

의학 분야 Web DB의 품질평가

- PubMed와 Embase를 대상으로 -

A Study on Quality Evaluation of Medical Web DBs : PubMed and Embase

김 상 준(Sang-Jun Kim)*

목 차

- | | |
|---------------------------|-----------------------------|
| 1. 서 론 | 4. PubMed와 Embase DB 품질의 비교 |
| 2. DB 품질평가 선행연구 및 | 평가 |
| PubMed와 Embase DB 품질 비교 | 5. 결 론 |
| 3. PubMed와 Embase DB 품질평가 | |
| 기준과 방법 설정 | |

초 록

본 연구는 의학 분야 Web DB의 품질평가를 위해 무료 PubMed와 상용 Embase DB를 선행연구 분석을 통해 품질평가 기준을 추출하여 객관적으로 측정하기 쉬운 평가항목을 중심으로 발견적·실무적 비교 평가를 실시한 것이다. 그 결과 PubMed는 최신성, 저널 형태의 논문 비중, 초록 보유 자료와 영어자료의 비중, 전자 저널 원문 링크 출판사 수, 자판 소장여부를 표시하는 LinkOut과 구입비용에서 우수하였다. 반면 Embase는 포괄성, 제약과 화학 분야, 수록정보의 다양한 형태에서 우수하였다. 따라서 상용 DB 도입의 타당성 검토와 이용자 교육 및 도서관 정보서비스는 이러한 두 DB의 품질평가 결과를 바탕으로 이루어져야 한다.

ABSTRACT

This study concerns DB quality of the Web databases which has been produced in medical science. For the quality evaluation of Embase and PubMed, 10 evaluating criteria were developed on the basis of literature review. The evaluation result showed that PubMed has superiority to the currentness, accuracy, completeness, consistency, ease of use, customer support, searchability, cost, and network & H/W on DB, but Embase has only superiority to the coverage. It is needed for purchasing, user training, and library service based on the above evaluation result.

키워드: 데이터베이스, 비교 평가, 정보시스템, 품질 비교, 품질평가

DB Evaluation, DB Quality, Embase, Medline, PubMed, Quality Evaluation, Web DB

* 한국생명공학연구원, 충남대학교 박사과정(sj.kim@kribb.re.kr)

논문접수일자 2004년 5월 14일

개재확정일자 2004년 6월 7일

1. 서 론

1. 1 연구의 필요성과 목적

최근 인터넷과 정보통신기술의 발달로 정보검색시스템과 DB의 구축 및 정보서비스가 급격하게 증가하고 있다. 인터넷의 등장과 함께 DB 검색은 도서관이나 정보센터는 물론 일반 국민들에게도 생활화 되어 가고 있다. 따라서 늘어나는 DB 중 검색목적이나 필요에 적합한 DB를 선정하여 효율적으로 이용하는 일이 점점 더 중요해지고 있다. 이와 같이 적합한 DB를 선정하여 효율적으로 이용하는 일에는 정보검색시스템이나 DB에 대한 올바른 품질평가가 전제되어야만 가능한 것이다. 그러나 정보검색시스템이나 DB에 대한 정립된 품질평가 기준과 방법이 없는 관계로 연구자들은 저마다 다양한 견해와 연구를 시도하고 있는 실정이다.

학술활동에 있어 생산성은 여러 요인에 의해 영향을 받지만 연구자를 둘러싼 정보시스템의 품질도 중요한 요인 중의 하나이다. 이러한 정보시스템의 가장 핵심적인 요소가 DB이며, DB의 품질은 정보시스템의 유용성에 결정적인 영향을 미친다(이제환 2002, 62). 국내에서도 ‘정보화’가 국정의 주요 목표로 등장하면서 다양한 형태의 학술 정보시스템이 구축되고 있다. 이 과정에서 처음에는 DB 구축에 집중하였지만 이제는 구축된 DB의 효율적인 서비스로 조금씩 초점을 옮겨지면서 정보검색시스템 및 DB의 품질평가가 중요해지고 그에 따른 연구도 점차 활성화되고 있다.

최근 몇 년 동안 급속히 증가하는 인터넷 정

보수요를 충족시키기 위해 정부는 물론 일반 DB 제공자는 질보다는 양적인 DB 구축에 더 힘을 기울이고 있다. 그러나 DB의 품질문제는 컴퓨터가 탄생한 이래 줄곧 이어져 온 고질적인 문제로 최근 DB 사용의 보편화·대중화 추세로 더 심각해지면서 일반 이용자들은 불량 데이터의 위험에 속수무책이 되고 있다(이용봉 외 2001, 110). 따라서 정보검색시스템의 핵심적인 품질이라 할 수 있는 DB의 효과성과 효율성의 측정은 DB, 정보검색엔진, 이용자 인터페이스, 이용자 만족도, 정보접근·전달 요소 등의 정보검색시스템 구성요소(유사라 1999, 158) 각각의 측면에서 품질평가가 이루어져야만 한다.

한 조사에 의하면(한국데이터베이스진흥센터 2003, 15-28) 국내 기관종사자들의 상용 온라인 DB 이용은 1998년을 정점으로 감소 추세이며, 2002년에는 57.5%의 이용률과 1주 일간 평균 7.9시간에 평균 39.7건을 사용하며, 문자 정보의 이용이 60% 정도인 것으로 나타나고 있다. 반면 개인이용자들은 상용 온라인 DB를 2002년에 37.6%의 이용률과 1주일간 평균 3.2시간에 평균 12.3건을 사용하며, 학술 DB 이용자는 13.6%이지만 전체 DB 중 학술 DB의 점유율은 1.2%에 불과하여 상용 온라인 DB 이용료에 대한 부담을 크게 느끼는 것으로 보인다.

본 연구는 Web 환경에서 제공되는 의학 분야 서지정보 DB의 대표 격인 무료 PubMed 와 상용(유료) Embase DB의 품질 비교 평가를 통해 탐색전문가는 물론 최종이용자들이 적합한 DB를 선택하여 최적의 정보를 탐색할 수 있는 토대를 마련하고, 분석된 문제점과 개선

방안을 통해 DB의 품질개선과 정보서비스 활성화에 보탬이 되고자 하는 것이 주요한 연구 목적이다.

그동안 국내에서 연구가 주로 이루어진 자체구축 DB의 품질평가 뿐만 아니라 본 연구와 같은 상용 DB의 품질평가도 동시에 발전된다면 높은 DB 품질, 정보이용의 향상과 새로운 정보 생산, 연구시간 단축, 연구자의 능률성과 효과성 향상, 생산성 있는 연구라는 높은 가치 발생의 순환과정을 거치게 되어 효과적인 도서관 정보서비스가 이루어질 수 있을 것이다.

1. 2 연구의 내용과 방법

최근 정보공유의 확산과 데이터 이용 기법의 고도화로 보다 정교한 데이터 관리가 요구되고, 이에 따라 DB의 품질 향상을 위한 DB의 품질평가 모델의 표준 정립과 표준 품질평가 모델에 따른 DB 품질평가가 필요한 실정이다. 특히 외국의 상용 DB 도입에는 웹 정보자원의 선정과 평가의 측면에서 심도 깊은 검토를 거쳐야 하지만, 국내 현실은 국내 DB 구독 대행사의 영세성과 문헌정보학자의 무관심과 도서관 실무자의 연구 활동 부족(김상준 2002)으로 정확한 DB 품질평가 과정 없이 상용 DB 가 도입되고 있는 실정이다. DB 컨소시엄 구독이 늘어나면서 DB 품질평가의 중요성이 더 높아지는 실정을 감안하면 정확한 DB 품질평가를 거쳐 올바른 이용자 교육의 방향과 정보서비스의 초점을 명확히 할 필요가 있다.

본 연구에서는 이러한 환경과 현실을 감안하여 DB 품질평가에 관한 국내외 선행연구를 종합적으로 조사·분석하여 검토한 뒤 표준모델

로 사용할만한 DB 품질평가 기준을 설정하고, 그 DB 품질평가 기준에 따라 의학 분야와 BT(생명공학) 분야 연구 현장에 서지정보 Web DB 형태로 널리 활용되고 있는 무료 PubMed 와 PubMed를 보완할 수 있는 유료 Embase DB의 품질평가를 비교해보고자 한다.

PubMed와 Embase DB의 품질 비교 평가를 위한 본 연구의 방법은 다음과 같다.

첫째, 정보시스템 및 DB 품질평가에 관한 국내외 선행연구와 PubMed와 Embase DB의 품질 비교연구에 대한 문헌분석을 실시한다.

둘째, 선행연구의 DB 품질평가에 관한 기준을 분석하여 본 연구에 적용할 품질평가 기준을 정립하고 그에 따른 품질평가 방법을 수립한다.

셋째, 품질평가를 수행하여 밝혀진 사실을 토대로 PubMed와 Embase DB의 도입과 이용자 교육 및 도서관 정보서비스를 위한 제안을 결론에서 다루고자 한다.

본 연구는 상용 DB의 품질평가에 대한 연구가 척박한 국내 현실에서 DB 수록대상 저널의 상당수가 중복되는 유사한 주제의 상용 DB와 무료 DB를 비교 평가 하여 상용 DB 도입 타당성의 근거를 확보하고, 평가결과에 따라 DB 가 도입된 이후 DB 검색서비스와 이용자 교육 및 DB의 활발한 활용을 위한 실무적 뒷받침을 할 수 있다는데 그 가치가 있다고 하겠다.

2. DB 품질평가 선행연구 및 PubMed와 Embase DB 품질 비교

DB의 품질평가와 관련된 국내외 선행연구

는 평가의 목적과 평가대상 및 평가주체에 따라 다양한 평가기준과 절차 및 방법이 개발되어 적용되지만, 크게 보아 불특정 다수 DB를 대상으로 객관적 평가기준과 계량적 측정방법을 이용한 ‘범용모델’과 특정 DB의 품질을 실증적으로 평가하고 개선방안을 마련하기 위해 사례별 평가기준을 설정하여 평가하는 ‘사례연구’로 구분할 수 있다(이제환 2002, 63).

사실상 DB 품질평가를 위한 ‘범용모델’의 개발에 가장 중요한 것은 보편성 있는 평가기준과 합리적인 측정방법 그리고 객관적인 등급 결정 메커니즘이다. 이 요소 중에 어느 하나라도 결여될 때 그 모델은 객관적이고 신뢰할 수 있는 평가결과를 기대하는 DB 사용자들의 기대와 요구에 부응하지 못하게 된다. DB의 개발자와 사용자 사이의 품질에 대한 인식의 차이와 같이 DB의 품질에 대한 인식의 차이 또한 ‘범용모델’의 개발을 어렵게 만드는 요인이다. 또 하나의 문제는 모든 사용자가 동일한 관점에서 DB의 품질을 판단하는 것이 아니라 는 사실이다. 급증하고 있는 상업용 DB의 경우, 사용자의 선택을 돋기 위해 객관적인 품질을 비교 평가하기 위한 ‘범용모델’의 개발이 이상적일지 모르지만, 특정기관에 의해 제작된 DB의 정기적인 품질 검증을 위해서는 해당 DB의 구축목적과 기능을 고려한 ‘사례별(혹은 특화된) 모델’이 보다 바람직할지 모른다(조순영 외 2002, 31-32).

이하에서는 DB의 품질평가와 관련된 국내외 선행연구를 외국과 국내연구로 나누어 발전 과정과 동향을 살펴보고, 본 연구의 연구대상인 PubMed와 Embase DB의 품질 비교연구에 대해 자세히 알아보고자 한다.

2. 1 외국의 DB 품질평가 선행연구

DB 품질평가에 관한 외국 선행연구의 동향과 발전과정을 국내 연구자의 연구결과를 중심으로 정리하면 다음과 같다(이명희 1998, 223-224; 이옹봉 1999, 31-32; 장혜란 2000, 85-87; 김석영 2002, 65-67; 김성희 2003, 19-25에서 각각 재인용).

60년대부터 이차자료의 품질에 대한 실증적인 연구가 시작되어 포괄성에 대한 고전적인 연구가 있은 이후 포괄성, 중복성, 유일성에 관한 다수의 연구가 레코드 분석이나 설문조사로 수행되었으며, 적합성, 일반성, 사용 용이성, 최신성 등의 요소가 색인의 질을 평가하기 위한 기준으로 언급되었다.

70년대에는 DB 14종의 포괄성 평가와 같이 유사한 분야·주제·자료유형에 따라 다수 DB를 상호 비교하는 연구가 수행되었다.

80년대에는 온라인 DB의 품질관리에 영향을 주는 5가지로 범위, 구조, 접근성, 정확성, 일관성을 제시한 연구나 DB 품질요소로 정확성에 대한 측정을 위해 철자 에러, 특정 에러 탐지, 전거파일 정정에 관한 연구가 있었다. 80년대 말에는 CD-ROM DB를 위해 정확성과 다큐멘테이션, 적합성, 이용 편의성, 권위와 신뢰성, 내용, 유사 DB와의 비교, 갱신, 가치 등 8가지 기준에 대한 연구와 같이 기존 평가요소와 더불어 이용자 인터페이스와 도서관 장서로서의 가치 등이 추가되었으며, 1989년 Finnish Society for Information Service는 평가 기준으로 시스템과 통신연결, 탐색언어와 기술, 탐색소프트웨어의 효과, DB 내용, 검색보조, 비용 등 6개 부문 109개 문항으로

구성하였다.

90년대 초반에는 미국 SCOUG(Southern California Online User Group)에 의해 일관성, 범위와 포괄성, 최신성, 정확성, 접근 가능성, 완전성, 출력형태, 디큐멘테이션, 이용자 교육과 지원, 비용대가치 등 10개 부문 디수 문항리스트로 된 평가기준이 마련되어 1993년 LA(The Library Association)와 UKOLUG(United Kingdom OnLine User Group)에 의해 창립된 영국 CIQM(Center for Information Quality Management)에 의해 신규 DB 등록기준으로 채택되면서 DB의 품질을 평가하기 위한 보편적인 기준(SCOUG의 9개 부문 이외에 품질관리에 대한 진술, 통계치, 레코드의 필드구성 추가)으로 활용되었다. 또한 CD-ROM DB 평가기준으로 상품명세, 기술명세, 디큐멘테이션/지원, DB, 이용자 인터페이스, 탐색, 출력, 신뢰성 등 8개 부문에 세목을 정해 점수화한 연구가 있었다.

90년대 후반에는 4개 평가 기준으로 정확성, 현행성, 완전성, 일관성을 제시한 AT&T 연구소의 연구(유사라 1999, 161 재인용)가 있고, 환경과학과 화학 분야 DB 평가를 위해 온라인 DB의 일반적 선정기준으로 데이터 소스의 크기, 시간당 온라인 탐색비용, 온라인 DB 개선, 타 미디어 이용성을 들고, 분야별 선정기준도 제시하면서 평가 적용 기준으로 비용, 시간, 질적 요소(수록범위, 재현율, 정도율, 데이터의 완결성과 정확성)를 듣 연구도 있었다. 한편 의학정보는 Downstream Filtering(이용자)와 Upstream Filtering(정보중개자)로 구분이 가능하므로 의학 분야를 위한 med-PICS(Platform for Internet

Contents Selection)를 개발한 연구(Eysenbach and Diepgen 1998)도 있었다. 또한 미국의 대학도서관 컨소시엄인 HELEN의 참고 사서들은 DB 평가 가이드라인으로 색인된 정보원, 색인된 학술지, 포괄성, 색인, 탐색기능, 탐색지원, 스크린 설계, 본문/그래픽 내용, 출력, 신뢰성, 접근성, 가격 등 12개 부문 54개 질문으로 된 기준을 구성하였다.

2000년대에는 캘리포니아 주립대학(CSU)의 EAR(Electronic Access to Information Resources) 위원회에서는 크게 5개 요소(DB 내용, 탐색 인터페이스, 이용자 지원서비스, 비용, 접근성, 기타)로 구분하여 1990년부터 그 동안 약 45개의 DB 평가를 실시하여 대학의 요구에 적합한 DB를 선정해오고 있다 (<http://seir.calstate.edu/acm/ear/index.shtml>).

이러한 외국의 선행연구를 요약하면 텍스트 DB의 품질에 관한 연구는 초기의 인쇄본 이차자료를 대상으로 시작하여 60년대 기계 가독형 DB 출현으로 정확성을 평가하는 연구가 발생하여 70년대에 비교 평가가 활발하였고, 80년대에 CD-ROM DB의 평가가 나타남과 동시에 대규모 평가기준이 나타났고, 90년대 인터넷을 활용한 탐색이 증가되면서 다양한 평가 모델이 시도되었고, 2000년대에는 웹 자원과 디지털 도서관의 사용성 평가가 시도되어 정보시스템 및 DB 품질평가가 치밀해지고 있다.

2. 2 국내의 DB 품질평가 선행연구

1995년에 국내 선행연구로 이국희 등(1995)이 한국데이터베이스진흥센터(DPC)에서 발간

한 연구보고서에 DB 품질평가에 관한 연구를 시작하여 품질평가 7개 기준(정확성, 완전성, 현행성, 일관성, 검색성, 사용 용이성, 사용자 지원성)을 발표하였다.

1996년에는 이응봉(1996)은 DB 품질에 관한 비평적 평가를 수행하였고, KORDIC(현재 KISTI) 구축 서지 DB의 품질을 평가하기 위한 이제환(1996)의 '사례연구'도 있었다.

1997년에 김선흥은 Saturn DB 등 KORDIC의 DB를 정확성, 일관성, 완전성, 현행성 등 4개 기준으로 평가하였고(김선흥 2002, 67 재인용), 이제환(1997)은 Saturn DB 등을 이용자 측면에서 정보자료(포괄성, 적합성, 최신성), 단위 레코드(레코드 구조의 일관성, 필드의 적합성, 데이터의 정확성, 갱신주기, 중복률), 서비스의 품질(편의성, 인터페이스의 적절성, 검색엔진의 능력, 이용자 지원)로 구분하여 평가하였다.

1998년에 이명희(1998)는 ERIC DB 중 ASKERIC을 선정해 정확성, 일관성, 완전성, 현행성의 4개 품질기준 측정을 수행하였다.

1999년에는 유사라(1999)가 KAIST, KORDIC, KINITI DB의 정보 품질평가를 위한 메타분석을 실시하였고, 이응봉(1999)은 CD-ROM DB의 품질평가를 위해 국립중앙도서관과 국회도서관 발행 CD-ROM에 대한 비교 평가를 실시하였다.

2000년에는 한국데이터베이스진흥센터(안계성, 함정훈 2000)가 '데이터베이스 품질평가 항목' 보고서에 17개 국내외 선행연구의 비교 분석을 통해 DB의 품질평가를 위한 범용모델의 기준을 마련했고, 장혜란(2000)은 DB 품질 평가를 위한 평가모형을 개발했고, 유사라

(2000, 137)는 정보서비스 품질평가 기준으로 검색성, 접근 용이성, 이용자 지원성으로 나누어 평가하고 있었다.

2001년에는 이응봉 등(2001)은 웹 기반 환경에서 Saturn DB 등의 과학기술 분야 DB 평가기준 18개를 개발하여 데이터의 품질에는 정확성, 완전성, 최신성, 수록범위, 전문성을 서비스의 품질에는 검색성, 용이성, 사용자 지원, 비용, 네트워크 및 하드웨어 등을 들고 있고, 문성빈 등(2001)은 BIST와 Saturn DB의 통합을 위한 평가 분석을 위해 데이터의 품질(정확성)과 서비스의 품질(검색성, 사용 용이성, 사용자 지원성)로 구분하여 평가한 연구가 있었다.

2002년에 한국교육학술정보원(조순영 외 2002)에서 해외 학술 DB 선정 및 평가모델 개발에 관한 연구를 수행하여 국내외 17개 선행연구의 품질평가 기준을 분석하여 품질 요소, 서비스 요소, 구매조건으로 나누어 기준을 설정하고 전문가 집단을 통해 1~4점 척도로 평균을 산출하되 AHP 기법을 활용하여 평가 요소에 대한 중요도 및 가중치를 측정하는 프로그램을 개발하였다. 문성빈 등(2002)은 과학 기술전문정보서비스 품질평가를 위해 KISTI의 BIST와 Saturn DB를 통합하기 위한 목적으로 비교 평가를 수행하였다. 그리고 김선흥(2002b)은 과학기술 분야 해외 온라인 DB 현황을 평가하고 조사하면서 618종 DB를 주제분야별, 발행처별, 갱신주기별, 온라인 이용방법별로 분석하여 간략하게 소개하였다.

2003년에는 김성희(2003)가 한국교육학술정보원(조순영 외 2002)에서 수행된 연구과정에 포함된 DB 평가모델 개발을 위한 해외사례 조

사연구에 대한 소개가 있었고, 이만수 등(2003)은 KERIS 종합목록(UNICAT)과 국립중앙도서관의 국가자료공동목록(KOLIS-NET) 서지 DB의 품질평가를 비교한 연구를 수행하였다. 또한 이춘열(2003, 9~11)은 데이터 품질평가 항목(완전성, 일관성, 최신성, 정확성), 데이터관리 프로세스 평가 항목(적절성, 운영성), 시스템 평가 항목(안정성, 보안성, 편의성)으로 구분하여 평가하고 있다. 한편 한국데이터베이스진흥센터에서는 2003년에도 DB 품질평가 모델의 확장 개발에 대한 위탁연구를 시행하였다.

지금까지 살펴본 바와 같이 그 동안 국내의 DB 품질평가 연구는 도서관에서 구축한 DB의 평가와 비교에 치중되었지만 구축비용 대비 효과에 관한 비용/편의 분석이 없는 실정이고, 상용 DB의 도입을 위한 상용 DB의 품질평가에 대한 연구도 거의 없는 실정이다.

2. 3 PubMed와 Embase DB의 품질 비교

유동숙(1998, 3, 64 재인용)의 연구에 소개된 두 DB에 관한 품질 비교연구로 John 등은 26개의 의약정보에 관한 질문을 대상으로 하여 Medlars와 Dialog를 질적·양적으로 비교한 결과 두 DB가 질적·양적으로 비슷한 검색(성)능을 나타냈고, Dialog의 Embase가 Medline에 비해 검색비용이 높은 것만이 차이점으로 나타났다고 한다. Odile 등은 9종의 온라인 서지사항 DB를 대상으로 10개 약물의 부작용에 대하여 검색(성)능을 질적·양적으로 평가한 결과 총 참고문헌 수에 있어서는 Embase가 (Core)Medline에 비해 월등히 우수했다고 한다. 한편 Pamela 등이 자신들이

근무하는 의약정보센터에 접수된 8개의 질문에 대해 Embase와 Medline의 검색(성)능을 평가한 결과 검색된 정보의 양은 Embase가 많았으나 연관성은 Medline이 우수했고 평균 검색비용은 Embase가 20배나 높은 것으로 나타났다. 또한 Marine 등이 약물상호작용을 검색대상으로 연구한 결과는 Embase가 자료량은 방대하지만 비용면이나 연관성면에서는 Medline이 우수하므로 Medline을 첫 번째로 추천하고 있었다.

그 밖에 Suarez-Almazor 등 외국의 많은 연구자와 도서관 사서들이 두 DB의 특성에 대해 Dialog 온라인 서비스 등을 활용해 비교 평가를 시도하였는데, 주로 PubMed(Medline)과 Embase DB의 중복률과 포괄성 및 두 DB가 검색하기에 더 우수한 주제 분야에 관한 사례적 연구가 많으며, 주요한 연구결과를 소개하여 정리하면 <표 1>과 같다.

한편, 국내 연구 중 유동숙(1998)은 Medline, Embase, IDIS, IPA CD-ROM DB를 대상으로 약물에 관한 정보에 국한하여 포괄성, 연관성, 상대적 검색(성)능, 문제해결 가능성, (원문)입수 가능성, 구독비용 대비 효용성 등 6가지 평가기준으로 나누어 부작용, 안정성, 약물상호작용, 중독, 최기성, 약동학의 6가지 주제군에 대해 총 110개의 검색주제로 검색한 11,184건의 참고문헌을 조사하여 비교 분석하였다. 연구결과 Medline은 상대적으로 연관성, (원문)입수 가능성, 구독비용 대비 효용성이 우수했고, Embase는 포괄성, 상대적 검색(성)능, 문제해결 가능성이 상대적으로 우수했다. 전반적으로 검색된 총 참고문헌 수, 연관 가능성 있는 참고문헌 수, 상대적 검색(성)능, 문

〈표 1〉 PubMed와 Embase DB를 비교한 외국 연구결과 정리표

발표연도	연구자	주요 연구 내용과 결과
1987	Collins 외	<ul style="list-style-type: none"> -Embase 온라인 사용기관 수가 유럽 78.6%, 미국 87% -Embase가 총 레코드 수는 비슷하지만 수록 저널 수와 초록의 양에서 많음 -이용자 8명 중 Embase 검색비가 높아 7명은 지불의사 없고, 2명은 포괄성(원문 입수 가능성 낮음) 위해 지불 용의 있고, 소급검색의 유용성도 7명만 선호해 결국, 사서와 이용자 모두 Medline 선호 -Medline은 질의어 작성이 단순하고 주제어 선정을 위한 색인시스템과 MeSH 사용의 편리성과 1-6개월 정도 최신성이 우수
1988	Parfrey	<ul style="list-style-type: none"> -Embase는 년간 40여만 건 논문 검토해 30여만 건 수록해 3,500여 저널 필터링 효과로 저널 구독비 절감이 가능하고 Medline과 15~40%의 중복률 -Embase는 저널 입수 후 30일내 색인이 되고 제약에 보다 전문적인 색인이 사용(Emtree가 효과적)과 CAS 등록번호로 약품 상품명과 제품명 정보 우수
1990	Yonker 외	<ul style="list-style-type: none"> -Ulrich's와 Index Medicus의 법의학 41종 저널 중 전문가가 선정한 5종 저널의 82~93년도 152개 논문의 2,061개 인용문현에서 78년 이후 499개 인용문현이 수록된 188종 저널 중 5종 저널의 104개 논문을 분석한 결과, Embase는 91%를 Medline(일부 저널 선별 수록)은 73%를 수록
1993	Gunn	<ul style="list-style-type: none"> -Embase는 주제의 특수성(제약, 독성학, 환경문제, 법의학, 대체의학)이 있지만 약 1/3 이 Medline과 중복되고 저널 입수 후 15~20일 뒤 DB에 수록 -Embase는 유럽지역 정보가 많이 포함되어 110개국 70%가 영어자료
1998	Brown	<ul style="list-style-type: none"> -Embase는 제약 분야(EMTAGS와 약품의 부작용, 상품명과 제품명 등)와 유럽 및 동아시아 저널 정보에서 우수 -Embase는 검색질의 10개 중 9개가 더 많이 검색되지만 Dialog 검색비용이 Medline 보다 3배정도 비싸므로 Dialog Web 기능을 이용해 비용 절감 가능
	Woods 외	<ul style="list-style-type: none"> -Embase는 제약 및 치료 분야에서 Medline에 충분한 정보가 없거나 포괄적 검색시 검색해야 하지만 검색결과 식별에 많은 시간 소요(중복 저널 34%) -21개 검색에서 Embase는 20개 적합한 결과 중 14개의 유일한 정보가 Medline은 17 개의 적합한 결과 중 2개의 유일한 정보가 포함
1999	McDonald 외	<ul style="list-style-type: none"> -Ulrich's 977종 정신의학 저널을 수록한 213종 초록/색인지를 4대 주요 DB로 비교한 결과 Embase는 67%를 Medline은 47%를 포함
	Hallam 외	<ul style="list-style-type: none"> -전체 색인정보의 45%를 차지하는 제약 분야에서 Embase는 Medline 보다 2배의 색인량과 유럽지역 연구자와 간호사가 유럽 정보를 상세히 수록해 선호 -Embase는 검색자의 70%가 수록범위(포괄성) 때문에 Medline 보다 더 선호하고 75%가 검색결과에도 만족하고, 59%가 Medline과 동일 문현을 검색하지만, 63%가 타 DB와 중복사용하고, 검색 후 도서관 이용하는 경향
	Watson 외	<ul style="list-style-type: none"> -RCTs of CBT for depression 분야 전문가의 수작업 정확율은 94%이지만 DB 이용한 검색에 Embase와 Medline은 표준색인어 검색에서 68% : 84%를 전문적 검색에서 76% : 97%로 Medline이 2개 검색에서 모두 우수
2000	Minozzi 외	<ul style="list-style-type: none"> -사회복귀 분야 90~97년 논문을 Dialog의 MeSH와 Emtree로 검색한 결과 Embase는 35%의 적합율을 Medline은 32%의 적합율로 중복률은 17%이므로 포괄적 조사에 2개 DB 모두 필요하고, 복잡한 검색어는 정보전문가(사서)와 임상의의 협조가 필수적
	Suarez-Almazor 외	<ul style="list-style-type: none"> -통제된 임상실험 분야 88~94년까지 논문을 96년 OVID DB에서 3개 주제 4,111건의 참고문현(83%가 영어자료)을 실험한 결과, Embase가 Medline보다 78%나 많은 문현이 검색되고, 30%가 두 DB에서 공통으로 검색되며, 비영어 자료의 색인량이 많아 임상의학에서는 Medline만으로 불충분하여 Embase와 같은 다른 DB와 수작업 검색을 병행해야 완벽한 검색이 가능

제해결 가능성은 Embase가 Medline 등 다른 DB 보다 월등히 높아서 가장 자료량이 방대함을 알 수 있고, 특히 한국에서 사용되나 미국에서는 사용되지 않는 약물에 대해서는 모든 검색 주제군에서 Embase의 자료량이 가장 많았지만, 비용대비 효용성면에서는 거의 대부분의 경우 Embase가 다른 DB들에 비해 고가여서 구독비용이 비싸다는 점과 Embase는 연관성이 약 30%로 원하는 자료를 얻기 위해서 검색결과를 살펴보는 과정에서 많은 시간과 노력을 들여야 하는 단점이 있었다.

또한 이춘실과 이은숙은 국내 의학 분야 연구업적 평가에 이용가능 한 DB의 포괄성에 대해 우리나라 연구자가 쓴 1990년~1995년 사이 국내외 저널의 논문을 검색하여 분석한 결과, 의학 분야 연구업적 평가를 위해 1종의 DB를 사용한다면 SciSearch가 적합하며, 2종을 사용한다면 SciSearch와 이를 보완할 Embase가 Medline보다 적합하다고 분석하였다(이은숙 1997; 이춘실, 이은숙 1998).

3. PubMed와 Embase DB 품질 평가 기준과 방법 설정

이하에서는 의학 분야에서 높은 사용빈도를 보이는 무료 PubMed와 유료 Embase의 현황을 알아보고, 2개 DB의 상호 비교를 위한 본 연구의 품질평가 기준과 방법 설정에 대해 알아본다.

3. 1 PubMed와 Embase DB의 현황

PubMed(<http://www.ncbi.nlm.nih.gov/>

[entrez/query.fcgi?db=PubMed](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed))는 미국 NIH(National Institutes of Health) 산하의 NLM(National Library of Medicine)에서 개발되어 NCBI(National Center for Biotechnology Information)의 Entrez 검색시스템에서 운영되는 대표적인 생의학 분야 서지 DB이며, NLM에서는 1997년 6월 26일부터 Medline을 무료로 개방하여 PubMed를 웹에서 무료로 IGM(Internet Grateful Med)과 함께 2 가지의 웹 서비스를 제공하고 있다. PubMed에 참여하는 출판사는 출판 이전에 전자적으로 NCBI에서 서지데이터(제목, 저자, 수록정보원, 초록 등)를 제출하면 전자저널 원문과 직접 연결되거나 LinkOut으로 표시할 수 있다. PubMed에는 Medline 데이터뿐만 아니라 색인하기로 결정하기 전의 저널과 일부 일반과학 저널과 NLM에 의해 질적인 평가를 거친 PMC(PubMed Central) 데이터도 수록되고 있다. 이 때문에 Medline에 수록되기 이전의 데이터도 PubMed를 통해서는 검색이 가능하다는 중요한 장점을 가진다.

Medline 데이터를 포함한 PubMed는 미국 외 70여국가의 4,600여종 생의학 저널을 수록 하며, 1,400여만 건의 서지정보를 66년부터(그 이전 데이터는 NLM's OldMedline DB 이용 가능) 수록하며, NLM의 MeSH Terms을 포함하며, 매일 갱신되며, 협력출판사에서 받은 데이터는 먼저 “in-process” 상태에서 서비스되다가 정식으로 색인 되면 “indexed for Medline” 상태로 변경된다(<http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html>).

Embase DB는 책자 형태의 Excerpta

Medica가 DB화 되어 Web DB 형태로 제공되며, 수록 출판물의 자료형태는 Article, Clinical Trial, Conference Paper, Erratum, Letter, Meta Analysis, Note, Randomized Controlled Trial, Review, Short Survey 등으로 PubMed보다 다양하다. 국내에서 한국의학도서관협의회를 통한 Embase.com의 공동구매는 2002년 1월 1일부터 시작하여 2002년에는 6개 기관(울산의대, 연세의대, 식품의약품안정청, 한국생명공학연구원, 삼양사, LG CI), 2003년에는 3개 기관(울산의대, 연세의대, 한국생명공학연구원), 2004년에는 4개 기관(울산의대, 식품의약품안정청, 건강보험심사평가원, 삼성제일병원)이 구매하여 3년간 계속 구독하는 기관은 울산의대뿐인 실정으로 구매기관이 줄어드는 현상을 보이고 있다(Elsvier Korea의 E-mail 자료 2004). 2004년도에는 국내에 Embase 구독대행사가 신규로 설정되고 판매정책도 조정되어 도서관 이용자 규모에 따라 \$3,000, \$4,500, \$6,000 등 3개 그룹으로 판매되고 있다.

외국의 연구에 의하면(김기홍 2000, 5-6 챕터 인용) Medline 이용자의 사용빈도는 64%로 민족도는 66%와 69%로 보고 되었다. 따라서 Medline의 Web 버전이라고 할 수 있는 PubMed는 전 세계적으로 매월 4천6백만 건 정도의 검색이 발생하여 의학 분야의 가장 기초적이면서 대표적인 문헌정보 DB이자 Web DB라고 할 수 있다. 또한 캘리포니아대학교(UCB) Melvyl 웹 목록의 이용을 분석한 결과 16개월 동안 7.4백만 건의 DB 선택 중 Medlars가 22%로 32%를 차지한 목록 DB에 이어 2위를 차지하며, 잡지와 정기간행물(18%)

은 물론 다른 과학기술 분야 Inspec(4%)과 Current Contents(4%) DB 보다도 월등히 높은 이용율을 보였다(Cooper 2001). 그리고 Medline은 북미지역에서 가장 널리 이용되지만, 유럽에서는 Embase가 자주 사용된다고 한다(Suarez-Almazor et al 2000, 47).

국내 의과대학 도서관, 병원 도서관, 의학 학회 등 146개 의학 관련 기관의 웹사이트를 분석한 결과 PubMed는 71개 기관(48.6%)에서 소개되어 있고, 미국과 스페인과 일본 및 독일의 99개 의학 관련 기관 중 57개(57.6%) 기관이 PubMed의 링크를 제공하여 국내외 총 245개 웹사이트 중 128개 기관(52.2%)이 링크하고 있다. 이를 의과대학 도서관으로 한정시키면 국내는 41개 기관 중 31개 기관(75.6%)이며, 외국은 35개 기관 중 31개 기관(88.6%)으로 PubMed는 국내외 의과대학 도서관의 평균 82% 이상의 사용 가능성을 보이고 있었다(김기호 2003, 32, 66).

이러한 PubMed와 Embase DB의 특성을 비교하여 정리하면 다음 <표 2>와 같고 PubMed와 Medline과 Embase의 관계를 그림으로 나타내면 <그림 1>과 같다.

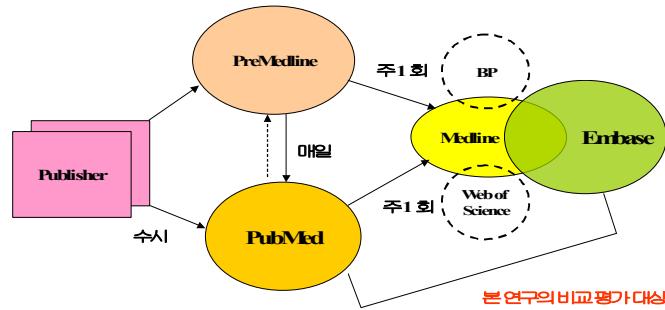
3. 2 PubMed와 Embase DB의 품질평가 기준을 위한 주요 연구 검토

국내 주요 선행연구의 내용을 중심으로 DB 평가 기준과 평가 방법을 종합하여 본 연구에 적용할 품질평가 기준을 설정하기 위해 다음처럼 주요 연구의 검토 결과를 정리해 볼 수 있다.

장혜란은 국내(이국희) 연구와 국외 4개(Finnish Society, SCOUG/CIQM, Harry

〈표 2〉 PubMed와 Embase DB 특성 비교

특성	PubMed	Embase
제작 기관	NLM	Elsevier Science
구독료	무료	FTE 별 1년 \$3,000/\$4,500/\$6,000
인쇄 색인지	Index Medicus, Index to Dental Literature, International Nursing Index	Excerpta Medica(42개 세부분야별)
검색 엔진	Entriz	EMBASE.com
갱신 빈도	매일	주 1회
자료 수록 기간	1966년 - 현재 (1953-1965 OldMedline)	1974년 - 현재 (1966년부터 Medline 포함)
수록 레코드 수	14,796,933건 (연간 56만건)	Embase 만 9,957,374건 (Medline 포함시 1,586만건)
색인 연속 간행물 수	70개국 4,600여종	70개국 4,600여종 (Medline 포함시 5,500여종)
저널 비율	북미 47% : 유럽 43%	북미 33% : 유럽 56%
초록 보유율	85 이후 69% 이상	80% 이상 저자 초록 보유
수록 및 색인	출판사 통보 후 목차는 5일내 (색인은 6개월 내 수록)	인쇄저널 도착 후 2주내 색인 (최대 5주내 수록)
원문 Link	KESLI 컨소시엄인 Elsevier와 Springer 출판사 등 1,324개 출판사	Elsevier, Ingenta, Springer 출판사 등
색인대상 주제	약물에 관한 자료 외에도 생명과학, 정신과학, 간호학, 치의학, 수의학, 마취과학 등 생물의학적(biomedical) 분야 전반에 걸친 자료가 수록되고 AIDSLINE, HealthSTAR, POPLINE, Toxline도 포함	생물, 의학, 제약학, 독성학을 주 대상 (화학 및 공학 분야는 PubMed보다 많이 포함, 이은숙 1997, 59) evidence-based medicine과 dietary supplement research에 우수
연속간행물 선정	LSTRC(Literature Selection Technical Review Committee)를 운영함으로써 엄격하게 논문 채택과정을 관리하고, 위원회가 논문의 목적, 신뢰성, 내용과 출판형태의 질적 수준까지 평가하여 채택	편집회의에서 심사하여 채택하고, 대부분 peer review 과정이 있는 연속간행물을 채택하여 conference의 초록을 제외하고는 저널에 수록된 모든 논문을 수록하는 원칙
서비스 플랫폼	BMN(BioMed Net), Biosis BP, Ebsco, Embase, HighWire, IdDB3, IGM(Internet Grateful Med), OCLC First Search, OVID, PML(ProQuest Medical Library), SciFinder, Web of Science, Dialog 등	Datastar, Dialog, DIMDI, Lexis-Nexis, OVID, STN
기타 특징	Single/Batch Citation Matcher, Tutorial, Clinical Querie, Journal DB 운영 LinkOut은 소장정보/전자저널과 연결가능한 1,324개 출판사 4,280종 저널 중 현재 179개 출판사 저널 연결 표시 가능 PubMedCentral등 무료데이터도 수록 MeSH Terms(2.2만 주요 개념과 12.5만 참조어) Cubby에 의한 Alert	Boolean, wildcard/truncation, field delimitation, subject/document limits Emtree Keywords(4.8만 descriptor, 20만 동의어) Quick/Advanced/Drug/Disease/Article Search Profile에 의한 Alert



〈그림 1〉 PubMed, Medline, Embase의 상호 관계

& Oppenheim, HELEN) 연구 등 5개 주요
선행연구를 분석하여 공통적인 사항을 포괄성,
완전성(색인 포함), 일관성, 최신성, 정확성으
로 정리하고, 이를 평가지표간의 중요성에 따
른 가중치는 6:5:4:3:2로 부여하여 지표별 평
가 척도를 만들었다(장혜란 2000, 88, 90-93).
이 평가모델은 평가지표간의 상대적 중요성을
반영한 결과 다양한 DB간의 평가결과 순위화
를 가능하게 하여 비교 평가를 손쉽게 했다는
장점이 있지만, 데이터 품질 영역에만 국한되
어 있고 실제로 평가를 위해 적용한 상태가 아
니므로 이 평가모델을 이용하여 실제 평가를
통해 설정된 평가모델의 유효성을 확인할 필요
가 절실하다고 하겠다.

한국데이터베이스진흥센터(2000, 20)는 1995년도에 수행한 DB 품질평가에 관한 연구를 2000년에 재 수행하면서 17개 국내외 선행연구를 분석하여 새로운 DB 품질평가 기준을 개발하였다. 이 연구는 국내 DB 품질평가 후속연구의 기본이 되면서 DB 품질평가의 기준에 대한 대강의 핵심의 도출의 기반이 되었다고 본다.

이용봉 등은 DB 평가에 관한 국내외 17개
선행연구를 종합 정리하여 DB 품질을 위해 데

이터 품질과 서비스 품질이라는 품질기준을 추출하였다(이용봉 외 2001, 114-115).

이제환은 KERIS 종합목록 DB의 품질평가를 위해 선행연구를 요약하여 품질평가 기준을 만들었고, 이 품질평가 기준에 따라 KERIS 종합목록 DB의 품질평가를 위해 상대적 유용성(포괄성, 배타성, 최신성)과 절대적 유용성(중복성, 일관성, 완전성)으로 나누어 평가지표를 만들었다(이제환 2002, 66, 69).

KERIS(조순영 등 2002)는 평가지표에 따라 전문가들이 평가한 값의 상대적 중요성을 AHP 기법을 이용한 프로그램으로 배분하여 DB의 종류에 관계없이 서비스 품질은 물론 비용과 구매요건을 포함하여 범용적으로 적용할 수 있는 모델을 개발하였다. KERIS의 평가지표를 프로그램을 이용해 전문가와 이용자들이 같이 평가한다면 이제까지의 다른 방법보다 더 객관적인 평가와 순위화가 가능할 것이다.

국내의 DB 품질평가에 관한 주요 연구 중 안계성과 함정훈의 한국데이터베이스진흥센터(2000, 20), 조순영 등의 한국교육학술정보원(2002, 30), 이응봉 등(2001, 115)의 19개 기준 평가, 11개 선행연구를 분석한 이제환(2002,

〈표 3〉 국내 주요 선행연구의 DB 품질평가 기준 정리표

구분	품질기준 (Criteria)	17개 선행연구 적용			이용봉 외 연구(2001)의 관련 항목	이재환(2002)과 김성희(2003) 연구의 관련 항목	이만수 외 연구(2003)의 관련 항목
		DPC	KERIS	이용봉			
데이터 품질	최신성 Currentness	15개	14개	14개	-데이터 갱신 빈도	-현행성 -신속성 -갱신주기 -타임 레그	-최근 데이터의 갱신 -신속한 갱신
	정확성 Accuracy	14개	13개	13개	-레코드 내 오류 데이터의 양 -레코드 작성 기준 유무 완전 성에도 포함됨)	-데이터의 신뢰성 -데이터 표현 정확성 -데이터 기술 정확성 (철자 오류, 누락 등)	-수록 데이터의 정확성과 일치 성 -데이터 표현의 정확성 -철자 오류와 누락
	완전성 Completeness	8개	11개	11개	-레코드 필드구조의 완전성과 일관성 -공표된 레코드와 실제 레코드 수의 차이	-레코드 구조 적합성 -데이터 기술 완전성 -레코드의 고유성 비중복성	-레코드 구조의 적합성 -데이터 기술의 완전성 -자료속성의 완전 표현 -고유 레코드(비중복)
	포괄성 Coverage 또는 수록범위 Scope	8개	9개	9개	-기타(수록범위)	-수록 자료의 범위 -수록 레코드의 수 -유형과 시기 망라성 -질의에 대한 응답율	-자료 범위/종류 다양 -수록 레코드의 수 -데이터의 유형과 시기 -가치있는 자료 수록량
	일관성 Consistency	8개	8개	8개	* 일관성 대신 전문성(Expert) 항목만 별도로 품질기준에 표 시	-레코드 구조 일관성 -데이터 표현 일관성 -데이터 기술 통일성	-레코드 구조의 일관성 -데이터 표현과 기술의 통일성
서비스 품질	편의성(사용 용이성) Ease of Use	7개	8개	8개	-메뉴검색 체계성/다양성 -키워드검색 용이성/다양성 -디자인 이용자 친화성 -검색결과 표시의 내용 측면별 다양성	-인터페이스의 편의성 -작업처리의 편의성 -접속의 편의성(* 김성희의 연구에는 이 항목만 제외)	-인터페이스의 편리성 -접속의 편리성 -결과의 순서운 활용
	지원성 (사용자 지원) Customer support	9개	7개	7개	-오픈 데이터 수정의 체계성 -DB 이용안내 제공 -Help 메뉴 제공 -Q & A 제공	-이용자 지원 및 교육 -도큐멘테이션의 유무 -도움(Help) 기능의 적절성 -원문제공의 편의성	-사용자 지원 및 교육 -도큐멘테이션, 도움말 제공 -원문접근의 용이성
	검색성 Searchability 또는 접근성 Searching	7개	5개	6개	-시소스스 관리 우수성 -다양한 탐색기법 제공 -복수 DB 검색 가능성 -검색식 1건의 평균 소요 시간 -검색결과 출력 다양성	-검색속도 -검색방법의 다양성 -출력기능의 다양성 -시소스스의 유무	-검색속도 -검색방법의 다양성 -출력기능의 다양성 -시소스스의 유무
기타	비용 Cost	7개	8개	8개	* 품질기준으로 표시되었으나 평가항목에 제외	-구입비용 -이용비용	-DB 이용비용의 적절성
	네트워크/하드웨어 Network & H/W	2개	5개	5개	* 품질기준으로 표시되었으나 평가항목에 제외	-접근성 -접근속도 -접속방식	
기타	일반적인 요구사항 General facts desired	11개	9개	10개	* 품질기준으로는 표시 없음	-권위(전문성) -통합성 -유일성 -객관성 -이용자 만족도 -데이터작성지침 준수	

* DPC(안계성, 함정훈 2000)의 기타는 통합성과 내용과 권리 포함하고, 자체제작 비율, 다큐멘테이션, 가치, 유일성, 마케팅, 객관성 등

** KERIS(조순영 외 2002)의 기타에는 자체제작 비율, 이용자 호응도, 신속성, 유일성 등

66), 4개 선행연구를 분석한 이만수 등(2003, 27)의 연구에 나타난 DB 품질평가 기준을 정리해보면 다음 <표 3>과 같이 나타낼 수 있다. 그러나 먼저 수행된 한국데이터베이스진흥센터 와 나중에 수행된 이웅봉과 한국교육학술정보 원의 동일한 17개 선행연구를 분석하여 표시한 품질기준 해당 선행연구의 건수가 각각 다르다는 사실이 품질 측정의 어려움을 단적으로 나타내주고 있다. 또한 품질평가 기준에 필요한 항목에는 대체적으로 합의가 이루어졌지만 세부 측정방안에 대해서는 서로 다르다.

3. 3 PubMed와 Embase DB의 품질평가를 위한 평가기준과 항목

본 연구에서는 국내 연구자의 선행연구를 기반으로 각 평가항목 중 쉽게 객관적으로 평가할 수 있는 항목을 중심으로 하여 Web 환경에서 제공되는 의학 분야 서지정보 DB인 무료 PubMed와 유료 Embase DB의 품질 비교 평가를 사례중심으로 발견적 평가를 통해 시도하고자 한다.

두 DB의 비교 평가를 위한 검색은 2004년 4월 14일부터 4월 16일 사이에 집중적으로 실시되었고, 보완을 위한 검색이 5월 18일부터 5월 25일 사이에 실시되었으며, 측정방법은 다음의 평가기준을 비교하기 위해 쉽게 확인할 수 있는 내용으로 한정했다.

이러한 평가기준의 종류와 정의는 본 연구의 평가기준을 위해 근거대상으로 삼은 <표 3>의 3~4개 선행연구와 동일하므로 서술은 생략하고 측정한 항목은 다음과 같다.

- 1) 최신성 : 2004년도 레코드 수, 발행빈도 가 빈번한 저널의 최근호 수록 현황
- 2) 정확성 : 입력 데이터의 오류 사례, 오류데이터의 양(영국식과 미국식 영어, 철자오류)
- 3) 완전성 : 공표된 레코드 수와 실제 레코드 현황, 레코드의 고유성이나 중복률
- 4) 포괄성 또는 수록범위 : 전체/년도별 레코드 수, 논문/초록/언어별 구성 비율
- 5) 일관성 : 저널명의 표기 방법의 통일성, 레코드 구조의 일관성(대소문자 구분, 색인어의 일관성, 띄어쓰기)
- 6) 편의성 또는 사용 용이성 : 중복데이터 제거, 시소리스 규모, 검색결과 활용의 편리성
- 7) 지원성 또는 사용자 지원성 : 전자저널 원문링크, 원문복사와의 연계
- 8) 검색성 또는 접근성 : 검색속도, 시소리스 유무, 최신 레코드의 초록/영어자료 보유 현황
- 9) 비용 : 구입비용과 상대적 타당성
- 10) 네트워크/하드웨어 : 인터넷 외 추가 소요 사항

4. PubMed와 Embase DB 품질의 비교 평가

출판사들이 데이터를 제공한 PubMed 데이터 중 NLM의 정책에 따라 선별된 저널의 논문만 Medline에 수록되므로 종전의 Medline과 무료로 Web에서 제공되는 PubMed는 데이터 선정 정책과 데이터의 양에서도 차이가

나고, Medline 데이터가 Embase에 통합되어 서비스 되면서 두 DB에 대한 선행연구의 품질 평가 결과와 본 연구의 품질평가 결과는 크게 차이가 날 수 있다.

4. 1 최신성(Currentness)

PubMed와 Embase DB 수록 레코드의 최신성을 알아보기 위해 각 DB의 2004년도 레코드 수를 검색해 본 결과 PubMed에서는 2004년도 레코드가 164,658건인데 비하여 Embase에서는 114,690건으로 총 레코드 수 와는 다르게 PubMed에서 더 많은 2004년도 자료의 레코드가 검색되어 Embase는 PubMed의 70% 미만이었다. 이는 Embase의 갱신주기가 검색결과에 영향을 미칠 수 있음을 감안하더라도 출판사 협력관계의 강화로 PubMed의 데이터 갱신주기가 매일이어서 갱신주기가 1주일인 Embase보다 신속(심지어는 출판사 사이트에서 본격적인 서비스 이전이라도 PubMed에서 검색가능 한 경우가 있어 출판사측 보다도 더 신속한 서비스가 되는 경우가 있음)하다고 할 수 있어서 최신성 측면에서 더 우수하다고 판단된다.

또한 Embase에는 Medline 데이터가 포함되어 중복레코드를 제거한채 서비스된다고 알려져 있지만 2004년도 데이터만 살펴보면, Medline 데이터가 포함된 Embase 데이터는 114,690건이고 Embase만의 단독 데이터는 104,354건으로 PubMed 레코드 수와 높은 중복률을 감안하면 PubMed 레코드 수에 비해 턱없이 부족한데, 그 이유는 Embase내 Medline 데이터의 갱신이 늦어지거나 Em-

base 자체 데이터와의 중복 레코드 제거 작업에 시간이 많이 소요되는 것으로 판단된다.

실제 PubMed에서는 “2004[dp] AND 2003[dp]”로 검색시 “Arch Women Ment Health. 2004 Apr;7(2):123-31. Epub 2003 Dec 15”와 같이 2003년도에 논문제재 허가가 나서 전자저널로 미리 서비스하는 논문인 경우 그 당시에도 서비스하지만 몇 달 뒤 정식 인쇄 저널이 발행되면 이 정보를 기준으로 데이터를 정비하게 되어 전자저널 발행년도와 인쇄저널 발행년도가 혼재하는 현상(이처럼 발행년도가 2개 년도에 걸쳐 혼재된 레코드가 총 20,270 건)이 발생한다. 그래서인지 각 년도별 레코드 수의 합계와 년도에 대한 범위를 주어서 검색한 레코드 수의 결과가 불일치하는 현상이 발생하고 있었다.

마지막으로 두 DB에 실록되어 있는 저널의 최신성을 시범적으로 비교해보기 위해 1년에 52개 이슈가 발행되는 Elsevier 출판사의 BBRC(Biochemical and Biophysical Research Communications)라는 저널의 가장 최근 이슈를 살펴보면, PubMed에는 317(3)권이 가장 최근 데이터로 MeSH 등이 부여되지 않은채로 서비스 되며, Elsevier 출판사 계열사의 제품인 Embase에는 색인은 되었지만 오히려 316(4)권이 가장 최신 정보여서, 조사대상 저널의 경우 PubMed와 Embase 사이에는 저널의 논문을 서비스 대상으로 수록하는 시점에서 약 3주 정도의 차이가 발생하고 있었고, 조사대상 저널의 원 출판사인 Elsevier 가 317(4)권을 현재 서비스하고 있어서 두 DB보다 신속하였다. 또한 Nature 저널의 경우에도 PubMed에는 428(6984)권이 가장 최

근이지만 Embase에서는 428(6981)권이 가장 최근이어서 두 DB의 최신성에는 약 3주의 시 간적 차이가 발생하였다.

4. 2 정확성(Accuracy)

레코드 내 오류나 데이터의 정확성 측면에 서 발견된 사실은 PubMed의 경우 출판사 측에서 데이터 수집에 협조를 받는 관계로 Clinical Radiology 저널의 경우 출판사 측의 작성 오류가 PubMed의 검증과정 없이 그대로 반영되어 데이터 12건의 발행년도가 2005년도로 검색되고 출판사 측에도 동일하게 2005년으로 잘못 표기되어 서비스되는 현상을 발견하였다.

영국식과 미국식 영어 구분을 위해 Color와

Labor 및 Encyclopedia를 검색해 본 결과 PubMed에서는 Color와 Labor에 대해서는 어휘통제가 되어 동일한 결과가 나왔지만 Embase는 전혀 통제되지 않았고, PubMed에서도 Encyclopedia에 대해서는 어휘통제가 되지 않아 PubMed는 부분적으로 어휘통제를 하고 있음이 밝혀졌다.

또한 철자 예러 정도를 확인하기 위해 February, Guatamala, Misssion, Goverment, Fransisco, Grammer, Recieve, Wensday, Seperate, Conditons, Bussiness, Infomation, Retrieval를 검색해 본 결과(Grammer는 대부분 인명이 포함되어서 이를 제외) Embase에는 모두 641건의 예러가 PubMed에는 모두 276건의 예러가 발견되어 Embase의 철자 예러율이 <표 4>처럼 높았다.

〈표 4〉 정확성 측정 결과

항 목	검색어	PubMed	Embase
영국식과 미국식 영어 구분	Color	65,085	63,493
	Colour		14,623
	Labor	83,024	69,591
	Labour		16,736
	Encyclopedia	292	250
	Encyclopaedia	33	46
철자 예러	Febuary	5	13
	Guatamala	4	10
	Misssion	0	0
	Goverment	21	57
	Fransisco	26	90
	Grammer	(354)7	(384)11
	Recieve	20	51
	Wensday	0	0
	Seperate	46	131
	Conditons	129	237
	Bussiness	0	0
	Infomation	12	31
	Retrival	6	10
	소 계	276	641

4. 3 완전성(Completeness)

공표된 레코드와 실제 레코드 수의 차이를 홈페이지 등의 자료와 실제 검색결과를 비교해서 조사해보면 두 DB 모두 공표된 레코드가 실제 레코드 수에 비해 46만~80만 여건이나 부족해 1년 수록분 이상의 차이가 나 홈페이지 등의 자료 갱신이 1년 이상 지연되고 있는 실정이다. 또한 공표된 자료에 의하면 PubMed에는 1966년 이후 Medline 기반 자료와 1953~1965년의 OldMedline 자료가 혼재되어 있다고 표시되었지만 실제 데이터는 1902년 데이터부터 존재하며, OldMedline 수록범위 이전인 1902~1952년까지의 자료가 무려 249,190 건이나 있었다. Embase도 공표된 자료에 의하면 1974년 이후부터 수록되었다고 하지만 실제로는 1967~1973년 사이의 자료가 200,833건이나 존재했다. 두 DB 모두 공표된 자료의 수록범위와 실제 검색결과 사이에 불일치가 존재하고 있다. 이러한 현상으로 두 DB 모두에 Medline 데이터만 있어야 할 1966~1973년 사이의 레코드 건수가 Embase가 PubMed보다 114,628건이나 더 많은데, 그 이유는

Embase에는 1973년 이전에 발행된 레코드도 자체적으로 수록하고 있기 때문이었다.

레코드의 고유성을 평가하기 위해서 PubMed와 Embase의 중복률을 조사해보면, 총 레코드 수에서는 중복률이 70% 정도이고 Embase에만 수록된 레코드의 비율은 30% 정도로 최근 연도로 올수록 중복 비율이 높아 80% 가량이 되어서 최근에는 두 DB 간에 수록 정보원의 차별성을 찾기가 어렵게 되었다. 또한 PubMed에서는 그러하지 않지만 Embase에서는 Embase에 없는 데이터를 Medline 데이터로 보충하면서 데이터 중복체크가 완벽하지 않아 일부 중복된 데이터를 포함하는 경우가 발생하고 있었다.

레코드의 필드구조를 평가하기 위해 두 DB의 필드명과 내용을 보면, 두 DB에 공통인 필드는 필드명이 다르더라도 내용이 같은 필드는 Author, Address, Abstract, Title, Language, Country of Author, Date of Entry, ID Number, ISSN, Number of References, Publication Type, Translated Title, Volume Issue 등이고, PubMed에만 있는 필드는 Named Person,

〈표 5〉 PubMed와 Embase DB의 수록년도별 중복 및 유일 레코드 현황

년도	PubMed 단독	Embase+ Medline	Embase 단독	2개 DB 중복 건수	Embase 중복률	Embase 유일레코드	Embase 비중복률
2004년만	164,658	114,690	104,354				
2000~2003	2,142,307	2,485,659	1,858,833	1,515,481	81.5%	343,352	18.5%
1997~2003	3,514,431	4,189,400	3,110,537	2,435,568	78.3%	674,969	21.7%
1985~2003	8,202,975	9,867,688	6,981,548	5,316,835	76.2%	1,664,713	23.8%
1974~2003	11,160,132	13,977,732	9,652,187	6,834,587	70.8%	2,817,600	29.2%
1966~2003	12,818,447	15,750,675	9,837,253	6,914,601	70.3%	2,922,652	29.7%
1966 이전	1,813,828						
전체 레코드	14,796,933	15,865,365	9,957,374	6,914,601	69.4%	2,922,652	29.4%

〈표 6〉 PubMed와 Embase DB의 수록년도별 논문, 초록, 영어, 영어초록 현황

년도	저널논문 건수 및 비율				초록보유 건수 및 비율			
	PubMed	비율	Embase	비율	PubMed	비율	Embase	비율
2004년만	155,167	94.2%	80,525	70.2%	131,847	80.1%	90,348	78.8%
2000~2004	2,101,140	91.1%	1,842,181	70.8%	1,803,605	78.2%	1,969,264	75.7%
1997~2004	3,343,183	90.9%	3,069,050	71.3%	2,841,527	77.2%	3,225,667	74.9%
1985~2004	7,657,537	91.5%	7,173,602	71.9%	5,983,350	71.5%	6,654,520	66.7%
1974~2004	10,492,206	92.6%	9,928,883	70.5%	7,228,419	63.8%	8,506,890	60.4%
1966~2004	12,147,856	93.6%	11,540,249	72.7%	7,233,116	55.7%	8,621,322	54.3%
전체 레코드	13,961,604	94.4%	11,540,249	72.7%	7,233,514	48.9%	8,621,322	54.3%
년도	영어자료 건수 및 비율				영어와 초록 건수 및 비율			
	PubMed	비율	Embase	비율	PubMed	비율	Embase	비율
2004년만	155,139	94.2%	106,498	92.9%	125,894	76.5%	85,568	74.6%
2000~2004	2,077,333	90.0%	2,263,410	87.0%	1,651,111	71.6%	1,754,840	67.5%
1997~2004	3,287,221	89.3%	3,706,644	86.1%	2,587,133	70.3%	2,842,996	66.1%
1985~2004	7,153,520	85.5%	8,154,926	81.7%	5,318,258	63.6%	5,767,406	57.8%
1974~2004	9,269,377	81.9%	10,840,508	76.9%	6,368,949	56.2%	7,200,033	51.1%
1966~2004	10,242,020	78.9%	11,861,034	74.8%	6,372,756	49.1%	7,279,018	45.9%
전체 레코드	11,105,349	75.1%	11,861,034	74.8%	6,373,104	43.1%	7,279,018	45.9%

* 주) 본 표는 전체 경향을 볼 목적이어서 시간 경과에 의한 필드구조의 가변성으로 상이할 가능성 있음

Contract/Grant Number, Enzyme Commission Number, Mission Name, Citation Owner, Record Creation Date 등이고, Embase에만 있는 필드는 Medical Device Trade and Manufacturer Names 등으로 Embase에는 약품정보와 관련된 독특한 필드명이 여럿 보였다.

4. 4 포괄성(Coverage) 또는 수록범위(Scope)

DB에 수록된 레코드의 총 건수는 PubMed의 14,796,933건에 비해 Embase는 Medline 데이터를 포함해서 15,865,365건으로 수록년도의 차이를 감안하더라도 Embase가 더 많이 수록하여 포괄성이 높다고 할 수 있다.

PubMed와 Embase에 수록되는 정보원은 주로 저널이고 대상 저널들도 상당수가 중복되는 것으로 선행연구에 보고 되고 있다. PubMed의 기본이 되는 Medline과 Embase에 수록된 2003년도 수록저널을 기준으로 중복정도를 평가해보면(Elsevier 2004), Medline에는 모두 4,584종이 있지만 2,563종이 중복되고 1,937종이 Medline에만 유일한 저널인 반면, Embase에는 모두 4,514종이 있지만 2,563종이 중복되고 1,881종이 Embase에만 유일한 저널이다. 중복되는 저널의 비율은 Medline의 55.9%와 Embase의 56.8%로 56% 내외이지만 Embase의 PubMed 레코드와의 중복률은 70% 정도로 저널 중복률보다 더 높다. 그렇지만 Embase에는 Medline

〈표 7〉 일관성 측정 결과

항 목	검색어	PubMed	Embase
대소문자 구분	DATABASE	53,329	44,078
	database		
	UNION	25,652	43,998
	union		
색인어의 일관성	databases	36,042	14,168
	database	53,329	44,078
	data base	32,031	54,049
	db	15,712	19,893
띄어쓰기	cd rom	1,169	906
	cdrom	760	16
	cd-rom	1,169	837
	on line	393,518	130,999
	online	12,561	18,954
	on-line	10,814	13,540

데이터가 포함되어서도 검색이 가능하므로 중복된 레코드만 완전히 정비된다면 이제 수록 저널이나 레코드의 중복 문제는 Embase와는 크게 상관이 없는 실정이다.

4. 5 일관성(Consistency)

저널명 표기의 일관성을 조사해보면, Embase에는 Medline 데이터와의 통합과정에서 중복데이터의 제거 문제가 완벽히 해결되지 않아 Medline 데이터에서는 “Biochim Biophys Acta 2003 1619:1 (1-8)”로 나타나고 Embase 데이터에는 “Biochim. Biophys. Acta Gen. Subj. 2003 1619:1 (1-8)”로 나타나 데이터의 중복과 함께 저널명 표기의 일관성이 보이지 않는 사례가 발견되었다.

일관성 항목 중 대소문자 구분은 Embase와 PubMed 모두 구분하지 않고 있었고, 색인어의 일관성을 알아보기 위해 databases, da-

tabase, data base, db로 검색해 본 결과 전혀 어휘통제가 이루어지지 않아 일관성이 나타나지 않았다(표 7 참조).

그리고 띄어쓰기에서 Embase는 ‘on line’으로 입력하면 이것을 ‘on AND line’ 방식으로 검색하여 띄어쓰기로 검색하면 PubMed보다 적은 결과가 나왔다.

4. 6 편의성 또는 사용 용이성(Ease of Use)

PubMed의 기본이 되는 Medline 데이터는 Embase는 물론 BMN(BioMed Net), BP(Biosis Preview), SciFinder 등 관련분야의 여러 DB에 통합되어 서비스되는 관계로 같이 검색할 수 있어서 이용자의 편의성을 높여주고 있다. 물론 통합 운영하는 DB에서 자체 DB와 Medline 데이터의 중복데이터 제거 관리에 따라 통합의 효과와 효율성이 좌우될 것이다.

또한 약품 및 화학 정보는 PubMed의

MeSH에는 7천개 정도가 수록되어 있지만 Embase에는 2만개 정도가 수록되어 Embase가 약품 및 화학 정보 검색에 더 편리할 수 있다. 그러나 PubMed는 Nucleotide, Protein Sequences, Protein Structures, Complete Genomes, Taxonomy, OMIM 등과 동일하게 Entrez 검색시스템에서 운영되므로 서지정보뿐만 아니라 추가적인 정보검색이 가능하다.

PubMed의 검색결과는 Text, File, Clipboard, E-mail, Order 등으로 활용할 수 있다. 반면에 Embase는 검색결과를 E-mail과 Infotrieve라는 원문복사서비스 체제를 통해 비소장 논문을 주문하는 Order와 text 형태의 변환은 물론 EndNote, ProCite, Reference Manager 등 3개의 개인용 논문 관리 프로그램 다운로드 형식으로도 다운이 가능하여 편리하고, 웹 자원의 DB인 Scirus(<http://www.scirus.com/>)로 연결하여 검색식을 곧바로 재사용 할 수 있어서 유용하다.

4. 7 지원성 또는 사용자 지원성(Customer support)

한편 원문 링크 측면에서도 PubMed에서는 1,324개 출판사의 4,280종 저널이 링크되고 있어서 전체 수록저널의 93% 정도가 전자저널 출판사 사이트와 직접 링크되어 있지만, Embase는 제작사인 Elsevier와 Ingenta 및 Springer 등 일부 출판사 전자저널만 원문 링크 되어서 검색 후 곧바로 원문을 볼 때 자판에서 구독하는 저널일지라도 원문 링크가 일부에 그쳐 불편한 측면이 있다.

PubMed 검색결과 중 전자저널이 링크되어 있고 자판에서 해당 전자저널을 구독하고 있으면 곧바로 원문으로 연결될 수 있다. 이는 자판의 소장정보와 함께 전자저널 구독 정보를 사전에 입력하여 두면 PubMed 검색결과를 보여주는 과정에 자판 구독 전자저널 여부를 쉽게 알 수 있는 자판의 등록된 아이콘을 표시하여 원문을 링크해 주는 LinkOut 기능으로, PubMed에서 제공되는 이 기능은 Embase 보다 원문 입수와 활용 측면에서 더 편리한 기능이다.

또한 PubMed에서는 검색 후 자판 비소장 자료에 대해서는 Loansome Doc이라는 DDS 서비스와 연계되어 있어서 북미지역의 경우 손쉽게 원문복사 신청을 할 수 있어서 서비스에가입한 도서관이라면 이용자에게 더 편리한 서비스라고 하겠다. 물론 무료로 이용가능 한 전자저널의 링크 서비스도 PMC를 포함해서 풍부하다.

4. 8 검색성(Searchability) 또는 접근성(Searching)

PubMed에는 MeSH가, Embase에는 Em-tree라는 역사를 자랑하는 시소러스가 있어서 이용자의 검색성을 도와줄 수 있다. 또한 검색어 제한(Limits), 검색 진행 과정(History), SDI 서비스(Alerts) 등은 두 DB에 모두 포함되어 있어서 본질적인 차이를 찾을 수 없었다.

PubMed에서는 출판년도로 검색시 년도가 범위를 벗어나도 검색결과를 보여 주지만 Embase는 범위를 벗어난 년도는 결과를 보여주는 화면으로 넘어가지 않고 검색식 입력화면에 계속 놓이게 된다. 그리고 무료 PubMed에 비해

Embase는 검색속도 면에서도 느린 편이다.

초록이 있는 레코드 건수가 전체로는 Embase가 많지만 최신 레코드인 2004년도 데이터 중에서는 PubMed에 수록된 초록이 더 많고 초록 보유비율도 80% 정도가 되어 Embase 보다 조금 더 높다. 영어자료의 보유율은 2004년의 경우 두 DB가 93~94% 정도여서 큰 차이를 보이지 않았다.

실제 검색성능을 평가하기 위한 사례로 2003년도에 발행된 'stem cell' 자료를 찾기 위해 필드를 제한하지 않고 PubMed에서 "stem cell" AND 2003[dp]'로 검색하면 4,212건이 검색되지만 Embase에는 동일한 의미의 검색어로 "stem cell'/de AND [2003]/py'로 검색하면 2,165건이 검색되어 Embase에서 더 많은 검색건수가 기대되었지만 결과가 더 적은 것은 아마도 검색시에 기본적으로 책정된 색인 대상 필드의 상이함 때문이라 판단된다. 다시 'stem cell'이라는 검색어를 자료의 제목으로 한정시켜 검색하면 PubMed가 1,458건이고 Embase에는 1,441건이 검색되어 비슷한 결과를 가져왔고, 이를 Embase 데이터에만 제한시켜 검색하면 1,197건으로 Medline 도입으로 인해 244건(17%) 정도의 레코드가 늘어난 효과를 보여주었다.

4. 9 비용(Cost)

비용 측면에서도 물론 유료 DB인 Embase 가 무료 DB인 PubMed보다 불리하다. 2004년 현재 국내 Embase 구독기관이 4개 도서관에 그치고 있는 사실은 Embase와 비슷한 규모의 구독료를 필요로 하는 유료 OVID Medline

DB의 구독기관이 15개 도서관이나 되고 있다 는 점에 비추어보면 구독비용만이 Embase의 이용을 저해하는 요인이라고 보이지 않는다.

더구나 PubMed를 널리 사용하는 분야에서는 국내에서 널리 인지되고 많이 사용되는 PubMed 외에 PubMed와 수록 저널의 중복도가 높은 Embase 등 다른 DB를 추가로 사용해야 한다면, 굳이 PubMed 외에 다른 DB를 사용해야만 획득 가능한 유일한 정보의 효과와 비용간의 근거가 다른 상용 DB 도입의 최대 관건이라고 보여 진다.

결국 무료 PubMed 이용이 활성화된 연구 현장에서 Embase의 도입가치는 70% 정도의 Medline과 중복된 데이터에 대해서는 얼마나 부가적이고 독자적인 추가 가공을 해 부가가치를 창출했느냐와 30% 정도의 Medline과 중복되지 않는 레코드의 독자적 가치와 그 우수성 및 자판의 필요성에 달려있다고 보여진다.

4. 10 네트워크/하드웨어(Network & H/W)

두 DB는 인터넷의 Web으로 제공이 가능한 DB여서 네트워크 및 하드웨어 설치에 대한 부담은 없다. 다만 실제로 측정을 하지는 않았지만 무료 PubMed에 비해 Embase는 검색을 위한 접근성과 접근속도 및 접속방식에서 일반적으로 더 느리고 불편한 편이다.

5. 결 론

지금까지 DB 품질평가에 관한 국내외 선행 연구 결과를 중심으로 DB의 품질평가 기준을

정립하고 그에 따라 쉽게 측정이 가능한 객관적 평가를 위한 입장에서 PubMed와 Embase DB의 품질을 비교 평가해 보았다.

이를 위한 본 품질평가 연구에서는 품질평가를 실무적으로 수행할 목적으로 선행연구를 적용하기 위해서는 평가 기준의 문제가 아니라 평가 방법의 문제와 구입비용의 타당성 확보에 어려움이 있음을 알 수 있다.

그 결과 PubMed는 출판사의 협력으로 데이터 갱신주기가 매일이어서 Embase보다 신속(심지어는 출판사 사이트에서 본격적인 서비스 이전이라도 PubMed에서 검색가능한 경우가 있음)하여 최신성이 우수하고, 4,600여종의 수록저널 중 4,280종 이상의 전자저널이 출판사와 직접 링크되어 있어서 검색 후 곧바로 원문을 볼 때 편리하고, 검색속도도 빠른 편이다. 또한 PubMed는 Embase에 비해 의학 분야의 가장 기본 정보원인 저널의 논문 비중이 높고, 초록과 영어자료의 보유 비율이 높고, 다른 Entrez 검색시스템과의 연결이나 호환이 우수하고, 검색 후 자판의 전자저널 보유여부를 표시하는 LinkOut 기능과 Loansome Doc이라는 원문복사 서비스와의 연계가 우수하고, DB 구입 비용면에서 우수하였다.

그러나 PubMed의 핵심 데이터인 Medline 데이터를 타 DB를 위해 제공하는 관계로 포괄성이 상대적으로 떨어지고, 최신성에 주안점을 둔 정책상 출판사측의 균원적인 오류데이터 발생 가능성, 인쇄저널과 전자저널 발행정보의 공존으로 인한 출판년도 개념의 혼재 등의 문제점이 발견되어 이에 대한 개선이 필요하다. 또한 PubMed는 누구나 무료로 이용할 수 있어서 PubMed만 사용하게 되면 의학정보 활

용 경쟁에서 다른 연구자와의 차별성이 없고, 다학제적 연구추세에 맞추어 수록 분야의 주제 범위와 저널의 종수를 다양하게 하기 위해서 유럽과 아시아 지역 저널과 함께 PMC와 같은 무료 아카이브 저널 수록을 획기적으로 확장할 필요성이 있다고 본다.

한편 Embase는 Medline 데이터의 포함으로 인해 저널과 논문에서 포괄성이 높고, 제약 분야와 화학 분야를 비롯한 PubMed에서 빈약한 주제 분야에 대한 상대적 우수성, 수록정보의 형태가 논문뿐만 아니라 다양한 장점이 있다. 그렇지만 Embase는 최신성이 떨어지고, 최신 Medline 데이터와 통합의 신속성이 떨어지며, Medline 데이터와의 중복데이터 관리가 어려운 문제점이 발견되었고, 어휘통제가 되지 않으며, 철자 에러율이 PubMed보다 높았다.

그리고 두 DB에서 공통적으로 발견된 점은 홈페이지나 안내 자료에 표시된 데이터의 양과 실제 데이터 양 간에는 1년분 이상의 커다란 격차가 발생하며, 각종 자료에 표시된 데이터의 수록년도와 실제 검색한 데이터의 수록년도에는 많은 불일치가 존재하여 DB의 수록범위에 대한 이용자의 혼란을 유발할 가능성이 있고, 어휘통제와 색인어의 일관성 및 띠어쓰기에서 DB 구축의 규칙성을 찾아보기 어렵다는 점이다.

그러므로 Embase라는 상용 DB의 도입 검토를 위해서는, Embase가 무료 PubMed만으로 불충분한 검색 분야에 제약 분야뿐만 아니라 법의학, 죽병학, 임상의학, 제약, 환경문제, Systematic Review 등의 분야가 있다는 선행연구의 결과를 감안해야 한다. 또한 Embase가 Medline 데이터를 통합하여 서비스하

면서 수록 레코드 수와 수록 저널의 포괄성과 제약분야 정보가 더 우수하므로, 이와 같은 중복성이 높은 경쟁 DB가 존재할 경우 Embase 와 같은 유료 DB의 구독여부는 유료 DB가 갖는 장점을 중심으로 가치판단과 의사결정이 이루어져야 한다. 따라서 의학 분야의 포괄적 정보검색이 필요하거나 Embase가 강점을 갖는 주제 분야 도서관에서는 무료 PubMed에만 의존할 것이 아니라 Embase와 같은 상용 DB의 도입도 적극적으로 검토할 필요가 있다고 본다.

결국 PubMed와 Embase는 의학 분야 핵심저널을 수록하므로 중요한 정보는 대부분 포함 하겠지만 14,000여종 의학 및 보건 관련 분야 출판물의 최대 33%인 4,600여종만 포함할 뿐이며, 임상실험과 같은 의학 분야의 정보는 잘못 사용하면 사람의 생사문제와 직결될 수 있어서 각별한 주의가 필요하므로 PubMed와 Embase DB에 대한 정밀한 품질평가를 전제로 도서관내 웹 자원으로 선정하여 이용자의 요구에 맞게 검색서비스가 이루어져야만 할 것이다. 그리고 상용 DB 도입의 타당성을 주장하기 위해 DB 품질평가에 근거한 도입과 이용자 교육 및 도서관 정보서비스가 이루어져야

효율적인 정보서비스가 가능할 것이다.

본 연구는 국내 선행연구의 DB 품질평가 기준을 중심으로 상용 DB 도입 검토를 위한 DB 평가에 표준적인 품질평가 기준을 적용했다는 점에 의의를 둔다. 따라서 본 연구에 적용된 품질평가 기준과 방법으로 다른 상용 DB의 품질평가 측정이 이루어 진다면 조금 더 쉽고 객관적인 DB의 품질평가가 가능할 것이다. 하지만 본 연구의 부족한 점을 보완하기 위해서는 상용 DB의 도입에 따른 비용/편익 분석, 무료 PubMed를 사용할 수 있는 상황에서 유료 Embase를 도입할 비용소요 대비 가치에 대한 객관적 평가, 이용자의 DB 서비스 품질에 대한 만족도 조사에 의한 종합적인 품질평가, 평가 기준 사이의 상대적 중요도나 점수화에 대한 방안의 보완, 데이터의 정확률이나 재현율과 같은 고도의 복잡한 품질 분석, 의학 분야나 BT 분야의 대표적 문헌정보 DB인 PubMed와 Embase DB는 물론 금년에 24개 기관이 구독하는 BP나 41개 기관이 구독하는 SciFinder나 21개 기관이 구독하는 Web of Science나 Elsevier 사에서 금년 말에 약 심적으로 출시할 예정인 Scopus DB 등과의 종합적인 비교 평가도 필요할 것이다.

참 고 문 헌

- 김기오. 2003. 『국내외 의학관련 웹사이트의 KoreaMed와 PubMed 인지도 비교 연구』, 석사학위논문. 숙명여자대학교 대학원.
- 김기홍. 2000. 『CD-ROM MEDLINE과

- PubMed의 검색 효용성에 대한 비교 연구』, 석사학위논문. 성균관대학교 대학원.
- 김나원. 2002. PubMed 이용법. 『한국의학도 서관』, 29(1-2): 40-46.

- 김상준. 2002. 문헌정보학 실무자와 실무중심의 연구활동. 『한국의학도서관』, 29(1-2): 12-25.
- 김석영. 2002. 과학기술분야 웹 정보원 평가 및 비교연구. 『한국도서관정보학회지』, 33(3): 133-152.
- 김석영. 2002. 과학기술분야 해외 온라인 데이터베이스 평가 연구. 『정보관리연구』, 33(1): 61-80.
- 김성희. 2003. 데이터베이스 평가모델 개발을 위한 해외사례 조사연구. 『국회도서관보』, 40(2): 15-33.
- 문성빈, 최인숙, 최상희, 이성숙, 장혜영. 2002. 과학기술전문정보서비스 품질평가 및 이용자 정보요구 분석. 『정보관리연구』, 33(1): 31-47.
- 박준식. 2002. 색인·초록 및 전문서비스의 최근 동향과 선택. 『국회도서관보』, 39(4): 59-108.
- 서은경. 1999. 정보검색시스템에서의 이용자 인터페이스 기능에 관한 분석적 고찰. 『정보관리학회지』, 16(4): 125-150.
- 안계성, 함정훈. 2000. 『데이터베이스 품질평가 항목』, 서울: 한국데이터베이스진흥센터.
- 연구개발정보센터. 1996. 『DB 선정 및 품질평가 기준에 관한 연구』, 과천: 과학기술처.
- 유동숙. 1998. 『의약정보관련 서지사항 데이터베이스의 비교 평가』, 석사학위논문. 숙명여자대학교 임상약학대학원.
- 유사라. 1999. 데이터베이스 정보 품질평가의 메타분석. 『정보관리학회지』, 16(1): 157-174.
- 유사라. 2000. Web-DB 정보서비스 평가를 위한 기준 측정지표 분석 I. 『한국문현정보학회지』, 34(3): 133-156.
- 유혜영. 1997. 『국내 제작 데이터베이스의 평가에 관한 연구』, 석사학위논문. 서울여자대학교 대학원.
- 윤대진, 이두영. 2002. 통합정보검색시스템 인터페이스 평가지표에 관한 연구. 『제9회 한국정보관리학회 학술대회 논문집』, 43-50.
- 이국희, 최창선, 임남홍, 김호중, 박옥순, 최민영. 1995. 『데이터베이스 품질평가에 관한 연구』, 서울: 한국데이터베이스진흥센터.
- 이만수, 이상렬, 신인수. 2003. 서지 데이터베이스의 품질평가에 관한 연구. 『정보관리연구』, 34(3): 21-39.
- 이명희. 1998. ASKERIC 데이터베이스의 품질에 관한 연구. 『한국비블리아』, 9: 221-240.
- 이명희, 이상렬. 2002. 국가대표전자도서관 웹사이트의 평가에 관한 연구: 한국, 일본, 미국의회도서관을 중심으로. 『국회도서관보』, 39(3): 3-21.
- 이은숙. 1997. 『Embase, Medline, Sci-Search에서 검색된 한국의학논문의 중복정도와 특성분석』, 석사학위논문. 숙명여자대학교 교육대학원.
- 이용봉. 1996. 데이터베이스 품질에 관한 비평적 평가. 『국회도서관보』, 33(4): 13-28.
- 이용봉. 1999. CD-ROM 데이터베이스의 품질평가 및 개선방안에 관한 연구. 『한국문현정보학회지』, 33(4): 29-46.

- 이용봉. 2002. 이용자 서비스의 품질 향상을 위한 웹사이트 사용성 평가에 관한 연구. 『한국 문헌정보학회지』, 36(4): 311-329.
- 이용봉, 류범종, 김우성, 김석동. 2002. 전자도서관 웹사이트의 품질향상을 위한 사용성 평가에 관한 연구. 『한국문헌정보학회지』, 36(1): 41-60.
- 이용봉, 조현양, 류범종, 최재황. 2001. 과학기술분야 데이터베이스의 품질향상을 위한 품질평가 연구. 『한국문헌정보학회지』, 35(2): 109-132.
- 이제환. 1996. 이용자의 관점에서 본 KRI-STAL System과 DB의 품질. 『KO-STI Proceedings』, 1: 240-267.
- 이제환. 1997. 과학기술분야 서지 DB의 품질 관리 및 평가방안: KORDIC의 KRI-STAL DB를 중심으로. 『한국문헌정보학회지』, 31(3): 109-134.
- 이제환. 2002. 공동목록 DB의 품질평가와 품질관리: KERIS의 종합목록 DB를 중심으로. 『한국문헌정보학회지』, 36(1): 61-89.
- 이제환. 2004. 디지털도서관의 품질평가에 관한 연구. 『한국문헌정보학회지』, 38(1): 143-172.
- 이춘실, 이은숙. 1998. Embase, Medline, SciSearch에서 검색된 한국의학논문의 중복정도 연구. 『정보관리학회지』, 15(2): 29-46.
- 이춘열. 2003. DB 품질평가 모형과 DB 품질 향상: 한국데이터베이스진흥센터 품질 평가 모형의 적용. 『2003 DB그랜드 컨퍼런스』, <http://www.dlc.or.kr/downl oad/6th/3-2.pdf>』
- 장혜란. 2000. 데이터베이스 품질평가를 위한 모형 개발: 텍스트 데이터베이스 내용을 중심으로. 『정보관리학회지』, 17(4): 83-97.
- 조순영, 김성희, 임석종, 한혜영, 박연희. 2002. 『해외학술DB 선정 및 평가모델 개발에 관한 연구』, 서울: 한국교육학술정보원.
- 최홍식. 2001. 정보검색시스템과 연계된 상호대차시스템 설계에 관한 연구: 의학 도서관을 중심으로. 『정보관리학회지』, 18(2): 165-185.
- 한국데이터베이스진흥센터. 1996. 『데이터베이스 표준화연구보고서』, 서울: 한국데이터베이스진흥센터.
- 한국데이터베이스진흥센터. 2003. 『데이터베이스 이용실태 및 정보수요조사 보고서』, 서울: 한국데이터베이스진흥센터.
- 함정훈. 2001. 데이터베이스 품질평가 기준 연구. 『Digital Contents』, 2001년 01월 호: 60-73.
- Bachmann, L. M., P. Estermann, C. Kronenberg, and G. ter Riet. 2003, "Identifying diagnostic accuracy studies in EMBASE." *J. Med. Libr. Assoc.*, 91(3): 341-346.
- Brown, C. M. 1998, "The benefits of searching EMBASE versus MEDLINE for pharmaceutical information." *Online & CD-ROM Review*, 22(1): 3-8.
- Collins, A. M. K. and S. Gass. 1987, "Searching EMBASE: do the re-

- sults justify the cost?" *Health Libraries Review*, 4(4): 225-231.
- Cooper, M. D. 2001, "Usage patterns of a web-based library catalog." *JASIST*, 52(2): 137-148.
- Elsevier. 2004, EMBASE-MEDLINE fact sheets(online). [Netherlands: Elsevier.]
- Gunn, A. 1993, "EMBASE or MEDLINE?: The impact of the BIDS EMBASE service on academic medical libraries and their users." *Health Libraries Review*, 10(3): 128-130.
- Hallam, E. and C. Plaice. 1999, "An evaluation of EMBASE within the NHS: findings of the Database Access Project working partnership to extend the knowledge base of healthcare." *Health Libraries Review*, 16(3): 192-203.
- Hay, P. J., C. E. Adams, and C. Lefebvre. 1996, "The efficiency of searches for randomized controlled trials in the International Journal of Eating Disorders: a comparison of hand-searching, EMBASE and PsycLIT." *Health Libraries Review*, 13(2): 91-96.
- Kleijnen, J. and P. Knipschild. 1992, "The comprehensiveness of Medline and Embase computer searches: Searches for controlled clinical trials of homeopathy, ascorbic acid for common cold and ginkgo biloba for cerebral insufficiency and intermittent claudication." *Pharmaceutisch Weekblad*, 14: 316 - 320.
- Mcdonald, S., L. Taylor, and C. Adams. 1999, "Searching the right database: a comparison of four databases for psychiatry journals." *Health Libraries Review*, 16(3): 151-156.
- Menz, H. B. 2001, "The case for multiple database searching in podiatric medicine." *J. Am. Podiatr. Med. Assoc.*, 91(2): 103-104.
- Minozzi, S., V. Pistotti, and M. Forni. 2000, "Searching for rehabilitation articles on MEDLINE and EMBASE: An example with cross-over design." *Arch Phys. Med. Rehabil.*, 81(6): 720-722.
- Odaka, T., A. Nakayama, K. Akazawa, M. Sakamoto, N. Kinukawa, T. Kamakura, Y. Nishioka, H. Itasaka, Y. Watanabe, and Y. Nose. 1992, "The effect of a multiple literature database search: a numerical evaluation in the domain of Japanese life science." *J. Med. Syst.*, 16(4): 177-181.
- Parfrey, W. 1988, "Searching Embase: do the results justify the cost?" *Health Libraries Review*, 5(3):

- 189–190.
- Sampson, M., N. J. Barrowman, D. Moher, T. P. Klassen, B. Pham, R. Platt, P. D. John, R. Viola, and P. Raina. 2003, “Should meta-analysts search Embase in addition to Medline?” *J. Clin. Epidemiol.*, 56 (10): 943–955.
- Suarez-Almazor, M., E. Belseck, J. Homik, M. Dorgan, and C. Ramos-Remus. 2000, “Identifying clinical trials in the medical literature with electronic databases: Medline alone is not enough.” *Controlled Clinical Trials*, 21(5): 476–487.
- Watson, R. J. and P. H. Richardson. 1999, “Identifying randomized controlled trials of cognitive therapy for depression: comparing the efficiency of Embase, Medline and PsycINFO bibliographic databases.” *Brit. J. Med. Psychol.*, 72(4): 535–542.
- Woods, D. and K. Trewhellar. 1998, “Medline and Embase complement each other in literature searches.” *Brit. Med. J.*, 316: 1166.
- Yonker, V. A., K. P. Young, S. K. Becham, S. Horwitz, and K. Cousin. 1990, “Coverage and overlaps in bibliographic databases relevant to forensic medicine: a comparative analysis of MEDLINE.” *Bull. Med. Libr. Assoc.*, 78(1): 49–56.

KCI