 1997-*, *Discovering Biography 1997-*, *Wilson Biographies plus Illustrated 1999-* 등). 전반적으로 보아 외국 인물데이터베이스들의 갱신빈도는 정보 요구를 반영하고 있는 것을 알 수 있다.

### 3.2 포괄성 분석

우리나라 인물데이터베이스 규모에 대하여 수록 레코드 수를 기초로 한 분석은 <표 3>과 같다. 레코드 수에 대한 데이터는 2001년 10월 현재의 것으로 직접 탐색을 통하여 혹은 해당 데이터베이스 관계자와 직접 통화에 의하여 수집되었다.

<표 3> 우리나라 인물 데이터베이스의 규모

수록 레코드 수	데이터베이스 수
5,000 미만	2
5,000 - 9,999	2
10,000 - 49,999	7
50,000 - 99,999	0
100,000이상	5
미상	3
계	19

수록 레코드 수의 범위는 5만건 미만과 10만건 이상으로 이분되어 있다. 50,000건 미만의 데이터베이스가 11종이나 되는데, 그 중 4종이 10,000건 미만의 소규모 주제별 인물데이터베이스이다. 100,000건 이상의 대규모 데이터베이스가 5종 있으며, 이들은 대부분 신문사에서 제작되고 있는 종합 인물데이터베이스 들이다.

외국의 인물데이터베이스 46종에 대하여 Gale Directory 엔트리에 나타나 있는 수록 레코드 수를 기초로 한 분석은 <표 4>와 같다.

<표 4> 외국 인물데이터베이스의 규모

수록 레코드 수	데이터베이스 수
50,000미만	20
50,000- 99,999	5
100,000-199,999	4
200,000-299,000	1
300,000-499,999	0
500,000-999,999	1
1,000,000이상	2
미상	13
계	46

수록 레코드 수의 범위 또한 매우 다양하며, 주제별 분야별 인물데이터베이스들은 규모가



작은 편이다. 50,000건 이하의 데이터베이스가 20종이나 되어 전체의 43%를 차지하고 있는데, 그 중 러시아 공보처에서 자국의 주요 인물을 영어로 소개하는 Who's Who in Russia...는 500건의 레코드를 수록하며 매일 갱신된다. 1,000,000건 이상의 대규모 데이터베이스는 모두 서지데이터베이스로 Biography Master Index는 4,000,000건을, World Biography Index는 2,000,000건을 수록하고 있다.

우리나라의 인구 규모와 수록 정보의 소급정도에 비추어 볼 때, 외국의 데이터베이스와 직접적인 비교는 의미가 없다. 디렉토리 데이터베이스 중에서 국가나 지역별 인물의 현황정보를 다루는 종합인물데이터베이스의 수록 건수와 비교해 보면, 독일 Dictionary of German National Biography 의 60,000건, 북미(미국·캐나다)의 유명인사를 수록하고 있는 Marquis Who's Who in America Regional Publications는 300,000건, 유럽의 저명인사를 수록하고 3개 국어로 출판되는 BioDoc은 41,000건으로, 우리나라 신문사에서 생산되는 대규모 종합인물 데이터베이스의 수록건수(100,000~170,000건)에 비추어 보면, 수록레코드 수의 적절성 견지에서 결코 부족하지 않은 것으로 판단된다.

다음으로 데이터베이스의 포괄성을 측정하였다. 우리나라의 분석대상 인물데이터베이스를 수록 인물의 직업/분야별 범주로 나누어 보면 경영, 기술, 교육 분야에 다수의 데이터베이스가 존재하는 것을 알 수 있다.

포괄성을 측정하기 위하여는 평가의 기준이 될 수치, 즉 해당 인물데이터베이스에 포함되어야 할 레코드 수가 확립되어야 한다. 경영, 기술, 교육 분야 중에서 다수의 데이터베이스에 수록되고, 기존의 데이터를 통해서 기준치를 알 수 있는 탐색가능한 직업군은 대학교수가 유일한 것으로 식별되었다. 따라서 본 분석에서는 대학교수를 포함하고 있는 데이터베이스를 대상으로 포괄성을 측정하였다.

수록범위에 대학교수를 포함하고 있는 데이터베이스는 연구인력(한국과학재단), 학술연구자 정보(한국교육학술정보원), 교수인명록(한국대학신문사), 여성인명사전(여성개발원), 그리고 동아일보 문화일보 중앙일보 조선일보 등 신문사 제작 데이터베이스와 연합뉴스 인물정보가 있다. 그중 연구인력은 과학·기술분야만을 다루므로 제외하였으며, 중앙일보 인물정보는 3,000건 이상의 탐색이 불가능하므로 분석에서 제외하였다.

수록 인물의 선정 기준에 대학교수가 명시되어 있고 동시에 직업별 탐색이 가능한 데이터베이스를 직접 탐색하여 얻은 수록 레코드 수와 포괄 정도는 <표 5>와 같다. 완전한 포괄에 대한 기준치는 교육부에서 출판한 “교육통계연보 1999”에 나와있는 통계를 취하였으며 교수 전체는 55,371로 여성교수전체는 8,477로 추정하였다.

&lt;표5&gt; 인물 데이터베이스의 포괄성 측정

데이터베이스명	수록레코드 수	포괄성(%)	비고
동아일보 인물정보	39,799	71.87	2001년 11월 탐색
문화일보 인물정보	22,847	41.26	2001년 11월 탐색
조선일보 인물정보	44,872	81.10	2001년 11월 탐색
연합뉴스 인물정보	37,665	63.80	2001년 11월 탐색
교수인명록	42,770	77.24	2001년 11월 탐색
학술연구자 정보	52,851	95.45	관계자와 면담
여성인명록	4,748	55.01	2001년 11월 탐색

대학교수의 직업을 가진 인물에 대한 포괄성의 범위는 41.26%~95.45%에 걸쳐 다양하게 나타났다. 따라서 단일 데이터베이스 탐색으로는 이용자를 만족시키기 어렵다. 학술진흥재단에서 제작하고 한국교육학술정보원에서 서비스하는 학술연구자 정보가 가장 높은 95.45%를 나타내고 있으며, 그 다음은 조선일보 인물정보(81.10%), 교수인명록(77.24%), 동아일보인물정보(71.87%)의 순서로 되어 있다.

<표 5>에 나타난 데이터를 보면, 대학교수의 직업을 가진 인물에 대한 정보가 지나치게 다수의 데이터베이스에 중복 생산되고 있는 것을 알 수 있다. 물론 각 데이터베이스의 레코드 내용과 탐색점이 동일한 것은 아니다. 교수인명록은 간단한 엔트리를 가지고 있고, 학술연구자 정보는 논문·저서 등 학술적 상세 정보를 포함하며, 신문사에서 제작된 데이터베이스의 내용은 보다 다양하다. 그러나 위에 열거된 데이터베이스들은 정형의 엔트리를 가지고 있는 디렉토리 데이터베이스이며 거의 유사한 기본정보를 제공하고 있다. <표 5>에 나타나 있지 않는 중앙일보 인물정보와 연구인력 등의 데이터베이스에도 막대한 수의 교수 인물레코드가 포함되어 있다. 또한 본 연구의 분석대상은 아니지만, 오프라인으로 제공되고 있는 국가인재DB(중앙인사위원회)의 경우도 마찬가지다.

### 3.3 완전성 분석

텍스트 데이터베이스는 데이터 자체와 데이터를 쉽게 탐색할 수 있도록 만든 관련색인으로 구성된다. 본 연구에서 완전성을 평가하기 위한 측정요소는 레코드 구성 내용과 탐색필드의 적절성이다. 완전성을 평가하기 위하여, 우리나라 인물데이터베이스 레코드 중에서 가장 많은 내용요소를 포함하고 있는 중앙일보 인물정보를 선택하여, 미국의 대표적인 디렉토리유

형 인물데이터베이스인 Marquis Who's Who와 수록 내용 및 미국의 Dialog서비스를 통하여 탐색가능한 요소를 분석하였다. 분석 결과는 <표 6>과 같다.

<표 6> 중앙일보 인물정보와 Marquis Who's Who의 수록내용 및 탐색 요소

구분	중앙일보 인물정보		Marquis Who's Who	
	내용요소	탐색여부	내용요소	탐색여부
성명	국문	O	성명	O
	영문	X	기타명	X
	한문	X		
	호	X		
	본관	X		
연락처	본적	X	주소	주택/직장 O
	원적	X		
	직장주소	X	주택주소	주명 O
	직장전화번호	X		지역명 O
	직장 FAX번호	X		도시명 O
	직장 e-mail	X		우편번호 O
				O
		주택전화번호	도시명 O	
직업	직업명	O	직업명	O
	직위	O	부업명	X
	소속기관명	O	직위	O
			소속기관명	O
			자격증	O
출생/사망	성별	X	성별	O
	출생년월일	X	출생년월일	연도 O
	출생지	O	출생지	국가명 O
				주명 O
				지역명 O
				도시명 O
	사망년월일	X	사망년월일	연도 O
	사인	X		
	장례일	X		
	발인장소	X		
장지	X			

(출처 : 중앙일보 인물정보 <http://people.joins.com>

Marquis Who's Who <http://library.dialog.com/bluesheets/htmlbl0234>)

12 한국도서관·정보학회지 (제 32권 제 4호)

구분	중앙일보 인물정보		Marquis Who's Who	
	내용요소	탐색가능여부	내용요소	탐색가능여부
가족/교육 관계	결혼기념일	X	부모명	X
	자녀수	X	배우자명	O
	부모명	X		
	배우자명	X		
	자녀명	X		
	교우	X		
종교/가입단체	종교	X	종교	X
	가입단체	X	정당가입	O
			가입단체	O
교육	학력	출신학교명 O	학력	출신학교명 O
				출신교지역명 O
				학위명 O
				수업/졸업년도 O
경력	경력	X	경력	X
저서/논문	저서/논문	X	저서/논문	저서/논문명 O
상훈	상훈	X	상훈	X
신상	신상	X		
	몸무게	X		
	혈액형	X		
기타	취미	X	관심주제/분야	O
	생활신조	X	군복무	X
	흡연량	X	인생관	X
	기호품	X		
레코드관련			생신일자	O
			생신년도	O

데이터베이스 레코드의 내용요소를 살펴보면, 중앙일보 인물정보가 호적, 사망, 신상, 기타 사항 등에서 상세한 정보를 제공하며, Marquis Who's Who는 연락처, 전문성, 사회활동 등에서 상세한 정보를 제공한다. 사회문화적 환경에 따라 차이는 있지만 제공정보 자체의 다양성에 대한 차이는 크지 않다. 그러나 탐색요소에서는 막대한 차이가 있다. 중앙일보 인물정보는 탐색요소가 6가지에 불과한 반면, Marquis Who's Who의 탐색 요소는 30가지나 된다. <표 6>에 나타난 바와 같이 Marquis Who's Who의 레코드는 내용요소가 거의 대부분 탐색 가능하며, 탐색요소가 더욱 세분화 되어 있다(예, 연락처 내의 주, 지역, 도시, 우편번호). 우리나라 디렉토리 인물데이터베이스는 대부분 소수의 탐색점을 제공하고 있으며, 탐색점의 범위는 2-8가지에 이른다. 인물데이터베이스 중에서 가장 많은 탐색요소를 제공하는 조선일보

인물정보의 경우에도 <표 6>에 나타난 중앙일보 인물정보의 탐색점 외에 출생년도와 경력 두 가지의 탐색요소가 추가될 뿐이다.

### 3.4 다양성 분석

우리나라에서 서비스되고 있는 인물데이터베이스는 모두 디렉토리 데이터베이스 유형에 속한다. 따라서 데이터베이스 유형에 대한 분석은 수행할 수 없었다. 우리나라 인물데이터베이스는 다양성 면에서 대단히 낙후되어 있는 것을 알 수 있다.

외국의 인물데이터베이스 46종에 Gale Directory 엔트리에 나타나 있는 유형을 분석한 결과가 <표 7>과 같다.

<표 7> 외국 인물데이터베이스의 유형

유형	데이터베이스 수
디렉토리	27
서지	3
전문	9
디렉토리+전문	2
서지+전문	1
수치+전문	1
전문+서지+이미지	3
계	46

외국의 인물데이터베이스는 다양한 유형으로 구성되어 있다. 일부 데이터베이스는 두 가지 이상 유형의 혼합이기도 하다. 디렉토리 유형이 29종으로 지배적이며 전체의 약 63%를 차지하고 있다. 전문 데이터베이스가 모두 16종, 서지 데이터베이스는 모두 7종으로 나타나 있으며, 수치 데이터베이스도 1종 있는데, 이것은 독일에서 서비스되고 있는 스포츠 분야의 데이터베이스이다. 이용자요구에 따라 다양한 인물정보의 레코드를 활용할 수 있다. 이용자가 찾고자하는 인물에 대한 객관적 정보를 필요로 할 때는 인물사전이나 백과사전과 같은 서술적으로 표현한 데이터베이스를 탐색하며, 개인의 이력에 대한 종합적인 기초데이터를 필요로 할 때는 항목별로 일정량의 정보를 제공하는 디렉토리 데이터베이스를 탐색한다. 인물에 대한 보다 깊이 있는 탐색, 즉 인물리서치를 위하여는 서지 데이터베이스가 필수적이다. 그리

고 인물정보 이상의 상세한 내용은(예, 생활양식, 사고방식, 정치성향 등등) 전문데이터베이스를 통하여 직접 문헌을 보고 얻을 수 있다.

#### 4. 결론 및 제언

본 연구에서는 우리나라 인물데이터베이스를 대상으로 온라인 상용 데이터베이스 이용자들의 불만족 요인으로 식별된, 최신성, 포괄성, 완전성, 다양성 등을 분석 평가하고, 가능한 경우에 외국의 인물데이터베이스 현황과 비교하였다. 분석결과를 요약하면 다음과 같다.

첫째, 우리나라 인물데이터베이스는 상당 부분이(63%) 불규칙한 갱신 빈도를 가지고 있으며 이러한 비계획적인 레코드 생산과 유지방법은 최신성의 결핍은 물론 정보에 대한 신뢰성을 크게 저하시키고 있다. 반면에 외국의 인물데이터베이스는 대부분 정기적으로 갱신되고 있으며(78%), 갱신빈도가 주제분야나 데이터베이스 유형에 따라 다양하므로 이용자 정보요구가 반영되어 있는 것을 알 수 있다.

둘째, 인구 규모나 수록 정보의 소급정도에 비추어 볼 때 우리나라 인물데이터베이스는 레코드 수의 견지에서는 외국의 인물데이터베이스보다 크게 부족하지 않는 것으로 나타났다. 그러나 보다 엄밀한 포괄성 측정결과를 보면, 그 범위가 다양하며(41.26%~95.45%) 지나치게 다수의 온라인 데이터베이스에(8종) 유사 정보가 중복하여 포함되어 있는 것을 알 수 있다.

셋째, 레코드 내용과 탐색가능 요소를 기초로 한 완전성 분석 결과를 보면, 우리나라와 외국의 인물데이터베이스는 레코드 구성 내용보다는 탐색 요소에 현격한 차이가 있는 것을(6 : 30) 알 수 있다. 우리나라 인물데이터베이스는 대부분의 내용요소가 탐색이 되지 않고 디스플레이만 되고 있으며, 제한된 탐색점은 일종의 known item 탐색만을 가능케 한다.

넷째, 우리나라 인물데이터베이스는 다양성 면에서 크게 낙후되어 있다. 인물데이터베이스는 모두 디렉토리 유형에 속한다. 데이터 수집과 조직 업무가 가장 단순한 유형에 치우쳐 데이터베이스가 생산된 것이다. 반면에 외국의 경우에는 텍스트 데이터베이스의 모든 유형이 망라되어 있으며, 이와 같은 다양성은 이용자 요구를 반영하고 있다.

이상과 같은 비교분석 결과를 토대로 우리나라 인물데이터베이스의 현황을 개선하기 위한 발전방안을 제시하면 다음과 같다.

우리나라의 인물데이터베이스는 다른 분야와 마찬가지로 다수의 데이터베이스가 구축되어 있으나, 정보시스템 구축 형태는 기존의 인쇄자료를 단순히 기계가독형으로 변화시켜 놓은

초기 단계에 머물러 있다. 온라인의 장점인 최신성과 데이터베이스의 장점인 다양한 탐색이 제공되고 있지 못하다. 이용자를 만족시키기 위하여는 정기적인 갱신이 이루어져야 하며, 이용자 요구에 따라 더욱 많은 탐색점을 제공하여 unknown item 탐색이 가능하도록 시스템이 변경 개발되어야 한다.

동일한 유형의 유사한 정보가 다양한 수준의 포괄성을 가진 다수의 데이터베이스에 중복되어 있는 현황은, 국가적으로는 중복투자 및 자원낭비, 이용자에게는 탐색의 불편과 비효율성을 초래하게 되었다. 중복투자를 막기 위하여는 현 단계에서라도 제작기관별 분업이나 조정이 이루어져야 하며, 이용자 탐색의 불편을 덜기 위하여는 통합데이터베이스 구축, 다수 데이터베이스 동시탐색 혹은 연계서비스가 확립되어야 할 것이다. 현재로는 천리안에서 유일하게 중앙일보, 문화일보, 조선일보 3종의 인물정보 동시탐색을 제공하고 있다.

다양성을 개선하기 위하여는 대규모의 투자와 창의적인 노력이 필요하다. 인물탐색은 필수 통계를 제공하는 디렉토리 데이터베이스 탐색으로 시작되지만, 이용자들은 보다 다양한 탐색 요구를 가지고 있으며 특히 서지데이터베이스의 구축은 전문적 인적자원을 필요로 한다. 한편 WWW기반 이용자들에게는 사실정보원으로서 에세이 형태의 인물정보가 제공되어야 할 것이다. 다양성과 관련하여 고려할 사항은 한국의 인물을 외국에 알릴 수 있는 영문 데이터베이스의 구축이다. 우리나라의 데이터베이스 산업이 국내 정보의 영문판 제작에 게으르다보니, 외국의 기업이 국내 정보를 데이터베이스로 제작하여 우리나라에 역수출하는 현상이 생기고 있다. 인물데이터베이스의 경우도 예외가 아니다(K.G. Saur사의 Korean Biological Archive). 세계화를 지향하는 정보사회에서 영문 데이터베이스 제작에 보다 적극적인 노력이 필요하다.

부록 1. 분석대상 우리나라 인물데이터베이스

1. 건설관련 주요인사 정보
2. 경영자 정보
3. 교수인명록
4. 동아일보 인물정보
5. 문화일보 인물정보
6. 북한인명록
7. 시민운동 인물정보
8. 언론인명 정보
9. 여성인명 사전
10. 연구인력
11. 연합뉴스인물정보
12. 인물정보(한국기업평가)
13. 인물정보(한국신용평가)
14. 전자산업인 정보
15. 조선일보 인물정보
16. 중앙일보 인물정보
17. 학술연구자 정보
18. IT 인물정보
19. KOLIS 법조인명록



## 부록 2. 분석대상 외국 인물 데이터베이스

1. Albi Awocatie Procurator
2. American Men and Women of Science
3. Artists in Canada
4. Aspen Law & Business Directory of Corporate Counsel
5. AP/Candidate Biography
6. ASTERI
7. Baseline Celebrity Bios
8. BEST Great Britain
9. BIODOC
10. Biography Master Index
11. Biography Resource Center
12. Bowker Biographical Directory
13. Congressional Staff
14. CQ Member Profiles
15. DBE
16. Dictionary of Literary Biography
17. Discovering Authors Modules
18. Discovering Biographies
20. KUKA
21. Kvinder pa Linien
22. MAPERS
23. MAPOP
24. Marquis Who's Who
25. Marquis Who's Who in American Regional Publications.
26. Marquis Who's Who in Topical Library
27. MASPORT
28. NYT Biographical File
29. People Library
30. Riksdags Information
31. Standard & Poors Register-Biographical

32. Technical Advisory Services for Attorney
33. Washington Post Biographical Stories
34. West's Legal Directory
35. Who' Who in American Art
36. Who's Who in American Politics
37. Who's Who in European Business and Industry
38. Who's Who in Germany
39. Who's Who in Russia and the Common Wealth of Independent States
40. Who's Who in Technology
41. Wilson Biographies
42. Wilson Biographies plus Illustrated
43. Wilson Biograhly Index
44. Wilson Current Biography(1940- )
45. Wilson Current Biographies(1983- )
46. World Biography Index

## 참 고 문 헌

- 교육부. 『교육통계연보 1999』. 서울: 교육개발원, 1999.
- 박준식, 박경호. “한국 웹기반 인물정보원의 분석적 연구”, 《한국문헌정보학회지》 35권, 1호 (2000). pp. 187-203.
- 장혜란. “데이터베이스 품질평가를 위한 모형개발 - 텍스트 데이터베이스 내용을 중심으로”, 《정보관리학회지》 17권, 4호(2000). pp. 83-97.
- 최명규. “국내 데이터베이스 이용자의 이용실태에 관한 연구”, 《정보관리연구》 29권, 2호 (1998). pp. 18-36.
- 한국데이터베이스진흥센터. 『알기 쉬운 한국의 데이터베이스 목록 1999』. 서울 : 한국데이터베이스진흥센터, 1999.
- 한국데이터베이스진흥센터. 『데이터베이스 이용실태 및 정보수요 조사 보고서』. 서울 : 한국데이터베이스진흥센터, 2001.
- 한상길. “국내인터넷 WWW 전자신문 비교 분석”, 《정보관리학회지》 13권, 2호 (1996). pp. 73-96.
- Basch, R. "Beyond Who's Who: Bibliographical Sources for Competitive Intelligence", *Proceedings of Online/CD-ROM '91 Conference*, (1991). pp. 10-14.
- Bates, J. M and Lu, Shaojun, 1997. "An Exploring Profile of Personal Home Pages: Content, Design, Metaphors", *Online & CD-Rom Review*, Vol. 21, No. 6(1997). pp. 331-340.
- Borasky, D. V. "Researching People on the Net", *Online*, Vol .22, No. 2(1998). pp. 55-59.
- Dialog. "Marquis who' s who".  
<http://library.dialog.com/bluesheets/html/bl0234.html>. [2001-08-12].
- Detemple, W. "Future Enhancements for Full Text-Graphics, Expert Systems and Front-End Software", *Proceedings of the International. Online Meeting*, (1988).
- Faerber, M. ed. *Gale Directory of Online Databases*. 2000 ed. Vol. 1. Detroit : Gale, 2000.
- Gordon-Till, J. "Discover: People Information", *Managing Information*, Vol. 5, No.10(1998). pp. 27-30.
- Feffer, C. "More than Just a Name: the Art of Prospect Research", *Aslib Proceedings*, Vol. 48, No. 4(1996). pp. 103-107.
- Lane, C. A. *Naked in Cyberspace*. Wilton, CT : Pemberton, 1997.
- Pack, T. "Finding Basic Information about People". *Database*(Apr. 1994). pp. 61-65.

20 한국도서관·정보학회지 (제 32권 제 4호)

Smith, L. C. and Lastra, S. "Ready Reference on the Internet", *Proceedings of the National Online Meeting*. (1999). pp. 487-

Wollschlaeger, H. "World Archives of Bygone Days".

<http://www.saur.de/bioarch/wollensch.htm> [2001-08-02]